



## Comparison of the Metabolite Profiles of HT-29 Colorectal Cancer Cells Treated with Curcumin, Cisplatin, 5-Fluorouracil and Doxorubicin in a Metabolomic Approach

Ade Arsianti<sup>1,2</sup>, Innas Widiasti<sup>3</sup>, Aryo Tedjo<sup>1,2</sup>, Shun Hirota<sup>4</sup>, Hiroki Tanimoto<sup>5</sup><sup>1</sup>Department of Medical Chemistry, Faculty of Medicine, Universitas Indonesia, Salemba Raya 6, Jakarta 10430, Indonesia<sup>2</sup>Drug Development Research Cluster, Indonesian Medical Education and Research Institute, Faculty of Medicine, Universitas Indonesia, Salemba Raya 6, Jakarta 10430, Indonesia<sup>3</sup>Biomedical Science, Faculty of Medicine, Universitas Indonesia, Salemba Raya 6, Jakarta 10430, Indonesia<sup>4</sup>Division of Materials Science, Graduate School of Science and Technology, Nara Institute of Science and Technology, 8916-5 Takayama, Ikoma, Nara 630-0192, Japan<sup>5</sup>Faculty of Pharmaceutical Sciences, University of Toyama, 2630 Sugitani, Toyama, 930-0194 Japan.

### ARTICLE INFO

#### Article history:

Received 17 December 2024

Revised 21 December 2024

Accepted 25 February 2025

Published online 01 April 2025

### ABSTRACT

Chemotherapy for colorectal cancer often leads to significant adverse effects on patients, underscoring the need for alternative treatments. Herbal medicines like curcumin are considered a valuable complementary therapy due to their low toxicity profile and potential to mitigate the side effects of chemotherapy. Curcumin's mechanism of action targets multiple pathways, with untargeted metabolomic analysis helping to understand its exact mechanisms and subsequent treatment response. The aim of this study was to compare HT-29 cancer cell metabolites after curcumin and chemotherapy drug interventions to identify metabolites that can predict similar mechanisms of action between these treatments. Principal Component Analysis (PCA) of Fourier transform infrared spectroscopy (FTIR) absorption spectrum showed similar metabolite profiles in HT-29 cell culture media treated with curcumin and the chemotherapeutic cisplatin. Five cell metabolomes emerged after additional gas chromatography mass spectrometry/mass spectrometry (GC-MS/MS) and MS-DIAL data annotation: 1-bromo-2-chloroethane, 2-cyanoacetamide, dimethylamine (DMA), 2-nitrobenzo acid, and butane. The confusion matrix of these five annotated metabolites could be distinguished in HT-29 cell cultures treated with curcumin, but not in control cell cultures or those treated with the drugs cisplatin, doxorubicin, or 5-fluorouracil (5-FU). 2-cyanoacetamide in particular can be used as a marker of HT-29 cells' response to treatment with curcumin based on a p-value of < 0.05. According to these findings, no metabolite can predict the resemblance of curcumin's mechanism of action to chemotherapeutic medicines. Further study should therefore focus on in vivo experimental validation and upgrading metabolomic analysis technologies to further establish the similarities in the metabolite profiles of curcumin and cisplatin treatments.

**Keywords:** Curcumin, Colorectal Cancer, Metabolomics

**Copyright:** © 2025 Arsianti *et al.* This is an open-access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

### Introduction

With a population of 250 million people, Indonesia has a fairly high incidence of colorectal cancer, estimated at 63,500 cases per year.<sup>1</sup> Conventional standard treatments for colorectal cancer are surgery, chemotherapy, and radiotherapy.<sup>1</sup> Radiation therapy and chemotherapy act as the main treatments, though both have serious side effects such as liver toxicity, nausea, and vomiting.<sup>2</sup> Medicinal plants are also used in complementary therapy to treat several types of cancer, including colorectal cancer, with relatively fewer and milder side effects.<sup>3</sup> Curcumin (*Curcuma longa* L.), derived from a medicinal plant, is a naturally occurring phenolic compound that has previously been studied as a drug for use in cancer treatment.<sup>4</sup> Curcumin's phenolic yellowish pigment contains curcuminoids (curcumin, demethoxycurcumin, and bisdemethoxycurcumin) that have been

\*Corresponding author. E mail: [arsi\\_ade2002@yahoo.com](mailto:arsi_ade2002@yahoo.com)

Tel +62-81312581253

Citation Arsianti A, Widiasti I, Tedjo A, Hirota S, Tanimoto H. Comparison of The Metabolite Profiles Of HT-29 Colorectal Cancer Cells Treated with Curcumin, Cisplatin, 5-Fluorouracil And Doxorubicin in a Metabolomic Approach. Trop J Nat Prod Res. 2025; 9(3): 1179 – 1186 <https://doi.org/10.26538/tjnpr/v9i3.38>

Official Journal of Natural Product Research Group, Faculty of Pharmacy, University of Benin, Benin City, Nigeria

linked to anti-microbial, anti-carcinogenic, anti-inflammatory, hypocholesterolemic, and hepatoprotective properties.<sup>5</sup> Its potential anti-cancer effects can induce apoptosis in cancer cells without triggering cytotoxicity in healthy cells, as well as overcome doxorubicin therapeutic resistance through antioxidation activity and p-glycoprotein inhibition.<sup>6</sup> Curcumin also affects the expression of various genes, such as the metallothionein gene, the tubulin gene, and p53, which is involved in colon carcinogenesis.<sup>6</sup> In addition, curcumin is known to decrease the cell viability and growth of HT-29 colorectal cancer cells. Curcumin can inhibit the activation of vascular endothelial growth factor,<sup>7</sup> matrix metalloproteinase, and protein activator-1<sup>8</sup> by inhibiting epithelial-mesenchymal transition and tumor angiogenesis. Moreover, curcumin can inhibit the self-renewal and differentiation of cancer stem cells, the contact and adhesion of tumor cells with blood vessels, and the metastasis of distant tumor cells and the formation of tumor cell microstasis and microspheres.<sup>4</sup>

Numerous studies have shown that curcumin-based cancer treatment is effective and has few to no side effects; however, its anti-cancer activity is hampered by limited absorption and poor solubility.<sup>9</sup> The active ingredients in traditional medicines have an integral mechanism of action to several targets. Therefore, metabolic analysis can help identify the overall mechanism of traditional medicines and their early diagnostic biomarkers, as well as explore disease-related processes and

monitor treatment responses.<sup>10</sup> Previous research highlighted a combination of active chemicals found in extracts and herbal medicines that automatically interact with those found in other medications to either boost or decrease their therapeutic impact.<sup>11</sup>

Four cancer drugs have been tested with two different mechanisms of action at inhibition concentrations (IC<sub>50</sub>) to fight against cancer cells.<sup>12</sup> Research has shown that infrared (IR) spectroscopy is a very accurate descriptor of how anti-cancer drugs work. Thus, the analysis of potential anti-cancer drugs using molecular fingerprints based on the Fourier transform infrared spectroscopy (FTIR) spectrum is invaluable to the discovery of new therapeutic molecules.<sup>13</sup> Gas chromatography mass spectrometry/mass spectrometry (GC-MS/MS) has also been widely used to determine the mechanisms underlying colorectal cancer disease and its biomarkers.<sup>14</sup> Metabolites, or the end products of cellular processes, represent a set of metabolites derived from cells, tissues, organs, or organisms.<sup>13</sup> The aim of this study was thus to compare metabolites between HT-29 cells treated with curcumin and chemotherapy drugs (5-fluorouracil [5-FU], doxorubicin, and cisplatin) to identify those that can predict similar mechanisms of action between treatment types. Novel to this study, FTIR allowed further classification of curcumin and the chemotherapy drugs based on the HT-29 cells' spectrum patterns after administering the four compounds. GC-MS/MS was then used to determine the compound or metabolite underlying these patterns.

## Material and Methods

### *HT-29 Cell Culture*

Human colon adenocarcinoma HT-29 cancer line cells (carrying Smad4 and p53 mutations) were cultured in McCoy 5A media supplemented with 10% fetal bovine serum, 2 mmol/L L- glutamine, 100 U/mL penicillin, and 100 µg/mL streptomycin. Cell cultures were maintained at 37°C in a humidified incubator containing 5% CO<sub>2</sub> and cultured every 3 days.<sup>15</sup>

### *In Vitro Testing of Cytotoxic Activity via the MTT Method*

The MTT method allowed in-vitro testing of cytotoxic activity against the HT-29 cells. The cells were bred using a complete medium containing Dulbecco's modified eagle medium (DMEM), fetal bovine serum 10%, streptomycin 1% as an antibiotic, and amphotericin B as an antifungal. The cells were incubated in a 5% CO<sub>2</sub> incubator and observed every 2–3 days, then harvested after growing up to 80% inside the flask with the addition of 0.25% trypsin EDTA. The cells were then incubated in a 5% CO<sub>2</sub> incubator for 3–5 minutes. The cells that detached from the flask were transferred to a centrifuge tube with 5 mL complete medium to stop the work of the trypsin enzyme. The cells were centrifuged for 5 minutes at 1500 rpm. The resulting supernatant was removed, and the pellets were added to 1 mL complete medium. Resuspension took place so that the cells were homogeneous.<sup>15</sup>

The HT-29 cells were examined with a hemocytometer. A total of 10 µL of cells were added to 10 µL trypan blue until mixed and piped onto the glass of the hemocytometer. The cells were observed under an inverted microscope. The known number and concentration of cells were then diluted according to the desired concentrations. The cells were then placed on 96-well plates (10,000 cells per well). The cells were incubated in a 5% CO<sub>2</sub> incubator for 24 hours, during which they were observed. After 24 hours, curcumin samples were added to the cells at different concentrations (3.12 µg/mL, 6.25 µg/mL, 12.5 µg/mL, 12.5 µg/mL, 100 µg/mL, and 200 µg/mL) dissolved in a complete medium. The cells were then incubated for another 24 hours.<sup>15</sup> The curcumin concentrations were removed from the cells so the cells could be added to an MTT substance at a concentration of 5 mg/mL (diluted ten times). The cells were given as much as 100 µL MTT and then incubated 3–4 hours. If the cells subsequently formed a purple formazan crystal, this indicated the presence of living cells. The resulting formazan crystals were dissolved with the addition of dimethyl sulfoxide and read using a microplate reader (Model 550, Bio-Rad, USA) with a wavelength of 590 nm. The absorbance obtained was used to curve the relationship between concentration and inhibition percentage to determine IC<sub>50</sub>.<sup>15</sup>

### *HT-29 Cell Culture Treatment*

Cell suspension in 6-well plates and a hemocytometer helped calculate the number of concentrations, with as many as 100,000 cells placed in each plate. The cells were incubated in a 5% CO<sub>2</sub> incubator for 24 hours, then combined with a test compound with an IC concentration of 50 (up to 1 mL). A compound treatment test was then carried out by adding curcumin compounds and chemotherapy drugs to individual HT-29 positive cell cultures. There were five types of treatments repeated five times each: a control culture that was not given any test compounds, a culture that was given cisplatin, a culture given 5-FU, a culture given doxorubicin, and a culture given curcumin compounds. The five treatment groups were each incubated for 24 hours.

### *Metabolite Extraction*

Metabolites were extracted from the treated and control HT-29 cell cultures by removing the well plates from the incubator at 37°C and immediately placing them on dry ice (–80°C) or ice (4°C). Additionally, 1 mL growth medium was collected from each well plate and transferred to a Eppendorf tube for centrifuging at 1,500 rpm for 5 minutes. The resulting supernatant was moved to a new Eppendorf tube.<sup>15</sup>

### *Metabolomics Analysis of HT-29 Cell Medium Using FTIR*

An HT-29 cell medium sample (as much as 1 ml) was placed on a diamond plate and mixed with 95 mg KBr, then compressed to form a tablet (3 mm) for FTIR spectrophotometry. The FTIR spectrophotometer (Nicolet™ iS50 FTIR Spectrometer, ThermoFisher, USA) was equipped with the detector deuterated triglycine sulfate. FTIR spectra were recorded in the region of 400–4,000 cm<sup>-1</sup>, in absorbance mode, at 32 scans/min and a resolution of 4 cm<sup>-1</sup>. Peak selection determined the samples' wavelength value, with the data saved in .pdf format. Although spectroscopy provides substantial insights, it is inadequate for determining precise chemical composition. This constraint led to pursuing an alternative approach for comprehensive compositional analysis: GC-MS, a sensitive analytical instrument utilized in metabolomics that can detect, characterize, and identify various chemical constituents and metabolites.<sup>16</sup>

### *Metabolomics Analysis of HT-29 Cell Medium Using GC-MS/MS<sup>15</sup>*

IR spectroscopy is a supplementary technique to GC-MS intended for the analysis of intricate variations in compounds.<sup>18</sup> The volatile constituents and comprehensive chemical distinctions between the curcumin- and chemotherapy drug-treated HT-29 cell cultures were thus determined using both IR spectroscopy and GC-MS (YL6900 GC/MS, YL Instrument, Korea). The culture samples were screened first with a syringe-driven filter unit. The filtered samples were pipetted with a micropipette into the GC-MS vial along with 200 µl methanol solvent and homogenized. The GC-MS vials were inserted into the GC-MS/MS injection site with column types GC 30 m, 0.25 mm, and 0.25 µm at 50°C for 5 minutes and MS 280°C for 20 minutes.<sup>17</sup>

### *Analysis of Raw FTIR Metabolomics Data Using Orange Data Mining*

The Fourier Transform Infrared Spectroscopy (FTIR) raw data were analyzed using the open-source software Orange Data Mining ver. 3.31. The collection of data absorbances were uploaded to Microsoft Excel and divided into five targets (one per treatment group). Absorbance spectrum data were pre-processed with Gaussian smoothing SD 0.5 and rubber band-type baseline correction, then cut according to the reference spectrum section of the bond data group. The Select Row widget allowed selection of a specific treatment. The absorbance spectrum data were then analyzed via multivariate principal component analysis and visualized with the Scatter Plot widget.<sup>14</sup>

### *Analysis of Raw GC-MS Metabolomics Data Using MS-DIAL*

The raw Gas Chromatography-Mass Spectrometry (GC-MS) data were analyzed using the open-source software MS-DIAL ver. 4.92, which is linked to the GCMS KovatsRI-VS3 database (retrieved from <http://prime.psc.riken.jp/compms/msdial/main.html>). The MS-DIAL

parameters were as follows: Data were collected in the mass range of 0–1000 Da. Peak detection was set at an average peak width of 20 scans and a minimum peak height of 1,000 amplitudes. A sigma window value of 0.5 with an electron ionization spectrum limit of 10 amplitudes was implemented for deconvolution. The identification settings were set to a retention index of 20, m/z tolerance of 0.5 Da, EI similarity limit at 70%, and identification score limit at 70%. Setting an alignment parameter yielded a retention index tolerance of 20, with an EI similarity tolerance of 70%. Metabolite annotations were performed by comparing the HT-29 cell culture samples' retention index and spectrum to those in the database.<sup>19</sup>

#### Data Accuracy Analysis with Machine Learning

Owing to the extensive array of metabolites analyzed by several devices, researchers often employ statistical methodologies such as principal component analysis (PCA). PCA, an unsupervised technique, is widely utilized in metabolomics to elucidate the distribution of many compounds following dimensional reduction. It has also been extensively utilized in metabolomics for biomarker discovery in human disorders, including cancer.<sup>20</sup> The current study's PCA data accuracy was tested via machine learning using the open-source software Orange Data Mining ver. 3.31., specifically the models support vector machine (SVM) and neural network (NN). The Test and Score widget with cross validation-type sample data processing displayed data accuracy, and the Confusion Matrix widget gave the proportion of prediction data as well as data from previous machine learning prediction results.<sup>14</sup> Machine learning can be used for metabolomic data to identify druggable targets or pathways in disease processes, as well as predictive metabolites that provide mechanistic inferences of target hypotheses. Target-agnostic drug discovery focuses on target pathways, identifying metabolites that differentiate individuals with specific diagnoses or illnesses. Discriminative metabolites can help develop larger target hypotheses for traditional drug discovery by profiling the metabolomes of individuals with specific diagnoses or illnesses.<sup>21</sup>

## Results and Discussion

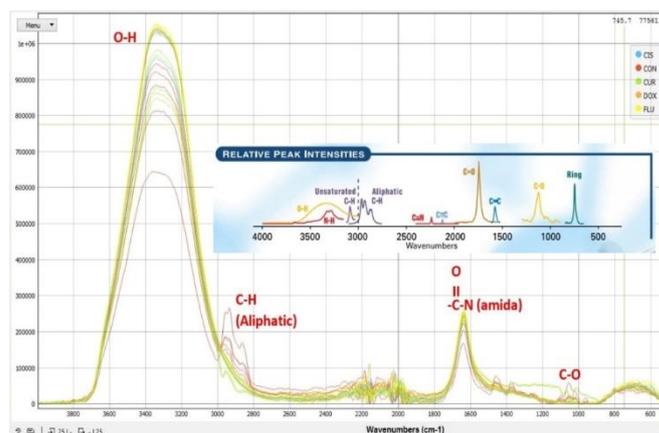
### Cytotoxic Activity

Curcumin compounds were tested for cytotoxicity (in  $IC_{50}$ ) via the 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) method as compared to three anti-cancer compounds commonly used in chemotherapy (doxorubicin, 5-FU, and cisplatin) using HT-29 colorectal cancer line cells. The  $IC_{50}$  value of each compound was obtained from the linear line equation  $y = ax + b$ , where  $y$  is % inhibition and  $x$  is the concentration log. The  $IC_{50}$  value, obtained by converting to anti-log  $x$ , is a 50% concentration of cell growth inhibition in  $\mu\text{g/mL}$  converted to  $\mu\text{M}$  after being multiplied by the molecular weight of each compound. The  $IC_{50}$  values of each compound were 102.44  $\mu\text{g/ml}$  for curcumin, 12.47  $\mu\text{g/ml}$  for cisplatin, 6.83  $\mu\text{g/ml}$  for 5-FU, and 16.21  $\mu\text{g/ml}$  for doxorubicin.

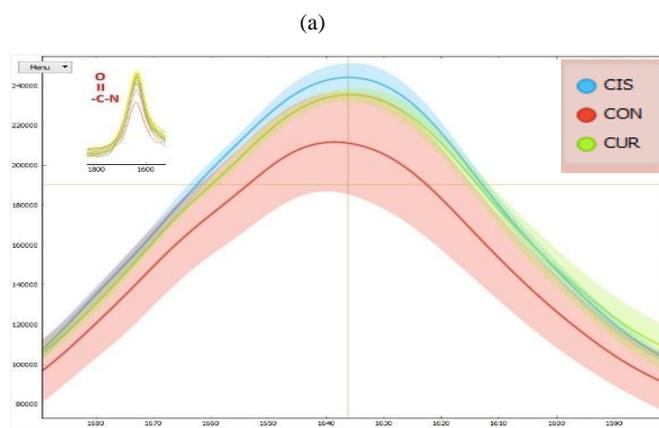
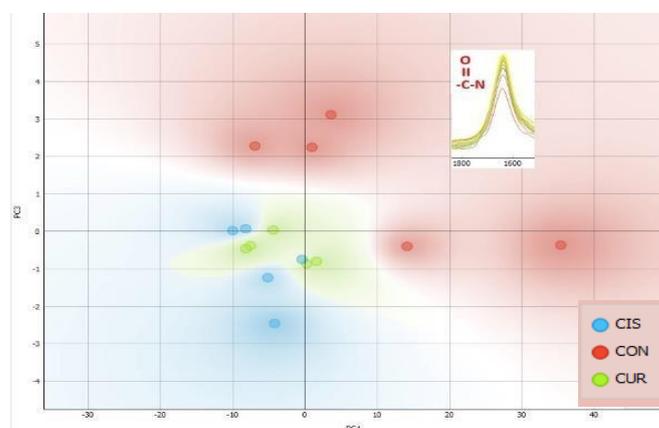
### Metabolite Data Processing with FTIR

Combining all infrared spectrum absorption data from four absorption waves (O-H, C-H, N-C=O, and C-O) of each HT-29 cell culture sample (curcumin, doxorubicin, 5-FU, cisplatin, and control) yielded extracellular metabolite data. The raw data were processed according to the Orange Data workflow. Figure 1 features the FTIR absorbance results of the pre-processed raw data as compared with the reference spectrum. **Figure 1.** Absorption wave spectrum compared to Relative Peak Intensities<sup>34</sup> The results of the pre-processed FTIR absorption data were compared with infrared absorption spectrum data from each functional group.<sup>22</sup> The obtained alcohol absorption spectrum was determined at wavelengths of 3,200–3,500  $\text{cm}^{-1}$ , while the aliphatic absorption spectrum appeared at wavelengths 2,800–3,000  $\text{cm}^{-1}$ , carbonyl absorption spectrum at wavelengths 1,300–1,700  $\text{cm}^{-1}$ , and amide at wavelengths 1,630–17,000  $\text{cm}^{-1}$ . The other functional groups appeared to be stacked, so they were not analyzed further. PCA of the FTIR absorbance data was then performed on each functional group. Absorption at wavelengths 3,200–3,500  $\text{cm}^{-1}$  represented the stretching

of the O-H bond. Absorption at wavelengths 2,800–3,000  $\text{cm}^{-1}$  was dominated by symmetrical stretching vibrations and asymmetry of the



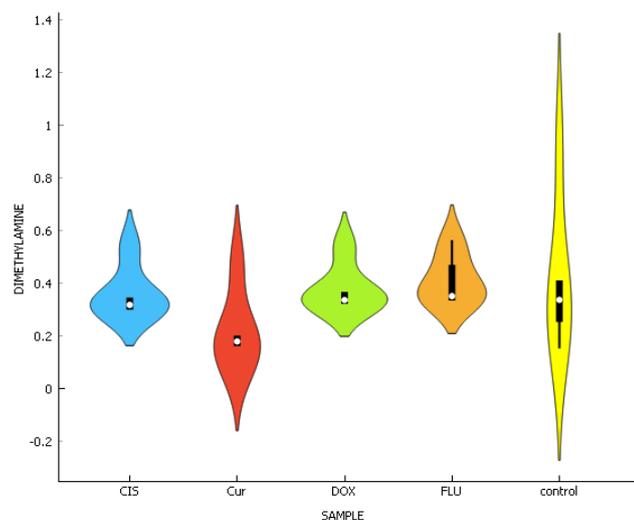
**Figure 1.** Absorption wave spectrum compared to Relative Peak Intensities<sup>34</sup>



**Figure 2.** (a) PCA profile similarities in N-C=O metabolite absorption spectra across control and treated HT-29 cell culture medium samples; (b) comparison of N-C=O metabolite absorption spectra across control and treated medium samples.

$\text{CH}_2$  and  $\text{CH}_3$  groups, especially in cell fatty acids. Absorption between 1,700 and 1,800  $\text{cm}^{-1}$  was characteristic of lipid vibrations. Wavelengths 1,300–1,700  $\text{cm}^{-1}$  indicated protein absorption. Specifically, amide I on the carbonyl stretching of peptide bonds was observed at 1,640  $\text{cm}^{-1}$ , sensitive to the secondary structure of the

protein. Deformation of the N-H amide protein bond (amide II) generated a signal at  $1,540\text{ cm}^{-1}$ . Absorption at  $1,153\text{ cm}^{-1}$  was mainly associated with C-O hydrogen and non-hydrogen bonds.<sup>23</sup>



**Figure 3.** GC-MS/MS data visualization after MS-DIAL processing

Peak stretching vibration at wavelength absorption areas of  $3,200\text{--}3,500\text{ cm}^{-1}$ , constituting the O-H absorption spectrum, underwent additional PCA. This revealed the separation of the metabolic profile of the curcumin-treated HT-29 cell culture medium compared to the control medium, as was also the case with all chemotherapy drug-treated media. Each sample thus had a typical metabolome profile, showing differences in different O-H wavelength absorption regions. Peak wavelength absorption at  $2,800\text{--}3,000\text{ cm}^{-1}$  indicated the C-H, or fatty acid (aliphatic), region, demonstrating lipogenesis and increased membrane lipid saturation. These are associated with the development of cancer because it can reduce membrane fluidity and cell permeability, thus making cancer cells less susceptible to lipid peroxidation and chemotherapy than normal cells.<sup>24</sup> A comparison of aliphatic uptake data showed the separation of the metabolomic profile of the curcumin-treated cell culture medium versus the control medium.

The overall PCA metabolomic profiles of the four treated HT-29 cell culture media showed different C-H absorption metabolites. The curcumin-treated medium had lower fatty acid levels compared to the three chemotherapy drug-treated media, indicating curcumin is better able to suppress fatty acid levels than the assessed chemotherapeutics. Fatty acid levels in colorectal cancer patients often increase due to the oxidation of fatty acids.<sup>25</sup> A previous comparison of fatty acid levels between the plasma of colorectal cancer patients with the plasma of healthy controls showed an increase in 2-methyl butyric acid-type fatty acids and propionic acid in the colorectal cancer patients.<sup>26</sup> PCA also yielded N-C=O absorption wavelength results in the metabolites of the control and treated HT-29 culture media, as summarized in Supplementary Data 1. This analysis showed further similarities in the profiles of the metabolites of cultures treated with curcumin compared to those treated with cisplatin (Figure 2).

**Figure 2.** (a) PCA profile similarities in N-C=O metabolite absorption spectrums across control and treated HT-29 cell culture medium samples; (b) comparison of N-C=O metabolite absorption spectrums across control and treated medium samples. As shown in Figure 2a, treatment with curcumin compared to cisplatin featured the same wavelength for amide absorption (N-C=O). This indicates that the processing of metabolomic data with PCA makes it easier for researchers to determine the separation or grouping of metabolite data from various samples. The complexity of the data generated by time spectroscopic analysis creates difficulties when comparing metabolite profiles between samples. However, metabolomic investigation requires meaningful data interpretation. PCA therefore helps examine entire iterative measurements of absorption wavelength over a single point in time for each sample. One study featured PCA examination of

metabolite variations between samples to determine whether there were systematic changes in the research group.<sup>27</sup> Untargeted MS also has the potential to generate large amounts of information and compare various

**Table 1:** GC-MS/MS analysis of five metabolites using t-test

Metabolite	HT-29 Cell Culture Treatment	Mean (SD)	t
1-Bromo-2-Chloroethana	Curcumin	0,47074 ± 0,10267	9,733 (p=0,000)
	Control	0,01513 ± 0,02033	
2-Cyanoacetamide	Curcumin	0,09153 ± 0,02392	2,812 (p=0,045)
	Control	0,25763 ± 0,12990	
Dimethylamine	Curcumin	0,21521 ± 0,13726	1,598 (p=0,185)
	Control	0,41510 ± 0,26503	
2-Asam Nitrobenzoic	Curcumin	0,03245 ± 0,01276	1,350 (p=0,228)
	Control	0,01147 ± 0,02634	
Butane	Curcumin	0,14017 ± 0,03535	2,347 (p=0,72)
	Control	0,29863 ± 0,14677	

complex metabolite datasets by using a correlation coefficient matrix to measure metabolite similarities between different samples.<sup>28</sup>

A comparison of the wavelengths of the amide absorbance spectrum (N-C=O) of the control, curcumin, and cisplatin culture medium samples revealed that the absorption of amide increased with curcumin and cisplatin treatments compared to the control. This shows that treatment with curcumin successfully inhibits the metabolism of HT-29 cell amide. Amide, namely L-glutamine, is the second source of nutrients for the growth and division of colon cancer cells.<sup>29</sup> Glutamine is considered fuel for the Krebs cycle through  $\alpha$ -ketoglutarate, which results in the synthesis of adenosine triphosphate. Glutamine plays an important role in cellular antioxidative processes, reducing oxidative stress by producing nicotinamide adenine dinucleotide phosphate and glutathione through biosynthesis. Glutamine can also control energy production, redox homeostasis, and intracellular signaling so that tumors are "glutamine addicted," indicating that glutamine and the enzymes involved in its route can be targeted in cancer treatment.<sup>30</sup>

#### Validation of PCA Data Accuracy with Machine Learning

PCA processing of the FTIR absorbance spectrums, especially the comparisons between the treatment and control HT-29 cell culture samples, validated the accuracy of the data, as well as the predicted proportion and actual confusion matrix values.

**Table 2:** PCA accuracy validation and confusion matrix of O-H absorption wave numbers

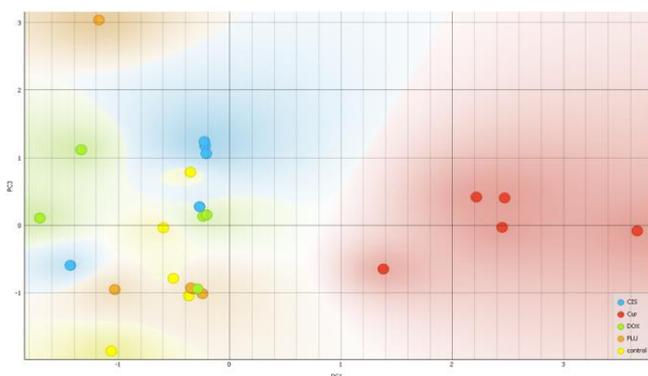
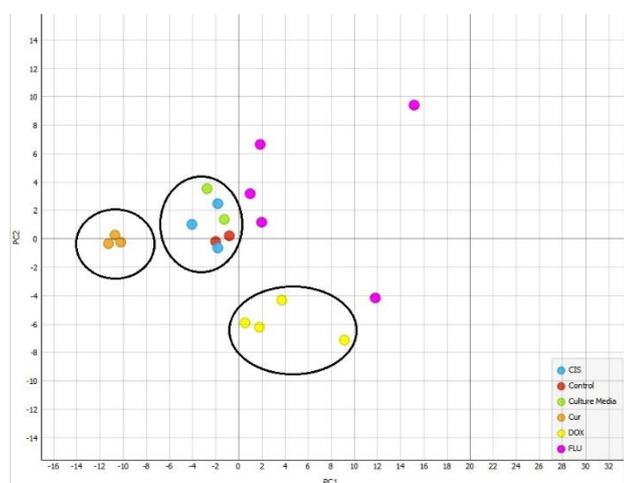
PCA Data	Spectrum Types of Machine Learning	Cross Validation Accuracy Score (Average for classes)	Confusion Matrix of Machine Learning Algorithm results
O-H	Support Vector Machine	0.96	[4 1] [0 5]
	Machine Learning	1.0	[5 0] [0 5]
		1.0	[5 0] [0 5]
		1.0	[5 0] [0 5]

**Table 3:** Validation of PCA accuracy and *confusion matrix* of C-H absorption wave number

PCA Spectrum Data C-H	Types of Machine Learning	Machine Learning	Cross Validation AccuracyRate Score (Average for classes)	Confusion Matrix of Machine Learning Algorithm results
CUR CON	Support Vector Machine	Vector	0.96	[4 1] [0 5]
CIS CON			1.0	[5 0] [0 5]
DOX CON			1.0	[5 0] [0 5]
FLU CON			1.0	[5 0] [0 5]

**Table 4:** Validation of PCA Accuracy and *confusion matrix* of N-C=O absorption wave numbers

PCA Spectrum Data N-C=O	Types of Machine Learning	Machine Learning	Cross Validation AccuracyRate Score (Average for classes)	Confusion Matrix of Machine Learning Algorithm results
CUR CON	Support Vector Machine	Vector	0.92	[3 2] [0 5]
CIS CON			1.0	[5 0] [0 5]
DOX CON			1.0	[5 0] [0 5]
FLU CON			1.0	[5 0] [0 5]

**Figure 4:** PCA data visualization after MS-DIAL processing**Figure 5:** PCA data visualization of unknown MS-DIAL metabolites. The blue dot represents the HT-29 cell sample treated with cisplatin, and the red dot represents the sample treated with curcumin.

Additional machine learning methods, including SVM and NN models, validated the accuracy of the PCA data of each treatment sample compared to the control. Internal validation was carried out five times in conjunction with SVM model and internal validation via cross-validation using training datasets to eliminate potential bias caused by random separation, as feature selection was not performed in this experiment and could therefore use features selected in previous data experiments. The limited sample size allowed a resampling test to be performed as well. The SVM accuracy analysis and validation of PCA FTIR absorbance spectrum data revealed an Area Under the Curve (AUC) value of  $> 0.92$  and Classification Accuracy (CA) value of  $> 0.80$  for all absorption spectrum samples (O-H, C-H, and N-C=O) of the curcumin-treated HT-29 cell cultures compared to the control showed based on their metabolite profiles (Tables 2–4). The accuracy of machine learning predictions can be measured with the confusion matrix. Tables 2–4 shows a separation between the metabolite profiles of curcumin-treated HT-29 cancer cells compared to the control as based on the confusion matrix. The suitability of the machine learning algorithms used plays a role in determining data accuracy as well. For example, for data validation between treatment and control cells, an SVM algorithm can be used for datasets that often cannot be completely separated. SVM will try to build a "soft margin" that minimizes data training points that are outside the classification limits while allowing some points to be misclassified. SVM can only distinguish between two classes, and due to its computational complexity, the algorithm does not scale well with very large data sets. In the case of metabolites, it is therefore often advantageous to perform feature selection before training multivariate algorithms.<sup>28</sup> Identifying multivariate statistical features from data helps distinguish between two separate groups within a high-dimensional feature space so that SVM can create an ideal hyper-plane that sets boundaries and maximizes margins between the two groups.<sup>31</sup>

#### Metabolite Data Processing with GC-MS/MS

Metabolite annotations were performed by comparing the retention index and sample spectrum of each HT-29 cell culture sample with those in the MS-DIAL database. This revealed five metabolites from each treatment group with similar chromatogram and ion peak patterns

to those in the MS-DIAL database. These five metabolites were 1-

		Predicted					
		CIS	Cur	DOX	FLU	control	Σ
Actual	CIS	0.0 %	0.0 %	0.0 %	100.0 %	0.0 %	5
	Cur	0.0 %	100.0 %	0.0 %	0.0 %	0.0 %	5
	DOX	60.0 %	0.0 %	0.0 %	20.0 %	20.0 %	5
	FLU	20.0 %	0.0 %	20.0 %	0.0 %	60.0 %	5
	control	60.0 %	0.0 %	20.0 %	20.0 %	0.0 %	5
Σ		7	5	2	7	4	25

**Figure 6:** PCA accuracy validation and confusion matrix of five metabolites found via GC-MS/MS.

bromo-2-chloroethana, 2-cyanoacetamide, dimethylamine (DMA), 2-nitrobenzoic acid, and butane. The annotation of the HT-29 cell culture metabolites also showed 300 metabolites that had nothing in common with the ones in the MS-DIAL database. Untargeted MS-based metabolomic analysis generates large datasets, making the identification of metabolites with high accuracy a fundamental difficulty. This could be because there are no candidate matches in the database or attributes (e.g., mass ratio and retention time pairings) that only show similarities to a large number of early structures.<sup>32</sup>

To date, there are no studies that support the existence of 1-bromo-2-chloroethana, 2-nitrobenzoic acid, or butane after cell metabolism in colorectal cancer cases. In contrast, 2-cyanoacetate is an organic volatile compound previously found in the urine of colorectal cancer patients using GC-MS.<sup>33</sup> 2-cyanoacetamide in the curcumin-treated HT-29 cell medium was significantly decreased compared to the control and other treated media. As shown in Table 1, 2-cyanoacetamide had a p-value of < 0.05, indicating a significant difference between the curcumin-treated and control media. This implies that 2-cyanoacetate is a suitable biomarker of HT-29 cells' response to curcumin treatment in targeted metabolomic *in vivo* analysis. **Figure 3.** GC-MS/MS data visualization after MS-DIAL processing. Dimethylamine (DMA) metabolites in the curcumin-treated HT-29 cell medium showed higher levels compared to media treated with all three chemotherapy drugs (Figure 3), suggesting that curcumin plays no role in reducing this metabolite in HT-29 cells. This is in contrast with the research of Bednars-Misa et al. (2020), who showed that DMA levels increase with malignancy stage in colorectal cancer patients.<sup>34</sup> Other research on the metabolomic profile of colorectal cancer metastases to spleen nodes found accumulated DMA in line with metastatic development versus normal tissue.<sup>25</sup> DMA is a simple aliphatic amine found in human urine and other bodily fluids such as plasma. The main source of DMA circulating in human urine is asymmetric DMA, which is released from the demethylated protein arginine product.<sup>35</sup> DMA can give rise to nitroso-DMA in acidic gastric juices in the presence of nitrates from carcinogenic foods with DNA alkylation activity. In humans, most DMA (95%) is excreted by the kidneys, while 1–3% of DMA is excreted as feces and exhaled air.<sup>36</sup> DMA's T-test yielded a p-value of > 0.05, showing a negligible difference between the curcumin-treated and control HT-29 cell cultures. As such, DMA cannot be used as a biomarker of HT-29 cell response to curcumin treatment. Analysis of extracellular metabolite data from the HT-29 culture media began with combining all raw intensity data from the four treatment samples' m/z. The results were then compared with MS-DIAL references and visualized in a boxplot. PCA of the five metabolites similar to the MS-DIAL references of each treatment and control group were visualized with Orange Data scatter plot software (Figure 4). The metabolite PCA showed no resemblance between the five samples and the MS-DIAL references (Figure 5). The overall PCA of the five metabolite profiles, unknown metabolites from the control sample, and all four treated HT-29 cell cultures showed different and typical metabolite profiles in the curcumin-treated medium. Figure 5 shows an unknown metabolite profile in the curcumin-treated HT-29 cell medium adjacent to that of the cisplatin-treated medium. This is in support of these media's similarity in the FTIR metabolite profiles. **Figure 4.** PCA data visualization after MS-DIAL processing. **Figure 5.** PCA data

visualization of unknown MS-DIAL metabolites. The blue dot represents the HT-29 cell sample treated with cisplatin, and the red dot represents the sample treated with curcumin. The PCA results of the five metabolites with the same time and retention index from each media sample validated the accuracy of the data, as well as the predicted proportion value and actual confusion matrix value, as shown in Figure 6. The validation analysis of the fifth PCA data metabolite from the MS-DIAL annotation of all treatment and control samples showed an AUC value of > 0.92, indicating that although the metabolites 2-nitrobenzoic acid, DMA, and butane have a p-value of > 0.05 with machine learning, the NN algorithm can show data separation. Subsequent evaluation of NN machine learning prediction performance with the confusion matrix showed that all five MS-DIAL annotation metabolites could be distinguished in the curcumin-treated HT-29 cell culture, but not in the control or other treatment groups. Figure 6. PCA accuracy validation and confusion matrix of five metabolites found via GC-MS/MS. PCA data validation for a large set of data uses appropriate NN algorithms that can process these data well. Units known as neurons form an artificial NN and combine many inputs to produce a single output. The network estimates the relationship between the input (e.g., absorbance spectrum) and the intended output (e.g., disease risk). Inputs, outputs, and intermediate layers, also referred to as hidden layers, form the neurons' organizational structure.<sup>37</sup> The neurons of the first hidden layer receive input from the variables of the layer input after being multiplied by a series of numbers called weights. Each neuron takes input and transforms it by applying a nonlinear activation function, such as a sigmoid or rectified linear unit, and adding bias to the result.<sup>38</sup>

## Conclusion

HT-29 cell culture media treated with curcumin and cisplatin displayed similarities in metabolite profiles based on their FTIR N-C=O and amide absorbance spectrums. Additionally, the metabolite 2-cyanoacetamide can be used as a biomarker of HT-29 cells' response to treatment with curcumin. However, the profiles of five metabolites resulting from the GC-MS/MS of all five culture samples do not predict similarities in the mechanism of action between curcumin and the chemotherapy drugs cisplatin, 5-FU, and doxorubicin. It is challenging to establish a relationship between these metabolite profiles and the proteins related to colorectal cancer due to the complexity of the cellular response mechanisms. Further research should therefore focus on *in vivo* experimental validation that upgrades metabolomic analysis tools to confirm the similar metabolite profiles between curcumin and cisplatin treatments, and to determine if any metabolites can predict similarities in the mechanism of action of curcumin to chemotherapy drugs.

## Conflicts Of Interest

The authors have declared that there is no conflict of interest

## Authors' Declaration

The authors hereby declare that the work presented in this article is original and that any liability for claims relating to the content of this article will be borne by them.

## Acknowledgement

The authors thank to Universitas Indonesia for PUTI Pascasarjana Research Grant Fiscal year 2022 with contract no NKB-126/UN2.RST/HKP.05.00/2022. The authors appreciate the technical and theoretical assistance from the Department of Medical Chemistry, Faculty of Medicine, Universitas Indonesia.

## References

1. Effendi-Ys R. Cancer Stem Cells and Molecular Biology Test in Colorectal Cancer: Therapeutic Implications. Vol. 49, Acta Med Indones-Indones J Intern Med. 2017;49(4):351-359.
2. Lin T, Liang C, Peng W, Qiu Y, Peng L. Mechanisms of core Chinese herbs against colorectal cancer: A study based on

- data mining and network pharmacology. *J Evid Based Complementary Altern Med.* 2020;8325076:1-15.
3. Bhutadiya VI, Mistry Kn. A Review On Bioactive Phytochemicals And It's Mechanism On Cancer Treatment And Prevention By Targeting Multiple Cellular Signaling Pathways. *Int J Pharm Pharm Sci.* 2021;13(12):1-19.
  4. Li P, Pu S, Lin C, He L, Zhao H, Yang C, Guo Z, Xu S, Zhou Z. Curcumin selectively induces colon cancer cell apoptosis and S cell cycle arrest by regulates Rb/E2F/p53 pathway. *J Mol Struct.* 2022;1263:1-9.
  5. Arya N, Prakash OM, Verma AK, Pant AK. Variation In Antioxidant Potential Of *Curcuma Longa* L. Collected From Different Ecological Niches Of Western Himalayan Region. *Int. J. Pharm. Pharm. Sci.* 2015; 7(7):1-6.
  6. War AR. Curcumin Co-Treatment Sensitizes Multi-Drug Resistant HT29 Colon Cancer Cell Line. *J. Can. Res. Immuno-onc.* 2018;4(2):1-8.
  7. Liu J, Li L, Zhang B, Xu ZP. MnO<sub>2</sub>-shelled Doxorubicin/Curcumin nanoformulation for enhanced colorectal cancer chemo-immunotherapy. *J Colloid Interface Sci.* 2022;617:315-325.
  8. Woo JH, Park JM, Jang JH, Yang H, Surh YJ, Na HK. Curcumin induces expression of 15-hydroxyprostaglandin dehydrogenase in gastric mucosal cells and mouse stomach in vivo: AP-1 as a potential target. *J. Nutr. Biochem.* 2020;85:1-11.
  9. Krishnamurthy G, Roy D, Kumar J. Curcumin, a natural golden drug and its anticancer aspects from synthesis to delivery: A review. *Int. J. Appl. Pharm. Innovare Academics Sciences Pvt. Ltd;* 2020; 12:70-84.
  10. Huang Y, Liu Z, Liu S, Song F, Jin Y. Studies on the mechanism of *Panax Ginseng* in the treatment of deficiency of vital energy dementia rats based on urine metabolomics. *J Chromatogr B Analyt Technol Biomed Life Sci.* 2022;1191:1-9.
  11. Gupta M. Synergistic Anti-Cancer Effects Of Natural Products And Their Mode Of Action. *Asian j. pharm. clin. res.* 2021;14:1-7.
  12. Altharawi A, Rahman M, Chan KLA. Towards identifying the mode of action of drugs using live-cell FTIR spectroscopy. *Analyst.* 2019;144(8):2725-2735.
  13. Na J, Hwang HJ, Shin MS, Kang M, Lee J, Bang G, Kim YJ, Hwang YJ, Hwang KA, Park YJ. Extract of radish (*R. Sativus Linn*) promotes anti-atherosclerotic effect using urine metabolomics in ApoE<sup>-/-</sup> mice. *J Funct Foods.* 2021;78:1-11.
  14. AJDA. Orange data mining Orange with Spectroscopy Add-on Workshop. [Online]. 2018 [cited 2022 Nov 25]. Available from: <https://orangedatamining.com/blog/orange-with-spectroscopy-add-on-workshop/>.
  15. Agarwal A, Kasinathan A, Ganesan R, Balasubramanian A, Bhaskaran J, Suresh S, Srinivasan R, Aravind KB, Sivalingan N. Curcumin induces apoptosis and cell cycle arrest via the activation of reactive oxygen species-independent mitochondrial apoptotic pathway in Smad4 and p53 mutated colon adenocarcinoma HT29 cells. *Nutrition Research.* 2018;51:67-81.
  16. He G, Yang S, Wang Y. An integrated chemical characterization based on FT-NIR, and GC-MS for the comparative metabolite profiling of 3 species of the genus *Amomum*. *Anal Chim Acta.* 2023;1280:341869.
  17. Okafor CE, Ijoma IK, Igboamalu CA, Ezebalu CE, Eze CF, Osita-Chikeze JC, Uzor CE, Ekwuekwue AL. Secondary metabolites, spectra characterization, and antioxidant correlation analysis of the polar and nonpolar extracts of *Bryophyllum pinnatum* (Lam) Oken. *Biotechnologia.* 2024;105(2):121-136.
  18. Zheng C, Li J, Liu H, Wang Y. Integrating infrared spectroscopy, metabolomics, and network pharmacology to investigate differential metabolites of volatile compounds in porcini mushrooms. *J. Food Comp. Anal.* 2025;139:107164.
  19. Tsugawa H, Cajka T, Kind T, Ma Y, Higgins B, Ikeda K, Kanazawa M, Vandergheynst J, Fiehn O, Arita M. MS-DIAL: Data-independent MS/MS deconvolution for comprehensive metabolome analysis. *Nat Methods.* 2015;12(6):523-526.
  20. Jiang L, Sullivan H, Wang B. Principal Component Analysis (PCA) Loading and Statistical Tests for Nuclear Magnetic Resonance (NMR) Metabolomics Involving Multiple Study Groups. *Anal Lett.* 2022;55(10):1648-1662.
  21. Martinelli DD. Machine learning for metabolomics research in drug discovery. *Intelligence-Based Medicine Elsevier B.* 2023;8:100101.
  22. UCSC. WPMUCDN Nuclear Magnetic Resonance (NMR). [Online]. 2015 [cited 2022 Nov 25]. Available from: <https://cpb-use1.wpmucdn.com/sites.ucsc.edu/dist/9/291/files/2015/11/108M-Lecture-3-NMR-Intro-Tables-2.pdf>
  23. Wu BB, Gong YP, Wu XH, Chen YY, Chen FF, Jin LT, Cheng BR, Hu F, Xiong B. Fourier transform infrared spectroscopy for the distinction of MCF-7 cells treated with different concentrations of 5-fluorouracil. *J Transl Med.* 2015;13(108):1-8.
  24. Li G, Fang S, Shao X, Li Y, Tong Q, Kong B, Chen L, Wang Y, Yang J, Yu H, Xie X, Zhang J. Curcumin reverses NNMT-induced 5-fluorouracil resistance via increasing ROS and cell cycle arrest in colorectal cancer cells. *Biomolecules.* 2021;11(9):1295.
  25. Zhang F, Zhang Y, Zhao W, Deng K, Wang Z, Yang C, Ma L, Openkova MS, Hou Y, Li K. Metabolomics for biomarker discovery in the diagnosis, prognosis, survival and recurrence of colorectal cancer: a systematic review. *Oncotarget.* 2017;8(21):35460-35472.
  26. Genua F, Mirković B, Mullee A, Levy M, Gallagher WM, Vodicka P, Hughes DJ. Association of circulating short chain fatty acid levels with colorectal adenomas and colorectal cancer. *Clin Nutr ESPEN.* 2021;46:297-304.
  27. Debik J, Sangermani M, Wang F, Madssen TS, Giskeødegård GF. Multivariate analysis of NMR-based metabolomic data. *NMR Biomed.* 2021:1-21.
  28. Kellogg JJ, Graf TN, Paine MF, McCune JS, Kvalheim OM, Oberlies NH, Cech NB. Comparison of Metabolomics Approaches for Evaluating the Variability of Complex Botanical Preparations: Green Tea (*Camellia sinensis*) as a Case Study. *J Nat Prod.* 2017; 80(5):1457-1466.
  29. Tran TQ, Hanse EA, Habowski AN, Li H, Ishak Gabra MB, Yang Y, Lowman Xh, Ooi Am, Liao SY, Edwards RA, Waterman ML, Kong M.  $\alpha$ -Ketoglutarate attenuates Wnt signaling and drives differentiation in colorectal cancer. *Nat Cancer.* 2020;1(3):345-358.
  30. Farhadi P, Yarani R, Dokaneheifard S, Mansouri K. The emerging role of targeting cancer metabolism for cancer therapy. *J. Tum. Bio.* 2020;42(10):1-18.
  31. Hammad A, Elshaer M, Tang X. Identification of potential biomarkers with colorectal cancer based on bioinformatics analysis and machine learning. *Math Biosci Eng.* 2021;18(6):8997-9015.
  32. Schrimpe-Rutledge AC, Codreanu SG, Sherrod SD, McLean JA. Untargeted Metabolomics Strategies—Challenges and Emerging Directions. *J Am Soc Mass Spectrom.* 2016;27(12):1897-1905.
  33. Arasaradnam RP, Mcfarlane MJ, Ryan-Fisher C, Westenbrink E, Hodges P, Thomas MG, Chambers S, Oconnell N, Bailey C, Harmston C, Nwokolo CU, Bardhan KD, Covington JA. Detection of colorectal cancer (CRC) by urinary volatile organic compound analysis. *PLoS One.* 2014;9(9):1-6.
  34. Bednarz-Misa I, Fleszar MG, Zawadzki M, Kapturkiewicz B, Kubiak A, Neubauer K, Witkiewicz W, Korpacka-Krzystek M. L-arginine/no pathway metabolites in colorectal cancer: Relevance as disease biomarkers and predictors of adverse

- clinical outcomes following surgery. *J Clin Med.* 2020;9(6):1–32.
35. Tsikas D. Urinary dimethylamine (DMA) and its precursor asymmetric dimethylarginine (ADMA) in clinical medicine, in the context of nitric oxide (no) and beyond. *J. Clin. Med.* 2020;9(6):1843.
36. Zhang N, Gao M, Wang Z, Zhang J, Cui W, Li J, Zhu X, Zhang H, Yang DH, Xu X. Curcumin reverses doxorubicin resistance in colon cancer cells at the metabolic level. *J Pharm Biomed Anal.* 2021;15:201.
37. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature.* 2015;521(7553):436–44.
38. Bradley M. ThermoFisher FTIR Basic Organic Functional Group Reference Chart. [Online]. 2015 [cited 2023 Feb 9]. Available from: <https://www.thermofisher.com/blog/materials/a-gift-for-you-an-ftir-basic-organic-functional-group-reference-chart/>