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

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

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 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 spec-

trometry (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 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 metabolomics and lipidomics fingerprinting approach could be of great value to reveal gastrointestinal disease-associated biomarkers and mechanisms.

Keywords: Cancer; Colon cell lines; Colon tissue; Q-Orbitrap-high resolution mass spectrometry; Comprehensive metabolomics/lipidomics profiling

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 correlat-

48 ed with evasion from apoptosis [5]. Therefore, exploring metabolic perturbations in cells offers
49 potential for devising novel prophylactic and therapeutic strategies in cancer [4, 5].

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

58 Untargeted metabolomics studies of different sample matrices have proven to be of great value as
59 a research tool, especially in cancer research, to study changes in the metabolome that is the most
60 accurate reflection of the cellular phenotype [9, 10]. Tissue is the principal location were patho-
61 logical processes occur and can therefore reveal the most interesting changes related to cancer
62 [11]. Nevertheless, sampling and extraction remains more challenging for tissue and cell lines as
63 opposed to biofluids [12]. For example, in cell lines cell detachment from the bottom of well
64 plates and quenching cellular metabolism at time of harvesting are critical points to ensure a ro-
65 bust analysis and special care must be taken to avoid metabolite leakage. In this context, several
66 studies have demonstrated that cell scraping is preferable above conventional trypsinization, since
67 the latter is time consuming and requires more wash and centrifugation steps [13-15]. In addition,
68 tissue is often composed of heterogeneous regions, e.g. presence of different cell types or zones
69 around tumors that are well-oxygenated or necrotic [16]. These factors may introduce additional
70 biological variation, indicating the need for optimization, validation and standardization of extrac-
71 tion parameters specifically for cell lines and tissue [16, 17]. Next to this, due to the broad physi-
72 co-chemical range of metabolites (polar and apolar) present in biological matrices, it is not possi-

ble 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 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

98 culture the cancer cell lines and CCD 841 CON in a humidified incubator at 37 °C and 5%
99 CO_2/95% air. FHC cells were cultured in DMEM:F12 supplemented with 10 mM HEPES, 0.005
100 mg mL^{-1} insulin, 0.005 mg mL^{-1} transferrin, 100 ng mL^{-1} hydrocortisone, 10% FBS and 1% P/S.
101 All cell reagents were purchased from Life Technologies (Ghent, Belgium). Twenty-four hours
102 before extraction cells were seeded in 6-well plates, making sure that on the day of extraction ±
103 80% confluency was reached. When the metabolomics/lipidomics protocol was applied to com-
104 pare the NT with the T state, each cell line was seeded in quintuplicate 24h before extraction (=
105 technical replicates) and this experiment was repeated three times with 1-week interval (= cell
106 line batches).

107 *2.1.2. Colon tissue*

108 For optimization and validation experiments, colon tissue samples were obtained from a local pig
109 slaughterhouse. The colon was washed with Phosphate Buffered Saline (PBS) to remove fecal
110 content and mucus. Small pieces of colon tissue (± 3 g) were dissected, aliquoted in 15 mL falcon
111 tubes and immediately snap frozen in liquid nitrogen. The metabolic fingerprinting of the T and
112 the NT state *in vivo* was performed on cancerous colon material and the corresponding healthy
113 tissue of 10 individuals, provided by Biobank@UZA (Antwerp, Belgium; ID: BE71030031000);
114 Belgian Virtual Tumorbank funded by the National Cancer Plan. The use of human colon tissue
115 samples in this study was ethically approved by the University Hospital of Antwerp (ECD
116 16/37/368 (Antwerp, Belgium)).

117

118

119 **2.2. UHPLC-Q-Orbitrap-UHRMS**

120 *2.2.1. Reagents and chemicals*

121 Analytical standards were purchased from Sigma-Aldrich (St-Louis, Missouri, USA), ICN Bio-
122 medicals Inc. (Ohio, USA), TLC Pharmchem (Vaughan, Ontario, Canada) or Cambridge Isotope
123 Laboratories Inc. (Tewksbury, Massachusetts, USA). More detailed information about chromato-

graphic 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 µm, 150 mm × 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 µL 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 µm, 150 mm × 2.1 mm, Waters), maintained at 40 °C, in combination with a Hypersil GOLD pre-column (1.9 µm, 50 mm × 2.1 mm, Thermo Fischer Scientific, San José, CA, USA) were used. The binary gradient program based on ultrapure water and methanol, both acidified with 3.5 mM NH_4Ac, 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].

150

151 **2.3. Polar metabolomics: optimization of sample extraction**

152 Thirty-one metabolites, belonging to 12 different chemical classes (i.e. polyols, hydroxylic acids,
153 multicarboxylic acids, imidazoles, carbohydrates, amino acids, amines, amides, other N-
154 compounds, fatty acids, sulfonic acids and phosphates) with a broad range of physicochemical
155 properties, were selected for statistical evaluation during chemometric optimization of the HT-29
156 cell line (Table S1) and colon tissue (Table S2) extraction. These compounds were chosen based
157 on their endogenous presence in the samples under investigation and their identity was confirmed
158 based on *m/z*-value, retention time and $^{13}C/^{12}C$ isotope ratio of the corresponding in-house analyt-
159 ical standard.

160 *2.3.1. HT-29 cell line*

161 Optimization parameters were adapted from a general extraction protocol for cell lines, as de-
162 scribed by Sapcariu *et al*. (2014) [23]. Different types of extraction solvent (ethanol, methanol
163 and acetonitrile), the ratio organic solvent *vs.* water (1:1 or 7:2, v/v), washing buffer (PBS or
164 0.9% NaCl, 1x washing or 2x washing) and drying of the cell extracts (flushing with N_2 or vacu-
165 um-drying) were assessed by a D-optimal statistical design (n=29), which determines the optimal
166 set of factors, i.e. extraction parameters, to be tested, in Modde 5.0. Finally, a response surface
167 methodology (RSM) model (n=18) was created to further determine the optimal values for quan-
168 titative variables, i.e. ratio extraction solvent/water (3:1, 2:1, 1:1, 1:2, v/v) and dilution volume of
169 cell extract after vacuum drying (125, 150, 500, 750 and 1000 μL).

170 *2.3.2. Colon pig tissue*

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

176

177 **2.4. Lipidomics: optimization of sample extraction**

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

186 *2.4.1. HT-29 cell line*

187 Dichloromethane, chloroform, methyl-tert-butyl-ether and butanol were assessed in different vol-
188 umes (1-3 mL). Also, vortex time (30-120 s), centrifugation time (5-20 min) and speed (368-
189 9,223 x g) were tested in a D-optimal statistical model (n=19). Additionally, solvent (2, 3.5 and 5
190 mL) and injection volume (1, 4.5 and 8 μL) were further optimized using an RSM model (n=11).

191 *2.4.2. Colon pig tissue*

192 Vortex time (30-120 s), centrifugation time (5-20 min) and speed (1,845-9,223 x g) were tested
193 using a D-optimal statistical model (n=11). Furthermore, an RSM model (n=17) was created to
194 further optimize vortex time (90, 120 and 150 s), centrifugation speed (184, 3,228 and 5,534 x g)
195 and volume of resuspension after drying with N_2 (500, 750 and 1000 μL).

196

197 **2.5. Validation of the polar metabolomics and lipidomics methods**

198 For validation purposes, the endogenously present analytes evaluated in both sample matrices
199 were evaluated as described for optimization (Table S5-S8). First, a dilution series (500-fold, 200-
200 fold, 100-fold, 50-fold, 20-fold, 10-fold, 5-fold, 2-fold dilution and undiluted) of the extracted
201 sample was made in mobile phase starting conditions to assess linearity in terms of coefficient of

202 determination (R^2). Based on these results, the dilution volume of the sample matrix was adjusted
203 and thereby assisted in the optimization process. Additionally, instrumental (injecting the same
204 sample 10 times), intra-assay and inter-assay (extraction and subsequent analysis of 10 samples in
205 parallel and of 2 series of 10 samples performed by two technicians at different days, respective-
206 ly) precision were evaluated based on the coefficients of variance (CV) of the selected targeted
207 analytes and untargeted components.
208

209 **2.6. Data analysis**

210 Targeted data processing was carried out with Xcalibur 3.0 software (Thermo Fischer Scientific,
211 San José, CA, USA). For the optimization experiments, peak areas of the targeted analytes or the
212 number of detected components in untargeted mode were imported as responses in Modde 5.0
213 (Umetrics AB, Umea, Sweden) to determine the most optimal extraction parameters. Components
214 were extracted from the full-scan data with Sieve™ 2.2 (Thermo Fischer Scientific, San José,
215 CA, USA) or the more recent available Compound Discoverer 2.1 software (Thermo Fisher Sci-
216 entific, San José, CA, USA), in which chromatographic peak alignment and retrieving of compo-
217 nents was performed. To this extent, the minimum peak intensity was set at 500,000 au, minimum
218 signal-to-noise ratio at 10, minimum number of isotopes at 2, retention time window at 0.75 min,
219 maximum retention time shift at 0.25 min and maximum mass shift (Δ ppm) at 6 ppm. In addi-
220 tion, blanks (i.e. acetonitrile for polar metabolomics and methanol for lipidomics,) were used for
221 substraction of noise peaks in the samples of interest. Compound Discoverer offers the interesting
222 capacity, in comparison to similar software such as Sieve, to handle both positive and negative
223 ionization modes together in one analysis. Therefore, the obtained data were analyzed using the
224 indicated software for the proof-of-concept metabolomics experiment. However, this was not
225 possible for the lipidomics experiment due to the presence of different scan ranges (67-1000 Da
226 and 1000-2300 Da) in the method and hence for the latter separate datasets for both ionization
227 modes were obtained. Normalization of the amount of analyzed cell/tissue material between sam-

228 ples was performed by dividing the abundance of each component by the total ion count (TIC) of
229 the respective sample. Indeed, Liu *et al.* (2014) have demonstrated that over 4 orders of magni-
230 tude, TIC normalized intensities accurately reflect intracellular metabolite levels [28].
231 For the proof-of-concept experiment, cellular extracts obtained from three different passages (1-
232 week interval) were analyzed immediately after harvesting and all cellular LC-MS data together
233 with tissue LC-MS data were processed simultaneously in Compound Discoverer or Sieve for the
234 polar metabolomics and lipidomics experiment, respectively. Components with CVs > 30% in
235 IQC samples, i.e. pools of samples that were run in duplicate after each batch of 10 samples, were
236 not stable during instrumental analysis and therefore removed from the dataset [29, 30]. Finally,
237 the relative abundance of each component was divided by the mean relative abundance of the two
238 following IQC samples to correct for potential instrumental drift [19]. Tissue components were
239 further evaluated using multilevel data analysis in Matlab 9.3, whereby effect of tissue origin for
240 each individual was separated from between individual variation. Next, unsupervised hierarchical
241 clustering (www.metaboanalyst.ca) and multivariate statistical analysis (SIMCA 14.1 software
242 (Umetrics AB, Umea, Sweden)) were performed, whereby PCA-X models were created to evalu-
243 ate natural clustering of samples and to check for outliers whereas OPLS-DA models were built to
244 discriminate between the NT versus T state in tissue and cell lines. For modeling, data were log-
245 transformed and Pareto-scaling was applied [31]. Validity of the obtained OPLS-DA models was
246 evaluated based on R_2Y (> 0.5), Q^2 (> 0.5), CV-Anova ($p < 0.05$) permutations testing ($n = 100$)
247 and VIP values (> 1.0) [32].
248

249 **2.7. Polar metabolomics: optimized extraction protocol**

250 *2.7.1. HT-29 cell line*

251 HT-29 cells were seeded in 6-well plates 24 h before extraction. At the day of extraction, medium
252 was aspirated and cells were washed with 0.9% NaCl. Next, 1 mL 50% methanol/ultrapure water
253 (on ice) and 27 µL working solution containing the internal standard (25 ng μL^{-1} of D-valine-d_8)

254 were added. This was followed by scraping and transferring the cells to an Eppendorf tube. Cells
255 were lysed by sonification (2 × 15 s, Soniprep 150, Beun-De Ronde, LA Abcoude, The Nether-
256 lands) and cell debris was removed by centrifugation (16,200 × g, 5 min, 4 °C). Next, a 100 µL
257 aliquot of supernatant was vacuum-dried, whilst the remaining cell suspension was stored at -80
258 °C for lipidomics analysis. The dried cell pellets were resolved in 1.5 mL of a solvent mixture at
259 HPLC starting conditions (0.1% formic acid in UP water and 0.1% formic acid in acetonitrile in
260 98/2 ratio, respectively) and a subfraction was transferred to a glass HPLC-vial and 10 µL was
261 injected.

262 *2.7.2. Colon pig tissue*

263 100 mg colonic tissue was weighed and 400 µL 50% methanol/ultrapure water (on ice) and 30 µL
264 internal standard (8 ng μL^{-1} D-valine-d_8) were added. The tissue was manually disintegrated with
265 scalpel and forceps, followed by vortexing (10 s) and centrifugation (21,161 x g, 7.5 min, 4 °C).
266 Afterwards, 100 µL supernatant was transferred to an Eppendorf tube and vacuum-dried. The
267 remaining tissue suspension was stored at -80 °C to perform lipidomics. Finally, the tissue pellets
268 were resolved in 2 mL solvent mixture at mobile phase starting conditions and transferred to a
269 glass HPLC-vial and 10 µL was injected.

270

271 **2.8. Lipidomics: optimized extraction protocol**

272 *2.8.1. HT-29 cell line*

273 The ISTDs (60 µL palmitic acid-d_{31} (25 ng μL^{-1}) and 30 µL phosphocholine-d_{54} (25 ng μL^{-1})),
274 900 µL cell suspension, 400 µL ultrapure water (4% trichloric acid) and 500 µL MeOH were add-
275 ed to 3 mL DCM supplemented with 0.01% butylhydroxytoluene (BHT). This mixture was vor-
276 texed (30 s), incubated (20 min, 20 °C) and centrifugated (9,223 x g, 5 min, 20 °C). Next, 2 mL
277 DCM phase was collected with a glass Pasteur pipette and dried at 30 °C under N_2. The pellet was
278 resuspended in 300 µL $CHCl_3$ and vortexed. Finally, 300 µL MeOH was added and the extract
279 was transferred to a glass HPLC vial and 10 µL was injected.

280 *2.8.2. Colon pig tissue*

281 First, 600 μL 50% MeOH/UP H_2O was added to 300 μL tissue suspension. Next, the same proto-
282 col as for the HT-29 cell line was performed, with slight adaptions of the centrifugation speed
283 (368 x g) and resuspension volume (350 μL $CHCl_3$ + 350 μL MeOH) after drying with N_2 and
284 finally, 6 μL was injected in the system.

285

286 ## 3. Results and discussion

287 ### 3.1. Polar metabolomics: optimization and validation

288 *3.1.1. HT-29 cell line*

289 First, different parameters (including washing of cells, type of organic solvent, ratio organic sol-
290 vent/ultrapure water and drying of extracts) for extraction of the polar metabolome of the HT-29
291 cell line were assessed and the most optimal conditions were included in the final protocol. In this
292 context, significant effects were observed for the washing step, the type of extraction solvent and
293 the ratio organic solvent/ultrapure water towards the extraction yield for 14 out of the 22 selected
294 endogenous metabolites, of which standards were present in our in-house database (Table 1).
295 Nonetheless, untargeted analysis in Sieve™ 2.2 revealed no important differences in terms of
296 number of detected components between the different extraction conditions (Table 1). Washing
297 with PBS, a washing solvent frequently used in untargeted metabolomics studies [33, 34], did not
298 improve nor deteriorate the extraction efficiency in comparison to rinsing with 0.9% NaCl solu-
299 tion although putrescine significantly benefitted from washing with the latter one (Table 1, Table
300 S1). To this end, we opted to wash with NaCl, supported by Sapcariu *et al.* (2014) wherein the
301 latter is recommended, since large phosphate peaks originating from PBS could contaminate the
302 obtained MS spectra [23]. Omitting the washing step was not considered in our experimental set-
303 up, as rinsing is necessary to eliminate residual media containing extracellular metabolites and
304 hence avoiding a biased higher metabolome coverage [35]. Due to the fact that no differences in

305 extraction yield could be observed between washing a single time and two times (Table 1), it
306 could be assumed that one washing step sufficed to remove unwanted media constituents. Also,
307 one rinsing step shortens the entire procedure reducing the risk for variability in extraction. Of the
308 tested organic solvents (added to ultrapure water), i.e. ethanol, methanol and acetonitrile, no re-
309 markable or significant differences could be observed in the number of components obtained
310 following untargeted analysis (Figure 1A, Table 1). However, extraction with acetonitrile resulted
311 in inconsistent results during targeted analysis, i.e. significantly higher or lower peak intensities
312 for 2 and 3 out of the 22 selected targeted endogenous metabolites, respectively (Table 1). As a
313 result, and because methanol is the most commonly used extraction solvent in metabolomics stud-
314 ies in literature [23, 33, 36, 37], this solvent was retained for further experiments. The ratio organ-
315 ic solvent/ultrapure water did not affect the extraction yield, i.e. number of components in untar-
316 geted analysis, but remarkably, 13 out of 22 selected targeted metabolites shared higher responses
317 when the 1:1 ratio was used (Table 1). A recent study has shown that an increase in ratio organic
318 solvent/ultrapure water results in a decreased response of compounds with a log P lower than 5
319 [38]. This is in concordance with our untargeted results, whereby components with a lower reten-
320 tion time, and thus lower log P (i.e. higher polarity), were more abundant following extraction
321 with the 1:1 as opposed the 7:2 ratio (Figure 2). Finally, since no effect of drying could be ob-
322 served and because of practical reasons, vacuum drying was preferred to drying with N_2 (Table
323 1). RSM methodology models were constructed to assess the optimal ratio methanol/ultrapure
324 water (3:1 - 1:2) (v/v) and the dilution volume after vacuum-drying (125 – 1000 μL). A lower
325 solvent ratio together with a lower dilution volume after vacuum drying improved the extraction
326 yield during untargeted analysis and the peak areas of 20 out of 22 selected endogenous metabo-
327 lites. Therefore, the ratio methanol/ultrapure water was set at 1:1 (v/v) and a dilution volume of
328 150 μL was used to efficiently extract most cellular metabolites.

329 Prior to validation, a nine-point dilution series (500-fold – not diluted sample) was established to
330 determine the optimal linear range for the detection of the selected endogenous targeted metabo-
331 lites and for the metabolome as a whole. To this extent, 150 µL of supernatant was vacuum dried
332 and the cell pellet was resuspended in the same volume, thus representing the non-diluted sample.
333 A minimum of five successive dilution levels were included to construct calibration curves that
334 obtained the highest coefficient of determination (R^2). With respect to the targeted assessment, 19
335 out of the 22 selected metabolites were present until the 500-fold dilution. For 14 selected metab-
336 olites, saturation of the MS instrument occurred, mostly at 20-fold dilution or less. An acceptable
337 ($R^2 \geq 0.950$) linear dynamic range was achieved for 20 out of 22 metabolites (Table 1). Addition-
338 ally, untargeted components present in the dilution series were extracted during data processing in
339 Sieve 2.2. Next, the R^2 of each component in every dilution range (n=15), containing a minimum
340 of five successive points, was determined and the dilution range that contained the highest num-
341 ber of components with $R^2 > 0.900$ was considered as optimal. As such, the best linearity could be
342 observed for dilution range 1/100 – 1/5 (77.89% of the components with $R^2 > 0.900$) (Table 1).
343 Furthermore, the average percentage of the total number of detected untargeted components was
344 still 98.72% for the 20-fold dilution in comparison to the non-diluted sample and was therefore
345 applied in the final extraction protocol, ensuring minimal loss of compounds. Nevertheless, in the
346 validation experiments and the final protocol we decided to dry 100 µL instead of 150 µL of su-
347 pernatant to reduce time and converted the dilution factor accordingly.
348 Finally, the method's precision was assessed in terms of instrumental, intra-assay and inter-day
349 repeatability. Since there are no general guidelines for the validation of untargeted extraction pro-
350 tocols in metabolomics, FDA recommendations were followed, whereby a CV less than 15% is
351 considered excellent and a CV less than 20% is still acceptable for values close to the limit of
352 quantification [19, 39]. However, for the validation of an untargeted method, an upper limit of
353 30% can be set [40-43]. The instrumental precision was excellent (≤ 15%) for 20 out of 22 en-
354 dogenously selected metabolites (Table 1). The CVs for intra-assay precision (n=10) and inter-

355 day repeatability (n = 20) were less than 20% for 19 and 18 out of 22 metabolites, respectively
356 (Table 1). The precision for the amines was poor (CVs > 39 %) and can be explained by their low
357 abundances close to the limit of detection in the cell line extract (Table 1, Table S5). However, in
358 total 80.00% of the untargeted components demonstrated a CV below 15% for instrumental varia-
359 tion, while 82.61% and 75.44% of the components obtained a CV below 30% for intra- and inter-
360 day precision, respectively (Table 1).

361 *3.1.2. Colon pig tissue*

362 Different conditions for extraction of colonic pig tissue were assessed, including tissue mass,
363 organic solvent volume, homogenization technique and centrifugation time, whereby only for the
364 latter no significant effects on extraction yield could be observed (Table 2). A higher tissue mass,
365 lower solvent volume and manual homogenization showed the best results for the extraction of 23
366 out of 29 of the selected compounds (Table 2). A possible reason for the inadequate performance
367 of the TissueLyzer as opposed to the manual homogenization technique could lay in the higher
368 heat formation leading to disintegration of metabolites caused by the TissueLyzer. However, it
369 could be assumed that manual tissue disruption would be less effective than the TissueLyzer.
370 Moreover, as Brown *et al.* (2012) demonstrated that extracting solid tissue with an aqueous meth-
371 anol solution with or without sample destruction generated similar results, it can be concluded
372 that the addition of methanol as such suffices and comprises a more defining factor than the me-
373 chanical disruption method itself for extracting metabolites. Indeed, in the study by Brown *et al.*
374 (2012), more than 92% of the components detected in the tissue-disrupted extracts, were also
375 present in the intact biopsy extracts [44].

376 An RSM model was created to find the optimal combination of quantitative extraction parame-
377 ters, i.e. tissue mass (50 – 100 mg), solvent volume (350 – 650 µL) and centrifugation time (5 –
378 10 min). For 21 out of the 29 selected metabolites, the highest tissue mass gave significantly
379 higher peak areas. For 15 out of 29 selected metabolites, the lowest solvent volume resulted in a
380 significant better extraction efficiency. Only 2 metabolites demonstrated significantly higher re-

381 sponses when using a higher solvent volume. The lowest centrifugation time resulted in signifi-
382 cantly better results for 11 selected metabolites, whilst for 3 metabolites significantly higher peak
383 areas were obtained with the highest centrifugation time. Upon untargeted evaluation, a higher
384 centrifugation time tended to provide a better, although not significant metabolome coverage. In
385 the final protocol, 100 mg tissue, 400 µL solvent and 7.5 min centrifugation time were applied for
386 optimal metabolite extraction. Upon linearity assessment, 27 out of 29 metabolites were detected
387 until the 500-fold dilution, suggesting excellent sensitivity of the developed methodology. For the
388 majority of selected metabolites, saturation already occurred at 20- fold dilution. Acceptable line-
389 arity ($R^2 \geq 0.950$) was observed for 26 out of 29 selected metabolites (Table 2). For the untargeted
390 analysis, the 1/100 – 1/5 dilution series provided the best linearity (86.67% of the components
391 had $R^2 > 0.900$) (Table 2) and the untargeted metabolome coverage was still 89.16% for the 20-
392 fold dilution as compared to the undiluted sample. Therefore, it was decided to dilute the samples
393 20-fold for validation purposes.
394 Instrumental precision was overall excellent (CV < 15%) for 28 out of 29 metabolites and the
395 CVs for the intra-assay and inter-day precision were less than 20% for 27 and 23 metabolites,
396 respectively (Table 2). Of the untargeted components, 70.56% had a CV < 15% for instrumental
397 precision and 92.68% and 42.68% had a CV < 30% for intra- and inter-day precision, respectively
398 (Table 2).
399

400 **3.2. Lipidomics: optimization and validation**

401 *3.2.1. HT-29 cell line*

402 A higher HT-29 cell line metabolome coverage was obtained upon extraction with dichloro-
403 methane as compared to chloroform, methyl-tert-butyl-ether and butanol (Table 3). A total of
404 2092, 2043, 2025 and 1934 components were obtained with the latter solvents, respectively (Fig-
405 ure 1B). These results are in agreement with a study of Masson *et al.* (2015), whereby liver sam-
406 ples obtained higher extraction efficiencies when using dichloromethane as compared to chloro-

407 form [18]. However, the majority (72.79%) of the total number of detected lipids were common
408 for all solvent types implying that the remaining 27.21% were absent in at least one solvent type
409 (Figure 1B). For example, 305 lipids were uniquely extracted with dichloromethane as opposed to
410 butanol and conversely, 137 lipids were specific for butanol in comparison to dichloromethane
411 (Figure 1B). Earlier studies have indeed demonstrated that the choice of extraction solvent greatly
412 influences lipidome coverage, e.g. methyl-tert-butyl-ether is better suitable for the extraction of
413 more polar lipids [45, 46]. Therefore, according to specific lipid classes of interest, the solvent
414 type needs to be adjusted. Nevertheless, since the aim of this study was to cover as much com-
415 pounds as possible, the use of dichloromethane was preferred. Next to this, a higher solvent vol-
416 ume resulted in better extraction efficiencies for 5 out of 27 endogenously selected lipids and for
417 the lipidome as a whole (Table 3). Vortex time, centrifugation time and centrifugation speed ex-
418 erted minimal effects on the detection of lipids (Table 3). Therefore, only solvent and injection
419 volumes were further optimized in an RSM model. A high solvent (3-5 mL) and injection volume
420 (4 – 8 µL) was demonstrated to be superior for most lipids.
421 For the linearity experiment, dried cell pellet reconstituted in 150 µL of chloroform and 450 µL of
422 methanol (ratio 1:3) was used as non-diluted sample. An acceptable R^2 (≥ 0.950) was obtained for
423 all reference lipids (Table 3). Twelve out of 27 lipids were detectable until 500-fold dilution and
424 only sphingosine led to saturation of the MS instrument in the undiluted sample. For the untarget-
425 ed analysis, the 1/20 – undiluted range showed the best linearity results, i.e. 68.89% of the com-
426 ponents had $R^2 > 0.900$ (Table 3), and the 2-fold dilution contained still 100% of the untargeted
427 components. Nevertheless, obtained CVs of the validation experiments were poor and upon in-
428 specting LC-MS vials after cloud formation could be observed. Therefore, it was decided to in-
429 crease the solvent ratio by adding chloroform and to adjust the dilution volume. This resulted in a
430 final solvent ratio of 1:1 chloroform/methanol and a resuspension volume of 600 µL.
431 Next, validation was repeated and all reference lipids had a CV less than 15% for instrumental
432 variation, 22 and 21 out of 27 had a CV less than 20% for intra-assay precision and intra-day

433 repeatability, respectively (Table 3). Nevertheless, only linoleic acid and α-linolenic acid had a
434 CV > 30% for the latter two parameters (Table S7), which could be explained by the higher sus-
435 ceptibility of polyunsaturated fatty acids to lipid peroxidation [47, 48]. Instrumental variation,
436 intra-assay and inter-day precision were good for 77.09% (CV < 15%), 28.91% and 32.38% (CV
437 < 30%) of the untargeted components, respectively (Table 3). Thus, the majority of the endoge-
438 nously selected targeted lipid showed good reproducibility in contrast to the untargeted analysis
439 and these results emphasize the crucial importance to proper validate -omics methods by as-
440 sessing all the detected components, i.e. knowns and unknowns. Also, to the best of our
441 knowledge, this is the first time that untargeted validation results of extraction for the cellular
442 lipidome are presented which makes comparison with other studies impossible.

443 *3.2.2. Colon pig tissue*

444 Almost no reference lipids were significantly altered in extraction efficiency as a result of vortex
445 and centrifugation time and speed (Table 4). Nevertheless, it was decided to further optimize
446 these quantitative parameters in an RSM experiment, together with the volume of resuspension
447 after N_2 drying. A lower resuspension volume resulted in better extraction efficiency for the ma-
448 jority of lipids and was applied in the finalized extraction protocol.

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

457 All reference lipids had good CV-values for instrumental (< 15%), intra-assay (< 20%) and inter-
458 day precision (< 20%), except PC (14:0/14:0) (Table 4, Table S8). The poor CV values for PC

459 (14:0/14:0) in colon pig tissue can be explained by low peak intensities close to the limit of detec-
460 tion. Indeed, its abundance was about 40 times lower in colonic pig tissue when compared to its
461 abundance in the HT-29 extracts. Next to this, 66.34%, 51.30% and 40.75% of the detected untar-
462 geted components had a CV less than 15% for instrumental and less than 30% for intra-assay and
463 inter-day precision, respectively (Table 4). These results, including those of the HT-29 cell line,
464 show that analysis of the apolar fraction is more susceptible to variation than the polar fraction in
465 untargeted mode. This is in concordance with previous research where CVs in aqueous extracts
466 were lower than in organic extracts [18]. Nevertheless, the opposite was observed in a study of
467 Leuthold *et al.* (2016) that performed polar metabolomics and lipidomic profiling of human kid-
468 ney tissue [49]. These inconsistent observations may be due to the use of different compositions
469 of solvents that has proven to influence the reproducibility of tissue extraction [18, 49].
470

471 **3.3. Metabolic fingerprinting of the NT and T state *in vitro* and *in vivo*.**

472 As proof-of-concept, T (HT-29, HCT-116, Caco-2, SW948 and SW480) and NT (CCD CON841
473 and FHC) colon cell lines, cancerous tissue samples and the corresponding healthy tissue samples
474 of 10 CRC patients were analyzed using the validated polar metabolomics and lipidomics plat-
475 form. LC-MS raw files of colon cell line and human colonic tissue samples were pre-processed
476 using Compound Discoverer 2.1. For the metabolomics and lipidomics experiment a total of 879
477 components and 17,432 components were obtained with CV < 30%, respectively. Remarkably,
478 the number of extracted lipids was much higher in human as opposed to pig colon tissue used in
479 the previous optimization and validation experiments (n = ± 4000). This can be explained by the
480 fact that pig and human colon tissue samples were, respectively, taken at the pig slaughter line
481 and during surgery of cancer patients, whereby blood perfusion in tissue is still guaranteed. Liter-
482 ature has shown that the human serum metabolome contains particularly lipids (n > 17000) and
483 therefore, the more perfused human tissue samples covered a larger part of the whole lipidome
484 [50, 51]. In addition, Williams *et al.* (2013) have demonstrated that metabolite levels are elevated

485 in cancerous tissue by about 33% as compared to healthy tissue. This could be due to enhanced
486 metabolism (e.g. lipid metabolism) in cancer cells and/or to the presence of more cancer cells per
487 tissue area [52]. In Simca 14.1, validated OPLS-DA models could be constructed for each matrix
488 and experiment with R^2Y and Q^2 higher than 0.5 (Table 5). Furthermore, CV-Anova ($p < 0.01$)
489 and permutations testing confirmed the validity the models (Table 5). This confirms the suitability
490 of our method, that was optimized and validated using healthy pig colon tissue obtained from the
491 slaughter line, to pinpoint metabolic pathways involved in human gastrointestinal diseases.
492 Unsupervised hierarchical cluster analysis using Pearson correlation coefficient and Ward linkage
493 of the complete polar metabolome and lipidome datasets combined revealed two main clusters,
494 i.e. NT and T state in colon cell lines (Figure 3 and 4A) and tissue (Figure 4B), which is in line
495 with the validated OPLS-DA models (Table 5). In human breast and pancreatic cell lines, discrim-
496 inating metabolic profiles between the NT and T state have been obtained before [53, 54]. Despite
497 significant variation between technical replicates (n = 5, Figure 3), hierarchical clustering of the
498 pooled data resulted in a clear separation of the different states (Figure 4A). The higher variability
499 between individual replicates as compared to other reports (e.g. Liu *et al.* (2014) [28]), is most
500 likely due to the fact that we have performed completely separate extractions at three different
501 time points with a one-week time interval, whereas usually extractions are performed in parallel
502 on the same day. Such true technical replicates also include variation that may arise from different
503 passage numbers. Therefore, we consider the resulting clustering at the batch level as more repre-
504 sentative of a true discriminatory effect between the T and NT state. Moreover, clustering of the
505 different transformed cell lines was more distinct when only the top 3000 components, retained
506 after filtering based on p-values, were taken into consideration (Figure 5). Indeed, two main sub-
507 clusters were formed, i.e. HT-29 together with Caco-2 and HCT116 together with SW480 and
508 SW948. In the second cluster, HCT116 was separated from the latter two cell lines (Figure 5).
509 Therefore, efforts should be undertaken to include different passage numbers as additional repli-
510 cates in future polar metabolomics and lipidomics experiments to obtain more reliable results.

511 Moreover, the results of the unsupervised hierarchical clustering strategy show that the use of two
512 different media (i.e. DMEM for the T cell lines and the NT CCD 841 CON cell line and
513 DMEM:F12 for the NT FHC cell line) did not overrule the ability to discriminate between the T
514 and NT state in colon cells, since the FHC cell line was not allocated as a separate cluster (Figure
515 4A). This emphasizes even more the application potential of our method, since it is known that
516 the composition of culture media may significantly influence the cellular metabolic profiles [55,
517 56].
518 Using our validated OPLS-DA models, 196 and 721 components (VIP > 1.0) were retained that
519 were discriminative for NT and T colon tissue, respectively, implying that metabolism in T tissue
520 is strongly enhanced compared to NT tissue. Of these 196 components specific for NT tissue, 41
521 and 110 were significantly up- and downregulated, respectively, in NT as opposed to T cell lines
522 (Figure 6A). Of the 721 components specific for T tissue, 110 and 55 were significantly up- and
523 downregulated in T as opposed to NT cell lines (Figure 6B). These results indicate that similar
524 trends could be observed at cell line level, although some discrepancies occurred. Indeed, Ertel *et*
525 *al.* (2006) have demonstrated that the amount of differentially expressed genes between cancer
526 cell lines and tumors are larger than between tumors and normal tissue [57]. This may be due to
527 accumulating genome changes, as passage numbers increase, cross contamination and changes in
528 gene expression due to the *in vitro* context [58, 59]. Therefore, a consequent bias on results ob-
529 tained with these cells may affect the reliability of cancer research [22]. The observed discrepan-
530 cies between tissue and cell lines emphasize the need to identify optimal colon cell lines as mod-
531 els for *in vivo* tissue.
532

533 **4. Conclusions**

534 We have successfully developed an untargeted holistic metabolic platform that profiles both polar
535 and apolar components in colon cell lines and tissue. This method is the first one that has proven
536 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.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website.

Eight tables are included listing detailed statistical output to evaluate optimization parameters and validation results.

AUTHOR INFORMATION

Corresponding Author

* E-mail: Lynn.Vanhaecke@UGent.be. Tel: +32 9 264 74 57. Fax: +32 9 264 74 92.

Author Contributions

This research was funded by the Institute for the Promotion and Innovation through Science and Technology in Flanders (IWT, project n° 151535) Vlaanderen. 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.

ACKNOWLEDGMENTS

The authors wish to thank S. Bos and L. De Beule for their contributions to this work.

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, Proceedings of the National Academy of Sciences of the United States of America, 99 (2002) 9433-9438.

[3] O. Warburg, F. Wind, E. Negelein, The Metabolism of Tumors in the Body, The Journal of general physiology, 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 metabolism, 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 journal of surgical oncology, 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: Metabonomics 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, Analytical chemistry, 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. Shockcor, N. Loftus, E. Holmes, J.K. Nicholson, Global metabolic profiling of animal and human tissues via UPLC-MS, Nature protocols, 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, Molecular nutrition & food research, (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.

616 [19] J. Vanden Bussche, M. Marzorati, D. Laukens, L. Vanhaecke, Validated High Resolution Mass Spectrometry-
617 Based Approach for Metabolomic Fingerprinting of the Human Gut Phenotype, Anal Chem, 87 (2015) 10927-10934.
618 [20] L. Van Meulebroek, E. De Paepe, V. Vercruysse, B. Pomian, S. Bos, B. Lapauw, L. Vanhaecke, Holistic
619 Lipidomics of the Human Gut Phenotype Using Validated Ultra-High-Performance Liquid Chromatography Coupled to
620 Hybrid Orbitrap Mass Spectrometry, Anal Chem, 89 (2017) 12502-12510.
621 [21] E. De Paepe, L. Van Meulebroek, C. Rombouts, S. Huysman, K. Verplanken, B. Lapauw, J. Wauters, L.Y.
622 Hemeryck, L. Vanhaecke, A validated multi-matrix platform for metabolomic fingerprinting of human urine, feces and
623 plasma using ultra-high performance liquid-chromatography coupled to hybrid orbitrap high-resolution mass
624 spectrometry, Analytica chimica acta, 1033 (2018) 108-118.
625 [22] N. Zhao, Y. Liu, Y. Wei, Z. Yan, Q. Zhang, C. Wu, Z. Chang, Y. Xu, Optimization of cell lines as tumour models
626 by integrating multi-omics data, Brief Bioinform, 18 (2017) 515-529.
627 [23] S.C. Sapcariu, T. Kanashova, D. Weindl, J. Ghelfi, G. Dittmar, K. Hiller, Simultaneous extraction of proteins and
628 metabolites from cells in culture, MethodsX, 1 (2014) 74-80.
629 [24] E. Fahy, S. Subramaniam, R.C. Murphy, M. Nishijima, C.R. Raetz, T. Shimizu, F. Spener, G. van Meer, M.J.
630 Wakelam, E.A. Dennis, Update of the LIPID MAPS comprehensive classification system for lipids, J Lipid Res, 50
631 Suppl (2009) S9-14.
632 [25] S.K. Jensen, Improved Bligh and Dyer extraction procedure Lipid Technology Volume 20, Issue 12, Lipid
633 Technology, 2008, pp. 280-281.
634 [26] L.P. de Jonge, R.D. Douma, J.J. Heijnen, W.M. van Gulik, Optimization of cold methanol quenching for
635 quantitative metabolomics of Penicillium chrysogenum, Metabolomics, 8 (2012) 727-735.
636 [27] T. Cajka, O. Fiehn, Comprehensive analysis of lipids in biological systems by liquid chromatography-mass
637 spectrometry, Trends Analyt Chem, 61 (2014) 192-206.
638 [28] X. Liu, Z. Ser, J.W. Locasale, Development and quantitative evaluation of a high-resolution metabolomics
639 technology, Anal Chem, 86 (2014) 2175-2184.
640 [29] E.J. Want, I.D. Wilson, H. Gika, G. Theodoridis, R.S. Plumb, J. Shockcor, E. Holmes, J.K. Nicholson, Global
641 metabolic profiling procedures for urine using UPLC-MS, Nature protocols, 5 (2010) 1005-1018.
642 [30] C. Brunius, L. Shi, R. Landberg, Large-scale untargeted LC-MS metabolomics data correction using between-
643 batch feature alignment and cluster-based within-batch signal intensity drift correction, Metabolomics : Official journal
644 of the Metabolomic Society, 12 (2016) 173-186.
645 [31] B. Li, J. Tang, Q. Yang, X. Cui, S. Li, S. Chen, Q. Cao, W. Xue, N. Chen, F. Zhu, Performance Evaluation and
646 Online Realization of Data-driven Normalization Methods Used in LC/MS based Untargeted Metabolomics Analysis,
647 Sci Rep-Uk, 6 (2016) 38881-38894.
648 [32] C. Rombouts, L.Y. Hemeryck, T. Van Hecke, S. De Smet, W.H. De Vos, L. Vanhaecke, Untargeted metabolomics
649 of colonic digests reveals kynurenine pathway metabolites, dityrosine and 3-dehydroxycarnitine as red versus white
650 meat discriminating metabolites, Sci Rep-Uk, 7 (2017) 42514-42527.
651 [33] Z. Ser, X.J. Liu, N.N. Tang, J.W. Locasale, Extraction parameters for metabolomics from cultured cells, Anal
652 Biochem, 475 (2015) 22-28.
653 [34] C. Muschet, G. Moller, C. Prehn, M.H. de Angelis, J. Adamski, J. Tokarz, Removing the bottlenecks of cell
654 culture metabolomics: fast normalization procedure, correlation of metabolites to cell number, and impact of the cell
655 harvesting method, Metabolomics, 12 (2016) 151-163.
656 [35] R.V. Kapoore, R. Coyle, C.A. Staton, N.J. Brown, S. Vaidyanathan, Influence of washing and quenching in
657 profiling the metabolome of adherent mammalian cells: a case study with the metastatic breast cancer cell line MDA-
658 MB-231, Analyst, 142 (2017) 2038-2049.
659 [36] C.L. Winder, W.B. Dunn, S. Schuler, D. Broadhurst, R. Jarvis, G.M. Stephens, R. Goodacre, Global metabolic
660 profiling of Escherichia coli cultures: an evaluation of methods for quenching and extraction of intracellular
661 metabolites, Analytical chemistry, 80 (2008) 2939-2948.
662 [37] C.A. Sellick, D. Knight, A.S. Croxford, A.R. Maqsood, G.M. Stephens, R. Goodacre, A.J. Dickson, Evaluation of
663 extraction processes for intracellular metabolite profiling of mammalian cells: matching extraction approaches to cell
664 type and metabolite targets, Metabolomics, 6 (2010) 427-438.
665 [38] A. Lindahl, S. Saaf, J. Lehtio, A. Nordstrom, Tuning Metabolome Coverage in Reversed Phase LC-MS
666 Metabolomics of MeOH Extracted Samples Using the Reconstitution Solvent Composition, Analytical Chemistry, 89
667 (2017) 7356-7364.
668 [39] A. Ipsen, E.J. Want, J.C. Lindon, T.M. Ebbels, A statistically rigorous test for the identification of parent-fragment
669 pairs in LC-MS datasets, Anal Chem, 82 (2010) 1766-1778.
670 [40] H.G. Gika, I.D. Wilson, G.A. Theodoridis, LC-MS-based holistic metabolic profiling. Problems, limitations,
671 advantages, and future perspectives, J Chromatogr B Analyt Technol Biomed Life Sci, 966 (2014) 1-6.
672 [41] D. Saigusa, Y. Okamura, I.N. Motoike, Y. Katoh, Y. Kurosawa, R. Saijyo, S. Koshiba, J. Yasuda, H. Motohashi, J.
673 Sugawara, O. Tanabe, K. Kinoshita, M. Yamamoto, Establishment of Protocols for Global Metabolomics by LC-MS
674 for Biomarker Discovery, PloS one, 11 (2016) 160555-160573.
675 [42] C. Hellmuth, O. Uhl, M. Standl, H. Demmelmair, J. Heinrich, B. Koletzko, E. Thiering, Cord Blood Metabolome
676 Is Highly Associated with Birth Weight, but Less Predictive for Later Weight Development, Obesity facts, 10 (2017)
677 85-100.

678 [43] C.L. Gibson, S.G. Codreanu, A.C. Schrimpe-Rutledge, C.L. Retzlaff, J. Wright, D.P. Mortlock, S.D. Sherrod, J.A.
679 McLean, R.D. Blakely, Global untargeted serum metabolomic analyses nominate metabolic pathways responsive to
680 loss of expression of the orphan metallo beta-lactamase, MBLAC1, Molecular omics, 14 (2018) 142-155.
681 [44] M.V. Brown, J.E. McDunn, P.R. Gunst, E.M. Smith, M.V. Milburn, D.A. Troyer, K.A. Lawton, Cancer detection
682 and biopsy classification using concurrent histopathological and metabolomic analysis of core biopsies, Genome Med,
683 4 (2012) 33-45.
684 [45] S.K. Byeon, J.Y. Lee, M.H. Moon, Optimized extraction of phospholipids and lysophospholipids for nanoflow
685 liquid chromatography-electrospray ionization-tandem mass spectrometry, Analyst, 137 (2012) 451-458.
686 [46] S. Tumanov, J.J. Kamphorst, Recent advances in expanding the coverage of the lipidome, Curr Opin Biotech, 43
687 (2017) 127-133.
688 [47] E. Niki, Lipid peroxidation: physiological levels and dual biological effects, Free radical biology & medicine, 47
689 (2009) 469-484.
690 [48] H. Yin, L. Xu, N.A. Porter, Free radical lipid peroxidation: mechanisms and analysis, Chemical reviews, 111
691 (2011) 5944-5972.
692 [49] P. Leuthold, E. Schaeffeler, S. Winter, F. Buttner, U. Hofmann, T.E. Murdter, S. Rausch, D. Sonntag, J. Wahrheit,
693 F. Fend, J. Hennenlotter, J. Bedke, M. Schwab, M. Haag, Comprehensive Metabolomic and Lipidomic Profiling of
694 Human Kidney Tissue: A Platform Comparison, J Proteome Res, 16 (2017) 933-944.
695 [50] S. Bouatra, F. Aziat, R. Mandal, A.C. Guo, M.R. Wilson, C. Knox, T.C. Bjorndahl, R. Krishnamurthy, F. Saleem,
696 P. Liu, Z.T. Dame, J. Poelzer, J. Huynh, F.S. Yallou, N. Psychogios, E. Dong, R. Bogumil, C. Roehring, D.S. Wishart,
697 The human urine metabolome, PloS one, 8 (2013) 73076-73104.
698 [51] N. Psychogios, D.D. Hau, J. Peng, A.C. Guo, R. Mandal, S. Bouatra, I. Sinelnikov, R. Krishnamurthy, R. Eisner,
699 B. Gautam, N. Young, J. Xia, C. Knox, E. Dong, P. Huang, Z. Hollander, T.L. Pedersen, S.R. Smith, F. Bamforth, R.
700 Greiner, B. McManus, J.W. Newman, T. Goodfriend, D.S. Wishart, The human serum metabolome, PloS one, 6 (2011)
701 16957-16980.
702 [52] M.D. Williams, R. Reeves, L.S. Resar, H.H. Hill, Jr., Metabolomics of colorectal cancer: past and current
703 analytical platforms, Anal Bioanal Chem, 405 (2013) 5013-5030.
704 [53] M. Watanabe, S. Sheriff, K.B. Lewis, J. Cho, S.L. Tinch, A. Balasubramaniam, M.A. Kennedy, Metabolic
705 Profiling Comparison of Human Pancreatic Ductal Epithelial Cells and Three Pancreatic Cancer Cell Lines using NMR
706 Based Metabonomics, Journal of molecular biomarkers & diagnosis, 3 (2012) S3-002.
707 [54] C.L. Silva, R. Perestrelo, P. Silva, H. Tomas, J.S. Camara, Volatile metabolomic signature of human breast cancer
708 cell lines, Sci Rep-Uk, 7 (2017) 43969-43977.
709 [55] Z. Huang, W. Shao, J. Gu, X. Hu, Y. Shi, W. Xu, C. Huang, D. Lin, Effects of culture media on metabolic
710 profiling of the human gastric cancer cell line SGC7901, Molecular bioSystems, 11 (2015) 1832-1840.
711 [56] E. Daskalaki, N.J. Pillon, A. Krook, C.E. Wheelock, A. Checa, The influence of culture media upon observed cell
712 secretome metabolite profiles: The balance between cell viability and data interpretability, Analytica chimica acta,
713 1037 (2018) 338-350.
714 [57] A. Ertel, A. Verghese, S.W. Byers, M. Ochs, A. Tozeren, Pathway-specific differences between tumor cell lines
715 and normal and tumor tissue cells, Mol Cancer, 5 (2006) 55-68.
716 [58] P. Hughes, D. Marshall, Y. Reid, H. Parkes, C. Gelber, The costs of using unauthenticated, over-passaged cell
717 lines: how much more data do we need?, Biotechniques, 43 (2007) 575, 577-578, 581-572 passim.
718 [59] J. Le Beyec, R. Xu, S.Y. Lee, C.M. Nelson, A. Rizki, J. Alcaraz, M.J. Bissell, Cell shape regulates global histone
719 acetylation in human mammary epithelial cells, Exp Cell Res, 313 (2007) 3066-3075.

720

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%)

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 µL/1 mL)	1 pos. affected with 500 µL	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:** No. of selected metabolites (n = 27)	**Untargeted evaluation:** 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 (368xg /9,223xg)	0 affected	No effect
Validation	**Targeted evaluation:** No. of selected metabolites (n = 27)	**Untargeted evaluation:** 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%)

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

Optimization	**Targeted evaluation:** No. of selected metabolites (n = 28)	**Untargeted evaluation:** 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 (368xg /9,223xg)	0 affected	No effect
Validation	**Targeted evaluation:** No. of selected metabolites (n = 28)	**Untargeted evaluation:** 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%)

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	**N° 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	**N° 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

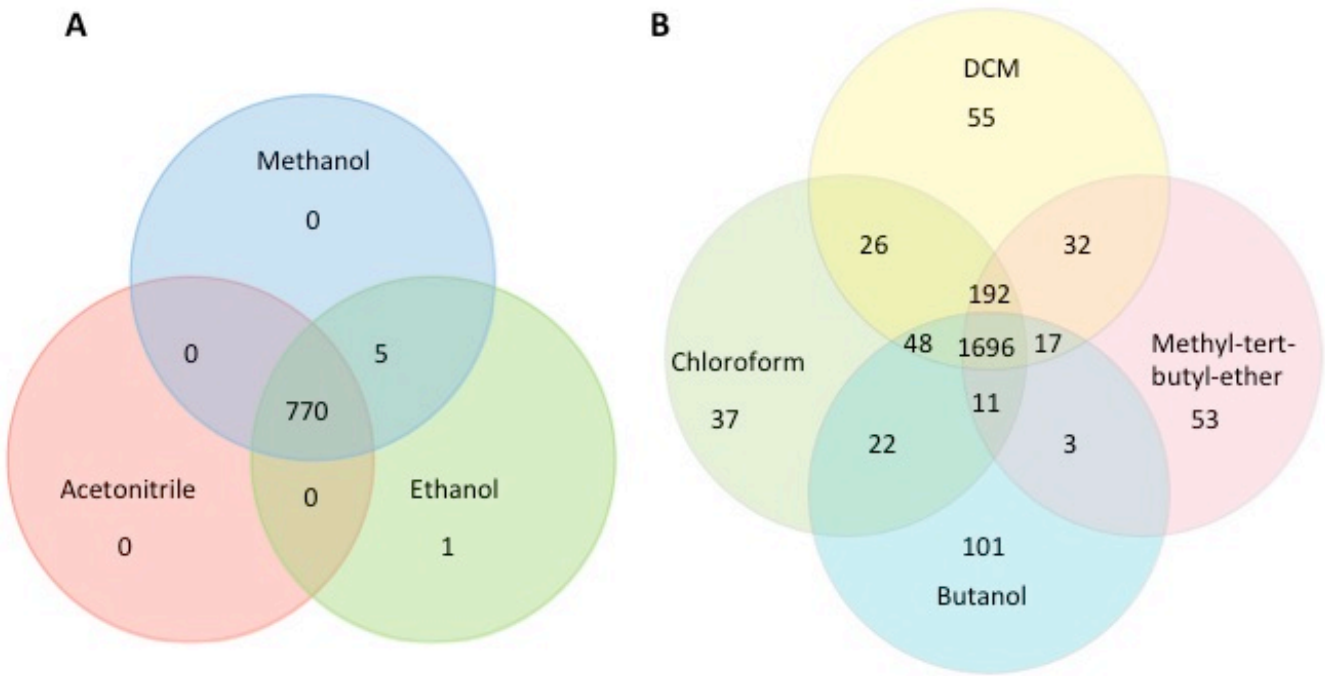

Figure 1. Venn diagram representing the number of detected components after extraction with different organic solvents during optimization of the polar metabolomics method (A) and the lipidomics method (B) for the HT-29 cell line.

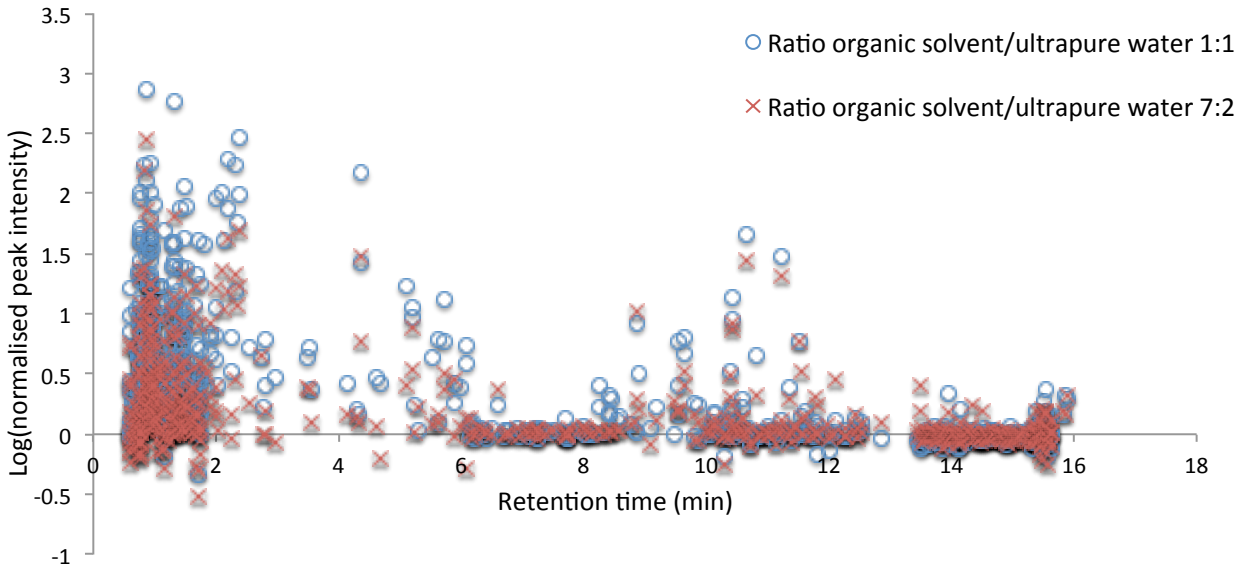

Figure 2. Scatter plot representing the results of the untargeted analysis after metabolomics extraction of the HT-29 cell line with different ratios of organic solvent/ultrapure water (1:1 or 7:2, v/v), whereby each point represent the log transformed normalized peak intensity of a component with a 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.

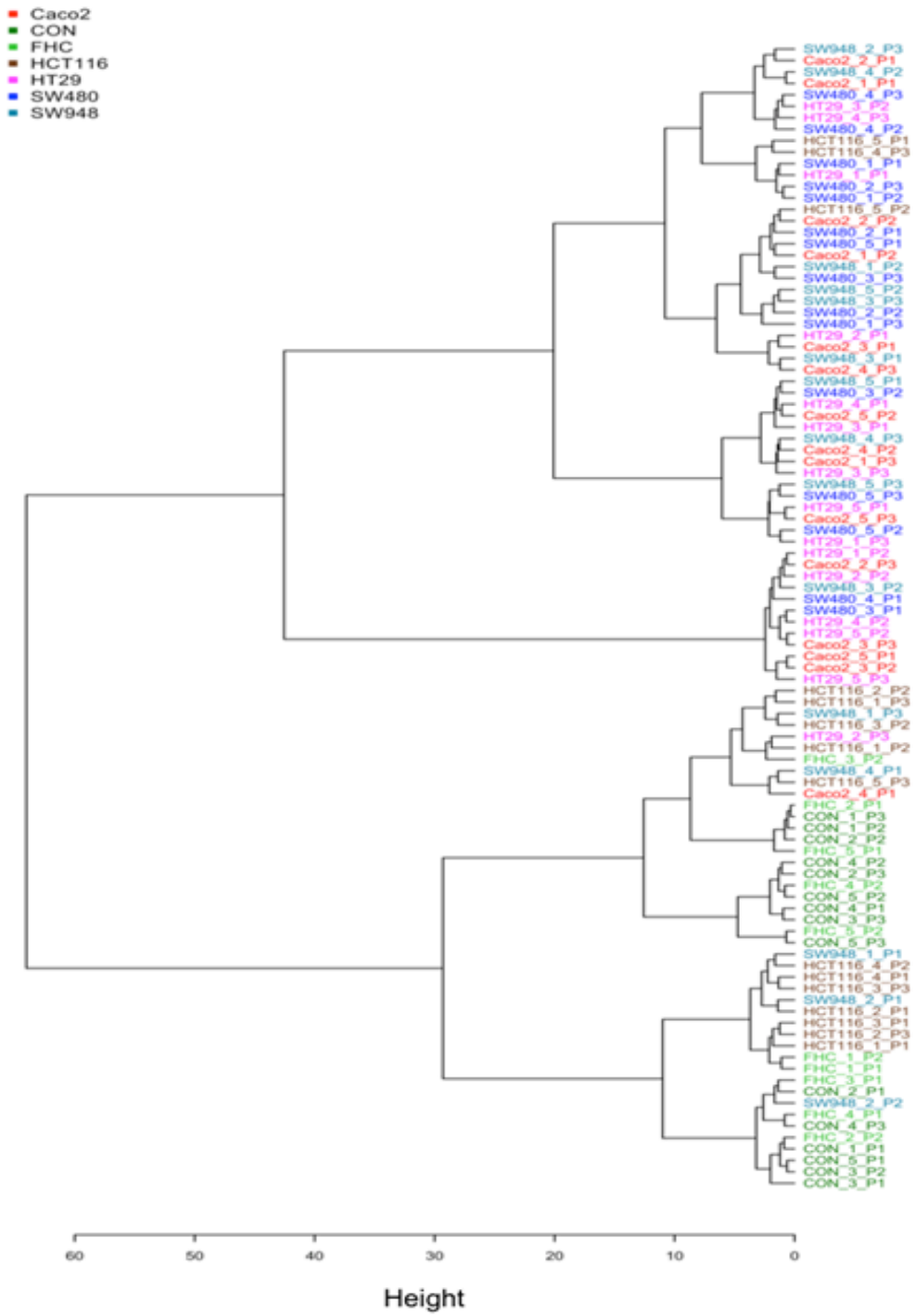

Figure 3. Hierarchical clustering of the technical replicates of the non-transformed (CCD CON841 and FHC) and transformed (Caco2, HT29, HCT116, SW480 and SW948) colon cell line samples using Pearson correlation coefficient and Ward linkage for the complete polar metabolome and lipidome datasets combined.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

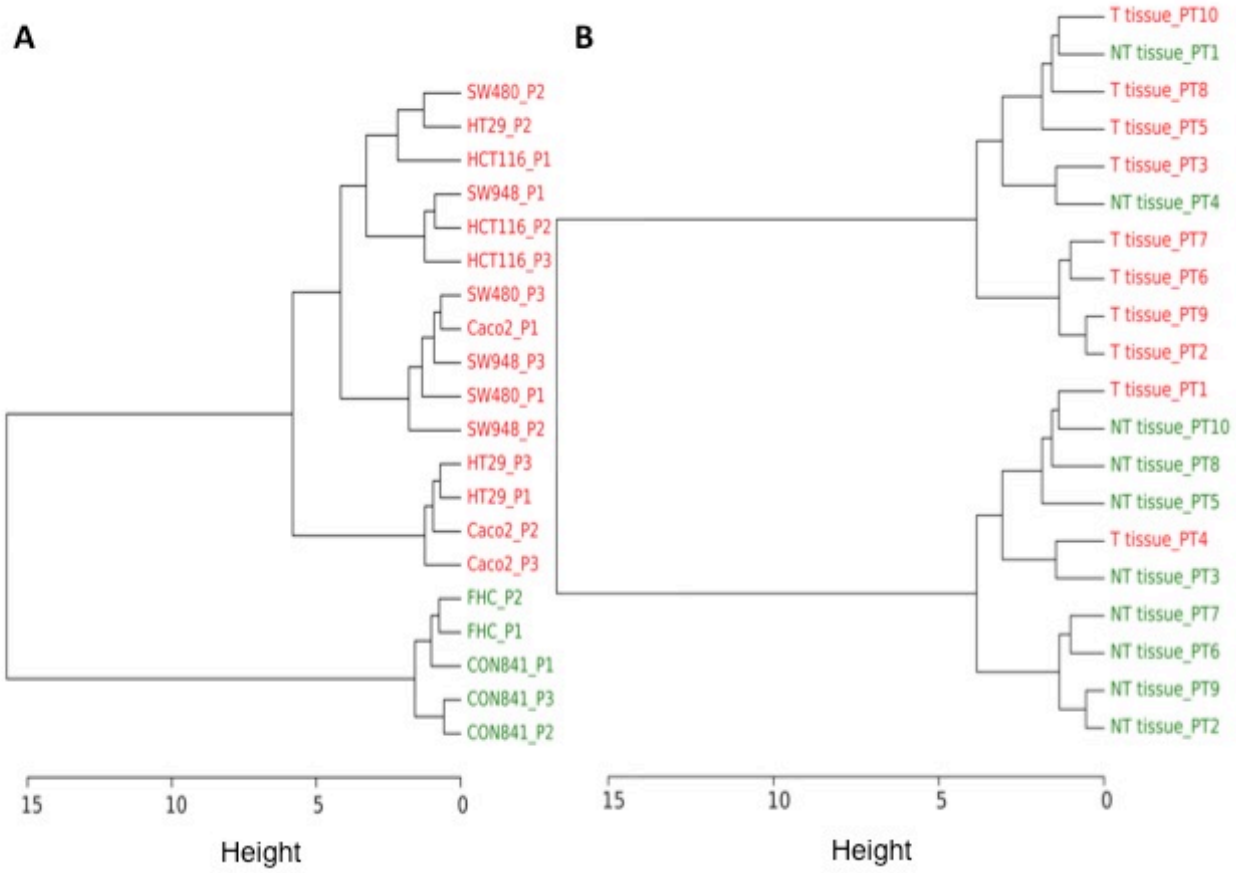

Figure 4. Hierarchical clustering of non-transformed (= green) and transformed (= red) colon cell line samples (A) and tissue (B) using Pearson correlation coefficient and Ward linkage for the complete polar metabolome and lipidome datasets combined. P1-3 = cell line passage 1-3 (1-week time interval), PT1-10 = colorectal cancer patient 1-10.

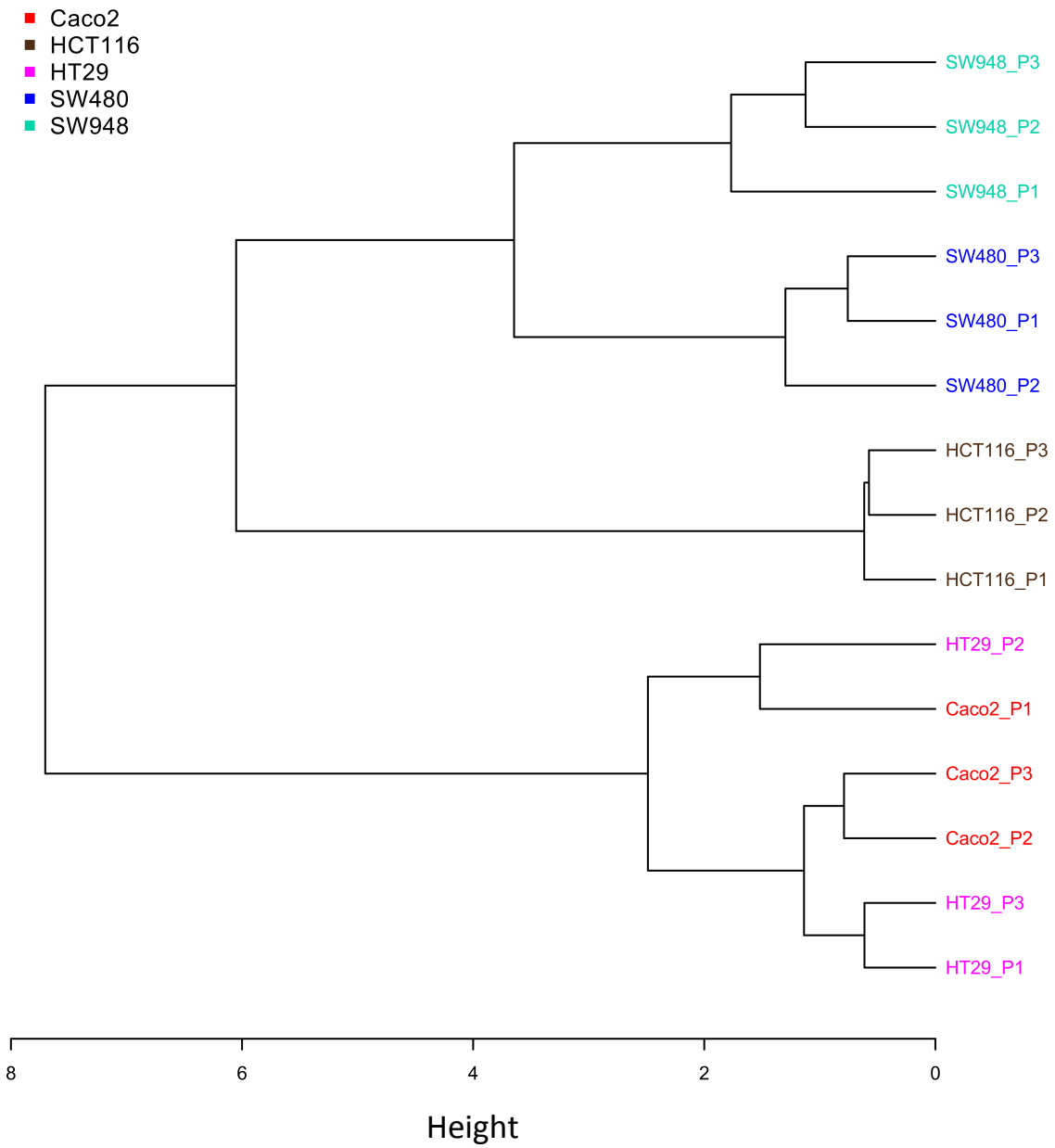

Figure 5. Hierarchical clustering of the transformed colon cell line samples using Pearson correlation coefficient and Ward linkage for the top 3000 components retained after filtering based on p-values (One-way Anova in MetaboAnalyst) for the complete polar metabolome and lipidome datasets combined.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

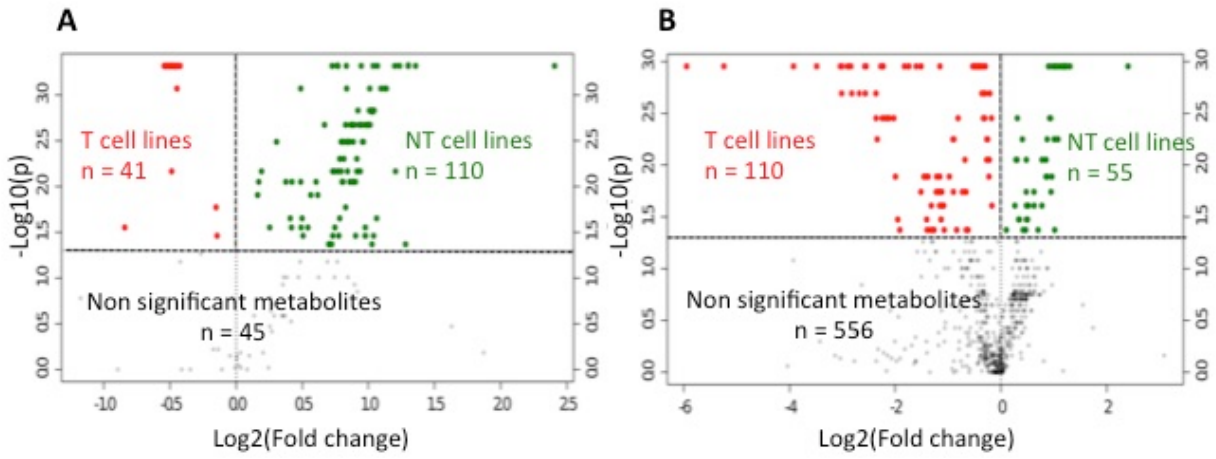

Figure 6. (A) Volcano plot displaying the behavior of the 196 components with VIP-value > 1.0, which were discriminating for the non-transformed (NT) state as opposed to the transformed (T) state at tissue level, in the cell line samples. Their up- or downregulation in the NT as opposed to the T cell 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 $p < 0.05$).