

OPEN

Comparative metabolomics of MCF-7 breast cancer cells using different extraction solvents assessed by mass spectroscopy

Mohammad H. Semreen^{1,2}, Hasan Y. Alniss^{1,2}, Stefan R. Grgic^{2,3}, Raafat A. El-Awady^{1,2}, Ahmed H. Almehdi⁴, Muath K. Mousa⁵ & Rifat A. Hamoudi^{2,6,7}

Metabolic profiling of cancer cells can play a vital role in revealing the molecular bases of cancer development and progression. In this study, gas chromatography coupled with mass spectrometry (GC-MS) was employed for the determination of signatures found in ER+/PR+ breast cancer cells derived from MCF-7 using different extraction solvents including: A, formic acid in water; B, ammonium hydroxide in water; C, ethyl acetate; D, methanol: water (1:1, v/v); and E, acetonitrile: water (1:1, v/v). The greatest extraction rate and diversity of metabolites occurs with extraction solvents A and E. Extraction solvent D showed moderate extraction efficiency, whereas extraction solvent B and C showed inferior metabolite diversity. Metabolite set enrichment analysis (MSEA) results showed energy production pathways to be key in MCF-7 cell lines. This study showed that mass spectrometry could identify key metabolites associated with cancers. The highest enriched pathways were related to energy production as well as Warburg effect pathways, which may shed light on how energy metabolism has been hijacked to encourage tumour progression and eventually metastasis in breast cancer.

Metabolic profiling of cancer cells and biomarkers identifications has drawn the attention of many researchers to understand the complexity and diversity of cancer cell biology. Metabolomics has helped researchers to reveal specific biomarkers for many diseases, and that might improve both personalized therapies and clinical outcomes after primary diagnosis¹. Breast cancer is the second leading cause of death and the most common cancer in women globally² and while there are many screening methods such as mammography³, Breast Self-Examination, and Clinical Breast Examination. However, high throughput and advanced molecular platforms are essential for the detection and characterization of novel biomarkers for disease-associated metabolic disorders and improved early diagnostic methods are desirable. Breast cancer is highly heterogeneous and consists of at least 10 different disease sub-types⁴. Breast cancer involve many different cellular pathways including metabolic and hypoxia related pathways such as the Warburg effect related pathways⁵. Warburg effect is when cancer cells keep a high glycolytic rate even in aerobic conditions indicating that glycolysis is vital for cancer survival and progression⁶.

Metabolomics is a promising analytical tool which is described as a large scale investigation of small molecules (metabolites) in a biological sample that are characterized using mass spectrometry or other suitable biophysical techniques. The metabolome is defined as the metabolites and their interactions with their targets within a biological system⁷. Due to the complexity of metabolome, there are three main analytical techniques currently applied in metabolomics: mass spectrometry (LC-MS)⁸, gas chromatography-mass spectrometry (GC-MS)^{9,10}, and nuclear magnetic resonance (NMR)¹¹⁻¹³. The huge advancement in the applications of these modern analytical techniques allow the analysis of large metabolomics data at different molecular levels, including the organism,

¹College of Pharmacy, University of Sharjah, P.O. Box 27272, Sharjah, United Arab Emirates. ²Sharjah Institute for Medical Research, University of Sharjah, P.O. Box 27272, Sharjah, United Arab Emirates. ³Faculty of Medicine, Department of Pharmacy, University of Novi Sad, P.O. Box 21001, Novi Sad, Serbia. ⁴Department of Chemistry, College of Sciences, University of Sharjah, P.O. Box 27272, Sharjah, United Arab Emirates. ⁵Research Institute of Science and Engineering, University of Sharjah, P.O. Box 27272, Sharjah, United Arab Emirates. ⁶College of Medicine, University of Sharjah, P.O. Box 27272, Sharjah, United Arab Emirates. ⁷Division of Surgery and Interventional Science, University College London, London, United Kingdom. Correspondence and requests for materials should be addressed to M.H.S. (email: msemreen@sharjah.ac.ae) or R.A.H. (email: rhamoudi@sharjah.ac.ae)

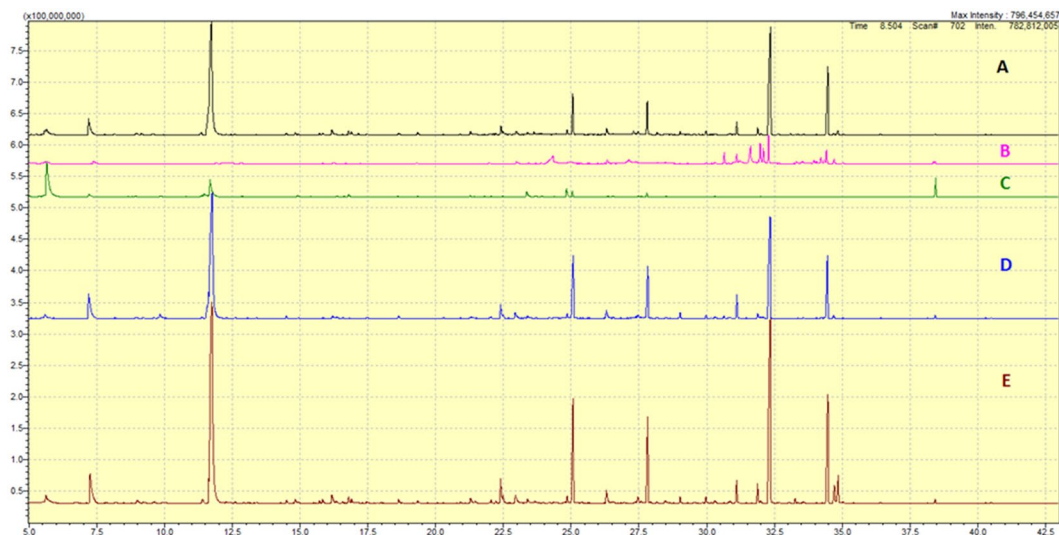


Figure 1. GC-MS chromatograms of the extracted cells generated using: (A) 0.2% formic acid in water; (B) 0.2% ammonium hydroxide in water; (C) ethyl acetate; (D) methanol/water (1:1, v/v); and (E) acetonitrile/water (1:1, v/v).

organ, tissue, cell and organelle levels. Such data can provide critical information about biomolecular function and the biochemical relationships of metabolites, which might lead to a better understanding of the biological system under investigation.

Gas chromatography–mass spectrometry (GC-MS) is the most commonly used and standardized technique in metabolomics, with more than 50 years of established protocols for metabolite characterization of amino acids¹⁴, hydroxyl acids¹⁵, fatty acids¹⁶, catecholamines¹⁷, sugars¹⁸, hormones¹⁹, and many other metabolic intermediates.

In the last four decades, mass spectra have been accumulated and stored in libraries under standard conditions of 70 eV electron ionization energy^{20–22}. More importantly, tremendous efforts have been made to computationally match mass spectra with experimental data to help in the identification of compounds via their unique fragmentation pattern^{23–25}.

In comparison, GC-MS/MS libraries are more comprehensive and updated compared to LC-MS/MS libraries. Furthermore, peak picking in GC is associated with true mass spectral deconvolution, which summarizes all fragment ions into a purified mass spectrum, whereas in LC-MS methods, the MS and data dependent MS-MS fragmentations are used separately. This was facilitated by the availability of an automated mass spectral deconvolution software (AMDIS), which has been successfully used in GC-MS for metabolomics studies since 1998^{26–28}. GC-MS is more selective, robust, rapid and easier to identify metabolites due to the availability of ready-made libraries. Therefore, it can be said that GC-mass spectrometry provides higher resolution for metabolite separation.

Tumour cells have altered metabolism, therefore their metabolic mapping is considered a promising tool in cancer diagnostics²⁹. Different pathways are subjected to alteration within the cancer cells during the development of the disease, which might affect the metabolism of cancer cells compared to the normal cells^{30–32}. Among the affected pathways is glycolysis, and its end product pyruvate, represents a key intermediate in several metabolic processes linked to energy production such as carbohydrate and fatty acid, which are important components to cell proliferation^{33,34}. However, little has been reported on the quantitation of metabolic intermediates involved in energy metabolism (such as metabolites of glycolysis, Krebs cycle and oxidative phosphorylation) in cultured breast cancer and normal cells^{35,36}. This is mainly due to the fact that most metabolites involved in energy metabolism are characterized by high polarity, low volatility, and thus poor detectability.

In this study, we explore the optimal metabolite extraction protocol from MCF-7 breast cancer cells³⁷ and the subsequent metabolic profiling utilizing gas chromatography coupled with tandem mass spectrometry (MS). GC-MS analysis of cultured MCF-7 tumour cells provided detailed information about their metabolic profiling.

Results

Assessment of metabolic extraction. The GC/MS chromatograms of the extracted cells using the five extraction solutions under investigation are shown in Fig. 1, the greatest extraction rate and diversity of metabolites is noticed when employing 0.2% formic acid and acetonitrile: water mixture, whereas 0.2% ammonium hydroxide and ethyl acetate showed inferior metabolite diversity, whereas methanol: water showed moderate extraction efficiency.

A total of 395 metabolites were detected and characterized in the extracts under investigation, as shown in Fig. 1 and Table 1.

Among the aqueous-based extraction solutions A (acidified with 0.2% formic acid) and B (0.2% ammonium hydroxide), A showed higher extraction efficiency in terms of metabolites detection and identification with overall coverage of 111 against 41. According to the data presented in Table 1, extraction solution E (acetonitrile/water), A (acidified with 0.2% formic acid) and D (methanol/water) demonstrated greatest superiority in overall

Metabolic Extract	Extraction Solvent	Detected metabolites	Identified metabolites	# of metabolites with RSD < 15%
A	0.2% formic acid in water (FA)	111	103	60
B	0.2% ammonium hydroxide in water (NH ₄ OH)	41	36	16
C	ethyl acetate (E.A)	76	66	15
D	methanol/water (MeOH: H ₂ O)	93	83	41
E	acetonitrile/water (MeCN: H ₂ O)	120	107	69

Table 1. Summary of metabolic extraction yield of the investigated extraction solutions. The data was generated from the analysis of triplicate samples. Supplementary Tables (S1–S5) has details of the solvent metabolite extraction data.

metabolome coverage, with number of metabolites identified 107, 103 and 83 respectively. The identification cut-off was based on the signal: noise ratio of the peaks. The reproducibility of the extraction efficiency was evaluated by calculating the RSD values of the identified metabolites. The results showed that the highest reproducibility was achieved using solvents E and A (RSD < 15% in 69 and 60 of the metabolites identified in E and A, respectively), while solvent D provided moderate reproducibility (RSD < 15% in 41 metabolites), and low reproducibility was observed with solvents B and C (RSD < 15% in 16 and 15 metabolites for B and C respectively).

Variety of fatty acids, hydroxyl acids, saccharides and their metabolites (glycolysis, Krebs cycle, polyol pathway) were detected by employing most of the solvents, excluding 0.2% ammonium hydroxide and ethyl acetate. However, optimal chromatographic conditions (resolution and peak intensity) were achieved by employing extraction solution A. Ribitol was found to be efficiently extracted with solution C (ethyl acetate) and D. Surprisingly, non-conventional metabolites (Stigmasterol and its metabolite) were detected and identified in the MCF-7 cells.

Five metabolites were identified and detected in all extraction solvents used in this study at different concentration levels. This includes Inositol sugar, which is a polar and neutral molecule; this may explain its existence in aqueous and organic solvents. Moreover, two fatty acids (Palmitic and stearic acids) were found in all solvents used in this study at different concentration levels and this is mainly related to the coexistence of two states (ionized and non-ionized) of these fatty acids, which is pH dependent.

Two molecules were detected (Octasiloxane-hexadecamethyl and Heptamethyl-3,3-bis(trimethylsiloxy)tetrasiloxane) in all extraction solvents, however, these species were generated as a by-product of the derivatization process, which is an essential step to enhance the lipophilicity of metabolites to improve peak symmetry and metabolite GC detection.

Venn diagram representation in Fig. 2, confirmed that the highest overlap of metabolite identification was achieved based on the polarity of the extraction solvents. In this case, the acetonitrile:water extracted 120 metabolites and formic acid (0.2%) extracted 111 metabolites compared to the other solvent extractions.

Association between different extraction methods on MCF-7 breast cancer cells. Metabolite data obtained from the five extraction protocols were subjected to unsupervised hierarchical clustering to identify patterns and clustering of metabolite recovery generated from the different extraction methods. Figure 3 shows the clustering patterns of each metabolic extract based on the relative amount of each metabolite in the extracted solutions (blue show decrease and red show increase in the metabolite concentration).

The clustering information in Fig. 3 show the five extracts are differentiated based on the polarity of the different extraction solvents used in this study. Two clear clusters, as predicted, were therefore noticed corresponding to ammonium hydroxide (NH₄OH) & ethyl acetate (cluster 1), and formic acid (FA), methanol: water, and acetonitrile: water (cluster 2). Taken together, the clustering showed that MeOH: water, MeCN:water and FA (0.2%) give closely related metabolite signature which is different from the closely related NH₄OH and ethyl acetate solvent extraction group. This show that the solvent extraction method can identify novel metabolites that may not be detected through other methodologies. This information was also validated by subjecting the metabolite data to Principle Component Analysis (PCA) (Fig. S1) and Venn diagram (see Fig. 2). Also, PCA revealed that the metabolic information generated from the aqueous solutions were more similar (i.e. less variable) than the other solvents (Fig. S2).

Pathways of MCF-7 breast cancer metabolome identified by GC-MS. The metabolites of (glycolysis, TCA and gluconeogenesis) were extracted and identified in most of the solvents used in this study, however, the concentration of the detected metabolites involved in these pathways were significantly affected by the polarity as well as the pH of the extraction solvents, for instance polar metabolites are more efficiently extracted in aqueous solution and the polarity of the metabolites is also affected by their ionization state which is pH and pKa dependent. However, more lipophilic metabolites of lipids and fatty acids tend to be more efficiently extracted in organic solvents (ethyl acetate), however, the ionization of these metabolites is significantly promoted in alkaline medium (NH₄OH) which justify the presence of these metabolites in this aqueous solvent.

MSEA results show enrichment in metabolic pathways for both extraction solvent groups with NH₄OH and ethyl acetate showing higher enrichment of the Warburg effect (p = 0.03) as shown in Fig. 4 and Tables (S6 & S7). In fact, Warburg effect is the second most enriched pathway in metabolism.

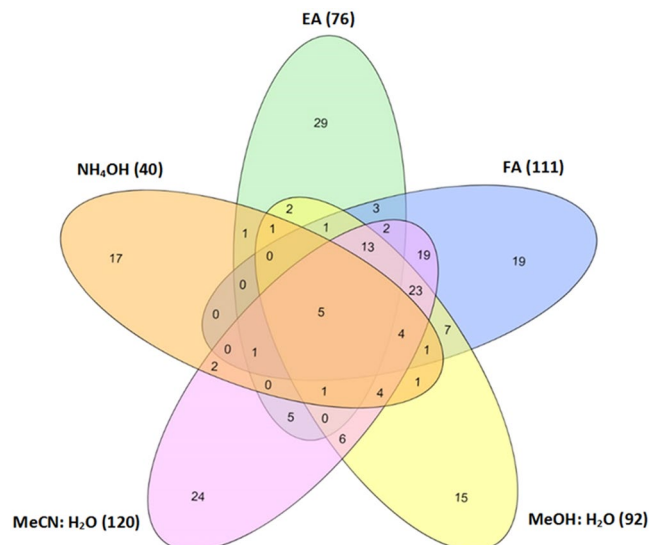


Figure 2. A Venn diagram showing the detected metabolites in ammonium hydroxide (NH₄OH), ethyl acetate (EA), Formic acid (FA), methanol: water (MeOH: H₂O) and acetonitrile: water (MeCN: H₂O) extracts.

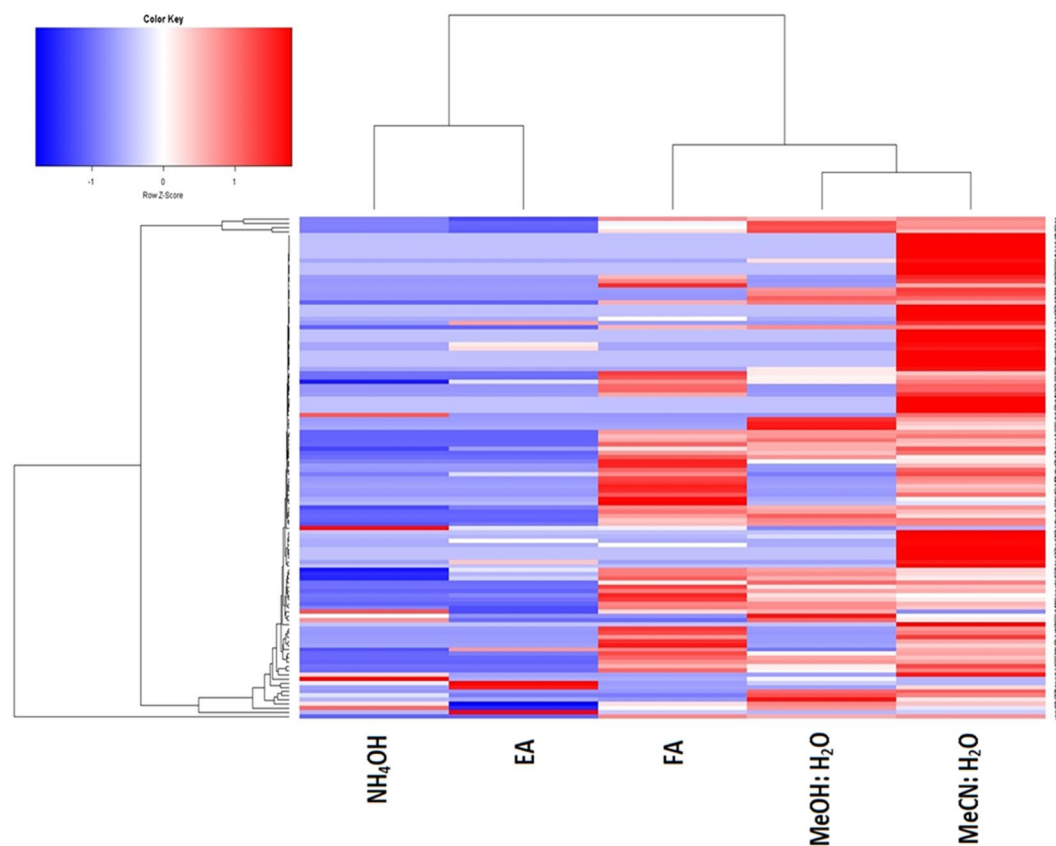


Figure 3. Unsupervised hierarchical clustering and heatmap of the identified metabolites in the extracted samples (rows) through different solvents (columns). NH₄OH (B), ethyl acetate (C), Formic acid (A), methanol: water (D) acetonitrile: water (E). Cell color reflects metabolite relative content. The metabolites that correspond to the row numbers shown on the heat map listed in Supplementary Table 8.

However in the FA, MeOH: H₂O and MeCN: H₂O group the Warburg is in the top 10 (Fig. 5). For metabolic pathways the top enriched pathways are related to glycolysis, TCA and oxidative phosphorylation pathways, all of which are related to energy production and regulation and hence enhancement of the tumour growth.

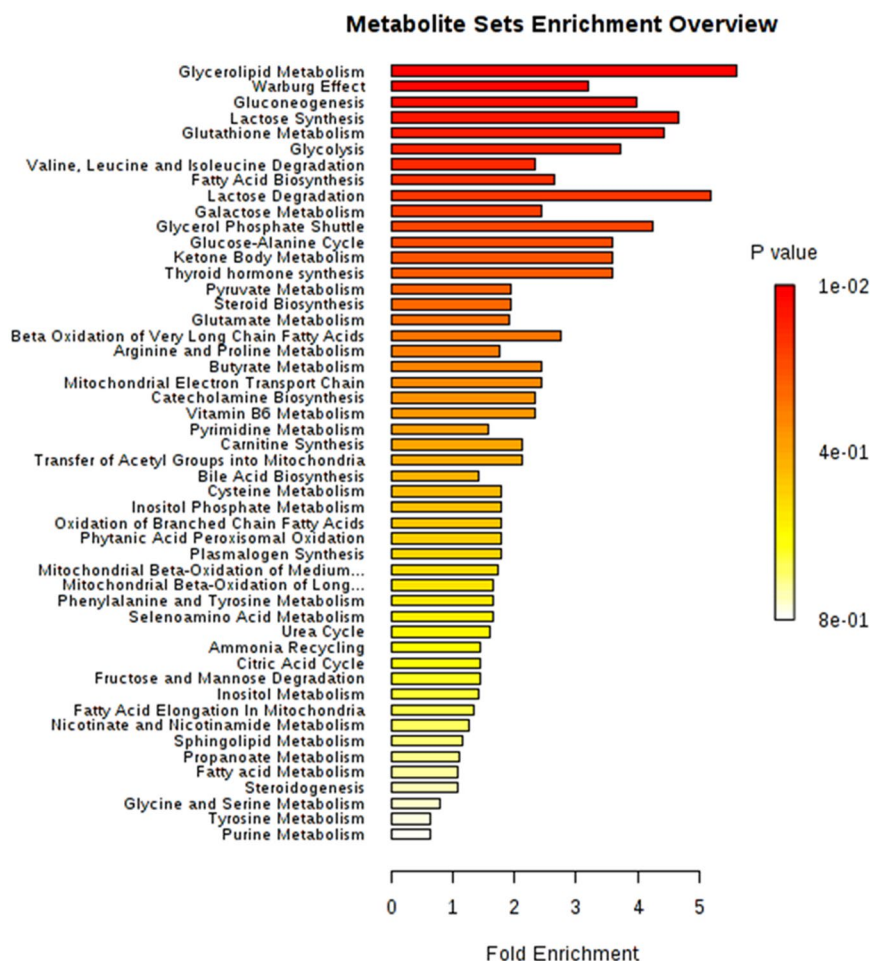


Figure 4. A diagram showing the MSEA for the identified metabolites in cluster 1: NH_4OH and ethyl acetate.

Pathways of MCF-7 breast cancer metabolome identified by GC-MS. Mapping of the metabolites from GC-MS to the genome SNP database, provided us with useful data regarding the possible genes involved with the metabolites. The enrichment pathways showed the involvement of Dynein cytoplasmic 2 heavy chain 1 gene (*DYNC2H1*), which is probably involved in protein transport via the chemi-osmotic potential and thus ATP production³⁸. In addition other genes enriched included *EVC*: Ellis-van Creveld syndrome (EvC) and *Weyers* acrodermal dysostosis which is linked to hedgehog pathway and is therefore related to Wnt signaling pathway involved in many cancers.

In addition, one of the genes identify through the rs1018503 SNP is *NCOA6* (Nuclear Receptor Coactivator 6) which amongst its related pathways are *BMAL1-CLOCK*, *NPAS2* that activates circadian gene expression³⁹.

This in turn activate the transcription of *AR* (androgen receptor) which regulate *KLK2* and *KLK3* genes amongst others. Gene Ontology (GO) annotations related to *NCOA6* gene include chromatin binding and transcription coactivator activity.

NCOA6 has been shown to be associated with Breast Cancer cells⁴⁰ which validates the finding from this study and show an interesting novel finding that relate metabolomics to specific genes related to breast cancer which may not be identified through genomic techniques and thus warrant further investigation.

Discussion

Metabolomics is an important diagnostic tool in tackling the machinery of tumor progression. Human breast cancers have a broad array of causes that modulate growth and metastasis. Therefore, biomarkers identification involved in certain pathways might reveal intrinsic causes of cancer and chemotherapy resistance. In this study, the solvent extraction efficiency of metabolites in MCF-7 cancer cells was evaluated.

The extraction results showed that 0.2% formic acid and acetonitrile: water mixture solutions, have showed the best extraction efficiency, whereas solutions B, C, and D (Table 1) showed inferior to moderate extraction efficiency. These differences in metabolite extraction efficiency might be due to both polarity and pH values of the extraction solvents, as well as the pKa's of the metabolites. Polarity of the extraction solutions decreases in following order: aqueous solutions > methanol > acetonitrile > ethyl acetate, whereas the pH value decreases according to this order: 0.2% ammonium hydroxide in water > ethyl acetate > methanol > acetonitrile > 0.2% formic acid in water. Solution A (FA 0.2%) and E (acetonitrile) showed the highest extraction efficiency in terms of both metabolites detection and identification (111 and 120 metabolites were detected respectively). This can be

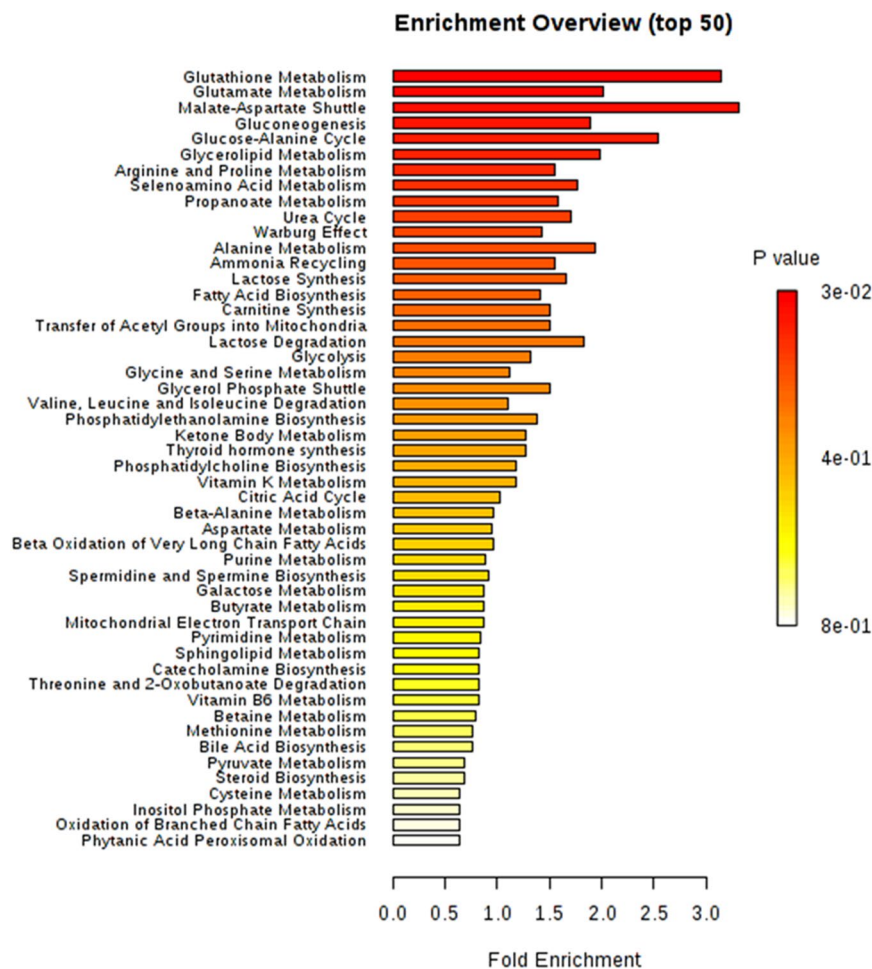


Figure 5. A diagram showing the MSEA of the identified metabolites in cluster 2: FA, MeOH: H₂O and MeCN:H₂O solvents.

explained by the biochemical properties of cancer cells themselves. Several studies have indicated that one of the characteristics that distinguishes cancer cells from normal cells is their low intracellular pH value^{41,42}, indicating the existence of chemically stable cancer-cell-specific metabolites. Therefore, acidified extracellular media, i.e. solvent with lower pH value, which in our case is the extraction solution A, provided greater extraction yield, due to the ionization of basic metabolites at low pH values or the existence neutral and polar metabolites, thus enabling the optimal conditions for diffusion of cancer-cell-specific metabolites to occur in the aqueous acidic medium. Moreover, a similar rationale can explain the extraction efficiency of acetonitrile: water solution, which is a slightly basic and polar solution, favoring mainly the extraction of polar neutral and acidic metabolites in higher concentrations. Certain metabolites have been extracted when solution B (0.2% ammonium hydroxide) was employed and this can be explained by the dominant existence of ionized acidic metabolites at alkaline pH. However, the extracted metabolites did not provide any significant information about the specific metabolic pathways occurring in MCF-7 cells. Moreover, each of the chromatograms and fragmentation patterns of the extracted solutions needs to be profoundly investigated in terms of metabolite peak intensity and characterization to evaluate the solvent effect, reproducibility and extraction efficiency.

Inferiority of the extraction solution C is based on the fact that the metabolites are mainly polar in their nature and the extraction is favored in aqueous medium, which is not the case for ethyl acetate as an organic solvent. However, ribitol which is a major metabolite of the pentose phosphate pathway in MCF-7, was efficiently extracted in the solutions C and D.

Surprisingly, non-conventional metabolites (Stigmasterol and its metabolite) were identified in the MCF-7 cells, and these metabolites are not regularly found in healthy normal cells. In this specific case, the detection of Stigmasterol (phytoestrogen) and its' metabolite implies an alteration in the cholesterol overall metabolism.

This study showed that mass spec can identify key metabolites that are associated with different diseases. In this case, it identified the NCOA6 gene as one of the genes linked to metabolites, which activates circadian gene expression and downstream-related pathways. However, this gene is not commonly associated with breast cancer but a recent paper⁴³ showed that it regulates the expression of kallikreins 2 and 3 in human breast cancer cell lines and can therefore be thought of as low penetrance gene that can explain one of the environmental link to breast cancer.

In addition to that, all the metabolites identified showed link to glycolysis and therefore may shed light on how energy metabolism has been hijacked to encourage tumour progression and eventually metastasis in breast cancer. Other studies reported on similar phenomena in various cancers including breast⁴⁴. However, this study showed that in addition to the energy metabolism pathways, it identified the Warburg effect as a key component in MCF-7 cell lines indicating that hypoxic conditions are key to its survival and proliferation. The Warburg effect can be induced by mitochondrial dysfunction or mutations in essential glycolytic enzymes such as 6-phosphofructo-1-kinase (PFK-1) triggered by hypoxia via the stimulation of the HIF-1 complex (hypoxia-inducible complex-1). HIF-1 complex is a transcriptional modulator that regulates the expression of most of hypoxia-related genes and is overexpressed in a variety of tumours. In addition, it is worth investigating the role of other genetic biomarkers identified through the GC-Mass spectroscopy data from this study, such as NCOA6 which may identify some of the environmental links to breast cancer such as stress or abnormal sleep patterns.

Taken together, using GC-Mass spectroscopy approach provides a different perspective in shedding light on the mechanism of breast cancer and identifying some of the genes that are associated with breast cancer pathogenesis (such as NCOA6) that could potentially be biomarkers for breast cancer.

Conclusions

This study showed that GC-MS could identify key metabolites that are associated with metabolic disorders and proved that the extraction efficiency is not limited to one specific solvent, due to the differences in physicochemical properties of the extracted metabolites.

Among the five solvents that have been used to investigate the extraction efficiency in this study, the greatest extraction rate and diversity of metabolites was obtained by employing 0.2% formic acid (solvent A) and acetonitrile: water mixture (solvent E). The fact that solvent A and E provided high reproducibility, selectivity and sensitivity for the extracted metabolites, is encouraging to exploit this efficient extraction methodology in future quantitative investigations on MCF-7 cancer cells. Our future plan is therefore to investigate the effects of treating triple positive and triple negative MCF-7 cells with different anticancer agents to characterize malfunctioning metabolic pathways.

Furthermore, the majority of the identified metabolites are linked to glycolysis and therefore this may shed light on how energy metabolism could be hijacked to encourage tumour progression and eventually metastasis in breast cancer.

Moreover, in addition to the energy metabolism pathways, the MSEA identified Warburg effect as a key component in MCF-7 breast cancer cells indicating that hypoxic conditions play a role in their survival and proliferation. On the other hand, the ability of this extraction methodology to reveal Warburg effect, which has been previously reported in the literature, proves the efficiency, selectivity and applicability of the developed extraction methodology.

Finally, mapping of the identified metabolites to the genome SNP database, identified NCOA6 gene as a key component involved in the activation of circadian gene expression and downstream-related pathways.

Materials and Methods

Metabolite extraction solvents were of LC-MS grade. Formic acid and methanol were obtained from Fisher Chemical (Loughborough, UK), whereas acetonitrile and ammonium hydroxide (30–33% grade) were obtained from Sigma Aldrich (Sr. Louis, MO, USA).

Reagents for derivatization included methoxyamine hydrochloride (98%) obtained from Sigma Aldrich, pyridine acquired from Merck (Darmstadt, Germany) and N-Methyl-N-(trimethylsilyl) trifluoroacetamide with 1% trimethylchlorosilane purchased from Sigma Aldrich.

As an internal standard for GC-MS analysis, Column Test Mix, containing methyl decanoate, methyl undecanoate, 2,3-butanediol, dicyclohexylamine, 2,6-dimethylaniline, 2,6-dimethylphenol, nonanal, 1-octanol, undecane, decane, hexachlorobenzene dissolved in n-hexane manufactured by Varian (Walnut Creek, CA, USA) was used for mass calibration.

Cell line and culture condition. The breast adenocarcinoma cells (MCF-7) were cultured in DMEM medium supplemented with 10% fetal calf serum and 1% penicillin/streptomycin solution and incubated at 37 °C in an atmosphere of 5% CO₂. In preparation for an experiment, 3 × 10⁶ cells were cultured in each of 15 tissue culture flasks (175 cm²) and cells were incubated for three days. When the cells reach 80% confluency, they were collected by trypsinization, counted in a cell counter and each 3 × 10⁶ cells were suspended in 1 ml phosphate-buffered saline in an Eppendorf tube.

Metabolite extraction and derivatization. A total of 5 triplicated MCF-7 cell culture samples were provided in Eppendorf vials dissolved in PBS, and stored at 4 °C for preservation purposes. Samples were centrifuged at 13000 rpm for 10 minutes at −4 °C. Supernatant was discarded, and cell pellets, each containing 3 million cells, were subjected to metabolomics analysis.

In order to evaluate the influence of the solvents on the extraction rate, samples were divided in five different extraction groups: A, 0.2% formic acid in water; B, 0.2% ammonium hydroxide in water; C, ethyl acetate; D, methanol/water (1:1, v/v); and E, acetonitrile/water (1:1, v/v).

Metabolic content extraction of the adenocarcinoma cells (MCF-7), applied in this research, was based on optimized method by our research group. Briefly, 300 µL of the extraction solvent was added to 3 million cell pellets, then vortexed for 2 minutes. To ensure the quantitative extraction of the metabolites, samples have been stored in ice for 1 hour, during which samples have been vortexed every 15 minutes. After this, cell insoluble matrix was centrifuged (13000 rpm, 10 minutes, −4 °C). Supernatant was collected and transferred to GC vial,

then dried using EZ-2 Plus (GeneVac, Ipswich, UK) at $37 \pm 1^\circ\text{C}$. Amino acids, saccharides, fatty acids and steroids cannot be analysed directly by gas chromatography due to their high polarity and low volatility. Hence, it was necessary to derivatize them prior to the GC-MS analysis. Having said that, dry samples were dissolved in 25 μL of a solution consisting 20 mg/mL methoxyamine hydrochloride in pyridine, followed by vortexing for 2 minutes, and storing for at least 6 hours at 25°C prior to the silylation step. Then 25 μL of MSTFA + 1% TMCS was added, followed by dissolving in 100 μL of pyridine, and vortexing for 2 minutes. For complete derivatization, it was necessary to incubate the samples at 50°C for 30 minutes. Due to low volume, all of the samples have been immediately transferred to 200 μL microinserts, and analysed by GC-MS.

Gas chromatography – mass spectrometric analysis of the samples. A GC/MS-QP 2010 Ultra System (Shimadzu, Kyoto, Japan) was employed for the metabolomic analysis, along with LabSolutions GC-MS software (v1.20). A Restek Rtx[®] – 5ms column (30.0 m \times 0.25 mm, 0.25 μm) was utilized for separation of the metabolites. Helium (99.9%) was utilized as the carrier gas at the flow rate of 1.0 mL/min. The initial oven temperature was set at 60°C and was held at this temperature for 2 minutes, then raised to 310°C by $50^\circ\text{C}/\text{min}$ and held at this temperature during the analysis. Both the auxiliary temperature at the interface and the ionization temperature were 250°C . Metabolites were analysed in full scan mode within the range of 50–650 amu. Total volume of 10 μL was injected in splitless mode, by employing AOC-20i Auto Injector (Shimadzu, Kyoto, Japan). GC total ion chromatograms (TIC) and fragmentation patterns of the metabolites identified using NIST/EPA/NIH Mass Spectral Library (NIST 14). Run time for each sample was 43.67 min.

Bioinformatics and functional genomic. Gas chromatography data was acquired as GCMS Data File (QGD) files and stored in text format. Mining of the metabolites data was carried out using various bioinformatics algorithms followed by mapping to genomics to acquire functional genomics data. All data including the raw QGD files, has been deposited to Metabolomics Workbench (<https://www.metabolomicsworkbench.org>). The data track id is 1759.

Unsupervised hierarchical clustering. The mass signal from the GCMS data was manually filtered to remove unidentified ions from the table. Signals that belong to the same category of ions were grouped together and the data normalised based on the total peak height. The table with metabolic data from all extraction replicates generated from the five solvents (triplicate for extraction) was then assessed in detail. Principle Component Analysis (PCA) was implemented using R programming language (version 3.5.2) using the *prcomp* library. Unsupervised hierarchical clustering was carried out using R programming language to explore possible grouping and patterns in metabolic coverage generated from the five extraction solutions.

MSEA – functional metabolomics. Metabolome coverage between the five different extracts was conducted using both MSEA and network-based metabolic pathway analysis (MPA)⁴⁵. MSEA tests, using over-representation analysis, the enrichment of metabolites in a metabolic pathway compared to the total of annotated metabolites in the same pathway. False Discovery Rate was calculated according to the Benjamini and Hochberg multiple testing correction statistical method⁴⁶. Metaboanalyst⁴³, was used to carry out the statistical analysis, MSEA and MPA. The enriched data was mapped to Single Nucleotide Polymorphism (SNP) databases to add biological meaning and aid in its interpretation.

References

- Kielkopf, C. L. *et al.* A structural basis for recognition of A.T and T.A base pairs in the minor groove of B-DNA. *Science* **282**, 111–115 (1998).
- Ferlay, J. *et al.* Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *International journal of cancer* **127**, 2893–2917 (2010).
- Vahabi, M. Breast cancer screening methods: a review of the evidence. *Health Care Women Int* **24**, 773–793 (2003).
- Curtis, C. *et al.* The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* **486**, 346–352 (2012).
- Niu, W. *et al.* BRD7 inhibits the Warburg effect and tumor progression through inactivation of HIF1 α /LDHA axis in breast cancer. *Cell Death Dis* **9**, 519 (2018).
- Potter, M., Newport, E. & Morten, K. J. The Warburg effect: 80 years on. *Biochem Soc Trans* **44**, 1499–1505 (2016).
- Wishart, D. S. *et al.* HMDB: the Human Metabolome Database. *Nucleic Acids Res* **35**, D521–526 (2007).
- Forcisi, S. *et al.* Liquid chromatography-mass spectrometry in metabolomics research: mass analyzers in ultra high pressure liquid chromatography coupling. *J Chromatogr A* **1292**, 51–65 (2013).
- Lee, M. S. *et al.* Metabolomics study with gas chromatography-mass spectrometry for predicting valproic acid-induced hepatotoxicity and discovery of novel biomarkers in rat urine. *Int J Toxicol* **28**, 392–404 (2009).
- Tan, B. *et al.* Metabonomics identifies serum metabolite markers of colorectal cancer. *J Proteome Res* **12**, 3000–3009 (2013).
- Dona, A. C. *et al.* A guide to the identification of metabolites in NMR-based metabonomics/metabolomics experiments. *Comput Struct Biotechnol J* **14**, 135–153 (2016).
- Moreno, A. & Arus, C. Quantitative and qualitative characterization of ¹H NMR spectra of colon tumors, normal mucosa and their perchloric acid extracts: decreased levels of myo-inositol in tumours can be detected in intact biopsies. *NMR Biomed* **9**, 33–45 (1996).
- Mairinger, T. & Hann, S. Implementation of data-dependent isotopologue fragmentation in (13)C-based metabolic flux analysis. *Anal Bioanal Chem* **409**, 3713–3718 (2017).
- Tornabene, T. G., Gelpi, E. & Oro, J. Identification of fatty acids and aliphatic hydrocarbons in *Sarcina lutea* by gas chromatography and combined gas chromatography-mass spectrometry. *J Bacteriol* **94**, 333–343 (1967).
- Kuksis, A. & Prioreshi, P. Isolation of Krebs cycle acids from tissues for gas chromatography. *Analytical biochemistry* **19**, 468–480 (1967).
- Niehaus, W. G. & Ryhage, R. Determination of double bond positions in polyunsaturated fatty acids by combination gas chromatography-mass spectrometry. *Analytical Chemistry* **40**, 1840–1847 (2017).
- Anggard, E. & Sedvall, G. Gas chromatography of catecholamine metabolites using electron capture detection and mass spectrophotometry. *Anal Chem* **41**, 1250–1256 (1969).
- DeJongh, D. C. *et al.* Analysis of trimethylsilyl derivatives of carbohydrates by gas chromatography and mass spectrometry. *Journal of the American Chemical Society* **91**, 1728–1740 (1969).

19. Green, K. Gas chromatography–mass spectrometry of O-methyloxime derivatives of prostaglandins. *Chem Phys Lipids* **3**, 254–272 (1969).
20. Babushok, V. I. *et al.* Development of a database of gas chromatographic retention properties of organic compounds. *J Chromatogr A* **1157**, 414–421 (2007).
21. Horai, H. *et al.* MassBank: a public repository for sharing mass spectral data for life sciences. *J Mass Spectrom* **45**, 703–714 (2010).
22. Kopka, J. *et al.* GMD@CSB.DB: the Golm Metabolome Database. *Bioinformatics* **21**, 1635–1638 (2005).
23. Ma, Y., Kind, T., Yang, D., Leon, C. & Fiehn, O. MS2Analyzer: A software for small molecule substructure annotations from accurate tandem mass spectra. *Anal Chem* **86**, 10724–10731 (2014).
24. Fiehn, O., Kopka, J., Trethewey, R. N. & Willmitzer, L. Identification of uncommon plant metabolites based on calculation of elemental compositions using gas chromatography and quadrupole mass spectrometry. *Anal Chem* **72**, 3573–3580 (2000).
25. Kumari, S., Stevens, D., Kind, T., Denkert, C. & Fiehn, O. Applying in-silico retention index and mass spectra matching for identification of unknown metabolites in accurate mass GC-TOF mass spectrometry. *Anal Chem* **83**, 5895–5902 (2011).
26. Kind, T. *et al.* FiehnLib: mass spectral and retention index libraries for metabolomics based on quadrupole and time-of-flight gas chromatography/mass spectrometry. *Anal Chem* **81**, 10038–10048 (2009).
27. Halket, J. M. *et al.* Deconvolution gas chromatography/mass spectrometry of urinary organic acids–potential for pattern recognition and automated identification of metabolic disorders. *Rapid Commun Mass Spectrom* **13**, 279–284 (1999).
28. Fiehn, O. *et al.* Metabolite profiling for plant functional genomics. *Nat Biotechnol* **18**, 1157–1161 (2000).
29. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144**, 646–674 (2011).
30. Denkert, C. *et al.* Mass spectrometry-based metabolic profiling reveals different metabolite patterns in invasive ovarian carcinomas and ovarian borderline tumors. *Cancer Res* **66**, 10795–10804 (2006).
31. Petrik, V., Loosemore, A., Howe, F. A., Bell, B. A. & Papadopoulos, M. C. OMICS and brain tumour biomarkers. *Br J Neurosurg* **20**, 275–280 (2006).
32. Yang, J. *et al.* Diagnosis of liver cancer using HPLC-based metabonomics avoiding false-positive result from hepatitis and hepatocirrhosis diseases. *J Chromatogr B Analyt Technol Biomed Life Sci* **813**, 59–65 (2004).
33. Vander Heiden, M. G. *et al.* Evidence for an alternative glycolytic pathway in rapidly proliferating cells. *Science* **329**, 1492–1499 (2010).
34. Beger, R. D. A review of applications of metabolomics in cancer. *Metabolites* **3**, 552–574 (2013).
35. Helmlinger, G., Sckell, A., Dellian, M., Forbes, N. S. & Jain, R. K. Acid production in glycolysis-impaired tumors provides new insights into tumor metabolism. *Clin Cancer Res* **8**, 1284–1291 (2002).
36. Nagana Gowda, G. A., Gowda, Y. N. & Raftery, D. Massive glutamine cyclization to pyroglutamic acid in human serum discovered using NMR spectroscopy. *Anal Chem* **87**, 3800–3805 (2015).
37. Ramadan, W. S. *et al.* Interplay between Epigenetics, Expression of Estrogen Receptor- alpha, HER2/ERBB2 and Sensitivity of Triple Negative Breast Cancer Cells to Hormonal Therapy. *Cancers (Basel)* **11** (2018).
38. Merrill, A. E. *et al.* Ciliary abnormalities due to defects in the retrograde transport protein DYNC2H1 in short-rib polydactyly syndrome. *Am J Hum Genet* **84**, 542–549 (2009).
39. Zhang, E. E. *et al.* A genome-wide RNAi screen for modifiers of the circadian clock in human cells. *Cell* **139**, 199–210 (2009).
40. Magklara, A., Brown, T. J. & Diamandis, E. P. Characterization of androgen receptor and nuclear receptor co-regulator expression in human breast cancer cell lines exhibiting differential regulation of kallikreins 2 and 3. *Int J Cancer* **100**, 507–514 (2002).
41. Zhang, Z. X. *et al.* Intracellular pH Regulates TRAIL-Induced Apoptosis and Necroptosis in Endothelial Cells. *J Immunol Res* **2017**, 1503960 (2017).
42. Amith, S. R., Wilkinson, J. M., Baksh, S. & Fliegel, L. The Na(+)/H(+) exchanger (NHE1) as a novel co-adjuvant target in paclitaxel therapy of triple-negative breast cancer cells. *Oncotarget* **6**, 1262–1275 (2015).
43. Xia, J., Psychogios, N., Young, N. & Wishart, D. S. MetaboAnalyst: a web server for metabolomic data analysis and interpretation. *Nucleic Acids Res* **37**, W652–660 (2009).
44. Chen, E. I. *et al.* Adaptation of energy metabolism in breast cancer brain metastases. *Cancer Res* **67**, 1472–1486 (2007).
45. Ibanez, C., Simo, C., Palazoglu, M. & Cifuentes, A. GC-MS based metabolomics of colon cancer cells using different extraction solvents. *Anal Chim Acta* **986**, 48–56 (2017).
46. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Royal Stat. Soc. B* **57**, 289–300 (1995).

Acknowledgements

We thank Ms Varsha Menon for providing the MCF-7 breast cancer cells. R.H. is funded by Breast Cancer Now (Grant No: 2014MaySP323) and Al-Jalila Foundation (Grant No: AJF201741).

Author Contributions

M.S., H.A., R.H. and S.G. designed the study and wrote the manuscript. R.H. carried out the bioinformatic analysis, M.S., H.A., R.H. and S.G. created the figures and interpreted the data. M.S., M.M. and S.G. performed GC-MS experiments, M.S., H.A., R.H., A.A., R.E. and S.G. interpreted the data and supervised the project.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-019-49509-y>.

Competing Interests: The authors declare no competing interests.

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



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019