# Targeted Serum Metabolite Profiling and Sequential Metabolite Ratio Analysis for Colorectal Cancer Progression Monitoring

Jiangjiang Zhu[1], Danijel Djukovic[1], Lingli Deng[1,2], Haiwei Gu[1], Farhan Himmati[1], Mohammad Abu Zaid[3], E. Gabriela Chiorean[3,4] and Daniel Raftery[1, 5*]

[1] Northwest Metabolomics Research Center, Department of Anesthesiology and Pain Medicine, University of Washington, 850 Republican St., Seattle, WA 98109

[2]Departments of Electronic Science and Communication Engineering, State Key Laboratory for the Physical Chemistry of Solid Surfaces, Xiamen University, Xiamen 361005, China

[3]Indiana University, Simon Cancer Center, 535 Barnhill Dr, Indianapolis, IN 46202

[4]University of Washington, 825 Eastlake Ave East,, Seattle, WA 98109

[5]Public Health Sciences Division, Fred Hutchinson Cancer Research Center, 1100 Fairview Ave. N., Seattle, WA 98109

* Corresponding Author:

Professor Daniel Raftery, PhD

Tel: 206-543-9709

Fax: 206-616-4819

Email: draftery@uw.edu

This is the author's manuscript of the article published in final edited form as:

Zhu, J., Djukovic, D., Deng, L., Gu, H., Himmati, F., Abu Zaid, M., … Raftery, D. (2015). Targeted serum metabolite profiling and sequential metabolite ratio analysis for colorectal cancer progression monitoring. *Analytical and Bioanalytical Chemistry*, 407(26), 7857–7863. http://doi.org/10.1007/s00216-015-8984-8

## Abstract

Colorectal cancer (CRC) is one of the most prevalent cancers worldwide, and a major cause of human morbidity and mortality. In addition to early detection, close monitoring of disease progression in CRC can be critical for patient prognosis and treatment decisions. Efforts have been made to develop new methods for improved early detection and patient monitoring; however, research focused on CRC surveillance for treatment response and disease recurrence using metabolomics has yet to be reported. In this proof of concept study, we applied a targeted LC-MS/MS metabolic profiling approach focused on sequential metabolite ratio analysis of serial serum samples to monitor disease progression from 20 CRC patients. The use of serial samples reduces patient to patient metabolic variability. A PLS-DA model using a panel of 5 metabolites (succinate, N2, N2-dimethylguanosine, adenine, citraconic acid, and 1-methylguanosine) was established, and excellent model performance (sensitivity=0.83, specificity = 0.94, AUROC=0.91) was obtained, which is superior to the traditional CRC monitoring marker carcinoembryonic antigen (sensitivity=0.75, specificity=0.76 AUROC=0.80). Monte Carlo cross validation was applied, and the robustness of our model was clearly observed by the separation of true classification models from the random permutation models. Our results suggest the potential utility of metabolic profiling for CRC disease monitoring.



**Keywords:**

Metabolomics, colorectal cancer, targeted metabolic profiling, sequential metabolite ratio analysis, serum metabolites, disease progression monitoring, serial patient samples

## Introduction

Colorectal cancer (CRC) is the third most commonly diagnosed cancer and the third leading cause of cancer death for both men and women. The American Cancer Society estimates that 132,700 people will be diagnosed in 2015 with colorectal cancer and 49,700 people will die of the disease in the US [1]. Although, the percentage of deaths due to CRC has steadily decreased over the years, the number of deaths is still unnecessarily high. Mass spectrometry (MS)-based metabolomics has also been shown recently to be a promising tool for analyzing metabolic alterations due to CRC, and can provide valuable information for diagnostics, pathogenesis clarification, and therapeutic targets for clinical treatments [2]. While these results are promising for CRC screening and prevention, research focused on CRC disease status surveillance has been less often reported. CRC patient mortality is primarily caused by aggressive metastatic disease progression, rather than the primary tumor [3]. Therefore, after primary surgery/initial treatments, CRC patients are usually monitored using blood tests and/or imaging to ensure that they remain disease free and are treated promptly with medical or surgical therapies upon relapse. Additionally, malignant disease progression is ultimately associated with drug resistance; therefore, monitoring disease progression can indicate therapeutic response and/or suggest the need for alternative therapies. In general, a monitoring test needs to be both sensitive and specific to ensure either initiation/continuation of beneficial therapies or discontinuation/replacement of ineffective treatments.

The most widely used CRC monitoring test is the carcinoembryonic antigen (CEA); CEA is a glycoprotein involved in cell adhesion that is normally produced during fetal development. Production of this protein ceases prior to birth and is, therefore, not typically present in the blood of healthy adults [4]. Elevated levels of CEA (>2.5 ng/mL) are most commonly used as a

biomarker for monitoring for relapse after CRC primary tumor resection, and for monitoring the response of metastatic CRC to systemic therapy. Ratio methods that compare sequential CEA measurements are also used, often with improved performance [5]. While CEA is FDA approved for these applications, elevated CEA levels are also associated with other types of carcinomas, such as gastric, pancreatic, lung, and breast, as well as non-malignant diseases like bronchitis, pneumonia and hypothyroidism, making it an unreliable biomarker solely for CRC cancer diagnosis or early cancer detection [6]. CEA levels correlate to recurrent CRC with a sensitivity of ~80% (range, 17–89%) and specificity of ~70% (range, 34–91%) [4,7], which is less than optimal.

In this study, we propose to monitor cancer progression by utilizing a targeted liquid chromatography tandem mass spectrometry (LC-MS/MS) approach to profile serum metabolites. Sequential metabolite ratio analysis of the serially acquired patient samples was used to identify metabolites that correlate with CRC patient disease status and reduce patient to patient variability. Individually, a number of metabolites showed a significant difference (*p*-values<0.05) in their sequential ratios between CRC patients with progressing disease and CRC patients with other disease status (includes non-progressing CRC patients in remission/stable disease while on treatment, or with no evidence of disease after resection). The individual performance of several of these metabolites was higher than for CEA alone. Partial least squares-discriminant analysis (PLS-DA) was performed using sequential patient sample ratios of these metabolite biomarkers, and high sensitivity and specificity were obtained for the differentiation of CRC patients with disease progression status compared to patients with stable disease or complete remission. The results show promise for the use of metabolite profile to monitor the CRC patient population.

## Materials and Methods

**Clinical samples:** Patient recruitment and sample collection protocols were approved by Institutional Review Boards at Purdue University and the Indiana University School of Medicine. Written informed consent was provided from all subjects in the study according to institutional guidelines. Longitudinal serum samples (49) were obtained from 20 CRC patients who participated in the Cancer Care Engineering project. A summary of patient information is shown in Table S2. The four major CRC disease statuses are defined as the following: At Diagnosis (AD) – the patient has just been diagnosed with cancer and has not yet received any form of treatment for it; Disease Progression (DP) – a patient has a growing tumor (determined either clinically or by imaging), and the patient is usually on treatment but can also be off treatment; Stable Disease (SD) – the patient has a tumor, may or may not be on treatment, and imaging studies/clinical exam suggest that his/her tumor is the same size as determined in previous visits; and Complete Remission (CR) – the patient had a tumor, may or may not be on treatment, and imaging studies/clinical exam suggest that he/she has no visible tumor anymore. All samples were evaluated for serum CEA values at the time of collection, and this information was also utilized for comparison in this study. Each blood sample was allowed to clot for 45 min and then centrifuged at 1500g for 10 min. Serum aliquots were then stored at -80 °C until experiments were performed.

**Chemicals, sample preparation and LC-MS/MS conditions:** Chemicals and methods related to serum sample preparation, and the targeted LC-MS/MS platform used for analysis were similar to utilized in our previous report [8], and are detailed in the Supplementary Information. All chemicals used in this study are LC-MS grade or above. Frozen samples were first thawed and then cold-methanol extraction of metabolites was performed. Chromatographic separations

were performed on an Agilent 1260 HPLC system installed with two hydrophilic interaction chromatography (HILIC) columns. Targeted mass spectrometry was performed in multiple-reaction-monitoring (MRM) mode using an AB Sciex QTrap 5500 instrument.

**Data analysis, model development, and cross validation:** The extracted MRM peaks were integrated, and the spectral data were exported using MultiQuant 2.1 software (AB Sciex). Sequential metabolite ratios (i.e., the ratio of the metabolite from one blood draw to that from the previous blood draw of the same patient) were calculated before applying other further analyses. The calculated ratio values were linked to the disease status at the time of the more recent blood draw, so it is possible to see the ratios from the same patient belong to two different disease status groups (such as a patient initially diagnosed as DP but who later transformed to SD status). As an example, four blood samples were collected from patient #1 (Table S2), three ratio values can be obtained for every detectable metabolite in this study (blood draw 2/ blood draw 1, 3/ 2, and 4/ 3). Three disease statuses from blood draw two, three and four (SD, DP and DP) were used to correspond to these three ratio values. There were only three groups of disease status (CR, DP and SD) remaining after the ratio calculation.

Both univariate and multivariate statistical analyses were applied for metabolite biomarker discovery and model development on a selected set of biomarker candidates. Mann-Whitney U-tests, generation of receiver operator characteristics (ROC) curves, and calculation of sensitivity, specificity, false discovery rate (FDR) and area under ROC curves (AUROCs) were calculated for each metabolite using JMP Pro10 (SAS Institute). Partial least squares-discriminant analysis (PLS-DA) and Monte Carlo Cross Validation (MCCV, developed using in-house scripts) were performed using Matlab software (Mathworks, Natick, MA) installed with the PLS toolbox (Eigenvector Research Inc., Wenatchee, WA). MCCV was applied with 100

iterations, using 70% of the data (randomly selected) as the training set while the remaining 30% served as the testing set for each iteration. Three specificities, 0.95, 0.85, and 0.75 for the training sets were used to determine the thresholds of PLS-DA predicted Y values. The same thresholds were then applied to the test set to determine sensitivities and specificities. The sample classification can be correctly assigned, termed "true class," or the sample class information can be randomly permuted, which is referred to as "random permutation." Pathway enrichment and topology analysis were achieved using the on-line server MetaboAnalyst 3.0 (http://www.metaboanalyst.ca/MetaboAnalyst/faces/home.xhtml) [9].

## Results and discussion

The targeted platform and HILIC chromatography-MS/MS method used in this study could detect of 162 metabolites (see Table S1), representing more than 20 different classes (such as amino acids, carboxylic acids, pyridines, etc.) from 25 important metabolic pathways (e.g., TCA cycle, amino acid metabolism, purine and pyrimidine metabolism, glycolysis, etc.). In this study, 131 metabolites were reproducibly detected in the 49 samples, with an average coefficient of variation (CV) of 7.1%.

CEA values were available for all 49 samples and the AUROC was calculated as 0.77. The sensitivity and specificity were 0.86 and 0.44, respectively, for the typical cutoff values of 2.5 ng/mL. At 5 ng/mL, the sensitivity and specificity were 0.86 and 0.67, respectively. Often, DP is better identified in CRC patients using CEA ratios calculated from sequential samples [5,10], and therefore we also evaluated the CEA performance for CRC DP monitoring based on its ratio of serial blood draws from the same patient, although with 29 ratios for the 20 patients, we did not have sufficient samples to calculate exponentially fitted slopes [10]. Using a ratio

cutoff value of 1.2, the sensitivity and specificity are 0.75 and 0.76, respectively (lower cutoff values can be used, which would increase the detection sensitivity but also decrease the specificity). The AUROC was 0.80 for the differentiation of DP from both complete remission and stable disease (CR + SD) groups (Figure 1A).

Metabolite data were then analyzed after calculating the sequential metabolite ratios for serial patient samples. Both univariate and multivariate statistical methods were used for metabolite biomarker selection. After applying the univariate Mann-Whitney U-test, 19 metabolites from different compound classes, such as monosaccharides, amino acids, carboxylic acids, and nucleosides, showed a significant statistical difference ($p<0.05$) between CRC DP and CR + SD. The *p*-values, fold changes and false discovery rate (FDR) for these metabolites are listed in Table 1. Furthermore, highly significant changes (defined as $p<0.01$) between CRC DP and CR+SD were found for six metabolites, namely succinate, N2, N2-dimethylguanosine, adenine, citraconic acid, methylmalonate, and 1-methylguanosine. We established the individual ROCs for each of these six metabolites for monitoring the CRC disease progression (Figure S1). Some of these metabolites had good AUROCs, such as 0.83 for succinate and 0.82 for N2, N2 dimethylguanosine, which represents better performance than CEA (or its sequential sample ratio) alone.

Furthermore, PLS-DA was utilized to identify the performance of multiple metabolite biomarkers in combination for monitoring CRC DP. Variable importance in projection (VIP) scores from the PLS-DA of all metabolites were calculated to evaluate those metabolites that contributed most to the differentiation of CRC DP from CR and SD (see Supplementary Table S3 for metabolites with VIP>1.5). A series of PLS-DA models was then established based on the different VIP thresholds (from 1.5 to 2); the model performances were evaluated and are listed in

Table S4. Interestingly, when the VIP threshold was set to 2, five out of the six metabolites (succinate, N2, N2-dimethylguanosine, adenine, citraconic acid and 1-methylguanosine) that had $p<0.01$ were again selected as important biomarkers for CRC DP monitoring.

A PLS-DA model using only these five core metabolite biomarkers was then applied to evaluate the performance of this approach for CRC DP monitoring, and the ROC curve generated for this metabolite model is shown in Figure 1B. This five-metabolite model demonstrated excellent performance with an AUROC of 0.91, a sensitivity of 0.83, a specificity of 0.94 and a FDR of 0.09. To further test the robustness of this model, MCCV was applied with three different specificities. The true classification models clearly outperformed the random permutation models (Figure 1C), suggesting that the five core metabolite biomarker model is reliable for the CRC DP monitoring. Adding the CEA ratio to the five metabolite model slightly improved performance (AUROC increased from 0.907 to 0.912, Figure S2), again with FDR of 0.09. While these results are promising, further studies are needed to determine whether the combination of both metabolite biomarkers and CEA provides the most robust utility for the close monitoring of patients for CRC DP. In order to evaluate model over-fitting, an additional permutation test were performed [11]. After randomizing the sample labels, individual metabolite performance were first calculated, and no metabolite showing significant disease monitoring ability was observed while tested with null data set (with random labels, which resulted in average AUROC < 0.53 and standard deviation < 0.13 for 100 runs); furthermore, the same variable selection and model setup procedure was performed 100 times. The null data sets resulted in an average AUROC of 0.76 with standard deviation of 0.04, indicating some overfitting that may due to the limitation of small sample size. Nevertheless this result is still significantly lower than our actual model performance 0.91.

A further refined analysis using a subset of our samples from stage IV patients was also conducted, as these patients are the most likely to experience disease progression. A comparison of DP patients and patients with CR and SD status resulted in a group of nine metabolite biomarkers with *p*-values less than 0.05 (1- methylguanosine, N2,N2-dimethylguanosine, adenine, succinate, pyruvate, methylmalonate, homogentisate, urate, and citraconic acid); the *p*-values, fold changes, and FDR are listed in Supplementary Table S5). Interestingly the five core metabolite biomarkers were among the metabolites in this list. A PLS-DA model using only stage IV patient sequential metabolite ratios of these five metabolites and CEA also showed and good performance (Figure S3) with AUROC=0.84, sensitivity of 0.91, specificity of 0.82 and FDR of 0.17. This result again suggests the potential usefulness of this targeted metabolite profiling approach for CRC DP monitoring, especially for stage IV patients who experience a high risk for disease progression.

Significantly changed metabolites discovered in this study are involved in multiple important metabolic pathways, such as the tricarboxylic acid (TCA) cycle, glycolysis, amino acid metabolism, purine metabolism, the urea cycle and their related pathways. In order to investigate the perturbation of CRC progression to multiple metabolic pathways in this study, pathway enrichment and pathway topology analysis were performed on the data using MetaboAnalyst [9], and a metabolome view is shown in Figure S4. Meanwhile, efforts were made to understand the possible connections among these serum metabolite perturbations. A metabolic pathway map showing significantly changed metabolites was also constructed (see Figure 2) using reference information obtained from the Kyoto Encyclopedia of Genes and Genomes website (www.genome.jp/kegg/). Six significantly altered metabolites in this study, namely galactose, pyruvate, cystathionine, 3 nitro-tyrosine, ornithine, and methyl succinate,

showed decreased levels in CRC DP serum samples in comparison with CR and SD. On the other hand, thirteen significantly altered metabolites, including: glucose 1, 6-bisphosphate (G16BP); fructose 1, 6-bisphosphate (F16BP); 1-methylguanosine; 2-aminoadipate; citraconic acid; N2, N2-dimethylguanosine; oxaloacetate; cis-aconitate; succinate; homogentisate; methylmalonate; adenine; and urate increased in the CRC DP group. Several of the biological discoveries in this study are consistent with previously published results. For example, G16BP and F16BP showed a significant increase in our CRC cancer progressing patients. Similarly, a recent gastric cancer study suggested that fructose-1, 6-bisphosphatase-2 (FBP2), the enzyme that catalyzes the hydrolysis of F16BP to fructose-6-phosphate and inorganic phosphate in glucose metabolism, was down regulated in gastric cancer patient tissue [12], which could lead to the accumulation of upstream F16BP. Pyruvate, the major downstream product of glycolysis, was significantly lower in CRC DP patients compared to CR and SD in this study, which matches the observations from several other studies [13,14]. Proposed mechanisms behind this observation are that CRC cancer cells try to 1) maintain low levels of pyruvate to avoid cell death caused by histone deacetylases (HDAC) [14], and 2) also overexpress pyruvate dehydrogenase kinase to increase drug resistance and early recurrence [13]. Increased levels of modified nucleosides, such as N2, N2-dimethylguanine, have been observed in urine from patients suffering from CRC [15,16], which is also in agreement with the current study. Increased levels of three TCA cycle metabolites (succinate, oxaloacetate and cis-aconitate) were observed in this study, which may suggest a typical metabolic fingerprint of mitochondrial dysfunction in hypoxic cells [17]; however, it is uncertain what causes the accumulation of these metabolites in CRC DP samples. While the results showing in this study are promising, other confounding factors such as the patient response to therapeutic treatments can also be the cause

of certain metabolite variations, which were not controlled in the experimental design due to its complexity in this group of patients, and was therefore not within the scope of this study.

## Conclusions

To the best of our knowledge, this is the first study in which an LC-MS/MS targeted serum metabolic profiling approach combined with sequential metabolite ratio analysis has been applied to distinguish CRC disease progression patients from CRC patients with complete remission and stable disease. Our results demonstrate that a panel of five core serum metabolites (succinate, N2, N2-dimethylguanosine, adenine, citraconic acid and 1-methylguanosine) can be used for sensitive and specific CRC disease status monitoring. Furthermore, with the enhancement of adding CEA to the model, this metabolic profiling approach and sequential ratio analysis can potentially serve as a novel tool for CRC disease status monitoring and provide useful information for many CRC related healthcare decisions. While these findings from a small samples size are promising, further studies with larger patient cohorts will be needed to substantiate the results, verify the important biological roles of these key metabolites, and determine any association of the derived metabolite markers with pathologically different CRC disease status. Considering their strong performance as biomarkers in the present study, these five core metabolites as well as larger profiles might be of particular interest for further validation studies.

## Acknowledgements

This work was supported by AMRMC grant W81XWH-10-1-0540. The authors would like to thank Dr. Li Yuan Bermel (Purdue University Cancer Center) for assistance with the Cancer

Care Engineering project sample collection bank. The China Scholarship Council is also gratefully acknowledged (Grant to LD).

## Potential Conflict of Interest

Daniel Raftery holds equity and an executive role at Matrix-Bio, Inc.

## References


1. Siegel RL, Miller KD, Jemal A (2015) Cancer statistics, 2015. CA Cancer J Clin 65 (1):5-29

2. Zhang A, Sun H, Yan G, Wang P, Han Y, Wang X (2014) Metabolomics in diagnosis and biomarker discovery of colorectal cancer. Cancer Letters 345 (1):17-20

3. Mook ORF, Frederiks WM, Van Noorden CJF (2004) The role of gelatinases in colorectal cancer progression and metastasis. Biochimica et Biophysica Acta (BBA) - Reviews on Cancer 1705 (2):69-89

4. Fletcher RH (1986) Carcinoembryonic Antigen. Ann Intern Med 104 (1):66-73

5. Staab HJ, Anderer FA, Stumpf E, Fischer R (1978) Slope analysis of the postoperative CEA time course and its possible application as an aid in diagnosis of disease progression in gastrointestinal cancer. The American Journal of Surgery 136 (3):322-327

6. Ludwig JA, Weinstein JN (2005) Biomarkers in Cancer Staging, Prognosis and Treatment Selection. Nat Rev Cancer 5 (11):845-856

7. Duffy MJ (2001) Carcinoembryonic Antigen as a Marker for Colorectal Cancer: Is It Clinically Useful? Clin Chem 47 (4):624-630

8. Zhu J, Djukovic D, Deng L, Gu H, Himmati F, Chiorean EG, Raftery D (2014) Colorectal Cancer Detection Using Targeted Serum Metabolic Profiling. Journal of Proteome Research 13 (9):4120-4130

9. Xia J, Mandal R, Sinelnikov IV, Broadhurst D, Wishart DS (2012) MetaboAnalyst 2.0—a comprehensive server for metabolomic data analysis. Nucleic Acids Res 40 (W1):W127-W133

10. Iwanicki-Caron I, Di Fiore F, Roque I, Astruc E, Stetiu M, Duclos A, Tougeron D, Saillard S, Thureau S, Benichou J, Paillot B, Basuyau JP, Michel P (2008) Usefulness of the Serum

299 Carcinoembryonic Antigen Kinetic for Chemotherapy Monitoring in Patients With Unresectable
300 Metastasis of Colorectal Cancer. J Clin Oncol 26 (22):3681-3686
301 11. Brereton RG (2006) Consequences of sample size, variable selection, and model validation
302 and optimisation, for predicting classification ability from analytical data. TrAC Trends in
303 Analytical Chemistry 25 (11):1103-1111
304 12. Li H, Wang J, Xu H, Xing R, Pan Y, Li W, Cui J, Zhang H, Lu Y (2013) Decreased fructose-
305 1,6-bisphosphatase-2 expression promotes glycolysis and growth in gastric cancer cells. Mol
306 Cancer 12 (1):110
307 13. Lu C-W, Lin S-C, Chien C-W, Lin S-C, Lee C-T, Lin B-W, Lee J-C, Tsai S-J (2011)
308 Overexpression of Pyruvate Dehydrogenase Kinase 3 Increases Drug Resistance and Early
309 Recurrence in Colon Cancer. The American Journal of Pathology 179 (3):1405-1414
310 14. Thangaraju M, Carswell KN, Prasad PD, Ganapathy V (2009) Colon cancer cells maintain
311 low levels of pyruvate to avoid cell death caused by inhibition of HDAC1/HDAC3. Biochem J
312 417 (1):379-389
313 15. Hsu W-Y, Chen WT-L, Lin W-D, Tsai F-J, Tsai Y, Lin C-T, Lo W-Y, Jeng L-B, Lai C-C
314 (2009) Analysis of urinary nucleosides as potential tumor markers in human colorectal cancer by
315 high performance liquid chromatography/electrospray ionization tandem mass spectrometry.
316 Clin Chim Acta 402 (1–2):31-37
317 16. Gehrke CW, Kuo KC, Waalkes TP, Borek E (1979) Patterns of Urinary Excretion of
318 Modified Nucleosides. Cancer Res 39 (4):1150-1153
319 17. Hirayama A, Kami K, Sugimoto M, Sugawara M, Toki N, Onozuka H, Kinoshita T, Saito N,
320 Ochiai A, Tomita M, Esumi H, Soga T (2009) Quantitative Metabolome Profiling of Colon and

Stomach Cancer Microenvironment by Capillary Electrophoresis Time-of-Flight Mass Spectrometry. Cancer Res 69 (11):4918-4925

**Table 1.** Summary of metabolites with low *p*-values ($p<0.05$) using sequential metabolite ratios in comparing DP vs. CR +SD.

| Metabolite | *p*-value | FC[a] | FDR |
|---|---|---|---|
| Succinate | 2.8E-03 | 1.33 | 0.23 |
| N2,N2-Dimethylguanosine | 3.7E-03 | 1.34 | 0.18 |
| Adenine | 4.9E-03 | 1.11 | 0.37 |
| Citraconic Acid | 4.9E-03 | 1.58 | 0.29 |
| Methylmalonate | 4.9E-03 | 1.31 | 0.31 |
| 1-Methylguanosine | 8.4E-03 | 1.25 | 0.29 |
| 3-Nitro-tyrosine | 1.2E-02 | 0.84 | 0.20 |
| Aconitate | 1.4E-02 | 1.45 | 0.31 |
| Cystathionine | 1.6E-02 | 0.65 | 0.20 |
| Urate | 1.8E-02 | 1.15 | 0.33 |
| Ornithine | 2.5E-02 | 0.91 | 0.18 |
| Homogentisate | 2.8E-02 | 1.19 | 0.14 |
| G16BP | 3.2E-02 | 1.05 | 0.17 |
| Galactose | 3.2E-02 | 0.36 | 0.11 |
| Methylsuccinate | 3.2E-02 | 0.80 | 0.41 |
| Oxaloacetate | 4.3E-02 | 1.30 | 0.33 |
| Pyruvate | 4.3E-02 | 0.64 | 0.33 |
| 2-Aminoadipate | 4.9E-02 | 1.38 | 0.31 |
| F16BP/F26BP | 4.9E-02 | 1.03 | 0.42 |

[a] Fold change represents the average metabolite ratio for disease progression samples compared to the average metabolite ratio of samples from complete remission and stable disease groups.

**Figure Legends:**

**Figure 1.** A) Receiver operator characteristic (ROC) curves for CEA sequential sample ratios (AUROC=0.80) comparing disease progression (DP) vs. complete remission (CR) and stable disease (SD); B). ROC of the PLS-DA model using five core metabolites (VIP>2) for DP vs. SD+CR. AUROC=0.907, sensitivity=0.833, specificity=0.941; C) Monte Carlo cross validation (MCCV) results from the PLS-DA models using metabolites with VIP >2. True, true class models; Random, random permutation model. From left to the right, the respective testing specificities were 0.95, 0.85 and 0.75.

**Figure 2.** Metabolic network of significantly changed metabolites in several important pathways (e.g., glycolysis, TCA, purine and pyrimidine metabolism). Bar chart: blue horizontal lines (left), disease progression (DP); red upward diagonal lines (right), complete remission and stable disease (CR+SD). Dashed arrows indicate multiple steps involved in pathway connections, which were simplified due to space restrictions. Y-axis represents the metabolite level ratios: *, $p<0.05$; **, $p<0.01$.

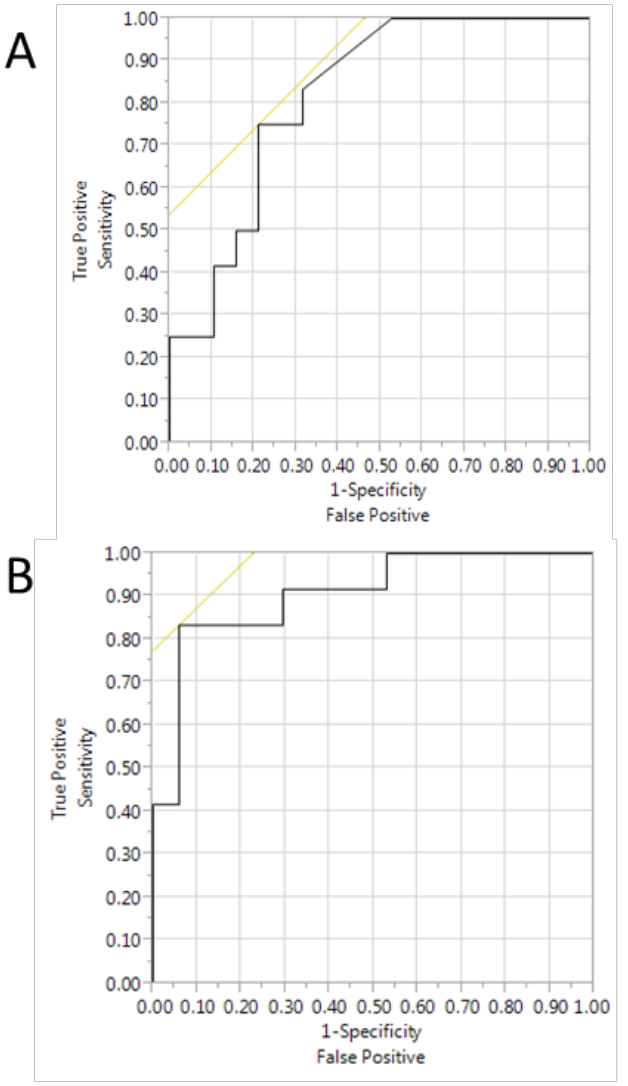

A
True Positive
Sensitivity
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
0.00 0.10 0.20 0.30 0.40 0.50 0.60 0.70 0.80 0.90 1.00
1-Specificity
False Positive
B
True Positive
Sensitivity
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
0.00 0.10 0.20 0.30 0.40 0.50 0.60 0.70 0.80 0.90 1.00
1-Specificity
False Positive
C
Sensitivity
1
0.8
0.6
0.4
0.2
0
0
0.1
0.2
0.3
1- Specificity
TRUE
RANDOM

**Figure 1**

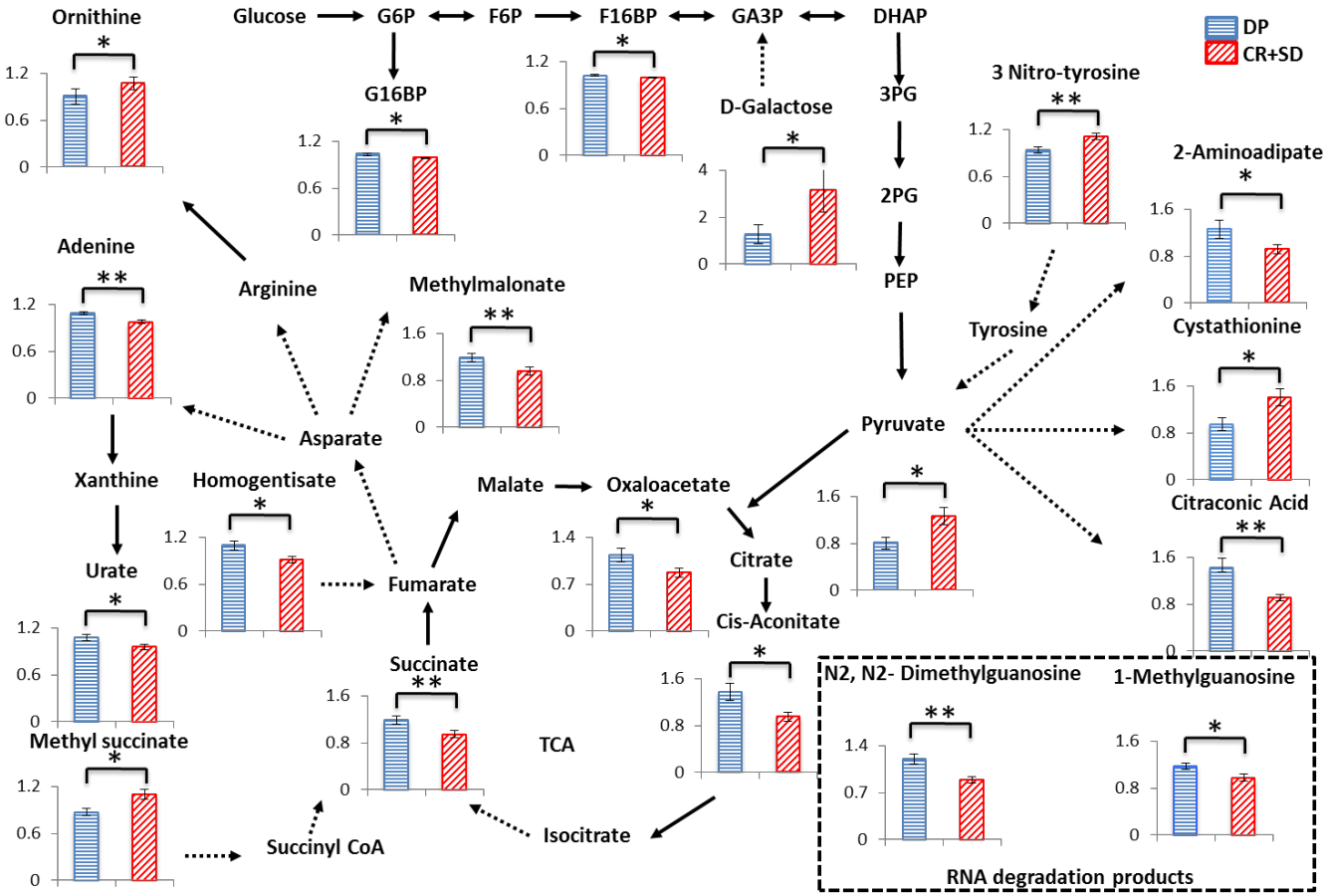

DP
CR+SD
Glucose
G6P
F6P
F16BP
GA3P
DHAP
3PG
2PG
PEP
Pyruvate
Ornithine
G16BP
D-Galactose
3 Nitro-tyrosine
2-Aminoadipate
Adenine
Arginine
Methylmalonate
Tyrosine
Cystathionine
Asparate
Xanthine
Homogentisate
Malate
Oxaloacetate
Citrate
Citraconic Acid
Urate
Fumarate
Cis-Aconitate
Succinate
TCA
Methyl succinate
Succinyl CoA
Isocitrate
N2, N2- Dimethylguanosine
1-Methylguanosine
RNA degradation products

**Figure 2**

# Supplementary Information:

## Targeted Serum Metabolite Profiling and Sequential Metabolite Ratio Analysis for Colorectal Cancer Progression Monitoring

Jiangjiang Zhu[1], Danijel Djukovic[1], Lingli Deng[1,2], Haiwei Gu[1], Farhan Himmati[1], Mohammad Abu Zaid[3], E. Gabriela Chiorean[3,4] and Daniel Raftery[1, 5*]

[1] Northwest Metabolomics Research Center, Department of Anesthesiology and Pain Medicine, University of Washington, 850 Republican St., Seattle, WA 98109

[2]Departments of Electronic Science and Communication Engineering, State Key Laboratory for the Physical Chemistry of Solid Surfaces, Xiamen University, Xiamen 361005, China

[3]Indiana University- Simon Cancer Center, 535 Barnhill Dr, Indianapolis, IN 46202

[4]University of Washington, 825 Eastlake Ave East,, Seattle, WA 98109

[5]Public Health Sciences Division, Fred Hutchinson Cancer Research Center, 1100 Fairview Ave. N., Seattle, WA 98109

* Corresponding Author:

Professor Daniel Raftery, PhD

Tel: 206-543-9709

Fax: 206-616-4819

Email: draftery@uw.edu

## Supplementary Materials and Methods:

**Chemicals and reagents:** LC-MS grade acetonitrile, ammonium acetate, and acetic acid were purchased from Fisher Scientific (Pittsburgh, PA). Standard compounds corresponding to the measured metabolites were purchased from Sigma-Aldrich (Saint Louis, MO) or Fisher Scientific (Pittsburgh, PA), and a list of these compounds can be found in the Supplementary Table S1. Stable isotope-labeled tyrosine and lactate (L-tyrosine-$^{13}C_2$ and sodium-L-lactate-$^{13}C_3$) were purchased from Cambridge Isotope Laboratories, Inc. (Tewksbury, MA). The purities of non-labeled standards were >95-99%, whereas the purities of the two $^{13}C$ labeled compounds were >99%.

**Sample preparation:** Frozen samples were first thawed at room temperature for approximately 45 min, and 50 µL of each sample was protein precipitated using two rounds of cold methanol extraction (150 µL and 300 µL, respectively) at -20 °C. The resulting supernatant containing desired metabolites was collected into a new Eppendorf vial, dried using a Vacufuge Plus evaporator (Eppendorf, Hauppauge, NY), and then reconstituted in a 500 µL solution (40% water / 60% acetonitrile with 5 mM ammonium acetate and 0.2% acetic acid) containing 5.13 µM L-tyrosine-$^{13}C_2$ and 22.5 µM sodium-L-lactate-$^{13}C_3$. The two isotope-labeled internal standards were added to monitor system performance. The samples were filtered through 0.45 µm PVDF filters (Phenomenex, Torrance, CA) prior to LC-MS analysis. A pooled human serum sample was extracted using the same procedure as above. This sample was used as the quality control (QC) sample and was analyzed once every ten serum samples. All patient samples were randomized before LC-MS analysis.

**LC-MS/MS system and conditions:** The LC system consisted of two Agilent 1260 binary pumps, an Agilent 1260 auto-sampler, and an Agilent 1290 column compartment containing a

column-switching valve (Agilent Technologies, Santa Clara, CA). Two separate injections (10 µL for analysis using negative ionization mode and 2 µL for analysis using positive ionization mode) were made for each sample. Chromatographic separations were performed using hydrophilic interaction chromatography (HILIC) on two SeQuant ZIC-cHILIC columns (150 x 2.1 mm, 3.0 µm particle size, Merck KGaA, Darmstadt, Germany) connected in parallel. This setup facilitates high-throughput analysis as it allows one column to perform the separation while the other column is being reconditioned for the next sample injection. The reconstituted serum samples were gradient-eluted at 0.300 mL/min using solvents A (5 mM ammonium acetate in 90% water / 10% acetonitrile + 0.2% acetic acid) and B (5 mM ammonium acetate in 90% acetonitrile / 10% water + 0.2% acetic acid). The auto-sampler temperature was kept at 4 °C, the column compartment was set at 40 °C, and the separation time for each ionization mode was 20 min. The gradient conditions for both separations were identical and are briefly summarized as follows: 75% B isocratic for 2 min, 75% B to 30% B in 3 min, 30% B isocratic for 4 min, back to 75% B in 2 min, and then remaining at 75% B for 9 min.

The metabolite identities were confirmed by spiking the pooled serum sample used for method development with mixtures of standard compounds (each mixture contained five standard metabolites). The few metabolites that could not be well separated and had similar m/z values (<1 Da) were integrated as single peaks (e.g., malonic acid and 3-hydroxybutyric acid).

The mass spectrometer setting was optimized and described as follows. Briefly, after the chromatographic separation, MS ionization and data acquisition were performed using an AB Sciex QTrap 5500 mass spectrometer (AB Sciex, Toronto, ON, Canada) equipped with an electrospray ionization (ESI) source. The instrument was controlled by Analyst 1.5 software (AB Sciex). Targeted data acquisition was performed in multiple-reaction-monitoring (MRM) mode.

We monitored 105 and 57 MRM transitions in negative and positive mode, respectively (162 transitions in total). The source and collision gas was $N_2$ (99.999% purity). The ion source conditions in negative/positive mode were as follows: curtain gas (CUR) = 25 psi, collision gas (CAD) = high, ion spray voltage (IS) = -3.8/3.8 KV, temperature (TEM) = 500 °C, ion source gas 1 (GS1) = 50 psi, and ion source gas 2 (GS2) = 40 psi. The optimized MS conditions for each compound were optimized with chemical standards.

## Supplementary data:

**Table S1**. List of targeted metabolites in this study (verified by chemical standards).

**Table S2.** Detailed patient and sample information.

**Table S3**. Summary of metabolites with high VIP score (VIP>1.5) using sequential metabolite ratios in comparing DP vs. CR + SD.

**Table S4.** Summary of PLS-DA model performance using different numbers of sequential metabolite ratios for the differentiation of DP vs. CR + SD.

**Table S5.** Summary of metabolites with p-values <0.05 using sequential metabolite ratios in comparing DP vs. CR +SD in stage IV patients.

**Figure S1.** Individual ROC curves for the top six metabolites with *p*-value<0.01 comparing DP with CR and SD using sequential metabolite ratios: (A) succinate, AUROC=0.83; (B) N2,N2-dimethylguanosine, AUROC=0.82; (C) citraconic acid, AUROC=0.81; (D) adenine, AUROC=0.81; (E) methylmalonate, AUROC=0.81; and (F) 1-methylguanosine, AUROC=0.79.

**Figure S2.** (A)ROC of PLS-DA model using five metabolites (with VIP>2) and CEA ratios for DP vs. CR + SD: AUROC= 0.912 (increased from 0.907, see Figure S3); sensitivity= 0.83; specificity= 0.94, FDR=0.09. (B) Monte Carlo cross validation (MCCV) PLS-DA results using the same metabolites: True, true class models; Random, random permutation model. The testing specificities were 0.95, 0.85, and 0.75. Error bars showing the standard deviation of 100 round of MCCV results.

**Figure S3.** ROC of PLS-DA model using five metabolites (with VIP>2) and CEA ratios for DP vs. CR + SD in stage IV patients: AUROC= 0.84, sensitivity= 0.91, specificity= 0.82, FDR=0.17.

**Figure S4.** A metabolome view showing all impacted metabolic pathways in this study analyzed using MetaboAnalyst(2.0), and using both scores from enrichment analysis (y axis) and from topology analysis (x axis). Due to space restriction only –log(p) > 3 and pathway impact score > 0.3 are labeled.

**Table S1**. List of targeted metabolites in this study (verified by chemical standards).

| | | |
|---|---|---|
| Glycine | Normetanephrin | Ribose-5-P |
| Trimethylamine-N-oxide | Histamine | Adenylosuccinate |
| Alanine | Pyruvate | D-Leucic acid |
| Aminoisobutyrate | Lactate | GDP |
| Choline | Acetoacetate | GTP |
| Dimethylglycine | Fumaric | DCDP |
| Serine | Succinate | Pyridoxal-5-P |
| Creatinine | Nicotinate | Gibberellin |
| Proline | Glutaric acid | Adipic acid |
| Valine | Malate | Maleic acid |
| Betaine | Hypoxanthine | Methylmalonate |
| Threonine | alpha-Ketoglutaric acid | DHAP |
| Taurine | Xanthine | Chenodeoxycholate |
| Creatine | PPA | G16BP |
| Hydroxyproline | Urate | F6P/F1P |
| Leucine/iso-Leucine | Homogentisate | Oxalic acid |
| Ornithine | PEP | Glyceraldehyde |
| Homocysteine | D-GA3P | Glycerate |
| Acetylcholine | Glycerol-3-P | N-Acetylglycine |
| Glutamine | Hyppuric acid | Guanidinoacetate |
| Glutamic acid | Glucose | Mevalonate |
| Methionine | 4-Pyridoxic acid | Allantoin |
| Cystamine | 2/3-Phosphoglyceric acid | Inositol |
| Histidine | Erythrose | Homovanilate |
| Carnitine | Cystathionine | Xanthurenate |
| Phenylalanine | G1P/G6P | Pentothenate |
| Arginine | Reduced glutatione | Biotin |
| Glucosamine | F16BP/F26BP | DCMP |
| Tyrosine | Sucrose | DUMP |
| Sorbitol | 5-Formyl THF | Geranyl pyrophosphate |
| Epinephrine | Oxidized glutathione | DTMP |
| Tryptophan | gamma-Aminobutyrate | CMP |
| 5-Hydroxytryptophan | Malonic acid/3HBA | Lactose |
| Uridine | Citraconic acid | cGMP |
| Phosphotyrosine | Adenine | AMP |
| Adenosine | Shikimic acid | IMP |
| Inosine | Aconitate | PGE |
| Guanosine | Citrulline | OMP |
| XMP | Citric acid | UDP |
| L-Kinurenine | Cystine | ADP |

| | | |
|---|---|---|
| Lysine | Xanthosine | Folic acid |
| Cytosine | Uracil | DUTP |
| Homoserine | OH-Phenylpyruvate | ATP |
| Niacinamide | Glycochenodeoxycholate | Taurocholate |
| 1-Methylhistamine | Glycocholate | Fructose |
| Asparagine | Dopamine | Aspartic acid |
| Salicylurate | Melatonin | Methylsuccinate |
| 2'-Deoxyuridine | Orotate | Myristic acid |
| 3-Hydroxykynurenine | Anthranilate | Margaric acid |
| Cytidine | Glucoronate | Linoleic acid |
| Pyroglutamic acid | Oxaloacetate | Linolenic acid |
| 1-Methyladenosine | Propionate | Galactose |
| 1-Methylguanosine | 2-Aminoadipate | |
| N2,N2-Dimethylguanosine | Kynorenate | |
| Aminolevulinic acid | 3-Nitro-tyrosine | |

104

105

**Table S2.** Detailed patient and sample information

| Patient | Gender | Age at Consent | BMI[a] | Diagnosis | Stage | # of blood draws | Disease Status [b] (follow the blood draw order) |
|---|---|---|---|---|---|---|---|
| 1 | M | 29 | 30.4 | Colon Cancer | Stage IV | 4 | DP, SD, DP, DP |
| 2 | M | 32 | 28.1 | Colon Cancer | Stage III | 2 | DP, DP |
| 3 | M | 37 | - | Colon Cancer | Stage IV | 2 | DP, DP |
| 4 | F | 42 | 18.9 | Colon Cancer | Stage IV | 3 | RD, DP, SD |
| 5 | F | 42 | - | Colon Cancer | Stage I/II | 2 | AD, CR |
| 6 | F | 45 | 28.1 | Colon Cancer | Stage IV | 2 | DP, DP |
| 7 | F | 45 | 20.4 | Colon Cancer | Stage IV | 3 | DP, DP, DP |
| 8 | F | 46 | 27.5 | Rectal Cancer | Stage IV | 2 | TR, CR |
| 9 | F | 50 | 22.7 | Colon Cancer | Stage IV | 2 | DP, DP |
| 10 | M | 51 | 29.0 | Colon Cancer | Stage III | 2 | AD, CR |
| 11 | F | 55 | 25.6 | Colon Cancer | Stage IV | 4 | LM, CR, CR, CR |
| 12 | M | 55 | - | Colon Cancer | Stage IV | 2 | DP, SD |
| 13 | M | 65 | 24.4 | Rectal Cancer | Stage III | 2 | CR, CR |
| 14 | F | 66 | - | Colon Cancer | Stage I/II | 3 | AD, CR, CR |
| 15 | F | 66 | 24.0 | Colon Cancer | Stage IV | 3 | DP, SD, DP |
| 16 | F | 68 | 22.3 | Rectal Cancer | Stage IV | 3 | DP, SD, DP |
| 17 | M | 68 | 27.6 | Colon Cancer | Stage IV | 2 | SD, SD |
| 18 | M | 73 | - | Colon Cancer | Stage IV | 2 | SD, SD |
| 19 | M | 77 | 32.2 | Colon Cancer | Stage III | 2 | SD, SD |
| 20 | F | 86 | 25.4 | Colon Cancer | Stage IV | 2 | DP, DP |

[a] BMI information was not obtained for five patients.

[b] Disease status: DP, Disease Progression; AD, At Diagnosis; SD, Stable Disease; CR, Complete Remission; RD, Recurrent Disease; TR, Tumor Response; LM, Liver Metastasis. There were only three groups of disease status (CR, DP and SD) remaining after the ratio calculation.

**Table S3**. Summary of metabolites with high VIP scores (VIP>1.5) using sequential metabolite ratios in comparing DP vs. CR + SD.

| Metabolite | VIP | FC* | FDR |
|---|---|---|---|
| N2,N2-Dimethylguanosine | 2.15 | 1.34 | 0.18 |
| Citraconic Acid | 2.04 | 1.58 | 0.29 |
| 1-Methylguanosine | 2.04 | 1.25 | 0.29 |
| Succinate | 2.01 | 1.33 | 0.23 |
| Adenine | 2.01 | 1.11 | 0.37 |
| Methylmalonate | 1.84 | 1.31 | 0.31 |
| 3-Nitro-tyrosine | 1.83 | 0.84 | 0.20 |
| Malonic Acid/3HBA | 1.77 | 5.99 | 0.14 |
| G16BP | 1.76 | 1.05 | 0.17 |
| Urate | 1.76 | 1.15 | 0.33 |
| Aconitate | 1.73 | 1.45 | 0.31 |
| Homogentisate | 1.66 | 1.19 | 0.14 |
| Methylsuccinate | 1.61 | 0.80 | 0.41 |
| 1-Methyladenosine | 1.61 | 1.15 | 0.17 |
| Cystathionine | 1.60 | 0.65 | 0.20 |
| Linolenic Acid | 1.58 | 1.77 | 0.40 |
| Cytidine | 1.57 | 1.39 | 0.36 |
| Pyruvate | 1.57 | 0.64 | 0.33 |
| Alanine | 1.55 | 0.80 | 0.25 |
| gamma-Aminobutyrate | 1.53 | 0.82 | 0.44 |

*Fold change represents the average metabolite ratio for disease progression samples compared to samples from complete remission and stable disease.

**Table S4.** Summary of PLS-DA model performance using different numbers of sequential metabolite ratios for the differentiation of DP vs. CR + SD.

| PLS-DA Models | Metabolites selection Threshold | # of Metabolites used in the model[a] | AUROC | Sensitivity | Specificity |
|---|---|---|---|---|---|
| DP vs. CR+SD (Metabolites only models) | VIP>1.5 | 20 | 0.92 | 0.92 | 0.88 |
| | VIP>1.8 | 7 | 0.90 | 0.83 | 0.94 |
| | VIP>2 | 5 | 0.91 | 0.83 | 0.94 |
| DP vs. CR+SD (Metabolites + CEA models) | VIP>1.5 | 20 | 0.92 | 0.92 | 0.88 |
| | VIP>1.8 | 7 | 0.89 | 0.83 | 0.94 |
| | VIP>2 | 5 | 0.91 | 0.83 | 0.94 |

[a] See Table S3 for metabolites and their corresponding VIP scores.

**Table S5.** Summary of metabolites with *p*-value <0.05 using sequential metabolite ratios in comparing DP vs. CR + SD in stage IV patients.

| Metabolites | *p*-value | FC* | FDR |
|---|---|---|---|
| 1-Methylguanosine | 7.1E-03 | 1.27 | 0.18 |
| N2,N2-Dimethylguanosine | 1.3E-02 | 1.33 | 0.18 |
| Adenine | 1.5E-02 | 1.12 | 0.31 |
| Succinate | 1.8E-02 | 1.33 | 0.25 |
| Pyruvate | 3.0E-02 | 0.62 | 0.27 |
| Methylmalonate | 3.0E-02 | 1.28 | 0.25 |
| Homogentisate | 3.0E-02 | 1.22 | 0.14 |
| Urate | 3.0E-02 | 1.18 | 0.25 |
| Citraconic Acid | 4.9E-02 | 1.47 | 0.25 |

*Fold change represents the average metabolite ratio for disease progression samples compared to samples from other groups.

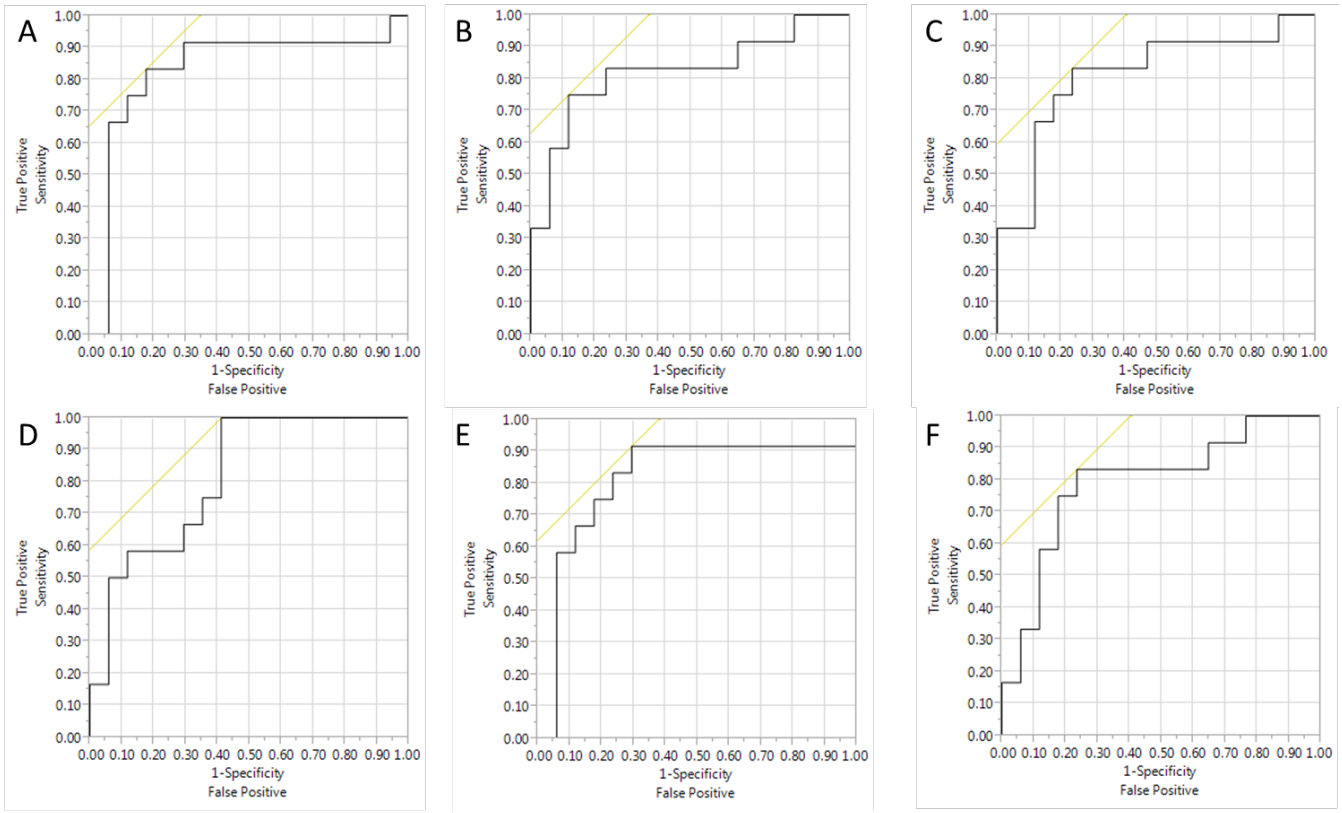
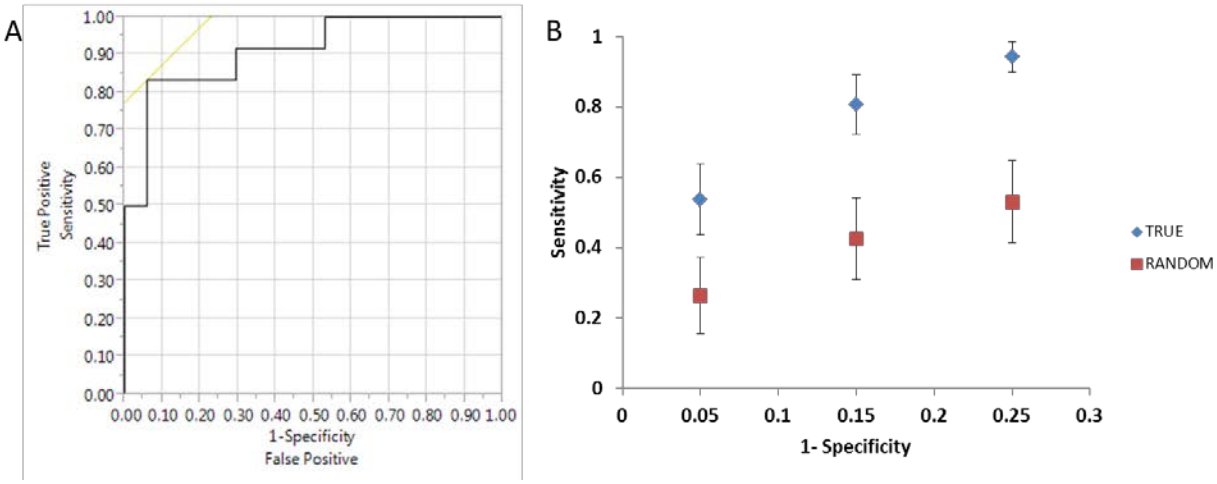

A
B
C
D
E
F
True Positive
Sensitivity
1-Specificity
False Positive
0.00 0.10 0.20 0.30 0.40 0.50 0.60 0.70 0.80 0.90 1.00

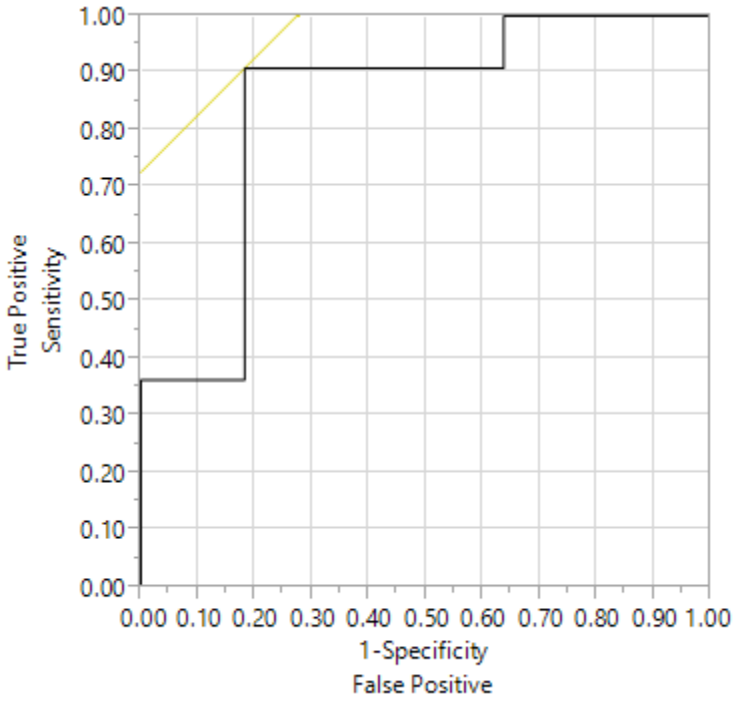
**Figure S1**

153




154 **Figure S2**

155




156

157 **Figure S3**

158

159

160

161

162

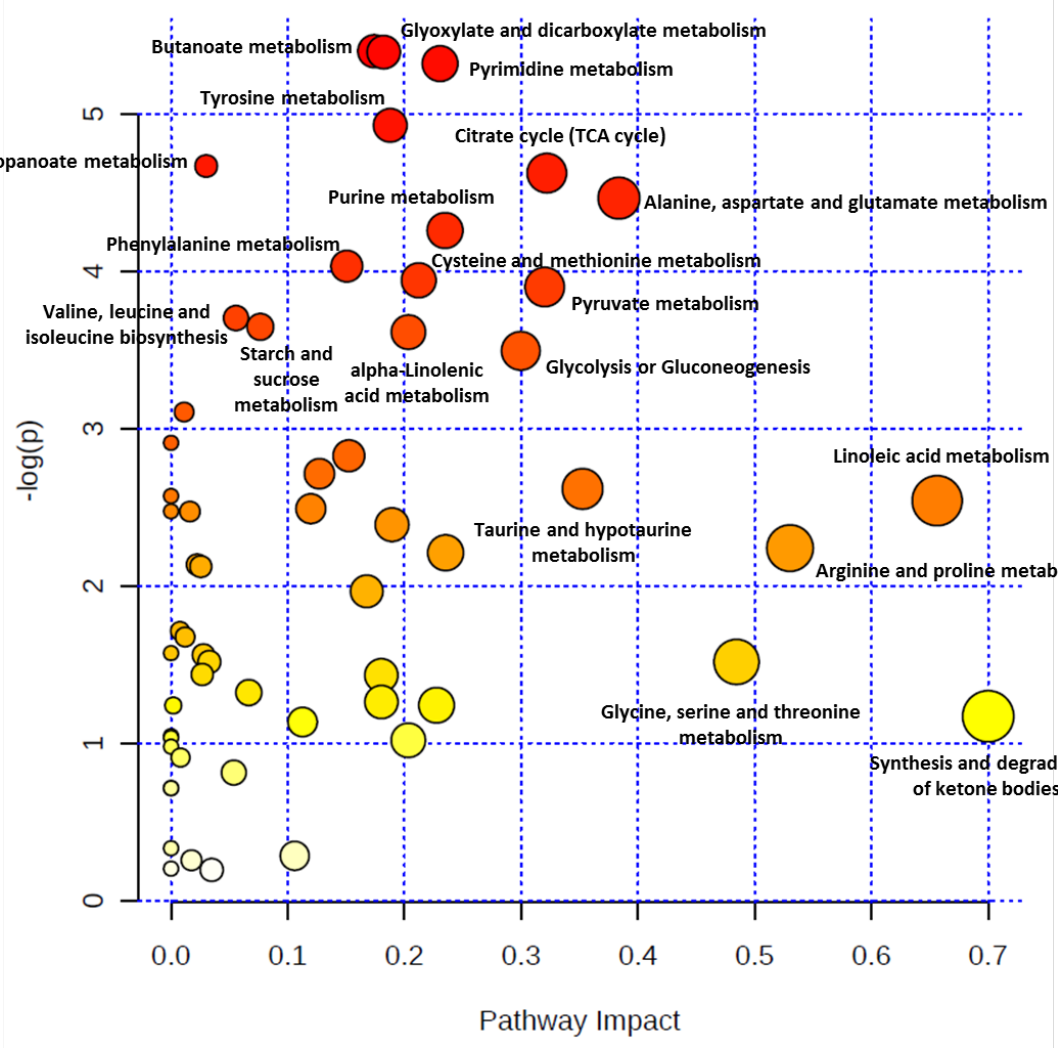

Glyoxylate and dicarboxylate metabolism
Butanoate metabolism
Pyrimidine metabolism
Tyrosine metabolism
Citrate cycle (TCA cycle)
Propanoate metabolism
Purine metabolism
Alanine, aspartate and glutamate metabolism
Phenylalanine metabolism
Cysteine and methionine metabolism
Pyruvate metabolism
Valine, leucine and isoleucine biosynthesis
Starch and sucrose metabolism
alpha-Linolenic acid metabolism
Glycolysis or Gluconeogenesis
Linoleic acid metabolism
Taurine and hypotaurine metabolism
Arginine and proline metabolism
Glycine, serine and threonine metabolism
Synthesis and degradation of ketone bodies
-log(p)
0
1
2
3
4
5
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
Pathway Impact

**Figure S4**