Urine Metabolic Signature of Pancreatic Ductal Adenocarcinoma by H-1 Nuclear Magnetic Resonance: Identification, Mapping, and Evolution

ARTICLE *in* JOURNAL OF PROTEOME RESEARCH · NOVEMBER 2011

Impact Factor: 4.25 · DOI: 10.1021/pr200960u · Source: PubMed

Journal of proteome research

ARTICLE

pubs.acs.org/jpr

Urine Metabolic Signature of Pancreatic Ductal Adenocarcinoma by ^{1}H Nuclear Magnetic Resonance: Identification, Mapping, and Evolution

Claudia Napoli,[†,‡] Nicola Sperandio,[§] Rita T. Lawlor,[§] Aldo Scarpa,[§,∥] Henriette Molinari,[†] and Michael Assfalg*[,†]

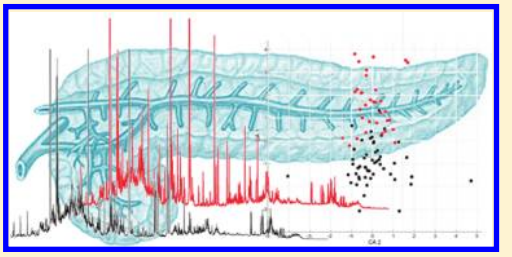
[†]Department of Biotechnology and [∥]Department of Pathology and Diagnostics, University of Verona, Verona, Italy

[‡]Bruker Italia, Milan, Italy

[§]ARC-NET Center for Applied Research on Cancer, Verona University Hospital, Verona, Italy

Supporting Information

ABSTRACT: Pancreatic ductal adenocarcinoma (PDAC) has a dismal prognosis and is highly chemoresistant. Early detection is the only means to impact long-term survival, but screening methods are lacking. Given the complex and heterogeneous nature of pancreatic cancer, unbiased analytical methods such as metabolomics by nuclear magnetic resonance (NMR) spectroscopy show promise to identify disease-specific molecular fingerprints. NMR profiles constitute a fingerprint of the biofluid, reporting quantitatively on all detectable small biomolecules. NMR spectroscopy was applied to investigate the urine metabolome of PDAC patients ($n = 33$) and to detect altered metabolic profiles in comparison with healthy matched controls ($n = 54$). The spectral data were analyzed using multivariate statistical techniques. Statistically significant differences were found between urine metabolomic profiles of PDAC and control individuals ($p < 10^{-5}$). Group discrimination was possible due to average concentration differences of several metabolite signals, pointing to a multimolecular signature of the disease. The robustness of the determined statistical model is confirmed by its predictive performance (sensitivity = 75.8%, specificity = 90.7%). Additionally, the method allowed for a neat separation between spectral profiles of individuals with intermediate and advanced pathologic staging, as well as for the discrimination of samples based on tumor localization. NMR spectroscopy analysis of urinary metabolic profiles proved successful in identifying a complex molecular signature of PDAC. Furthermore, results of a descriptive-level analysis show the possibility to follow disease evolution and to carry out tumor site mapping. Given the high reproducibility and the noninvasive nature of the analytical procedure, the described method bears potential to impact large-scale screening programs.



KEYWORDS: pancreatic ductal adenocarcinoma, metabolomics, NMR spectroscopy, urine

INTRODUCTION

Pancreatic ductal adenocarcinoma (PDAC) is the fifth cause of cancer death in Europe and is highly aggressive and resistant to conventional and targeted therapeutic agents, resulting in a dismal 5-year survival rate of 5%.[1] Of all patients <5% can be treated by surgery, and most will live only 1 year after resection. At diagnosis, 80% have metastatic or locally advanced disease.[2] The current biochemical markers are of little help in addressing diagnosis of this disease. No tumor marker specific for pancreatic cancer exists.

Research performed during the past few decades has clearly revealed that cancer cells, in conjunction with significant aberrations in genome and proteome, possess unique metabolic phenotypes, as the result of multiple altered signaling pathways.[3] It is thus clear that perturbed molecular mechanisms determine differences in the central metabolic pathways operating in malignant tissues, as an adaptation response required to support cell growth and survival.[4]

Metabolomics, emerging as one of the core disciplines in systems biology,[5,6] addresses the global study of metabolites, the ultimate downstream indicators of perturbations to the normal status of biological systems[7] and constitutes an established tool to investigate functional changes associated with diseases such as cancer.[8] While tissue metabolomics is of great value in understanding tumor biology, the study of biofluids provides an opportunity for broad-spectrum screening of early tumor-associated perturbations in metabolism.[9] The two most widely used analytical platforms in the field are mass spectrometry (MS) and nuclear magnetic resonance spectroscopy (NMR), the latter being particularly suited in clinical screening programs.

Pancreatic cancer is known to have associated metabolic changes and thus appears to be well-suited for this kind of investigation. For example, a solid-state NMR investigation of cancerous rat pancreases has described unique metabolomic profiles associated with the disease, together with a differential lipid content in pancreatic cancer versus chronic pancreatitis.[10]

Received: September 23, 2011
Published: November 08, 2011

Although there has been little work performed on the metabolomic profiling of pancreatic cancer, the interest in metabolomics as an appropriate tool for the investigation of PDAC is confirmed by a growing number of papers appearing recently in the literature. In a pilot study, covering a number of about 50 patients, the NMR serum metabolomic profile of patients with pancreatic cancer has been investigated, and it was shown to differ significantly from that of patients with benign disease.[11] Furthermore, changes in plasma lipid levels have been recorded in pancreatic cancer patients (n = 100) compared to healthy controls (n = 90),[12] and a number of metabolic classifiers of PDAC have been identified from MS analysis of blood plasma in a small cohort study.[13] Interestingly, even the salivary metabolomic profile was shown to be significantly altered in patients with pancreatic cancer (n = 18).[14]

Although urine is the most frequently used biofluid in preclinical toxicological applications of metabolomics, limited experience exists with urinary metabolic markers for cancer diagnosis.[15] However, promising results have been obtained, for example, in the fields of colorectal,[16] ovarian, and breast cancers.[17] One of the biggest challenges for global urine metabolic profiling in the clinic is its high variability depending on phenotype, life style, diet, and several other environmental factors.[6,18−20] From a methodological point of view, however, urine represents an ideal biofluid for NMR metabolomics analysis because it contains the highest number of water-soluble metabolites, it can be sampled frequently and noninvasively, and it requires minimal sample preparation. From a clinical point of view urine analysis represents a screening method easily implementable on a wide scale.

We have conducted and present here a case-control study employing a ^{1}H NMR approach to examine the urine metabolic profiles of 33 patients with PDAC and 54 healthy matched controls. Due to the inherent sensitivity of urine composition to a plethora of perturbing factors, in order to limit possible gender-related variability we purposely focused the study on male individuals in a limited age range. Multivariate statistical analysis of the spectral data was applied to obtain group discrimination and selection of the disease-specific metabolic signature.

METHODS

Study Population

Thirty-three patients and fifty-four healthy controls, referred to the University Hospital of Verona, were enrolled in the study using an informed consent approved by the local Ethics Committee. The consent included collection of a urine sample, background and clinical data from patients such as age, weight, weight loss, diagnosis, hospitalization, history of diabetes and pancreas inflammation, medical treatment, and some additional lifestyle information provided in a short questionnaire. Patients (controls) were males in the age range of 62 ± 6 (61 ± 6) years at the time of the study. Patients included in the study had a confirmed diagnosis of PDAC .

Sample Collection and Preparation

The samples were all collected prior to surgery and after the patient had been fasting overnight. The urine was collected in a sterile pot and was the first urine of the morning, with the first drops discarded. A stock solution of Protease Inhibitor Cocktail tablet, EDTA-free (Roche), was added to every 50 mL of collected urine, to a final concentration of 1 μg/mL (producing no detectable NMR signal in the used experimental conditions). The pots were stored at 4 °C until transportation to the laboratory for processing. Once in the laboratory the samples were aliquoted and then frozen in −80 °C freezers within 5 h of collection.

Frozen samples were thawed at room temperature and shaken before use. Aliquots of each urine sample (630 μL) were added to 70 μL of potassium phosphate buffer (1.5 M K_2HPO_4 in 100% ^2H_2O, pH 7.4), to minimize variations in metabolite NMR chemical shifts arising from differences in urinary pH, also containing 0.1% sodium 3-(trimethylsilyl)-[2,2,3,3-$^2H_4$]propionate (TSP) and 2 mM sodium azide. Samples were centrifuged at 14,000 × g for 5 min at 4 °C to remove any solid debris, and then 600 μL of the supernatant was placed in a 5-mm o.d. NMR tube. All chemicals were from Sigma.

NMR Spectroscopy

All measurements were performed on a Bruker Avance III NMR spectrometer (Bruker BioSpin, Rheinstetten, Germany) operating at 600.13 MHz for ^{1}H observation, equipped with TCI cryoprobe incorporating a z axis gradient coil and automatic tuning-matching (ATM). Experiments were run in automation mode after loading individual samples on a Bruker Automatic Sample Changer, interfaced with the software IconNMR (Bruker). A time delay of 5 min was set between sample injection and preacquisition calibrations to ensure complete temperature equilibration (300 K). Measurements were repeated once in random order after completion of the first entire set. For each sample, a one-dimensional NOESY experiment including solvent signal saturation during relaxation and mixing time and a spoil gradient (Bruker noesygppr1d) was acquired using 32 free induction decays (FIDs), 64k data points, a spectral width of 12,019 Hz, an acquisition time of 2.7 s, a relaxation delay of 4 s, and a mixing time of 10 ms. A very high resolution was achieved, with the TSP signal line width resulting always <1 Hz. The FIDs were multiplied by an exponential weighting function corresponding to a line broadening of 0.3 Hz before Fourier transformation, phasing, and baseline correction. Two-dimensional J-resolved ^{1}H and ^{1}H,^{13}C-HSQC NMR experiments were run on representative samples to aid in signal identification. All spectra were referenced to the TSP signal (δ = 0.00 ppm). NMR data were processed using TopSpin 2.1 (Bruker). Metabolite assignments were performed using Amix 3.9 (Bruker) in combination with the Bruker NMR Metabolic Profiling Database as well as on the basis of reference compound spectra available in databases such as the Human Metabolome Database[21] and the Biological Magnetic Resonance Data Bank.[22]

Data Processing and Multivariate Statistical Analysis

^{1}H NMR spectra were segmented in rectangular buckets of fixed 0.04 ppm width and integrated using Amix 3.9 (Bruker). The spectral region between 4.5 and 6.0 ppm was discarded because of the variability (though limited) in the suppression of the water signal and variations in the urea signal caused by partial cross-solvent saturation through solvent-exchanging protons. The remaining buckets in the range 0.5−9.5 ppm were then normalized to the total area to minimize differences in urine concentration between samples and subsequently mean-centered.

The preprocessed data were subjected to principal component analysis (PCA) using an available function in Matlab 7.3 (Mathworks Inc.). PCA involves projecting the original X-matrix (N, objects × p, variables) onto a d-dimensional space ($d \leq p$) using a projection matrix ($\mathbf{L}_{p\times d}$, loadings) and creating object

Table 1. Patient Information

no.	age	staging	grading	site	diabetes	pancreatitis	other conditions
13	67	III	G2	head			
18	62	III	G2	head			
30	62	III	G3	body			
31	62	IIB	G2	uncinate			
44	63	III	G2	head		yes	
45	69	III	G2	body			
58	66	III	G2	body			
62	54	III	G2	body			
64	58	IIB	G2	head	yes		
72	70	IIB	G2	head			
79	60	III	G2	head			
80	70	III	G2	head	yes		prostate cancer
91	52	III	G2	uncinate			
95	70	IIB	G2	head			
101	60	III	G2	head	yes		
104	62	IV	G2	head			
121	51	III	G3	head			
133	59	III	G2	head			
135	53	IIB	G2	head	yes		
138	62	IIB	G2	head			hypercholesterolemia
155	71	III	G2	body	yes		
162	72	III	G2	head			
165	67	IV	G2	body			gallstones
171	67	IV	G2	head		yes	hypertension
178	63	IV	G2	uncinate			
192	65	IV	G2	uncinate			
196	63	IV	G2	uncinate			hypertension
205	56	IV	G2	head		yes	
222	53	IV	G2	head			gallstones
226	63	IV	G2	head			
236	56	IV	G2	body	yes		
248	52	IV	G2	head			
251	60	IV	G2	tail			

coordinates ($\mathbf{T}_{N\times d}$, scores) in a new reference system. The influence of the original variables on the principal components (PC) is determined on the basis of the maximum variance criterion. The first PC lies in the direction describing maximum variance in the original data. Each subsequent principal component lies in an orthogonal direction of maximum variance that has not been considered by the former components.

The second stage of the data analysis involved multivariate analysis of variance (MANOVA),[23] as implemented in Matlab. This procedure provides the dimensionality of the group means and corresponding probabilities. Furthermore, it provides results from a canonical analysis (CA), which searches for linear combinations (canonical variables) of the original variables that reflect the largest possible separation between groups. Again a matrix of scores ($\mathbf{C}_{N\times d}$) and loadings ($\mathbf{E}_{d\times d}$) are generated. The part of the objects responsible for group discrimination can be reconstructed by back-projection from the relevant CA subspace representation into the original space bearing direct physical meaning.

In summary, to identify differences between diseased and control individuals, the bucketed spectra were assembled in a matrix $\mathbf{X}_{87\times 189}$ and projected into a 28-dimensional PCA subspace representing 99% of the variance. The $\mathbf{T}_{87\times 28}$ matrix was then projected into a 1-dimensional CA subspace. Permutation testing was performed using in-house written scripts by assigning random class memberships, repeating the full PCA/CA procedure, and evaluating the probabilities corresponding to the estimated dimensionality of the data in the discriminating subspace. Leave-one-out cross-validation in conjunction with a five-nearest neighbor classification method was used to evaluate the self-prediction ability of the statistical model. As statistical measures of performance, the specificity (true negative rate), sensitivity (true positive rate), and accuracy (correct prediction rate) of the model were computed. The interpretation of the contribution of the original variables in discriminating groups was based on the calculation of combined loadings from both the PCA and CA transformations. The combined loadings coefficients were standardized by evaluating the pooled within-group covariance matrix.[23]

An additional PCA/CA analysis was performed only on PDAC samples to explore the possibility of metabolic profile classification based on staging and tumor site. In the case of staging a matrix $\mathbf{T}_{33\times 28}$ was projected into a 1-dimensional CA

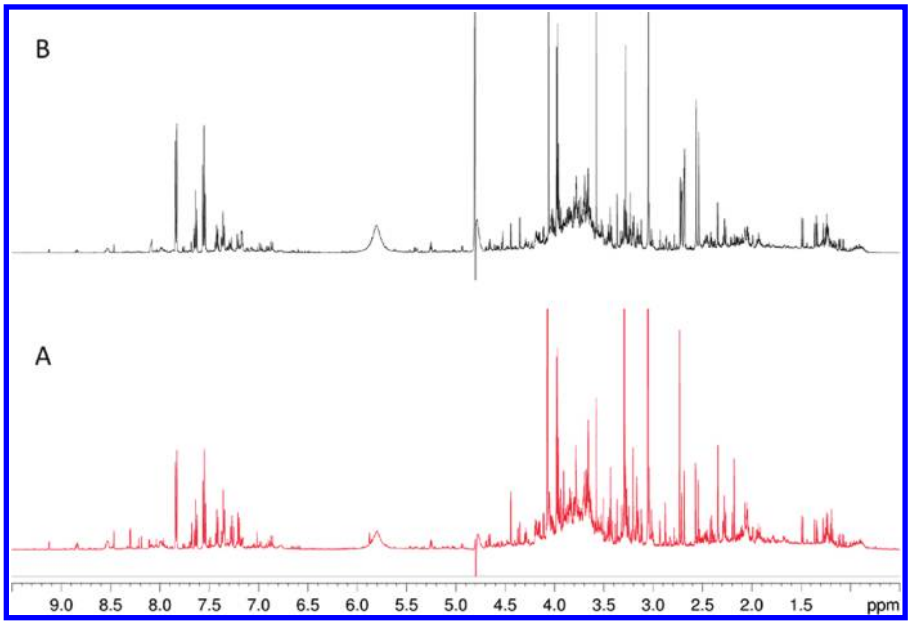


Figure 1. Representative 600 MHz one-dimensional NOESY ^{1}H NMR spectra of urine samples measured at 27 °C, collected from (A) a PDAC patient and (B) a matched control.

subspace, whereas for tumor site classification a matrix $\mathbf{T}_{32\times29}$ was projected into a 1-dimensional CA subspace.

RESULTS

All patients were males in the age range of 62 ± 6 years at the time of the study (Table 1). All patients had a confirmed diagnosis of PDAC, 14 situated in the head of the pancreas, 7 in the body, 5 in the uncinate process, and 1 in the tail. Of the 33 patients, 6 had a staging of IIB (lymph node involvement and a pT of 1−3), 15 had a staging of III (infiltration to the celiac axis or superior mesenteric artery), and 12 had a staging of IV (liver metastasis). Of the 33 patients, 31 had a grading of G2, and two had a grading of G3. Only four had been diagnosed with diabetes and three with pancreatitis. Six had other pathologic conditions (hypertension, gallstones). Six were on medication, four for diabetes and two for other conditions. Eight had suffered rapid weight loss of more than 10 kg, 6 of between 6 and 10 kg, and 7 of less than 5 kg. Eight of these experienced the weight loss in less than 3 months. Control individuals were males in the age range of 61 ± 6 years at the time of the study with no declared pathological condition.

Figure 1 shows representative ^{1}H NMR spectra recorded on urine samples of a pancreatic cancer patient and a healthy individual. Each spectrum represents the overall envelope of signals originated by small biomolecules constituting the urine metabolome and present in the test tube at concentrations higher than about 1−10 μM. The position of the signals reflects the chemical structures, the peaks of aromatics ($\delta \approx 6.5-7.5$ ppm) and amides ($\delta \approx 6.5-9.0$ ppm) falling in the left part of the spectrum, those of the aliphatics in the extreme right ($\delta \approx 0.0-2.0$ ppm), and carbohydrates in the middle-right region ($\delta \approx 3.0-4.0$ ppm). Due to the inherent quantitative nature of the analytical method, it is possible to compare intensities of peaks appearing in the corresponding positions in different spectra to derive information on the underlying metabolite concentrations. By visual inspection of the whole data set no striking differences in signal intensities between the metabolic profiles of the diseased and healthy individuals can be found.

The collected experimental data are multivariate in nature and can be effectively analyzed by a wide range of multivariate statistical techniques. We applied PCA, i.e., an unsupervised pattern recognition approach, to explore the spectral data set in terms of the general variability of signal intensities among different samples. The first 28 principal components (linear combinations of the original spectral buckets) described 99% of the total variance. The spectral data are visualized in the subspace of the first two principal components, as well as of the third and fourth PCs (Figure 2). The metabolites responsible for the spread of objects in the PC space, identified from the PCA loadings' plots (Figure S1, Supporting Information), include hippurate, creatinine, trimethylamine-*N*-oxide, and a number of signals in the crowded 3.2−3.9 ppm region. Besides this natural variability in urine metabolite composition, no indication of group clustering of the samples according to the health state emerges in any PC dimension. However, it is well-known that this approach provides excellent results in highlighting the overall variability of a data set, but it is not a classification method.

To determine whether the NMR spectra could be used to differentiate PDAC from control individuals, we introduced class information in the construction of a statistical model based on the maximization of the differences between groups. A canonical analysis was performed on the PCA transformed data. This method consists of a new projection of scores data in a subspace with a lower dimensionality that takes into account only the correlation between variables and group separation. The samples from the two known groups appeared indeed significantly different ($p < 10^{-5}$), a result that is also evident from Figure 3A, which displays the objects in the space of the first two canonical variables, CA 1 and CA 2 (the latter being included only for visualization purposes). The samples appear represented in two distinct and partially overlapping clusters according to health status. By increasing the number of principal components used in the subsequent canonical analysis a complete separation

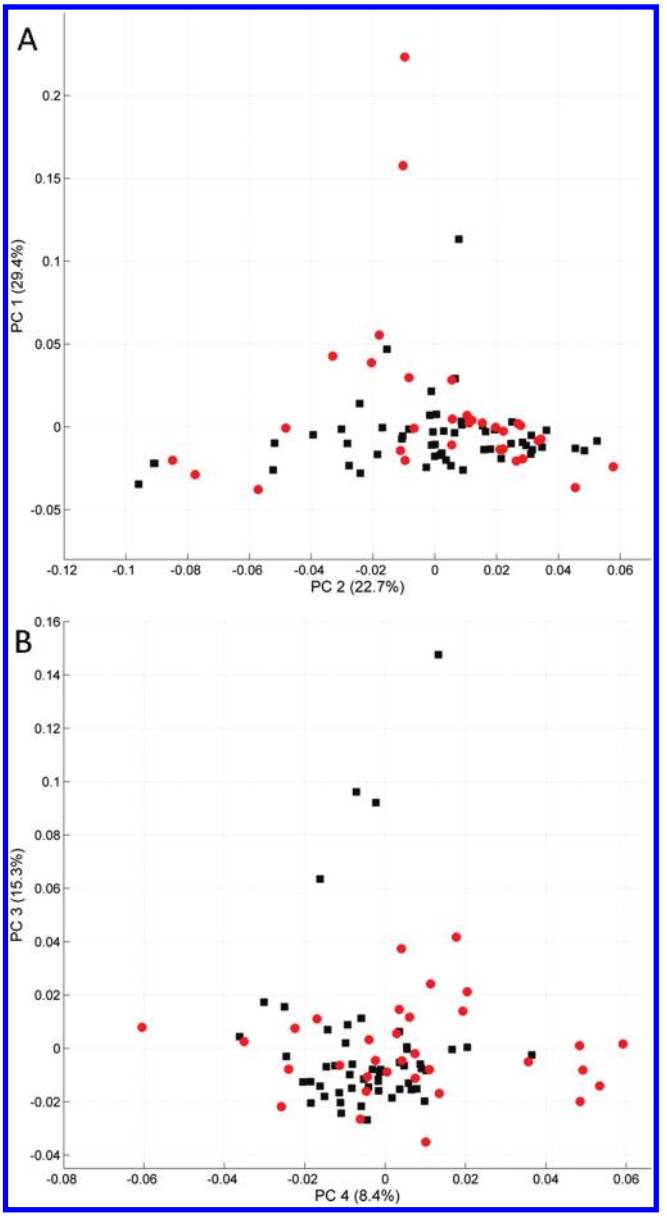


Figure 2. Unsupervised principal component analysis scores plot of ^{1}H NMR spectra of urine from PDAC patients (red filled circles) and healthy control individuals (black squares). (A) Projection of the spectral profiles into the subspace of the first two principal components. (B) Projection of the spectral profiles into the subspace of the third and fourth principal components. The percentage of explained variance is reported on each principal component axis.

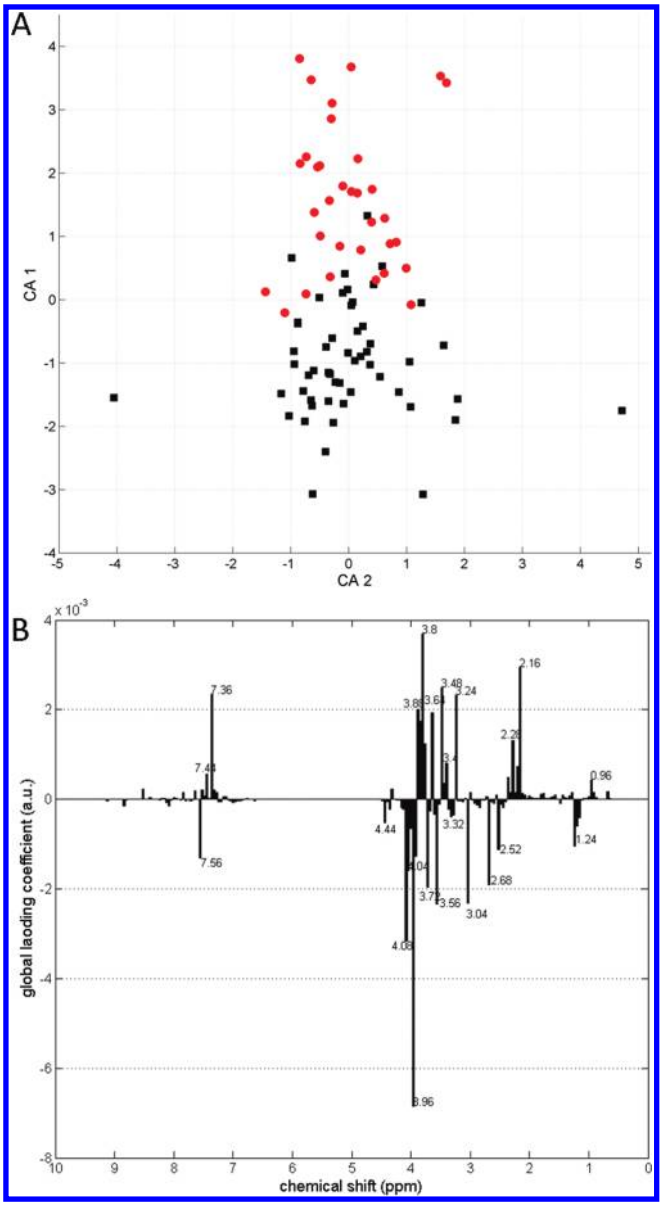


Figure 3. Supervised pattern recognition. Projection of the ^{1}H NMR spectral profiles into the first two dimensions of the PCA/CA subspace. Red filled circles: samples of PDAC patients; black squares: samples of control individuals. (A) Canonical analysis applied to the data reduced in 28 principal components representing 99.0% of the total variance. By application of MANOVA to the transformed data a dimensionality of the group means of 1 was found (two distinct classes). CA 1 represents the dimension with maximum group discrimination, and CA 2 is included only for visualization purposes. (B) Global coefficients representing the contribution of the different spectral regions to group discrimination. The values are obtained from combinations of eigenvectors determined from both the PCA and CA analyses. Positive (negative) values indicate increased (decreased) intensities in samples from PDAC patients compared to the levels found for healthy controls. The chemical shifts of a number of data points are reported explicitly.

of the groups was achieved (including 55 PCs explaining 99.9% of the total variance in the data) (Figure S2, Supporting Information). Nevertheless we preferred to retain a smaller number of PCs because no significant improvement at the predictive level could be attained including more variables, as well as to avoid adding complexity to the interpretation of metabolite level contributions. Permutation testing ensured that the introduced specific classification was significantly better than any other random classification in two arbitrary groups. Only in 3.5% of 999 runs of permutation tests, two distinct classes were found with statistical significance, but the computed Wilk's lambda value resulted always worse than in the correct classification case. Finally, the bucketed spectra represented in the CA discriminating subspace were back-projected into the original space and comparison between original and reconstructed data sets confirmed that no artificial features were introduced by data reduction and transformation procedures.

Table 2. Comparison of Metabolite Levels in PDAC Urine Samples with Respect to Healthy Controls

metabolite	level change	spectral bucket[a] (ppm)	assignments[b] (ppm)
acetoacetate	↑	2.28	2.27s
acetylated compounds	↑	2.16	2.16s
adenine	↔		8.19s, 8.17s
alanine	↔		1.47d
bile salts	↔		0.68s
citrate	↓	2.52, 2.68	2.68d, 2.54d
creatinine	↓	3.04, 4.08	4.07s, 3.04s
formate	↔		8.45s
glucose	↑	3.20−3.90	5.23d, 4.64d, 3.90−3.23m
glycine	↓	3.56	3.56s
hippurate	↓	3.96, 7.56	8.52s, 7.81m, 7.62m, 7.54m, 3.96d
2-hydroxyisobutyrate	↔		1.35s
3-hydroxyisovalerate	↓	1.24	2.36s, 1.26s
4-hydroxyphenylacetate	↔		7.15m, 6.84m, 3.44s
isobutyrate	↔		1.06d
lactate	↔		1.32d
leucine	↑	0.96	0.96d + d
dimethylamine	↔		2.71s
trimethylamine-*N*-oxide	↔		3.27s
3-methylhistidine	↔		7.66s, 7.00s, 3.68s, 3.06m
1-methylnicotinamide	↔		9.26s, 8.95d, 8.88d, 8.17t, 4.46s
2-phenylacetamide	↑	7.36, 7.44	7.44m, 7.36m
trigonelline	↓	4.44	9.11s, 8.82m, 8.07m, 4.43s
unknown molecule	↓	3.72	3.72s
valine	↔		1.03d, 0.98d

[a] Only spectral buckets showing significant intensity changes between PDAC and controls are listed. [b] Signals identified in the spectra are reported and indicated by their central chemical shift (s = singlet, d = doublet, t = triplet, m = multiplet).

After having established the possibility to discriminate metabolic profiles of PDAC from controls, we challenged the described statistical model for its predictive ability. A cross-validation procedure was used to assess the performances, and the following statistical measures were computed: sensitivity = 75.8% (indicating the proportion of diseased individuals with a positive test), specificity = 90.7% (indicating the proportion of nondiseased individuals with a negative test), and accuracy = 85.1% (indicating the rate of correct tests).

The contribution of the original spectral intensities to the separation of groups was evaluated. To this purpose we determined the global standardized loading coefficients of the PCA/CA analysis. The global coefficients for the first canonical variable (the one explaining differences between groups) are displayed in Figure 3B. In the figure, high positive values indicate that the signals in the corresponding spectral regions are more intense for the diseased group, while negative values indicate signals that are less intense compared to the values found in healthy controls. Although a single spectral bucket may contain signals from different metabolites, some of the most intense and/or resolved signals can be safely attributed to specific molecules. With the aid of two-dimensional NMR experiments and comparison with reference spectral databases, 24 metabolites could be identified in all of the spectra. Signals attributed to these metabolites that fell in the determined discriminating buckets were used to label the corresponding metabolite levels as increased, decreased, or unaltered in the PDAC group (Table 2). The levels of acetoacetate, leucine, glucose, 2-phenylacetamide, and some acetylated compounds appear elevated, and those of citrate, creatinine, glycine, hippurate, 3-hydroxyisovalerate, trigonelline, and an unknown signal at 3.72 ppm are lower in the samples of diseased individuals compared to those of the controls.

In order to keep the analysis as general as possible, sample collection was not based on a selection of specific disease features such as staging, grading, or tumor site. However, we performed an exploratory analysis to understand the possibility of subclassification within the set of samples from PDAC individuals. Grading was not investigated because the large majority of cases corresponded to grade G2. A supervised analysis was instead performed on the stage of disease after merging stages IIB and III into a single group. Twenty-eight principal components, corresponding to 99.9% of total variance, were required to achieve group separation with statistical significance (p = 0.019), as shown in Figure 4A. As for the classification based on tumor site, we excluded the sample corresponding to a tail localization, being the only case, and carried out the analysis on the three remaining classes, namely, head (n = 20), body (n = 7), and uncinate (n = 5). Here 29 principal components, corresponding to 99.9% of total variance, were used in the canonical analysis. The dimensionality of the group means resulted to be 1 (two groups, p = 0.0052), with an evident separation of the uncinate tumors from the other two tumor classes (Figure 5A). Although graphically the body tumors cluster separately from head tumors, the multivariate statistics indicated a poor related probability (p = 0.4). Given the relatively small number of cases in each group we did not perform an additional predictivity test on the staging and the site

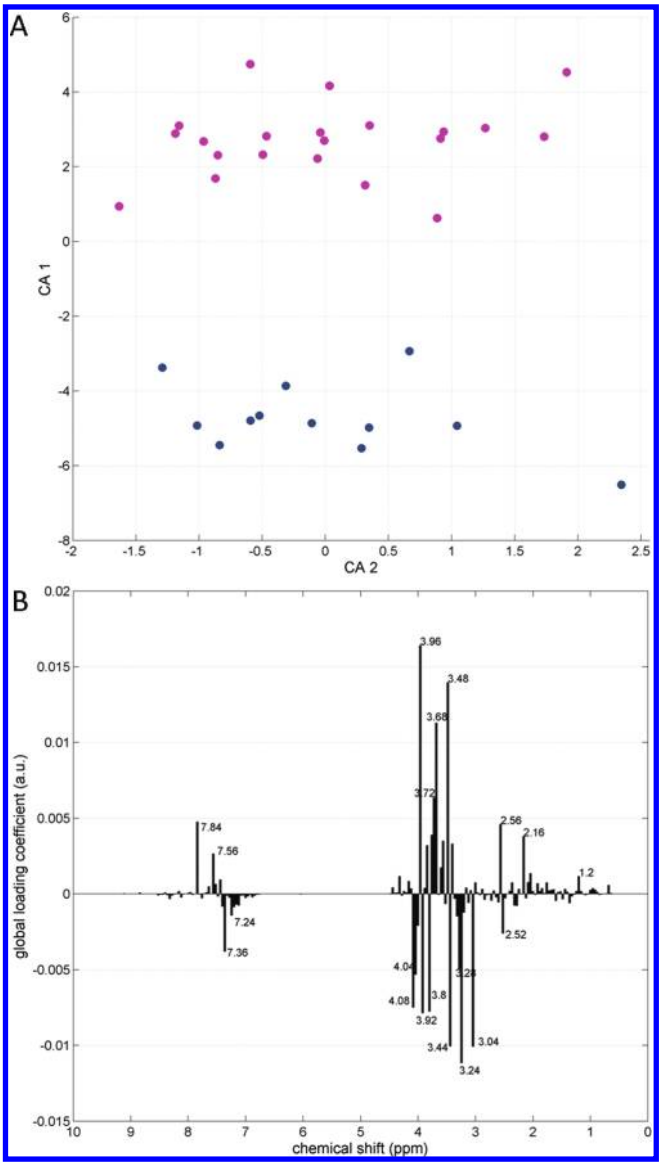


Figure 4. Sample classification based on pathologic staging. (A) Projection of the ^{1}H NMR spectral profiles into the first two dimensions of the PCA/CA subspace. Magenta: samples of patients with staging IIB and III; dark blue: samples of patients with staging IV. All 33 PDAC samples are included. The canonical analysis was applied to the data reduced in 28 principal components representing 99.9% of the total variance. The dimensionality of the group means was 1 (two distinct classes). CA 1 represents the dimension with maximum group discrimination. (B) Global coefficients representing the contribution of the different spectral regions to group discrimination. Positive (negative) values indicate increased (decreased) intensities in samples from patients with tumor stages of IIB and III, compared to patients with tumor stage IV. The chemical shifts of a number of data points are reported explicitly.

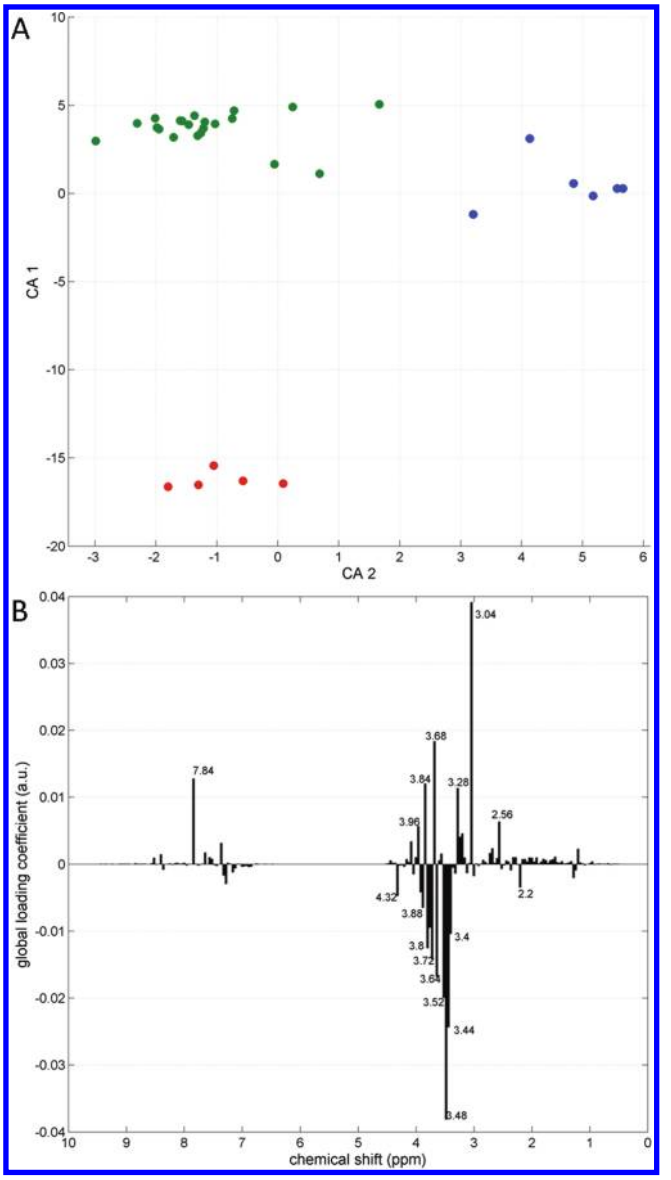


Figure 5. Sample classification based on tumor localization. Projection of the ^{1}H NMR spectral profiles into the first two dimensions of the PCA/CA subspace. Dark green: head; blue: body; dark red: uncinate. Thirty-two PDAC samples are included. The canonical analysis was applied to the data reduced in 29 principal components representing 99.9% of the total variance. The dimensionality of the group means was one (two distinct classes). CA 1 represents the dimension with maximum group discrimination. (B) Global coefficients representing the contribution of the different spectral regions to group discrimination. Positive (negative) values indicate increased (decreased) intensities in samples from patients with head and body tumor localization compared to the levels found for patients with uncinate tumor localization.

classification models. By inspection of the global loadings coefficients (Figures 4B, 5B) and with reference to resonance assignments reported in Table 2, it can be derived that a number of high-concentration metabolites (e.g., hippuric acid and other aromatics, citrate) contribute to class discrimination in both types of analyses. Most noteworthy are differences in the metabolic profiles of the various groups observed in the crowded carbohydrate region, possibly as a consequence of the involvement of exocrine pancreas (from which PDAC arises) in the processing of these metabolites.

DISCUSSION

We have analyzed the urine metabolic profiles of male individuals with pancreatic adenocarcinoma in comparison with age-matched controls. The metabolic profiles of PDAC patients differ from those of the controls with statistical significance,

demonstrating that a urine metabolic signature of the disease exists. A separation of the data-reduced profiles based on straightforward unsupervised pattern recognition is not possible, a result that is not unexpected, given the high intrinsic variability of the small molecule complement in this biofluid. However, class separation becomes apparent after application of a supervised analysis, and the robustness of the statistical model is confirmed by an encouraging predictive ability (accuracy ~85%). It is worth noting that this predictivity is comparable with that of the most commonly used marker for pancreatic cancer, CA19-9 or syalylated Lewis (a) blood group antigen (sensitivity: ~79%, specificity: ~82%).[24] These figures become important considering that the use of the latter marker as diagnostic tool has several limitations. For example, elevated CA19-9 levels are often observed in benign obstructive jaundice, chronic pancreatitis, liver cirrhosis, and cholangitis; increased CA19-9 levels are observed in multiple types of adenocarcinoma, especially in advanced gastrointestinal cancers; CA19-9 is not expressed in subjects with Lewis a- b- genotype; there is lack of sensitivity for early or small-diameter pancreatic cancer; and finally, poorly differentiated pancreatic cancers also appear to produce less CA19-9 than either moderately or well-differentiated cancers.[24−26]

The present untargeted metabolomics approach allows the simultaneous unbiased investigation of a broad range of molecular classes. The described results add to previous findings obtained from the analysis of samples of different biological sources. A very recent ^{1}H NMR study described the serum metabolic signature of patients with pancreatic cancer and showed the possibility to distinguish benign from malignant pancreatic lesions.[11] Elevated levels of glutamate, acetone, and 3-hydroxybutyrate were observed in sera of cancer patients, while increased glucose, phenylalanine, formate, and mannose were found to correlate with both disease and age. Differences between serum metabolic profiles of pancreatic cancer patients and healthy controls were also reported,[27] although the investigation was limited at the level of a principal component analysis. In a different study, mass spectrometry metabolic profiling was applied to plasma samples of pancreatic cancer patients.[13] A number of candidate biomarkers were proposed, including the amino acids *N*-methylalanine, lysine, glutamine, phenylalanine, the fatty acid arachidonic acid, the lipids lysoPC(18:2), PC(34:2), PE(26:0), and the bile acids tauroursodeoxycholic acid, taurocholic acid, deoxycholylglycine and cholylglycine. The plasma of pancreatic cancer patients also was shown to have an altered lipid profile based on a ^{1}H NMR investigation of lipophilic blood extracts.[12] Our data on urine confirm the association with the disease of metabolites produced in ketogenesis such as acetoacetate, as well as of a decreased level of citrate which has often been correlated with a down-regulated TCA cycle as consequence of the so-called "Warburg effect".[28] In addition, several spectral regions are found to display differential signal intensities between PDAC and controls, pointing to a multimolecular biomarker as the signature of disease. Given this result, it may be predicted that metabolomics by NMR will not suffer from the limitations of single-marker methods often attributable to cancer heterogeneity.[29,30]

In order to explore the possibility to identify fine structures of the PDAC metabolic fingerprint, we further carried out a preliminary analysis on the spectral correlations with pathologic staging and tumor localization. This information is of value to determine the most appropriate treatment and prognostic groups. Application of a PCA/CA analysis to the NMR spectral profiles allowed distinguishing samples of individuals in intermediate (lymph node involvement and a pT of 1−3, or infiltration to the celiac axis or superior mesenteric artery) from advanced staging (liver metastasis). Also the different tumor site resulted in a separation of metabolic profiles, with an evident clustering of uncinate tumors distant from both head and body tumors. Both of these analyses were kept at the descriptive level, without attempting to evaluate the predictive ability of the models due to limited group size availability. This successful investigation however points to the possible application of the NMR metabolomics method to differentiate different tumor localization and disease staging.

The notion that the presence, localization, and progression of a pancreatic tumor may be identified using statistical analysis of NMR spectra of urine is relevant in many respects; however, caution is needed when attempting to link these metabolic signatures to the underlying biochemistry. In general terms, multiple molecular mechanisms, both intrinsic and extrinsic, converge to alter core cellular metabolism in cancerous tissues. Tumor cells reprogramme their metabolic pathways to meet their needs during the process of tumor progression. The abnormal tumor microenvironment, such as hypoxia, pH, and low glucose concentrations, also has a major role in determining the metabolic phenotype of the cell.[4] In addition, the metabolic profile reports on a global metabolic state that incorporates but is not confined to tumor cell metabolism. More specifically, pancreatic cancer shares many hallmarks with cancers of different origin but is unique for its ability to form large tumors, the rapid progression, and its invasiveness.[31] An aberrant activation of at least seven developmental pathways during PDAC progression has been reported.[32] Recent results also suggest a link between PDAC and neurodegenerative diseases, while PDAC-associated secretory proteome revealed elevated levels of metabolic enzymes operational during inflammation.[33] All of these molecular pathways elicit broad effects that combine to produce the specific pathology of PDAC. The complexity of involved metabolic alterations and the interference of associated diseases suggest that, starting from our positive results, a more penetrating analysis of pancreatic cancer biology requires their integration with in vivo investigations of appropriate model systems.

In conclusion, the present study and the above-mentioned investigations show the possibility to isolate distinct features of pancreatic cancer in the metabolome of human systemic biofluids, reporting on a global metabolic response of the host to the pathological insult. The available data are still not sufficient to allow establishing a different suitability of urine, blood, or saliva in the diagnosis of PDAC, and more extensive analyses of all of these biofluids are needed. In the case of urine, a study analogous to the one here described could be performed on samples obtained from female individuals and add another dimension to the understanding of gender-related metabolic markers, while discrimination of metabolite profiles from blood or saliva on the basis of pathologic staging and tumor localization could also be considered. In the latter respect, metabolic profiling of pancreatic juice is expected to provide more direct correlations with tumor cell metabolism, however, with the disadvantage of surgical intervention. NMR spectroscopy confirms a powerful tool for the noninvasive and reproducible identification of metabolic signatures with a large potential to become a leading technique in programs for early detection and population screening. At present, the delivery of a definitive diagnostic tool warrants

further refinement of the proposed metabolic signature of pancreatic cancer. Several strategies bear the potential to contribute to an improvement in diagnostic accuracy, such as the combined investigation of urine and serum profiles of the same panel of patients and controls, the integrated NMR and MS analysis of the specimen, the extension of the proposed approaches to very large populations, and the collection of multiple samples from the same individual to minimize the influence of perturbing factors not related to the disease.

■ ASSOCIATED CONTENT

Supporting Information

Figures S1 and S2. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*Phone: +39 045 8027949. Fax: + 39 045 8027929. E-mail: michael.assfalg@univr.it.

■ ACKNOWLEDGMENT

This work was supported by the University of Verona and Bruker Italia (Joint Research Project 2007); Fondazione Cariverona, Verona, Italy (young investigator grant 2007); Associazione Italiana Ricerca Cancro (AIRC); Fondazione CariParo, Padova, Italy; Ministero della Salute e Ministero dell'Università, Roma, Italy. Maintenance of the NMR spectrometer was funded in part by the Department of Biotechnology of the University of Verona.

■ ABBREVIATIONS

CA, canonical analysis; HSQC, heteronuclear single quantum coherence; MANOVA, multivariate analysis of variance; NMR, nuclear magnetic resonances; NOESY, nuclear Overhauser effect spectroscopy; PC, principal component; PCA, principal component analysis; PDAC, pancreatic ductal adenocarcinoma

■ REFERENCES

(1) Ferlay, J.; Autier, P.; Boniol, M.; Heanue, M.; Colombet, M.; Boyle, P. Estimates of the cancer incidence and mortality in Europe in 2006. *Ann. Oncol.* **2007**, *18* (3), 581–92.

(2) Key, C. Cancer of the pancreas. In *Seer Survival Monograph: Cancer Survival among Adults: U.S. Seer program, 1988—2001, patient and tumor characteristics*; Ries, L.; Young, J., Keel, G.; Eisner, M.; Lin, Y.; Horner, M. J.; National Cancer Institute: Bethesda, 2007.

(3) Kroemer, G.; Pouyssegur, J. Tumor cell metabolism: cancer's Achilles' heel. *Cancer Cell.* **2008**, *13* (6), 472–82.

(4) Cairns, R. A.; Harris, I. S.; Mak, T. W. Regulation of cancer cell metabolism. *Nat. Rev. Cancer* **2011**, *11* (2), 85–95.

(5) Fiehn, O. Metabolomics—the link between genotypes and phenotypes. *Plant Mol. Biol.* **2002**, *48* (1—2), 155–71.

(6) Nicholson, J. K.; Wilson, I. D. Opinion: understanding 'global' systems biology: metabonomics and the continuum of metabolism. *Nat. Rev. Drug Discovery* **2003**, *2* (8), 668–76.

(7) Goodacre, R.; Vaidyanathan, S.; Dunn, W. B.; Harrigan, G. G.; Kell, D. B. Metabolomics by numbers: acquiring and understanding global metabolite data. *Trends Biotechnol.* **2004**, *22* (5), 245–52.

(8) Griffin, J. L.; Shockcor, J. P. Metabolic profiles of cancer cells. *Nat. Rev. Cancer* **2004**, *4* (7), 551–61.

(9) Spratlin, J. L.; Serkova, N. J.; Eckhardt, S. G. Clinical applications of metabolomics in oncology: a review. *Clin. Cancer Res.* **2009**, *15* (2), 431–40.

(10) Fang, F.; He, X.; Deng, H.; Chen, Q.; Lu, J.; Spraul, M.; Yu, Y. Discrimination of metabolic profiles of pancreatic cancer from chronic pancreatitis by high-resolution magic angle spinning 1H nuclear magnetic resonance and principal components analysis. *Cancer Sci.* **2007**, *98* (11), 1678–82.

(11) Bathe, O. F.; Shaykhutdinov, R.; Kopciuk, K.; Weljie, A. M.; McKay, A.; Sutherland, F. R; Dixon, E.; Dunse, N.; Sotiropoulos, D.; Vogel, H. J. Feasibility of Identifying Pancreatic Cancer based on Serum Metabolomics. *Cancer Epidemiol. Biomarkers Prev.* **2011**, *20* (1), 140–47.

(12) Beger, R. D.; Schnackenberg, L. K.; Holland, R. D.; Li, D.; Dragan, Y. Metabonomic models of human pancreatic cancer using 1D proton NMR spectra of lipids in plasma. *Metabolomics.* **2006**, *2* (3), 125–34.

(13) Urayama, S.; Zou, W.; Brooks, K.; Tolstikov, V. Comprehensive mass spectrometry based metabolic profiling of blood plasma reveals potent discriminatory classifiers of pancreatic cancer. *Rapid Commun. Mass Spectrom.* **2010**, *24* (5), 613–20.

(14) Sugimoto, M.; Wong, D. T.; Hirayama, A.; Soga, T.; Tomita, M. Capillary electrophoresis mass spectrometry-based saliva metabolomics identified oral, breast and pancreatic cancer-specific profiles. *Metabolomics.* **2010**, *6* (1), 78–95.

(15) Serkova, N. J.; Glunde, K. Metabolomics of cancer. *Methods Mol. Biol.* **2009**, *520*, 273–95.

(16) Qiu, Y.; Cai, G.; Su, M.; Chen, T.; Liu, Y.; Xu, Y.; Ni, Y.; Zhao, A.; Cai, S.; Xu, L. X.; Jia, W. Urinary metabonomic study on colorectal cancer. *J. Proteome Res.* **2010**, *9* (3), 1627–34.

(17) Slupsky, C. M.; Steed, H.; Wells, T. H.; Dabbs, K.; Schepansky, A.; Capstick, V.; Faught, W.; Sawyer, M. B. Urine metabolite analysis offers potential early diagnosis of ovarian and breast cancers. *Clin. Cancer Res.* **2010**, *16* (23), 5835–41.

(18) Assfalg, M.; Bertini, I.; Colangiuli, D.; Luchinat, C.; Schäfer, H.; Schütz, B.; Spraul, M. Evidence of different metabolic phenotypes in humans. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105* (5), 1420–24.

(19) Rezzi, S.; Ramadan, Z.; Fay, L. B.; Kochhar, S. Nutritional metabonomics: applications and perspectives. *J. Proteome Res.* **2007**, *6* (2), 513–25.

(20) Bollard, M. E.; Stanley, E. G.; Lindon, J. C.; Nicholson, J. K.; Holmes, E. NMR-based metabonomic approaches for evaluating physiological influences on biofluid composition. *NMR Biomed.* **2005**, *18* (3), 143–62.

(21) Wishart, D. S.; Knox, C.; Guo, A. C.; Eisner, R.; Young, N.; Gautam, B.; Hau, D. D.; Psychogios, N.; Dong, E.; Bouatra, S.; Mandal, R.; Sinelnikov, I.; Xia, J.; Jia, L.; Cruz, J. A.; Lim, E.; Sobsey, C. A.; Shrivastava, S.; Huang, P.; Liu, P.; Fang, L.; Peng, J.; Fradette, R.; Cheng, D.; Tzur, D.; Clements, M.; Lewis, A.; De Souza, A.; Zuniga, A.; Dawe, M.; Xiong, Y.; Clive, D.; Greiner, R.; Nazyrova, A.; Shaykhutdinov, R.; Li, L.; Vogel, H. J.; Forsythe, I. HMDB: a knowledgebase for the human metabolome. *Nucleic Acids Res.* **2009**, *37* (Database issue), D603–10.

(22) Ulrich, E. L.; Akutsu, H.; Doreleijers, J. F.; Harano, Y.; Ioannidis, Y. E.; Lin, J.; Livny, M.; Mading, S.; Maziuk, D.; Miller, Z.; Nakatani, E.; Schulte, C. F.; Tolmie, D. E.; Kent Wenger, R.; Yao, H.; Markley, J. L. BioMagResBank. *Nucleic Acids Res.* **2008**, *36* (Database issue), D402–08.

(23) Rencher, A. *Methods of Multivariate Analysis*, 2nd ed.; J. Wiley: New York, 2002.

(24) Goonetilleke, K. S.; Siriwardena, A. K. Systematic review of carbohydrate antigen (CA 19-9) as a biochemical marker in the diagnosis of pancreatic cancer. *Eur. J. Surg. Oncol.* **2007**, *33* (3), 266–70.

(25) Duffy, M. J. CA 19-9 as a marker for gastrointestinal cancers: a review. *Ann. Clin. Biochem.* **1998**, *35* (Pt 3), 364–70.

(26) Lamerz, R. Role of tumour markers, cytogenetics. *Ann. Oncol.* **1999**, *10* (Suppl 4), 145–49.

(27) OuYang, D.; Xu, J.; Huang, H.; Chen, Z. Metabolic profiling of serum from human pancreatic cancer patients using ^{1}H NMR spectroscopy

and principal component analysis. *Appl. Biochem. Biotechnol.* **2011**, *165*, 148–54.

(28) Garber, K. Energy boost: the Warburg effect returns in a new theory of cancer. *J. Natl. Cancer Inst.* **2004**, *96* (24), 1805–06.

(29) Jones, S.; Zhang, X.; Parsons, D. W.; Lin, J. C.; Leary, R. J.; Angenendt, P.; Mankoo, P.; Carter, H.; Kamiyama, H.; Jimeno, A.; Hong, S. M.; Fu, B.; Lin, M. T.; Calhoun, E. S.; Kamiyama, M.; Walter, K.; Nikolskaya, T.; Nikolsky, Y.; Hartigan, J.; Smith, D. R.; Hidalgo, M.; Leach, S. D.; Klein, A. P.; Jaffee, E. M.; Goggins, M.; Maitra, A.; Iacobuzio-Donahue, C.; Eshleman, J. R.; Kern, S. E.; Hruban, R. H.; Karchin, R.; Papadopoulos, N.; Parmigiani, G.; Vogelstein, B.; Velculescu, V. E.; Kinzler, K. W. Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. *Science* **2008**, *321* (5897), 1801–06.

(30) Collisson, E. A.; Sadanandam, A.; Olson, P.; Gibb, W. J.; Truitt, M.; Gu, S.; Cooc, J.; Weinkle, J.; Kim, G. E.; Jakkula, L.; Feiler, H. S.; Ko, A. H.; Olshen, A. B.; Danenberg, K. L.; Tempero, M. A.; Spellman, P. T.; Hanahan, D.; Gray, J. W. Subtypes of pancreatic ductal adenocarcinoma and their differing responses to therapy. *Nat. Med.* **2011**, *17* (4), 500–03.

(31) Bardeesy, N.; DePinho, R. A. Pancreatic cancer biology and genetics. *Nat. Rev. Cancer* **2002**, *2* (12), 897–09.

(32) Rhim, A. D.; Stanger, B. Z.. Molecular biology of pancreatic ductal adenocarcinoma progression: aberrant activation of developmental pathways. *Prog. Mol. Biol. Transl. Sci.* **2010**, *97*, 41–78.

(33) Vareed, S. K.; Bhat, V. B.; Thompson, C.; Vasu, V. T.; Fermin, D.; Choi, H.; Creighton, C. J.; Gayatri, S.; Lan, L.; Putluri, N.; Thangjam, G. S.; Kaur, P.; Shabahang, M.; Giri, J. G.; Nesvizhskii, A. I.; Asea, A. A.; Cashikar, A. G.; Rao, A.; McLoughlin, J.; Sreekumar, A. Metabolites of purine nucleoside phosphorylase (NP) in serum have the potential to delineate pancreatic adenocarcinoma. *PLoS One* **2011**, *6* (3), e17177.