

Serum Colorectal Cancer Biomarkers Unraveled by NMR Metabolomics: Past, Present, and Future

Ana M. Salmerón, Ana I. Tristán, Ana C. Abreu,* and Ignacio Fernández*

Cite This: *Anal. Chem.* 2022, 94, 417–430

Read Online

ACCESS |

Metrics & More

Article Recommendations

CONTENTS

Metabolomics Undergoing NMR in Clinical Studies	418
Setting a NMR-Based Clinical Metabolomics Study	418
NMR Data Acquisition	419
Metabolomics Analysis Strategy	420
NMR Spectral Processing	420
Baseline Correction, Phase Correction, and Calibration to Reference	420
Normalization	420
Bucketing	420
Statistical Data Approaches	421
Pathway Analysis	422
NMR Analysis of Serum Samples Obtained from Patients with Colorectal Cancer	422
Correlation of Main Serum Biomarkers to CRC	425
Multivariate Data Analysis Methods	425
Analytical Platforms, Acquisition Parameters, and Processing	427
Other Matrices Analyzed through NMR Metabolomics in the Quest of CRC Biomarkers	427
Conclusions	427
Author Information	428
Corresponding Authors	428
Authors	428
Funding	428
Notes	428
Biographies	428
References	429

Colorectal cancer (CRC) is a malignant growth, known as polyp, located on the inner lining of the large intestine, which is constituted by the colon and rectum. Its development begins when cells start growing continuously, modifying their shape, size, and other related characteristics.¹ This can lead to cancer over time, and it can be presented in two ways, as adenomatous polyps (adenomas), which are able to turn into cancer, and as inflammatory or hyperplastic polyps, which are more frequent and in general are not precancerous.²

In 2020, colorectal cancer represented the cancer with the third highest incidence worldwide, after breast and lung cancer, estimating more than 1.9 million new cases, and ranked in second place in terms of mortality with a total of 935 173 deaths according to the Global Cancer Observatory GLOBOCAN 2020.^{3–5} The most important predictor of

CRC survival is its stage at diagnosis: if an early diagnosis of colon cancer can be provided, a relative 5-year survival rate has been proven to be around 90% for patients diagnosed with localized-stage disease, declining to approximately 71% and 14% for those diagnosed with regional and distant stages, respectively.¹ Therefore, preventive CRC screening tests are generally recommended to be carried out in the population over 50 years old due to a higher risk of enduring this disease over this age.

Currently, other causes that can lead to the development of these tumors are related to unhealthy lifestyle habits such as being overweight or obese, smoking, intaking processed meat, having a sedentary lifestyle, and excessively consuming alcohol. In addition to these factors, some others that do not depend on the individual can be classified into the following categories: (1) having a predisposition to diseases and conditions such as the presence of polyps in the colon and/or rectum and inflammatory diseases such as Crohn's disease and ulcerative colitis, (2) having previously suffered from colorectal cancer, which increases the risk of subsequent cancer, and (3) having genetic factors, such as Lynch syndrome and familial adenomatous polyposis (FAP), or family factors, since the incidence has been shown to be higher in those with relatives who have developed colorectal cancer.⁶

Nowadays, the detection of cancerous polyps is carried out by visual analysis of the structure of the colon and rectum with colonoscopy and sigmoidoscopy being the first and the most widely used techniques with the highest sensitivity and detection rate of this kind of pathology. However, colonoscopy procedures have the clear disadvantage of being highly invasive, a fact that implies an increase of certain risks, such as intraperitoneal or extraperitoneal perforation of the colon, along with the requirement of a tedious and not pleasant preparation process prior to the procedure as well as possible anxiety effects.^{7–9} For this reason, noninvasive techniques included in the “omics” sciences, such as genomics, proteomics, transcriptomics, or metabolomics, are in full swing in the field of biotechnology, contributing in a

Special Issue: Fundamental and Applied Reviews in Analytical Chemistry 2022

Published: November 22, 2021



fundamental way to the understanding and prediction of basic biological issues such as cancer diseases.^{10,11} Multiple omics approaches, either alone or in combination, can be applied to explore the heterogeneity of a certain disease and even in a patient's response to treatment.

Clinical metabolomics aims to identify small molecule metabolites present in patient-derived samples and has attracted much attention to support the discovery of novel biomarkers, which can assess not only the choice of the best treatment for each patient but also the ideal personalized dose regimen. Clinical biomarkers can be dynamic and static. The first ones are commonly employed in patient care and for treatment assessment since they help to define disease progression and the patient's response to the treatment. Static biomarkers are prognostic and aim to predict the clinical response and typically reflect aspects of the physiological state of a patient related to drug treatment response or disease progression dynamics.¹² Overall, metabolomics is involved in a range of clinical applications, including the identification of diagnostic biomarkers of certain diseases, the elucidation of illness mechanisms, the discovery of novel drug targets, and the prediction of treatment reactions. Along with the response of patients to therapies, it is relevant to be able to create more personalized options, contributing to precision medicine. Thus, pharmacometabolomic studies are showing promising results for predicting drug efficacy and toxicity.¹³ Metabolomics supports the relevance of viewing each individual as a different combination of biochemical, physiological, and environmental interactions.¹⁴

We perform herein a review of the published investigations dedicated to the research of the metabolic changes produced in blood serum samples of patients with colorectal cancer, mainly involving nuclear magnetic resonance (NMR) alone and/or in combination with other analytical platforms, with the aim of providing a cutting-edge list of potential biomarkers of this disease. Due to the scarce contributions in the most recent years that dealt with the main topic of this Review, the covered time period consisted of 12 years from July 2009 up to June 2021. This Review is also intended to adequately summarize some of the advances made in the subject and to provide a generic application guide for future studies. Some important considerations on how to perform a NMR-based metabolomics project in a clinical setting are also given. Moreover, a brief section mentioning additional NMR-based metabolomics studies on other biological matrices different from serum, such as urine or feces, were included in order to demonstrate the potential and versatility of this technique.

■ METABOLOMICS UNDERGOING NMR IN CLINICAL STUDIES

Omic sciences attempt to comprehensively study and interpret the complex interactions between molecules in biological systems. As technological advances progress, omic sciences are becoming more notable in the clinical setting,¹⁵ allowing the development of earlier personalized diagnoses to patients and in some instances preventing the progress of the disease.¹⁶

Metabolomics allows one to obtain a picture of the final state of an organism, offering current information on cellular activity.¹⁷ Until recently, metabolomics has been applied less broadly than other omics, and it has sometimes been referred to as a "complementary analysis" to the rest of them.¹⁸ This is the case for many reasons: high subject-to-subject and intrasubject variability, a limited number of annotated

metabolites, dependence on the use of complex and usually expensive analytical platforms, and the fact that interpretation of the results requires a special combination of technical, statistical, and biological or physiological knowledge.^{12,19,20} Although having controversial factors, this omic science has emerged in recent years as a powerful tool in the search for potential biomarkers associated with diseases,²¹ such as colorectal cancer,^{22,23} and as a source of classification and/or prediction models. Some of the main advantages of this discipline are listed herein: (1) it allows to obtain a significantly smaller data set in comparison to other omics, simplifying data processing; (2) the obtained data (profile of metabolites) is able to faithfully reflect multiple aspects of cellular physiology and the current status of the organisms; (3) the identification and quantification of these metabolites are reliable and reproducible and allow one to correlate the fluctuations on their concentration levels and metabolic fluxes with phenotype information.^{24,25}

Metabolomics studies are supported by different high-resolution analytical platforms, such as mass spectrometry (MS) hyphenated to separation methods such as gas (GC) or liquid (LC) chromatography and of course NMR,²⁶ enabling to reach the set of metabolites involved in several cellular processes in a certain biological system. NMR is one of the most widely used techniques and is presented as a robust and versatile platform that performs the measurement, identification, and quantification of a large number of metabolites, even in complex mixtures, in a reliable and repetitive way. It simultaneously provides quantitative and, when needed, structural information. In recent years, it has achieved a drastic gain in sensitivity (signal-to-noise ratio) thanks to the use of cryogenically cooled NMR probes, the so-called cryoprobes. NMR has been demonstrated to overcome many of the disadvantages of other analytical techniques, for instance: (1) it performs a nondestructive and noninvasive analysis of the sample, (2) it does not require a previous separation or derivatization step, (3) it does not depend on the ionization of the analytes, (4) there is no need to use a mass analyzer since there is no dependence on the mass-to-charge ratio (m/z),^{27,28} (5) it has no matrix effect, and (6) the quantification of the metabolites does not rely on calibration curves to quantify the concentration and recovery because only one internal standard is usually added for quantification purposes.^{29–31}

Routine "omic" NMR spectroscopy suffers from several drawbacks, but probably, the most important one is the fact that ¹H NMR complex spectra may be inevitably crowded, which hampers the identification and quantification of metabolites.³² Nevertheless, there are strategies that overcome signal overlap, which can be obtained by spreading the resonances in a second dimension using 2D NMR spectroscopy or by applying specific filters, such as Carr-Purcell-Meiboom-Gill (CPMG) or diffusion modules.^{33,34} This and other important considerations will be briefly detailed in the next section.

■ SETTING A NMR-BASED CLINICAL METABOLOMICS STUDY

NMR metabolomics has fully expanded the understanding of cellular and physiological metabolism, helping researchers to identify multiple unexpected biochemical associations in different conditions and diseases, for example, in cancer,^{35–37} autism,^{38,39} infertility,^{40,41} major depression disorder,^{42,43} anorexia nervosa,^{44,45} etc. Since metabolite levels within an

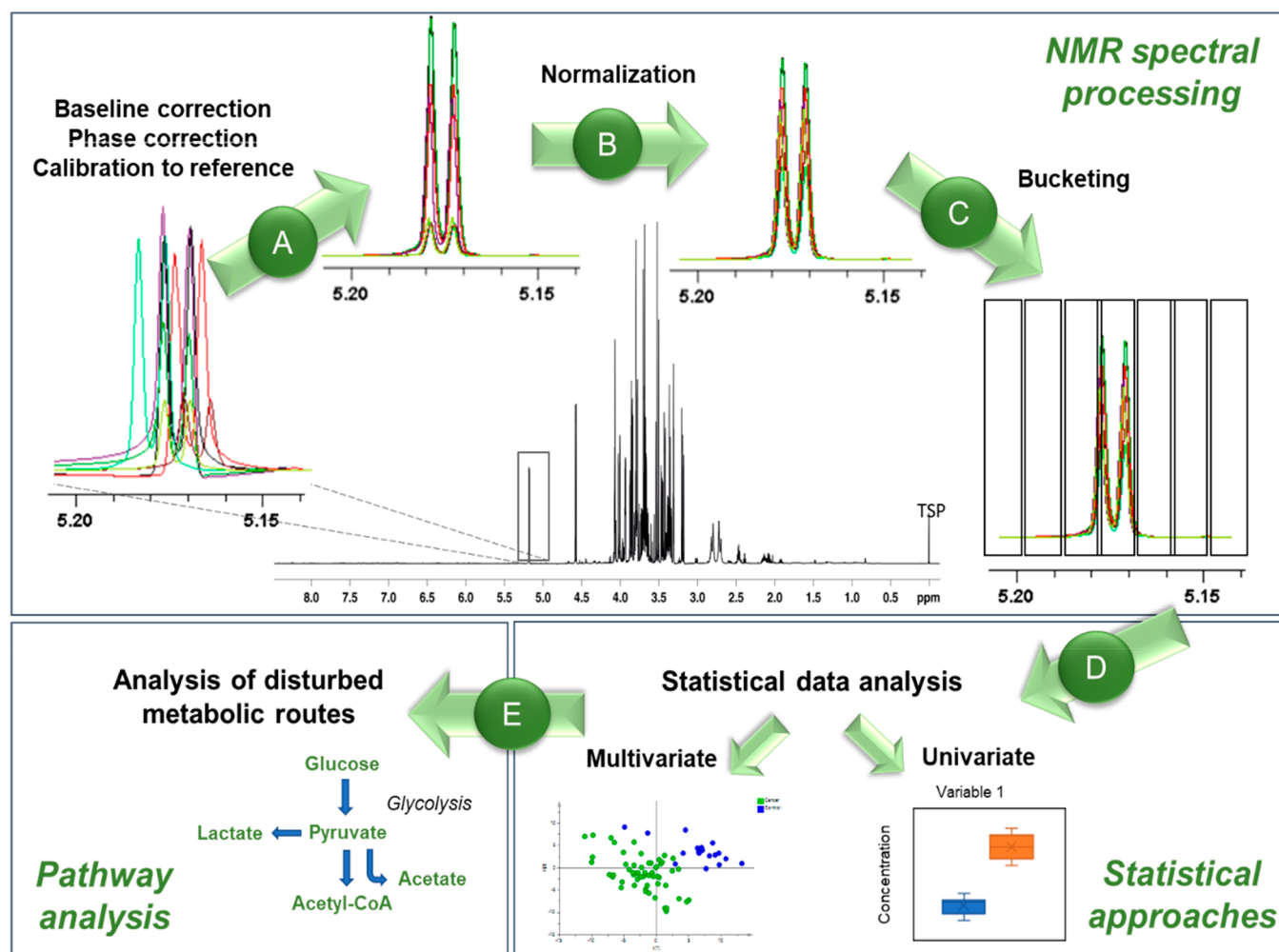


Figure 1. Basic steps of a NMR-based metabolomics study through the fingerprinting method: (A) baseline correction, phase correction, and calibration to the reference, which is usually placed at 0 ppm, (B) NMR spectra normalization to total intensity, (C) bucketing of the spectral data, obtaining a data matrix, which can be subjected to scaling and centering, (D) statistical data analysis, distinguishing between multivariate and univariate approaches, and (E) the analysis of disturbed metabolic routes.

individual vary over time, in order to obtain the most satisfactory results and to detect associations with a disease, clinical metabolomics studies must attend some specific considerations for controlling and decreasing within-individual and technical variability through an adequate study design.

Commonly, NMR-based metabolomics studies employ well-known biofluids collected from the patients with the most common being blood serum or urine, where the location of the main peaks and the compositions are already almost established. Some other fluids include cerebrospinal fluid, bile, eye humor, and saliva. Also, tissue extracts are being extensively studied although they eventually give worse results than those from intact tissues directly based on the fact that metabolic changes are usually more concentrated in the tissue itself.^{30,46}

In a metabolomics study, the experiment design and sampling method are of utmost importance and must be accurately defined. There are some patient-specific factors that may affect interindividual variation, such as gender, age, weight, or lifestyle, so appropriate selection criteria of the individuals must be applied.²⁵ Furthermore, when the aim of the research is the discovery of biomarkers for case-control situations, it can lead to inappropriate oversimplifications,

contributing to the presence of intra- and interindividual variability in metabolic signatures. For this reason, the collection of repeated samples in triplicate is of great relevance.¹² Other important factors that may be taken into consideration in a metabolomics study involve the sample storage, sampling size, and time.¹² Regarding the former, an adequate temperature for clinical samples, which is $-80\text{ }^{\circ}\text{C}$, is required, allowing the steadiness of the metabolome at least for 6 months.^{47,48}

In this section, we describe some relevant aspects to take into consideration when developing a NMR-based metabolomics project in a clinical setting to allow for the optimization of the process.

NMR Data Acquisition. When measuring biological samples by NMR, it is frequently necessary to perform the suppression of water, especially in body fluids, since they usually present a substantial difference in water concentration regarding their own metabolites. Therefore, experiments such as the presaturation of the solvent signal employing a continuous wave pulse are generally performed. For instance, 1D-NOESY PRESAT, in combination with the presaturation module, introduces a 90° triple pulse sequence that effectively eliminates such signals without causing increased distortions in

adjacent signals.⁴⁹ Also, other methods such as excitation sculpting or WATERGATE solvent suppression are occasionally employed.⁵⁰

Furthermore, it may also become necessary to ensure the elimination of signals based on molecular weight for which the so-called diffusion filters are used.⁴⁹ Those employ a combination of radiofrequency pulses and magnetic field gradients, whether monopolar or bipolar, that manages to attenuate signals of smaller molecules, usually those from the solvent employed, although they evidence a disadvantage: the rest of the metabolites also suffer attenuation in their signals to a greater or lesser extent depending on their size. Transverse relaxation time (T_2) filters such as CPMG are also applied, which constitutes the most employed sequence in the study of serum samples⁵¹ and eliminates signals with a small T_2 that are usually associated with systems with long correlation times generally present in macromolecules such as proteins.⁵²

The profiling process, which is in fact, the identification of metabolites in one-dimensional NMR spectra, is carried out through direct assignment using multiplicities and chemical shifts with the help of databases such as the Human Metabolome Database (HMDB), multiple tools available such as the Chenomx software, and some packages available for R such as BATMAN or ASICS. Thus, the confirmation of the assignments is performed using different homonuclear bidimensional spectra such as ^1H , ^1H -COSY or ^1H , ^1H -TOCSY and heteronuclear spectra such as ^1H , ^{13}C -HMQC, ^1H , ^{13}C -HSQC, ^1H , ^{13}C -HMBC, ^1H , ^{15}N -HMQC/HMBC, and ^1H , ^{31}P -HMQC/HMBC, where even more detailed information on the structure of the metabolites is obtained.^{53,54}

Metabolomics Analysis Strategy. Similarly to the rest of the omic sciences, metabolomics studies require large numbers of samples and generate a large amount of data, so reducing their size is a special need in order to obtain a more adequate and correct interpretation of the results. For this purpose, chemometrics methods are employed.⁵⁵ Chemometrics is the discipline that combines mathematical and statistical procedures to extract the most relevant information from the experimental data set, thus improving the process of interpretation and providing quality results.^{56,57} Nowadays, chemometrics techniques are primarily used in chemistry for signal processing, experimental designs, variable reduction, data exploration, multivariate data analysis, and pattern recognition.^{58,59}

Metabolomics analyses, in global terms, can be divided according to whether there is some type of prior knowledge about the metabolites of interest or whether there is no information about them.⁶⁰ The first one, targeted metabolomics, focuses on the monitoring of previously selected compounds based on known metabolic pathways or pays attention to those biomarkers strongly associated with the study condition. Thus, these metabolites must be appropriately assigned and quantified in the samples. The second one drives untargeted metabolomics analyses and therefore focuses on the unbiased study of the spectral profile as a whole, considering every single signal present in the sample.⁶¹ To do this, two basic approaches can be utilized: (a) the fingerprinting and/or (b) the profiling method.^{62,63} The former performs a rapid evaluation of the total metabolites present in the spectra by transforming them into data matrices using the bucketing method (or binning), where small spectral regions or "buckets" with a width between 0.02 and 0.04 ppm are taken and are later used to carry out the pertinent statistical analyses and

perform classifications.⁶⁴ The latter consists of studying the entire spectrum using specific peak alignment algorithms and is used to determine the concentrations of all quantifiable metabolites in biological samples, providing useful information from a biochemical point of view.⁶⁵

The process of analysis through the fingerprinting method follows a series of steps, which will be collected and briefly discussed in the following sections: (1) **NMR Spectral Processing** (Figure 1A–C), (2) **Statistical Data Approaches** (Figure 1D), and (3) **Pathway Analysis** (Figure 1E).⁶⁶

NMR Spectral Processing. *Baseline Correction, Phase Correction, and Calibration to Reference.* This step involves the transformation of the spectral data into their optimal version for the subsequent statistical analysis. It includes the following: checking for the absence of data; adjusting the baseline; referencing the spectrum so that the signal from the internal standard is located at the same chemical shift in all spectra; multiplying the spectrum by functions that soften or accentuate the spectral resolution; applying algorithms that minimize fluctuation in the chemical shifts as a consequence of variations in temperature; suppressing defective spectral regions or areas where there are signal shifts, usually coming from acidic groups or exchangeable protons.^{30,64}

Normalization. In NMR analysis, identical sample volumes are usually acquired to make all samples comparable with each other. However, in the case of samples corresponding to biofluids, there are multiple external variables that can affect the concentration of metabolites, such as the hydration status of each individual or even possible experimental inaccuracies or technical errors.^{53,67} In order to obtain comparable volumes and concentrations, a normalization step should be applied, which manages the correction of these dilution or concentration factors between samples. In metabolomics, a series of methods are used although normalization is generally achieved by considering the intensity of the total area of the spectrum⁶⁸ and by dividing the values of the peak integration of the buckets by the sum of all of them, so that the sum of all these divisions must provide a value equal to unity.^{30,67,69,70}

Bucketing. As mentioned before, the total of the metabolites in the NMR spectra are assessed through their transformation into data matrices constituted by minor spectral areas (between 0.02 and 0.04 ppm of the width) called "buckets". Once the "buckets" table has been obtained, the multivariate statistical analysis of the data is carried out. For this, the purpose of the study must be kept in mind, which may be (a) the visualization of the general differences between samples, such as trends or correlations, (b) the detection of statistical significant differences between groups, (c) the highlight of spectral regions that contribute the most to the observed differences, and/or (d) the construction of a predictive model for the correct classification of new samples.⁶⁴

After this step of bucketing, multivariate analysis techniques are usually used to extract information from the data with the aim of providing biological knowledge on the studied matter.⁷¹ This data analysis focus on the spectral profile and any information on biological variation can overlap, so centering through the mean of the data is a fairly common step, since it enables to compensate for this problem, focusing on biological variation and the possible differences and similarities between the samples. However, those metabolites that are more abundant in the samples will show higher values in the data table, so they will end up contributing more to the model that is generated later.⁶⁹ In order to avoid this bias, scaling methods

Table 1. List of Metabolomics Studies of Colorectal Cancer Covered in This Manuscript until June 2021 in Chronological Order^{a,c}

reference	technique	NMR probe-head	NMR sample size	sample volume	multivariate analysis	aim	conclusions ^c
Ludwig et al. ⁹¹ (2009)	¹ H NMR (800 MHz)	HCN cryogenic probe	38 CRC, 8 adenoma, 19 controls	-	PCA, PLS-DA	identification of CRC and adenoma-positive biomarkers	higher levels of Lac, Pyr, Acet, Aco, and 3-HB in CRC samples
Backshall et al. ⁹² (2011)	¹ H NMR (600 MHz)	5 mm triple resonance TXI probe	52 patients	200 μ L	PLS-DA	prediction of toxic effects on patients undergoing treatment with capecitabine	differences in the serum metabolic profile between pretreatment and subsequent toxicity; the lipid profile provides information on the risk and possible severity of toxicity
Bertini et al. ⁹³ (2012)	¹ H NMR (600 MHz)	CPTPI probe	153 metastatic CRC, 139 controls	300 μ L	PLS-DA, SVM, CA	identification of metastatic CRC biomarkers and predict overall patient survival	lower serum levels in CRC samples of Ala, Cit, Leu, Pyr, Tyr, and Val and higher content of 3-HB, Acet, Form, Glo, lipids, GlyP, Phe, and Pro; this could lead to a signature to predict overall survival
Farshidfar et al. ⁹⁴ (2012)	GC-MS, ¹ H NMR (600 MHz)	5 mm triple resonance TXI probe	42 colorectal CRC, 45 liver-only CRC, 25 extra-hepatic metastases	-	PCA, OPLS-DA	identification of occult metastases biomarkers in CRC samples	serum metabolic profile changes with metastases and with sites of disease
Zamani et al. ⁹⁵ (2014)	¹ H NMR (500 MHz)	-	33 CRC, 33 controls	600 μ L	PCA, PLS-DA	identification of CRC biomarkers	decrease in the levels of Pyx, Oro, AdeH, Pya, Glyc, β -Leu, MetCy, Tau, 3-HB, AcChol, 3-HV, Fuc, Chol, and PCar and increase in Gly in CRC samples; the LDA/DCA ratio is considered a possible CRC biomarker
Chen et al. ⁹⁶ (2015)	¹ H NMR (500 MHz)	HCN cryogenic probe	44 polyps, 58 controls	530 μ L	seemingly unrelated regression	identification of polyps-associated biomarkers	eleven groups of metabolites showed alterations between patients with polyps and healthy controls
Deng et al. ⁹⁷ (2016)	LC-MS, ¹ H NMR (500 MHz)	5 mm triple resonance TXI probe	28 CRC, 44 polyps, 55 controls	530 μ L	PLS-DA	identification of CRC- and polyps-associated biomarkers	higher levels of Glc, lower levels of Ade, and alterations in Pyr and Gln levels in CRC samples; decrease in Ort and increase in Ade in polyp samples
Vahabi et al. ⁹⁸ (2017)	¹ H NMR (500 MHz)	-	8 CRC stage 0-I, 8 CRC stage I-IV	600 μ L	OPLS-DA	identification of CRC biomarkers in different disease stages (0 to I and II to VI)	alterations in Fum, Cit, Oxa, Lin, and lipids in both groups
Gu et al. ⁷ (2019)	¹ H NMR (600 MHz)	-	40 CRC, 32 polyps, 39 controls	300 μ L	PCA, PLS-DA, OPLS-DA, RF, SVM	identification of CRC and polyps-associated biomarkers	lower levels of Pyx, increase in Gly, Cholt, Tau, CholEs, and Den in the CRC group (II to IV stages)
Di Donato et al. ⁹⁹ (2021)	¹ H NMR (600 MHz)	-	65 CRC relapse free, 29 relapsed, 75 metastatic CRC	1 mL	PCA-CA-kNN	prediction of prognosis and identification of metastatic CRC biomarkers	increase in the levels of Lac, Gly, Ser, Chol, and 3-HB and decrease of Cit and Suc in CRC samples; higher levels of Lac, Glut, Chol, PUFA, and NaGly and lower levels of Acet, Glo, Gln, Ala, and Asp in polyps samples Acet/Glo and Lac/Cit ratios considered as possible polyps and CRC biomarkers, respectively lower levels of Gln and His in the metastatic CRC group

^aSymbol: “-” denotes studies where the specific information was not provided. ^bColorectal cancer. ^cIn samples corresponding to colorectal cancer, abbreviations are as follows. Acet: acetate; Aco: acetoacetate; AcChol: 7-acetcholesterol; Ade: adenosine; AdeH: s-adenosylhomocysteine; Ala: alanine; Asp: aspartate; Chol: cholesterol; CholEs: cholesteryl ester; Chol: choline; Cit: citrate; DCA: deoxycholic acid; Den: deoxyinosine; For: formate; Fuc: 3-fucose; Fum: fumarate; Gln: glutamine; Glo: glycerol; Gly: glycine; Glyc: glycocholic acid; GlyP: glycoproteins; His: histidine; 3-HB: 3-hydroxybutyrate; 3-HV: 3-hydroxyisovaleric acid; Lac: lactate; Leu: leucine; Lin: linolenic acid; LDA: lithocholic acid; MetCy: 5-methyltyridine; NaGly: N-acetyl glycoproteins; Ort: orotate; Oro: orotidine; Oxa: oxaloacetate; PCar: l-palmitoylcarnitine; Phe: phenylalanine; Pro: proline; PUFA: polyunsaturated fatty acids; Pya: pyridoxamine; Pyx: pyridoxine; Pyr: pyruvate; Ser: serine; Suc: succinate; Tau: taurocholic acid; Tyr: tyrosine.

visualization of the most shared keywords provided by the VOSviewer software⁸⁸ when applied to this data set.

As it is observed, some clusters of words were established to emphasize, for example, the relevance of multivariate analysis techniques in the metabolomics research of CRC (through the terms “curve”, “auc”, “roc”, “opls-da”, “pls-da”, “accuracy”, “specificity”, “sensitivity”, “value”, “area”, and “test”), the importance of NMR spectroscopy and mass spectrometry in this topic (with the terms “NMR spectroscopy” and “flight mass spectrometry”), and the description of some colorectal cancer biomarkers and disturbed pathways reported by several metabolomics studies of this disease (such as “glycolysis”, “lactate”, “glutamate”, “glucose”, “arginine”, and “lipid metabolism”).

When a second search is performed by introducing “NMR” or “nuclear magnetic resonance” to the same previous keywords, the resulting output was reduced to a set of 146 contributions published from 2004 to October 2021. The following graphic illustrates a positive trend in which a considerable increase in the number of publications on this topic occurred starting in 2009. As previously discussed, analytical techniques have progressed and improved over the last years, especially NMR, which has increased its sensitivity by up to a factor of 5 mostly due to the development of cryoprobes, which have contributed to the development of metabolomics as an increasingly applied field of research, as the upward trend of Figure 3 reflects.

An analysis of some of these publications was further performed using the CitNetExplorer software⁸⁹ to obtain a citation network including the articles, reviews, and book chapters involved (Figure 4). As it could be observed, all of these publications conjoin in the popular publication of 1956 by Warburg et al. (marked in green)⁹⁰ in which the Warburg hypothesis, which explains the alleged root cause of cancer, was formulated. The publications that employ serum samples for the NMR-based metabolomics analysis of colorectal cancer, which constitutes the main topic of the current review, appear in orange, while the articles marked in blue are those focused on the analysis of tissue, feces, and urine samples, some of which will be also mentioned further below. A total of 10 contributions were obtained for the former case, including Ludwig et al.,⁹¹ Backshall et al.,⁹² Bertini et al.,⁹³ Farshidfar et al.,⁹⁴ Zamani et al.,⁹⁵ Chen et al.,⁹⁶ Deng et al.,⁹⁷ Vahabi et al.,⁹⁸ Gu et al.,⁷ and Di Donato et al.⁹⁹

Table 1 highlights some of the most important metabolomics investigations dedicated to the study of colorectal cancer specifically in serum samples through NMR with the general aim of obtaining biomarkers of this disease.

The study by Ludwig et al.⁹¹ employed multidimensional 2D-¹H,¹H-TOCSY NMR spectra for the metabolomics analysis of 38 serum samples positive in colorectal cancer, 8 of them also being confirmed in adenoma, and 19 control samples. The application of PCA and PLS-DA models to NMR data allowed them to conclude that the cancerous samples showed higher levels of lactate, pyruvate, and ketone scaffolds (acetate, acetoacetate, and 3-hydroxybutyrate) than the control ones. In this work, they were able to significantly reduce the measurement time of the ¹H,¹H-TOCSY spectra due to the substitution of the Fourier transform by the Hadamard transform, thus being able to deconvolute crowded NMR spectra.⁹¹

In 2011, the team of Keun⁹² accomplished the first study to report the capacity of ¹H NMR metabolomics to predict

adverse effects and toxicity severity associated with the administration of the chemotherapy medication capecitabine in serum samples of patients with colorectal cancer. For this purpose, a generated PLS-DA model was able to correlate the presence of higher levels of polyunsaturated fatty acids and choline phospholipids with higher grade toxicity over the treatment period; however, this model did not reach significance by cross-validation.⁹²

Later, Bertini et al.⁹³ employed ¹H NMR to study the metabolic profile of 153 samples of metastatic colorectal cancer serum and 139 control samples with the aim of obtaining valid biomarkers and predicting patient survival. First, they applied PLS-DA for dimension reduction, followed by a canonical analysis (CA) evaluation that revealed good discrimination and a support vector machine (SVM) model for classification. In this analysis, they observed lower and higher levels of 6 (alanine, citrate, leucine, pyruvate, tyrosine, valine) and 8 (3-hydroxybutyrate, acetate, formate, glycerol, lipids, glycoproteins, phenylalanine, and proline) metabolites, respectively, in metastatic colorectal cancer samples, leading to a possible metabolic signature, which may offer an independent tool to predict overall survival. In addition, they verified different metabolic shifts between patients with shorter and longer survivals.⁹³

In the same year, Farshidfar et al.⁹⁴ conducted research employing GC-MS and ¹H NMR with the aim of distinguishing the stage of colorectal cancer in 42 serum samples of patients with colorectal colorectal cancer (cancer stages II and III), 45 samples of patients with liver-only metastases, and 25 samples of patients with extrahepatic metastases (cancer stage IV in both cases). A PCA exploratory analysis followed by an OPLS-DA model allowed them to differentiate serum metabolic profiles of patients with metastases and between metastases appearing in different organs.⁹⁴

In 2014, Zamani et al.⁹⁵ carried out a metabolomics study using ¹H NMR of 33 serum samples corresponding to a positive group in colorectal cancer and 33 control samples with the aim of obtaining a prediction model and possible biomarkers. The application of PCA and PLS models to ¹H NMR data showed a positive discrimination between both groups, caused by a decrease in the levels of pyridoxine, orotidine, *s*-adenosylhomocysteine, pyridoxamine, glycocholic acid, β -leucine, 5-methylcytidine, taurocholic acid, 3-hydroxybutyric acid, 7-acetocholesterol, 3-hydroxyisovaleric acid, l-fucose, cholesterol, and l-palmitoylcarnitine for the cancer group together with an increase in glycine. In addition, they highlighted the ratio of lithocholic acid/deoxycholic acid as a possible biomarker of colon cancer.⁹⁵

Furthermore, Chen et al.⁹⁶ conducted an investigation into the ¹H NMR metabolic profile of 44 patient samples with colon polyps and 58 control samples along with numerous demographic parameters, performing seemingly unrelated regression (SUR) for the correlation of the metabolites and the biological groups. They were able to obtain valine as a slightly significant metabolite for patients with polyps, as they had a reduced sample size, but could report 11 groups of metabolites that were significantly different between polyps and control samples.⁹⁶

A year later, in 2016, the group of Raftery⁹⁷ carried out a metabolomics study using LC-MS and ¹H NMR on serum samples from a positive group for colorectal cancer of 28 subjects, a total of 44 individuals with polyps, and a third group of 55 controls. They generated an algorithm where all variables

were examined, removing one of them in each iteration and employing the remaining ones for PLS-DA. Variables with the highest prediction accuracy for the test samples in Monte Carlo Cross Validation (MCCV) were kept for the subsequent iteration, resulting in a 30% test set and a 70% training set, the portion where PLS-DA was performed to predict the classification of the test set samples. Colon cancer samples displayed higher levels of glucose, lower levels of adenosine, and alterations in the levels of pyruvate and glutamine, while a decrease in orotate and an increase in adenosine were found in the group positive for polyps. Alterations in the levels of amino acids, fumarate, citrate, oxaloacetate, linolenic acid, and lipids were observed for both cancer and polyp groups compared to the controls.⁹⁷

Vahabi et al.⁹⁸ investigated the differences in the ¹H NMR metabolic profile of 16 colorectal cancer samples between 0–I stages (8 samples) and I–IV stages (8 samples), similarly to Farshidfar et al.⁹⁴ An OPLS-DA model showed decreased levels of pyridoxine and increased contents of glycine, cholesterol, taurocholic acid, cholesteryl, and deoxyinosine for the II–IV stages of colorectal cancer samples.⁹⁸

In 2019, Gu et al.⁷ conducted a ¹H NMR metabolomics analysis of 40 serum samples from colon cancer patients, 32 samples positive for polyps, and 38 controls, and several models [PCA, PLS-DA, OPLS-DA, random forest (RF), and SVM methods] were applied to the ¹H NMR data to identify possible biomarkers. A total of 23 metabolites were elucidated, reporting an increase in the levels of lactate, glycine, serine, choline, and 3-hydroxybutyrate and a decrease of citrate and succinate for colorectal cancer samples. Also, higher levels of lactate, glutamate, choline, polyunsaturated fatty acids, and N-acetyl glycoproteins and lower levels of acetate, glycerol, glutamine, alanine, and aspartate were found for the polyps samples. Furthermore, they could determine that the acetate/glycerol and lactate/citrate ratios were important biomarkers for the presence of polyps and colorectal cancer, respectively.

Recently, Di Donato et al.⁹⁹ hypothesized that NMR-based metabolic fingerprinting could improve risk stratification in patients with early colorectal cancer and investigated serum samples of 94 elderly patients with early stage colorectal cancer (65 relapse free and 29 relapsed after a 5 year median followup) and 75 elderly patients with metastatic colorectal cancer. Prognosis was assessed using Kaplan–Meier curves, and a PCA-based kNN analysis was able to distinguish between relapse free and metastatic colorectal cancer groups, mainly due to lower levels of glutamine and histidine in patients with metastatic colorectal cancer.⁹⁹

Correlation of Main Serum Biomarkers to CRC. In general, it has been proven that the most relevant metabolites found to be important biomarkers associated with colorectal cancer are widely related to carbohydrate metabolism, involving gluconeogenesis^{96,93} and specially glycolysis,^{7,91,94,96,97,100,101} since an increase in activity in this pathway can lead to an increase in malignant tumors, known as the Warburg effect.⁹⁰ This process involves an abnormal accumulation of glucose, pyruvate, and lactate (initial, intermediate, and final metabolites of glycolysis, respectively), as reported in many of the studies previously discussed.^{7,97,91,93} Furthermore, lower amounts of other metabolites related to glycolysis, such as citrate and succinate, were also indicated as part of this Warburg effect.^{7,90} Citrate is also involved in the citric acid cycle in combination with fumarate and oxaloacetate, and its levels were found to change in the

serum of individuals with CRC cancer and polyps when compared to the control ones.^{7,97}

In turn, a high demand for amino acids by the growing tissues can also cause alterations in the metabolic routes associated with these compounds, and a consequent decrease in their levels in carcinogenic samples has been reported, e.g., arginine, glutamine, proline, alanine, aspartate, and glutamate,^{7,95,97,96,98} accompanied by an accumulation of ketone scaffolds such as acetate, acetoacetate, and 3-hydroxybutyrate.^{95,91} Additionally, Deng et al.,⁹⁷ Gu et al.,⁷ and Zamani et al.⁹⁵ reported a decrease in the levels of unsaturated and polyunsaturated fatty acids, possibly due to perturbations in the metabolisms of glycerolipids and fatty acids. Moreover, the biosynthesis of primary bile acids and the metabolism of vitamin B6 were referenced among others (cyanoamino acid, thymine, methane, glutathione, fucose, and mannose metabolisms) by Zamani et al.⁹⁵ and Vahabi et al.⁹⁸ Alternatively, Farshidfar et al.,⁹⁴ focused on the comparison among serum samples from individuals with colorectal and liver-only metastases, reporting an accelerated galactose metabolism being involved in colorectal cancer samples. Also, changes in the metabolism of purine were commonly observed by Vahabi et al.⁹⁸ and Deng et al.⁹⁷ as well as in choline metabolism by Gu et al.⁷

Multivariate Data Analysis Methods. Most of the studies reviewed herein applied the unsupervised technique of PCA as a first step for the discrimination between groups and to obtain potential biomarkers of colorectal cancer with the exception of Bertini et al.⁹³ who employed for this purpose another technique of this sort, canonical analysis (CA). It is worth mentioning the research developed by Di Donato et al.,⁹⁹ in which a PCA in combination with CA and kNN was applied for the discrimination between groups. On the contrary, Backshall et al.,⁹² Chen et al.,⁹⁶ Deng et al.,⁹⁷ and Vahabi et al.⁹⁸ did not employ PCA in their research. Further, for the application of unsupervised techniques, supervised linear multivariate techniques such as PLS-DA^{7,95,97,91,93,92} and OPLS-DA^{7,94,98} were generally applied, and the associated biomarkers were determined using different methods. Some of these studies^{7,97,94} selected the variables associated with the discrimination between disease and healthy individuals based on their variable importance in projection (VIP) index values given by supervised models, in which those variables with a VIP value greater than 1 were considered statistically significant for the model. Chen et al.⁹⁶ employed other statistical approaches like seemingly unrelated regression for the identification of significant biomarkers.

Some studies also included nonlinear methods for the classification and identification of the most discriminant metabolites, for instance, the study by Bertini et al.,⁹³ in which a SVM model was implemented to the PLS scores by applying the nonparametric Kruskal–Wallis rank-sum test for the continuous variables and the Fisher exact test for the categorical ones. Moreover, Gu et al.⁷ implemented an RF classifier in combination with the correlation coefficients of several discriminating metabolites found from a previous OPLS analysis and selected the most important biomarkers according to their frequency of being chosen by the algorithm, and later, a SVM model was applied in order to validate the obtained results. The results were supported by employing the area under the ROC curve.

In order to validate the linear models of PLS-DA and OPLS-DA, several of the studies employ the parametric cross-

Table 2. List of Metabolomics Studies of Colorectal Cancer in Other Matrices until March 2019 in Chronological Order^a

reference	technique	matrix	NMR probehead	NMR sample size	sample volume	multivariate analysis	aim	conclusions ^c
Bezabeh et al. ¹⁰² (2009)	¹ H NMR (400 MHz)	fecal extracts	flow probe	111 CRC, ^b 412 controls	500 μ L	LDS (linear discriminant analysis)	noninvasive detection of CRC	identified spectral differences important for the detection of colorectal cancer
Chan et al. ¹⁰⁵ (2009)	¹ H NMR (600 MHz), GC-MS	colon mucosae	HR-MAS probe	31 patients	10 mg	OPLS-DA	description of metabolic signatures that discriminate malignant from normal mucosae	higher levels of Lac, Gly, PC, PE, Urd, fecal bile acids, and Cholt and lower levels of Glc, Ara, Mal, Fum, and lipids in CRC samples
Piotto et al. ¹⁰⁶ (2009)	¹ H NMR (500 MHz)	tumoral and healthy tissues	4 mm double resonance (¹ H, ¹³ C) gradient HR-MAS probe	44 patients	15/20 mg	PCA, PLS-DA	characterization of the metabolic fingerprint of tumoral and healthy tissue samples from patients affected by primary colorectal adenocarcinoma	elevated amounts of Tau, Glu, Asp, and Lac in colorectal biopsy tissues and a high amount of MI and β -in healthy tissues
Monleon et al. ¹⁰³ (2009)	¹ H NMR (600 MHz)	fecal water extracts (stool samples)	triple-resonance ¹ H/ ¹³ C/ ¹⁵ N probe	21 CRC, 11 controls	500 μ L	PCA	identification of potential diagnostic markers of CRC	a low SCFA (Acet and But) concentration seems to be the most effective marker for CRC; Pro and Cys also displayed a correlation with the disease
Jiménez et al. ¹⁰⁷ (2013)	¹ H NMR (400 MHz)	tumor and adjacent mucosa	TBI HR-MAS probe	26 patients	10 mg	OPLS-DA	find potential biomarkers of CRC in both matrices	discrimination between T- and N-stages; increased levels of Tau, Isogh, Chol, Lac, Phe, and Tyr, and decrease in lipids and triglycerides for tumor tissues
Wang et al. ¹⁰⁸ (2017)	¹ H NMR (400 MHz)	urine	-	55 CRC, 40 controls	600 μ L	PCA, OPLS-DA	find metabolic variations between CRC and healthy controls and differentiate between CRC stages	discrimination between T- and N-stages and accurate predictive capability of 5-year survival for adjacent mucosa
Kim et al. ¹⁰⁹ (2019)	¹ H NMR (500 MHz)	urine	-	92 CRC, 156 controls	500 μ L	OPLS-DA	explore the potential of urine NMR metabolomics as a diagnostic tool for early detection of CRC, compare patients with CRC neoplasia at various stages and healthy controls	16 potential biomarkers were identified for CRC; metabolic profiles from early stage CRC and esophageal cancer patients were also distinguishable

^aSymbol: “-” denotes studies where the specific information was not provided. ^bColorectal cancer. ^cIn samples corresponding to colorectal cancer, abbreviations are as follows. Ala: alanine; 2-AB: 2-aminobutyrate; 3-AIB: 3-aminoisobutyrate; Ara: arachidonic acid; Acet: acetate; Asp: aspartate; But: butyrate; Chol: choline; Cholt: cholesterol; Cit: citrate; Cr: creatine; Cys: cysteine; Fum: fumarate; Glc: glucose; Gln: glutamine; Gly: glycine; Hip: hippurate; Isogh: isoglutamine; Kyn: kynurenate; Lac: lactate; Mal: malate; MI: myo-inositol; Myr: myristate; *p*-cre: *p*-cresol; PEG: polyethylene glycol; PC: phosphocholine; PE: phosphoethanolamine; Phe: phenylalanine; Pro: proline; Put: putrescine; SCFA: short-chain fatty acids; Tau: taurine; TCA: tricarboxylic acids; Tyr: tyrosine; Urd: uridine.

validation test CV-ANOVA, while the precision of the nonlinear models was evaluated using AUC-ROC curves, which make it possible to verify the proportion of false positives derived from confusion matrices in combination with their respective confidence intervals (CIs). Lastly, most of the studies carried out metabolomics pathway analysis by performing the Holm-Bonferroni correction with the aim of determining the most enriched ones.

Analytical Platforms, Acquisition Parameters, and Processing. In general, the ^1H NMR-based metabolomic studies mentioned in this Review demonstrate a great power of prediction, classification, and selection of biomarkers associated with colorectal cancer; however, although the current trend continues to increase, the number of articles that relied solely on ^1H NMR is still lower. Indeed, some of the studies included in this Review performed the analysis of the metabolic profiles in combination with other analytical techniques. For example, in addition to NMR, Deng et al.⁹⁷ applied LC-MS, while Farshidfar et al.⁹⁴ utilized GC-MS. It is of great importance to highlight these differences since, depending on the equipment used, different metabolites can be detected depending on the sensitivity and specificity of each platform.

Concerning the publications using ^1H NMR, the data acquisition and processing parameters employed could also be a cause of variability of the results. In general, most of the discussed studies in this Review implemented the sequence CPMG to suppress resonances involving high-molecular weight molecules,^{91–93,95–99} some of them solely or in combination with the presaturation of the water signal, as a method of erasing the water signal from the serum samples.^{7,91,92,94,96,97,99} Regarding the data processing parameters, including the type of normalization, scaling, and/or transformation applied, these studies showed a general lack of consensus among them and generally applied different statistical approaches. In this sense, depending on the scaling method used, several types of signals can be prioritized, which can lead to misleading conclusions. Therefore, except for Bertini et al.,⁹³ most articles presented quite a small sample size, an aspect that could have conditioned many of the statistical results achieved, resulting in the possible variation of some of the metabolites envisaged as biomarkers.

Differences in the selection of research participants, targeted population groups, and sampling procedures described in each research study should be also considered. Generally, there is a tendency to study differential metabolites between a group of colorectal cancer, overall involving metastases, and a control group. Nevertheless, some of the studies mentioned herein also focused on the study of metabolic differences between samples from patients with polyps and controls. In contrast, the study by Farshidfar et al.⁹⁴ focused equally on the location of the cancer, making a distinction between coloproctorectal, liver-only, and extrahepatic as well as on the stage of this disease, a factor that the team of Vahabi et al.⁹⁸ also studied. On the other hand, Backshall et al.⁹² employed samples derived from patients before being treated with capecitabine in order to relate their profiles to subsequent treatment toxicity, while Di Donato et al.⁹⁹ distinguished between early colorectal cancer patients with and without relapse and elderly patients with metastatic disease.

■ OTHER MATRICES ANALYZED THROUGH NMR METABOLOMICS IN THE QUEST OF CRC BIOMARKERS

Apart from serum, there are plenty of studies about colorectal cancer that apply NMR metabolomics in other matrices. Table 2 shows seven of the most relevant studies with the principal aims and results of each one being described. These include three main types of matrices: fecal, tissue, and urine samples.

In the first group using fecal samples, two of these articles need to be highlighted: Bezabeh et al.¹⁰² and Monleón et al.,¹⁰³ who investigated metabolic differences between feces samples from healthy controls and colorectal cancer patients. Both used NMR operating at different frequencies, and while Bezabeh et al.¹⁰² only found spectral differences between groups (using a 400 MHz spectrometer), Monleón et al.¹⁰³ were able to find some biomarkers associated with colorectal cancer (using a 600 MHz spectrometer), such as acetate or butyrate. For NMR analyses, feces samples are not difficult to prepare,^{102–104} and therefore, it could be a simple way to study this disease in a less invasive way.

The second group related to tissue samples includes the works of Chan et al.,¹⁰⁵ Piotta et al.,¹⁰⁶ and Jiménez et al.¹⁰⁷ (Table 2), who applied high resolution magic angle spinning (HR-MAS) NMR spectroscopy to study samples coming from tumors or adjacent normal mucosae obtained through biopsies. An advantage of using HR-MAS is that a smaller amount of sample is needed for the analysis and that the measurements are carried out directly in the solid state. However, handling tissue samples implies the use of an invasive collection method that contradicts one of the main advantages of applying NMR in metabolomics studies. HR-MAS analysis in tissue samples was able to find principally lactate and glucose, among others, as biomarkers for the disease.^{105–107} As previously discussed, some of these same biomarkers were also elucidated in the serum of patients with CRC^{7,91,97} but with the advantage that a noninvasive collection method was applied.

The third group includes the studies of Wang et al.¹⁰⁸ and Kim et al.¹⁰⁹ with urine samples. They employed NMR operating at 400 and 500 MHz, respectively, to assess metabolic changes in urine samples and were able to find some specific biomarkers, such as taurine, alanine, and 3-aminoisobutyrate, as well as distinguish between early stages of colorectal cancer and esophageal cancer.^{108,109} There are some studies, such as the one of Vahabi et al.,⁹⁸ where they were also able to distinguish between different stages of colorectal cancer employing serum samples.

From Table 2, it is deduced that there is some variability on NMR operation frequencies as a function of the matrix chosen, but in terms of unraveled biomarkers, there is a trend that agrees with the tendency observed in serum samples: the Warburg effect is emphasized once again due to the increase of lactate and the decrease in glucose detected in most matrices.

In Figure 5, we have summarized the different metabolites found as biomarkers as a function of the matrix under study. In addition, we have illustrated the different sets of metabolites with an arrow pointing upward or downward depending on whether the biomarker increases or decreases for the colorectal cancer group, respectively.

■ CONCLUSIONS

NMR spectroscopy is presented as a powerful technique for the identification of specific metabolites even in complex

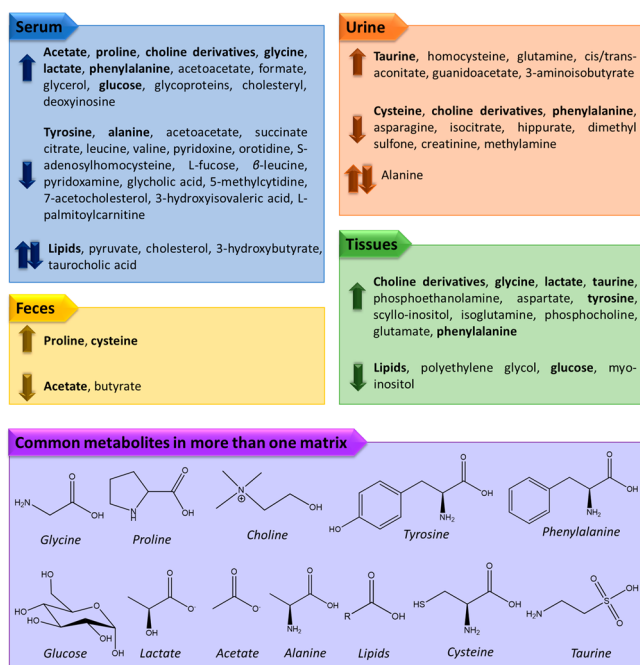


Figure 5. Biomarkers as a function of the matrix under study for colorectal cancer. Arrows pointing up or down reveal upregulated and downregulated biomarkers, respectively. Arrows pointing up and down simultaneously illustrate contradictory results on biomarkers that were reported to increase or decrease in different studies. Metabolites in bold have been detected in more than one matrix, and their molecular structures are given in the box below.

mixtures, showing great applicability in the field of clinical metabolomics. There are currently scarce contributions of NMR metabolomics approaches in the study of colorectal cancer serum samples although, in general, all of them show promising outcomes. On the basis of the multiple statistical methods employed by each study, it can be concluded that there is no standard procedure among them for the identification of relevant biomarkers, which can lead to multiple conclusions, since data processing and data preparation are crucial steps to achieve correct interpretation of the results. Nevertheless, most studies discussed in this Review agreed on the role of colorectal cancer metabolites involved in glycolysis, specifically referring to the Warburg effect, which is a characteristic of carcinogenic samples. Also, alterations in the amino acids content and in the metabolism of glycerolipids and fatty acids were reported in most of the studies.

It would be of great interest to continue exploring the associated serum metabolic profiles to different stages of the disease, consolidating sample sizes, aims, and interest groups, and to increase the low number of studies in NMR metabolomics addressing this topic. Also, researchers should take into account the presence of other variables, such as the patient's age, the occurrence of other diseases, and the physical state of the individual, since some investigations with those aspects have also shown encouraging outcomes.^{106,108} Finally, research in this field should be stimulated and correctly driven to understand and predict basic biological issues associated with colorectal cancer.

AUTHOR INFORMATION

Corresponding Authors

Ignacio Fernández – Department of Chemistry and Physics, Research Centre CIAIMBITAL, University of Almería, 04120 Almería, Spain; orcid.org/0000-0001-8355-580X; Phone: +34 950214465; Email: ifernan@ual.es

Ana C. Abreu – Department of Chemistry and Physics, Research Centre CIAIMBITAL, University of Almería, 04120 Almería, Spain; Email: acabreu@ual.es

Authors

Ana M. Salmerón – Department of Chemistry and Physics, Research Centre CIAIMBITAL, University of Almería, 04120 Almería, Spain

Ana I. Tristán – Department of Chemistry and Physics, Research Centre CIAIMBITAL, University of Almería, 04120 Almería, Spain

Complete contact information is available at: <https://pubs.acs.org/10.1021/acs.analchem.1c04360>

Funding

This research has been funded by the State Research Agency of the Spanish Ministry of Science, Innovation and Universities (program CTQ2017-84334-R), Junta de Andalucía (programs 102C2000004, UAL2020-AGR-B1781, P20_01041, and CV20-78799), Gobierno de España MCIN/AEI/10.13039/501100011033/Unión Europea “Next GenerationEU”/PRTR (programs PDC2021-121248-I00 and PLEC2021-007774), and University of Almería (programs UAL TransferE 2019, 2020, and 2021). A.I.T. and A.C.A. thank Junta de Andalucía and University of Almería for their respective predoctoral and HIPATIA postdoctoral fellowships.

Notes

The authors declare no competing financial interest.

Biographies

Ana M. Salmerón studied Chemistry at the University of Almería (Spain); subsequently, she obtained her Master's degree from the Open University of Catalonia along with the University of Barcelona in Bioinformatics and Biostatistics. She is currently part of the NMR metabolomics research group of Prof. Dr. Ignacio Fernández, where she has recently started her Ph.D. degree. Her research interests include metabolomics analysis of biological matrices through the application of analytical techniques such as NMR.

Ana I. Tristán studied Chemistry at the University of Almería (Spain) where she obtained her Master's degree in Chemistry's Advanced Laboratory and started her Ph.D. degree in 2020. She has been involved in numerous research projects as a part of the NMR metabolomics research group of Prof. Dr. Ignacio Fernández. Her current research activities include the application of analytical techniques, such as NMR and IR, in plant and clinical samples for metabolomics purposes.

Ana C. Abreu studied Bioengineering at the Faculty of Engineering of University of Porto (FEUP, Portugal). She completed her Ph.D. with Cum Laude distinction (02/2012–01/2017) in Chemical Engineering with a focus on NMR-based metabolomics analysis of plants with antibacterial properties at FEUP and in collaboration with the group of Dr. Young Hae Choi (01/2014–01/2016) at the Institute of Biology of Leiden (Netherlands). During this time, she was also a guest researcher at the Erasmus Medical Center of Rotterdam (Netherlands). In March 2017, she became a postdoctoral researcher studying NMR-based metabolomics in the group of Prof. Dr. Ignacio

Fernández at the University of Almería (Spain). In May 2020, she was given a HIPATIA fellowship from the University of Almería.

Ignacio Fernández is Full Professor at the University of Almería since July 2018. He enjoyed two predoctoral stays at Bath University with Matthew G. Davidson and at Wurzburg Universität (Germany) with Dietmar Stalke. In 2003, he moved to Zurich (ETHZ) as a postdoctoral researcher in the group of Prof. Paul S. Pregosin. He stayed in Zurich for three years until he was awarded with a Ramón y Cajal research contract in 2006 to return to Almería. After a few months in Spain, he became a visiting professor at Cornell University (New York, USA) in Paul J. Chirik's group, where he stayed for almost a year until he returned in 2008 to the University of Almería. He was also a visiting professor in 2008, 2010, and 2013 at the Karlsruhe Institute of Technology (KIT, Germany). He has published more than 110 articles and four book chapters related to NMR topics and is an inventor of four international patents and two software packages related to molecular diffusion. Since 2014, he has been the head of the research group of Advanced NMR Methods and Metal-Based Catalysts (www.nmrmbc.com). His research interests focus on (i) the application of multinuclear NMR to chemistry problems, (ii) the development of new structural platforms based on transition metals and lanthanides for application in synthetic and catalytic processes, (iii) NMR metabolomics studies in agri-food, clinical, and materials science fields, and (iv) mathematical algorithms applied to NMR molecular diffusion.

REFERENCES

- (1) American Cancer Society. *Colorectal Cancer Facts & Figures 2020–2022*; ACS: Atlanta, GA, 2020.
- (2) Shussman, N.; Wexner, S. D. *Curr. Gastroenterol. Rep.* **2014**, *2*, 1–15.
- (3) Bray, F.; Ferlay, J.; Soerjomataram, I.; Siegel, R. L.; Torre, L. A.; Jemal, A. *Ca-Cancer J. Clin.* **2018**, *68*, 394–424.
- (4) Sung, H.; Ferlay, J.; Siegel, R. L.; Laversanne, M.; Soerjomataram, I.; Jemal, A.; Bray, F. *Ca-Cancer J. Clin.* **2021**, *71*, 209–249.
- (5) Sawicki, T.; Ruskowska, M.; Danielewicz, A.; Niedźwiedzka, E.; Arlukowicz, T.; Przybyłowicz, K. E. *Cancers* **2021**, *13*, 2025.
- (6) Hagggar, F.; Boushey, R. *Clin. Colon Rectal Surg.* **2009**, *22*, 191–197.
- (7) Gu, J.; Xiao, Y.; Shu, D.; Liang, X.; Hu, X.; Xie, Y.; Lin, D.; Li, H. *Dis. Markers* **2019**, *2019*, 3491852.
- (8) Yang, C.; Sriranjana, V.; Abou-Setta, A. M.; Poluha, W.; Walker, J. R.; Singh, H. *Am. J. Gastroenterol.* **2018**, *113*, 1810–1818.
- (9) Gatto, N. M.; Frucht, H.; Sundararajan, V.; Jacobson, J. S.; Grann, V. R.; Neugut, A. I. *J. Natl. Cancer Inst.* **2003**, *95*, 230–236.
- (10) Pinu, F. R.; Goldansaz, S. A.; Jaïne, J. *Metabolites* **2019**, *9*, 108.
- (11) Manzoni, C.; Kia, D. A.; Vandrovicova, J.; Hardy, J.; Wood, N. W.; Lewis, P. A.; Ferrari, R. *Briefings Bioinf.* **2018**, *19*, 286–302.
- (12) Kohler, I.; Hankemeier, T.; van der Graaf, P. H.; Knibbe, C. A. J.; van Hasselt, J. G. C. *Eur. J. Pharm. Sci.* **2017**, *109*, S15–S21.
- (13) Pang, H.; Jia, W.; Hu, Z. *Clin. Pharmacol. Ther.* **2019**, *106*, 544–556.
- (14) Ashrafian, H.; Sounderajah, V.; Glen, R.; Ebbels, T.; Blaise, B. J.; Kalra, D.; Kultima, K.; Spjuth, O.; Tenori, L.; Salek, R. M.; Kale, N.; Haug, K.; Schober, D.; Rocca-Serra, P.; O'Donovan, C.; Steinbeck, C.; Cano, I.; De Atauri, P.; Cascante, M. *Med. Prin. Pract.* **2021**, *30*, 301–310.
- (15) Pinu, F. R.; Beale, D. J.; Paten, A. M.; Kouremenos, K.; Swarup, S.; Schirra, H. J.; Wishart, D. *Metabolites* **2019**, *9*, 76.
- (16) Tounta, V.; Liu, Y.; Cheyne, A.; Larrouy-Maumus, G. *Mol. Omics* **2021**, *17*, 376–393.
- (17) Frigolet, M. E.; Gutiérrez-Aguilar, R. *Rev. Digit. Univ.* **2017**, *18*, 18.
- (18) Roberts, L. D.; Souza, A. L.; Gerszten, R. E.; Clish, C. B. *Curr. Protoc. Mol. Biol.* **2012**, *98*, 30.32.31–30.32.24.
- (19) Tolstikov, V.; Moser, A. J.; Sarangarajan, R.; Narain, N. R.; Kiebish, M. A. *Metabolites* **2020**, *10*, 224.
- (20) Nagana Gowda, G.; Raftery, D. *Curr. Metabolomics* **2013**, *1*, 227–240.
- (21) Bracewell-Milnes, T.; Saso, S.; Abdalla, H.; Nikolau, D.; Norman-Taylor, J.; Johnson, M.; Holmes, E.; Thum, M. Y. *Hum. Reprod. Update* **2017**, *23*, 723–736.
- (22) Mohd Yusof, H.; Ab-Rahim, S.; Suddin, L. S.; Ahmad Saman, M. S.; Mazlan, M. *Malays. J. Med. Sci.* **2018**, *25*, 16–34.
- (23) Amir Hashim, N. A.; Ab-Rahim, S.; Wan Ngah, W. Z.; Nathan, S.; Ab Mutalib, N. S.; Sagap, I.; Jamal, A. R. A.; Mazlan, M. *BiolImpacts* **2021**, *11*, 33–43.
- (24) Yu, L.; Li, K.; Zhang, X. *Oncotarget* **2017**, *8*, 115774–115786.
- (25) Johnson, C. H.; Gonzalez, F. J. *J. Cell. Physiol.* **2012**, *227*, 2975–2981.
- (26) Miggliels, P.; Wouters, B.; van Westen, G. J. P.; Dubbelman, A.-C.; Hankemeier, T. *TrAC, Trends Anal. Chem.* **2019**, *120*, 115323.
- (27) Abreu, A. C.; Fernández, I. In *Recent Trends in Biofilm Science and Technology*; Academic Press, 2020; pp 23–70.
- (28) Dettmer, K.; Aronov, P. A.; Hammock, B. D. *Mass Spectrom. Rev.* **2007**, *26*, 51–78.
- (29) Gowda, G. N.; Zhang, S.; Gu, H.; Asiago, V.; Shanaiah, N.; Raftery, D. *Expert Rev. Mol. Diagn.* **2008**, *8*, 617–633.
- (30) Emwas, A. H.; Roy, R.; McKay, R. T.; Tenori, L.; Saccenti, E.; Gowda, G. A. N.; Raftery, D.; Alahmari, F.; Jaremko, L.; Jaremko, M.; Wishart, D. S. *Metabolites* **2019**, *9*, 123.
- (31) Emwas, A. H. M. In *Methods in Molecular Biology*; Springer: New York, 2015; pp 161–193.
- (32) Tang, H.; Wang, Y.; Nicholson, J. K.; Lindon, J. C. *Anal. Biochem.* **2004**, *325*, 260–272.
- (33) Abreu, A. C.; Fernández, I. *Molecules* **2020**, *25*, 3738.
- (34) Bingol, K.; Brüschweiler, R. *Anal. Chem.* **2014**, *86*, 47–57.
- (35) Hu, J. M.; Sun, H. T. *Radiat. Oncol.* **2018**, *13*, 40.
- (36) Michálková, L.; Horník, Š.; Sýkora, J.; Habartová, L.; Setnička, V. *Analyst* **2018**, *143*, 5974–5978.
- (37) Ranjan, R.; Sinha, N. *NMR Biomed.* **2019**, *32*, No. e3916.
- (38) Abreu, A. C.; Morales, M.; Pérez, C.; Sánchez-Santed, F.; Fernández, I. *ACS Chem. Biol.* **2021**, *16*, 753–765.
- (39) Toczyłowska, B.; Zieminska, E.; Senator, P.; Lazarewicz, J. W. *Mol. Neurobiol.* **2020**, *57*, 3089–3105.
- (40) Neto, F. T. L.; Marques, R. A.; De Freitas Cavalcanti Filho, A.; Araujo, L. C. N.; Lima, S. V. C.; Pinto, L.; Silva, R. O. *J. Assist. Reprod. Genet.* **2020**, *37*, 2233–2247.
- (41) Botros, L.; Sakkas, D.; Seli, E. *Mol. Hum. Reprod.* **2008**, *14*, 679–690.
- (42) Hung, C. I.; Lin, G.; Chiang, M. H.; Chiu, C. Y. *Sci. Rep.* **2021**, *11*, 15608.
- (43) Zheng, H.; Zheng, P.; Zhao, L.; Jia, J.; Tang, S.; Xu, P.; Xie, P.; Gao, H. *Clin. Chim. Acta* **2017**, *464*, 223–227.
- (44) Salehi, M. A.; Nilsson, I. A. K.; Figueira, J.; Thornton, L. M.; Abdulkarim, I.; Pålsson, E.; Bulik, C. M.; Landén, M. *Eur. Neuropsychopharmacol.* **2021**, *49*, 1–10.
- (45) Monteleone, P.; Monteleone, A. M.; Troisi, J.; Dalle Grave, R.; Corrivetti, G.; Calugi, S.; Scala, G.; Patriciello, G.; Zanetti, A.; Maj, M. *Mol. Psychiatry* **2021**, *26*, 3980.
- (46) Zhang, S.; Nagana Gowda, G. A.; Ye, T.; Raftery, D. *Analyst* **2010**, *135*, 1490.
- (47) Kanani, H.; Chrysanthopoulos, P. K.; Klapa, M. I. *J. Chromatogr. B: Anal. Technol. Biomed. Life Sci.* **2008**, *871*, 191–201.
- (48) Vigor, C.; Bertrand-Michel, J.; Pinot, E.; Oger, C.; Vercauteren, J.; Le Faouder, P.; Galano, J.-M.; Lee, J. C.-Y.; Durand, T. *J. Chromatogr. B: Anal. Technol. Biomed. Life Sci.* **2014**, *964*, 65–78.
- (49) Esturau, N.; Espinosa, J. F. *J. Org. Chem.* **2006**, *71*, 4103–4110.
- (50) Emwas, A.-H.; Roy, R.; McKay, R. T.; Ryan, D.; Brennan, L.; Tenori, L.; Luchinat, C.; Gao, X.; Zeri, A. C.; Gowda, G. A. N.; Raftery, D.; Steinbeck, C.; Salek, R. M.; Wishart, D. S. *J. Proteome Res.* **2016**, *15*, 360–373.

- (51) Djukovic, D.; Nagana Gowda, G. A.; Raftery, D. In *Proteomic and Metabolomic Approaches to Biomarker Discovery*; Elsevier, 2013; pp 279–297.
- (52) Aiello, F.; Gerretzen, J.; Simons, M. G.; Davies, A. N.; Dani, P. *Magn. Reson. Chem.* **2021**, *59*, 172–186.
- (53) Crook, A. A.; Powers, R. *Molecules* **2020**, *25*, 5128.
- (54) Dona, A. C.; Kyriakides, M.; Scott, F.; Shephard, E. A.; Varshavi, D.; Veselkov, K.; Everett, J. R. *Comput. Struct. Biotechnol. J.* **2016**, *14*, 135–153.
- (55) Peris-Díaz, M. D.; Krężel, A. *TrAC, Trends Anal. Chem.* **2021**, *135*, 116157.
- (56) Brereton, R. G.; Jansen, J.; Lopes, J.; Marini, F.; Pomerantsev, A.; Rodionova, O.; Roger, J. M.; Walczak, B.; Tauler, R. *Anal. Bioanal. Chem.* **2017**, *409*, 5891–5899.
- (57) Kumar, N.; Bansal, A.; Sarma, G. S.; Rawal, R. *Talanta* **2014**, *123*, 186–199.
- (58) Sarker, S. D.; Nahar, L. In *Evidence-Based Validation of Herbal Medicine*; Mukherjee, P. K., Ed.; Elsevier: Boston, 2015; pp 405–425.
- (59) Ebrahimi, P.; Viereck, N.; Bro, R.; Engelsen, S. B. In *Modern Magnetic Resonance*; Webb, G. A., Ed.; Springer International Publishing: Cham, 2018; pp 1649–1668.
- (60) Pereira Braga, C.; Adamec, J. In *Encyclopedia of Bioinformatics and Computational Biology*; Ranganathan, S., Gribskov, M., Nakai, K., Schönbach, C., Eds.; Academic Press: Oxford, 2019; pp 463–475.
- (61) Zhang, X.; Zhu, X.; Wang, C.; Zhang, H.; Cai, Z. *Oncotarget* **2016**, *7*, 63437–63448.
- (62) Barderas, M. G.; Laborde, C. M.; Posada, M.; De La Cuesta, F.; Zubiri, I.; Vivanco, F.; Alvarez-Llamas, G. *J. Biomed. Biotechnol.* **2011**, *2011*, 1–9.
- (63) Nalbantoglu, S. In *Molecular Medicine*; IntechOpen, 2019.
- (64) Vignoli, A.; Ghini, V.; Meoni, G.; Licari, C.; Takis, P. G.; Tenori, L.; Turano, P.; Luchinat, C. *Angew. Chem., Int. Ed.* **2019**, *58*, 968–994.
- (65) Sahoo, N. K.; Tejaswini, G.; Madhusmita, S.; Muralikrishna, K. *S. Int. J. Pharm. Sci.* **2020**, *6*, 016–020.
- (66) Lavine, B.; Workman, J. *Anal. Chem.* **2008**, *80*, 4519–4531.
- (67) Dieterle, F.; Ross, A.; Schlotterbeck, G.; Senn, H. *Anal. Chem.* **2006**, *78*, 4281–4290.
- (68) Wu, Y.; Li, L. *J. Chromatogr. A* **2016**, *1430*, 80–95.
- (69) Kohl, S. M.; Klein, M. S.; Hochrein, J.; Oefner, P. J.; Spang, R.; Gronwald, W. *Metabolomics* **2012**, *8*, 146–160.
- (70) Worley, B.; Powers, R. *Curr. Metabolomics* **2012**, *1*, 92–107.
- (71) Saccenti, E.; Hoefsloot, H. C. J.; Smilde, A. K.; Westerhuis, J. A.; Hendriks, M. M. W. B. *Metabolomics* **2014**, *10*, 361–374.
- (72) Van Den Berg, R. A.; Hoefsloot, H. C.; Westerhuis, J. A.; Smilde, A. K.; Van Der Werf, M. J. *BMC Genomics* **2006**, *7*, 142.
- (73) Perez De Souza, L.; Alseikh, S.; Brotman, Y.; Fernie, A. R. *Expert Rev. Proteomics* **2020**, *17*, 243–255.
- (74) Percival, B.; Gibson, M.; Leenders, J.; Wilson, P. B.; Grootveld, M. In *Theoretical and Computational Chemistry Series*; Royal Society of Chemistry, 2020; pp 1–40.
- (75) Tarca, A. L.; Carey, V. J.; Chen, X.-W.; Romero, R.; Drăghici, S. *PLoS Comput. Biol.* **2007**, *3*, No. e116.
- (76) Turi, K. N.; Romick-Rosendale, L.; Ryckman, K. K.; Hartert, T. *V. J. Allergy Clin. Immunol.* **2018**, *141*, 1191–1201.
- (77) Kourou, K.; Exarchos, T. P.; Exarchos, K. P.; Karamouzis, M. V.; Fotiadis, D. I. *Comput. Struct. Biotechnol. J.* **2015**, *13*, 8–17.
- (78) Berry, M.; Mohamed, A. *Supervised and Unsupervised Learning for Data Science*; Springer, 2020;.
- (79) Kosmides, A. K.; Kamisoglu, K.; Calvano, S. E.; Corbett, S. A.; Androulakis, I. P. *Crit. Rev. Biomed. Eng.* **2013**, *41*, 205–221.
- (80) Smith, P. F.; Zheng, Y. *Front. Neurol.* **2021**, *12*, 627294–627294.
- (81) Xia, J.; Broadhurst, D. I.; Wilson, M.; Wishart, D. S. *Metabolomics* **2013**, *9*, 280–299.
- (82) Hanczar, B.; Hua, J.; Sima, C.; Weinstein, J.; Bittner, M.; Dougherty, E. R. *Bioinformatics* **2010**, *26*, 822–830.
- (83) Zhu, C.; Vitali, B.; Donders, G.; Parolin, C.; Li, Y.; Laghi, L. *Metabolites* **2019**, *9*, 15.
- (84) Blakesley, R. E.; Mazumdar, S.; Dew, M. A.; Houck, P. R.; Tang, G.; Reynolds, C. F., 3rd; Butters, M. A. *Neuropsychology* **2009**, *23*, 255–264.
- (85) Lee, S.; Lee, D. K. *Korean J. Anesthesiol.* **2018**, *71*, 353–360.
- (86) Marco-Ramell, A.; Palau-Rodriguez, M.; Alay, A.; Tulipani, S.; Urpi-Sarda, M.; Sanchez-Pla, A.; Andres-Lacueva, C. *BMC Bioinf.* **2018**, *19*, 1.
- (87) Clarivate Analytics. *Web Of Science*; Retrieved from <https://www.webofscience.com/wos/alldb/basic-search> (accessed October 2021).
- (88) *VOSviewer Software*, v.1.6.15; Centre for Science and Technology Studies, Leiden University: The Netherlands, 2020.
- (89) *CitNetExplorer Software*, v.1.0.0; Centre for Science and Technology Studies, Leiden University: The Netherlands, 2020.
- (90) Warburg, O. *Science* **1956**, *123*, 309.
- (91) Ludwig, C.; Ward, D. G.; Martin, A.; Viant, M. R.; Ismail, T.; Johnson, P. J.; Wakelam, M. J. O.; Günther, U. L. *Magn. Reson. Chem.* **2009**, *47*, S68–S73.
- (92) Backshall, A.; Sharma, R.; Clarke, S. J.; Keun, H. C. *Clin. Cancer Res.* **2011**, *17*, 3019–3028.
- (93) Bertini, I.; Cacciatore, S.; Jensen, B. V.; Schou, J. V.; Johansen, J. S.; Kruhøffer, M.; Luchinat, C.; Nielsen, D. L.; Turano, P. *Cancer Res.* **2012**, *72*, 356–364.
- (94) Farshidfar, F.; Weljie, A. M.; Kopciuk, K.; Buie, W. D.; Maclean, A.; Dixon, E.; Sutherland, F. R.; Molckovsky, A.; Vogel, H. J.; Bathe, O. F. *Genome Med.* **2012**, *4*, 42.
- (95) Zamani, Z.; Arjmand, M.; Vahabi, F.; Eshaq Hosseini, S. M.; Fazeli, S. M.; Irvani, A.; Bayat, P.; Oghalayee, A.; Mehrabanfar, M.; Haj Hosseini, R.; Tashakorpour, M.; Tafazzoli, M.; Sadeghi, S. *Biochem. Res. Int.* **2014**, *2014*, 1–7.
- (96) Chen, C.; Deng, L.; Wei, S.; Nagana Gowda, G. A.; Gu, H.; Chiorean, E. G.; Abu Zaid, M.; Harrison, M. L.; Pekny, J. F.; Loehrer, P. J.; Zhang, D.; Zhang, M.; Raftery, D. *J. Proteome Res.* **2015**, *14*, 2492–2499.
- (97) Deng, L.; Gu, H.; Zhu, J.; Nagana Gowda, G. A.; Djukovic, D.; Chiorean, E. G.; Raftery, D. *Anal. Chem.* **2016**, *88*, 7975–7983.
- (98) Vahabi, F.; Sadeghi, S.; Arjmand, M.; Mirkhani, F.; Hosseini, E.; Mehrabanfar, M.; Hajhosseini, R.; Irvani, A. *Iran J. Basic Med. Sci.* **2017**, *835*–840.
- (99) Di Donato, S.; Vignoli, A.; Biagioni, C.; Malorni, L.; Mori, E.; Tenori, L.; Calamai, V.; Parnofello, A.; Di Pierro, G.; Migliaccio, I.; Cantafio, S.; Baraghini, M.; Mottino, G.; Becheri, D.; Del Monte, F.; Miceli, E.; Mccartney, A.; Di Leo, A.; Luchinat, C.; Biganzoli, L. *Cancers* **2021**, *13*, 2762.
- (100) Wu, J.; Wu, M.; Wu, Q. *J. Clin. Lab. Anal.* **2020**, *34*, No. e23333.
- (101) Qiu, Y.; Cai, G.; Su, M.; Chen, T.; Zheng, X.; Xu, Y.; Ni, Y.; Zhao, A.; Xu, L. X.; Cai, S.; Jia, W. *J. Proteome Res.* **2009**, *8*, 4844–4850.
- (102) Bezabeh, T.; Somorjai, R.; Dolenko, B.; Bryskina, N.; Levin, B.; Bernstein, C. N.; Jeyarajah, E.; Steinhart, A. H.; Rubin, D. T.; Smith, I. C. P. *NMR Biomed.* **2009**, *22*, 593–600.
- (103) Monleón, D.; Morales, J. M.; Barrasa, A.; López, J. A.; Vázquez, C.; Celda, B. *NMR Biomed.* **2009**, *22*, 342–348.
- (104) Kostidis, S.; Kokova, D.; Dementeva, N.; Saltykova, I. V.; Kim, H. K.; Choi, Y. H.; Mayboroda, O. A. *BMC Infect. Dis.* **2017**, *17*, 275.
- (105) Chan, E. C. Y.; Koh, P. K.; Mal, M.; Cheah, P. Y.; Eu, K. W.; Backshall, A.; Cavill, R.; Nicholson, J. K.; Keun, H. C. *J. Proteome Res.* **2009**, *8*, 352–361.
- (106) Piotto, M.; Moussallieh, F. M.; Dillmann, B.; Imperiale, A.; Neuville, A.; Brigand, C.; Bellocq, J. P.; Elbayed, K.; Namer, I. J. *Metabolomics* **2009**, *5*, 292.
- (107) Jiménez, B.; Mirnezami, R.; Kinross, J.; Cloarec, O.; Keun, H. C.; Holmes, E.; Goldin, R. D.; Ziprin, P.; Darzi, A.; Nicholson, J. K. *J. Proteome Res.* **2013**, *12*, 959–968.
- (108) Wang, Z.; Lin, Y.; Liang, J.; Huang, Y.; Ma, C.; Liu, X.; Yang, J. *Oncotarget* **2017**, *8*, 105819–105831.
- (109) Kim, E. R.; Kwon, H. N.; Nam, H.; Kim, J. J.; Park, S.; Kim, Y. H. *Sci. Rep.* **2019**, *9*, 4786.