Data Analysis Report

Metabolomics Analysis of Serum from Healthy Individuals and Colon Cancer Patients

Client: Panome Bio

Contact: info@panomebio.com

Contents

Project Summary

Sample Description

Twenty human serum samples were received. Ten of these samples were from colorectal cancer patients (CRC, Stage IV). The remaining ten samples were from non-affected individuals (referred to as healthy). Individuals in both the CRC and healthy groups were between 40 and 80 years of age at the time of sample collection.

Goal

As CRC risk is correlated to lifestyle and environmental factors, metabolomics offers an opportunity to identify early disease biomarkers from exogenous-derived compounds and endogenous metabolite changes that reflect biochemical re-wiring in CRC. Accordingly, the primary objective of this project is to identify and characterize the specific metabolite changes in serum associated with CRC.

Assay Summary

Global metabolomics assays were performed on all samples, including both polar and lipid fractions analyzed with both positive and negative ionization.

Analysis Summary

Global metabolomics data was processed with unbiased peak characterization and metabolite identification. Statistical analysis was performed to identify metabolites with differential abundance between the CRC and healthy serum samples. Results were aggregated through pathway analysis.

Results Summary

1,616 metabolites were profiled in the serum samples, leading to clear separation and clustering of healthy and CRC samples in unsupervised analyses. Statistical analysis found 20 metabolites with differential abundance in CRC. Interpretation of these metabolites provided multiple insights into dietary differences between healthy and CRC individuals, relationship with known changes in inflammatory bowel disease (IBD), increased inflammatory markers in CRC, exposure compounds that may be a risk factor for developing CRC, and specific molecular changes related to vitamin C and cysteine metabolism. The diversity of these insights underscores the value of metabolomics profiling to CRC. To provide further insights into these changes we recommend complementing this analysis with global proteomics, expanding the cohort to increase statistical power (potentially including serum from individuals with IBD and early-stage CRC), profiling metabolic changes occurring within primary tissue, and conducting in-vitro analyses to validate and elucidate the mechanism of potential toxicants identified in the untargeted metabolomics analysis.

Results

Metabolite Summary

In the comprehensive metabolomic profiling of the samples, we employed four distinct assays. These assays detected and characterized a total of 1,616 metabolites in the dataset, as illustrated in Figure 1. Out of these, 1,115 compounds were confidently identified within the serum samples. We quantified the relative abundance of each metabolite across the samples. The polar fraction of the samples exhibited a diverse range of metabolites, including small peptides (amino acid-like molecules, dipeptides, etc.), saccharides, and fatty acids. The lipid data revealed the presence of various glycerophospholipids, glycerolipids, and fatty acyls. A comprehensive categorization of these measured compound classes is presented in Figure 2 below.

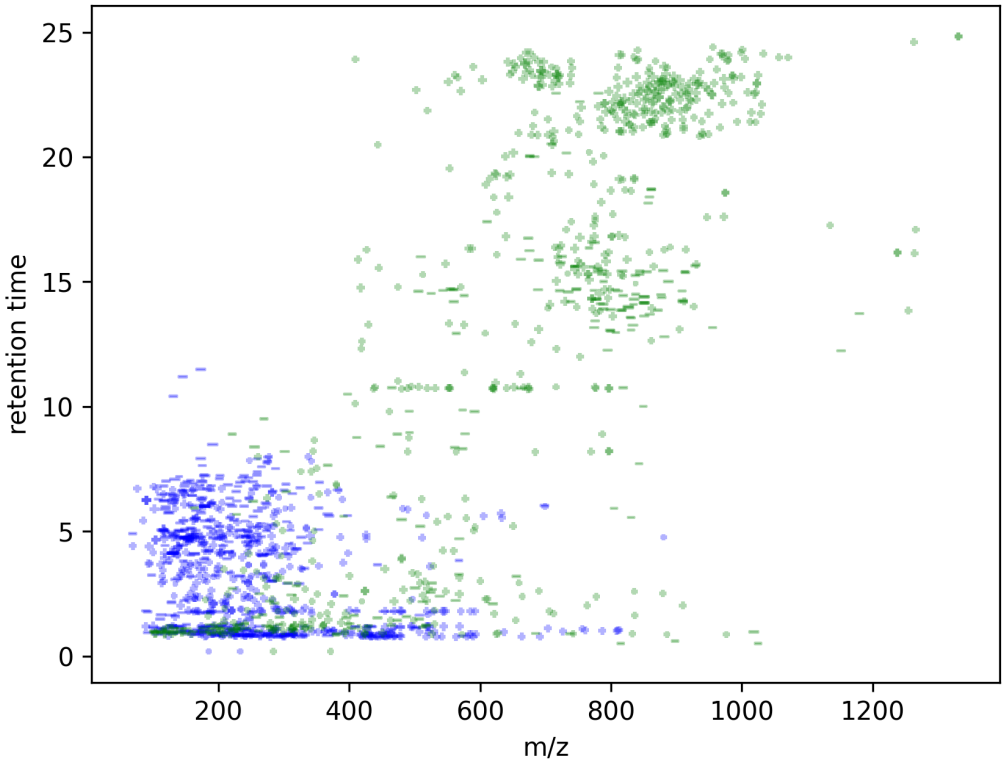

Figure 1: Detected metabolite signals. Each metabolite signal is represented by a point on the plot representing the m/z and retention time coordinates of the feature. Color of dots corresponds to the assay type (green = lipid, blue = polar) with the ionization mode indicated by the marker ("-" = negative mode, "+" = positive mode). The size of the points corresponds to the mean intensity of the signal.

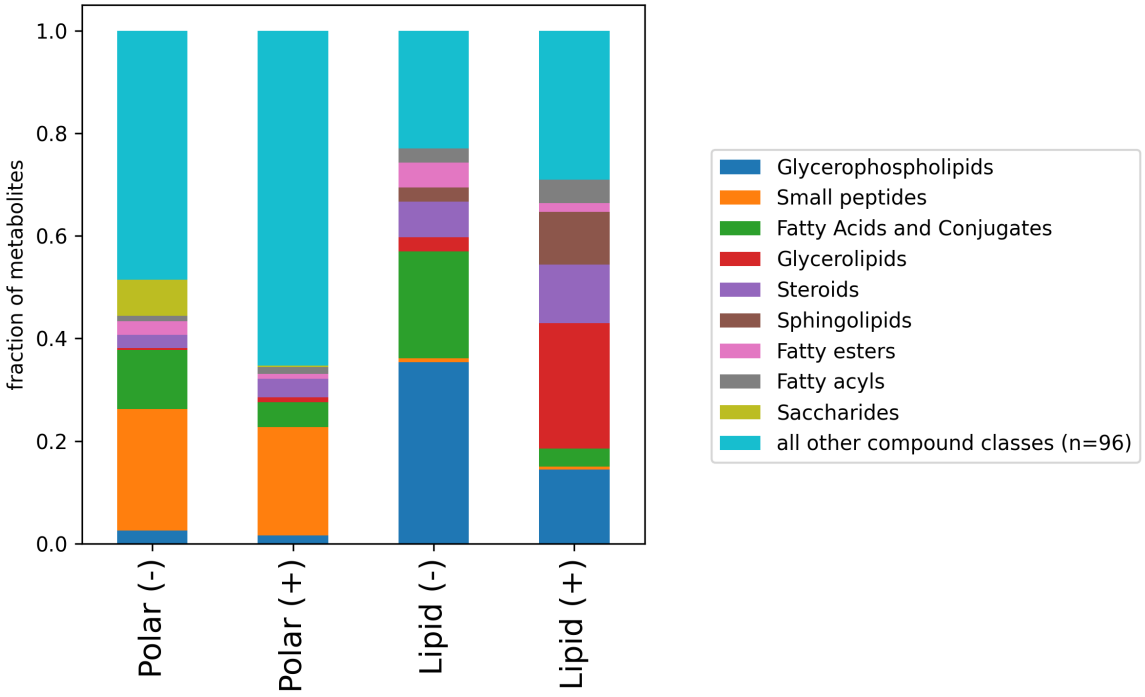

Figure 2: Chemical composition of identified metabolites. The chemical class composition is shown for each assay type. The top 9 most prominent compound classes are listed of the 90+ total metabolite classes assessed. The remaining compound classes are grouped and listed as “all other compound classes.”

Technical Variation

A summary of the technical variation across the assays employed is shown below. Across all four assays, coefficient of variation (CV) values was generally less than 10% (Figure 3). Less than 10.0% of all measurements were missing values that required imputation. Missing values are driven by signals that are below the limit of detection in a sample. Many of these missing values were due to the exogenous compounds (e.g., drugs) detected in only a small subset the serum samples.

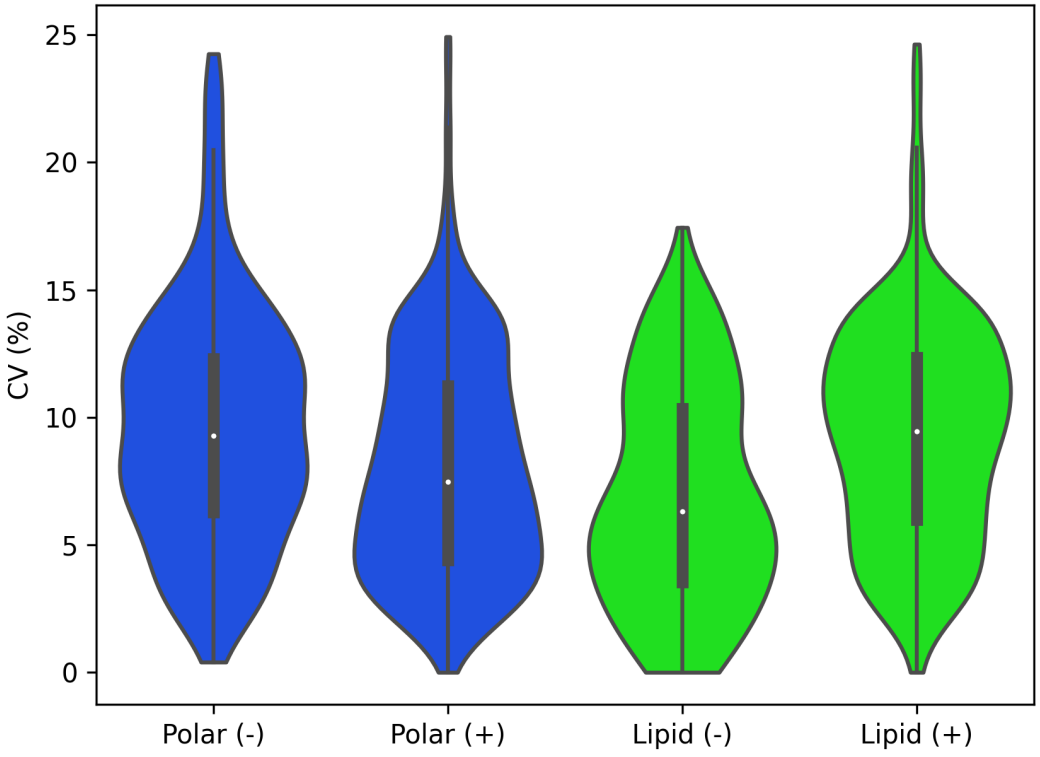

Figure 3: Coefficients of variation for profiled metabolites. The technical variation present in each assay is shown in the violin plot. The coefficient of variation (CV) is plotted from analysis of quality control samples.

Metabolite Associations

A comprehensive understanding of the sample metabolite profiles is achieved through principal component analysis (PCA), depicted in Figure 4, and hierarchical clustering analysis visualized as a heatmap in Figure 5. These visualizations clearly illustrate the distribution and relative abundance of metabolites across the samples. The analyses consider the comparison between CRC and healthy groups. Figure 6 features a scatter plot that highlights variations in common polar metabolite abundance. Each data point is color-coded based on its metabolic pathway, as denoted in the legend. By categorizing metabolites based on their associated pathways, the visuals offer insights into the metabolic shifts among the groups. In this study, amino acids levels were generally decreased in CRC samples compared to healthy. Nucleotides, glycolytic, and TCA-cycle metabolites, however, did not show differences between CRC and healthy samples. Figure 7 presents a visual representation of the log2 fold change in lipid category abundances, differentiating the sample groups using color, and using healthy samples as the reference. Here we see generally lower lipid levels in CRC when compared to healthy. Both the heatmap and the scatter plot represent fold changes with the healthy sample group as the reference.

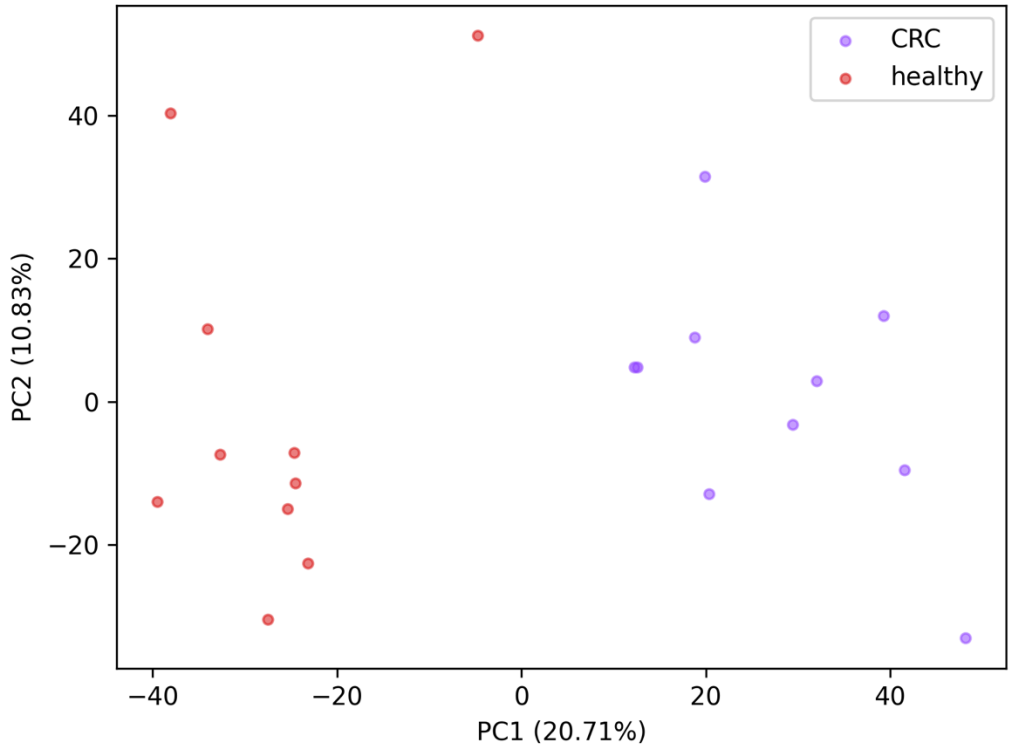

Figure 4: Principal components analysis of metabolomics data. The combined metabolic profiles from all assays for each sample are visualized in the scatter plot above. Samples are color-coded based on their respective groups. The specific colors for each group are provided in the legend.

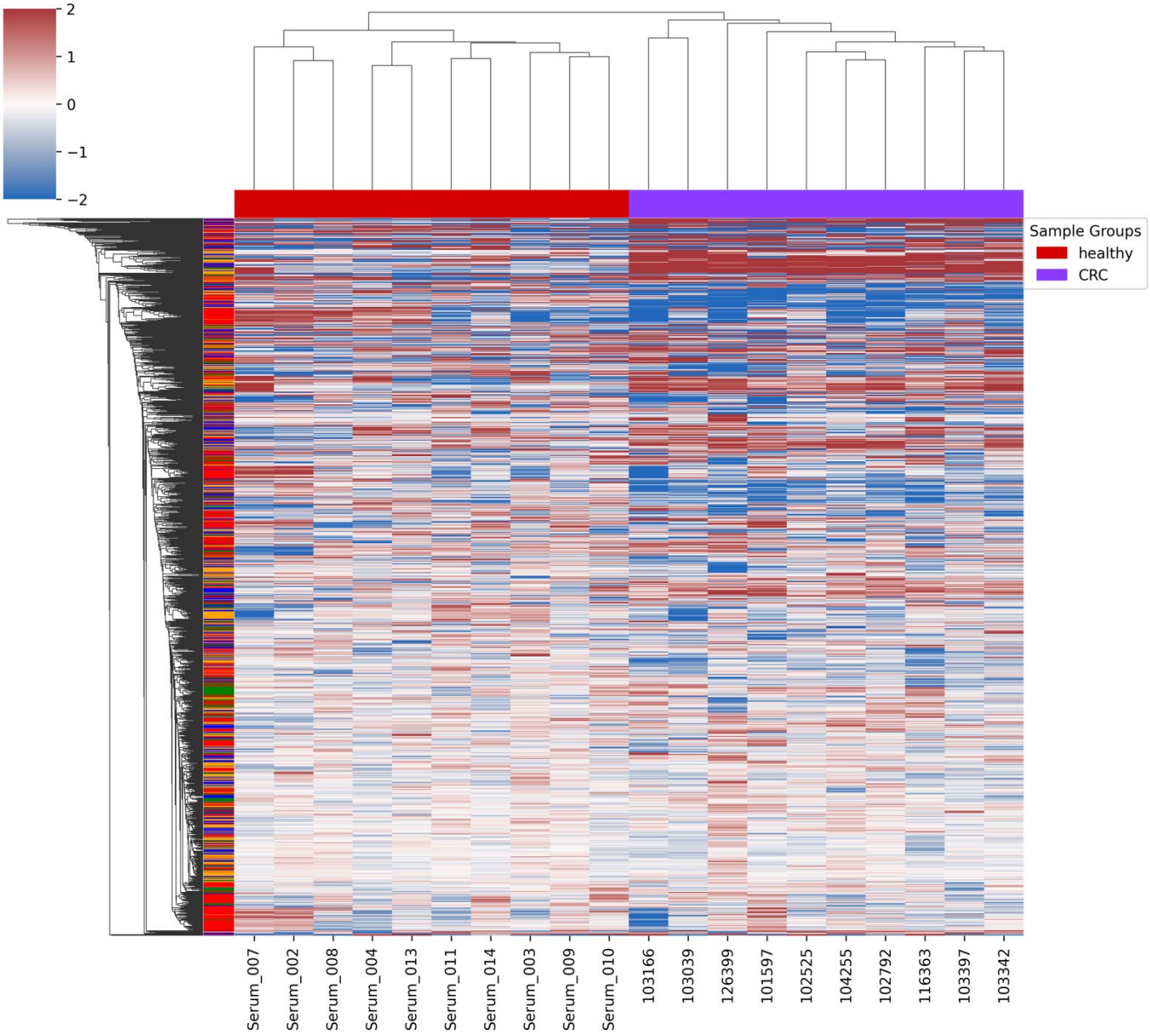

Figure 5: Heatmap of metabolomics data. The metabolic profiles (all assay results combined) for each sample are visualized in the heatmap above. Each column represents a sample, and each row represents a metabolite. Columns are colored according to experimental type. Rows are colored according to assay type (polar (-) = blue, polar (+) = orange, lipid (-) = green, lipid (+) = red). The color of each cell indicates the log2(fc) relative to the mean level of each metabolite in the healthy group.

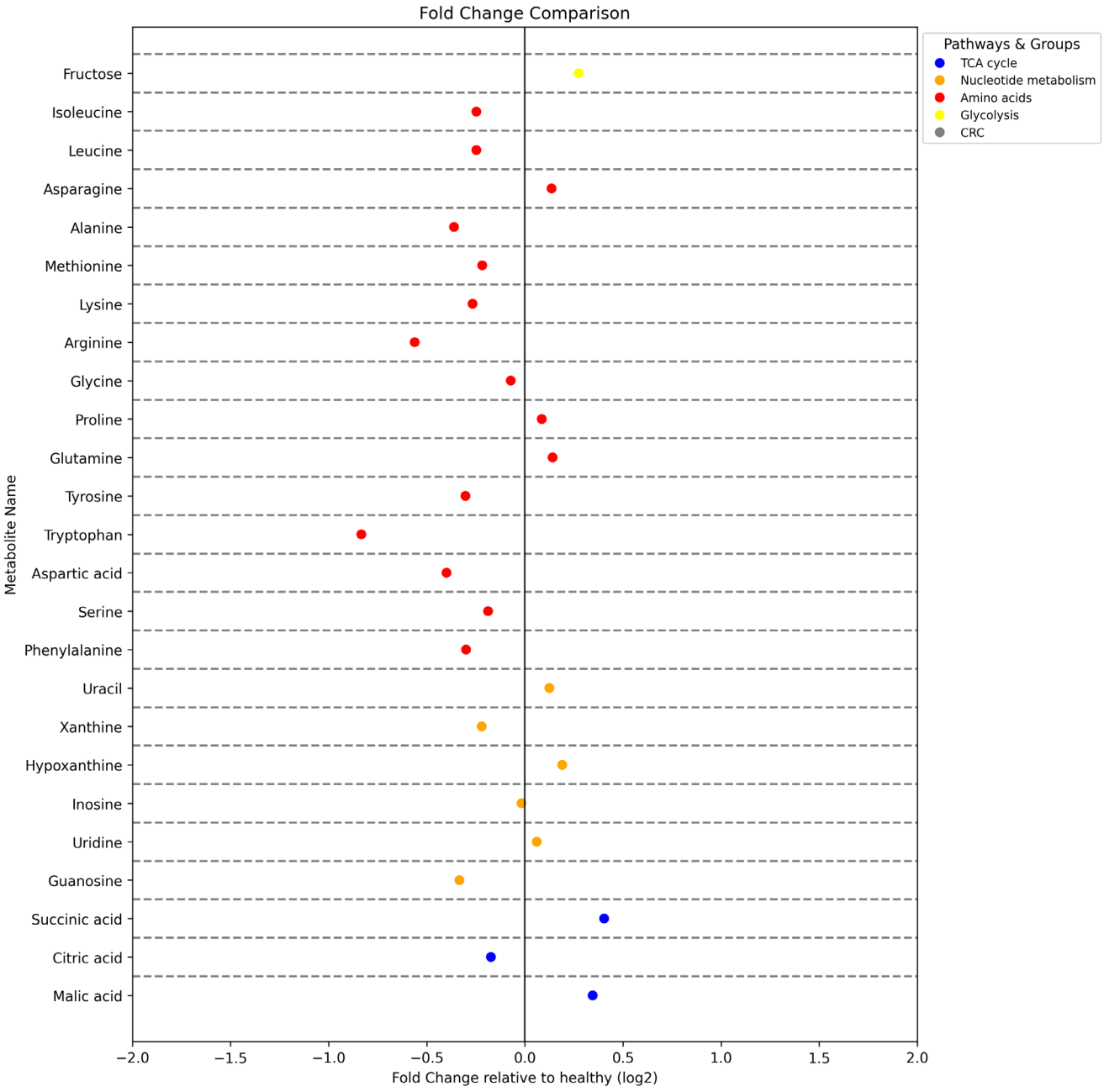

Figure 6: Fold change plot of common metabolites. The scatter plot showcases the log2 fold change in metabolite abundance, using the healthy group as a reference. Each point represents a metabolite. Points are color-coded by their metabolic pathway, and different shapes denote the sample groups. Both the color codes for pathways and the shapes for the groups are detailed in the legend. Metabolites within the plot are clustered based on their associated pathways, offering insights into the metabolic shifts among the groups.

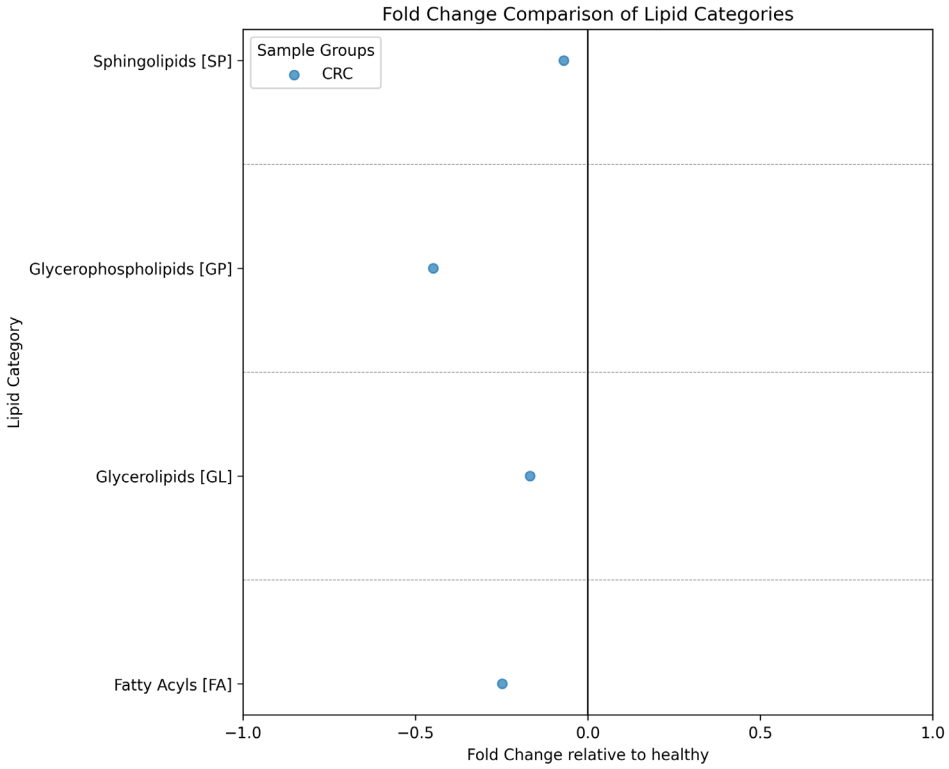

Figure 7: Lipid Categories Fold Change Overview. This scatter plot depicts the log2 fold change in lipid category abundances, comparing CRC to the healthy group, which serves as the reference. Each dot represents a lipid category's fold change, with categories on the y-axis and fold changes on the x-axis. Colors differentiate the sample groups. The vertical line at x=0 signifies no change from the healthy group. The legend shows the sample groups by color.

To determine specific differences in metabolite levels between the healthy and CRC samples, a one-way ANOVA was conducted on all profiled metabolites. This analysis identified 20 metabolites with highly confident metabolite identifications that exhibited statistically significant changes. Each of these metabolites had a corrected p-value (q) of less than 0.05 and a log2(fc) greater than 1 (Figure 8). The statistically significant identified metabolites are presented in Figure 9. Gemcabene, one of the most increased metabolites, is a common anti-inflammatory drug.

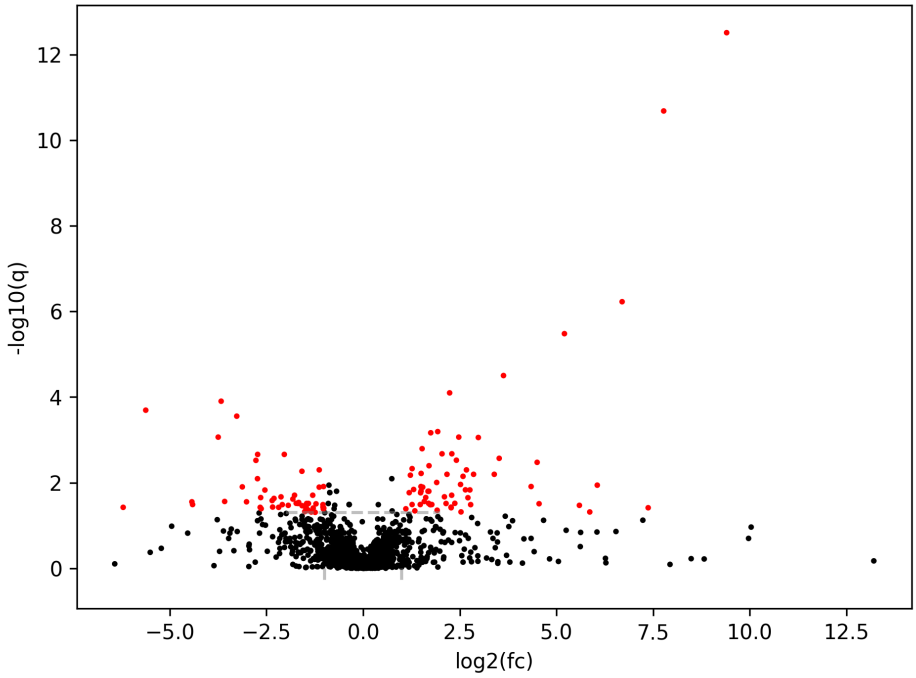

Figure 8: Volcano plot of metabolites. The q-values (q) and $\log_2$ fold-changes of the metabolite levels between healthy and CRC samples are plotted against each other in the plot above. Red dots indicate features that pass the $q < 0.05$ cutoff and have an absolute log2(fc) greater than 1.0.

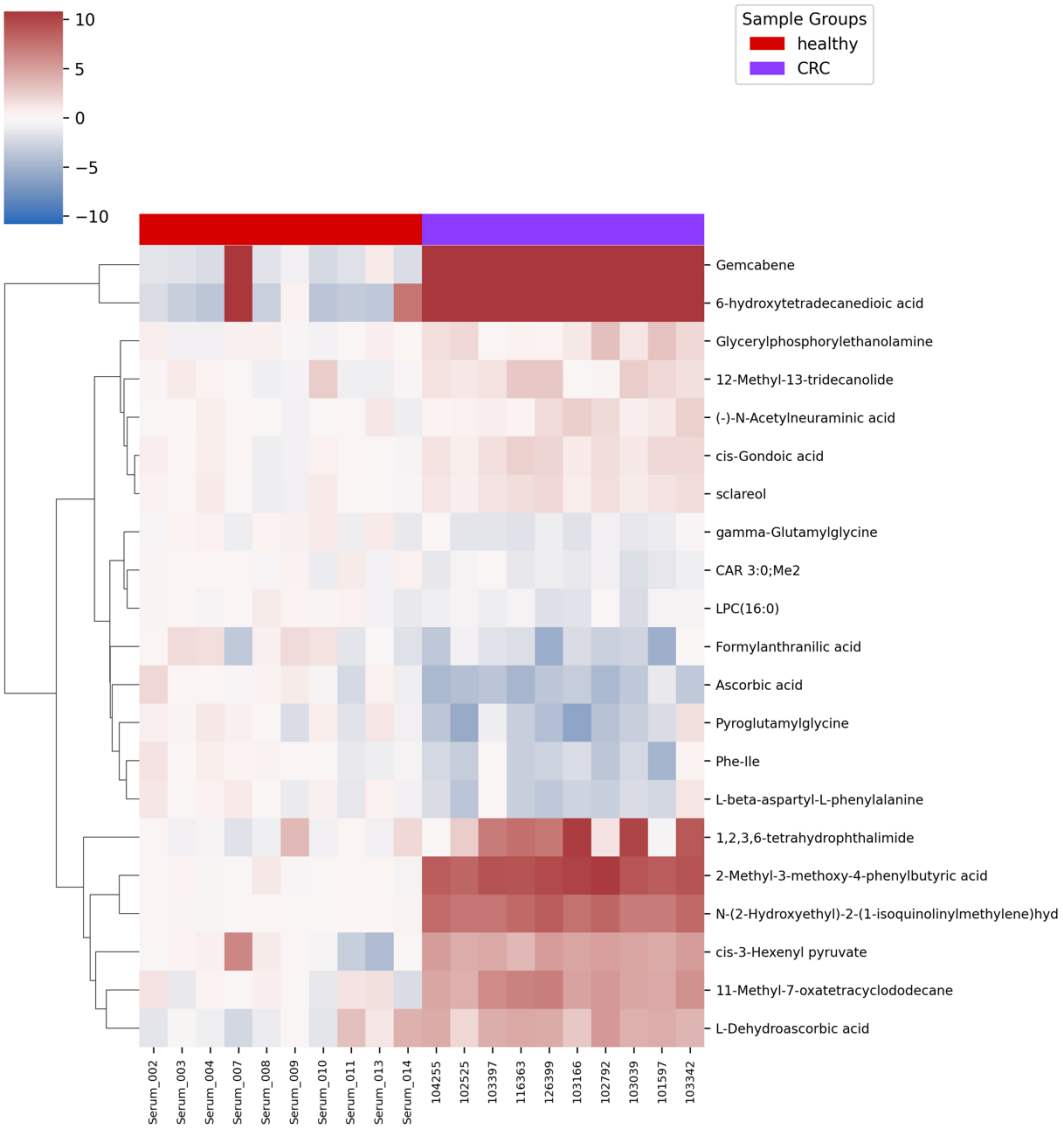

Figure 9: Heatmap of significantly altered metabolites. The normalized abundance of significantly altered metabolites are plotted for each experimental group.

Interpretation

To gain biological insights from the dysregulated metabolites, an over-representation analysis was conducted. This identified key metabolic pathways and lipid classes that were altered between the healthy and CRC samples (Figure 10). This analysis identified seven metabolic pathways and lipid classes that were enriched for metabolites showing statistically significant changes in CRC. The diversity of these pathways underscores the breadth of metabolic dysregulation in CRC. Individual heatmaps for each pathway that had significant enrichment is provided in Figures 11-17.

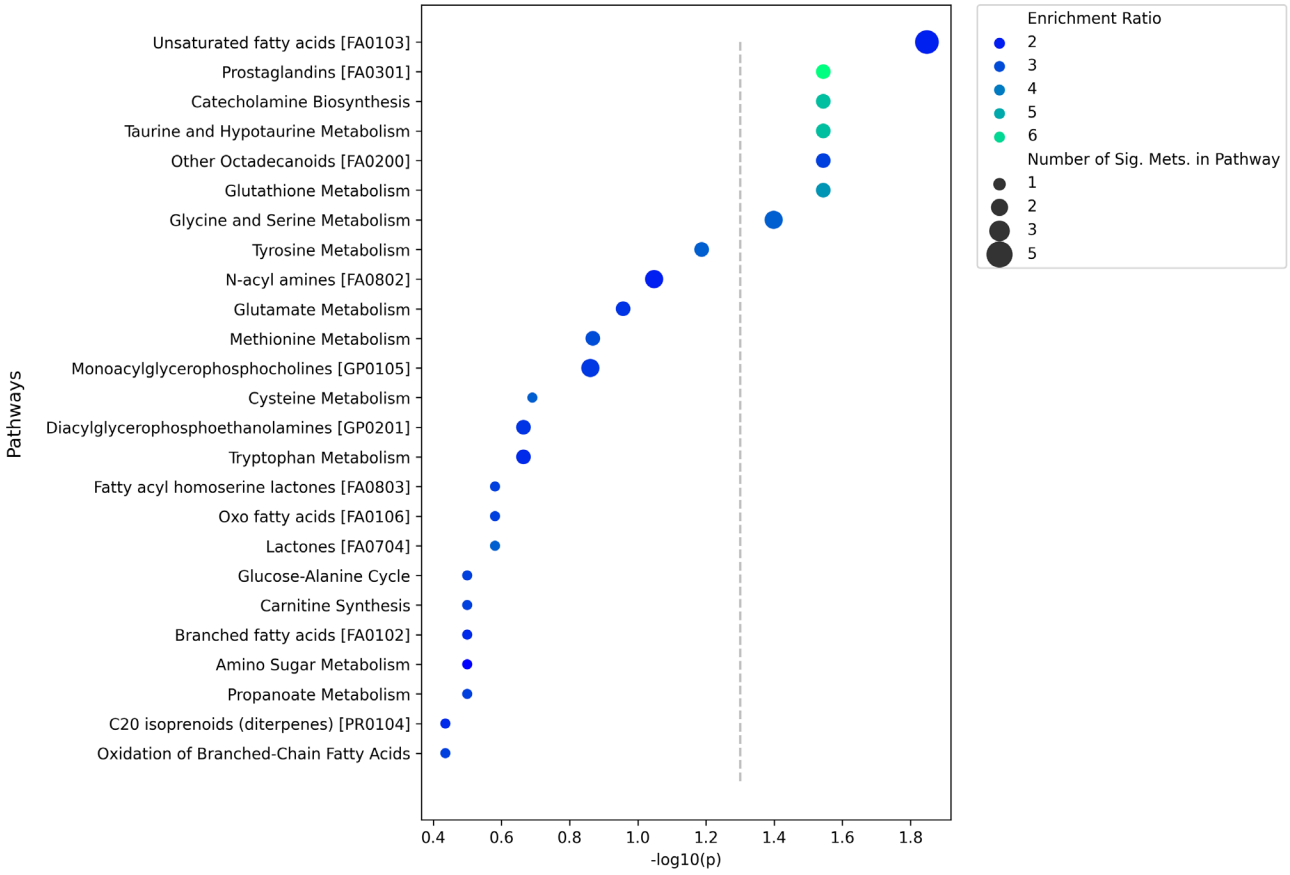

Figure 10: Pathway analysis of significant metabolites. The dot plot displays the enriched pathways identified in the analysis, with their corresponding significance levels represented on the x-axis. The significance threshold (p = 0.05) is denoted by the grey dashed line. The size of the dots reflects the number of metabolites that showed statistical differences between the sample groups. Additionally, the color of the dots represents the enrichment ratio, which indicates the extent to which the observed number of statistically significant metabolites in a pathway deviate from what would be expected by random chance.

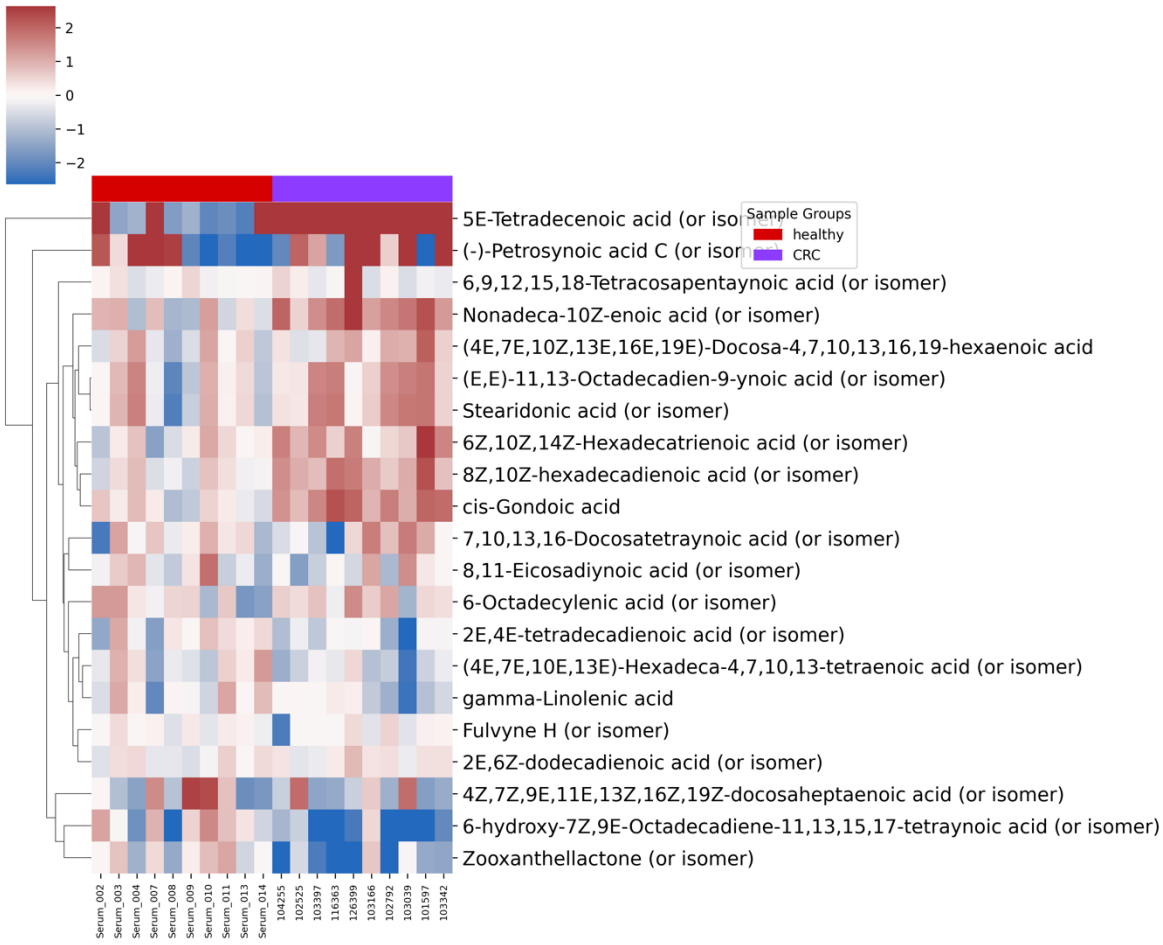

Figure 11: Heatmap of the unsaturated fatty acid levels across healthy and diseased samples. Each row represents a metabolite within the pathway, while columns correspond to the experimental conditions.

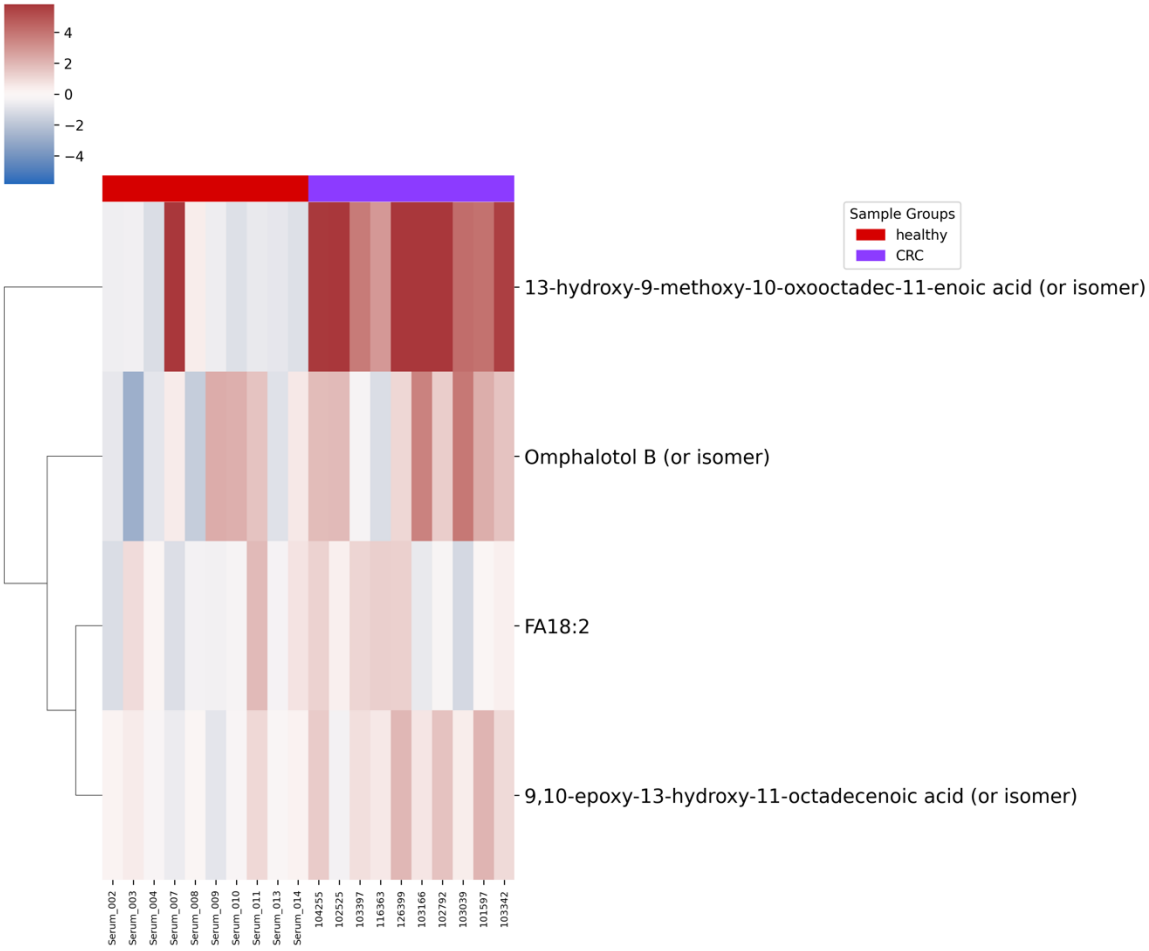

Figure 12: Heatmap of other octadecanoids across healthy and diseased samples. Each row represents a metabolite within the pathway, while columns correspond to the experimental conditions.

21 unsaturated fatty acids were measured in the serum samples. Of those, five showed a significant abundance difference. These fatty acids cannot be produced endogenously and are therefore derived from exogenous sources such as nuts, oils, and plant material that are consumed. Differences in the circulating levels of these compounds suggests dramatic dietary differences from disease. This is further supported by the changes in octadecanoids levels (Figure 12), which are another class of fatty acids that are derived from plants and have important signaling effects in humans. A related hypothesis is that perturbation induced by CRC to the gut microbiome and digestion changes the ability of these dietary compounds to enter circulation.

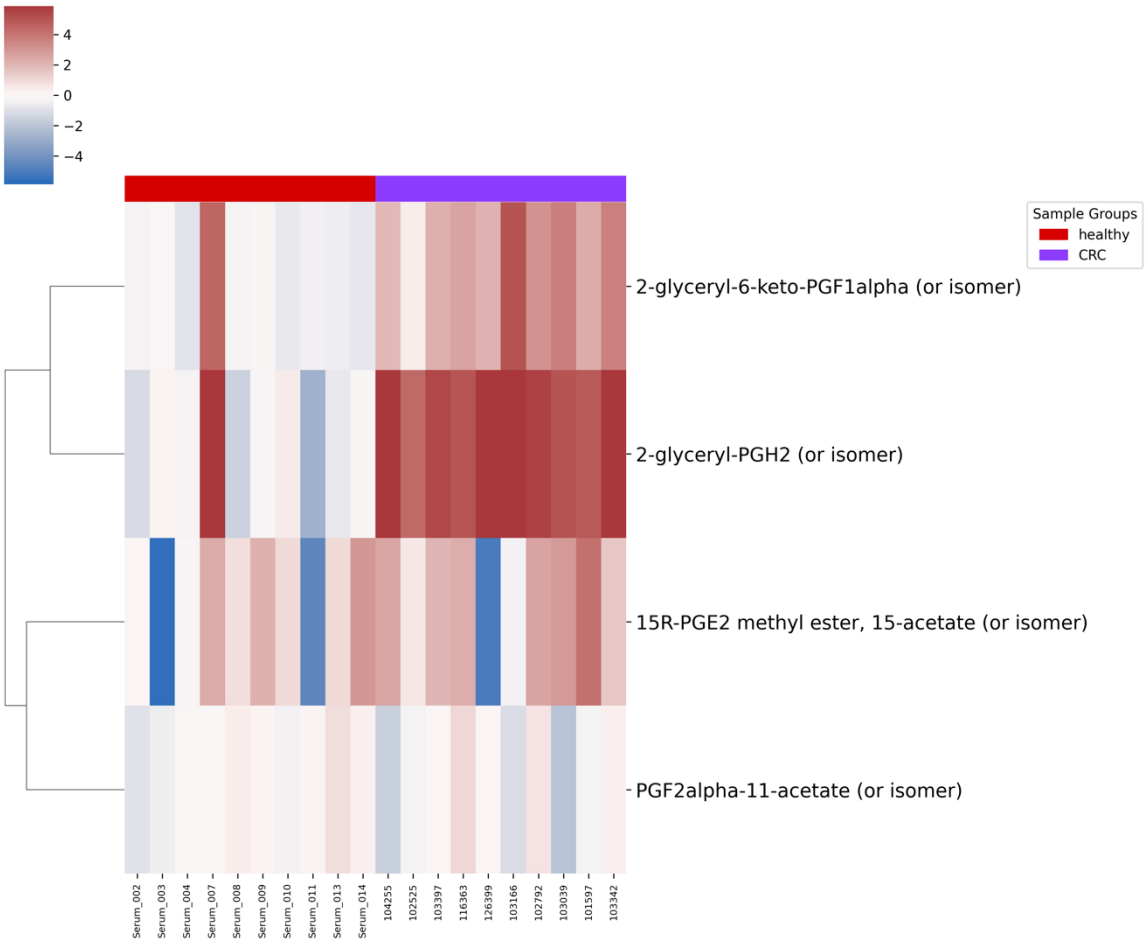

Figure 13: Heatmap of prostaglandin levels across healthy and diseased samples. Each row represents a metabolite within the pathway, while columns correspond to the experimental conditions.

Of the four prostaglandins measured in the serum samples, two showed a statistically significant increase. Prostaglandins are synthetized from the unsaturated fatty acid, arachidonate, which must be derived from the diet. Given the global changes to unsaturated fatty acid levels, this could explain, in part, the changes observed in prostaglandin levels. But, Prostaglandins are also known to play important roles in signaling and immune pathways. In particular, elevated levels of prostaglandins are often concordant with an increased inflammatory state of the individual, which would be expected in cancer patients. Octadecanoids, which were also increased in CRC, have also been implicated in inflammation, providing further evidence of an increased inflammatory state across the CRC cohort in this study. Additionally, the presence of anti-inflammatory drugs, such as gemcabene, were also detected in many CRC samples through our analysis, further suggesting a strong pro-inflammatory phenotype in this CRC population.

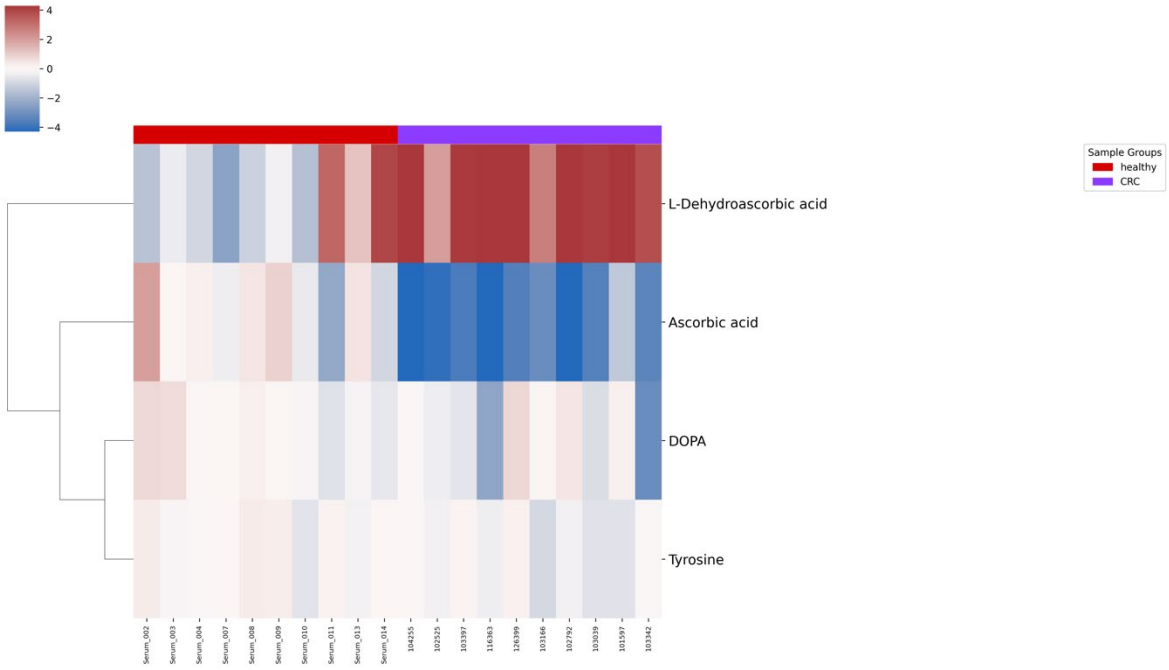

Figure 14: Heatmap of catecholamine biosynthesis-related metabolites across healthy and diseased samples. Each row represents a metabolite within the pathway, while columns correspond to the experimental conditions.

Some of the largest metabolite changes were related to catecholamine biosynthesis. In particular, ascorbic acid (vitamin C) levels were 14x lower on average in CRC samples when compared to healthy. A derivative of ascorbic acid, L-dehydroascorbic acid, on the other hand showed a 6x increase in CRC relative to healthy. Ascorbic acid is a substrate for the metabolic reaction dopamine beta-hydroxylase, which consumes dopamine, ascorbic acid, and oxygen and produces norepinephrine, L-dehydroascrobic acid, and water. The paired reactant/product changes that are observed suggest changes in the activity of this reaction. The intermediate in the synthesis pathway of dopamine from tyrosine, DOPA, shows a non-significant change, providing evidence that these changes are not due to differences in dopamine availability. While the changes in ascorbic acid levels could be due to differences in vitamin C intake, there is a known SNP in dopamine beta-hydroxylase that is over-represented in inflammatory bowel disease (IBD) patients (PMID: 30817802). IDB is a known risk factor for developing CRC.

The enrichment of taurine and hypotaurine metabolism, glutathione metabolism, and glycine and serine metabolism were all driven by statistically significant changes in three metabolites: the elevated levels of cysteine and taurine and the decreased level of alanine. Cysteine can be catabolized into both taurine and alanine, but through different metabolic routes utilizing distinct metabolic enzymes. In both routes, however, cysteine is first converted into cysteinesulfinate, which can then be converted to hypotaurine and taurine or directly converted to alanine. The increased levels of hypotaurine and taurine and the decreased levels of alanine suggest that cysteine catabolism to taurine is potentially preferred in CRC. As these processes occur primarily in the liver, these changes

may reflect systemic metabolic dysregulation. However, the taurine transporter has been shown to be over-expressed in CRC and is currently being explored as a diagnostic and therapeutic target (PMID: 24781822). The elevated level of cysteine itself may be the result of up-stream changes in methionine metabolism (methionine can be catabolized to cysteine), which is often seen during oxidative stress.

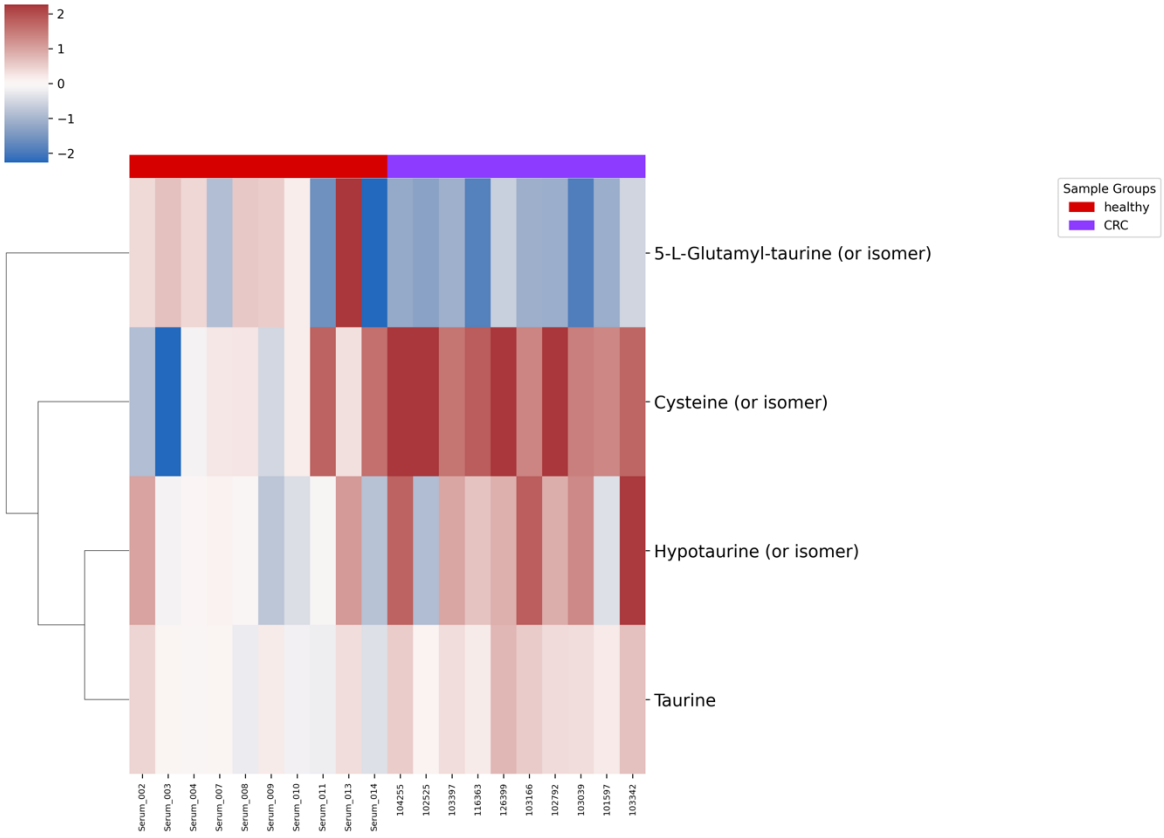

Figure 15: Heatmap of taurine and hypotaurine metabolism-related metabolites across healthy and diseased samples. Each row represents a metabolite within the pathway, while columns correspond to the experimental conditions.

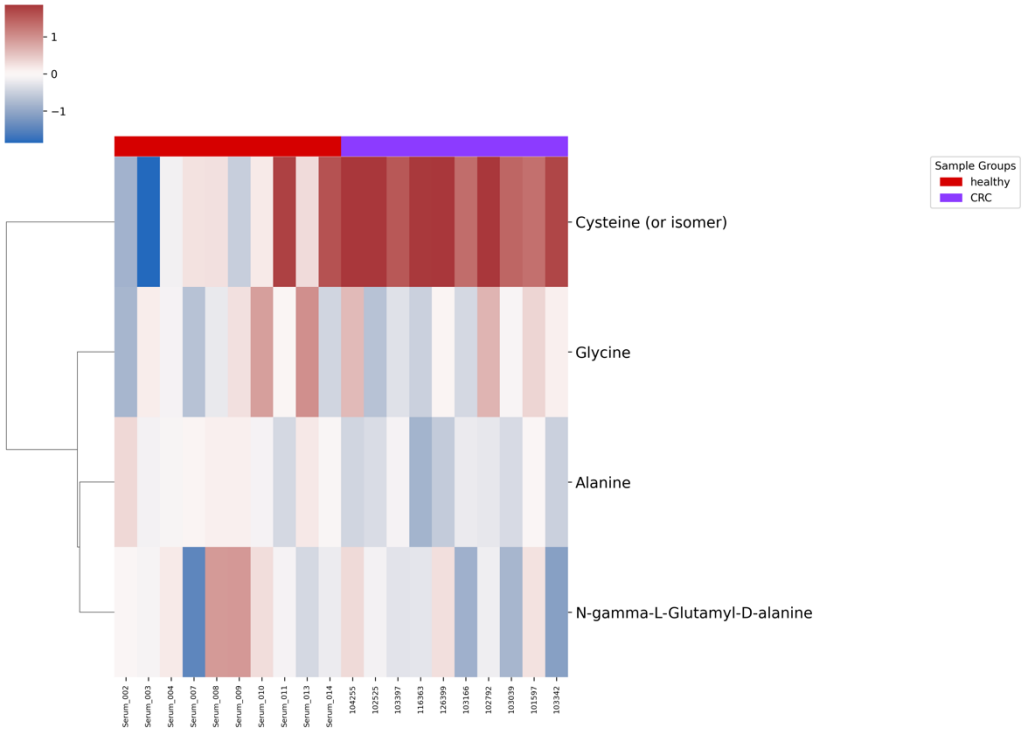

Figure 16: Heatmap of glutathione metabolism related-metabolites across healthy and diseased samples. Each row represents a metabolite within the pathway, while columns correspond to the experimental conditions.

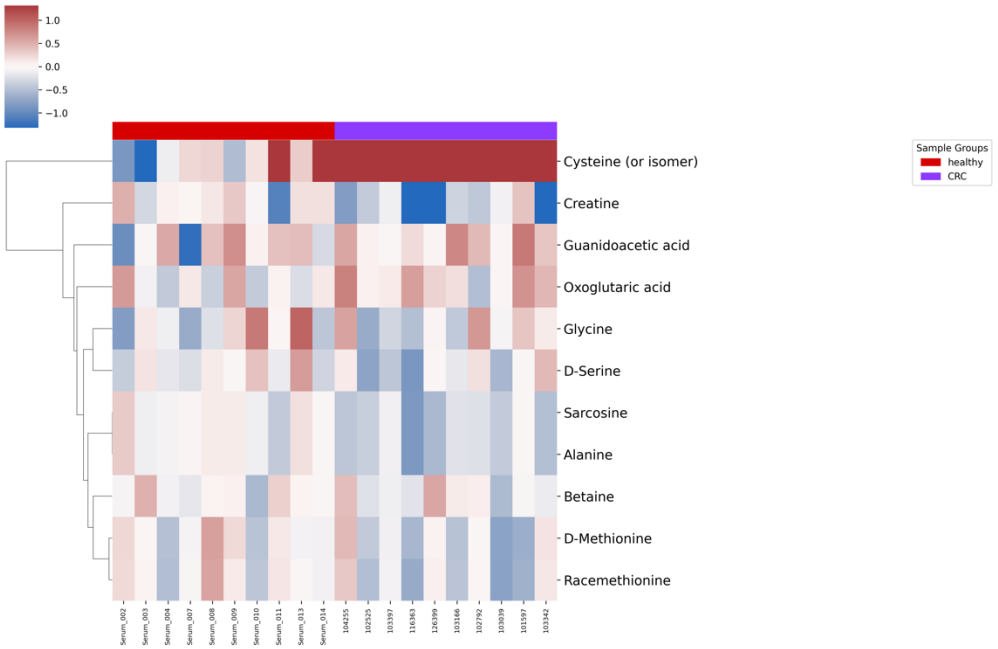

Figure 17: Heatmap of the levels of glycine and serine metabolism compounds in healthy and diseased samples. Each row represents a metabolite within the pathway, while columns correspond to the experimental conditions.

When examining the top three metabolites that are increased the highest in CRC relative to the healthy samples, we see three exogenous compounds that all have a massive increased concentration in CRC. The highest, 2-methyl-3-methoxy-4-phenylbutyric acid is elevated over 500x when compared to healthy. Little is known about this compound, but it has been detected in blood previously. The structure of this compound is similar to that of the second most increased compound, N-(2-Hydroxyethyl)-2-(1-isoquinolinylmethylene) hydrazinecarbothioamide, as they both contain benzene rings. Given the drastic difference in abundance of these compounds in CRC and the similarity of their chemical structures, they are likely secondary metabolites of therapeutics given to the CRC patients.

11-Methyl-7-oxatetracyclododecane, on the other hand, is a known flavor additive to many foods and is not well studied. However, predictions based on the structure of this compound suggest it may have biological roles in membrane formation and fatty acid metabolism. While further validation is required, these results suggest that ingestion of 11-Methyl-7-oxatetracyclododecane may be a risk factor for developing CRC.

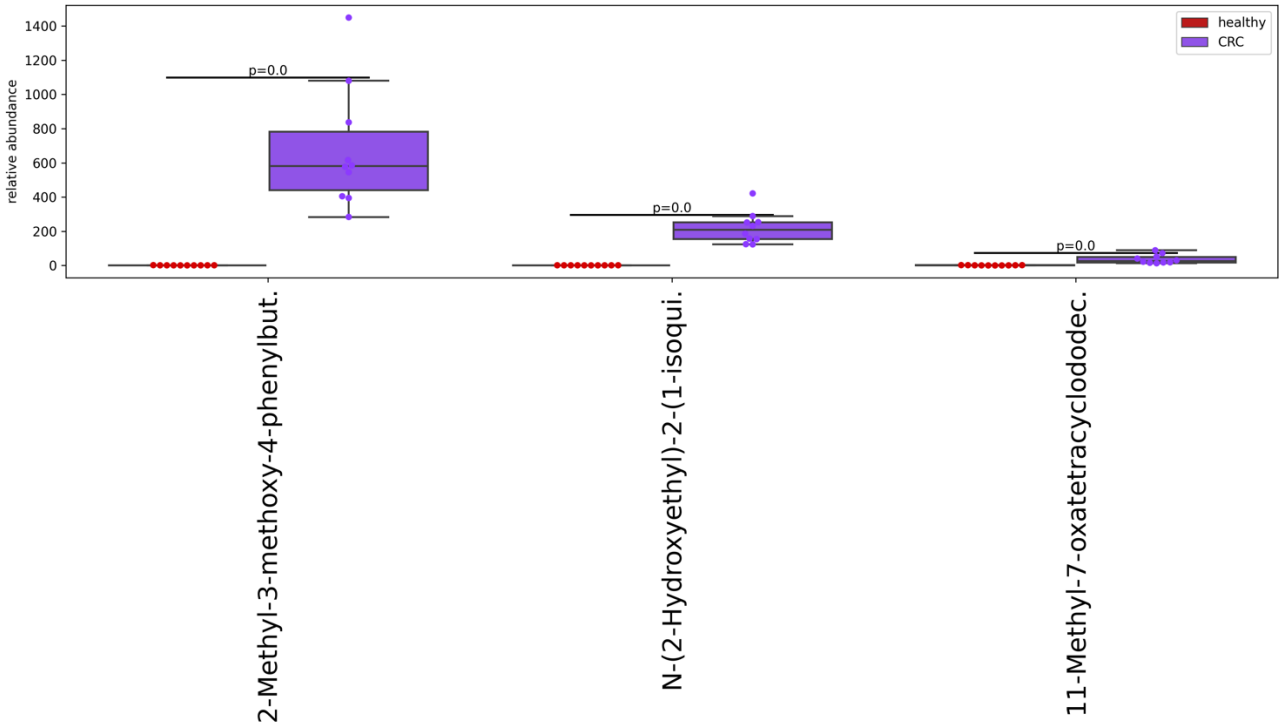

Figure 18: Top 3 metabolites showing the greatest increase in abundance in CRC. The boxplots above show the levels of the top 3 most metabolites identified at MSI Level 1 and 2 with the most increased abundance in CRC vs healthy.

When looking at the metabolites that showed the greatest decrease in CRC when compared to healthy samples, ascorbic acid (Vitamin C) was the most reduced. In addition to the connection with IBD discussed above, ascorbic acid is a known endogenous tyrosine kinase inhibitor. Multiple tyrosine kinase inhibitors are in development for CRC treatment and one tyrosine kinase inhibitor has been approved by the FDA for treatment of CRC.

The second most decreased metabolite is the short-chain acylcarnitine CAR 3:0:Me2 (also referred to as pivaloylcarnitine). Short-chain acylcarnitines are the most abundant class of acylcarnitine in humans and are involved with the transport of acetyl units into the mitochondria for oxidation. Short-chain acylcarnitines have also been implicated in IBD (a risk factor for CRC) as many of these short-chain acyl units are derived from the gut (PMID: 33141762).

The third most decreased metabolite is mevalolactone. Mevalolactone is rapidly interconverted with mevalonic acid, which is an intermediate in steroid biosynthesis. Additionally, mevalonic acid inhibits HMG-CoA reductase activity. HMG-CoA reductase inhibition is often accomplished pharmacologically with statins to decrease cholesterol levels. Therefore, the differences in mevalolactone levels suggests changes to cholesterol metabolism and associated SHEBP-2-mediated signaling pathways.

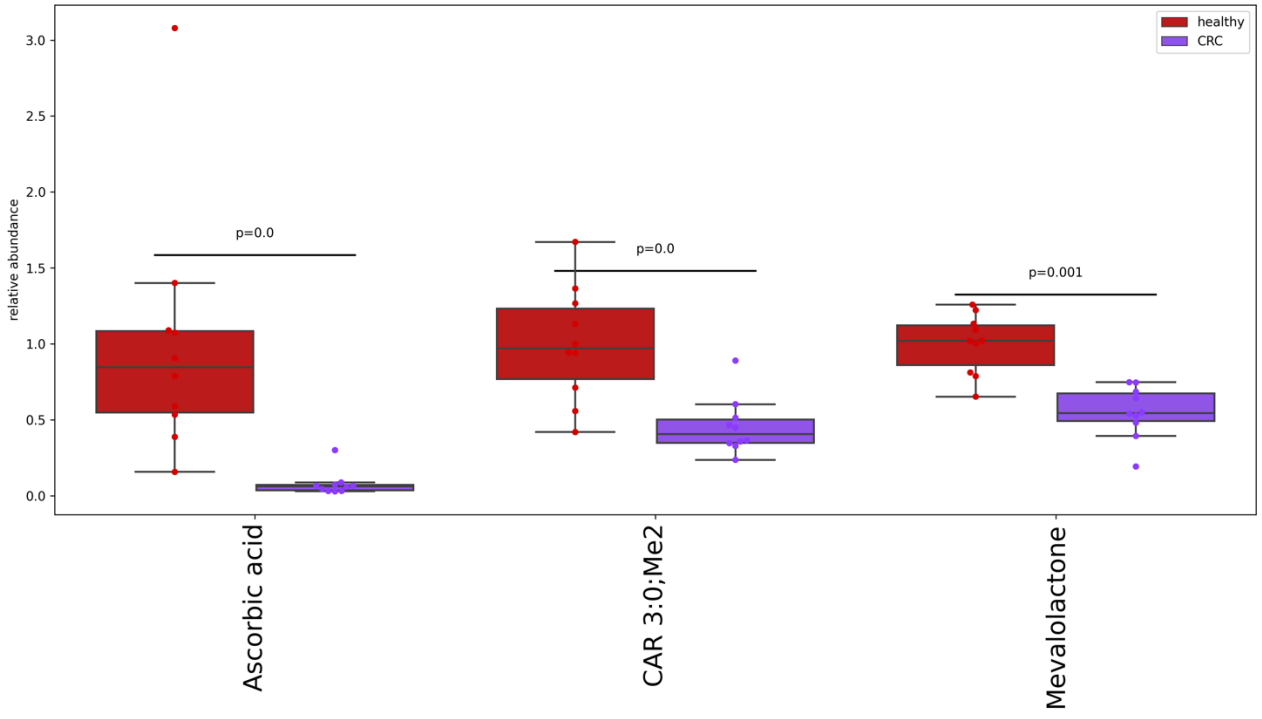

Figure 19: Top 3 metabolites showing the greatest decrease in abundance in CRC. The boxplots above show the levels of the top 3 most metabolites identified at MSI Level 1 and 2 with the most decreased abundance in CRC vs healthy.

Conclusions

In summary, the metabolomics study of healthy and CRC serum samples revealed multiple modules of metabolic dysregulation. Over, 1,100 identified metabolites were measured in the serum samples. The overall metabolic profiles of these samples clearly segregate into healthy and CRC groups. When examining the specific metabolite abundance differences between the two sample groups, multiple endogenously and exogenously derived metabolites showed statistically significant differences, demonstrating the value of untargeted profiling for CRC. The primary findings are summarized below:

- Changes in the lipid profile suggest strong dietary or digestive differences between CRC and healthy individuals:

  Unsaturated fatty acids and octadecanoids were enriched for statistically significant metabolites. These molecules cannot be produced endogenously and are derived from ingestion of nuts, oils, and other foods. The dysregulation of unsaturated fatty indicates substantial changes in eating habits between healthy and CRC patients or the ability to digest these nutrients due to gut perturbations from CRC. While derived from exogenous sources, many of these fatty acids have important signaling roles, serve as energy substrates, and are related to changes in inflammation-related pathways. One of metabolites with the lowest abundance in CRC relative to healthy samples was an acylcarnitine which is an intermediate of the fatty acid oxidation pathway. As CRC incidence is known to be affected by eating patterns and other lifestyle factors, understanding the influence of unsaturated fatty acids on CRC progression is worthy of further investigation. Mevalolactone, one of the metabolites with the lowest abundance in CRC relative to healthy individuals, has a key role in controlling lipid and cholesterol metabolism.

- IBD-related findings:

  IBD is a known risk factor for developing CRC. Accordingly, many of the metabolic alterations identified in this analysis have previously been implicated in IBD. Two of the strongest metabolite changes were in ascorbic acid (vitamin C) and dehydroascorbic acid. These metabolites are involved with the conversion of dopamine to norepinephrine. The enzyme responsible for this conversion, has a known SNP that is over-represented in IBD patients and reduces enzyme abundance and activity. As ascorbic acid is an endogenous tyrosine kinase inhibitor (TKI) and multiple TKIs are in development for CRC, understanding the levels of ascorbic acid and mutation status may be critical factors for guiding CRC treatment with TKIs. Beyond ascorbic acid, a short-chain acylcarnitine was one of the most decreased metabolites in CRC. Short-chain acylcarnitines have been implicated in IBD previously.

- Inflammatory signature in CRC:

  Multiple enriched metabolic pathways relate to inflammation. Octadecanoids and prostaglandins were both increased in CRC samples when compared to healthy. This is indicative of greater systemic inflammation in CRC. A related finding was the presence of anti-inflammatories, such as gemcabene, present in higher levels in the CRC samples.

- Exposures measured in the serum samples suggest potential risk factors for CRC:

Many of the metabolites showing the highest fold-change in CRC relative to healthy are exposure derived compounds. The two highest are likely derived from therapeutics given to the CRC patients. However, 11-Methyl-7-oxatetracyclododecane, the third most increased compound, is a flavor molecule added to many foods. CRC is distinct from many other cancers in that its incidence is increasing in recent years. One explanation for this is toxic compounds in the environment that cause pathology. This molecule is one such candidate that is worthy of further investigation.

- Cysteine Catabolism

Cysteine was highly elevated in CRC samples. This may be the result of up-stream changes in methionine metabolism, which is often seen during oxidative stress. Cysteine can be catabolized into both taurine and alanine through different metabolic pathways. However, alanine showed a decreased abundance in CRC while taurine was increased. Both of these pathways primarily take place in the liver, suggesting potential dysregulation of peripheral tissue due to tumor-induced systemic metabolic changes. The taurine transporter has also been shown to be over-expressed in CRC and is currently being explored as a diagnostic and therapeutic target.

Next Steps

To better contextualize the metabolite abundance changes captured in this metabolomics study we recommend the follow-on experiments listed below:

- Proteomics analysis of the same serum samples to provide insight into the mechanistic changes that are occurring, specifically related to ascorbate and cysteine metabolism. Measurement of inflammatory proteins in particular would provide greater clarity for which metabolic changes are a result of system inflammation as opposed to tumor metabolism.

- Profiling (both metabolomic and proteomic) of serum from patients with IBD, but not CRC, would help stratify the metabolic changes from IBD with CRC.

- Evaluation of identified exogenous compounds on CRC cells in vitro through a dose-response analysis providing different levels of compound (such as 11-Methyl-7-oxatetracyclododecane) and measuring the response of metabolites (and/or proteins) to treatment with the compound.

- Profiling of primary tumor tissue and healthy tissue to determine which metabolic alterations are specific to the tumor as opposed to systemic changes.

- Expanding the sample set to a larger cohort that includes early-stage patients to enable identification of early markers of disease to improve outcomes.

Experimental Methods

Sample Preparation

Sample handling and storage

Samples were frozen at -80°C after receipt.

Polar Metabolite Extraction

A 55 mL aliquot of serum was transferred onto a solid phase extraction (SPE) system and 5 mL was taken from each sample to form the pooled quality-control sample. 200 mL of 1:1 ACN:MeOH was added to each well and shaken for 1 min at room temperature at 360 rpm, followed by a 10 min incubation at 4°C. Next, 150 mL of 2:2:1 MeOH:ACN:H20 was added to each well and shaken again for 10 min at room temperature. Polar metabolites were then eluted into the 96-well collection plate by using the positive pressure manifold. This is repeated with 150 mL of 2:2:1 MeOH:ACN:H20 into the same collection plate. Polar eluates were then covered and stored at -80°C until LC/MS analysis.

Lipid Metabolite Extraction

The SPE plates from the polar metabolite extraction were washed twice with 500 mL 1:1 MTBE:MeOH to elute lipids into a new collection plate with a positive pressure manifold. The combined eluates were dried under a stream of nitrogen at room temperature and reconstituted with 200 mL 1:1 IPA:MeOH prior to LC/MS analysis.

Metabolomics Assays

LC/MS analysis of polar metabolites

LC/MS mobile phases A and B were prepared as follows: A) 20mM ammonium bicarbonate, 0.1% ammonium hydroxide, 5% ACN, 2.5 mM medronic acid and B) 95% ACN.

A 4 mL aliquot of polar metabolite extract was analyzed with HILIC/MS (HILICON iHILIC (P)-Classic, 2.1 x 100 mm) by using the following linear gradient at a flow rate of 250 mL/min: 0-1 min: 90% B, 1-12 min: 90-35% B, 12-12.5 min: 35-25% B, 12.5min-14.5 min: 25% B. The column was re-equilibrated with 20 column volumes of 90% B. Mass spectrometry analysis was completed with a mass range of 50-1500 m/z on an Agilent 6546 QTOF in both positive and negative mode.

MS/MS data was acquired in a data-dependent iterative fashion with a 1.3 m/z isolation window.

LC/MS analysis of lipid metabolites

LC/MS mobile phases A and B were prepared as follows: A) 5:3:2 H2O:ACN:IPA, 10 mM ammonium formate, and 5 mM Agilent deactivator additive and B) 1:9:90 H2O:ACN:IPA, 10 mM ammonium formate.

A 4 mL aliquot of lipid metabolite extract was analyzed with RPLC/MS (Waters Acquity Premier HSS T3, 2.1 x 100 mm) by using the following linear gradient at a flow rate of 250 mL/min: 0-2 min: 30% B, 2-17 min: 30-75% B, 17-20 min: 75-85% B, 20-23 min: 85-100% B, 23-26 min: 100% B, 26-27 min: 100-30% B. The column was re-equilibrated for 5 min. Mass spectrometry analysis was completed with a mass range of 100-1700 m/z with 1 scan/sec.

MS/MS data was acquired in a data-dependent iterative fashion with a 1.3 m/z isolation window.

Data Pre-Processing

Metabolite detection and identification

Metabolite signals (features) were detected in the LC/MS data with in-house software. Features were aligned across samples and features with intensities greater than 1/3 of the corresponding intensity in the QC sample were classified as contaminates and removed from future analysis. Feature degeneracy (isotopes, adducts, fragments, etc.) were identified through clustering and ion assignment.

Metabolites were structurally identified through searching against a database composed of known metabolites found in RefMet, LipidMaps, and HMDB and comparing isotope patterns and MS/MS fragmentation data (when available). Metabolite identifications were classified based on the Metabolomics Standards Initiative scoring scheme with values ranging from Level 1 to Level 4. Level 1 identifications are supported by experimental retention times and MS/MS spectra from authentic standards and an isotope pattern match. Level 2 identifications were made on based on isotope pattern matches of greater than 90% similarity (reverse dot-product) and an MS/MS match of greater than 50% similarity (entropy similarity) along with a predicted retention time match of <2 min. Level 3 identifications were made based on an isotope pattern and predicted retention time match with the same cutoffs as defined above. Level 4 identifications are signals with matches to the database, but no matches meeting the matching criteria listed above. Compounds with no match to our database were removed from downstream analysis to ensure actionable hits. For lipid compounds, the best matching lipid species is reported. However, the location of double bonds within acyl chains of a lipid species is not discernable from the LC/MS methods utilized. Thus, the double bond positions listed are not meaningful. Panome Bio can confirm any selected metabolite identifications for compounds for which an authentic standard is available.

Data normalization and curation

Metabolomics data from all assays were concatenated. Metabolite signals were discarded if the coefficient of variation (CV) amongst the quality control samples was greater than 25% for Level 1-2 metabolites or greater than 15% for Level 3-4 metabolites. Missing values were imputed by using half of the minimum detected intensity for each metabolite. Batch effects and other technical variability were normalized by fitting a random forest model to predict and correct for the drift in intensity estimated based on analytical batch, run order, and bioinformatics processing batch for each metabolite. Metabolite intensities were log2 transformed prior to statistical analysis.

Metabolite identifications and intensities were manually reviewed for concordance and accuracy.

Statistical Analysis

Hierarchical clustering analyses was performed based on log2(fc) transformed values and Euclidian distance. Clusters were calculated with the UPGMA algorithm.

Over-representation analysis was performed using a Fisher's Exact test, which compares the expected number of metabolites identified from a particular comparison in each pathway to the actual number identified.

Appendix

Supplementary Files

1. Raw_data.zip: raw data in an open (.mzML) format
2. SupplementaryTables.xlxs: excel file containing sample metadata, technical variation metrics for each metabolite, all detected unique metabolite features and their raw intensities and normalized intensities, the metabolite identifications and confidence for each feature, the detailed results of the statistical analyses.
3. Plots.zip: high-resolution images for all figures in this report as well as boxplots showing the intensity of each metabolite across all experimental conditions. Additionally, extracted ion chromatograms (EICs) for all signals are provided to visualize the signal in the raw data. MS/MS fragmentation data matches are also included.

Glossary

1. LC/MS: Liquid chromatography couple to mass spectrometry. The analytical technique used for the metabolite assays.
2. Retention time: The elution time of metabolites in the LC portion of the analysis. Retention time is determined by the chemical properties of the metabolite (e.g., hydrophobicity)
3. m/z: The mass (m) to charge (z) ratio of the measured ions in the MS portion of the analysis. In metabolomics, the charge of the metabolites is typically one or two. Thus, this can be thought of as just the molecular mass of the ions.
4. Features/metabolites signals: Signals in the raw data that represent unique compounds or metabolites measured in the LC/MS experiment.
5. Intensity: The relative abundance of the compound/metabolite in the sample.
6. Quality control sample: A sample repeatedly analyzed to measure system stability and determine signal drift.

panomebio

Contact us to initiate a project

www.panomebio.com

info@panomebio.com

4340 Duncan Avenue, St. Louis, MO 63110