



# Paraffin-Embedded Tissue as a Novel Matrix in Metabolomics Study: Optimization of Metabolite Extraction Method

Magdalena Buszewska-Forajta<sup>1</sup> · Małgorzata Patejko<sup>1</sup> · Szymon Macioszek<sup>1</sup> · Dawid Sigorski<sup>2</sup> · Ewa Iżycka-Świeszewska<sup>2</sup> · Michał J. Markuszewski<sup>1</sup>

Received: 16 January 2019 / Revised: 13 June 2019 / Accepted: 10 July 2019 / Published online: 18 July 2019  
© The Author(s) 2019

## Abstract

Tissue is considered as a highly informative matrix in omics studies. Most commonly used tissues in metabolomics are fresh-frozen specimens. However, an alternative metabolomics matrix, up to now most often applies in immunohistochemistry or histopathology, is formalin-fixed, paraffin-embedded tissue (FFPE). The main advantages of FFPE tissues over fresh-frozen ones are the possibility to be stored at ambient temperature which is a cost-effective method and, consequently, their higher availability in biobanks. The main aim of the study was to develop and optimize the preparation protocol for FFPE samples in metabolomics, using cancer tissue and corresponding control tissue sections. Several conditions were evaluated: type and volume of deparaffinization solvent, number of purification cycles, solvent used in purification step, extraction solvent, and thickness of FFPE samples. Samples were analyzed by gas chromatography coupled with mass spectrometry (GC–MS). As a result, a sample preparation protocol of FFPE tissues for metabolomics was developed. Moreover, applicability of the proposed method was tested with the use of FFPE tissues obtained from prostate cancer patient and non-cancer controls.

**Keywords** Paraffin-embedded tissue · Metabolomics · Prostate cancer · GC–MS · Sample preparation · AMDIS

## Introduction

Metabolomics is a field of study focusing on quantitative and qualitative analysis of products of the metabolic pathways [1]. The aim of metabolomics is to provide the information about the composition of metabolome and changes within it under various conditions. In metabolomics, three different approaches are utilized: metabolic fingerprinting (determination of global metabolome), metabolic profiling (quantitative analysis of selected set of compounds that are related by the similar function, properties or belong to the same biochemical pathway), and targeted analysis (determination

of particular compounds). Taking into account the object of the research, human [2–4], plant [5–7], or microbial [8–10] metabolomics is distinguished [11]. Mostly, human metabolomics concentrates on pathophysiological disorders, either on evaluating potential biomarkers or explaining pathomechanism. Consequently, in metabolomics, different types of biological matrices are utilized, mainly urine, serum, and tissue [12]. Many diseases, especially cancer, initialize in tissue and because of that, the slightest qualitative and quantitative changes in tissue metabolic profile can be observed at the early stage of disease development. Consequently, tissue seems to be an adequate matrix when the aim is to investigate the mechanism of the disease [12]. In genomics or proteomics of cancer, formalin-fixed, paraffin-embedded (FFPE) tissues are commonly used samples, whereas they were rarely described in metabolomics research [13–17]. Preparation of FFPE samples includes three steps: formalin fixation, dehydration, and paraffin embedding. Formalin fixation is performed with the use of neutral-buffered formalin solution (3.7–4.0% formaldehyde buffered with phosphate at a pH of 7) [18]. The mechanism of formalin fixation consists of two steps. In the first step, formaldehyde reacts with amino group (for example from proteins) to create a

Published in the topical collection *The 41st Symposium Chromatographic Methods of Investigating Organic Compounds* with guest editor Michał Daszykowski.

✉ Michał J. Markuszewski  
michal.markuszewski@gumed.edu.pl

<sup>1</sup> Department of Biopharmaceutics and Pharmacodynamics, Medical University of Gdańsk, Al. Gen. J. Hallera 107, 80-416 Gdańsk, Poland

<sup>2</sup> Department of Pathology and Neuropathology, Medical University of Gdańsk, Dębinki 1, 80-211 Gdańsk, Poland

Schiff base, which, in the second step, reacts with another amino group and creates a stable crosslink by methylene bridge [18, 19]. The aim of formalin fixation is to preserve the tissue by the inhibition of proteolytic enzymes as well as the growth of microorganisms [20, 21]. Consequently, fixed tissue should remain biologically stable through the time of storage [21].

The next step of FFPE sample preparation is paraffin embedding, followed by the dehydration. Dehydration includes washing the specimen with ethanol and xylene. It is performed to remove the water from tissue, because paraffin, which is the most often used embedding factor, cannot penetrate through wet-fixed matrix [21, 22]. The final step is paraffin embedding. Melted paraffin is added to the dry tissue, which is then left to solidify at room temperature [18, 21]. The aim of embedding is to improve the preservation as well as to ease further sample sectioning.

The main advantage of FFPE samples is that tissue preservation through formalin fixation and paraffin embedding provides the stability of tissue morphology, and enables to store the sample at ambient temperature. It is considered to be a more cost and space effective choice in comparison to fresh-frozen tissues that must be kept at special conditions (e.g., temperature  $-80\text{ }^{\circ}\text{C}$  or liquid nitrogen) to maintain biological stability. However, comparing to fresh-frozen tissue, application of FFPE samples requires more complex sample preparation protocol, including deparaffinization and purification step. Furthermore, specimen preservation by formalin fixation may cause loss of some group of compounds due to the creation of cross-links between proteins and other biomolecules. This aspect is critical for FFPE utilization in metabolomics, especially in untargeted approach, that focuses on the identification of as many different chemical class of metabolites as possible in analyzed biological matrix. Nonetheless, previously published research indicated that FFPE tissues can be successfully implemented in metabolomics; however, there is a lack of data concerning optimization of sample preparation step. Application of FFPE was performed with several analytical techniques and the goal was to evaluate the influence of preservation procedure on the quality of obtained metabolic profiles [13–17, 23, 24]. It was proved that formalin fixation and paraffin embedding may cause loss of specific group of metabolites such as amino acids, peptides, carbohydrates, steroids, or fatty acids esters. Nevertheless, it was possible to observe statistically significant differentiation between healthy and prostate cancer FFPE tissues based on the levels of detected compounds.

However, application of FFPE was reported in several studies, detailed optimization of sample preparation step was not conducted. Due to the fact that the fixation as well as deparaffinization may lead the loss of metabolites, sample preparation step seems to be crucial for the final results of analysis. For this reason, the main aim of the study was to

develop reliable and effective sample preparation protocol for FFPE samples, which could be applied for untargeted metabolomics approach. There are previously published work focusing on the developing of sample preparation, but they mainly cover DNA and RNA extraction from FFPE samples [25–28], whereas, to our knowledge, there is no such research in metabolomics that would focus on the comparison of different sample preparation protocols. Described sample preparation method consisted of four parts: deparaffinization, purification, extraction, and derivatization. Among the study, deparaffinization and purification methods were developed. Two deparaffinization methods were compared: based on xylene and hot water. In addition, several conditions were evaluated: type and volume of deparaffinization solvent, number of purification cycles as well as type of solvent used in purification step, extraction solvent, and thickness of FFPE samples. All samples were analyzed with the use of gas chromatography hyphenated with triple quadrupole mass spectrometer (GC–QqQ/MS) technique. Application of GC–MS required a derivatization step to convert metabolites to their volatile derivatives. Developed method was applied for the untargeted metabolomics analysis of FFPE tissue samples obtained from prostate cancer patients and healthy prostatic tissue.

## Experimental

### Chemicals

Pentadecanoic acid standard was purchased from Sigma-Aldrich, USA. The derivatizing agents, methoxyamine hydrochloride, pyridine, *N,O*-bis(trimethylsilyl)trifluoroacetamide (BSTFA), and chlorotrimethylsilane (TMCS), were obtained from Sigma-Aldrich, St. Louis, MO, USA. Heptane, ethanol and xylene were purchased from Chem-solve, Witko, Poland. Chloroform was purchased from POCH, Gliwice, Poland, and deionized water was obtained with Milli-RO and Milli-QPlus instrumentation from Millipore, Switzerland.

### Sample Collection

The research was approved by the Independent Commission for Bioethics Research, Medical University of Gdańsk, NKBBN/448/2015 and NKBBN/432/2016. The FFPE samples were obtained, preserved, and stored in the Department of Pathology and Neuropathology, Medical University of Gdańsk, Poland, according to routine protocol used in pathology laboratories. All methods used to carry out the study were performed in accordance with the relevant guidelines and regulations. All the participants were recruited at the Department of Pathology and Neuropathology, Medical

University of Gdańsk, Poland by a single physician—a specialist in pathology and all the patients agree to use the collected tissue for scientific study. In the following study, 67 samples were collected. Samples were divided according to developed step as follows: 9 samples were used to choose the proper deparaffinization method, 24 samples for tissue samples purification, 25 samples for investigation of the influence of FFPE thickness on metabolic profile, and 9 samples for the application study. Samples were stored in an ambient temperature and dark place up to the day of analysis.

## Developing of Sample Preparation

Sample preparation method was developed regarding several conditions: type and volume of deparaffinizing solvent, number of purification cycles, type of solvent for sample purification, solvent used for extraction, and thickness of samples. Each tissue sample was weighed after the deparaffinization procedure, what was used for normalization. First, two deparaffinizing agents: hot water and xylene, were compared in terms of deparaffinization efficiency. In addition, thickness of sample was tested. In reference to the literature, thickness of utilized paraffin-embedded tissue applied in clinical settings varies between 7 and 40  $\mu\text{m}$ , while in metabolomics, most common is 40  $\mu\text{m}$  [13, 14]. In the presented research, five thicknesses of samples, namely, 10  $\mu\text{m}$ , 15  $\mu\text{m}$ , 20  $\mu\text{m}$ , 30  $\mu\text{m}$ , and 40  $\mu\text{m}$ , were tested. Afterward, purification step was optimized. Three purification methods were evaluated: utilization of no purification procedure, one-step purification with the use of ethanol, and two-step purification using ethanol followed by acetone. Moreover, the number of purification cycles was compared Paragraph 2 to describe final protocol of optimized FFPE sample preparation procedure.

### Sample Deparaffinization and Purification Procedure

1 mL of xylene was added to each specimen and sample was vortex-mixed for 1 min. Samples were centrifuged for 5 min (16 657 g, 25 °C). After that, xylene was discarded and procedure was repeated three times. Next step included purification procedure. Xylene residue was removed with the use of 0.7 mL of 96% EtOH. Samples were vortex-mixed and centrifuged, and this step was repeated three times. Thereafter, tissues were slightly dried. After deparaffinization and purification, each deparaffinized prostate tissue was weighed.

### Sample Extraction Procedure

Metabolites extraction from prostate tissues was performed as described by Yuan et al. [29], with modification. Tissue samples were homogenized using 150  $\mu\text{L}$  of methanol: water mixture (1:1, v/v), preceded by addition of 5  $\mu\text{L}$  of internal

standard (IS, 1 mg mL<sup>-1</sup>, *n*-pentadecanoic acid in methanol). The procedure included 1 min of homogenization and 1 min of cooling on ice, and was repeated three times. Afterward, samples were centrifuged (10 min, 16 657g; 4 °C). 120  $\mu\text{L}$  of supernatant was collected. The residue was then homogenized with 150  $\mu\text{L}$  of mixture chloroform:methanol (3:1, v/v) and the whole procedure was repeated. Finally, 120  $\mu\text{L}$  of organic fraction was collected and pooled with aqueous phase. Pooled fraction was evaporated to dryness with the use of vacuum concentrator (Genevac Inc., Valley Cottage, NY, USA) for 2 h at 36 °C. After that, derivatization was performed.

### Derivatization Procedure

Two-step derivatization procedure was applied to enhance thermal stability and volatility of analyzed compounds. First step was methoxylation performed by the addition of 30  $\mu\text{L}$  of methoxyamine dissolved in pyridine (15 mg mL<sup>-1</sup>) to the dry mass. Samples were then kept for 16 h in a dark place, at room temperature. The next step was silylation that included the addition of 30  $\mu\text{L}$  of mixture of BSTFA and TMCS (99:1, v/v). Samples were vortex-mixed for 5 min and kept for 60 min at 70 °C. Derivatized samples were diluted in 90  $\mu\text{L}$  of heptane and vortex-mixed for 5 min. Samples were introduced to the GC–MS system.

### Preparation of Alkane Standard Mixture

To calculate retention index of analyzed compounds, alkane standard mixture in concentration of 10  $\mu\text{g mL}^{-1}$  was prepared by dilution in heptane and introduced to GC/MS system.

### Instrumentation

Analyses were performed with the use of gas chromatography–tandem mass spectrometry (GC/MS/MS) system from Shimadzu (Kyoto, Japan) 8030 series composed of autosampler (AOC-20S), autoinjector (AOC-20i), and gas chromatograph (GC 2010 plus) coupled with a triple quadrupole mass spectrometer. Post-analysis processing of the obtained data was conducted using Automated Mass Spectral Deconvolution and Identification System software (AMDIS, <http://chemdata.nist.gov/mass-spc/amdis/>) and Mass Profiler Professional B. 02.01 software (Agilent Technologies, Waldbronn, Germany) as well as Mass GC/MS Solution Software version 4.01 (Shimadzu, Kyoto, Japan). Statistical analysis was performed with the use of Metaboanalyst 4.0 software (<http://www.metaboanalyst.ca/>), MatLab (Mathworks, Natick, Ma, USA), and Simca P+ 13.0.3 software (Umetrics, Umea, Sweden).

## GC/MS Analysis

Derivatized samples were introduced to gas chromatograph equipped with Zebron ZB-5MS (30 m × 0.25 mm i.d., 0.25 μm film thickness) capillary column (Phenomenex, Torrance, CA, USA). Helium was used as a carrier gas, with the constant pressure mode (65.2 kPa) and total flow 10 mL/min. 1 μL of sample was injected in a splitless mode. The temperature of injection port was maintained at 320 °C. Compounds were separated using temperature gradient program. The oven temperature was set at 70 °C for 1 min, and then, temperature increased to 320 °C with grade 8 °C min<sup>-1</sup> and held at 320 °C for 10 min to elute any impurities with high boiling point. The temperature of MS interface was set at 320 °C.

The analyses were performed in a scan mode with mass range 50–1000 *m/z*. Electron impact (EI) ion source with temperature set at 220 °C was used. The time of each analysis was 43 min. All samples were analyzed in one sequence in randomized order. At the beginning of the sequence, standard mixture of alkanes was analyzed.

## Data Preprocessing

Obtained chromatograms were transferred from GC/MS Solution Software Postrun Analysis (Shimadzu, Kyoto, Japan) to AMDIS with NIST spectra library to perform deconvolution and identification steps. At first, retention indices for all compounds present in each sample were calculated based on retention time ( $t_R$ ) and retention index (RI) of alkane mixture. The retention time of each compound was normalized by retention time and retention index of the nearest eluting alkane. Deconvolution procedure of raw data was applied to reduce noise signals and matrix interferences from the background. To deconvolute spectra in whole *m/z* range, following parameters were used: width = 15, adjacent peak subtraction = two, resolution = low, sensitivity = low, and shape requirements = low.

## Statistical Analysis

The post-analysis processing included data alignment regarding the retention time and data filtration based on QA criteria and presence in the compared groups. Data processing was conducted using Mass Profiler Professional B.02.01 software (Agilent Technologies, Waldbronn, Germany). Only compounds present in at least 80% of all samples with CV lower than 30 were used for further statistical calculations. The filtered data set was normalized using intensity of IS (n-pentadecanoic acid, Sigma-Aldrich). Univariate and multivariate statistical comparisons were conducted using Matlab 2016b software (Mathworks, Natick, Ma, USA) to evaluate compounds contributing the most into group

discrimination. Identification of the statistically significant metabolites was based on RI, RT and mass spectra of compounds detected in tissue samples and compared with hits included in the NIST 11 and in-house spectral libraries.

## Results and Discussion

### Optimization of Deparaffinization Procedure

#### Type of the Deparaffinization

In the current study, the deparaffinization procedure of formalin-fixed, paraffin-embedded tissue samples was evaluated. Two methods of deparaffinization were tested: with the use of xylene and hot water.

Results showed that after the deparaffinization step with the use of hot water, the paraffin still covered the biological material. The main concern included the possibility of tissue destruction after performing too many cycles with the use of hot water to obtain efficient deparaffinization step. Consequently, this method was found as problematic to control and deparaffinization was not efficient and reliable. For this reason, method based on hot water was not applied for further metabolic investigation. Second, deparaffinization method was compared based on xylene. Deparaffinization with the utilization of xylene was efficient as the paraffin was easily dissolved and removed, and no paraffin residues were observed on the surface of the tissue. Deparaffinization was easy to control and did not provide the destruction of biological matrix.

It is worth highlighting that in genomics, both tested methods were previously successfully applied [26, 28, 30–32]. For instance, Sengüven et al. reported the use of xylene-free deparaffinization protocol [26]. Furthermore, Kalantari et al. [28] compared hot water and xylene as two methods for deparaffinization in case of DNA extraction. Obtained results showed that both methods are similarly effective. However, authors indicated that application of hot water seems to be a shorter, safer, and easier one. Nonetheless, in proteomics, the most commonly used deparaffinization method is based on xylene [27, 33, 34].

Results of our study proved that the deparaffinization with the use of hot water was not efficient and resulted in contamination of tissue samples; the paraffin residues were still observed, also on the tissue surface. On the other hand, in the case of xylene, deparaffinization was found as more efficient. Xylene caused that the paraffin melted and could be removed from the sample.

The first application of paraffin-embedded tissue in metabolomics was reported by Kelly et al. [17] in 2011. Targeted LC–MS analysis was carried out to determine polar metabolites in FFPE sarcoma tissues. However, in

their study, deparaffinization step was done with the use of 80% methanol solution. Samples were incubated in 70 °C for 45 min. In reference to our results, utilization of high temperature, as in the method with hot water, can cause protein denaturation. In addition, application of the high temperature and methanol may accelerate the extraction step and cause the significant loss of some compounds. On the other hand, application of high temperature and methanol as a solvent may comprise two steps of protocol, namely deparaffinization and purification in one step. Because of that, it can be assumed that method based on application of xylene can be found as more efficient for metabolomics. Furthermore, this method was previously reported in metabolomics study, for instance by Wojakowska et al. [13, 14]. However, it should be highlighted that influence on the deparaffinization step with the use of several methods has not been previously reported. Wojakowska et al. [13] used untargeted metabolomics approach to compare three types of mouse kidney tissue samples: fresh-frozen; formalin-fixed; and formalin-fixed, paraffin-embedded. Samples were analyzed by gas chromatography–mass spectrometry (GC–MS) technique. In the case of paraffin-embedded tissue, it was proved that embedding impacts on extraction efficiency. In the case of fresh-frozen specimen, the profile was more complex in reference to paraffin-embedded tissue. As a result, 75% of all identified metabolites were common for three types of samples. However, paraffinization can be considered as alternative method for tissue preparation.

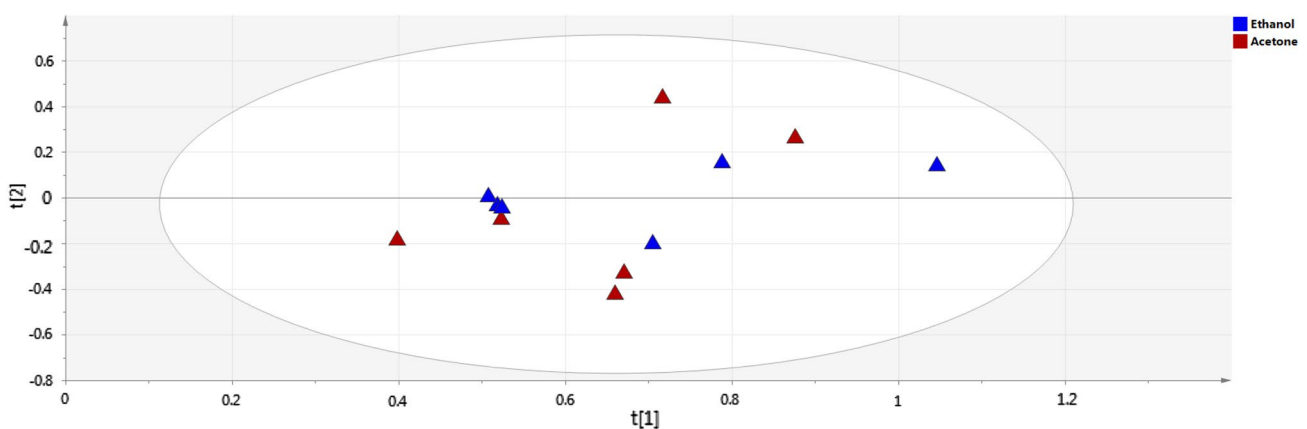
### Type of Solvent Used for Purification

To evaluate the influence of the purification procedure on the quality of analyzed samples, two methods were tested. First method included three cycles of purification with the use of 0.7 mL 96% ethanol, whereas second method was enriched

by the addition of two cycles of purification with 1 mL of acetone. Twelve samples were analyzed; six samples for each analyzed group. Samples underwent deparaffinization, extraction, and derivatization procedures as described in “Sample deparaffinization and purification procedure”, “Sample extraction procedure”, and “Preparation of alkane standard mixture” sections, respectively.

Obtained data were processed according to protocol described in “Data Preprocessing” section. First step of statistical analysis included normalization procedure. Signals were normalized by peak intensity of internal standard (*n*-pentadecanoic acid) and weight of each tissue sample. Obtained data were filtrated according to quality assurance (QA) criteria (coefficient of variation, CV, lower than 30%) and presence in group (presence in at least 80% of samples from one group). After filtration, 29 compounds were chosen for further statistical analysis. Multivariate statistical methods were performed to check if addition of acetone provides statistically significant differences between profiles obtained after two compared purification procedures. In the first step of multivariate analysis, principal component analysis (PCA) was applied to evaluate general trends in the data. PCA models were built using Simca P+ 13.0.3 software. Obtained PCA model is presented in Fig. 1.

As it can be observed in Fig. 1, samples corresponding to two different purification procedures are not separated. It can be assumed that addition of purification step with the use of acetone does not improve sample clean-up. In addition, samples are spread on the graph, what can be explained by the high heterogeneity of biological samples. Afterwards, obtained data were analyzed with the use of parametric or nonparametric statistical tests to prove the assumption about the influence of acetone on complexity of metabolic profile as well as to select statistically significant compounds. First step included checking the normality of the data distribution



**Fig. 1** PCA model built on data obtained with GC–QqQ/MS. The calculated  $R^2$  factor was  $R^2=0.899$ . Charts correspond to two tested purification procedures. Protocol based on ethanol is marked by blue

triangles and purification procedure enriched with acetone is presented by red triangles

by Shapiro–Wilk test. For normally distributed variables, with homogeneity of variances, the standard *t* test was used. In the case of the lack in homoscedasticity, Welch's *t* test was applied. Mann–Whitney *U* test was applied for variables without normal distribution. The multiple comparison correction of calculated *p* values was performed with the use of Benjamini–Hochberg false discovery rate (FDR) test. Analyses were carried out with the use of Matlab 2016b software. Results are presented in Table 1.

As it can be observed in Table 1, calculated *p* value was under  $p < 0.05$  for three compounds, namely butanoic acid, tetradecanoic acid, and octadecanoic acid. However, application of *p* value correction excluded these compounds as statistically significant. Results show that sample purification procedure based on enrichment of acetone does not enhance the purity of sample. In addition, comparison of profiles obtained with the use of two tested methods

**Table 1** Univariate statistical analysis performed on samples analyzed with GC–EI–QQQ/MS

No.	Compound	RT [min]	<i>P</i> value	<i>P</i> value after FDR correction
1.	Butanoic acid	9.08	0.009	0.133
2.	Tetradecanoic acid	20.40	0.015	0.133
3.	Octadecanoic acid	25.34	0.018	0.126
4.	Heptadecanoic acid	24.18	0.065	0.271
5.	Hexadecanoic acid	28.96	0.068	0.315
6.	Octadecanoic acid	30.81	0.091	0.315
7.	Threonine	11.56	0.093	0.271
8.	Butanedioic acid	11.94	0.099	0.315
9.	Ethanedioic acid	8.55	0.123	0.358
10.	Threonic acid	16.05	0.138	0.370
11.	Sucrose	29.37	0.144	0.337
12.	Lactic acid	7.24	0.201	0.485
13.	Hexanoic acid	7.51	0.224	0.485
14.	Proline	15.48	0.240	0.495
15.	Acetic acid	7.52	0.263	0.485
16.	Nonanoic acid	12.80	0.264	0.485
17.	Propanoic acid	7.22	0.331	0.578
18.	Tagatose	20.60	0.394	0.627
19.	Aspartic acid	13.87	0.394	0.627
20.	Tetradecanoic acid	20.40	0.436	0.615
21.	Benzoic acid	10.77	0.654	0.789
22.	13-Octadecanoic acid	25.04	0.666	0.789
23.	Glycerol	11.18	0.681	0.789
24.	Octadecanoic acid	25.09	0.687	0.822
25.	1,4-Butanediol	8.80	0.705	0.822
26.	Glycine	11.78	0.734	0.803
27.	Hexadecanoic acid	22.98	0.774	0.818
28.	Myo-inositol	23.46	0.794	0.895
29.	Propanetricarboxylic acid	19.90	0.886	0.940

shows that addition of acetone does not influence their complexity. To our knowledge, this is a first time when the influence of acetone on the efficiency of FFPE sample purification was tested for metabolomics approach. It was evaluated by the comparison of metabolic profiles of samples prepared with the use of two different clean-up methods. Obtained results proved that there is no need to enlarge sample purification by the addition of acetone.

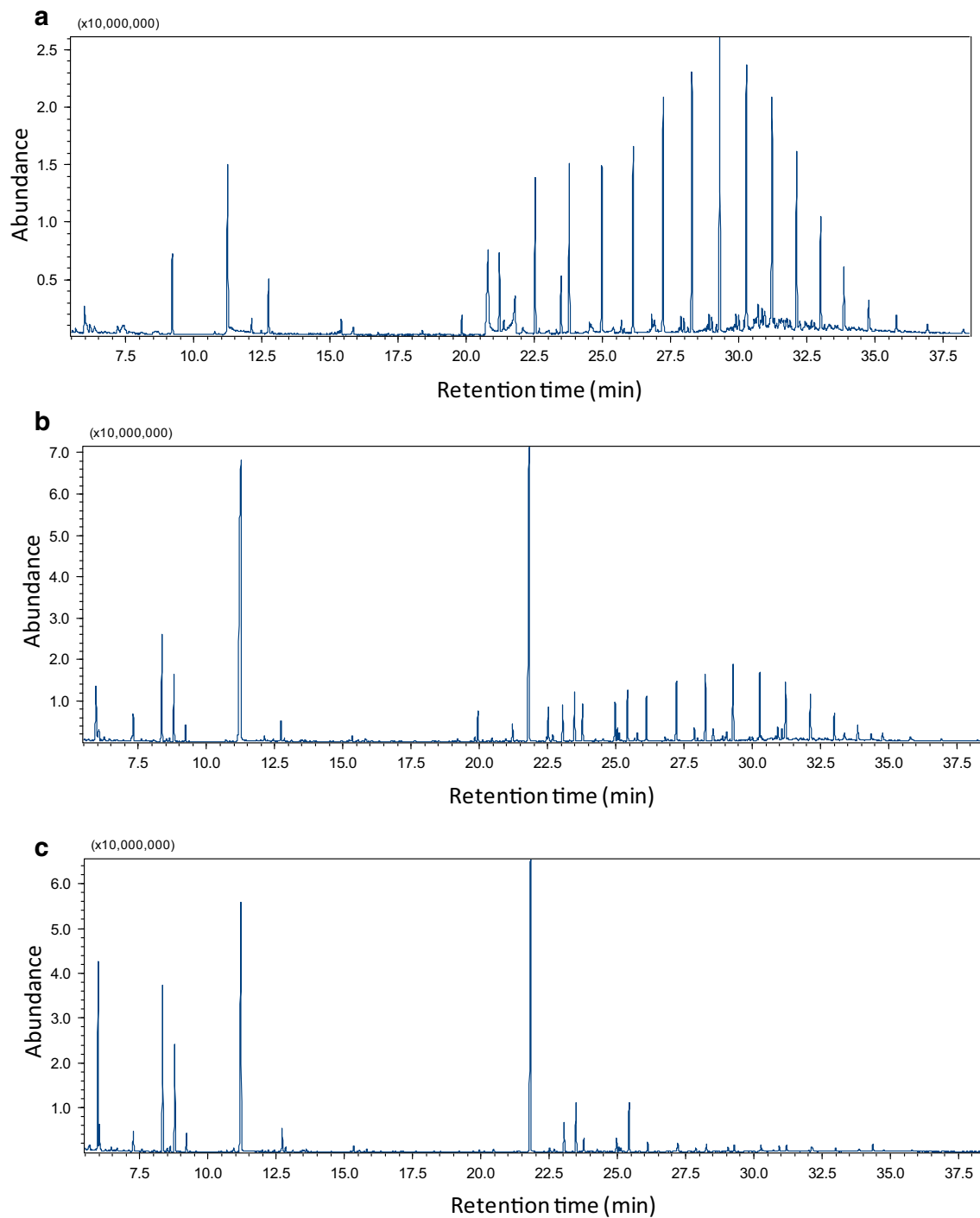
### Number of Purification Cycle

During the method optimization, the numbers of the purification cycles were taken into account. Three sets of samples were prepared:

- (A) samples deparaffinized without purification step;
- (B) deparaffinization followed by two cycles of purification process with the use of 1 mL of ethanol for each cycle;
- (C) deparaffinization followed by three cycles of purification process with the use of 0.7 mL of ethanol for each cycle.

Further steps of sample preparation procedure and data processing were performed as described above (“[Sample deparaffinization and purification procedure](#)”, [Sample extraction procedure](#)”, [Preparation of alkane standard mixture](#)”, and [Data Preprocessing](#)” sections). Exemplary metabolic profiles obtained with the use of GC–EI–QQQ/MS are presented in Fig. 2.

As it was observed, in the case of no purification step, paraffin is not fully removed. Sample is contaminated by the paraffin residue which is seen on the last part of chromatogram as regular peaks characteristic for the alkanes (Fig. 2a). On the other hand, on chromatogram presented in Fig. 2b, it is seen that application of two-step purification with higher volume of ethanol significantly reduced the amount of the paraffin impurities; however, the paraffin is still present in the sample. The most pure profile was obtained in the case of three-step purification covered by 0.7 mL of ethanol (Fig. 2c). As a result, three-step purification processes was chosen for further analysis. Previously reported protocols applied two cycles of purification with the use of ethanol. However, based on theoretical knowledge, division of the same volume of the solvent enhances the efficiency of the sample purification. Although the procedure proposed here may extend the time of sample preparation step, it enables to obtain more pure and complex metabolic profile instead.



**Fig. 2** Chromatograms obtained from the GC–MS analysis of samples prepared with modified purification step to control the influence of the number of purification cycles on the sample purity. **a** Tissue

samples without purification step, **b** two cycles using 1 mL of ethanol, and **c** purification composed of three cycles with the use of 0.7 mL of ethanol

### The Influence of the Paraffin Specimen Thickness

In presented study, different thickness values of specimen were tested: 10  $\mu\text{m}$ , 15  $\mu\text{m}$ , 20  $\mu\text{m}$ , 30  $\mu\text{m}$ , and 40  $\mu\text{m}$ . Five replicates were prepared for each thickness. Sample

preparation and data processing were performed as described in Experimental section. Obtained data were filtered based on QA criteria and presence in the group. Only compounds present in at least 80% of samples with CV lower than 30% were used for further statistical calculations. As a result,

28 compounds were chosen for further statistics. Filtered data set was normalized using abundance of IS (*n*-penta-decanoic acid) and calculated normalization factor-tissue weight divided by the thickness of the specimen. This type of normalization excluded the influence of weight on metabolic profile quality. As a result, observed differences are a consequence of various sample thicknesses. Multivariate statistical calculations were applied to evaluate how metabolic profiles differed between each other depending on the thickness of used paraffin specimen. For this reason, analyzed samples were divided into five groups consisting of five replicates (Group 1—10  $\mu\text{m}$ , Group 2—15  $\mu\text{m}$ , Group 3—20  $\mu\text{m}$ , Group 4—30  $\mu\text{m}$ , and Group 5—40  $\mu\text{m}$ ). Twenty-eight selected analytical signals were used to build PCA model. Results are presented in Fig. 3.

No clustering was observed in the presented model what provides the assumption about negligible influence of the paraffin thickness on the differentiation of the metabolic profiles. Statistical analysis included ANOVA test or Kruskal–Wallis depending on normality of data and the homogeneity of the variances. Normal distribution was tested by Shapiro–Wilk test, while homogeneity of the variances was tested by Levene test. Post hoc test was carried out with the use of Fisher test or Wilcoxon test after ANOVA or Kruskal–Wallis, respectively. Calculations were covered by RStudio [35] and MetaboAnalyst 3.6

As a result, four compounds, namely: citric acid, glycine, inositol, and 5-oxo-proline, demonstrated statistically significant differences between five tested groups. Average abundances of chosen metabolites are shown in Fig. 4.

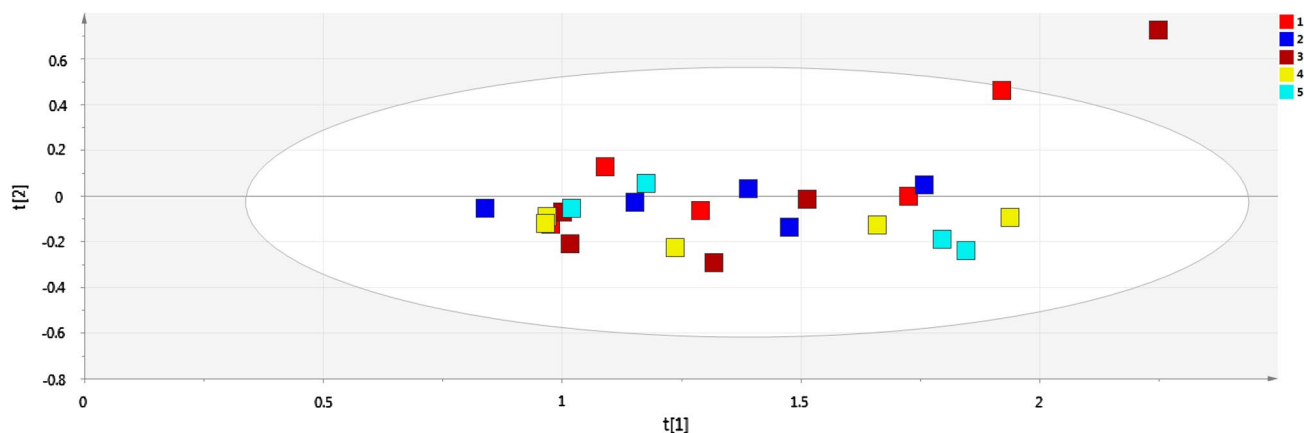
Selected compounds are presented in Table 2. Post hoc test provides the information about observed differences between groups.

Statistical analysis proved that the differences between obtained profiles are not statistically significant in the case

of paraffin tissue samples with 10  $\mu\text{m}$ , 15  $\mu\text{m}$ , and 20  $\mu\text{m}$  thick. These results were similar for each compound taken into account. However, multivariate statistical analysis enabled to select four compounds for which differences in profiles depending on tissue thickness were observed. Post hoc test, performed for glycine, shows that only profiles of metabolites extracted from paraffin-embedded tissue 40  $\mu\text{m}$  thick differed in statistically significant way from other tested thickness of samples. Results obtained for the rest three compounds, namely inositol, citric acid, and proline, showed that profiles obtained from samples with 30  $\mu\text{m}$  and 40  $\mu\text{m}$  of thickness are different from profiles obtained from samples with 10  $\mu\text{m}$ , 15  $\mu\text{m}$ , and 20  $\mu\text{m}$  thick. Results of post hoc test provided information that changes are on infinitesimal level in the case of paraffin-embedded tissue with thickness between 10 and 20  $\mu\text{m}$ . It can be assumed that the most reliable, complex, and repeatable profiles are obtained in the case of paraffin-embedded tissues with thickness not lower than 30  $\mu\text{m}$ . However, Kelly et al. [17] as well as Wojakowska et al. [13, 14] used 40  $\mu\text{m}$  thickness of paraffin-embedded tissue. Nonetheless, there is lack of knowledge in the field of the optimization regarding tissue thickness and results presented here provide information that tissue thickness should be at least 30  $\mu\text{m}$ .

### Application of the Optimized Sample Preparation Procedure

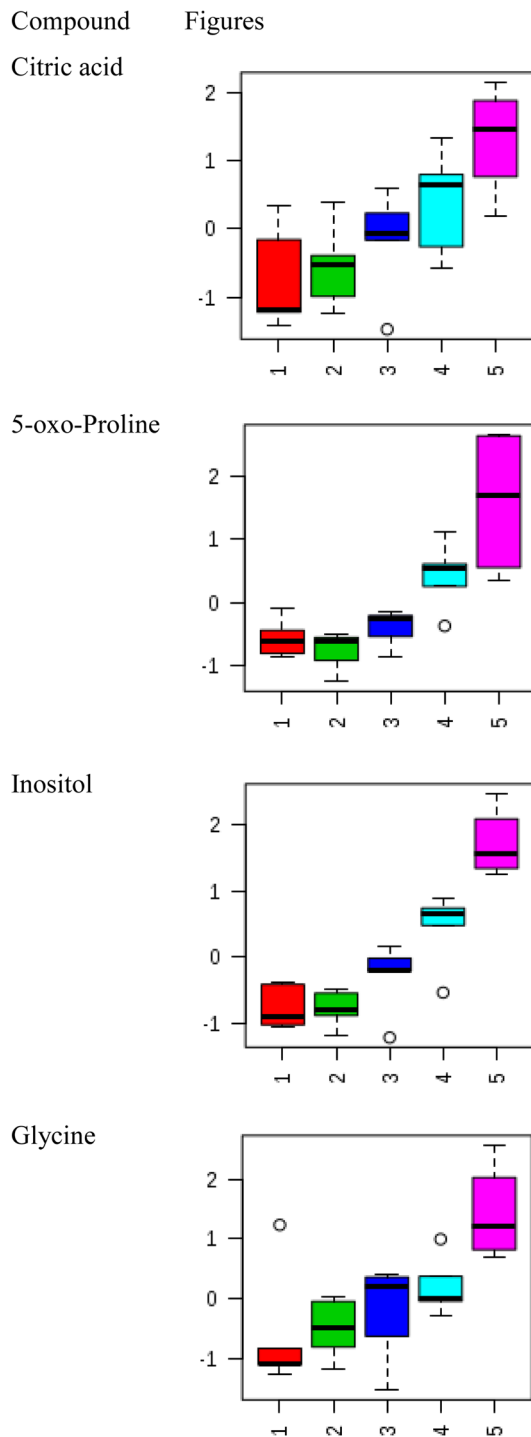
Optimized sample preparation procedure was utilized for the analysis of 9 FFPE prostate tissue samples. Four samples were obtained during prostatectomy surgery from patient with diagnosed prostate cancer and five were control group composed of tissue obtained during prostatectomy from non-cancer patient. 30  $\mu\text{m}$ -thick samples were chosen for the analysis. Samples were prepared according to the optimized



**Fig. 3** PCA models built on data obtained with GC–QqQ/MS. The calculated  $R^2$  was 0.994. Boxes correspond to five groups differing in thickness of paraffin-embedded tissue. Group 1—10  $\mu\text{m}$  is marked by

red boxes, Group 2—15  $\mu\text{m}$  by dark blue boxes, Group 3—20  $\mu\text{m}$  by brown color, Group 4—30  $\mu\text{m}$  by yellow color, and Group 5—40  $\mu\text{m}$  by light blue boxes





**Fig. 4** Compounds differentiating compared groups at statistical significant level and box plots representing their average abundance. Each number corresponds to group characterized by different thickness of FFPE tissue (Group 1—10  $\mu\text{m}$ , Group 2—15  $\mu\text{m}$ , Group 3—20  $\mu\text{m}$ , Group 4—30  $\mu\text{m}$ , and Group 5—40  $\mu\text{m}$ )

procedure as described above and consisted of three main steps: deparaffinization (described in “[Sample deparaffinization and purification](#)” section), extraction (described in

“[Sample extraction procedure](#)” section), and derivatization (described in “[Derivatization procedure](#)” section). Selected chromatograms from cancer and control group are presented in Fig. 5.

As a result, 157 compounds were identified. Identification of 157 compounds was based on the calculated retention index, retention time, and comparison of obtained mass spectrum to the library (in-house library based on NIST). According to the Fiehn protocol [36], the usual number of metabolites identified by GC–MS in biological matrices varies between 120 (cell culture) and 200 (urine). Matsuo et al. [37] reported that with the use of GC–MS analysis, 500–1000 chromatographic peaks can be obtained in a single run; however, only 100–200 metabolites (20%) can be identified. It has to be highlighted that the number of identified compounds in GC–MS strongly depends on the applied library. The number of 157 identified compounds is comparable with results obtained by Wojakowska et al. [13].

Obtained data were processed as shown in “[Data Preprocessing](#)” section. Processed and aligned data were filtered based on the presence of the group. Compounds present in at least 80% of samples were used for further statistical calculations. As a result, 24 compounds were selected for statistical analysis. Restrict filtration was carried out to avoid selection of false-positive features. Next, normalization was performed based on the intensity of the IS as well as weight of the tissue. Normalized data were used to build PCA model (Fig. 6).

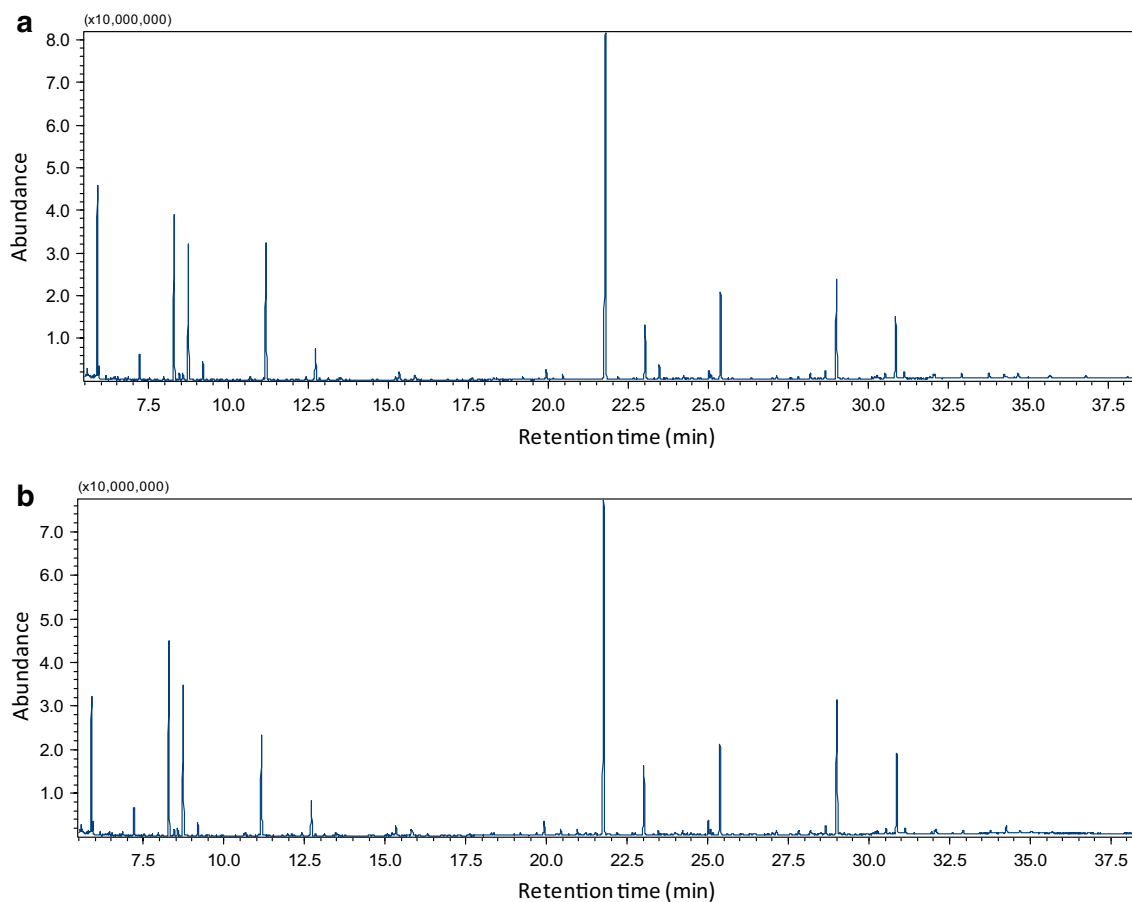
Although PCA belongs to unsupervised methods, some trends in data grouping can already be observed. Furthermore, data underwent univariate statistical analysis. In the first step, the normality of the data distribution and equality of variances were tested by Shapiro–Wilk and Levene test, respectively. Based on the result, either parametric *t* student or Mann–Whitney *U* test was applied. Calculated *p* values were corrected by Benjamini–Hochberg false discovery rate (FDR) test. Analyses were performed with the use of Matlab and Simca P+ software. Obtained results are presented in Table 3.

Table 3 shows ten compounds found to differ between two tested groups at statistically significant level (*p* value after Benjamini–Hochberg correction under  $p < 0.05$ ). The main goal of applicability as well as further filtration and statistical analysis was to assess whether developed method allows identifying significant compounds that would be biologically reliable and have ever been described in prostate cancer tissues.

As can be observed, most of statistically significant metabolites belong to fatty acids. Among them, citric acid was also identified. It was calculated that amount of fatty acids in prostate cancer group was up-regulated versus the amount determined in control samples. The opposite phenomenon was observed in the case of citric acid. In prostate

**Table 2** Statistically significant compounds, selected by ANOVA or Kruskal–Wallis methods

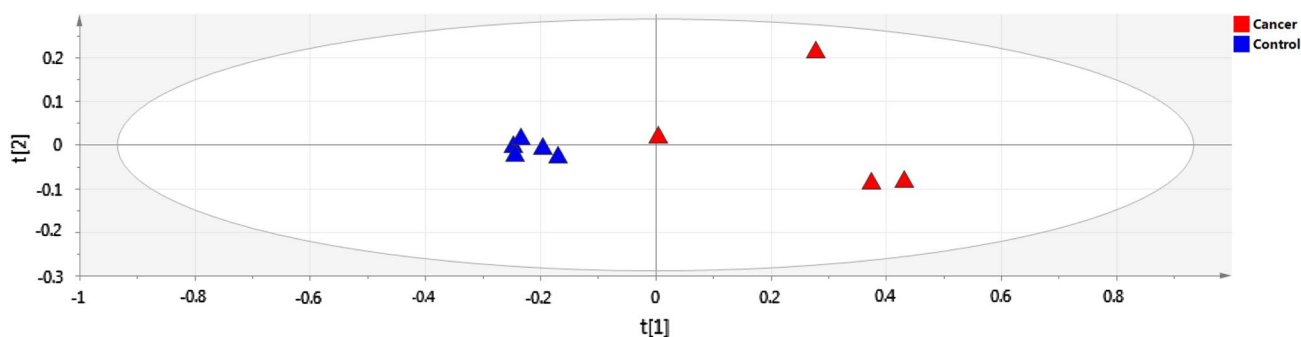
No.	Compound	Retention time	Statistical method	<i>p</i> value	Post hoc test	Post hoc results
1.	Citric acid	19.93	ANOVA	0.005	Fisher's test	4–1; 5–1; 5–2; 5–3
2.	5-Oxo-proline	15.53	Kruskal–Wallis	0.002	Wilcoxon	4–1; 5–1; 4–2; 5–2; 4–3; 5–3
3.	Inositol	23.46	ANOVA	5.94E–7	Fisher's test	4–1; 5–1; 4–2; 5–2; 4–3; 5–3; 5–4
4.	Glycine	11.81	ANOVA	0.006	Fisher's test	5–1; 5–2; 5–3; 5–4

**Fig. 5** Example of samples chromatograms obtained from GC–MS analysis of 30  $\mu\text{m}$ -thick FFPE prostate tissues samples. **a** FFPE tissue from prostate cancer patient; **b** FFPE prostate tissue from non-cancer patient

cancer samples, the concentration of citric acid was down-regulated in comparison with control group. 2-amino-2-methyl-1,3-propanediol and benzamide are assumed to be microbiota origins. Moreover, benzamide is considered a diet or drug metabolite. Kyoto Encyclopedia of Genes and Genomes reports participation of benzamide in two biochemical pathways: aminobenzoid degradation and microbial metabolism in diverse environments [<http://www.genom.e.jp/kegg/>].

FFPE tissues were also applied in untargeted metabolomic studies previously. Wojakowska et al. [13] used GC–MS technique to compare metabolic profiles obtained

from formalin-fixed, fresh-frozen, and FFPE mouse kidney tissue specimens. As a result, 60 compounds were detected in FFPE samples. Recently, Cacciatore et al. [16] performed untargeted metabolomic analysis, based on ultraperformance liquid chromatography–mass spectrometry (UPLC–MS) and GC–MS techniques, to compare two types of samples: fresh-frozen and FFPE cell lines of prostate cancer, and fresh-frozen, and FFPE prostate cancer tissues. As a result, 252 metabolites were detected in FFPE cell line specimens, while 140 metabolites were identified in FFPE prostate cancer tissues. Detected compounds belonged to various chemical classes, including energy metabolites, nucleotides, lipids,



**Fig. 6** PCA model for tissue samples obtained from prostate cancer patients (red triangles) and control group (blue triangles). The calculated  $R^2$  factor was  $R^2=0.977$

**Table 3** Results of univariate statistical analysis of tissue samples obtained from prostate cancer patients and healthy volunteers

No.	Compound	RT [min]	<i>P</i> value	<i>P</i> value after FDR correction	Chemical group
1.	Heptadecanoic acid	24.22	2.40E−04	0.002	Fatty acid
2.	3-octadecanoic acid	30.86	0.016	0.030	Fatty acid
3.	2-Amino-2-methyl-1,3-propanediol	11.41	0.016	0.030	Amino alcohol
4.	Nonanoic acid	12.83	0.001	0.033	Fatty acid
5.	13-Octadecenoic acid	25.09	0.004	0.041	Fatty acid
6.	Hexanoic acid	7.51	0.004	0.041	Fatty acid
7.	Citric acid	19.93	0.016	0.049	Organic acid
8.	Succinic acid	11.98	0.009	0.049	Organic acid
9.	Hexadecanoic acid	23.02	0.017	0.049	Fatty acid
10.	Benzamide	8.74	0.009	0.049	Aromatic amide

amino acids, carbohydrates, xenobiotics, and peptides, but the number of identified metabolites within every class was between 2 and 50 times smaller when compared to fresh-frozen specimens. The number of metabolites was similar to results obtained within presented study (157 compounds before filtration). However, the results presented by Cacciatore et al. [16] were obtained with the use of two complementary techniques, whereas, in presented research, only one technique was applied. Moreover, authors did not distinguish how many compounds were detected by every applied technique. According to results, 48 metabolites were found as statistically significant, what constitutes ca. 34% of all identified metabolites. Moreover, Kelly et al. [17] applied LC–MS/MS in Single Reaction Monitoring (SRM) mode to detect 249 compounds in FFPE tissues. Based on filtration, 119 compounds were selected for further investigations. As a result, averagely 106 metabolites were identified in analyzed biological matrix. As it can be seen, application of data preprocessing reduced in around 50% the number of metabolites. In presented manuscript, data preprocessing reduced the number of metabolites from 157 to 24 compounds (ca. 85%). These 24 compounds underwent statistical analysis.

As a result, ten compounds were found as statistically significant among studied groups. Since presented research is a retrospective study, fresh-frozen samples collected from the same patients as FFPE specimens could not be obtained. However, some previously published studies covered the comparison of metabolic profiles obtained from FFPE tissues and fresh-frozen specimens. According to the literature, formalin fixation and paraffin embedding may cause loss of about 25% of detected metabolites when compared to fresh-frozen samples, especially such groups as amino acids, peptides, or carbohydrates [13, 15, 16]. Fortunately, the goal of our study was to analyze fatty compound synthesis and metabolism, which were found to be stable during FFPE sample preparation procedure.

## Conclusions

In the presented study, optimized method in reference to several factors for direct extraction of metabolites from FFPE tissue samples is proposed. Although obtained protocol was prepared based on one type of samples, it is considered as

universal and may be applied for various types of tissue specimens as well as different approaches in metabolomics (untargeted but also targeted analysis). To our knowledge, it was the first time when the method was fully optimized and applied for untargeted metabolomics in prostate cancer. Doubtlessly, paraffin-embedded tissue is an attractive, alternative matrix for metabolomics. Application of the developed and optimized sample preparation method of the FFPE tissue for metabolic fingerprinting by GC–MS confirmed its potential in CaP development investigation. As a result, ten metabolites were found as contributing the most into differentiating tested cancer and non-cancerous prostatic tissue groups. Tissue metabolites, with observed altered levels in cancer group, are involved in such biochemical pathways like TCA cycle and fatty acid metabolism. Obtained results may provide new insight into the mechanism of prostate cancer development. However, it should be underlined that they are preliminary results, derived from pilot study, and because of that require future validation on larger set of samples.

**Acknowledgements** Authors would like to thank Shimpol A. M. Borzymowski Company for the opportunity to carry out analysis with the use of their GC–MS 8030TQ System. Authors would also like to thank Renata Wawrzyniak for support in the field of analytical study and Agnieszka Borsuk-De Moor for support in the field of statistical analysis.

**Funding** The study was funded by the National Science Center (NCN, Poland), Grant No. UMO-2016/21/D/ST4/03730 received by Magdalena Buszewska-Forajta.

## Compliance with Ethical Standards

**Conflict of interest** Magdalena Buszewska-Forajta has received research grant from National Science Center, Małgorzata Patejko declares she has no conflict of interest, Szymon Macioszek declares he has no conflict of interest, Dawid Sigorski declares he has no conflict of interest, Ewa Iżycka-Świeszewska declares she has no conflict of interest, and Michał J. Markuszewski declares he has no conflict of interest.

**Ethical Approval** Described research is a retrospective study and was approved by the Independent Commission for Bioethics Research, Medical University of Gdansk, NKBBN/448/2015 and NKBBN/432/2016.

**Informed Consent** All human tissue samples used in this study were obtained after the subjects gave written informed consent.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- Johnson CH, Ivanisevic J, Siuzdak G (2016) Metabolomics: beyond biomarkers and towards mechanisms. *Nat Rev Mol Cell Biol* 17:451–459
- Andrisi L, Dudzik D, Barbas C, Milkovic L, Grune T, Zarkovic N (2018) Short overview on metabolomics approach to study pathophysiology of oxidative stress in cancer. *Redox Biol* 14:47–58
- Armitage EG, Southam AD (2016) Monitoring cancer prognosis, diagnosis and treatment efficacy using metabolomics and lipidomics. *Metabolomics* 12(9):146
- Gowda GAN, Zhang S, Gu H, Asiago V, Shanaiah N, Raftery D (2008) Metabolomics-based methods for early disease diagnostics. *Expert Rev Mol Diagn* 8(5):617–633
- Kim HK, Verporte R (2010) Sample preparation for plant metabolomics. *Phytochem Anal* 21(1):4–13
- Wolfender JL, Rudaz S, Choi YH, Kim HK (2013) Plant metabolomics: from holistic data to relevant biomarkers. *Curr Med Chem* 20(8):1056–1090
- De Vos RCH, Moco S, Lommen A, Keurentjes JJB, Bino RJ, Hall RD (2007) Untargeted large-scale plant metabolomics using liquid chromatography coupled to mass spectrometry. *Nat Protoc* 2(4):778–791
- Mashego MR, Rumbold K, De Mey M, Vandamme E, Soetaert W, Heijnen JJ (2007) Microbial metabolomics: past, present and future methodologies. *Biotechnol Lett* 29(1):1–16
- Patejko M, Jacyna J, Markuszewski MJ (2017) Sample preparation procedures utilized in microbial metabolomics: an overview. *J Chromatogr B Anal Technol Biomed Life Sci* 1043:150–157
- Xu YJ, Wang C, Ho WE, Ong CN (2014) Recent developments and applications of metabolomics in microbiological investigations. *Trends Anal Chem* 56:37–48
- Dunn WB (2008) Current trends and future requirements for the mass spectrometric investigation of microbial, mammalian and plant metabolomes. *Phys Biol* 5(1):011001
- Chetwynd AJ, Dunn WB, Rodriguez-Blanco G (2017) In: Susulini A (ed) *Metabolomics: from fundamentals to clinical applications, advances in experimental medicine and biology*. Springer, Cham
- Wojakowska A, Marczak Ł, Jelonek K, Polanski K, Widlak P, Pietrowska M (2015) An optimized method of metabolite extraction from formalin-fixed paraffin-embedded tissue for GC/MS analysis. *PLoS One* 10(9):e0136902
- Wojakowska A, Chekan M, Marczak Ł, Polanski K, Lange D, Pietrowska M, Widlak P (2015) Detection of metabolites discriminating subtypes of thyroid cancer: molecular profiling of FFPE samples using the GC/MS approach. *Mol Cell Endocrinol* 417:149–157
- Buck A, Ly A, Balluff B, Sun N, Gorzalka K, Feuchtinger A, Janssen KP, Kuppen PJK, Van De Velde CJH, Weirich G, Erlmeier F, Langer R, Aubele M, Zitzelsberger H, Aichler M, Walch A (2015) High-resolution MALDI-FT-ICR MS imaging for the analysis of metabolites from formalin-fixed, paraffin-embedded clinical tissue samples. *J Pathol* 237(1):123–132
- Cacciatore S, Zadra G, Bango C, Penney KL, Tyekucheva S, Yanes O, Loda M (2017) Metabolic profiling in formalin-fixed and paraffin-embedded prostate cancer tissues. *Mol Cancer Res* 15(4):439–447
- Kelly AD, Breitkopf SB, Yuan M, Goldsmith J, Spentzos D, Asara JM (2011) Metabolomic profiling from formalin-fixed, paraffin-embedded tumor tissue using targeted LC/MS/MS: application in sarcoma. *PLoS One* 6(10):e25357

18. Donczo B, Guttman A (2018) Biomedical analysis of formalin-fixed, paraffin-embedded tissue samples: the holy grail for molecular diagnostics. *J Pharm Biomed Anal* 155:125–134
19. Thavarajah R, Mudimbaimannar VK, Elizabeth J, Rao UK, Ranganathan K (2012) Chemical and physical basics of routine formaldehyde fixation. *J Oral Maxillofac Pathol* 16(3):400–405
20. Huang BQ, Yeung EC (2015) In: Yeung ECT, Stasolla C, Sumner MJ, Huang BQ (eds) *Plant microtechniques and protocols*. Springer, Cham
21. Canene-Adams K (2013) In: Abelson JN, Simon MI (eds) *Methods in enzymology*. Elsevier, San Diego
22. Stasolla C, Yeung EC (2015) In: Yeung ECT, Stasolla C, Sumner MJ, Huang BQ (eds) *Plant microtechniques and protocols*. Springer, Cham
23. Richter S, Peitzsch M, Rapizzi E, Lenders JW, Qin N, De Cubas AA, Schiavi F, Rao JU, Beuschlein F, Quinkler M, Timmers HJ, Opocher G, Mannelli M, Pacak K, Robledo M, Eisenhofer G (2014) Krebs cycle metabolite profiling for identification and stratification of pheochromocytomas/paragangliomas due to succinate dehydrogenase deficiency. *J Clin Endocrinol Metab* 99(10):3903–3911
24. Ly A, Buck A, Balluff B, Sun N, Gorzolka K, Feuchtinger A, Janssen KP, Kuppen PJK, Van De Velde CJH, Weirich G, Erlmeier F, Langer R, Aubele M, Zitzelsberger H, McDonnell L, Aichler M, Walch A (2016) High-mass-resolution MALDI mass spectrometry imaging of metabolites from formalin-fixed paraffin-embedded tissue. *Nat Protoc* 11(8):1428–1443
25. Carturan E, Tester DJ, Brost BC, Basso C, Thiene G, Ackerman MJ (2008) Postmortem genetic testing for conventional autopsy-negative sudden unexplained death: an evaluation of different DNA extraction protocols and the feasibility of mutational analysis from archival paraffin-embedded heart tissue. *Am J Clin Pathol* 129(3):391–397
26. Sengüven B, Baris E, Oygur T, Berktaş M (2014) Comparison of methods for the extraction of DNA from formalin-fixed, paraffin-embedded archival tissues. *Int J Med Sci* 11(5):494–499
27. Sato Y, Sugie R, Tsuchiya B, Kameya T, Natori M, Mukai K (2001) Comparison of the DNA extraction methods for polymerase chain reaction amplification from formalin-fixed and paraffin-embedded tissues. *Diagn Mol Pathol* 10(4):265–271
28. Kalantari N, Bayani M, Ghaffari T (2016) Deparaffinization of formalin-fixed paraffin-embedded tissue blocks using hot water instead of xylene. *Anal Biochem* 507:71–73
29. Yuan M, Breitkopf SB, Yang X, Asara JM (2012) A positive/negative ion-switching, targeted mass spectrometry-based metabolomics platform for bodily fluids, cells, and fresh and fixed tissue. *Nat Protoc* 7(5):872–881
30. Patel PG, Selvarajah S, Boursalie S, How NE, Ejdelman J, Guerard KP, Bartlett JM, Lapointe J, Park PC, Okello JBA, Berman DM (2016) Preparation of formalin-fixed paraffin-embedded tissue cores for both RNA and DNA extraction. *J Vis Exp* 114:e54299
31. Urban C, Buck A, Siveke JT, Lordick F, Luber B, Walch A, Aichler M (1862) PAXgene fixation enables comprehensive metabolomic and proteomic analyses of tissue specimens by MALDI MSI. *Biochim Biophys Acta* 1:51–60
32. Kokkat TJ, Patel MS, McGarvey D, Livolsi VA, Baloch ZW (2013) Archived formalin-fixed paraffin-embedded (FFPE) blocks: a valuable underexploited resource for extraction of DNA, RNA, and protein. *Biopreserv Biobank* 11(2):101–106
33. Schwarzfischer P, Reinders J, Dettmer K, Kleo K, Dimitrova L, Hummel M, Feist M, Kube D, Szczepanowski M, Klapper W, Taruttis F, Engelmann JC, Spang R, Gronwald W, Oefner PJ (2017) Comprehensive metaboloproteomics of burkitt's and diffuse large b-cell lymphoma cell lines and primary tumor tissues reveals distinct differences in pyruvate content and metabolism. *J Proteom Res* 16(3):1105–1120
34. Zhong H, Liu Y, Talmor M, Wu B, Hui P (2013) Deparaffinization and lysis by hydrothermal pressure (pressure cooking) coupled with chaotropic salt column purification: a rapid and efficient method of DNA extraction from formalin-fixed paraffin-embedded tissue. *Diagn Mol Pathol* 22(1):52–58
35. R Core Team (2016) A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. [www.r-project.org/](http://www.r-project.org/)
36. Fiehn O (2016) Metabolomics by gas chromatography-mass spectrometry: combined targeted and untargeted profiling. *Curr Protoc Mol Biol* 114:30.4.1–30.4.32
37. Matsuo T, Tsugawa H, Miyagawa H, Fukusaki E (2017) Integrated strategy for unknown EI-MS identification using quality control calibration curve, multivariate analysis, EI-MS spectral database, and retention index prediction. *Anal Chem* 12:6766–6773

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.