Fecal metabolomics: assay performance and association with colorectal cancer

James J. Goedert*, Joshua N. Sampson, Steven C. Moore, Qian Xiao, Xiaoxin Xiong1, Richard B. Hayes2, Jiyoung Ahn3, Jianxin Shi and Rashmi Sinha

Epidemiology and Biostatistics Program, Division of Cancer Epidemiology and Genetics, National Cancer Institute, 9609 Medical Center Drive, Bethesda, MD 20892-9704, USA, 1Information Management Services, 6110 Executive Boulevard, Rockville, MD 20852, USA and 2Division of Epidemiology, Department of Population Health, New York University School of Medicine, 650 First Avenue, #518, New York, NY 10016, USA

*To whom correspondence should be addressed. Tel: +1 240 276 7103; Fax: +1 240 276 7806; Email: goedertj@mail.nih.gov

Metabolomic analysis of feces may provide insights on colorectal cancer (CRC) if assay performance is satisfactory. In lyophilized feces from 48 CRC cases, 102 matched controls, and 48 masked quality control specimens, 1043 small molecules were detected with a commercial platform. Assay reproducibility was good for 527 metabolites [technical intraclass correlation coefficient (ICC)>0.7 in quality control specimens], but reproducibility in 6-month paired specimens was lower for the majority of metabolites (within-subject ICC <0.5). In the CRC cases and controls, significant differences (false discovery rate ≤0.10) were found for 41 of 1043 fecal metabolites. Direct cancer association was found with three fecal heme-related molecules [covariate-adjusted 90th versus 10th percentile odds ratio (OR) = 17–345], 18 peptides/amino acids (OR = 3–14), palmitoyl-sphingomyelin (OR = 14), mandelate (OR = 3) and p-hydroxy-benzaldehyde (OR = 4). Conversely, cancer association was inverse with acetylaminophen metabolites (OR <0.1), tocopherols (OR = 0.3), sitostanol (OR = 0.2), 3-dehydrocarnitine (OR = 0.4), pterin (OR = 0.3), conjugated-linoleate-18-2N7 (OR = 0.2), N2-furfuryl-glucine (OR = 0.3) and p-aminobenzoate (PABA, OR = 0.2). Correlations suggested an independent role for palmitoyl-sphingomyelin and a central role for PABA (which was stable over 6 months, within-subject ICC 0.67) modulated by p-hydroxy-benzaldehyde. Power calculations based on ICC's indicate that only 45% of metabolites with a true relative risk 5.0 would be found in prospectively collected, prediagnostic specimens from 500 cases and 500 controls. Thus, because fecal metabolites vary over time, very large studies will be needed to reliably detect associations of many metabolites that potentially contribute to CRC.

Introduction

Metabolomics technology can measure hundreds of small molecules in a biospecimen, which may advance understanding of mechanisms and facilitate early diagnosis of disease. Colorectal cancer (CRC) is a good candidate, as it arises as a consequence of genetic mutations that accumulate and are driven by carcinogens related to diet (1). By convention, biomarker assays have been developed for serum or urine, but this approach may be limited because these fluids are anatomically remote from the gut mucosa in which CRC arises. Metabolomic study of feces may prove to be more effective, because feces is in close proximity to the colorectal mucosa and is a product of interactions between dietary components and the microbiota. In addition to host cell metabolites that may be shed into the gut lumen, the gut microbiota is thought to contribute to colorectal neoplasia through several immunologic and metabolic pathways (2,3). The gut microbiota produces many important metabolites including short-chain fatty acids, biotin and vitamin K (4,5) and may affect the metabolism of suspected dietary carcinogens, such as benzo(a)pyrene and acetaldheyde (6–10).

To date, metabolic analyses of fecal samples have mostly been restricted to experimental studies in animals and small cross-sectional studies in humans (2,5,11). A few have considered CRC (12–16), but comprehensive identification of human CRC-associated metabolic end products is lacking. Thus, the current study had three objectives. We quantified technical, between-subject and within-subject variabilities in fecal metabolites as determined by ultra high-performance liquid phase chromatography and gas chromatography coupled with tandem mass spectrometry (HPLC-GC/MS-MS). Then, in a case–control study, we compared fecal metabolomic profiles in a small group of well-characterized CRC cases and matched controls. Finally, we estimated expected power for a larger case–control study nested within a prospective cohort to detect true associations between a fecal metabolite and a subsequent disease event.

Materials and methods

Study participants and specimens

The current project used data and stored frozen specimens from a CRC case–control study of fecal mutagens that was reviewed and approved by an Institutional Review Board at the National Cancer Institute (17,18). Briefly, during 1985–89 at three Washington DC area hospitals, following signed informed consent, patients suspected to have CRC were recruited before surgery or initiation of treatment. Only newly diagnosed, histologically confirmed cases of adenocarcinoma of the colon or rectum were retained. Likewise, contemporaneous patients awaiting elective surgery for non-oncologic, non-gastrointestinal conditions at these hospitals were recruited as controls. A median of 6 days (interquartile range 3–13 days) prior to hospitalization and surgery, participants completed diet and demographic questionnaires and provided 2-day fecal samples that were frozen at home on dry ice and subsequently lyophilized. The 2-day lyophilates were pooled, mixed and stored at −40°C. Of 69 cases and 114 controls in the original study (17,18), the case–control analysis included 48 cases and 102 controls for whom at least 100 mg of lyophilized feces was available. Controls were frequency matched to cases by gender and body mass index. Tumors in the cases were classified by stage and site in the large intestine.

To quantify technical variability of the metabolomics assay, 48 identical, masked aliquots of lyophilized feces were prepared from the same patients, including 11 from each of 2 controls, and 26 from 7 other controls. To quantify within-subject variability over time, half of the 48 replicate control specimens were collected 6 months after the baseline specimens, using identical collection and handling methods.

Laboratory methods

A range of small molecules (most <1000 Daltons) was detected in the lyophilized fecal specimens by HPLC-GC/MS-MS (Metabolon, Durham, NC) as described previously (19,20). Briefly, non-targeted single methanol extraction was performed, followed by protein precipitation. Individual molecules and their relative levels were identified from the mass spectral peaks compared with a chemical reference library generated from 2500 standards, based on mass spectral peaks, retention times and mass-to-charge ratios. The molecules include, but are not limited to, amino acids, carbohydrates, fatty acids, androgens and xenobiotics. Volatile molecules, such as short-chain fatty acids, may be lost during lyophilization or extraction. However, such loss is generally equivalent across specimens, and lyophilization is optimal for fecal specimens to assure equal loading of dry weight.

Assay performance statistical methods

As we have previously evaluated overall reliability and validity of this platform for serum and urine (21), herein we assessed the technical, within-subject and between-subject variability of the fecal metabolite data by calculating intra-class correlation coefficients (ICC) for the fecal quality control samples, as follows. We decomposed the total variance of each metabolite, $\sigma^2$, into three
different components: the between-subject variance, $\sigma_w^2$, which represents the variance of the ‘usual’ level for subjects in a population; the within-subject variance, $\sigma_t^2$, which represents the variability over time around the ‘usual’ level within an individual and the technical variance, $\sigma_e^2$, which is the variance introduced by measurement error in the laboratory procedures.

From these three variance components, we defined the following additional quantities:

1) Technical ICC: the proportion of the total variance that is attributed to biological variance, as opposed to random laboratory variation. High technical ICC indicates high laboratory reproducibility.

$$ICC = \frac{\sigma_t^2}{\sigma_t^2 + \sigma_w^2 + \sigma_e^2}$$

2) Within-subject ICC: the ratio of between-subject variance/total variance. Higher $\pi_t$ indicates higher stability over time in vivo and thus higher power to act as a marker for long-term risk analysis.

$$\pi_t = 1 - \frac{\sigma_e^2}{\sigma_t^2 + \sigma_w^2 + \sigma_e^2}$$

Case–control statistical methods

Demographic data for the cases and controls were compared by Fisher’s exact test for categorical variables. For this pilot study of fecal metabolites, our primary analysis modeled the association between each metabolite and CRC by unconditional logistic regression, adjusting for body mass index, age, gender, race and hospital. Metabolite values below the level of detection were assigned the minimum observed value for that metabolite. For metabolites present in <80% of the individuals, we categorized the metabolite as present or absent (i.e. dichotomous) and report the associated odds ratio (OR) and its 95% confidence interval. For metabolites present in at least 80% of individuals, we performed a standard pathway analysis. Specifically, we evaluated whether the metabolites within predefined pathways were associated with CRC in the case–control study. Statistical power then was applied to selected metabolites that were observed to be associated with CRC in the case–control study.

All analyses were performed with SAS software version 9.1.3 (SAS Institute, Cary, NC) and the R statistical language version 3.0.1.

Results

Assay performance

In the fecal specimens, there were 1043 small molecules detected. These included 773 characterized and 270 uncharacterized (‘X’) molecules. Of the 579 molecules detected in at least 10% of the fecal specimens, overall laboratory reproducibility in masked replicate specimens was high, with the technical ICCs exceeding 0.7 for 527 (91%) of the metabolites (Figure 1A).

In addition to technical reproducibility, within-subject and between-subject variance contribute to total variance. Considering the between-subject/total variance ratio, which is equivalent to within-subject ICC, of variability. This average probability, or the average power, indicates the proportion of true metabolite–disease associations that we expect to discover in a given prospective study. Statistical power then was applied to selected metabolites that were observed to be associated with CRC in the case–control study.

Fig. 1. Distributions of ICCs across 579 metabolites detected in at least 10% of fecal specimens. Percentage beside each asterisk indicates the proportion of metabolites with at least the indicated ICC. (A) Technical ICCs, a measure of laboratory variability; 91% of metabolites have technical ICC $\geq0.7$. (B) Within-subject ICCs, a measure of stability over 6 months; 44% of metabolites have within-subject ICC $\geq0.5$. 
Figure 1B shows that within-subject ICC was relatively low. Only 44% of the metabolites had a within-subject ICC ≥0.5. Only 5% of the metabolites had a within-subject ICC ≥0.7.

**CRC case–control study**

The 48 CRC cases tended to be older than the 102 frequency-matched controls (mean 62.9 versus 58.3 years, \( P = 0.06 \)) but they did not differ by sex (60% male), body mass index, race, smoking history, attained education or hospital of recruitment (Supplementary Table 1, available at Carcinogenesis Online). The primary CRC tumors arose in approximately equal proportions in the proximal colon, distal colon and rectum (29, 33 and 27%, respectively). The cases included 35% with metastases at diagnosis (Dukes’ stage C/D), 42% with local invasion but no known metastases (Dukes’ stage B) and 21% with only localized disease (Dukes’ stage A).

Global assessment indicated that many of the 1043 fecal metabolites differed significantly between cases and controls (Supplementary Figure 1, available at Carcinogenesis Online). The prevalence of all 1043 fecal metabolites in cases and controls is presented in Supplementary Table 2, available at Carcinogenesis Online. Of these, 41 (3.9%) were significantly associated with CRC at FDR = 0.10. As summarized in Supplementary Table 3, available at Carcinogenesis Online, CRC was overrepresented with cofactors and vitamins (\( P = 0.032 \), especially tocopherol-related), xenobiotics (\( P = 0.006 \), especially drugs and food components/plants) and marginally with lipids (\( P = 0.064 \) and uncharacterized metabolites (\( P = 0.074 \)).

Heme was detected in 29% of cases and 2% of controls (OR = 16.55, Table I). Heme met the Bonferroni threshold for statistical significance.

### Table I. Fecal metabolites associated with CRC at FDR 0.10

<table>
<thead>
<tr>
<th>Metabolite</th>
<th>Prevalence (%)</th>
<th>( P ) value</th>
<th>OR (CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Heme-related</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heme</td>
<td>29</td>
<td>1.5E-05</td>
<td>16.55 (3.46–79.09)</td>
</tr>
<tr>
<td>X_18565</td>
<td>67</td>
<td>&lt;1.0E-10</td>
<td>345.3 (35.43–3364.92)</td>
</tr>
<tr>
<td>X_19549</td>
<td>48</td>
<td>1.8E-10</td>
<td>29.81 (7.99–111.14)</td>
</tr>
<tr>
<td><strong>Cofactors and vitamins</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>α-Tocopherol</td>
<td>96</td>
<td>6.0E-03</td>
<td>0.25 (0.08–0.74)</td>
</tr>
<tr>
<td>γ-Tocopherol</td>
<td>98</td>
<td>1.8E-03</td>
<td>0.26 (0.1–0.64)</td>
</tr>
<tr>
<td>Pterin</td>
<td>90</td>
<td>4.0E-03</td>
<td>0.33 (0.15–0.73)</td>
</tr>
<tr>
<td><strong>Xenobiotics</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4-Acetamidophenol</td>
<td>92</td>
<td>9.1E-06</td>
<td>0.04 (0.01–0.27)</td>
</tr>
<tr>
<td>2-Hydroxyacetaminophen sulfate</td>
<td>0</td>
<td>1.4E-03</td>
<td>0 (0–Infinity)</td>
</tr>
<tr>
<td>3-Cystein-S-YL-acetaminophen</td>
<td>2</td>
<td>7.6E-04</td>
<td>0.07 (0.01–0.58)</td>
</tr>
<tr>
<td>ρ-Acetamidophenylglycuronide</td>
<td>0</td>
<td>1.9E-03</td>
<td>0 (0–Infinity)</td>
</tr>
<tr>
<td>PABA</td>
<td>98</td>
<td>8.8E-04</td>
<td>0.22 (0.08–0.57)</td>
</tr>
<tr>
<td>N-2-Furoyl-glycine</td>
<td>98</td>
<td>3.8E-05</td>
<td>0.26 (0.1–0.69)</td>
</tr>
<tr>
<td>Sitostanol</td>
<td>90</td>
<td>3.0E-04</td>
<td>0.20 (0.08–0.5)</td>
</tr>
<tr>
<td>ρ-Hydroxybenzaldehyde</td>
<td>100</td>
<td>1.6E-03</td>
<td>3.96 (1.56–10.05)</td>
</tr>
<tr>
<td>Mandelate</td>
<td>96</td>
<td>5.1E-03</td>
<td>3.32 (1.36–8.08)</td>
</tr>
<tr>
<td><strong>Lipids</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Palmitoyl-sphingomyelin</td>
<td>98</td>
<td>2.1E-06</td>
<td>13.6 (3.9–47.41)</td>
</tr>
<tr>
<td>Conjugated linoleate-18-2N7</td>
<td>96</td>
<td>2.3E-04</td>
<td>0.16 (0.06–0.47)</td>
</tr>
<tr>
<td>3-Dehydroacarnitine</td>
<td>92</td>
<td>1.6E-03</td>
<td>0.35 (0.18–0.7)</td>
</tr>
<tr>
<td><strong>Amino acids</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Histidine</td>
<td>100</td>
<td>2.7E-05</td>
<td>7.46 (2.56–21.8)</td>
</tr>
<tr>
<td>Cis-Urocanate</td>
<td>100</td>
<td>6.6E-04</td>
<td>5.19 (1.88–14.37)</td>
</tr>
<tr>
<td><strong>Peptides</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tryptophyl-glycine</td>
<td>100</td>
<td>1.7E-06</td>
<td>14.34 (4.09–50.35)</td>
</tr>
<tr>
<td>Leucyl-tryptophan</td>
<td>100</td>
<td>5.6E-05</td>
<td>7.96 (2.58–24.56)</td>
</tr>
<tr>
<td>Alanyl-histidine</td>
<td>60</td>
<td>4.1E-05</td>
<td>5.28 (2.29–12.16)</td>
</tr>
<tr>
<td>Histidyl-glycine</td>
<td>67</td>
<td>3.7E-05</td>
<td>4.94 (2.23–10.94)</td>
</tr>
<tr>
<td>Tyrosyl-glutamine</td>
<td>96</td>
<td>5.7E-04</td>
<td>4.97 (1.84–13.45)</td>
</tr>
<tr>
<td>Histidyl-alanine</td>
<td>92</td>
<td>3.3E-04</td>
<td>6.87 (2.22–21.28)</td>
</tr>
<tr>
<td>Valyl-aspartate</td>
<td>100</td>
<td>3.6E-04</td>
<td>7.10 (2.18–23.12)</td>
</tr>
<tr>
<td>Pyro-glutamyl-glycine</td>
<td>19</td>
<td>1.8E-03</td>
<td>7.12 (1.98–25.68)</td>
</tr>
<tr>
<td>Alanyl-leucine</td>
<td>100</td>
<td>2.9E-03</td>
<td>4.94 (1.63–14.95)</td>
</tr>
<tr>
<td>Alanyl-tryptophan</td>
<td>100</td>
<td>3.4E-03</td>
<td>4.09 (1.51–11.08)</td>
</tr>
<tr>
<td>Histidyl-phenylalanine</td>
<td>96</td>
<td>1.7E-03</td>
<td>5.33 (1.76–16.19)</td>
</tr>
<tr>
<td>Leucyl-glutamate</td>
<td>100</td>
<td>2.5E-03</td>
<td>5.12 (1.67–15.65)</td>
</tr>
<tr>
<td>Leucyl-serine</td>
<td>100</td>
<td>2.1E-03</td>
<td>5.34 (1.7–16.79)</td>
</tr>
<tr>
<td>α-Glutamyl-valine</td>
<td>100</td>
<td>4.6E-03</td>
<td>3.39 (1.39–8.3)</td>
</tr>
<tr>
<td>Prolyl-alanine</td>
<td>100</td>
<td>2.3E-03</td>
<td>3.21 (1.45–7.08)</td>
</tr>
<tr>
<td>Valyl-histidine</td>
<td>90</td>
<td>3.0E-03</td>
<td>4.43 (1.61–12.22)</td>
</tr>
<tr>
<td><strong>Uncharacterized molecules</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>X_19558</td>
<td>92</td>
<td>9.3E-05</td>
<td>9.11 (2.76–30.11)</td>
</tr>
<tr>
<td>X_16343</td>
<td>96</td>
<td>4.6E-03</td>
<td>3.77 (1.44–9.86)</td>
</tr>
<tr>
<td>X_17749</td>
<td>98</td>
<td>7.0E-04</td>
<td>0.22 (0.09–0.57)</td>
</tr>
<tr>
<td>X_19232</td>
<td>96</td>
<td>4.9E-04</td>
<td>0.22 (0.09–0.55)</td>
</tr>
<tr>
<td>X_19136</td>
<td>77</td>
<td>1.8E-03</td>
<td>0.19 (0.06–0.55)</td>
</tr>
</tbody>
</table>

\(^{a}\)Fraction above the limit of detection.

\(^{b}\)Ors, corresponding confidence intervals (CI) and \( P \) values were calculated by logistic regression, adjusted for continuous age, body mass index, sex, race and hospital. ORs compared 10th versus 90th percentiles for common metabolites (prevalence >80%), else presence versus absence for less common metabolites.
Fig. 2. Mean levels, by case–control status, of 11 CRC-associated metabolites in feces. These metabolites’ pairwise correlations and associations with CRC risk are presented in Figure 3.

Fig. 3. Pairwise correlations of 11 CRC-associated metabolites in feces of CRC cases (above) and matched controls (below). Double-headed arrows (↔) indicate direct (positive) correlations; blocked lines (→) indicate inverse (negative) correlations. Green indicates inverse correlation with CRC; red indicates direct correlation with CRC. Line weight indicates strength of correlation coefficient. Cases’ metabolites are generally more correlated than controls’ metabolites. Mean normalized levels of these metabolites are presented in Figure 2.
Fecal metabolomics with colon cancer

As shown in Table I, the CRC risk was reduced with four acetaminophen metabolites. Pairwise correlations of the four acetaminophen metabolites ranged $R = 0.29$–$0.82$ (median $R = 0.40$). CRC risk also was low with 2-hydroxyhippurate (salicylate), which did not meet FDR ($OR = 0.27$, $P = 0.004$), but not with salicylate, salicyluric-gluconuride, ibuprofen or hydroxy-ibuprofen ($P = 0.023$–$0.081$, data not presented).

Eighteen fecal peptides/amino acids were associated with increased CRC risk, and none was associated with decreased risk. These 18 peptides/amino acids were highly correlated with each other (median $R = 0.51$, interquartile range $R = 0.33$–$0.67$). Tryptophylglycine had very strong correlations ($R = 0.73$–$0.88$) with eight other dipeptides. One uncharacterized molecule (X16343) had very strong correlations ($R = 0.70$–$0.85$) with tryptophylglycine and five other dipeptides. The strongest inverse correlation observed was between tryptophylglycine and sitostanol ($R = -0.45$).

The 11 other CRC-associated molecules, which may functionally contribute to the disease, included pterin, 2-tocopherol, 5 xenobiotics and 3 lipids. Of these 11, eight were associated with lower CRC risk and three with higher risk (Table I). Figure 2 presents the mean levels of the eight reduced-risk (green) and three increased-risk (red) metabolites. Figure 3 presents pairwise correlations of these metabolites. Nearly all correlations were positive (arrows). In both cases and controls, the metabolite network was centered around $p$-aminobenzoate (PABA). Cases had lower levels (Figure 2) but many more and stronger correlations (Figure 3) of reduced-risk metabolites compared with controls.

Among potential candidate molecules (23), usredoxecholate was marginally associated with lower CRC risk (OR = 0.32, $P = 0.03$), but six other fecal bile acids were unrelated ($P = 0.30$–$0.98$, data not presented). CRC was not associated with 3-aminoisobutyrate, 3-aminoisobutyrate or 3-methyl-2-oxobutyrate ($P = 0.03$–$0.74$). The five uncharacterized CRC-associated molecules included a pair (X19232 and X17749) associated with very low CRC risk that had the strongest direct correlation that was observed ($R = 0.93$).

None of the metabolites that were very strongly associated with CRC differed significantly by tumor site (Supplementary Table 4A, available at Carcinogenesis Online). Two of them differed by stage of the tumor (Supplementary Table 4B, available at Carcinogenesis Online). Prevalence of heme was directly related to stage: 10% with non-invasive cases, 20% of locally invasive cases and 47% of metastatic cases. Better performance, particularly higher sensitivity in preinvasive cases, would be expected with fecal immunochemical tests that are licensed and currently in widespread use (24) and perhaps with detection of methylated or uncharacterized CRC-associated oncogenes in feces (25). We observed higher sensitivity, as well as good specificity, with our two partially characterized, heme-related molecules X19549 and especially X18565, suggesting that further investigations with them are warranted.

Second, four acetaminophen-related molecules in feces were associated with a very low risk of CRC. This probably reflects confounding, with frequent use of acetaminophen among controls (30.4% with 4-acetaminophen sulfate in feces, versus 12.5% in cases), approximately half of whom were awaiting orthopedic surgery (17). Detection of salicylate and ibuprofen did not differ between cases and controls, despite prior evidence in favor of risk reduction (26).

Third, CRC risk was positively associated with 18 fecal peptides/amino acids; there were no inverse associations with peptides/amino acids. These metabolites included histidine, a metabolite of histidine (cis-Urocanate), and five histidine-containing dipeptides. This is consistent with previous reports that amino acids were more abundant in CRC tumor tissue compared with normal colonic mucosa (27–30). Paradoxically, two of our dipeptides, histidyl-alanine and histidyl-glycine, were reported to inhibit the growth of three different cancer cell lines (31). Moreover, histidine was not among the 11 amino acids reported as higher in previous CRC case–control analyses of feces, although histidine may not have been readily detected with the older methods that were used (12–15). In our study, tryptophyl-glycine had the highest risk and strong correlations with other CRC-associated dipeptides, and it was reported to have moderately high activity in the Salmonella typhimurium mutagenesis assay (32). It is likely, however, that the peptide/amino acid associations are not etiologic. Rather, the breadth of CRC-associated amino acids and dipeptides may reflect shedding into cases’ feces due to high levels of cell division, cell death and protein degradation.

Fourth, unlike heme, acetaminophen and peptides/amino acids, the 11 other CRC-associated molecules cannot be readily dismissed as the result of fecal shedding or confounding. Their pairwise correlations are presented in Figure 3. Compared with controls, the correlations in cases were stronger, more numerous, centered around PABA and included strong inverse associations of $p$-hydroxy-benzaldehyde with linoleate-18-2N7 and N-2-furoyl-glycine. Linoleate-18-2N7, a lipid that can be generated by bacteria (33), was low in our CRC cases.
and was reported to inhibit cancer in animal models by modulating inflammation (34,35). Obstruction of the protective effect of linoleate by \(\beta\)-hydroxy-benzaldehyde is suggested by the inverse correlation that we observed (\(R = -0.46\) in cases versus \(R = -0.21\) in controls). 2-Furoyl-glycine, also low in our cases, is a xenobiotic found in cigarette smoke and especially in coffee (36). Coffee consumption is inversely associated with CRC risk (37), and our data suggest the same association with fecal 2-Furoyl-glycine, possibly reduced by \(\beta\)-hydroxy-benzaldehyde (Figure 3).

PABA can be processed by certain bacteria into acetyl coenzyme A for citric acid and fatty acid metabolism (38). Along with pterin, PABA also is centrally related to folate-mediated one-carbon metabolism (39). Folate was not detected in our stool specimens. However, CRC risk was reported to be increased with altered metabolism of folate (40,41), inconsistently associated with level of folate in plasma (42), and perhaps slightly increased with folate-deficient diet (43). Animal models suggest that folate repletion or supplementation upregulates inflammation pathways (43). The reports that CRC tissue has reduced levels of PABA, lipids and glucose (27–29) complement our findings.

Palmitoyl-sphingomyelin was strongly associated with a \(>10\)-fold increased risk of CRC. An independent effect is suggested by its inverse correlation with \(\beta\)-hydroxy-benzaldehyde in the cases (Figure 3). Sphingomyelinase signaling produces ceramide, a messenger that modulates both cell proliferation and apoptosis through critical pathways, including the mitochondrial and c-Jun N-terminal kinase systems that can generate reactive oxygen species (44). Sphingolipid alterations are common in cancer and may contribute to CRC risk through the WNT/\(\beta\)-catenin pathway (45). Moreover, sphingomyelinase activity in stool has been proposed as a screening test for CRC (46).

Finally, carnitine, sitostanol and the two tocopherols were low in cases and correlated with each other. Sitostanol is a vegetable oil derivative that may reduce absorption of cholesterol and perhaps tocopherols from the gut (47–49). Tocopherols, which are vitamin E, have antioxidant properties and may reduce the formation of nitrosamines from dietary nitrates (50). In a double blind, placebo-controlled trial, supplementation with \(\alpha\)-tocopherol was associated with a 20% lower incidence of CRC, although this was not statistically significant (51). Carnitine appears to reduce colonic neoplasia by reducing inflammation, inhibiting proliferation and increasing apoptosis (52–54).

Strengths of our study include careful processing and preservation of the fecal specimens, and our quantification of within-subject ICC, from which we could estimate statistical power with our cutting-edge fecal metabolomics platform. Our platform had high sensitivity and technical reproducibility, but it has limited ability to detect some volatile and larger molecules, such as calprotectin that has been associated with colonic inflammation and cancer (12–15,55). Our pilot study’s major limitations are its small size and cross-sectional, hospital-based case-control design. It provided no assessment of temporality and could only detect very strong associations with CRC. Because large prospective studies have not collected fecal samples, we conducted the pilot study to generate hypotheses that can be tested in the future.

Our calculations of statistical power for larger prospective studies to detect 2.5- to 5-fold RRs across metabolites with a range of ICCs is an additional strength.

In summary, complementing CRC-associated differences in fecal microbial diversity and composition that we found in the same cases and controls (56), herein we found that CRC was associated with differences in a diverse array of many small molecules in feces. Some of these, such as heme and peptides, are likely shed from the tumor and might facilitate earlier diagnosis. Others appear to represent differences in diet, medications, microbes (56,57) and the host, with PABA modulated by \(\beta\)-hydroxy-benzaldehyde playing a central role. Fecal metabolomics may prove to be a useful new tool for studies of
Fecal metabolomics with colon cancer

Cancer and other diseases. However, instability of metabolomic profiles over time mandate large, prospective studies with prediagnostic specimens to clearly identify targets for prevention, diagnosis and intervention.

Supplementary material
Supplementary Tables 1–4 and Figure 1 can be found at http://carcin.oxfordjournals.org/

Funding
Intramural Research Program of the National Cancer Institute, National Institutes of Health (Z01-CP-010214).

Acknowledgements
We thank Dr M.Schiffman for his visionary study that made this project possible, as well as the patients who participated.

Conflict of Interest Statement: None declared.

References
38. Butler,J.E. et al. (2007) Genomic and microarray analysis of aromatics degradation in Geobacter metallireducens and comparison to a Geobacter isolate from a contaminated field site. BMC Genomics, 8, 180.
2095


Received December 3, 2013; revised May 6, 2014; accepted May 30, 2014