Annals of Surgery

DOI: 10.1097/SLA.0000000000005164

Lipidomic Profiling of Colorectal Lesions for Real-Time Tissue Recognition and Risk-Stratification using Rapid Evaporative Ionisation Mass Spectrometry

Sam E Mason MBBS[a]; Eftychios Manoli BSc[a], James L Alexander PhD[b], Liam Poynter MBBS[a], Lauren Ford PhD[a], Petra Paizs BSc[b], Afeez Adebesin BSc[a], James S McKenzie PhD[b], Francesca Rosini[c], Rob Goldin PhD[b], Ara Darzi PhD[a], Zoltan Takats PhD[b], James M Kinross PhD[a]

[a]Department of Surgery and Cancer, Imperial College London

[b]Department of Metabolism, Digestion and Reproduction, Imperial College London

[c]Imperial College London NHS Trust, London

Corresponding author:

Professor Zoltan Takats

Department of Metabolism, Digestion and Reproduction

SAF Building

Imperial College London

London

SW7 2AZ

Email: z.takats@imperial.ac.uk

Sources of Funding - infrastructure support for this research was provided by the NIHR Imperial Biomedical Research Centre (BRC). Project grants were received from an NIHR Invention for Innovation grant and from Waters Corporation. SM is funded on a Cancer Research UK PhD fellowship. JLA is the recipient of a NIHR Academic Clinical Lectureship. JLA receives funding for his Clinical Lectureship from Imperial College London and The Joyce and Norman Freed Charitable Trust.

Conflicts of Interest - ZT acts as a consultant for Waters and is the inventor of REIMS. No other potential conflicts of interest exist.

Running Head – Lipidomic profiling of colorectal lesion

INTRODUCTION

Powerful metabolomic techniques such as mass spectrometry (MS) and nuclear magnetic resonance have started to reveal that intrinsic relationships exist between cellular metabolism and clinical phenotype. This is being applied to better understand the process of carcinogenesis, with profiling of metabolites within a tissue (the ultimate consequence of all metabolic pathways) used to define metabolic 'fingerprints' that are unique to cancer. This gives insight into how cellular metabolism transitions through the colorectal normal-adenoma-carcinoma sequence.[1-3]

Translating such technologies into diagnostics and therapeutics for colorectal disease requires effectively leveraging the power of metabolomic datasets towards clinically orientated outcomes, ultimately to support clinical decision making.[4] Such an example is performing tissue recognition in real-time, where a tool assists the operator in accurately distinguishing neoplastic from benign tissue when no macroscopic difference exists. This can be applied to reducing the rate of cancer-involved margins during local excision of rectal lesions, ensuring an R0 resection during endoscopic mucosal resection for polyps and diagnosis for indeterminate lesions encountered endoscopically or intra-abdominally.[5, 6] In addition, technologies are required for risk-stratification of lesions (such as by determining the subtype or grade of dysplasia in colorectal adenomas), allowing precision management decisions in real-time. Innovation is required to overcome the insufficient accuracy of existing tissue recognition technologies (such as narrow band imaging), routine requirement for timely histopathology and the lack of a deeper biological insight being used for personalised healthcare.[7, 8]

Metabolomic technologies such as liquid chromatography MS and nuclear magnetic resonance require a laboratory setting and extensive sample preparation; negating use for real-time diagnostics in the clinical setting. Rapid Evaporative Ionisation Mass Spectrometry (REIMS) chemically analyses the surgical aerosol (smoke) generated when an energy source is applied to tissues and is commonly coupled to electrosurgical instruments such as monopolar diathermy.[9, 10] It benefits from having an analytical time of 2 seconds and does not require any interruption to the surgical workflow. Explosive cell rupture caused by tissue heating releases water droplets containing cellular metabolic components, the relative abundance of which can be used to define a tissue's metabolome.[9] Typically measuring negative ions over the 600-1000*m/z* range (largely consisting of complex lipids and referred herein as the 'lipidome'), pilot datasets of REIMS have demonstrated an accuracy of 70-98% in differentiating malignant from benign tissues across colorectal, breast, ovarian, cervical, liver and lung.[6, 11-14] REIMS provides advantages over alternative MS technologies currently being studied for intra-operative tissue recognition (such as the MasSpec Pen and SpiderMass), which suffer from issues including long acquisition times (10-30s), lack of data for colorectal tissues and low total intensities.[15-17]

The aim of this project was to create a powerful, clinically annotated spectral database through the *ex vivo* REIMS analysis of colorectal tissues. The primary outcome was to assess

the accuracy of REIMS in differentiating colorectal carcinoma, adenoma and normal mucosa; describing the lipidomic variation seen throughout this transition. Secondary outcomes were the ability of REIMS to risk stratify colorectal cancers and adenomas by clinically relevant histological features.

METHODS

Study Design and Patient Recruitment

This was a prospective observational cohort study conducted of patients at three London hospitals who had tissue recruited and collected for biobanking between November 2014 and September 2019. All adult patients undergoing endoscopic or surgical resection of colorectal tissues were eligible for inclusion. Preoperatively, patients were required to give written and informed consent for their tissue to be sampled for research, performed as a subcollection of the Imperial College Healthcare Tissue Bank (reference 17/WA/0161 under HTA license 12275) or under a dedicated ethical application (reference 14/EE/0024, approved by the London research ethics committee). Patients were excluded if they suffered with inflammatory bowel disease, a hereditary polyposis syndrome or were less than 18 years of age. A patient advocate was consulted and advised on optimising the patient experience during study recruitment and tissue sampling.

Sampling of Colorectal Tissues

Following resection, tissues were immediately transported fresh to the histopathology department, where approximately 500mg of normal mucosa (10cm from any lesion), carcinoma and adenoma (where available) were sampled for research supervised by a consultant histopathologist. These were bio-banked at -80^{o}C until MS analysis.

REIMS Ex Vivo Mass Spectrometric Analysis

Lipidomic profiling of tissues using REIMS was conducted *ex vivo* using a Xevo G2S QTof mass spectrometer (Waters Corporation, UK), coupled to a modified handheld diathermy pencil (figure 1). 50-100mg tissue pieces were freshly thawed and analysed with monopolar diathermy at 20W, with the surgical aerosol immediately aspirated into the MS via PTFE tubing (driven by a Venturi interface). The MS was configured for the detection of negatively charged ions with data collected continuously over a mass:charge (*m/z*) range of 150-1200. A solution of isopropanol with the standard leucine-enkephalin was co-aspirated at 0.2ml/min. Following analysis, tissue samples underwent histopathological validation by experts in gastrointestinal pathology (RG and FR) applying standard diagnostic criteria.

Tandem MS (MS/MS) acquired spectra over the 50-1200*m/z* range, using quadrupole tuning for the *m/z* of the pre-selected parent ion. Nitrogen was used as the collision gas, with collision energies ranging from 30-60meV.

Statistical Analyses

Data were pre-processed and analysed using an in-house pipeline, starting with lock-mass correction to leucine-enkephalin (at *m/z* 554.2615), normalisation and background subtraction (Abstract Model Builder v1.0.2055, Waters Research Centre, Hungary). Binning was performed at 0.1, with a single spectrum created for each application of diathermy and analyses conducted over the 600-1000*m/z* range. Low quality spectra were excluded prior to modelling, based on pre-defined criteria: signal:noise ratio <1000 (defined as the median intensity of the top 20 peaks divided by the median intensity of all peaks), sampling of non-mucosal tissue, non-diagnostic validation section or <15% containing the tissue of interest and spectra involved in a technical batch effect due to mass drift.

Exploratory modelling included principal components analysis (PCA) with supervised classification modelling using linear discriminant analysis (LDA) and orthogonal partial least squares discriminant analysis (OPLS-DA); conducted in Abstract Model Builder and SIMCA v15 (Umetrics, Sweden) respectively. Analysis of variance (ANOVA) was used for univariate comparisons of relative intensity between groups in RStudio v1.1.419, with *p* value correction using the Bonferroni technique. A corrected p value of less than 0.05 was deemed statistically significant.

The molecular structure of lipid metabolites was identified from characteristic fragmentation patterns during MS/MS, with reference to publicly available databases including METLIN and LIPID MAPS.[18, 19]

RESULTS

161 patients were included in the study, generating 1013 mass spectra from the analysis of colorectal carcinoma (n = 346), adenoma (n = 247) and normal mucosa (n = 420). The demographics of included patients and the sampling methodology are presented in supplementary data 1, http://links.lww.com/SLA/D371. Histological validation demonstrated that REIMS was conducted on 109 tissue samples from 79 suspected carcinomas, 134 tissue samples from 56 polyps and 137 pieces of normal mucosa. Of the 56 polyps analysed by REIMS, 50 were validated as adenomas, 5 were carcinomas and one was a fibro-epithelial polyp. Of the 79 suspected carcinomas sampled, 75 were histologically validated as carcinomas, 3 were adenomas and 1 only sampled normal tissue. Of the adenocarcinomas, 18% were defined as mucinous, with carcinomas demonstrating a variety of prognostic histological features as shown in supplementary data 2, http://links.lww.com/SLA/D371. The mean adenoma size was 33mm (5-80mm), with 17% validated as having high-grade dysplasia (supplementary data 2, http://links.lww.com/SLA/D371).

Lipidomic Profile of Colorectal Tissues

Representative mass spectra from normal, adenoma and carcinoma tissues are presented in figure 2. This reveals that REIMS coupled to monopolar diathermy detects metabolites across

many lipid classes, including fatty acids (FAs, 210-305*m/z*), glycerophospholipids (GPLs; 630-850*m/z*) and triglycerides (TGs; 850-950*m/z*). The most abundant single peaks appear to be FAs such as palmitic acid (C16:0 at 255.23*m/z*) and oleic acid (C18:1 at 281.25*m/z*), however, the GPL region appears to contain the most biological data when considering numbers and size of peaks. Comparing spectra of the tissue types demonstrates a high degree of similarity in the lipidome between them, with the same peaks (and their corresponding metabolites) present across all tissue types. As such, the abundance of no single lipid metabolite is capable of acting as a biomarker of disease and comparisons require multivariate modelling to assess differences in relative abundance across a large number of metabolites.

Tissue Recognition using REIMS

Multivariate modelling was conducted to assess the ability of REIMS in differentiating colorectal tissue types based on the lipidome over the 600-1000*m/z* range (largely GPLs and TGs). A Principal Components Analysis (PCA) plot is presented in figure 3.A, demonstrating that the greatest source of variation between spectra does not appear to be tissue subtype. This is represented by large intra- compared to inter-group variance, with group clusters overlapping. Annotating the plot by demographic data (age, gender) or experimental factors (instrument used, date of analysis) did not demonstrate any batch effects (data not shown).

The first 60 principal components underwent a Linear Discriminant Analysis (LDA), with the plot presented in figure 3.B. This demonstrates clear clustering by tissue type, a graphical representation that the model can differentiate between colorectal carcinoma, adenoma and normal mucosa based on the abundance of the underlying lipid metabolites. Leave-one-patient-out (LOPO) cross validation revealed that REIMS can perform tissue recognition with accuracy for each spectrum of 91.1%, sensitivity of 86.4%, specificity of 95.4%, positive predictive value (PPV) of 90.7% and negative predictive value (NPV) of 93.1% for carcinoma (figure 4.A). When exploring the sources of false negatives for carcinoma (the sampled lipidome of carcinoma being recognised normal or adenoma), this was most common when sizeable tissue samples were analysed and a small proportion of a relatively large number of REIMS analyses were incorrect. When REIMS is used to predict the histological subtype using all spectra generated from a tissue sample (defined as the most frequently predicted tissue subtype) the accuracy increases to 93.1%, sensitivity of 91.5%, specificity of 95.7%, PPV of 90.7% and NPV of 96.1% for carcinoma (figure 4.B).

A further source of misclassifications was the lipidome of adenomas predicted to be carcinoma. In these cases, it was identified that 78% (21 of 27) of adenoma samples were validated to contain high-grade dysplasia (HGD) or were sampled immediately adjacent to carcinoma (an adenoma within which a carcinoma originated). From a clinical perspective, both adenoma and carcinoma requires excision and therefore these may be considered as one group for modelling purposes. REIMS was able to recognise neoplasia (carcinoma or adenoma) from normal mucosa for each spectrum with an accuracy of 93.5%, a sensitivity of 91.2%, specificity of 96.7%, PPV of 97.5% and NPV of 88.6%. When making predictions for

each tissue sample analysed, the accuracy was 96.0%, sensitivity of 94.9%, specificity of 97.8%, PPV of 98.7% and NPV of 91.8%.

Learning curve simulations were used to assess whether the dataset was adequately powered to differentiate carcinoma, adenoma and normal mucosa; or whether analysis of further tissue samples was required. This revealed that maximum classification accuracy was reached for any tissue type with no more than 140 spectra included in the modelling, markedly below the 247 included in the smallest group in this dataset (example shown in supplementary data 3, http://links.lww.com/SLA/D371). Furthermore, it appears that no more than 50 variables (metabolite abundances selected with the smallest *p* values) were required to reach maximum accuracy.

Lipidomic Features of Colorectal Cancer

Univariate and multivariate techniques were used to identify the bins which were most responsible for differentiating colorectal carcinoma, adenoma and normal mucosa; which could then be explored further to establish the molecular structure of the underlying metabolites. This process allowed 61 lipid metabolites to be structurally identified based on their fragmentation patterns, covering GPLs, sphingolipids and DGs. 15 of these were ^{13}C isotopes, with the relative abundance of the remaining 46 metabolites by tissue type presented in supplementary data 4, http://links.lww.com/SLA/D371.

The relative abundance of lipid metabolite subclasses was compared between colorectal carcinoma, adenoma and normal tissues; to identify associations with the carcinogenic sequence (figure 5). In the majority of cases, the relative abundance of lipids by subclass increases at the initiation of adenoma formation, with the exception of phosphatidylcholines, which significantly decrease. Changes in metabolite abundance as adenomas progress to carcinomas follow no clear general pattern. It does appear that the abundance of phosphatidylglycerols progressively and statistically significantly increase throughout carcinogenesis, with a similar pattern seen with sphingomyelins (however the increase from normal to adenoma tissue was not statistically significant). Unsupervised hierarchical clustering demonstrated that the relative abundance of lipid subclasses appears to change independently of each other and that those with similar biosynthetic pathways (for example, ceramides and sphingomyelins) show little similarity in their changes (supplementary data 5, http://links.lww.com/SLA/D371). The relative abundance of GPLs containing monounsaturated fatty acids (MUFA) increases throughout carcinogenesis, with polyunsaturated fatty acids (PUFA) only appearing to increase in adenomas (figure 5).

REIMS in Risk-Stratification of Colorectal Cancers

The ability of REIMS in predicting the presence of histological features of poor prognosis in colorectal cancer is presented in table 1, demonstrating that there is some degree of association between these features and the mucosal lipidome. REIMS appears to have the

highest accuracy when predicting the presence of nodal micrometastasis and has the same NPV as pre-operative cross-sectional imaging (78.0 and 78.8% respectively).

REIMS in Risk-Stratification of Colorectal Adenomas

REIMS was applied to differentiate colorectal adenomas with high-grade dysplasia (HGD) and low-grade dysplasia in real-time based on analysis of the mucosal lipidome across 246 spectra. This reveals that unique lipidomic fingerprints exist for each polyp phenotype, with a high predictive accuracy of 93.5%, sensitivity of 80.0%, specificity of 95.7%, PPV of 75.7% and NPV of 96.7% for HGD (with an OPLS-DA plot shown in supplementary data 6, http://links.lww.com/SLA/D371). Such a distinction does not appear to exist when comparing villous and tubular adenomas subtype (with tubulovillous excluded due mixed nature), were REIMS has an accuracy of 66.9%, sensitivity of 71.1%, specificity of 57.9%, PPV of 78.7% and NPV of 47.8% for the villous subtype.

It was not possible to assess other features suggested by the British Society of Gastroenterology for the definition of an advanced colorectal polyp including adenomas greater than 10mm, the presence of serrated features with dysplasia or without dysplasia if greater than 10mm.[20] This was due to imbalances in the composition of the polyps sampled for research, such that there were insufficient numbers in groups for comparison.

DISCUSSION

This study represents the most comprehensive lipidomic profiling of colorectal tissues using REIMS, demonstrating that unique lipidomic fingerprints exist for colorectal carcinoma, adenoma and normal mucosa; allowing strong associations to be drawn between cellular metabolism and clinical phenotype.

Clinical Impact of Findings

When coupled to monopolar diathermy, REIMS is able to accurately differentiate colorectal tissues based on cellular lipid composition, a feat which can be conducted in real-time, without interfering with the surgical workflow. Such a tissue recognition capability has the potential to be applied in healthcare as a point of care diagnostic system during endoscopy, but it also has applications as a platform to improve the quality and safety of therapeutic interventions. Accurately diagnosing the histological subtype of colorectal polyps endoscopically would support use of the strategies '*diagnose and leave*' (where benign lesions are not removed, particularly hyperplastic polyps of the recto-sigmoid) and '*resect and discard*' (low-risk adenomas can be discarded without pathological analysis); resulting in a reduction in unnecessary polypectomies, procedure time and histology processing costs.[6, 21, 22] REIMS appears to have superior diagnostic potential compared to existing optical endoscopic technologies (such as narrow-band imaging), with a NPV for neoplasia (adenoma or carcinoma) of 91.8% compared to 85.7% for optical technologies.[7] This surpasses the 90% NPV threshold for routine clinical use from the American Society of Gastrointestinal

Endoscopy, however higher thresholds may be necessary for widespread acceptance from clinicians and patients.[8, 22] REIMS also has the potential to be used in parallel with other technologies, including the emerging role of computer-aided diagnosis, augmenting diagnostic ability through provision of biological data.[23] Therapeutic roles for REIMS are likely to focus on the detection of cancer- or adenoma-involved margins, during procedures such as trans-anal minimally invasive surgery (TAMIS) and endoscopic mucosal resection (EMR).[5, 6] An interface which alerts the endoscopist to an adenoma-involved margin in real-time during EMR allows an additional mucosal resection to be performed, reducing the polyp recurrence rate (currently estimated to be 15-20% for non-pedunculated lesions).[24] Similarly during TAMIS, the ability to remove early rectal cancers *en bloc* as an R0 resection directly contributes to an improved oncological outcome.[25]

Lipidomic features of Carcinogenesis

The metabolome can be considered the ultimate consequence of all metabolic reactions and is therefore directly impacted by cellular functioning driven by genomic, transcriptomic and proteomic factors.[26] These are well described to be dysregulated during colorectal carcinogenesis, which underlines the power of REIMS – its ability to explore cellular phenotype through measuring metabolite abundance, indirectly incorporating features usually requiring a range of '-omics' technologies. However, the cellular dysregulation through the normal-adenoma-carcinoma transition is generally considered a progression of multiple small steps and applying (sometimes arbitrary) thresholds related to clinical outcomes is likely to cause difficulty when making classification predictions using REIMS.[27] For example, the greatest source of misclassification error during tissue recognition was adenomas predicted to be carcinomas, where significant similarity is known to exist between these tissues on a genomic, transcriptomic and proteomic level.[28-30] Additionally, cohorts of adenoma cells may have the metabolic dysregulation sufficient for invasion but have yet to extend through the muscularis mucosae to be defined as a cancer, making it challenging to metabolically define the moment the lesion becomes neoplastic. This is also likely to contribute to the relative difficulties REIMS had in risk-stratifying CRCs based on histological features of poor prognosis. The exact definition of the ordinal categories used (for example, exactly what constitutes T1-T4; and the interpretation of which classes are 'early' or 'advanced' cancers) may be a source of error, where the relevant metabolic changes are not contemporaneous with the morphology described pathologically. Alternatively, it may be that other factors are required to accurately risk-stratify CRCs, such as small metabolites, time since lesion inception or the mucosal microbiome.[31] The paucity of literature associating the metabolome and histological features demonstrates the challenge here.

This study presents the most comprehensive insight into the lipidomic fingerprints associated with colorectal normal, adenoma and carcinoma tissue using REIMS, presenting both a highly powered dataset and a large number of structurally identified metabolites. The fact that there does not appear to be a single biomarker for colorectal carcinogenesis and instead the relative abundance of metabolites is required to differentiate tissues is consistent with findings across breast,[32] prostate,[33] lung,[34] liver[11] and many more.[35] It is evident that there are

some cases where lipid subclass alone has a strong association with carcinogenesis. Despite only constituting 1-2mol% of cellular lipids and having roles that are largely described in relation to mitochondria and the endoplasmic reticulum,[36, 37] PGs increase in abundance throughout this transition. They have been demonstrated to selectively activate nuclear protein kinase C bII during the G2 phase of the cell cycle, which is required for mitosis;[38] and is therefore a potential mechanism upon which they are directly tumourigenic. However, the acyl chain saturation status of GPLs appeared to be more significantly associated with tissue phenotype, with this classification sufficient for unsupervised clustering to identify the similarity between adenoma and carcinoma. Increase in the relative abundance of MUFA-containing GPLs is seen when normal colorectal mucosa becomes malignant, with previous studies also identifying such increases in the serum and tissue as free fatty acids.[39, 40] The converse appears to be the case in oesophageal adenocarcinoma,[41] with the reason for this difference (despite such similar tissue types) requiring future exploration. The lipidomic feature of MUFA-containing GPLs increasing in carcinogenesis would be consistent with over-expression of delta-9 desaturase (SCD-1), which is responsible for desaturating palmitic acid (C16:0) following *de novo* fatty acid synthesis.[42] This overexpression has been well described in colorectal cancer and SCD-1 activity is correlated with proliferation of cancer cells, poorer outcomes from CRC and has been made a target of anti-cancer therapies.[3, 43, 44] This link from lipidomic signature detected in real-time using REIMS, through cancer-specific metabolic dysregulation to clinical outcome needs to be confirmed in future studies, as it could be applied to clinical problems such as acting as a biomarker of disease response in anti-SCD treatments.

Limitations

A selection bias exists due to the requirement for not compromising clinical diagnosis when sampling tissue for research, as potentially invasive margins (deep and lateral) require preservation for microscopy. This decreases the yield of smaller and flatter lesions, which particularly impacts polyps (the mean size in this study was 33mm, whereas clinically 81% are less than 10mm).[45] The time between devascularisation and sample freezing was not collected, however, previous work demonstrates no difference in abundance of complex lipids for at least 90 minutes and therefore this is unlikely to be a confounding factor.[46] Sampling was also conducted cross-sectionally, with the most insight likely to be generated from longitudinal sampling of the same lesions throughout carcinogenesis, however, this is not currently viable in humans due to the need for removing adenomas when diagnosed. Performing REIMS analysis on tissues prospectively collected for the biobank was such that this was not a trial of consecutive patients, however, no conscious selection of tissue was undertaken and therefore this is deemed unlikely to have imparted a bias.

Challenges arose from using histopathological validation as the reference standard. When heterogenous large tissue samples were analysed in multiple places, it was not possible to consistently determine the pathological subtype at the exact location of each REIMS analysis and as such, all spectra were annotated as tumour when in fact, some would have been in histologically normal or dysplastic tissue. This was a likely source of false negatives and was

managed in the later batches by using smaller tissue samples with fewer REIMS analyses. In addition, analyses that were conducted in overwhelmingly normal tissue, with a small proportion of tumour cells present, may have caused false negatives as the lipidomic signal of cancer may have been sufficiently diluted. Inter-observer variability between histopathologists has been reported, such as a disagreement in 4% of polyp sections regarding the presence of adenoma, increasing to 17% when asked to differentiate advanced from non-advanced adenomas.[47] Such limitations in the reference standard make it challenging to optimise REIMS to achieve close to 100% accuracies. REIMS analysis is typically conducted over a $1mm^3$ volume, however the lipid composition at this specific location may not be representative of the entire lesion due to the presence of intra-tumour heterogeneity. This is a well-documented phenomenon in CRC, with spatial variation in clonal populations and the microenvironment impacting clinical phenotype of cancer cells.[48, 49] Incomplete lipidomic sampling is likely to be a source of error, particularly when attempting to predict histological features such as invasion depth, where it is not possible to sample the lipidome at the invasive margin. Unfortunately, imaging MS approaches which can spatially resolve metabolites are not suitable for intra-operative translation.[50, 51] Finally, this application of REIMS targets complex lipids that are prone to deprotonation, however, this is not the case for lipids such as TGs (which have a preponderance to form $[M+NH_4]^+$ ions). Therefore, lipids with these electrochemical properties are likely to be under-sampled in negative mode, in addition to other metabolites which have smaller masses; particularly those intrinsically related to cellular metabolism, carcinogenesis and signalling; including arachidonic acid, Kreb's cycle products, amino acids and prostaglandins.[52, 53] Even within the complex lipid range, it was not possible to identify all relevant metabolites due to mixtures within bins (a consequence of no chromatography), isomers and lack of characteristic fragments during MS/MS.

Future Work

The unique lipidomic features associated with carcinogenesis described in this work have been derived from a single dataset and as such, the generalisability of the findings needs to be confirmed by prospectively validating the findings on a truly independent, external cohort of colorectal tissue. This will help to mitigate any biases in this data such as it being based in a tertiary cancer centre, the largely Caucasian population and the statistical workflow used. The power of REIMS for clinical applications (such as the examples discussed above) can now be interrogated using this powerful, high-quality *ex vivo* dataset as the basis for *in vivo* tissue recognition. Colorectal REIMS proof-of-principle *in vivo* translation has been successfully achieved for EMR and TAMIS applications,[5, 6] with a clinical study currently ongoing to prospectively determine the diagnostic accuracy of REIMS for endoscopically diagnosing dysplasia in polyps. These studies must be focussed on clinically relevant endpoints, understanding how REIMS can augment clinical decision-making and the challenges that come with implementation of such an analytical device in a clinical environment. Integration of such a disruptive technology into the patient pathway will be reliant on such studies, in addition to robust guidance on how and where they may be beneficial. In addition, the REIMS platform will need technological development to more easily collect the necessary high-quality data, with a device that is smaller and more portable, quieter, and capable of

complete operation (with problem-solving) by staff with minimal training. Such progress is likely to require ongoing multi-disciplinary input, particularly from clinicians working closely with engineers and mass spectrometry manufacturers.

CONCLUSION

The colorectal lipidome analysed using REIMS is closely associated to tissue phenotype, allowing accurate differentiation of cancer, adenoma and normal mucosa in real-time, in addition to the risk-stratification of adenomas by grade of dysplasia. Unique lipidomic fingerprints exist throughout colorectal carcinogenesis, with changes seen in relative abundance by structurally identified metabolites, lipid class such as phosphatidylglycerols and MUFA chain inclusion in phospholipids. This comprehensive *ex vivo* dataset will be used to power *in vivo* REIMS clinical applications such as real-time colorectal tissue recognition of polyps during endoscopy.

REFERENCES

1. Gao P, Zhou C, Zhao L, et al. Tissue amino acid profile could be used to differentiate advanced adenoma from colorectal cancer. J Pharm Biomed Anal. 2016;**118**:349-355.
2. Farshidfar F, Weljie AM, Kopciuk KA, et al. A validated metabolomic signature for colorectal cancer: exploration of the clinical value of metabolomics. Br J Cancer. 2016;115(7):848-857.
3. Pakiet A, Kobiela J, Stepnowski P, et al. Changes in lipids composition and metabolism in colorectal cancer: a review. Lipids Health Dis. 2019;18(1):29.
4. Kinross JM, Mason SE, Mylonas G, et al. Next-generation robotics in gastrointestinal surgery. Nat Rev Gastroenterol Hepatol. 2020;17(7):430-440.
5. Mason S, Manoli E, Poynter L, et al. Mass spectrometry transanal minimally invasive surgery (MS-TAMIS) to promote organ preservation in rectal cancer. Surg Endosc. 2020;34(8):3618-3625.

6. Alexander J, Gildea L, Balog J, et al. A novel methodology for in vivo endoscopic phenotyping of colorectal cancer based on real-time analysis of the mucosal lipidome: a prospective observational study of the iKnife. Surg Endosc. 2017;31(3):1361-1370.

7. Mason SE, Poynter L, Takats Z, et al. Optical Technologies for Endoscopic Real-Time Histologic Assessment of Colorectal Polyps: A Meta-Analysis. Am J Gastroenterol. 2019;114(8):1219-1230.

8. Sakata S, Lee AHS, Kheir AO, et al. Patient acceptance of the optical diagnosis and misdiagnosis of diminutive colorectal polyps. Gastrointest Endosc. 2017;86(2):372-375 e372.

9. Balog J, Szaniszlo T, Schaefer KC, et al. Identification of biological tissues by rapid evaporative ionization mass spectrometry. Anal Chem. 2010;82(17):7343-7350.

10. Manoli E, Mason S, Ford L, et al. Validation of Ultrasonic Harmonic Scalpel for Real-Time Tissue Identification Using Rapid Evaporative Ionization Mass Spectrometry. Anal Chem. 2021;93(14):5906-5916.

11. Balog J, Sasi-Szabo L, Kinross J, et al. Intraoperative tissue identification using rapid evaporative ionization mass spectrometry. Sci Transl Med. 2013;5(194):194ra193.

12. Phelps DL, Balog J, Gildea LF, et al. The surgical intelligent knife distinguishes normal, borderline and malignant gynaecological tissues using rapid evaporative ionisation mass spectrometry (REIMS). Br J Cancer. 2018;118(10):1349-1358.

13. St John ER, Balog J, McKenzie JS, et al. Rapid evaporative ionisation mass spectrometry of electrosurgical vapours for the identification of breast pathology: towards an intelligent knife for breast cancer surgery. Breast Cancer Res. 2017;19(1):59.

14. Tzafetas M, Mitra A, Paraskevaidi M, et al. The intelligent knife (iKnife) and its intraoperative diagnostic advantage for the treatment of cervical disease. Proc Natl Acad Sci U S A. 2020;117(13):7338-7346.

15. Zhang J, Rector J, Lin JQ, et al. Nondestructive tissue analysis for ex vivo and in vivo cancer diagnosis using a handheld mass spectrometry system. Sci Transl Med. 2017;9(406).

16. Fatou B, Saudemont P, Leblanc E, et al. In vivo Real-Time Mass Spectrometry for Guided Surgery Application. Sci Rep. 2016;6:25919.

17. Schafer KC, Balog J, Szaniszlo T, et al. Real time analysis of brain tissue by direct combination of ultrasonic surgical aspiration and sonic spray mass spectrometry. Anal Chem. 2011;83(20):7729-7735.

18. Fahy E, Subramaniam S, Murphy RC, et al. Update of the LIPID MAPS comprehensive classification system for lipids. J Lipid Res. 2009;50 Suppl:S9-14.

19. Tautenhahn R, Cho K, Uritboonthai W, et al. An accelerated workflow for untargeted metabolomics using the METLIN database. Nat Biotechnol. 2012;30(9):826-828.

20. Rutter MD, East J, Rees CJ, et al. British Society of Gastroenterology/Association of Coloproctology of Great Britain and Ireland/Public Health England post-polypectomy and post-colorectal cancer resection surveillance guidelines. Gut. 2020;69(2):201-223.

21. Kessler WR, Imperiale TF, Klein RW, et al. A quantitative assessment of the risks and cost savings of forgoing histologic examination of diminutive polyps. Endoscopy. 2011;43(8):683-691.

22. Committee AT, Abu Dayyeh BK, Thosani N, et al. ASGE Technology Committee systematic review and meta-analysis assessing the ASGE PIVI thresholds for adopting real-time endoscopic assessment of the histology of diminutive colorectal polyps. Gastrointest Endosc. 2015;81(3):502 e501-502 e516.

23. Mori Y, Kudo SE, Misawa M, et al. Real-Time Use of Artificial Intelligence in Identification of Diminutive Polyps During Colonoscopy: A Prospective Study. Ann Intern Med. 2018;169(6):357-366.

24. Belderbos TD, Leenders M, Moons LM, et al. Local recurrence after endoscopic mucosal resection of nonpedunculated colorectal lesions: systematic review and meta-analysis. Endoscopy. 2014;46(5):388-402.

25. Martin-Perez B, Andrade-Ribeiro GD, Hunter L, et al. A systematic review of transanal minimally invasive surgery (TAMIS) from 2010 to 2013. Tech Coloproctol. 2014;18(9):775-788.

26. Fiehn O. Metabolomics--the link between genotypes and phenotypes. Plant Mol Biol. 2002;48(1-2):155-171.

27. Nguyen HT, Duong HQ. The molecular characteristics of colorectal cancer: Implications for diagnosis and therapy. Oncol Lett. 2018;16(1):9-18.

28. Takayama T, Miyanishi K, Hayashi T, et al. Colorectal cancer: genetics of development and metastasis. J Gastroenterol. 2006;41(3):185-192.

29. Yang Y, Junjie P, Sanjun C, et al. Long non-coding RNAs in Colorectal Cancer: Progression and Future Directions. J Cancer. 2017;8(16):3212-3225.

30. Roblick UJ, Hirschberg D, Habermann JK, et al. Sequential proteome alterations during genesis and progression of colon cancer. Cell Mol Life Sci. 2004;61(10):1246-1255.

31. Nakatsu G, Li X, Zhou H, et al. Gut mucosal microbiome across stages of colorectal carcinogenesis. Nat Commun. 2015;6:8727.

32. Hilvo M, Denkert C, Lehtinen L, et al. Novel theranostic opportunities offered by characterization of altered membrane lipid metabolism in breast cancer progression. Cancer Res. 2011;71(9):3236-3245.

33. Wang X, Han J, Hardie DB, et al. Metabolomic profiling of prostate cancer by matrix assisted laser desorption/ionization-Fourier transform ion cyclotron resonance mass spectrometry imaging using Matrix Coating Assisted by an Electric Field (MCAEF). Biochim Biophys Acta Proteins Proteom. 2017;1865(7):755-767.

34. Lee GK, Lee HS, Park YS, et al. Lipid MALDI profile classifies non-small cell lung cancers according to the histologic type. Lung Cancer. 2012;76(2):197-203.

35. Ifa DR, Eberlin LS. Ambient Ionization Mass Spectrometry for Cancer Diagnosis and Surgical Margin Evaluation. Clin Chem. 2016;62(1):111-123.

36. Uphoff A, Hermansson M, Haimi P, et al. Chapter 11 - Analysis of complex lipidomes. In: Vékey K, Telekes A, Vertes A, editors. Medical Applications of Mass Spectrometry. Amsterdam: Elsevier; 2008. p. 223-249.

37. Morita SY, Terada T. Enzymatic measurement of phosphatidylglycerol and cardiolipin in cultured cells and mitochondria. Sci Rep. 2015;5:11737.

38. Murray NR, Fields AP. Phosphatidylglycerol is a physiologic activator of nuclear protein kinase C. J Biol Chem. 1998;273(19):11514-11520.

39. Kondo Y, Nishiumi S, Shinohara M, et al. Serum fatty acid profiling of colorectal cancer by gas chromatography/mass spectrometry. Biomark Med. 2011;5(4):451-460.

40. Pickens CA, Lane-Elliot A, Comstock SS, et al. Altered Saturated and Monounsaturated Plasma Phospholipid Fatty Acid Profiles in Adult Males with Colon Adenomas. Cancer Epidemiol Biomarkers Prev. 2016;25(3):498-506.

41. Abbassi-Ghadi N, Antonowicz SS, McKenzie JS, et al. De Novo Lipogenesis Alters the Phospholipidome of Esophageal Adenocarcinoma. Cancer Res. 2020;80(13):2764-2774.

42. Green CD, Ozguden-Akkoc CG, Wang Y, et al. Role of fatty acid elongases in determination of de novo synthesized monounsaturated fatty acid species. J Lipid Res. 2010;51(7):1871-1877.

43. Liao C, Li M, Li X, et al. Trichothecin inhibits invasion and metastasis of colon carcinoma associating with SCD-1-mediated metabolite alteration. Biochim Biophys Acta Mol Cell Biol Lipids. 2020;1865(2):158540.

44. Vargas T, Moreno-Rubio J, Herranz J, et al. ColoLipidGene: signature of lipid metabolism-related genes to predict prognosis in stage-II colon cancer patients. Oncotarget. 2015;6(9):7348-7363.

45. Turner K, Genta R, Sonnenberg A. Lesions of All Types Exist in Colon Polyps of All Sizes. The American Journal of Gastroenterology. 2017;113.

46. Haukaas TH, Moestue SA, Vettukattil R, et al. Impact of Freezing Delay Time on Tissue Samples for Metabolomic Studies. Front Oncol. 2016;6:17.

47. van Putten PG, Hol L, van Dekken H, et al. Inter-observer variation in the histological diagnosis of polyps in colorectal cancer screening. Histopathology. 2011;58(6):974-981.

48. Prasetyanti PR, Medema JP. Intra-tumor heterogeneity from a cancer stem cell perspective. Mol Cancer. 2017;16(1):41.

49. Marusyk A, Polyak K. Tumor heterogeneity: causes and consequences. Biochim Biophys Acta. 2010;1805(1):105-117.

50. Inglese P, McKenzie JS, Mroz A, et al. Deep learning and 3D-DESI imaging reveal the hidden metabolic heterogeneity of cancer. Chem Sci. 2017;8(5):3500-3511.

51. Balluff B, Frese CK, Maier SK, et al. De novo discovery of phenotypic intratumour heterogeneity using imaging mass spectrometry. J Pathol. 2015;235(1):3-13.

52. Koundouros N, Karali E, Tripp A, et al. Metabolic Fingerprinting Links Oncogenic PIK3CA with Enhanced Arachidonic Acid-Derived Eicosanoids. Cell. 2020;181(7):1596-1611 e1527.

53. Williams MD, Zhang X, Belton AS, et al. HMGA1 drives metabolic reprogramming of intestinal epithelium during hyperproliferation, polyposis, and colorectal carcinogenesis. J Proteome Res. 2015;**14**(3):1420-1431.

FIGURE LEGENDS

Figure 1 – workflow schematic for the REIMS analysis of tissue *ex vivo*.

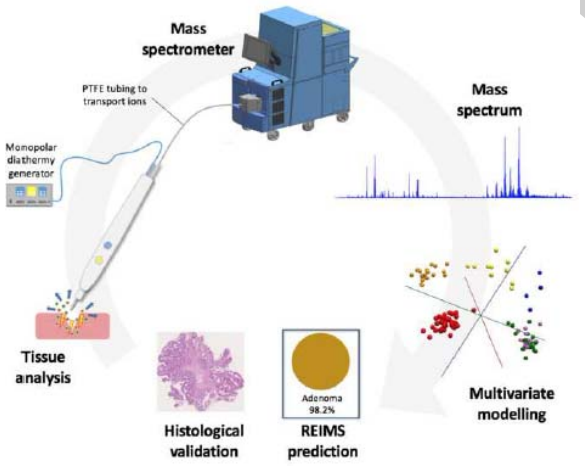

Figure 2 – representative spectra from REIMS analysis of colorectal normal mucosa, adenoma and carcinoma over the 150-1000*m/z* range. Peaks corresponding to lipid metabolite across different classes are annotated, in addition to the standard leucine enkephalin. Example histological sections corresponding to the tissue types are shown. FA – fatty acid; PA – phosphatidic acid; PE – phosphatidylethanolamine; TG – triglyceride.

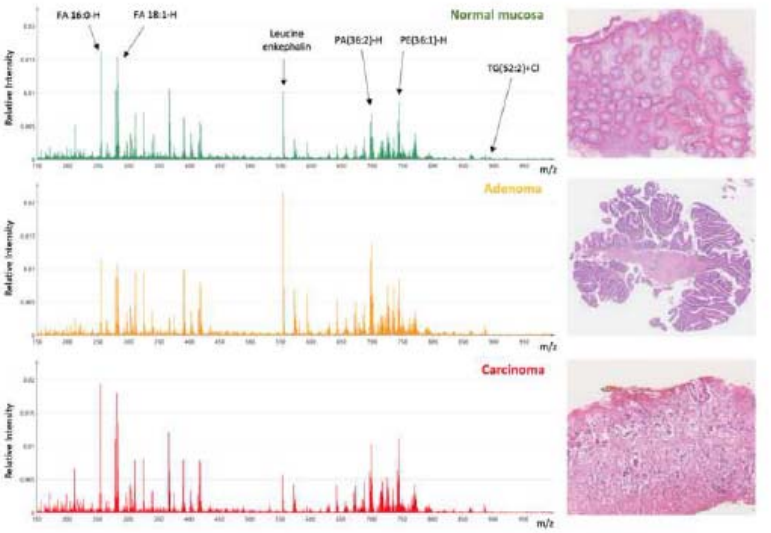

Figure 3 - PCA plot (**A**) and LDA plot (**B**) of the 1013 spectra generated from the *ex vivo* analysis of colorectal tissues over 600-1000*m/z* mass range. PC1, PC2 and PC3 are responsible for 43.3, 21.7 and 4.9% of the variation respectively.

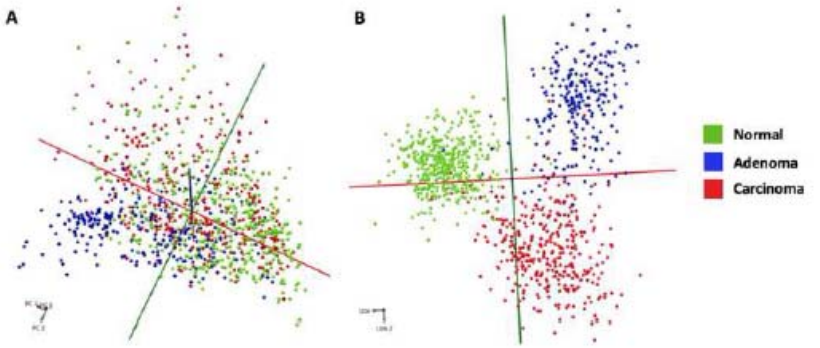

Figure 4 – confusion matrices for the accuracy of REIMS in differentiating colorectal tissue types using each spectrum (**A**) or for the overall tissue sample (**B**).

A

True Subtype	Predicted Subtype For Each Spectrum		
	Normal	Adenoma	Carcinoma
Normal	410	6	4
Adenoma	6	214	27
Carcinoma	28	19	299

B

True Subtype	Predicted Subtype For Each Tissue Sample		
	Normal	Adenoma	Carcinoma
Normal	135	3	-
Adenoma	2	107	11
Carcinoma	5	5	107

Figure 5 – box and whisker plots of the total relative abundance of 7 lipid classes and the saturation status of acyl chains within glycerophospholipids across normal, adenoma and carcinoma colorectal tissue. The box reflects the 25-75th percentiles with a median line and the whiskers are 1.5 x inter-quartile range. A jitter plot of the raw data is presented (grey).

PUFA – polyunsaturated fatty acid; MUFA – monounsaturated fatty acid.

* statistically significant difference in adenoma compared to normal tissue.

** statistically significant difference in carcinoma compared to adenoma tissue.

*** statistically significant difference in carcinoma compared to normal tissue.

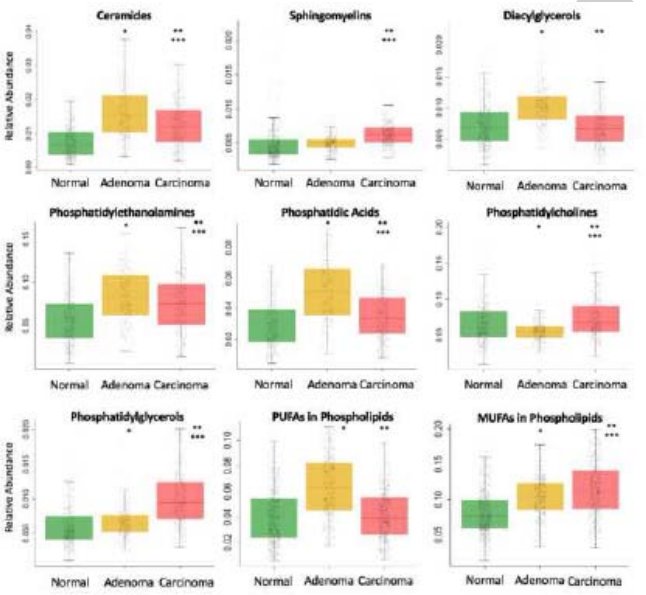

Table 1 – the ability of REIMS to predict the presence of histological features associated with poor prognosis in colorectal carcinomas. The accuracy and NPV were determined from LOPO CV of an LDA model, with the NPV with respect to the second group defined. The Q^2 was determined from an OPLS-DA model with 1 orthogonal and 2 X-Y components. Mucin-producing carcinomas with a tumour volume of less than 50% were included in the mucinous category. NPV – negative predictive value; AJCC – American Joint Committee on Cancer.

Histological Feature	Group Definitions	Accuracy (%)	NPV (%)	Q^2
Tumour Stage	T1/2 *vs* T3/4	52.6	25.3	0.231
Nodal Micrometastasis	N0 *vs* N1/2	62.3	78.0	0.348
AJCC Stage	I/II *vs* III/IV	60.9	77.0	0.326
Extramural Venous Invasion	Absent *vs* present	53.8	67.3	0.285
Lymphovascular Invasion	Absent *vs* present	57.6	61.5	0.389
Tumour Budding	Absent *vs* present	58.2	38.2	0.354
Mucinous Subtype	Non-mucinous *vs* mucinous	49.7	80.0	0.368