



Validated comprehensive metabolomics and lipidomics analysis of colon tissue and cell lines

Caroline Rombouts^{a, b, c}, Margot De Spiegeleer^a, Lieven Van Meulebroek^a, Winnok H. De Vos^{b, c}, Lynn Vanhaecke^{a, d, *}

^a Ghent University, Faculty of Veterinary Medicine, Department of Veterinary Public Health and Food Safety, Laboratory of Chemical Analysis, Salisburylaan 133, B-9820, Merelbeke, Belgium

^b Ghent University, Faculty of Bioscience Engineering, Department of Molecular Biotechnology, Cell Systems & Imaging, Coupure Links 653, 9000, Ghent, Belgium

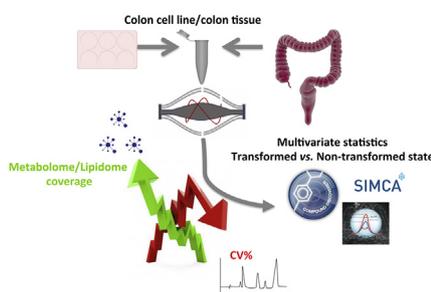
^c Antwerp University, Faculty of Veterinary Medicine, Department of Veterinary Sciences, Laboratory of Cell Biology & Histology, Universiteitsplein 1, 2610, Wilrijk, Belgium

^d Institute for Global Food Security, School of Biological Sciences, Queen's University, Belfast, Northern Ireland, United Kingdom

HIGHLIGHTS

- A validated Q-Orbitrap-HRMS based polar metabolomics and lipidomics method is presented for colon cell lines and tissue.
- Both selected targeted analyte's peak areas and number of components (= metabolome coverage) were assessed.
- More than 15,000 components with CV < 30% were detected in human colon tissue and cell lines.
- Discriminative metabolic profiles were obtained between the non-transformed and transformed state in both matrices.
- This method can be of paramount importance in gaining new insights in gastro-intestinal diseases.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:

Received 26 November 2018

Received in revised form

4 March 2019

Accepted 5 March 2019

Available online 12 March 2019

Keywords:

Cancer

Colon cell lines

Colon tissue

ABSTRACT

Current untargeted approaches for metabolic fingerprinting of colon tissue and cell lines lack validation of reproducibility and/or focus on a selection of metabolites as opposed to the entire metabolome. Yet, both are critical to ensure reliable results and pursue a fully holistic analysis. Therefore, we have optimized and validated a platform for analyzing the polar metabolome and lipidome of colon-derived cell and tissue samples based on a consecutive extraction of polar and apolar components. Peak areas of selected targeted analytes and the number of untargeted components were assessed. Analysis was performed using ultra-high performance liquid-chromatography (UHPLC) coupled to hybrid quadrupole-Orbitrap high-resolution mass spectrometry (HRMS). This resulted in an optimized extraction protocol using 50% methanol/ultrapure water to obtain the polar fraction followed by a dichloromethane-based lipid extraction. Using this comprehensive approach, we have detected more than 15,000 components

* Corresponding author. Ghent University, Faculty of Veterinary Medicine, Department of Veterinary Public Health and Food Safety, Laboratory of Chemical Analysis, Salisburylaan 133, B-9820, Merelbeke, Belgium.

E-mail address: lynn.vanhaecke@ugent.be (L. Vanhaecke).

Q-orbitrap-high resolution mass spectrometry
Comprehensive metabolomics/lipidomics profiling

with CV < 30% in internal quality control (IQC) samples and were able to discriminate the non-transformed (NT) and transformed (T) state in human colon tissue and cell lines based on validated OPLS-DA models ($R^2Y > 0.719$ and $Q^2 > 0.674$). To conclude, our validated polar metabolomics and lipidomics fingerprinting approach could be of great value to reveal gastrointestinal disease-associated biomarkers and mechanisms.

© 2019 Elsevier B.V. All rights reserved.

1. Introduction

Colorectal cancer (CRC) is at present the second and third most common cancer for women and men, respectively. Furthermore, this disease is ranked on the fourth place of deadliest cancers, only preceded by lung, liver and stomach cancer [1]. Research into the molecular pathways underlying cancer development and progression, revealed a strong relationship with mutations in the adenomatous polyposis (APC) oncogene, K-ras oncogene and the TP53 tumor suppressor gene [2]. Next to this, a number of important metabolic alterations have been recognized to be associated with cancer, such as the “Warburg effect”, which is defined as an enhanced anaerobic glycolysis and high production of lactate [3]. For a long time, it was even assumed that metabolic reprogramming was a consequence of genetic mutations to comply with the higher energy requirement of cancer cells. However, new findings suggest that an altered metabolic state itself could promote cell proliferation [4]. Moreover, besides the association of certain metabolites (e.g. hydroxybutyrate and lactate) with higher growth rates of cancer cells, some have also been correlated with evasion from apoptosis [5]. Therefore, exploring metabolic perturbations in cells offers potential for devising novel prophylactic and therapeutic strategies in cancer [4,5].

Colon-derived cell lines are of paramount importance to gather basic fundamental, translational and clinical information that could improve the diagnosis and prognosis of CRC [6]. They are widely accessible, easy to culture, and represent a fairly homogeneous cell population with reproducible responses to targeted perturbations [7,8]. These characteristics especially apply for human cancer-derived cell lines as opposed to cell lines obtained from normal, healthy tissue that are more difficult to purchase and to maintain in culture. Therefore, human cancer-derived cell lines are not only used to study cancer-associated mechanisms, but also to investigate physiological processes occurring in normal tissue [6].

Untargeted metabolomics studies of different sample matrices have proven to be of great value as a research tool, especially in cancer research, to study changes in the metabolome that is the most accurate reflection of the cellular phenotype [9,10]. Tissue is the principal location where pathological processes occur and can therefore reveal the most interesting changes related to cancer [11]. Nevertheless, sampling and extraction remains more challenging for tissue and cell lines as opposed to biofluids [12]. For example, in cell lines cell detachment from the bottom of well plates and quenching cellular metabolism at time of harvesting are critical points to ensure a robust analysis and special care must be taken to avoid metabolite leakage. In this context, several studies have demonstrated that cell scraping is preferable above conventional trypsinization, since the latter is time consuming and requires more wash and centrifugation steps [13–15]. In addition, tissue is often composed of heterogeneous regions, e.g. presence of different cell types or zones around tumors that are well-oxygenated or necrotic [16]. These factors may introduce additional biological variation, indicating the need for optimization, validation and standardization of extraction parameters specifically for cell lines and tissue

[16,17]. Next to this, due to the broad physico-chemical range of metabolites (polar and apolar) present in biological matrices, it is not possible to cover the metabolome with a single analytical platform. Nevertheless, as human tissue material is usually obtained in small quantities, a simultaneous extraction of both polar and apolar (i.e. lipids) metabolites from one sample is preferred over parallel analysis. In this context, consecutive extraction of the polar fraction followed by the apolar fraction (i.e. two-step extraction) has proven to be superior to the simultaneous (i.e. one-step) extraction of both polarities by using a biphasic solvent mixture in both esophageal and liver tissue [12,18].

An optimized and validated two-step extraction protocol that allows for the comprehensive coverage of a broad range of polar and apolar components and ensures robust untargeted metabolomics studies for colon tissue and cell lines has not been described earlier. Therefore, we have developed and validated such a protocol, whereby extracts were analyzed using a polar metabolomics and lipidomics UHPLC-HRMS methodology that was recently developed for feces, plasma and urine [19–21]. In this work, we have first fine-tuned the extraction parameters of both the polar metabolome and the lipidome in targeted and untargeted mode. Next, as a proof-of-concept, our optimized and validated method was applied to obtain discriminating fingerprints between the NT and T state of colon cell lines and human colon tissue biopsies.

2. Materials and methods

2.1. Biological samples

2.1.1. Cell lines

The human transformed (T) colorectal cell lines HT-29, Caco-2, HCT-116, SW480 and SW948 and the non-transformed (NT), immortalized colon cell lines FHC and CCD 841 CON were obtained from ATCC, Manassas, VA, U.S.A. The HT-29 cell line was used for optimization and validation of the cellular metabolomics and lipidomics methods, since it is the most commonly used cell line in cancer research [22]. Dulbecco's Modified Eagle's Medium (DMEM), supplemented with 10% Fetal Bovine Serum (FBS) and 1% penicillin/streptomycin (P/S), was used to culture the cancer cell lines and CCD 841 CON in a humidified incubator at 37 °C and 5% CO₂/95% air. FHC cells were cultured in DMEM:F12 supplemented with 10 mM HEPES, 0.005 mg mL⁻¹ insulin, 0.005 mg mL⁻¹ transferrin, 100 ng mL⁻¹ hydrocortisone, 10% FBS and 1% P/S. All cell reagents were purchased from Life Technologies (Ghent, Belgium). Twenty-four hours before extraction cells were seeded in 6-well plates, making sure that on the day of extraction ±80% confluency was reached. When the metabolomics/lipidomics protocol was applied to compare the NT with the T state, each cell line was seeded in quintuplicate 24 h before extraction (= technical replicates) and this experiment was repeated three times with 1-week interval (= cell line batches).

2.1.2. Colon tissue

For optimization and validation experiments, colon tissue samples were obtained from a local pig slaughterhouse. The colon

was washed with Phosphate Buffered Saline (PBS) to remove fecal content and mucus. Small pieces of colon tissue (± 3 g) were dissected, aliquoted in 15 mL falcon tubes and immediately snap frozen in liquid nitrogen. The metabolic fingerprinting of the T and the NT state *in vivo* was performed on cancerous colon material and the corresponding healthy tissue of 10 individuals, provided by Biobank@UZA (Antwerp, Belgium; ID: BE71030031000); Belgian Virtual Tumorbank funded by the National Cancer Plan. The use of human colon tissue samples in this study was ethically approved by the University Hospital of Antwerp (ECD 16/37/368 (Antwerp, Belgium)).

2.2. UHPLC-Q-orbitrap-UHRMS

2.2.1. Reagents and chemicals

Analytical standards were purchased from Sigma-Aldrich (St-Louis, Missouri, USA), ICN Biomedicals Inc. (Ohio, USA), TLC Pharmchem (Vaughan, Ontario, Canada) or Cambridge Isotope Laboratories Inc. (Tewksbury, Massachusetts, USA). More detailed information about chromatographic and mass spectrometric features of the reference polar and lipid compounds can be consulted in De Paepe et al. (2018) and Van Meulebroek et al. (2018), respectively [20,21]. Stock solutions of standards were made in appropriate solvents (methanol or ultrapure water) according to polarity at a concentration of 5–10 mg mL⁻¹ (polar analytes) or 1 mg mL⁻¹ (apolar analytes). Solvents were of LC-MS grade for extraction purposes and obtained from Fisher Scientific (Loughborough, UK) and VWR International (Merck, Darmstadt, Germany). Ultrapure water was obtained by usage of a purified-water system (VWR International, Merck, Darmstadt, Germany).

2.2.2. Polar metabolomics method

The chromatographic separation was performed with an Ultimate 3000 XRS UHPLC system (Thermo Fisher Scientific, San José, CA, USA). An Acquity HSS T3 C18 column (1.8 μ m, 150 mm \times 2.1 mm, Waters), kept at 45 °C, was used. A binary solvent system with ultrapure water and acetonitrile, both acidified with 0.1% formic acid, was established, accompanied by a gradient program with a flow rate of 400 μ L and a duration time of 18 min according to De Paepe et al. (2018) [21].

A Q-Exactive™ stand-alone bench top Orbitrap mass spectrometer (MS) (Thermo Fisher Scientific, San José, CA, USA), equipped with a heated electrospray ionization source (HESI II) operating in polarity switching mode, was used for analysis as described earlier [21].

2.2.3. Lipidomics method

LC was achieved with an Ultimate 3000 XRS UHPLC system (Thermo Fisher Scientific, San José, CA, USA). An Acquity UPLC BEH Phenyl column (1.7 μ m, 150 mm \times 2.1 mm, Waters), maintained at 40 °C, in combination with a Hypersil GOLD pre-column (1.9 μ m, 50 mm \times 2.1 mm, Thermo Fisher Scientific, San José, CA, USA) were used. The binary gradient program based on ultrapure water and methanol, both acidified with 3.5 mM NH₄Ac, was the same as described previously [20]. The MS instrumentation used for lipidomics analysis was the same as described for polar metabolomics analysis. However, operating settings were adjusted according to Van Meulebroek et al. (2018) [20].

2.3. Polar metabolomics: optimization of sample extraction

Thirty-one metabolites, belonging to 12 different chemical classes (i.e. polyols, hydroxylic acids, multicarboxylic acids, imidazoles, carbohydrates, amino acids, amines, amides, other N-compounds, fatty acids, sulfonic acids and phosphates) with a broad

range of physicochemical properties, were selected for statistical evaluation during chemometric optimization of the HT-29 cell line (Table S1) and colon tissue (Table S2) extraction. These compounds were chosen based on their endogenous presence in the samples under investigation and their identity was confirmed based on *m/z*-value, retention time and ¹³C/¹²C isotope ratio of the corresponding in-house analytical standard.

2.3.1. HT-29 cell line

Optimization parameters were adapted from a general extraction protocol for cell lines, as described by Sapcariu et al. (2014) [23]. Different types of extraction solvent (ethanol, methanol and acetonitrile), the ratio organic solvent vs. water (1:1 or 7:2, v/v), washing buffer (PBS or 0.9% NaCl, 1x washing or 2 x washing) and drying of the cell extracts (flushing with N₂ or vacuum-drying) were assessed by a D-optimal statistical design (n = 29), which determines the optimal set of factors, i.e. extraction parameters, to be tested, in Modde 5.0. Finally, a response surface methodology (RSM) model (n = 18) was created to further determine the optimal values for quantitative variables, i.e. ratio extraction solvent/water (3:1, 2:1, 1:1, 1:2, v/v) and dilution volume of cell extract after vacuum drying (125, 150, 500, 750 and 1000 μ L).

2.3.2. Colon pig tissue

In a first experiment (full factorial design, n = 19), tissue mass (25 or 100 mg), organic solvent methanol/ultrapure water volume (50:50, v/v) (500 μ L or 1 mL), homogenization technique (manually with forceps and scalpel or TissueLyzer) and centrifugation time (5 or 15 min) were optimized. Additionally, an RSM model was used to further optimize tissue mass (50, 75 and 100 mg), solvent volume (350, 500 and 650 μ L) and centrifugation time (5, 7.5 and 10 min) (n = 19).

2.4. Lipidomics: optimization of sample extraction

For the development of the lipidomics method, 28 reference lipids from 6 different lipid classes (fatty acyls, glycerolipids, glycerophospholipids, sphingolipids, steroids, and prenol lipids) were chosen based on the same criteria as described for the polar metabolomics method (Table S3, Table S4). According to The International Lipid Classification and Nomenclature Committee (ILCNC) there are 8 primary lipid classes, but saccharolipids and polyketides represent only a small proportion of the lipidome (1%). Both the aforementioned lipid classes were not detected in the HT-29 cell line, neither in the colonic tissue [24]. The optimization parameters as discussed below were chosen based on literature, using a modified Folch procedure [25–27].

2.4.1. HT-29 cell line

Dichloromethane, chloroform, methyl-tert-butyl-ether and butanol were assessed in different volumes (1–3 mL). Also, vortex time (30–120 s), centrifugation time (5–20 min) and speed (368–9223 \times g) were tested in a D-optimal statistical model (n = 19). Additionally, solvent (2, 3.5 and 5 mL) and injection volume (1, 4.5 and 8 μ L) were further optimized using an RSM model (n = 11).

2.4.2. Colon pig tissue

Vortex time (30–120 s), centrifugation time (5–20 min) and speed (1845–9223 \times g) were tested using a D-optimal statistical model (n = 11). Furthermore, an RSM model (n = 17) was created to further optimize vortex time (90, 120 and 150 s), centrifugation speed (184, 3228 and 5534 \times g) and volume of resuspension after drying with N₂ (500, 750 and 1000 μ L).

2.5. Validation of the polar metabolomics and lipidomics methods

For validation purposes, the endogenously present analytes evaluated in both sample matrices were evaluated as described for optimization (Tables S5–S8). First, a dilution series (500-fold, 200-fold, 100-fold, 50-fold, 20-fold, 10-fold, 5-fold, 2-fold dilution and undiluted) of the extracted sample was made in mobile phase starting conditions to assess linearity in terms of coefficient of determination (R^2). Based on these results, the dilution volume of the sample matrix was adjusted and thereby assisted in the optimization process. Additionally, instrumental (injecting the same sample 10 times), intra-assay and inter-assay (extraction and subsequent analysis of 10 samples in parallel and of 2 series of 10 samples performed by two technicians at different days, respectively) precision were evaluated based on the coefficients of variance (CV) of the selected targeted analytes and untargeted components.

2.6. Data analysis

Targeted data processing was carried out with Xcalibur 3.0 software (Thermo Fischer Scientific, San José, CA, USA). For the optimization experiments, peak areas of the targeted analytes or the number of detected components in untargeted mode were imported as responses in Modde 5.0 (Umetrics AB, Umea, Sweden) to determine the most optimal extraction parameters. Components were extracted from the full-scan data with Sieve™ 2.2 (Thermo Fischer Scientific, San José, CA, USA) or the more recent available Compound Discoverer 2.1 software (Thermo Fisher Scientific, San José, CA, USA), in which chromatographic peak alignment and retrieving of components was performed. To this extent, the minimum peak intensity was set at 500,000 au, minimum signal-to-noise ratio at 10, minimum number of isotopes at 2, retention time window at 0.75 min, maximum retention time shift at 0.25 min and maximum mass shift (Δ ppm) at 6 ppm. In addition, blanks (i.e. acetonitrile for polar metabolomics and methanol for lipidomics) were used for subtraction of noise peaks in the samples of interest. Compound Discoverer offers the interesting capacity, in comparison to similar software such as Sieve, to handle both positive and negative ionization modes together in one analysis. Therefore, the obtained data were analyzed using the indicated software for the proof-of-concept metabolomics experiment. However, this was not possible for the lipidomics experiment due to the presence of different scan ranges (67–1000 Da and 1000–2300 Da) in the method and hence for the latter separate datasets for both ionization modes were obtained. Normalization of the amount of analyzed cell/tissue material between samples was performed by dividing the abundance of each component by the total ion count (TIC) of the respective sample. Indeed, Liu et al. (2014) have demonstrated that over 4 orders of magnitude, TIC normalized intensities accurately reflect intracellular metabolite levels [28].

For the proof-of-concept experiment, cellular extracts obtained from three different passages (1-week interval) were analyzed immediately after harvesting and all cellular LC-MS data together with tissue LC-MS data were processed simultaneously in Compound Discoverer or Sieve for the polar metabolomics and lipidomics experiment, respectively. Components with CVs > 30% in IQC samples, i.e. pools of samples that were run in duplicate after each batch of 10 samples, were not stable during instrumental analysis and therefore removed from the dataset [29,30]. Finally, the relative abundance of each component was divided by the mean relative abundance of the two following IQC samples to correct for potential instrumental drift [19]. Tissue components were further evaluated using multilevel data analysis in Matlab 9.3,

whereby effect of tissue origin for each individual was separated from between individual variation. Next, unsupervised hierarchical clustering (www.metaboanalyst.ca) and multivariate statistical analysis (SIMCA 14.1 software (Umetrics AB, Umea, Sweden)) were performed, whereby PCA-X models were created to evaluate natural clustering of samples and to check for outliers whereas OPLS-DA models were built to discriminate between the NT versus T state in tissue and cell lines. For modeling, data were log-transformed and Pareto-scaling was applied [31]. Validity of the obtained OPLS-DA models was evaluated based on R^2Y (> 0.5), Q^2 (> 0.5), CV-Anova ($p < 0.05$) permutations testing ($n = 100$) and VIP values (> 1.0) [32].

2.7. Polar metabolomics: optimized extraction protocol

2.7.1. HT-29 cell line

HT-29 cells were seeded in 6-well plates 24 h before extraction. At the day of extraction, medium was aspirated and cells were washed with 0.9% NaCl. Next, 1 mL 50% methanol/ultrapure water (on ice) and 27 μ L working solution containing the internal standard (25 ng μ L⁻¹ of D-valine-d₈) were added. This was followed by scraping and transferring the cells to an Eppendorf tube. Cells were lysed by sonification (2 \times 15 s, Soniprep 150, Beun-De Ronde, LA Abcoude, The Netherlands) and cell debris was removed by centrifugation (16,200 \times g, 5 min, 4 °C). Next, a 100 μ L aliquot of supernatant was vacuum-dried, whilst the remaining cell suspension was stored at -80 °C for lipidomics analysis. The dried cell pellets were resolved in 1.5 mL of a solvent mixture at HPLC starting conditions (0.1% formic acid in UP water and 0.1% formic acid in acetonitrile in 98/2 ratio, respectively) and a subtraction was transferred to a glass HPLC-vial and 10 μ L was injected.

2.7.2. Colon pig tissue

100 mg colonic tissue was weighed and 400 μ L 50% methanol/ultrapure water (on ice) and 30 μ L internal standard (8 ng μ L⁻¹ D-valine-d₈) were added. The tissue was manually disintegrated with scalpel and forceps, followed by vortexing (10 s) and centrifugation (21,161 \times g, 7.5 min, 4 °C). Afterwards, 100 μ L supernatant was transferred to an Eppendorf tube and vacuum-dried. The remaining tissue suspension was stored at -80 °C to perform lipidomics. Finally, the tissue pellets were resolved in 2 mL solvent mixture at mobile phase starting conditions and transferred to a glass HPLC-vial and 10 μ L was injected.

2.8. Lipidomics: optimized extraction protocol

2.8.1. HT-29 cell line

The ISTDs (60 μ L palmitic acid-d₃₁ (25 ng μ L⁻¹) and 30 μ L phosphocholine-d₅₄ (25 ng μ L⁻¹)), 900 μ L cell suspension, 400 μ L ultrapure water (4% trichloric acid) and 500 μ L MeOH were added to 3 mL DCM supplemented with 0.01% butylhydroxytoluene (BHT). This mixture was vortexed (30 s), incubated (20 min, 20 °C) and centrifuged (9223 \times g, 5 min, 20 °C). Next, 2 mL DCM phase was collected with a glass Pasteur pipette and dried at 30 °C under N₂. The pellet was resuspended in 300 μ L CHCl₃ and vortexed. Finally, 300 μ L MeOH was added and the extract was transferred to a glass HPLC vial and 10 μ L was injected.

2.8.2. Colon pig tissue

First, 600 μ L 50% MeOH/UP H₂O was added to 300 μ L tissue suspension. Next, the same protocol as for the HT-29 cell line was performed, with slight adaptations of the centrifugation speed (368 \times g) and resuspension volume (350 μ L CHCl₃ + 350 μ L MeOH) after drying with N₂ and finally, 6 μ L was injected in the system.

3. Results and discussion

3.1. Polar metabolomics: optimization and validation

3.1.1. HT-29 cell line

First, different parameters (including washing of cells, type of organic solvent, ratio organic solvent/ultrapure water and drying of extracts) for extraction of the polar metabolome of the HT-29 cell line were assessed and the most optimal conditions were included in the final protocol. In this context, significant effects were observed for the washing step, the type of extraction solvent and the ratio organic solvent/ultrapure water towards the extraction yield for 14 out of the 22 selected endogenous metabolites, of which standards were present in our in-house database (Table 1). Nonetheless, untargeted analysis in Sieve™ 2.2 revealed no important differences in terms of number of detected components between the different extraction conditions (Table 1). Washing with PBS, a washing solvent frequently used in untargeted metabolomics studies [33,34], did not improve nor deteriorate the extraction efficiency in comparison to rinsing with 0.9% NaCl solution although putrescine significantly benefitted from washing with the latter one (Table 1, Table S1). To this end, we opted to wash with NaCl, supported by Sapcariu et al. (2014) wherein the latter is recommended, since large phosphate peaks originating from PBS could contaminate the obtained MS spectra [23]. Omitting the washing step was not considered in our experimental set-up, as rinsing is necessary to eliminate residual media containing extracellular metabolites and hence avoiding a biased higher metabolome coverage [35]. Due to the fact that no differences in extraction yield could be observed between washing a single time and two times (Table 1), it could be assumed that one washing step sufficed to remove unwanted media constituents. Also, one rinsing step shortens the entire procedure reducing the risk for variability in extraction. Of the tested organic solvents (added to ultrapure water), i.e. ethanol, methanol and acetonitrile, no remarkable or significant differences could be observed in the number of components obtained following untargeted analysis (Fig. 1A, Table 1). However, extraction with acetonitrile resulted in inconsistent results during targeted analysis, i.e. significantly higher or lower peak intensities for 2 and 3 out of the 22 selected targeted endogenous metabolites, respectively (Table 1). As a result, and because methanol is the most commonly used extraction solvent in metabolomics studies in literature [23,33,36,37], this solvent was retained for further experiments. The ratio organic solvent/ultrapure water did not affect the extraction yield, i.e. number of components in untargeted analysis, but remarkably, 13 out of 22 selected targeted metabolites shared higher responses when the 1:1 ratio was used

(Table 1). A recent study has shown that an increase in ratio organic solvent/ultrapure water results in a decreased response of compounds with a log P lower than 5 [38]. This is in concordance with our untargeted results, whereby components with a lower retention time, and thus lower log P (i.e. higher polarity), were more abundant following extraction with the 1:1 as opposed the 7:2 ratio (Fig. 2). Finally, since no effect of drying could be observed and because of practical reasons, vacuum drying was preferred to drying with N₂ (Table 1). RSM methodology models were constructed to assess the optimal ratio methanol/ultrapure water (3:1–1:2) (v/v) and the dilution volume after vacuum-drying (125–1000 μL). A lower solvent ratio together with a lower dilution volume after vacuum drying improved the extraction yield during untargeted analysis and the peak areas of 20 out of 22 selected endogenous metabolites. Therefore, the ratio methanol/ultrapure water was set at 1:1 (v/v) and a dilution volume of 150 μL was used to efficiently extract most cellular metabolites.

Prior to validation, a nine-point dilution series (500-fold – not diluted sample) was established to determine the optimal linear range for the detection of the selected endogenous targeted metabolites and for the metabolome as a whole. To this extent, 150 μL of supernatant was vacuum dried and the cell pellet was resuspended in the same volume, thus representing the non-diluted sample. A minimum of five successive dilution levels were included to construct calibration curves that obtained the highest coefficient of determination (R²). With respect to the targeted assessment, 19 out of the 22 selected metabolites were present until the 500-fold dilution. For 14 selected metabolites, saturation of the MS instrument occurred, mostly at 20-fold dilution or less. An acceptable (R² ≥ 0.950) linear dynamic range was achieved for 20 out of 22 metabolites (Table 1). Additionally, untargeted components present in the dilution series were extracted during data processing in Sieve 2.2. Next, the R² of each component in every dilution range (n = 15), containing a minimum of five successive points, was determined and the dilution range that contained the highest number of components with R² > 0.900 was considered as optimal. As such, the best linearity could be observed for dilution range 1/100–1/5 (77.89% of the components with R² > 0.900) (Table 1). Furthermore, the average percentage of the total number of detected untargeted components was still 98.72% for the 20-fold dilution in comparison to the non-diluted sample and was therefore applied in the final extraction protocol, ensuring minimal loss of compounds. Nevertheless, in the validation experiments and the final protocol we decided to dry 100 μL instead of 150 μL of supernatant to reduce time and converted the dilution factor accordingly.

Finally, the method's precision was assessed in terms of instrumental, intra-assay and inter-day repeatability. Since there

Table 1
Optimization and validation results of the HT-29 cell line polar metabolomics method.

Optimization	Targeted evaluation: No. of selected metabolites (n = 22)	Untargeted evaluation: Metabolome coverage (n = 776)
Washing solution (0.9% NaCl/PBS)	1 pos. affected with 0.9% NaCl	No effect
Number of wash steps (1x/2x)	0 affected	No effect
Solvent ethanol	1 pos. affected	No effect
Solvent methanol	0 affected	No effect
Solvent acetonitrile	2 pos. and 3 neg. affected	No effect
Drying (N ₂ /Gyrovap)	0 affected	No effect
Ratio solvent/water (1:1/7:2, v/v)	13 pos. affected with ratio 1:1	No effect
Validation	Targeted evaluation: No. of selected metabolites (n = 22)	Untargeted evaluation: Metabolome coverage (n = 746)
Linearity (R ²)	20 (R ² > 0.950)	77.89% (R ² > 0.900)
Instrumental precision (CV%)	20 (CV < 15%)	80.00% (CV < 15%)
Intra-assay precision (CV%)	19 (CV < 20%)	82.61% (CV < 30%)
Inter-day precision (CV%)	18 (CV < 20%)	75.44% (CV < 30%)

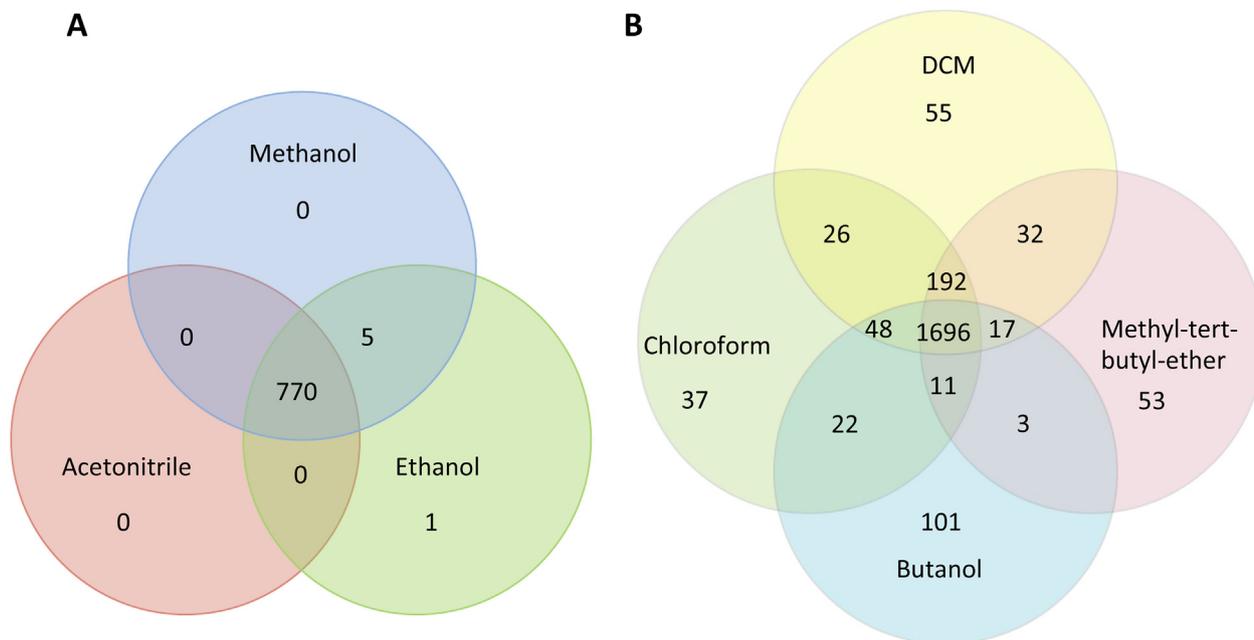


Fig. 1. Venn diagram representing the number of detected components after extraction with different organic solvents during optimization of the polar metabolomics method (A) and the lipidomics method (B) for the HT-29 cell line.

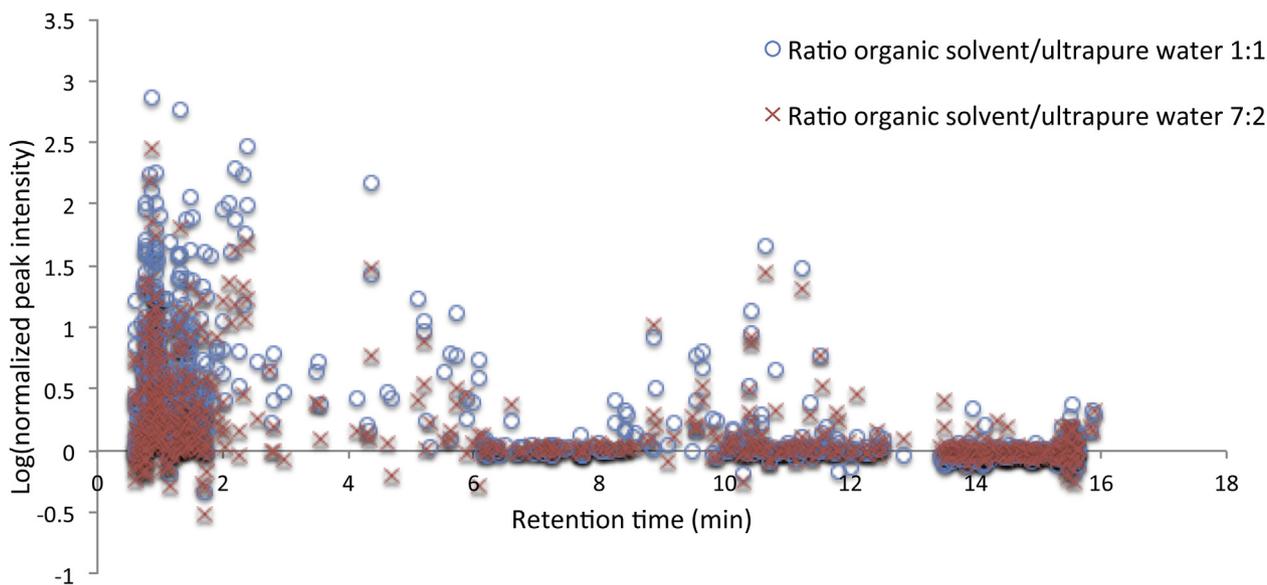


Fig. 2. Scatter plot representing the results of the untargeted analysis after metabolomics extraction of the HT-29 cell line with different ratios of organic solvent/ultrapure water (1:1 or 7:2, v/v), whereby each point represent the log transformed normalized peak intensity of a component with a specific retention time. Normalization was performed by dividing the peak area (expressed by arbitrary units) of a specific component in a sample by the mean peak area of that component in the next two IQC samples that were run after each batch of ten samples.

are no general guidelines for the validation of untargeted extraction protocols in metabolomics, FDA recommendations were followed, whereby a CV less than 15% is considered excellent and a CV less than 20% is still acceptable for values close to the limit of quantification [19,39]. However, for the validation of an untargeted method, an upper limit of 30% can be set [40–43]. The instrumental

precision was excellent ($\leq 15\%$) for 20 out of 22 endogenously selected metabolites (Table 1). The CVs for intra-assay precision ($n = 10$) and inter-day repeatability ($n = 20$) were less than 20% for 19 and 18 out of 22 metabolites, respectively (Table 1). The precision for the amines was poor (CVs $> 39\%$) and can be explained by their low abundances close to the limit of detection in the cell line

extract (Table 1, Table S5). However, in total 80.00% of the untargeted components demonstrated a CV below 15% for instrumental variation, while 82.61% and 75.44% of the components obtained a CV below 30% for intra- and inter-day precision, respectively (Table 1).

3.1.2. Colon pig tissue

Different conditions for extraction of colonic pig tissue were assessed, including tissue mass, organic solvent volume, homogenization technique and centrifugation time, whereby only for the latter no significant effects on extraction yield could be observed (Table 2). A higher tissue mass, lower solvent volume and manual homogenization showed the best results for the extraction of 23 out of 29 of the selected compounds (Table 2). A possible reason for the inadequate performance of the TissueLyzer as opposed to the manual homogenization technique could lay in the higher heat formation leading to disintegration of metabolites caused by the TissueLyzer. However, it could be assumed that manual tissue disruption would be less effective than the TissueLyzer. Moreover, as Brown et al. (2012) demonstrated that extracting solid tissue with an aqueous methanol solution with or without sample destruction generated similar results, it can be concluded that the addition of methanol as such suffices and comprises a more defining factor than the mechanical disruption method itself for extracting metabolites. Indeed, in the study by Brown et al. (2012), more than 92% of the components detected in the tissue-disrupted extracts, were also present in the intact biopsy extracts [44].

An RSM model was created to find the optimal combination of quantitative extraction parameters, i.e. tissue mass (50–100 mg), solvent volume (350–650 μ L) and centrifugation time (5–10 min). For 21 out of the 29 selected metabolites, the highest tissue mass gave significantly higher peak areas. For 15 out of 29 selected metabolites, the lowest solvent volume resulted in a significant better extraction efficiency. Only 2 metabolites demonstrated significantly higher responses when using a higher solvent volume. The lowest centrifugation time resulted in significantly better results for 11 selected metabolites, whilst for 3 metabolites significantly higher peak areas were obtained with the highest centrifugation time. Upon untargeted evaluation, a higher centrifugation time tended to provide a better, although not significant metabolome coverage. In the final protocol, 100 mg tissue, 400 μ L solvent and 7.5 min centrifugation time were applied for optimal metabolite extraction. Upon linearity assessment, 27 out of 29 metabolites were detected until the 500-fold dilution, suggesting excellent sensitivity of the developed methodology. For the majority of selected metabolites, saturation already occurred at 20-fold dilution. Acceptable linearity ($R^2 \geq 0.950$) was observed for 26 out of 29

selected metabolites (Table 2). For the untargeted analysis, the 1/100–1/5 dilution series provided the best linearity (86.67% of the components had $R^2 > 0.900$) (Table 2) and the untargeted metabolome coverage was still 89.16% for the 20-fold dilution as compared to the undiluted sample. Therefore, it was decided to dilute the samples 20-fold for validation purposes.

Instrumental precision was overall excellent (CV < 15%) for 28 out of 29 metabolites and the CVs for the intra-assay and inter-day precision were less than 20% for 27 and 23 metabolites, respectively (Table 2). Of the untargeted components, 70.56% had a CV < 15% for instrumental precision and 92.68% and 42.68% had a CV < 30% for intra- and inter-day precision, respectively (Table 2).

3.2. Lipidomics: optimization and validation

3.2.1. HT-29 cell line

A higher HT-29 cell line metabolome coverage was obtained upon extraction with dichloromethane as compared to chloroform, methyl-tert-butyl-ether and butanol (Table 3). A total of 2092, 2043, 2025 and 1934 components were obtained with the latter solvents, respectively (Fig. 1B). These results are in agreement with a study of Masson et al. (2015), whereby liver samples obtained higher extraction efficiencies when using dichloromethane as compared to chloroform [18]. However, the majority (72.79%) of the total number of detected lipids were common for all solvent types implying that the remaining 27.21% were absent in at least one solvent type (Fig. 1B). For example, 305 lipids were uniquely extracted with dichloromethane as opposed to butanol and conversely, 137 lipids were specific for butanol in comparison to dichloromethane (Fig. 1B). Earlier studies have indeed demonstrated that the choice of extraction solvent greatly influences lipidome coverage, e.g. methyl-tert-butyl-ether is better suitable for the extraction of more polar lipids [45,46]. Therefore, according to specific lipid classes of interest, the solvent type needs to be adjusted. Nevertheless, since the aim of this study was to cover as much compounds as possible, the use of dichloromethane was preferred. Next to this, a higher solvent volume resulted in better extraction efficiencies for 5 out of 27 endogenously selected lipids and for the lipidome as a whole (Table 3). Vortex time, centrifugation time and centrifugation speed exerted minimal effects on the detection of lipids (Table 3). Therefore, only solvent and injection volumes were further optimized in an RSM model. A high solvent (3–5 mL) and injection volume (4–8 μ L) was demonstrated to be superior for most lipids.

For the linearity experiment, dried cell pellet reconstituted in 150 μ L of chloroform and 450 μ L of methanol (ratio 1:3) was used as non-diluted sample. An acceptable R^2 (≥ 0.950) was obtained for all

Table 2
Optimization and validation results of the colon tissue polar metabolomics method.

Optimization	Targeted evaluation: No. of selected metabolites (n = 29)	Untargeted evaluation: Metabolome coverage (n = 2985)
Tissue mass (25 mg/100 mg)	22 pos. affected with 100 mg	Increase with 100 mg
Solvent volume (500 μ L/1 mL)	1 pos. affected with 500 μ L	No effect
Manual homogenization	12 pos. affected	Increase
Homogenization with TissueLyzer	1 pos. affected	No effect
Centrifugation time (5 min/15 min)	0 affected	No effect
Validation	Targeted evaluation: No. of selected metabolites (n = 29)	Untargeted evaluation: Metabolome coverage (n = 3117)
Linearity (R^2)	26 ($R^2 > 0.950$)	86.67% ($R^2 > 0.900$)
Instrumental precision (CV%)	28 (CV < 15%)	70.56% (CV < 15%)
Intra-assay precision (CV%)	27 (CV < 20%)	92.68% (CV < 30%)
Inter-day precision (CV%)	23 (CV < 20%)	42.68% (CV < 30%)

Table 3
Optimization and validation results of the HT-29 cell line lipidomics method.

Optimization	Targeted evaluation:	Untargeted evaluation:
	No. of selected metabolites (n = 27)	Metabolome coverage (n = 2330)
Solvent dichloromethane	17 pos. affected	Increase
Solvent chloroform	3 neg. affected	No effect
Solvent butanol	10 neg. affected	No effect
Solvent methyl-tert-butyl-ether	2 neg. affected	No effect
Amount of solvent (1 mL/3 mL)	5 pos. affected with 3 mL	Increase with 3 mL
Vortex time (30 s/120 s)	1 pos. affected with 120 s	No effect
Centrifugation time (5 min/20 min)	1 pos. affected with 20 min	No effect
Centrifugation speed (368 × g/9,223 × g)	0 affected	No effect
Validation	Targeted evaluation:	Untargeted evaluation:
	No. of selected metabolites (n = 27)	Metabolome coverage (n = 2249)
Linearity (R ²)	27 (R ² > 0.950)	68.89% (R ² > 0.900)
Instrumental precision (CV%)	27 (CV < 15%)	77.09% (CV < 15%)
Intra-assay precision (CV%)	22 (CV < 20%)	28.91% (CV < 30%)
Inter-day precision (CV%)	21 (CV < 20%)	32.38% (CV < 30%)

reference lipids (Table 3). Twelve out of 27 lipids were detectable until 500-fold dilution and only sphingosine led to saturation of the MS instrument in the undiluted sample. For the untargeted analysis, the 1/20 – undiluted range showed the best linearity results, i.e. 68.89% of the components had R² > 0.900 (Table 3), and the 2-fold dilution contained still 100% of the untargeted components. Nevertheless, obtained CVs of the validation experiments were poor and upon inspecting LC-MS vials after cloud formation could be observed. Therefore, it was decided to increase the solvent ratio by adding chloroform and to adjust the dilution volume. This resulted in a final solvent ratio of 1:1 chloroform/methanol and a resuspension volume of 600 µL.

Next, validation was repeated and all reference lipids had a CV less than 15% for instrumental variation, 22 and 21 out of 27 had a CV less than 20% for intra-assay precision and intra-day repeatability, respectively (Table 3). Nevertheless, only linoleic acid and α -linolenic acid had a CV > 30% for the latter two parameters (Table S7), which could be explained by the higher susceptibility of polyunsaturated fatty acids to lipid peroxidation [47,48]. Instrumental variation, intra-assay and inter-day precision were good for 77.09% (CV < 15%), 28.91% and 32.38% (CV < 30%) of the untargeted components, respectively (Table 3). Thus, the majority of the endogenously selected targeted lipid showed good reproducibility in contrast to the untargeted analysis and these results emphasize the crucial importance to properly validate -omics methods by assessing all the detected components, i.e. knowns and unknowns. Also, to the best of our knowledge, this is the first time that untargeted validation results of extraction for the cellular lipidome

are presented which makes comparison with other studies impossible.

3.2.2. Colon pig tissue

Almost no reference lipids were significantly altered in extraction efficiency as a result of vortex and centrifugation time and speed (Table 4). Nevertheless, it was decided to further optimize these quantitative parameters in an RSM experiment, together with the volume of resuspension after N₂ drying. A lower resuspension volume resulted in better extraction efficiency for the majority of lipids and was applied in the finalized extraction protocol.

All reference lipids had acceptable linearity (R² ≥ 0.95), except for α -tocopherol (Table 4, Table S8). The majority (n = 26) of lipids were detectable until the 500-fold dilution, however 20 of them showed saturation in the undiluted sample that was reconstituted in 150 µL of chloroform and 450 µL of methanol. The optimal dilution range for untargeted analysis was between 1/50–1/2 dilution (70.53% of the components had R² > 0.900) and the 5-fold dilution still covered 99.31% of the undiluted lipidome (Table 4). Nevertheless, in analogy with the HT-29 cell line, solvent ratio and dilution volume was adjusted to 1:1 chloroform/methanol and 700 µL, respectively, to account for cloud formation.

All reference lipids had good CV-values for instrumental (< 15%), intra-assay (< 20%) and inter-day precision (< 20%), except PC (14:0/14:0) (Table 4, Table S8). The poor CV values for PC (14:0/14:0) in colon pig tissue can be explained by low peak intensities close to the limit of detection. Indeed, its abundance was about 40 times lower in colonic pig tissue when compared to its abundance

Table 4
Optimization and validation results of the colon tissue lipidomics method.

Optimization	Targeted evaluation:	Untargeted evaluation:
	No. of selected metabolites (n = 28)	Metabolome coverage (n = 4227)
Vortex time (30 s/120 s)	0 affected	No effect
Centrifugation time (5 min/20 min)	0 affected	No effect
Centrifugation speed (368 × g/9,223 × g)	0 affected	No effect
Validation	Targeted evaluation:	Untargeted evaluation:
	No. of selected metabolites (n = 28)	Metabolome coverage (n = 3886)
Linearity (R ²)	27 (R ² > 0.950)	70.53% (R ² > 0.900)
Instrumental precision (CV%)	27 (CV < 15%)	66.34% (CV < 15%)
Intra-assay precision (CV%)	27 (CV < 20%)	51.30% (CV < 30%)
Inter-day precision (CV%)	27 (CV < 20%)	40.75% (CV < 30%)

in the HT-29 extracts. Next to this, 66.34%, 51.30% and 40.75% of the detected untargeted components had a CV less than 15% for instrumental and less than 30% for intra-assay and inter-day precision, respectively (Table 4). These results, including those of the HT-29 cell line, show that analysis of the apolar fraction is more susceptible to variation than the polar fraction in untargeted mode. This is in concordance with previous research where CVs in aqueous extracts were lower than in organic extracts [18]. Nevertheless, the opposite was observed in a study of Leuthold et al. (2016) that performed polar metabolomics and lipidomic profiling of human kidney tissue [49]. These inconsistent observations may be due to the use of different compositions of solvents that has proven to influence the reproducibility of tissue extraction [18,49].

3.3. Metabolic fingerprinting of the NT and T state *in vitro* and *in vivo*

As proof-of-concept, T (HT-29, HCT-116, Caco-2, SW948 and SW480) and NT (CCD CON841 and FHC) colon cell lines, cancerous tissue samples and the corresponding healthy tissue samples of 10 CRC patients were analyzed using the validated polar metabolomics and lipidomics platform. LC-MS raw files of colon cell line and human colonic tissue samples were pre-processed using Compound Discoverer 2.1. For the metabolomics and lipidomics experiment a total of 879 components and 17,432 components were obtained with CV < 30%, respectively. Remarkably, the number of extracted lipids was much higher in human as opposed to pig colon tissue used in the previous optimization and validation experiments ($n = \pm 4000$). This can be explained by the fact that pig and human colon tissue samples were, respectively, taken at the pig slaughter line and during surgery of cancer patients, whereby blood perfusion in tissue is still guaranteed. Literature has shown that the human serum metabolome contains particularly lipids ($n > 17000$) and therefore, the more perfused human tissue samples covered a larger part of the whole lipidome [50,51]. In addition, Williams et al. (2013) have demonstrated that metabolite levels are elevated in cancerous tissue by about 33% as compared to healthy tissue. This could be due to enhanced metabolism (e.g. lipid metabolism) in cancer cells and/or to the presence of more cancer cells per tissue area [52]. In Simca 14.1, validated OPLS-DA models could be constructed for each matrix and experiment with R^2Y and Q^2 higher than 0.5 (Table 5). Furthermore, CV-Anova ($p < 0.01$) and permutations testing confirmed the validity of the models (Table 5). This confirms the suitability of our method, that was optimized and validated using healthy pig colon tissue obtained from the slaughter line, to pinpoint metabolic pathways involved in human gastrointestinal diseases.

Unsupervised hierarchical cluster analysis using Pearson correlation coefficient and Ward linkage of the complete polar metabolome and lipidome datasets combined revealed two main clusters,

i.e. NT and T state in colon cell lines (Figs. 3 and 4A) and tissue (Fig. 4B), which is in line with the validated OPLS-DA models (Table 5). In human breast and pancreatic cell lines, discriminating metabolic profiles between the NT and T state have been obtained before [53,54]. Despite significant variation between technical replicates ($n = 5$, Fig. 3), hierarchical clustering of the pooled data resulted in a clear separation of the different states (Fig. 4A). The higher variability between individual replicates as compared to other reports (e.g. Liu et al. (2014) [28]), is most likely due to the fact that we have performed completely separate extractions at three different time points with a one-week time interval, whereas usually extractions are performed in parallel on the same day. Such true technical replicates also include variation that may arise from different passage numbers. Therefore, we consider the resulting clustering at the batch level as more representative of a true discriminatory effect between the T and NT state. Moreover, clustering of the different transformed cell lines was more distinct when only the top 3000 components, retained after filtering based on p-values, were taken into consideration (Fig. 5). Indeed, two main subclusters were formed, i.e. HT-29 together with Caco-2 and HCT116 together with SW480 and SW948. In the second cluster, HCT116 was separated from the latter two cell lines (Fig. 5). Therefore, efforts should be undertaken to include different passage numbers as additional replicates in future polar metabolomics and lipidomics experiments to obtain more reliable results. Moreover, the results of the unsupervised hierarchical clustering strategy show that the use of two different media (i.e. DMEM for the T cell lines and the NT CCD 841 CON cell line and DMEM:F12 for the NT FHC cell line) did not overrule the ability to discriminate between the T and NT state in colon cells, since the FHC cell line was not allocated as a separate cluster (Fig. 4A). This emphasizes even more the application potential of our method, since it is known that the composition of culture media may significantly influence the cellular metabolic profiles [55,56].

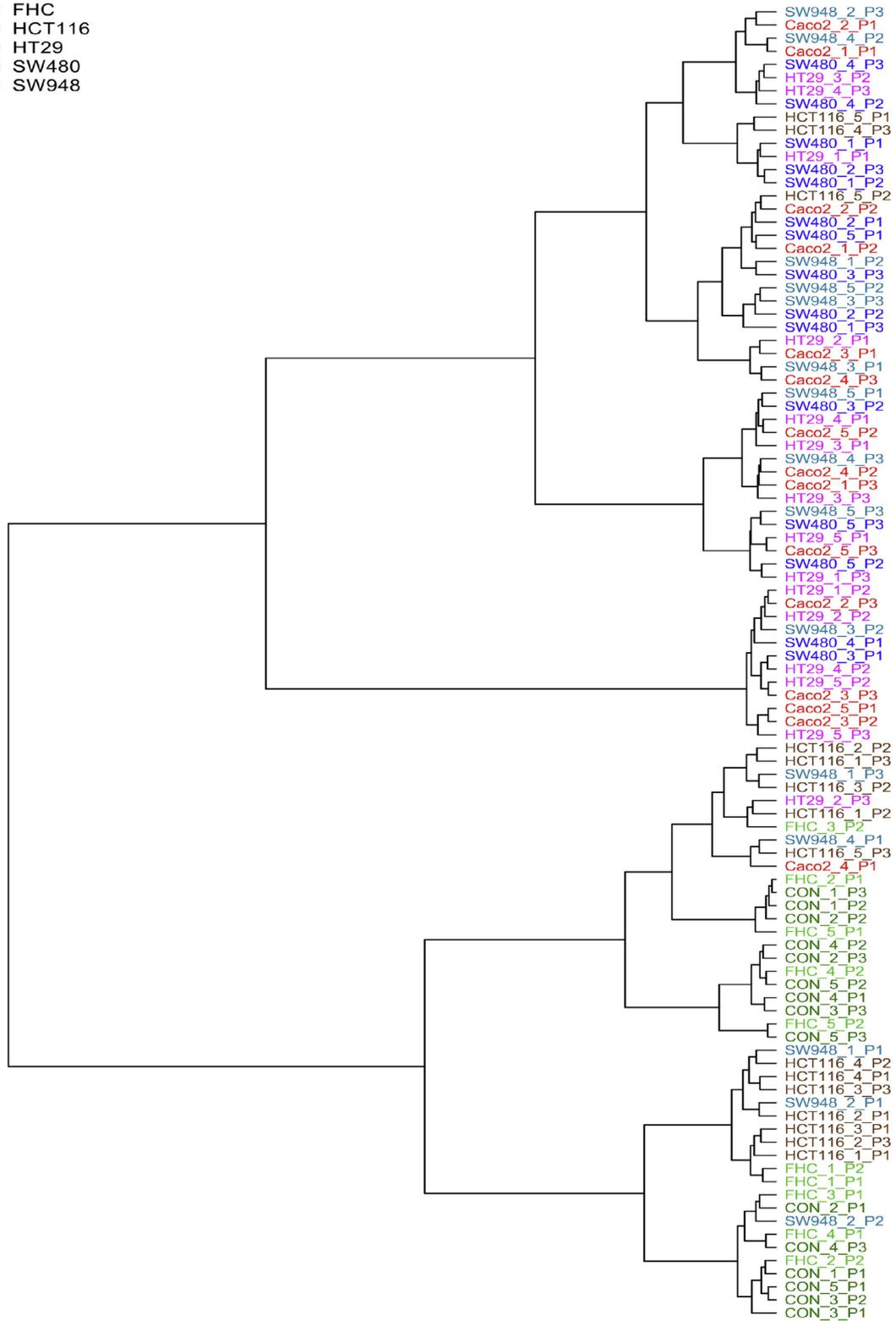
Using our validated OPLS-DA models, 196 and 721 components (VIP > 1.0) were retained that were discriminative for NT and T colon tissue, respectively, implying that metabolism in T tissue is strongly enhanced compared to NT tissue. Of these 196 components specific for NT tissue, 41 and 110 were significantly up- and downregulated, respectively, in NT as opposed to T cell lines (Fig. 6A). Of the 721 components specific for T tissue, 110 and 55 were significantly up- and downregulated in T as opposed to NT cell lines (Fig. 6B). These results indicate that similar trends could be observed at cell line level, although some discrepancies occurred. Indeed, Ertel et al. (2006) have demonstrated that the amount of differentially expressed genes between cancer cell lines and tumors are larger than between tumors and normal tissue [57]. This may be due to accumulating genome changes, as passage numbers increase, cross contamination and changes in gene expression due to the *in vitro* context [58,59]. Therefore, a consequent bias on results

Table 5

Validation parameter values of OPLS-DA models discriminating between the non-transformed and transformed state in colon tissue and cell line samples. IM = ionization mode.

Colon tissue samples	No of components	R^2Y	Q^2	CV-Anova	Permutations testing
Polar metabolomics (+ and – IM)	1 + 1 + 0	0.978	0.938	< 0.001	OK
Lipidomics (+ IM)	1 + 1 + 0	0.939	0.830	< 0.001	OK
Lipidomics (- IM)	1 + 2 + 0	0.897	0.713	< 0.001	OK
Colon cell line samples	No of components	R^2Y	Q^2	CV-Anova	Permutations testing
Polar metabolomics (+ and – IM)	1 + 1 + 0	0.962	0.943	< 0.001	OK
Lipidomics (+ IM)	1 + 1 + 0	0.719	0.674	< 0.001	OK
Lipidomics (- IM)	1 + 1 + 0	0.962	0.943	0.0054	OK

- Caco2
- CON
- FHC
- HCT116
- HT29
- SW480
- SW948



60 50 40 30 20 10 0

Height

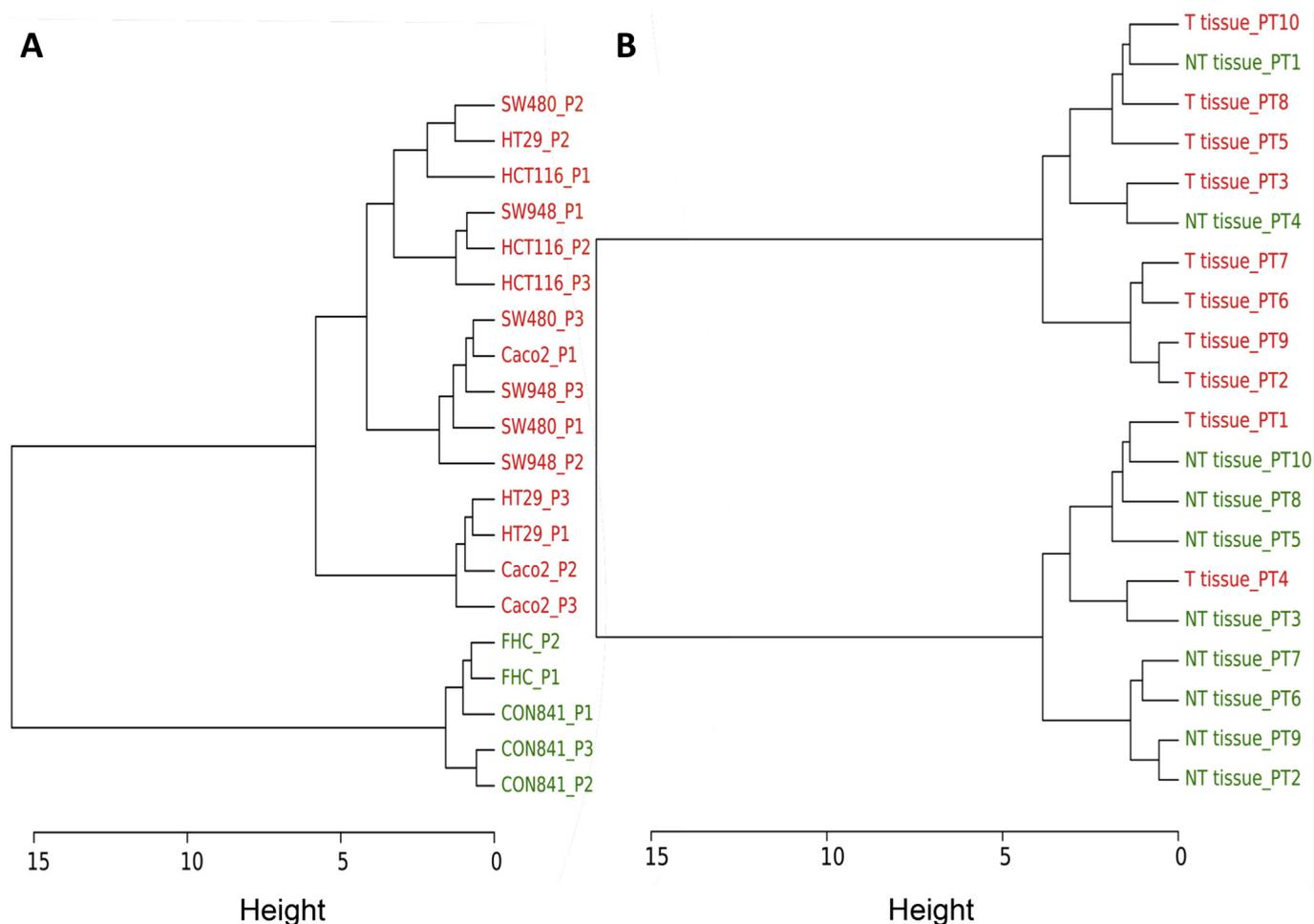


Fig. 4. Hierarchical clustering of non-transformed (= green) and transformed (= red) colon cell line samples (A) and tissue (B) using Pearson correlation coefficient and Ward linkage for the complete polar metabolome and lipidome datasets combined. P1-3 = cell line passage 1–3 (1-week time interval), PT1-10 = colorectal cancer patient 1–10. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

obtained with these cells may affect the reliability of cancer research [22]. The observed discrepancies between tissue and cell lines emphasize the need to identify optimal colon cell lines as models for *in vivo* tissue.

4. Conclusions

We have successfully developed an untargeted holistic metabolic platform that profiles both polar and apolar components in colon cell lines and tissue. This method is the first one that has proven to be ‘fit-for-purpose’ based on different criteria, including linearity, instrumental, inter- and intra assay precision. Uniquely, the number of reproducibly extracted and detected untargeted components was evaluated in parallel with various targeted analytes belonging to a broad range of physicochemical classes. Despite that the analysis of the whole lipidome encountered larger issues in terms of reproducibility as opposed to the metabolomics method, we were able to obtain a tremendous amount of lipids ($n = 17,432$) with $CV < 30\%$ in IQC samples after analysis of NT and T colon tissue

and cell lines. Moreover, this technology, using the hybrid quadrupole Q-Exactive Orbitrap, was able to discriminate between the NT and T state in colon tissue and cell lines and therefore, this platform could be a promising tool in further colon cancer research.

Author contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Notes

The authors declare no competing financial interests.

Declaration of interest statement

The authors declare no competing financial interests.

Fig. 3. Hierarchical clustering of the technical replicates of the non-transformed (CCD CON841 and FHC) and transformed (Caco2, HT29, HCT116, SW480 and SW948) colon cell line samples using Pearson correlation coefficient and Ward linkage for the complete polar metabolome and lipidome datasets combined. P1-3 = cell line passage 1–3 (1-week time interval).

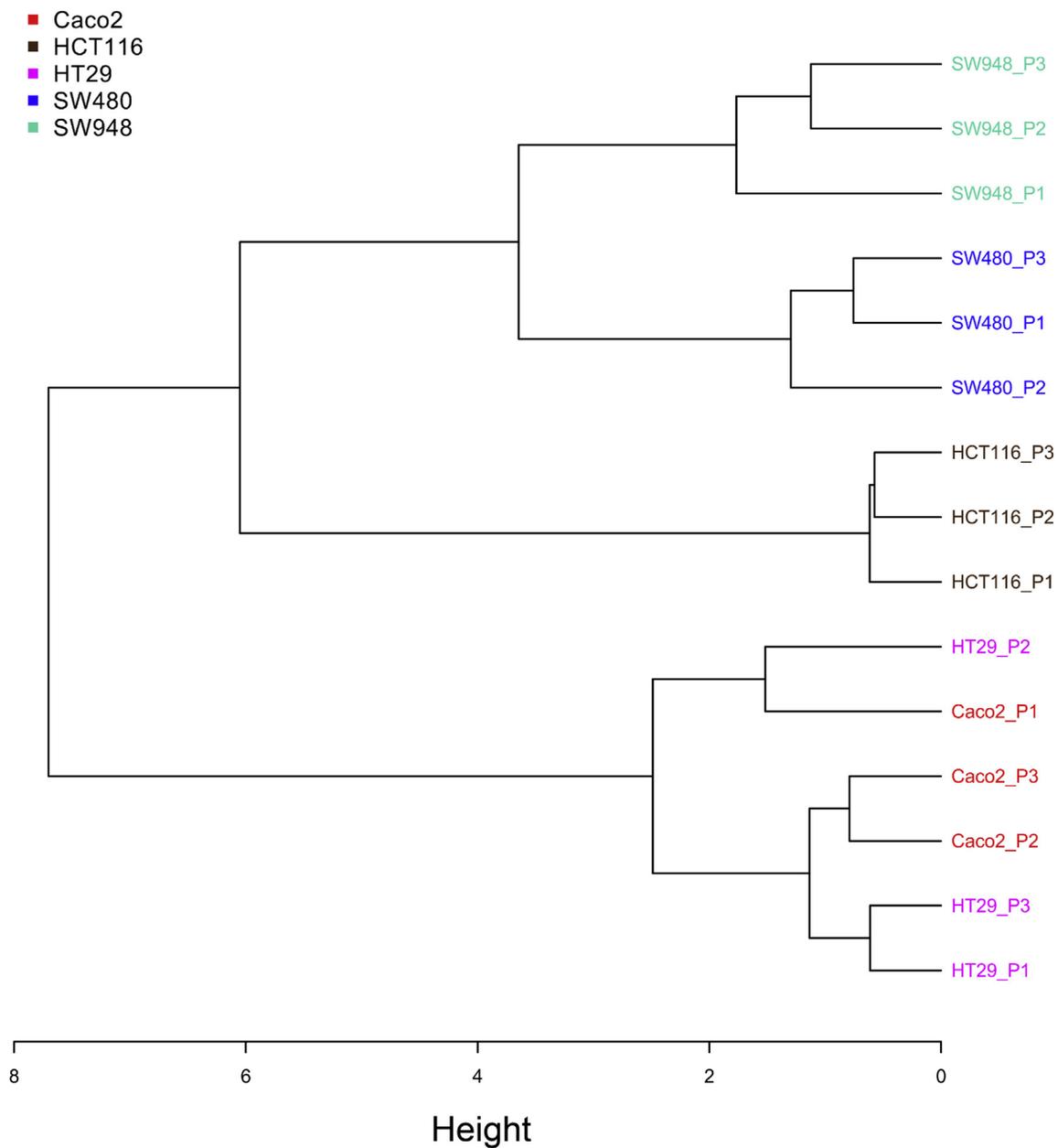


Fig. 5. Hierarchical clustering of the transformed colon cell line samples using Pearson correlation coefficient and Ward linkage for the top 3000 components retained after filtering based on p-values (One-way Anova in MetaboAnalyst) for the complete polar metabolome and lipidome datasets combined. P1-3 = cell line passage 1–3 (1-week time interval).

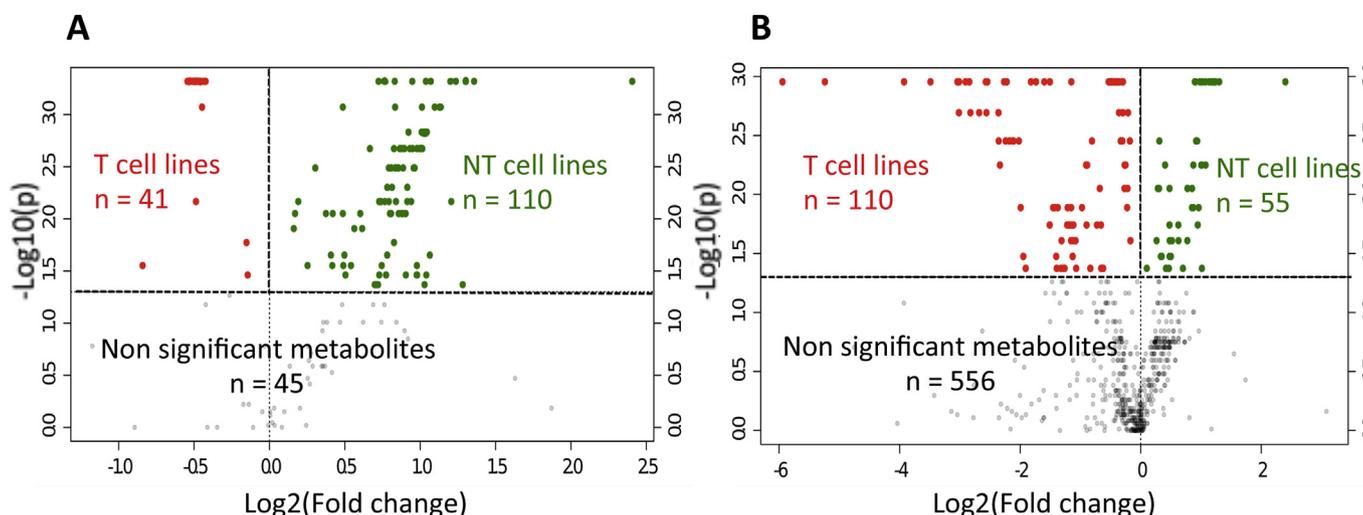


Fig. 6. (A) Volcano plot displaying the behavior of the 196 components with VIP-value > 1.0, which were discriminating for the non-transformed (NT) state as opposed to the transformed (T) state at tissue level, in the cell line samples. Their up- or downregulation in the NT as opposed to the T cell lines was statistically evaluated (T-test in MetaboAnalyst 4.0, significance level at $p < 0.05$). (B) Volcano plot displaying the behavior of the 721 components with VIP-value > 1.0, which were discriminating for the T state as opposed to the NT state at tissue level, in the cell line samples. Their up- or downregulation in the NT as opposed to the T cell lines was statistically evaluated (T-test in MetaboAnalyst 4.0, significance level at $p < 0.05$).

Acknowledgments

Caroline Rombouts is supported by the Institute for Promotion and Innovation through Science and Technology in Flanders, Belgium (IWT, project no 151535) (IWT merged with the Flemish Institute for Scientific Research (FWO) in 2016). The authors wish to thank S. Bos and L. De Beule for their contributions to this work.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.aca.2019.03.020>.

References

- [1] I. Marmol, C. Sanchez-de-Diego, A. Pradilla Dieste, E. Cerrada, M.J. Rodriguez Yoldi, Colorectal carcinoma: a general overview and future perspectives in colorectal cancer, *Int. J. Mol. Sci.* 18 (2017) 197–236.
- [2] G. Smith, F.A. Carey, J. Beattie, M.J. Wilkie, T.J. Lightfoot, J. Coxhead, R.C. Garner, R.J. Steele, C.R. Wolf, Mutations in APC, Kirsten-ras, and p53-alternative genetic pathways to colorectal cancer, *Proc. Natl. Acad. Sci. U. S. A.* 99 (2002) 9433–9438.
- [3] O. Warburg, F. Wind, E. Negelein, The metabolism of tumors in the body, *J. Gen. Physiol.* 8 (1927) 519–530.
- [4] A. Moussaieff, M. Rouleau, D. Kitsberg, M. Cohen, G. Levy, D. Barasch, A. Nemirovski, S. Shen-Orr, I. Laevsky, M. Amit, D. Bomze, B. Elena-Herrmann, T. Scherf, M. Nissim-Rafinia, S. Kempa, J. Itskovitz-Eldor, E. Meshorer, D. Aberdam, Y. Nahmias, Glycolysis-mediated changes in acetyl-CoA and histone acetylation control the early differentiation of embryonic stem cells, *Cell Metabol.* 21 (2015) 392–402.
- [5] C. Li, G. Zhang, L. Zhao, Z. Ma, H. Chen, Metabolic reprogramming in cancer cells: glycolysis, glutaminolysis, and Bcl-2 proteins as novel therapeutic targets for cancer, *World J. Surg. Oncol.* 14 (2016) 15–22.
- [6] J.F. Pereira, N.T. Awatade, C.A. Loureiro, P. Matos, M.D. Amaral, P. Jordan, The third dimension: new developments in cell culture models for colorectal research, *Cell. Mol. Life Sci.* 73 (2016) 3971–3989.
- [7] W.C. van Staveren, D.Y. Solis, A. Hebrant, V. Detours, J.E. Dumont, C. Maenhaut, Human cancer cell lines: experimental models for cancer cells in situ? For cancer stem cells? *Biochim. Biophys. Acta* 1795 (2009) 92–103.
- [8] K. Lenaerts, F.G. Bouwman, W.H. Lamers, J. Renes, E.C. Mariman, Comparative proteomic analysis of cell lines and scrapings of the human intestinal epithelium, *BMC Genomics* 8 (2007) 91–105.
- [9] J.K. Nicholson, I.D. Wilson, Understanding 'global' systems biology: metabolomics and the continuum of metabolism, *Nat. Rev. Drug Discov.* 2 (2003) 668–676.
- [10] R. Goodacre, S. Vaidyanathan, W.B. Dunn, G.G. Harrigan, D.B. Kell, Metabolomics by numbers: acquiring and understanding global metabolite data, *Trends Biotechnol.* 22 (2004) 245–252.
- [11] R. Mirnezami, K. Spagou, P.A. Vorkas, M.R. Lewis, J. Kinross, E. Want, H. Shion, R.D. Goldin, A. Darzi, Z. Takats, E. Holmes, O. Cloarec, J.K. Nicholson, Chemical mapping of the colorectal cancer microenvironment via MALDI imaging mass spectrometry (MALDI-MSI) reveals novel cancer-associated field effects, *Mol. Oncol.* 8 (2014) 39–49.
- [12] H. Wang, J. Xu, Y. Chen, R. Zhang, J. He, Z. Wang, Q. Zang, J. Wei, X. Song, Z. Abliz, Optimization and evaluation strategy of esophageal tissue preparation protocols for metabolomics by LC-MS, *Anal. Chem.* 88 (2016) 3459–3464.
- [13] Q. Teng, W.L. Huang, T.W. Collette, D.R. Ekman, C. Tan, A direct cell quenching method for cell-culture based metabolomics, *Metabolomics* 5 (2009) 199–208.
- [14] K. Dettmer, N. Nurnberger, H. Kaspar, M.A. Gruber, M.F. Almstetter, P.J. Oefner, Metabolite extraction from adherently growing mammalian cells for metabolomics studies: optimization of harvesting and extraction protocols, *Anal. Bioanal. Chem.* 399 (2011) 1127–1139.
- [15] H.C. Bi, K.W. Krausz, S.K. Manna, F. Li, C.H. Johnson, F.J. Gonzalez, Optimization of harvesting, extraction, and analytical protocols for UPLC-ESI-MS-based metabolomic analysis of adherent mammalian cancer cells, *Anal. Bioanal. Chem.* 405 (2013) 5279–5289.
- [16] E.J. Want, P. Masson, F. Michopoulos, I.D. Wilson, G. Theodoridis, R.S. Plumb, J. Shockor, N. Loftus, E. Holmes, J.K. Nicholson, Global metabolic profiling of animal and human tissues via UPLC-MS, *Nat. Protoc.* 8 (2013) 17–32.
- [17] M.M. Ulaszewska, C.H. Weinert, A. Trimigno, R. Portmann, C. Andres Lacueva, R. Badertscher, L. Brennan, C. Brunius, A. Bub, F. Capozzi, M. Cialie Rosso, C.E. Cordero, H. Daniel, S. Durand, B. Egert, P.G. Ferrario, E.J.M. Feskens, P. Franceschi, M. Garcia-Aloy, F. Giacomoni, P. Giesbertz, R. Gonzalez-Dominguez, K. Hanhineva, L.Y. Hemeryck, J. Kopka, S.E. Kulling, R. Llorach, C. Manach, F. Mattivi, C. Migne, L.H. Munger, B. Ott, G. Picone, G. Pimentel, E. Pujos-Guillot, S. Riccadonna, M.J. Rist, C. Rombouts, J. Rubert, T. Skurk, P.S.C. Sri Harsha, L. Van Meulebroek, L. Vanhaecke, R. Vazquez-Fresno, D. Wishart, G. Vergeres, Nutrimetabolomics: an integrative action for metabolomic analyses in human nutritional studies, *Mol. Nutr. Food Res.* (2018) 1800384–1800422.
- [18] P. Masson, A.C. Alves, T.M. Ebbels, J.K. Nicholson, E.J. Want, Optimization and evaluation of metabolite extraction protocols for untargeted metabolic profiling of liver samples by UPLC-MS, *Anal. Chem.* 82 (2010) 7779–7786.
- [19] J. Vanden Bussche, M. Marzorati, D. Laukens, L. Vanhaecke, Validated high resolution mass spectrometry-based approach for metabolomic fingerprinting of the human gut phenotype, *Anal. Chem.* 87 (2015) 10927–10934.
- [20] L. Van Meulebroek, E. De Paepe, V. Verbruyse, B. Pomian, S. Bos, B. Lapauw, L. Vanhaecke, Holistic lipidomics of the human gut phenotype using validated ultra-high-performance liquid chromatography coupled to hybrid orbitrap mass spectrometry, *Anal. Chem.* 89 (2017) 12502–12510.
- [21] E. De Paepe, L. Van Meulebroek, C. Rombouts, S. Huysman, K. Verplanken, B. Lapauw, J. Wauters, L.Y. Hemeryck, L. Vanhaecke, A validated multi-matrix platform for metabolomic fingerprinting of human urine, feces and plasma using ultra-high performance liquid-chromatography coupled to hybrid orbitrap high-resolution mass spectrometry, *Anal. Chim. Acta* 1033 (2018) 108–118.
- [22] N. Zhao, Y. Liu, Y. Wei, Z. Yan, Q. Zhang, C. Wu, Z. Chang, Y. Xu, Optimization of cell lines as tumour models by integrating multi-omics data, *Briefings Bioinf.*

- 18 (2017) 515–529.
- [23] S.C. Sapcariu, T. Kanashova, D. Weindl, J. Ghelfi, G. Dittmar, K. Hiller, Simultaneous extraction of proteins and metabolites from cells in culture, *Methods (Orlando)* 1 (2014) 74–80.
- [24] E. Fahy, S. Subramaniam, R.C. Murphy, M. Nishijima, C.R. Raetz, T. Shimizu, F. Spener, G. van Meer, M.J. Wakelam, E.A. Dennis, Update of the LIPID MAPS comprehensive classification system for lipids, *J. Lipid Res.* 50 (Suppl) (2009) S9–S14.
- [25] S.K. Jensen, Improved bligh and dyer extraction procedure lipid technology, *Lipid Technol.* 20 (12) (2008) 280–281.
- [26] L.P. de Jonge, R.D. Douma, J.J. Heijnen, W.M. van Gulik, Optimization of cold methanol quenching for quantitative metabolomics of *Penicillium chrysogenum*, *Metabolomics* 8 (2012) 727–735.
- [27] T. Cajka, O. Fiehn, Comprehensive analysis of lipids in biological systems by liquid chromatography-mass spectrometry, *Trends Anal. Chem.* 61 (2014) 192–206.
- [28] X. Liu, Z. Ser, J.W. Locasale, Development and quantitative evaluation of a high-resolution metabolomics technology, *Anal. Chem.* 86 (2014) 2175–2184.
- [29] E.J. Want, I.D. Wilson, H. Gika, G. Theodoridis, R.S. Plumb, J. Shockcor, E. Holmes, J.K. Nicholson, Global metabolic profiling procedures for urine using UPLC-MS, *Nat. Protoc.* 5 (2010) 1005–1018.
- [30] C. Brunius, L. Shi, R. Landberg, Large-scale untargeted LC-MS metabolomics data correction using between-batch feature alignment and cluster-based within-batch signal intensity drift correction, *Metabolomics*, Official journal of the Metabolomic Society 12 (2016) 173–186.
- [31] B. Li, J. Tang, Q. Yang, X. Cui, S. Li, S. Chen, Q. Cao, W. Xue, N. Chen, F. Zhu, Performance evaluation and online realization of data-driven normalization methods used in LC/MS based untargeted metabolomics analysis, *Sci Rep-Uk* 6 (2016) 38881–38894.
- [32] C. Rombouts, L.Y. Hemeryck, T. Van Hecke, S. De Smet, W.H. De Vos, L. Vanhaecke, Untargeted metabolomics of colonic digests reveals kynurenine pathway metabolites, dihydroxy and 3-dehydroxycarnitine as red versus white meat discriminating metabolites, *Sci Rep-Uk* 7 (2017) 42514–42527.
- [33] Z. Ser, X.J. Liu, N.N. Tang, J.W. Locasale, Extraction parameters for metabolomics from cultured cells, *Anal. Biochem.* 475 (2015) 22–28.
- [34] C. Muschet, G. Moller, C. Prehn, M.H. de Angelis, J. Adamski, J. Tokarz, Removing the bottlenecks of cell culture metabolomics: fast normalization procedure, correlation of metabolites to cell number, and impact of the cell harvesting method, *Metabolomics* 12 (2016) 151–163.
- [35] R.V. Kapoore, R. Coyle, C.A. Staton, N.J. Brown, S. Vaidyanathan, Influence of washing and quenching in profiling the metabolome of adherent mammalian cells: a case study with the metastatic breast cancer cell line MDA-MB-231, *Analyst* 142 (2017) 2038–2049.
- [36] C.L. Winder, W.B. Dunn, S. Schuler, D. Broadhurst, R. Jarvis, G.M. Stephens, R. Goodacre, Global metabolic profiling of *Escherichia coli* cultures: an evaluation of methods for quenching and extraction of intracellular metabolites, *Anal. Chem.* 80 (2008) 2939–2948.
- [37] C.A. Sellick, D. Knight, A.S. Croxford, A.R. Maqsood, G.M. Stephens, R. Goodacre, A.J. Dickson, Evaluation of extraction processes for intracellular metabolite profiling of mammalian cells: matching extraction approaches to cell type and metabolite targets, *Metabolomics* 6 (2010) 427–438.
- [38] A. Lindahl, S. Saaf, J. Lehtio, A. Nordstrom, Tuning metabolome coverage in reversed phase LC-MS metabolomics of MeOH extracted samples using the reconstitution solvent composition, *Anal. Chem.* 89 (2017) 7356–7364.
- [39] A. Ipsen, E.J. Want, J.C. Lindon, T.M. Ebbels, A statistically rigorous test for the identification of parent-fragment pairs in LC-MS datasets, *Anal. Chem.* 82 (2010) 1766–1778.
- [40] H.G. Gika, I.D. Wilson, G.A. Theodoridis, LC-MS-based holistic metabolic profiling. Problems, limitations, advantages, and future perspectives, *J Chromatogr B Analyt Technol Biomed Life Sci* 966 (2014) 1–6.
- [41] D. Saigusa, Y. Okamura, I.N. Motoike, Y. Katoh, Y. Kurosawa, R. Saijyo, S. Koshihara, J. Yasuda, H. Motohashi, J. Sugawara, O. Tanabe, K. Kinoshita, M. Yamamoto, Establishment of protocols for global metabolomics by LC-MS for biomarker discovery, *PLoS One* 11 (2016) 160555–160573.
- [42] C. Hellmuth, O. Uhl, M. Standl, H. Demmelmaier, J. Heinrich, B. Koletzko, E. Thiering, Cord blood metabolome is highly associated with birth weight, but less predictive for later weight development, *Obesity facts* 10 (2017) 85–100.
- [43] C.L. Gibson, S.G. Codreanu, A.C. Schrimpe-Rutledge, C.L. Retzlaff, J. Wright, D.P. Mortlock, S.D. Sherrod, J.A. McLean, R.D. Blakely, Global untargeted serum metabolomic analyses nominate metabolic pathways responsive to loss of expression of the orphan metallo beta-lactamase, MBLAC1, *Molecular omics* 14 (2018) 142–155.
- [44] M.V. Brown, J.E. McDunn, P.R. Gunst, E.M. Smith, M.V. Milburn, D.A. Troyer, K.A. Lawton, Cancer detection and biopsy classification using concurrent histopathological and metabolomic analysis of core biopsies, *Genome Med.* 4 (2012) 33–45.
- [45] S.K. Byeon, J.Y. Lee, M.H. Moon, Optimized extraction of phospholipids and lysophospholipids for nanoflow liquid chromatography-electrospray ionization-tandem mass spectrometry, *Analyst* 137 (2012) 451–458.
- [46] S. Tumanov, J.J. Kamphorst, Recent advances in expanding the coverage of the lipidome, *Curr. Opin. Biotechnol.* 43 (2017) 127–133.
- [47] E. Niki, Lipid peroxidation: physiological levels and dual biological effects, *Free Radic. Biol. Med.* 47 (2009) 469–484.
- [48] H. Yin, L. Xu, N.A. Porter, Free radical lipid peroxidation: mechanisms and analysis, *Chem. Rev.* 111 (2011) 5944–5972.
- [49] P. Leuthold, E. Schaeffeler, S. Winter, F. Buttner, U. Hofmann, T.E. Murrter, S. Rausch, D. Sonntag, J. Wahrheit, F. Fend, J. Hennenlotter, J. Bedke, M. Schwab, M. Haag, Comprehensive metabolomic and lipidomic profiling of human kidney tissue: a platform comparison, *J. Proteome Res.* 16 (2017) 933–944.
- [50] S. Bouatra, F. Aziat, R. Mandal, A.C. Guo, M.R. Wilson, C. Knox, T.C. Bjorn Dahl, R. Krishnamurthy, F. Saleem, P. Liu, Z.T. Dame, J. Poelzer, J. Huynh, F.S. Yallou, N. Psychogios, E. Dong, R. Bogumil, C. Roehring, D.S. Wishart, The human urine metabolome, *PLoS One* 8 (2013) 73076–73104.
- [51] N. Psychogios, D.D. Hau, J. Peng, A.C. Guo, R. Mandal, S. Bouatra, I. Sinelnikov, R. Krishnamurthy, R. Eisner, B. Gautam, N. Young, J. Xia, C. Knox, E. Dong, P. Huang, Z. Hollander, T.L. Pedersen, S.R. Smith, F. Bamforth, R. Greiner, B. McManus, J.W. Newman, T. Goodfriend, D.S. Wishart, The human serum metabolome, *PLoS One* 6 (2011) 16957–16980.
- [52] M.D. Williams, R. Reeves, L.S. Resar, H.H. Hill Jr., Metabolomics of colorectal cancer: past and current analytical platforms, *Anal. Bioanal. Chem.* 405 (2013) 5013–5030.
- [53] M. Watanabe, S. Sheriff, K.B. Lewis, J. Cho, S.L. Tinch, A. Balasubramaniam, M.A. Kennedy, Metabolic profiling comparison of human pancreatic ductal epithelial cells and three pancreatic cancer cell lines using NMR based metabolomics, *J. Mol. Biomark. Diagn.* 3 (2012). S3-002.
- [54] C.L. Silva, R. Perestrelo, P. Silva, H. Tomas, J.S. Camara, Volatile metabolomic signature of human breast cancer cell lines, *Sci Rep-Uk* 7 (2017) 43969–43977.
- [55] Z. Huang, W. Shao, J. Gu, X. Hu, Y. Shi, W. Xu, C. Huang, D. Lin, Effects of culture media on metabolic profiling of the human gastric cancer cell line SGC7901, *Mol. Biosyst.* 11 (2015) 1832–1840.
- [56] E. Daskalaki, N.J. Pilon, A. Krook, C.E. Wheelock, A. Checa, The influence of culture media upon observed cell secretome metabolite profiles: the balance between cell viability and data interpretability, *Anal. Chim. Acta* 1037 (2018) 338–350.
- [57] A. Ertel, A. Verghese, S.W. Byers, M. Ochs, A. Tozeren, Pathway-specific differences between tumor cell lines and normal and tumor tissue cells, *Mol. Canc.* 5 (2006) 55–68.
- [58] P. Hughes, D. Marshall, Y. Reid, H. Parkes, C. Gelber, The costs of using unauthenticated, over-passaged cell lines: how much more data do we need? *Biotechniques* 43 (575) (2007) 577–578, 581–572 passim.
- [59] J. Le Beyec, R. Xu, S.Y. Lee, C.M. Nelson, A. Rizki, J. Alcaraz, M.J. Bissell, Cell shape regulates global histone acetylation in human mammary epithelial cells, *Exp. Cell Res.* 313 (2007) 3066–3075.