Title: The potential of nanoflow liquid chromatography-nano electrospray ionisation-mass spectrometry for global profiling the faecal metabolome

Authors: Andrew J. Chetwynd, Lesley A Ogilvie, Jonathan Nzakizwanayo, Filip Pazdirek, Jiří Hoch, Cinzia Dedi, Duncan Gilbert, Alaa Abdul-Sada, Brian V Jones, Elizabeth M Hill

PII: S0021-9673(19)30396-6
DOI: https://doi.org/10.1016/j.chroma.2019.04.028
Reference: CHROMA 360174

To appear in: *Journal of Chromatography A*

Received date: 16 April 2018
Revised date: 1 April 2019
Accepted date: 11 April 2019

Please cite this article as: Chetwynd AJ, Ogilvie LA, Nzakizwanayo J, Pazdirek F, Hoch J, Dedi C, Gilbert D, Abdul-Sada A, Jones BV, Hill EM, The potential of nanoflow liquid chromatography-nano electrospray ionisation-mass spectrometry for global profiling the faecal metabolome, *Journal of Chromatography A* (2019), https://doi.org/10.1016/j.chroma.2019.04.028

The potential of nanoflow liquid chromatography-nano electrospray ionisation-mass spectrometry for global profiling the faecal metabolome

Andrew J. Chetwynd[1,2], Lesley A Ogilvie[3], Jonathan Nzakizwanayo[3], Filip Pazdirek[4], Jiří Hoch[4], Cinzia Dedi[3], Duncan Gilbert[5], Alaa Abdul-Sada[1], Brian V Jones[3,6], Elizabeth M Hill[1*]

[1] School of Life Sciences, University of Sussex, Falmer, Brighton, BN1 9QG, U.K

[2] School of Geography, Earth and Environmental Sciences, University of Birmingham, Edgbaston, B15 2TT

[3] School of Pharmacy and Biomolecular Sciences, University of Brighton, Brighton BN2 4GJ, UK

[4] Surgery Department, 2nd Medical Faculty of Charles University and University Hospital Motol, Prague, Czech Republic

[5] Sussex Cancer Centre, Royal Sussex County Hospital, Brighton BN2 5DA

[6] Department of Biology and Biochemistry, University of Bath, Bath BA2 7AY, UK

*To whom correspondence should be addressed

Tel: +44 1273678742

Fax: +44 1273677196

e.mail: e.m.hill@sussex.ac.uk

Highlights

- Metabolites in faecal extracts were profiled using nanospray-nanoscale LC-MS
- nUHPLC-nESI-TOFMS analyses gave repeatable peak retention times and intensities
- A pilot metabolomics study revealed differences between CRC and healthy patients
- Signalling metabolites were some key classes discriminating between patient groups

1.0 Abstract

Faeces are comprised of a wide array of metabolites arising from the circulatory system as well as the human microbiome. A global metabolite analysis (metabolomics) of faecal extracts offers the potential to uncover new compounds which may be indicative of the onset of bowel diseases such as colorectal cancer (CRC). To date, faecal metabolomics is still in its infancy and the compounds of low abundance present in faecal extracts poorly characterised. In this study, extracts of faeces from healthy subjects were profiled using a sensitive nanoflow-nanospray LC-MS platform which resulted in highly repeatable peak retention times (<2% CV) and intensities (<15% CV). Analysis of the extracts revealed wide coverage of the faecal metabolome including detection of low abundant signalling compounds such as sex steroids and eicosanoids, alongside highly abundant pharmaceuticals and tetrapyrrole metabolites. A small pilot study investigating differences in metabolomics profiles of faecal samples obtained from 7 CRC, 25 adenomatous polyp and 26 healthy groups revealed that secondary bile acids, conjugated androgens, eicosanoids, phospholipids and an unidentified haem metabolite were potential classes of metabolites that discriminated between the CRC and control sample groups. However, much larger follow up studies are needed to confirm which components of the faecal metabolome are associated with actual CRC disease rather than dietary influences. This study reveals the potential of nanospray-nanoflow LC-MS profiling of faecal samples from large scale cohort studies for uncovering the role of the faecal metabolome in colorectal disease formation.

Key words: metabolomics; nanoLC; nanoESI; Colon cancer, LC-MS; Faeces;

2.0 Introduction

The use of metabolomic analysis for the discovery of early markers of disease has been successfully achieved using matrices such as urine [1], plasma/serum [2], cerebrospinal fluid [3] and saliva.[4] However, faecal metabolomics is still very much in its infancy despite offering huge potential for biomarker discovery for diseases of the gut. [5] [6] To date, faecal metabolomic profiling has been applied to Crohn's disease [7], ulcerative colitis [7, 8], liver cirrhosis [9], colorectal cancer (CRC)[10, 11] and nutritional studies. [12] In addition, faecal analysis provides an insight into the health of the bowel microbiome via the analysis of microbial metabolites. [6] Changes in microbial metabolites such as secondary bile acids and tetrapyrroles may indicate changes in the population distribution or function of the microbiome which, in turn, may signal a disease state or be a potential cause of disease. [13, 14]

One potential application for faecal metabolomics is to investigate metabolite profiles associated with CRC and pre-cancerous polyps in order to aid understanding of disease formation and in the early diagnosis of the malignancy. CRC has the third highest morbidity and mortality rate of any cancer in the world, with a predicted 135,000 new cases in North America in 2017.[14] Early diagnosis of CRC significantly reduces overall mortality by allowing early intervention. To date, several metabolomic studies of CRC have been carried out using tumour tissue samples[15, 16], urine [17]and serum from diseased subjects. [18, 19] However the collection of these samples can be invasive and, in the case of urine and serum, distant from the tumour site. Faecal samples offer a non-invasive alternative which has the advantage of providing a more direct representation of the colonic environment and microbial activities in this compartment. Several CRC metabolomic studies of faeces have been undertaken, however, for the most part these have utilised nuclear magnetic resonance [20] or gas chromatography-mass spectrometry. [11, 21] Any LC-MS analysis of CRC in faeces has typically utilised a targeted metabolomic approach, and very few have employed LC-MS for a global faecal metabolomic analysis of CRC. [6, 21] One of the challenges of global metabolomics is to analyse as much of the metabolome as possible in the sample, including low abundance compounds alongside those present in high concentrations. This is particularly important in studies of CRC where changes in the concentrations of signalling metabolites such as secondary bile acids, sex steroids and eicosanoids have been associated with onset of the disease. [22-24]

A recent development in global metabolomics is the introduction of highly sensitive nanoflowUHPLC-nanosprayESI-TOFMS (nUHPLC-nESI-TOFMS) methodologies.[25, 26] nUHPLC technology uses columns with 100 µm internal diameters or smaller and flow rates of <1000 nL/min resulting in reduced chromatographic dilution and greatly improved ionisation efficiency, enabling low and high abundant molecules to be analysed simultaneously in complex matrices. This approach has been applied to urine [27, 28], plasma [29, 30] and cell cultures [31]. In these studies, both retention time and the mass spectral signal have proven to be highly repeatable and reproducible, while at the same time providing a highly sensitive metabolomic analysis. The properties of the nUHPLC-nESI-HRMS platform enable the analysis and separation of trace level metabolites which would normally not be detected using conventional LC-MS methodologies. Furthermore, advances in sample preparation methods, which include solid phase extraction (SPE) and pre-concentration, have enabled the injection of small sample volumes that are required for nanoscale LC-MS. Previous work has shown that hydrophilic/lipophilic balance (HLB) SPE phases are capable of good recoveries of a wide range of polar and lipophilic molecules including eicosanoids, bile acids, amino acids, nucleotides, free and conjugated steroids and a variety of pharmaceuticals and other xenobiotics, thus indicating that SPE is a suitable sample preparation method for metabolomic analysis [29, 32-34]. SPE is also often used

to remove salts from faecal water prior to LC-MS analyses for metabolomics [35, 36]. In this study, we used the SPE methods based upon our previous SPE nLC-nESI-MS studies where we examined the reproducibility, sensitivity and recovery of our sample preparation methods. In these studies we found the method to be repeatable and suitable for metabolomics profiling with recoveries of a variety of signalling compounds including eicosanoids, free and conjugated sex steroids and bile acids of between 60 and 85 % with a repeatability of between 1-13% [27, 29] .

The aim of this study was to assess the potential of nUHPLC-nESI-TOFMS to profile the metabolome, including signalling compounds, from faecal extracts from cohort studies. An offline SPE sample preparation and nUHPLC-nESI-TOFMS analysis methodology were used and the repeatability of the metabolomic analyses was assessed. A metabolomic investigation of a small pilot study comprising samples from recently diagnosed CRC, adenomatous polyp and healthy control subjects was performed to determine the potential of trace nUHPLC-nESI-TOFMS profiling to detect components of the faecal metabolome which could be associated with the onset of CRC.

3.0 Materials and methods

3.1 Materials

HPLC grade solvents were purchased from Rathburn Chemicals Ltd (Walkerburn, Scotland, UK) and UHPLC grade solvents were purchased from Fisher Scientific UK (Loughborough, UK). Strata X (HLB sorbent) 60 mg/3ml SPE 96-well plates were purchased from Phenomenex (Macclesfield, U.K). A mix of deuterated compounds were used as internal standards (IS); 17β-estradiol 2,4,16,16-d4 sodium 3-sulfate (E2-d4-S, >99% D atom), carbamazepine (ring-d10), and diclofenac (phenyl-d4) were purchased from Cambridge Isotope Laboratories Inc. (MA, USA). Progesterone-2,2,4,6,6,17R,21,21,21-d9 (P-d9, 98% D atom) was purchased from CDN isotopes (Quebec, Canada). All other standards and reagent chemicals were purchased from Sigma-Aldrich Company Ltd., Dorset, U.K.

3.2 Sample collection

In total 7 CRC, 26 healthy and 25 adenomatous polyp faecal samples were collected shortly following diagnosis and prior to the initiation of any treatment to ensure any differences detected between groups were not attributed to differences in treatment. Samples were collected at the Royal Sussex County Hospital, Brighton, UK and the University Hospital Motol, Prague, Czech Republic. All CRC samples came from one hospital (UK) and, with the exception of one sample, were from male subjects. Eligible subjects gave written formal consent for collection and use of their faecal samples. Ethical approval for the study was awarded by the South East Coast – Brighton and Sussex NRES Committee (REC Reference number 10/H1107/39) and the Ethics Committee for Multi-Centric Clinical Trials of the University Hospital Motol (Reference number EK-555/12).

No patients were taking antibiotics, probiotics, or pre-biotics at the time of sampling or in the 3 months preceding sample collection. Patients were matched for age, weight and body mass index (BMI) and no significant differences were observed for these parameters between the disease groups when analysed according to country of origin or gender status. However, for all subjects the average height was significantly shorter ($p<0.05$) in healthy compared with polyp groups (Table 1).

3.3 Sample preparation

Whole stool samples were homogenised and frozen at -80 C. Faecal samples were extracted according to a modified methanol extraction method, [34] followed by SPE to remove salts and concentrate the sample prior to nUHPLC-nESI-TOFMS analysis. From each faecal sample, a 10 g sub sample was taken and homogenised in 20 mL phosphate buffer (PBS) (Ultra Turrax, IKA, Oxford, UK). PBS was used as the initial extraction buffer as an aliquot of the samples was also collected for analysis of the gut microbiome (to be published). To a 400 mg sub sample of this homogenate, 1.4 mL PBS and 3.5 µL of 0.1 mg/mL IS mix were added and the sample vortexed then centrifuged at 10,000 g for 10 minutes. The supernatant was collected and the pellet extracted twice more with methanol (1.4ml). From each extract, 50µL was taken and pooled to produce 150 µL (1 part PBS: 2 parts MeOH) of final extract for each faecal sample which was diluted with 1.9 mL 0.1% formic acid for HLB SPE extraction. Strata X 96-well plates were conditioned with 1 mL methanol and washed with 1 mL water. The acidified samples were loaded and the SPE washed with 4 mL of water and the metabolites eluted with 0.5 mL methanol and 0.5 mL ethyl acetate. The solvent was removed under vacuum and the samples reconstituted in 20 µL 70:30 H2O: MeOH to give a 7.5-fold concentration of the original 150 µL faecal extract. Samples were injected on nUHPLC-nESI-TOFMS immediately following sample preparation. All extraction solvents used were HPLC grade >99% purity.

3.4 nUHPLC-nESI-TOFMS analysis

Extracts (0.5 µL, comprising 175 µg equivalent of faeces) were injected onto a Waters nanoAcquity UHPLC and separated using a Waters nanoAcquity HSS-T3 (100mm x 100 µm x 1.8 µm, 100 Å) column. Chromatographic separation was carried out at 700 nL/min using UHPLC grade (>99.99% purity) water and acetonitrile as mobile phases A and B respectively, both modified with 0.01% formic acid. A gradient elution was used: 0 mins 10% B, 4 mins 30% B, 18 mins 50% B, 30 mins 100% B, 100% B maintained for 10 minutes then equilibrated in initial conditions for a further 15 minutes. Metabolites were detected in positive and negative nESI modes using a Waters Xevo G2 TOFMS tuned to a mass resolution of 20,000 and equipped with a nano ESI source using homemade pulled fused silica emitters.[26] A sample from each group and a QC sample was also analysed using MS^E for collisional- induced dissociation with an energy ramp of 10-40 eV to obtain fragment information for identification of metabolite structures.

3.5 Quality control

Quality control (QC) samples were produced by pooling 5 µL of each faecal extract from all subjects to create a composite mixture. Prior to analysis in both ionisation modes, 5 QC samples were injected to condition the nESI source. Injections of the QC sample were carried out every 10 injections throughout the analysis to determine instrumental drift in intensity response and retention time data. The peak area variability of the QC samples was determined where areas of all peaks present in at least 80% of the QC samples were integrated and the % coefficient of variation (CV) of the mean peak areas were calculated The % of peaks where the CV of mean peak area was below recommended levels of 15% and 30% for metabolomics data was calculated [37, 38].

The repeatability of the peak retention time of the nano platform was determined as the % CV of the mean retention time for the 100 most abundant peaks (50 in each ionisation mode) in all of the samples and QCs analysed. Previous reports suggest that a target % CV should be below 2%.[39]

3.6 Statistical analysis

The MS datasets were deisotoped, mass spectral peaks deconvoluted, and the datasets binned and normalized to the total spectral area using Waters MarkerLynx software and which is a method previously shown to be applicable to nanospray datasets [40]. After exclusion of outliers, the variation (CV) of the total spectral area used in normalisation of datasets was 11%. Binned datasets were exported to Simca v13.0 software (Umetrics Ltd, Crewe, UK) for multivariate analyses. All data were log transformed and Pareto scaled prior to principal components analysis (PCA). Due to the small sample sizes in this study, it was not possible to fully validate supervised analyses such as partial least-squares discriminant analysis or orthogonal partial least-squares discriminant analysis (OPLS-DA). However, OPLS-DA models were used to identify metabolites that may potentially drive discrimination between sample groups. These metabolites were detected using an 'S' plot of the orthogonal partial least-squares discriminant analysis (OPLS-DA) model which is a plot of reliability (correlation) of the loading variables versus their covariance (contribution to the model). [41]

Further statistical analysis was carried out using GraphPad Prism version 6.05 for Windows (GraphPad Software, San Diego California USA, www.graphpad.com). The normalised mass intensities for discriminatory metabolites highlighted by S-plot analysis were tested for normal distribution (D'Agostino's K squared test) and were then tested for significance at $P<0.05$ using a one-way ANOVA with a Holm-Sidak multiple comparison test to determine p-values. Non-normally distributed data were tested using the Kruskal-Wallis with a Dunn's multiple comparison test to determine p-values. Data was also tested for homoscedasticity using the Levenes test. Sample groups with unequal variances were analysed using the Brown-Forsythe ANOVA followed by the Games-Howell post hoc test to analyse for significant differences between sample groups. Statistically significant metabolites

were tested for their predictability using the Receiver-Operator Characteristic (ROC) to plot the false positive rate against the true positive rate and the area under the curve was calculated to give the probability of a positive result being a true positive.

Metabolite identities were determined from their accurate mass, isotopic fit, and comparison of fragmentation data with *in-silico* fragmentation from Waters Mass Fragment software, authentic standards or with Metlin [42], Human Metabolome Database [43] and MycompoundID [44] databases.

4.0 Results and discussion

4.1 Repeatability of the metabolomic analyses

The repeatability of the metabolome peak area was calculated by determining the percentage of peaks common to 80% of the QC samples which had a mean peak area CV of < 15% and <30% [37]. In this study the mean area of 85.5% (-nESI mode) and 89.5% (+nESI mode) of peaks had a CV <30%, whereas 61.6% (-nESI) and 60.7% (+nESI) of peaks gave a mean peak area CV of <15%. This data indicated that the repeatability of the analysis of the metabolome peak area was reliable.

To assess retention time stability, the % CV for the 50 most abundant metabolites in each ionisation mode was determined using data from all samples and the QCs analysed. The % CV for median retention time was between 0.16 and 1.99%, i.e. between 1.9 and 23.5 seconds, and this is within the 2% guideline proposed by Theodoridis *et al* [39] (see Figure S1). In total 93% of the 100 peaks investigated had a % CV of <1.5 for retention time stability with 70 of these peaks returning a CV <1%. The peaks with the greatest variability tended to be sulphated bile acids or tetrapyrroles which were present at very high levels, and typically with broader peaks and varying peak apex retention time compared to the other metabolites investigated. The repeatability data in this study compare well with previous nUHPLC-nESI-TOFMS analyses of biofluid metabolomes, indicating that the nanospray platform was acceptable for profiling faecal extracts.[26, 27, 29]

4.2 Profiles of metabolites in faecal extracts

In total, 5387 metabolite signals were detected in +nESI mode and 4851 signals in -nESI mode and of these 2256 and 3032 features respectively were putatively annotated using PutMedID [45]. This represents a 2-fold increase in metabolite signals that had been reported in previous profiling studies of the faecal metabolome using conventional uHPLC-TOF-MS platforms [10, 46].

Representative base peak intensity (BPI) chromatograms resulting from nUHPLC-nESI-TOFMS analyses of extracts of QC samples obtained from a composite of faecal extracts from all study subjects are given in Figure 1. Details of structural identification of xenobiotics and metabolites are given in Tables S1 and S2. In +ESI mode, highly polar metabolites including amino acids and dipeptides were detected in the unretained section of the chromatogram, eluting between 5.2 and 5.8 mins.

Tetrapyrrole metabolites, bile acids and lipid metabolites, including sphingosine, pharmaceuticals and other xenobiotics also dominated the metabolomic profile. The tetrapyrrole metabolites include urobilins and urobilinogens which are the endpoints of haemoglobin and bilirubin metabolism. [9, 12] As >90% of the bilirubin is metabolised and excreted in the faeces, then this accounts for the very large peak observed for these compounds in the BPI chromatogram. [47] Analyses of selected ion chromatograms for the different tetrapyrroles revealed the co-elution of at least four metabolites which were identified from high resolution molecular formula and published fragmentation details [48, 49] (Figure S2 and Table S2). These were D-Urobilinogen, formed from the deconjugation of bilirubin diglucuronide and the downstream metabolites I-urobilinogen, L-urobilin and I-urobilin/D-urobilinogen.

Major peaks corresponding to two protease inhibitors, darunavir and ritonavir, were present in +nESI BPI profiles. These inhibitors are typically prescribed to patients infected with the human immunodeficiency virus (HIV) (Table S1). A total of 20 pharmaceuticals or their metabolites were detected in the sample extracts as well as dietary metabolites such as solanidine and piperine. The detection of a range of dietary and pharmaceutical metabolites in the metabolome highlight the need for detailed knowledge of patients in terms of diet and pharmaceutical intake and the possibility for patient stratification for metabolomic studies.

In -nESI mode the most prevalent class of metabolites detected were free and conjugated bile acids (Figure 1). Bile acids have long been thought to play a role in colorectal diseases such as CRC, especially the secondary bile acids formed as a result of C7 dehydroxylation by the intestinal microflora. [50-52] However, the role of bile acids in the initiation of cancer remains to be clearly established, although their signalling properties via interactions with nuclear receptors are a possible mechanism [52]. Profiles of bile acids in the QC samples were further investigated using selected ion chromatograms to determine the overall coverage of bile acid metabolism (Figure 2, Table S2). A total of 32 bile acids were detected of which 18 identities were determined using genuine standards. These included the primary bile acids cholic acid, chenodeoxycholic acid, taurocholic acid, taurochenodeoxycholic acid, glycocholic acid and glycochenodeoxycholic acid in addition to the major secondary bile acids lithocholic acid and deoxycholic acid. The detection of all the major primary and secondary bile acids, in addition to the wide array of other less abundant bile acid metabolites is promising for studying the role of the metabolome in bowel diseases and the composition of the bacterial microbiome responsible for C7 dehydroxylation and deconjugation of taurine or glycine groups [53, 54].

The importance of sex steroids in diseases of the gut is well established, and changes in circulating levels of androgens, which act via nuclear receptors, have been linked to CRC [55, 56],

Crohn's disease[57], ulcerative colitis[57] and irritable bowel syndrome.[58] Hence, the ability to detect faecal steroids during global metabolomics profiling studies may further help to elucidate their function in such conditions and understand the role of the gut microbiome in sex steroid metabolism. [59] Nine conjugated androgens were detected in the faecal metabolome, (Figure 3) as well as conjugated progestogens, pregnenolone and estrogens (Figure S3). Free steroids tend to be excreted via urine, and the only unconjugated steroids detected in faecal extracts were those of progesterone and its hydroxylated metabolite, hydroxyprogesterone, together with two other unidentified progesterone-like steroids (Figure S4).

The use of highly sensitive nUHPLC-nESI-TOFMS has enabled the detection of eicosanoids, small molecule signalling mediators of inflammation, in untargeted metabolic analyses. [27] Eicosanoids such as prostaglandin E2 (PGE2) are inflammatory markers of cancer and non-steroidal anti-inflammatory drugs (NSAIDs) are thought to reduce the risk of some cancers, such as CRC, by inhibiting the cyclooxygenase enzymes responsible for eicosanoid production. [60] In this study, PGE2 was detected in the faecal metabolome alongside a number of other unidentified eicosanoid peaks (Figure S5). Our studies reveal that low abundance signalling compounds such as prostaglandins, progesterones, and sex steroids were detected during profiling of the faecal metabolome using the nanospray platform. These metabolites are normally present at sub nM concentrations in blood or faeces [61] and were not detected in previous conventional LC-MS profiling studies of the faecal metabolome [62, 63] and have only previously been detected in faecal extracts using targeted LC-MS analyses [64]. Low abundance signalling metabolites in the faecal metabolome would not have been detected using other analytical platforms such as NMR. However it should be recognised that no one analytical platform can analyse the full complement of the metabolome., and that semi- volatile or polar metabolites present in faecal extracts are not retained by LC columns and are usually analysed by GC-MS [36].

4.3 nUHPLC-nESI-TOFMS profiling of faecal extracts from disease groups

A pilot study comprising samples of faecal extracts from 7 CRC, 25 polyp and 26 healthy patient groups were profiled by nUHPLC-nESI-TOFMS in both +nESI and -nESI modes and the datasets analysed by multivariate modelling. Principal component score plots were constructed for the datasets (Figures S6 and S7). Tight clustering of the QC samples in both ionisation modes further indicated that the metabolomic analyses were highly repeatable. However, using unsupervised analyses no discrimination between sample sources (Prague versus UK) or between disease groups was apparent, possibly due to the high amount of inter-individual variation in the faecal metabolome. Due to the small sample groups in this work it was not possible to fully validate supervised analysis such as PLS-DA or OPLS-DA, however OPLS-DA models were used to identify metabolites that may

potentially drive discrimination between sample groups. Analysis of the S-plots of the loading variables of models tentatively revealed ten metabolites driving discrimination between the sample groups. Further univariate and ROC analysis of these metabolites highlighted in the S-plots were used to determine significant differences in concentrations between sample groups (Table 2). Box plots of these putative discriminating metabolites were also constructed (Supplementary Figure S8), and demonstrate the significant biological variation in the faecal metabolome. Discriminating metabolites included a putative porphyrin metabolite which was detected at >40-fold higher concentrations in the CRC group compared with healthy or polyp groups. Haem metabolites are known to be markers of CRC disease [10] and the identity of this metabolite warrants further investigation as all of the patients in this study were referred for colonoscopy due to a positive result in a faecal occult blood test of which only seven were diagnosed with CRC. In addition, concentrations of a potential cholesterol oxidation product, two other unidentified nitrogen containing lipids, a phospholipid and a phytyl phosphate type metabolite were significantly higher in CRC compared to healthy and polyp groups.

In a recent metabolomics study, both 7 and 12-ketodeoxycholic acid have been observed to be up-regulated in CRC tissue[15]. Similarly in our pilot study, concentrations of two unidentified keto bile acid structures were significantly higher in the CRC patient group. A comparison with the retention time of standards revealed that these secondary bile acids were not 7 and 12-ketodeoxycholic acids or other known keto bile acids associated with the gut metabolome (i.e. 3-oxocholic acid, 3,7 dihydroxy-12-oxocholic acid or 7-ketolithocholic acid). The identity and role of low abundant keto bile acids in bowel diseases warrants further study. A further analysis of the peak areas of the major bile acids detected in our study was undertaken and revealed that there were no significant differences in the levels of the primary bile acids and other secondary bile acids between CRC and either the healthy or polyp sample groups (data not shown).

A 3.1 fold increase in abundance of trace levels of a dinor prostaglandin metabolite was observed in the CRC group which may be a reflection of inflammatory processes. PGE2 and thromboxane B2 have been found to be increased in tumours from CRC patients.[60, 65, 66] It is possible that a larger study with more CRC samples may uncover additional eicosanoids that have become disrupted as a result of cancer and polyp development. This finding warrants further investigation due to the potential benefits of NSAID consumption in reducing CRC and polyp incidence.

This metabolomics study also revealed that concentrations of an unidentified androgen sulfate metabolite were significantly reduced in samples from CRC subjects compared to the healthy group. Reduced androgen levels have previously been reported in male CRC patients possibly as a result of reduced circulating levels of the sex hormone binding globulin in these subjects.[55, 56]

The 10 metabolites associated with the CRC group were re-analysed by univariate statistics following the exclusion of all patients from Prague and female subjects to eliminate and bias introduced by hospital and gender. Concentrations of all the metabolites, with the exception of the androgen sulfate, were still significantly different between healthy and CRC groups. However, concentrations of metabolites of m/z 381.3117, 443.2875 and the octadecyl lysophosphatidic acid like metabolite were no longer significant between CRC and polyp groups.

The discriminating metabolites detected in this pilot study require validation in a larger follow-up study, with tight dietary controls to remove inter-individual variability. Furthermore, although some of the metabolites (e.g. the prostaglandin and androgen sulfate) that were detected in this pilot study were not obviously dietary in source, it is possible diet may influence the CRC-associated metabolome and thus a diet-controlled study would further validate any future findings.

5.0 Conclusions

This study was the first to incorporate nanoscale UHPLC-TOFMS in a faecal metabolomic study and, to our knowledge, the first to use this technology to provide an overview of selected low and high abundance compounds represented in the faecal metabolome. Metabolite profiling of faecal extracts revealed the presence of trace levels of eicosanoid and sex steroid signalling compounds alongside compounds of high abundance including xenobiotics, tetrapyrrole and the major bile acid metabolites. Profiling of faecal extracts using the nUHPLC-nESI-TOFMS platform was highly repeatable. A small pilot study revealed that signalling metabolites as well other key metabolic pathways are potentially implicated in the CRC sample group. However, due to the small scale of this work and the variability associated with dietary influences, more studies are required with larger sample sets to improve statistical power to validate and uncover discriminating metabolites for CRC disease. In future larger studies it would also be prudent to standardise the patient's diet prior to sampling in order to control for variation in metabolite profiles due to food. In addition, normalising datasets to the water content of the sample may also improve the analytical method [6]. Nevertheless, our study reveals the potential of the nanoscale platform to be used in larger scale studies to identify the role of the metabolome in the onset of colorectal diseases.

Acknowledgements:

This study was supported by funding from the Medical Research Council (G0901553). We are grateful to Mr Jeremy Clark and staff at the Clinical Investigations Research Unit and the Sussex Cancer Centre, Royal Sussex County Hospital, for support in collection of stool samples.

References

[1] S. Bouatra, F. Aziat, R. Mandal, A.C. Guo, M.R. Wilson, C. Knox, T.C. Bjorndahl, R. Krishnamurthy, F. Saleem, P. Liu, Z.T. Dame, J. Poelzer, J. Huynh, F.S. Yallou, N. Psychogios, E. Dong, R. Bogumil, C. Roehring, D.S. Wishart, The human urine metabolome, Plos One, 8 (2013) e73076.
[2] N. Psychogios, D.D. Hau, J. Peng, A.C. Guo, R. Mandal, S. Bouatra, I. Sinelnikov, R. Krishnamurthy, R. Eisner, B. Gautam, N. Young, J. Xia, C. Knox, E. Dong, P. Huang, Z. Hollander, T.L. Pedersen, S.R. Smith, F. Bamforth, R. Greiner, B. McManus, J.W. Newman, T. Goodfriend, D.S. Wishart, The Human Serum Metabolome, Plos One, 6 (2011).
[3] D.S. Wishart, M.J. Lewis, J.A. Morrissey, M.D. Flegel, K. Jeroncic, Y. Xiong, D. Cheng, R. Eisner, B. Gautam, D. Tzur, S. Sawhney, F. Bamforth, R. Greiner, L. Li, The human cerebrospinal fluid metabolome, J Chromatogr B Analyt Technol Biomed Life Sci, 871 (2008) 164-173.
[4] Q. Wang, P. Gao, X. Wang, Y. Duan, The early diagnosis and monitoring of squamous cell carcinoma via saliva metabolomics, Sci. Rep., 4 (2014).
[5] J.K. Nicholson, E. Holmes, J. Kinross, R. Burcelin, G. Gibson, W. Jia, S. Pettersson, Host-Gut Microbiota Metabolic Interactions, Science, 336 (2012) 1262-1267.
[6] O. Deda, A.C. Chatziioannou, S. Fasoula, D. Palachanis, N. Raikos, G.A. Theodoridis, H.G. Gika, Sample preparation optimization in fecal metabolic profiling, J. Chromatogr. B Analyt. Technol. Biomed. Life Sci., 1047 (2017) 115-123.
[7] J. Bjerrum, Y. Wang, F. Hao, M. Coskun, C. Ludwig, U. Günther, O. Nielsen, Metabonomics of human fecal extracts characterize ulcerative colitis, Crohn's disease and healthy individuals, Metabolomics, 11 (2015) 122-133.
[8] G. Le Gall, S.O. Noor, K. Ridgway, L. Scovell, C. Jamieson, I.T. Johnson, I.J. Colquhoun, E.K. Kemsley, A. Narbad, Metabolomics of fecal extracts detects altered metabolic activity of gut microbiota in ulcerative colitis and irritable bowel syndrome, J Proteome Res, 10 (2011) 4208-4218.
[9] H.-j. Huang, A.-y. Zhang, H.-c. Cao, H.-f. Lu, B.-h. Wang, Q. Xie, W. Xu, L.-J. Li, Metabolomic analyses of faeces reveals malabsorption in cirrhotic patients, Dig Liver Dis, 45 (2013) 677-682.
[10] J.J. Goedert, J.N. Sampson, S.C. Moore, Q. Xiao, X.Q. Xiong, R.B. Hayes, J. Ahn, J.X. Shi, R. Sinha, Fecal metabolomics: assay performance and association with colorectal cancer, Carcinogenesis, 35 (2014) 2089-2096.
[11] T.L. Weir, D.K. Manter, A.M. Sheflin, B.A. Barnett, A.L. Heuberger, E.P. Ryan, Stool Microbiome and Metabolome Differences between Colorectal Cancer Patients and Healthy Adults, PLoS One, 8 (2013).
[12] A. Jimenez-Giron, C. Ibanez, A. Cifuentes, C. Simo, I. Munoz-Gonzalez, P.J. Martin-Alvarez, B. Bartolome, M.V. Moreno-Arribas, Faecal Metabolomic Fingerprint after Moderate Consumption of Red Wine by Healthy Subjects, J Proteome Res, 14 (2015) 897-905.
[13] J.M. Ridlon, D.J. Kang, P.B. Hylemon, Bile salt biotransformations by human intestinal bacteria, J. Lipid Res., 47 (2006) 241-259.
[14] R.L. Siegel, K.D. Miller, S.A. Fedewa, D.J. Ahnen, R.G.S. Meester, A. Barzi, A. Jemal, Colorectal cancer statistics, 2017, CA-Cancer J. Clin., 67 (2017) 177-193.
[15] M.D. Williams, X. Zhang, J.J. Park, W.F. Siems, D.R. Gang, L.M.S. Resar, R. Reeves, H.H. Hill, Characterizing metabolic changes in human colorectal cancer, Anal. Bioanal. Chem, 407 (2015) 4581-4595.
[16] M. Mal, P.K. Koh, P.Y. Cheah, E.C.Y. Chan, Metabotyping of human colorectal cancer using two-dimensional gas chromatography mass spectrometry, Anal. Bioanal. Chem, 403 (2012) 483-493.
[17] Y.P. Qiu, G.X. Cai, M.M. Su, T.L. Chen, Y.M. Liu, Y. Xu, Y. Ni, A.H. Zhao, S.J. Cai, L.X. Xu, W. Jia, Urinary Metabonomic Study on Colorectal Cancer, J Proteome Res, 9 (2010) 1627-1634.
[18] F. Farshidfar, A.M. Weljie, K. Kopciuk, W.D. Buie, A. MacLean, E. Dixon, F.R. Sutherland, A. Molckovsky, H.J. Vogel, O.F. Bathe, Serum metabolomic profile as a means to distinguish stage of colorectal cancer, Genome Med, 4 (2012) 42.

[19] S. Nishiumi, T. Kobayashi, A. Ikeda, T. Yoshie, M. Kibi, Y. Izumi, T. Okuno, N. Hayashi, S. Kawano, T. Takenawa, T. Azuma, M. Yoshida, A Novel Serum Metabolomics-Based Diagnostic Approach for Colorectal Cancer, Plos One, 7 (2012) e40459.
[20] Y. Lin, C.C. Ma, C.K. Liu, Z.N. Wang, J.R. Yang, X.M. Liu, Z.W. Shen, R.H. Wu, NMR-based fecal metabolomics fingerprinting as predictors of earlier diagnosis in patients with colorectal cancer, Oncotarget, 7 (2016) 29454-29464.
[21] M.D. Williams, R. Reeves, L.S. Resar, H.H. Hill, Metabolomics of colorectal cancer: past and current analytical platforms, Anal. Bioanal. Chem, 405 (2013) 5013-5030.
[22] P. Louis, G.L. Hold, H.J. Flint, The gut microbiota, bacterial metabolites and colorectal cancer, Nat. Rev. Microbiol., 12 (2014) 661-672.
[23] M.H.K. Roshan, A. Tambo, N.P. Pace, The role of testosterone in colorectal carcinoma: pathomechanisms and open questions, Epma Journal, 7 (2016) 10.
[24] A. Lasry, A. Zinger, Y. Ben-Neriah, Inflammatory networks underlying colorectal cancer, Nat. Immunol., 17 (2016) 230-240.
[25] A.J. Chetwynd, A. David, A review of nanoscale LC-ESI for metabolomics and its potential to enhance the metabolome coverage, Talanta, 182 (2018) 380-390.
[26] A.J. Chetwynd, A. David, A. Abdul-Sada, E.M. Hill, Evaluation of analytical performance and reliability of direct nanoLC-nanoESI-high resolution mass spectrometry for profiling the (xeno)metabolome, J. Mass Spec, 49 (2014) 1063-1069.
[27] A.J. Chetwynd, A. Abdul-Sada, E.M. Hill, Solid-Phase Extraction and Nanoflow Liquid Chromatography-Nanoelectrospray Ionization Mass Spectrometry for Improved Global Urine Metabolomics, Anal. Chem., 87 (2015) 1158-1165.
[28] A.J. Chetwynd, A. Samarawickrama, J.H. Vera, S.A. Bremner, A. Abdul-Sada, Y. Gilleece, S.G. Holt, E.M. Hill, Nanoflow-Nanospray Mass Spectrometry Metabolomics Reveals Disruption of the Urinary Metabolite Profiles of HIV-Positive Patients on Combination Antiretroviral Therapy, Jaids, 74 (2017) E45-E53.
[29] A. David, A. Abdul-Sada, A. Lange, C.R. Tyler, E.M. Hill, A new approach for plasma (xeno)metabolomics based on solid-phase extraction and nanoflow liquid chromatography-nanoelectrospray ionisation mass spectrometry, J. Chromatogr. A, 1365 (2014) 72-85.
[30] A. David, A. Lange, A. Abdul-Sada, C.R. Tyler, E.M. Hill, Disruption of the Prostaglandin Metabolome and Characterization of the Pharmaceutical Exposome in Fish Exposed to Wastewater Treatment Works Effluent As Revealed by Nanoflow-Nanospray Mass Spectrometry-Based Metabolomics, Environ. Sci. Technol., 51 (2017) 616-624.
[31] D.R. Jones, Z.P. Wu, D. Chauhan, K.C. Anderson, J.M. Peng, A Nano Ultra-Performance Liquid Chromatography-High Resolution Mass Spectrometry Approach for Global Metabolomic Profiling and Case Study on Drug-Resistant Multiple Myeloma, Anal. Chem., 86 (2014) 3667-3675.
[32] M. Calderon-Santiago, F. Priego-Capote, M.D.L. de Castro, Enhancing detection coverage in untargeted metabolomics analysis by solid-phase extraction on-line coupled to LC-MS/MS, Electrophoresis, 36 (2015) 2179-2187.
[33] K. Skov, N. Hadrup, J. Smedsgaard, H. Frandsen, LC–MS analysis of the plasma metabolome—A novel sample preparation strategy, J Chromatogr B Analyt Technol Biomed Life Sci, 978–979 (2015) 83-88.
[34] O. Deda, H.G. Gika, I.D. Wilson, G.A. Theodoridis, An overview of fecal sample preparation for global metabolic profiling, J Pharm Biomed Anal (2015).
[35] J. Jansson, B. Willing, M. Lucio, A. Fekete, J. Dicksved, J. Halfvarson, C. Tysk, P. Schmitt-Kopplin, Metabolomics Reveals Metabolic Biomarkers of Crohn's Disease, PLoS One, 4 (2009) 10.
[36] N. Karu, L. Deng, M. Slae, A.C. Guo, T. Sajed, H. Huynh, E. Wine, D.S. Wishart, A review on human fecal metabolomics: Methods, applications and the human fecal metabolome database, Anal. Chim. Acta, 1030 (2018) 1-24.
[37] FDA, Analytical Procedures and Methods Validation for Drugs and Biologics Guidance for Industry. MD USA, https://www.fda.gov/downloads/drugs/guidances/ucm386366.pdf (2015).

[38] E.J. Want, I.D. Wilson, H. Gika, G. Theodoridis, R.S. Plumb, J. Shockcor, E. Holmes, J.K. Nicholson, Global metabolic profiling procedures for urine using UPLC-MS, Nat. Protoc, 5 (2010) 1005-1018.
[39] G.A. Theodoridis, H.G. Gika, E.J. Want, I.D. Wilson, Liquid chromatography-mass spectrometry based global metabolite profiling: A review, Anal. Chim. Acta, 711 (2012) 7-16.
[40] A.J. Chetwynd, A. Abdul-Sada, S.G. Holt, E.M. Hill, Use of a pre-analysis osmolality normalisation method to correct for variable. urine concentrations and for improved metabolomic analyses, J. Chromatogr. A, 1431 (2016) 103-110.
[41] S. Wiklund, E. Johansson, L. Sjostrom, E.J. Mellerowicz, U. Edlund, J.P. Shockcor, J. Gottfries, T. Moritz, J. Trygg, Visualization of GC/TOF-MS-based metabolomics data for identification of biochemically interesting compounds using OPLS class models, Anal. Chem., 80 (2008) 115-122.
[42] R. Tautenhahn, K. Cho, W. Uritboonthai, Z.J. Zhu, G.J. Patti, G. Siuzdak, An accelerated workflow for untargeted metabolomics using the METLIN database, Nat. Biotechnol., 30 (2012) 826-828.
[43] D.S. Wishart, T. Jewison, A.C. Guo, M. Wilson, C. Knox, Y.F. Liu, Y. Djoumbou, R. Mandal, F. Aziat, E. Dong, S. Bouatra, I. Sinelnikov, D. Arndt, J.G. Xia, P. Liu, F. Yallou, T. Bjorndahl, R. Perez-Pineiro, R. Eisner, F. Allen, V. Neveu, R. Greiner, A. Scalbert, HMDB 3.0-The Human Metabolome Database in 2013, Nucleic Acids Res, 41 (2013) D801-D807.
[44] L. Li, R.H. Li, J.J. Zhou, A. Zuniga, A.E. Stanislaus, Y.M. Wu, T. Huan, J.M. Zheng, Y. Shi, D.S. Wishart, G.H. Lin, MyCompoundID: Using an Evidence-Based Metabolome Library for Metabolite Identification, Anal. Chem., 85 (2013) 3401-3408.
[45] M. Brown, D.C. Wedge, R. Goodacre, D.B. Kell, P.N. Baker, L.C. Kenny, M.A. Mamas, L. Neyses, W.B. Dunn, Automated workflows for accurate mass-based putative metabolite identification in LC/MS-derived metabolomic datasets, Bioinformatics, 27 (2011) 1108-1112.
[46] R. Girlanda, A.K. Cheema, P. Kaur, Y. Kwon, A. Li, J. Guerra, C.S. Matsumoto, M. Zasloff, T.M. Fishbein, Metabolomics of Human Intestinal Transplant Rejection, Am. J. Transplant., 12 (2012) S18-S26.
[47] P.D. Berk, R.B. Howe, J.R. Bloomer, N.I. Berlin, Studies of bilirubin kinetics in normal adults, J. Clin. Investig., 48 (1969) 2176-2190.
[48] A. Walker, B. Pfitzner, S. Neschen, M. Kahle, M. Harir, M. Lucio, F. Moritz, D. Tziotis, M. Witting, M. Rothballer, M. Engel, M. Schmid, D. Endesfelder, M. Klingenspor, T. Rattei, W.z. Castell, M.H. de Angelis, A. Hartmann, P. Schmitt-Kopplin, Distinct signatures of host-microbial meta-metabolome and gut microbiome in two C57BL/6 strains under high-fat diet, ISME J., 8 (2014) 2380-2396.
[49] K.D. Quinn, N.Q.T. Nguyen, M.M. Wach, T.D. Wood, Tandem mass spectrometry of bilin tetrapyrroles by electrospray ionization and collision-induced dissociation, Rapid Commun. Mass Spectrom, 26 (2012) 1767-1775.
[50] H. Ajouz, D. Mukherji, A. Shamseddine, Secondary bile acids: an underrecognized cause of colon cancer, World J Surg Oncol 12 (2014) 164.
[51] F.M. Nagengast, M.J. Grubben, I.P. van Munster, Role of bile acids in colorectal carcinogenesis, Eur J Cancer, (1995) 1067-1070.
[52] C. Degirolamo, S. Modica, G. Palasciano, A. Moschetta, Bile acids and colon cancer: Solving the puzzle with nuclear receptors, Trends Mol Med, 17 (2011) 564-572.
[53] M. Baptissart, A. Vega, S. Maqdasy, F. Caira, S. Baron, J.-M.A. Lobaccaro, D.H. Volle, Bile acids: From digestion to cancers, Biochimie, 95 (2013) 504-517.
[54] S. Hirano, R. Nakama, M. Tamaki, N. Masuda, H. Oda, Isolation and characterization of thirteen intestinal microorganisms capable of 7 alpha-dehydroxylating bile acids, Appl. Environ. Microbiol., 41 (1981) 737-745.
[55] J.H. Lin, E. Giovannucci, Sex Hormones and Colorectal Cancer: What Have We Learned So Far?, J Natl Cancer Inst 102 (2010) 1746-1747.
[56] J.H. Lin, S.M. Zhang, K.M. Rexrode, J.E. Manson, A.T. Chan, K. Wu, S.S. Tworoger, S.E. Hankinson, C. Fuchs, J.M. Gaziano, J.E. Buring, E. Giovannucci, Association between Sex Hormones and Colorectal Cancer Risk in Men and Women, Clin. Gastroenterol. Hepatol, 11 (2013) 419-424.e411.

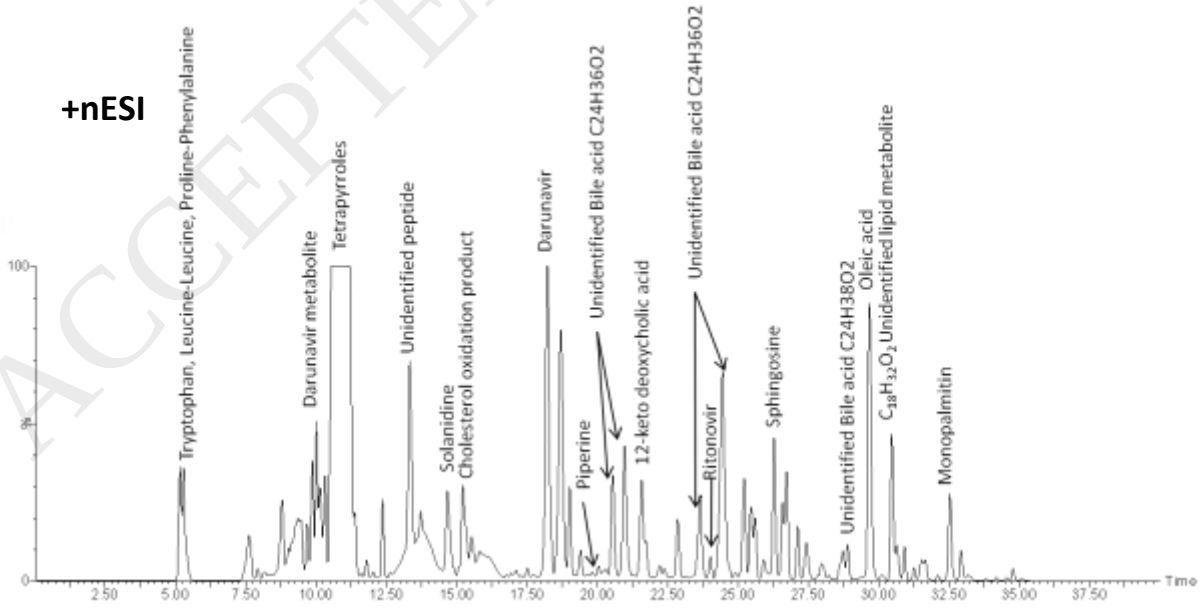
[57] H. Khalili, A.N. Ananthakrishnan, G.G. Konijeti, L.M. Higuchi, C.S. Fuchs, J.M. Richter, S.S. Tworoger, S.E. Hankinson, A.T. Chan, Endogenous levels of circulating androgens and risk of Crohn's disease and ulcerative colitis among women: a nested case-control study from the nurses' health study cohorts, Inflamm Bowel Dis, 21 (2015) 1378-1385.
[58] A. Mulak, Y. Taché, M. Larauche, Sex hormones in the modulation of irritable bowel syndrome, World J Gastroenterol., 20 (2014) 2433-2448.
[59] P. Lombardi, B. Goldin, E. Boutin, S.L. Gorbach, Matabolism of androgens and estrogens by human fecal microorganisms., J. Steroid Biochem. Mol. Biol., 9 (1978) 795-801.
[60] Y.S. Liu, H. Sun, M. Hu, Y. Zhang, S.L. Chen, S. Tighe, Y.T. Zhu, The Role of Cyclooxygenase-2 in Colorectal Carcinogenesis, Clin. Colorectal Canc., 16 (2017) 165-172.
[61] D.S. Wishart, Y.D. Feunang, A. Marcu, A.C. Guo, K. Liang, R. Vazquez-Fresno, T. Sajed, D. Johnson, C.R. Li, N. Karu, Z. Sayeeda, E. Lo, N. Assempour, M. Berjanskii, S. Singhal, D. Arndt, Y.J. Liang, H. Badran, J. Grant, A. Serra-Cayuela, Y.F. Liu, R. Mandal, V. Neveu, A. Pon, C. Knox, M. Wilson, C. Manach, A. Scalbert, HMDB 4.0: the human metabolome database for 2018, Nucleic Acids Res., 46 (2018) D608-D617.
[62] H.C. Cao, H.J. Huang, W. Xu, D.Y. Chen, J. Yu, J. Li, L.J. Li, Fecal metabolome profiling of liver cirrhosis and hepatocellular carcinoma patients by ultra performance liquid chromatography-mass spectrometry, Anal. Chim. Acta, 691 (2011) 68-75.
[63] E. De Paepe, L. Van Meulebroek, C. Rombouts, S. Huysman, K. Verplanken, B. Lapauw, J. Wauters, L.Y. Hemeryck, L. Vanhaecke, A validated multi-matrix platform for metabolomic fingerprinting of human urine, feces and plasma using ultra-high performance liquid-chromatography coupled to hybrid orbitrap high-resolution mass spectrometry, Anal. Chim. Acta, 1033 (2018) 108-118.
[64] Y.L. Song, Q.Q. Song, J. Li, J. Zheng, C. Li, Y. Zhang, L.L. Zhang, Y. Jiang, P.F. Tu, An integrated platform for directly widely-targeted quantitative analysis of feces part II: An application for steroids, eicosanoids, and porphyrins profiling, J. Chromatogr. A, 1460 (2016) 74-83.
[65] E.R. Greene, S. Huang, C.N. Serhan, D. Panigrahy, Regulation of inflammation in cancer by eicosanoids, Prostaglandins Other Lipid Mediat, 96 (2011) 27-36.
[66] S. Tuncer, S. Banerjee, Eicosanoid pathway in colorectal cancer: Recent updates, World J. Gastroenterol., 21 (2015) 11748-11766.

Figure captions.

Figure 1: Base Peak Intensity chromatograms of faecal QC samples analysed in +nESI and -nESI modes. Equivalent of 175 µg wet weight of faeces injected onto the nUHPLC-nESI-TOFMS. -Mass spectrum data of the identified xenobiotics and metabolites are given in Supplementary Tables S1 and S2.

Figure 2: Major primary and secondary bile acids present in the faecal metabolome (- nESI mode). Selected ion chromatograms of all primary and secondary bile acids present in the faecal metabolomes, the latter being formed by the bowel microbiome via C7 hydroxylation of amine conjugates. All peaks detected using a mass window of 5PPM. Primary and secondary bile acids confirmed by comparison of retention time and fragmentation patterns with genuine standards (Supplementary Table S2). 1-4 correspond to the structure $C_{24}H_{40}O_5$ and isomers of cholic acid, 5 is glycohydodeoxycholic acid confirmed by a genuine standard. Peak 6 corresponds to the molecular formula $C_{24}H_{40}O_4$ a glycine conjugated bile acid. Peaks 7 and 8 are $C_{24}H_{40}O_4$ isomers of deoxycholic acid. Peaks 9 and 10 correspond to ursodeoxycholic acid and hyodeoxycholic acid, respectively, and peak 11 to allolithocholic acid.

Figure 3: Faecal conjugated androgens detected in negative nESI mode in a pooled quality control sample. Overlaid selected ion chromatograms of faecal conjugated androgens detected with a mass window of 5PPM, 1). Here the 383.1527/C19H28O6S peaks are dihydroxyandrosterone sulfate like metabolites, 2) Androsterone and dihydrotestosterone sulfate like with the mass/formula 369.1738 C19H30O5S, 3) The 367.1577 peak at 13.9 min corresponds to dihydroepiandrosterone as confirmed by a pure standard, 4) Androsterone and dihydrotestosterone glucuronide structures with the mass M-H 465.2485 and molecular formula C25H38O8. All peaks identified as conjugated androgens based upon accurate mass measurement and in the case of sulphates HSO3 and HSO4 fragments (Supplementary Table S2).



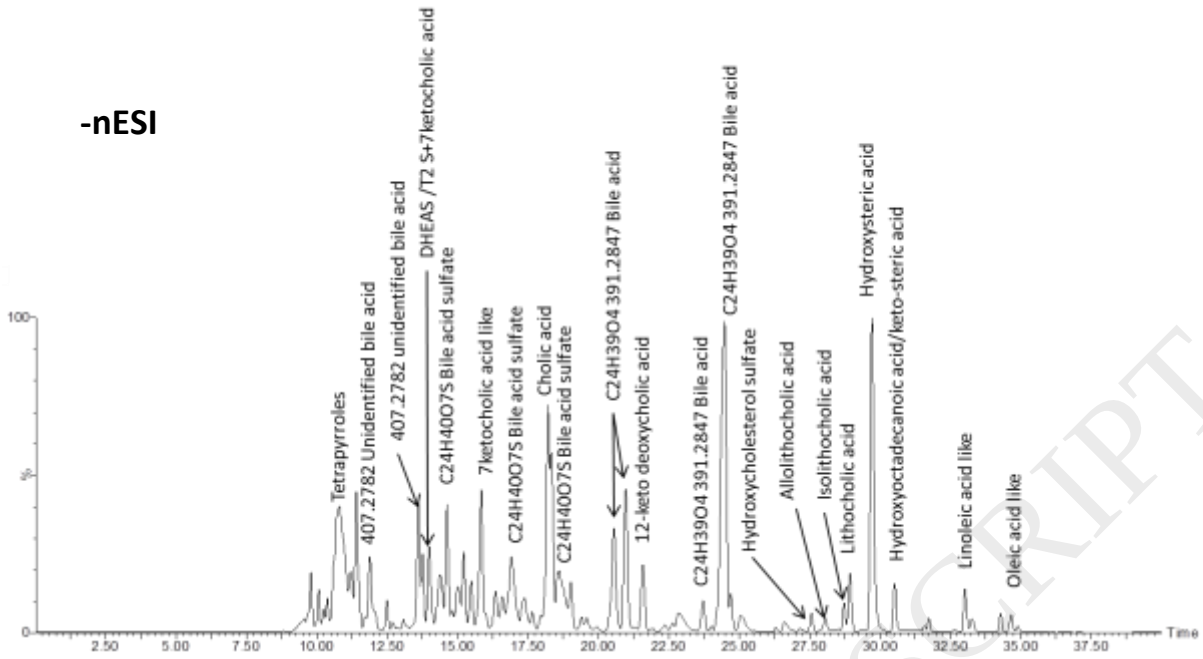


Figure 1: Base Peak Intensity chromatograms of faecal QC samples analysed in +nESI and -nESI modes

Equivalent of 175 µg wet weight of faeces injected onto the nUHPLC-nESI-TOFMS. -Mass spectrum data of the identified xenobiotics and metabolites are given in Supplementary Tables S1 and S2.

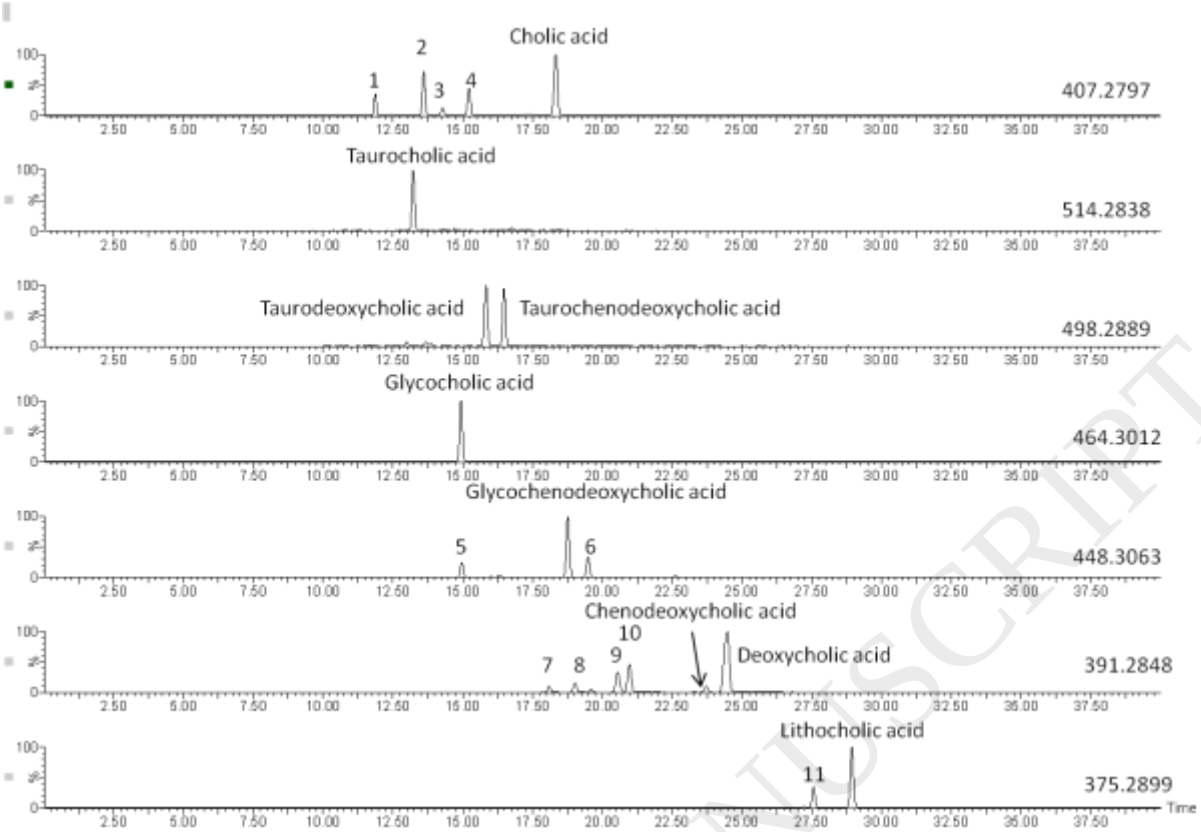


Figure 2: Major primary and secondary bile acids present in the faecal metabolome (- nESI mode)

Selected ion chromatograms of all primary and secondary bile acids present in the faecal metabolomes, the latter being formed by the bowel microbiome via C7 hydroxylation of amine conjugates. All peaks detected using a mass window of 5PPM. Primary and secondary bile acids confirmed by comparison of retention time and fragmentation patterns with genuine standards (Supplementary Table S2). 1-4 correspond to the structure $C_{24}H_{40}O_5$ and isomers of cholic acid, 5 is glycohydodeoxycholic acid confirmed by a genuine standard. Peak 6 corresponds to the molecular formula $C_{24}H_{40}O_4$ a glycine conjugated bile acid. Peaks 7 and 8 are $C_{24}H_{40}O_4$ isomers of deoxycholic acid. Peaks 9 and 10 correspond to ursodeoxycholic acid and hyodeoxycholic acid, respectively, and peak 11 to allolithocholic acid.

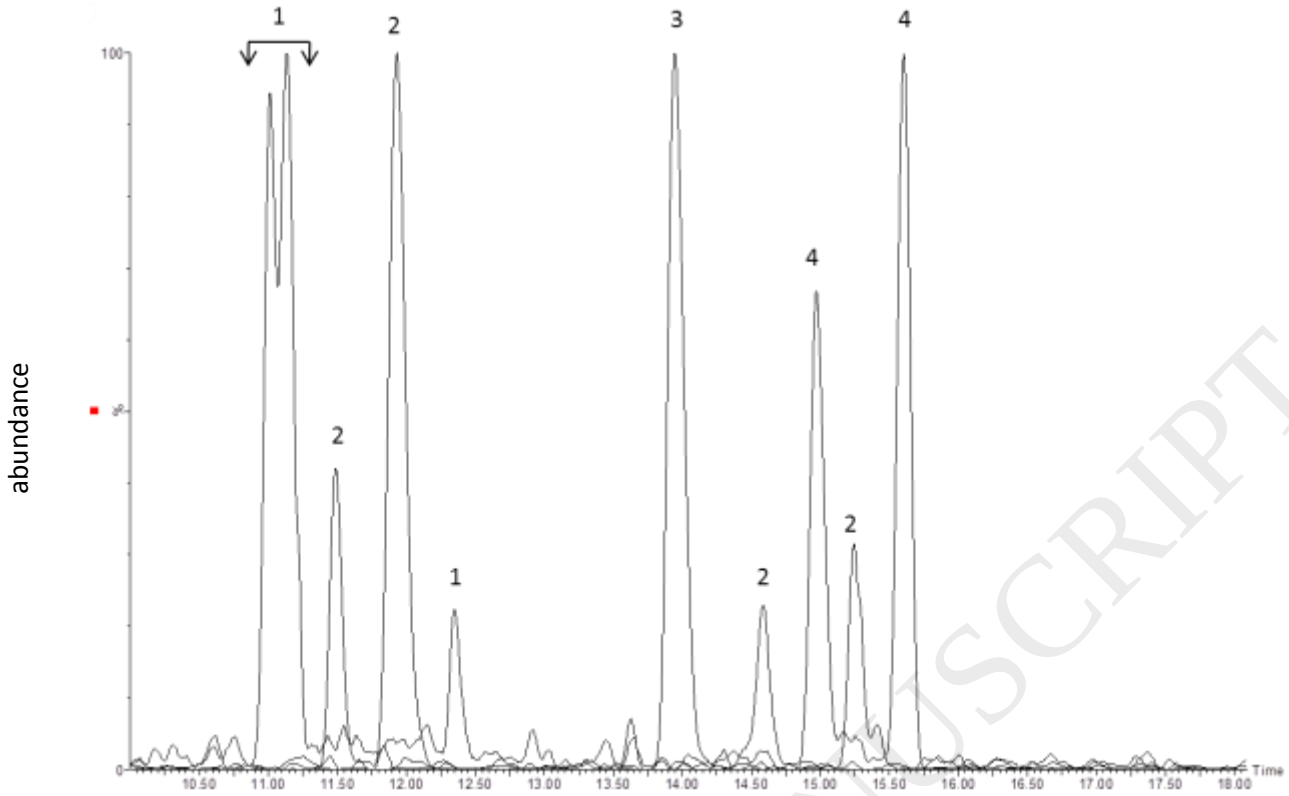


Figure 3: Faecal conjugated androgens detected in negative nESI mode in a pooled quality control sample

Overlaid selected ion chromatograms of faecal conjugated androgens detected with a mass window of 5PPM, 1). Here the 383.1527/C19H28O6S peaks are dihydroxyandrosterone sulfate like metabolites, 2) Androsterone and dihydrotestosterone sulfate like with the mass/formula 369.1738 C19H30O5S, 3) The 367.1577 peak at 13.9 min corresponds to dihydroepiandrosterone as confirmed by a pure standard, 4) Androsterone and dihydrotestosterone glucuronide structures with the mass M-H 465.2485 and molecular formula C25H38O8. All peaks identified as conjugated androgens based upon accurate mass measurement and in the case of sulphates HSO3 and HSO4 fragments (Supplementary Table S2).

Table 1: Patient sex, age and anthropometric data according to country of origin (mean ± standard deviation).

Sample group	Numbers of male subjects	Number of female subjects	Age (yrs)	Height (m)	Weight (kg)	BMI (kg/m^2)
Healthy -all subjects	**13**	**13**	**60.3±12.0**	**1.69±0.12***	**77.0±17.0**	**27.1±5.7**
Healthy -Prague only	4	2	60.5±10.9	1.71±0.11	87.3± 23.7	29.7±6.3
Healthy -UK only	9	11	61.0± 13.0	1.68±0.11	73.3± 13.5	26.1±4.5
Healthy -all males	13		61.69±9.8	1.76±0.076	87.1±13.8	28.1±4.6
Healthy- all females		13	60.2±15.1	1.60±0.085	66.1±13.0	25.7±5.4
Polyp- all subjects	**19**	**6**	**63.5±11.2**	**1.76±0.09**	**84.7±15.9**	**27.4±4.9**
Polyp -Prague only	11	2	61.1±10.0	1.76±0.09	87.3± 17.6	28.2±5.8
Polyp-UK only	8	4	66.2± 12.1	1.76±0.09	81.5± 13.8	26.4±3.6
Polyp- all males	19		65.1±10.9	1.78±0.077	90.1±13.1	28.4±4.6
Polyp- all females		6	58.8±11.4	1.68±0.087	67.8±11.4	24.2±4.6
CRC- all subjects (UK)	**6**	**1**	**64.4±7.2**	**1.76±0.04**	**81.6±26.0**	**26.0±7.4**
CRC-all males	6		65.3±7.5	1.77±0.037	87.9±21.9	27.8±6.1
CRC- female		1	59	1.70	43.7	15.1

*Height of healthy patients (all subjects) were significantly lower compared to the polyp but not CRC all subject group as assessed by the Holm-Sidak multiple comparison test. No other differences were observed for comparisons between healthy, polyp or CRC classes when subjects were grouped either according to country (both sexes together) or gender (both countries together).

Table 2: Putative identities of significantly (p<0.05) discriminating metabolites between CRC, polyp and healthy samples.

Measured Mass (relative intensity after CID)	Formula	PPM	RT	Fragments (relative intensity after CID)	Identity	Statistical measure	CRC vs Healthy	CRC vs Polyp	Healthy vs Polyp
Discriminating metabolite detected in positive nESI mode									
381.3117 (73)	$C_{22}H_{40}N_2O_3$	0.0	17.25	363.3026 (100), 359.2176 (30)	Unknown	p-value *Fold change* **AUC**	p=0.05 *5.6x higher in CRC* **0.8489**	p=0.05 *5.5x higher in CRC* **0.8333**	ns
379.2991 (59)	$C_{27}H_{38}O$	2.6	15.19	$^{+}$361.2863 (100), $^{+}$275.1763 (3), $^{+}$265.1628 (72)	Cholesterol oxidation product	p-value *Fold change* **AUC**	p=0.0014 *4.7x higher in CRC* **0.9231**	p=0.0032 *4.6x higher in CRC* **0.9226**	ns
443.2875	$C_{19}H_{43}N_2O_7P$	2.5	20.39	nd	Unknown	p-value *Fold change*	p=0.0026 *3.0x higher in CRC*	p=0.0045 *1.8x higher in CRC*	ns

						AUC	**0.8462**	**0.8214**	
						p-value	p=0.0002	p=0.0001	
689.3567	$C_{38}H_{48}N_4O_8$	2.5	8.37	nd	Porphyrin metabolite (+C_2H_4)	*Fold change*	*42.9x higher in CRC*	*55.6x higher in CRC*	ns
						AUC	0.8297	0.8333	
Discriminating metabolite detected in negative nESI mode									
						p-value	p=0.0006	p=0.0029	
423.2879 (100)	$C_{21}H_{45}O_6P$	0.7	15.1	$^+$405.2811, (22)	Octadecyl lysophosphatidic Acid like metabolite	*Fold change*	*8.8x higher in CRC*	*12.7x higher in CRC*	ns
						AUC	**0.9396**	**0.9345**	
				96.9601 (4)		p-value	p=0.040	p=0.042	
375.2662 (100)	$C_{20}H_{41}O_4P$	0.5	14.4	79.9579 (1)	Phytyl phosphate	*Fold change*	*4.6x higher in CRC*	*4.8x higher in CRC*	ns
						AUC	**0.8736**	**0.8512**	
				$^+$165.0572, (4)		p-value	p=0.0063		
405.2647 (100)	$C_{24}H_{38}O_5$	0.2	13.3	$^+$145.0680, (2)	Keto bile acid structure	*Fold change*	*2.4x higher in CRC*	ns	ns
				$^+$143.0469		**AUC**	**0.8571**		

389.2694 (100)	$C_{24}H_{38}O_4$	0.5	21.7	(2) $^{+}$371.2618, (1) $^{+}$343.2656, (1) $^{+}$325.2545 (1)	Keto bile acid structure	p-value *Fold change* **AUC**	p=0.0538 (ns) *1.8x higher in CRC* **0.7747**	ns	p=0.0123 *1.6 higher in polyp* **0.7420**
325.2012 (63)	$C_{18}H_{30}O_5$	0.9	24.7	307.187, (100) 289.195 (55) $^{+}$275.1964 (1)	Dinor prostaglandin	p-value *fold change* **AUC**	p=0.0133 *2.6x higher in CRC* **0.8187**	p=0.0035 *3.1 higher in CRC* **0.9048**	ns
369.1737 (100)	$C_{19}H_{30}O_5S$	0.3	11.7	96.9603 (3) 78.9600 (1)	Androgen sulfate	p-value *Fold change* **AUC**	p=0.0464 *5.6x higher in healthy* **0.8159**	ns	ns

AUC values refer to receiver operator characteristic area under the curve as a measure of the false discovery rate, values greater than 0.7 are typically highly reliable markers, the closer to 1 the greater the predictability of the metabolite marker. Fragments denoted with $^{+}$ represent observed fragments which agree with predicted fragments generated using Waters Mass Fragment software. Numbers in brackets refer to relative percentage intensity for fragment and parent ions after collision-induced dissociation, CID.