Imperial College London

A next-generation *-omics* analysis of the radiotherapy response in rectal cancer

Liam Robert Poynter

A thesis submitted for the award of the degree of Doctor of Philosophy (PhD)

Department of Surgery & Cancer
Imperial College London

March 2021

"Biology is Chemistry…

Chemistry is Physics…

Physics is Maths…

…and Maths is actually Philosophy"

– Anon.

Declaration

This thesis, and all work presented herein, is the result of my own work and original ideas unless specifically stated otherwise. Elements of Chapter 4 of this work have already been published in a peer-reviewed journal with myself as first author; this manuscript can be reviewed in Appendix F. All figures are my own original artwork (except where explicitly stated in the text or legends) and were created under license with Biorender.com. All tables, except where explicitly stated in the text or legends, are my own original work.

Histological tissue analysis was performed in partnership with Professor Robert Goldin in the Department of Pathology at St Mary's Hospital, Imperial College London. Data analysis for the experimental work using Desorption Electrospray Ionisation Mass Spectrometry Imaging (DESI MSI) was performed using MatLab code and an in-house bioinformatic package created and maintained by Dr James McKenzie (Department of Surgery & Cancer, Imperial College London) with code available as open source on GitHub. Next-Generation Whole Genome Sequencing Data were acquired in collaboration with Professor Trevor Graham and Dr Ann-Marie Baker (both of The Barts Cancer Institute, Queen Mary University of London) and assistance with data analysis was provided by Dr Salpie Nowinski (Barts Cancer Institute, Queen Mary University of London).

The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution - Non-Commercial 4.0 International Licence (CC BY-NC). Under this licence, you may copy and redistribute the material in any medium or format. You may also create and distribute modified versions of the work. This is on the condition that: you credit the author and do not use it, or any derivative works, for a commercial purpose. When reusing or sharing this work, ensure you make the licence terms clear to others by naming the licence and linking to the licence text. Where a work has been adapted, you should indicate that the work has been changed and describe those changes.

Please seek permission from the copyright holder for uses of this work that are not included in this licence or permitted under UK Copyright Law.

For Stacey; without you and your boundless patience,
none of this would have been possible

Abstract

Rectal cancer (RC) accounts for over a third of all cases of colorectal cancer (CRC), itself the second-leading cause of cancer-related death worldwide. Fundamental to the treatment armamentarium in locally advanced RC is the application of neoadjuvant radiotherapy (RT). The exact response to this therapy is highly individualised and unpredictable. As we enter an era where the role for RT in RC is set to increase, the critical unmet need remains both the lack of a predictive or prognosticating RT biomarker, and an incomplete understanding of the mechanisms that govern radioresistance in RC. Multiple, Next-Generation '-omics' platforms now provide us with the tools to obtain a more holistic snapshot of the biology at play, in ways that until very recently were not feasible.

This body of work represents the design and implementation of a novel multi-omics pipeline to address a greater understanding of the contributing and complementary mechanisms at play in RC radioresistance. Firstly, I have demonstrated the value in application of machine-learning algorithms to the vast amount of existing experimental data to reveal previously untapped areas likely to be of biological significance. I have subsequently utilised a soft-ionisation mass spectrometry imaging platform to gain maximally representative biological insights into the tumour lipidome in RC in response to RT. I have further complemented this with a holistic whole-genome sequencing (WGS) approach of pre- and post-RT RC specimens, with a targeted view on the likely genomic aberrations that may be at play. My findings have demonstrated that:

(1) ***Gene Ontology (GO)-based network mapping*** of published, investigated, candidate biomarkers demonstrates areas of significant promise in the research of this problem. GO domains of *cellular metabolism*, *response to stimuli* and *cell communication* are not only the areas most implicated as being significantly correlated to RT response, but they are all heavily influenced by the actions of complex cellular lipids.

(2) ***Gene Set Enrichment Analysis (GSEA)*** demonstrates previously unidentified enrichment in gene sets associated with glycerophospholipid (GPL) metabolism in radioresistant RC in a legacy microarray dataset

(3) ***Desorption electrospray ionisation mass spectrometry imaging (DESI MSI)*** reveals increased abundances of phosphatidylethanolamines (PE), phosphatidylserines (PS) and phosphatidylglycerols (PG) in radioresistant RC, with increased integration of polyunsaturated fatty acids (FAs) into these GPLs as compared to RT-naïve tissues

(4) ***Low-pass whole-genome sequencing (lpWGS)*** reveals a significant burden of somatic copy number aberration (SCNA) in pre-RT RC, correlating with a CMS2 or CMS4 phenotype. Furthermore, there is significant heterogeneity in SCNA burden in separately sequenced macroscopic regions of tumour prior to RT, and an increased genomic divergence correlates with a superior response to RT.

Acknowledgements

I remain eternally grateful and indebted to Professor Ara Darzi (Paul Hamlyn Chair of Surgery, Institution of Global Heath Innovation, Department of Surgery and Cancer), Professor Zoltan Takáts (Head of Systems Medicine; Department of Metabolism, Digestion and Reproduction) and Mr Reza Mirnezami (Clinical Lecturer, Department of Surgery and Cancer) for their guidance, support and expert scientific supervision. Each of them has brought unique insights to the project and have been instrumental in crafting the direction of this work.

Achieving the aims and scope of this project, as with any pilot work, would only be possible with valuable input and collaboration from a number of other contributors. I should specifically like to thank Professor Alexander Mirnezami and Miss Nadia Peppa (University of Southampton) for kindly providing access to their institutional tissue bank to enable this work to progress. I am also completely indebted to Mr James Kinross, Clinical Senior Lecturer in the Department of Surgery and Cancer at Imperial College London, for providing the links and opportunities to further pursue avenues allied to this work, through the Cancer Research UK (CRUK) Imperial Centre. This has been of particular benefit in the setting up of the multi-omics platform which has brought together a number of internationally recognised collaborators from across the CRUK landscape. Expert assistance has been willingly provided to me by Professor Trevor Graham, Dr Ann-Marie Baker and Dr Salpie Nowinski (Barts Cancer Institute; Queen Mary, University of London (QMUL)) and I thank them all for their guidance and practical help with Whole Genome Sequencing and the subsequent data analysis,

and also to Professor John Marshall and Dr Caroline Sproat (also of the Barts Cancer Institute, QMUL) for the pilot work with the Hyperion® CyTOF analyses. I am grateful to Professor Robert Goldin (Professor of GI Pathology; Department of Metabolism, Digestion and Reproduction) for giving his valuable time to guide me through annotations of histological specimens and to Hiromi Kudo for technical assistance in staining and creating high-quality digital photographic images of tissue sections for analysis.

Teaching a Colorectal Surgery trainee to near-independence in the practicalities of operating, maintaining and in some cases, repairing an Orbitrap mass spectrometer is no mean feat. For this, and all the support, expert advice (and, above all, patience) they provided, I am extremely grateful to the following post-doctoral colleagues: Dr Renata Soares, Dr James McKenzie and Dr Luísa Doria (Department of Surgery, Imperial College London). I should also like to thank my day-to-day laboratory compatriots Mr Pranav Patel, Mr Sam Mason and Miss Anna Mróz (PhD candidates, Imperial College London) with whom we battled through the highs and lows of mass spectrometry research together. Lastly, to my colleagues from office 1092 on the 10th Floor at St Mary's, with whom I have formed lasting friendships.

This work has received funding from a number of sources. Grant funding was awarded to Mr Reza Mirnezami by National Institute of Health Research (NIHR) Biomedical Research Centre (BRC) at Imperial College London. Grant funding was also awarded to Liam Poynter during the course of this work by both the Bowel Disease Research Foundation (BDRF) and The Royal College of Surgeons/Saven Research and Development Programme. I additionally give thanks to the American Association for Cancer Research (AACR), for a 2018 Scholar-In-Training Award.

Contents

DECLARATION 3

ABSTRACT 6

ACKNOWLEDGEMENTS 8

CONTENTS 10

LIST OF FIGURES 14

LIST OF TABLES 24

PUBLICATIONS, PRESENTATIONS AND PRIZES ARISING FROM THIS WORK 27

ABBREVIATIONS 29

CHAPTER 1 31

INTRODUCTION TO RECTAL CANCER 31

1.1 A background to rectal cancer 32
1.1.1 Rectal cancer as a distinct entity within the colorectal cancer landscape 32
1.1.2 Molecular basis for rectal cancer 34
1.1.3 Clinicopathological features of rectal cancer 37
1.1.3.1 Local spread 41
1.1.3.2 Lymphatic spread 41
1.1.3.3 Haematogenous spread 42
1.1.3.4 Transcoelemic spread 43
1.1.4 Prognosticating histological features and adjuvant radiotherapy for rectal cancer 43
1.1.5 The multimodality approach to rectal cancer treatment 45
1.1.5.1 Surgery for rectal cancer 47
1.1.5.2 Neoadjuvant radiotherapy for rectal cancer 49

CHAPTER 2 54

THE -*OMICS* LANDSCAPE IN DISCOVERY SCIENCE - A TECHNICAL INTRODUCTION TO THE RELEVANT MODALITIES 54

2.1 An overview of realised and potential solid-tissue '-omics' approaches in rectal cancer research 55
2.1.1 Genomics 55
2.1.2 Transcriptomics 56
2.1.3 Proteomics 56
2.1.4 Metabolomics 57
2.1.5 The gut microbiome – "microbiomics" 59

2.2 Metabolism and metabolic disruption in cancer 62
2.2.1 The Warburg effect 62
2.2.2 Lipids in cell biology, colon and rectal cancer 65

2.3 Mass spectrometry-based lipidomic profiling 70
2.3.1 Introduction 70
2.3.2 Basic principles of mass spectrometry 71
2.3.3 Fundamentals of the Orbitrap mass spectrometer 73
2.3.4 Electrospray ionisation (ESI) and desorption electrospray ionisation (DESI) 74

2.4 Data handling approaches for biological mass spectrometry datasets 78

2.5 The role for a lipidomics approach as part of a multi-omics analysis of the radiotherapy response in rectal cancer 81

CHAPTER 3 83

AIMS AND HYPOTHESES 83
3.1 An *in silico* analysis of the radiation response in rectal cancer 84
3.2 Mapping cellular lipidomics in response to neoadjuvant radiotherapy with Desorption Electrospray Ionisation Mass Spectrometry Imaging (DESI MSI) 85
3.3 Analysis of tumour genomic heterogeneity pre- and post-neoadjuvant radiotherapy 86

CHAPTER 4 87

AN *IN SILICO* STUDY OF THE RESPONSE TO RADIOTHERAPY IN RECTAL CANCER – A SYSTEMATIC REVIEW WITH NETWORK MAPPING AND GENE SET ENRICHMENT ANALYSIS 87

4.1 Background and rationale 88

4.2 A systematic review and 'knowledge network mapping' of biomarkers implicated in radiotherapy response in rectal cancer 91
4.2.1 Search methodology 91
4.2.2 Statistical data processing and analysis 92
4.2.3 Results – study inclusion 94
4.2.4 GO Biological Function - Cell cycle and cell proliferation 96
4.2.5 GO Biological Function - Cell death 99
4.2.6 GO Biological Function - Response to stimuli, stress and DNA damage 100
4.2.7 GO Biological Function - Cell signalling 107
4.2.8 GO Biological Function - Cellular metabolism – nucleic acids and lipid products 113
4.2.9 GO 'knowledge network' mapping 116
4.2.10 Discussion 121

4.3 Gene Set Enrichment Analysis of a legacy gene expression dataset 124
4.3.1 Introduction and rationale 124
4.3.2 Search methodology 125
4.3.3 Data processing methodology 126
4.3.4 Results 128
4.3.5 Discussion 131

4.4 Summary and conclusions 135

CHAPTER 5 137

MAPPING CELLULAR LIPID MODULATION IN RESPONSE TO RADIOTHERAPY IN RECTAL CANCER 137

5.1 Introduction 138
5.2 Tissue collection protocol 141
5.3 Patient cohort and clinical metadata 143
5.4 Tissue preparation for DESI-MSI 144
5.5 DESI MSI experimental setup 145
5.6 Histopathology and raw data pre-processing 149
5.7 Image alignment and tissue type annotation 150
5.8 Multiple sample analysis 150
5.9 Compound identification with DESI MS/MS 151
5.10 Results – Clinicopathological data 152
5.11 Results – DESI MSI Spectral acquisition 154
5.12 Results - histopathology quality control 157
5.12 Results – annotation and discrimination of in-sample tissue types 158
5.13 Results – multiple sample comparative analysis 161
5.14 Results – the lipids responsible for the classification 165
5.15 Lipid identification with DESI MS/MS 166
5.16 Discussion 170
5.17 Conclusions 179

CHAPTER 6 181

THE GENOMIC LANDSCAPE IN RADIATION RESPONSE – NEXTGEN SEQUENCING OF PRE- AND POST-RADIOTHERAPY SPECIMENS IN RECTAL CANCER 181
6.1 Introduction 182
6.2 Methods – Patient cohort and tissue collection protocol 184
6.3 Methods – Extraction of genomic DNA from FFPE (pre-RT) rectal cancer tissue samples 184
6.4 Methods – Extraction of genomic DNA from fresh frozen (post-RT) rectal cancer tissue samples 187
6.5 Methods – DNA quality control, library preparation and NextGen WGS sequencing 187
6.6 Methods – WGS data analysis 188
6.7 Results – DNA extraction and sequencing run QC 189
6.8 Results – Analysis of the burden of SCNAs on radiotherapy response 190
6.9 Results – Analysis of the impact of divergence of genomic aberration on tumour regression grade 200
6.10 Discussion 203
6.11 Conclusions 211

CHAPTER 7 212

COMPLETING A HOLISTIC MULTI-OMICS WORKFLOW FOR DISCOVERY SCIENCE IN RECTAL CANCER RADIORESISTANCE 212
7.1 Overview of the multi-omics platform 213
7.2 Multiplexed, quantitative, imaging mass cytometry with Hyperion™ 218
7.3 Evaluation of the on-tumour microbiome in radioresistant rectal cancer 223
7.4 Summary 225

CHAPTER 8 226

SUMMARY AND CONCLUSIONS 226

8.1 The rationale for this research 227

8.2 Summary of findings 229
8.2.1 *In silico* analysis of the response to radiotherapy in rectal cancer – mining the current knowledge landscape 229
8.2.2 Mapping cellular lipid modulation in response to radiotherapy with DESI MSI 231
8.2.3 Assessment of the whole tumour genome pre- and post-radiotherapy 232

8.3 Promising areas for future work 234

REFERENCES 236

APPENDIX A: BIOMARKER EXTRACTION AND GO ANNOTATIONS 265

APPENDIX B: KEGG GPL METABOLISM PATHWAY 273

APPENDIX C: APPLIED DESI MS/MS SETTINGS FOR THERMO® *LTQ X*L 274

APPENDIX D: DNA EXTRACTION PROTOCOLS FROM BOTH FFPE AND FRESH FROZEN TISSUE 275

APPENDIX E: DNA EXTRACTION QC, WGS CELLULARITY ESTIMATES AND ALL-SAMPLES SCNA MAPPING 278

APPENDIX F: PUBLISHED PAPER 282

List of figures

Figure 1.1 – Summary figure of CMS distribution and core genetic features of tumours according to embryological development of the colon and rectum. A mixed CMS phenotype may be present in up to 13% of cases of CC and RC. Adapted from Dekker et al.

Figure 1.2 - The adenoma-adenocarcinoma sequence; (1) diminutive, benign polyp (tubular or tubulovillous adenoma; (2) high grade dysplasia through (3) carcinoma in-situ (polyp cancer); (4) invasive adenocarcinoma. The Duke's Classification was, for many years, in common usage to describe staging of colon and rectal cancer in a concise manner. Duke's A, tumour extending to - but not breaching - the muscularis propria; Duke's B, tumour extending to the serosa; Duke's C, tumour breaching serosa with involvement of local lymph nodes. Further subdivided into Duke's C1 - involving local draining mesorectal/mesocolic nodes – and Duke's C2 when apical node involved. "Duke's D" (not shown, and not originally described) implies ***D***istant metastasis. This classification has undergone many alterations and reiterations since its original description in 1932 and has now largely been entirely superseded by the TNM system, following extensive clinical research into the prognosticating features of individual tumours in relation to their location and depth of invasion.

Figure 2.1 – Schematic workflow of a typical microbial 16S rRNA sequencing experiment.

Figure 2.2 - The Warburg effect. Cancer cells exhibit a more rapid consumption of available glucose due to a drive towards aerobic glycolysis, as opposed to oxidative phosphorylation through the TCA cycle. This occurs even in the presence of sufficient oxygen, and therefore differs from the standard (i.e., hypoxic) conditions which drive anaerobic fermentation, which accounts for lactate generation in normal tissues.

Figure 2.3 - Metabolic pathways in cancer cells. Increased fatty acid, lipid and nucleotide synthesis are key endpoints in the metabolic cascade that follows. GLUT1, Glucose transporter 1; TCA, tricarboxylic acid cycle; ETC, electron transport chain; ROS, reactive oxygen species; HIF-1, hypoxia inducible factor 1; PDK, pyruvate dehydrogenase kinase; FASN, fatty acid synthase. NADP, nicotinamide adenine dinucleotide phosphate.

Figure 2.4: Structure of a typical membrane phospholipid, in this case a phosphatidylethanolamine (PE) comprising both 16:0 (palmitate) and 18:1 (*cis*-9, oleate) fatty acyl chains (black), typically found on the cytoplasmic surface of the lipid bilayer. The hydrophilic head group (blue) in this case denotes the *-ethanolamine* component but could be equally plausible with other head groups; for example, methylation of the amine group here would lead to a phosphatidylcholine (PC). The abbreviated nomenclature for the above (*1-hexadecanoyl-2(9Z-octadecanoyl)-sn-glycero-3-phosphoethanolamine*) is PE(16:0/18:1(9Z)).

Figure 2.5 - Schematic of intracellular fatty acyl (FA) and glycerophospholipid (PL) production. Cytoplasmic citrate is exported from the TCA cycle and hence into FA generation by action of Type I Fatty Acid Synthase (FASN) on Acetyl CoA. In addition to the Warburg effect previously outlined, cancer cells exhibit increased concentrations of mitochondrial membrane citrate carriers (CIC), driving a higher proportion of citrate into the cytoplasmic compartment and hence into FA generation. The ultimate fate of this FA pool is either lyso- or phosphatidic acid (LPA, PA respectively) generated by the endoplasmic reticulum or onwards into PL generation via diacylglycerol (DAG) generation. DAG generation is further fuelled by triglycerides (TG) drawn from intracytoplasmic lipid droplets. From here, DAG may enter either the CDP-DAG or Kennedy pathways with flux between the two, dependent on the ultimate PL generated. ACL, ATP citrate lysase; ACS, fatty acid CoA ligase; GPAT, G3P-acyl transferase; AGPAT, acyl-G3P transferase; PAP, phosphatidic acid phosphohydrolase; DGAT, DAG acyltransferase.

Figure 2.6 – Example mass spectrum in negative ionisation mode across the m/z range 150-2000. Each peak corresponds to a m/z value for a detected ion, and this is graded according to relative abundance on the *y* axis.

Figure 2.7 – Schematic representation of a typical DESI MSI experiment. Adjustments can be made in geometry of the sprayer assembly, voltage, ionisation mode and solvent of choice. The sample is typically mounted onto a PTFE or glass slide – mounting on a standard glass microscopy slide permits downstream H&E staining and other histological modalities for comparison. Figure from Takáts *et al* 2004

Figure 2.8 - A typical example of the spatially resolved chemometric data acquired from a DESI MSI run using colorectal tissue. Left, post-DESI H&E-stained colorectal cancer tissue

section; right, ion map (with post-processing RGB colour intensity grading) of an ion with m/z 337.3112 (FA 24:1, i.e., 24 carbon chain with one double bond). Note the relative abundance in varying degrees in adenocarcinoma tissue compared to its absence in submucosa and muscularis layers (unpublished data).

Figure 4.1 – The Hue, Saturation, Brightness (HSB) colour space allows conversion of a numerical value into a standardised colour representation that can be used in a quantitative manner and far more simplistically than RGB.

Figure 4.2 – PRISMA flow outlining search methodology and study exclusion/inclusion for quantitative synthesis

Figure 4.3 – Summary schematic of stress response pathways, amalgamating and linking the implicated biomarkers from this review which fall under the GO biological function for response to stress and DNA damage. Pathways A, B and C are complementary and work simultaneously. Note critical role of lipid mediators in inflammatory process following release of fatty acids form the cell membrane.

Figure 4.4 - Static knowledge network visualisation of biomarkers clustered according to Gene Ontology terminology, as per data from table 1. Each node represents a Gene Ontology (GO) Term, clustered under group headings (spokes). Size of nodes was calculated based upon the number of biomarkers under that GO term, with hue/saturation/brightness mapped to aggregate significance of multiple statistical tests employed across studies.

Figure 4.5 – GSEA Enrichment plot of the Snipstad et al array as per the KEGG GPL metabolism gene set in MSigDB v5.2. Peak enrichment was more significant for depleted genes in RT.

Figure 4.6: Heatmap of gene expression in the Snipstad et al microarray series following GSEA, collapsed to the KEGG GPL metabolism gene set in MSigDB v5.2. Phenotype labels RC RT+ (grey) and RC RT- (yellow)

Figure 4.7 - Summary map of enriched and depleted genes in GPL metabolism in the Snipstad dataset. Heat maps superimposed onto genes as per expression data and log^2fold changes. KEGG codes (e.g., *PPAP* 3.1.3.4) further subdivided into gene isoform expression data – e.g., PPAP2A and PPAP2B enriched in RT+ but PPAP2C enriched in RT-, and vice versa.

Figure 5.1 The classical 'top down' approach to investigation of biological processes. The limitation is the degree to which it considers modifications at each level and the resultant metabolites of interest

Figure 5.2. A possible proposed workflow of a solid-tissue, multi-omics approach; complementarity with sequencing and proteomics approaches but with a more holistic systems biology overview. Integrative bioinformatics will be critical to processing these combined datasets rather than analysis in silos. DESI MSI is just one modality for evaluating the metabolome - in this context for the tumour microenvironment lipidome – but other forms of biological MS and indeed nuclear magnetic resonance (NMR) studies would apply. Equally, 'RNASeq' broadly covers a wide range of Next Generation RNA sequencing technologies, including spatially-resolved transcriptomics.

Figure 5.3. Tissue sample characteristics. From a total of 26 patients there were paired tumour and normal samples from each which were harvested at surgical resection. These snap-frozen specimens were suitable for DESI MSI as their fresh-frozen nature provides a metabolic 'snapshot' which is then unaltered by the mechanics of ionisation in DESI.

Figure 5.4 – (Top) DESI MSI interface as connected to Thermo *Exactive* Orbitrap mass spectrometer. (A) Windows PC running Thermo™ Xcalibur software, spectral acquisition in-progress; (B) Thermo *Exactive* MS; (C) Heated sample inlet; (D) Home-built sample stage; (E) electrospray capillary assembly; (F) solvent pump. (Bottom) (G) nitrogen gas inlet to sprayer; (H) electrospray ionisation sprayer; (I) inlet capillary; (J) mounted sample slide (attached with double-sided tape); (K) moving stage

Figure 5.5 – Full mass range 150-2000m/z in negative ionisation mode for CRA283-12 N S2 row 32, time index 1.1 minutes. On-tissue, this demonstrates significant abundance in the phospholipid range 650-900 m/z with corresponding fatty acyls and triglycerides (normal colorectal mucosa)

Figure 5.6 – Raw single-ion, intensity-graded imzML images from CRA532-15 S7 TUMOUR; A, m/z 737.500; B, m/z 810.500.

Figure 5.7 - Raw spectral profiles (non-centroided data) taken from representative rows from RT+ and RT- cancer samples and matched normal samples. Full mass range acquisition has here been narrowed to compare relative abundances amongst compounds in the GPL mass range (i.e., m/z 600-9. Sample spectra here selected as relatively well-matched on histological grounds: RT-, CRA247-13: T3N1 EMVI positive; RT+, CRA550-15: T3N0 EMVI positive, TRG3.

Figure 5.8 – Stromal component and residual tumour glands still visible after DESI MSI of section CRA306-13 S7 TUMOUR (left) but higher definition of tissue features including gland/stromal interface completely preserved in the adjacent section with near-identical architectural appearances due to 10μm distance in z axis.

Figure 5.9 – Within the MatLab toolbox the optical/ion images were aligned according to the protocol as outlined. Histology-guided annotation of tissue types was performed by first delimiting the optical H&E image to pixels with equivalent size to the DESI MSI pixel size (100μm x 100μm). Within the toolbox a colour-coded classification was then employed to either select individual pixels/cells or to draw marquee areas over homogenous regions: blue, normal rectal mucosa; pink, normal submucosa; green, muscularis. White marquees are drawn over the background to differentiate from solvent background in tissue classification. Next, unsupervised analysis was performed with PCA with loadings plots for PC1 and PC2 shown in mass range 150-950 m/z above a cluster plot which demonstrates separation of mucosa, submucosa and muscularis in PC1 and PC2.

Figure 5.10 – Representative H&E-stained tissue section of normal colorectal mucosa from patient CRA268-13 (A), demonstrating normal mucosa and submucosa. (B) Composite, all-spectra DESI-MSI image with background (i.e., solvent) filtered out and represented as dark blue– signal intensity graded from white (low) through light blue to red. Selecting ions of interest in RGB image mode, DESI-MSI demonstrates differing lipid compounds found almost exclusively in colorectal mucosa (C – PI 18:0/20:4 m/z 885.50) and normal submucosa (D – 331.20 m/z, unidentified fatty amide)

Figure 5.11 – Annotated, spatially aligned, H&E stained and DESI-MSI images for a T3N0M0 rectal cancer treated with LCCR (Mandard TRG 3). MMC cross-validation demonstrates significant separation between adenocarcinoma (*green*) and desmoplastic stroma (*pink*) pixels. MSI demonstrates significant differences in lipid abundances between adenocarcinoma glands that have proven resistant to radiotherapy vs the associated desmoplastic stroma. Furthermore, in the RGB image (top right), selected m/z values are demonstrated as being more abundant in adenocarcinoma (red, m/z 788.55) and desmoplastic stroma (blue, m/z 331.26). m/z 196.04 (green) is found exclusively in scatted adenocarcinoma glands.

Figure 5.12 – For all annotated rectal adenocarcinoma pixels across the dataset, classified as either RT+ or RT-. (A) PCA full mass range with Eigenvalues plot and (B) MMC with mass range delimited to 600-950Da, cross validated with LOOCV model in (C). Predicted vs. actual representation in confusion matrix represents PPV and NPV.

Figure 5.13 – Analysis of RT+ samples only grouped according to Tumour Regression Grade (TRG). (A) MMC plot with cross validation in confusion matrix (B). Selected discriminating compounds in the model shown in (C) plotted as $\log^2$ fold changes.

Figure 5.14 – Analysis of RT+ vs. RT- samples with MMC clustering in (A) and ANOVA volcano plot in (B) windowed to highlight compounds with significant $\log^2$ fold change. A total of 11 GPLs were significantly more abundant in RT+ vs RT-, with C-F demonstrating example box plots from these selected outlying ions of interest, (C) m/z 776.5602, (D) m/z 778.5748, (E) m/z 773.5329 and (F) m/z 810.5290, p-values $7.1531E^{-20}$.

Figure 5.15 - Example MS/MS spectra (in this case both MS2 and MS3 of the major fragment) of m/z 810.5281 obtained at DESI MSI of a RT+ tumour (CRA550-15). DESI MS/MS performed on the same patient sample. Clear fragmentation pattern corresponding to an 18:0/20:4 PS – annotated peaks correspond to either the ionised FA chain itself (i.e., at 283.00, 18:0) or remnant fragment after neutral or ketene loss of the FA chain (i.e., the peak at m/z 419.00 corresponds to loss of the 20:4 *sn2* chain but the ionised *sn2* fragment itself was not clearly seen).

Figure 5.16 - A more complex sample demonstrating a mixture of (at least) the two GPLs shown, which both have a m/z of 774.5440. The unaccounted peak at m/z 308.7 likely corresponds to a 20:1 fatty acyl chain and therefore there is likely to be a third compound in this mixture.

Figure 5.17 - Truncated view of GPL synthesis, demonstrating the interplay between the CDP-DAG (blue) and Kennedy (green) pathways. Fatty acyls are liberated from GPLs by the action of various isoforms of phospholipase A2 (PLA2). PA is the gatekeeper to both pathways, whereby PG, phosphatidylinositol (PI) and cardiolipins are purely products of the CDP-DAG pathway. There is intimate crossover between PC, PE and PS, with PSS1 able to act on CDP-DAG to provide linkage into the Kennedy pathway via PS and PE. PPAP2, phosphatidic acid phosphatase 2; LP(x)AT variously, lysophospho(-*x*) acyltransferase, where '*x*' is -choline, -ethanolamine, -serine, -glycerol, -cardiolipin or -inositol; PSS, phosphatidylserine synthetase (-1, -2); EPT1, ethanolamine phosphotransferase.

Figure 5.18 - Schematic view of ELOVL activity in humans. Omega-3 and omega-6 fatty acids have a distinct chain of derivatives through ELOVL5, -2 and -4. Saturated fatty acids elongate exclusively through ELOVL6 and -1. 'Omega' prefix refers to position of first double bond, i.e., ω-3 is 3 carbons away from terminal methyl group whilst ω-6 is at 6 carbons.

Figure 5.19 – Structural changes in, for example, PS 18:0/18:1 (diacyl-) to ether (O-; alkylacyl-) and vinyl ether (P-; alkenylacyl-) lipids. The *sn2* chain remains unchanged in both, whereas in PS (P-18:0/18:1) there is a double bond placed in the *sn1* as an alk-1-enyl group

Figure 6.1 – Example of H&E annotation of the multiple biopsies taken pre-RT from CRA550-15. Separate sections taken from macroscopically distinct regions of tumour and embedded into a single paraffin wax block then cut up and slide-mounted as above. Sections 1-5 validated as containing rectal adenocarcinoma glands with varying degrees of stroma. Unlabelled sections are either, for example, normal tissue (*) or necrotic fragments and therefore not suitable for sequencing.

Figure 6.2 – All samples (*n* = 65), including pre- and post-RT. For those with minimal to no recorded aberrant bins, examining cellularity of samples revealed very small amounts of

tumour either due to necrosis or stromal contamination. The majority of these were in the post-RT cohort whereas the matched pre-RT multi-region samples demonstrate good cellularity

Figure 6.3 – Violin plots of all samples (*n* = 65) - including pre- and post-RT - as per figure 5.2, plotted against percentage of genome aberrated as measured by frequency of SCN gains or losses

Figure 6.4 – Total SCNAs plotted as losses (blue) and gains (red) from this cohort as compared to available annotated datasets within The Cancer Genome Atlas; hallmark changes (e.g., gains on Chr7, Chr13, Chr20, losses on Chr4) are present, demonstrating validity of dataset

Figure 6.5 – Histology image from slide-scrape dissection and matched copy number (CN) plot from CRA306-13 (T3 N0 M0, 0/23 lymph nodes involved, EMVI positive, no tumour perforation, no radiotherapy given). All tissue on slide is tumour and so all digested for sequencing. For copy number plot, blue bars indicate net CN gains and red bars indicate losses.

Figure 6.6 – Histology image from slide-scrape dissection and matched copy number (CN) plot from CRA174-12 (T3 N0 M0, 0/11 nodes involved, EMVI positive, no tumour perforation, neoadjuvant RT and CapOX given, TRG2). Four distinct macroscopic regions identified as being composed of sufficient tumour for separate sequencing as annotated. Significant variation in SCNA burden noted between macroscopic regions of primary tumour. Blue bars indicate net CN gains, red bars indicate net CN losses.

Figure 6.7 – Histology image from slide-scrape dissection and matched copy number (CN) plot from CRA174-12 (T3 N0 M0, 0/11 nodes involved, EMVI positive, no tumour perforation, neoadjuvant RT and CapOX given, TRG2) tissue taken at resection after RT. Areas of tumour annotated by outlines, asterisks (*) denote areas of stroma which were excluded from dissection and digestion. Bland SCNA pattern despite significant SCNA burden pre-RT. Cellularity of data estimated at 0.3. Blue bars indicate net CN gains, red bars indicate net CN losses.

Figure 6.8 – Histology image from slide-scrape dissection and matched copy number (CN) plot from CRA532-15 (T3 N0 M0, 0/28 nodes involved, EMVI positive, no tumour perforation, neoadjuvant RT and CapOX given, TRG4) biopsies taken prior to RT. Two tumour-bearing areas sequenced separately with remainder being excluded as necrotic/stroma. Biopsy 2 demonstrates a higher burden of SCN gains, notably across Chr 1q, Chr 6 and Chr 12 compared to biopsy 1. Blue bars indicate net CN gains, red bars indicate net CN losses.

Figure 6.9 – Histology image from slide-scrape dissection and matched copy number (CN) plot from CRA532-15 (T3 N0 M0, 0/28 nodes involved, EMVI positive, no tumour perforation, neoadjuvant RT and CapOX given, TRG4) resection following RT. Outlined areas on H&E image denotes tumour which was sequenced whereas asterisk (*) denotes area of necrosis which was not included. Pattern of SCNA mirrors the pattern of CN gains in biopsy 2 with gains on Chr 6, Chr 8, Chr 12 and Chr 17. Blue bars indicate net CN gains, red bars indicate net CN losses

Figure 6.10 – Pairwise comparison of divergence between differing macroscopic regions of RC prior to administration of RT. None of the post-RT samples are included in this analysis and this therefore represents the 'core' genomic makeup prior to commencing RT. X-axis = sample IDs compared in each block – all combinations accounted for across all TRGs represented in the cohort. Bar height represents aggregated differences in SCN loss and gain. There is significant divergence in the genome within each multi-region set; TRG 2 appears more homogeneous versus the blander SCNA profile in TRG 3-4.

Figure 6.11 – Violin plots of divergence measured as in Figure 6.10 – again, each point represents a measure of divergence in a pairwise sample comparison between all possible combinations. At the pre-treatment stage, RCs that ultimately have a good TRG (1-2) have a more heterogeneous genomic landscape with a higher aggregate proportion of SCN losses and gains. Poor responders (TRG3-4) have a more homogenous genome with a far lower incidence of SCNAs. In this sample set, the TRG3 cohort appears to behave more in-line with those in TRG4; although with significance set at $p < 0.01$ there is still a statistically significant difference between the two groups albeit not to the order of magnitude seen in the cohort comparisons between either TRG2 and TRG3 or TRG2 and TRG4.

Figure 7.1 - The proposed workflow for a holistic multi-omics approach to the unrealised understanding of the radiation response in rectal cancer. Potentially the most important step is the bioinformatics strategy to resolve integration of these varying datasets.

Figure 7.2 – Detailed pipeline of how I have applied the workflow proposed in figure 6.1 to my work. The resultant sequencing and IMC studies can be either histology-driven or metabolite-driven, based on ion imaging acquired through DESI MSI

Figure 7.3 – Complementary assessment of archival FFPE pre-RT rectal cancer biopsies. Arrows denote 5μm thickness/distance between samples. (A) H&E histology assessment guides (B) regions bearing tumour for separate DNA extraction and sequencing as discussed in Chapter 5. The same approach can be applied for RNA extractions for RNASeq from a further immediately adjacent section. (C) denotes selected regions of interest for Hyperion™ CyTOF® further selected and annotated to allow for heterogeneity of sample – coloured boxes select normal rectal mucosa (yellow), rectal cancer (green), high grade dysplasia (white) and lymphoid follicle (black)

Figure 7.4 – Schematic outline of the Hyperion™ IMC system – adapted from the materials provided by Fluidigm®

Figure 7.5 – Example IMC images from selected regions of interest of normal rectal mucosa as part of optimisation steps for the multiplexed antibody panel. Selected antibody targets shown which demonstrate differential expression profiles according to the tissue compartment in which they are located, for example a-SMA being exclusively located in the basement membrane and submucosa.

Figure 7.6 – Log-transformed box and whiskers abundance plots of taxa identified in 16S rRNA sequencing *on tumour* responsible for clustering in Principal Co-ordinates Analysis (PCoA). In-line with previous pilot work by the group, there is relative overexpression/survival of oral pathobionts on-tumour in radioresistance. Whether this demonstrates true symbiosis or merely survival of an already-established highly pathogenic colony on-tumour remains to be established.

List of tables

Table 1.1 - Consensus molecular phenotypes (CMS) for colon and rectal cancers as postulated by Guinney et al. MSI, microsatellite instability; SCNA, somatic copy number alterations; CIMP, CpG methylator island phenotype; TGFβ, transforming growth factor-beta; DFS, disease-free survival; OS, overall survival

Table 1.2 – Current iteration (8th Edition) of the *Tumour-Node-Metastasis* final pathological staging system for colon and rectal cancer, as per UICC TNM8. Post-neoadjuvant radiotherapy there is a further 'y' prefix, i.e., ypT3 and a supplementary note on tumour regression grading (TRG).

Table 2.1 – Selected, recently published studies employing metabolomics-based platforms for phenotyping of colon and rectal cancers and development of (predominantly) lipidomics-guided tissue type classifications. Commonality is seen between studies in their findings, particularly with regard to the role of glycerophospholipids (GPL) and fatty acids (FA). LOOCV, Leave One (Patient) Out Cross-Validation; PC, phosphatidylcholine; LPC, lysophosphatidylcholine; AUC, area-under-the-curve. MALDI MSI, matrix-assisted laser desorption ionisation mass spectrometry imaging; REIMS, rapid evaporative ionisation mass spectrometry; DESI MSI, desorption electrospray ionisation mass spectrometry imaging; HR-MAS NMR, high-resolution magic angle spinning nuclear magnetic resonance spectroscopy.

Table 4.1 - Association of cell-surface clusters of differentiation (CD) molecules with response to LCCR. In the study by Guedj et al, CD184 labelled as CXCR4 (C-X-C chemokine receptor 4). CD340 is more commonly referred to as HER-2. OR = odds ratio

Table 4.2 - Biomarkers classified according to Gene Ontology hierarchical tree (biological functions). Node size was calculated based upon number of biomarkers within each GO term, accounting for sample size in each study. The number of references for each biomarker reflects the number of studies retrieved. Significance was mapped to the HSB colour space to reflect overall significance of each node in the resulting network, with a lower relative value reflecting greater significance level based upon multiple statistical tests used across studies.

The HSB colour space (Hue, Saturation, Brightness) was selected to run from 0^{o} (red) through 120^{o} (green)

Table 4.3 – Definitions of the Tumour Regression Grading (TRG) systems used across the studies identified in this review

Table 4.4 - Selected genes within KEGG GPL gene set demonstrating significant differential expression between cohorts. In the RT- samples, the above reported genes demonstrate core enrichment .

Table 5.1. Comparison of the two TRG systems applied to the sample cohort, for those patients who received RT

Table 5.2 – Summary characteristics of final pathological staging samples from both cohorts. For the purposes of comparison between groups, AJCC staging has been used.

Table 5.3 – Full clinicopathological metadata for the sample cohort (*n* = 26). TNM reported as TNM as per the core dataset at the time of sample harvest.

Table 5.4 - Results of DESI MS/MS on representative cancer tissue sections subjected to RT. Major fragments as listed compared to known fragmentation patterns in LipidMaps; where only the dominant fragment was visible in MS2, MS3 performed to fragment the fatty acyl chains as shown. Not all fragments shown can be accounted for by a single GPL; where possible the various GPLs in the mixture are identified. For m/a 810.5281, the intense peak at 528.67 implies loss of a 22:4 acyl chain as a ketene, possibly from a PE, but no other clear fragments were seen. Fatty acyl chains with an O- prefix are ether lipids, whilst those with a P-prefix are plasmalogens. The PE ID for 778.5749 is obtained by neutral water loss in addition to protonation as shown. PA = phosphatidic acid; PE = phosphatidylethanolamine; PG = phosphatidylglycerol; PS = phosphatidylserine.

Table 6.1 – Comparison between microarrays and *NextGen* sequencing technologies. NGS serves to rectify a number of the limitations of microarrays especially with regard to reproducibility and data limitations

Table 6.2 – Selected genes coding for enzymes in the FA elongation, PE/PS interplay within the Kennedy pathway and peroxisomal catalysis of ether lipids cascades.

Table 7.1 – Conjugated antibodies utilized in the multiplexed 23 antibody panel on pre-RT RC biopsies on the Hyperion^{TM} IMC platform. Concentrations separately validated and all antibodies validated prior to use with Western blot data.

Publications, presentations and prizes arising from this work

In peer-reviewed journals:

Poynter LR, Galea D, Veselkov K, Mirnezami A, Kinross JM, Takáts Z, Nicholson J, Darzi A, Mirnezami R. Network mapping of molecular biomarkers influencing radiation response in rectal cancer. *Clinical Colorectal Cancer.* 2019;18(2):e210-e222 (Appendix A)

First-author presentations to International Meetings:

Poynter LR, Mirnezami R, Soares R, McKenzie J, Patel P, Peppa N, Mirnezami A, Kinross J, Darzi A, Takáts Z. Modulation of cellular phospholipids correlates with tumor regression grade and radio resistance in rectal cancer. *109th AACR Annual Meeting, Chicago USA. April 2018.* (Appendix B)

Poynter LR, Mirnezami R, McKenzie J, Soares R, Mroz A, Peppa N, Mirnezami A, Kinross J, Takáts Z, Darzi A. Mass spectrometry imaging reveals lipidomic biomarkers of radioresistance in rectal cancer. *11th European Colorectal Congress, St Gallen, Switzerland. December 2017*

Poynter LR, Veselkov K, Galea D, Kinross J, Mirnezami A, Nicholson J, Takáts Z, Mirnezami R, Darzi A. Network-driven analytics of published tissue-based biomarkers to predict response to neoadjuvant therapy in rectal cancer. *108th AACR*

Annual Meeting, Washington DC, USA. April 2017. Published as Abstract #839, AACR 2017 Proceedings: Abstracts 1-3062

Prizes Awarded for Presented Works:

American Association for Cancer Research (AACR) Scholar-In-Training Award. *109th AACR Annual Meeting (2018), Chicago IL, USA*

Winner; Best Research Presentation. European Colorectal Congress 2017. *St Gallen, Switzerland. December 2017*

Winner; Best Research Presentation (Imperial College Division of Surgery Annual Research Afternoon 2017). *For the presentation: "Lipidomics of radioresistance in rectal cancer – a multi-centre concept study"*

Grant funding received for this work:

Awarding body: **Royal College of Surgeons of England/Saven Research and Development Programme**

Applicant: Mr Liam Poynter

Grant value: £22,000 (for consumables and project costs)

Awarding body: **Bowel Disease Research Foundation (BDRF)**

Applicant: Mr Liam Poynter

Grant value: £12,000 (for consumables and project costs)

Abbreviations

AJCC	-	American Joint Committee on Cancer
APE	-	Abdomino-perineal excision
CC	-	Colon cancer (adenocarcinoma of the colon)
CIMP	-	CpG island methylator phenotype
CMS	-	Consensus molecular subtypes
CRC	-	Colorectal cancer
DAG	-	Diacylglycerol
DESI MSI	-	Desorption electrospray ionisation mass spectrometry imaging
DNA	-	Deoxyribonucleic acid
ELAPE	-	Extralevator abdomino-perineal excision
EMVI	-	Extramural venous invasion
FFPET	-	Formalin-fixed, paraffin-embedded tissue
GSEA	-	Gene Set Enrichment Analysis
GPL	-	Glycerophospholipid
HESI	-	Heated electrospray ionisation
KEGG	-	Kyoto Encyclopaedia of Genes and Genomes
LCCR	-	‘Long course’ chemoradiotherapy
LGI	-	Lower gastrointestinal (i.e., of the anus, rectum and colon)
LPL	-	Lysophospholipid
LVI	-	Lymphovascular invasion
miRNA	-	Micro-ribosomal nucleic acid
MRI	-	Magnetic resonance imaging
MSI	-	Microsatellite instability

PA	-	Phosphatidic acid
PC	-	Phosphatidylcholine
PE	-	Phosphatidylethanolamine
PG	-	Phosphatidylglycerol
PG	-	Phosphatidylinositol
PNI	-	Perineural invasion
PS	-	Phosphatidylserine
RC	-	Rectal cancer (adenocarcinoma of the rectum)
RNA	-	Ribosomal nucleic acid
SCN	-	Somatic copy number
SCNA	-	Somatic copy number aberrations
SCRT	-	'Short course' radiotherapy
STS	-	Straight-to-surgery
TCA	-	Tricarboxylic acid (in context of TCA - or Krebs' - cycle)
TGFβ	-	Transforming growth factor beta
TNM	-	Tumour, node, metastasis staging system
TNT	-	Total neoadjuvant therapy
TPE	-	Total pelvic exenteration
UICC	-	Union for International Cancer Control
WES	-	Whole exome sequencing
WGS	-	Whole genome sequencing

Chapter 1
Introduction to Rectal Cancer

A BACKGROUND TO RECTAL CANCER 32

1.1.1 Rectal cancer as a distinct entity within the colorectal cancer landscape 32

1.1.2 Molecular basis for rectal cancer 34

1.1.3 Clinicopathological features of rectal cancer 37

1.1.3.1 Local spread 41

1.1.3.2 Lymphatic spread 41

1.1.3.3 Haematogenous spread 42

1.1.3.4 Transcoelemic spread 43

1.1.4 Prognosticating histological features and adjuvant radiotherapy for rectal cancer 43

1.1.5 The multimodality approach to rectal cancer treatment 45

1.1.5.1 Surgery for rectal cancer 47

1.1.5.2 Neoadjuvant radiotherapy for rectal cancer 49

1.1 A background to rectal cancer

1.1.1 Rectal cancer as a distinct entity within the colorectal cancer landscape

In many clinical contexts, rectal cancer (RC) is generally considered as being merely part of the continuum of colorectal cancer (CRC). CRC has recently become the second leading cause of cancer-related death, accounting for almost 10% of worldwide mortality from malignancy in 2018. [1] In the United Kingdom, over 40,000 new cases of CRC are diagnosed annually, and the European age-adjusted incidence has remained largely steady over the course of the last 20 years, at around 70 cases per 100,000 population. [2] One suggested reason for this, is the established uptake of lower gastrointestinal (LGI) fibreoptic endoscopy (sigmoidoscopy and colonoscopy), which has permitted the earlier detection of dysplastic lesions and concurrent polypectomy. Consequently - and in line with CRC as a whole - survival from RC in both the UK and equivalent high- to middle-income countries has gradually improved over the last decades, up to a 5-year survival of around 65%. [3] [4] The growing and aging population in the United States, for example, is believed to be a key factor in the estimate that the number of survivors from RC will increase by up to 25% in the next decade. [4]

The underlying principle of the natural history of RC is that it is - for most individuals - a series of aberrant events whose trigger is not pre-defined by inherited genetic susceptibility. Unlike certain cancer subtypes (for example, lung cancer), no single causative environmental factor has been identified, although epidemiological studies have validated the increased risk posed by a number of lifestyle pressures. [5] Obesity, the current major health epidemic, now accounts for a stepwise increase in

CRC incidence in successive birth cohorts. [6] Other established risk factors include diabetes, [7] male gender, [8] smoking, alcohol, [9] excessive red meat consumption and inadequate dietary fibre. [10] In the Western world, many of these risks will co-exist, making univariate correlation challenging. [11] More recently, an expansion in our understanding of the implications of individual variation in the gut microbiome has demonstrated the effect this appears to have on not only the development of CRC, but also modulation of response to both surgical and oncological therapies. [12-16]

The presence of inflammatory bowel disease (ulcerative colitis and Crohn's colitis) and established family history of CRC both confer significantly increased relative risks. [17] [18] Furthermore, despite most cases of CRC being sporadic events, twin and family studies provide estimates of between 12 and 35% for some element of heritability. [19] [20] Approximately 5% will be affected by a defined hereditary CRC syndrome, with *hereditary non-polyposis colorectal cancer* (HNPCC; Lynch Syndrome) and *familial adenomatous polyposis* (FAP) being the commonest accountable genetically inherited conditions. [21]

As a 'surgically curable' cancer, the anatomical distribution of CRC is classically considered in relation to its blood supply and hence, the operative resection that might be performed to effect complete oncological clearance. Broadly, cancers of the colorectal tract are referred to as either right colon cancers (CC), left colonic cancers, sigmoid colon cancers or rectal cancers (RC). Current data demonstrate that up to a third of all cases of CRC occur in the rectum, despite the incidence of right-sided CC increasing steadily over the past four decades. [22] [23] Owing to its inherent oncogenesis, natural history and the anatomical challenges it poses for surgical clearance, RC is

increasingly being considered as a separate entity from CC. It is these characteristic features of the natural history of RC which largely guide the current multimodal approach to the therapeutic strategy.

1.1.2 Molecular basis for rectal cancer

The development of RC results as a sequence of genetic and epigenetic alternations in the glandular epithelium of the rectal mucosa, which transform normal epithelial cells within to cancer. This process occurs through a series of stages of aberrant cellular growth and behaviour which manifest initially as a diminutive adenomatous polyp, thence though successive degrees of dysplasia to invasive cancer. Taking place over the course of (on average) 10 years - and equally applicable to CC - this *adenoma-adenocarcinoma sequence* has historically been held as the core principle in the pathophysiology of sporadic cases of RC. The initiating event is believed to be a mutation in the adenomatous polyposis coli (*APC*) gene, which is present in around 70% of all sporadic colon and rectal adenomas. [24] Further promotion of *KRAS* and p53 mutations, in combination with chromosomal instability (CIN), drive the adenoma-adenocarcinoma sequence towards invasive cancer.

Increased understanding of the molecular biology of both CC and RC, through aggregated bioinformatics-based analysis across multiple CRC gene datasets, has led to the description of four, distinct, consensus molecular subtypes (CMS), summarised in **table 1.1**. The landmark work by Guinney and colleagues demonstrated the above-described pathway (driven by mutations in *APC*, marked activation of WNT and MYC signalling and high levels of somatic copy number

alteration (SCNA)) as the ‘canonical’ CMS2 subtype, accounting for approximately 38% of all colon and rectal cancer. CMS1 tumours account for 14%, displaying high levels of microsatellite instability (MSI), *BRAF* mutations and marked immune infiltration, and are further characterised by genomic hypermethylation which results in ‘switching off’ of a variety of tumour suppressors, described as the CpG island methylator phenotype (CIMP). [25] CMS3 tumours (13%) are characterised by significant metabolic dysregulation. *KRAS* mutations are almost ubiquitous, but they display mixed MSI status and low CIMP. Finally, the CMS4 subtype (23%) is characterised by high stromal infiltrates, transforming growth factor-beta (TGFβ) activation and significant angiogenesis. CMS1-4 are referred to as MSI immune, canonical, metabolic and mesenchymal respectively. [26] Furthermore, these subtypes tend to demonstrate anatomical distribution, with RC being more commonly of the CMS2 or CMS4 phenotypes **(figure 1.1)**.

CMS subtype	CMS 1 'MSI immune'	CMS 2 'Canonical'	CMS 3 'Metabolic'	CMS 4 'Mesenchymal'
Percentage of CRC cases	**14%**	**37%**	**13%**	**23%**
Genomic features	MSI High CIMP Hypermutated	High SCNA	Mixed MSI status Low SCNA Low CIMP	High SCNA
Hallmark mutations	*BRAF* mutations	-	*KRAS* mutations	-
Cellular processes and micro-environment	Immune infiltration and activation	WNT and MYC activation	Dysregulation of metabolic processes	Stromal infiltrates Angiogenesis TGFβ activation
Clinico-pathological features	Worse OS in relapse	-	-	Worse DFS Worse OS

Table 1.1 - Consensus molecular phenotypes (CMS) for colon and rectal cancers as postulated by Guinney et al. MSI, microsatellite instability; SCNA, somatic copy number alterations; CIMP, CpG methylator island phenotype; TGFβ, transforming growth factor-beta; DFS, disease-free survival; OS, overall survival

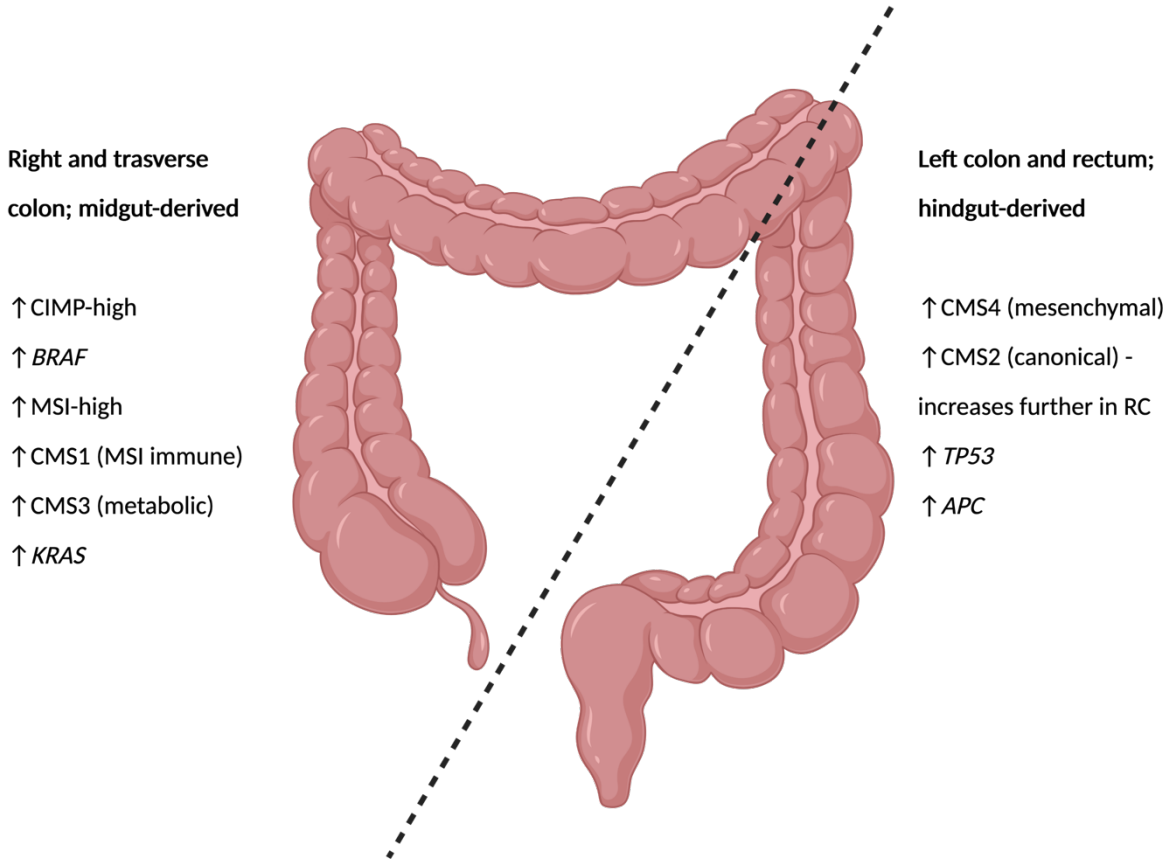


Figure 1.1 – Summary figure of CMS distribution and core genetic features of tumours according to embryological development of the colon and rectum. A mixed CMS phenotype may be present in up to 13% of cases of CC and RC. Adapted from Dekker et al. [22]

1.1.3 Clinicopathological features of rectal cancer

Once adenocarcinoma has been established, the patient's symptoms may be so subtle as to delay presentation for many months. The classically described, cardinal symptoms of rectal cancer include bleeding *per rectum,* a change in bowel habit (usually to diarrhoea, and typically early in the morning) and a feeling of incomplete evacuation (tenesmus). [27] Delays in presentation are often marked by more advanced disease, whereby the cancer has progressed beyond the confines of the rectal wall.

The Dukes' staging system, despite its common usage to contextualise the extent of disease in the clinical context, is now mostly a historical system and does not have a significant role in the stratification of rectal cancers in modern management (**figure 1.2**). The Tumour-Node-Metastasis (currently UICC TNM8) classification is now the internationally recognised standard as the optimum classification for histopathological tumour staging (**table 1.2**). [28] Furthermore, accurate pre-operative MRI staging guides multimodal therapies. In addition to features such as extramural venous invasion (EMVI), tumour deposits and involvement (or threatened involvement) of the circumferential resection margin (CRM), staging MRI can identify other prognosticating features not described in TNM8. In particular, with regards to T-stage, MRI assessment permits the subdivision of T3 tumours dependent on the depth of invasion beyond the muscularis propria as follows:

- T3a - <1mm extension beyond the muscularis propria
- T3b – 1-5mm extension beyond the muscularis propria
- T3c – 5-15mm extension beyond the muscularis propria
- T3d – 15mm extension beyond the muscularis propria

Although this subclassification is not incorporated into any version of TNM to date, it has been shown to be of clinical value with regards to survival outcomes. Early reports suggest that there is significant variability in the natural history and outcomes of T3 stage III rectal cancer compared to T2 stage III rectal cancer. [29] Repeated meta-analysis of outcomes following subdivision of T3 tumours into T3a-b and T3c-d demonstrates prognostic relevance to overall survival, disease-free survival and cancer-specific survival. [30 31]

Locally advanced rectal tumours (i.e., those breaching the rectal wall), demonstrate invasion and spread according to a number of means, and these features impact significantly upon the clinical treatment pathway.

Primary tumour (pT)

pTx Primary tumour cannot be assessed
pT0 No evidence of primary tumour
pT1 Tumour invades submucosa
pT2 Tumour invades muscularis propria
pT3 Tumour invades into subserosa or into non-peritonealised pericolic or perirectal tissues
pT4 Tumour breaches serosa and:
- **4a** perforates visceral peritoneum and/or
- **4b** directly invades other organs or structures

Regional lymph nodes (pN)

pNx Regional lymph nodes cannot be assessed
pN0 No regional lymph node metastatic disease
pN1 Metastatic disease in 1–3 regional lymph nodes
- **pN1a** Metastasis in 1 regional lymph node
- **pN1b** Metastases in 2–3 regional lymph nodes
- **pN1c** Tumour deposit(s) in the subserosa, or in non-peritonealised pericolic or perirectal soft tissue without regional lymph node metastatic disease

pN2 Metastatic disease in 4 or more regional lymph nodes
- **pN2a** Metastases in 4–6 regional lymph nodes
- **pN2b** Metastases in 7 or more regional lymph nodes

Distant metastatis (pM)

pM0 No evidence of distant metastatic disease
pM1 Distant metastatic disease
- **pM1a** Metastasis confined to one organ without peritoneal metastases
- **pM1b** Metastases in more than one organ
- **pM1c** Metastases to the peritoneum with or without other organ involvement

Table 1.2 – Current iteration (8th Edition) of the *Tumour-Node-Metastasis* final pathological staging system for colon and rectal cancer, as per UICC TNM8. Post-neoadjuvant radiotherapy there is a further ‘y’ prefix, i.e. ypT3 and a supplementary note on tumour regression grading (TRG). [28]

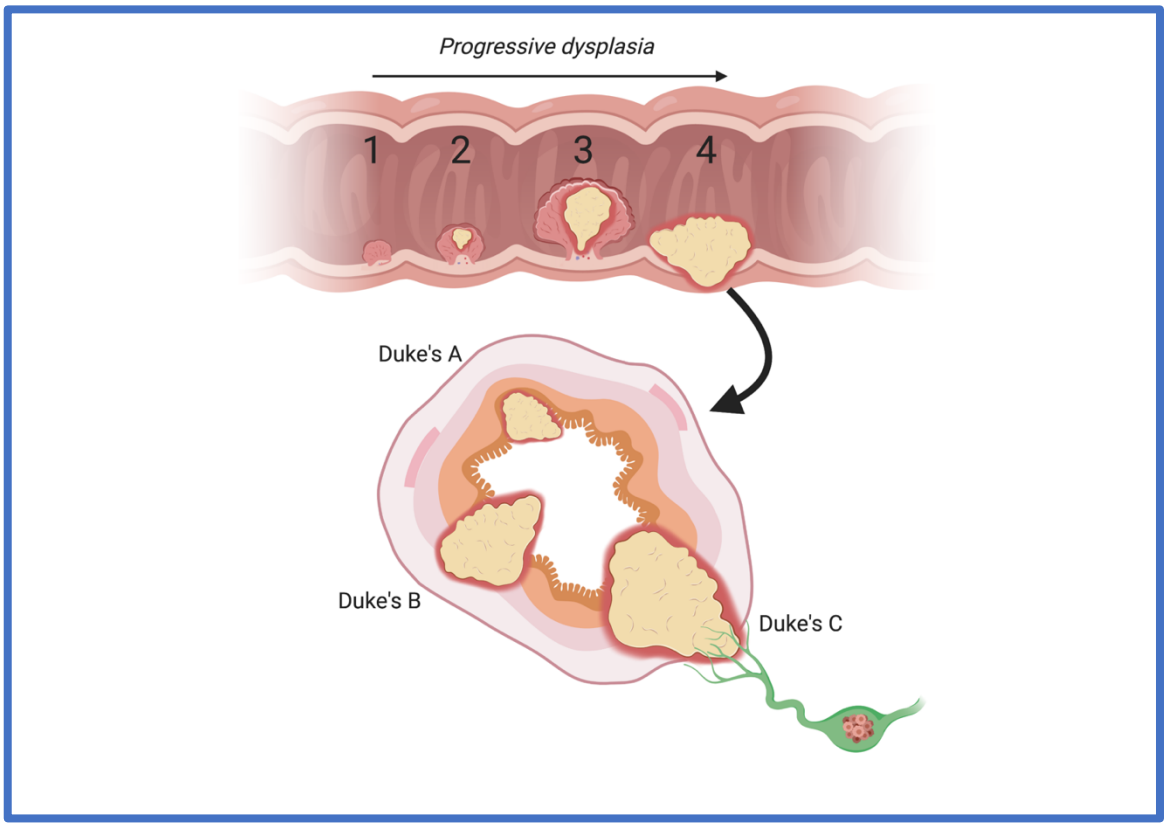


Figure 1.2 - The adenoma-adenocarcinoma sequence; (1) diminutive, benign polyp (tubular, villous or tubulovillous adenoma; (2) high grade dysplasia through (3) carcinoma in-situ (polyp cancer); (4) invasive adenocarcinoma. The Duke's Classification was, for many years, in common usage to describe staging of colon and rectal cancer in a concise manner. Duke's A, tumour extending to - but not breaching - the muscularis propria; Duke's B, tumour extending to the serosa; Duke's C, tumour breaching serosa with involvement of local lymph nodes. Further subdivided into Duke's C1 - involving local draining mesorectal/mesocolic nodes – and Duke's C2 when apical node involved. "Duke's D" (not shown, and not originally described) implies ***D***istant metastasis. This classification has undergone many alterations and reiterations since its original description in 1932 and has now largely been entirely superseded by the TNM system, following extensive clinical research into the prognosticating features of individual tumours in relation to their location and depth of invasion.

1.1.3.1 Local spread

Whereas CC exhibits both circumferential and longitudinal patterns of spread, spread of RC occurs circumferentially. [32] Following T4 breach of the seromuscular coat the tumour spreads into the surrounding mesorectum, thereafter breaching the visceral peritoneum in T4a or direct invasion of adjacent organs in T4b. Initially, this degree of spread is controlled by the fascial coverings of the mesorectum, but further invasion beyond this envelope may involve a number of surrounding anatomical structures, presenting challenges in complete surgical clearance of the tumour.

1.1.3.2 Lymphatic spread

Lymphatic drainage of the mesorectum follows the accompanying veins, and almost invariably occurs in an upwards direction towards the lymphatic chain accompanying the inferior mesenteric vascular pedicle. Tumours within the territory of the middle rectal vessels may exhibit radial spread, with the middle and inferior rectal lymphatics following the internal iliac chain to the pelvic sidewall. [33] Indeed, the Japanese guidelines (JSCCR) for management of tumours below the peritoneal reflection includes lateral lymph node dissection. [34] Metastasis via the inferiormost lymphatics to the groin nodes is very infrequently seen. [35] Any degree of lymphatic invasion has historically been considered an independent, adverse prognostic feature of the tumour. Clinical detection of this in the pre-treatment phase is challenging and generally only heralded by the identification of pathological lymph nodes on dedicated MRI. Although MRI detection of nodal spread at the diagnostic stage invariably

mandates neoadjuvant therapy, MRI staging of lymph nodes is increasingly seen as being prognostically inaccurate. [36] [37] There is now increasing evidence, due to improvements in MRI protocol and image interpretation, that lymphatic and nodal involvement may in fact not be as critical a prognostic feature as, for example, tumour deposits (TD) and extramural venous invasion (EMVI) detected by MRI (i.e., mrEMVI). [38] [39]

1.1.3.3 Haematogenous spread

The venous drainage of the rectum forms a plexus within the mesorectal fascia, with the majority of the drainage from the upper and middle thirds passing into the inferior mesenteric vein and thence to the liver via the portal circulation. As noted previously, the middle and inferior rectal veins (the latter first via the internal pudendal vein) drain into the systemic circulation via the internal iliac vein. Hence, it follows that the commonest sites for blood-borne metastasis are the liver (34%), lungs (22%) and adrenals (11%). EMVI is now recognised as an independent prognosticator of RC; many published datasets now demonstrate this to be a more important factor than lymph node (N) status. [38-41] Furthermore, mrEMVI is increasingly amenable to detection by dedicated protocols. However, despite this emerging evidence as to impact on prognosis, TNM8 does not take EMVI into account and so is not a core data item *per se* when guiding decisions for pre-operative radiotherapy.

1.1.3.4 Transcoelemic spread

This does not tend to be a typical feature of rectal cancers, except for those that arise in the upper rectum and demonstrate cephalad extension above the peritoneal reflection. When it does occur, however, spread may follow a route throughout the visceral and/or parietal peritoneum if a T4 primary tumour were to breach the peritonealised surface. Management of these tumours is challenging, and if primary surgery is even feasible may involve multivisceral resection or total pelvic exenteration (TPE), including cytoreductive surgery and use heated intraperitoneal chemotherapy (HIPEC) in highly selected cases. [42]

1.1.4 Prognosticating histological features and adjuvant radiotherapy for rectal cancer

Following surgical resection, there follows a core dataset of reporting measures on the histological specimen which both determines the final pathological staging of the cancer, and whether further (adjuvant) treatment is required in the form of systemic chemotherapy. In the United Kingdom, this core dataset is set out by The Royal College of Pathologists (RCPath) and is based upon consensus data from the Union for International Cancer Control (UICC) TNM system – currently in its 8th iteration, TNM8 – and the American Joint Committee on Cancer (AJCC). [43] The core data items are as those previously summarised in **table 1.2**. In addition to the histological features which are independently associated with a poor prognostic outcome, a number of molecular profiling approaches are now used at this stage to guide appropriateness of

targeted monoclonal antibody agents (typically, anti-epidermal growth factor (EGFR) therapy). These include mismatch repair (MMR) status (measured by either microsatellite instability (MSI) or MMR immunohistochemistry (IHC)), as well as mutational testing for specific codons of the *KRAS, NRAS, BRAF* and *PIK3CA* genes. Although not core data items, several other biomarkers have been identified as having prognostic value and are variably reported between units. Examples include the caudal-related homeobox transcription factor 2 (*CDX2*), which has been demonstrated to be linked to CRC outcomes. [44 45] Decreased or absent expression of its associated CDX2 protein is linked to advanced tumours features (advanced T stage, vascular invasion and distant metastasis). Furthermore, a marked benefit to 5-year survival has been demonstrated in patients with CDX2-negative stage II colon cancer, when these patients were treated with adjuvant chemotherapy. [46] Presence of tumour-infiltrating lymphocytes, tumour budding and tumour stromal volume are also variably reported in addition to the core data items stipulated by both TNM and the Royal College of Pathologists (RCPath) in the UK, having been demonstrated to be of prognostic significance. [47-52] *Radiomics* approaches (i.e., extraction of quantitative features from staging MRI scans) have also demonstrated promise in detection of these features at the pre-treatment stage. [53 54]

1.1.5 The multimodality approach to rectal cancer treatment

In the UK, contemporary management of RC takes place within a multidisciplinary team (MDT) structure consisting typically of colorectal surgeons, medical and clinical oncologists, radiologists, pathologists, specialist nurses and palliative care physicians. Diagnosis is confirmed by means of clinical examination, flexible lower GI endoscopy and biopsies for histopathological analysis. TNM staging is then established by both CT scanning of the chest, abdomen and pelvis to rule out presence of distant metastases, and MRI scanning (for preference) of the rectum. MRI permits high-definition assessment of the loco-regional extent of disease and defines the predicted circumferential resection margin (CRM). In some challenging instances, this may be contraindicated (for example with certain types of pacemaker or implanted metallic devices), in which case endo-rectal ultrasound scanning may need be used as an alternative, with isolated studies demonstrating equivalence to MRI in expert hands. [55-57]

The presence of distant metastases will prompt the involvement of other agencies - for example liver or thoracic surgeons – if consensus dictates a chance for surgical clearance. However, for the majority of cases of RC, treatment options will fall into three categories:

(1) **Straight to surgery (StS)**: where no distant metastases are identified, and the radiological tumour characteristics are suggestive of a RC confined either the submucosa or muscularis propria (T1 and T2 respectively) and hence no involvement of threatening of the CRM. In many centres, for low-volume and low-risk T3 tumours tend to be treated in this manner. Adjuvant chemotherapy may be indicated after

surgery, based on final histopathologic analysis of the resected specimen which takes into account a final assessment of lymph node metastases and adverse molecular features of the tumour.

(2) **Neoadjuvant therapy**: The accurate definition of CRM involvement has been demonstrated as a critical stratifier for treatment in rectal cancer – the MERCURY study demonstrated that identification of a clear CRM at staging MRI means that primary surgery alone will confer good outcomes, whereas CRM involvement mandates neoadjuvant radiotherapy for downstaging. [58] The mainstay of neoadjuvant therapy is fractionated radiotherapy, delivered most commonly in the UK, USA and some European countries as 'long course' chemoradiotherapy (LCCR). This entails 50.4Gy delivered in daily fractions with a 5-fluorouracil (5-FU) -based chemotherapy agent used in tandem as a radiosensitising agent. In comparison to 'short course' radiotherapy (SCRT – 5Gy/day, 5 days), the treatment goal is regression of the tumour such that the CRM no longer appears involved or threatened on post-treatment MRI. This consequently decreases the likelihood of local recurrence, as demonstrated in a number of trials. [59 60] The degree of regression varies considerably between individuals, with some cancers disappearing completely (pathological complete response, pCR) and others progressing locally or metastasising in spite of treatment, such that further therapies can no longer be delivered with curative intent. By contrast, in the UK SCRT is selected for patients deemed unlikely to tolerate the prolonged treatment of LCCR or to decrease the risk of local recurrence where formal downstaging is thought to be unnecessary, although in Scandinavia and much of Europe SCRT remains the neoadjuvant strategy of choice.

In recent years, there has been increased interest in the application of long course chemoradiotherapy, especially as a mainstay of treatment in organ-preserving strategies such as the 'Watch and Wait' strategies as employed by Habr-Gama and colleagues, [61-66] or SCRT with local excision and avoidance of conventional TME surgery as in both the reported TREC trial and the awaited STAR-TREC trial. [67 68]

(3) **Palliation**: in the context of such disseminated disease such that curative treatment is not feasible. Such an approach may still involve surgery (faecal diversion to prevent obstruction, by way of colostomy formation), chemotherapy (to attempt to decrease the systemic tumour burden) and radiotherapy (for localised tumour control, and control of symptoms such as rectal bleeding from large exophytic tumours).

1.1.5.1 Surgery for rectal cancer

Whether as a primary course of action or following LCCR, the mainstay of curative treatment for rectal cancer remains resectional surgery. Contemporary minimally invasive approaches continue to employ the principles of total mesorectal excision (TME) as popularised by Heald, for complete clearance of the mesorectal lymph nodes within their fascial envelope. [69-71] This approach allows for optimal oncological clearance whilst also retaining the ability to anastomose the proximal and distal resection margins and preserve the anorectal sphincter complex. Tumours at the dentate line have traditionally been managed with complete excision of the anal canal and sphincters with formation of a permanent end colostomy (abdomino-perineal excision, APE). However, multiple groups have reported favourable oncologic and functional outcomes with intersphincteric resections (ISR) for these tumours,

coupled with hand-sewn colo-anal anastomosis, in order to avoid permanent end colostomy. [72-74] Many surgeons will still advocate APE for these tumours, and certainly this is the accepted approach for RC distal to the dentate line; they are too distal to be able to provide adequate oncological clearance and still preserve the external anal sphincter. This may be the case even where LCCR has been employed in an attempt to downstage the tumour. In cases of more lateral spread beyond the mesorectal fascia or involving the levators, an extralevator approach may be employed (ELAPE) to improve the chances of an intact specimen with clear resection margins. [75-77] Involvement of other components of the pelvis (bladder, vagina, pelvis sidewalls) need careful consideration for total pelvic exenteration (TPE) with *en bloc* removal of all affected compartments.

Laparoscopic surgery has generally superseded the traditional open surgical approach in a vast number of cases, with regards to shorter inpatient stays and lower morbidity when combined with modern anaesthetic strategies and enhanced recovery (ERAS) protocols. [78-80] Although a 2014 Cochrane review attempted to address ongoing controversy regarding non-inferiority compared to open TME as regards to oncological outcomes, [70] in both the ACOSOG Z6051 and ALaCaRT trials the use of laparoscopic TME versus open TME failed to satisfy the criteria for non-inferiority with regards to pathological outcomes. However, the short- and long-term follow-up data from the landmark MRC-CLASICC, COLOR II and COREAN trials all reported clear benefits to the approach (although it should be noted that MRC-CLASICC included colon as well as rectal cancers). [70 81-88] With regards to robotic TME, the ROLARR randomised control trial has only shown non-inferiority as compared to laparoscopic

surgery, with no significant benefits demonstrated in either patient recovery, morbidity or oncological clearance. [89]

More recently there has been increased interest in less invasive, organ-preserving strategies for rectal cancer treatment. The true incidence of pathological complete response (pCR - such that no residual tumour is identifiable neither on post-treatment MRI nor on histological examination following resection) has been variously reported as being in the region of between 10% to 31% in both experimental data and meta-analysed registry data, albeit with the question raised about selection bias in experimental data cohorts. [90-95] There is ongoing debate as to the best way to manage such patients, particularly in the context of there being no validated, prognosticating biomarkers to either (a) predict response to LCCR, nor (b) to exonerate involvement of the mesorectal lymph nodes, nor (c) to accurately monitor absence of disease. In the context of T1 and T2 tumours, some groups have reported on outcomes of minimally invasive transanal approaches such as transanal endoscopic microsurgery (TEMS) or transanal minimally invasive surgery (TAMIS). [68] The long-term outcomes of these cases are eagerly awaited from further high-quality trials such as STAR-TREC, which are yet to report.

1.1.5.2 Neoadjuvant radiotherapy for rectal cancer

The mainstay of neoadjuvant therapy for RC centres around delivery of radiotherapy by linear accelerators, which precisely target the tumour with minimal involvement of surrounding normal tissues. As discussed previously, the primary aims of providing LCCR are to downstage the tumour to facilitate complete surgical

resection and to reduce the risk of pelvic recurrence, whereas with SCRT the aim is to ‘sterilise’ the surgical field. Both Cochrane and other systematic reviews have repeatedly demonstrated that both pre-operative and post-operative radiotherapy reduces the risk of local (pelvic) recurrence compared to surgery alone. [96-98] In the USA, early clinical trials focussed on postoperative (i.e. adjuvant) chemoradiation. [99] In Europe, both Stockholm I and The Swedish Rectal Cancer Trial reported improved outcomes in a cohort of 1168 patients undergoing SCRT. [100] [101] These early trials preceded the dramatic improvements which were being made in surgical technique following the proposition of TME surgery as standard by Heald, and a number of landmark trials followed. The subsequent Dutch TME and MRC CR07 trials confirmed superiority of pre-operative (neoadjuvant) chemoradiotherapy compared to post-operative (adjuvant) chemoradiotherapy, and a decrease in local recurrence rates with all forms of neoadjuvant radiotherapy (SCRT and LCCR). [59] [102] [103] The addition of concurrent 5-FU agents to long course radiotherapy has also demonstrated a reduction in local recurrence compared to radiotherapy alone. [104] [105] The recently reported RAPIDO trial reported a decreased probability of disease-related treatment failure in the experimental arm which concluded that the provision of pre-operative chemotherapy in addition to RT is thought to convey benefit by early treatment of micrometastatic disease (given that late, distant metastases continue to be the major cause of death following treatment for rectal cancer). [106]

The optimal timing of subsequent surgery after LCCR or SCRT remains the subject of discussion. Following LCCR, a period of 6-8 weeks is generally observed, and still largely based upon data from the Lyon R90-01 trial conducted in 1999. [107] [108] However, a meta-analysis of 13 more recent trials confirmed a long-held suspicion

that a cumulative effect continues to be observed after the standard 6 to 8-week delay prior to surgery, with pCR observed in 19.5% vs 13.7% in the control (standard interval) group, and with one study employing a post-LCCR delay of up to 14 weeks - albeit in a small cohort (*n* = 33) in a prospective series. [109] [110]

The traditional regime for SCRT in the landmark trials is an administration over the course of 5 days to 'sterilise' the surgical field, with surgical resection of the tumour the following week. Classically this approach does not yield much in terms of histological tumour regression at subsequent analysis. However, similarly to the improvements in TRG and pCR rate seen in the LCCR studies above, the Stockholm III trial has reported improved tumour regression and oncological outcomes in the SCRT arm, where a delay of 4-8 weeks to surgery was included, with a pCR of 11.8% (vs. 1.7% when following the standard SCRT regime). [111]

These significant improvements in pCR rate following a longer delay after administration of RT highlight a number of possibilities which have been the focus of a number of UK trials in recent years, including OnCORe, TREC and STAR-TREC. [67] [68] [112] It stands to reason therefore that the real source of interest in neoadjuvant RT is the concept for adding personalised rectal cancer treatment strategies to the management armamentarium - notably organ-preservation - be that through total neoadjuvant therapy (TNT) or by localised excision of tumour following LCCR (as proposed for tumours up to T3b in the STAR-TREC trial). [113] Pooled meta-analysis also demonstrated the higher rates of pCR with LCCR versus RT alone. The CARTS study reported 92% of the 51 patients recruited with T1-T3 tumours at staging to have a response sufficient for subsequent TEMS, with 64% found to subsequently be ypT0-

T1 with negative margins. The question of pCR has further been a focus of the previous GRECCAR-2 [114] [115] and GRECCAR-6 trials, [116] with the upcoming GRECCAR-12 due to examine TNT in T2N0 patients without subsequently offering TME. However, as in the various series reported by the Habr-Gama 'Watch & Wait' groups, crossover from the local excision (i.e. TEMS) to formal TME was observed, in the order of 15% in the acute setting. [66] [117] [118] This further increased following rigorous surveillance and subsequent late recurrences. [119]

Critical to any decision to stratify patients to such strategies is a robust understanding of the likely response to RT. This serves two aims: firstly, that patients who are likely to experience a pCR are appropriately selected and counselled, and secondly that those who are set to receive minimal benefit from an extended period of LCCR and a long post-RT delay do not have definitive TME surgery postponed, thereby putting them at risk of early disease progression or distant metastasis whilst undergoing LCCR. Furthermore, for those patients who experience a pCR or a response so significant as to permit scar excision by TEMS or TAMIS, there needs to be a consensus strategy for surveillance, coupled with a validated, prognosticating biomarker or biomarkers to confirm absence of disease.

This thesis aims to explore further the underlying facets of the biology of response to radiotherapy from a basic science standpoint; the critical unmet need has, as yet, not been solved by traditional approaches and so in this work I have set out to employ a multi-omics approach to the problem in such a way that has not yet been attempted. In the following chapter I provide an overview of some of these various -omics approaches, justify a case for using the tumour lipidome as a surrogate endpoint

for cellular processes and describe in detail some of the technical aspects of the technologies that will become relevant in the experimental sections of my work.

Chapter 2
The *-omics* landscape in discovery science - a technical introduction to the relevant modalities

2.1 AN OVERVIEW OF REALISED AND POTENTIAL '-OMICS' APPROACHES IN RECTAL CANCER RESEARCH 55

2.1.1 Genomics 55

2.1.2 Transcriptomics 56

2.1.3 Proteomics 56

2.1.4 Metabolomics 57

2.1.5 The gut microbiome – "microbiomics" 59

2.2 METABOLISM AND METABOLIC DISRUPTION IN CANCER 62

2.2.1 The Warburg effect 62

2.2.2 Lipids in cell biology, colon and rectal cancer 65

2.3 MASS SPECTROMETRY-BASED LIPIDOMIC PROFILING 70

2.3.1 Introduction 70

2.3.2 Basic principles of mass spectrometry 71

2.3.3 Fundamentals of the Orbitrap mass spectrometer 73

2.3.4 Electrospray ionisation (ESI) and desorption electrospray ionisation (DESI) 74

2.4 DATA ANALYSIS APPROACHES FOR BIOLOGICAL MASS SPECTROMETRY DATASETS 78

2.5 THE ROLE FOR A LIPIDOMICS APPROACH AS PART OF A MULTI-OMICS ANALYSIS OF THE RADIOTHERAPY RESPONSE IN RECTAL CANCER 81

2.1 An overview of realised and potential solid-tissue '-omics' approaches in rectal cancer research

Traditionally, cancer research has been undertaken in a top-down approach (genes, transcriptome, proteins). Whereas findings from these approaches have yielded significant leaps in our understanding of the genetic basis for cancer - in particular, oncogenesis and dysregulation of cellular cycle processes and proliferation - subsequent attempts to connect the resulting outputs in a biologically meaningful way has proven challenging. This is especially true of experiments performed at different timepoints and under conditions that limit the validity of translating findings to an *in vivo* model. However, with the introduction of a full range of *-omics* approaches in recent years, we can now study every timepoint of cellular processes (and, crucially, obtain a more biologically relevant measure of the endpoints of cellular metabolism products) in far greater detail than ever before. These approaches are far-reaching and reliant on technological advances to extract quantitative features from biological specimens, be they biofluids, solid tissue or even radiological images in the field of radiomics. [120] It would be beyond the scope of this work to cover these all in full. As the basis of the experimental work that follows is on solid tissue, the current complementary modalities fall broadly into the categories that follow.

2.1.1 Genomics

Next-generation sequencing technologies for analysis of DNA currently fall broadly into either *whole exome sequencing* (WES) or *whole genome sequencing* (WGS). [121] WES is used to identify variance in the coding portion of the genome, i.e.,

the exomes. WGS on the other hand, covers the entire genome and is better placed to identify non-coding variants in addition to exomal sequences. [4] [122] [123] Whereas WES already has a role in human disease clinical practice, the costs of WGS have historically been prohibitive. However, as WGS pipelines have become increasingly streamlined, high-throughput, low-coverage sequencing can now be performed which strikes a balance between meaningful data output and lower cost.

2.1.2 Transcriptomics

Next-generation RNASeq modalities are set to replace microarrays as the modality of preference for high-throughput transcriptomic analysis. This is based on the principles that RNASeq are more reproducible and rely on replication/sequencing rather than the hybridisation approach in microarrays, which can be highly variable dependent on the platform used. [124] More recent advances have linked fluorescent in-situ hybridisation (FISH) approaches to visualise mRNA within cancer tissues, such as with the BaseScope and RNAScope technologies. [125] [126]

2.1.3 Proteomics

Protein identification has historically been inherently more complex than either genomics or transcriptomics, due in part to the existence of a multitude of post-transcriptional modifications, thus rendering structural elucidation challenging. Methods for the detection of known proteins - including structural elements and enzymes - broadly include immunoassays, metabolomics-based platforms (nuclear magnetic resonance spectroscopy (NMR) and mass spectrometry (MS) approaches)

or antigen microarrays. More recently, advances have been made to link immunoassays to mass spectrometry platforms. Of note here is the Hyperion™ Imaging Mass Cytometry (IMC) system from Fluidigm, which relies on conjugating lanthanide metal-ions to the antibody probes, which are then detected by a time-of-flight (TOF) mass spectrometer after tissue ablation and laser desorption. This has demonstrated promise in performing multiplexed (in the order of 40 unique antibodies per analysis, per sample), quantitative immunoassays for enzyme and structural protein detection. This has demonstrated a paradigm shift in discovery science and a number of high-impact studies have recently been published based on the insights obtained. [127] [128]

2.1.4 Metabolomics

Generally accepted as an interchangeable term with *metabonomics*, this is defined as the systematic study of dynamic metabolic shifts that occur in any living system, be that as a response to drugs, environmental stimuli or disease. Metabolomics therefore involves the detection and measurement of products of cellular metabolic processes. [129] This field has evolved around the use of analytical chemistry techniques with the ability to detect and identify a vast range of small molecules including amino acids and lipids. The Human Metabolome Database (HMD) currently contains records for over 114,000 human metabolites (known, expected and predicted). [130] This compares to the circa. 20,000 protein-coding genes in the human genome. [131] Due to post-translational modification, it is estimated there could be as many as tens of millions of end products in the human proteome. [132] [133] Compared to proteomics, this makes the prospect of metabolomics as a modality for the study of

human disease and treatment response attractive, even if only on the basis of data processing. As an approach to understanding systems biology, metabolomics can be seen as more than just another field within a top-down overview. Rather, acting as it does as a metabolic 'footprint' as it measures the endpoint of cellular metabolic processes, it takes into account not only the function of the implicated genes but also all post-translational modifications. In the context of understanding the mechanisms of response and resistance to therapies, therefore, it has the potential to provide a unique insight into a hostile and completely aberrated cellular environment resulting from both the activity of the primary cancer and the therapy applied. This has, hitherto, remained unstudied in rectal cancer research.

In recent years, established platforms such as nuclear magnetic resonance spectroscopy (NMR) and mass spectrometry (MS) have been modified significantly for the study of *in vivo* and *ex vivo* human tissues and biofluids to include analysis of solid cancer tissue (e.g., magic angle spinning NMR, MAS NMR) and spatially-aligned analysis of tissue sections (e.g., desorption electrospray ionisation mass spectrometry imaging, DESI MSI and matrix-assisted laser desorption ionisation, MALDI). These techniques complement the more traditional approaches of liquid and gas chromatography MS (LCMS and GCMS respectively). Both MALDI and DESI MSI present attractive options for the study of metabolism in solid human cancer tissue, providing a snapshot of a metabolic endpoint in a tissue section that can be analysed in context of histological architecture and associated metadata. The role for the consideration and analysis of cellular lipids ('lipidomics') are explored further in section 2.3 and a technical overview of the MS platforms follows in section 2.4.

2.1.5 The gut microbiome – "microbiomics"

The human gut comprises a diverse and metabolically active ecosystem in its own right, harbouring in the region of 3.8 x10^{13} bacteria. [134] In recent years there has been a surge in studies pertaining to the human-associated microbiome and the metagenome in relation to human disease, including cancer. In contrast to the approximately 20,000 genes in the human genome – which is both inherited and relatively static throughout life – the gut microbiome in particular is a continually changing and evolving landscape, comprising as many as 20 million genes. Whereas non-pathogenic gut bacteria were previously believed to be effective bystanders, it is now increasingly understood that the metagenome of the gut microbiome may directly influence the regulation of the host genome, and thus play a significant role in physiology and the development of human disease, not least cancer. [135] [136]

Studies that have attempted to directly link diet to CRC development have required very large cohorts of subjects, with the implication being that is it the effect of the host gut microbiota that exerts the environmental pressure to drive towards higher susceptibility in Western populations, with direct reference to a diet lower in composition of dietary fibre and relatively high levels of red meat. [137] [138] Protective effects are thought to be influenced by fermentation products produced by the activity of the gut microbiota on dietary matter that reaches the colon, notably by short-chain fatty acids (SCFAs) produced from fibre fermentation. [139] [140] In addition to providing a preferred bacterial energy source for colonocytes, these metabolites (notably, butyrate) in turn impart a number of anti-inflammatory effects on colonic microenvironment through inhibition of NFκB. [141] Indeed, further studies have demonstrated that a reduction in a beneficial bacterial composition results in a

decreased butyrate load which is inversely correlated with rate of tumour generation. [138] [142] [143] Of the specific taxa that have been consistently demonstrated to exert a drive towards CRC oncogenesis, the anaerobic oral pathobiont *Fusobacterium nucleatum* has been most consistently and significantly associated with RC. [144-146] Of interest there appears to be evidence of a symbiotic interplay between established CRC and such pathobionts, whereby following surgical resection of CRC, a shift back towards a 'normal/healthy' gut microbiome is observed. [147]

A plausible argument can therefore be made for the role of gut microbiota in therapy response modulation and resistance. To date a few studies have extrapolated the above model of colonic gut health and carcinogenesis to evaluate response to chemotherapy agents, with relative inhibition of those butyrate-producing colonies and upregulation of *Fusobacteria spp.* correlating with increased degrees of chemoresistance. [148] [149] There remains, however, a complete paucity of published studies evaluating the role for the gut microbiome modulation of the response to radiotherapy.

In the era of next-generation sequencing and to account for diverse microbial populations, classification of the gut microbiome from biopsies of colonic mucosa and from stool remains centred around sequencing of 16S ribosomal RNA (rRNA), a highly conserved ~1.6kB region with 9 hypervariable regions which allows for identification of bacterial taxonomy. [150] [151] However, the ubiquity of a significant proportion of the gene has allowed for the development of universal primers which bind to conserved regions proximal and distal to the hypervariable regions. [152] Whereas Sanger sequencing allowed for longer reads of base pairs, current NextGen methods produce

reads of 200-300bp with the trade-off being that of a much higher throughput and decreased cost (**figure 2.1**). [153]

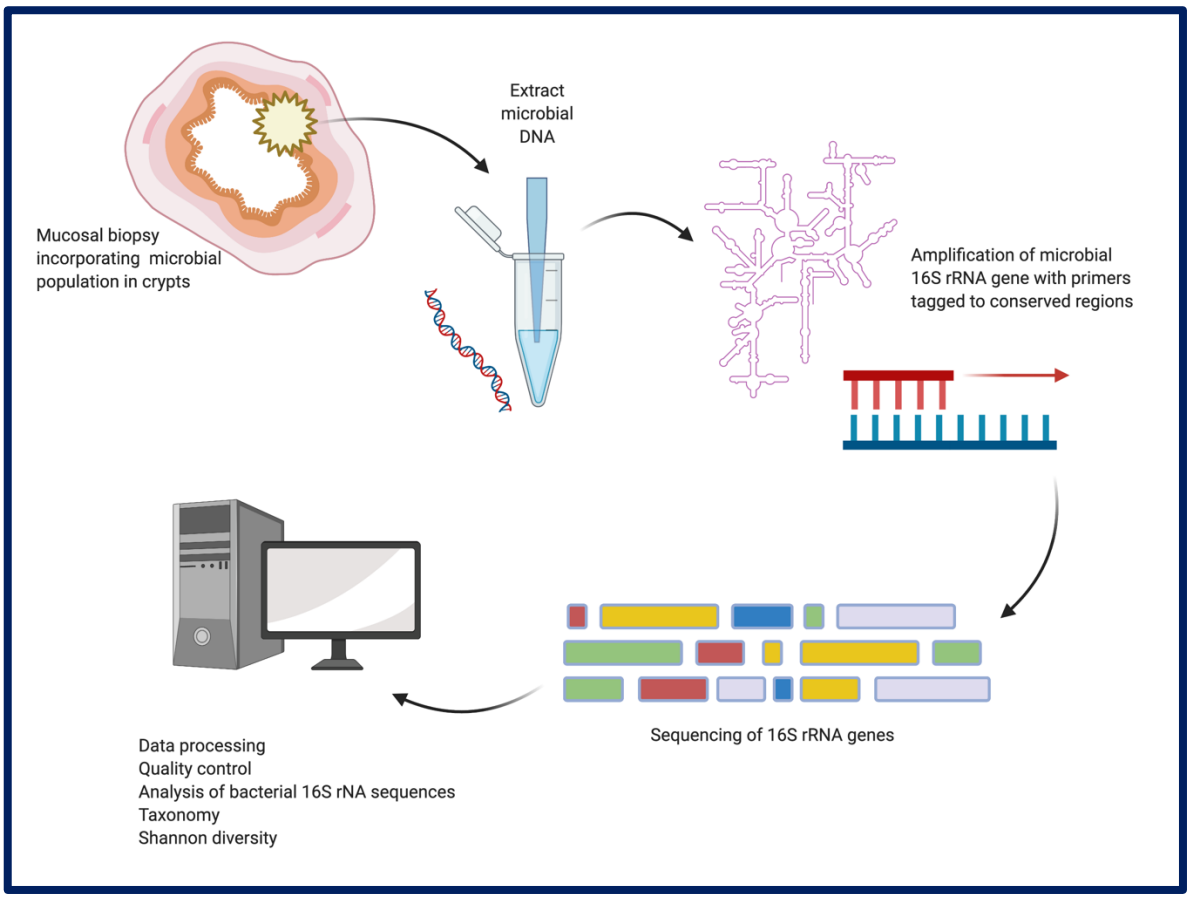


Figure 2.1 – Schematic workflow of a typical microbial 16S rRNA sequencing experiment.

2.2 Metabolism and metabolic disruption in cancer

2.2.1 The Warburg effect

Significantly altered metabolic processes were first described as a fundamental feature of cancer cells by Otto Warburg in the 1920s. His observation was that cancer cells would preferentially undergo aerobic glycolysis (thus converting glucose to lactate) rather than aerobic respiration, even in the presence of sufficient oxygen tensions and fully functional mitochondria **(figure 2.2)**. [154] [155] Given that only around 4 mol of adenosine triphosphate (ATP) are liberated per mol of glucose by this means (as approximately 5% of pyruvate in tumour cells still enters the TCA cycle), as opposed to the 36 mol of ATP liberated per mol of glucose through aerobic respiration, this means that available glucose is consumed by tumour cells at a far higher rate than their normal counterparts. The resultant *Warburg effect* therefore produces vast amounts of lactate in comparison the usual carbon dioxide resultant from the tricarboxylic acid (TCA) cycle which, amongst other features, contributes to a tumour microenvironment that is inherently more acidic and hostile. Additionally, the rate of turnover of glucose in cancer cells – sometimes as high as 100 times that of normal cells - results in increased diversion of both glucose-6-phosphate and glycerol-3-phosphate into the pentose phosphate pathway (for nucleotide synthesis) and lipid biosynthetic pathways respectively, as shown in **figure 2.3**.

These observations have resulted in several theories as to whether the Warburg effect is a key initiator and player in tumour aggressivity, or merely a bystander. Multiple hypotheses have been put forward, notably linking the Warburg

effect with *de novo* lipid biosynthesis, [156 157] the conditions and regulation of the tumour microenvironment [158 159] and cell signalling. [158 160 161] Almost 100 years on, the exact degree to which the Warburg effect explains the nature of tumour development and invasion remains unclear, albeit in an age of significantly renewed interest as we are now able to study tumour metabolism in greater detail than ever before. As discussed in section 1.2, the adaptation of mass spectrometry and NMR spectroscopy platforms now permits a hitherto unseen ability to study tumour metabolism, including its modulation in response to cancer therapies.

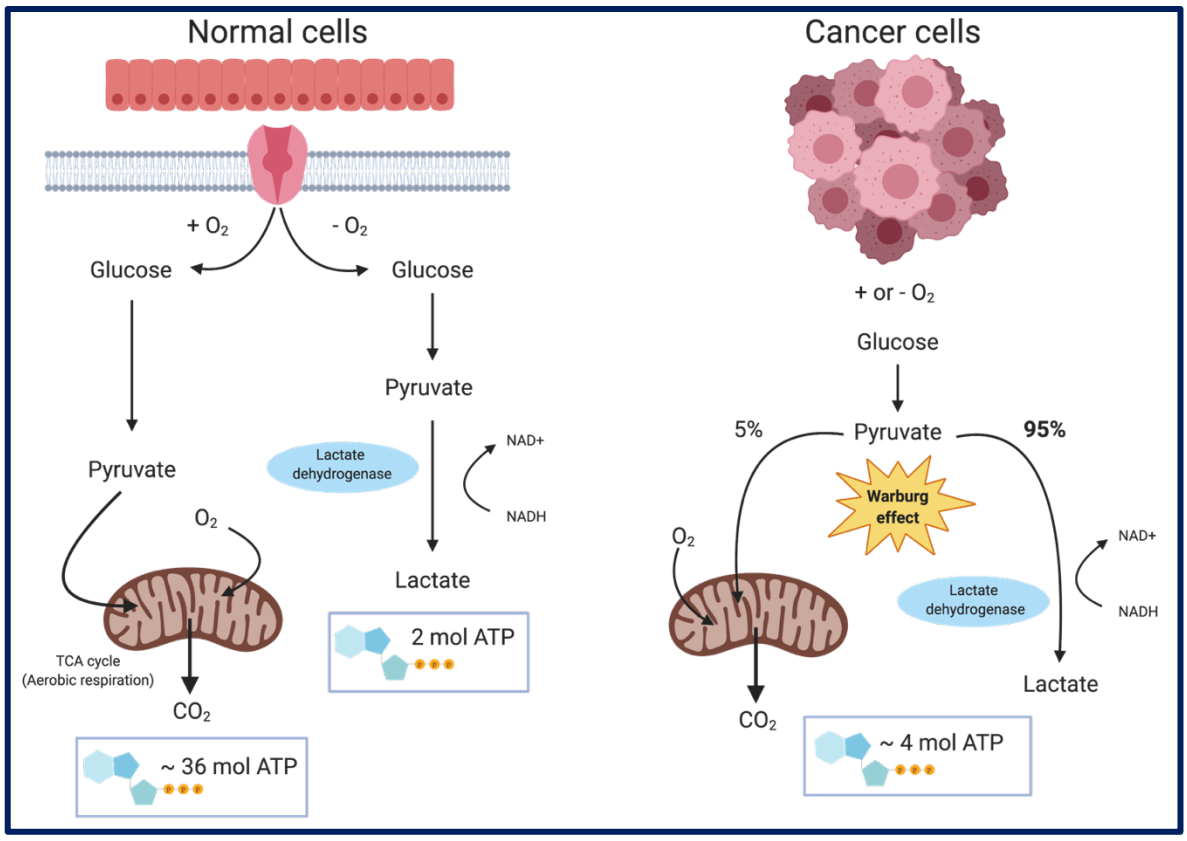


Figure 2.2 - The Warburg effect. Cancer cells exhibit a more rapid consumption of available glucose due to a drive towards aerobic glycolysis, as opposed to oxidative phosphorylation through the TCA cycle. This occurs even in the presence of sufficient oxygen, and therefore differs from the standard (i.e., hypoxic) conditions which drive anaerobic fermentation, which accounts for lactate generation in normal tissues.

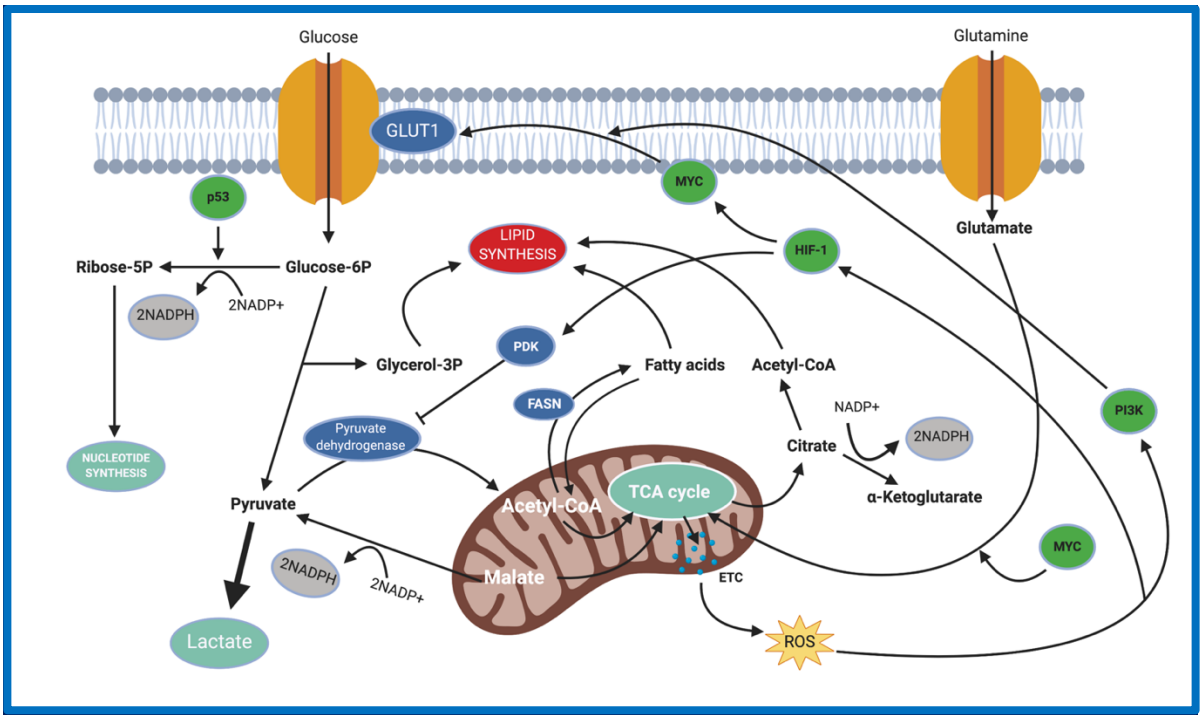


Figure 2.3 - Metabolic pathways in cancer cells. Increased fatty acid, lipid and nucleotide synthesis are key endpoints in the metabolic cascade that follows. GLUT1, Glucose transporter 1; TCA, tricarboxylic acid cycle; ETC, electron transport chain; ROS, reactive oxygen species; HIF-1, hypoxia inducible factor 1; PDK, pyruvate dehydrogenase kinase; FASN, fatty acid synthase. NADP, nicotinamide adenine dinucleotide phosphate.

2.2.2 Lipids in cell biology, colon and rectal cancer

Lipids represent an important class of bioactive molecules with a broad variety of cellular functions. This is exemplified in the dynamic interplay between cell membrane structure and function (as a phospholipid bilayer), mitochondrial pathways, the endoplasmic reticulum and lipid droplets. [162] [163] Of the eight subclasses of lipids described in the open access LIPID MAPS database, fatty acyls (FA), glycerolipids (GL), glycerophospholipids (GPL), sphingolipids (SL) and sterols (ST) are the most biologically important in human cellular homeostasis and metabolism. [164] [165]

The backbone of human cell membranes is comprised of a bilayer of GPLs, so orientated as to maintain a hydrophobic interior with their hydrophilic polar heads lying at the extracellular and cytosolic surfaces. The basic structure of a phospholipid is that shown in **figure 2.4**, of two fatty acyl chains (*sn1* and *sn2*) with a glycerol backbone (diacylglycerol; DAG), with a phosphorylated carbon-based moiety attached. It follows, therefore, that fatty acyls themselves are the essential building blocks of cell membranes, in addition to acting as signalling molecules and key suppliers of energy. [166]

The action of type-I fatty acid synthase (FASN) on citrate (and hence acetyl CoA) exported from the TCA cycle is ubiquitous across all eukaryotes and is the gatekeeper step to FA generation. FA CoA may subsequently re-enter the mitochondria and undergo beta-oxidation, but subsequent cytoplasmic formation of DAG is the precursor step to phospholipid formation - through either the CDP-DAG or Kennedy pathway - which are then exported to the cell membranes or utilised in cell signalling (**figure 2.5**). [165] To this end, it is increasingly evident that a significant number of cellular signalling processes are mediated by the cellular *lipidome*. Despite

the existence of over 1000 biologically plausible cellular lipids, they can be individually detected and characterised by the intra- and extracellular environment with exquisite sensitivity. [167]

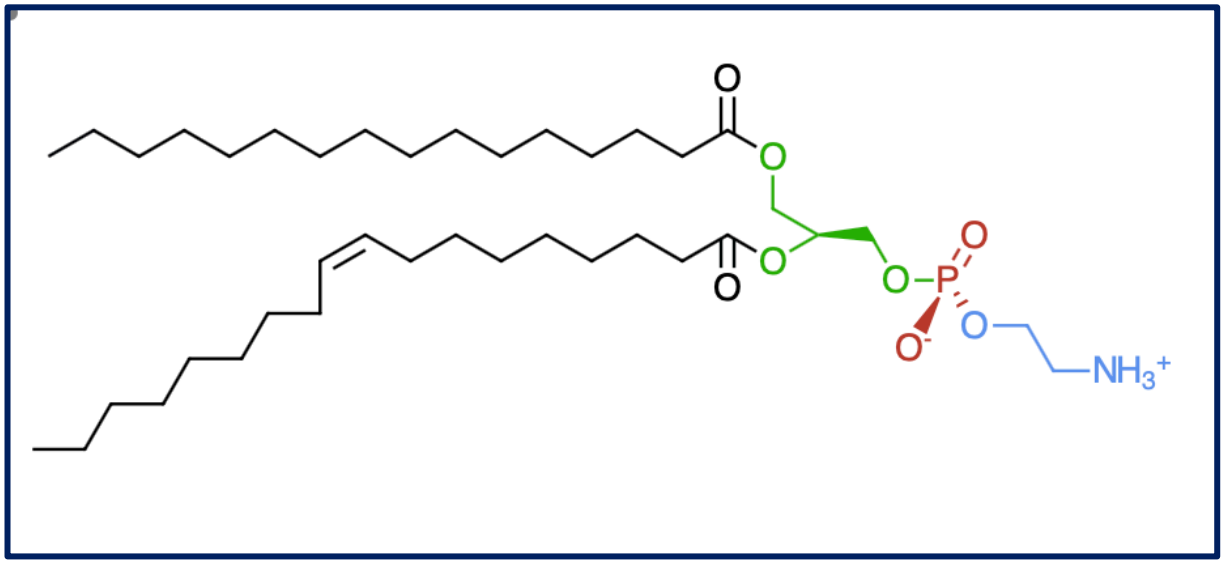


Figure 2.4: Structure of a typical membrane phospholipid, in this case a phosphatidylethanolamine (PE) comprising both 16:0 (palmitate) and 18:1 (cis-9, oleate) fatty acyl chains (black), typically found on the cytoplasmic surface of the lipid bilayer. The hydrophilic head group (blue) in this case denotes the *-ethanolamine* component but could be equally plausible with other head groups; for example, methylation of the amine group here would lead to a phosphatidylcholine (PC). The abbreviated nomenclature for the above (*1-hexadecanoyl-2(9Z-octadecanoyl)-sn-glycero-3-phosphoethanolamine*) is PE (16:0/18:1(9Z)).

The increased abundance of particular subclasses of cellular lipids in cancer is well-documented. [168] With particular reference to colorectal cancer, both fatty acids and glycerophospholipids are the focus of interest with particular relevance to how the understanding of their roles and interplay might unlock further chemotherapeutic targets. [169] As shown, their synthesis is directly influenced by the processes of

oncogenesis and progression of the tumour microenvironment. It is therefore reasonable to hypothesise that they play a vital role in a tumour's in-built defence mechanism against both chemo- and radiotherapeutic modalities. However, to date it has only been possible to focus on specific subclasses of these complex and important molecules, and the data are generated from varied sources (i.e., solid tissue and biofluids, and many with extensive sample preparation for analysis). [170] [171] Furthermore, although the role for ceramide and sphingomyelin metabolism has been implicated in resistance to 5-FU, there is paucity of lipidomics-based studies expressly examining the effects of the tumour lipidome on radiotherapy resistance. [172]

There have, however, been significant advances in the ability to generate biologically representative lipidomic data from human cancer tissue in the course of the last few years. These approaches by their very definition involve minimal sample preparation and permit the analysis of solid tumour tissue harvested from either endoscopic biopsy or surgical resection. In the case of a surgical phenotyping centre - such as that introduced at St Mary's Hospital, Imperial College London – this can be performed within minutes of sample acquisition. However, high quality analysis is also possible when samples are immediately snap-frozen in liquid nitrogen. A snapshot of these studies - with particular reference to the study of lipidomics in colorectal cancer - is summarised in **table 2.1**, and a further technical overview of the techniques relevant to this work follow in section 2.3. It should be noted that, whereas the below table is highly specific to colon and rectal cancer, the modalities employed have been validated across a wide range of tumour subtypes including lung, prostate, ovarian and breast cancers.

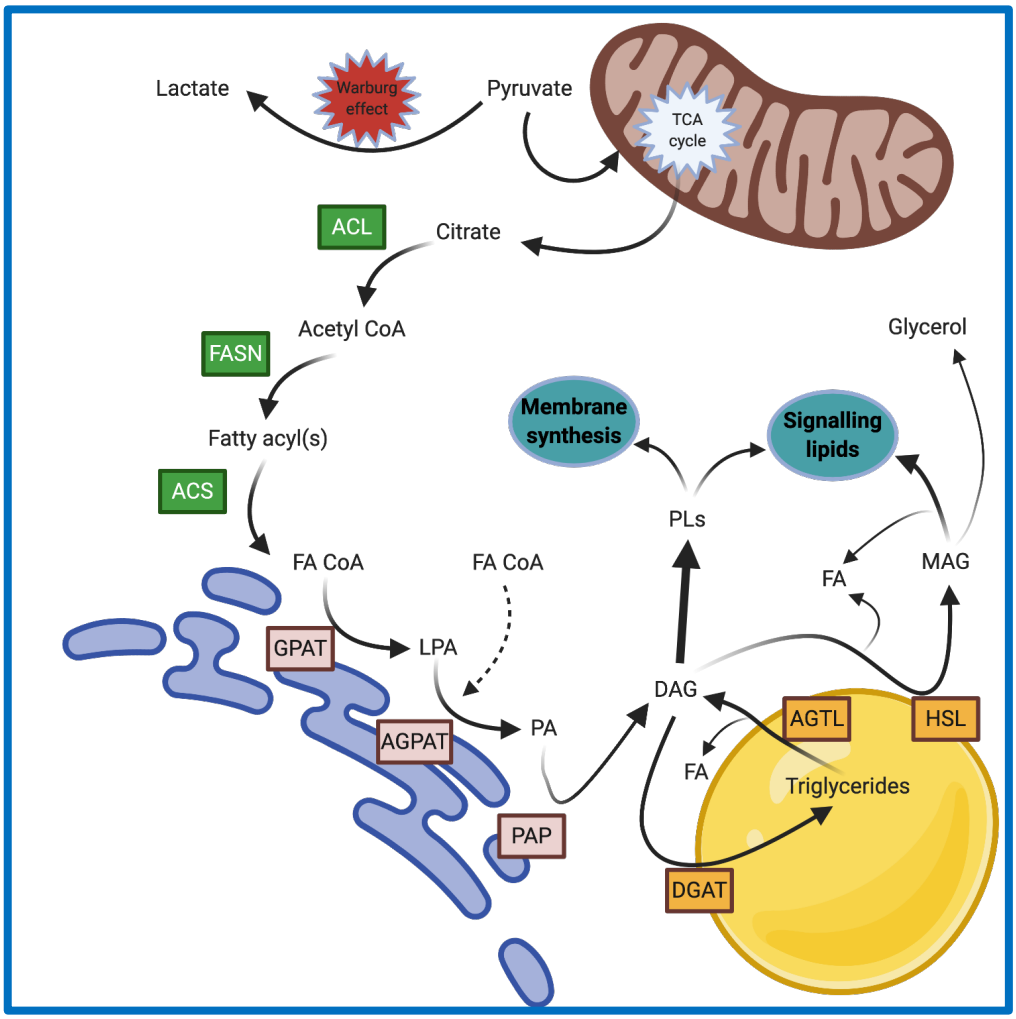


Figure 2.5 - Schematic of intracellular fatty acyl (FA) and glycerophospholipid (PL) production. Cytoplasmic citrate is exported from the TCA cycle and hence into FA generation by action of Type I Fatty Acid Synthase (FASN) on Acetyl CoA. In addition to the Warburg effect previously outlined, cancer cells exhibit increased concentrations of mitochondrial membrane citrate carriers (CIC), driving a higher proportion of citrate into the cytoplasmic compartment and hence into FA generation. The ultimate fate of this FA pool is either lyso- or phosphatidic acid (LPA, PA respectively) generated by the endoplasmic reticulum or onwards into PL generation via diacylglycerol (DAG) generation. DAG generation is further fuelled by triglycerides (TG) drawn from intracytoplasmic lipid droplets. From here, DAG may enter either the CDP-DAG or Kennedy pathways with flux between the two, dependent on the ultimate PL generated. ACL, ATP citrate lysase; ACS, fatty acid CoA ligase; GPAT, G3P-acyl transferase; AGPAT, acyl-G3P transferase; PAP, phosphatidic acid phosphohydrolase; DGAT, DAG acyltransferase. Adapted from Currie et al. [162]

Study	Analyte & sample size	Platform	Results	Metabolic processes implicated
Gerbig *et al* 2012 [173]	Human colorectal tissue – 70 samples	DESI MSI	High variability of GPL abundance in colorectal cancer due to intra-tumoural and inter-patient variability. High abundance of PI in invasive colorectal adenocarcinoma	GPL synthesis FA metabolism/ synthesis
Balog *et al*, 2013 [174]	Human colorectal tissue – 156 patients	REIMS	Accurate discrimination (93.1 – 97.5% on LOOCV across different anatomical locations) between cancer, adenoma and normal colon and rectal tissue on the basis of the composition profile of GPLs	GPL synthesis FA metabolism/ synthesis
Jimenez *et al*, 2013 [175]	Human colorectal tissue – 26 patients	HR-MAS NMR	Accurate discrimination between tumour and non-tumour tissue from within same patients, in addition to metabolic 'field effects' of cancer in tumour-adjacent tissue. Triglycerides depleted within cancer tissue whereas PCs more abundant. Taurine, isogluatmine, choline, lactate, phenylalanine and tyrosine found to be abundant in tumour tissue	Lipid pathways Triglyceride metabolism Cellular regeneration Microbial cell wall breakdown Warburg effect
Mirnezami *et al* 2014 [176]	Human colorectal tissue – 44 patients	HR-MAS NMR	In addition to the accurate discrimination between normal and cancerous colorectal tissue, distinct lipidomic profiles observed for T1/2, T3 and T4 tumours. Depleted triglycerides a feature of T4 tumours whilst T3 tumours demonstrated an increased lipid profile compared to normal tissue and early tumours	GPL synthesis FA metabolism/ synthesis Triglyceride metabolism
Mirnezami *et al*, 2014 [177]	Human colorectal tissue – 12 patients	MALDI MSI	Elevated concentrations of PC (16:0-18:1), LPC (16:0) and LPC (18:1) in cancerous vs. adjacent normal tissue	GPL synthesis and metabolism
Alexander *et al* 2016 [178]	Human colorectal tissue – 28 patients	REIMS	Lipidomic profile generated distinguishes between cancer and non-cancerous tissue (AUC 0.96). Raised levels of long-chain PS and bacterial PGs characterised cancer with triglycerides being more abundant in normal tissue. Ceramides were over-represented in adenomas.	GPL synthesis Bacterial metabolic processes FA metabolism/ synthesis Triglyceride metabolism

Table 2.1 – Selected, recently published studies employing metabolomics-based platforms for phenotyping of colon and rectal cancers and development of (predominantly) lipidomics-guided tissue type classifications. Commonality is seen between studies in their findings, particularly with regard to the role of glycerophospholipids (GPL) and fatty acids (FA). LOOCV, Leave One (Patient) Out Cross-Validation; PC, phosphatidylcholine; LPC, lysophosphatidylcholine; AUC, area-under-the-curve. MALDI MSI, matrix-assisted laser desorption ionisation mass spectrometry imaging; REIMS, rapid evaporative ionisation mass spectrometry; DESI MSI, desorption electrospray

ionisation mass spectrometry imaging; HR-MAS NMR, high-resolution magic angle spinning nuclear magnetic resonance spectroscopy.

2.3 Mass spectrometry-based lipidomic profiling

2.3.1 Introduction

There are several benefits to the use of ambient mass spectrometry (MS) imaging modalities over MAS NMR as an investigative tool in the context of studying human rectal cancer. HAS NMR produces highly reproducible data but lacks the sensitivity of MS applications. By contrast, MS platforms can employ a wide variety of ionisation techniques and modes (i.e. positive or negative ionisation modes) to tailor an experiment to detect compounds with a high degree of sensitivity and specificity. [179] This is not to say that MS is without limitations, and these will be covered in the coming section.

The prime goal of a lipidomics study in human cancer research is to ensure that the compounds that are being detected are biologically representative of the underlying mechanisms at play. Furthermore, given the fact that cancer is not a homogenous disease process, and that there will be regions within the same tumour which will be both genetically and metabolically distinct, the prospect of a platform that can provide spatially-aligned chemometric data that can be correlated back to histological features is an attractive one. Coupled with limited sample preparation, the *Desorption Electrospray Ionisation Mass Spectrometry Imaging* (DESI MSI) interface presents itself as the ideal tool for a hypothesis-driven shotgun lipidomics assay of human rectal cancer tissue with the specific goal of identifying lipidomic features that

correlate with (or rather, are representative of) a biological process underpinning radiotherapy resistance.

2.3.2 Basic principles of mass spectrometry

Mass spectrometry is the identification and quantification of molecular structures by measurement of their mass-to-charge ratios (*m/z*). This involves measurement of ionised molecules both within a vacuum environment and whilst subjected to a magnetic field. It further follows that unique compounds will exhibit measurable differences based on their molecular mass (*m)* and that charged particles immersed in a magnetic field will experience forces exerted by the electric and magnetic fields as per the Lorentz Force (**F,** measured in newtons) phenomenon [180] :

$$F = q\,(E + v \times B)$$

Where q is the electric charge of the particle (in coulombs), E is the electric field (in volts per metre), v is the velocity of the particle (in milliseconds) and B is the magnetic field (in tesla). Furthermore, as per Newton's second law, with force being equal to the product of mass (m) and acceleration (a), the above equation can be rewritten thus [180]

$$\frac{m}{q}\,a = E + v \times B$$

In the context of mass spectrometry, q is expressed as a charge number (z) by dividing by the elementary charge constant, e (1.602 x 10^{-19}); thus, as is standard parlance in mass spectrometry, mass-to-charge ratio is noted as *m/z.*

Therefore, it stands to reason that in order to achieve the conditions required for the above, the following form the key components of any mass spectrometer platform:

1. An *ionisation source* – this ionises the sample (be it a gas, liquid or solid) thus producing both molecular ions and ionised fragments
2. A *vacuum system* to prevent interference from or detection of atmospheric molecules
3. A *mass analyser*, with a perpendicular electromagnetic field. The magnetic field will exert a differing influence on ionised molecules dependent on their mass and these are further sorted by the mass analyser
4. A *detector* – which measures the *m/z* of each ion strike

The resultant output is displayed as a *mass spectrum,* in which each peak or bar on the x axis represents a given *m/z* and its y axis value it's relative abundance in the sample (**figure 2.6**). [181]

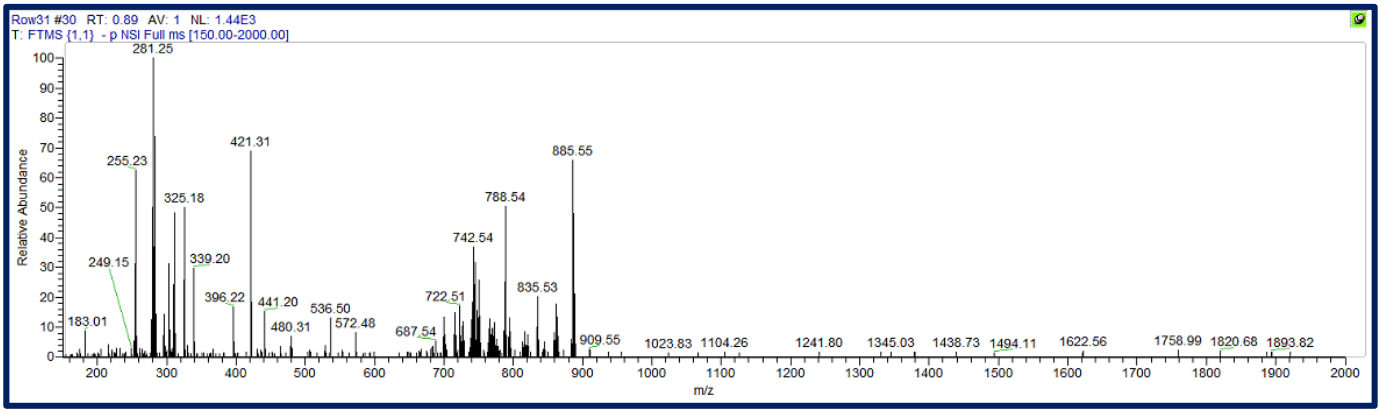


Figure 2.6 – Example mass spectrum in negative ionisation mode across the m/z range 150-2000. Each peak corresponds to a m/z value for a detected ion, and this is graded according to relative abundance on the *y* axis.

2.3.3 Fundamentals of the Orbitrap mass spectrometer

Once charged ions pass through the inlet they are accelerated (in the case of this work) into an Orbitrap mass spectrometer (*Exactive*^{TM}, Thermo Fisher Scientific, MA USA). The operation of this instrument differs from the traditional time-of-flight (TOF) or quadrupole instruments. Whereas the ionisation method and mode will have the greatest effect on the ions produced for analysis from a given sample, there is a significant variability in the physics of the various available detectors and mass analysers which ultimately determines the quality and reliability of the analysis. The Orbitrap belongs to the *Fourier Transform Mass Spectrometry* (FTMS) family and are noted for their superior mass accuracy (measured in a ppm shift in *m/z* from its true value) and high resolution. The Orbitrap functions on the principle of trapping ions in a strong magnetic field combined with a weak electric field; the resultant image produced by the excited (trapped) ions is detected and digitized before being

converted into readable mass spectra by Fourier transform. [182] [183] This differs from other forms of mass spectrometry in that the ions do not strike a detector plate. The fundamental physics behind the instrument are similar to that of a cyclotron particle accelerator, thus:

$$\omega_c = \frac{qB}{m} \quad ,$$

where ω_c is the angular cyclotron frequency, q = ion charge (furthermore transformed to z as per dividing by the elementary charge constant as above), B is the magnetic field strength and m is the ion mass.

Upon passing into the inlet, ions are injected into the space between central and outer electrodes whereby the applied radial electric current bends the trajectory of the ions towards the central electrode. The opposing centrifugal force traps the ions in a circular spiral (hence, 'Orbitrap') and it is from here that the harmonic oscillations of the ions is detected and spectra generated by Fourier transform. [184] [185]

2.3.4 Electrospray ionisation (ESI) and desorption electrospray ionisation (DESI)

There are multiple variables with regards to ionisation source and mass analyser platform albeit the above core principles remain common to all. The DESI MSI interface employed in my work works on the adapted principle of *electrospray ionisation* (ESI).

ESI is an attractive option for analysing complex biological samples due to its ability to ionise minute volumes of thermally labile and non-volatile molecules. It uses

electrical energy to charge solvent droplets and assist in the transfer from liquid to gaseous phase prior to MS analysis. This process involves three steps: first, the charged droplets need to be delivered as a fine spray – usually through a stainless steel or quartz silica tubing with a high voltage applied (usually between 2 - 5kV). Secondly, the carrier solvent has to evaporate away and finally ions ejected from the charged droplets. [186] This spray is often nebulised in an inert gas such as nitrogen to enhance the flow rate. Heating (HESI) or further drying with a secondary nitrogen flow accelerates the evaporation process which reaches a point where the electric field strength within the shrinking droplet passes a critical point when surface ions can be ejected into a gaseous phase, sampled by the skimmer cone array at the aperture to the spectrometer and accelerated into the mass analyser. [187] A key feature of ESI is its ability to produce a cloud of multiple ions in addition to being able to simultaneously run over a wide *m/z* range, thus creating a richer analysis. [188]

The DESI interface was developed by Professor Zoltán Takáts in 2004 and lifts the vacuum constraints of classical MS approaches. The technique is an adaptation of the ESI principle, but in this case the solvent in question is a bland ethanol:water mixture which is charged and fired at high velocity to impact upon the sample of interest (**figure 2.7**). [189] In the case of cancer research, the obvious application is to microsectioned, snap-frozen cancer tissue obtained from endoscopic biopsy or surgical resection. The required adaptation for creating a 3-dimensional chemical image of a tissue section is to be able to move the sample in space beneath DESI source – this action is achieved with a 3D printing stage that in the case of the Orbitrap mass spectrometer is an original home-built device within our laboratory, although a

commercially available interface now exists from Prosolia® and the Waters Corporation®. [190-192]

This ability to move a solid sample in space beneath an ionisation source means that a pixel-by-pixel image can be built up consisting of thousands of individual mass spectra even for a small tissue sample, with a typical pixel size of 100 x 100 microns (μm). Bearing in mind the volume of a human enterocyte is in the order of 1400μm^3, a DESI MSI experiment such as this can provide a high-resolution chemical map of cancer and have the resolving power to discriminate between normal tissue and islands of cancer cells even before the spectral data is taken into account. This degree of topographic resolving power means that even slight chemical variations within a sample can be accounted for and overlaid correspondingly to the H&E-stained tissue section. The ability for DESI MSI to distinguish between tissue features in the same sample has already been established across several tissue types.

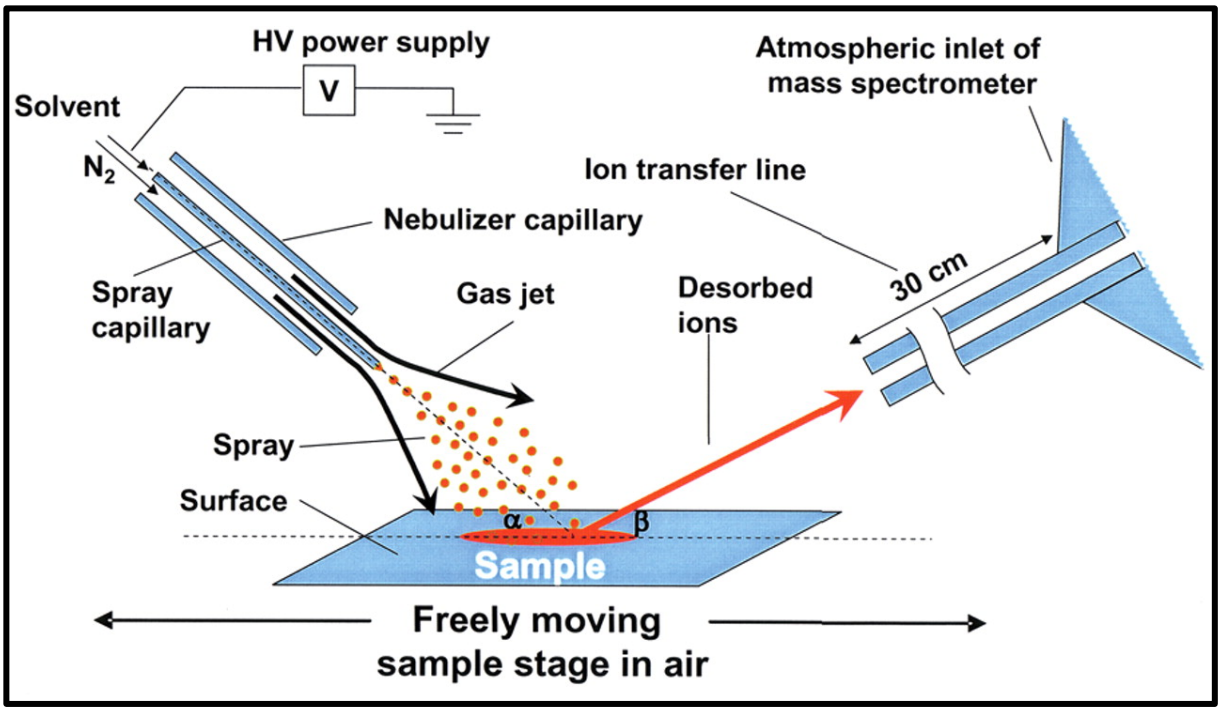


Figure 2.7 – Schematic representation of a typical DESI MSI experiment. Adjustments can be made in geometry of the sprayer assembly, voltage, ionisation mode and solvent of choice. The sample is typically mounted onto a PTFE or glass slide – mounting on a standard glass microscopy slide permits downstream H&E staining and other histological modalities for comparison. Figure from Takáts *et al* 2004. [189]

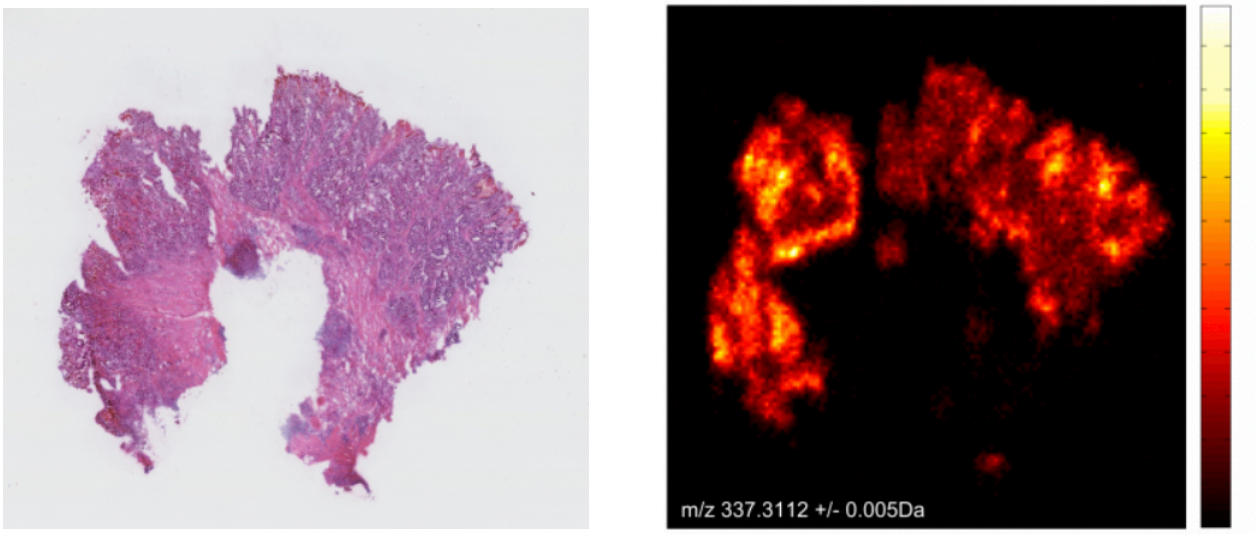


Figure 2.8 - A typical example of the spatially resolved chemometric data acquired from a DESI MSI run using colorectal tissue. Left, post-DESI H&E-stained colorectal cancer tissue section; right, ion map (with post-processing RGB colour intensity grading) of an ion with m/z 337.3112 (FA 24:1, i.e., 24 carbon chain with one double bond). Note the relative abundance in varying degrees in adenocarcinoma tissue compared to its absence in submucosa and muscularis layers (unpublished data).

2.4 Data handling approaches for biological mass spectrometry datasets

MSI experiments by their definition produce a vast quantity of multi-dimensional data, and the development of appropriate data handling strategies is critical to interpretation of the data in a biologically plausible manner. In the context of MSI, the field of bioinformatics has made tremendous strides in in the ability to apply machine learning algorithms to the measurement of compound distributions and relative abundances with direct correlation to histological features in tissues (**figure 2.8**). [177] [193-195] The raw spectral data generated in DESI MSI as per figure 2.6 equates to one such spectrum for each 100 x 100μm pixel of tissue – therefore, a typical 6 x 6mm

tissue section will generate 3600 spectra for that sample alone, all of which need to be pre-processed and histologically aligned prior to multivariate statistical analysis.

In order to tackle this problem, a variety of chemo-informatics packages have been developed which focus on the following components:

1) **Pre-processing of raw imaging data** – critical to downstream modelling, MSI data needs to be comparable and manageable whilst also accounting for true biological variabilities. Initial pre-processing involves filtering out of background 'noise'; in most cases this incorporates a machine-learning algorithm that selects spectral peaks and then identifies relevant m/z values. In so doing, the data passed on for analysis is filtered of spectral peaks that are likely to be irrelevant to the analysis, i.e. related to background noise generated by solvent or atmospheric contaminants or to non-representative background signal. Data are then normalised by scaling according to a data-specific factor – for the majority of our pipelines we use *total ion count* (TIC) normalisation. This is performed whereby the sum of all intensities for each spectrum is calculated and each m/z intensity divided by the sum value – this enables a more representative comparison of intensities between spectra across the sample. This normalisation step is especially important for MSI as the ambient conditions and long analysis times can lead to in-sample variation due to extrinsic experimental (i.e., non-biological) factors. [196] [197]
2) **Histological/Ion image co-registration** algorithms within an in-house chemometric toolbox that has been previously described. Of note, once

alignment has been performed the toolbox permits dual image magnification for detailed annotation and correlation of features to single-pixel resolution. [193]

3) **Multivariate statistical analysis** is then used to establish relationships across multiple variables in a dataset and are broadly split into two groups: unsupervised (in this case, principal component analysis (PCA)) and supervised analysis (which for this work I have employed recursive maximal margin criteria (RMMC)). RMMC has been consistently validated for use in DESI MSI datasets across both our group and other published work. [194] The transformation of the high-dimensionality of the raw spectral data to a low-dimensional space, such that the data is workable and yet retains meaningful properties reflective of the underlying biology in both cases is referred to as *dimensionality reduction.* Both PCA and RMMC represent linear transformations of the high-dimensionality space with the RMMC model requiring classification of each spectrum into cohorts to determine differentially abundant m/z values. The Eigenvalue (characteristic value/root) decomposition for PCA is summarised thus:

$$[S_t]w = \lambda w$$

where S_t is overall dataset variation and λw is the derived multivariate model. This maximised the overall dataset variation, without specific regards to between-class differences. For the RMMC model, the between-class variation is maximised with a loss of within-class variation thus:

$$[S_b - S_w]w = \lambda w$$

where S_b is between-class scatter and S_w is within-class scatter. Ultimately this model generates a data matrix from which univariate statistics can be applied – in this case, analysis of variance (ANOVA) – for extraction of the m/z values of between-class discriminating components. [198]

2.5 The role for a lipidomics approach as part of a multi-omics analysis of the radiotherapy response in rectal cancer

Ambient mass spectrometry has already demonstrated an ability to provide unique insight into both the metabolic footprint and the lipidomics of colorectal cancer. Whereas this has only previously been applied in the context of tissue classification and in the mapping of distinct lipidomic profiles in colorectal cancer, transition zones/tumour adjacent tissue and normal colorectal mucosa, the fact that such significant lipidomic perturbation exists between normal colorectal tissue and colorectal cancer implies that lipid metabolism plays a significant role in the biology of colon (and in the case of this work, rectal) cancer. It stands to reason that if shifts in lipid metabolism can play such a key role in tumorigenesis and progression, they are also likely to be implicated in the biology of these tumours with regards to their response to therapies and specifically radiotherapy for rectal cancers. There is paucity of mechanistic work on the basis of the radiotherapy response in rectal cancer, although a great deal of effort has been made to identify clinically translatable, predictive biomarkers. The aim of my work herein is to explore the role of a more holistic approach to the contributing factors at play in radiotherapy resistance in rectal cancer, and to justify a lipidomics standpoint as a starting block in this process. In

Chapter 3, I outline my aims and objectives for this work and more precisely, how I plan to test these hypotheses.

Chapter 3
Aims and Hypotheses

The original work that I have undertaken and present in this thesis represents several pilot studies and focuses on the following workflows:

3.1 An *in silico* analysis of the radiation response in rectal cancer

In this section of my work the following hypotheses were formulated and tested:

1. Current attempts at mechanistic insights into radiotherapy response in rectal cancer do not fully appreciate the required holistic view of the biological problem
2. The radiotherapy response is likely to be governed by a multi-modal mechanism involving exaggerated stress response mechanisms
3. These stress response mechanisms are likely to be heavily influenced by perturbations in lipid metabolism

Specific objectives designed to test the above hypotheses:

1. To undertake an extensive systematic review of published, peer-reviewed, original experimental work on biomarkers predictive of radioresistance and/or radiosensitivity
2. To meta-analyse this data using an original, holistic approach by way of classifying biomarkers using Gene Ontology biological functions.
3. To map this data in such a way as to demonstrate significance of various biological functions in the context of radiotherapy response
4. To undertake a Gene Set Enrichment Analysis (GSEA) on a publicly available dataset of rectal cancer samples, and specifically to map the expression data to KEGG pathways implicated in lipid metabolism

3.2 Mapping cellular lipidomics in response to neoadjuvant radiotherapy with Desorption Electrospray Ionisation Mass Spectrometry Imaging (DESI MSI)

In this section of my work the following hypotheses were formulated and tested:

1. DESI MSI can be successfully utilised to build a complete, untargeted lipidomic overview of rectal cancer and associated normal tissues even in the context of radiotherapy and chemoradiotherapy
2. DESI MSI can identify and distinguish between rectal cancer samples subjected to radiotherapy and those which did not receive radiotherapy, based on their bioactive lipid chemistry
3. Perturbation in lipid metabolism - observed as differences in abundancies of various bioactive lipids – correlates with both radioresistant tumour cells and also, holistically, with the clinical tumour regression grade of the macroscopic tumour at resection

Specific objectives designed to test the above hypotheses:

1. To employ a validated, untargeted DESI MSI approach on human rectal cancer tissue specimens with associated normal tissue to build a spectral model of the rectal cancer microenvironment directly, topographically correlated with *same section* H&E histological analysis
2. To build statistical models that can differentiate radioresistant rectal cancer tissue, radiotherapy naïve rectal cancer tissue and adjacent normal tissue from the same patients, based on spectra acquired from DESI MSI
3. To conduct compound identification with *in situ* DESI MS/MS on immediately adjacent tissue microsections

3.3 Analysis of tumour genomic heterogeneity pre- and post-neoadjuvant radiotherapy

In this section of my work the following hypotheses were formulated and tested:

1. Pre-treatment biopsies from rectal cancer at time of diagnosis exhibit a degree of heterogeneity that is linked to response (TRG) to radiotherapy
2. Next-generation sequencing methods can demonstrate a more radioresistant genotype in post-treatment rectal cancer tissue and
3. This genotype evolves as a response to the therapy applied

Specific objectives designed to test the above hypotheses:

1. To perform high-quality DNA extractions from matched pre-treatment biopsies and post-treatment rectal cancer specimens to pass into a Next-Generation Sequencing pipeline
2. To separately sequence discrete regions of macroscopic, pre-RT tumour using a Whole Genome Sequencing (WGS) approach, to evaluate the degree of tumour genetic heterogeneity with relation to somatic copy number aberrations (SCNAs)
3. To employ a bioinformatics-based approach to analyse and compare aggregated SCNAs both within-patient (pre- and post-RT) and between patients to establish the role of genomic heterogeneity in radiotherapy response.

Chapter 4

An *in silico* study of the response to radiotherapy in rectal cancer – a systematic review with network mapping and Gene Set Enrichment Analysis

4.1 BACKGROUND AND RATIONALE 88

4.2 A SYSTEMATIC REVIEW AND 'KNOWLEDGE NETWORK MAPPING' OF BIOMARKERS IMPLICATED IN RADIOTHERAPY RESPONSE IN RECTAL CANCER 91

4.2.1 Search methodology 91

4.2.2 Statistical data processing and analysis 92

4.2.3 Results – study inclusion 94

4.2.4 GO Biological Function - Cell cycle and cell proliferation 96

4.2.5 GO Biological Function - Cell death 99

4.2.6 GO Biological Function - Response to stimuli, stress and DNA damage 100

4.2.7 GO Biological Function - Cell signalling 107

4.2.8 GO Biological Function - Cellular metabolism – nucleic acids and lipid products 113

4.2.9 GO 'knowledge network' mapping 116

4.2.10 Discussion 121

4.3 GENE SET ENRICHMENT ANALYSIS OF A LEGACY GENE EXPRESSION DATASET 124

4.3.1 Introduction and rationale 124

4.3.2 Search methodology 125

4.3.3 Data processing methodology 126

4.3.4 Results 128

4.3.5 Discussion 131

4.4 SUMMARY AND CONCLUSIONS 135

4.1 Background and rationale

Rectal cancer presents unique therapeutic challenges as a consequence of unforgiving anatomical location and propensity for loco-regional recurrence. The mainstay of treatment for patients with operable disease is radical surgery by total mesorectal excision (TME) which has led to significantly improved local control over the past three decades. [69] Several landmark randomised control trials have demonstrated that the addition of pre-operative radiotherapy (RT) reduces rates of local recurrence in advanced rectal cancer (T3/4 and/or node positive tumours). [59 102 111 199] Pre-operative therapy can be administered in one of two ways: short course radiotherapy (SCRT) consisting of 25Gy of radiation delivered as 5Gy/fraction over 5 days, followed by immediate surgery, or long-course chemo-radiotherapy (LCCR) consisting of 45–50.4Gy, delivered at 1.8–2Gy/fraction, usually with concurrent 5-fluorouracil based chemotherapy for enhanced radiosensitivity.

In the majority of patients, RT leads to clinically meaningful tumour regression. However, there is considerable variability in terms of tumour response; some tumours exhibit almost no response at all, while up to 20% will exhibit a complete pathological response (pCR). The remainder have an intermediate response, with varying degrees of residual viable tumour seen in the post-operative resection specimen. Correspondingly, patients will fall broadly into three groups: (1) those in whom RT has led to complete tumour destruction; in this group of patients, there is growing interest in the feasibility of 'watch and wait' organ-preserving treatment strategies; [200 201] (2) those in whom RT will result in at least a moderate degree of tumour regression, which in turn may enhance the likelihood of R0 resection and/or sphincter preservation and (3) those in whom pre-operative RT will lead to negligible tumour shrinkage.

Paradoxically, for patients in the last category, the delay in proceeding to tumour excision whilst completing RT may increase the likelihood of tumour dissemination beyond the pelvis. Thus, the development of reliable methods to predict response at the pre-treatment phase represents a critical unmet need in order to personalise treatment algorithms in rectal cancer.

Current understanding of response to RT is limited by the lack of a unifying interpretation of the molecular pathways implicated in response mechanisms. Additionally, previous reviews in this context have served little to no role in expanding current understanding of this topic as they typically evaluate data as a narrative series of disconnected molecular entities, with varying experimental methodology and clinical outcome measures and no wider appreciation of biological interconnectivity. The mechanisms governing tissue response to RT are likely to involve scores of biological pathways and processes at genomic, transcriptomic and metabolomic level are likely to have importance. Thus, it is desirable to map potential biomarkers more holistically, in such a way as to allow integration of heterogeneous molecular data and intuitive visualisation of the resulting network in order to identify broad areas of biological function with greatest translational potential.

In this chapter of my thesis, I have conducted a systematic review of the literature of biomarkers implicated in rectal cancer radiosensitivity. I have furthermore performed a novel analysis of the identified biomarkers, by classifying these according to Gene Ontology in an attempt to identify biological domains that hold promise for research. This aims to map the key molecular processes already identified (or otherwise), and permit visualisation of how these are potentially interconnected. The

goal here was to ascertain a role for lipid metabolism as an under-represented area of research but also one that is justified in being examined further. Finally, I have analysed a legacy microarray dataset using a Gene Set Enrichment Analysis (GSEA), mapping the outputs to key lipid biosynthesis pathways, to further demonstrate a foundation for pursuing a lipidomic and metabolism-based approach. Both of these methodologies will serve to elucidate the degree to which lipids may be implicated in the radiotherapy response and to direct the work in the subsequent chapters of this thesis.

4.2 A systematic review and 'knowledge network mapping' of biomarkers implicated in radiotherapy response in rectal cancer

4.2.1 Search methodology

An electronic literature search of MEDLINE and Embase databases was performed for English-language articles published between January 2000 and November 2015, utilising Boolean logic and controlled vocabularies (Medical Subject Headings (MeSH) and EMTREE). The following keywords were used, with wildcards used to capture maximum possible articles: '*rectal', 'cancer', 'neoplasm', 'response', 'pathological complete response', 'tumour response', 'radiotherapy', 'chemoradio*', 'neoadjuvant', 'predict*', 'biomarker', '*omic', 'RNA'* and *'DNA'*. Both human *in vivo* and human colorectal cancer cell-based *in vitro* studies were included. Inclusion criteria were studies examining potential biomarkers (specific genes, protein products or metabolites) correlated with a documented tumour regression grade (TRG) after LCCR or SCRT. Studies were excluded at this stage if they evaluated targeted monoclonal antibody therapy, gene expression panels obtained through microarrays, micro-RNAs where the gene targets were not defined, or pure animal model derived data. Microarray studies were further interrogated for the work in section 4.3.

Patient numbers, processed data and statistical methods were extracted from each study for each biomarker assessed, including those biomarkers found to have no significant association with response to RT. Biomarkers were then classified based on biological function according to the Gene Ontology (GO) database and furthermore molecular function. Data were entered into a datasheet with descending ontology terms clustered with reference to previous work (see Appendix A). [202]

4.2.2 Statistical data processing and analysis

The following information was extracted from each publication: (a) studied biomarker; (b) the analytical methodology employed; (c) patient numbers; and (d) the statistical parameters used by the authors to categorize discovery as significant, or not, according to correlation with tumour regression grade (TRG) (including p-values and odds ratios for univariate tests, and sensitivity and specificity for multi-variate tests). Each biomarker was then classified according to the system proposed by the Gene Ontology (GO) Consortium into one of five domains of molecular function or biological purpose (cell cycle, cell death, response to stimuli, cell signalling and cellular metabolism). [203] The GO approach represents a relatively recent development in biological computational science that offers a means of annotating and grouping molecular products of upstream gene expression on the basis of functional/biological commonality. We chose to utilise a GO approach here in order to generate a network of biological processes likely to be involved in governing radiation response that was accessible and straightforward to interpret.

Following statistical review, depending on the statistical parameter used in the original study, significance was binarised (i.e., significant or not) based on p-values ($p<0.05$), odds ratios ($OR > 1$), or sensitivity and specificity (>60%). A composite score summarising the significance of a given biomarker was in turn calculated as a ratio of significant biomarkers relative to the total number of biomarkers under a given ontology heading. Non-unique biomarkers reported by different studies were included in this scoring as the magnitude of supporting or conflicting evidence for a given biomarker by different studies requires appropriate representation. For each GO node, a second score was computed to represent the total number of biomarkers and

publications investigating this domain. This was calculated as the number of unique biomarkers for a given GO term (unlike in the first score where non-unique biomarkers were considered). These two scores (each ranging between 0-1) represent two dimensions on the reported biomarkers we identified through systematic review and were mapped as 'relative significance' and 'popularity'. Subsequent node size, therefore, reflects the number of unique biomarkers for a given GO term, and the node colour represents relative significance, as mapped to the Hue/Saturation/Brightness colour map from 0 to 120^{O} (see **figure 4.1**).

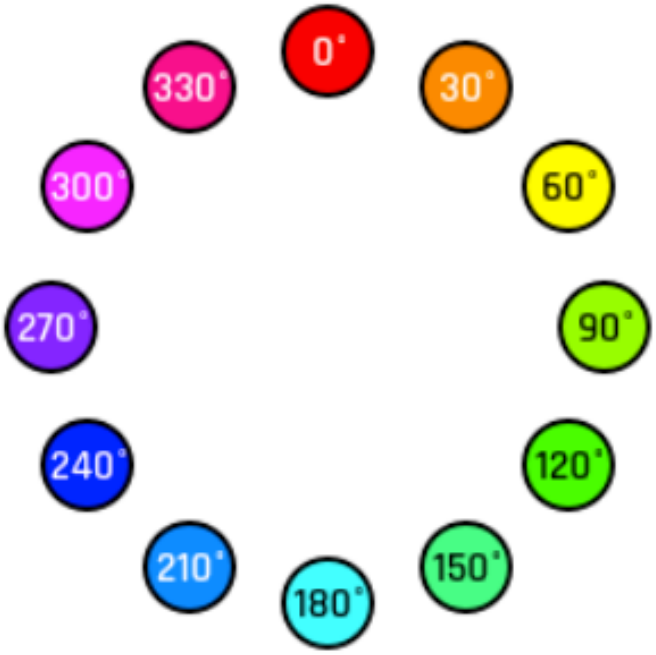


Figure 4.1 – The Hue, Saturation, Brightness (HSB) colour space allows conversion of a numerical value into a standardised colour representation that can be used in a quantitative manner and far more simplistically than RGB.

4.2.3 Results – study inclusion

Four hundred and fifty-two studies were identified following removal of articles which did not include full-text (mostly conference proceedings) and manual de-duplication. The abstracts of these articles were further screened and 373 were excluded as they did not address the question of tissue biomarkers predicting response to RT in rectal adenocarcinoma (i.e., plasma/serum factors, gene microarray profiling), or based on insufficient information on methodology or statistical analysis. A further 11 excluded articles comprised systematic reviews and meta-analyses, which were hand-searched, providing a further 5 articles. This left 74 articles which provided sufficient methodological and statistical detail for both systematic review and to enable integration into a node-based biological network of tissue-based biomarkers. Finally, the reference lists from these articles were reviewed but no further eligible studies were identified for inclusion (many reference lists included the studies included here). The flow for study selection as per PRISMA is shown in **figure 4.2**.

For ease of discussion biomarkers evaluated in these studies were analysed based on biological and molecular function according to Gene Ontology (GO), and clustered under broad, highest level biological function using controlled language as previously described in a figure generated by Hu et al in a review aimed at categorising gene function. [203 204] For biomarkers with multiple GO annotations, clustered GO terms fell under different domains of biological function and so were recorded as separate entries in a datasheet of hierarchical ontology terms for each biomarker **(Appendix A)**. It should be noted, however, that many of the biomarkers assessed have multiple functions and GO annotations – this is taken into account in the network-driven

analytics. In the sections that follow (4.2.4 to 4.2.8 inclusive), the genes and proteins behind these putative biomarkers are discussed in more detail.

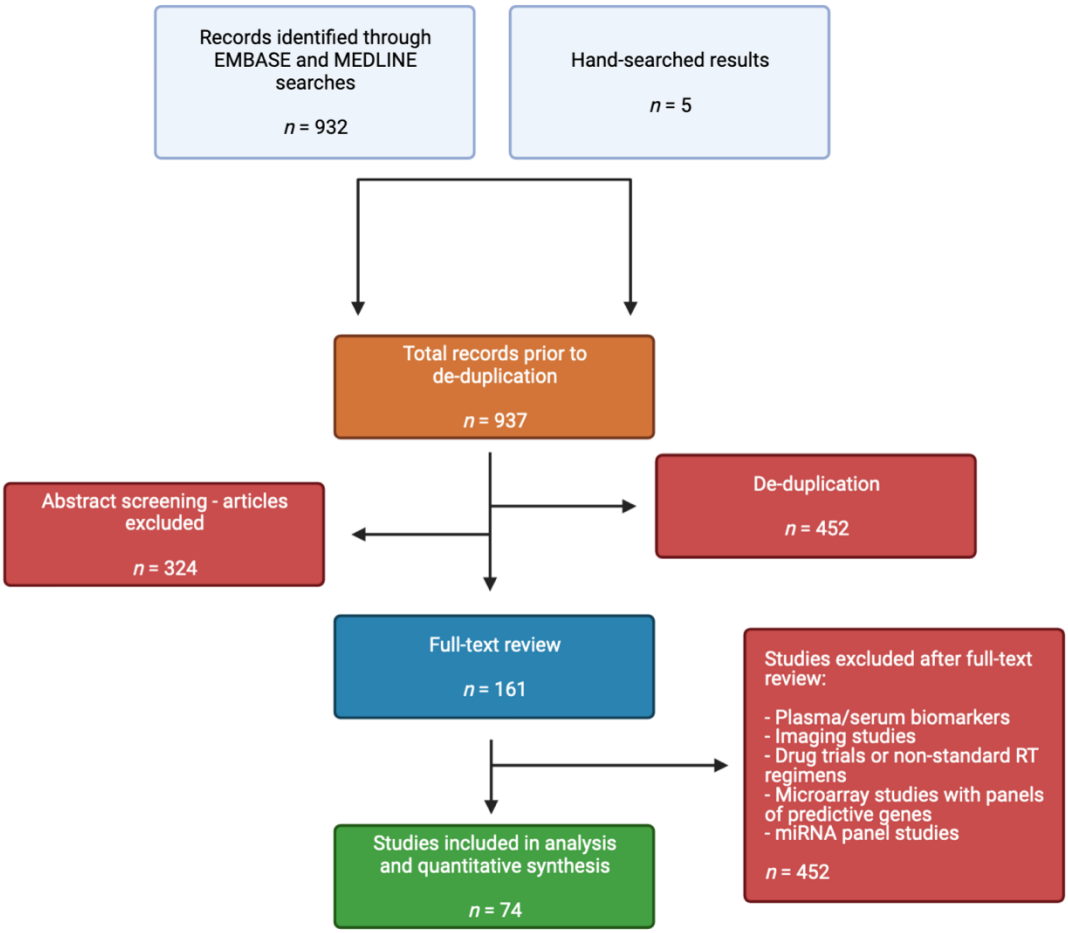


Figure 4.2 – PRISMA flow outlining search methodology and study exclusion/inclusion for quantitative synthesis

4.2.4 GO Biological Function - Cell cycle and cell proliferation

Following GO annotation of biomarkers with biological functions, all identified biomarkers from review were clustered into five key domains. Studies of cell cycle and potentiators of cell proliferation are core to the understanding of oncogenesis and have also acted as starting blocks for therapeutic targets. Numerous genes implicated in the regulation of the cell cycle and cell proliferation have long been identified as key players in the development of cancers. Amongst these, the p53 tumour suppressor network has been extensively studied and is ubiquitously implicated in almost all cancer subtypes. At its core, p53 can be broadly considered a tumour suppressor activated in response to stress, with subsequent effects on apoptosis, cell-cycle arrest and senescence, with loss of p53 function being associated with decreased apoptosis. [205] [206] Although certain radioresistant CRC cell lines demonstrate decreased expression of p53, multiple immunohistochemical (IHC) studies aimed at identifying it as a candidate biomarker for radioresistance have not shown a significant association between isolated p53 expression and TRG. [206-211] One 2014 study in which IHC was performed for a panel of biomarkers including p53, revealed correlation between low p53 expression and pCR, but not with TRG, [212] whereas Kelley et al found that intact p53 with deficiency in Caspase-8 expression predicted inferior TRG. [213]

These conflicting findings may in part relate to the complex interactions of the p53 network. In a study of 112 patients, both univariate and multivariate analysis found that p53 expression in pre-treatment biopsy specimens did not correlate with RT response, whereas p21 expression – the product of which is a CDK inhibitor acting through p53-dependent and independent pathways – was significantly associated with

both TRG and disease-free survival (DFS). [210] By contrast, two further studies examining both p21 and p53 did not demonstrate any correlation between p21 expression and TRG. [206 214]

Downstream, a cross-regulatory network exists between p53 and protein kinase B (Akt), balancing the apoptotic regulation by p53 with the effects of the Akt signalling pathway. Phosphorylated Akt exerts subsequent effects on cell cycle progression, metabolism and inhibition of apoptotic pathways through Bad, [215] although its expression was not found to be predictive of response to LCCR when evaluated with immunohistochemistry. [216] Other CDK inhibitors, notably p27, have also failed to demonstrate any correlation with response to RT. [206 217] Markers of cell-cycle progression are, however, useful surrogate markers of proliferation, and both Ki67 and securin (pituitary tumour transforming gene 1, *PTTG1*) have been evaluated for correlation with response to RT. Whereas Ki67 is detected in all active phases of the cell cycle, securin peaks in the G2/M phases. Significant correlation has been demonstrated between Ki67 expression and response to RT. [218-220] Although securin did not demonstrate a predictive capacity for TRG, expression levels in general were shown to fall following RT; furthermore, for those tumours with high securin expression following treatment, overall prognosis was poorer. [221]

The Regenerating/regenerating islet-derived gene 4 (REG4) is one of a family of genes (*Reg*) belonging to the calcium-dependent lectin superfamily, encoding a group of small secretory proteins. These have been found to be constitutively expressed in cell lines of the GI tract with characteristic upregulation in rectal cancer at both mRNA and peptide levels. [222 223] In cellular models, it has been shown to activate cell

proliferation, in addition to *in vitro* migration and invasion of colorectal cancer cells. [224] Increased expression of REG4 in radioresistant rectal cancer cell lines led to a subsequent upregulation of NEIL2 and BIRC5 (survivin), suggesting a network correlation of expression of these genes and their products with radioresistance. [225] Additionally, IHC for REG4 expression in 172 clinical specimens revealed increased REG4 expression to be a surrogate marker for radioresistance. [226] BIRC5 acts as both a controller of mitotic progression and inhibition of apoptosis, and has been demonstrated to play a role in the molecular pathogenesis of CRC. [227] Conflicting evidence exists, however, for survivin as a putative biomarker for radioresistance. Whereas McDowell *et al* did not find survivin alone to be a predictor of response when assessed by both mRNA and protein expression, [228] other studies have found a significant correlation, in both *in vivo* and *in vitro* models with established radioresistant cell lines including SW480 and HCT-15. [225 229 230]

Although the full role and molecular function of the glycoprotein YKL-40 (chitinase-3-like-1) is yet to be elucidated, it is understood to regulate a number of cellular processes, including cellular proliferation and by exerting effects on the tumour microenvironment to stimulate angiogenesis and tissue remodelling in CRC. [231 232] A multicentre study utilising immunohistochemistry to verify YKL-40 expression in biopsy specimens found a high positive predictive value for radioresistance, associated with a lack of pCR in 87% of cases. This predictive value was increased to 94% when evaluating for biomarker co-expression with the c-Met proto-oncogene. [233]

4.2.5 GO Biological Function - Cell death

Cell death covers a range of biological processes by which the cell ceases to function, either due to an intrinsic, programmed pathway, or in response to extrinsic stressful stimuli such as exposure to cytotoxic agents or RT. A number of mediators of the apoptotic pathway have shown potential as predictors of radioresistance.

A critical step in apoptotic signalling is the release of cytochrome-c from mitochondria, which in turn activates apoptosis protease-activating factor 1 (APAF-1/apoptosome). Following the activation of the initiator Caspase 9, a series of reactions ensue that result in apoptosis, and which are regulated by a balance of pro- and anti-apoptotic mediators. APAF-1 expression in pre-treatment biopsies has been shown to correlate significantly with an improved TRG. Evidence for downstream mediators such as Bax (pro-apoptotic) and Bcl2 (anti-apoptotic) as predictive biomarkers has been inconclusive, [206 210 234] although one group found a significant association between Bcl-2 and response on both uni- and multivariate analysis. [211] Furthermore, programmed cell death 4 (PDCD4) was first shown to be upregulated in apoptosis by Shibahara et al and is now known to act as a nuclear mediator of apoptosis having interactions with factors downstream of Akt; [235] in a recent study PDCD4 expression was shown to correlate positively with TRG on multivariate analysis. [236]

Although some correlation has been observed between apoptotic rate in pre-treatment CRC biopsies and response to RT, [228 237] expression levels of M30 (caspase-cleaved keratin 18 cytoskeletal protein) have not been shown to be predictive of response to RT. [238] Similarly, Annexin V - an established marker for apoptosis as it binds to phosphatidylserine molecules on the cell surface (an event

observed in apoptotic cells) - was not found to be predictive of response. [239] In spite of this, there is increasing evidence that the presence of higher levels of complex lipids is not only characteristic of the tumour microenvironment but may also contribute to radioresistance. Levels of phosphatidylethanolamine binding protein 4 (hPEBP4) have been shown to correlate strongly with increased resistance to RT in a cohort of prospectively recruited patients who received SCRT. [240] Functionally, hPEBP4 is a secreted glycoprotein that has been shown to be involved in the activation of Akt. [241] Overexpression of hPEBP4 in SW480 cell lines has been shown to enhance clonogenic survival, a finding which was mirrored in mouse models subjected to ionising radiation. Inhibition of Akt activation additionally appeared to reverse the radioresistance effect of hPEBP4. [242]

4.2.6 GO Biological Function - Response to stimuli, stress and DNA damage

Ionising radiation leads to DNA strand breaks leading to cell death; induction chemotherapy (commonly, 5-flurouracil-based agents with or without leucovorin) are administered to augment this effect by sensitising tumour cells to RT. Mediators involved in maintaining cellular homeostasis and responding to changes in the tumour microenvironment have shown variation in expression levels in relation to radiosensitivity. The presence of relative tumour hypoxia has been implicated in radioresistance in a number of malignancies, and this theory is increasingly applied in advanced imaging techniques such as diffusion-weighted MRI and PET-CT in characterising tumour biology. Central to the regulation of over 100 genes involved in the adaptive response to hypoxia is hypoxia-inducible factor 1 (HIF-1). [243] [244] The effects of this transcription factor, stabilised at low oxygen tensions, include regulation

of apoptosis, angiogenesis, glycolysis and the cell cycle. A number of these downstream genes and products, have subsequently been investigated as putative biomarkers for radioresistance. Using quantitative PCR to evaluate HIF-1 expression in pre-treatment tumour biopsies, Toiyama et al demonstrated not only significantly lower levels of HIF-1 in those with significant tumour regression, but also of downstream induced factors EGFR and VEGF on both univariate and multivariate analyses. [245] Furthermore, decreased levels of these markers had a sensitivity and specificity of between 72.7 and 89.5% (dependent on marker) for predicting TRG.

The effects of hypoxia include further biochemical changes in the tumour microenvironment. CA9, coding for carbonic anhydrase isoenzyme 9, plays an important role in regulating extracellular pH under both hypoxic and normoxic conditions. Using IHC, this was found to be present at significantly lower levels in patients with a pCR or with microscopic foci of residual tumour only. [246] This study broadly examined several mechanisms of the adaptive response of CRC to hypoxia, additionally evaluating glucose transporter-1 expression. GLUT-1 expression had been shown in an earlier study by Brophy et al to be significantly associated with a poor response to RT; conversely, GLUT-1 negative tumours were found to have a 70% probability of a good response to RT. These findings were not corroborated by Guedj et al, however, who found no significant association between GLUT-1 expression and TRG. [209 246] It is worth noting that these two studies employed differing intervals between RT and surgery and used different TRG grading systems. Theoretically, GLUT-1 may have promise as a biomarker for CRC, as it is not constitutively expressed by normal colonic mucosa but is found in up to 80% of colorectal cancers. It is believed to be a late feature in the adenoma-carcinoma

sequence and has previously been demonstrated as a surrogate marker for tumour hypoxia in both bladder and cervical cancers. [247] [248]

A number of these mechanisms are summarised in **figure 4.3** – these are factors that have been found to correlate with response to RT, based on individual variations. Pathway A outlines how HIF-1 relates to other markers identified though these studies in the presence of hypoxia. At low oxygen tensions, HIF-1 is upregulated, with corresponding increased expression of downstream co-factors; EGFR is upregulated on the cell surface; carbonic anhydrases are released in order to stabilise extracellular pH, with CA9 particularly shown to be an independent predictor for RR; VEGF stimulates angiogenesis in the tumour microenvironment. GLUT-1 is not expressed in normal rectal mucosa but has been shown to be present in up to 80% of colon and rectal tumours .

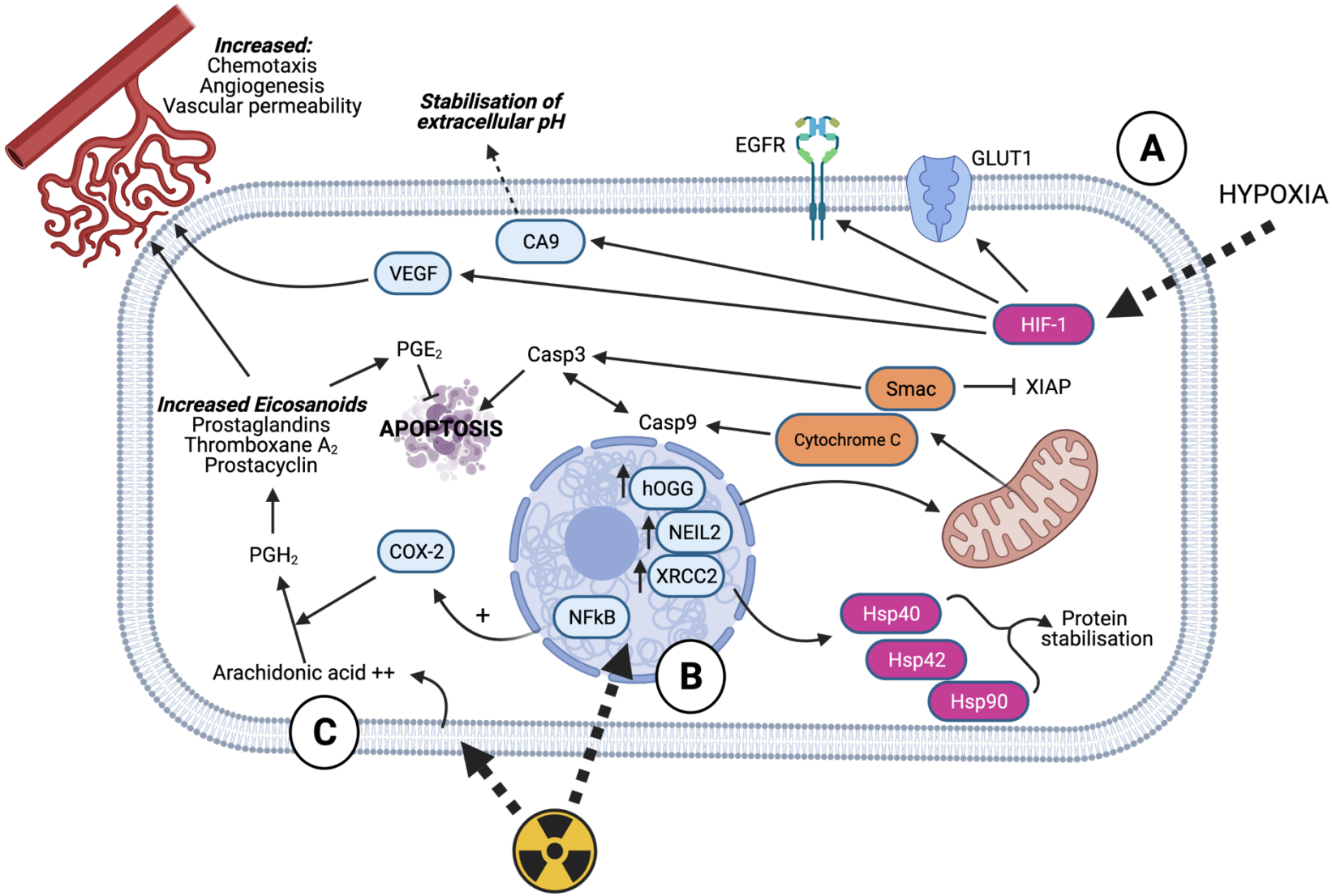


Figure 4.3 – Summary schematic of stress response pathways, amalgamating and linking the implicated biomarkers from this review which fall under the GO biological function for response to stress and DNA damage. Pathways A, B and C are complementary and work simultaneously. Note critical role of lipid mediators in inflammatory process following release of fatty acids form the cell membrane.

The pathways directed at mounting a local response to ionising radiation and chemotherapy, repairing DNA double strand breaks and stabilising protein structures, or activating apoptosis, are summarised in pathways B and C. In (B), direct DNA damage leads to activation of a number of DNA repair factors (hOGG, MTFHR, NEIL2) in addition to increased levels of heat-shock proteins (particularly Hsp40 (DNAJ), Hsp42 and Hsp90), increased levels of all of which have shown a direct relationship with RR. DNA damage stimulates release of cytochrome c and activated Smac proteins from the mitochondria, and subsequent activation of the apoptosome and direct inhibition of XIAP leads to apoptosis under normal circumstances. XIAP expression in turn has been demonstrated as an independent predictor of radioresistance. [249] Decreased relative levels of Smac were shown to be associated with RR. There may be an additional interplay with Smac contributing to the direct activation of caspase-3. Finally in pathway (C), ionising radiation leads to increased degradation of complex lipids in the cell membrane and subsequent release of higher levels of arachidonic acid. Higher levels of COX-2 have been shown to correlate with RR, with subsequent increased levels of eicosanoids and their downstream effects in both the intra- and extracellular environment. Within the cytosol, both COX-2 and eicosanoids have been shown to inhibit NF-kB, whilst in the tumour microenvironment PGs and thromboxane A_2 have well-established effects on chemotaxis, vascular permeability and angiogenesis as part of the local and systemic inflammatory process

The primary aim of RT is to cause lethal double-stranded DNA (DS-DNA) breaks resulting in cell death. A number of genes involved in base-excision repair pathways have already been well-studied. *NEIL2* belongs to a family of DNA glycosylases which initiate the first step in base exicision repair. Kobunai et al

demonstrated that NEIL2 was not only upregulated in radioresistant colorectal cancer cell lines, but also in pre-treatment clinical biopsies of patients who had a poor response to RT. [225] Equally implicated in the base excision repair pathway, polymorphisms within the *hOGG* gene - specifically the *hOGG1 1245C>G* polymorphism - have been shown to be associated with inferior TRG when extracted from peripheral blood and rectal cancer tissue in 238 rectal cancer patients. [250] It is thought that the 1245C>G polymorphism codes for a low-functioning isoform of the protein resulting in less DS breaks as part of excision repair, leading to a smaller proportion of cells undergoing cell death as part of lethal DS-DNA breaks. Additionally in this study, all patients were concurrently treated with fluoropyrimidine-based chemotherapy, with the *MTHFR 677C>T* polymorphism being found to be significantly associated with inferior TRG. *MTHFR* codes for methylenetetrahydrofolate reductase, the critical enzyme in the folate cycle which irreversibly reduces 5,10-methylenetetrahydrofolate to 5-methyltetrahydrofolate. This step involves methylation of deoxyuridylate to deoxythymidylate, and hence is susceptible to 5-FU. Polymorphisms in this gene have been shown variably to both improve and inhibit response to 5-FU across a number of cancers. Furthermore, in combination with the presence of the 1298C>A polymorphism, the haplotype *MTFHR 677T-1298A* was found to be an independent predictor of poor TRG, although this genomic data was acquired from peripheral blood only. [251]

The x-ray repair cross-complementing proteins 2 and 3 (*XRCC2 and XRCC3)* are involved in a separate DNA repair pathway (homologous recombinant repair) and have been separately evaluated. Both seem to have a potential role as a predictive marker of radioresistance mechanisms. Knockdown of *XRCC2 in vitro* resulted in

impaired repair of DS-DNA breaks and Qin et al concluded that intact XRCC2 proteins resulted in radioresistance owing to a functional DNA repair mechanism. [252] Agostini et al modelled a biological network centred on XRCC3 which accounted for a more complex interplay between itself and gene-targeting microRNA (miRNA) expression. This study indicates a possible role for XRCC3 in radioresistance mechanisms, in the context of a biological network, which had not previously been demonstrated when evaluating polymorphisms (SNPs) within the XRCC3 gene in isolation. [253]

In addition to nuclear repair mechanisms, DNA damage signals the release of cytochrome-c from the mitochondria, with subsequent activation of the apoptosome, as previously described. Concurrently, the injurious insult leads to activation of Smac proteins which are released from the mitochondria in conjunction with cytochrome-c. As well as preventing the inhibition of apoptotic promoters, it is postulated that Smac could also play a role in both activation of caspase-3 and enhance its catalytic effects. Correspondingly, increased expression of Smac has been found to correlate with improved TRG. [220 254] Within the cytosol, a number of small stabilising proteins (termed heat-shock proteins - Hsp) have been found to to have roles in stabilising protein structures and maintaining cellular homeostasis. The gene *DNAJC12* codes for a 40kDa Hsp, increased expression of which was shown to be significantly associated not only with with inferior TRG, but also increased vascular and perineural invasion and post-treatment nodal involvement. [255] Two further studies have reported significant correlation between increased expression of selected Hsps and poor TRG. Increased Hsp42 expression, verified by protein separation with 2D gel electrophoresis and matrix-assisted laser desorption/ionisation time of flight (MALDI-TOF) mass spectrometry was associated with radioresistance in 17 clinical specimens

prior to RT. [239] Another study reported significantly increased expression of hsp90 in pre-treatment biopsies of 33 of 64 patients recruited who had a poor response to RT. [216]

4.2.7 GO Biological Function - Cell signalling

Many of the factors already discussed play a role in both cell-cell signalling as well as intracellular signal cascade mechanisms. Cell-surface signalling molecules have long been established as molecular fingerprints for immunophenotyping of cancer subtypes, such as the family of clusters of differentiation (CD). This large class of molecules act as receptors or ligands, which either initiate a signalling cascade once activated, or play a role in processes such as cell-cell adhesion, [210 230 246 256-258] although the evidence for these correlating with radioresistance has been inconclusive (see **table 4.1**). CXCL10 (CXC motif chemokine 10) is the ligand for the CXCR3 receptor, although this is found predominantly on T-cells. [259] Presence of increased expression of CXCL10 mRNA in pre-treatment rectal cancer biopsies was shown to be predictive of sensitivity to RT, although IHC for the chemokine itself was not significantly correlated. [260] CXCL10 has been shown to act as an angiostatic, inhibiting VEGF, [260] but the correlation between mRNA expression and protein expression in this case, and hence the implication for CXCL10 to act as a surrogate biomarker, is not clear.

Although numerous adhesion molecules on the cell surface also play a major role in tumour-tumour microenvironment interactions, only the glycoprotein thrombospondin 2 (THBS2) has been shown to be significantly associated with

response to radiotherapy. [210] THBS2 demonstrates anti-angiogenic effects upon interaction with stromal endothelial cells, and low or absent levels of THBS2 expression in colorectal cancer biopsies was demonstrated by Lin et al to be linked to poor tumour regression, although 118 of the 172 specimens examined fell into an intermediate TRG (Dworak TRG 2-3), with a near 50-50 split between high and low expression of THBS2 on IHC. [261]

Biomarker	Function	Cell types	Studies	Association with LCCR response
CD34	Cell-surface glycoprotein Cell-cell adhesion	Haematopoietic stem cells	*Min et al 2008*; IHC of pre-treatment biopsies [256]	No significant correlation
CD44	Cell-surface glycoprotein Cell-cell adhesion and migration	Some cancer stem cells Most mammalian cell lines	*Huh et al 2014*; RT-PCR and Western Blots for 13 markers including those for angiogenesis, apoptosis, proliferation, cell adhesion and collagenases [230] *Sim et al 2014*; IHC for markers of cell cycle, proliferative index, apoptosis, cell adhesion and response to hypoxia [210]	Low CD44 correlates with improved TRG (OR 4.694, p=0.027) No significant correlation
CD133	Cell-surface glycoprotein Uncertain function	Haematopoietic stem cells Cancer stem cell cells (brain, breast, liver, medulloblastoma, melanoma) Glandular tissue cells (including GI tract)	*Shinto et al 2011*; IHC for CD133, COX-2, p53, p27, p21, EGFR in pre-treatment biopsies of patients undergoing short-course LCCR (20Gy/5 fractions) [257] *Sim et al 2014*; IHC for markers of cell cycle, proliferative index, apoptosis, cell adhesion and response to hypoxia [210]	CD133 expression correlates with RR (OR 3.32, p=0.03) No significant correlation with pCR
CD166	Cell-surface glycoprotein Putative roles in cell-cell adhesion and activation of matrix metalloproteinase cascades	Neurons, fibroblasts, endothelial cells, keratinocytes	*Sim et al 2014*; IHC for markers of cell cycle, proliferative index, apoptosis, cell adhesion and response to hypoxia [210]	High CD166 expression associated with poor prognosis (DFS) but not specifically with TRG
CD184 (CXCR4)	Chemokine receptor Lymphocyte chemotaxis	Haematopoietic stem cells Over 20 cancer subtypes (including colorectal)	*Guedj at al 2011;* IHC for panel of biomarkers by function; CXCR4 believed to have role in tumour cell migration and metastatic potential [246]	No significant correlation with TRG
CD340 (HER-2, ERBB2)	EGFR family cell-surface protein Activation of signal transduction pathways, including MAPK and PI3K/Akt	Important biomarker for breast cancer Upregulation of *ERBB2* gene associated with certain breast, lung, gastric, uterine and ovarian cancer cells	*Drebber et al 2011;* IHC for both *Her-2/neu* and beta-catenin and correlated with TRG [262] *Meng et al 2014*; IHC and FISH for HER-2 in pre-treatment biopsies [258]	No association of *Her2* expression with TRG, although *Her2* positivity associated with a survival benefit No association between HER-2 expression in tumours and TRG; positive HER-2 status associated with poorer 5 year survival

Table 4.1 - Association of cell-surface clusters of differentiation (CD) molecules with response to LCCR. In the study by Guedj et al, CD184 labelled as CXCR4 (C-X-C chemokine receptor 4). CD340 is more commonly referred to as HER-2. OR = odds ratio

As with the cellular response mechanisms to stress previously described, expression levels of a number of receptor tyrosine kinases (RTKs) have been shown to increase in CRC and may be implicated in radioresistance. The downstream effects of activation of RTKs and the MAPK pathway on gene regulation, cellular proliferation and apoptosis as described, have stimulated interest in investigating the expression of these receptors and their associated proteins in the context of response to RT. Holck et al examined IQGAP1 as a potential biomarker for response on the basis of it being an essential scaffold protein in ERK signalling, which is a key step in the MAPK cascade. [263] It is known that MAPKs are activated - and MAPK signalling upregulated - in response to a variety of stimuli. [264] Although this study did not evaluate levels of IQGAP1 in pre-treatment biopsies as a stand-alone predictor of response to RT, the authors found that apical cell IQGAP1 IHC staining increased most dramatically between pre- and post-treatment in those patients with a poor TRG (<50% regression). [265] Components of other signalling pathways activated by membrane receptor binding (notably, the Wnt and TGF-β signalling cascades) have failed to demonstrate conclusive results. SMAD3 transcriptionally regulates TGF-β target genes upon accumulation in the nucleus, and SMAD3 mutations in murine models have been linked to the development of CRC. [266] The phosphorylated isoform of SMAD3 is considered as a hallmark for SMAD3 activation, increased expression of which has been shown to be associated with a poor TRG. [267] β-catenin can be detected at the membranous, cytoplasmic and nuclear levels on IHC, representing its function within the Wnt cascade, and counts activation of genes linked to the *APC* gene amongst its downstream effects. Wang et al demonstrated a shift in outcome dependent on nuclear vs. cytoplasmic expression of β-catenin, with increased nuclear expression being association with radioresistance. [268] However, conflicting evidence was provided

in two other studies; no significant association was found between membranous β-catenin expression and the Dworak TRG, but when correlated with the Cologne grading system statistical significance was reached on univariate analysis; [262] a further multivariate analysis of 130 samples by IHC also demonstrated no significant correlation of β-catenin with Mandard TRG. [211] By contrast, in a single study, expression of T-cell factor 4 (TCF4) - another downstream mediator of the Wnt cascade and implicated in the adenoma-adenocarcinoma sequence – demonstrated inverse correlation with TRG. [269]

As previously described, epithelial growth factor receptor (EGFR) is one of a group of RTKs which are activated and upregulated by the stabilisation of HIF-1. As with several other studies, IHC evaluation has provided conflicting results, with only the study by Carlomagno et al (out of a total of three IHC studies identified) demonstrating a link between low EGFR expression and pCR. [210 214 218] Evaluation of gene expression levels with qPCR by Toiyama et al corroborated these results, also demonstrating a significant association between low expression levels of the *EGFR* gene and increased tumour regression. [245] Interestingly, a high *EGFR* gene copy number was found to be predictive of improved TRG when combined with the *k-RAS* wild type, but only in the context of cetuximab administration with standard 5-FU/50.4Gy chemoradiotherapy. [270]

In addition to RTKs, expression of both transporter molecules and G-protein coupled receptors, namely ATP-binding cassette subfamily C member 4 (ABCC4) and growth hormone releasing-hormone receptor (GHRH-R) have been evaluated. Functionally, ABCC4 has been shown to be implicated in the transport of cyclic AMP

(cAMP) and glutathione conjugates and is additionally known as multi-drug resistance protein 4 (MRP4). It has previously been implicated in resistance to a multitude of nucleotide analogue agents and has been associated with a poorly prognostic phenotype in both CRC cell lines and microarray analysis. Previous work by Yu et al screened high-throughput gene chips of CRC cell lines and identified *ABCC4* as a key differentiating gene between responders and non-responders, with subsequent work employing IHC staining for ABCC4 demonstrating a significant correlation between upregulated ABCC4 expression and radioresistance, on both uni- and multivariate analysis. [271] Furthermore, not only was high MRP3 expression found to correlate with a poor TRG on IHC of 144 pre-treatment rectal cancer biopsies, but equally knockdown of *MRP3* using siRNAs in radioresistant CRC cell lines improved both chemo- and radiosensitivity, confirmed in murine models. [272]

When bound to its receptor, GHRH stimulates local tissue proliferation via release of insulin-like growth factor 1 (IGF-1) and has been demonstrated as a feature of tumour aggressiveness, with effects on both malignant transformation and metastasis. [273] Work on CRC cell lines has demonstrated that when administered in combination with a cytotoxic agent, GHRH antagonists can induce S-phase arrest and thus apoptosis in part by inhibition of these mechanisms. [274] Correspondingly, expression of GHRH-R in pre-treatment rectal cancer biopsies has been found to correlate with response to RT, with increased expression being predictive of a poor response, and vice versa. [216]

4.2.8 GO Biological Function - Cellular metabolism – nucleic acids and lipid products

The cytotoxic backbone of the majority of LCCR regimens is 5-FU, which functions as an antimetabolite and pyrimidine analogue, irreversibly inhibiting the action of thymidylate synthase (TS). TS, in turn, plays a key role in nucleic acid metabolism, catalysing the conversion of deoxyuridine monophosphate (dUMP) to deoxyuridine TMP, a key nucleoside required in DNA replication. Most groups have found an inverse correlation between TS expression and TRG, [219] [230] [275-277] although three studies did not find any significant correlation. [210] [214] [278] Furthermore, microarray analyses of published rectal cancer transcriptomic datasets have identified candidate genes involved in both nucleotide and amino acid metabolism which may be implicated in radioresistance, including CPS1 (carbonyl phosphate synthetase 1, an enzyme involved in glutamine metabolism and in catalysing the initial steps in ammonia detoxification). CPS1 was identified not only as the most significantly upregulated gene in the glutamine pathway in radioresistant CRC, but additionally IHC staining of clinical specimens demonstrated significant correlation between CPS1 expression and radioresistance. [279]

Accumulation of both aspartate and glutamine in asparagine synthetase (ASNS)-deficient tissues leads to an increase in nucleotide synthesis and thus cellular proliferation. [280] Consequently, a significant correlation between low ASNS expression and radioresistance was observed in clinical rectal cancer specimens, following identification of both the *ASNS* and *PAH* genes as being differentially expressed between responders and non-responders in a transcriptomic dataset. [281]

Aberrant gene methylation, notably pathways linked to CIMP, is one of numerous genetic alterations, which may predispose to tumourigenesis and alter the biology of invasive cancer. Hyper- or hypomethylation can lead to the silencing of tumour suppressor genes or activation of oncogenes respectively. A study evaluating the methylation status of 24 tumour suppressor genes, found no consistent evidence to implicate gene methylation as a surrogate marker for response to RT, although there was a significant association between the methylated *TIMP3* (TIMP metalloproteinase inhibitor 3) gene and TRG. [282] However, the levels of methylated *TIMP3* did not subsequently correlate with TIMP3 protein expression, nor did TIMP3 protein expression correlate with TRG, such that the authors concluded that it is unlikely to serve as a principal biomarker for tumour response.

As described, the distinguishing features of invasive cancer involve a multitude of genetic and proteomic anomalies which in turn have profound downstream effects on normal cellular metabolism and bioenergetics. Metabolic phenotyping approaches (metabonomics/metabolomics) are showing increasing promise in elucidating the metabolic factors that govern radiosensitivity. Modifications to lipid biosynthetic pathways and lipid metabolism are now established as hallmark features of cancer, [94] and intracellular lipid mediators have now been implicated in mechanisms of response to ionising radiation and cytotoxicity. Through analysis of the gene expression database of rectal cancer tissue published by Watanabe et al, *HSD17B2* and *HMGCS2* were identified as key lipid biosynthesis-associated genes with a predictive capacity for response to RT. [283] The enzyme product HSD17B2 (hydroxysteroid 17-beta dehydrogenase 2) is known to catalyse the reduction of E2 oestrogens, testosterone to androstenedione and 5α-androstenediol to dehydroepiandrosterone.

[284] HMGCS2 (3-hydroxy-3-Methylglutaryl-CoA Synthase 2) catalyses the formation of HMG-CoA from both acetyl-CoA and acetoacetyl CoA, a key early step in the malevonate and ketogenesis pathways. This is in-keeping with the Warburg phenomenon consistently displayed by solid cancers, where in addition to the (inefficient) aerobic glycolysis there also exists a propensity for anaerobic fermentation of ketone bodies, even in the presence of abundant oxygen. [285] Overexpression of both genes and their products was found to be associated with poor response to RT. Previous work by Yeo et al had additionally demonstrated a significant association between increased HMGCS2 expression and poor TRG on IHC, although knockdown of *HMGCS2* with siRNAs in radioresistant CRC cell lines did not result in improved radiosensitivity. [286]

Cyclo-oxygenase-2 (COX-2; prostaglandin endoperoxidase synthase) converts arachidonic acid to prostaglandin-H_2, which is the precursor to all other prostanoids which act principally as inflammatory mediators. [287 288] For this reason, and due to the fact that arachidonic acid is released from the cell membrane following direct damage by ionising radiation, a number of groups have evaluated the role of COX-2 in radioresistance and its ability to act as a predictive RT biomarker. [206 208 211 256 257 289 290] In addition to the functional effects of these bioactive lipids (angiogenesis, increased vascular permeability and chemotaxis - Figure 2), COX-2 has also been shown to directly stimulate VEGF synthesis. [256] This supports a theory of COX-2 playing an important role in the response to the cellular injury sustained following RT, although whether this is purely due to an increased amount of substrate (i.e., arachidonic acid) because of cell membrane damage or an additional upregulated pathway is yet to be fully elucidated. A cell-wide inflammatory response mechanism,

portended by upregulation of nuclear factor kappa-B (NFκB) in the nucleus which in turn acts as a transcription factor for, amongst others, COX-2, TGF- β, VEGF and EGFR, has been proposed. [291] However, the two studies identified through this review which evaluated NFκB as a predictive biomarker revealed conflicting results. Nuclear IHC staining for NFκB in 74 pre-treatment samples revealed a correlation between upregulated NFκB and poorer overall survival, but not with TRG. [292] This may have been partly due to Type I error however, as a larger study in 2014 found that lower NFκB expression also correlated with a better response to RT. [230] Interestingly, although a number of groups have identified annexin 1 (ANXA1) as a both a potent anti-inflammatory mediator, and a calcium ion/phospholipid binder that has effects on tumourigenesis and progression, Sheu et al found that high levels of expression of ANXA1 in fact correlated with inferior TRG, despite previous evidence to suggest that ANXA1 is an endogenous inhibitor of NFκB. [293] [294]

4.2.9 GO 'knowledge network' mapping

The data obtained from the review process were extracted in such a way as to assign an identifier to each biomarker assessed by each study. This generated a list of biomarkers with multiple entries for those which had been the subject of investigation for more than one study. Equally, to account for statistical power, separate entries were made for each statistical test (uni- or multivariate) applied to each biomarker. The full data extraction table is provided in **Appendix A**.

By using GO terms, a hierarchical classification for each biomarker was built which allowed for clustering as per the subheadings ("*Biological Functions"*) used in

the above sections. A network model was then developed in both static (**Figure 4.4**) and html format. In the model below, nodes are annotated according to their GO term and are clustered in spokes around the root term ‘Biological Process’. As per the methods, nodes are sized according to a significance score for that Ontology term, calculated by taking into account the various statistical tests applied to each study (most commonly, chi-squared association between biomarker and TRG in univariate analysis, and multivariate odds ratios). The size of the node represents the number of biomarkers under that GO term that have been investigated, and the colour scale indicates the relative aggregate significance level. Therefore, a large, red node would represent a GO term with both a large number of biomarkers and high statistical significance. A summary table of biomarkers, their annotations, subsequent node size and the HSB colour space level is shown in **table 4.2**.

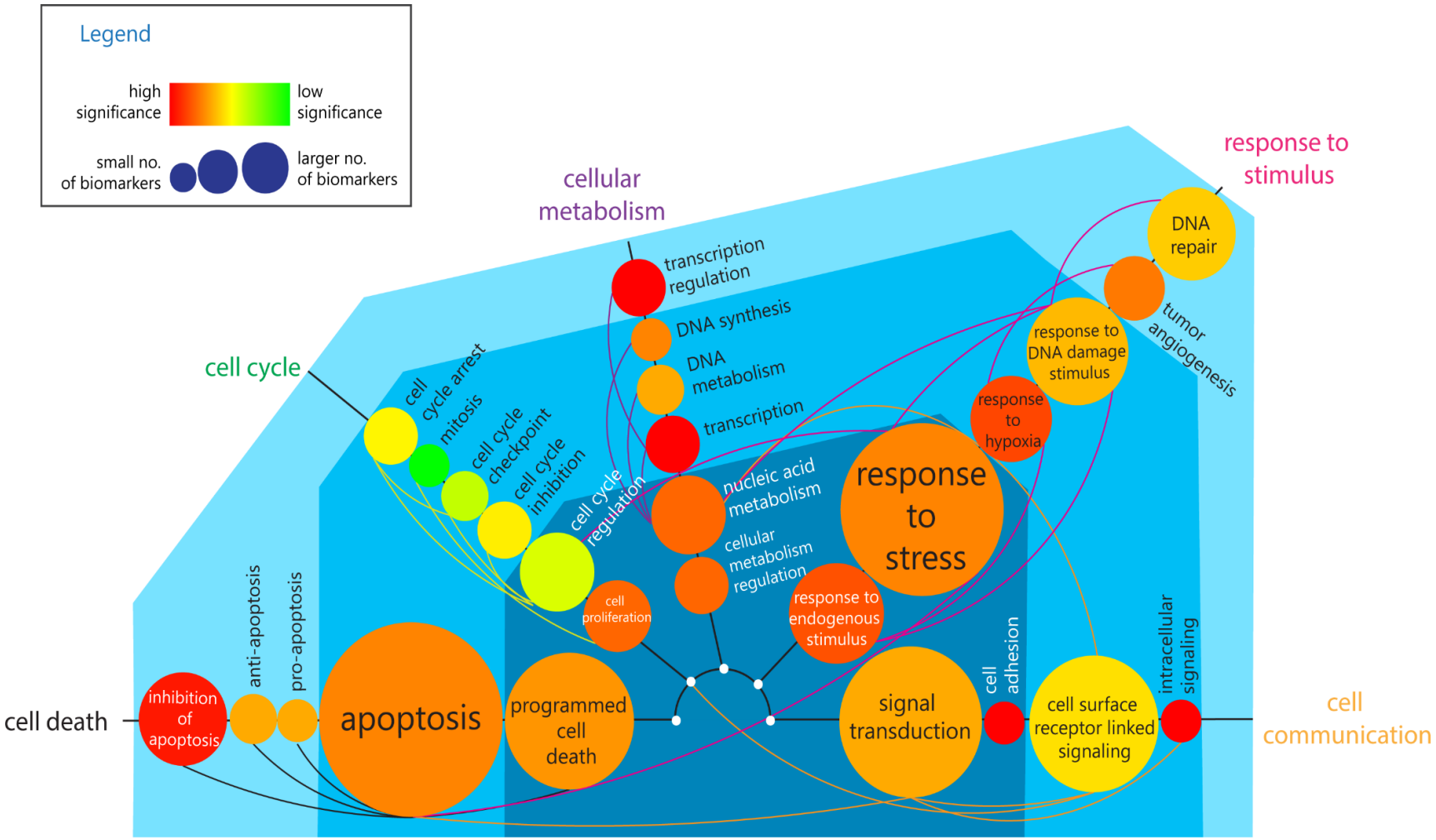


Figure 4.4 - Static knowledge network visualisation of biomarkers clustered according to Gene Ontology terminology, as per data from table 1. Each node represents a Gene Ontology (GO) Term, clustered under group headings (spokes). Size of nodes was calculated based upon the number of biomarkers under that GO term, with hue/saturation/brightness mapped to aggregate significance of multiple statistical tests employed across studies.

GO Cluster	GO Node	Biomarkers	Node size	Node significance (HSB)
Cell cycle & cell proliferation	Cell cycle arrest	BIRC5 [225 228 229 295] , p21 [206 210 214], p53 [206-210 214]	29.42857	57.77778
	Cell cycle checkpoint	p21 [206 210 214], p53 [206-210 214]	25.71429	74.28571
	Cell cycle inhibition	p27 [206 217], BIRC5 [225 228 229 295]	29.42857	29.42857
	Cell cycle regulation	BIRC5 [225 228 229 295], p21 [206 210 214], p27 [206 217], p53 [206-210 214], 119 phosphor-Akt [216]	40.57143	68.8
	Cellular proliferation	Ki67 [218-220], REG4 [225 226], Securin [221], YKL-40 [233]	36.85714	24
	Mitosis	β-tubulin [239]	22	120
Cell death	Programmed cell death	Apoptotic index [228 237], Bax [206 210 234], BIRC5 [225 228 229 295], Caspase-8 [213], hPEBP4 [240 242], M30 [238], MIB1 [214], p53 [206-210 214], PDCD4 [236], phospho-Akt [216], XIAP [249], YKL-40 [233]	70.28571	35
	Apoptosis	Apoptotic index [228 237], Bax [206 210 234], BIRC5 [225 228 229 295], Caspase-8 [213], hPEBP4 [240 242], M30 [238], MIB1 [214], p53 [206-210 214], PDCD4 [236], [119] phospho-Akt [216], XIAP [249], YKL-40 [233], CD34 [256], Calsenilin [239], Smac [220 254]	100	31.11111
	Pro-apoptosis	Bax [206 210 234]	22	40
	Inhibition of apoptosis	BIRC5 [225 228 229 295], hPEBP4 [240 242], [119] phospho-Akt [216], XIAP [249], YKL-40 [233], HIF-1 [245], HIF-1a [210]	48	5.714286
Response to stimuli, stress & DNA damage	Response to endogenous stimulus	CA9 [246], c-met [233], CXCL10 [260], MSH2 [214], [119] phospho-Akt [216], Thymidine phosphorylase [256], VEGF [206 211 212 214 220 245 256 296 297], YKL-40 [233]	51.71429	18.94737
	Response to stress	ANXA1 [293], BIRC5 [225 228 229 295], DIABLO [249], DNAJC12 [255], HIF-1 [245], HIF-1a [210], hOGG1-1245C>G [250], MTHFR 677T-1298A [251], MTFHR-677C>T [250], Hsp42 [239], Hsp90 [216], m-TIMP3 [282], NEIL2 [225], NFkB [230 292], p21 [206 210 214], p53 [206-210 214], sHsp16.2 [216], Smac [220 254], XRCC2 [252], Tropomodulin [239]	88.85714	31.11111
	Response to hypoxia	HIF-1 [245], HIF-1a [210], CA9 [246], c-met [233], CXCL10 [260], [119] phospho-Akt [216], Thymidine phosphorylase [256], VEGF [206 211 212 214 220 245 256 296 297], YKL-40 [233]	44.28571	16.47059
	Response to DNA damage stimulus	MSH2 [214], MTHFR 677T-1298A [251], BIRC5 [225 228 229 295], hOGG1-1245C>G [250], MTFHR-677C>T [250], NEIL2 [225], NFkB [230 292], p53 [206-210 214], XRCC2 [252]	55.42857	43.47826
	DNA repair	MSH2 [214], BIRC5 [225 228 229 295], hOGG1-1245C>G [250], MTFHR-677C>T [250], NEIL2 [225], p53 [206-210 214], XRCC2 [252]	48	48
	Tumour angiogenesis	VEGF [206 211 212 214 220 245 256 296 297], [119] phospho-Akt [216], Thymidine phosphorylase [256], YKL-40 [233]	33.14286	28.57143
Cell communication	Signal transduction	ABCC4 [298], Bcl-2 [206 211], β-catenin [211 262 268], CD44 [210 230], CD133 [210 257], CD166 [210], c-met [233], CXCR4 [246], EGFR [210 214 218 245], EpCAM [210], GHRH-R [216], HER-2 [211 258], IQGAP1 [265], MRP3 [272], TGF-β1 [267], THBS2 [261]	77.71429	40
	Cell-surface receptor signalling	ABCC4 [298], CD34, CD44 [210 230], CD133 [210 257], CD166 [210], CXCR4 [246], EGFR [210 214 218 245], EpCAM [210] , GHRH-R [216], HER-2 [211 258], MRP3 [272], TGF-β1 [267], Calsenilin [239]	70.28571	52.63158
	Intracellular signalling	IQGAP1 [265]	22	0
	Cell adhesion	THBS2 [261]	22	0

Cellular metabolism	Regulation of cellular metabolism	COX-2 [206 208 211 256 257 289], ALDH1 [210 299], HMGCS2 [283 286], HSD17B2 [283], MMP9 [296], [120] phospho-Akt [216], GLUT-1 [209 246], CPS1 [279], Asparagine synthetase [281]	29.42857	24
	Nucleic acid metabolism	p-SMAD3 [267], SMAD3 [267], TCF4 [269], Thymidine phosphorylase [256], Thymidylate synthase [210 214 218 219 230 275-278]	40.57143	24
	DNA synthesis	Thymidylate synthase [210 214 218 219 230 275-278]	22	31.11111
	DNA metabolism	Thymidine phosphorylase [256]	25.71429	40
	Transcription	p-SMAD3 [267], SMAD3 [267], TCF4 [269]	29.42857	0
	Regulation of transcription	p-SMAD3 [267], SMAD3 [267], TCF4 [269]	29.42857	0

Table 4.2 - Biomarkers classified according to Gene Ontology hierarchical tree (biological functions). Node size was calculated based upon number of biomarkers within each GO term, accounting for sample size in each study. The number of references for each biomarker reflects the number of studies retrieved. Significance was mapped to the HSB colour space to reflect overall significance of each node in the resulting network, with a lower relative value reflecting greater significance level based upon multiple statistical tests used across studies. The HSB colour space (Hue, Saturation, Brightness) was selected to run from 0^{o} (red) through 120^{o} (green)

4.2.10 Discussion

The generation of a knowledge-based network analysis is a novel way of displaying data acquired from systematic review. The methodology employed here is the first such analysis of its type to be utilised for the research question presented and offers an intuitive method of visually appraising statistical data gathered from published experimental work, by mapping identified biomarkers using Gene Ontology terms. One benefit of such analysis is that it can consider all published data, including non-significant findings, and aid in identifying gaps in knowledge and potential targets for future work. Clearly, the true value of such a network is in interactive, live format that can be freely accessed as an online resource and continually updated.

Previous narrative reviews have drawn minimal conclusions on this topic, not least owing to the heterogeneity of studies available. In part, this is influenced by qualitative clinicopathological means of assessment. For example, in this review, 5 different clinical tumour regression grading (TRG) systems were used across studies, whilst some studies instead assessed for pCR or incomplete response only. Even allowing for similarities in the fundamentals of these regression grading systems, there was further disparity in the way the regression grades were dichotomised into 'good response' and 'poor response' across groups (see **table 4.3**). The use of pCR as the target endpoint may not be justified in the current practice of delivery of RT as a treatment modality; given that the aim of RT is to downsize the tumour and sterilise the circumferential resection margin (rather than an attempt at being outright curative), it would be difficult to justify not offering neoadjuvant therapy to patients with LARC on the basis that they would only achieve 50% regression rather than pCR.

Dworak et al. [300]	Mandard et.al [301]	Rödel et al. [302]	Ryan et al. [303]	Wheeler et al. [304]
0. No regression 1. Predominantly tumour with significant fibrosis and/or vasculopathy 2. Predominantly fibrosis with scattered tumour cells (slightly recognisable histologically) 3. Only scattered tumour cells in the space of fibrosis with/without acellular mucin 4. No vital tumour cells detectable	1. Complete regression (fibrosis without detectable tissue of tumour - pCR) 2. Fibrosis with scattered tumour cells 3. Fibrosis and tumour cells with preponderance of fibrosis 4. Fibrosis and tumour cells with preponderance of tumour cells 5. Tissue of tumour without changes of regression	0. No regression 1. Regression of <25% of tumour mass 2. Regression of 25-50% of tumour mass 3. Regression of >50% of tumour mass 4. Complete regression (pCR)	1. No viable cancer cells 2. Residual cancer outgrown by fibrosis 3. Significant fibrosis outgrown by cancer	1. Sterilisation or only microscopic foci of adenocarcinoma remaining, with marked fibrosis 2. Marked fibrosis but macroscopic disease present 3. Little or no fibrosis, with abundant macroscopic disease

Table 4.3 – Definitions of the Tumour Regression Grading (TRG) systems used across the studies identified in this review

Another limitation when comparing studies is the fact that the majority of those included were based on immunohistochemistry, which remains a subjective assessment even when two blinded assessors are used. The disparity in results between studies may be in part due to difficulties in the reproducibility of immunohistochemistry, particularly when staining for novel biomarkers. Reviews of Western blot data for antibodies demonstrate that targeted binding can vary widely between manufacturer and antibody preparation. Indeed, for those studies which employed more than one experimental method, there was occasional disparity between IHC and more quantitative methods such as gene expression and qPCR.

The summary network figure demonstrates several features. Firstly, based on the reviewed studies, genes and protein products involved in regulation of cell cycle have a minimal role to play in the modulation of RT response, nor do expression profiles appear to be correlated significantly with the clinical response to RT. A larger number of pathways involved in response to stress and ultimately cell death have been found to be significant, and there is significant crossover between these pathways. Intracellular signalling and cell adhesion with downstream cellular metabolism represent the areas with relative paucity of data but demonstrate promise with the biomarkers assessed being shown to demonstrate higher levels of aggregated significance. That these GO root processes (Cellular Metabolism, Cell Communication) and their associated biosynthetic pathways are so influenced by complex lipids justifies further investigation. In the section that follows, I will perform a further *in silico* analysis of a legacy microarray dataset identified as part of the literature review for this section to elucidate a yet unidentified niche for lipid metabolism as part of the modulation of the RT response.

4.3 Gene Set Enrichment Analysis of a legacy gene expression dataset

4.3.1 Introduction and rationale

The review and meta-synthesis of biomarker data as laid out in the previous section provides insight into the current landscape of knowledge regarding possible mechanisms of radioresistance in RC. The Gene Ontology mapping analysis further establishes that the 'tree' biological functions of 'response to stimuli' and 'cellular metabolism' to be significant in their correlation to a putative mechanism of radioresistance. Given the role for bioactive lipids in both these domains, the next phase of this work was to establish whether legacy sequencing datasets might also demonstrate a role for lipid metabolism and complex lipids in the radioresistance process.

Beginning in 1990, the Human Genome Project set out to sequence and map all genes that comprise *Homo sapiens*. Technological advancement since those early days has led to the development of ever-more sophisticated approaches, broadly termed the *NextGen* sequencing technologies. Emerging in the late-1990s, microarray platforms provide a means of measuring expression levels of mRNA from thousands of genes across multiple samples. [305] Furthermore - and in-line with the spirit of the Human Genome Project - a condition of publication of microarray experiments in peer-reviewed journals has been to make the dataset publicly available on platforms such as the National Centre for Biotechnology Information (NCBI). With reference to cancer, The Cancer Genome Atlas (TCGA) further represents a publicly accessible repository of gene expression data for interrogation with an appropriate hypothesis.

The inherent challenge in working with gene expression data is the extraction of meaningful biological insight, without generating the false positives common with univariate analysis of such massive datasets. In the context of cancer research, the Gene Set Enrichment Analysis (GSEA) method laid out by the Broad Institute of Harvard and the Massachusetts Institute of Technology (MIT) provides a powerful analytical method which can determine whether an *a prioiri* defined set of genes demonstrates statistically significant differences between disease states or phenotypes – in this case, the presence or absence of ionising radiation as a treatment modality. [306-308] It also provides for data analysis in the context of known biological pathways; for example, those laid out in the Kyoto Encyclopaedia of Genes and Genomes (KEGG). [309] In this next stage of my pilot work, I aimed to analyse a legacy microarray dataset to look for enrichment of candidate genes implicated in lipid metabolism, between irradiated and non-irradiated rectal cancer tissue.

4.3.2 Search methodology

As part of the literature search outlined in section 4.2.1, appropriate legacy gene expression studies were interrogated where they employed a microarray sequencing technology. Inclusion limits were set to a minimum of 10 patients per cohort and these were classified as rectal cancer with neoadjuvant radiotherapy delivered (RC RT+), rectal cancer without neoadjuvant radiotherapy (RC RT-), normal matched rectal mucosa in the radiotherapy field (N RT+) and normal matched rectal mucosa in the non-irradiated patients (N RT-). The Gene Expression Omnibus (GEO) on NCBI was further searched again using the search parameters and Boolean logic as set out in 3.2.1. [310] The study by Snipstad et al 2010 fulfilled these criteria and was

based on GSE15781. These samples were frozen macro-dissected and microarray analysis run with The Human Genome Survey Microarray v.2.0 (Applied Biosystems Inc., USA) with 32,878 probes for 29,098 genes chip. [311] [312]

4.3.3 Data processing methodology

The GSEA platform from the Broad Institute was used for the bioinformatic analysis of the microarray data. This provides the software and computational method for the analysis and is run within the online resource. Following identification and selection of the above dataset, the software requires the end-user to classify the data into a number of predefined files as follows:

1. **Expression dataset** – the entire dataset comprising multiple samples within the GSE file (GSM396309 – GSM396350 inclusive) where each sample ID is a separate tissue sample with corresponding features (i.e., genes) and expression values. In the case of GSE15781, this dataset was based on 21 patients; 9 who had nRT and 13 without. Matched normal tissue was taken from each patient yielding a total of 42 samples.
2. **Phenotype labels** – in the case of the original study and the GEO data file these were classified under two factors; rectal tumours and normal rectal tissue, and further subdivided by treatment protocol (i.e., CRT or control). For consistency throughout this work the labels were amended on classification of the datafiles for GSEA into RC RT+, RC RT-, N RT+, N RT-
3. **Gene sets** – the Molecular Signatures Database (MSigDB – v5.2 accessed and used for this analysis) represents a collection of annotated gene sets for use with the GSEA. It is divided into 9 major collections including curated,

canonical biological pathways - for the purpose of this analysis the KEGG pathways for glycerophospholipid metabolism, arachidonic acid metabolism and fatty acid synthesis were interrogated in the KEGG online resource and selected for use. These were downloaded as separate .gmt files and used for independent analyses.

4. **Microarray annotations ('chip annotations')** – This file contains a list of the probes and their corresponding Human Genome Organisation (HUGO) symbols. This enables alignment of the microarray to the gene set as probe positions can vary between chips. Again, this was downloaded from the GSEA resources for use with *MSigDB v5.2*.

The GSEA method thereafter comprises three steps: calculation of an enrichment score (ES), estimation of the significance of the ES and adjustment for multiple hypothesis testing. In addition to the method as proposed by Subramanian et al the following adjustments were made to ranking metrics as per the optimisation study by Zyla et al: [307 313]

1. Weighted signal enrichment (WSE)
2. Exclusion of sets <100 and >500
3. Normalisation by enrichment scores (NES)
4. Genes ranked by *'Signal2Noise'* metric
5. False-discovery rate (FDR) of <0.25 as per index publication
6. Bonferroni correction made to set a significance level of $1.5E^{-6}$ by dividing by ty the number of degrees of freedom (i.e., 32,878 probes)

The groups of interest were set as RC RT+ *versus* RC RT- and N RT+ *versus* N RT- and comparative analysis was performed with Mann-Whitney U-test due to the data being deemed non-parametric. Heat maps were generated for genes mapped to the above selected KEGG pathways for visualisation of expression values.

4.3.4 Results

A test analysis was initially run by collapsing the expression dataset in *MSigDB v5.2* into gene symbols across the entire curated KEGG pathway gene set (c2.cp.kegg.v5.2), permuting by phenotypes RC RT+ and RC RT-. For RC RT+, 31 gene sets were significant at FDR <25%, with 19 gene sets significantly enriched with $p < 0.05$. In RC RT-, there were no gene sets significantly enriched at FDR < 25% but 7 significantly enriched with $p < 0.05$. Enriched gene sets were further examined for biological feasibility in rectal cancer carcinogenesis, progression and therapeutics. The GPL metabolism pathway demonstrated enrichment and targeted analysis was performed.

The dataset was then collapsed into the KEGG_ GLYCEROPHOSPHOLIPID_ METABOLISM gene set with the parameters as above, permuted by phenotypes RC RT+ and RC RT-, with RC RT- demonstrating enrichment across the GPL metabolism gene set. The enrichment plot is shown in **figure 4.5**. The genes were ranked by NES and an expression heatmap generated as per **figure 4.6**. Significantly upregulated genes were mapped onto the full KEGG GPL pathway for visual representation within the biological pathway **(Appendix B)**. Significantly upregulated and downregulated genes were further investigated by review of encyclopaedic databases with reference

to human colorectal cancer. [314] Log-fold changes were calculated for median expression values of selected genes of interest between RC RT+ and RC RT- and paired T-tests applied to estimate significance, the results shown in **table 4.4**. For genes shown, those depleted in RT+ to RT- demonstrate core enrichment when modelled RT- to RT+.

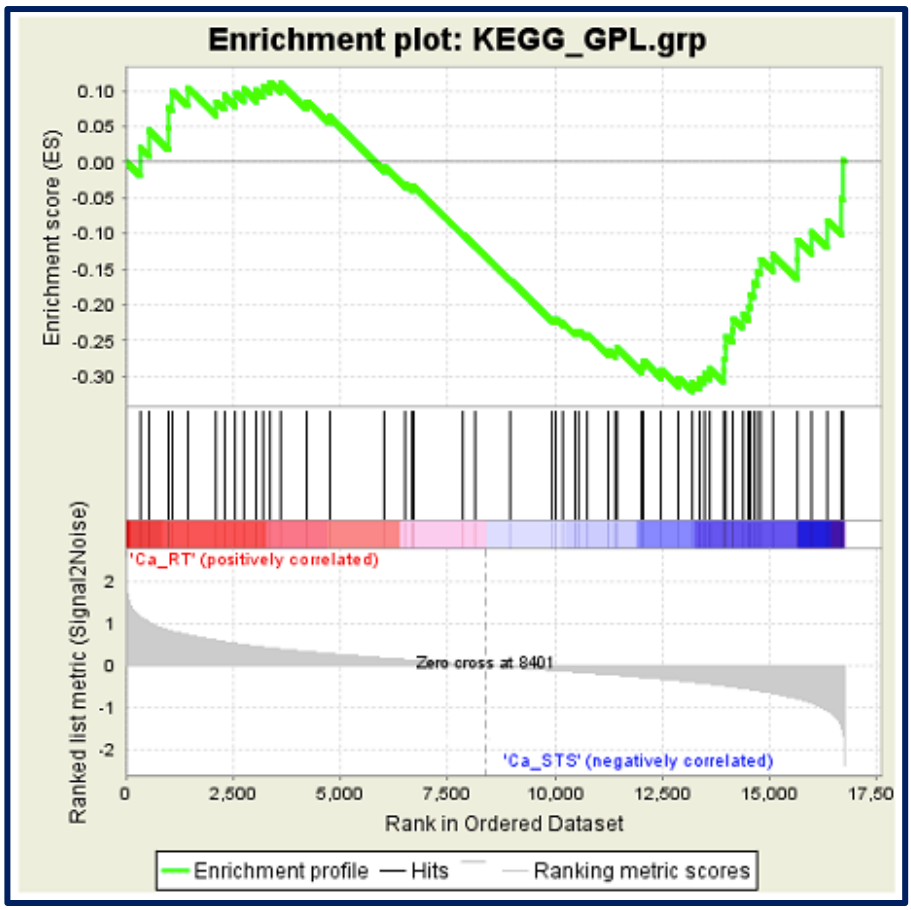


Figure 4.5 – GSEA enrichment plot for cohorts RT+ (Ca_RT) and RT- (Ca_STS) mapped to KEGG GPL metabolism. Peak enrichment was more significant for depleted genes in RT.

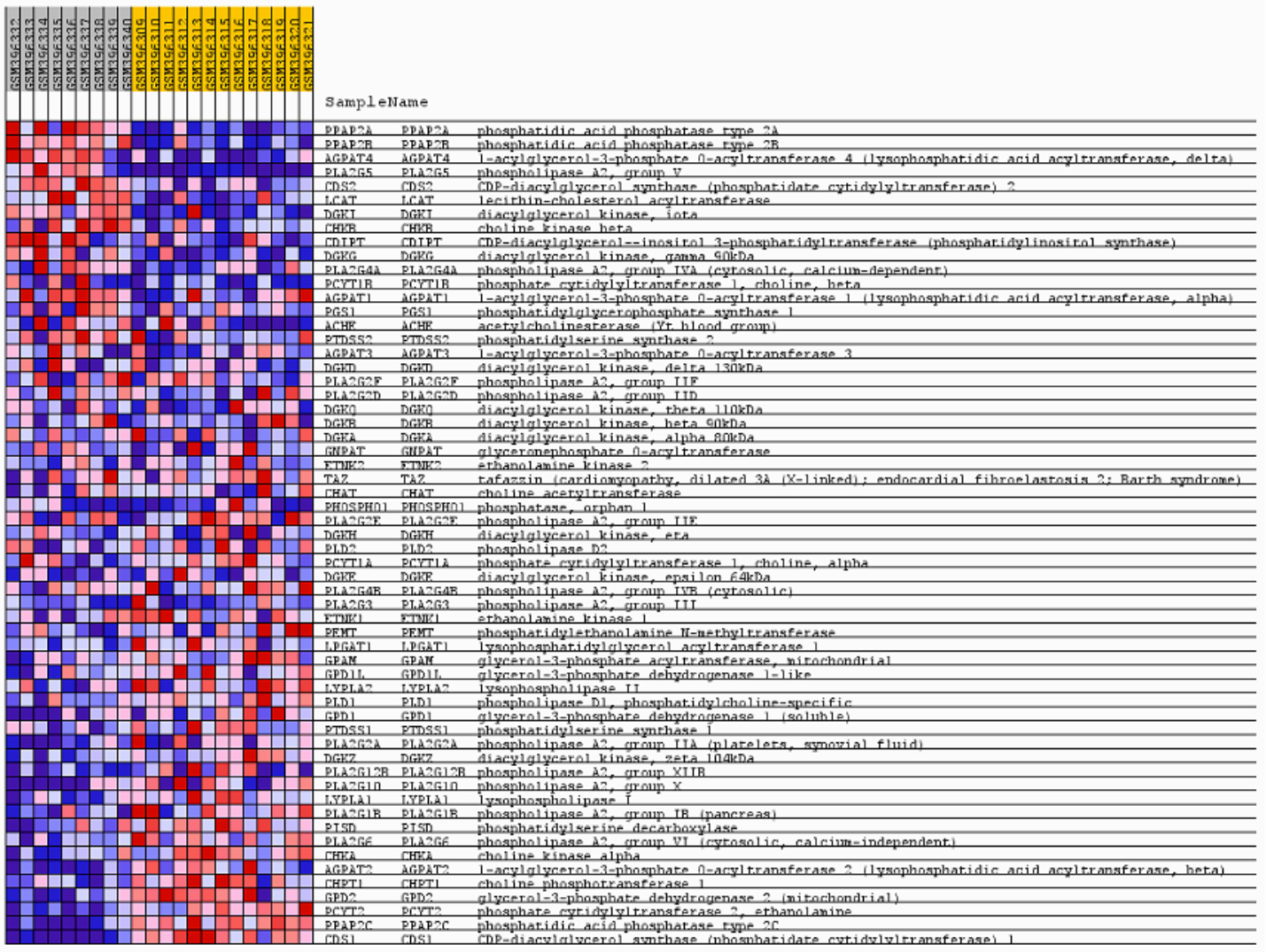


Figure 4.6: Heatmap of gene expression in the Snipstad et al microarray series following GSEA, collapsed to the KEGG GPL metabolism gene set in MSigDB v5.2. Phenotype labels RC RT+ (grey) and RC RT- (yellow)

Gene	Expression value RT+	Expression value RT-	Enriched/depleted RT+ vs. RT-	log^2 fold change	p-value (dataset)
PPAP2A	68429.500	16118.900	Enriched	2.08	$8.19E^{-4}$
PPAP2B	33111.030	10574.800	Enriched	1.64	$5.44E^{-4}$
AGPAT4	1496.830	572.320	Enriched	1.38	$1.17E^{-2}$
PLA2G5	4452.060	610.600	Enriched	2.87	$1.59E^{-2}$
CDS2	6433.120	4693.130	Enriched	0.45	$3.47E^{-4}$
CDIPT	7259.590	3816.370	Enriched	0.92	$1.89E^{-2}$
CDS1	401.890	3240.610	Depleted	-3.01	$3.36E^{-6}$
PPAP2C	6188.790	60957.820	Depleted	-3.30	$8.1E^{-7}$
PCYT2	2979.280	5207.020	Depleted	-0.81	$1.03E^{-4}$
GPD2	1260.620	3436.740	Depleted	-1.45	$2.66E^{-4}$
CHPT1	4611.270	8245.530	Depleted	-0.83	$8.35E^{-4}$
PLA2G6	574.290	809.370	Depleted	-0.50	$6.59E^{-3}$
AGPAT2	7507.610	16965.190	Depleted	-1.18	$2.96E^{-3}$

Table 4.4 - Selected genes within KEGG GPL gene set demonstrating significant differential expression between cohorts. In the RT- samples, the above reported genes demonstrate core enrichment.

4.3.5 Discussion

The index publication for this dataset found significant enrichment in KEGG gene sets associated with cellular adhesion molecule expression and cell signalling, specifically leucocyte transendothelial migration. [311] These findings were replicated in an initial analysis run mapping to the entire curated library of KEGG pathway gene sets. These pathways themselves involve crossover with other biosynthetic pathways as discussed in 4.2.6 and 4.2.7 with bioactive lipids playing a significant role in these mechanisms at both a membrane and intracytoplasmic signalling level. In line with my hypothesis, the justification for a deeper analysis and interrogation of the KEGG GPL gene set was made by way of discovering enrichment in the KEGG GPL gene set on whole library analysis.

As expected, the genes reported above represent significant gatekeeper steps in glycerophospholipid metabolism. **Figure 4.7** summarises the key steps in the GPL metabolism pathway where these genes are interconnected, acting on a subset of the pathway between phosphatidylcholine (PC) and phosphatidylethanolamine (PE). The pathway is further interlinked to phosphatidylglycerol (PG) by way of the action of PGS1 on CDP-DAG; again, this demonstrated core enrichment in the RT- cohort.

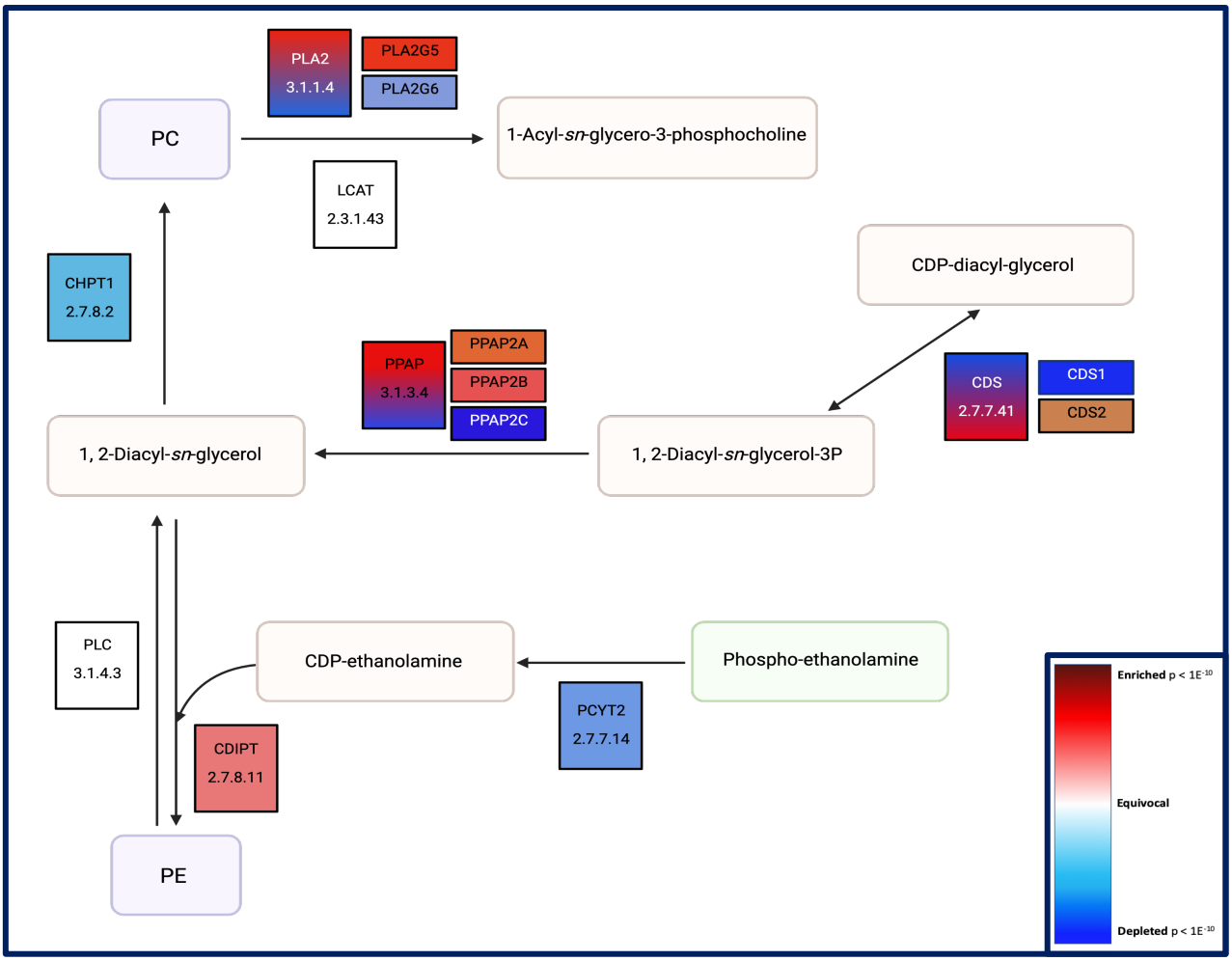


Figure 4.7 - Summary map of enriched and depleted genes in GPL metabolism in the Snipstad dataset. Heat maps superimposed onto genes as per expression data and $\log^2$fold changes. KEGG codes (e.g., *PPAP* 3.1.3.4) further subdivided into gene isoform expression data – e.g., PPAP2A and PPAP2B enriched in RT+ but PPAP2C enriched in RT-, and vice versa.

As a class of complex cellular lipids, GPLs are known to have critical bioactive functions in addition to their key role in the bilayer architecture of the cell membrane. [315] PCs are by far the most abundant of the complex lipids in humans, generally accounting for >50% of the total. As such, they are the major component of the outer leaflet of the cell membrane bilayer. Furthermore, they are a major component of circulating free plasma lipids, in particular of lipoproteins such as high-density lipoprotein. PEs additionally play a key structural role in cell membranes, with ratios of PE to PC in membranes thought to have implications on cellular functions. Furthermore, in contrast to PCs, PEs tend to be concentrated (with PIs) on the inner leaflet of the plasma membrane and as such are key mediators between cytosolic functions and interplay between the ER, mitochondria and cell membrane. Indeed, the highest proportions of PEs are found within mitochondrial membranes. PEs will bind to cytosolic phosphatidylethanolamine-binding proteins (PEBPs) with downstream cellular regulatory and signalling functions. As discussed in 3.2.5, one possible role for PEs in rectal cancer radioresistance has been reported, with hPEBP4 being shown to be correlated to enhanced clonogenic survival by inhibition of Akt-mediated apoptotic pathways. [242]

PGs are found at a lower abundance than other GPLs in the cell membrane, where their mechanism of action includes activation of lipid-gated ion channels. Through combination with CDP-DAG, PGs are key intermediates in the biosynthesis of cardiolipin, which is in turn a critical component of the inner mitochondrial membrane and required for the functioning of enzymes involved in oxidative phosphorylation. It is also a major component of bacterial membranes, which is worth

consideration in the context of the interaction between the gut microbiome and the human host. [316 317]

The dataset used for this analysis is subject to inherent limitations, in line with other experimental work on this topic to date. The experimental limitation is the availability of a dataset compatible for use with the GSEA platform. The sample set is by definition quite small (only 9 irradiated tumours) compared to datasets that are focussed on sequencing of tumour and normal tissue cohorts without therapies. Review of the unsupervised (PCA) and supervised (PLS) analysis of the differentially expressed genes in the Snipstad et al study demonstrated significant heterogeneity amongst the RT+ tumour samples in this cohort, whereas the RT- and matched normal cohorts were more tightly clustered with good separation. This raises the question of the limitation of such a small cohort of irradiated tumours – of note on the heat maps, some of the differentially expressed genes in the GPL pathway (particularly those with $\log^2$ fold changes between 0.5-1.5) are 'all or nothing' with regards to expression values, thus likely affecting the overall fold changes in expression over the cohort. Nonetheless, even with these limitations taken into account, appropriate FDR and significance levels mean that this analysis still demonstrates a plausible role for modulation of GPLs in response to radiotherapy, worthy of further investigation.

4.4 Summary and conclusions

This *in silico* analysis represents both the first report of a GO-based annotation strategy for homogenising experimental radiotherapy resistance biomarker data in rectal cancer and also a novel insight into the role of GPL metabolism in this mechanism by way of targeted GSEA of a legacy dataset. These findings are complementary and yield deeper insight into the role of complex lipids in modulating RT response and also justify further experimental work.

The aim at the outset was not to identify individual genes or proteins of interest, as what is evident is that a single biomarker is unlikely to govern the entire response mechanism in the highly heterogenous molecular landscape that is rectal cancer. Rather, the goal here was to identify biological domains of interest and to justify the hypothesised role for complex lipids that interact at every level of the biological pathways involved. To this end, although the subsequent GSEA has demonstrated enrichment in the GPL metabolism pathway, the answer to the problem is likely to be more elusive than that demonstrated at a gene expression level, given the number of post-translational modifications that can occur and the overlap of differing homologs of particular classes of genes (e.g., the three enriched homologs of phosphatidic acid phosphatase: PPAP2A,PPAP2B and PPAP2C). The limitations of the work as outlined above notwithstanding, the enriched genes of interest lie within a connected segment of the wider GPL metabolism blueprint, connecting phosphatidylcholines (PC), phosphatidylethanolamines (PE) and phosphatidylglycerols (PG). There is also an apparent role for the release of increased levels of fatty acyls from their constituent GPLs through the action of phospholipase 2 (PLA2), although this is a diverse and

complex group of enzymes with both membranous or cytosolic and calcium-dependent or -independent isoforms (e.g., Group V, PLA2G5 and Group VI, PLA2G6).

The implication that GPLs may be implicated in the response to a therapy in rectal cancer is not entirely unsurprising. There is evolving evidence for the changes in compositions of cellular membranes and abundances of certain classes of GPLs in various cancer subtypes versus their normal tissue counterparts. Far from being inert structural components, modulation in the composition of important molecules have already been identified as hallmark features of colorectal cancer and the myriad degrees of variability in their fatty acyl chain structures mean that even small modifications to commonly observed GPLs may have significant implications to the tumour microenvironment.

In summary, I have demonstrated a role for the modulation of complex lipids (and the GPL pathway in particular) in response to radiotherapy in rectal cancer, both on the basis of extensive systematic review and meta-synthesis of published experimental data and a GSEA of a publicly available dataset. These findings justify a lipidomics approach to the research question which follows in Chapter 5.

Chapter 5
Mapping cellular lipid modulation in response to radiotherapy in rectal cancer

5.1 INTRODUCTION 138

5.2 Tissue collection protocol 141

5.3 Patient cohort and clinical metadata 143

5.4 Tissue preparation for DESI-MSI 144

5.5 DESI MSI experimental setup 145

5.6 Histopathology and raw data pre-processing 149

5.7 Image alignment and tissue type annotation 150

5.8 Multiple sample analysis 150

5.9 Compound identification with DESI MS/MS 151

5.10 Results – Clinicopathological data 152

5.11 Results – DESI MSI Spectral acquisition 154

5.12 Results - histopathology quality control 157

5.12 Results – annotation and discrimination of in-sample tissue types 158

5.13 Results – multiple sample comparative analysis 161

5.14 Results – the lipids responsible for the classification 165

5.15 Lipid identification with DESI MS/MS 166

5.16 Discussion 170

5.17 Conclusions 179

5.1 Introduction

Allowing for some variation on an international level, neoadjuvant radiotherapy (RT) is generally delivered as a 'long course' comprising of between 45 and 50.4Gy ionising radiation, divided into 25 fractions, with concomitant 5-flurouracil based chemotherapy to act as a radiosensitizer (LCCR). The decision to offer LCCR or short-course radiotherapy (SCRT) is primarily guided by the interpretation of cross-sectional imaging; in this case, by interpreting features of magnetic resonance imaging (MRI) such as TNM stage and assessment of extramural venous invasion (mrEMVI). However, other than the routine histopathological analysis of the index biopsy - usually obtained during a lower gastrointestinal (LGI) endoscopy – the lack of an understanding of the exact mechanisms of radioresistance or radiosensitivity means that routine, individualised phenotyping is not performed at this stage to guide this decision.

As discussed in section 4.1, the vast majority of investigative studies aimed at identifying candidate biomarkers or understanding response pathways have been largely based around genomic assays (typically, microarrays or qPCR) or proteomic-level analyses, most typically with immunohistochemistry (IHC). The majority of the reviewed biomarker studies focus on a single point in a simplified biological schema as represented in **figure 5.1**. This approach works on the assumption that such determinants of response are either genome-led (either by way of copy number or core mutations, detected by sequencing platforms) or as a result of post-translational modification and protein or enzymatic activity (as detected by IHC). It is interesting to note that some groups are already taking the standpoint that a gene signature may

not exist as the ultimate predictor for radiation response in rectal cancer. [318] Furthermore, despite its currently deployed clinical utility, IHC remains a relatively qualitative methodology, relying on the eye of the trained histopathologist to formulate a decision regarding degree of immunochemical staining. Although validated in clinical practice for established markers which determine the ability to deploy, for example, biological therapies in the adjuvant setting (i.e., kRAS status to predict efficacy of adjuvant cetuximab therapy), the ability to detect subtlety in a discovery science setting might be debatable.

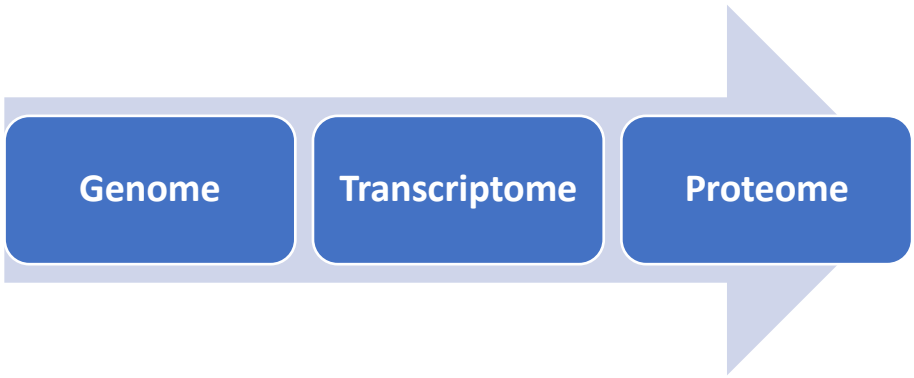


Figure 5.1 The classical 'top down' approach to investigation of biological processes. The limitation is the degree to which it considers modifications at each level and the resultant metabolites of interest.

Furthermore, given the current paucity of translational data emerging from classical studies using PCR, microarrays and IHC, it is plausible to consider that an important step in understanding radiation response is to develop a greater understanding of the metabolic footprint left after therapy has been applied. Coupled with increasing advancements in the analytical techniques for tissues *ex vivo* (and, increasingly, *in vivo*), a contrasting schematic workflow to the previous figure is demonstrated in **figure 5.2**, as to how a retrograde approach to this problem has been proposed for the purposes of this work. In order to better understand whether response is governed by either (a) the core genetic makeup of the tumour or (b) is an adaptive

response following application of the therapy, the core approach to this work was employing a ‘bottom-up’ approach using mass spectrometry imaging of solid rectal cancer and adjacent normal tissue, in order to elucidate the metabolic footprint as a consequence of radiotherapy in the neoadjuvant setting. In this section of my work, I have opted to obtain rectal cancer tissue that has been frozen at resection, thus preserving as close a metabolic ‘snapshot’ as possible. I have then analysed these tissue sections using desorption electrospray mass spectrometry imaging (DESI MSI), justified on the basis of the minimal sample preparation and without the need for the matrix fixation required for MALDI MSI. DESI MSI has already been demonstrated to provide biologically meaningful insights into colorectal cancer lipidomics and the platform has been optimised in previous work to obtain reliable spectral data in the GPL mass range.

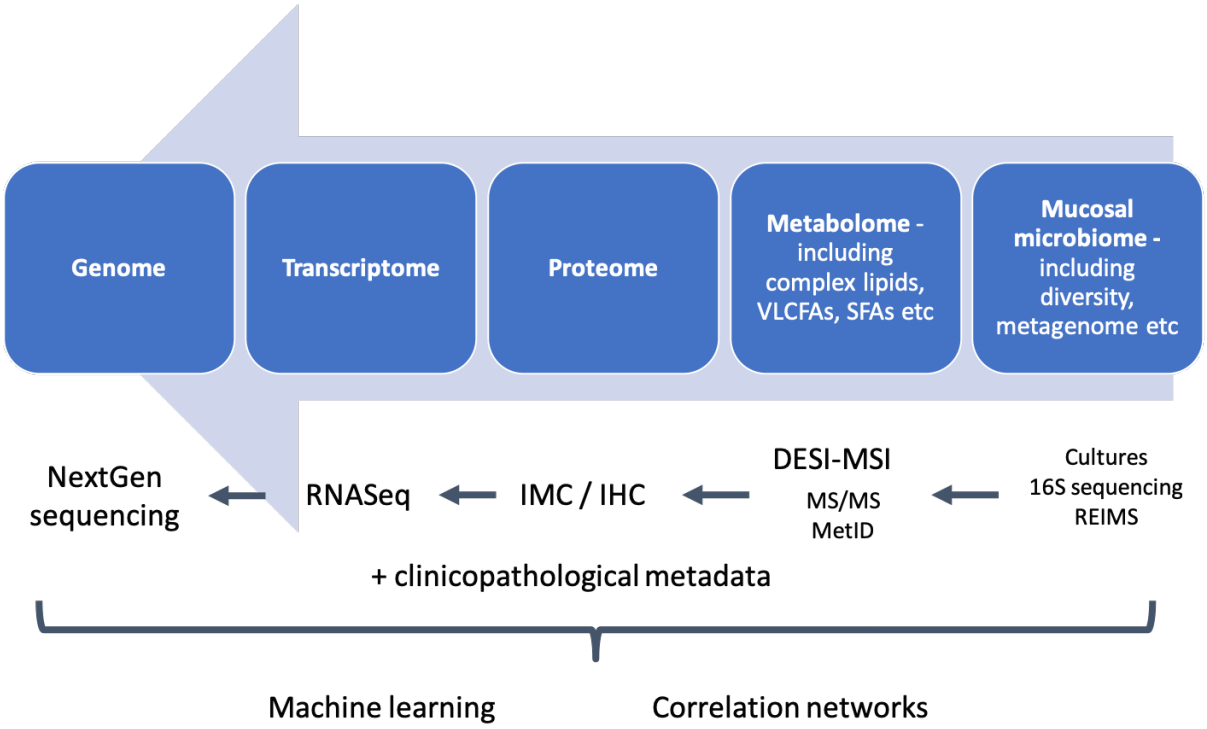


Figure 5.2. A possible proposed workflow of a solid-tissue, multi-omics approach; complementarity with sequencing and proteomics approaches but with a more holistic systems biology overview. Integrative bioinformatics will be critical to processing these combined datasets rather than analysis in silos. DESI MSI is just one modality for evaluating the metabolome - in this context for the tumour microenvironment lipidome – but other forms of biological MS and indeed nuclear magnetic resonance (NMR) studies would apply. Equally, 'RNASeq' broadly covers a wide range of Next Generation RNA sequencing technologies, including spatially-resolved transcriptomics.

5.2 Tissue collection protocol

Favourable ethical approvals for this work were obtained from the East Midlands NHS Research Ethics Committee (REC reference 15/EM/0385). Tissue was subsequently obtained via a material transfer agreement into Imperial College from the University of Southampton and University Hospitals Southampton NHS Trust Tissue Bank (MTA-in reference 16338/59). Following local consent processes to the

tissue bank, upon removal of the surgical specimen at surgery, the specimen was opened in the operating theatres and a 5 x 5mm section obtained from the tumour edge (avoiding the area of central necrosis). A further 5 x 5mm section of 'normal adjacent tissue' was obtained 10cm away from the tumour edge in the same manner. Both were immediately wrapped in Parafilm® M (Bemis® MA, Oshkosh WI 54904, USA), and subsequently placed into 2mL cryovials (Nalgene-Nunc^{TM}, Rochester, NY 14625 USA) labelled with both a unique study identifier and either 'tumour' or 'normal' and placed into a liquid nitrogen cylinder to instantaneously snap-freeze. Samples were then transferred to a -80^{o}C freezer prior to transfer to Imperial College via time-critical courier on dry ice.

5.3 Patient cohort and clinical metadata

A total of 52 fresh-frozen tissue samples from patients treated with surgical resection were acquired via the MTA-in and were all included in this study. In addition, matched pre-treatment FFPE biopsy samples were acquired from those patients who underwent LCCR. The sample characteristics are as shown in **figure 5.3**. Additionally, a complete set of clinicopathological metadata was compiled for each patient sample.

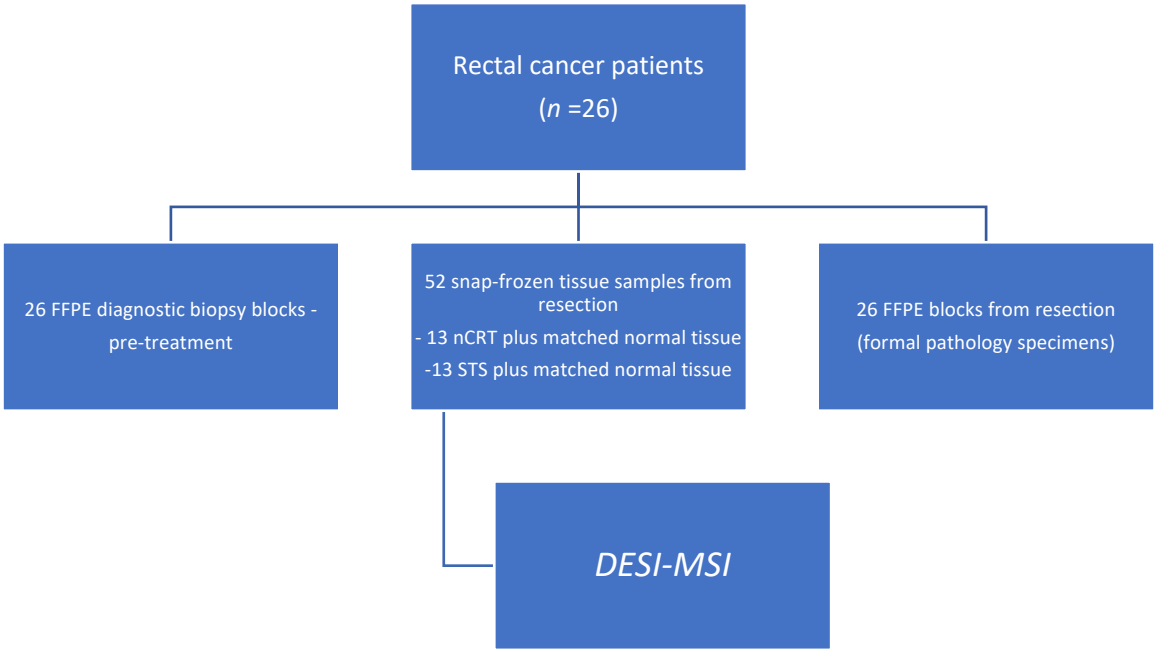


Figure 5.3. Tissue sample characteristics. From a total of 26 patients there were paired tumour and normal samples from each which were harvested at surgical resection. These snap-frozen specimens were suitable for DESI MSI as their fresh-frozen nature provides a metabolic 'snapshot' which is then unaltered by the mechanics of ionisation in DESI.

Mandard et.al [301]	Ryan et al. [303]
1. Complete regression (fibrosis without detectable tissue of tumour - pCR)	1. No viable cancer cells
2. Fibrosis with scattered tumour cells	2. Residual cancer outgrown by fibrosis
3. Fibrosis and tumour cells with preponderance of fibrosis	3. Significant fibrosis outgrown by cancer
4. Fibrosis and tumour cells with preponderance of tumour cells	
5. Tissue of tumour without changes of regression	

Table 5.1. Comparison of the two TRG systems applied to the sample cohort, for those patients who received RT

Data were collected on TNM stage, Duke's classification, EMVI, tumour perforation, R status, metastases, LCCR regime, tumour regression grading and survival. Tumour regression grading was performed according to either the 5-point system proposed by Mandard et al and in some cases additionally by the compressed

3-point system as described by Ryan et al. [301 319] A comparison of these two regression grading systems is summarised in **table 5.1**.

5.4 Tissue preparation for DESI-MSI

Tissue samples were stored in a -80°C freezer prior to processing. A cryomicrotome was used for sectioning frozen tissue sections for analysis. Samples were transferred from the -80°C freezer directly to the cold chamber of a Leica cryomicrotome. For each sample, the tissue pellet was mounted onto a cold chuck

onto which a few drops of Optima™ LC-MS grade water (Fisher® Scientific, Leicestershire UK) were placed using a Pasteur pipette to facilitate mounting. 10µm sections were then taken from each tissue pellet, thaw-mounting 2 sections onto each labelled SuperFrost glass slide (Thermo Scientific™). This enabled a concurrent, immediately adjacent, haematoxylin and eosin (H&E) stain to be made following mass spectrometry analysis in cases of tissue desiccation due to DESI MSI. Slides were dried to remove excess surrounding water ice in the cold chamber using a low-pressure nitrogen gas gun. A total of 16 adjacent sections was taken from each tissue samples (8 slides each) in this manner and the slides returned to the -80°C freezer prior to DESI MSI. The slides not used for DESI were utilised in experiments discussed later.

5.5 DESI MSI experimental setup

A home-built DESI MSI setup was employed as described in previous work by the group. [173 190 320] A Thermo Scientific, Fourier transform, Orbitrap mass spectrometer (*Exactive™)* was utilised, with a charged electrospray capillary solvent/gas assembly orientated above a moving stage, onto which the slides were mounted (**figure 4.4)**. This was coupled to a standalone computer terminal running the Thermo™ Xcalibur software. As previously described, a solvent mixture of 95% UPLC-MS grade methanol and 5% Optima™ LC-MS grade water (Fisher® Scientific, Leicestershire UK) was used as the ionising solvent, and stored in sterile vials.

Calibration of the instrument was performed using a heated electrospray, direct injection interface (HESI) and Pierce™ Calibration Solution mixtures for ESI, containing known standards across the full mass range in both positive and negative

ionisation modes (Thermo Scientific™). Calibration solutions were introduced separately using a gas-tight syringe pump and capillary tubing, with a flow rate set to 0.005mL/min. After signal stabilisation, a calibration run was performed, and the signal-to-noise ratio measured, with an upper-limit cut-off root-means-square of 0.3/0.3 parts per million (ppm). The calibration settings were saved, and the run repeated for the second ionisation mode.

To calibrate the geometry of the sprayer-sample-inlet interface, a 10µm section of porcine liver was used, owing to the high abundance of phospholipids in mammalian liver tissue. A single piece of porcine liver was frozen at -80^{o}C and cryomicrosectioned into over 200 10µm sections and mounted onto SuperFrost glass slides using the same method described in section 3.12. These samples were used for all instrument calibrations as described in this chapter. To optimise the ionisation and minimise noise obtained using a low-flow, DESI experiment, sprayer-sample-inlet geometry was set as per the index publication, i.e. distance from DESI sprayer tip to sample was 1.5mm, from the sprayer tip to MS inlet 14mm, from inlet capillary to sample 0.1mm; the sprayer angle was 80^{O}. [320] Care was taken to ensure that the stage was level and that the sprayer tip was aligned with the centre of the MS inlet. Solvent was loaded into a 1mL glass Hamilton gas-tight syringe, and the flow rate was set to 1.5µL/min. The DESI method was run over a full mass range from *m/z* 150 - 2000. A sample run was built into Xcalibur utilising geometric data from the moving stage software, with the top-left corner of the sample being aligned with the sprayer to start the run and a pixel size of 100µm x 100µm.

For calibration purposes, between 3-5 rows of porcine liver sample were analysed with real-time assessment of the acquired spectra. A total ion count (TIC) of greater than $7x10^3$ with phospholipid ratios (*m/z* range 600-950) of over 50% was deemed acceptable to proceed with the experiment. Raw spectral data were saved for longitudinal quality control purposes. Mass accuracy was also checked by way of interrogating spectra for three highly abundant compounds with known exact masses: de-protonated arachidonic acid (AA - m/z 303.2), PG 16:1/20:5 (m/z 766.48) and PI 18:0/20:4 (885.50). Provided the error margin was less than 10ppm for each case, the experiment would proceed. Samples were subsequently analysed in negative ionisation mode for optimisation of lipid detection (for complex/glycerophospholipids, this is most commonly due to deprotonation (M-H) or water loss ($M-H_2O$).

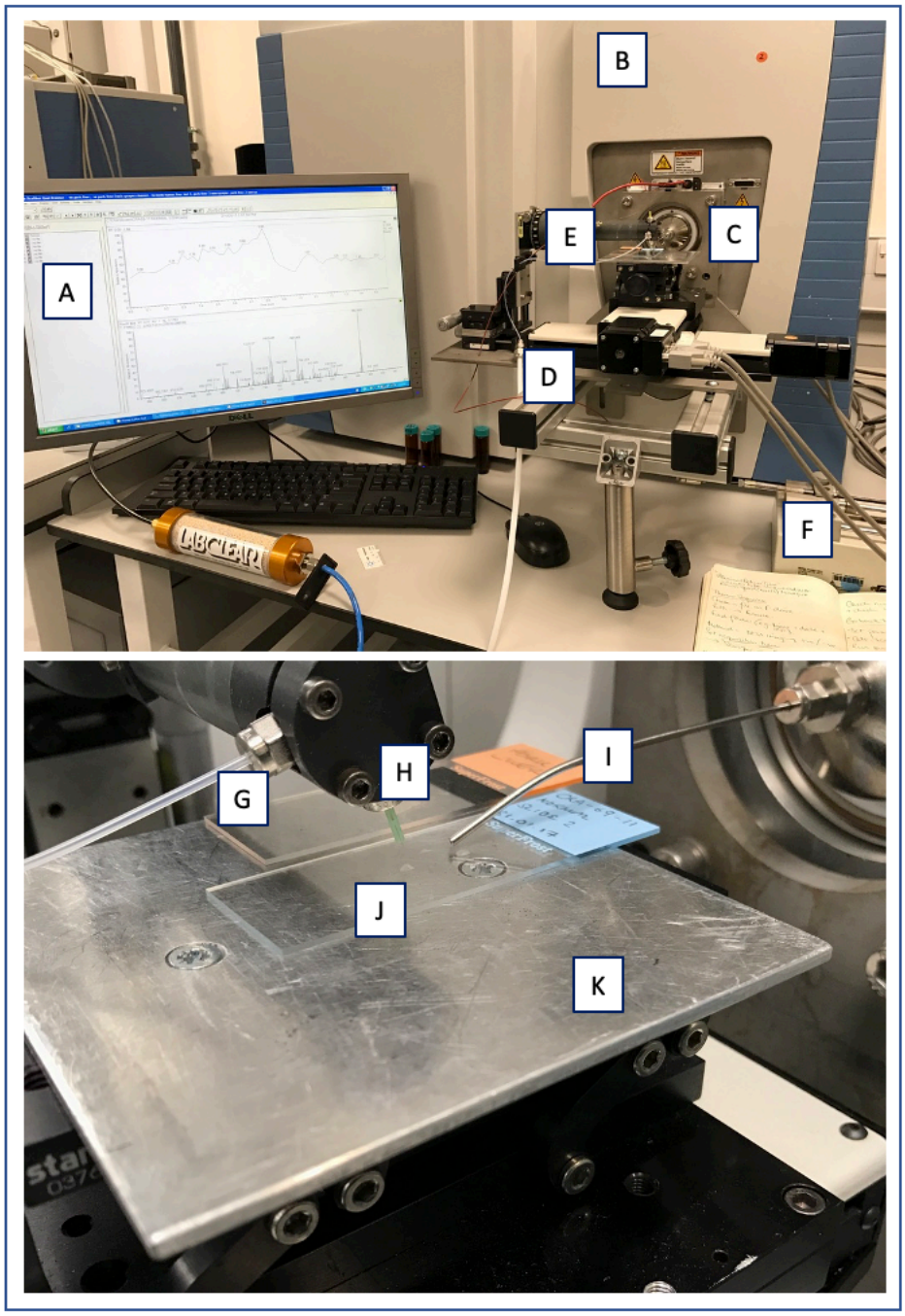


Figure 5.4 – (Top) DESI MSI interface as connected to Thermo *Exactive* Orbitrap mass spectrometer. (A) Windows PC running Thermo™ Xcalibur software, spectral acquisition in-progress; (B) Thermo *Exactive* MS; (C) Heated sample inlet; (D) Home-built sample stage; (E) electrospray capillary assembly; (F) solvent pump. (Bottom) (G) nitrogen gas inlet to sprayer; (H) electrospray ionisation sprayer; (I) inlet capillary; (J) mounted sample slide (attached with double-sided tape); (K) moving stage

5.6 Histopathology and raw data pre-processing

Following DESI MSI, each slide was stained by the standardised haematoxylin and eosin (H&E) method, although due to the usage of fresh frozen tissue the usual deparaffinisation step was not required. Briefly, this proceeded with rehydration in deionised water, then a hematoxylin stain for 3 minutes, rinsing in deionised water, 5 minutes in tap water, fast dip in acid ethanol, 2x 1-minute rinses in tap water and a 2-minute rinse in deionised water. The eosin stain was applied for 30 seconds followed by 3x 5-minute washes in 95% ethanol, 3x 5-minute washes in 100% ethanol and then 3x 15-minute washes in xylene to dehydrate prior to covering with a coverslip. Digital photographs of the tissue sections were then obtained using a high-resolution slide scanner (Hammamatsu). These images were subsequently reviewed by an expert histopathologist (Professor Robert Goldin, Division of Surgery and Cancer, Imperial College London) and all tissue features were hand-annotated (including - but not limited to – adenocarcinoma, tumour stroma, normal mucosa, normal submucosa, normal muscularis, lymphoid follicles and blood vessels).

Raw spectral data was centroided and saved as imzML files prior to creating QC composite ion images using the *Myrion* application. [320] Individual imzML files were subsequently uploaded into an in-house chemometric toolbox in MatLab for spatially aligned annotations of the tissue sections and ion images.

5.7 Image alignment and tissue type annotation

For each patient sample (tumour and matched normal), both the imzML file with centroided mass spectrometry data and the corresponding H&E image were uploaded into the MatLab toolbox. The full mass range of spectra was imported for analysis and the spectra within the m/z range 600-950 were aligned with a peak-matching procedure with a tolerance of +/- 8ppm. The optical and ion images were then aligned to permit accurate tissue type annotation. Next, in order to correlate mass spectral profiles with particular tissue types within each sample, tissue pixels were colour-annotated using code linked to the RGB colour space with a unique colour for each tissue type i.e., dark blue for normal mucosa, dark green for adenocarcinoma etc. Solvent background areas (i.e., away from the tissue edges) were annotated accordingly. The same colours and labels were used for all samples. Finally, samples were identified as being either 'Tumour – nCRT', 'Normal – nCRT', 'Tumour – STS' (straight to surgery) or 'Normal – STS'. Principle component analysis (PCA) was performed to analyse in-sample heterogeneity between tissue types and the resulting annotated file was saved as a .mat file for multiple sample analysis.

5.8 Multiple sample analysis

Annotated .mat files were aggregated and imported into a second MatLab toolbox for processing. Data within each file contained all annotated pixels for all tissue types and files were labelled as being either from samples which received nCRT or those which were STS. Spectral data were initially normalised by total ion current (TIC) across the full mass range (*m/z* 150-1500; minimal spectra were acquired above *m/z*

1500). Aggregated spectra were then analysed for peak shift by inspecting true measured values for ubiquitous compounds of high abundance and with a known exact *m/z* value, for example 885.50 (ubiquitous PI 18:0/20:4) and corrected at a bioinformatics level as necessary. At this point, any anomalous data identified prompted a repeated DESI MSI analysis of a section from the corresponding patient.

Following data clean-up, loadings plots were generated from aggregated data. Next, PCA was performed on both an all-pixels and median tissue values model using both all tissue classes and a targeted approach for adenocarcinoma tissue. Supervised analysis (maximum margin criteria – MMC) was then performed and both leave-one-out (all pixels) and a leave-patient-out cross validation model created (LOOCV). Confusion matrices were generated from both models.

Univariate ANOVA was then performed for extraction of compounds with discriminating power in the models. Volcano plots were generated with a negative $\log^2$ fold change applied and outlying *m/z* values identified which were either responsible for the nCRT pixel prediction or the STS pixel prediction based on all spectral data across tissue types. *m/z* values were extracted and subsequently used to guide tandem DESI MS (MS/MS) for compound identification.

5.9 Compound identification with DESI MS/MS

For classification of lipid species identified by ANOVA in the models described above, the same DESI interface was connected to a Thermo Scientific *LTQ-XL™* mass spectrometer for collision-induced dissociation (CID) MS/MS experiments. The

same DESI method (solvent composition, flow rate, ionisation settings) were employed. Ion images of all samples were examined and samples with high abundance of the identified, discriminating lipid compounds were selected based on the MSI data. Chamber retention times and collision energies were adjusted manually for each m/z value to optimise fragmentation. Major peaks identified were further fragmented in MS3 and MS4 runs. Resulting fragmentation patterns were then compared to known standards in publicly-available repositories within *LipidMaps* and *METLIN*. [315 321]

5.10 Results – Clinicopathological data

For the purposes of comparing the two groups of tumours, both straight to surgery (STS) and neoadjuvant therapy (nCRT, LCCR), the AJCC staging for each cohort is summarised in **table 5.2**, as reported by the final histopathology at resection. Median age was 68 years. 8 out of 26 patients underwent APER (31%), 6 of whom were within the nCRT cohort. The remainder underwent laparoscopic or open anterior resection. A complete summary of the clinicopathological data for the sample cohort is shown in **table 5.3**. Each cancer specimen had a matched adjacent normal tissue sample as described in section 4.2.

Sample cohort	AJCC Rectal Cancer Stage						
	I	II			III		
	I	IIA	IIB	IIC	IIIA	IIIB	IIIC
nCRT (n=13)	2	5	0	0	1	5	0
STS (n=13)	5	2	0	3	2	1	0

Table 5.2 – Summary characteristics of final pathological staging samples from both cohorts. For the purposes of comparison between groups, AJCC staging has been used.

Patient ID (pseudo-anonymised)	Age at treatment	Procedure	Differentation	Dukes' stage	T stage 5th edition	N stage 5th edition	M Stage 5th edition	Nodes Positive	EMVI	Tumour Perforation	CRM	Neo adjuvant radiotherapy	Chemo	Radiation	TRG	Post-Op Mets	Date of Death
CRA-226-13	62	APER	WELL-MODERATE	C	T1	N1	Mx	2/18	NO	NO	R0	NO	N/A	N/A	N/A	NO	N/A
CRA-247-13	73	ANTERIOR RESECTION	MODERATE	C1	T3	N1	Mx	1/10	YES	NO	R0	NO	N/A	N/A	N/A	NO	N/A
CRA-278-13	68	ANTERIOR RESECTION	WELL-MODERATE	A	T1	N0	Mx	0/44	NO	NO	R0	NO	N/A	N/A	N/A	NO	N/A
CRA-306-13	69	ANTERIOR RESECTION	WELL-MODERATE	B	T3	N0	Mx	0/23	YES	NO	R0	NO	N/A	N/A	N/A	NO	N/A
CRA-336-13	65	ANTERIOR RESECTION	WELL-MODERATE	A	T2	N0	Mx	0/10	NO	NO	R0	NO	N/A	N/A	N/A	NO	N/A
CRA-506-15	83	ANTERIOR RESECTION	WELL-MODERATE	C	T2	N1	Mx	1/6	YES	NO	R0	NO	N/A	N/A	N/A	NO	N/A
CRA-534-15	74	ANTERIOR RESECTION	WELL-MODARATE	A	T2	N0	Mx	0/13	NO	NO	R0	NO	N/A	N/A	N/A	NO	N/A
CRA-543-15	86	HARTMANS	WELL-MODERATE	C1	T2	N1	Mx	1/6	NO	NO	R0	NO	N/A	N/A	N/A	NO	N/A
CRA-579-15	69	ANTERIOR RESECTION	WELL-MODERATE	C1	T3	N2	Mx	8/37	YES	NO	R1	NO	N/A	N/A	N/A	NO	N/A
CRA-602-15	79	ANTERIOR RESECTION	WELL-MODERATE	C1	T4b	N2	Mx	6/18	NO	NO	R0	NO	N/A	N/A	N/A	NO	N/A
CRA-608-15	45	APER	WELL-MODERATE	A	T2	N0	Mx	0/33	NO	NO	R0	NO	N/A	N/A	N/A	NO	N/A
CRA-613-15	68	ANTERIOR RESECTION	WELL-MODERATE	A	T2	N0	Mx	0/28	NO	NO	R0	NO	N/A	N/A	N/A	NO	N/A
CRA-069-11	58	ANTERIOR RESECTION	WELL-MODERATE	C1	T3	N1	Mx	1/43	NO	NO	R0	YES	CapOx	YES	RCRG2/TRG3	NO	N/A
CRA-085-12	75	APER	WELL-MODERATE	B	T2	N0	Mx	0/10	NO	NO	R0	YES	CapOx	YES	TRG3	NO	N/A
CRA-136-12	62	APER	WELL-MODERATE	B	T3	N0	Mx	0/28	NO	NO	R0	YES	CapOx	YES	RCRG1	LIVER	N/A
CRA-174-12	62	HARTMANS	WELL-MODERATE	B	T3	N0	Mx	0/11	YES	NO	R1	YES	CapOx	YES	TRG2	NODE,LIVER, LUNG	25/04/2013
CRA-268-13	62	ANTERIOR RESECTION	WELL-MODERATE	A	T2	N0	Mx	0/14	NO	NO	R0	YES	YES	YES	TRG2	NO	N/A
CRA-231-13	59	ANTERIOR RESECTION	WELL-MODERATE	C	T3	N1	Mx	2/24	NO	NO	R0	YES	CapOx	YES	TRG5/RCRG3	NO	08/06/2014
CRA-281-13	78	ANTERIOR RESECTION	POOR	C1	T3	N2	Mx	14/53	YES	NO	R0	YES	YES	YES	TRG4	NO	N/A
CRA-532-15	88	APER	POOR	B	T3	N0	Mx	0/28	YES	NO	R0	YES	CapOx	YES	TRG4	NO	N/A
CRA-612-15	79	APER	WELL-MODERATE	C	T3	N1	Mx	1/8	YES	NO	R1	YES	YES	YES	TRG4	NO	N/A
CRA-550-15	64	ANTERIOR RESECTION	POOR	B	T3	N0	Mx	0/27	YES	NO	R0	YES	CapOx	YES	TRG3	NO	N/A
CRA-483-14	85	APER	WELL-MODERATE	A	T2	N0	Mx	0/13	NO	NO	R0	YES	CapOx	YES	TRG4	NO	N/A
CRA-588-15	61	ANTERIOR RESECTION	WELL-MODERATE	C1	T3	N1	Mx	1/12	YES	NO	R0	YES	CapOx	YES	TRG3	NO	N/A

Table 5.3 – Full clinicopathological metadata for the sample cohort (*n* = 26). TNM reported as TNM as per the core dataset at the time of sample harvest.

5.11 Results – DESI MSI Spectral acquisition

Following cryomicrosectioning sample slides were racked and randomised to mitigate further compounding of batch effect arising from cutting – analysis time. Sample run times were directly dependent on sample size and ranged from approximately 90 minutes up to 280 minutes. Regular visual quality control checks were made to DESI spectra within each experiment to check for signal quality. On-sample base peak intensity (normalisation level, NL) of >1.0E^3 was generally considered acceptable in the context of colorectal tissue – during calibration of the instrument with porcine liver, an on-tissue intensity of >7.0E^3 was aimed for. Occasionally due to atmospheric changes or requirements to repair, recalibrate or replace components of the assembly, this threshold was not attained; in these instances, interface servicing or instrument cleaning and calibration was performed. **Figure 5.5** demonstrates an example of acceptable, snapshot mass spectrum. Following acquisition of the entire MSI run, data were centroided and converted into imzML file formats whereby individual ion abundances could be visualised for QC purposes prior to export into the MatLab toolbox (**figure 5.6**)

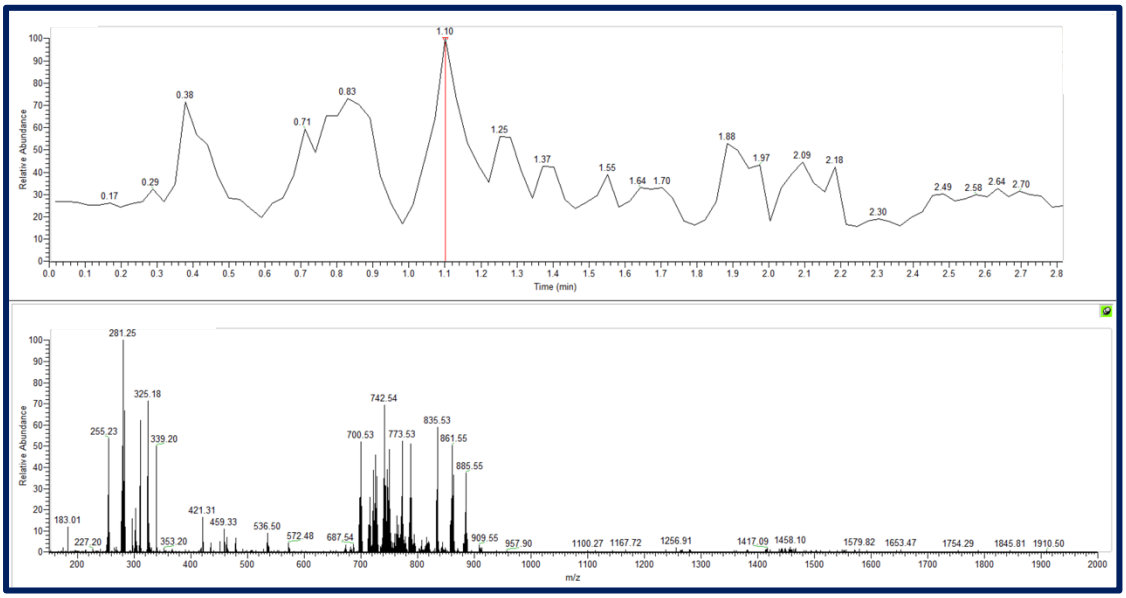


Figure 5.5 – Full mass range 150-2000m/z in negative ionisation mode for CRA283-12 N S2 row 32, time index 1.1 minutes. On-tissue, this demonstrates significant abundance in the phospholipid range 650-900 m/z with corresponding fatty acyls and triglycerides (normal colorectal mucosa)

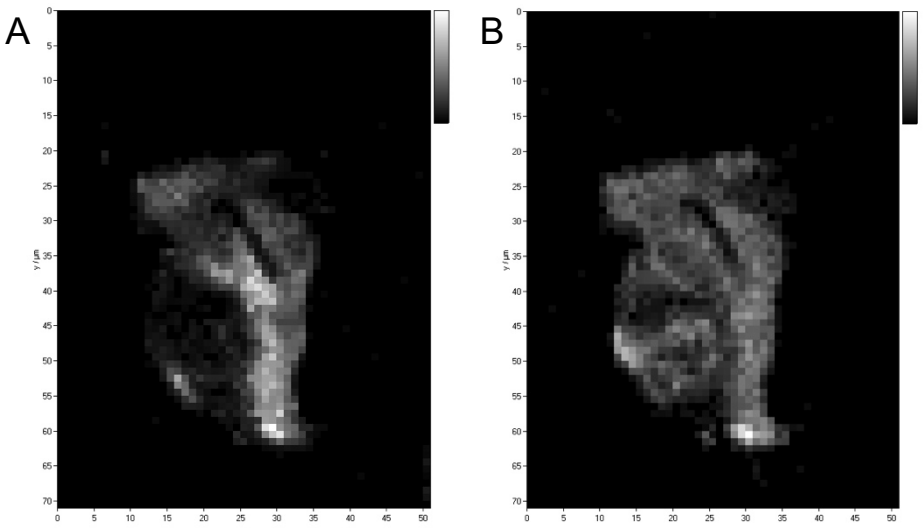


Figure 5.6 – Raw single-ion, intensity-graded imzML images from CRA532-15 S7 TUMOUR; A, m/z 737.500; B, m/z 810.500.

Visual review of raw spectral profiles also demonstrated shifts in relative ion abundances, both ‘in patient’ between cancer and matched normal rectal tissue and also between RT+ and RT- (‘STS’) tissue as shown in **figure 5.7**.

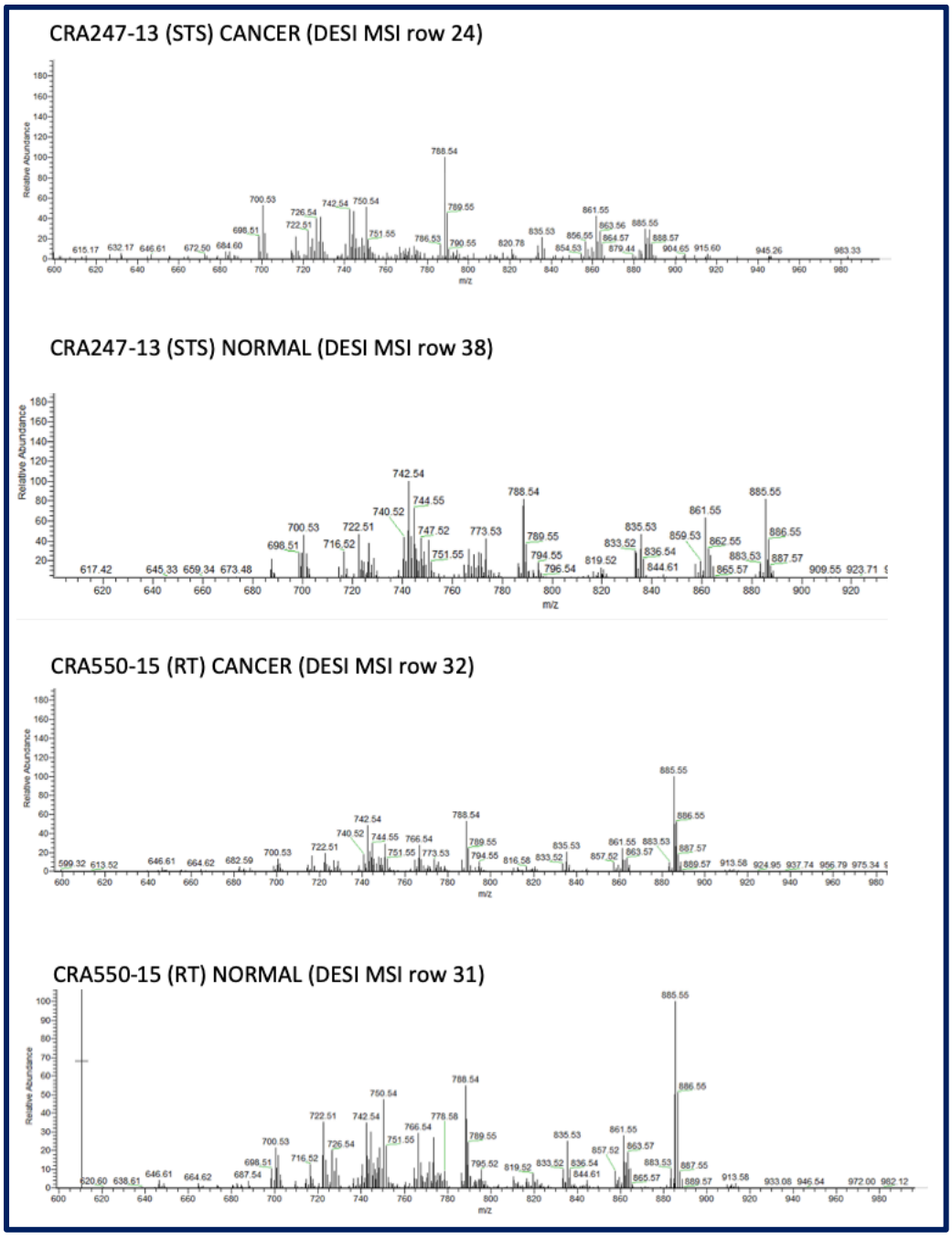


Figure 5.7 - Raw spectral profiles (non-centroided data) taken from representative rows from RT+ and RT- cancer samples and matched normal samples. Full mass range acquisition has here been narrowed to compare relative abundances amongst compounds in the GPL mass range (i.e., m/z 600-9. Sample spectra here selected as relatively well-matched on histological grounds: RT-, CRA247-13: T3N1 EMVI positive; RT+, CRA550-15: T3N0 EMVI positive, TRG3.

5.12 Results - histopathology quality control

All the supplied tissue for analysis was of sufficient volume and quality to facilitate cryosectioning in water and subsequent analysis with DESI MSI. Following H&E staining of analysed tissue sections, some degree of desiccation of the sections exposed to DESI MSI was identified. Annotation of tissue features, however, was still possible with the assistance of an expert histopathologist – finer detail was correlated with the adjacent aligned tissue section mounted on the same slide. This was of particular importance in cancer sections with a good response to nCRT, e.g., TRG1-2. These sections typically demonstrated a few scattered adenocarcinoma glands within islands of fibrosis, necrosis and desmoplastic stroma (**figure 5.8**). Analysis of all cancer tissue sections demonstrated a histological diagnosis of adenocarcinoma in all but one case where the (very minimal) residual cancer tissue amongst dense necrotic tissue was found to be squamous cells (CRA483-14). This was correlated with two further re-runs of this particular patient specimen.

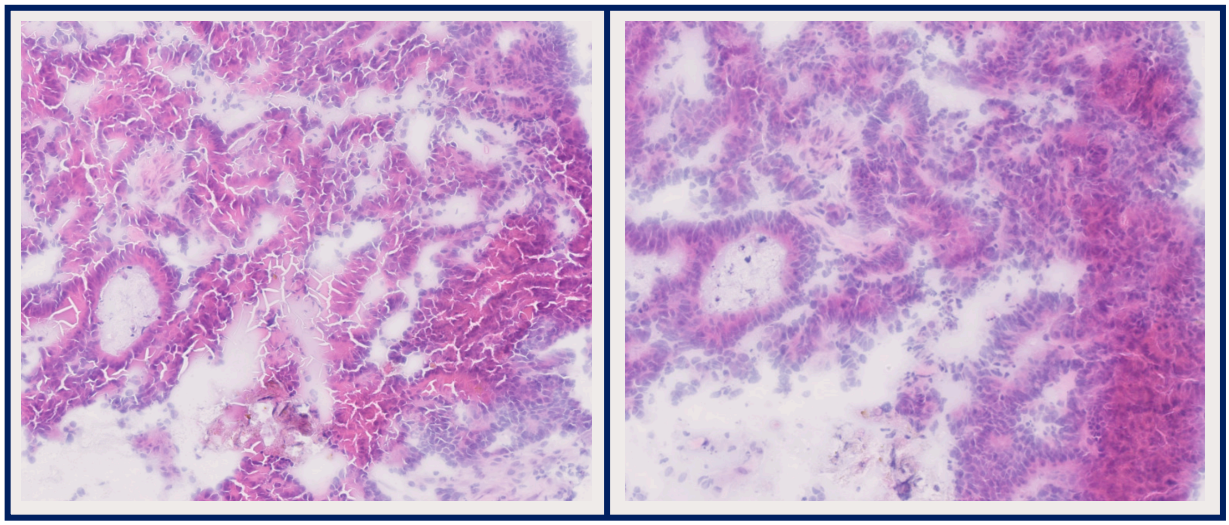


Figure 5.8 – Stromal component and residual tumour glands still visible after DESI MSI of section CRA306-13 S7 TUMOUR (left) but higher definition of tissue features including gland/stromal interface completely preserved in the adjacent section with near-identical architectural appearances due to 10μm distance in z axis.

5.12 Results – annotation and discrimination of in-sample tissue types

Due to the mucosal tissue sampling protocol, identified tissue types in cancer samples were relatively homogenous with either adenocarcinoma glands or desmoplastic stroma. Post-nCRT samples also demonstrated variable degrees of tumour necrosis and fibrosis. Review of spectral profiles and centroided imzML files from these necrotic regions found these to be relatively bland in the phospholipid m/z range 600-950. Normal rectal mucosa in both nCRT and STS patients demonstrated a richer tissue architecture whereby annotation of mucosa, submucosa, muscularis, blood vessels and lymphoid follicles was possible. Following optical/ion image alignment in the MatLab toolbox, DESI MSI was able to demonstrate discrimination of these tissue type features on unsupervised analysis (PCA), with loadings plots

weighted to the phospholipid, triglyceride and fatty acyls ranges in the first and second principal components (**figure 5.9**). Single ion selection and histological co-localisation demonstrates example lipid species responsible for this drive in the PCA model as shown in **figure 5.10**. Cross-validated supervised analysis with maximal margin criteria (MMC) on cancer samples also demonstrated separation of spectral clusters between residual tumour glands and desmoplastic stroma (**figure 5.11**)

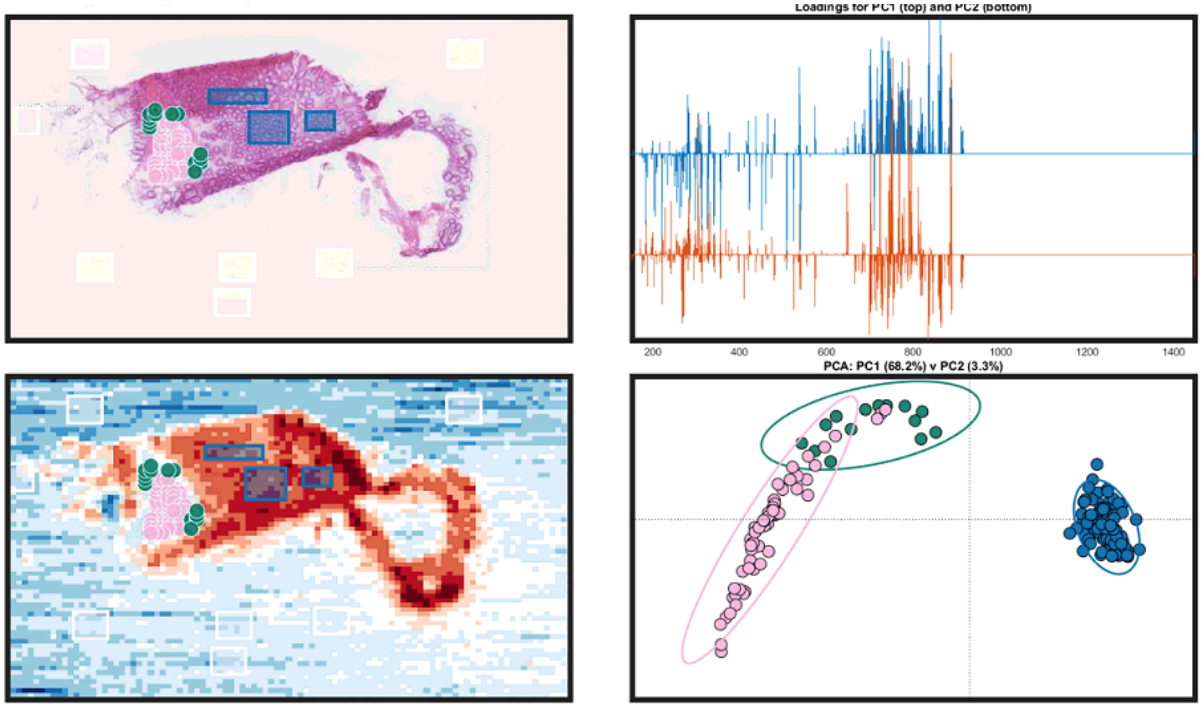


Figure 5.9 – Within the MatLab toolbox the optical/ion images were aligned according to the protocol as outlined. Histology-guided annotation of tissue types was performed by first delimiting the optical H&E image to pixels with equivalent size to the DESI MSI pixel size (100µm x 100µm). Within the toolbox a colour-coded classification was then employed to either select individual pixels/cells or to draw marquee areas over homogenous regions: blue, normal rectal mucosa; pink, normal submucosa; green, muscularis. White marquees are drawn over the background to differentiate from solvent background in tissue classification. Next, unsupervised analysis was performed with PCA with loadings plots for PC1 and PC2 shown in mass range 150-950 m/z above a cluster plot which demonstrates separation of mucosa, submucosa and muscularis in PC1 and PC2.

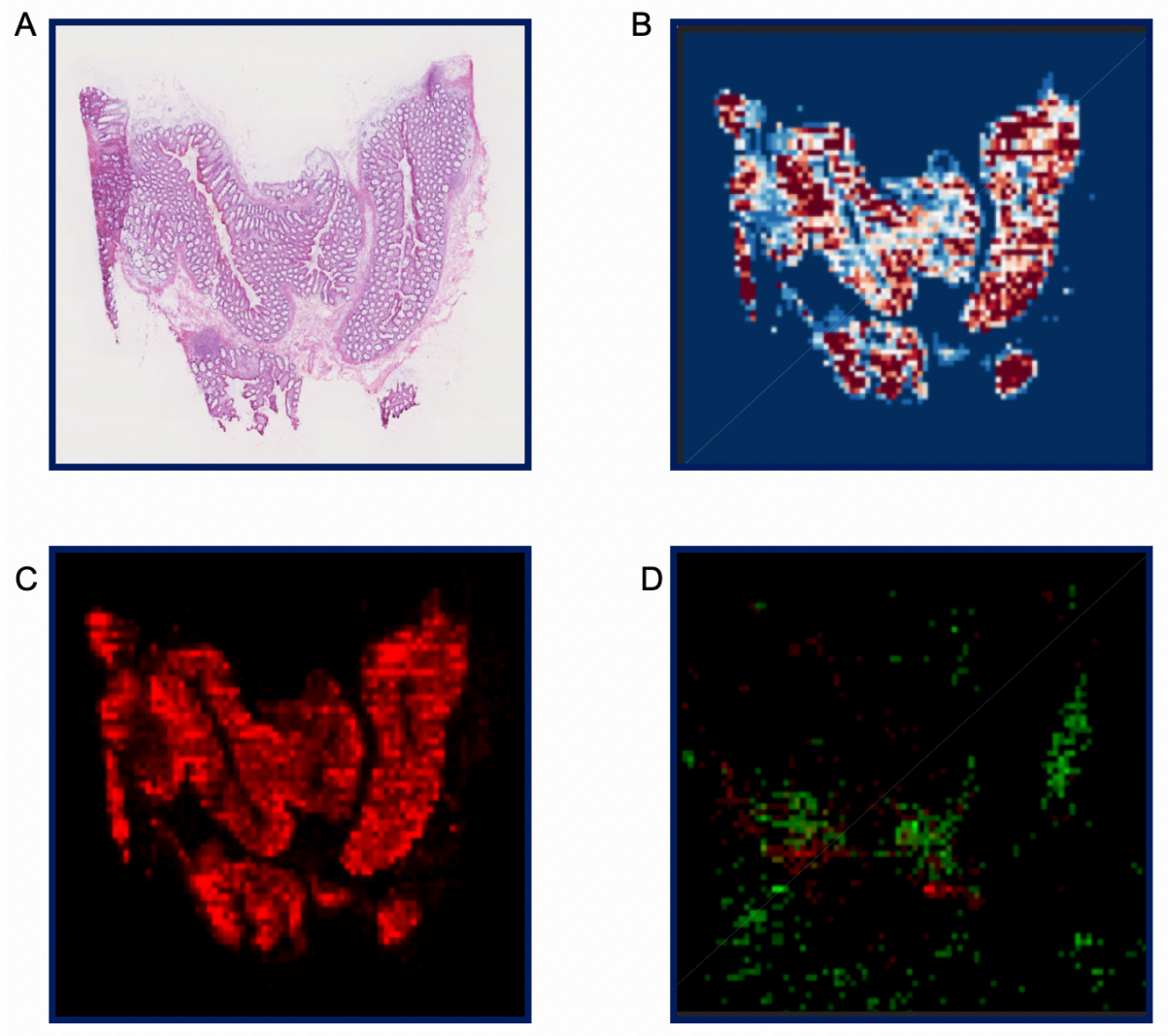


Figure 5.10 – Representative H&E-stained tissue section of normal colorectal mucosa from patient CRA268-13 (A), demonstrating normal mucosa and submucosa. (B) Composite, all-spectra DESI-MSI image with background (i.e., solvent) filtered out and represented as dark blue– signal intensity graded from white (low) through light blue to red. Selecting ions of interest in RGB image mode, DESI-MSI demonstrates differing lipid compounds found almost exclusively in colorectal mucosa (C – PI 18:0/20:4 m/z 885.50) and normal submucosa (D – 331.20 m/z, unidentified fatty amide)

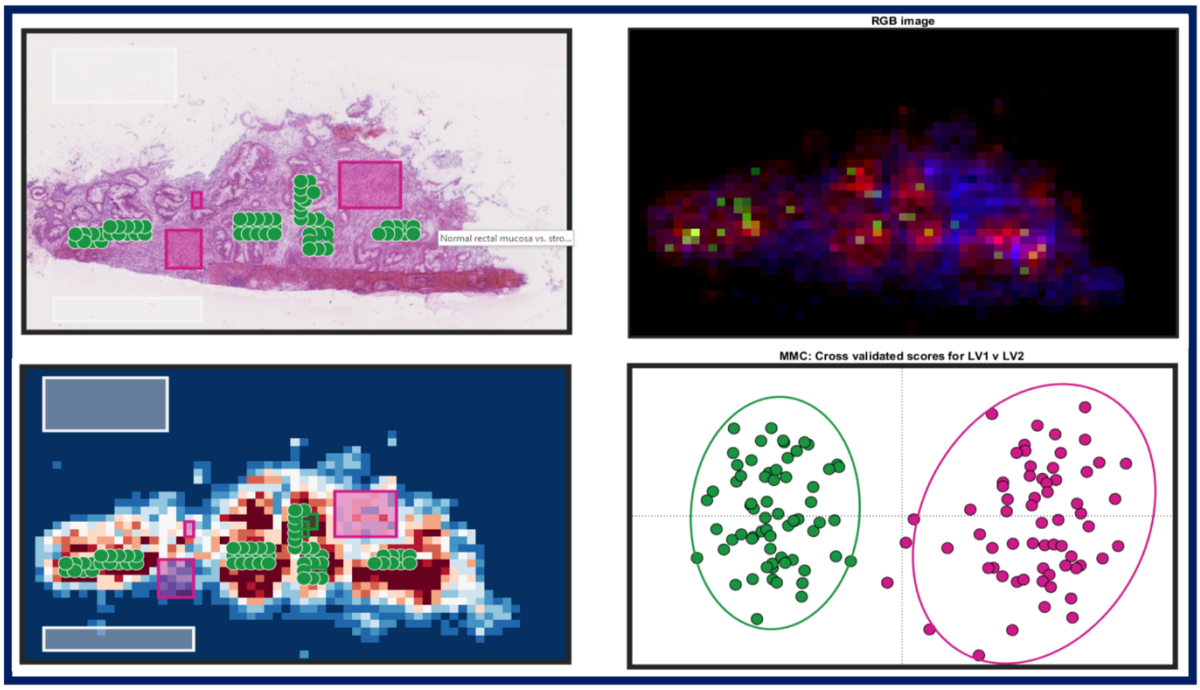


Figure 5.11 – Annotated, spatially aligned, H&E stained and DESI-MSI images for a T3N0M0 rectal cancer treated with long-course nCRT (Mandard TRG 3). MMC cross-validation demonstrates significant separation between adenocarcinoma (*green*) and desmoplastic stroma (*pink*) pixels. MSI demonstrates significant differences in lipid abundances between adenocarcinoma glands that have proven resistant to radiotherapy vs the associated desmoplastic stroma. Furthermore, in the RGB image (top right), selected m/z values are demonstrated as being more abundant in adenocarcinoma (red, m/z 788.55) and desmoplastic stroma (blue, m/z 331.26). m/z 196.04 (green) is found exclusively in scatted adenocarcinoma glands.

5.13 Results – multiple sample comparative analysis

Each sample was annotated in the MatLab toolbox and saved as a separate annotated file classified with both the sample ID and as either RC RT+, RC RT-, N RT+ and N RT-. The ability for DESI MSI to distinguish between tissue types within samples and also to discriminate lipidomic features of cancerous vs. normal colorectal tissue has been the subject of other published work and so is not explicitly repeated here. [173 190 195 320 322] As per the findings borne out in section 5.12, equivalence was

observed in both unsupervised and supervised analyses with regards to predicting class of tissue types across both normal and cancerous tissue, in line with other work published by the group.

The next step was to analyse the centroided and annotated spectral data ‘in cancer’ but to assess for lipidomic shifts between RT+ and RT-. All 26 cancer samples were imported into this analysis and data delimited according to pixels annotated as either adenocarcinoma (RT+) or adenocarcinoma (RT-). Multivariate analysis was performed as per **figure 5.12**. For supervised RMMC analysis, the mass range was limited to the phospholipid range m/z 600-950. This focused approach (i.e., not including stroma and normal tissue) was utilised so as to maximise the chance of the principal components being due to true changes in the lipidome of tumour glands, rather than being influenced by differences in tissue types. By only accounting for adenocarcinoma-annotated pixels, the model focuses on lipidomic shifts between the RT naïve glands and those that have been subjected to RT and yet persist, meaning by definition they exhibit radioresistance.

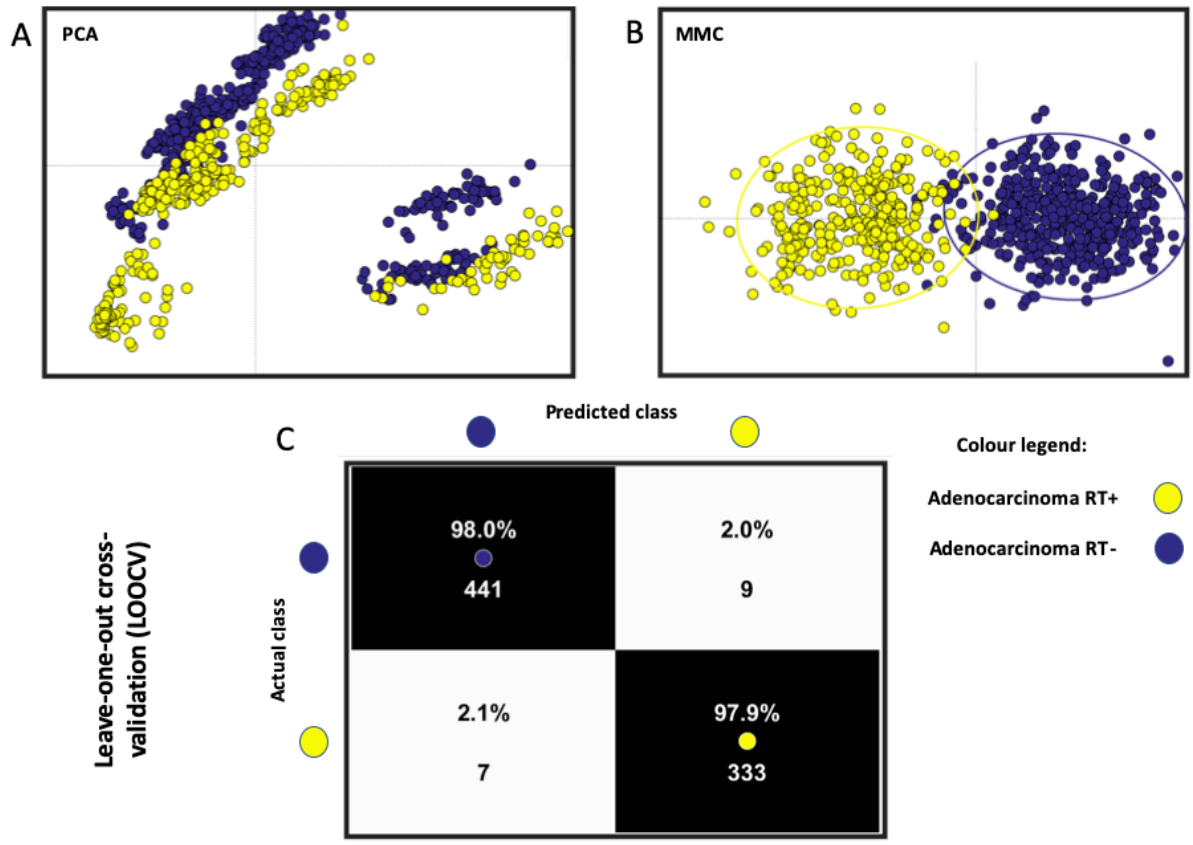


Figure 5.12 – For all annotated rectal adenocarcinoma pixels across the dataset, classified as either RT+ or RT-. (A) PCA full mass range with Eigenvalues plot and (B) MMC with mass range delimited to 600-950Da, cross validated with LOOCV model in (C). Predicted vs. actual representation in confusion matrix represents PPV and NPV.

As the figure demonstrates, there was good separation between RT+ and RT- in PCA although in this dimension there are two distinct clusters which separate out individually. This persisted despite re-runs of samples and there was no clear explanation within the metadata as to why these samples might otherwise be subject to batch effect due to their sampling. Review of peak shift data enabled a degree of correction prior to performing MMC which following leave-one-out cross validation demonstrates high predictive accuracy in the model for differentiating between RT+

and RT- tumour glands. The next step was to repeat the LOOCV as a leave-one-patient-out model with a means value for the annotated pixels taken – this lowers the predictive accuracy to 62% with poor clustering and is likely due to the low sample numbers in each cohort. For this reason, the above MMC model was used for further ANOVA to identify the lipids responsible for the classification.

Finally, and allowing for the caveat of the low sample numbers, a pilot analysis was undertaken to see if there was any clustering of samples according to TRG. In this model (**figure 5.13**), only the RT+ samples were used (*n* = 13). For the purposes of grouping, TRG 1-2 and TRG 4-5 were grouped together on the basis of overall 'good-complete response' and 'poor response – progression'. TRG 3 was grouped separately to examine whether at a lipidomic level it might demonstrate similarity to TRG 2 or TRG 4, or if indeed it stood alone as a distinct metabolic entity.

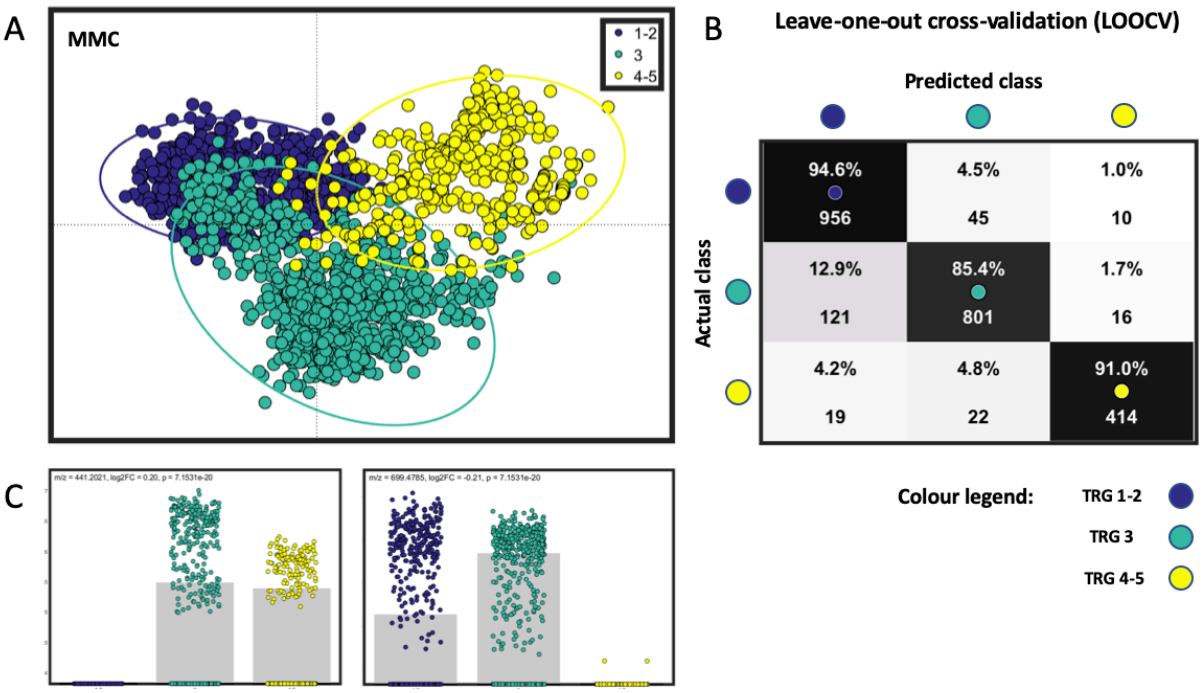


Figure 5.13 – Analysis of RT+ samples only grouped according to Tumour Regression Grade (TRG). (A) MMC plot with cross validation in confusion matrix (B). Selected discriminating compounds in the model shown in (C) plotted as $\log^2$ fold changes.

5.14 Results – the lipids responsible for the classification

For the cross-validated MMC model of RT+ versus RT- shown above, the next step was to apply ANOVA to the model to determine the lipids with significant shifts in abundance at cellular level that were responsible for driving the model and thus capable of differentiating radiotherapy naïve rectal cancer from radioresistant rectal cancer. A volcano plot was generated to demonstrate -log^{10} p-values of differentially abundant GPLs in the range 600-950Da with a p/q value set to 0.05 and log^{2} fold change to 0.1, and box plots generated from pixel spectra selected during aligned ion image/histology annotation (**figure 5.14**).

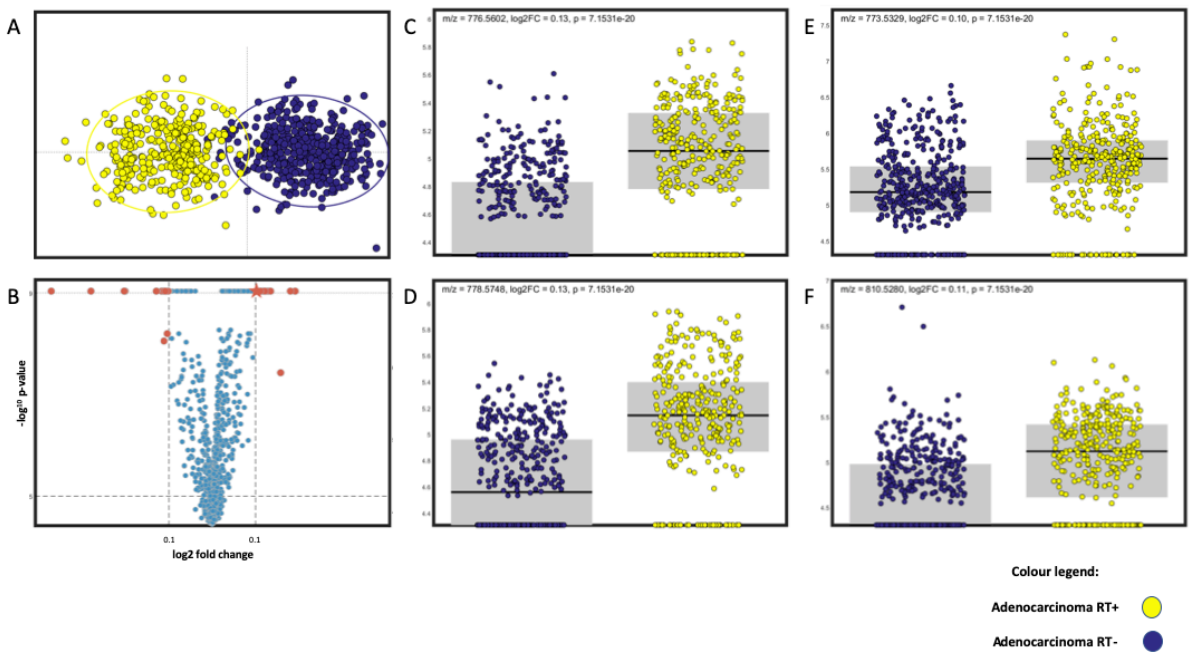


Figure 5.14 – Analysis of RT+ vs. RT- samples with MMC clustering in (A) and ANOVA volcano plot in (B) windowed to highlight compounds with significant log^{2} fold change. A total of 11 GPLs were significantly more abundant in RT+ vs RT-, with C-F demonstrating example box plots from these selected outlying ions of interest, (C) m/z 776.5602, (D) m/z 778.5748, (E) m/z 773.5329 and (F) m/z 810.5290, p-values 7.1531E^{-20}.

The 11 ions with significantly increased abundance in RT+ were assigned putative IDs following identification in encyclopaedic searching based on the ionisation method. This search confirmed likely IDs to all be phospholipids, with m/z values corresponding to a classic [M-H]- ionisation pattern in negative mode. These lipids were further investigated to formally identify them with tandem MS using the same DESI interface as described. For completion, the model was also used to identify ions which were found to discriminate RT- from RT+ adenocarcinoma; in full mass range acquisition, these are smaller molecules and have not been fully classified.

5.15 Lipid identification with DESI MS/MS

The centroided ion images (imzML) to identify representative samples with areas of high abundance of the lipids identified in 5.14. The *LTQ XL* instrument was adjusted and optimised using the settings laid out in Appendix B. For MS/MS, the definitions scan rate (turbo), scan type (full), mass range (normal) were employed with the range set to m/z 200-1000. DESI MS/MS was achieved for each lipid ion identified in 5.14 – the results are shown in **table 5.4**. For each parent ion, the MS2 fragmentation is demonstrated and was compared with known fragmentation patterns of standards in LipidMaps. Further fragmentation of the dominant fragments was performed in MS3; e.g., following loss of an 87Da fragment corresponding to the serine residue in a PS, the subsequent PA would be further fragmented in MS3 to visualise the fatty acyl chain fragments. Example fragmentation pattern spectra acquired from these experiments are shown in **figures 5.15** and **5.16**.

Measured m/z	CID	Major fragment(s) m/z	m/z CID	MS3	GPL IDs
749.5320	18.50	705.00, 665.00, 483.00 465.00, 455.00	-	-	**PA (20:5/20:0)** **PG (16:0/18:0)**
750.5435	17.50	464.33, 303.00, 259.67	-	-	**PE (P-18:0/20:4)**
771.5176	18.00	697.00, 281.33, 256.00	-	-	**PG (18:2/18:1)**
773.5330	22.00	718.00, 462.67, 418.00 352.33, 309.30, 281.33 253.67, 227.67	-	-	**PG (18:1/18:1)** **PG (16:0/20:1)**
774.5440	18.00	687.00, 464.67, 446.67 327.33, 281.00, 308.70	m/z 688 15.00	405.57 423.00	**PS (O-18:0/18:1)** **PE (P-18:0/22:6-H)**
775.5484	19.00	687.00, 511.33 283.00, 281.33	-	-	**PG (18:0/18:1)** **? PS (O-18:0-18:1) with C13**
776.5603	18.50	510.33, 283.00 329.33, 285.33	-	-	**PE (P-18:0/22:5)** **PE (P-18:1/22:4)**
778.5749	19.00	311.27, 303.00 279.67, 255.33	-	-	**PE (P-20:0/20:4)** **PE (18:2/22:1)** by [M-H_2O-H]-
790.5455	17.50	703.33, 437.30 419.35, 283.60	-	-	**PS (18:0/18:0)**
810.5281	18.00	723.00, 528.67 † 419.33, 283.33	m/z 723 14.50	437.67 419.00 283.00	**PS (18:0/20:4)**
819.5184	18.60	509.00, 327.00 283.00	-	-	**PG (18:1/22:6)**

Table 5.4 - Results of DESI MS/MS on representative cancer tissue sections subjected to RT. Major fragments as listed compared to known fragmentation patterns in LipidMaps; where only the dominant fragment was visible in MS2, MS3 performed to fragment the fatty acyl chains as shown. Not all fragments shown can be accounted for by a single GPL; where possible the various GPLs in the mixture are identified. For m/a 810.5281, the intense peak at 528.67 implies loss of a 22:4 acyl chain as a ketene, possibly from a PE, but no other clear fragments were seen. Fatty acyl chains with an O- prefix are ether lipids, whilst those with a P-prefix are plasmalogens. The PE ID for 778.5749 is obtained by neutral water loss in addition to protonation as shown. PA = phosphatidic acid; PE = phosphatidylethanolamine; PG = phosphatidylglycerol; PS = phosphatidylserine.

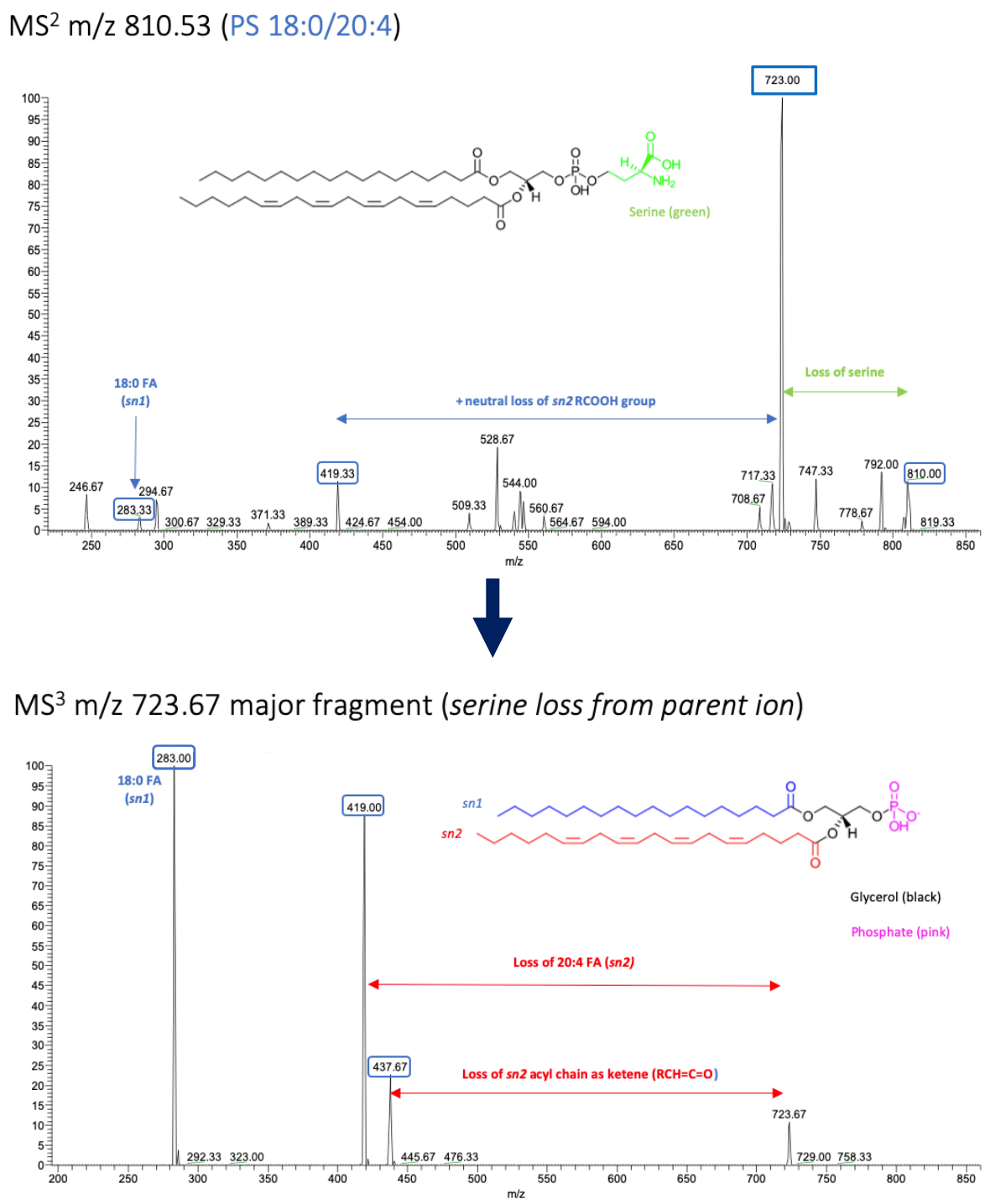


Figure 5.15 - Example MS/MS spectra (in this case both MS2 and MS3 of the major fragment) of m/z 810.5281 obtained at DESI MSI of a RT+ tumour (CRA550-15). DESI MS/MS performed on the same patient sample. Clear fragmentation pattern corresponding to an 18:0/20:4 PS – annotated peaks correspond to either the ionised FA chain itself (i.e., at 283.00, 18:0) or remnant fragment after neutral or ketene loss of the FA chain (i.e., the peak at m/z 419.00 corresponds to loss of the 20:4 *sn2* chain but the ionised *sn2* fragment itself was not clearly seen).

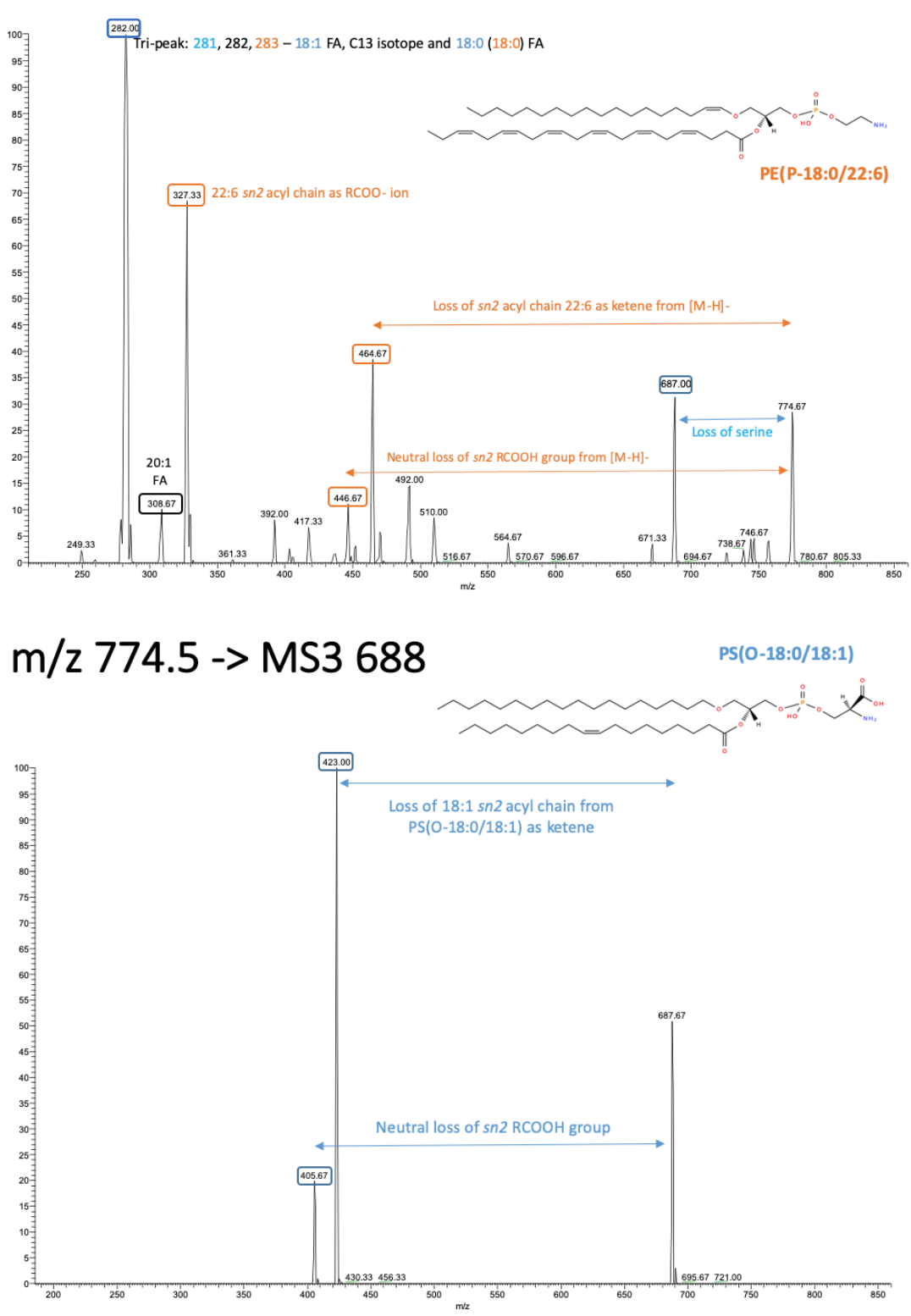


Figure 5.16 - A more complex sample demonstrating a mixture of (at least) the two GPLs shown, which both have a m/z of 774.5440. The unaccounted peak at m/z 308.7 likely corresponds to a 20:1 fatty acyl chain and therefore there is likely to be a third compound in this mixture.

5.16 Discussion

Bioactive complex lipids remain an under-investigated niche of rectal cancer metabolism with regard to response to therapies. Over the past 50 years we have moved in from observations of the 'Warburg effect', through to contemporary scientific reports on ever-more biologically representative lipidomics studies. What is evident now is that there is a more clearly-defined role for these molecules in both cytosolic and extracellular signalling as well as membrane dynamics, in defining the clinicopathological phenotype of colon and rectal cancers. [154] [177] [191] [285] [323] Mass spectrometry-based research in the field of colorectal cancer has demonstrated a significant increase in abundances of fatty acids (FAs) and glycerophospholipids (GPLs) in cancers, with trends in abundance correlating to tumour stage and pathogenicity. [324] [178] It therefore stands to reason that the same classes of lipid might have a gatekeeper role to play in the pathways associated with radioresistance in rectal cancer. The work presented in Chapter 3 further qualified this approach by identifying previously unchecked enrichment in GPL metabolism in radioresistant RC.

In this novel pilot work, DESI MSI was employed specifically to provide further insights into this problem. This is the first such study to employ a soft ionisation technique coupled with the ability to topographically correlate abundances of individual lipids to the level of persistent – and therefore resistant – rectal cancer glands as compared to matched RT-naïve tissue. The ability of DESI MSI to be able to classify histological features of rectal tissue (both normal and cancerous) to the order of >96% predictive value in unsupervised models further justified this approach. [325] [320] [173] [178]

Using this approach, DESI MSI has demonstrated an ability to provide valuable chemometric data that can be considered a truly representative biological snapshot of the rectal cancer lipidome in response to RT. In both unsupervised and supervised analytical models, it is possible to classify RT+ from RT- rectal adenocarcinoma glands purely on the basis of GPL subtypes and relative abundances. Eleven ions were found to be responsible for discriminating in this model, which were overabundant in MMC modelling in RT+ cancers. As demonstrated by the subsequent MS/MS experiment, some of these m/z values represent ionic mixtures, but all were found to be subtypes of GPLs; notably, PA, PE, PG and PS.

Phosphatidic acid is an umbrella term for all derivatives of 1,2-diacyl-*sn*-glycero-3-phosphate (DAG); that is, a molecule with 2 fatty acyl chains, and a glycerol plus phosphate head group. It is the fundamental backbone of all other GPLs, having a gatekeeper role in that it acts as the entry point into both the CDP-DAG and Kennedy pathways for GPL biosynthesis. [326] Additionally, it has essential functions in both cellular signalling and influence on membrane dynamics; mTOR and its subsequent signalling cascade is a major target of PA, and it is known to regulate activities of protein kinases and membrane trafficking through binding to such proteins. From a GPL biosynthesis standpoint, the ultimate fate of PA can be in one of either PC, PE or PS via the Kennedy pathway. [327] [328] From a colorectal cancer standpoint, LPA is known to exert effects on Peroxisome proliferator-activated receptor gamma (PPARγ), inducing proliferation and survival in cancer cells. [329]

PS acts as a further link between pathways, in that it can be synthesised by the action of PSS1 on CDP-DAG as well as on PC. PSS2 catalyses the conversion of PE

to PS (**figure 5.17**). In normal cells, PS species are predominantly found on the cytosolic surface of the lipid bilayer, whilst a migration of PS to the extracellular surface is a hallmark characteristic of apoptosis and for cellular recognition by tissue macrophages. [330] In the non-apoptotic state, increased abundance of extracellular PS has been variously linked to increased tumour aggressivity and preponderance for metastasis. [331-334]

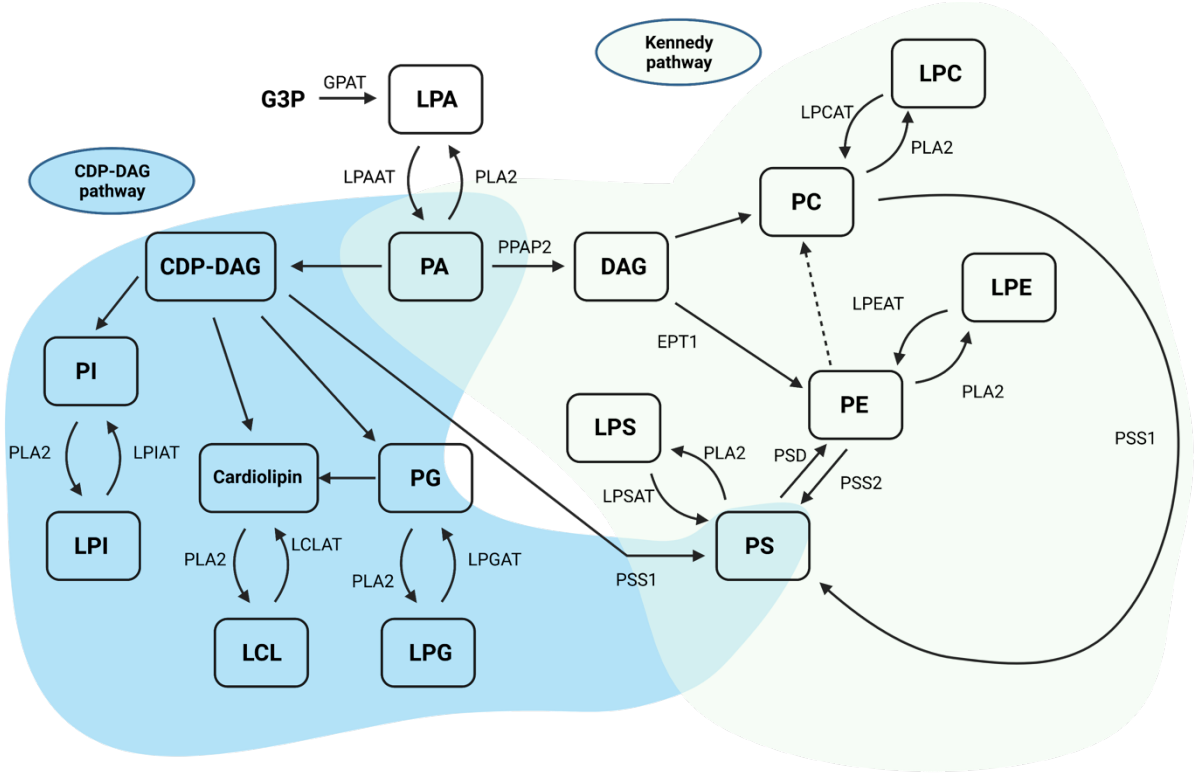


Figure 5.17 - Truncated view of GPL synthesis, demonstrating the interplay between the CDP-DAG (blue) and Kennedy (green) pathways. Fatty acyls are liberated from GPLs by the action of various isoforms of phospholipase A2 (PLA2). PA is the gatekeeper to both pathways, whereby PG, phosphatidylinositol (PI) and cardiolipins are purely products of the CDP-DAG pathway. There is intimate crossover between PC, PE and PS, with PSS1 able to act on CDP-DAG to provide linkage into the Kennedy pathway via PS and PE. PPAP2, phosphatidic acid phosphatase 2; LP(x)AT variously, lysophospho(-*x*) acyltransferase, where '*x*' is -choline, -ethanolamine, -serine, -glycerol, -cardiolipin or -inositol; PSS, phosphatidylserine synthetase (-1, -2); EPT1, ethanolamine phosphotransferase.

Phosphatidylethanolamines (PE) are the second-most abundant class of GPL in eukaryotic cells and is especially enriched on the cytosolic surface of the cell membrane. [326] Their importance is highlighted by the fact that there are four biosynthetic pathways with the CDP-ethanolamine and the PS-decarboxylase pathways taking place in different subcellular compartments. [335] Interestingly, there is such minimal redundancy in these pathways such that normal cellular function is not possible with non-functioning in any of the major PE-producing pathways as shown in **figure 5.17**. [336] Moreover, the structure of PEs confers specific properties, including modulating cell membrane events and post-translational modifications to protein structures. [337] [338] In cancer, stepwise increases in PE abundance have been demonstrated to be a hallmark of increasing tumour grade in breast cancer, but its overall role in cancer remains relatively under-reported. [339] [340] Finally, PGs (specifically PG 18:0/16:0) have been demonstrated to be overabundant in plasma of CRC patients. Although not studied in detail in rectal cancer, recent work in oesophageal adenocarcinoma has shown a stepwise increase in abundance of polyunsaturated PGs in line with increasing tumour stage. [341] [342]

In this work, the modulation and increased abundance of glycerophospholipids of the classes PA, PE, PG and PS were identified as being hallmarks of RT+ rectal cancer tissue. In addition to the GPL metabolism pathways and the possible implications this has on these results, the other factor worthy of consideration is the composition of the fatty acid (FA) chains that form these complex lipids. My findings have demonstrated increased abundances of elongated and polyunsaturated fatty acid *sn2* chains. In humans, stepwise elongation of fatty acid chains is governed by the ELOVL family of fatty acid elongases (**figure 5.18**). [343] Although fatty acid

elongation can occur in three intracellular compartments (cytosol, mitochondria and endoplasmic reticulum (ER)), the cytosol is considered the primary site for *de novo* FA synthesis by the action of fatty acid synthase (FASN), whereas microsomal FA elongation is the predominant pathway for increasing chain length beyond 12 carbons in humans. [168] As shown in the figure, the ELOVLs implicated depend on the substrate, including the degree of saturation. [344]

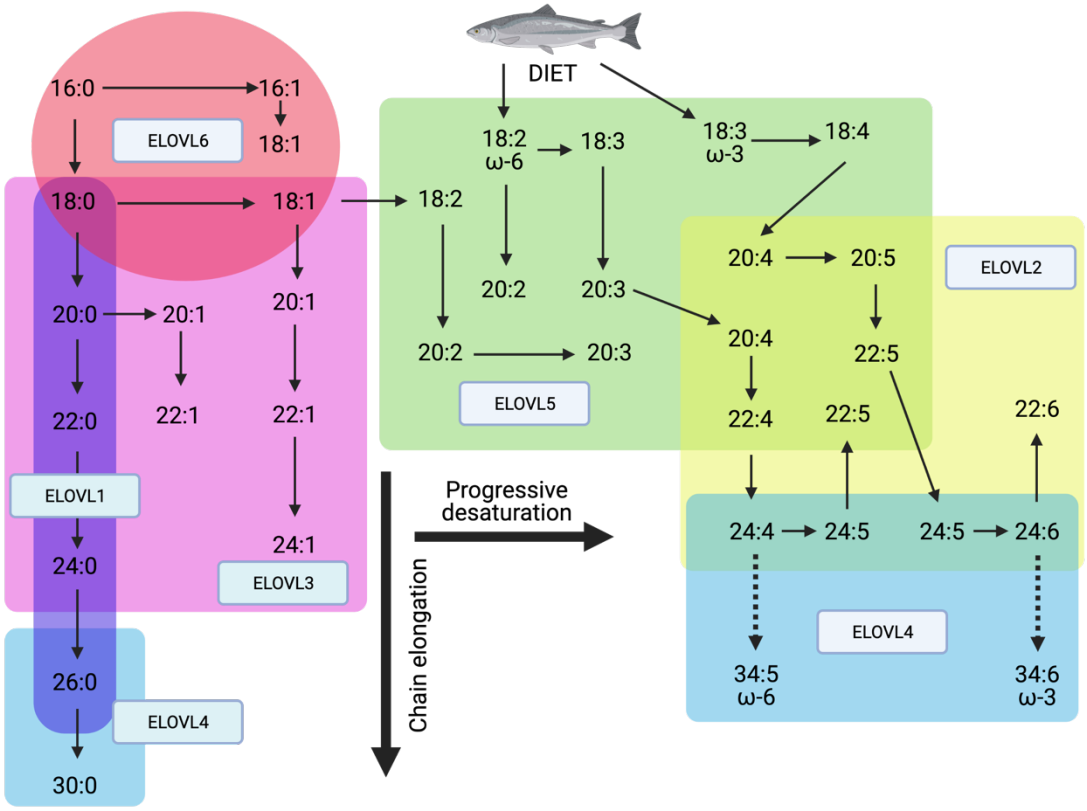


Figure 5.18 - Schematic view of ELOVL activity in humans. Omega-3 and omega-6 fatty acids have a distinct chain of derivatives through ELOVL5, -2 and -4. Saturated fatty acids elongate exclusively through ELOVL6 and -1. 'Omega' prefix refers to position of first double bond, i.e., ω-3 is 3 carbons away from terminal methyl group whilst ω-6 is at 6 carbons. Adapted from Yamashita et al [344]

As the data obtained from DESI MSI demonstrates, a number of the identified GPLs overabundant in RT+ RC, a number of the species comprise *sn2* chains that must be derived from both ω-3 and ω-6 fatty acids. PE (P-18:0/20:4), PE (P-18:0/22:5),

PE(P-18:1/22:4), PE (P-20:0/20:4), PS (18:0/20:4) and PG (18:1/22:6) all comprise a *sn2* chain that – if the above patterns of elongation by action of ELOVLs holds true – must be derived from action of ELOVL5, ELOVL2 and ELOVL4 on essential fatty acids. This seems at first glance something of a paradox given the established role of ω-3 and ω-6 as exerting anti-inflammatory and generally 'healthy' homeostatic effects, particularly in the context of cardiovascular disease and carcinogenesis. However, given that eicosanoids (mediators and regulators of cellular inflammation) are generated from C20 polyunsaturated FAs (PUFAs), it is reasonable to postulate an interplay between the actions of PLA2 and lysophospho-acyltransferases on these identified GPLs in regulating the inflammatory process in radioresistant RC cells. [345] Moreover, progressive desaturation of the FA pool and hence uptake into GPLs has recently been demonstrated as a hallmark of cancer progression in oesophageal adenocarcinoma, again using DESI MSI as a starting block for topographically-aligned lipid detection. [342] These findings are complementary with other recent breakthrough studies demonstrating the role of polyunsaturated FAs in determining the phenotype of gastric, clear cell renal and prostate cancers, with particular reference to the actions of upregulated ELOVL5. [346-348]

The final consideration regarding the structure of these GPLs is the relatively high proportion of ether lipids. Historically considered to be little more than a biological novelty, they are now known to comprise almost 20% of the cellular phospholipidome. [315] The basic chemistry that underpins the structural difference of these GPLs is the inclusion of a glycerol ether bond on the *sn1* chain; either 'ether' (alkyl) or 'vinyl' (alk-1-enyl). The nomenclature in the results section note these with the prefix P- for the alk-1-enyl *(vinyl)* GPLs and the prefix O- for the alkyl (ether) GPLs (**figure 5.19**).

Synthesis of these compounds is mediated by peroxisomal pathways with the enzyme *fatty acyl Co-A reductase 1* (FAR1) acting as the rate-limiting step before passing into the peroxisome and onto the ER.

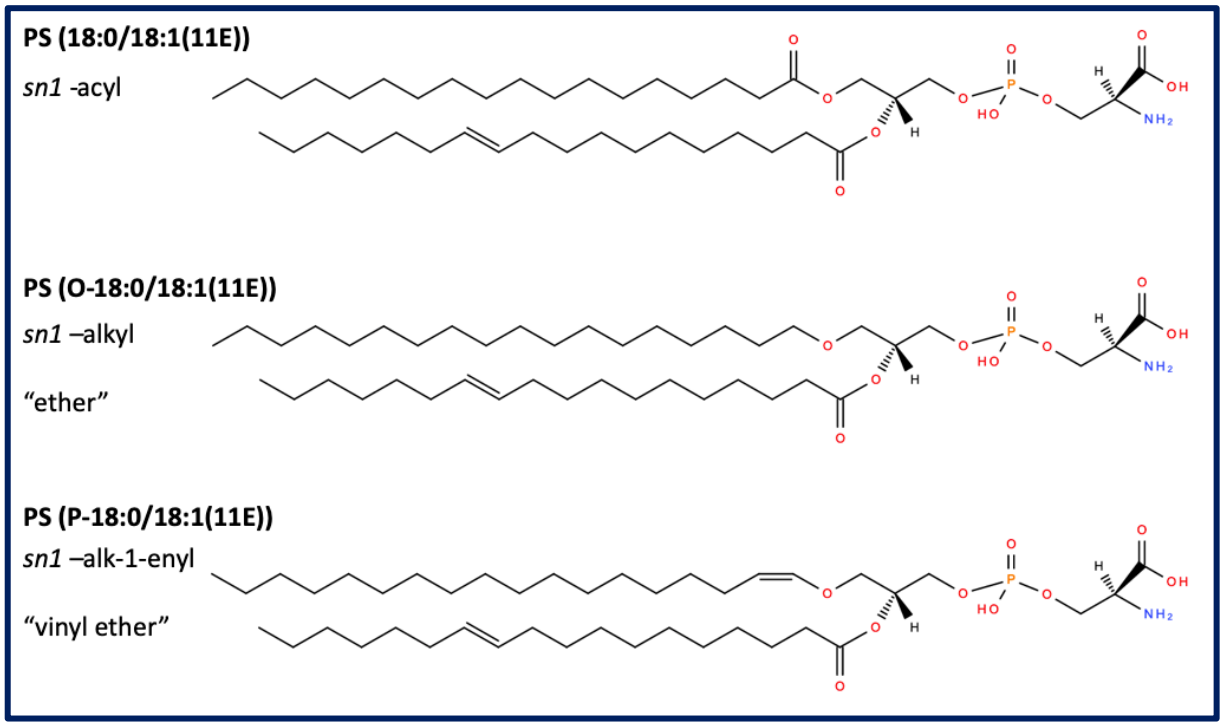


Figure 5.19 – Structural changes in, for example, PS 18:0/18:1 (diacyl-) to ether (O-; alkylacyl-) and vinyl ether (P-; alkenylacyl-) lipids. The *sn2* chain remains unchanged in both, whereas in PS (P-18:0/18:1) there is a double bond placed in the *sn1* as an alk-1-enyl group

Increased abundances of plasmalogens (P-) GPLs have effects on cell signalling and regulation of inflammatory cascades in addition to their impact on cell membrane dynamics. [349] Plasmalogen-containing cell membranes are less fluid than their predominantly diacyl- counterparts, forming more compressed and rigid structures. Coupled with their ability to act as natural antioxidants, this may in turn confer a degree of membrane-mediated protection against RT in RC, although these hypotheses remain under investigation. [350]

As a pilot study, this work is subject to number of inherent limitations. The sample cohort is quite small, and although the amount of data generated from DESI-MSI is vast, there is still the risk of in-patient bias in the statistical modelling. Running a 'leave one patient out' model did not produce adequately high predictive power to extract compounds of interest on ANOVA. On the other hand this work has aimed to establish the hallmarks of the radioresistant RC lipidome, and by definition in samples displaying residual tumour there is a degree of heterogeneity within the sample. It was therefore considered feasible in this pilot work to annotate discrete regions of tumour displaying histological radioresistance and include all such annotations in the statistical model. Furthermore, the DESI MS/MS methodology employed here does not represent a gold standard in fragmentation and lipid ID discovery as this still requires the controlled conditions employed by chromatography-linked MS (UPLC-MS, for example). However, given the highly heterogeneous nature of the clinical problem, it was deemed important to be able to fragment the exact ions obtained from cancer tissue using an equivalent ionisation method rather than using laboratory standards or liquid tissue extracts. Using the DESI MS/MS approach has enabled identification of biologically plausible lipids that not only fit with the hypothesis generated from the previous *in silico* exploration of GPL pathway perturbation, but also lend themselves to an equally plausible biological mechanism that may be partly responsible for radioresistance.

Nonetheless, a key limitation to this particular workflow is the fact that only negative ionisation mode was used for DESI MSI. Although this does not directly impact the identification of the GPLs that were found to be represented in negative mode, it does mean that this particular assessment does not provide a complete overview of GPLs

as zwitterionic phospholipids (notably phosphatidylcholine (PC), LPC and sphingomyelin (SM)) are more efficiently ionised in the positive ionisation mode. [351] With regards to PCs, this is especially important to take into consideration given their location in the Kennedy pathway and their interrelation with PE and PS. Time and instrument constraints prevented proceeding to repeat the entire analysis in positive mode, although it is interesting to note the abundance of identified PEs (which is also zwitterionic) in negative mode as demonstrated here. It would be remiss to dismiss the role of PC as merely 'bricks and mortar' in the cell membrane without further analysis, as it has been demonstrated to contribute to a number of cell signalling processes. [352] A further limitation of note is that the DSI MSI approach cannot determine subtle differences in isobaric molecules using the method described; such differences may have biological relevance and cannot be overlooked. Clearly, therefore, this work would require further validation studies and further MS/MS using cellular extracts for UPLC-MS and more robust structural analysis. Increasing the sample cohort would lend to the power of such a study. Developing confidence in the ID of the GPLs found to be enriched in radioresistance would then naturally lead on to pathway elucidation, although there are already some plausible candidate pathways that emerge from this work.

Finally, all metabolic profiling studies including those performed with MS are subject to possible 'batch effects'. These can be due to minute shifts in environmental conditions or because of the condition of either the instrument or the ionisation interface. Bioinformatics approaches with peak shift correction can be useful in improving data at the pre-processing stage. To mitigate these effects, the geometry of the sprayer was always kept consistent, and the instrument cleaned fully every 72

hours. Calibration was performed every day the instrument was in use. Theoretically the possibility of 'in-sample' batch effect may exist due to the long analysis times (in some cases, over 4 hours per sample). This is why stringency was applied when analysing individual and pooled samples before accepting data for analysis.

5.17 Conclusions

The first two aims of this section of my work were to employ a DESI MSI strategy to build a spectral model of the rectal cancer microenvironment that could account for shifts in the RC lipidome in response to radiotherapy, and to further build models that could define the key players in this modulation of the RC lipidome. As demonstrated in my work, I have shown that ions present in cancer glands in the GPL mass range can discriminate radioresistant RC from RT-naïve RC tissue.

The final aim was to identify the lipids responsible for the discrimination in the model and to postulate a model for how these might arise. DESI MS/MS has identified these GPLs as being part of the subclasses PG, PE, PS and to a lesser extent PA. Some of these molecules have also been demonstrated on MS/MS to have fragmentation patterns consistent with ether lipids, which have a putative role in alteration of cell membrane fluid dynamics by way of increased rigidity and through their action as antioxidants. Furthermore, I have demonstrated that the involved acyl chains in the identified GPLs demonstrate a significant degree of saturation, which may in turn have interplay with eicosanoid-mediated cellular inflammatory responses. This pilot work has demonstrated significant and important original data which lends

new insights into modulation of cellular GPLs as being hallmarks for radioresistance in rectal cancer, worthy of further investigation in larger-scale studies.

Chapter 6
The genomic landscape in radiation response – NextGen sequencing of pre- and post-radiotherapy specimens in rectal cancer

6.1 Introduction 182

6.2 Methods – Patient cohort and tissue collection protocol 184

6.3 Methods – Extraction of genomic DNA from FFPE (pre-RT) rectal cancer tissue samples 184

6.4 Methods – Extraction of genomic DNA from fresh frozen (post-RT) rectal cancer tissue samples 187

6.5 Methods – DNA quality control, library preparation and NextGen WGS sequencing 187

6.6 Methods – WGS data analysis 188

6.7 Results – DNA extraction and sequencing run QC 189

6.8 Results – Analysis of the burden of SNCAs on radiotherapy response 190

6.9 Results – Analysis of the impact of divergence of genomic aberration on tumour regression grade 200

6.10 Discussion 203

6.11 Conclusions 211

6.1 Introduction

The *in-silico* work in Chapter 3 utilised a legacy microarray sequencing dataset to reveal previously unreported enrichment in glycerophospholipid (GPL) metabolism pathways in rectal cancer (RC) that may represent a hallmark feature of radioresistance. Over the course of the last decade, microarrays have become largely superseded by the introduction of Next-Generation Sequencing (NGS) approaches whereby RNA-Seq has overtaken the role of microarrays for gene/probe-based sequencing. The umbrella term for the technology also includes approaches capable of taking a holistic view of the entire genome, such as whole-genome sequencing (WGS) and whole-exome sequencing (WES). The key differential step in these technologies is that microarrays work on the basis of hybridization of genes to probes (and as such is affected by the probe set), whereas NGS technologies work on the basis of DNA/RNA amplification.

Despite the power and value of microarrays, they do have a number of intrinsic limitations which are summarised in **table 6.1**, shown alongside with how NGS-based approaches can rectify these. [124] Whereas initially these approaches were prohibitively expensive, with the advent of high-throughput platforms the costs of NGS are now within reach of discovery science projects. Nonetheless, the routine use of genome sequencing in translational medicine remains controversial, mostly on the grounds of cost.

In order to prevent replication of work and to provide a more holistic overview of the genomic landscape in rectal cancer radiotherapy response, it was decided to employ a whole-genome sequencing method as a starting point. In order to investigate

the underlying genome of radioresistant phenotypes, both pre-RT and post-RT rectal cancer tissues were sequenced to evaluate the role for tumour heterogeneity. This work should serve as a suitable starting block for deeper genomic interrogation at an RNASeq level and will enable interrogation of the whole RC genome to establish the degree of somatic copy number aberration pre- and post-RT and how this might differ with regard to clinicopathological tumour regression grade (TRG). Recent work has optimised the workflow for low-pass whole genome sequencing with coverage in the order of 0.1x. Even at this level, very useful data can be obtained on SCNAs whilst balancing against the cost of sequencing, which increases significantly with depth of sequencing and number of samples.

	Microarrays	*NextGen sequencing*
Genome features	Microarray design requires *a priori* knowledge of the genome or specific features Incorrect or outdated features at risk of affecting array effectiveness	Genome annotation not essential; unsupervised *de novo* assembly is possible
Experimental bias	Cross-hybridization between similar sequences can be an issue Array is restricted to non-repetitive components of the genome and analysis of related genes or features is challenging	Direct sequencing of material within the pipeline removes experimental bias and cross-hybridization problems Single-nucleotide resolution means differential expression from gene alleles that differ by as little as one nucleotide is possible
Technical limitations	High signal-to-noise ratios Competitive hybridization Detection of low-abundance sequences can be challenging Quantitative resolution of highly-represented sequences equally difficult	Unlimited, fully-quantitative dynamic range of signal Adept at detecting changes in both rare and highly-expressed sequences
Material requirements	Micrograms (µg) of DNA required for hybridization Separate-step PCR amplification of material can be a source of bias	Nanograms (ng) of DNA required PCR amplification generally not required
Data output and reproducibility	Wide variability in microarray formats Difficulties with reproducibility of data	All platforms have same raw data output, therefore reproducibility is much improved

Table 6.1 – Comparison between microarrays and *NextGen* sequencing technologies. NGS serves to rectify several of the limitations of microarrays especially with regard to reproducibility and data limitations

6.2 Methods – Patient cohort and tissue collection protocol

The same patient cohort (*n* = 26) was used as described in Chapter 5. In addition to the snap-frozen RC tissue harvested at resection (which was analysed with DESI MSI), the matched, archival, formalin-fixed, paraffin-embedded (FFPE) biopsy blocks were also obtained. These biopsy blocks contained between 1 and 12 variously sampled, macroscopically distinct regions of the primary tumours are were obtained for both the RT+ and RT- patients.

As part of the frozen tissue preparation prior to DESI MSI outlined in section 4.4, multiple serial 10µm sections had already been mounted and frozen at -80^{O}C. For the archival FFPE biopsy blocks, these were serially sectioned onto slides in the conventional manner for clinical histopathology preparation to a thickness of 5µm. Multiple sections were taken, again mounting two serial sections per slide. An index H&E staining was performed on the first FFPE biopsy slide in the series to direct the tissue microdissection outlined in the next section. H&E index slides had already been made as part of the DESI MSI work for the ion/histological image alignments.

6.3 Methods – Extraction of genomic DNA from FFPE (pre-RT) rectal cancer tissue samples

Following H&E staining of an index slide, samples were examined with the assistance of an expert histopathologist (Prof Robert Goldin, Imperial College London). Each biopsy fragment within a sample was interrogated and annotated to ensure adequate tumour content. Preparation of genomic DNA from the FFPE sections largely followed the protocol as laid out in the High Pure FFPET DNA Isolation

Kit (Roche Life Sciences, Germany), with a number of modifications as follows. The slides for sequencing were then prepared for slide-scrape microdissection by first performing a de-waxing procedure by immersing the slide in xylene for 5 minutes before refreshing with new xylene and repeating. The tissue on the slide was then rehydrated through graded ethanols (100% for 3 minutes, 90% for 3 minutes, 70% for 3 minutes) and then in Optima™ LC-MS grade water (Fisher® Scientific, Leicestershire UK). Following this, targeted slide scrapes using a 20µL sterile, DNA-grade, filtered pipette tip were performed, which after acquiring tissue was dipped and flushed into the PCR grade 1.5mL reaction tube (LoBind; Eppendorf) containing pre-loaded 180µL of lysis buffer. Only one annotated region of interest was scraped per reaction tube and labelled as a distinct region number within the sample as per **figure 6.1**. 70µL of Proteinase K was added for digestion and the tube vortexed briefly in several intervals. The tubes were then transferred to a Thermomixer® (Eppendorf) for the first incubation step, at 56^{O}C for 60 minutes at 600rpm. Following this, any tubes without clear lysate were incubated for a further 60 minutes at the above settings. For all samples, the incubation temperature was then increased to 90^{O}C and again left on the Thermomixer® for 60 minutes at 600rpm. The remaining extraction protocol was followed as per the kit protocol as shown in Appendix C, with the final modification to the elution step being to add 20µL of Elution Buffer to the filter tube, incubating for 5 minutes at room temperature (15-25^{O}C) and centrifuging at 8,000 x g for 1 minute before the repeating the elution step with a further 20µL of Elution Buffer. Carrier RNA was not used. Samples were subsequently stored at -80^{O}C prior to sequencing.

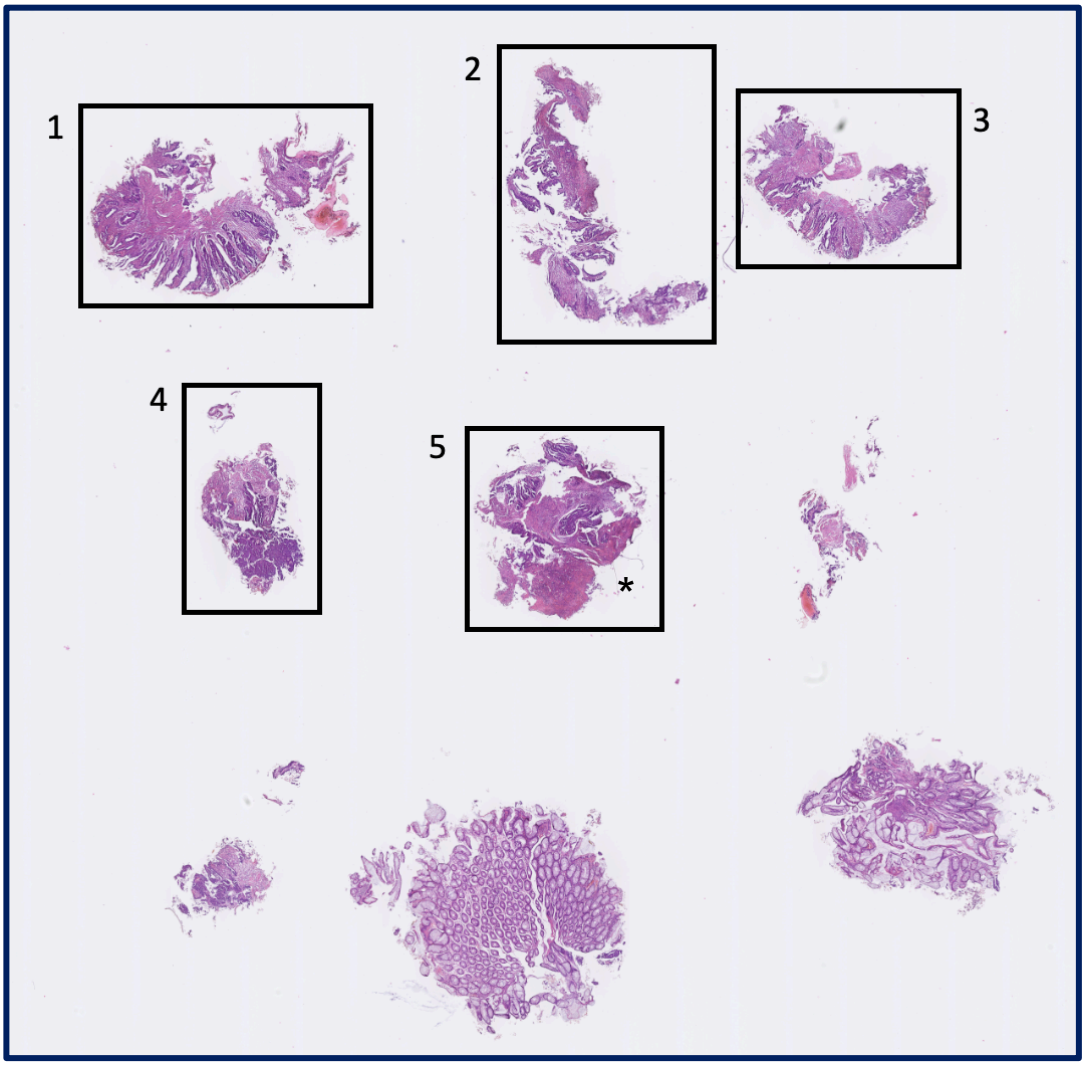


Figure 6.1 – Example of H&E annotation of the multiple biopsies taken pre-RT from CRA550-15. Separate sections taken from macroscopically distinct regions of tumour and embedded into a single paraffin wax block then cut up and slide-mounted as above. Sections 1-5 validated as containing rectal adenocarcinoma glands with varying degrees of stroma. Unlabelled sections are either, for example, normal tissue (*) or necrotic fragments and therefore not suitable for sequencing.

6.4 Methods – Extraction of genomic DNA from fresh frozen (post-RT) rectal cancer tissue samples

The fresh frozen nature of the post-RT tissue samples (i.e., those harvested at surgical resection, *n* = 13) were embedded in water ice (rather than OCT) and so merely required thawing on the benchtop for 5 minutes following removal from the -80^{O}C freezer. The slide scrape microdissection then proceeded in a similar fashion to that outlined above in 5.1, with care taken to avoid regions of dense stroma and necrosis in the sections. Again, microdissection was guided by annotated H&E slides – this time from the index DESI MSI experiments. For these samples, the entire region of tumour was sequenced as a whole as all remaining glands were by their definition radioresistant. The Qiagen QIAamp DNA Micro kit was used for DNA extraction from these samples. The protocol was followed as per the kit, using a Thermomixer® (Eppendorf) for the incubation phases and with an amendment at the eluting step - similar to the FFPE extractions – whereby 20µL of Buffer AE was added at the elution step, left for 5 minutes on the benchtop then centrifuged at 20,000 x g for 1 minute. This step was repeated once more to yield a total eluate of 40µL. Samples were transferred to a -80^{O}C freezer prior to sequencing.

6.5 Methods – DNA quality control, library preparation and NextGen WGS sequencing

DNA extract quality was first verified by GAPDH PCR as described in the publication by van Beers et al; this provides a quantitative evaluation of a 'good' versus a 'bad' quality sample and is broadly based on PCR reaction with four primer sets, producing 100, 200, 300 and 400 bp fragments from non-overlapping regions in the

GADPH gene on Chromosome 12. [353] This was repeated for both FFPE (pre-RT) and fresh frozen (post-RT) DNA sets. Poor quality samples were identified, and extraction protocols were repeated on serial sections. Next, library preparations were performed with the NEBNext® Ultra™ II DNA Library Prep kit (New England BioLabs, MA, USA). A repair step was used for the FFPE biopsy specimens using the NEBNext® FFPE DNA repair mix. Protocols for these kits are shown in Appendix D. For library preparation, a fragmentation time of 6 minutes was used for FFPE samples and 13 minutes for fresh frozen samples. Libraries were dual indexed, and between 4-14 cycles of PCR amplification was used, depending on DNA input and quality. Libraries were equimolar pooled and subsequently sequenced on the Illumina NextSeq using 75bp paired end reads.

6.6 Methods – WGS data analysis

For sequencing data quality control, per-base quality, GC content, and adapter contamination of the raw read sequences in each fastq file were assessed using FastQC (versions 0.11.5). Read clipping in a subset of samples that were found to have adapter contamination was performed using the Skewer tool (versions 0.2.1). A minimum read length of 35bp after trimming was required. Reads were aligned with BWA (versions 0.7.15 using default parameters) to UCSC human genome reference (version 19).

To check the quality of the alignments, a combination of Picard tools (*CollectWgsMetrics, CollectInsertSizeMetrics, CollectAlignmentSummaryMetrics,* all versions 2.18) was used. Duplicate reads were marked using Picard *MarkDuplicates*.

Of note, it was ensured that the percentage of reads removed was not excessive (<35%) and that the target coverage and depth were obtained.

For copy number aberration (SCNA) calling in shallow whole genome sequencing data, the package QDNAseq was used to call SCNAs, an open-source analytical package described in the index publication by Sheinin et al. [354] QDNASeq is implemented in R (R Core Team 2020) and is available in Bioconductor. [355] The QDNAseq package bins reads from the input bam file into a user-specified series of 500Kb bins, apply filtering (loess residual and blacklisting), correct the bins for GC content and mappability, then normalises read counts per bin using the 'median' method and smooths outlier counts. QDNAseq was also used to segment per-bin read counts into 'segments' of equal copy number. The 'sqrt' function was used to transform the data, bin counts normalized with the 'normalizeSegmentedBins' function.

All statistical comparisons were performed in the R statistical programming language. The Copy number frequency plot was made using the R package *Copynumber*.

6.7 Results – DNA extraction and sequencing run QC

In both FFPE and fresh frozen cohorts, all samples extracted had adequate concentrations (per µL) and overall volumes of DNA (>10ng) to proceed to WGS. GAPDH PCRs were run for randomly selected samples and an example image from gel electrophoresis is shown in Appendix E. For one sample (CRA543-15) there were discrete regions of tumour and high-grade dysplasia. These were separately

sequenced. None of the samples in either cohort failed sequencing on technical grounds, and were all included in subsequent analysis. Median coverage achieved from the fresh frozen cohort was 0.56x; a summary of per-sample reads is shown in **Appendix E**. For the FFPE biopsy samples, median coverage was 0.3x.

6.8 Results – Analysis of the burden of SCNAs on radiotherapy response

All samples were processed using the QDNASeq software package as described for analysis of reads and SCNAs. A sample summary is shown in **figures 6.2** and **6.3**, with plots to demonstrate the number of aberrant bins, colour coded to SCN losses or gains. To estimate validity of the sequencing data against a benchmark dataset, all samples were aggregated (both pre-RT FFPE, post-RT fresh frozen and RT-naïve fresh frozen; *n* = 65) and a combined SCNA plot mapped against the available data on rectal cancers in The Cancer Genome Atlas (TCGA) as in **figure 6.4**. A comprehensive summary of SCNA profiles alongside the clinicopathological features of each sample are shown in **Appendix E**.

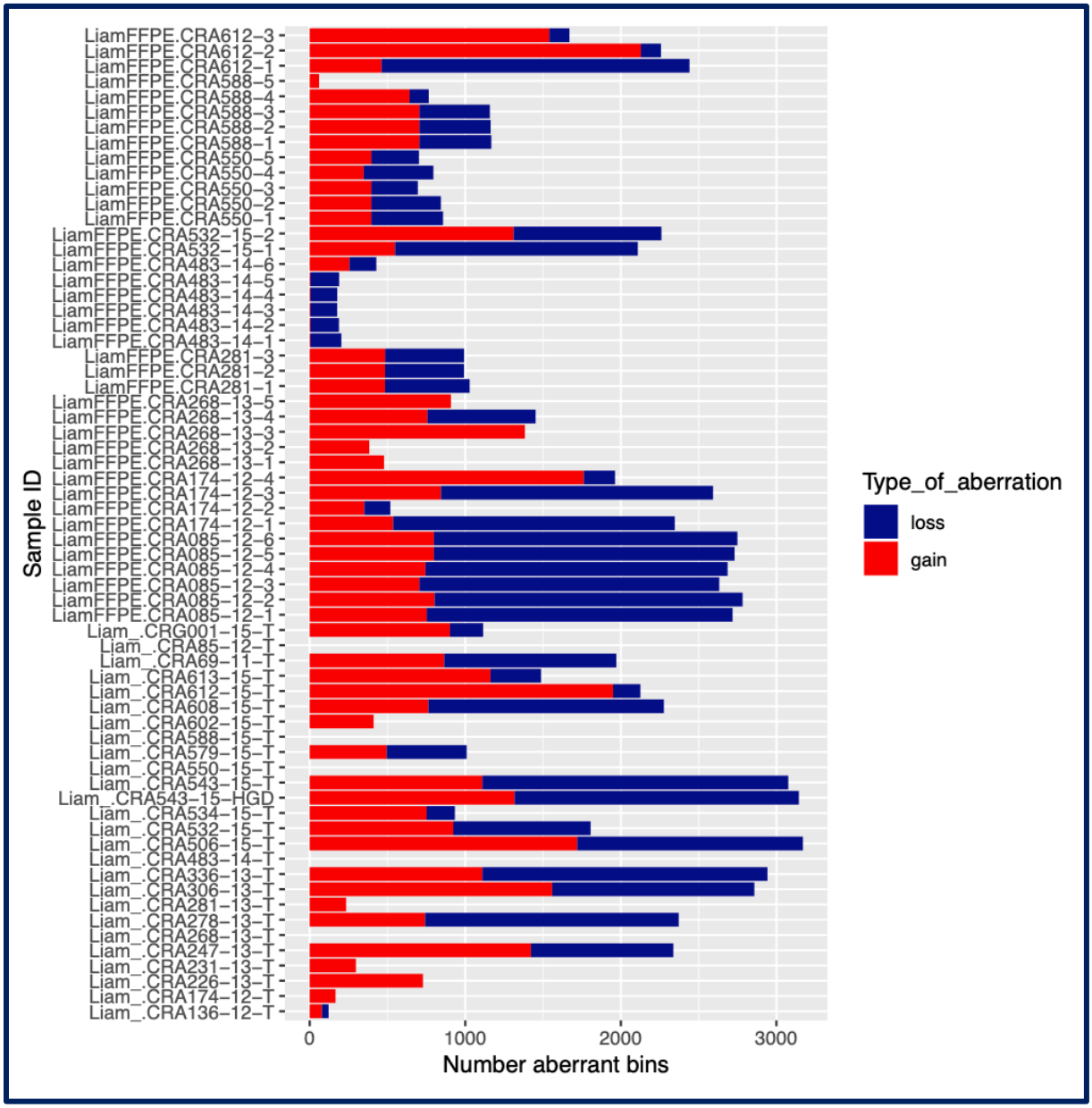


Figure 6.2 – All samples (*n* = 65), including pre- and post-RT. For those with minimal to no recorded aberrant bins, examining cellularity of samples revealed very small amounts of tumour either due to necrosis or stromal contamination. The majority of these were in the post-RT cohort whereas the matched pre-RT multi-region samples demonstrate good cellularity

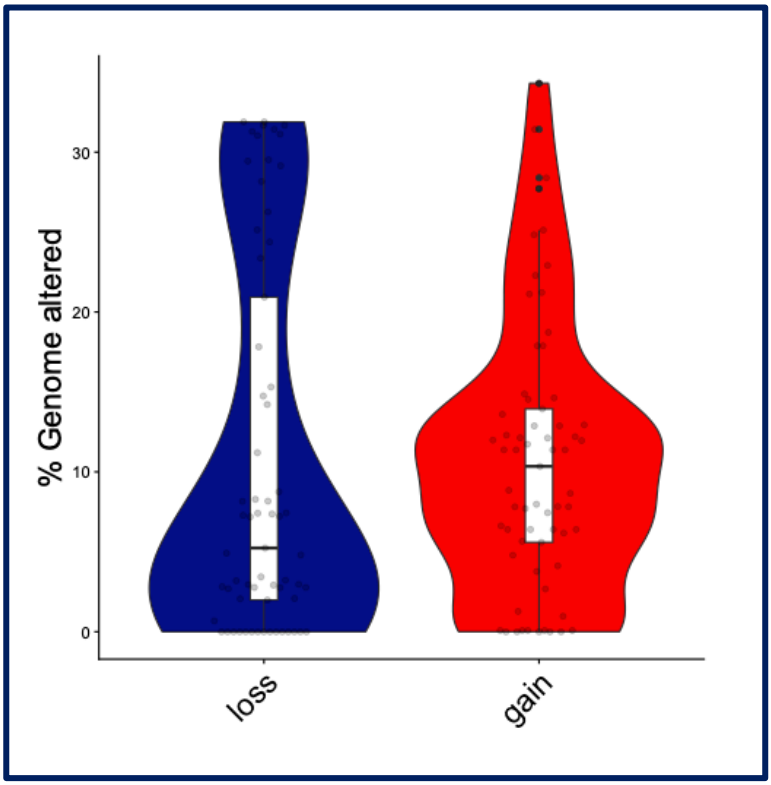


Figure 6.3 – Violin plots of all samples (*n* = 65) - including pre- and post-RT - as per figure 5.2, plotted against percentage of genome aberrated as measured by frequency of SCN gains or losses

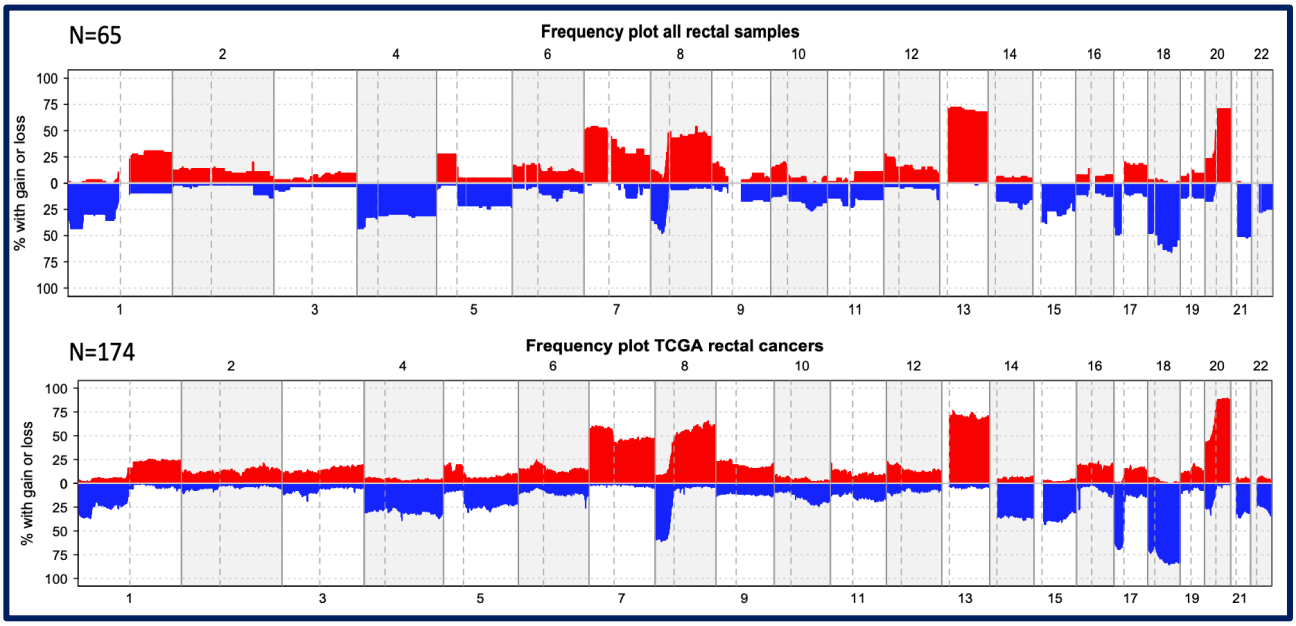


Figure 6.4 – Total SCNAs plotted as losses (blue) and gains (red) from this cohort as compared to available annotated datasets within The Cancer Genome Atlas; hallmark changes (e.g., gains on Chr7, Chr13, Chr20, losses on Chr4) are present, demonstrating validity of dataset

As shown in the figures, there is significant blandness in some of the post-RT samples despite histological appearances of definitive islands of cancer tissue in all samples at slide microdissection. However, on all-samples comparison to the rectal cancer TCGA dataset, this cohort of sequencing data appears to demonstrate representative features correlating to SCNAs (characteristic gains on Chr13q, Chr7p and Chr7q and losses on Chr4p and Chr4q). Individual SCNA analysis of representative samples from the dataset above is shown in **figures 6.5** – **6.9**. As shown, even with the small biopsy fragments in some samples the sequencing run was able to count in excess of $9E^6$ reads.

The five figures that follow compare three histologically comparable tumours from the cohort: CRA306-13 (pT3 pN0 M0, 0/23 lymph nodes involved, EMVI positive, no tumour perforation, no radiotherapy given); pre- and post-RT samples from CRA174-12 (ypT3 ypN0 M0, 0/11 nodes involved, EMVI positive, no tumour perforation, neoadjuvant RT and CapOX given, TRG2) and pre- and post-RT samples from CRA532-15 (ypT3 ypN0 M0, 0/28 nodes involved, EMVI positive, no tumour perforation, neoadjuvant RT and CapOX given, TRG4). CRA306-13 demonstrates an example of a RT-naïve tumour displaying a significant SCNA burden. Following a 'good' TRG after RT in CRA174-12 (TRG2 = fibrosis with scattered tumour cells) the landscape of SCNAs is much blander than in any of the 4 pre-RT biopsies. Although this could be explained by poor cellularity (i.e., excessive stromal contamination), the cellularity count of this particular sample was 0.3 which – although low – is not as low as some of the TRG 1-2 samples, where cellularity was in some cases as low as 0.1 (see **Appendix E**). What is interesting to note that across the 4 pre-RT biopsies there

is significant heterogeneity in the degree and variation of SCNA across the macroscopic landscape of the tumour.

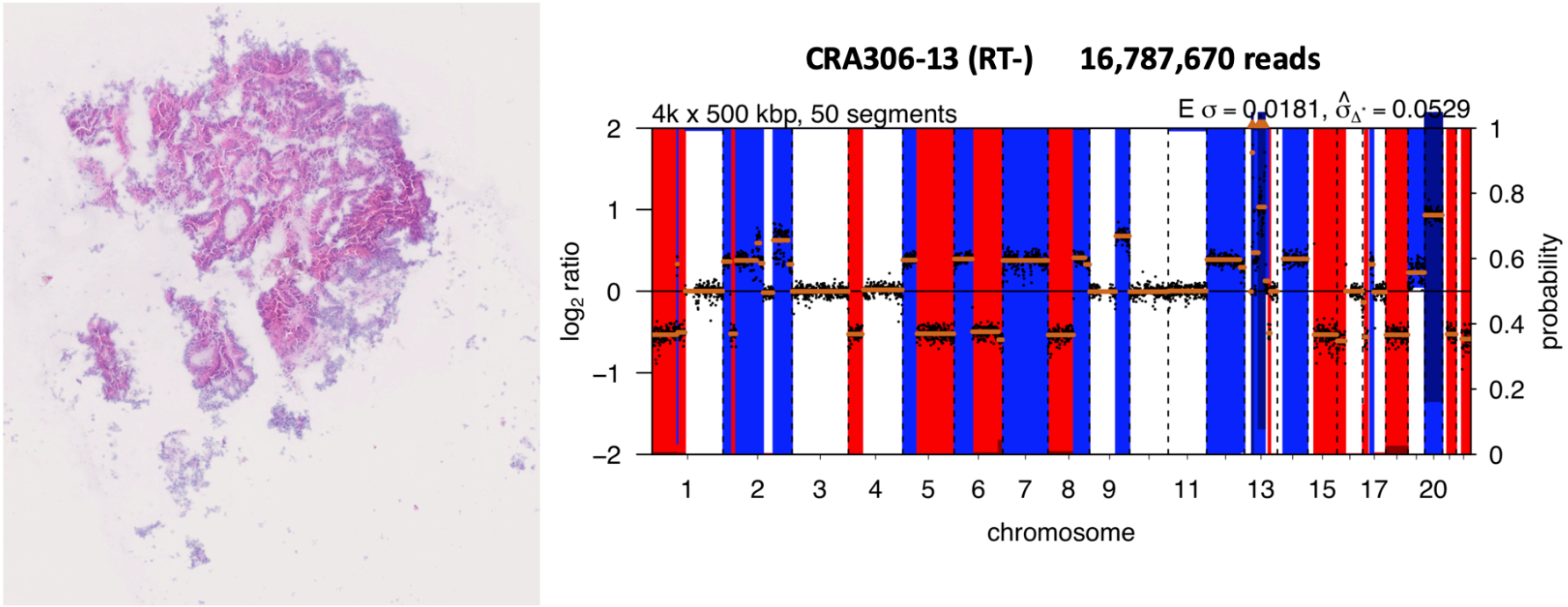


Figure 6.5 – Histology image from slide-scrape dissection and matched copy number (CN) plot from CRA306-13 (T3 N0 M0, 0/23 lymph nodes involved, EMVI positive, no tumour perforation, no radiotherapy given). All tissue on slide is tumour and so all digested for sequencing. For copy number plot, blue bars indicate net CN gains and red bars indicate losses.

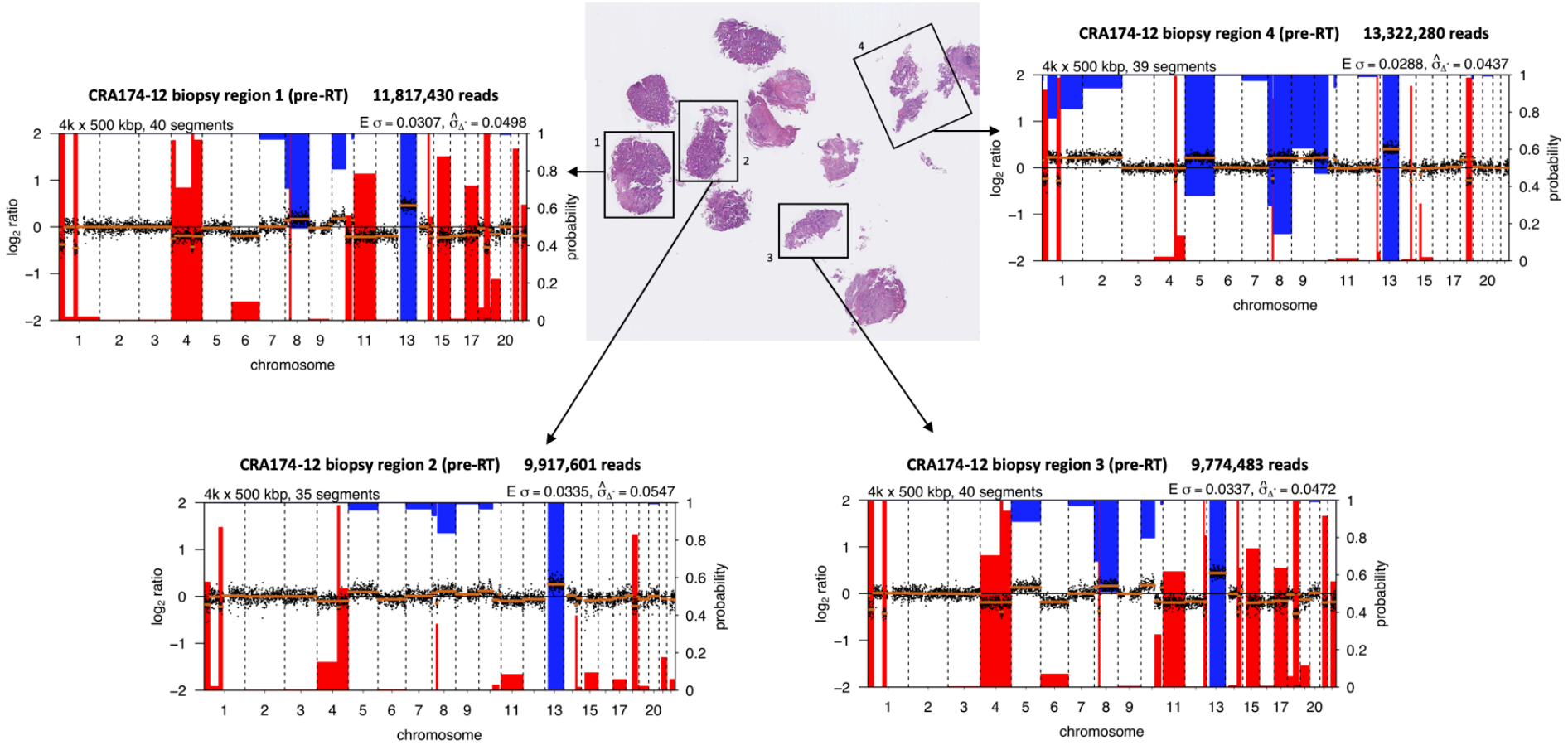


Figure 6.6 – Histology image from slide-scrape dissection and matched copy number (CN) plot from CRA174-12 (T3 N0 M0, 0/11 nodes involved, EMVI positive, no tumour perforation, neoadjuvant RT and CapOX given, TRG2). Four distinct macroscopic regions identified as being composed of sufficient tumour for separate sequencing as annotated. Significant variation in SCNA burden noted between macroscopic regions of primary tumour. Blue bars indicate net CN gains, red bars indicate net CN losses.

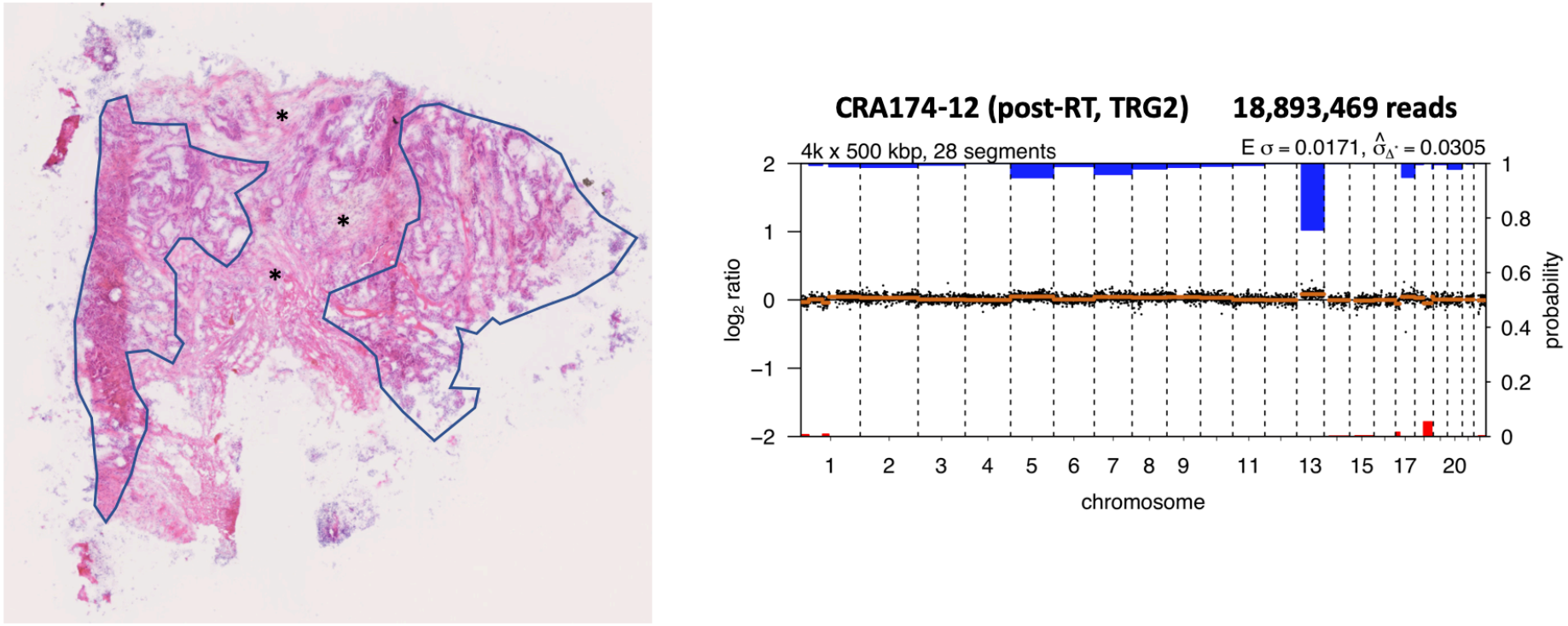


Figure 6.7 – Histology image from slide-scrape dissection and matched copy number (CN) plot from CRA174-12 (T3 N0 M0, 0/11 nodes involved, EMVI positive, no tumour perforation, neoadjuvant RT and CapOX given, TRG2) tissue taken at resection after RT. Areas of tumour annotated by outlines, asterisks (*) denote areas of stroma which were excluded from dissection and digestion. Bland SCNA pattern despite significant SCNA burden pre-RT. Cellularity of data estimated at 0.3. Blue bars indicate net CN gains, red bars indicate net CN losses.

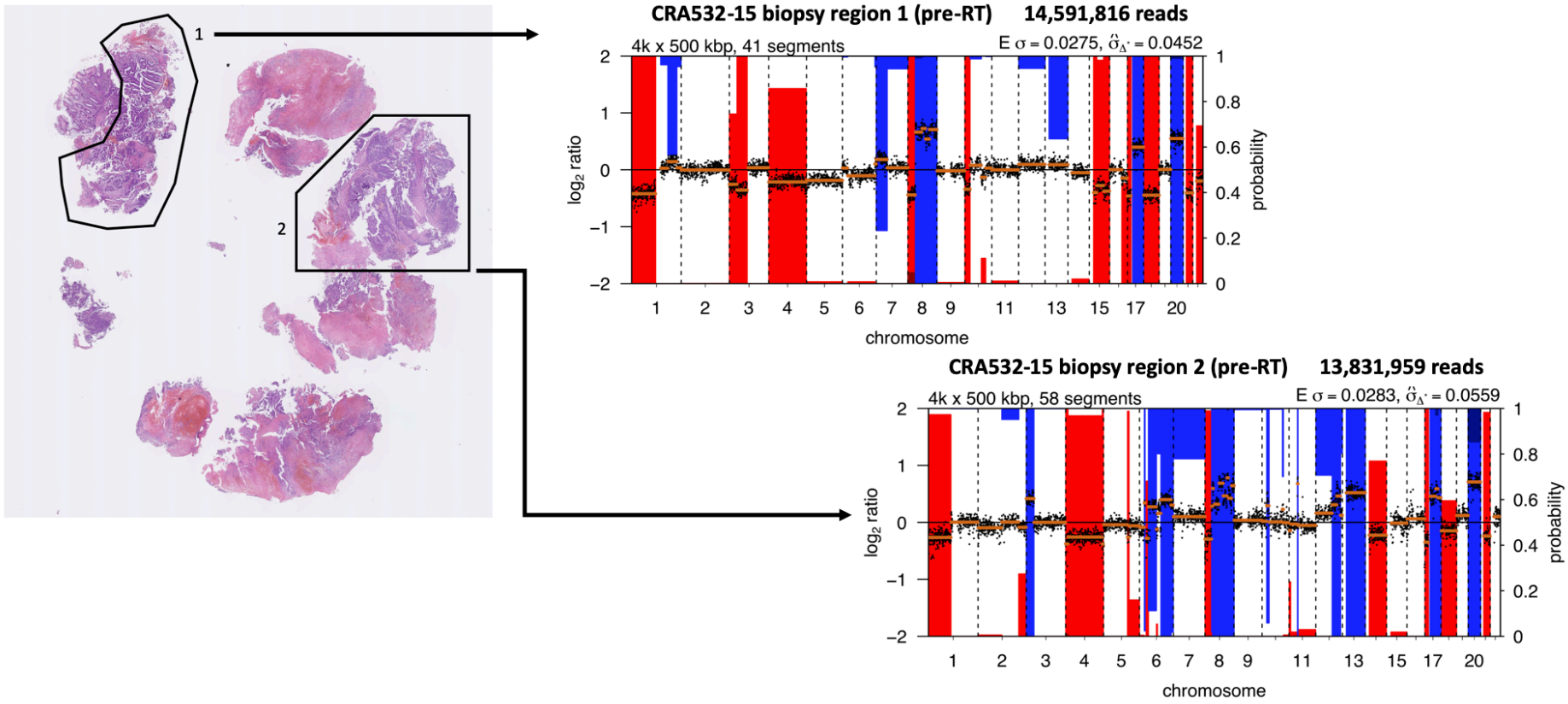


Figure 6.8 – Histology image from slide-scrape dissection and matched copy number (CN) plot from CRA532-15 (T3 N0 M0, 0/28 nodes involved, EMVI positive, no tumour perforation, neoadjuvant RT and CapOX given, TRG4) biopsies taken prior to RT. Two tumour-bearing areas sequenced separately with remainder being excluded as necrotic/stroma. Biopsy 2 demonstrates a higher burden of SCN gains, notably across Chr 1q, Chr 6 and Chr 12 compared to biopsy 1. Blue bars indicate net CN gains, red bars indicate net CN losses.

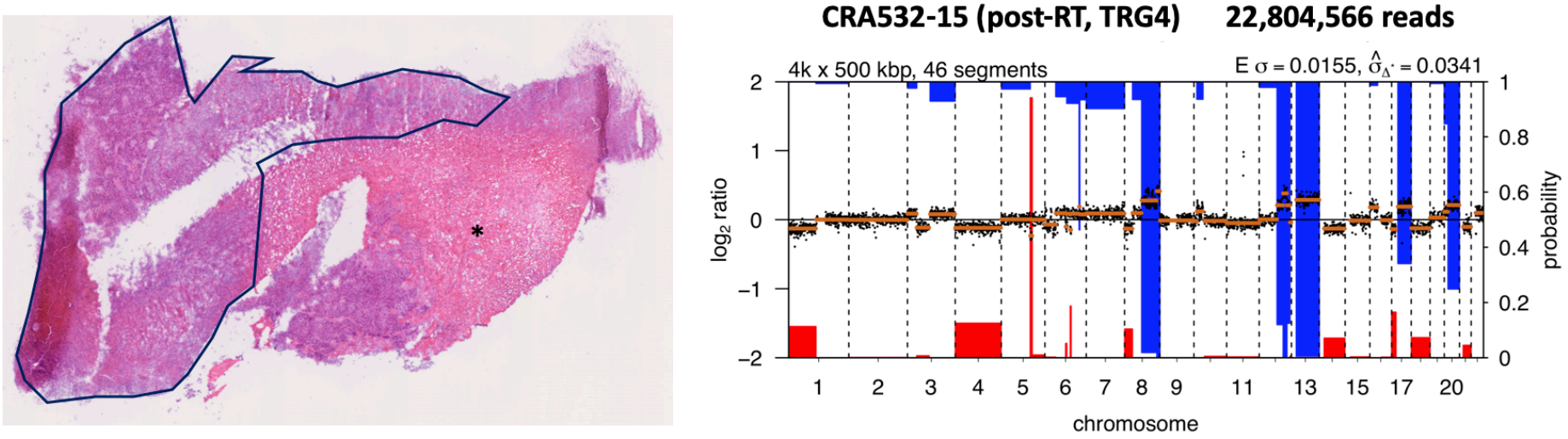


Figure 6.9 – Histology image from slide-scrape dissection and matched copy number (CN) plot from CRA532-15 (T3 N0 M0, 0/28 nodes involved, EMVI positive, no tumour perforation, neoadjuvant RT and CapOX given, TRG4) resection following RT. Outlined areas on H&E image denotes tumour which was sequenced whereas asterisk (*) denotes area of necrosis which was not included. Pattern of SCNA mirrors the pattern of CN gains in biopsy 2 with gains on Chr 6, Chr 8, Chr 12 and Chr 17. Blue bars indicate net CN gains, red bars indicate net CN losses.

By contrast, for the radioresistant RC specimen in figure 5.8 and 5.9, the SCNA profile is significant and present in both biopsy regions but with key differences – in biopsy 2, for example, there are visibly significant $\log^2$ fold changes in SCNA gains on Chromosomes 6, 8, 12 and 13 which are either not present at all or to the same degree in biopsy 1. This is despite equal cellularity (0.6) in both and equivalent numbers of reads. Following RT, the tumour histologically displays evidence of poor tumour regression (TRG 4 – fibrosis and tumour cells but with preponderance of tumour cells) but the overall SCNA profile is more akin to that found in biopsy 2. For example, net SCNA on Chromosome 6 in biopsy 1 is a small amount of loss on the long arm, whereas the residual tumour demonstrates net gains on the same arm post-RT.

6.9 Results – Analysis of the impact of divergence of genomic aberration on tumour regression grade

To account for the fact that some of the post-RT specimens demonstrated poor cellularity after RT (notably, those with TRG 1-2 and selected TRG 3 specimens), the next step was to examine the pre-RT specimens alone. **Figure 6.10** demonstrates a pairwise comparison of divergence grouped according to TRG. In this model, all combinations are accounted for across TRGs and aggregated differences in SCN losses and gains were compared. These findings are further summarised in the violin plots shown in **figure 6.11**. This shows that there is a significantly higher degree of divergence in the genome of pre-RT RC samples in those cancers that ultimately have a good response to radiotherapy. By contrast, those RCs that ultimately have a poorer response to RT have a far more homogeneous genomic landscape at the point of diagnosis.

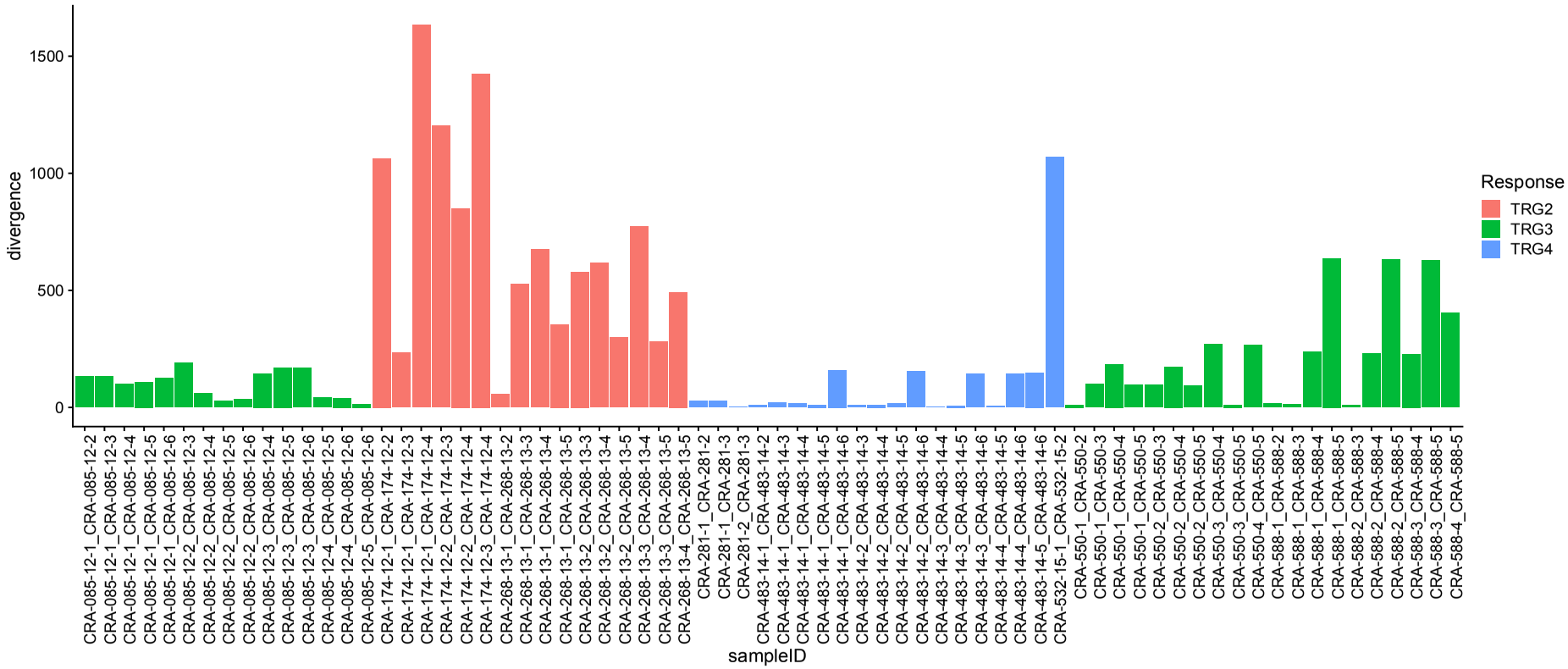


Figure 6.10 – Pairwise comparison of divergence between differing macroscopic regions of RC prior to administration of RT. None of the post-RT samples are included in this analysis and this therefore represents the 'core' genomic makeup prior to commencing RT. X-axis = sample IDs compared in each block – all combinations accounted for across all TRGs represented in the cohort. Bar height represents aggregated differences in SCN loss and gain. There is significant divergence in the genome within each multi-region set; TRG 2 appears more homogeneous versus the blander SCNA profile in TRG 3-4.

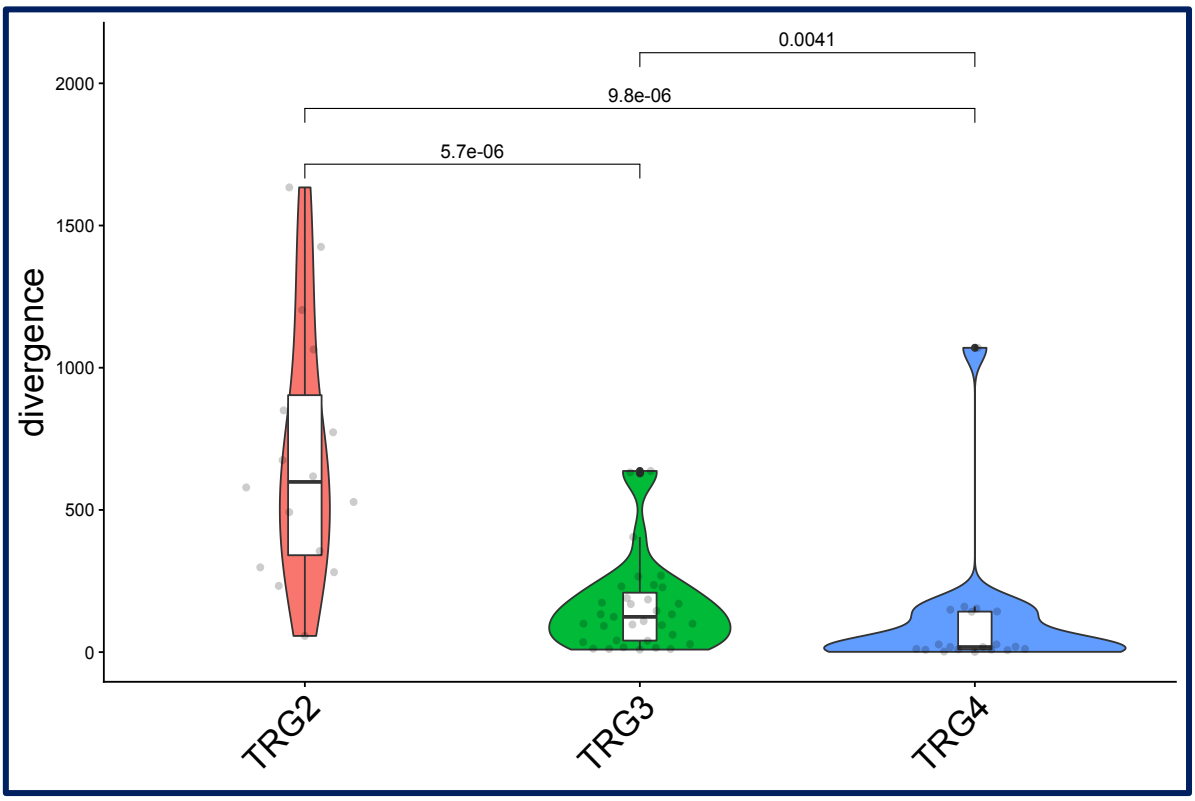


Figure 6.11 – Violin plots of divergence measured as in Figure 5.10 – again, each point represents a measure of divergence in a pairwise sample comparison between all possible combinations. At the pre-treatment stage, RCs that ultimately have a good TRG (1-2) have a more heterogeneous genomic landscape with a higher aggregate proportion of SCN losses and gains. Poor responders (TRG3-4) have a more homogenous genome with a far lower incidence of SCNAs. In this sample set, the TRG3 cohort appears to behave more in-line with those in TRG4; although with significance set at $p < 0.01$ there is still a statistically significant difference between the two groups albeit not to the order of magnitude seen in the cohort comparisons between either TRG2 and TRG3 or TRG2 and TRG4.

6.10 Discussion

At the molecular level, carcinogenesis of rectal (and colon) cancer is an incredibly complex series of events characterised by a series of genomic, DNA-methylomic and transcriptomic dysregulation. The majority of colon and rectal cancers (CC and RC respectively) have now been determined to fall under the consensus molecular subtypes (CMS) 2 and 4. CMS2, or the 'canonical' subtype is characterised by high-SCNA, *Wnt* and *myc* activation, and accounts for approximately 37% of all colorectal cancers. CMS4 cancers ('mesenchymal') account for around 23% and are also characterised by high SCNA, but additionally feature stromal infiltration, TGFβ activation, prolific angiogenesis and matrix remodelling as hallmarks. [26] [356] [357] The metabolic CMS3 subtype is the remaining subtype characterised by chromosomal instability and features low SCNA, low CpG island methylator phenotype (CpG), *KRAS-mut*, metabolic deregulation and an epithelial signature. These tumours comprise 13% of all CC and RC; however, approximately a further 13% will demonstrate mixed features. Given that up to 60% of all rectal cancers therefore will demonstrate some degree of SCNA, investigating this in the context of a wider multi-omics approach to radiation resistance mechanisms is validated. By contrast, the role for an ultra-mutated tumour genome as the prime driver in radioresistance is less likely, by virtue of the fact that only around 14% of all CC and RCs fall under the CMS1 phenotype, characterised by a hypermutated state, microsatellite instability, CIMP-high and *BRAF-mut.* Furthermore, these molecular distinctions have been shown to reflect the differing pathogenesis of right-sided (i.e., CC) vs left-sided (CC and RC) cancers, with a greater proportion of right-sided colon cancers being determined as CMS1 or CMS3. [358] [359]

The quest to find a robust genomic signature that can predict response to RT in RC has gathered pace, following the move from hybridization techniques for genomics analysis to high-throughput NextGen sequencing with the associated computing power brought by bioinformatics and machine learning. At the time of writing, a number of groups have sought to answer this question by a number of NGS methods, although the majority have been based on whole exome sequencing and a drive to identify mutational signatures that correlate with the radioresistant phenotype. [360] [361] Despite this work, the genomic alterations that occur in response to RT for RC remain poorly understood. As far as can be ascertained, this work is the first to use WGS as a modality for studying the issue in more detail.

It stands to reason therefore that for the majority of (microsatellite stable) RC, SCNAs are frequent events, with up to 30% of the genome aberrated in some cases. In many cases, these aberrations can lead to the amplification or silencing of tumour suppressor genes, oncogenes or ultimately alter the fate of gene expression by modification of non-coding RNAs. [362] In general, CN gains tend toward overexpression of genes whereas losses lead to relative under-expression, with gains on Chr7p, 8q, 13q and 20q and losses on Chr1p, 4, 5q, 8p, 14q, 15q and 18 being hallmark features of SCNA in CC and RC. [363] [364] The relationship of such SCNAs on gene expression is complex and has not yet been fully defined. Nonetheless, particular patterns of SCNA have been investigated where shown to be linked to poorer prognosis, with gains in Chr7q31 being linked to alterations in *Wnt* signalling due to overexpression of *WNT2,* and gains on Chr20q being correlated to a catalogue of genes involved in mechanisms that promote the adenoma-carcinoma sequence and invasion processes. [365-367] Furthermore, with specific reference to RC, CN gains in-line with the above patterns

have previously been demonstrated in an array-based approach to be correlated with poor response to RT. In addition to the hallmark SCNA pattern, gene expression affected by gains on Chr12 and 15 were found to be associated with poor TRG, albeit these findings have not as yet been replicated with NextGen sequencing. [368]

In this pilot study I performed whole genome sequencing on paired, multi-region RC samples pre- and post-RT. Multiple pre-radiotherapy biopsy samples from macroscopically distinct areas of the primary tumour were, wherever possible, sequenced separately to assess for the degree in genomic heterogeneity and the impact this had on the eventual response to RT as measured by clinical tumour regression grade (TRG). What has been demonstrated is that RC as a distinct entity amongst the colorectal cancer landscape is characterised by significant burden of SCNA. This is likely to correspond to a CMS2, -4 or mixed phenotype that is in-line with the molecular basis of rectal carcinogenesis versus colonic carcinogenesis (especially as compared to right-sided colon cancers). It should, however, be noted that CMS profiling would require RNASeq here to confirm these phenotypes; these data were not available at the time of writing this thesis. However, the presence of such high SCNA burden would be consistent with such an assumption.

Furthermore, there is significant heterogeneity across the macroscopic landscape of the primary RC with regards to SCNA, with some regions displaying significant gains and losses additional to the hallmark SCNA pattern described above. This is particularly evident in the pre-treatment biopsies of RC whose fate is a good RT response (TRG1-2). This would suggest that at this stage in tumour development - even when matched under standard staging and phenotyping parameters such as

TNM, EMVI and tumour budding – increased diversity in the tumour genome is predictive of a superior response to RT. The inverse is also borne out of this dataset – the degree of genomic diversity in tumours fated to have a poorer response (TRG3-5) is significantly lower than for TRG1-2, suggesting that by this point the radioresistant phenotype has already evolved. Individual SCNA plots do indeed demonstrate a more homogeneous SCNA profile across regions, and of interest are the regions that are not consistently described as hallmark features of CC and RC SCNA – notably, consistent gains on Chr6, 7, 12 and 17q with losses on Chr21q and 22q.

In Chapter 4, I demonstrated a novel analysis of the lipidomic landscape of post-RT RC which revealed a significant shift in abundance of certain classes of glycerophospholipid (GPL) in RC glands that persisted after RT. It stands to reason to theorise whether expression of any of the genes that could be implicated in the shifts in glycerophospholipid abundances that were observed in the post-RT are likely to be over- or under-expressed as a direct result of SCNA in the primary tumour prior to RT. Given that in this low-pass WGS run the genome was divided up into 500kbp bins, a preliminary analysis of SCNAs was mapped to the possible regions of interest that correspond to the loci of the gatekeeper genes that may have a role to play in the increased abundances of the GPLs seen in the DESI models. **Table 5.2** summarises some of these key genes, with specific reference to the *ELOVL* family and gatekeepers of both the CDP-DAG pathway and the interrelation between PEs and PSs. Although this is very preliminary data and warrants deeper sequencing, it is interesting to note in particular that the three *ELOVLs* that rely on essential FAs (*ELOVL2*, *ELOVL4*, *ELOVL6*) are coded by regions that have significant SCN gains in biopsies from tumours that ultimately demonstrated radioresistance. Although this data does not

account for any post-translational modification that may occur with the gene transcripts, it is feasible to postulate a link between SCNA and effects on – at least - the coding regulation of these genes. Furthermore, aberrations are noted in the chromosomal regions coding for the PE/PS interplay in the Kennedy pathway.

Gene symbol	Description	Encoded protein functions	Locus	SCNA pattern on locus in RT+ RC (poor responders >TRG4)
ELOVL1	Fatty Acid Elongase 1	Acts on saturated and monounsaturated acyl-CoA substrates, with the highest activity towards C22:0 acyl-CoA	**1p34.2**	Variable SCN losses on 1p
ELOVL2	Fatty Acid Elongase 2	Catalyses the synthesis of polyunsaturated very long chain fatty acid (C20- and C22-PUFA), acting specifically toward polyunsaturated acyl-CoA with the higher activity toward C20:4 (*n-6*) acyl-CoA	**6p24.2**	SCN gains on Chr6 versus minimal SNCA in RT- / TRG1-3
ELOVL3	Fatty Acid Elongase 3	Condensing enzyme that exhibits activity toward saturated and unsaturated acyl-CoA substrates with higher activity toward C18 acyl-CoAs, especially C18:0 acyl-CoAs	**10q24.32**	No consistent SCNA pattern on Chr10
ELOVL4	Fatty Acid Elongase 4	Catalyses the synthesis of very long chain saturated (VLC-SFA) and polyunsaturated (PUFA) fatty acids	**6q14.1**	SCN gains on Chr6 versus minimal SNCA in RT- / TRG1-3
ELOVL5	Fatty Acid Elongase 5	Important substrates are both omega-3 (ω-3) and ω-6 FA; both through catalysing addition of double bonds to ω-3 or ω-6 or elongation of carbon chains	**6p12.1**	SCN gains on Chr6 versus minimal SNCA in RT- / TRG1-3
ELOVL6	Fatty Acid Elongase 6	Condensing enzyme that elongates unsaturated FAs with 12, 14 and 16 carbons with higher activity toward C16:0 acyl-CoAs.	**4q25**	Significant SCN losses on Chr4
PSS1	Phosphatidylserine synthetase 1	Catalyses the formation of phosphatidylserine from either phosphatidylcholine or phosphatidylethanolamine	**8q22.1**	Significant gains on Chr8, especially 8q. Less common in TRG1-2, variable in TRG3
PSS2	Phosphatidylserine synthetase 2	As for *PSS1*	**11p15.5**	Variable, high-intensity SCN losses and gains on Chr11
PLPP1 (PAP1)	Phosphatidic acid phosphatase 1	Converts phosphatidic acid to diacylglycerol, and function in synthesis of glycerolipids and in phospholipase D-mediated signal transduction	**5q11.2**	Variable SCN losses on 5q – heterogeneity across multi-region biopsies in TRG3-5
PLPP2 (PAP2)	Phosphatidic acid phosphatase 2	As for *PLPP1*	**19p13.3**	SCN losses on Chr19
GPAT1	Glycerol-3-Phosphate Acyltransferase 1 (Mitochondrial)	Esterifies acyl-group from acyl-ACP to the sn-1 position of glycerol-3-phosphate	**10q25.2**	No consistent SCNA pattern on Chr10
GPAT2	Glycerol-3-Phosphate Acyltransferase 1 (Mitochondrial)	As for *GPAT1*	**2q11.2**	No consistent SCNA pattern on Chr2p
SELENOI (EPT1)	Ethanolamine phosphotransferase 1	Catalyzes the transfer of phosphoethanolamine from CDP-ethanolamine to diacylglycerol to produce phosphatidylethanolamine	**2p23.3**	No consistent SCNA pattern on Chr2p
GNPAT	Glyceronephosphate O-Acyltransferase	Encodes an enzyme located in the peroxisomal membrane which is essential to the synthesis of ether phospholipids	**1q42.2**	SNCA gains on 1q
AGPS	Alkylglycerone Phosphate Synthase	Encodes a protein that catalyzes the second peroxisomal step of ether lipid biosynthesis in which DHAP is converted to alkyl-DHAP	**2q31.2**	Some regions with significant SCN loss on 2q
DGAT1	Diacylglycerol O-Acyltransferase 1	Catalyzes the terminal and only committed step in triacylglycerol synthesis by using diacylglycerol and fatty acyl CoA as substrates. In ether lipid synthesis, uses 1-monoalkylglycerol (1-MAkG) as an acyl acceptor for the synthesis of monoalkyl-monoacylglycerol (MAMAG)	**8q24.3**	Significant gains on Chr8, especially 8q. Less common in TRG1-2, variable in TRG3

Table 6.2 – Selected genes coding for enzymes in the FA elongation, PE/PS interplay within the Kennedy pathway and peroxisomal catalysis of ether lipids cascades.

A further finding of interest is the behaviour of the TRG3 samples within this cohort. In line with what was found on pilot, three-cohort clustering analysis with DESI MSI based on lipidomic diversity, TRG3 emerges as a highly heterogenous cohort which does not appear to consistently demonstrate a phenotype of its own. Some tumours in this cohort have SCNA profiles and divergence more in-line with those tumours that were fated to be TRG1-2, whilst others were more in-line with the poor responders. This is shown in the pairwise comparisons in **figure 6.10** where two clusters emerge, not dissimilar in behaviour to the lipidomic data borne out in Chapter 4. This raises the important question on how to approach these tumours in the context of their underlying biology. After all, TRG is a clinical assessment of tumour regression and on a 3- or 5-point system there will always be a mid-point. Biologically, it would stand to reason that this would never hold entirely true as a true 50:50 response is unlikely and these tumours (which tended to comprise the majority) have demonstrated a unique 'in-group' diversity towards one or other end of the regression spectrum.

This experimental work is subject to a number of limitations. Following sequencing, the majority of the post-RT tumour samples which demonstrated a good response revealed low cellularity in the samples, making meaningful SCNA calls unreliable. This is likely to be due to either a low proportion of tumour or dilution with desmoplastic stroma in the tumour microenvironment, which by definition is diploid and therefore masks the sequencing calls from tumour DNA. [362] Equally, the high stromal component of these tumours may implicate them as being part of the CMS4 phenotype and as such this may be a truly representative outcome, but this would require a combination of laser-microdissection of the residual glands in the specimens

to obtain single-cellular DNA for sequencing and also a consideration of sequencing of the stromal compartment. What holds true, however, is the data obtained in the nested divergence analysis of pre-RT samples and the fact that heterogeneity appears to decrease as one progresses through TRG2-4. Again, this remains a pilot finding but is interesting insofar as it appears almost paradoxical in the clinical context; inherently, one would imagine a tumour with a greater degree of heterogeneity to be more aggressive. In this case, what appears to be the case is one of non-dominance of the radioresistant clone at the diagnostic stage. If the end-point of therapy is to obtain a *pathological complete response* (pCR)*,* clearly the persistence of any tumour obviates this, and the question remains which of the tumour islands is responsible for the radioresistant clones post-RT. However, based on the question of *degree* of response and clinical down-staging, the presence of heterogeneity to a degree that radiosensitive clones must still be present is a positive finding for the patient at the pre-treatment stage with regard to ensuring a clear resection margin at TME surgery.

Although the number of patients is small in this cohort, additional weight is gained in this study by accounting for discrete macroscopic regions of the tumour in the pre-treatment biopsies. With regards to assessing as to whether the argument of divergence in the pre-RT genome being correlated with TRG holds true, a test set would be indicated to run against this cohort as a training set. This would undoubtedly increase the strength and validity of these findings. Deeper sequencing would also yield further insights into mutational burden, although as discussed it may yield less insight than is hoped for on the basis of previous studies. My collaborators at the Barts Cancer Institute are currently in the process of designing a bioinformatics-based approach that can mine low-pass WGS for mutational signatures. If validated, this

would be a more feasible next step before committing to deeper (and inherently more expensive) sequencing, particularly considering the depth of coverage achieved in this study.

6.11 Conclusions

In this chapter I have presented a novel, pilot Whole Genome Sequencing (WGS)-based analysis of pre- and post-radiotherapy (RT) rectal cancer tissue, for assessment of the burden and impact of somatic copy number aberrations (SCNAs) on the radiotherapy response. My results have revealed that the burden of SCNAs across this cohort of rectal cancers was generally high, and in the region expected of tumours conforming to the CMS2 and CMS4 phenotypes. In-tumour sampling of macroscopically distinct regions at diagnostic biopsy prior to RT demonstrates significant heterogeneity across the tumour in most cases. This is especially evident in RC whose fate is a poor response to radiotherapy. The divergence in the genome is further correlated to eventual TRG, with good responders having far greater divergence and heterogeneity within samples compared to poor responders. This raises the possibility of a SCNA-mediated radioresistant phenotype being pre-programmed even at the time of diagnosis, although the exact downstream effects on the genome and functional coding remains to be elucidated.

Chapter 7
Completing a holistic multi-omics workflow for discovery science in rectal cancer radioresistance

7.1 **Overview of the multi-omics platform** **213**

7.2 **Multiplexed, quantitative, imaging mass cytometry with Hyperion™** **218**

7.3 **Evaluation of the on-tumour microbiome in radioresistant rectal cancer** **223**

7.4 **Summary** **225**

7.1 Overview of the multi-omics platform

In the introductory chapters and the in-silico analysis, I made the case for a new approach to the issue of establishing new insights into the mechanism of radioresistance in RC, for which classical array-based and proteomics studies had not been able to fully address due to inherent biological complexity of introducing a cytotoxic therapy into the arena. Accounting for both the rapid advances in multiple -omics technologies and the challenges faced by previous efforts, my work has set out to establish a novel outlook on the landscape of radioresistance in rectal cancer. Whereas the bulk of the experimental data (other than the *in silico* work) has focused on DESI MSI for determination of the lipidomics of radioresistance and on whole-genome sequencing (WGS) for assessment of the impact of somatic copy number aberrations (SCNAs) and how these might impact the regulation and expression of glycerophospholipid (GPL) pathway-coding genes, this has all been part of a wider portfolio of work that I have set up and undertaken whilst in the Department of Surgery and Cancer at Imperial College London. Whereas the workstreams that follow are entirely complementary to the work presented in this thesis and part of my ongoing research, full exploration and analysis of the data from these is beyond the scope of this thesis and therefore remains a fundamental part of the next phases of this wider research theme.

A holistic approach to pathway elucidation and biological discovery science in cancer was postulated in Chapter 4 and the figure is repeated here in **figure 7.1**. This demonstrates the multiple layers of a multi-omics approach from the bottom up, i.e., from metabolites back through to the core genome. Although not implicated in a linear fashion, the important addition to this is the influence of the gut microbiome, especially

in gut pathology as with RC in this case. It is now accepted that there is a clear link between a 'high risk' metagenome and the development of CC and RC, and therefore it stands to reason that the actions of bacterial metabolic processes both in the gut lumen and on the surface of the tumour may contribute to radioresistance mechanisms through their action on the tumour microenvironment. In the subsequent **figure 7.2** and **figure 7.3** I have summarised the workflow for maximising the output from a 3-dimensional multi-omics model of RC, annotating the various experimental methods. The fact that the outputs from these modalities can be directly correlated back to histology-driven tumour features is a particular strength and takes into account the significant heterogeneity that has been demonstrated in the work so far at both genomic and metabolic levels.

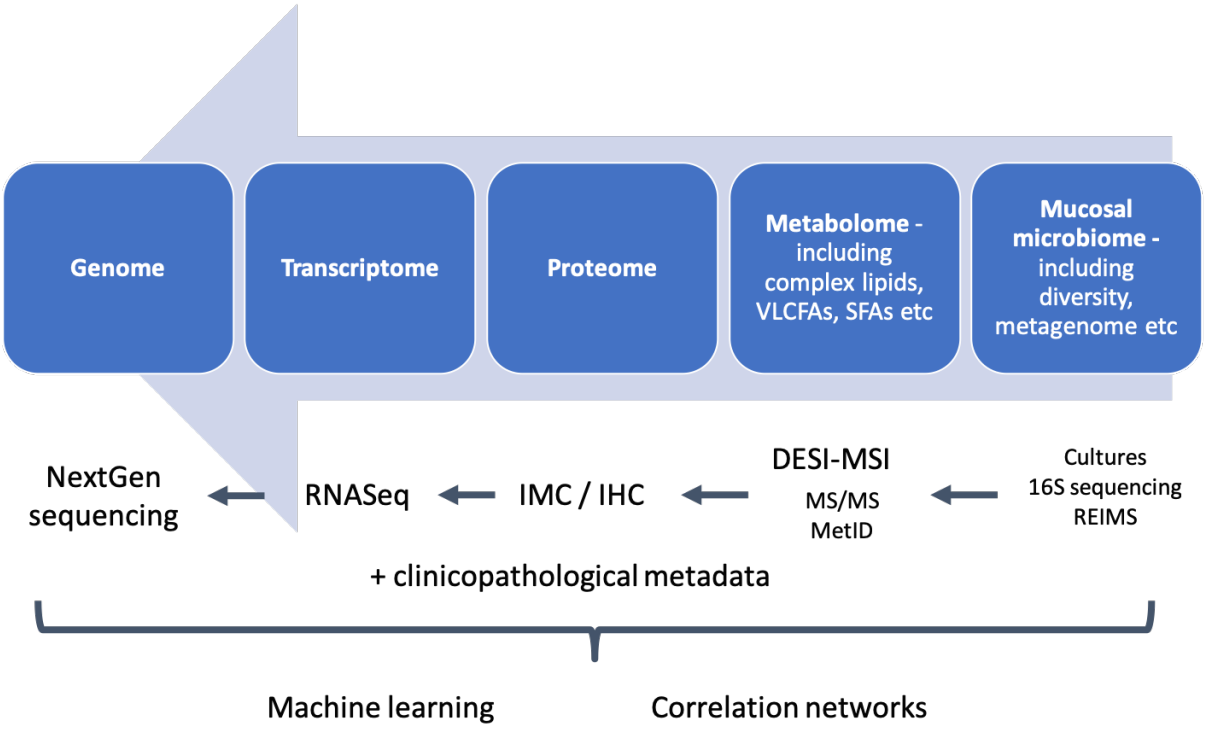


Figure 7.1 - The proposed workflow for a holistic multi-omics approach to the unrealised understanding of the radiation response in rectal cancer. Potentially the most important step is the bioinformatics strategy to resolve integration of these varying datasets.

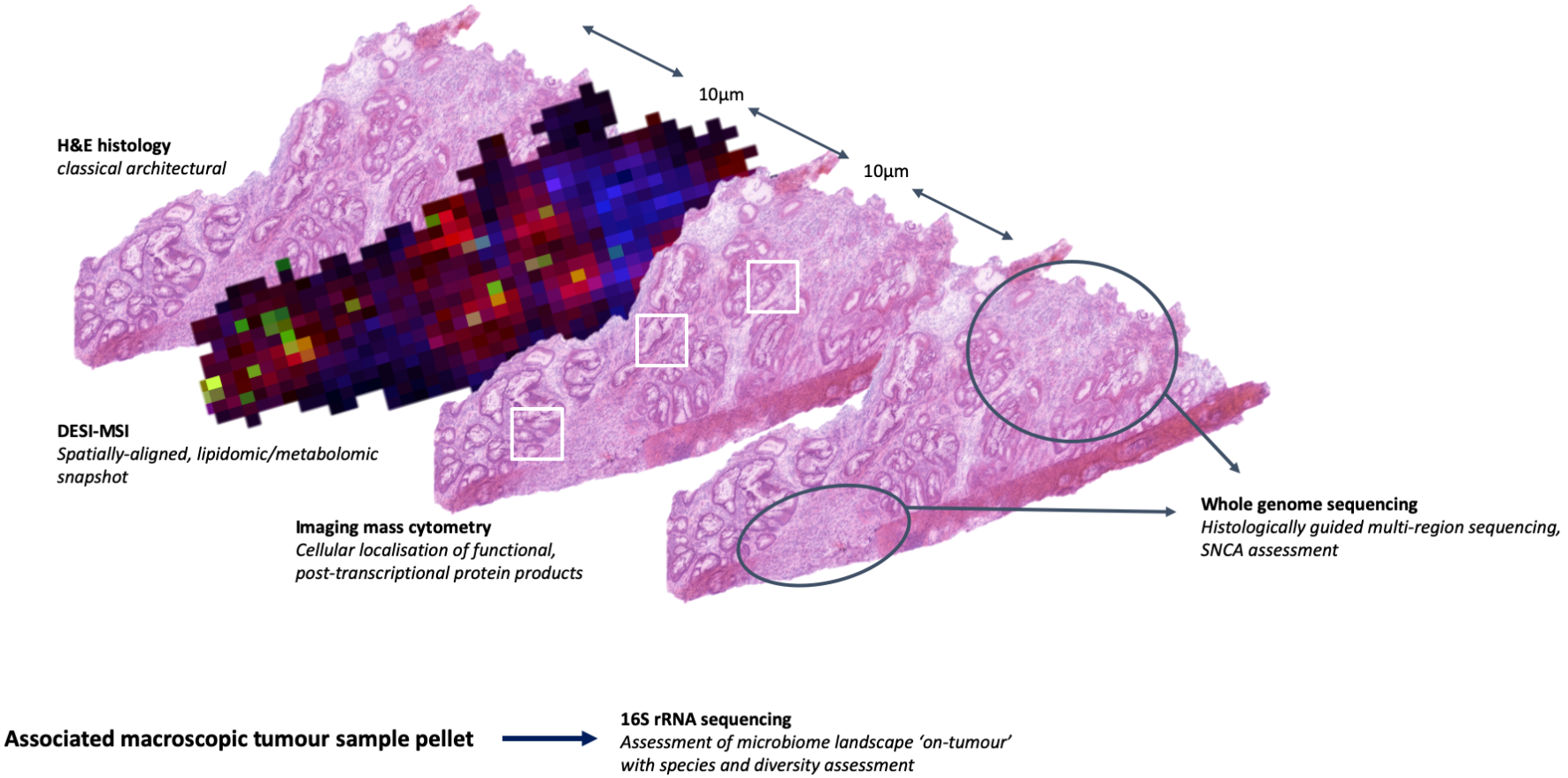


Figure 7.2 – Detailed pipeline of how I have applied the workflow proposed in figure 6.1 to my work. The resultant sequencing and IMC studies can be either histology-driven or metabolite-driven, based on ion imaging acquired through DESI MSI

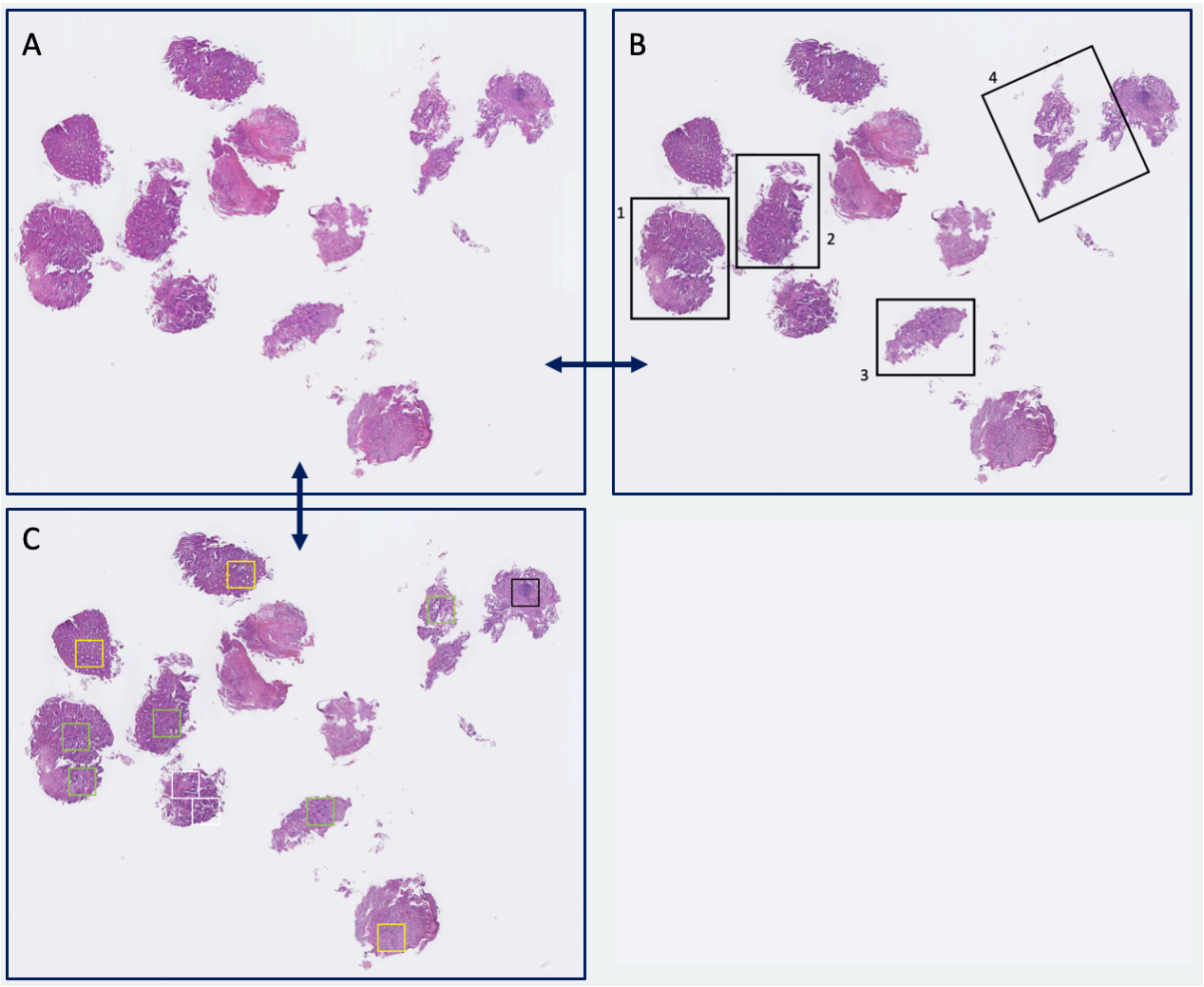


Figure 7.3 – Complementary assessment of archival FFPE pre-RT rectal cancer biopsies. Arrows denote 5µm thickness/distance between samples. (A) H&E histology assessment guides (B) regions bearing tumour for separate DNA extraction and sequencing as discussed in Chapter 5. The same approach can be applied for RNA extractions for RNASeq from a further immediately adjacent section. (C) denotes selected regions of interest for Hyperion™ CyTOF® further selected and annotated to allow for heterogeneity of sample – coloured boxes select normal rectal mucosa (yellow), rectal cancer (green), high grade dysplasia (white) and lymphoid follicle (black)

As demonstrated, the workflow has incorporated the matched pre- and post-RT tissue in addition to the associated full clinicopathological metadata which has been previously shown. There are some key differences in the two workflows which are guided by the validity of the data outputs from the various -omics platforms. Although sectioning in pure water ice is considerably more difficult than in Optimal Cutting

Temperature compound (OCT) or paraffin blocks, it still allows for standard H&E staining. However, what it subsequently permits is DESI MSI of the post-RT tissues without creating any interference, masking or reaction with possible biologically representative compounds. Whereas the use of DESI MSI on FFPE tissue for peptide and metabolite detection has been described, the concern with using FFPE sections for lipidomics work is that the deparaffinisation step with xylenes potentially elutes or destroys significant quantity of complex lipids as to render the landscape unrepresentative of the tumour biology. [369 370] It is for this reason, and the lack of fresh frozen tissue from pre-RT rectal cancer in this cohort, that DESI MSI was not employed on the pre-treatment cohort. On the other hand, use of fresh frozen tissue for DESI MSI provides, in the authors opinion, the most biologically relevant snapshot of lipidomic data in a histology-aligned manner, owing to the lack of solvent or embedding material interference and the lack of a matrix that would otherwise be required for MALDI. [371 372]

Use of FFPE tissues has become routine for work involving extraction of genomic DNA and RNA, including in the NextGen Sequencing (NGS) era. In reality, in order for NGS to evolve as a translational tool in precision medicine, it is essential for the technology to be able to work with archival tissues which remain the most abundantly available human tissues associated with a disease process and with accompanying metadata. [373 374] Fortunately, it is now broadly accepted that the fixation and embedding processes do not affect the ability of NGS platforms to produce meaningful sequencing data from FFPE-derived DNA and RNA, and a major study has demonstrated high correlation of expression profiles between FFPE and fresh frozen tissues. [375]

Furthermore, most antibody conjugation work (notably IHC) continues to be conducted on FFPE tissues and these are the conditions that most commercially available antibodies are optimised for. However, as discussed in Chapter 3, one of the major limitations about studies utilising IHC as an investigative modality in discovery science is that, even in the hands of an expert histopathologist, the interpretation remains subjective. [376] In the clinical arena, IHC is most useful for evaluating an 'all or nothing scenario'; for example, looking for presence of a KRAS- or BRAF-wild type versus mutant. In order for the approach to have value in discovery science, a quantitative (and, preferably, multiplexed) platform would be preferable. This is exactly what is provided in a new mass spectrometry-based imaging platform which combines cytometry with time-of-flight mass spectrometry (CyTOF®) with the Hyperion™ imaging system from Fluidigm® (Fluidigm, CA, USA) which forms a complimentary layer to the 3D multi-omics model comprising H&E, DESI MSI, CyTOF® targeted WGS (and ultimately RNASeq). The final layer is not derived from tissue microsection but is instead obtained by harvesting a section from the solid tissue pellet – this provides adequate number of microbial species for digestion and metagenomic sequencing for an appreciation of the on-tumour microbiome. A brief overview and the progress with these approaches are elaborated on further in the following sections.

7.2 Multiplexed, quantitative, imaging mass cytometry with Hyperion™

As postulated above, the previous critical unmet need in been to understand complex single-cell phenotypes within the broader landscape of the tumour microenvironment. The Hyperion™ Imaging System is the first and only commercially available *Imaging Mass Cytometry* platform that permits multiplexed antibody-directed

imaging with (in principle) between 4 to 37 protein markers in fixed tissue sections simultaneously. Briefly, antibodies of interest are conjugated to heavy rare earth metals (lanthanide series) and it is these ions that are detected using the time-of-flight (TOF) mass cytometry (hence, CyTOF®). [377] The slide of interest first needs an index H&E and regions of interest defined. The ablation chamber takes a high-resolution digital photograph of the slide before a laser beam is focused onto the first 'pixel' measuring 1µm x 1µm and sample proteins stained with the metal-tagged antibodies are ablated and the resultant aerosol carried into the ICP torch where they are vaporized and ionised in plasma before filtration of low-mass ions and passage into the TOF analyser (**figure 7.4**). [378]

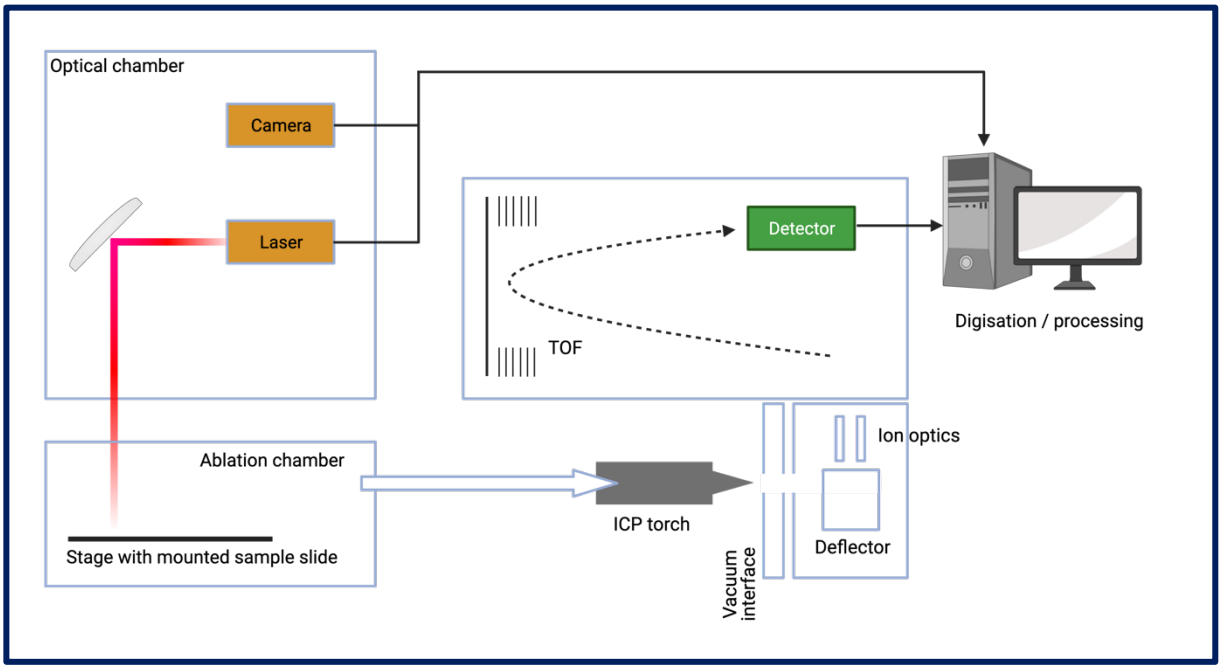


Figure 7.4 – Schematic outline of the Hyperion™ IMC system – adapted from the materials provided by Fluidigm®

To date, I have coordinated an effort within my collaboration to optimise the platform for use with the thicker, fresh-frozen sections that were used for DESI MSI in order to provide complementary, metabolic images that can be co-localised to the CyTOF® IMC data. However, due the existing optimization with FFPE we have already studied the pre-treatment RC biopsies in great detail, using an in-house, pre-defined multiplexed panel of 23 validated antibodies as outlined in **table 7.1**. This panel aims to provide deep insights into the various cellular compartments and features of the tumour microenvironment. Following a trial panel ablation on normal rectal mucosa (the first use of the instrument on human colon or rectal tissues) as shown in **figure 7.5**, we have now performed histology-guided ablations on over 160 regions of interest across multiple discrete regions of the biopsy samples as shown in figure 6.3. In addition to the valuable insights and complementarity of data that this will yield in the question of response to radiotherapy in RC, many of the samples analysed include an adenoma-adenocarcinoma threshold within the same samples, yielding us with a unique opportunity to study the *within-patient* adenoma-adenocarcinoma sequence in a completely novel way. This work remains in-progress and the equally important work towards developing a bioinformatics stream to combine the two datasets in a biologically meaningful way is also ongoing.

Compartment	Target	Clone	Tag	Concentration	Lane
Structural	**panCK**	C11	148-Nd	1:100	1
Structural	**E-Cad**	24E10	158-Gd	1:50	2
Structural	**αSMA**	1A4	141-Pr	1:200	3
Structural	**Vim**	D21H3	143-Nd	1:100	4
Immune	**CD3**	poly (C-terminal)	170-Er	1:100	5
Immune	**CD20**	H1	161-Dy	1:400	6
Immune	**CD45**	D9M8I	152-Sm	1:100	7
Immune	**CD68**	KP1	159-Tb	1:50	8
Mitosis / Apoptosis	**Ki67**	B56	168-Er	1:50	9
Mitosis / Apoptosis	**Cleaved Caspase-3 [Asp175]**	5A1E	172-Yb	1:50	10
Structural	**CD31 (PECAM-1)**	EPR3094	151-Eu	1:100	11
Immune	**CD8a**	D8A8Y	162-Dy	1:50	12
Immune	**FoxP3**	236A/E7	155-Gd	1:50	13
Signalling	**VEGF**	G153-694	163-Dy	1:100	15
Signalling	**pERK1/2**	D13.14.4E	171-Yb	1:50	16
Signalling	**pTyr-100**	P-Tyr-100	144-Nd	1:50	17
Signalling	**pS6 [S235/236]**	N7-548	175-Lu	1:50	18
Immune	**CD4**	EPR6855	156-Gd	1:200	19
Immune	**CD16**	EPR16784	146-Nd	1:50	20
Immune	**granzyme B**	EPR20129	167-Er	1:100	21
Immune	**CD11b**	EPR1344	149-Sm	1:100	22
Structural	**Col I**	poly	169-Tm	1:300	23
Nuclear	**HH3**	D1H2	176-Yb	1:300	24

Table 7.1 – Conjugated antibodies utilized in the multiplexed 23 antibody panel on pre-RT RC biopsies on the Hyperion™ IMC platform. Concentrations separately validated and all antibodies validated prior to use with Western blot data.

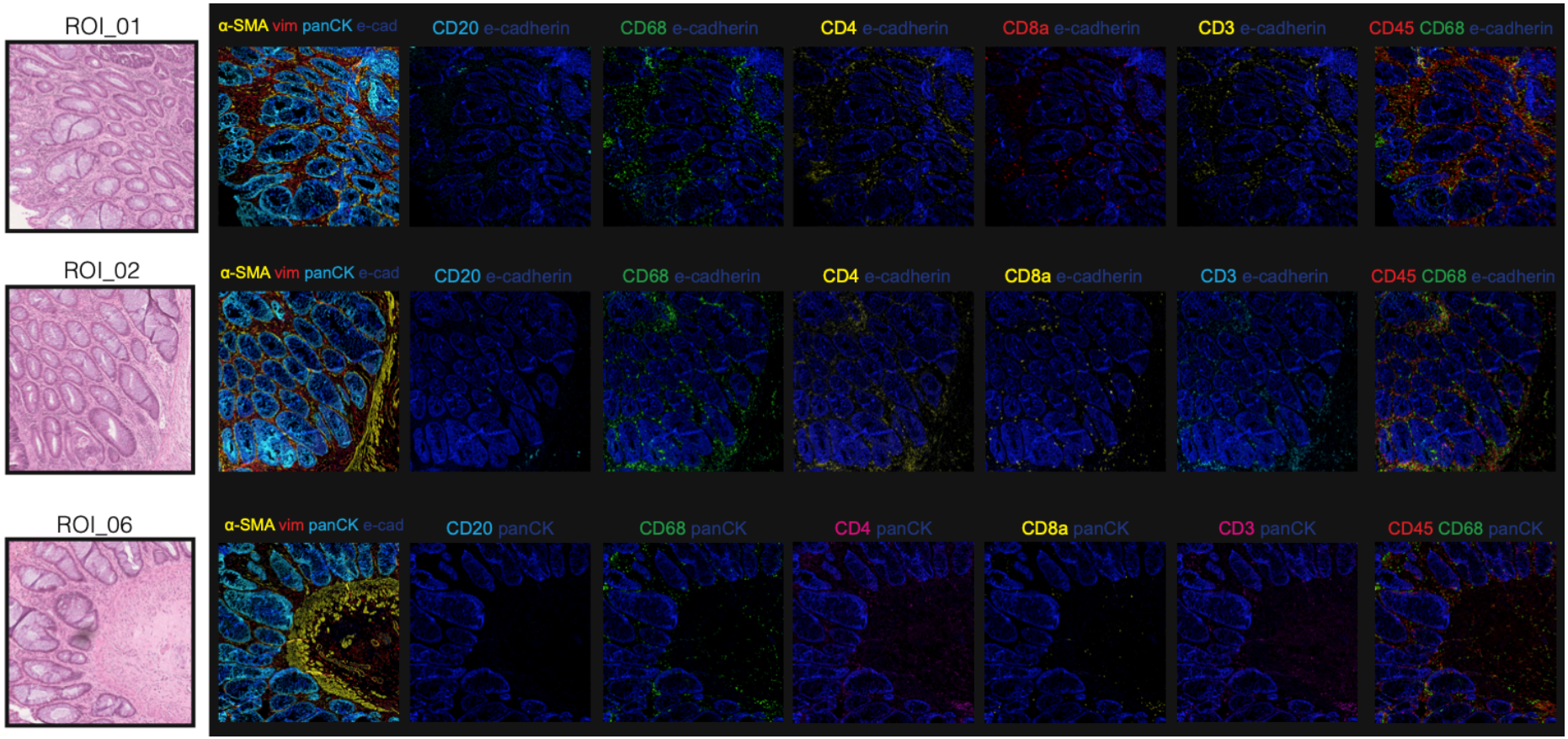


Figure 7.5 – Example IMC images from selected regions of interest of normal rectal mucosa as part of optimisation steps for the multiplexed antibody panel. Selected antibody targets shown which demonstrate differential expression profiles according to the tissue compartment in which they are located, for example a-SMA being exclusively located in the basement membrane and submucosa.

7.3 Evaluation of the on-tumour microbiome in radioresistant rectal cancer

As discussed in the introduction and figure 7.1 above, it stands to reason that the individual biology of RCs will be significantly moulded by the activity of the on-tumour microbiome. In tandem with the work presented in this thesis, I have set up and co-ordinated supervision of an MRes project to assess the impact of the microbiome and metagenome in rectal cancer radiotherapy response. As previously outlined, numerous human studies have found associations between RC and the gut microbiome. The underlying mechanisms suggested include (but are not limited to) influence on inflammatory processes and genomic stability, and modulation to host gene expression. [379] It has been further postulated that, in the development of RC, microbial-induced pro-inflammatory changes could be sufficient to constitute the second 'hit' in the Knudson's two-hit hypothesis. [380] Given these possible mechanisms and the effects that they exert on the tumour microenvironment, it is plausible to consider a role for variation in the gut microbiome in response mechanisms of the primary tumour to RT. A previous nested analysis of a small cohort from the work by Alexander, Poynter et al demonstrated some preliminary data suggesting persistence of certain pathobionts on radioresistant rectal cancer, with particular enrichment of *Bacteroides fragilis* and *Parvimonas micra* on TRG4-5 patients versus TRG1-2 (conference data). [381] [382] Preliminary analysis from my cohort has demonstrated similarity in the data in our pilot work; clearly, this is subject to deeper interrogation and integration in the broader context of the other multi-omics datasets (**figure 7.6**)

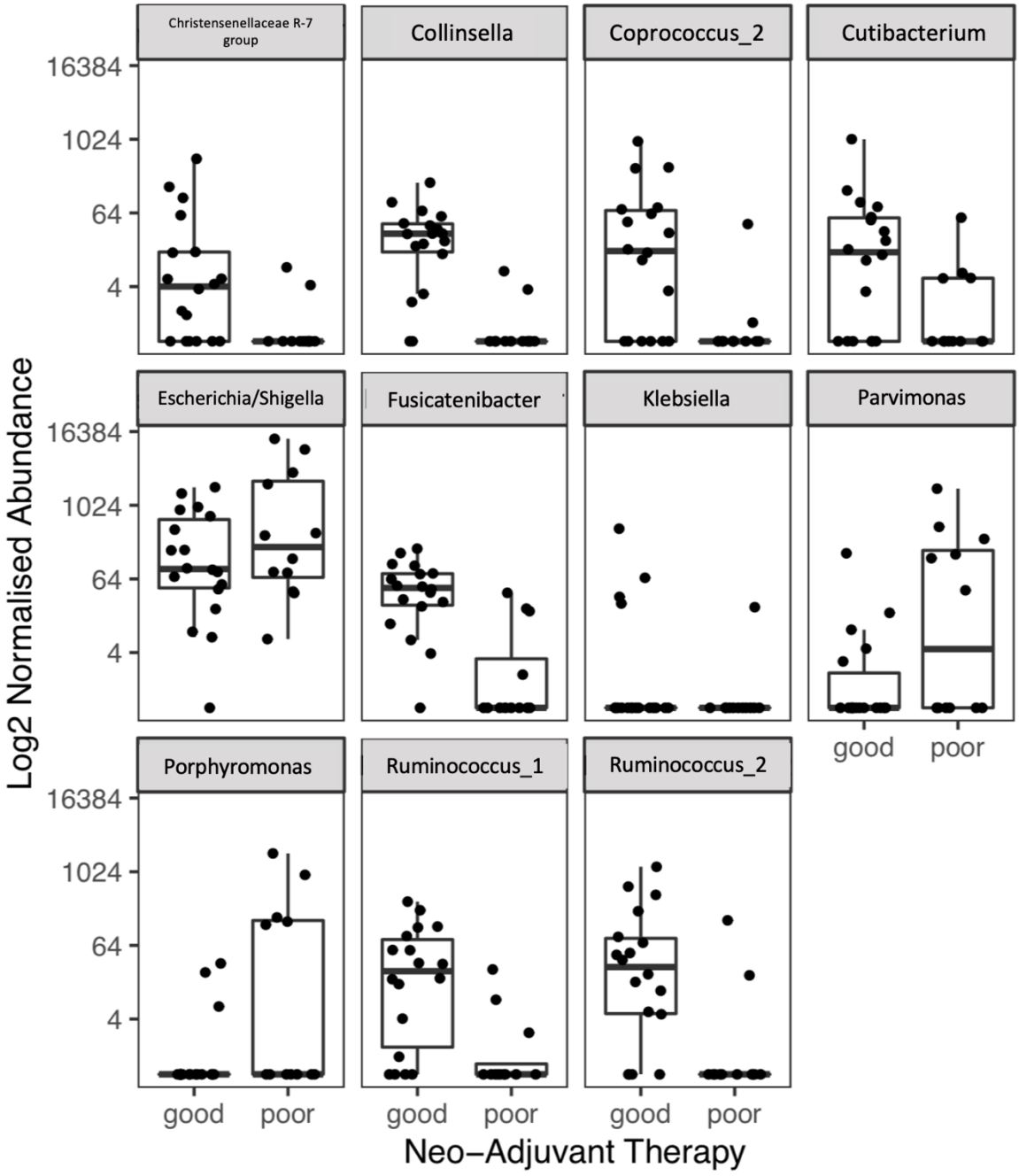


Figure 7.6 – Log-transformed box and whiskers abundance plots of taxa identified in 16S rRNA sequencing *on tumour* responsible for clustering in Principal Co-ordinates Analysis (PCoA). In-line with previous pilot work by the group, there is relative overexpression/survival of oral pathobionts on-tumour in radioresistance. Whether this demonstrates true symbiosis or merely survival of an already-established highly pathogenic colony on-tumour remains to be established.

7.4 Summary

The problem of radioresistance in rectal cancer is one that has not, to date, been tackled. There remain considerable gaps in our knowledge despite the wealth of experimental work and published data that have attempted to model a pathway by which individual RCs exert radioresistant effects. The answer is unlikely to be straightforward nor linear and therefore a new approach in the multi-omics era has been put forward in this work. This is especially important as clinical management approaches are evolving, potentially to include primary RT and watch-and-wait strategies, whereby an understanding of the biology and potential biomarkers of response and remission will be essential in their safe application.

In addition to the complementary findings from *in silico*, spatially-aligned lipidomic and whole genomic datasets, in this chapter I have presented a summary of my ongoing workstreams that complement the results I have demonstrated in my data chapters. The results of the data analysis from the imaging mass cytometry (IMC) are particularly eagerly awaited; this approach has the potential to streamline the efficiency of protein and enzyme analysis in a biologically representative human cancer system that is not possible with standard immunohistochemistry (IHC). Moreover, the combination of precise localisation of product expression and quantitative output means that the potential applications of this approach in pathway elucidation and discovery science are near-limitless. The integration of these heterogeneous datasets into a biologically plausible representation remains the major hurdle to overcome in the bioinformatics sphere and remains a work in-progress.

Chapter 8
Summary and Conclusions

8.1 THE RATIONALE FOR THIS RESEARCH 227

8.2 SUMMARY OF FINDINGS 229

8.2.1 *In silico* analysis of the response to radiotherapy in rectal cancer – mining the current knowledge landscape 229

8.2.2 Mapping cellular lipid modulation in response to radiotherapy with DESI MSI 231

8.2.3 Assessment of the whole tumour genome pre- and post-radiotherapy 232

8.3 PROMISING AREAS FOR FUTURE WORK 234

8.1 The rationale for this research

The background to this work stems from the distinct therapeutic challenges posed by rectal cancer as a distinct entity within the landscape of colorectal cancer (CRC). As originally described in the landmark work by Heald, critical to the aim of complete tumour clearance and reduction in local disease recurrence is the complete removal of both the primary rectal tumour and its associated mesentery and draining nodes in an intact fascial envelope (TME). [69] For locally advanced rectal cancers, this can be challenging or near impossible to achieve without effective downstaging radiotherapy, particularly when factoring in the additional anatomical challenges of a narrow android pelvis. For those patients who are demonstrated to have a threatened circumferential resection margin on pre-operative MRI staging, delivery of external beam, neoadjuvant radiotherapy in some form is almost universally indicated. Given that 30% of all colorectal cancers arise in the rectum, this is an ongoing clinical issue of great importance. [383]

In the UK, the decision to offer radiotherapy remains one guided by multi-disciplinary input and ultimately radiological assessment of involvement of the TME plane by the advancing front of the tumour and presence of mesorectal lymph nodes, but not (currently) extramural venous invasion. Tumour regression is a both a radiological clinical assessment following therapy but also determined at final pathological assessment (ypT status). The relevance and impact of both TRG and pathological complete response (pCR) has been variously assessed, with recent studies continuing to demonstrate that pCR is an independent predictor of superior disease-free survival, with mixed results regarding ‘good’ versus ‘bad’ TRG on disease recurrence. [384-386] Unfortunately, it currently remains only a minority of patients who

actually experience a pCR (between 10% and 31%, dependent on varying studies), and therefore the importance of understanding the mechanisms of radioresistance such that they might be made the subject of targeted therapies or new radiosensitisers is paramount. [90-95]

Although as surgical oncologists we are always striving to find the 'big answers' to cancer questions, it would be foolhardy to think the scope of this thesis would be to discover a clinically translatable biomarker where the efforts of so many other pre-eminent individuals and institutions have thus far failed to do. Therefore, throughout my research journey I have taken this work to very much be a way of introducing disruptive methodologies to the more well-trodden approaches in cancer discovery science, in the hope that it would yield new insights and create viable new directions of enquiry for future work. In the introductory chapter of this thesis, I argued the rationale for a more holistic approach to the clinical and biological problem posed by response of the primary RC to such radiotherapy. I hypothesised that a bottom-up approach, from the assessment of the metabolic footprint left behind after radiotherapy in radioresistant RC glands to the core genome of the tumour prior to RT, and everything in-between, was required for a more complete picture of this inherently complex oncological dilemma. I subsequently proceeded to demonstrate a number of key findings from my various workflows, which are summarised as follows.

8.2 Summary of findings

8.2.1 *In silico* analysis of the response to radiotherapy in rectal cancer – mining the current knowledge landscape

The objectives of this section of my thesis in Chapter 4 were (1) to undertake an extensive systematic review of published, peer-reviewed, original experimental work on biomarkers predictive of radioresistance and/or radiosensitivity; (2) to meta-analyse this data using an original, holistic approach by way of classifying biomarkers using Gene Ontology biological functions; (3) to map this data in such a way as to demonstrate significance of various biological functions in the context of radiotherapy response, and (4) to undertake a Gene Set Enrichment Analysis (GSEA) on a publicly available dataset of rectal cancer samples, and specifically to map the expression data to KEGG pathways implicated in lipid metabolism. My analysis demonstrated novel insights into existing published data by attempting to homogenise experimental data in a way that could allow for hypothesis generation. By utilising a publicly available, controlled language categorisation (Gene Ontology terms), and accounting for all statistical data reported in the reviewed studies, aggregated significance of all investigated genes and protein products becomes possible. By clustering these according to their biological function, what has emerged are the areas of the biological process that should direct future radioresistance mechanism studies. Not only was this work positively received by both the *American Association for Cancer Research (AACR) 2017 Annual Scientific Congress* and through peer-reviewed publication in *Clinical Colorectal Cancer*, but it is also entirely in-keeping with the approach that should be followed when various avenues of research into a clinical or biological

problem have failed to yield translational results. With the perennial increases in global investment in biomedical research (a total of US$240 billion in 2010), it is increasingly important that time, resources and capital are not wasted through failure to systematically appraise and critique the existing evidence in a meaningful way. [387] [388] The approach taken in this methodology does just this. The results of this work were that the biological processes 'response to stimuli', 'cellular metabolism' and 'cell communication' are area which have shown promise and have reached statistical significance in individual experimental studies, whereas elements of the cell cycle and carcinogenesis pathways are much less likely to be representative of mechanisms governing response to RT. Within these domains, the pathways that are intrinsically regulated by or influenced by complex cellular lipids were of particular significance, and yet with a relative paucity of experimental data outside of in vitro studies.

An added layer of this work, therefore, was to take a legacy microarray dataset with a representative patient cohort and to perform a Gene Set Enrichment Analysis of the expression data, mapped to the KEGG glycerophospholipid (GPL) metabolism pathway. This demonstrated significant enrichment and depletion in genes in this pathway between irradiated (yet persistent) RC and RT-naïve tumours, notably those centred around the actions on CDP-DAG entering the Kennedy pathway and subsequent generation of PA, PE and LPC. Both of these works not only served to provide a new insight into existing data, but also justified the lipidomics approach taken in the subsequent chapter.

8.2.2 Mapping cellular lipid modulation in response to radiotherapy with DESI MSI

In Chapter 5, I set out to utilise DESI MSI to profile fresh frozen RC tissue in two cohorts; those had had been subjected to radiotherapy and exhibited a degree of tumour regression (yet with persistent and therefore, by definition, radioresistant tumour glands), and matched RC tissue that had not been subjected to radiotherapy. This is the first time that the technology has been employed for a hypothesis-driven lipidomics study of response to therapy in RC, although the experimental approach was to perform untargeted data acquisition across the full mass range before homing in on GPLs during the analysis. Using DESI MSI, I was able to generate lipidomic profiles in the negative ionisation mode that could not only accurately differentiate both normal tissue compartments and features of the tumour microenvironment on the basis of the spatially-aligned mass spectra generated, but could also differentiate between irradiated, radioresistant RC and RT-naïve RC in both unsupervised and supervised statistical models. Further analysis demonstrated 11 GPLs responsible for the discriminatory power in the model, and these were further classified using same-standards DESI MS/MS. GPL IDs were confirmed as being a mixture of PEs, PGs and PSs, with some ether-lipid variants. Fatty acid (FA) *sn2* chains were demonstrated to have high incidence of polyunsaturation, suggesting a role for fatty acid elongases ELOVL2, ELOVL4 and ELOVL5 on essential (ω-3 and ω-6) fatty acids in their composition. This is in line with other recently published work on more aggressive phenotypes of oesophageal cancer published by the group. [342] Preliminary nested analysis according to tumour regression grade demonstrates clustering of RC gland spectra according to TRG, with complete absence of lipids in TRG1-2 that are abundant in TRG4-5 and vice-versa. TRG3 emerges as a heterogeneous group that demonstrates high variance between individual tumours, with some having a

phenotype more akin to good responders and others with a lipidomic profile similar to poor responders.

A significant strength of the DESI MSI approach in this case is that it gives the closest possible snapshot of the representative biology, matched in real time to histological features of the tumour microenvironment. It is also important to note that the GPLs detected are far more likely to be biologically plausible owing to the complete lack of preparation steps used for the tissue sample and the absence of tissue surface disruption by vacuums or matrices inherent to other mass spectrometry imaging techniques.

8.2.3 Assessment of the whole tumour genome pre- and post-radiotherapy

In Chapter 6, I set out to move away from microarray-based technology to undertake both a more holistic and more reproducible overview of the RC genome in this patient cohort. The objectives for this work were (1) to perform high-quality DNA extractions from matched pre-treatment biopsies and post-treatment rectal cancer specimens to pass into a Next-Generation Sequencing pipeline; (2) to separately sequence discrete regions of macroscopic tumour prior to RT using Whole Genome Sequencing (WGS) to evaluate the degree of tumour genetic heterogeneity with relation to somatic copy number aberrations (SCNAs) and (3) to employ a bioinformatics-based approach to analyse and compare aggregated SCNAs both within-patient (pre- and post-RT) and between patients to establish the role of genomic heterogeneity in radiotherapy response. Multi-region dissections resulted in 65 distinct tumour regions across pre-RT, post-RT and RT-naïve RC. The primary outcome in

this work was the assessment of somatic copy number aberration (SCNA) across multiple macroscopically distinct regions of pre-RT biopsies and following RT.

Comparative analysis to the 175 RC datasets in The Cancer Genome Atlas (TCGA) demonstrated the sample set to be representative. All tumours demonstrated a high degree of SCNA, suggestive of a CMS2 or CMS4 phenotype. Multi-region sequencing in pre-RT biopsies demonstrated significant degrees of divergence in the genome, with this heterogeneity being far more pronounced in those tumours fated to have a good response to RT (TRG1-2). The SCNA profile of those tumours fated to have a poor response (TRG4-5) demonstrated a far more homogeneous genome across the macroscopic landscape of the tumour. In this small cohort, nested analysis of pre-RT TRG3 tumours demonstrates a SCNA profile similar to that of TRG4-5. Finally, analysis of the regions of the genome with significant SCN gains and losses demonstrates a degree of correlation with loci on which implicated genes in the FA elongation and GPL pathways responsible for the changes seen in DESI MSI are present. This pilot data requires deeper analysis and validation with a test set but yields new insights into a role for SCNAs rather than core mutation in radioresistance; if proven, this would more in-keeping with the biology of CMS2 and CMS4 phenotypes as described.

8.3 Promising areas for future work

Although this research has generated valuable, novel data on radioresistance mechanisms in RC, it really has only just scratched the surface in terms of the deeper dive into the comprehensive multi-omics phenotyping that is required. It therefore should be emphasised that the work presented herein represents a pilot study, albeit with a vast amount of biologically useful data. A great deal of the ongoing work has already been covered in Chapter 6, but the following represent specific additional objectives in this field for the future:

- Analysis of pre- and post-RT FFPE rectal cancer tissue using Imaging Mass Cytometry, with antibodies targeted to key steps within the GPL metabolism cascade. This would be specifically targeted to elements of the CDP-DAG and Kennedy pathways responsible for the generation of PE, PG and PS as has been borne out of both the *in silico* and DESI MSI data. Additional investigation of the actions of the FA elongases that act on essential fatty acids would be valuable.

- A further test set of pre-RT, multi-region RC biopsies to validate the findings from the WGS SCNA analysis. Deeper sequencing for mutational analysis may then be indicated. However, the missing link in the chain with regards to NGS will be RNASeq to account for post-translational expression of the coding RNAs for the above enzymes implicated in GPL and FA metabolism.

- To further validate the 16S rRNA dataset with regards to the impact of the ‘on tumour’ microbiome and its role in promoting radioresistance.

- Fusion of metabolic imaging, histological, genomic and metagenomic datasets to develop more precise classification models and to begin to crystallise the complex multi-level interplays that are likely to be responsible for radioresistance. These desired advances in bioinformatics remain major biomedical research priorities and further work and investment in this area will be critical if the full value of these individual datasets is to be realised.

References

1. GLOBOCAN. Estimated cancer incidence, mortality and prevalence worldwide in 2018: WHO; 2018 [Available from: https://gco.iarc.fr/today/data/factsheets/cancers/10_8_9-Colorectum-fact-sheet.pdf accessed 17th August 2020.

2. (CRUK) CRU. Bowel cancer statistics 2017 [Available from: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/bowel-cancer accessed 17th August 2020.

3. Brenner H, Bouvier AM, Foschi R, et al. Progress in colorectal cancer survival in Europe from the late 1980s to the early 21st century: the EUROCARE study. *Int J Cancer* 2012;131(7):1649-58. doi: 10.002/ijc.26192. Epub 2011 Aug 3.

4. Miller KD, Siegel RL, Lin CC, et al. Cancer treatment and survivorship statistics, 2016. *CA Cancer J Clin* 2016;66(4):271-89. doi: 10.3322/caac.21349. Epub 2016 Jun 2.

5. Brenner H, Kloor M, Pox C. Colorectal cancer. *Lancet* 2014;383(9927) doi: 10.1016/S0140-6736(13)61649-9

6. Sung H, Siegel RL, Rosenberg PS, et al. Emerging cancer trends among young adults in the USA: analysis of a population-based cancer registry. *Lancet Public Health* 2019;1(18):30267-6.

7. Saleh R, Sasidharan Nair V, Murshed K, et al. Transcriptome of CD8(+) tumor-infiltrating T cells: a link between diabetes and colorectal cancer. *Cancer Immunol Immunother* 2021;13(10):021-02879.

8. Hultcrantz R. Aspects of colorectal cancer screening, methods, age and gender. *J Intern Med* 2020;14(10):13171.

9. O'Sullivan DE, Sutherland RL, Town S, et al. Risk Factors for Early-Onset Colorectal Cancer: A Systematic Review and Meta-analysis. *Clin Gastroenterol Hepatol* 2021;29(21):00087-2.

10. Zhou E, Rifkin S. Colorectal Cancer and Diet: Risk Versus Prevention, Is Diet an Intervention? *Gastroenterol Clin North Am* 2021;50(1):101-11. doi: 10.1016/j.gtc.2020.10.012.

11. Keum N, Giovannucci E. Global burden of colorectal cancer: emerging trends, risk factors and prevention strategies. *Nat Rev Gastroenterol Hepatol* 2019;16(12):713-32. doi: 10.1038/s41575-019-0189-8. Epub 2019 Aug 27.

12. Taghinezhad SS, Mohseni AH, Fu X. Intervention on gut microbiota may change the strategy for management of colorectal cancer. *J Gastroenterol Hepatol* 2020;8(10):15369.

13. Mori G, Pasca MR. Gut Microbial Signatures in Sporadic and Hereditary Colorectal Cancer. *Int J Mol Sci* 2021;22(3):1312. doi: 10.3390/ijms22031312.

14. Heo G, Lee Y, Im E. Interplay between the Gut Microbiota and Inflammatory Mediators in the Development of Colorectal Cancer. *Cancers (Basel)* 2021;13(4):734. doi: 10.3390/cancers13040734.

15. Coleman OI, Haller D. Microbe-Mucus Interface in the Pathogenesis of Colorectal Cancer. *Cancers (Basel)* 2021;13(4):616. doi: 10.3390/cancers13040616.

16. Chattopadhyay I, Dhar R, Pethusamy K, et al. Exploring the Role of Gut Microbiome in Colon Cancer. *Appl Biochem Biotechnol* 2021;25(10):021-03498.

17. Fiorillo C, Schena CA, Quero G, et al. Challenges in Crohn's Disease Management after Gastrointestinal Cancer Diagnosis. *Cancers (Basel)* 2021;13(3):574. doi: 10.3390/cancers13030574.

18. Fantini MC, Guadagni I. From inflammation to colitis-associated colorectal cancer in inflammatory bowel disease: Pathogenesis and impact of current therapies. *Dig Liver Dis* 2021;1(21):00018-9.

19. Lichtenstein P, Holm NV, Verkasalo PK, et al. Environmental and heritable factors in the causation of cancer--analyses of cohorts of twins from Sweden, Denmark, and Finland. *N Engl J Med* 2000;343(2):78-85. doi: 10.1056/NEJM200007133430201.

20. Czene K, Lichtenstein P, Hemminki K. Environmental and heritable causes of cancer among 9.6 million individuals in the Swedish Family-Cancer Database. *Int J Cancer* 2002;99(2):260-6. doi: 10.1002/ijc.10332.

21. Syngal S, Brand RE, Church JM, et al. ACG clinical guideline: Genetic testing and management of hereditary gastrointestinal cancer syndromes. *Am J Gastroenterol* 2015;110(2):223-62; quiz 63. doi: 10.1038/ajg.2014.435. Epub 15 Feb 3.

22. Dekker E, Tanis PJ, Vleugels JLA, et al. Colorectal Cancer. *The Lancet* 2019;394:1467-80. doi: https://doi.org/10.1016/S0140-6736(19)32319-0

23. Recio-Boiles A, Kashyap S, Tsoris A, et al. Rectal Cancer. 2021

24. Network CGA. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* 2012;487(7407):330-7. doi: 10.1038/nature11252.

25. Nazemalhosseini Mojarad E, Kuppen PJ, Aghdaei HA, et al. The CpG island methylator phenotype (CIMP) in colorectal cancer. *Gastroenterol Hepatol Bed Bench* 2013;6(3):120-8.

26. Guinney J, Dienstmann R, Wang X, et al. The consensus molecular subtypes of colorectal cancer. *Nat Med* 2015;21(11):1350-6. doi: 10.038/nm.3967. Epub 2015 Oct 12.

27. Williams NSB, C. J. K.; O'Connell, P. R. Bailey & Love's Short Practice of Surgery. 2013

28. Brierley JD, Gospodarowicz MK, Wittekind C. Colon and Rectum. TNM Classification of Malignant Tumours. 8 ed. Oxford, UK: WIley-Blackwell 2017.

29. Hotta T, Takifuji K, Yokoyama S, et al. Survival of patients with T3 stage III rectal cancer is significantly worse than T2 stage III rectal cancer. A new proposal of a subdivision of stage III rectal cancer. *Dig Surg* 2009;26(1):69-74. doi: 10.1159/000194948. Epub 2009 Jan 23.

30. Siddiqui MRS, Simillis C, Bhoday J, et al. A meta-analysis assessing the survival implications of subclassifying T3 rectal tumours. *Eur J Cancer* 2018;104:47-61.(doi):10.1016/j.ejca.2018.07.131. Epub 18 Oct 12.

31. Zinicola R, Pedrazzi G, Haboubi N, et al. The degree of extramural spread of T3 rectal cancer: an appeal to the American Joint Committee on Cancer. *Colorectal Dis* 2017;19(1):8-15. doi: 10.1111/codi.13565.

32. Balbaa MA, Elkady N, Abdelrahman EM. Predictive Factors of Positive Circumferential and Longitudinal Margins in Early T3 Colorectal Cancer Resection. doi: 10.1155/2020/6789709

33. Otero de Pablos J, Mayol J. Controversies in the Management of Lateral Pelvic Lymph Nodes in Patients With Advanced Rectal Cancer: East or West?

34. Watanabe T, Muro K, Ajioka Y, et al. Japanese Society for Cancer of the Colon and Rectum (JSCCR) guidelines 2016 for the treatment of colorectal cancer. *Int J Clin Oncol* 2018;23(1):1-34. doi: 10.1007/s10147-017-1101-6. Epub 2017 Mar 27.

35. Hagemans JAW, Rothbarth J, van Bogerijen GHW, et al. Treatment of Inguinal Lymph Node Metastases in Patients with Rectal Adenocarcinoma. *Ann Surg Oncol* 2019;26(4):1134-41. doi: 10.245/s10434-019-07191-4. Epub 2019 Feb 6.

36. Lord A, D'Souza N, Shaw A, et al. The Current Status of Nodal Staging in Rectal Cancer. doi: 10.1007/s11888-019-00441-3

37. Lord AC, D'Souza N, Shaw A, et al. MRI-Diagnosed Tumour Deposits and EMVI Status Have Superior Prognostic Accuracy to Current Clinical TNM Staging in Rectal Cancer. *Ann Surg* 2020;15(10):0000000000004499.

38. Lord AC, Knijn N, Brown G, et al. Pathways of spread in rectal cancer: a reappraisal of the true routes to distant metastatic disease. *Eur J Cancer* 2020;128:1-6.(doi):10.1016/j.ejca.2019.12.025. Epub 20 Mar 5.

39. Bhoday J, Balyasnikova S, Wale A, et al. How Should Imaging Direct/Orient Management of Rectal Cancer? *Clin Colon Rectal Surg* 2017;30(5):297-312. doi: 10.1055/s-0037-1606107. Epub 2017 Nov 27.

40. Siddiqui MRS, Simillis C, Hunter C, et al. A meta-analysis comparing the risk of metastases in patients with rectal cancer and MRI-detected extramural vascular invasion (mrEMVI) vs mrEMVI-negative cases. *Br J Cancer* 2017;116(12):1513-19. doi: 10.038/bjc.2017.99. Epub 17 Apr 27.

41. Lord AC, D'Souza N, Pucher PH, et al. Significance of extranodal tumour deposits in colorectal cancer: A systematic review and meta-analysis. *Eur J Cancer* 2017;82:92-102.(doi):10.1016/j.ejca.2017.05.027. Epub 17 Jun 23.

42. Tuech JJ, Pinson J, Nouhaud FX, et al. Total Pelvic Exenteration, Cytoreductive Surgery, and Hyperthermic Intraperitoneal Chemotherapy for Rectal Cancer with Associate Peritoneal Metastases: Surgical Strategies to Optimize Safety. *Cancers (Basel)* 2020;12(11):3478. doi: 10.390/cancers12113478.

43. Loughrey MBQ, P.; Shepherd, N. A. G049 Dataset for histopathological reporting of colorectal cancer. *The Royal College of Pathologists* 2018

44. Graule J, Uth K, Fischer E, et al. CDX2 in colorectal cancer is an independent prognostic factor and regulated by promoter methylation and histone deacetylation in tumors of the serrated pathway. *Clin Epigenetics* 2018;10(1):120. doi: 10.1186/s13148-018-0548-2.

45. Tomasello G, Barni S, Turati L, et al. Association of CDX2 Expression With Survival in Early Colorectal Cancer: A Systematic Review and Meta-analysis. *Clin Colorectal Cancer* 2018;17(2):97-103. doi: 10.1016/j.clcc.2018.02.001. Epub 18 Feb 15.

46. Dalerba P, Sahoo D, Paik S, et al. CDX2 as a Prognostic Biomarker in Stage II and Stage III Colon Cancer. *N Engl J Med* 2016;374(3):211-22. doi: 10.1056/NEJMoa1506597.

47. Chen W, Farchoukh L, Seigh L, et al. Combined histopathological risk score using TP53 protein expression, CD8(+) T cell density and intratumoral budding is an independent predictor of neoadjuvant therapy response in rectal adenocarcinoma. *Histopathology* 2021;79(5):826-35. doi: 10.1111/his.14430. Epub 2021 Sep 3.

48. Georges NDF, Oberli B, Rau TT, et al. Tumour budding and CD8(+) T cells: 'attackers' and 'defenders' in rectal cancer with and without neoadjuvant chemoradiotherapy. *Histopathology* 2021;78(7):1009-18. doi: 10.111/his.14319. Epub 2021 Apr 19.

49. Park JH, van Wyk H, Roxburgh CSD, et al. Tumour invasiveness, the local and systemic environment and the basis of staging systems in colorectal cancer. *Br J Cancer* 2017;116(11):1444-50. doi: 10.038/bjc.2017.108. Epub 17 Apr 20.

50. Shin JK, Park YA, Huh JW, et al. Tumor Budding as a Prognostic Marker in Rectal Cancer Patients on Propensity Score Analysis. *Ann Surg Oncol* 2021;28(13):8813-22. doi: 10.1245/s10434-021-286-6. Epub 2021 Jul 7.

51. Yang Y, Tian W, Su L, et al. Tumor-Infiltrating Cytotoxic T Cells and Tumor-Associated Macrophages Correlate With the Outcomes of Neoadjuvant Chemoradiotherapy for Locally Advanced Rectal Cancer. *Front Oncol* 2021;11:743540.(doi):10.3389/fonc.2021.743540. eCollection 2021.

52. Zinovkin DA, Kose SY, Nadyrov EA, et al. Potential role of tumor-infiltrating T-, B-lymphocytes, tumor-associated macrophages and IgA-secreting plasma cells in long-term survival in the rectal adenocarcinoma patients. *Life Sci* 2021;286:120052.(doi):10.1016/j.lfs.2021.120052. Epub 2021 Oct 14.

53. Jeon SH, Lim YJ, Koh J, et al. A radiomic signature model to predict the chemoradiation-induced alteration in tumor-infiltrating CD8(+) cells in locally advanced rectal cancer. *Radiother Oncol* 2021;162:124-131.(doi):10.1016/j.radonc.2021.07.004. Epub 21 Jul 12.

54. Li Z, Chen F, Zhang S, et al. The feasibility of MRI-based radiomics model in presurgical evaluation of tumor budding in locally advanced rectal cancer. *Abdom Radiol* 2021;21(10):021-03311.

55. Ye D, Zhu Z, Chen F, et al. Correlation Between Endorectal Ultrasound and Magnetic Resonance Imaging for Predicting the Circumferential Resection Margin in Patients With Mid-Low Rectal Cancer Without Preoperative Chemoradiotherapy. *J Ultrasound Med* 2020;39(3):569-77. doi: 10.1002/jum.15135. Epub 2019 Oct 16.

56. Burdan F, Sudol-Szopinska I, Staroslawska E, et al. Magnetic resonance imaging and endorectal ultrasound for diagnosis of rectal lesions. *Eur J Med Res* 2015;20(1):4. doi: 10.1186/s40001-014-0078-0.

57. Samdani T, Garcia-Aguilar J. Imaging in rectal cancer: magnetic resonance imaging versus endorectal ultrasonography. *Surg Oncol Clin N Am* 2014;23(1):59-77. doi: 10.1016/j.soc.2013.09.011.

58. Taylor FG, Quirke P, Heald RJ, et al. Preoperative high-resolution magnetic resonance imaging can identify good prognosis stage I, II, and III rectal cancer best managed by surgery alone: a prospective, multicenter, European study. *Ann Surg* 2011;253(4):711-9. doi: 10.1097/SLA.0b013e31820b8d52.

59. van Gijn W, Marijnen CA, Nagtegaal ID, et al. Preoperative radiotherapy combined with total mesorectal excision for resectable rectal cancer: 12-year follow-up of the multicentre, randomised controlled TME trial. *Lancet Oncol* 2011;12(6):575-82. doi: 10.1016/S470-2045(11)70097-3. Epub 2011 May 17.

60. Braendengen M, Tveit KM, Berglund A, et al. Randomized phase III study comparing preoperative radiotherapy with chemoradiotherapy in nonresectable rectal cancer. *J Clin Oncol* 2008;26(22):3687-94. doi: 10.1200/JCO.2007.15.3858.

61. Sun W, Al-Rajabi R, Perez RO, et al. Controversies in Rectal Cancer Treatment and Management. *Am Soc Clin Oncol Educ Book* 2020;40:1-11.(doi):10.1200/EDBK_279871.

62. São Julião GP, Karagkounis G, Fernandez LM, et al. Conditional Survival in Patients With Rectal Cancer and Complete Clinical Response Managed by Watch and Wait After Chemoradiation: Recurrence Risk Over Time. *Ann Surg* 2020;272(1):138-44. doi: 10.1097/SLA.0000000000003286.

63. Habr-Gama A, São Julião GP, Vailati BB, et al. Organ Preservation Among Patients With Clinically Node-Positive Rectal Cancer: Is It Really More Dangerous? *Dis Colon Rectum* 2019;62(6):675-83. doi: 10.1097/DCR.0000000000001337.

64. Habr-Gama A, São Julião GP, Fernandez LM, et al. Achieving a Complete Clinical Response After Neoadjuvant Chemoradiation That Does Not Require Surgical Resection: It May Take Longer Than You Think! *Dis Colon Rectum* 2019;62(7):802-08. doi: 10.1097/DCR.0000000000001338.

65. Fernandez LM, São Julião GP, Figueiredo NL, et al. Conditional recurrence-free survival of clinical complete responders managed by watch and wait after neoadjuvant chemoradiotherapy for rectal cancer in the International Watch & Wait Database: a retrospective, international, multicentre registry study. *Lancet Oncol* 2021;22(1):43-50. doi: 10.1016/S470-2045(20)30557-X. Epub 2020 Dec 11.

66. Dattani M, Heald RJ, Goussous G, et al. Oncological and Survival Outcomes in Watch and Wait Patients With a Clinical Complete Response After Neoadjuvant Chemoradiotherapy for Rectal Cancer: A Systematic Review and Pooled Analysis. *Ann Surg* 2018;268(6):955-67. doi: 10.1097/SLA.0000000000002761.

67. Rombouts AJM, Al-Najami I, Abbott NL, et al. Can we Save the rectum by watchful waiting or TransAnal microsurgery following (chemo) Radiotherapy versus Total mesorectal excision for early REctal Cancer (STAR-TREC study)?: protocol for a multicentre, randomised feasibility study. *BMJ Open* 2017;7(12):e019474. doi: 10.1136/bmjopen-2017-74.

68. Bach SP, Gilbert A, Brock K, et al. Radical surgery versus organ preservation via short-course radiotherapy followed by transanal endoscopic microsurgery for early-stage rectal cancer (TREC): a randomised, open-label feasibility study. *Lancet Gastroenterol Hepatol* 2021;6(2):92-105. doi: 10.1016/S2468-1253(20)30333-2. Epub 2020 Dec 11.

69. Heald RJ, Husband EM, Ryall RD. The mesorectum in rectal cancer surgery--the clue to pelvic recurrence? *Br J Surg* 1982;69(10):613-6.

70. Vennix S, Pelzers L, Bouvy N, et al. Laparoscopic versus open total mesorectal excision for rectal cancer. *Cochrane Database Syst Rev* 2014(4):CD005200. doi: 10.1002/14651858.CD005200.pub3.

71. Eltair M, Hajibandeh S, Nuno A, et al. Meta-analysis and trial sequential analysis of robotic versus laparoscopic total mesorectal excision in management of rectal cancer. *Int J Colorectal Dis* 2020;35(8):1423-38. doi: 10.007/s00384-020-3655-2. Epub 2020 Jun 19.

72. Collard M, Lefevre JH. Ultimate Functional Preservation With Intersphincteric Resection for Rectal Cancer. *Front Oncol* 2020;10:297.(doi):10.3389/fonc.2020.00297. eCollection 2020.

73. Piozzi GN, Baek SJ, Kwak JM, et al. Anus-Preserving Surgery in Advanced Low-Lying Rectal Cancer: A Perspective on Oncological Safety of Intersphincteric Resection. *Cancers (Basel)* 2021;13(19):4793. doi: 10.3390/cancers13194793.

74. Yamada K, Saiki Y, Takano S, et al. Long-term results of intersphincteric resection for low rectal cancer in Japan. *Surg Today* 2019;49(4):275-85. doi: 10.1007/s00595-018-1754-4. Epub 2019 Jan 2.

75. Lehtonen T, Räsänen M, Carpelan-Holmström M, et al. Oncological outcomes before and after the extralevator abdominoperineal excision era in rectal cancer patients treated with abdominoperineal excision in a single centre, high volume unit. *Colorectal Dis* 2019;21(2):183-90. doi: 10.1111/codi.14468. Epub 2019 Jan 3.

76. Papp G, Dede K, Bursics A. Short-term advantages of ELAPE over APR. *Acta Chir Belg* 2021;121(5):327-32. doi: 10.1080/00015458.2020.1778265. Epub 2020 Jun 12.

77. Tao Y, Han JG, Wang ZJ. Extralevator abdominoperineal excision for advanced low rectal cancer: Where to go. *World J Gastroenterol* 2020;26(22):3012-23. doi: 10.748/wjg.v26.i22.12.

78. Milone M, Manigrasso M, Burati M, et al. Surgical resection for rectal cancer. Is laparoscopic surgery as successful as open approach? A systematic review with meta-analysis. *PLoS One* 2018;13(10):e0204887. doi: 10.1371/journal.pone.87. eCollection 2018.

79. Hsieh C, Cologne KG. Laparoscopic Approach to Rectal Cancer—The New Standard?

80. Group EC. The Impact of Enhanced Recovery Protocol Compliance on Elective Colorectal Cancer Resection: Results From an International Registry. *Ann Surg* 2015;261(6):1153-9. doi: 10.097/SLA.0000000000001029.

81. van der Pas MH, Haglind E, Cuesta MA, et al. Laparoscopic versus open surgery for rectal cancer (COLOR II): short-term outcomes of a randomised, phase 3 trial. *Lancet Oncol* 2013;14(3):210-8. doi: 10.1016/S470-2045(13)70016-0. Epub 2013 Feb 6.

82. Kang SB, Park JW, Jeong SY, et al. Open versus laparoscopic surgery for mid or low rectal cancer after neoadjuvant chemoradiotherapy (COREAN trial): short-term outcomes of an open-label randomised controlled trial. *Lancet Oncol* 2010;11(7):637-45. doi: 10.1016/S470-2045(10)70131-5. Epub 2010 Jun 16.

83. Jeong SY, Park JW, Nam BH, et al. Open versus laparoscopic surgery for mid-rectal or low-rectal cancer after neoadjuvant chemoradiotherapy (COREAN trial): survival outcomes of an open-label, non-inferiority, randomised controlled trial. *Lancet Oncol* 2014;15(7):767-74. doi: 10.1016/S470-2045(14)70205-0. Epub 2014 May 15.

84. Jayne DG, Guillou PJ, Thorpe H, et al. Randomized trial of laparoscopic-assisted resection of colorectal carcinoma: 3-year results of the UK MRC CLASICC Trial Group. *J Clin Oncol* 2007;25(21):3061-8. doi: 10.1200/JCO.2006.09.7758.

85. Guillou PJ, Quirke P, Thorpe H, et al. Short-term endpoints of conventional versus laparoscopic-assisted surgery in patients with colorectal cancer (MRC CLASICC trial): multicentre, randomised controlled trial. *Lancet* 2005;365(9472):1718-26. doi: 10.016/S0140-6736(05)66545-2.

86. Stevenson AR, Solomon MJ, Lumley JW, et al. Effect of Laparoscopic-Assisted Resection vs Open Resection on Pathological Outcomes in Rectal Cancer: The ALaCaRT Randomized Clinical Trial. *JAMA* 2015;314(13):1356-63. doi: 10.001/jama.2015.12009.

87. Son GM, Kye BH, Kim MK, et al. Reconsideration of the Safety of Laparoscopic Rectal Surgery for Cancer. *Ann Coloproctol* 2019;35(5):229-37. doi: 10.3393/ac.2019.10.16. Epub 19 Oct 31.

88. Fleshman J, Branda M, Sargent DJ, et al. Effect of Laparoscopic-Assisted Resection vs Open Resection of Stage II or III Rectal Cancer on Pathologic Outcomes: The ACOSOG Z6051 Randomized Clinical Trial. *JAMA* 2015;314(13):1346-55. doi: 10.001/jama.2015.10529.

89. Jayne D, Pigazzi A, Marshall H, et al. Effect of Robotic-Assisted vs Conventional Laparoscopic Surgery on Risk of Conversion to Open Laparotomy Among Patients Undergoing Resection for Rectal Cancer: The ROLARR Randomized Clinical Trial. *JAMA* 2017;318(16):1569-80. doi: 10.001/jama.2017.7219.

90. Zorcolo L, Rosman AS, Restivo A, et al. Complete pathologic response after combined modality treatment for rectal cancer and long-term survival: a meta-analysis. *Ann Surg Oncol* 2012;19(9):2822-32. doi: 10.1245/s10434-011-2209-y. Epub 2012 Mar 21.

91. Yoon SM, Kim DY, Kim TH, et al. Clinical parameters predicting pathologic tumor response after preoperative chemoradiotherapy for rectal cancer. *Int J Radiat Oncol Biol Phys* 2007;69(4):1167-72. doi: 10.016/j.ijrobp.2007.04.047.

92. Maggiori L, Bretagnol F, Aslam MI, et al. Does pathologic response of rectal cancer influence postoperative morbidity after neoadjuvant radiochemotherapy and total mesorectal excision? *Surgery* 2014;155(3):468-75. doi: 10.1016/j.surg.2013.10.020. Epub 13 Oct 22.

93. Lorimer PD, Motz BM, Kirks RC, et al. Pathologic Complete Response Rates After Neoadjuvant Treatment in Rectal Cancer: An Analysis of the National Cancer Database. *Ann Surg Oncol* 2017;24(8):2095-103. doi: 10.1245/s10434-017-5873-8. Epub 2017 May 22.

94. Garland ML, Vather R, Bunkley N, et al. Clinical tumour size and nodal status predict pathologic complete response following neoadjuvant chemoradiotherapy for rectal cancer. *Int J Colorectal Dis* 2014;29(3):301-7. doi: 10.1007/s00384-013-1821-7. Epub 2014 Jan 14.

95. Fernández-Aceñero MJ, Estrada Muñoz L, Sastre Varela J, et al. Prognostic influence of histopathological regression patterns in rectal adenocarcinoma receiving neoadjuvant therapy. *J Gastrointest Oncol* 2017;8(1):49-54. doi: 10.21037/jgo.2017.01.02.

96. Wong RK, Tandan V, De Silva S, et al. Pre-operative radiotherapy and curative surgery for the management of localized rectal carcinoma. *Cochrane Database Syst Rev* 2007(2):CD002102. doi: 10.1002/14651858.CD002102.pub2.

97. Glimelius B, Grönberg H, Järhult J, et al. A systematic overview of radiation therapy effects in rectal cancer. *Acta Oncol* 2003;42(5-6):476-92. doi: 10.1080/02841860310012301.

98. Cammà C, Giunta M, Fiorica F, et al. Preoperative radiotherapy for resectable rectal cancer: A meta-analysis. *JAMA* 2000;284(8):1008-15. doi: 10.01/jama.284.8.08.

99. Health NIf. NIH consensus conference. Adjuvant therapy for patients with colon and rectal cancer. *JAMA* 1990;264(11):1444-50.

100. Cedermark B, Johansson H, Rutqvist LE, et al. The Stockholm I trial of preoperative short term radiotherapy in operable rectal carcinoma. A prospective randomized trial. Stockholm Colorectal Cancer Study Group. *Cancer* 1995;75(9):2269-75. doi: 10.1002/97-0142(19950501)75:9<2269::aid-cncr2820750913>3.0.co;2-i.

101. Cedermark B, Dahlberg M, Glimelius B, et al. Improved survival with preoperative radiotherapy in resectable rectal cancer. *N Engl J Med* 1997;336(14):980-7. doi: 10.1056/NEJM199704033361402.

102. Sebag-Montefiore D, Stephens RJ, Steele R, et al. Preoperative radiotherapy versus selective postoperative chemoradiotherapy in patients with rectal cancer (MRC CR07 and NCIC-CTG C016): a multicentre, randomised trial. *Lancet* 2009;373(9666):811-20. doi: 10.1016/S0140-6736(09)60484-0.

103. Kapiteijn E, Marijnen CA, Nagtegaal ID, et al. Preoperative radiotherapy combined with total mesorectal excision for resectable rectal cancer. *N Engl J Med* 2001;345(9):638-46. doi: 10.1056/NEJMoa010580.

104. Bosset JF, Collette L, Calais G, et al. Chemotherapy with preoperative radiotherapy in rectal cancer. *N Engl J Med* 2006;355(11):1114-23. doi: 10.056/NEJMoa060829.

105. Gérard JP, Conroy T, Bonnetain F, et al. Preoperative radiotherapy with or without concurrent fluorouracil and leucovorin in T3-4 rectal cancers: results of FFCD 9203. *J Clin Oncol* 2006;24(28):4620-5. doi: 10.1200/JCO.2006.06.7629.

106. Bahadoer RR, Dijkstra EA, van Etten B, et al. Short-course radiotherapy followed by chemotherapy before total mesorectal excision (TME) versus preoperative chemoradiotherapy, TME, and optional adjuvant chemotherapy in locally advanced rectal cancer (RAPIDO): a randomised, open-label, phase 3 trial. *Lancet Oncol* 2021;22(1):29-42. doi: 10.1016/S470-2045(20)30555-6. Epub 2020 Dec 7.

107. Glehen O, Chapet O, Adham M, et al. Long-term results of the Lyons R90-01 randomized trial of preoperative radiotherapy with delayed surgery and its effect on sphincter-saving surgery in rectal cancer. *Br J Surg* 2003;90(8):996-8. doi: 10.1002/bjs.4162.

108. Francois Y, Nemoz CJ, Baulieux J, et al. Influence of the interval between preoperative radiation therapy and surgery on downstaging and on the rate of sphincter-sparing surgery for rectal cancer: the Lyon R90-01 randomized trial. *J Clin Oncol* 1999;17(8):2396. doi: 10.1200/JCO.999.17.8.2396.

109. Stein DE, Mahmoud NN, Anne PR, et al. Longer time interval between completion of neoadjuvant chemoradiation and surgical resection does not improve downstaging of rectal carcinoma. *Dis Colon Rectum* 2003;46(4):448-53. doi: 10.1007/s10350-004-6579-0.

110. Petrelli F, Sgroi G, Sarti E, et al. Increasing the Interval Between Neoadjuvant Chemoradiotherapy and Surgery in Rectal Cancer: A Meta-analysis of Published Studies. *Ann Surg* 2016;263(3):458-64. doi: 10.1097/SLA.0000000000000368.

111. Erlandsson J, Holm T, Pettersson D, et al. Optimal fractionation of preoperative radiotherapy and timing to surgery for rectal cancer (Stockholm III): a multicentre, randomised, non-blinded, phase 3, non-inferiority trial. *Lancet Oncol* 2017;18(3):336-46. doi: 10.1016/S470-2045(17)30086-4. Epub 2017 Feb 10.

112. Renehan AG, Malcomson L, Emsley R, et al. Watch-and-wait approach versus surgical resection after chemoradiotherapy for patients with rectal cancer (the OnCoRe project): a propensity-score matched cohort analysis. *Lancet Oncol* 2016;17(2):174-83. doi: 10.1016/S470-2045(15)00467-2. Epub 2015 Dec 17.

113. Petrelli F, Trevisan F, Cabiddu M, et al. Total Neoadjuvant Therapy in Rectal Cancer: A Systematic Review and Meta-analysis of Treatment Outcomes. *Ann Surg* 2019;15(10):0000000000003471.

114. Rullier E, Vendrely V, Asselineau J, et al. Organ preservation with chemoradiotherapy plus local excision for rectal cancer: 5-year results of the GRECCAR 2 randomised trial. *Lancet Gastroenterol Hepatol* 2020;5(5):465-74. doi: 10.1016/S2468-1253(19)30410-8. Epub 2020 Feb 7.

115. Sun Myint A, Dhadda A, Rao C, et al. Will GRECCAR 2 be a game changer for the management of rectal cancer? *Colorectal Dis* 2020;22(12):2330. doi: 10.1111/codi.15304. Epub 2020 Aug 30.

116. Lefèvre JH, Mineur L, Cachanado M, et al. Does A Longer Waiting Period After Neoadjuvant Radio-chemotherapy Improve the Oncological Prognosis of Rectal Cancer?: Three Years' Follow-up Results of the Greccar-6 Randomized Multicenter Trial. *Ann Surg* 2019;270(5):747-54. doi: 10.1097/SLA.0000000000003530.

117. Habr-Gama A, Perez RO. No Surgery After Chemoradiation Is Not Equal to Nonoperative Management After Complete Clinical Response and Chemoradiation. *J Clin Oncol* 2016;34(33):4051. doi: 10.1200/JCO.2016.67.9001. Epub 2016 Oct 31.

118. Perez RO, Habr-Gama A, Lynn PB, et al. Transanal endoscopic microsurgery for residual rectal cancer (ypT0-2) following neoadjuvant chemoradiation therapy: another word of caution. *Dis Colon Rectum* 2013;56(1):6-13. doi: 10.1097/DCR.0b013e318273f56f.

119. Habr-Gama A, Gama-Rodrigues J, São Julião GP, et al. Local recurrence after complete clinical response and watch and wait in rectal cancer after neoadjuvant chemoradiation: impact of salvage therapy on local disease control. *Int J Radiat Oncol Biol Phys* 2014;88(4):822-8. doi: 10.1016/j.ijrobp.2013.12.012. Epub 14 Feb 1.

120. van Timmeren JE, Cester D, Tanadini-Lang S, et al. Radiomics in medical imaging—"how-to" guide and critical reflection. doi: 10.1186/s13244-020-00887-2

121. Slatko BE, Gardner AF, Ausubel FM. Overview of Next-Generation Sequencing Technologies. *Curr Protoc Mol Biol* 2018;122(1):e59. doi: 10.1002/cpmb.59.

122. Ng SB, Turner EH, Robertson PD, et al. Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* 2009;461(7261):272-6. doi: 10.1038/nature08250. Epub 2009 Aug 16.

123. Belkadi A, Bolze A, Itan Y, et al. Whole-genome sequencing is more powerful than whole-exome sequencing for detecting exome variants. *Proc Natl Acad Sci U S A* 2015;112(17):5473-8. doi: 10.1073/pnas.1418631112. Epub 2015 Mar 31.

124. Hurd PJ, Nelson CJ. Advantages of next-generation sequencing versus the microarray in epigenetic research. *Briefings in Functional Genomics* 2009;8(3):174-83. doi: 10.1093/bfgp/elp013

125. Jessen E, Liu Y, Davila J, et al. Determining mutational burden and signature using RNA-seq from tumor-only samples. *BMC Med Genomics* 2021;14(1):65. doi: 10.1186/s12920-021-00898-y.

126. Baker AM, Huang W, Wang XM, et al. Robust RNA-based in situ mutation detection delineates colorectal cancer subclonal evolution. *Nat Commun* 2017;8(1):1998. doi: 10.038/s41467-017-02295-5.

127. Jackson HW, Fischer JR, Zanotelli VRT, et al. The single-cell pathology landscape of breast cancer. *Nature* 2020;578(7796):615-20. doi: 10.1038/s41586-019-1876-x. Epub 2020 Jan 20.

128. Aoki T, Chong LC, Takata K, et al. Single-Cell Transcriptome Analysis Reveals Disease-Defining T-cell Subsets in the Tumor Microenvironment of Classic Hodgkin Lymphoma. *Cancer Discov* 2020;10(3):406-21. doi: 10.1158/2159-8290.CD-19-0680. Epub 2019 Dec 19.

129. Nicholson JK, Lindon JC, Holmes E. 'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. *Xenobiotica; the fate of foreign compounds in biological systems* 1999;29(11):1181-9. doi: 10.1080/004982599238047 [published Online First: 1999/12/22]

130. Wishart DS, Feunang YD, Marcu A, et al. HMDB 4.0: the human metabolome database for 2018. *Nucleic acids research* 2018;46(D1):D608-d17. doi: 10.1093/nar/gkx1089 [published Online First: 2017/11/16]

131. Pertea M, Shumate A, Pertea G, et al. CHESS: a new human gene catalog curated from thousands of large-scale RNA sequencing experiments reveals extensive transcriptional noise. *Genome Biol* 2018;19(1):208. doi: 10.1186/s13059-018-1590-2.

132. Smith LM, Kelleher NL. Proteoform: a single term describing protein complexity. *Nature methods* 2013;10(3):186-7. doi: 10.1038/nmeth.2369 [published Online First: 2013/02/28]

133. Schwenk JM, Omenn GS, Sun Z, et al. The Human Plasma Proteome Draft of 2017: Building on the Human Plasma PeptideAtlas from Mass Spectrometry and Complementary Assays. *J Proteome Res* 2017;16(12):4299-310. doi: 10.1021/acs.jproteome.7b00467 [published Online First: 2017/09/25]

134. Sender R, Fuchs S, Milo R. Revised Estimates for the Number of Human and Bacteria Cells in the Body. *PLoS biology* 2016;14(8):e1002533. doi: 10.1371/journal.pbio.1002533 [published Online First: 2016/08/20]

135. Lloyd-Price J, Mahurkar A, Rahnavard G, et al. Strains, functions and dynamics in the expanded Human Microbiome Project. *Nature* 2017;550(7674):61-66. doi: 10.1038/nature23889 [published Online First: 2017/09/28]

136. Scott AJ, Alexander JL, Merrifield CA, et al. International Cancer Microbiome Consortium consensus statement on the role of the human microbiome in carcinogenesis. *Gut* 2019;68(9):1624-32. doi: 10.1136/gutjnl-2019-318556 [published Online First: 2019/05/17]

137. Houlston RS, Cheadle J, Dobbins SE, et al. Meta-analysis of three genome-wide association studies identifies susceptibility loci for colorectal cancer at 1q41, 3q26.2, 12q13.13 and 20q13.33. *Nature genetics* 2010;42(11):973-7. doi: 10.1038/ng.670 [published Online First: 2010/10/26]

138. Figueiredo JC, Hsu L, Hutter CM, et al. Genome-wide diet-gene interaction analyses for risk of colorectal cancer. *PLoS genetics* 2014;10(4):e1004228. doi: 10.1371/journal.pgen.1004228 [published Online First: 2014/04/20]

139. Parada Venegas D, De la Fuente MK, Landskron G, et al. Short Chain Fatty Acids (SCFAs)-Mediated Gut Epithelial and Immune Regulation and Its Relevance for Inflammatory Bowel Diseases. *Front Immunol* 2019;10:277.(doi):10.3389/fimmu.2019.00277. eCollection 2019.

140. Matthews GM, Howarth GS, Butler RN. Short-chain fatty acids induce apoptosis in colon cancer cells associated with changes to intracellular redox state and glucose metabolism. *Chemotherapy* 2012;58(2):102-9. doi: 10.1159/000335672. Epub 2012 Apr 6.

141. Canani RB, Costanzo MD, Leone L, et al. Potential beneficial effects of butyrate in intestinal and extraintestinal diseases. *World J Gastroenterol* 2011;17(12):1519-28. doi: 10.3748/wjg.v17.i12.1519.

142. O'Keefe SJ. Diet, microorganisms and their metabolites, and colon cancer. *Nat Rev Gastroenterol Hepatol* 2016;13(12):691-706. doi: 10.1038/nrgastro.2016.165 [published Online First: 2016/11/17]

143. Gomes SD, Oliveira CS, Azevedo-Silva J, et al. The Role of Diet Related Short-Chain Fatty Acids in Colorectal Cancer Metabolism and Survival: Prevention and Therapeutic Implications. *Current medicinal chemistry* 2018 doi: 10.2174/0929867325666180530102050 [published Online First: 2018/06/01]

144. Scott AJ, Alexander JL, Merrifield CA, et al. International Cancer Microbiome Consortium consensus statement on the role of the human microbiome in carcinogenesis. *Gut* 2019;68(9):1624-32. doi: 10.136/gutjnl-2019-318556. Epub 2019 May 15.

145. Li YY, Ge QX, Cao J, et al. Association of Fusobacterium nucleatum infection with colorectal cancer in Chinese patients. *World J Gastroenterol* 2016;22(11):3227-33. doi: 10.748/wjg.v22.i11.27.

146. Castellarin M, Warren RL, Freeman JD, et al. Fusobacterium nucleatum infection is prevalent in human colorectal carcinoma. *Genome Res* 2012;22(2):299-306. doi: 10.1101/gr.126516.111. Epub 2011 Oct 18.

147. Sze MA, Baxter NT, Ruffin MTt, et al. Normalization of the microbiota in patients after treatment for colonic lesions. *Microbiome* 2017;5(1):150. doi: 10.1186/s40168-017-0366-3 [published Online First: 2017/11/18]

148. Zheng DW, Dong X, Pan P, et al. Phage-guided modulation of the gut microbiota of mouse models of colorectal cancer augments their responses to chemotherapy. *Nature biomedical engineering* 2019;3(9):717-28. doi: 10.1038/s41551-019-0423-2 [published Online First: 2019/07/25]

149. Yu T, Guo F, Yu Y, et al. Fusobacterium nucleatum Promotes Chemoresistance to Colorectal Cancer by Modulating Autophagy. *Cell* 2017;170(3):548-63.e16. doi: 10.1016/j.cell.2017.07.008 [published Online First: 2017/07/29]

150. FOX GE, PECHMAN KR, WOESE CR. Comparative Cataloging of 16S Ribosomal Ribonucleic Acid: Molecular Approach to Procaryotic Systematics. *International Journal of Systematic and Evolutionary Microbiology* 1977;27(1):44-57. doi: https://doi.org/10.1099/00207713-27-1-44

151. Clarridge JE, 3rd. Impact of 16S rRNA gene sequence analysis for identification of bacteria on clinical microbiology and infectious diseases. *Clinical microbiology reviews* 2004;17(4):840-62, table of contents. doi: 10.1128/cmr.17.4.840-862.2004 [published Online First: 2004/10/19]

152. Ames NJ, Ranucci A, Moriyama B, et al. The Human Microbiome and Understanding the 16S rRNA Gene in Translational Nursing Science. *Nursing research* 2017;66(2):184-97. doi: 10.1097/nnr.0000000000000212 [published Online First: 2017/03/03]

153. Vincent AT, Derome N, Boyle B, et al. Next-generation sequencing (NGS) in the microbiological world: How to make the most of your money. *Journal of microbiological methods* 2017;138:60-71. doi: 10.1016/j.mimet.2016.02.016 [published Online First: 2016/03/21]

154. Warburg O. On the origin of cancer cells. *Science* 1956;123(3191):309-14. doi: 10.1126/science.123.3191.309.

155. Liberti MV, Locasale JW. The Warburg Effect: How Does it Benefit Cancer Cells? *Trends Biochem Sci* 2016;41(3):211-18. doi: 10.1016/j.tibs.2015.12.001. Epub 16 Jan 5.

156. Ward PS, Thompson CB. Metabolic reprogramming: a cancer hallmark even warburg did not anticipate. *Cancer cell* 2012;21(3):297-308. doi: 10.1016/j.ccr.2012.02.014 [published Online First: 2012/03/24]

157. Vander Heiden MG, Cantley LC, Thompson CB. Understanding the Warburg effect: the metabolic requirements of cell proliferation. *Science* 2009;324(5930):1029-33. doi: 10.1126/science.1160809 [published Online First: 2009/05/23]

158. Estrella V, Chen T, Lloyd M, et al. Acidity generated by the tumor microenvironment drives local invasion. *Cancer Res* 2013;73(5):1524-35. doi: 10.1158/0008-5472.Can-12-2796 [published Online First: 2013/01/05]

159. Gatenby RA, Gawlinski ET. A reaction-diffusion model of cancer invasion. *Cancer Res* 1996;56(24):5745-53. [published Online First: 1996/12/15]

160. Wellen KE, Thompson CB. A two-way street: reciprocal regulation of metabolism and signalling. *Nat Rev Mol Cell Biol* 2012;13(4):270-6. doi: 10.1038/nrm3305 [published Online First: 2012/03/08]

161. Wellen KE, Hatzivassiliou G, Sachdeva UM, et al. ATP-citrate lyase links cellular metabolism to histone acetylation. *Science* 2009;324(5930):1076-80. doi: 10.1126/science.1164097 [published Online First: 2009/05/23]

162. Currie E, Schulze A, Zechner R, et al. Cellular fatty acid metabolism and cancer. *Cell metabolism* 2013;18(2):153-61. doi: 10.1016/j.cmet.2013.05.017 [published Online First: 2013/06/25]

163. Santos AL, Preta G. Lipids in the cell: organisation regulates function. *Cellular and molecular life sciences : CMLS* 2018;75(11):1909-27. doi: 10.1007/s00018-018-2765-4 [published Online First: 2018/02/11]

164. Fahy E, Subramaniam S, Murphy RC, et al. Update of the LIPID MAPS comprehensive classification system for lipids. *Journal of lipid research* 2009;50 Suppl(Suppl):S9-14. doi: 10.1194/jlr.R800095-JLR200 [published Online First: 2008/12/23]

165. Yang Y, Lee M, Fairn GD. Phospholipid subcellular localization and dynamics. *The Journal of biological chemistry* 2018;293(17):6230-40. doi: 10.1074/jbc.R117.000582 [published Online First: 2018/03/29]

166. de Carvalho CCCR, Caramujo MJ. The Various Roles of Fatty Acids. *Molecules (Basel, Switzerland)* 2018;23(10) doi: 10.3390/molecules23102583 [published Online First: 2018/10/12]

167. van Meer G. Cellular lipidomics. *The EMBO journal* 2005;24(18):3159-65. doi: 10.1038/sj.emboj.7600798 [published Online First: 2005/09/03]

168. Pakiet A, Kobiela J, Stepnowski P, et al. Changes in lipids composition and metabolism in colorectal cancer: a review. *Lipids in health and disease* 2019;18(1):29. doi: 10.1186/s12944-019-0977-8 [published Online First: 2019/01/28]

169. Yan G, Li L, Zhu B, et al. Lipidome in colorectal cancer. *Oncotarget* 2016;7(22):33429-39. doi: 10.18632/oncotarget.7960.

170. Taylor LA, Arends J, Hodina AK, et al. Plasma lyso-phosphatidylcholine concentration is decreased in cancer patients with weight loss and activated inflammatory status. *Lipids Health Dis* 2007;6:17.(doi):10.1186/476-511X-6-17.

171. Kurabe N, Hayasaka T, Ogawa M, et al. Accumulated phosphatidylcholine (16:0/16:1) in human colorectal cancer; possible involvement of LPCAT4. *Cancer Sci* 2013;104(10):1295-302. doi: 10.111/cas.12221. Epub 2013 Jul 30.

172. Jung JH, Taniguchi K, Lee HM, et al. Comparative lipidomics of 5-Fluorouracil–sensitive and –resistant colorectal cancer cells reveals altered sphingomyelin and ceramide controlled by acid sphingomyelinase (SMPD1). *Scientific reports* 2020;10(1):6124. doi: 10.1038/s41598-020-62823-0

173. Gerbig S, Golf O, Balog J, et al. Analysis of colorectal adenocarcinoma tissue by desorption electrospray ionization mass spectrometric imaging. *Anal Bioanal Chem* 2012;403(8):2315-25. doi: 10.1007/s00216-012-5841-x. Epub 2012 Mar 25.

174. Balog J, Sasi-Szabó L, Kinross J, et al. Intraoperative tissue identification using rapid evaporative ionization mass spectrometry. *Sci Transl Med* 2013;5(194):194ra93. doi: 10.1126/scitranslmed.3005623.

175. Jiménez B, Mirnezami R, Kinross J, et al. 1H HR-MAS NMR spectroscopy of tumor-induced local metabolic "field-effects" enables colorectal cancer staging and prognostication. *J Proteome Res* 2013;12(2):959-68. doi: 10.1021/pr3010106. Epub 2013 Jan 16.

176. Mirnezami R, Jiménez B, Li JV, et al. Rapid diagnosis and staging of colorectal cancer via high-resolution magic angle spinning nuclear magnetic resonance (HR-MAS NMR) spectroscopy of intact tissue biopsies. *Ann Surg* 2014;259(6):1138-49. doi: 10.097/SLA.0b013e31829d5c45.

177. Mirnezami R, Spagou K, Vorkas PA, et al. Chemical mapping of the colorectal cancer microenvironment via MALDI imaging mass spectrometry (MALDI-MSI) reveals novel cancer-associated field effects. *Mol Oncol* 2014;8(1):39-49. doi: 10.1016/j.molonc.2013.08.010. Epub 13 Sep 14.

178. 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):1361-70. doi: 10.007/s00464-016-5121-5. Epub 2016 Aug 8.

179. Emwas A-HM. Metabonomics: Methods and Protocols. In: Bjerrum JT, ed.: Humana Press, New York, NY, 2015:161 - 93.

180. Lindon J. C. NJK, Holmes E. The Handbook of Metabonomics. 1 ed: Elsevier 2007.

181. Bjerrum JT. Metabonomics: Methods and Protocols. In: Bjerrum JT, ed. Metabonomics: Methods and Protocols: Humana Press, New York, NY, 2015:1 - 14.

182. Scigelova M, Hornshaw M, Giannakopulos A, et al. Fourier transform mass spectrometry. *Mol Cell Proteomics* 2011;10(7):M111.009431. doi: 10.1074/mcp.M111.31.

183. Marshall AGV, Francis R. Fourier Transforms in NMR, Optical, and Mass Spectrometry: Elsevier Science 1990.

184. Michalski A, Damoc E, Lange O, et al. Ultra high resolution linear ion trap Orbitrap mass spectrometer (Orbitrap Elite) facilitates top down LC MS/MS and versatile peptide fragmentation modes. *Mol Cell Proteomics* 2012;11(3):O111.013698. doi: 10.1074/mcp.O111.98. Epub 2011 Dec 9.

185. Zubarev RA, Makarov A. Orbitrap mass spectrometry. *Anal Chem* 2013;85(11):5288-96. doi: 10.1021/ac4001223. Epub 2013 May 13.

186. Ho CS, Lam CW, Chan MH, et al. Electrospray ionisation mass spectrometry: principles and clinical applications. *Clin Biochem Rev* 2003;24(1):3-12.

187. Bruins AP. Mechanistic aspects of electrospray ionization. *Journal of Chromatography A* 1998:345 - 57. doi: https://doi.org/10.1016/S0021-9673(97)01110-2

188. Ashton DS, Beddell CR, Green BN, et al. Rapid validation of molecular structures of biological samples by electrospray-mass spectrometry. *FEBS Lett* 1994;342(1):1-6. doi: 10.1016/0014-5793(94)80572-5.

189. Takáts Z, Wiseman JM, Gologan B, et al. Mass spectrometry sampling under ambient conditions with desorption electrospray ionization. *Science* 2004;306(5695):471-3. doi: 10.1126/science.1104404.

190. Abbassi-Ghadi NJ, E. A.; Gomez-Romero, M.; Golf, O.; Kumar, S.; Huang, J.; Kudo, H.; Goldin, R. D.; Hanna, G. B.; Takats, Z. . A Comparison of DESI-MS and LC-MS for the Lipidomic Profiling of Human Cancer Tissue. *J Am Soc Mass Spectrom* 2015 doi: 10.1007/s13361-015-1278-8

191. Tillner J, Wu V, Jones EA, et al. Faster, More Reproducible DESI-MS for Biological Tissue Imaging. *J Am Soc Mass Spectrom* 2017;28(10):2090-98. doi: 10.1007/s13361-017-1714-z. Epub 2017 Jun 15.

192. Abbassi-Ghadi N, Golf O, Kumar S, et al. Imaging of Esophageal Lymph Node Metastases by Desorption Electrospray Ionization Mass Spectrometry. *Cancer Res* 2016;76(19):5647-56. doi: 10.1158/0008-5472.CAN-16-0699. Epub 2016 Jun 30.

193. Veselkov KA, Mirnezami R, Strittmatter N, et al. Chemo-informatic strategy for imaging mass spectrometry-based hyperspectral profiling of lipid signatures in colorectal cancer. *Proc Natl Acad Sci U S A* 2014;111(3):1216-21. doi: 10.073/pnas.1310524111. Epub 2014 Jan 7.

194. Takats Z, Strittmatter N, McKenzie JS. Ambient Mass Spectrometry in Cancer Research. *Adv Cancer Res* 2017;134:231-256.(doi):10.1016/bs.acr.2016.11.011. Epub 16 Dec 29.

195. 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):3500-11. doi: 10.1039/c6sc03738k. Epub 2017 Feb 21.

196. Deininger SO, Cornett DS, Paape R, et al. Normalization in MALDI-TOF imaging datasets of proteins: practical considerations. *Anal Bioanal Chem* 2011;401(1):167-81. doi: 10.1007/s00216-011-4929-z. Epub 2011 Apr 12.

197. Marczyk M, Drazek G, Pietrowska M, et al. Modeling of Imaging Mass Spectrometry Data and Testing by Permutation for Biomarkers Discovery in Tissues. *Procedia Computer Science* 2015;51:693-702. doi: https://doi.org/10.1016/j.procs.2015.05.186

198. Takats Z, Strittmatter N, McKenzie JS. Advances in Cancer Research2017:231-256.

199. Kreis ME, Ruppert R, Ptok H, et al. Use of Preoperative Magnetic Resonance Imaging to Select Patients with Rectal Cancer for Neoadjuvant Chemoradiation--Interim Analysis of the German OCUM Trial (NCT01325649). *J Gastrointest Surg* 2016;20(1):25-32; discussion 32-3. doi: 10.1007/s11605-015-3011-0. Epub 2015 Nov 10.

200. Creavin B, Ryan E, Martin ST, et al. Organ preservation with local excision or active surveillance following chemoradiotherapy for rectal cancer. *Br J Cancer* 2016;20(10):417.

201. Habr-Gama A, Sabbaga J, Gama-Rodrigues J, et al. Watch and wait approach following extended neoadjuvant chemoradiation for distal rectal cancer: are we getting closer to anal cancer management? *Dis Colon Rectum* 2013;56(10):1109-17. doi: 10.097/DCR.0b013e3182a25c4e.

202. Hu P, Bader G Fau - Wigle DA, Wigle Da Fau - Emili A, et al. Computational prediction of cancer-gene function. *Nat Rev Cancer* 2007(1474-175X (Print))

203. AmiGO 2 - The Gene Ontology database [Website]. [Available from: http://amigo.geneontology.org/amigo.

204. Hu P, Bader G, Wigle DA, et al. Computational prediction of cancer-gene function. *Nat Rev Cancer* 2007;7(1):23-34. Epub 2006 Dec 14.

205. Perez RE, Shen H, Duan L, et al. Modeling the Etiology of p53-mutated Cancer Cells. *The Journal of biological chemistry* 2016;291(19):10131-47. doi: 10.1074/jbc.M116.724781 [published Online First: 2016/03/30]

206. Edden Y, Wexner SD, Berho M. The use of molecular markers as a method to predict the response to neoadjuvant therapy for advanced stage rectal adenocarcinoma. *Colorectal Disease* 2012;14(5):555-61.

207. Lopez-Crapez E, Bibeau F, Thezenas S, et al. p53 status and response to radiotherapy in rectal cancer: A prospective multilevel analysis. *British Journal of Cancer* 2005;92(12):2114-21.

208. Kobayashi H, Hashiguchi Y, Ueno H, et al. Absence of cyclooxygenase-2 protein expression is a predictor of tumor regression in rectal cancer treated with preoperative short-term chemoradiotherapy. *Diseases of the colon and rectum* 2007;50(9):1354-62.

209. Brophy S, Sheehan KM, McNamara DA, et al. GLUT-1 expression and response to chemoradiotherapy in rectal cancer. *Int J Cancer* 2009;125(12):2778-82. doi: 10.1002/ijc.24693

210. Sim SH, Kang MH, Kim YJ, et al. P21 and CD166 as predictive markers of poor response and outcome after fluorouracil-based chemoradiotherapy for the patients with rectal cancer. *BMC Cancer* 2014;14 (1) (no pagination)(241)

211. Garcia-Florez LJ, Gomez-Alvarez G, Frunza AM, et al. Predictive markers of response to neoadjuvant therapy in rectal cancer. *Journal of Surgical Research* 2015;194(1):120-26.

212. Hur H, Kim NK, Min BS, et al. Can a biomarker-based scoring system predict pathologic complete response after preoperative chemoradiotherapy for rectal cancer? *Diseases of the colon and rectum* 2014;57(5):592-601.

213. Kelley ST, Coppola D, Yeatman T, et al. Tumor response to neoadjuvant chemoradiation therapy for rectal adenocarcinoma is mediated by p53-dependent and caspase 8-dependent apoptotic pathways. *Clinical colorectal cancer* 2005;5(2):114-18.

214. Bertolini F, Bengala C, Losi L, et al. Prognostic and predictive value of baseline and posttreatment molecular marker expression in locally advanced rectal cancer treated with neoadjuvant chemoradiotherapy. *Int J Radiat Oncol Biol Phys* 2007;68(5):1455-61. doi: 10.1016/j.ijrobp.2007.02.018

215. Carnero A, Paramio JM. The PTEN/PI3K/AKT Pathway in vivo, cancer mouse models. *Frontiers in Oncology* 2014;4 doi: 10.3389/fonc.2014.00252

216. Farkas R, Pozsgai E, Schally AV, et al. Possible predictors of histopathological response to neoadjuvant chemoradiotherapy for rectal cancer. *Journal of Cancer Research and Clinical Oncology* 2012;138(3):387-95.

217. Gunther K, Dimmler A, Rodel F, et al. P27 does not predict histopathological response to radiochemotherapy in rectal cancer. *Journal of Surgical Research* 2003;113(2):179-88.

218. Carlomagno C, Pepe S, D'Armiento FP, et al. Predictive factors of complete response to neoadjuvant chemoradiotherapy in patients with rectal cancer. *Oncology* 2010;78(5-6):369-75.

219. Jakob C, Liersch T, Meyer W, et al. Predictive value of Ki67 and p53 in locally advanced rectal cancer: Correlation with thymidylate synthase and histopathological tumor regression after neoadjuvant 5-FU-based chemoradiotherapy. *World Journal of Gastroenterology* 2008;14(7):1060-66.

220. Yan H, Wang R, Yu J, et al. Predictive value of Smac, VEGF and Ki-67 in rectal cancer treated with neoadjuvant therapy. *Oncology Letters* 2010(1):641-47. doi: 10.3892/ol_00000113

221. Avoranta ST, Korkeila EA, Minn HRI, et al. Securin identifies a subgroup of patients with poor outcome in rectal cancer treated with long-course (chemo)radiotherapy. *Acta Oncologica* 2011;50(8):1158-66.

222. Bishnupuri KS, Luo Q, Sainathan SK, et al. Reg IV regulates normal intestinal and colorectal cancer cell susceptibility to radiation-induced apoptosis. *Gastroenterology* 2010;2010 Feb;138(2):616-26.

223. Matsumoto S, Konishi H, Maeda R, et al. Expression analysis of the regenerating gene (Reg) family members Reg-IIIbeta and Reg-IIIgamma in the mouse during development. *J Comp Neurol* 2012;520(3):479-94. doi: 10.1002/cne.22705.

224. Rafa L, Dessein AF, Devisme L, et al. REG4 acts as a mitogenic, motility and pro-invasive factor for colon cancer cells. *Int J Oncol* 2010;36(3):689-98.

225. Kobunai T, Watanabe T, Fukusato T. REG4, NEIL2, and BIRC5 gene expression correlates with gamma-radiation sensitivity in patients with rectal cancer receiving radiotherapy. *Anticancer Res* 2011;31(12):4147-53.

226. He HL, Lee YE, Shiue YL, et al. Overexpression of REG4 confers an independent negative prognosticator in rectal cancers receiving concurrent chemoradiotherapy. *Journal of Surgical Oncology* 2014;110(8):1002-10.

227. Kim PJ, Plescia J, Clevers H, et al. Survivin and molecular pathogenesis of colorectal cancer. *Lancet* 2003;362(9379):205-09. doi: 10.1016/S0140-6736(03)13910-4

228. McDowell DT, Smith FM, Reynolds JV, et al. Increased spontaneous apoptosis, but not survivin expression, is associated with histomorphologic response to neoadjuvant chemoradiation in rectal cancer. *Int J Colorectal Dis* 2009;24(11):1261-9. doi: 10.007/s00384-009-0755-6. Epub 2009 Jul 11.

229. Rodel F, Hoffmann J, Distel L, et al. Survivin as a radioresistance factor, and prognostic and therapeutic target for radiotherapy in rectal cancer. *Cancer Res* 2005;65(11):4881-7.

230. Huh JW, Lee JH, Kim HR. Pretreatment expression of 13 molecular markers as a predictor of tumor responses after neoadjuvant chemoradiation in rectal cancer. *Annals of Surgery* 2014;259(3):508-15.

231. Shao R. YKL-40 acts as an angiogenic factor to promote tumor angiogenesis. *Frontiers in Physiology* 2013;4 doi: 10.3389/fphys.2013.00122

232. Kawada M, Seno H, Kanda K, et al. Chitinase 3-like 1 promotes macrophage recruitment and angiogenesis in colorectal cancer. *Oncogene* 2012;31(26):3111-23. doi: 10.1038/onc.2011.498

233. Senetta R, Duregon E, Sonetto C, et al. YKL-40/c-met expression in rectal cancer biopsies predicts tumor regression following neoadjuvant chemoradiotherapy: A multi-institutional study. *PLoS ONE* 2015;10 (4) (no pagination)(e0123759)

234. Chang HJ, Jung KH, Kim DY, et al. Bax, a predictive marker for therapeutic response to preoperative chemoradiotherapy in patients with rectal carcinoma. *Human Pathology* 2005;36(4):364-71.

235. Shibahara K, Asano M, Ishida Y, et al. Isolation of a novel mouse gene MA-3 that is induced upon programmed cell death. *Gene* 1995;166(2):297-301. doi: http://dx.doi.org/10.1016/0378-1119(95)00607-9

236. Dou X, Wang R-B, Meng X-J, et al. PDCD4 as a Predictor of Sensitivity to Neoadjuvant Chemoradiotherapy in Locally Advanced Rectal Cancer Patients. *Asian Pacific Journal of Cancer Prevention* 2014;15(2):825-30. doi: 10.7314/apjcp.2014.15.2.825

237. Rodel C, Grabenbauer GG, Papadopoulos T, et al. Apoptosis as a cellular predictor for histopathologic response to neoadjuvant radiochemotherapy in patients with rectal cancer. *International Journal of Radiation Oncology Biology Physics* 2002;52(2):294-303.

238. Saigusa S, Inoue Y, Tanaka K, et al. Lack of M30 expression correlates with factors reflecting tumor progression in rectal cancer with preoperative chemoradiotherapy. *Molecular and Clinical Oncology* 2014;2(1):99-104.

239. Allal AS, Kahne T, Reverdin AK, et al. Radioresistance-related proteins in rectal cancer. *Proteomics* 2004;4(8):2261-9.

240. Cao WT, Zhou ZY, Deng YH, et al. [Clinical value of MR diffusion weighted imaging in prediction of pathological complete response of rectal cancer after neoadjuvant therapy]. *Zhonghua Wei Chang Wai Ke Za Zhi* 2013;16(12):1164-8.

241. He H, Liu D, Lin H, et al. Phosphatidylethanolamine binding protein 4 (PEBP4) is a secreted protein and has multiple functions. *Biochim Biophys Acta* 2016;1863(7 Pt A):1682-9. doi: 10.1016/j.bbamcr.2016.03.022

242. Qiu J, Yang G, Lin A, et al. Human phosphatidylethanolamine-binding protein 4 promoted the radioresistance of human rectal cancer by activating Akt in an ROS-dependent way. *PLoS One* 2014;9(3):e90062. doi: 10.1371/journal.pone.0090062. eCollection 2014.

243. Moeller BJ, Dewhirst MW. HIF-1 and tumour radiosensitivity. *Br J Cancer* 2006;95(1):1-5. Epub 2006 May 30.

244. Moeller BJ, Richardson RA, Dewhirst MW. Hypoxia and radiotherapy: opportunities for improved outcomes in cancer treatment. *Cancer and Metastasis Reviews* 2007;26(2):241-48. doi: 10.1007/s10555-007-9056-0

245. Toiyama Y, Inoue Y, Saigusa S, et al. Gene expression profiles of epidermal growth factor receptor, vascular endothelial growth factor and hypoxia-inducible factor-1 with special reference to local responsiveness to neoadjuvant chemoradiotherapy and disease recurrence after rectal cancer surgery. *Clinical Oncology* 2010;22(4):272-80.

246. Guedj N, Bretagnol F, Rautou PE, et al. Predictors of tumor response after preoperative chemoradiotherapy for rectal adenocarcinomas. *Human Pathology* 2011;42(11):1702-09.

247. Airley RE, Loncaster J, Raleigh JA, et al. GLUT-1 and CAIX as intrinsic markers of hypoxia in carcinoma of the cervix: relationship to pimonidazole binding. *Int J Cancer* 2003;104(1):85-91.

248. Hoskin PJ, Sibtain A, Daley FM, et al. GLUT1 and CAIX as intrinsic markers of hypoxia in bladder cancer: relationship with vascularity and proliferation as predictors of outcome of ARCON. *Br J Cancer* 2003;89(7):1290-7.

249. Moussata D, Amara S, Siddeek B, et al. XIAP as a radioresistance factor and prognostic marker for radiotherapy in human rectal adenocarcinoma. *Am J Pathol* 2012;181(4):1271-8. doi: 10.016/j.ajpath.2012.06.029. Epub 12 Aug 4.

250. Cecchin E, Agostini M, Pucciarelli S, et al. Tumor response is predicted by patient genetic profile in rectal cancer patients treated with neo-adjuvant chemo-radiotherapy. *Pharmacogenomics Journal* 2011;11(3):214-26.

251. Terrazzino S, Agostini M, Pucciarelli S, et al. A haplotype of the methylenetetrahydrofolate reductase gene predicts poor tumor response in rectal cancer patients receiving preoperative chemoradiation. *Pharmacogenetics and Genomics* 2006;16(11):817-24.

252. Qin CJ, Song XM, Chen ZH, et al. XRCC2 as a predictive biomarker for radioresistance in locally advanced rectal cancer patients undergoing preoperative radiotherapy. *Oncotarget* 2015;6(31):32193-204. doi: 10.18632/oncotarget.4975.

253. Agostini M, Zangrando A, Pastrello C, et al. A functional biological network centered on XRCC3: a new possible marker of chemoradiotherapy resistance in rectal cancer patients. *Cancer Biol Ther* 2015;16(8):1160-71. doi: 10.1080/15384047.2015.1046652

254. Yan H, Yu J, Wang R, et al. Prognostic value of Smac expression in rectal cancer patients treated with neoadjuvant therapy. *Medical Oncology* 2012;29(1):168-73.

255. He HL, Lee YE, Chen HP, et al. Overexpression of DNAJC12 predicts poor response to neoadjuvant concurrent chemoradiotherapy in patients with rectal cancer. *Experimental and Molecular Pathology* 2015;98(3):338-45.

256. Min BS, Choi YJ, Pyo HR, et al. Cyclooxygenase-2 expression in pretreatment biopsy as a predictor of tumor responses after preoperative chemoradiation in rectal cancer. *Arch Surg* 2008;143(11):1091-7; discussion 97. doi: 10.1001/archsurg.143.11.1091

257. Shinto E, Hashiguchi Y, Ueno H, et al. Pretreatment CD133 and cyclooxygenase-2 expression as the predictive markers of the pathological effect of chemoradiotherapy in rectal cancer patients. *Diseases of the colon and rectum* 2011;54(9):1098-106.

258. Meng X, Wang R, Huang Z, et al. Human epidermal growth factor receptor-2 expression in locally advanced rectal cancer: Association with response to neoadjuvant therapy and prognosis. *Cancer Science* 2014;105(7):818-24.

259. Karin N, Wildbaum G, Thelen M. Biased signaling pathways via CXCR3 control the development and function of CD4+ T cell subsets. *J Leukoc Biol* 2016;99(6):857-62. doi: 10.1189/jlb.2MR0915-441R. Epub 2015 Dec 9.

260. Li C, Wang Z, Liu F, et al. CXCL10 mRNA expression predicts response to neoadjuvant chemoradiotherapy in rectal cancer patients. *Tumour Biol* 2014;35(10):9683-91. doi: 10.1007/s13277-014-2234-0

261. Lin CY, Chang IW, Sheu MJ, et al. Low thrombospondin 2 expression is predictive of low tumor regression after neoadjuvant chemoradiotherapy in rectal cancer. *American Journal of Translational Research* 2015;7(11):2423-32.

262. Drebber U, Madeja M, Odenthal M, et al. beta-catenin and Her2/neu expression in rectal cancer: Association with histomorphological response to neoadjuvant therapy and prognosis. *International Journal of Colorectal Disease* 2011;26(9):1127-34.

263. Jameson KL, Mazur PK, Zehnder AM, et al. IQGAP1 scaffold-kinase interaction blockade selectively targets RAS-MAP kinase-driven tumors. *Nat Med* 2013;19(5):626-30. doi: 10.1038/nm.3165. Epub 2013 Apr 21.

264. Munshi A, Ramesh R. Mitogen-Activated Protein Kinases and Their Role in Radiation Response. *Genes Cancer* 2013;4(9-10):401-8. doi:10.1177/1947601913485414.

265. Holck S, Nielsen HJ, Hammer E, et al. IQGAP1 in rectal adenocarcinomas: localization and protein expression before and after radiochemotherapy. *Cancer letters* 2015;356(2 Pt B):556-60.

266. Taketo MM, Takaku K. Gastrointestinal tumorigenesis in Smad4 (Dpc4) mutant mice. *Hum Cell* 2000;13(3):85-95.

267. Huang MY, Lin CH, Huang CM, et al. Relationships between SMAD3 expression and preoperative fluoropyrimidine-based chemoradiotherapy response in locally advanced rectal cancer patients. *World journal of surgery* 2015;39(5):1257-67.

268. Wang L, Zhang XM, Li Z, et al. Overexpression of nuclear beta-catenin in rectal adenocarcinoma is associated with radioresistance. *World J Gastroenterol* 2013;19(40):6876-82. doi: 10.3748/wjg.v19.i40.6876

269. Dou X, Wang R, Meng X, et al. The prognostic role of TCF4 expression in locally advanced rectal cancer patients treated with neoadjuvant chemoradiotherapy. *Cancer Biomarkers* 2015;15(2):187-94.

270. Bengala C, Bettelli S, Bertolini F, et al. Epidermal growth factor receptor gene copy number, K-ras mutation and pathological response to preoperative cetuximab, 5-FU and radiation therapy in locally advanced rectal cancer. *Annals of Oncology* 2009;20(3):469-74.

271. Yu ZQ, Zhang C, Wang H, et al. Downregulation of ATP-binding cassette subfamily c member 4 increases sensitivity to neoadjuvant radiotherapy for locally advanced rectal carcinoma. *Diseases of the Colon and Rectum* 2013;56(5):600-08.

272. Yu Z, Zhang C, Wang H, et al. Multidrug resistance-associated protein 3 confers resistance to chemoradiotherapy for rectal cancer by regulating reactive oxygen species and caspase-3-dependent apoptotic pathway. *Cancer Lett* 2014;353(2):182-93. doi: 10.1016/j.canlet.2014.07.025. Epub 14 Aug 1.

273. Schally AV, Varga JL, Engel JB. Antagonists of growth-hormone-releasing hormone: an emerging new therapy for cancer. *Nat Clin Pract Endocrinol Metab* 2008;4(1):33-43.

274. Rick FG, Seitz S, Schally AV, et al. GHRH antagonist when combined with cytotoxic agents induces S-phase arrest and additive growth inhibition of human colon cancer. *Cell Cycle* 2012;11(22):4203-10. doi:10.161/cc.22498.

275. Saw RPM, Morgan M, Koorey D, et al. p53, deleted in colorectal cancer gene, and thymidylate synthase as predictors of histopathologic response and survival in low, locally advanced rectal cancer treated with preoperative adjuvant therapy. *Diseases of the Colon and Rectum* 2003;46(2):192-202.

276. Jakob C, Aust DE, Meyer W, et al. Thymidylate synthase, thymidine phosphorylase, dihydropyrimidine dehydrogenase expression, and histological tumour regression after 5-FU-based neo-adjuvant chemoradiotherapy in rectal cancer. *Journal of Pathology* 2004;204(5):562-68.

277. Jakob C, Liersch T, Meyer W, et al. Immunohistochemical analysis of thymidylate synthase, thymidine phosphorylase, and dihydropyrimidine dehydrogenase in rectal cancer (cUICC II/III): Correlation with histopathologic tumor regression after 5-fluorouracil-based long-term neoadjuvant chemorodiotheropy. *American Journal of Surgical Pathology* 2005;29(10):1304-09.

278. Conradi LC, Bleckmann A, Schirmer M, et al. Thymidylate synthase as a prognostic biomarker for locally advanced rectal cancer after multimodal treatment. *Annals of Surgical Oncology* 2011;18(9):2442-52.

279. Lee YY, Li CF, Lin CY, et al. Overexpression of CPS1 is an independent negative prognosticator in rectal cancers receiving concurrent chemoradiotherapy. *Tumour Biology* 2014;35(11):11097-105.

280. Balasubramanian MN, Butterworth EA, Kilberg MS. Asparagine synthetase: regulation by cell stress and involvement in tumor biology. *Am J Physiol Endocrinol Metab* 2013;304(8):E789-99. doi: 10.1152/ajpendo.00015.2013. Epub 2013 Feb 12.

281. Lin CY, Sheu MJ, Li CF, et al. Deficiency in asparagine synthetase expression in rectal cancers receiving concurrent chemoradiotherapy: Negative prognostic impact and therapeutic relevance. *Tumor Biology* 2014;35(7):6823-30.

282. Molinari C, Casadio V, Foca F, et al. Gene methylation in rectal cancer: Predictive marker of response to chemoradiotherapy? *Journal of Cellular Physiology* 2013;228(12):2343-49.

283. Lee YE, He HL, Shiue YL, et al. The prognostic impact of lipid biosynthesis-associated markers, HSD17B2 and HMGCS2, in rectal cancer treated with neoadjuvant concurrent chemoradiotherapy. *Tumor Biology* 2015;36(10):7675-83.

284. Labrie F, Luu-The V, Lin S, et al. Role of 17beta-hydroxysteroid dehydrogenases in sex steroid formation in peripheral intracrine tissues. *Trends in Endocrinology & Metabolism* 2000;11(10):421-27. doi: http://dx.doi.org/10.1016/S1043-2760(00)00342-8

285. Liberti MV, Locasale JW. The Warburg Effect: How Does it Benefit Cancer Cells? *Trends Biochem Sci* 2016;41(3):211-8. doi: 10.1016/j.tibs.2015.12.001. Epub 16 Jan 5.

286. Yeo SG, Kim DY, Kim KH, et al. Hydroxymethylglutaryl-coenzyme a synthase 2 expression is associated with chemoradiotherapy responses in colorectal cancer. *Diseases of the colon and rectum* 2012;55(6):686-94.

287. Tuncer S, Banerjee S. Eicosanoid pathway in colorectal cancer: Recent updates. *World J Gastroenterol* 2015;21(41):11748-66. doi: 10.3748/wjg.v21.i41.48.

288. Yarla NS, Bishayee A, Sethi G, et al. Targeting arachidonic acid pathway by natural products for cancer prevention and therapy. *Semin Cancer Biol* 2016;40-41:48-81.(doi):10.1016/j.semcancer.2016.02.001. Epub 16 Feb 4.

289. Smith FM, Reynolds JV, Kay EW, et al. COX-2 overexpression in pretreatment biopsies predicts response of rectal cancers to neoadjuvant radiochemotherapy. *International Journal of Radiation Oncology, Biology, Physics* 2006;64(2):466-72.

290. Bouzourene H, Yan P, Sandmeier D, et al. The role of COX-2 in rectal cancer treated with preoperative radiotherapy. *Virchows Archiv* 2008;452(5):499-505. doi: 10.1007/s00428-008-0606-9

291. Di Maggio FM, Minafra L, Forte GI, et al. Portrait of inflammatory response to ionizing radiation treatment. *Journal of Inflammation* 2015;12(1):14. doi: 10.1186/s12950-015-0058-3

292. Berardi R, Maccaroni E, Mandolesi A, et al. Nuclear factor-kappaB predicts outcome in locally advanced rectal cancer patients receiving neoadjuvant radio-chemotherapy. *Digestive and Liver Disease* 2012;44(7):617-22.

293. Sheu MJ, Li CF, Lin CY, et al. Overexpression of ANXA1 confers independent negative prognostic impact in rectal cancers receiving concurrent chemoradiotherapy. *Tumour Biol* 2014;35(8):7755-63. doi: 10.1007/s13277-014-2032-8

294. Zhang Z, Huang L Fau - Zhao W, Zhao W Fau - Rigas B, et al. Annexin 1 induced by anti-inflammatory drugs binds to NF-kappaB and inhibits its activation: anticancer effects in vitro and in vivo. (1538-7445 (Electronic)) doi: D - NLM: NIHMS172356

D - NLM: PMC2953961 EDAT- 2010/03/11 06:00 MHDA- 2010/04/27 06:00 CRDT- 2010/03/11 06:00 AID - 0008-5472.CAN-09-4204 [pii] AID - 10.1158/0008-5472.CAN-09-4204 [doi] PST - ppublish

295. Huh J, Lim S, Kim H, et al. Pretreatment 13 molecular markers expression as a predictor of tumor responses after neoadjuvant chemoradiation in rectal cancer. *Diseases of the colon and rectum* 2011;Conference:Annual Meeting of the American Society of Colon and Rectal Surgeons, ACSRS 2011 Vancouver, BC Canada. Conference Start: 20110514 Conference End: 18. Conference Publication: (var.pagings). 54 (5) (pp e163-e164).

296. Kurt A, Yanar F, Asoglu O, et al. Low Mmp 9 and VEGF levels predict good oncologic outcome in mid and low rectal cancer patients with neoadjuvant chemoradiation. *BMC Clinical Pathology* 2012;12 (no pagination)(27)

297. Zlobec I, Steele R, Compton CC. VEGF as a predictive marker of rectal tumor response to preoperative radiotherapy. *Cancer* 2005;104(11):2517-21.

298. Yu Z, Zhang C, Chai R, et al. Prognostic significance and molecular mechanism of ATP-binding cassette subfamily C member 4 in resistance to neoadjuvant radiotherapy of locally advanced rectal carcinoma. *PLoS One* 2014;9(1):e85446. doi: 10.1371/journal.pone.0085446. eCollection 2014.

299. Avoranta ST, Korkeila EA, Ristamaki RH, et al. ALDH1 expression indicates chemotherapy resistance and poor outcome in node-negative rectal cancer. *Hum Pathol* 2013;44(6):966-74. doi: 10.1016/j.humpath.2012.10.003. Epub 13 Jan 17.

300. Dworak O, Keilholz L, Hoffmann A. Pathological features of rectal cancer after preoperative radiochemotherapy. *Int J Colorectal Dis* 1997;12(1):19-23.

301. Mandard AM, Dalibard F, Mandard JC, et al. Pathologic assessment of tumor regression after preoperative chemoradiotherapy of esophageal carcinoma. Clinicopathologic correlations. *Cancer* 1994;73(11):2680-6.

302. Rodel C, Martus P, Papadoupolos T, et al. Prognostic significance of tumor regression after preoperative chemoradiotherapy for rectal cancer. *Journal of Clinical Oncology* 2005;23(34):8688-96.

303. Ryan R, Gibbons D, Hyland JMP, et al. Pathological response following long-course neoadjuvant chemoradiotherapy for locally advanced rectal cancer. *Histopathology* 2005;47(2):141-46.

304. Wheeler JM, Warren BF, Mortensen NJ, et al. Quantification of histologic regression of rectal cancer after irradiation: a proposal for a modified staging system. *Diseases of the colon and rectum* 2002;45(8):1051-6.

305. Page G.P. ZSO, Kim K., Mehta T., Chen L., Zhang K. . Topics in Biostatistics. Methods in Molecular Biology. In: W.T. A, ed. Topics in Biostatistics Methods in Molecular Biology: Humana Press 2007.

306. Mootha VK, Lindgren CM, Eriksson K-F, et al. PGC-1α-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. doi: 10.1038/ng1180

307. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 2005;102(43):15545-50. doi: 10.1073/pnas.0506580102. Epub 2005 Sep 30.

308. Broad Institute UoSD, USA. Gene Set Enrichment Analysis [Available from: https://www.gsea-msigdb.org/gsea/index.jsp2017.

309. Kyoto Encyclopaedia of Genes and Genomes (KEGG) pathway database [Available from: https://www.genome.jp/kegg/pathway.html accessed 1st July 2017.

310. Gene Expression Omnibus, National Centre for Biotechnology Information 2002 [cited 2020. Available from: https://www.ncbi.nlm.nih.gov/geo/2017.

311. Snipstad K, Fenton CG, Kjæve J, et al. New specific molecular targets for radio-chemotherapy of rectal cancer. *Molecular Oncology* 2010;4(1):52-64. doi: https://doi.org/10.1016/j.molonc.2009.11.002

312. Paulssen R.H. FCG, Snipstad K., Kjæve J., Anderssen E. New specific molecular targets for radiochemotherapy in colorectal cancer. Gene Expression Omnibus (GEO), 2009.

313. Zyla J, Marczyk M, Weiner J, et al. Ranking metrics in gene set enrichment analysis: do they matter? *BMC Bioinformatics* 2017;18 doi: 10.1186/s12859-017-1674-0

314. The Human Protein Atlas [Available from: https://www.proteinatlas.org accessed January 2021.

315. LIPID MAPS [Available from: https://www.lipidmaps.org/2018.

316. Cronan JE. Bacterial membrane lipids: where do we stand? *Annual Review of Microbiology* 2003;57:203-24. doi: 10.1146/annurev.micro.57.030502.090851

317. Dugail I, Kayser BD, Lhomme M. Specific roles of phosphatidylglycerols in hosts and microbes. *Biochimie* 2017;141:47-53. doi: https://doi.org/10.1016/j.biochi.2017.05.005

318. Perez RO, Habr-Gama A, São Julião GP, et al. Should We Give Up The Search for a Clinically Useful Gene Signature for the Prediction of Response of Rectal Cancer to Neoadjuvant Chemoradiation? *Dis Colon Rectum* 2016;59(9):895-7. doi: 10.1097/DCR.0000000000000620.

319. Ryan R, Gibbons D, Hyland JM, et al. Pathological response following long-course neoadjuvant chemoradiotherapy for locally advanced rectal cancer. *Histopathology* 2005;47(2):141-6. doi: 10.1111/j.365-2559.005.02176.x.

320. Oetjen J, Veselkov K, Watrous J, et al. Benchmark datasets for 3D MALDI- and DESI-imaging mass spectrometry. *Gigascience* 2015;4:20.(doi):10.1186/s13742-015-0059-4. eCollection 2015.

321. Institute S. METLIN [Available from: https://metlin.scripps.edu/landing_page.php?pgcontent=mainPage2018.

322. Mirnezami R, Strittmatter N, Veselkov K, et al. DESI-MSI profiling of the colorectal cancer microenvironment reveals subtle modifications in lipid metabolism that correlate with pathological phenotype: SP068. *Colorectal Disease July* 2014;2:21.

323. Zalba S, Ten Hagen TL. Cell membrane modulation as adjuvant in cancer therapy. *Cancer Treat Rev* 2017;52:48-57.(doi):10.1016/j.ctrv.2016.10.008. Epub 16 Nov 9.

324. 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):3618-25. doi: 10.1007/s00464-019-7140-y. Epub 2019 Oct 15.

325. Veselkov KA, Mirnezami R, Strittmatter N, et al. Chemo-informatic strategy for imaging mass spectrometry-based hyperspectral profiling of lipid signatures in colorectal cancer. *Proc Natl Acad Sci U S A* 2014;111(3):1216-21. doi: 10.1073/pnas.1310524111

326. Jiang X-C. Lipid Homeostasis on Cell Membrane. doi: 10.1007/978-981-13-9077-7_1

327. Schenkel LC, Bakovic M. Formation and regulation of mitochondrial membranes. *Int J Cell Biol* 2014;2014:709828.(doi):10.1155/2014/709828. Epub 2014 Jan 22.

328. Kennedy EP, Weiss SB. The function of cytidine coenzymes in the biosynthesis of phospholipides. *J Biol Chem* 1956;222(1):193-214.

329. Tsukahara T. PPAR γ Networks in Cell Signaling: Update and Impact of Cyclic Phosphatidic Acid. *J Lipids* 2013;2013:246597.(doi):10.1155/2013/246597. Epub 2013 Feb 7.

330. Davis HW, Vallabhapurapu SD, Chu Z, et al. Enhanced phosphatidylserine-selective cancer therapy with irradiation and SapC-DOPS nanovesicles. *Oncotarget* 2019;10(8):856-68. doi: 10.18632/oncotarget.26615. eCollection 2019 Jan 25.

331. Huynh ML, Fadok VA, Henson PM. Phosphatidylserine-dependent ingestion of apoptotic cells promotes TGF-beta1 secretion and the resolution of inflammation. *J Clin Invest* 2002;109(1):41-50. doi: 10.1172/JCI11638.

332. Ran S, Downes A, Thorpe PE. Increased exposure of anionic phospholipids on the surface of tumor blood vessels. *Cancer Res* 2002;62(21):6132-40.

333. Ran S, Thorpe PE. Phosphatidylserine is a marker of tumor vasculature and a potential target for cancer imaging and therapy. *Int J Radiat Oncol Biol Phys* 2002;54(5):1479-84. doi: 10.016/s0360-3016(02)03928-7.

334. Desai TJ, Toombs JE, Minna JD, et al. Identification of lipid-phosphatidylserine (PS) as the target of unbiasedly selected cancer specific peptide-peptoid hybrid PPS1. *Oncotarget* 2016;7(21):30678-90. doi: 10.18632/oncotarget.8929.

335. Calzada E, Onguka O, Claypool SM. Phosphatidylethanolamine Metabolism in Health and Disease. *Int Rev Cell Mol Biol* 2016;321:29-88.(doi):10.1016/bs.ircmb.2015.10.001. Epub 15 Oct 31.

336. Steenbergen R, Nanowski TS, Beigneux A, et al. Disruption of the phosphatidylserine decarboxylase gene in mice causes embryonic lethality and mitochondrial defects. *J Biol Chem* 2005;280(48):40032-40. doi: 10.1074/jbc.M506510200. Epub 2005 Sep 28.

337. Becker T, Horvath SE, Böttinger L, et al. Role of phosphatidylethanolamine in the biogenesis of mitochondrial outer membrane proteins. *J Biol Chem* 2013;288(23):16451-9. doi: 10.1074/jbc.M112.442392. Epub 2013 Apr 26.

338. Dowhan W, Bogdanov M. Lipid-dependent membrane protein topogenesis. *Annu Rev Biochem* 2009;78:515-40.(doi):10.1146/annurev.biochem.77.060806.91251.

339. Beckonert O, Monnerjahn J, Bonk U, et al. Visualizing metabolic changes in breast-cancer tissue using 1H-NMR spectroscopy and self-organizing maps. *NMR Biomed* 2003;16(1):1-11. doi: 10.1002/nbm.797.

340. Podo F. Tumour phospholipid metabolism. *NMR Biomed* 1999;12(7):413-39. doi: 10.1002/(sici)99-492(199911)12:7<413::aid-nbm587>3.0.co;2-u.

341. Shen S, Yang L, Li L, et al. A plasma lipidomics strategy reveals perturbed lipid metabolic pathways and potential lipid biomarkers of human colorectal cancer. *Journal of Chromatography B*;1068-1069:41-48. doi: https://doi.org/10.1016/j.jchromb.2017.10.004

342. Abbassi-Ghadi N, Antonowicz SS, McKenzie JS, et al. De Novo Lipogenesis Alters the Phospholipidome of Esophageal Adenocarcinoma. *Cancer Res* 2020;80(13):2764-74. doi: 10.1158/0008-5472.CAN-19-4035. Epub 2020 Apr 28.

343. Jump DB. Mammalian fatty acid elongases. *Methods Mol Biol* 2009;579:375-89.(doi):10.1007/978-1-60761-322-0_19.

344. Yamashita Y, Nishiumi S, Kono S, et al. Differences in elongation of very long chain fatty acids and fatty acid metabolism between triple-negative and hormone receptor-positive breast cancer. *BMC Cancer* 2017;17(1):589. doi: 10.1186/s12885-017-3554-4.

345. Ibarguren M, López DJ, Escribá PV. The effect of natural and synthetic fatty acids on membrane structure, microdomain organization, cellular functions and human health. *Biochimica et Biophysica Acta (BBA) - Biomembranes* 2014;1838(6):1518-28. doi: https://doi.org/10.1016/j.bbamem.2013.12.021

346. Lucarelli G, Ferro M, Loizzo D, et al. Integration of Lipidomics and Transcriptomics Reveals Reprogramming of the Lipid Metabolism and Composition in Clear Cell Renal Cell Carcinoma. *Metabolites* 2020;10(12):509. doi: 10.3390/metabo10120509.

347. Lee JY, Nam M, Son HY, et al. Polyunsaturated fatty acid biosynthesis pathway determines ferroptosis sensitivity in gastric cancer. *Proc Natl Acad Sci U S A* 2020;117(51):32433-42. doi: 10.1073/pnas.2006828117. Epub 2020 Dec 7.

348. Centenera MM, Scott JS, Machiels J, et al. ELOVL5 is a critical and targetable fatty acid elongase in prostate cancer. *Cancer Res* 2021;5(10):0008-5472.

349. Fernandes A, Messias MCF, Duarte GHB, et al. Plasma Lipid Profile Reveals Plasmalogens as Potential Biomarkers for Colon Cancer Screening. *Metabolites* 2020;10(6):262. doi: 10.3390/metabo10060262.

350. Benjamin DI, Cozzo A, Ji X, et al. Ether lipid generating enzyme AGPS alters the balance of structural and signaling lipids to fuel cancer pathogenicity. *Proc Natl Acad Sci U S A* 2013;110(37):14912-17. doi: 10.1073/pnas.1310894110

351. Ivanova PT, Cerda BA, Horn DM, et al. Electrospray ionization mass spectrometry analysis of changes in phospholipids in RBL-2H3 mastocytoma cells during degranulation. *Proc Natl Acad Sci U S A* 2001;98(13):7152-7. doi: 10.1073/pnas.131195098.

352. Furse S, de Kroon AI. Phosphatidylcholine's functions beyond that of a membrane brick. *Mol Membr Biol* 2015;32(4):117-9. doi: 10.3109/09687688.2015.1066894. Epub 2015 Aug 25.

353. van Beers EH, Joosse SA, Ligtenberg MJ, et al. A multiplex PCR predictor for aCGH success of FFPE samples. *Br J Cancer* 2006;94(2):333-7. doi: 10.1038/sj.bjc.6602889.

354. Scheinin I, Sie D, Bengtsson H, et al. DNA copy number analysis of fresh and formalin-fixed specimens by shallow whole-genome sequencing with identification and exclusion of problematic regions in the genome assembly. *Genome Res* 2014;24(12):2022-32. doi: 10.1101/gr.175141.114. Epub 2014 Sep 18.

355. Gentleman RC, Carey VJ, Bates DM, et al. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* 2004;5(10):R80. doi: 10.1186/gb-2004-5-10-r80. Epub 04 Sep 15.

356. An N, Yang X, Cheng S, et al. Developmental genes significantly afflicted by aberrant promoter methylation and somatic mutation predict overall survival of late-stage colorectal cancer. *Scientific reports* 2015;5 doi: 10.1038/srep18616

357. Müller MF, Ibrahim AE, Arends MJ. Molecular pathological classification of colorectal cancer. *Virchows Arch* 2016;469(2):125-34. doi: 10.1007/s00428-016-1956-3. Epub 2016 Jun 20.

358. Lee MS, Menter DG, Kopetz S. Right Versus Left Colon Cancer Biology: Integrating the Consensus Molecular Subtypes. *J Natl Compr Canc Netw* 2017;15(3):411-19. doi: 10.6004/jnccn.2017.0038.

359. Bae JM, Kim JH, Kang GH. Molecular Subtypes of Colorectal Cancer and Their Clinicopathologic Features, With an Emphasis on the Serrated Neoplasia Pathway. *Arch Pathol Lab Med* 2016;140(5):406-12. doi: 10.5858/arpa.2015-0310-RA.

360. Toomey S, Gunther J, Carr A, et al. Genomic and Transcriptomic Characterisation of Response to Neoadjuvant Chemoradiotherapy in Locally Advanced Rectal Cancer. *Cancers (Basel)* 2020;12(7):1808. doi: 10.3390/cancers12071808.

361. Kamran SC, Lennerz JK, Margolis CA, et al. Integrative Molecular Characterization of Resistance to Neoadjuvant Chemoradiation in Rectal Cancer. *Clin Cancer Res* 2019;25(18):5561-71. doi: 10.1158/078-0432.CCR-19-908. Epub 2019 Jun 28.

362. Alonso MH, Aussó S, Lopez-Doriga A, et al. Comprehensive analysis of copy number aberrations in microsatellite stable colon cancer in view of stromal component. *British Journal of Cancer* 2017;117:421-31. doi: 10.1038/bjc.2017.208

363. Goossens-Beumer IJ, Oosting J, Corver WE, et al. Copy number alterations and allelic ratio in relation to recurrence of rectal cancer. *BMC Genomics* 2015;16 doi: 10.1186/s12864-015-1550-0

364. Ashktorab H, Schäffer AA, Daremipouran M, et al. Distinct genetic alterations in colorectal cancer. *PLoS One* 2010;5(1):e8879. doi: 10.1371/journal.pone.0008879.

365. Wang H, Liang L, Fang JY, et al. Somatic gene copy number alterations in colorectal cancer: new quest for cancer drivers and biomarkers. *Oncogene* 2016;21:2011-19. doi: 10.1038/onc.2015.304

366. Sillars-Hardebol AH, Carvalho B, Tijssen M, et al. TPX2 and AURKA promote 20q amplicon-driven colorectal adenoma to carcinoma progression. *Gut* 2012;61(11):1568-75. doi: 10.136/gutjnl-2011-301153. Epub 2011 Dec 29.

367. Sillars-Hardebol AH, Carvalho B, Beliën JA, et al. BCL2L1 has a functional role in colorectal cancer and its protein expression is associated with chromosome 20q gain. *J Pathol* 2012;226(3):442-50. doi: 10.1002/path.2983. Epub 011 Oct 18.

368. Chen Z, Liu Z, Li W, et al. Chromosomal copy number alterations are associated with tumor response to chemoradiation in locally advanced rectal cancer. *Genes Chromosomes Cancer* 2011;50(9):689-99. doi: 10.1002/gcc.20891. Epub 2011 May 16.

369. Buck A, Heijs B, Beine B, et al. Round robin study of formalin-fixed paraffin-embedded tissues in mass spectrometry imaging. *Anal Bioanal Chem* 2018;410(23):5969-80. doi: 10.1007/s00216-018-1216-2. Epub 2018 Jul 3.

370. Inglese P, Correia G, Pruski P, et al. Colocalization Features for Classification of Tumors Using Desorption Electrospray Ionization Mass Spectrometry Imaging. *Anal Chem* 2019;91(10):6530-40. doi: 10.1021/acs.analchem.8b05598. Epub 2019 May 1.

371. Mas S, Torro A, Fernández L, et al. MALDI imaging mass spectrometry and chemometric tools to discriminate highly similar colorectal cancer tissues. *Talanta* 2020;208:120455.(doi):10.1016/j.talanta.2019.120455. Epub 2019 Oct 10.

372. Wang Y, Hinz S, Uckermann O, et al. Shotgun lipidomics-based characterization of the landscape of lipid metabolism in colorectal cancer. *Biochim Biophys Acta Mol Cell Biol Lipids* 2020;1865(3):158579. doi: 10.1016/j.bbalip.2019.79. Epub 2019 Nov 30.

373. Nagahashi M, Shimada Y, Ichikawa H, et al. Formalin-fixed paraffin-embedded sample conditions for deep next generation sequencing. *J Surg Res* 2017;220:125-132.(doi):10.1016/j.jss.2017.06.077. Epub 17 Jul 27.

374. McDonough SJ, Bhagwate A, Sun Z, et al. Use of FFPE-derived DNA in next generation sequencing: DNA extraction methods. *PLoS One* 2019;14(4):e0211400. doi: 10.1371/journal.pone.00. eCollection 2019.

375. Hedegaard J, Thorsen K, Lund MK, et al. Next-generation sequencing of RNA and DNA isolated from paired fresh-frozen and formalin-fixed paraffin-embedded samples of human cancer and normal tissue. *PLoS One* 2014;9(5):e98187. doi: 10.1371/journal.pone.0098187. eCollection 2014.

376. Matos LL, Trufelli DC, de Matos MG, et al. Immunohistochemistry as an important tool in biomarkers detection and clinical practice. *Biomark Insights* 2010;5:9-20.(doi):10.4137/bmi.s2185.

377. Hartmann FJ, Simonds EF, Vivanco N, et al. Scalable Conjugation and Characterization of Immunoglobulins with Stable Mass Isotope Reporters for Single-Cell Mass Cytometry Analysis. *Methods Mol Biol* 2019;1989:55-81.(doi):10.1007/978-1-4939-9454-0_5.

378. Canada F. Hyperion Imagain System - User Guide 2019 [Available from: https://www.fluidigm.com/binaries/content/documents/fluidigm/resources/hyperion-imaging-system-ug-400311-a2/hyperion-imaging-system-ug-400311-a2/fluidigm%3Afile accessed February 2021.

379. Viaud S, Daillère R, Boneca IG, et al. Gut microbiome and anticancer immune response: really hot Sh t! *Cell Death Differ* 2015;22(2):199-214. doi: 10.1038/cdd.2014.56. Epub 14 May 16.

380. Knudson AG, Jr. Mutation and cancer: statistical study of retinoblastoma. *Proc Natl Acad Sci U S A* 1971;68(4):820-3. doi: 10.1073/pnas.68.4.820.

381. Alexander J, Poynter L, Scott A, et al. A prospective multi-centered analysis of the rectal cancer mucosal microbiome during neoadjuvant long course chemoradiotherapy. Annual Scientific Meeting of the American Society of Colon and Rectal Surgeons. Seattle, WA, USA, 2017:E118.

382. Alexander JL, Wilson ID, Teare J, et al. Gut microbiota modulation of chemotherapy efficacy and toxicity. *Nat Rev Gastroenterol Hepatol* 2017;14(6):356-65. doi: 10.1038/nrgastro.2017.20. Epub 17 Mar 8.

383. Feeney G, Sehgal R, Sheehan M, et al. Neoadjuvant radiotherapy for rectal cancer management. *World J Gastroenterol* 2019;25(33):4850-69. doi: 10.3748/wjg.v25.i33.4850.

384. Abdul-Jalil KI, Sheehan KM, Kehoe J, et al. The prognostic value of tumour regression grade following neoadjuvant chemoradiation therapy for rectal cancer. *Colorectal Dis* 2014;16(1):O16-25. doi: 10.1111/codi.12439.

385. Sakin A, Sahin S, Sengul Samanci N, et al. The impact of tumor regression grade on long-term survival in locally advanced rectal cancer treated with preoperative chemoradiotherapy. *J Oncol Pharm Pract* 2020;26(7):1611-20. doi: 10.177/1078155219900944. Epub 2020 Feb 11.

386. Song C, Chung JH, Kang SB, et al. Impact of Tumor Regression Grade as a Major Prognostic Factor in Locally Advanced Rectal Cancer after Neoadjuvant Chemoradiotherapy: A Proposal for a Modified Staging System. *Cancers (Basel)* 2018;10(9):319. doi: 10.3390/cancers10090319.

387. Hinrichs-Krapels S, Grant J. Exploring the effectiveness, efficiency and equity (3e's) of research and research impact assessment. *Palgrave Communications* 2016;2 doi: 10.1057/palcomms.2016.90

388. Chalmers I, Bracken MB, Djulbegovic B, et al. How to increase value and reduce waste when research priorities are set. *Lancet* 2014;383(9912):156-65. doi: 10.1016/S0140-6736(13)62229-1. Epub 2014 Jan 8.

Appendix A: Biomarker extraction and GO annotations

Study no.	n	Statistical test	Results (if p, <0.05 significant)	Biomarker	Function (broad)	Function (level 1)	Function (level 2)	Function (level 3)	Function (level 4)	Function (level 5)	Function (level 6)
LP109	121	Chi-squared	Univariate: <0.001	ABCC4	Biological process	Cellular process	Cellular physiological process	Cell communication	Signal transduction	Cell-surface receptor linked signalling	
LP109	121	Multivariate analysis	Coefficient 1.443, SE 0.405, Wald 12.688, p <0.001, OR 4.232	ABCC4	Biological process	Cellular process	Cellular physiological process	Cell communication	Signal transduction	Cell-surface receptor linked signalling	
LP083	64		NS	ALDH1	Biological process	Cellular process	Cellular physiological process	Cellular metabolism	Regulation of cellular metabolism		
LP129	112	Logistic regression	NS	ALDH1	Biological process	Cellular process	Cellular physiological process	Cellular metabolism	Regulation of cellular metabolism		
LP005	17	Mann-Whitney U	NS	Annexin V	Biological process	Cellular process	Cellular physiological process	Cell death	Programmed cell death	Apoptosis	
LP127	172	Chi-squared	0.009	ANXA1	Biological process	Response to stimulus	Response to stress				
LP070	152	Chi-squared	0.001	APAF-1	Biological process	Cellular process	Cellular physiological process	Cell death	Programmed cell death	Apoptosis	
LP138	130	Chi-squared	0.003	APAF-1	Biological process	Cellular process	Cellular physiological process	Cell death	Programmed cell death	Apoptosis	
LP138	130	Multivariate stepwise	NS	APAF-1	Biological process	Cellular process	Cellular physiological process	Cell death	Programmed cell death	Apoptosis	
LP001	44	Mann-Whitney U	0.003	Apoptotic index	Biological process	Cellular process	Cellular physiological process	Cell death	Programmed cell death	Apoptosis	
LP029	36	Mann-Whitney U	0.0051	Apoptotic index	Biological process	Cellular process	Cellular physiological process	Cell death	Programmed cell death	Apoptosis	
LP120	172	Chi-squared	<0.001	Asparagine synthase	Biological process	Cellular process	Cellular physiological process	Cellular metabolism			
LP120	172	Chi-squared	<0.001	Asparagine synthase	Biological process	Cellular process	Cellular physiological process	Cellular metabolism			
LP007	130	Odds ratio	4.754 95%CI=1.037-21.776	Bax	Biological process	Cellular process	Cellular physiological process	Cell death	Programmed cell death	Apoptosis	Pro-apoptosis
LP007	130	Chi-squared	0.017	Bax	Biological process	Cellular process	Cellular physiological process	Cell death	Programmed cell death	Apoptosis	Pro-apoptosis
LP070	152	Chi-squared	NS	Bax	Biological process	Cellular process	Cellular physiological process	Cell death	Programmed cell death	Apoptosis	Pro-apoptosis
LP073	29	Chi-squared	0.003	Bax	Biological process	Cellular process	Cellular physiological process	Cell death	Programmed cell death	Apoptosis	Pro-apoptosis
LP129	112	Logistic regression	NS	Bax	Biological process	Cellular process	Cellular physiological process	Cell death	Programmed cell death	Apoptosis	Pro-apoptosis
LP070	152	Chi-squared	NS	Bcl-2	Biological process	Cellular process	Cellular physiological process	Cell communication	Signal transduction	Apoptosis	Inhibition of apoptosis
LP138	130	Chi-squared	0.005	Bcl-2	Biological process	Cellular process	Cellular physiological process	Cell communication	Signal transduction	Apoptosis	Inhibition of apoptosis
LP138	130	Multivariate stepwise	Wald 4.02	Bcl-2	Biological process	Cellular process	Cellular physiological process	Cell communication	Signal transduction	Apoptosis	Inhibition of apoptosis
LP049	54	Chi-squared	NS	beta-catenin	Biological process	Cellular process	Cellular physiological process	Cell communication	Signal transduction		
LP107	136	Chi squared	<0.001	Beta-catenin	Biological process	Cellular process	Cellular physiological process	Cell communication	Signal transduction		
LP107	136	Univariate logistic regression	<0.001	Beta-catenin	Biological process	Cellular process	Cellular physiological process	Cell communication	Signal transduction		
LP107	136	Multivariate logistic regression	OR 6.375, p <0.001	Beta-catenin	Biological process	Cellular process	Cellular physiological process	Cell communication	Signal transduction		
LP138	130	Chi-squared	0.027	beta-catenin	Biological process	Cellular process	Cellular physiological process	Cell communication	Signal transduction		
LP138	130	Multivariate stepwise	NS	beta-catenin	Biological process	Cellular process	Cellular physiological process	Cell communication	Signal transduction		
LP005	17	Mann-Whitney U	NS	beta-tubulin	Biological process	Cellular process	Cellular physiological process	Cell proliferation	Mitosis		
LP012	59	Student's t test	0.03	BIRC5 (survivin)	Biological process	Cellular process	Cellular physiological process	Cell death	Programmed cell death	Apoptosis	Inhibition of apoptosis
LP012	59	Student's t test	0.03	BIRC5 (survivin)	Biological process	Cellular process	Cellular physiological process	Cell cycle	Regulation of cell cycle	Cell cycle inhibition	Cell cycle arrest
LP012	59	Student's t test	0.03	BIRC5 (survivin)	Biological process	Response to stimulus	Response to stress	Response to DNA damage stimulus	Impaired post-irradiation DNA repair		
LP029	36	Mann-Whitney U	NS	BIRC5 (survivin)	Biological process	Cellular process	Cellular physiological process	Cell death	Programmed cell death	Apoptosis	Inhibition of apoptosis
LP029	36	Mann-Whitney U	NS	BIRC5 (survivin)	Biological process	Cellular process	Cellular physiological process	Cell cycle	Regulation of cell cycle	Cell cycle inhibition	Cell cycle arrest
LP029	36	Mann-Whitney U	NS	BIRC5 (survivin)	Biological process	Response to stimulus	Response to stress	Response to DNA damage stimulus	Impaired post-irradiation DNA repair		
LP056	22	Welch's 2-tailed t test	<0.05	BIRC5 (survivin)	Biological process	Cellular process	Cellular physiological process	Cell death	Programmed cell death	Apoptosis	Inhibition of apoptosis
LP056	22	Welch's 2-tailed t test	<0.05	BIRC5 (survivin)	Biological process	Cellular process	Cellular physiological process	Cell cycle	Regulation of cell cycle	Cell cycle inhibition	Cell cycle arrest
LP056	22	Welch's 2-tailed t test	<0.05	BIRC5 (survivin)	Biological process	Response to stimulus	Response to stress	Response to DNA damage stimulus	Impaired post-irradiation DNA repair		
LP115	123	Chi-squared/Fisher's exact	0.019	BIRC5 (survivin)	Biological process	Cellular process	Cellular physiological process	Cell death	Programmed cell death	Apoptosis	Inhibition of apoptosis
LP115	123	Chi-squared/Fisher's exact	0.019	BIRC5 (survivin)	Biological process	Cellular process	Cellular physiological process	Cell cycle	Regulation of cell cycle	Cell cycle inhibition	Cell cycle arrest

LP115	123	Chi-squared/Fisher's exact	0.019	BIRC5 (survivin)	Biological process	Response to stimulus	Response to stress	Response to DNA damage stimulus	Impaired post-irradiation DNA repair		
LP052	36	Chi-squared	0.012	CA9	Biological process	Response to stimulus	Response to endogenous stimulus	Response to hypoxia			
LP005	17	Mann-Whitney U	NS	Calsenilin	Biological process	Cellular process	Cell communication	Cell surface receptor signalling	Apoptosis		
LP010	50	Fisher's exact test	0.03	Caspase 8	Biological process	Cellular process	Cellular physiological process	Cell death	Programmed cell death	Apoptosis	
LP062	96	Chi-squared	0.03	CD133	Biological process	Cellular process	Cellular physiological process	Cell communication	Signal transduction	Cell-surface receptor linked signalling	
LP062	96	Multivariate logistic regression	OR = 3.32	CD133	Biological process	Cellular process	Cellular physiological process	Cell communication	Signal transduction	Cell-surface receptor linked signalling	
LP129	112	Logistic regression	NS	CD133	Biological process	Cellular process	Cellular physiological process	Cell communication	Signal transduction	Cell-surface receptor linked signalling	
LP129	112	Logistic regression	NS	CD166	Biological process	Cellular process	Cellular physiological process	Cell communication	Signal transduction	Cell-surface receptor linked signalling	
LP024	30	Chi-squared	NS	CD34	Biological process	Cellular process	Cell communication	Cell surface receptor signalling			
LP115	123	Chi-squared/Fisher's exact	0.027	CD44	Biological process	Cellular process	Cellular physiological process	Cell communication	Signal transduction	Cell-surface receptor linked signalling	
LP129	112	Logistic regression	NS	CD44	Biological process	Cellular process	Cellular physiological process	Cell communication	Signal transduction	Cell-surface receptor linked signalling	
LP156	81	Chi-squared	0.01	c-met	Biological process	Response to stimulus	Response to endogenous stimulus	Response to hypoxia			
LP156	81	Chi-squared	0.01	c-met	Biological process	Cellular process	Cellular physiological process	Cell communication	Signal transduction		
LP156	81	Predictive value	Sensitivity 74%, specificity 62%; PPV 86%, NPV 43%, accuracy 71%	c-met	Biological process	Cellular process	Cellular physiological process	Cell communication	Signal transduction		
LP015	49	Chi-squared	0.026	COX-2	Biological process	Cellular process	Cellular physiological process	Cellular metabolism			
LP019	52	Chi-squared	0.021	COX-2	Biological process	Cellular process	Cellular physiological process	Cellular metabolism			
LP019	52	Chi-squared	0.0031	COX-2	Biological process	Cellular process	Cellular physiological process	Cellular metabolism			
LP024	30	Chi-squared	0.003	COX-2	Biological process	Cellular process	Cellular physiological process	Cellular metabolism			
LP062	96	Chi-squared	0.0014	COX-2	Biological process	Cellular process	Cellular physiological process	Cellular metabolism			
LP062	96	Multivariate logistic regression	OR = 7.84	COX-2	Biological process	Cellular process	Cellular physiological process	Cellular metabolism			
LP070	152	Chi-squared	0.003	COX-2	Biological process	Cellular process	Cellular physiological process	Cellular metabolism			
LP138	130	Chi-squared	NS	COX-2	Biological process	Cellular process	Cellular physiological process	Cellular metabolism			
LP138	130	Multivariate stepwise	NS	COX-2	Biological process	Cellular process	Cellular physiological process	Cellular metabolism			
LP117	172	Chi-squared	0.004	CPS1	Biological process	Cellular process	Cellular physiological process	Cellular metabolism			
LP119	95	Chi-squared	0.01	CXCL10	Biological process	Response to stimulus	Response to endogenous stimulus	Response to hypoxia			
LP052	36	Chi-squared	NS 0.30	CXCR4	Biological process	Cellular process	Cellular physiological process	Cell communication	Signal transduction	Cell-surface receptor linked signalling	
LP076	38	Mann-Whitney U	0.02	DIABLO	Biological process	Response to stimulus	Response to stress	Apoptosis			
LP140	172	Chi-squared	0.009	DNAJC12	Biological process	Cellular process	Response to stimulus	Response to stress	Cellular homeostasis		
LP017	91	Chi-squared	NS	EGFR	Biological process	Cellular process	Cellular physiological process	Cell communication	Signal transduction	Cell-surface receptor linked signalling	Cell cycle
LP017	91	Chi-squared	NS	EGFR	Biological process	Cellular process	Cellular physiological process	Cell communication	Signal transduction	Cell-surface receptor linked signalling	Cellular metabolism
LP017	91	Chi-squared	NS	EGFR	Biological process	Cellular process	Cellular physiological process	Cell communication	Signal transduction	Apoptosis	Inhibition of apoptosis

LP034	46	Chi-squared	0.007	EGFR	Biological process	Cellular process	Cellular physiological process	Cell communication	Signal transduction	Cell-surface receptor linked signalling	Cell cycle
LP034	46	Chi-squared	0.007	EGFR	Biological process	Cellular process	Cellular physiological process	Cell communication	Signal transduction	Cell-surface receptor linked signalling	Cellular metabolism
LP034	46	Odds ratio	12.9 95%CI=1.44-115.5	EGFR	Biological process	Cellular process	Cellular physiological process	Cell communication	Signal transduction	Apoptosis	Inhibition of apoptosis
LP039	40	Odds ratio	OR=14.8 [95%CI=2.2-100.6]	EGFR	Biological process	Cellular process	Cellular physiological process	Cell communication	Signal transduction	Cell-surface receptor linked signalling	Cell cycle
LP039	40	Logistic regression	0.006	EGFR	Biological process	Cellular process	Cellular physiological process	Cell communication	Signal transduction	Cell-surface receptor linked signalling	Cellular metabolism
LP039	40	Logistic regression	0.006	EGFR	Biological process	Cellular process	Cellular physiological process	Cell communication	Signal transduction	Apoptosis	Inhibition of apoptosis
LP129	112	Logistic regression	NS	EGFR	Biological process	Cellular process	Cellular physiological process	Cell communication	Signal transduction	Apoptosis	
LP025	39	Odds ratio	25.6 95%CI 2.76-239.9	EGFR with k-RAS wild type	Biological process	Cellular process	Cellular physiological process	Cell communication	Signal transduction	Cell-surface receptor linked signalling	Cell cycle
LP025	39	Logistic regression	0.0012	EGFR with k-RAS wild type	Biological process	Cellular process	Cellular physiological process	Cell communication	Signal transduction	Cell-surface receptor linked signalling	Cellular metabolism
LP025	39	Logistic regression	0.0012	EGFR with k-RAS wild type	Biological process	Cellular process	Cellular physiological process	Cell communication	Signal transduction	Apoptosis	Inhibition of apoptosis
LP129	112	Logistic regression	NS	EpCAM	Biological process	Cellular process	Cellular physiological process	Cell communication	Signal transduction	Cell-surface receptor linked signalling	
LP071	69	Chi-squared	0.00006	GHRH-R	Biological process	Cellular process	Cellular physiological process	Cell communication	Signal transduction	Cell-surface receptor linked signalling	
LP026	69	Predictive value	Sensitivity 45%, specificity 87%	GLUT-1	Biological process	Cellular process	Cellular physiological process	Cellular metabolism			
LP052	36	Chi-squared	NS 0.98	GLUT-1	Biological process	Cellular process	Cellular physiological process	Cellular metabolism			
LP122	119	Chi-squared	NS	HER-2	Biological process	Cellular process	Cellular physiological process	Cell communication	Signal transduction	Cell-surface receptor linked signalling	
LP138	130	Chi-squared	NS	HER-2	Biological process	Cellular process	Cellular physiological process	Cell communication	Signal transduction	Cell-surface receptor linked signalling	
LP138	130	Multivariate stepwise	NS	HER-2	Biological process	Cellular process	Cellular physiological process	Cell communication	Signal transduction	Cell-surface receptor linked signalling	
LP039	40	Odds ratio	OR=10 [95%CI=1.8-56.2]	HIF-1	Biological process	Response to stimulus	Response to stress	Apoptosis	Anti-apoptosis		
LP039	40	Logisitic regression	0.008	HIF-1	Biological process	Response to stimulus	Response to stress	Apoptosis	Anti-apoptosis		
LP129	112	Logistic regression	NS	HIF1a	Biological process	Response to stimulus	Response to stress	Apoptosis	Anti-apoptosis		
LP082	45	Multivariate logistic regression	0.026	HMGCS2	Biological process	Cellular process	Cellular physiological process	Cellular metabolism	Regulation of cellular metabolism		
LP082	45	Odds ratio	7.778 [95%CI=1.280-47.383]	HMGCS2	Biological process	Cellular process	Cellular physiological process	Cellular metabolism	Regulation of cellular metabolism		
LP146	172	Mann-Whitney U	<0.01	HMGCS2	Biological process	Cellular process	Cellular physiological process	Cellular metabolism	Regulation of cellular metabolism		
LP046	238	Fisher's exact test	0.49 [95%CI=0.25-0.94]	hOGG1-1245C>G	Biological process	Response to stimulus	Response to stress	Response to DNA damage stimulus	DNA repair		
LP046	238	Fisher's exact test	<0.05	hOGG1-1245C>G	Biological process	Response to stimulus	Response to stress	Response to DNA damage stimulus	DNA repair		
LP104	86	Multiple logistic regression	Univariate 0.000; Multivariate OR 10.570, p 0.001	hPEBP4	Biological process	Cellular process	Cellular physiological process	Cell death	Programmed cell death	Apoptosis	Inhibition of apoptosis
LP146	172	Mann-Whitney U	<0.01	HSD17B2	Biological process	Cellular process	Cellular physiological process	Cellular metabolism	Regulation of cellular metabolism		
LP005	17	Mann-Whitney U	0.03	Hsp42	Biological process	Cellular process	Response to stimulus	Response to stress	Cellular homeostasis		
LP071	69	Chi-squared	0.00002	Hsp90	Biological process	Cellular process	Response to stimulus	Response to stress	Cellular homeostasis		
LP141	30	General linear model	0.01	IQGAP1	Biological process	Cellular process	Cellular physiological process	Cell communication	Signal transduction	Intracellular signalling	
LP023	40	Chi-squared	<0.05	Ki67	Biological process	Cellular process	Cellular physiological process	Cell proliferation			

LP071	69	Chi-squared	0.00002	Hsp90	Biological process	Cellular process	Response to stimulus	Response to stress	Cellular homeostasis			
LP141	30	General linear model	0.01	IQGAP1	Biological process	Cellular process	Cellular physiological process	Cell communication	Signal transduction	Intracellular signalling		
LP023	40	Chi-squared	<0.05	Ki67	Biological process	Cellular process	Cellular physiological process	Cell proliferation				
LP034	46	Chi-squared	0.05	Ki67	Biological process	Cellular process	Cellular physiological process	Cell proliferation				
LP041	40	Chi-squared	0.034	Ki67	Biological process	Cellular process	Cellular physiological process	Cell proliferation				
LP116	81	Chi-squared	0.01	Ki67	Biological process	Cellular process	Cellular physiological process	Cell proliferation				
LP138	130	Chi-squared	NS	Ki67	Biological process	Cellular process	Cellular physiological process	Cell proliferation				
LP138	130	Multivariate stepwise	NS	Ki67	Biological process	Cellular process	Cellular physiological process	Cell proliferation				
LP126	72	Chi-squared	NS	M30	Biological process	Cellular process	Cellular physiological process	Cell death	Programmed cell death	Apoptosis		
LP017	91	Chi-squared	NS	MIB1	Biological process	Cellular process	Cellular physiological process	Cell death	Programmed cell death	Apoptosis		
LP073	29	Chi-squared	0.04	MMP9	Biological process	Cellular process	Cellular physiological process	Cellular metabolism				
LP133	144	Logistic regression	0.002	MRP3	Biological process	Cellular process	Cellular physiological process	Cell communication	Signal transduction	Cell-surface receptor linked signalling		
LP017	91	Chi-squared	NS	MSH2	Biological process	Response to stimulus	Response to endogenous stimulus	Response to DNA damage stimulus	DNA repair			
LP046	238	Fisher's exact test	<0.05	MTFHR-677C>T	Biological process	Response to stimulus	Response to stress	Response to DNA damage stimulus	DNA repair			
LP016	142	Odds ratio	3.85 95%CI=1.41-6.75	MTHFR gene haplotype 677T-1298A	Biological process	Cellular process	Cellular physiological process	Response to DNA damage stimulus	Nucleic acid metabolism	DNA metabolism		
LP016	142	Fisher's exact test	<0.05	MTHFR gene haplotype 677T-1298A	Biological process	Cellular process	Cellular physiological process	Response to DNA damage stimulus	Nucleic acid metabolism	DNA metabolism		
LP046	238	Fisher's exact test	0.50 [95%CI=0.25-0.97]	MTHFR-677C>T	Biological process	Response to stimulus	Response to stress	Response to DNA damage stimulus	DNA repair			
LP101	74	ANOVA	<0.05	m-TIMP3	Biological process	Response to stimulus	Response to stress					
LP056	22	Welch's 2-tailed t test	<0.05	NEIL2	Biological process	Response to stimulus	Response to stress	Response to DNA damage stimulus	DNA repair			
LP115	123	Chi-squared/Fisher's exact	0.019	NFkB	Biological process	Response to stimulus	Response to stress	Response to DNA damage stimulus				
LP067	74	Chi-squared	NS	NFkB	Biological process	Response to stimulus	Response to stress	Response to DNA damage stimulus				
LP017	91	Chi-squared	NS	p21	Biological process	Cellular process	Cellular physiological process	Cell cycle	Regulation of cell cycle	Cell cycle checkpoint	Cell aycle arrest	
LP017	91	Chi-squared	NS	p21	Biological process	Response to stimulus	Response to stress	Cell cycle	Regulation of cell cycle	Cell cycle inhibition	Cell aycle arrest	
LP070	152	Chi-squared	NS	p21	Biological process	Cellular process	Cellular physiological process	Cell cycle	Regulation of cell cycle	Cell cycle checkpoint	Cell aycle arrest	
LP070	152	Chi-squared	NS	p21	Biological process	Response to stimulus	Response to stress	Cell cycle	Regulation of cell cycle	Cell cycle inhibition	Cell aycle arrest	
LP116	81	Chi-squared	0.01	p21	Biological process	Cellular process	Cellular physiological process	Cell cycle	Regulation of cell cycle	Cell cycle checkpoint	Cell aycle arrest	
LP129	112	Logistic regression	0.021	p21	Biological process	Cellular process	Cellular physiological process	Cell cycle	Regulation of cell cycle	Cell cycle checkpoint	Cell aycle arrest	
LP129	112	Logistic regression	0.021	p21	Biological process	Response to stimulus	Response to stress	Cell cycle	Regulation of cell cycle	Cell cycle inhibition	Cell aycle arrest	
LP129	112	(Multivariate) Odds ratio	0.127 95%CI=0.022-0.729	p21	Biological process	Cellular process	Cellular physiological process	Cell cycle	Regulation of cell cycle	Cell cycle checkpoint	Cell aycle arrest	
LP129	112	(Multivariate) Odds ratio	0.127 95%CI=0.022-0.729	p21	Biological process	Response to stimulus	Response to stress	Cell cycle	Regulation of cell cycle	Cell cycle inhibition	Cell aycle arrest	
LP002	42	Mann-Whitney U	NS	p27	Biological process	Cellular process	Cellular physiological process	Cell cycle	Regulation of cell cycle	Cell cycle inhibition		

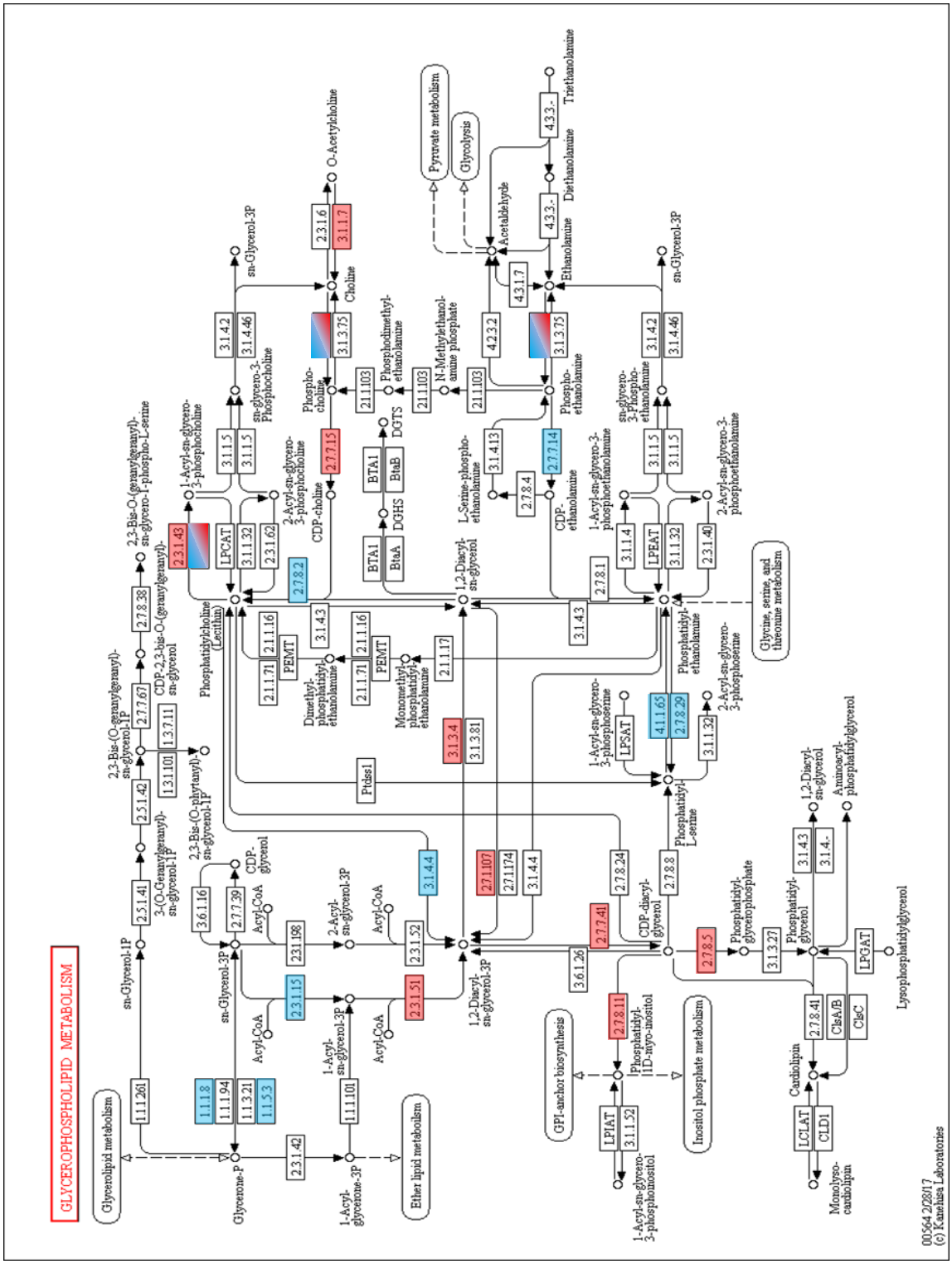
LP070	152	Chi-squared	NS	p27	Biological process	Cellular process	Cellular physiological process	Cell cycle	Regulation of cell cycle	Cell cycle inhibition	
LP010	50	Fisher's exact test	0.03	p53	Biological process	Response to stimulus	Response to stress	Response to DNA damage stimulus	DNA repair		
LP010	50	Fisher's exact test	0.03	p53	Biological process	Cellular process	Cellular physiological process	Cell cycle	Regulation of cell cycle	Cell cycle checkpoint	Cell aycle arrest
LP010	50	Fisher's exact test	0.03	p53	Biological process	Cellular process	Cellular physiological process	Cell death	Programmed cell death	Apoptosis	
LP011	70	Chi-squared	NS	p53	Biological process	Response to stimulus	Response to stress	Response to DNA damage stimulus	DNA repair		
LP011	70	Chi-squared	NS	p53	Biological process	Cellular process	Cellular physiological process	Cell cycle	Regulation of cell cycle	Cell cycle checkpoint	Cell aycle arrest
LP011	70	Chi-squared	NS	p53	Biological process	Cellular process	Cellular physiological process	Cell death	Programmed cell death	Apoptosis	
LP017	91	Chi-squared	NS	p53	Biological process	Response to stimulus	Response to stress	Response to DNA damage stimulus	DNA repair		
LP017	91	Chi-squared	NS	p53	Biological process	Cellular process	Cellular physiological process	Cell cycle	Regulation of cell cycle	Cell cycle checkpoint	Cell aycle arrest
LP017	91	Chi-squared	NS	p53	Biological process	Cellular process	Cellular physiological process	Cell death	Programmed cell death	Apoptosis	
LP019	52	Chi-squared	NS	p53	Biological process	Response to stimulus	Response to stress	Response to DNA damage stimulus	DNA repair		
LP019	52	Chi-squared	NS	p53	Biological process	Cellular process	Cellular physiological process	Cell death	Programmed cell death	Apoptosis	
LP019	52	Chi-squared	NS	p53	Biological process	Cellular process	Cellular physiological process	Cell cycle	Regulation of cell cycle	Cell cycle checkpoint	Cell aycle arrest
LP026	69	Chi-squared	NS	p53	Biological process	Response to stimulus	Response to stress	Response to DNA damage stimulus	DNA repair		
LP026	69	Chi-squared	NS	p53	Biological process	Cellular process	Cellular physiological process	Cell cycle	Regulation of cell cycle	Cell cycle checkpoint	Cell aycle arrest
LP026	69	Chi-squared	NS	p53	Biological process	Cellular process	Cellular physiological process	Cell death	Programmed cell death	Apoptosis	
LP070	152	Chi-squared	NS	p53	Biological process	Response to stimulus	Response to stress	Response to DNA damage stimulus	DNA repair		
LP070	152	Chi-squared	NS	p53	Biological process	Cellular process	Cellular physiological process	Cell cycle	Regulation of cell cycle	Cell cycle checkpoint	Cell aycle arrest
LP070	152	Chi-squared	NS	p53	Biological process	Cellular process	Cellular physiological process	Cell death	Programmed cell death	Apoptosis	
LP116	81	Chi-squared	0.03	p53	Biological process	Response to stimulus	Response to stress	Response to DNA damage stimulus	DNA repair		
LP116	81	Chi-squared	0.03	p53	Biological process	Cellular process	Cellular physiological process	Cell cycle	Regulation of cell cycle	Cell cycle checkpoint	Cell aycle arrest
LP116	81	Chi-squared	0.03	p53	Biological process	Cellular process	Cellular physiological process	Cell death	Programmed cell death	Apoptosis	
LP129	112	Logistic regression	NS	p53	Biological process	Response to stimulus	Response to stress	Response to DNA damage stimulus	DNA repair		
LP129	112	Logistic regression	NS	p53	Biological process	Cellular process	Cellular physiological process	Cell cycle	Regulation of cell cycle	Cell cycle checkpoint	Cell aycle arrest
LP129	112	Logistic regression	NS	p53	Biological process	Cellular process	Cellular physiological process	Cell death	Programmed cell death	Apoptosis	
LP138	130	Chi-squared	NS	p53	Biological process	Response to stimulus	Response to stress	Response to DNA damage stimulus	DNA repair		
LP138	130	Chi-squared	NS	p53	Biological process	Cellular process	Cellular physiological process	Cell cycle	Regulation of cell cycle	Cell cycle checkpoint	Cell aycle arrest
LP138	130	Chi-squared	NS	p53	Biological process	Cellular process	Cellular physiological process	Cell death	Programmed cell death	Apoptosis	
LP138	130	Multivariate stepwise	NS	p53	Biological process	Response to stimulus	Response to stress	Response to DNA damage stimulus	DNA repair		
LP138	130	Multivariate stepwise	NS	p53	Biological process	Cellular process	Cellular physiological process	Cell cycle	Regulation of cell cycle	Cell cycle checkpoint	Cell aycle arrest

LP138	130	Multivariate stepwise	NS	p53	Biological process	Cellular process	Cellular physiological process	Cell death	Programmed cell death	Apoptosis	
LP112	88	Chi-squared	0.02	PDCD4	Biological process	Cellular process	Cellular physiological process	Cell death	Programmed cell death	Apoptosis	
LP112	88	(Multivariate) Odds ratio	3.38 95%CI=1.79-7.02	PDCD4	Biological process	Cellular process	Cellular physiological process	Cell death	Programmed cell death	Apoptosis	
LP071	69	Chi-squared	NS	phospho-Akt	Biological process	Cellular process	Cellular physiological process	Cellular metabolism			
LP071	69	Chi-squared	NS	phospho-Akt	Biological process	Cellular process	Cellular physiological process	Cell death	Programmed cell death	Apoptosis	Inhibition of apoptosis
LP071	69	Chi-squared	NS	phospho-Akt	Biological process	Cellular process	Cellular physiological process	Cell cycle	Regulation of cell cycle		
LP071	69	Chi-squared	NS	phospho-Akt	Biological process	Response to stimulus	Response to endogenous stimulus	Response to hypoxia	Tumour angiogenesis		
LP142	86	Multivariate logistic regression	0.015	p-SMAD3	Biological process	Cellular process	Cellular physiological process	Cellular metabolism	Nucleic acid metabolism	Transcription	Regulation of transcription
LP142	86	Logistic regression	<0.001	p-SMAD3	Biological process	Cellular process	Cellular physiological process	Cellular metabolism	Nucleic acid metabolism	Transcription	Regulation of transcription
LP056	22	Welch's 2-tailed t test	<0.05	REG4	Biological process	Cellular process	Cellular physiological process	Cell proliferation			
LP114	172	Chi-squared	0.001	REG4	Biological process	Cellular process	Cellular physiological process	Cell proliferation			
LP043	65	ANOVA	NS	Securin	Biological process	Cellular process	Cellular physiological process	Cell proliferation			
LP071	69	Chi-squared	NS	sHsp16.2	Biological process	Cellular process	Response to stimulus	Response to stress	Cellular homeostasis		
LP041	40	Chi-squared	0.011	Smac	Biological process	Response to stimulus	Response to stress	Apoptosis			
LP081	98	Chi-squared	0.007	Smac	Biological process	Response to stimulus	Response to stress	Apoptosis			
LP081	98	Odds ratio	OR 9.5 [95%CI=1.34-47.8]	Smac	Biological process	Response to stimulus	Response to stress	Apoptosis			
LP142	86	Multivariate logistic regression	NS	SMAD3	Biological process	Cellular process	Cellular physiological process	Cellular metabolism	Nucleic acid metabolism	Transcription	Regulation of transcription
LP142	86	Logistic regression	0.003	SMAD3	Biological process	Cellular process	Cellular physiological process	Cellular metabolism	Nucleic acid metabolism	Transcription	Regulation of transcription
LP137	96	Chi-squared	0.03	TCF4	Biological process	Cellular process	Cellular physiological process	Cellular metabolism	Nucleic acid metabolism	Transcription	Regulation of transcription
LP142	86	Multivariate logistic regression	NS	TGF-b1	Biological process	Cellular process	Cellular physiological process	Cell communication	Signal transduction	Cell-surface receptor linked signalling	
LP147	172	Chi-squared	0.015	THBS2	Biological process	Cellular process	Cellular physiological process	Cell communication	Cell adhesion		
LP024	30	Chi-squared	NS	Thymidine phosphorylase	Biological process	Cellular process	Cellular physiological process	Cellular metabolism	Nucleic acid metabolism	DNA metabolism	
LP024	30	Chi-squared	NS	Thymidine phosphorylase	Biological process	Response to stimulus	Response to endogenous stimulus	Response to hypoxia	Tumour angiogenesis		
LP004	60	Chi-squared	0.047	Thymidylate synthase	Biological process	Cellular process	Cellular physiological process	Cellular metabolism	Nucleic acid metabolism	DNA synthesis	
LP006	14	Chi-squared	0.0179	Thymidylate synthase	Biological process	Cellular process	Cellular physiological process	Cellular metabolism	Nucleic acid metabolism	DNA synthesis	
LP009	40	Chi-squared	0.04	Thymidylate synthase	Biological process	Cellular process	Cellular physiological process	Cellular metabolism	Nucleic acid metabolism	DNA synthesis	
LP017	91	Chi-squared	NS	Thymidylate synthase	Biological process	Cellular process	Cellular physiological process	Cellular metabolism	Nucleic acid metabolism	DNA synthesis	
LP023	40	Chi-squared	<0.05	Thymidylate synthase	Biological process	Cellular process	Cellular physiological process	Cellular metabolism	Nucleic acid metabolism	DNA synthesis	
LP034	46	Chi-squared	0.002	Thymidylate synthase	Biological process	Cellular process	Cellular physiological process	Cellular metabolism	Nucleic acid metabolism	DNA synthesis	
LP034	46	Odds ratio	16.7 95%CI=1.8-150.9	Thymidylate synthase	Biological process	Cellular process	Cellular physiological process	Cellular metabolism	Nucleic acid metabolism	DNA synthesis	
LP047	167	Spearman correlation coefficient	NS	Thymidylate synthase	Biological process	Cellular process	Cellular physiological process	Cellular metabolism	Nucleic acid metabolism	DNA synthesis	

LP115	123	Chi-squared/Fisher's exact	0.027	Thymidylate synthase	Biological process	Cellular process	Cellular physiological process	Cellular metabolism	Nucleic acid metabolism	DNA synthesis	
LP129	112	Logistic regression	NS	Thymidylate synthase	Biological process	Cellular process	Cellular physiological process	Cellular metabolism	Nucleic acid metabolism	DNA synthesis	
LP005	17	Mann-Whitney U	0.01	Tropomodulin	Biological process	Cellular process	Cell structure				
LP013	59	Logistic regression	0.0172	VEGF	Biological process	Response to stimulus	Response to endogenous stimulus	Response to hypoxia	Tumour angiogenesis		
LP017	91	Chi-squared	NS	VEGF	Biological process	Response to stimulus	Response to endogenous stimulus	Response to hypoxia	Tumour angiogenesis		
LP024	30	Chi-squared	NS	VEGF	Biological process	Response to stimulus	Response to endogenous stimulus	Response to hypoxia	Tumour angiogenesis		
LP039	40	Odds ratio	OR=11.7 [95%CI=1.8-74.2]	VEGF	Biological process	Response to stimulus	Response to endogenous stimulus	Response to hypoxia	Tumour angiogenesis		
LP039	40	Logistic regression	0.009	VEGF	Biological process	Response to stimulus	Response to endogenous stimulus	Response to hypoxia	Tumour angiogenesis		
LP041	40	Chi-squared	0.01	VEGF	Biological process	Response to stimulus	Response to endogenous stimulus	Response to hypoxia	Tumour angiogenesis		
LP070	152	Chi-squared	0.009	VEGF	Biological process	Response to stimulus	Response to endogenous stimulus	Response to hypoxia	Tumour angiogenesis		
LP073	29	Chi-squared	0.05	VEGF	Biological process	Response to stimulus	Response to endogenous stimulus	Response to hypoxia	Tumour angiogenesis		
LP116	81	Chi-squared	0.03	VEGF	Biological process	Response to stimulus	Response to endogenous stimulus	Response to hypoxia	Tumour angiogenesis		
LP138	130	Chi-squared	0.048	VEGF	Biological process	Response to stimulus	Response to endogenous stimulus	Response to hypoxia	Tumour angiogenesis		
LP138	130	Multivariate stepwise	NS	VEGF	Biological process	Response to stimulus	Response to endogenous stimulus	Response to hypoxia	Tumour angiogenesis		
LP076	38	Mann-Whitney U	0.04	XIAP	Biological process	Cellular process	Cellular physiological process	Cell death	Programmed cell death	Apoptosis	Inhibition of apoptosis
LP153	67	Chi-squared	0.0002	XRCC2	Biological process	Response to stimulus	Response to stress	Response to DNA damage stimulus	DNA repair		
LP156	81	Chi-squared	0.006	YKL-40	Biological process	Cellular process	Cellular physiological process	Cell proliferation			
LP156	81	Chi-squared	0.006	YKL-40	Biological process	Cellular process	Cellular physiological process	Cell death	Programmed cell death	Apoptosis	Inhibition of apoptosis
LP156	81	Chi-squared	0.006	YKL-40	Biological process	Response to stimulus	Response to endogenous stimulus	Response to hypoxia	Tumour angiogenesis		
LP156	81	Predictive value	Sensitivity 71%, specificity 66%; PPV 87%, NPV 41%, accuracy 70%	YKL-40	Biological process	Cellular process	Cellular physiological process	Cell proliferation			
LP157	82	Predictive value	Sensitivity 71%, specificity 66%; PPV 87%, NPV 41%, accuracy 70%	YKL-40	Biological process	Cellular process	Cellular physiological process	Cell death	Programmed cell death	Apoptosis	
LP158	83	Predictive value	Sensitivity 71%, specificity 66%; PPV 87%, NPV 41%, accuracy 70%	YKL-40	Biological process	Response to stimulus	Response to endogenous stimulus	Response to hypoxia	Tumour angiogenesis		

Appendix A – Full biomarker experimental data extraction table with hierarchical GO annotations and aggregated statistical data from each published study

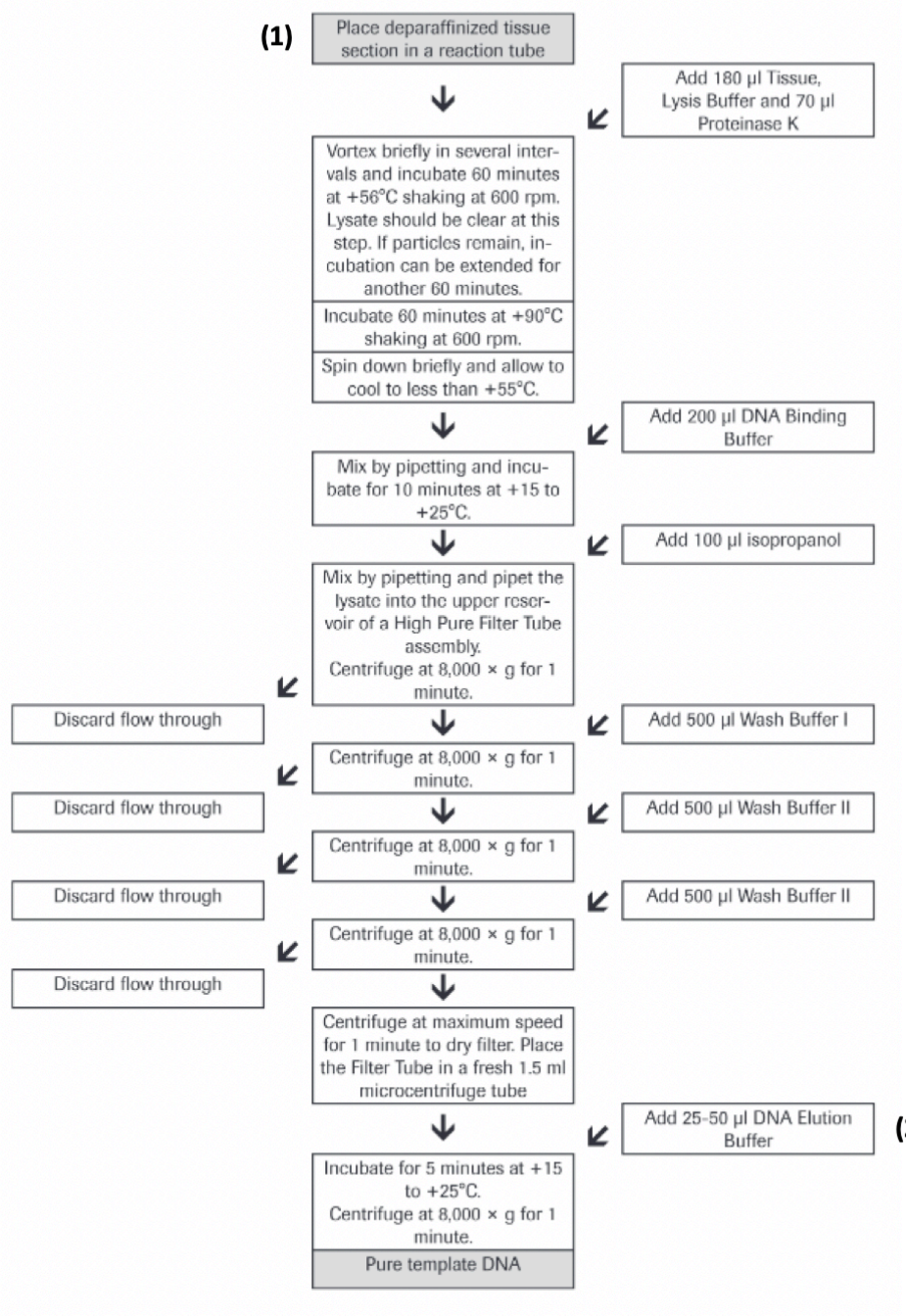
Appendix B: KEGG GPL metabolism pathway



Appendix C: Applied DESI MS/MS settings for Thermo® *LTQ XL*

	Variable	Value
NSI source	Sheath flow rate	0.15
	Aux gas flow rate	0.15
	Sweep gas flow rate	0
	Spray voltage	4.5kV
	Spray current	0.16µA
	Capillary temperature	200^{O}C
	Capillary voltage	-44V
	Tube lens	-120V
Ion optics	Multipole 00 offset	4V
	Lens 0 Voltage	4.2V
	Multipole 0 offset	4.5V
	Lens 1 voltage	15.00V
	Gate lens voltage	35.00V
	Multipole 1 offset	8.00V
	Multipole RF amplitude	400(V p-p)
	Front lens	5.25V

Appendix D: DNA extraction protocols from both FFPE and fresh frozen tissue



Appendix D1 – Experimental workflow taken from the Roche LifeSciences High Pure FFPET DNA Isolation Kit. Numbered brackets denote variance from this protocol; (1) dewaxing procedure as stated and (2) amendments to elution step for improving yields as per methods section 5.2. Taken from the Roche LifeSciences website https://lifescience.roche.com/en_gb/products/

Appendix D2 - QIAamp DNA Micro Kit: procedure for DNA extraction from tissues

Taken from QIAGEN website; https://www.qiagen.com/gb/products

This protocol was selected as it is for isolation of genomic DNA from less than 10 mg tissue:

1. Transfer a tissue sample of less than 10 mg in weight to a 1.5 ml microcentrifuge tube.

2. Immediately add 180 µl Buffer ATL and equilibrate to room temperature (15–25°C).

3. Add 20 µl proteinase K and mix by pulse-vortexing for 15 s.

4. Place the 1.5 ml tube in a thermomixer or heated orbital incubator and incubate at 56°C overnight until the sample is completely lysed. For small amounts of tissue, lysis is complete in 4–6 h, but best results are achieved after overnight lysis.

5. Add 200 µl Buffer AL, close the lid, and mix by pulse-vortexing for 15 s. To ensure efficient lysis, it is essential that the sample and Buffer AL are thoroughly mixed to yield a homogenous solution.

6. Add 200 µl ethanol (96–100%), close the lid, and mix thoroughly by pulse-vortexing for 15 s. Incubate for 5 min at room temperature (15–25°C). Note: If room temperature exceeds 25°C, cool the ethanol on ice before adding to the tube.

7. Briefly centrifuge the 1.5 ml tube to remove drops from inside the lid.

8. Carefully transfer the entire lysate from step 7 to the QIAamp MinElute column (in a 2ml collection tube) without wetting the rim. Close the lid, and centrifuge at 6000 x *g* (8000 rpm) for 1 min. Place the QIAamp MinElute column in a clean 2ml collection tube and discard the collection tube containing the flow-through. If the lysate has not completely passed through the membrane after centrifugation, centrifuge again at a higher speed until the QIAamp MinElute column is empty.

9. Carefully open the QIAamp MinElute column and add 500 µl Buffer AW1 without wetting the rim. Close the lid, and centrifuge at 6000 x *g* (8000 rpm) for 1 min. Place the QIAamp MinElute column in a clean 2 ml collection tube and discard the collection tube containing the flow-through.

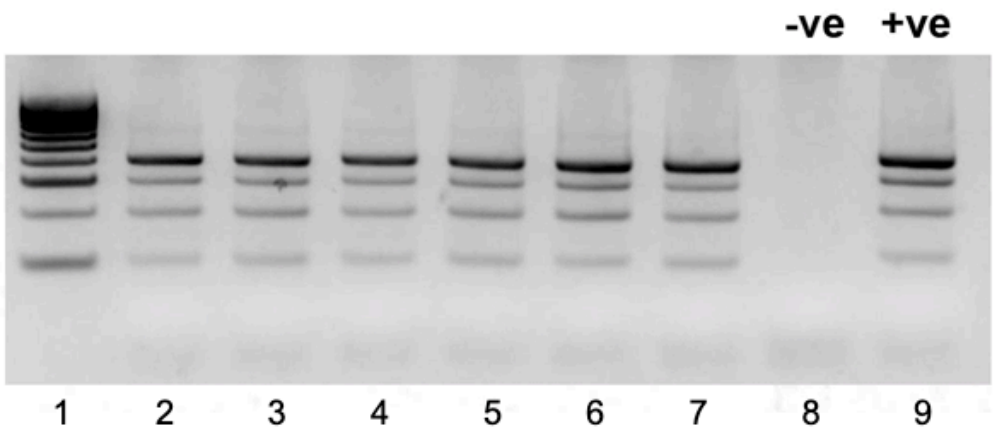
10. Carefully open the QIAamp MinElute column and add 500 µl Buffer AW2 without wetting the rim. Close the lid, and centrifuge at 6000 x *g* (8000 rpm) for 1 min. Place the QIAamp MinElute column in a clean 2 ml collection tube and discard the collection tube containing the flow-through. Contact between the QIAamp MinElute column and the flow-through should be avoided. Some centrifuge rotors may vibrate upon deceleration, resulting in the flow-through, which contains ethanol, coming into contact with the QIAamp MinElute column. Take care when removing the QIAamp MinElute column and collection tube from the rotor, so that flow-through does not come into contact with the QIAamp MinElute column.

11. Centrifuge at full speed (20,000 x *g*; 14,000 rpm) for 3 min to dry the membrane completely. This step is necessary, since ethanol carryover into the eluate may interfere with some downstream applications.

12. Place the QIAamp MinElute column in a clean 1.5 ml microcentrifuge tube and discard the collection tube containing the flow-through. Carefully open the lid of the QIAamp MinElute column and apply 20–100 µl Buffer AE or distilled water to the centre of the membrane. If high pH or EDTA affects sensitive downstream applications, use water for elution (see page 12). Important: Ensure that Buffer AE or distilled water is equilibrated to room temperature (15–25°C). If using small elution volumes (<50 µl), dispense Buffer AE or distilled water onto the centre of the membrane to ensure complete elution of bound DNA. QIAamp MinElute columns provide flexibility in the choice of elution volume. Choose a volume according to the requirements of the downstream application. Remember that the volume of eluate will be up to 5 µl less than the volume of the solution applied to the column.

13. Close the lid and incubate at room temperature (15–25°C) for 1 min. Centrifuge at full speed (20,000 x *g*; 14,000 rpm) for 1 min. Incubating the QIAamp MinElute column loaded with Buffer AE or water for 5 min at room temperature before centrifugation generally increases DNA yield.

Appendix E: DNA extraction QC, WGS cellularity estimates and all-samples SCNA mapping

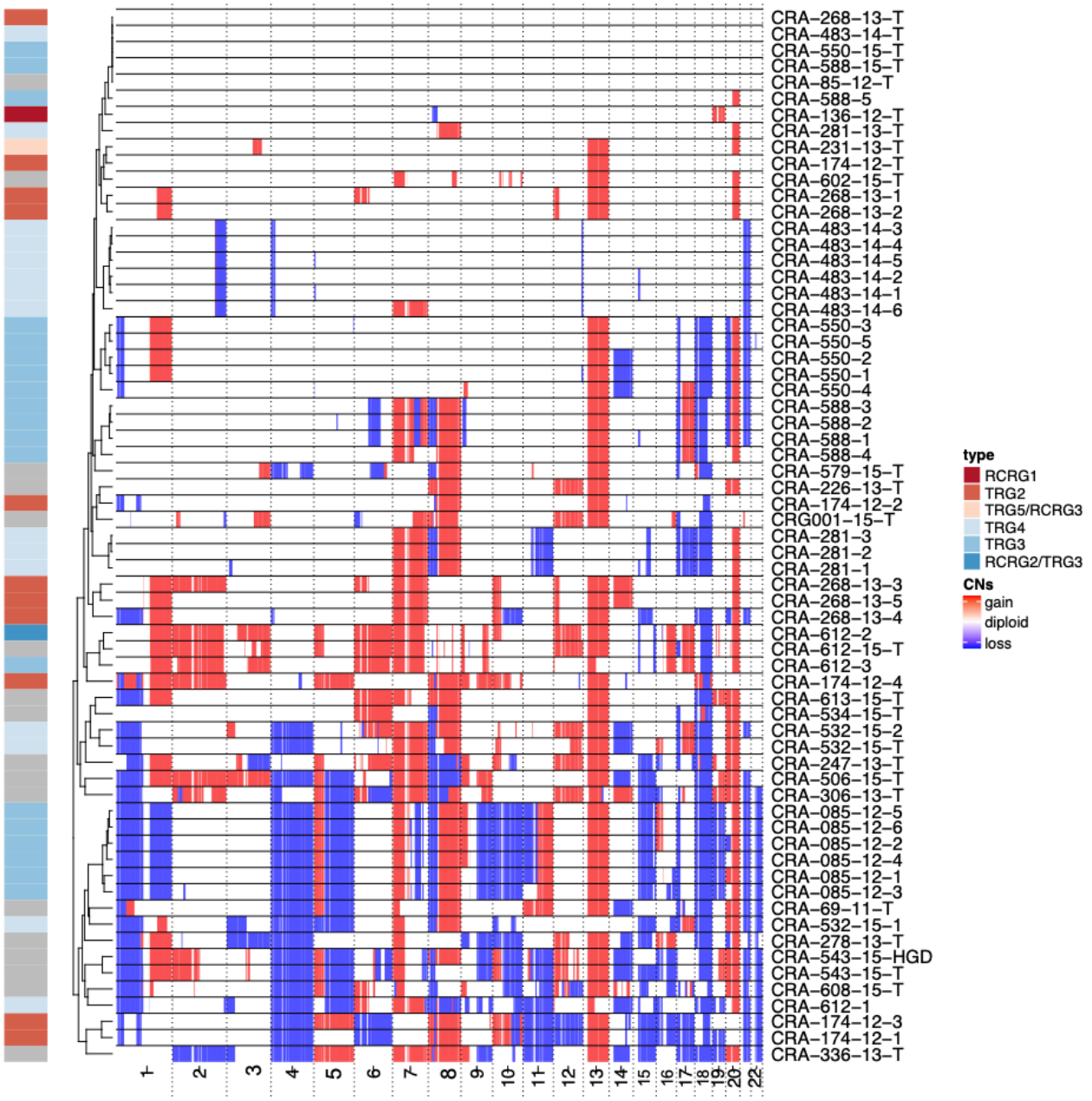


Appendix E1 – GAPDH PCR gel electrophoresis image for selected QC of DNA preparations prior to lpWGS. Lane 1 = molecular size standard with first four bars representing 100, 200, 300 and 400bp fragments. Samples in lanes 2-7 demonstrating presence of all size fragments (CRA069-11, CRA085-12, CRA226-13, CRA247-13, CRA306-13, CRA336-13). Positive and negative controls in lanes 8 and 9 respectively. Data on primers used and conditions as per index publication by van Beers et al. [353]

Sample_ID	Cellularity	Notes
CRA-136-12-T	0.6	
CRA-174-12-T	0.3	Very little tumour
CRA-226-13-T	0.7	
CRA-231-13-T	0.6	
CRA-247-13-T	0.7	
CRA-268-13-T	0.1	Little-no tumour
CRA-278-13-T	0.7	
CRA-281-13-T	0.3	Little tumour
CRA-306-13-T	0.7	
CRA-336-13-T	0.7	
CRA-483-14-T	0.1	Little-no tumour
CRA-506-15-T	0.7	
CRA-532-15-T	0.5	
CRA-534-15-T	0.7	
CRA-543-15-HGD	0.7	
CRA-543-15-T	0.7	
CRA-550-15-T	0.1	Little-no tumour
CRA-579-15-T	0.7	
CRA-588-15-T	0.1	Little-no tumour
CRA-602-15-T	0.6	
CRA-608-15-T	0.7	
CRA-612-15-T	0.7	
CRA-613-15-T	0.7	
CRA-69-11-T	0.9	
CRA-85-12-T	0.1	Little-no tumour
CRG001-15-T	0.7	
CRA-085-12-1	0.7	
CRA-085-12-2	0.7	
CRA-085-12-3	0.7	
CRA-085-12-4	0.7	
CRA-085-12-5	0.7	
CRA-085-12-6	0.6	
CRA-174-12-1	0.5	
CRA-174-12-2	0.6	
CRA-174-12-3	0.6	
CRA-174-12-4	0.4	
CRA-268-13-1	0.4	
CRA-268-13-2	0.2	Little tumour
CRA-268-13-3	0.4	
CRA-268-13-4	0.4	
CRA-268-13-5	0.6	
CRA-281-13-1	0.6	
CRA-281-13-2	0.6	
CRA-281-13-3	0.6	
CRA-483-14-1	0.6	
CRA-483-14-2	0.6	
CRA-483-14-3	0.7	
CRA-483-14-4	0.7	
CRA-483-14-5	0.6	

CRA-483-14-6	0.6	
CRA-532-15-1	0.6	
CRA-532-15-2	0.6	
CRA-550-15-1	0.6	
CRA-550-15-2	0.6	
CRA-550-15-3	0.6	
CRA-550-15-4	0.6	
CRA-550-15-5	0.6	
CRA-588-15-1	0.4	
CRA-588-15-2	0.6	
CRA-588-15-3	0.6	
CRA-588-15-4	0.5	
CRA-588-15-5	0.3	Little tumour
CRA-612-15-1	0.7	
CRA-612-15-2	0.7	
CRA-612-15-3	0.7	

Appendix E2 – Cellularity estimates for all (pre- and post-RT) lpWGS sequencing runs



Appendix E2 – Comprehensive mapping of all samples sequenced with lpWGS and estimated SCNA gains and losses plotted. From the pre-treatment, multi-region samples, clustering occurs in relation to eventual TRG.

Appendix F: Published paper

ARTICLE IN PRESS

Review

Network Mapping of Molecular Biomarkers Influencing Radiation Response in Rectal Cancer

Liam Poynter,[1] Dieter Galea,[2] Kirill Veselkov,[2] Alexander Mirnezami,[3] James Kinross,[1] Jeremy Nicholson,[2] Zoltán Takáts,[2] Ara Darzi,[1] Reza Mirnezami[1,4]

Abstract

Preoperative radiotherapy (RT) plays an important role in the management of locally advanced rectal cancer (RC). Tumor regression after RT shows marked variability, and robust molecular methods are needed to help predict likely response. The aim of this study was to review the current published literature and use Gene Ontology (GO) analysis to define key molecular biomarkers governing radiation response in RC. A systematic review of electronic bibliographic databases (Medline, Embase) was performed for original articles published between 2000 and 2015. Biomarkers were then classified according to biological function and incorporated into a hierarchical GO tree. Both significant and nonsignificant results were included in the analysis. Significance was binarized on the basis of univariate and multivariate statistics. Significance scores were calculated for each biological domain (or node), and a direct acyclic graph was generated for intuitive mapping of biological pathways and markers involved in RC radiation response. Seventy-two individual biomarkers across 74 studies were identified. On highest-order classification, molecular biomarkers falling within the domains of response to stress, cellular metabolism, and pathways inhibiting apoptosis were found to be the most influential in predicting radiosensitivity. Homogenizing biomarker data from original articles using controlled GO terminology demonstrated that cellular mechanisms of response to RT in RC—in particular the metabolic response to RT—may hold promise in developing radiotherapeutic biomarkers to help predict, and in the future modulate, radiation response.

Clinical Colorectal Cancer, Vol. ■, No. ■, ■-■ © 2019 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Keywords: Neoadjuvant therapy, Radiation tolerance, Radiotherapy, Rectal neoplasms

Introduction

Total mesorectal excision represents the surgical standard of care in rectal cancer (RC) and has led to significant improvements in local control over the past 3 decades.[1] The additive benefit of radiotherapy (RT) in reducing local recurrence in locally advanced RC (T3/T4 and/or node-positive tumors) has been confirmed by several landmark randomized controlled trials (RCTs).[2-5] In the majority of patients, RT leads to clinically meaningful tumor regression. However, there is considerable variability in terms of tumor response, such that up to 20% of tumors undergo minimal regression, while an estimated 15% exhibit a complete radiologic response. The remainder have an intermediate response, with varying degrees of residual viable tumor seen in the postoperative resection specimen. Correspondingly, patients will fall broadly into 3 groups: (1) those in whom RT has led to complete tumor destruction; in this group of patients, there is growing interest in the feasibility of organ preservation[6,7]; (2) those in whom RT will result in at least a moderate degree of tumor regression, which in turn may enhance the likelihood of R0 resection and/or sphincter preservation; and (3) those in whom preoperative RT will lead to negligible tumor shrinkage. Paradoxically, for patients in the last category, the delay in proceeding to tumor excision while completing RT may increase the likelihood of distant metastases. Thus, the development of reliable methods to predict response at the pretreatment phase represents a critical unmet need in order to personalize locally advanced RC treatment algorithms.

Current understanding of response to RT is limited by the lack of a unifying interpretation of the molecular pathways implicated in response mechanisms. Additionally, systematic reviews in this context have served little to no role in expanding current

[1]Department of Surgery & Cancer, Imperial College London, London, UK
[2]Computational & Systems Medicine, Imperial College London, London, UK
[3]Department of Surgical Oncology, University of Southampton, Southampton, UK
[4]St Mark's Hospital and Academic Institute, Harrow, London, UK

Submitted: Jun 19, 2018; Revised: Nov 12, 2018; Accepted: Jan 23, 2019

Address for correspondence: Reza Mirnezami, PhD, FRCS, Department of Surgery & Cancer, Imperial College London, 10th Floor QEQM Building, St Mary's Hospital, South Wharf Road, London W2 1NY, United Kingdom
E-mail contact: r.mirnezami@imperial.ac.uk

1533-0028/© 2019 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
https://doi.org/10.1016/j.clcc.2019.01.004

understanding of this topic, as they typically evaluate data as a narrative series of disconnected molecular entities, with varying experimental methodology and clinical outcome measures and no wider appreciation of biological interconnectivity. While biomarkers, commonly oncogenes, are important in cancer diagnosis, such genes do not occur in isolation, and therefore mechanisms such as those governing tissue response to RT are likely to involve multiple biological pathways and processes at genomic, transcriptomic, and metabolomic levels. Thus, it is desirable to map potential biomarkers more holistically, in such a way as to allow integration of heterogeneous molecular data and intuitive visualization of the resulting network in order to identify areas with greatest translational potential.

Here we present a systematic review and network analysis of biomarkers implicated in RC radiosensitivity, with the aim of mapping key molecular processes involved, statistically estimating their relative importance, and permitting visualization of how these are interconnected.

Methods

An electronic literature search of the Medline and Embase databases was performed for English-language articles published between January 2000 and November 2015, utilizing Boolean logic and controlled vocabularies (Medical Subject Headings [MeSH] and Emtree). No review protocol was published in advance of this review. The following keywords were used, with wildcards used to maximize article capture: rectal, cancer, neoplasm, response, pathological complete response, tumor response, radiotherapy, chemoradio*, neoadjuvant, predict*, biomarker, *omic, RNA, DNA. Both human in vivo and human colorectal cancer (CRC) cell-based in vitro studies were included. Inclusion criteria were studies examining potential biomarkers (specific genes, protein products, or metabolites) correlated with a documented tumor regression grade (TRG) after RT (either long-course chemoradiotherapy [LCCR] or short-course RT [SCRT]). Studies were excluded if they evaluated targeted monoclonal antibody therapy, gene expression panels obtained through microarrays, micro-RNAs (miRNA) where the gene targets were not defined, or pure animal model–derived data. miRNA studies were initially screened but were excluded from this review primarily because the corresponding gene targets of respective miRNAs were not defined in any of the studies. Irrespective of this, with the potential for a single miRNA to target multiple genes, the inability to assign each miRNA to a single gene classifier was deemed incompatible with the objectives of the present study. Similarly, there was poor congruence between studies that used tissue microarrays, and the resultant outputs from these were almost universally reported as panels of multiple genes providing an in silico means for predicting response, and as such, these data were noncompliant with our objectives.

The following information was extracted from each publication: (1) studied biomarker, (2) analytical methodology used, (3) patient numbers, and (4) statistical parameter used by the authors to categorize discovery as significant (or not), according to correlation with TRG (including P values and odds ratios [ORs] for univariate tests, and sensitivity and specificity for multivariate tests). Each biomarker was then classified according to the system proposed by the Gene Ontology (GO) Consortium into 1 of 5 domains of molecular function or biological purpose (cell cycle, cell death, response to stimuli, cell signaling, and cellular metabolism).[8] The GO approach represents a relatively recent development in biological computational science that offers a means of annotating and grouping molecular products of upstream gene expression according to functional/biological commonality. We chose to utilize a GO approach here in order to generate a biologically more interpretable panorama of processes governing radiation response.

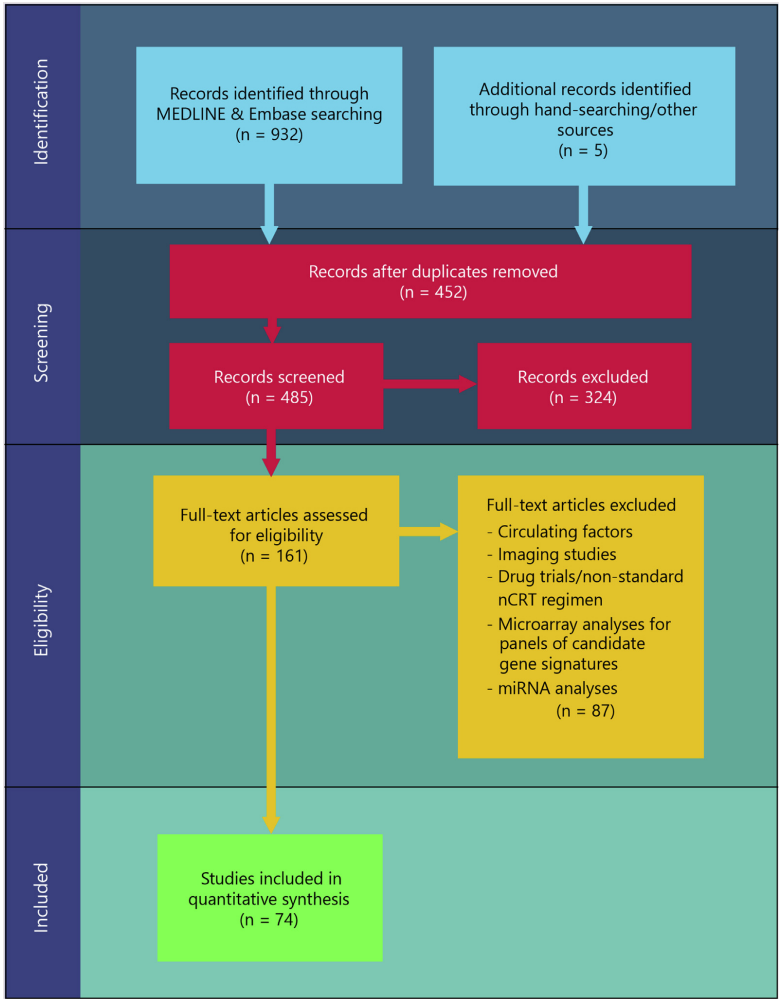
Depending on the statistical parameter used in the original study, significance was binarized (significant or not) based on P values ($P < .05$), OR (> 1), or sensitivity and specificity ($> 60\%$). A score summarizing the significance of a given biomarker was in turn calculated as a ratio of significant biomarkers relative to the total number of biomarkers under a given ontology heading. Nonunique biomarkers reported by different studies were included in this scoring, as the magnitude of supporting or conflicting evidence for a given biomarker by different studies needs to be represented. For each ontology node, a second score was computed to represent the total number of biomarkers and publications investigating this domain. This was calculated as the number of unique biomarkers for a given GO term (unlike in the first score, where nonunique biomarkers were considered). The 2 scores (each ranging between 0 and 1) represent 2 dimensions on the reported biomarkers we identified through systematic review: relative significance and popularity.

Results

A total of 452 studies were identified after removal of articles that did not include full text (mostly conference proceedings) and manual deduplication. The abstracts of these articles were further screened, and 373 were excluded because they did not address the question of tissue biomarkers predicting response to RT in RC, or because there was insufficient information on methodology or statistical analysis. A further 11 excluded articles comprised systematic reviews and meta-analyses, the bibliographies of which were hand searched, providing a further 5 articles (Figure 1). This left 74 articles that provided sufficient methodologic and statistical detail for both systematic review and to enable integration into a node-based biological network of tissue-based biomarkers. For ease of discussion, biomarkers evaluated in these studies were analyzed on the basis of biological and molecular function according to GO, and clustered under broad, highest-level biological function using controlled language, as previously described.[8,9] For many biomarkers with multiple GO annotations, clustered GO terms fell under different domains of biological function and so were recorded as separate entries in a data sheet of hierarchical ontology terms for each biomarker. It should be noted, however, that many of the biomarkers assessed have multiple functions and GO annotations; this is taken into account in the network-driven analytics.

The data obtained from the review process were extracted in such a way as to assign an identifier to each biomarker assessed by each study (Table 1). This generated a list of biomarkers with multiple entries for those that had been the subject of investigation for more than one study. Equally, to account for statistical power, separate entries were made for each statistical test (univariate or multivariate) applied to each biomarker. The full data extraction table is provided in the Supplemental Material.

Figure 1 Search Strategy (PRISMA) Outlining Search Methodology and Article Inclusion



By using GO terms, we were able to build a hierarchical classification for each biomarker that allowed for clustering under 1 of 5 subheadings (cell cycle, cell death, response to stimuli, cell signaling, and cellular metabolism). A network model was then developed (Figure 2). In this model, nodes are annotated according to their GO term and are clustered in spokes around the root term "biological process." Nodes are sized according to a significance score for that ontology term, calculated by taking into account the various statistical tests applied to each study (most commonly chi-square association or OR between biomarker and TRG in univariate and multivariate analysis, respectively). The size of the node represents the number of biomarkers investigated under that ontology term, and the color scale indicates the relative significance level. Therefore, a large red node would represent an ontology term with both a large number of biomarkers and high statistical significance.

The GO-based analytical approach used here shows that when the vast number of assessed biomarkers in the literature are

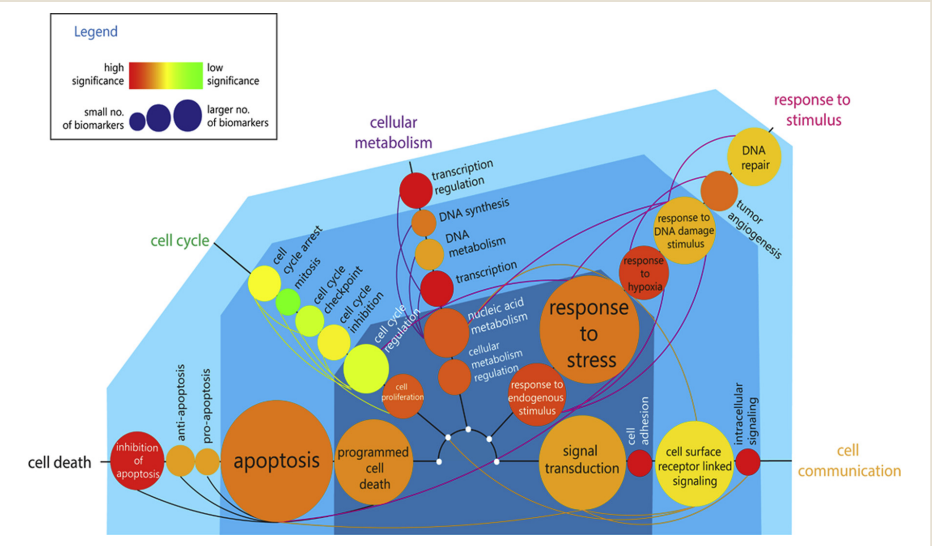
Table 1 Biomarkers Classified According to GO Hierarchical Tree (Biological Functions)

GO Cluster (Biological Function)	GO Node	Biomarker and References	Node Size	Node Significance (HSB Color Space)
Cell cycle and cell proliferation	Cell cycle arrest	***BIRC5***,[10-13] p21,[14-16] p53[14-19]	29.42857	57.77778
	Cell cycle checkpoint	p21,[14-16] p53[14-19]	25.71429	74.28571
	Cell cycle inhibition	p27,[16-20] BIRC5[10-13]	29.42857	29.42857
	Cell-cycle regulation	BIRC5,[10-13] p21,[14-16] p27,[16,20] p53,[14-19] phospho-Akt[21]	40.57143	68.8
	Cellular proliferation	Ki-67,[22-24] REG4,[11,25] Securin,[26] YKL-40[27]	36.85714	24
	Mitosis	β-Tubulin[28]	22	120
Cell death	Programmed cell death	Apoptotic index,[12,29] Bax,[14,16,30] BIRC5,[10-13] caspase-8,[31] hPEBP4,[32,33] M30,[34] MIB1,[15] p53,[14-19] PDCD4,[35] phospho-Akt,[21] XIAP,[36] YKL-40[27]	70.28571	35
	Apoptosis	Apoptotic index,[12,29] Bax,[14,16,30] BIRC5,[10-13] caspase-8,[31] hPEBP4,[32,33] M30,[34] MIB1,[15] p53,[14-19] PDCD4,[35] phospho-Akt,[21] XIAP,[36] YKL-40,[27] CD34,[37] calsenilin,[28] Smac[24,38]	100	31.11111
	Proapoptosis	Bax[14,16,30]	22	40
	Inhibition of apoptosis	BIRC5,[10-13] hPEBP4,[32,33] phospho-Akt,[21] XIAP,[36] YKL-40,[27] HIF-1,[39] HIF-1α[14]	48	5.714286
Response to stimuli, stress and DNA damage	Response to endogenous stimulus	CA9,[40] c-met,[27] CXCL10,[41] MSH2,[15] phospho-Akt,[21] thymidine phosphorylase,[37] VEGF,[15,16,24,37,39,42-45] YKL-40[27]	51.71429	18.94737
	Response to stress	ANXA1,[46] BIRC5,[10-13] DIABLO,[36] DNAJC12,[47] HIF-1,[39] HIF-1α,[14] hOGG1-1245C > G,[48] MTHFR 677T-1298A,[49] MTFHR-677C > T[48] Hsp42,[28] Hsp90,[21] m-TIMP3,[50] NEIL2,[11] NF-κB,[10,51] p21,[14-16] p53,[14-19] sHsp16.2,[21] Smac,[24,38] XRCC2,[52] tropomodulin[28]	88.85714	31.11111
	Response to hypoxia	HIF-1,[39] HIF-1α,[14] CA9,[40] c-met,[27] CXCL10,[41] phospho-Akt,[21] thymidine phosphorylase,[37] VEGF,[15,16,24,37,39,42-45] YKL-40[27]	44.28571	16.47059
	Response to DNA damage stimulus	MSH2,[15] MTHFR 677T-1298A,[49] BIRC5,[10-13] hOGG1-1245C > G,[48] MTFHR-677C > T,[48] NEIL2,[11] NF-κB,[10,51] p53,[14-19] XRCC2[52]	55.42857	43.47826
	DNA repair	MSH2,[15] BIRC5,[10-13] hOGG1-1245C > G,[48] MTFHR-677C > T,[48] NEIL2,[11] p53,[14-19] XRCC2[52]	48	48
	Tumor angiogenesis	VEGF,[15,16,24,37,39,42-45] phospho-Akt,[21] thymidine phosphorylase,[37] YKL-40[27]	33.14286	28.57143
Cell communication	Signal transduction	ABCC4,[53] Bcl-2,[16,45] β-catenin,[45,54,55] CD44,[10,14] CD133,[14,56] CD166,[14] c-met,[27] CXCR4,[40] EGFR,[14,15,22,39] EpCAM,[14] GHRH-R,[21] HER-2,[45,57] IQGAP1,[58] MRP3,[59] TGF-β1,[60] THBS2[61]	77.71429	40
	Cell-surface receptor signaling	ABCC4,[53] CD34[37], CD44,[10,14] CD133,[14,56] CD166,[14] CXCR4,[40] EGFR,[14,15,22,39] EpCAM,[14] GHRH-R,[21] HER-2,[45,57] MRP3,[59] TGF-β1,[60] calsenilin[28]	70.28571	52.63158
	Intracellular signaling	IQGAP1[58]	22	0
	Cell adhesion	THBS2[61]	22	0
Cellular metabolism	Regulation of cellular metabolism	COX-2,[16,17,37,45,56,62] ALDH1,[14,63] HMGCS2,[64,65] HSD17B2,[65] MMP9,[43] phospho-Akt,[21] GLUT-1,[18,40] CPS1,[66] asparagine synthetase[67]	29.42857	24
	Nucleic acid metabolism	p-SMAD3,[60] SMAD3,[60] TCF4,[68] thymidine phosphorylase,[37] thymidylate synthase[10,14,15,22,23,69-72]	40.57143	24
	DNA synthesis	Thymidylate synthase[10,14,15,22,23,69-72]	22	31.11111
	DNA metabolism	Thymidine phosphorylase[37]	25.71429	40
	Transcription	p-SMAD3,[60] SMAD3,[60] TCF4[68]	29.42857	0
	Regulation of transcription	p-SMAD3,[60] SMAD3,[60] TCF4[68]	29.42857	0

Node size was calculated based on number of biomarkers within each GO term, accounting for sample size in each study. Number of references for each biomarker reflects number of studies retrieved. Significance was mapped to HSB color space to reflect overall significance of each node in resulting network, with lower relative value reflecting greater significance level based on multiple statistical tests used across studies.
Abbreviations: GO = Gene Ontology; HSB = hue, saturation, brightness.

clustered according to controlled language, it is possible to begin to identify patterns that may aid direction of future study by highlighting areas where there has been success to date. Figure 2 indicates that the biological function spoke "cell cycle" demonstrates a number of biomarkers that have shown negative results in experimental work. Thus, despite this being a fundamental component of cancer regulation, cell-cycle–based molecular dynamics may have little to do with radiation response. We

Figure 2 Static Knowledge Network Visualization of Biomarkers Clustered According to GO Terminology. As Per Data From Table 1. Each Node Represents GO Term, Clustered Under Group Headings (Spokes). Size of Nodes Was Calculated Based on Number of Biomarkers Under That GO Term, With Color/Hue/Saturation/Brightness Mapped to Aggregate Significance of Multiple Statistical Tests Used Across Studies



found that the greatest numbers of significant biomarkers lie in the domains of "response to stimulus" and the interplay of this domain with "cell death." In addition, while a relatively small number of biomarkers responsible for cellular metabolism have been evaluated to date, many of these demonstrate a high level of correlation with TRG.

Discussion

Cell Cycle

Numerous genes involved in cell-cycle regulation and cellular proliferation have been implicated in cancer development and progression. The p53 tumor suppressor network has been extensively studied and is ubiquitously implicated in almost all cancer subtypes. At its core, p53 can be broadly considered a tumor suppressor activated in response to stress, with subsequent effects on apoptosis, cell-cycle arrest, and senescence, with loss of p53 function being associated with decreased apoptosis.[16,73] Although certain radioresistant CRC cell lines demonstrate decreased expression of p53, multiple immunohistochemical (IHC) studies aimed at identifying it as a candidate biomarker for radioresistance have not shown a significant association between isolated p53 expression and TRG.[14,16-19,45] One 2014 study in which IHC was performed for a panel of biomarkers including p53, revealed a correlation between low p53 expression and pathologic complete response, but not with TRG,[42] whereas Kelley et al[31] found that intact p53 with deficiency in caspase-8 expression predicted inferior TRG.

These conflicting findings may in part relate to the complex interactions of the p53 network. In a study of 112 patients, both univariate and multivariate analysis found that p53 expression in pretreatment biopsy specimens did not correlate with RT response, whereas p21 expression—the product of which is a cyclin-dependent kinase (CDK) inhibitor acting through p53-dependent and independent pathways—was significantly associated with both TRG and disease-free survival.[14] By contrast, 2 further studies examining both p21 and p53 did not demonstrate any correlation between p21 expression and TRG.[15,16]

Downstream, a cross-regulatory network exists between p53 and protein kinase B (Akt), balancing the apoptotic regulation by p53 with the effects of the Akt signaling pathway. Phosphorylated Akt exerts subsequent effects on cell-cycle progression, metabolism, and inhibition of apoptotic pathways through Bad,[74] although its expression was not found to be predictive of response to RT when evaluated with IHC.[21] Other CDK inhibitors, notably p27, have also failed to demonstrate any correlation with response to RT.[16,20] Markers of cell-cycle progression are, however, useful surrogate markers of proliferation, and both Ki-67 and securin (pituitary tumor transforming gene 1, *PTTG1*) have been evaluated for correlation with response to RT. Whereas Ki-67 is detected in all active phases of the cell cycle, securin peaks in the G_2/M phases. Significant correlation has been demonstrated between Ki-67 expression and response to RT.[22-24] Although securin did not demonstrate a predictive capacity for TRG, expression levels in general were shown

to decrease after RT; furthermore, for those tumors with high securin expression after treatment, overall prognosis was poorer.[26]

The regenerating islet-derived gene 4 (*REG4*) is one of a family of genes (*Reg*) belonging to the calcium-dependent lectin superfamily, encoding a group of small secretory proteins. These have been found to be constitutively expressed in cell lines of the gastrointestinal tract with characteristic up-regulation in RC at both messenger RNA (mRNA) and peptide levels.[75,76] In cellular models, it has been shown to activate cell proliferation, in addition to in vitro migration and invasion of CRC cells.[77] Increased expression of *REG4* in radioresistant RC cell lines led to a subsequent up-regulation of *NEIL2* and survivin (*BIRC5*), suggesting a network correlation of expression of these genes and their products with radioresistance.[11] Additionally, IHC for *REG4* expression in 172 clinical specimens revealed increased *REG4* expression to be a surrogate marker for radioresistance.[25] Survivin acts as both a controller of mitotic progression and inhibition of apoptosis, and it has been demonstrated to play a role in the molecular pathogenesis of CRC.[78] Conflicting evidence exists, however, for survivin as a putative biomarker for radioresistance. Whereas McDowell et al[12] did not find survivin alone to be a predictor of response when assessed by both mRNA and protein expression, other studies have found a significant correlation in both in vivo and in vitro models with established radioresistant cell lines, including SW480 and HCT-15.[10,11,13]

Although the full role and molecular function of the glycoprotein YKL-40 (chitinase-3-like-1) is yet to be elucidated, it is understood to regulate a number of cellular processes, including cellular proliferation, by exerting effects on the tumor microenvironment to stimulate angiogenesis and tissue remodeling in CRC.[79,80] A multicenter study utilizing IHC to verify YKL-40 expression in biopsy specimens found a high positive predictive value for radioresistance. This predictive value was increased to 94% when evaluating for biomarker coexpression with the c-Met protooncogene.[27]

Cell Death

Cell death covers a range of biological processes by which the cell ceases to function, either as a result of an intrinsic, programmed pathway or in response to extrinsic stressful stimuli such as exposure to cytotoxic agents or RT. A number of mediators of the apoptotic pathway have shown potential as predictors of radiosensitivity.

A critical step in apoptotic signaling is the release of cytochrome *c* from mitochondria, which in turn activates apoptosis protease-activating factor 1 (APAF-1/apoptosome). After the activation of the initiator caspase-9, a series of reactions ensue that result in apoptosis and that are regulated by a balance of pro- and anti-apoptotic mediators. APAF-1 expression in pretreatment biopsy samples has been shown to correlate significantly with an improved TRG. Evidence for downstream mediators such as Bax (proapoptotic) and Bcl-2 (antiapoptotic) as predictive biomarkers has been inconclusive,[14,16,30] although one group found a significant association between Bcl-2 and TRG in both univariate and multivariate analysis.[45] Furthermore, programmed cell death 4 (PDCD4) was first shown to be up-regulated in apoptosis by Shibahara et al[81] and is now known to act as a nuclear mediator of apoptosis having interactions with factors downstream of Akt. In a recent study PDCD4 expression was shown to correlate positively with TRG on multivariate analysis.[35]

Although some correlation has been observed between apoptotic rate in pretreatment CRC biopsy samples and response to RT,[12,29] expression levels of M30 (caspase-cleaved keratin 18 cytoskeletal protein) have not been shown to be predictive of response to RT.[34] Similarly, annexin V—an established marker for apoptosis, as it binds to phosphatidylserine molecules on the cell surface (an event observed in apoptotic cells)—was not found to be predictive of response.[28] In spite of this, there is increasing evidence that the presence of higher levels of complex lipids is not only characteristic of the tumor microenvironment but may also contribute to radioresistance. Levels of phosphatidylethanolamine binding protein 4 (hPEBP4) have been shown to correlate strongly with increased resistance to RT in a cohort of prospectively recruited patients who received SCRT.[33] Functionally, hPEBP4 is a secreted glycoprotein that has been shown to be involved in the activation of Akt.[82] Overexpression of hPEBP4 in SW480 cell lines has been shown to enhance clonogenic survival, a finding mirrored in mouse models subjected to ionizing radiation. Inhibition of Akt activation additionally appeared to reverse the radioresistance effect of hPEBP4.[32]

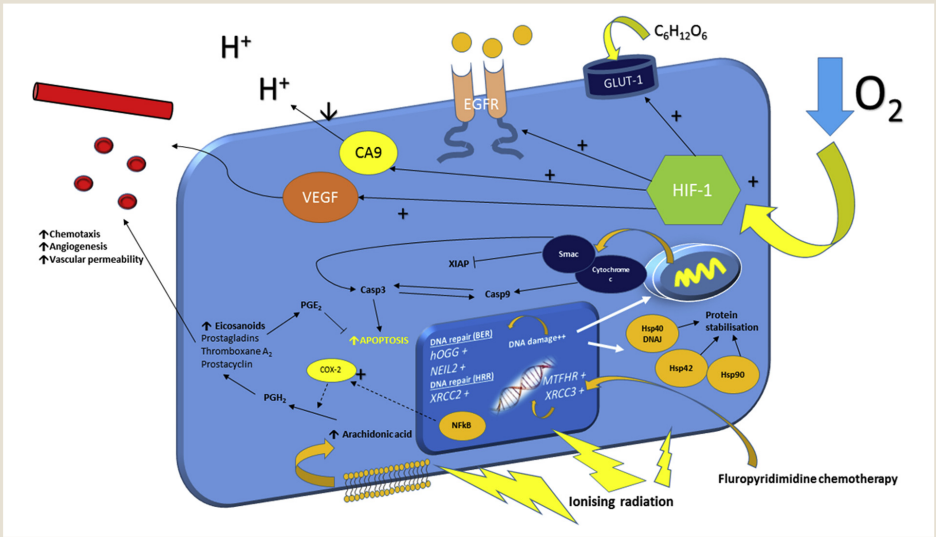
Response to Stimuli

Ionizing radiation leads to DNA strand breaks and ultimately cell death. Mediators involved in maintaining cellular homeostasis and responding to changes in the tumor microenvironment have shown variation in expression levels in relation to radiosensitivity. The presence of relative tumor hypoxia has been implicated in radioresistance in a number of malignancies. Central to the regulation of over 100 genes involved in the adaptive response to hypoxia is hypoxia-inducible factor 1 (HIF-1).[83,84] The effects of this transcription factor, stabilized at low oxygen tensions, include regulation of apoptosis, angiogenesis, glycolysis, and the cell cycle. A number of these downstream genes and products have subsequently been investigated as putative biomarkers for radiation response. Using quantitative PCR to evaluate HIF-1 expression in pretreatment tumor biopsy samples, Toiyama et al[39] demonstrated not only significantly lower levels of HIF-1 in those with significant tumor regression but also downstream induced factors epidermal growth factor receptor (EGFR) and vascular endothelial growth factor (VEGF) on both univariate and multivariate analyses.

The effects of hypoxia include further biochemical changes in the tumor microenvironment. CA9 (carbonic anhydrase 9), coding for carbonic anhydrase isoenzyme 9, plays an important role in regulating extracellular pH under both hypoxic and normoxic conditions, and has been found to be underexpressed in tumors with a pathologic complete response.[40] This study broadly examined a number of mechanisms of the adaptive response of RC to hypoxia, including glucose transporter 1 (GLUT-1) expression, which had been shown in an earlier study to be significantly associated with radioresistance. Conversely, GLUT-1−negative tumors were found to have a 70% probability of a good response to RT. These findings[18] were not corroborated by Guedj et al,[40] however, who found no significant association between GLUT-1 expression and TRG. It is worth noting that these 2 studies used differing intervals between RT and surgery, and also used different TRG grading systems.

Figure 3 summarizes the factors identified through this review that have been found to correlate with response to RT, based on

Figure 3 **Factors Implicated in Response to Cellular Damage Due to Unfavorable Tumor Microenvironmental Conditions (Acidosis, Hypoxia) and RT. Individual Variation in These Pathways Are Those Shown to Be Implicated in Differing Responses to RT. (A) At Low Oxygen, Tension HIF-1 Is Up-regulated, With Corresponding Increased Expression of Downstream Cofactors; EGFR Is Up-regulated on Cell Surface; Carbonic Anhydrases Are Released in Order to Stabilize Extracellular pH, With CA9 Particularly Shown to Be Independent Predictor for RR; VEGF Stimulates Angiogenesis in Tumor Microenvironment. GLUT-1 Is Not Expressed in Normal Colonic Mucosa But Has Been Shown to Be Present in up to 80% of Colorectal Tumors. (B) Direct DNA Damage Leads to Activation of Number of DNA Repair Factors (hOGG, MTFHR, NEIL2) in Addition to Increased Levels of Heat Shock Proteins (Particularly Hsp40 [DNAJ], Hsp42 and Hsp90), Increased Levels of All of Which Have Shown Direct Relationship With RR. DNA Damage Stimulates Release of Cytochrome *c* and Activated Smac Proteins From Mitochondria, and Subsequent Activation of Apoptosome and Direct Inhibition of XIAP Leads to Apoptosis Under Normal Circumstances. XIAP Expression in Turn Has Been Demonstrated as Independent Predictor of Radioresistance.[37] Decreased Relative Levels of Smac Were Shown to Be Associated With RR. There May Be Additional Interplay With Smac Contributing to Direct Activation of Caspase-3. (C) Ionizing Radiation Leads to Increased Degradation of Complex Lipids in Cell Membrane and Subsequent Release of Higher Levels of Arachidonic Acid. Higher Levels of COX-2 Have Been Shown to Correlate With RR, With Subsequent Increased Levels of Eicosanoids and Their Downstream Effects in Both Intra- and Extracellular Environment. Within Cytosol, Both COX-2 and Eicosanoids Have Been Shown to Inhibit NF-κB, While in Tumor Microenvironment PGs and Thromboxane A_2 Have Well-established Effects on Chemotaxis, Vascular Permeability, and Angiogenesis as Part of Local and Systemic Inflammatory Process. Arrows Denote Signaling Pathway or Up-regulation; Hammerheads Denote Inhibition**



Abbreviations: BER = base excision repair; CA9 = carbonic anhydrase 9; Casp3/Casp9 = caspase 3/9; COX-2 = cyclo-oxygenase 2; EGFR = endothelial growth factor receptor; GLUT-1 = glucose transporter 1; HIF-1 = hypoxia-inducible factor 1; HRR = homologous recombinant repair; Hsp = heat shock protein; PGE2/PGH2 = prostaglandin E_2/H_2; RT = radiotherapy; VEGF = vascular endothelial growth factor; XIAP = X-linked inhibitor of apoptosis protein.

individual variations. Pathway A outlines how HIF-1 relates to other markers identified though these studies in the presence of hypoxia. The pathways directed at mounting a local response to ionizing radiation, repairing DNA double-strand breaks and stabilizing protein structures, or activating apoptosis are summarized in pathways B and C.

The primary aim of RT is to cause lethal double-stranded DNA (DS-DNA) breaks resulting in cell death. A number of genes involved in base-excision repair pathways have already been well studied. *NEIL2* belongs to a family of DNA glycosylases that initiates the first step in base exicision repair. *NEIL2* has been shown to be up-regulated in radioresistant RC cell lines as well as in pretreatment clinical biopsy samples of patients whose disease had a poor response to RT.[11] Equally implicated in the base excision repair pathway, polymorphisms within the *hOGG* gene—specifically the *hOGG1 1245C > G* polymorphism—have been shown to be associated with inferior TRG when extracted from peripheral blood and RC tissue in 238 RC patients.[48] It is thought that the 1245C > G polymorphism codes for a low-functioning isoform of the protein, resulting in less DS breaks as part of excision repair, leading to a smaller proportion of cells undergoing cell death as part of lethal DS-DNA breaks.

The X-ray repair cross-complementing proteins 2 and 3 (XRCC2 and XRCC3) are involved in a separate DNA repair pathway

(homologous recombinant repair) and have been separately evaluated as markers of radiosensitivity. Knockdown of *XRCC2* in vitro results in impaired repair of DS-DNA breaks, and Qin et al[52] concluded that intact XRCC2 proteins resulted in radioresistance by facilitating DNA repair after RT. In addition to nuclear repair mechanisms, DNA damage signals the release of cytochrome *c* and activated Smac proteins from the mitochondria, with subsequent activation of the apoptosome, as previously described. As well as preventing the inhibition of apoptotic promoters, it is postulated that Smac could also play a role in both activation and enhancement of caspase-3. Correspondingly, increased expression of Smac has been found to correlate with improved TRG.[24,38] Within the cytosol, a number of small stabilizing proteins (heat-shock proteins, Hsps) have been found to to have roles in stabilizing protein structures and maintaining cellular homeostasis. The gene *DNAJC12* codes for a 40 kDa Hsp, increased expression of which was shown to be significantly associated with inferior TRG as well as with increased vascular and perineural invasion and posttreatment nodal involvement,[47] with selected Hsps being correlated with radioresistance. Increased Hsp42 expression, verified by protein separation with 2-D gel electrophoresis and matrix-assisted laser desorption/ionization time-of-flight mass spectrometry, was associated with radioresistance in 17 clinical specimens sampled before RT.[28] Similarly, increased expression of Hsp90 in pretreatment biopsy samples has been shown to correlate with a poor response to RT.[21]

Cell Signaling

Many of the factors already discussed play a role in both cell−cell signaling and intracellular signal cascade mechanisms. Cell-surface signaling molecules have long been established as molecular fingerprints for cancer immunotyping, such as the family of clusters of differentiation (CD). This large class of molecules act as receptors or ligands, which either initiate a signaling cascade once activated or play a role in processes such as cell−cell adhesion,[10,14,37,40,56,57] although the evidence for these correlating with radiosensitivity has been inconclusive (Table 2). CXCL10 (CXC motif chemokine 10) is the ligand for the CXCR3 receptor, although this is found predominantly in T cells.[85] Presence of increased expression of CXCL10 mRNA in pretreatment RC biopsy samples was shown to be predictive of sensitivity to RT, although IHC did not corroborate this.[41] CXCL10 has been shown to act as an angiostatic, inhibiting VEGF,[41] but the correlation between mRNA expression and protein expression in this case, and hence the implication for CXCL10 to act as a surrogate biomarker is not clear.

Although numerous cell-surface adhesion molecules play a major role in the interaction between tumor and neighboring microenvironment, only the glycoprotein thrombospondin 2 (THBS2) has been shown to be significantly associated with response to RT.[14] THBS2 demonstrates antiangiogenic effects upon interaction with stromal endothelial cells, and low or absent levels of THBS2 expression in CRC biopsy samples has been shown to be linked to inferior TRG. In this study, however, 118 of the 172 specimens examined fell into an intermediate TRG (Dworak TRG 2-3), with a near 50-50 split between high and low expression of THBS2 on IHC.[61]

As with the cellular response mechanisms to stress, expression levels of a number of receptor tyrosine kinases (RTKs) have been shown to increase in RC and may be implicated in radioresistance. The downstream effects of activation of RTKs and the mitogen-activated protein kinase (MAPK) pathway on gene regulation, cellular proliferation and apoptosis as described, have highlighted these pathways as potential players in radioresistance. IQGAP1 (Ras GTPase-activating-like protein) has been evaluated as a potential biomarker on the basis of its being an essential scaffold protein in extracellular signal-regulated kinase signaling, which is a key step in the MAPK cascade.[86] It is known that MAPKs are activated—and MAPK signaling up-regulated—in response to a variety of stimuli.[87] Although this study did not evaluate levels of IQGAP1 in pretreatment biopsy samples as a stand-alone predictor of response to RT, the authors found that apical cell IQGAP1 expression increased most dramatically before and after treatment in those patients with a poor TRG (< 50% regression).[58]

Components of other signaling pathways activated by membrane receptor binding (notably the Wnt and transforming growth factor [TGF]-β signaling cascades) have failed to demonstrate conclusive results. SMAD3 transcriptionally regulates TGF-β target genes upon accumulation in the nucleus, and SMAD3 mutations in murine models have been linked to the development of CRC.[88] The phosphorylated isoform of SMAD3 is considered as a hallmark for SMAD3 activation, increased expression of which has been shown to be associated with a poor TRG.[60] β-Catenin can be detected at the membranous, cytoplasmic, and nuclear levels, representing its function within the Wnt cascade, and counts activation of genes linked to the *APC* gene among its downstream effects. A shift in treatment response has been demonstrated that is dependent on nuclear versus cytoplasmic expression of β-catenin, with increased nuclear expression being association with radioresistance.[55] However, conflicting evidence was provided in 2 other studies; no significant association was found between membranous β-catenin expression and Dworak TRG, but when correlated with the Cologne grading system, statistical significance was reached on univariate analysis.[54] A further multivariate analysis of 130 samples also demonstrated no significant correlation of β-catenin with Mandard TRG.[45] By contrast, in a single study, expression of T-cell factor 4 (TCF4)—another downstream mediator of the Wnt cascade and implicated in the adenoma−adenocarcinoma sequence—demonstrated inverse correlation with TRG.[68]

EGFR is one of a group of the RTKs that are activated and up-regulated by the stabilization of HIF-1. As with a number of other studies, conflicting results abound, with only one IHC study out of 3 identified demonstrating a link between low EGFR expression and complete pathologic response.[14,15,22] Evaluation of gene expression levels with quantitative PCR corroborated this finding, demonstrating a significant association between low expression levels of the *EGFR* gene and increased tumor regression.[39]

Growth-hormone-releasing hormone (GHRH), when bound to its receptor, stimulates local tissue proliferation via release of insulin-like growth factor 1 (IGF-1) and has been demonstrated as a feature of tumor aggressiveness, with effects on both malignant transformation and metastasis.[89] Work on RC cell lines has demonstrated that when administered in combination with a cytotoxic agent, GHRH antagonists can induce S-phase arrest and thus apoptosis in part by inhibition of these mechanisms.[90] Correspondingly, increased expression of the GHRH receptor in pretreatment RC biopsy samples has been found to correlate with radioresistance.[21]

(homologous recombinant repair) and have been separately evaluated as markers of radiosensitivity. Knockdown of *XRCC2* in vitro results in impaired repair of DS-DNA breaks, and Qin et al[52] concluded that intact XRCC2 proteins resulted in radioresistance by facilitating DNA repair after RT. In addition to nuclear repair mechanisms, DNA damage signals the release of cytochrome *c* and activated Smac proteins from the mitochondria, with subsequent activation of the apoptosome, as previously described. As well as preventing the inhibition of apoptotic promoters, it is postulated that Smac could also play a role in both activation and enhancement of caspase-3. Correspondingly, increased expression of Smac has been found to correlate with improved TRG.[24,38] Within the cytosol, a number of small stabilizing proteins (heat-shock proteins, Hsps) have been found to to have roles in stabilizing protein structures and maintaining cellular homeostasis. The gene *DNAJC12* codes for a 40 kDa Hsp, increased expression of which was shown to be significantly associated with inferior TRG as well as with increased vascular and perineural invasion and posttreatment nodal involvement,[47] with selected Hsps being correlated with radioresistance. Increased Hsp42 expression, verified by protein separation with 2-D gel electrophoresis and matrix-assisted laser desorption/ionization time-of-flight mass spectrometry, was associated with radioresistance in 17 clinical specimens sampled before RT.[28] Similarly, increased expression of Hsp90 in pretreatment biopsy samples has been shown to correlate with a poor response to RT.[21]

Cell Signaling

Many of the factors already discussed play a role in both cell−cell signaling and intracellular signal cascade mechanisms. Cell-surface signaling molecules have long been established as molecular fingerprints for cancer immunotyping, such as the family of clusters of differentiation (CD). This large class of molecules act as receptors or ligands, which either initiate a signaling cascade once activated or play a role in processes such as cell−cell adhesion,[10,14,37,40,56,57] although the evidence for these correlating with radiosensitivity has been inconclusive (Table 2). CXCL10 (CXC motif chemokine 10) is the ligand for the CXCR3 receptor, although this is found predominantly in T cells.[85] Presence of increased expression of CXCL10 mRNA in pretreatment RC biopsy samples was shown to be predictive of sensitivity to RT, although IHC did not corroborate this.[41] CXCL10 has been shown to act as an angiostatic, inhibiting VEGF,[41] but the correlation between mRNA expression and protein expression in this case, and hence the implication for CXCL10 to act as a surrogate biomarker is not clear.

Although numerous cell-surface adhesion molecules play a major role in the interaction between tumor and neighboring microenvironment, only the glycoprotein thrombospondin 2 (THBS2) has been shown to be significantly associated with response to RT.[14] THBS2 demonstrates antiangiogenic effects upon interaction with stromal endothelial cells, and low or absent levels of THBS2 expression in CRC biopsy samples has been shown to be linked to inferior TRG. In this study, however, 118 of the 172 specimens examined fell into an intermediate TRG (Dworak TRG 2-3), with a near 50-50 split between high and low expression of THBS2 on IHC.[61]

As with the cellular response mechanisms to stress, expression levels of a number of receptor tyrosine kinases (RTKs) have been shown to increase in RC and may be implicated in radioresistance. The downstream effects of activation of RTKs and the mitogen-activated protein kinase (MAPK) pathway on gene regulation, cellular proliferation and apoptosis as described, have highlighted these pathways as potential players in radioresistance. IQGAP1 (Ras GTPase-activating-like protein) has been evaluated as a potential biomarker on the basis of its being an essential scaffold protein in extracellular signal-regulated kinase signaling, which is a key step in the MAPK cascade.[86] It is known that MAPKs are activated—and MAPK signaling up-regulated—in response to a variety of stimuli.[87] Although this study did not evaluate levels of IQGAP1 in pretreatment biopsy samples as a stand-alone predictor of response to RT, the authors found that apical cell IQGAP1 expression increased most dramatically before and after treatment in those patients with a poor TRG (< 50% regression).[58]

Components of other signaling pathways activated by membrane receptor binding (notably the Wnt and transforming growth factor [TGF]-β signaling cascades) have failed to demonstrate conclusive results. SMAD3 transcriptionally regulates TGF-β target genes upon accumulation in the nucleus, and SMAD3 mutations in murine models have been linked to the development of CRC.[88] The phosphorylated isoform of SMAD3 is considered as a hallmark for SMAD3 activation, increased expression of which has been shown to be associated with a poor TRG.[60] β-Catenin can be detected at the membranous, cytoplasmic, and nuclear levels, representing its function within the Wnt cascade, and counts activation of genes linked to the *APC* gene among its downstream effects. A shift in treatment response has been demonstrated that is dependent on nuclear versus cytoplasmic expression of β-catenin, with increased nuclear expression being association with radioresistance.[55] However, conflicting evidence was provided in 2 other studies; no significant association was found between membranous β-catenin expression and Dworak TRG, but when correlated with the Cologne grading system, statistical significance was reached on univariate analysis.[54] A further multivariate analysis of 130 samples also demonstrated no significant correlation of β-catenin with Mandard TRG.[45] By contrast, in a single study, expression of T-cell factor 4 (TCF4)—another downstream mediator of the Wnt cascade and implicated in the adenoma−adenocarcinoma sequence—demonstrated inverse correlation with TRG.[68]

EGFR is one of a group of the RTKs that are activated and up-regulated by the stabilization of HIF-1. As with a number of other studies, conflicting results abound, with only one IHC study out of 3 identified demonstrating a link between low EGFR expression and complete pathologic response.[14,15,22] Evaluation of gene expression levels with quantitative PCR corroborated this finding, demonstrating a significant association between low expression levels of the *EGFR* gene and increased tumor regression.[39]

Growth-hormone-releasing hormone (GHRH), when bound to its receptor, stimulates local tissue proliferation via release of insulin-like growth factor 1 (IGF-1) and has been demonstrated as a feature of tumor aggressiveness, with effects on both malignant transformation and metastasis.[89] Work on RC cell lines has demonstrated that when administered in combination with a cytotoxic agent, GHRH antagonists can induce S-phase arrest and thus apoptosis in part by inhibition of these mechanisms.[90] Correspondingly, increased expression of the GHRH receptor in pretreatment RC biopsy samples has been found to correlate with radioresistance.[21]

Table 2 Association of Cell-Surface Clusters of Differentiation Molecules With Response to NCRT

Biomarker	Function	Cell Type	Study	Association With NCRT Response
CD34	Cell-surface glycoprotein, cell—cell adhesion	Hematopoietic stem cells	Min et al 2008; IHC of pretreatment biopsy samples[37]	No significant correlation
CD44	Cell-surface glycoprotein	Some cancer stem cells	Huh et al 2014; RT-PCR and Western blot analysis for 13 markers including those for angiogenesis, apoptosis, proliferation, cell adhesion, and collagenases[10]	Low CD44 correlates with improved TRG (OR 4.694, $P = .027$)
	Cell—cell adhesion and migration	Most mammalian cell lines	Sim et al 2014; IHC for markers of cell cycle, proliferative index, apoptosis, cell adhesion and response to hypoxia[14]	No significant correlation
CD133	Cell-surface glycoprotein	Hematopoietic stem cells	Shinto et al 2011; IHC for CD133, COX-2, p53, p27, p21, EGFR in pretreatment biopsy samples of patients undergoing short-course NCRT (20 Gy/5 fractions)[56]	CD133 expression correlates with RR (OR 3.32, $P = .03$)
	Uncertain function	Cancer stem cells (brain, breast, liver, medulloblastoma, melanoma); glandular tissue cells (including GI tract)	Sim et al 2014; IHC for markers of cell cycle, proliferative index, apoptosis, cell adhesion and response to hypoxia[14]	No significant correlation with pCR
CD166	Cell-surface glycoprotein; putative roles in cell—cell adhesion and activation of matrix metalloproteinase cascades	Neurons, fibroblasts, endothelial cells, keratinocytes	Sim et al 2014; IHC for markers of cell cycle, proliferative index, apoptosis, cell adhesion and response to hypoxia[14]	High CD166 expression associated with poor prognosis (DFS) but not specifically with TRG
CD184 (CXCR4)	Chemokine receptor; lymphocyte chemotaxis	Hematopoietic stem cells; over 20 cancer subtypes (including colorectal)	Guedj at al 2011; IHC for panel of biomarkers by function; CXCR4 believed to have role in tumor cell migration and metastatic potential[40]	No significant correlation with TRG
CD340 (HER-2, ERBB2)	EGFR family cell-surface protein	Important biomarker for breast cancer	Drebber et al 2011; IHC for both HER-2/neu and β-catenin and correlated with TRG[54]	No association of HER-2 expression with TRG, although HER-2 positivity associated with survival benefit
	Activation of signal transduction pathways, including MAPK and PI3K/Akt	Up-regulation of ERBB2 gene associated with certain breast, lung, gastric, uterine and ovarian cancer cells	Meng et al 2014; IHC and FISH for HER-2 in pretreatment biopsy samples[57]	No association between HER-2 expression in tumors and TRG; positive HER-2 status associated with poorer 5-year survival

In study by Guedj et al, CD184 labeled as CXCR4 (C-X-C chemokine receptor 4). CD340 is more commonly referred to as HER-2.
Abbreviations: DFS = disease-free survival; EGFR = epidermal growth factor receptor; FISH = fluorescence in-situ hybridization; GI = gastrointestinal; HER-2 = human epidermal growth factor receptor 2; IHC = immunohistochemistry; NCRT = neoadjuvant chemotherapy; OR = odds ratio; pCR = pathologic complete response; RT-PCR = reverse transcriptase PCR; TRG = tumor regression grade.

Cellular Metabolism

The cytotoxic backbone of the majority of chemo-radiotherapeutic regimens is 5-fluorouracil (5-FU), which functions as an antimetabolite and pyrimidine analog, irreversibly inhibiting the action of thymidylate synthase. Thymidylate synthase in turn plays a key role in nucleic acid metabolism, catalyzing the conversion of deoxyuridine monophosphate (dUMP) to deoxyuridine TMP, a key nucleoside required in DNA replication. Most groups have found a correlation between low-level thymidylate synthase expression and enhanced TRG,[10,23,69,70,72] although this has not been universally reported.[14,15,71] Microarray analysis of published RC transcriptomic data sets has identified candidate genes involved in both nucleotide and amino acid metabolism that may be implicated in radioresistance, including CPS1 (carbonyl phosphate synthetase 1, an enzyme involved in glutamine metabolism and in catalyzing the initial steps in ammonia detoxification). CPS1 was identified as the most significantly up-regulated gene in the glutamine pathway in radioresistant RC; additionally, IHC staining of clinical specimens demonstrated significant correlation between CPS1 expression and radioresistance.[66]

Another enzymatic marker being explored is asparagine synthetase (ASNS), which has recently been found to serve functionality beyond amino acid synthesis; accumulation of both aspartate and glutamine in ASNS-deficient tissues leads to an increase in nucleotide synthesis and thus cellular proliferation.[91] Consequently, a significant correlation between low ASNS expression and radioresistance was observed in RC specimens, after identification of both the *ASNS* and *PAH* genes as being differentially expressed between response and nonresponse in a transcriptomic data set.[67]

As described, the distinguishing features of invasive cancer involve a multitude of genetic and proteomic anomalies, which in turn have profound downstream effects on normal cellular metabolism and bioenergetics. Metabolic phenotyping approaches (metabonomics/metabolomics) are showing increasing promise in

elucidating the metabolic factors that govern radiosensitivity. Modifications to lipid biosynthetic pathways and lipid metabolism are now established as hallmark features of cancer,[92] and intracellular lipid mediators have now been implicated in mechanisms of response to ionizing radiation and cytotoxicity. Through analysis of the gene expression database of RC tissue, *HSD17B2* and *HMGCS2* were identified as key lipid biosynthesis-associated genes with a predictive capacity for response to RT.[65] The enzyme product HSD17B2 (hydroxysteroid 17-beta dehydrogenase 2) is known to catalyze the reduction of E_2 estrogens, testosterone to androstenedione, and 5α-androstenediol to dehydroepiandrosterone.[93] HMGCS2 (3-hydroxy-3-methylglutaryl-CoA synthase 2) catalyses the formation of HMG-CoA from both acetyl-CoA and aceto-acetyl coenzyme A (CoA), a key early step in the malonate and ketogenesis pathways. This is in keeping with the Warburg phenomenon consistently displayed by solid cancers, where a propensity for anaerobic fermentation of ketone bodies is displayed, even in the presence of abundant oxygen.[94] Overexpression of both genes and their products was found to be associated with poor response to RT.

Cyclooxygenase-2 (COX-2; prostaglandin endoperoxidase synthase) converts arachidonic acid to prostaglandin H_2, which is the precursor to all other prostanoids that act principally as inflammatory mediators.[92,95] For this reason, and because arachidonic acid is released from the cell membrane after direct damage by ionizing radiation, a number of groups have evaluated the role of COX-2 in radiosensitivity.[16,17,37,45,56,62,96] In addition to the functional effects of these bioactive lipids (angiogenesis, increased vascular permeability, and chemotaxis; Figure 2), COX-2 has also been shown to directly stimulate VEGF synthesis.[37] This supports a theory of COX-2 playing an important role in the response to the cellular injury sustained after RT, although whether this is purely due to an increased amount of substrate (ie, arachidonic acid) as a result of cell membrane damage or an additional up-regulated pathway remains to be fully elucidated. Moreover, a cell-wide inflammatory response mechanism, portended by up-regulation of nuclear factor kappa B (NF-κB) in the nucleus, which in turn acts as a transcription factor for (among others) COX-2, TGF- β, VEGF, and EGFR, has been proposed.[97] However, evaluation of NF-κB as a predictive biomarker reveals conflicting results. Nuclear IHC staining for NF-κB in 74 pretreatment samples demonstrated up-regulation of NF-κB correlated with poorer overall survival, but not with TRG.[51] This may have been partly due to a type I error, however, as a subsequent larger study found that lower NF-κB expression correlated with improved TRG.[10] Furthermore, despite previous evidence that annexin 1 (ANXA1) is an endogenous inhibitor of NFκB, Sheu et al[46] found that high levels of expression of ANXA1 correlated with a poor TRG. ANXA1 exerts its effects as both a potent anti-inflammatory mediator and a calcium ion/phospholipid binder that has effects on tumourigenesis and progression.

Conclusion

Locally advanced RC represents an excellent case model for personalized multimodality therapy, although to date molecular phenotyping has had little direct impact on the RC treatment algorithm. Currently an estimated 15% to 20% of patients with locally advanced RC experience complete tumor regression after up-front RT. The value of radical surgery in these patients is increasingly challenged, and selected patients in this group could instead be offered organ-preserving treatment, such as local excision or a watch-and-wait approach. A model predictive of the RT response, applied at the pretreatment stage, will be critical to personalizing RC treatment and would facilitate organ preservation, perhaps even in patients for whom up-front RT would not otherwise be routinely considered. Molecular biomarker discovery offers the strongest opportunity for development of a radiation response predictive model. However, the inherent complexities of radiation oncobiology, coupled with the vast heterogeneity of existing works in the literature, have made it difficult to precisely define the molecular drivers of radiosensitivity. In the present study, we sought to present the RC radiation response molecular landscape as holistically as possible; we used a GO approach to group biomarkers together according to biological commonality for ease of interpretation. A more targeted search for relevant biomarkers in this context is essential, and as this work appears to demonstrate, there are several key areas that appear to be showing genuine promise and others—perhaps surprisingly—that do not.

Insofar as analyzing a heterogenous landscape of biomarkers, the generation of a knowledge-based GO network is a concise way of displaying data acquired from a comprehensive systematic review while providing both an appreciation of the statistical analyses of the original studies and insight into potential overall biological impact. We believe the work presented here is the first such analysis of its type and offers an intuitive method of visually appraising statistical data gathered from published experimental work by mapping identified biomarkers using GO terms. One benefit of such analysis is that it can take into account all published data, including nonsignificant findings, and can aid in identifying gaps in knowledge and potential targets for future work. We have shown that by clustering biomarkers according to GO terms and integrating them into a network analysis based on biological function, the ontologies under "stress response" and "cellular metabolism" appear to be of greatest significance. Conversely, biomarkers of cell-cycle regulation and progression have been found to be of low relative significance despite large numbers of studies. The work presented here should allow for more targeted direction of enquiry for future RT biomarker discovery studies, and the biostatistical methodology we used will undoubtedly be applicable to other clinical questions in which a biomarker-driven end point is the ultimate aim.

The data presented herein are subject to a number of inherent limitations. Although the methodology used has sought to mitigate the impact of these, it is important to highlight them. First, it is possible that subtle differences exist in terms of the molecular mechanisms that govern response to SCRT (25 Gy, 5 Gy in 5 fractions) and LCCR (50.4 Gy, 1.8 Gy in 28 fractions), and studies evaluating both approaches have been included in the present review. Unfortunately, the current literature on biomarkers in RC radiation response does not permit these distinctions to be defined. However, despite differences in timing and fractionation, the recent Stockholm III RCT demonstrated equivalent oncologic outcomes after SCRT and immediate surgery, SCRT and delayed surgery, and LCCR.[4] This suggests that in spite of methodologic variability, different approaches lead to relatively homogeneous oncologic end points, which most likely is because the central constituent of each

Table 3 Definitions of Tumor Regression Grading Systems as Used by Studies in This Review

Dworak et al[98]	Mandard et.al[99]	Rödel et al[100]	Ryan et al[101]	Wheeler et al[102]
0. No regression.	1. Complete regression (fibrosis without detectable tissue of tumor—pCR).	0. No regression.	1. No viable cancer cells.	1. Sterilisation or only microscopic foci of adenocarcinoma remaining, with marked fibrosis.
1. Predominantly tumor with significant fibrosis and/or vasculopathy.	2. Fibrosis with scattered tumor cells.	1. Regression of <25% of tumor mass.	2. Residual cancer outgrown by fibrosis.	2. Marked fibrosis but macroscopic disease present.
2. Predominantly fibrosis with scattered tumor cells (slightly recognizable histologically).	3. Fibrosis and tumor cells with preponderance of fibrosis.	2. Regression of 25-50% of tumor mass.	3. Significant fibrosis outgrown by cancer.	3. Little or no fibrosis, with abundant macroscopic disease.
3. Only scattered tumor cells in the space of fibrosis with/ without acellular mucin.	4. Fibrosis and tumor cells with preponderance of tumor cells.	3. Regression of >50% of tumor mass.		
4. No vital tumor cells detectable.	5. Tissue of tumor without changes of regression.	4. Complete regression (pCR).		

Many of these were not originally determined for rectal cancer but have been validated for such use.
Abbreviation: pCR = pathologic complete response.

approach is pelvic RT. We suggest that this fundamental radiobiological commonality will compensate for differences in approach, and therefore all eligible studies evaluating LCCR (n = 67), SCRT (n = 5), and a combination of these (n = 2) were included in the present study. Finally, multiple tumour regression grading systems were used across studies, and this lack of congruence may lead to difficulties in interpreting data (Table 3)

In the context of LCCR, 5-FU–based chemotherapy was almost universally used for radiosensitization, combined with fractionated long-course RT. However, we acknowledge that a minority of included studies also included use of oxaliplatin and/or leucovorin as induction agents.[10,22,24,30,35,37,57,64] Although we appreciate that this represents a further source of data heterogeneity, it is interesting to note that a number of the biomarkers assessed in these papers (eg, EGFR, Smac, and Ki-67) demonstrated congruent results with other studies where 5-FU alone was used as the radiation sensitizer.[22,24,35,57] In this context, we believe that the data were sufficiently comparable for collective inclusion in the present review.

Although previous studies have suggested that radiation response per se is not a reliable surrogate end point in recent clinical trials, we would argue that the development of reliable methods with which to predict radiation response will allow more precise treatment planning in the future for locally advanced RC.[103,104] It is feasible to surmise that were a favorable phenotype to be revealed through biomolecular profiling, an individual would be offered pelvic RT followed by radical surgery (conventional practice) or organ-preserving treatment (in the case of a complete response). In the patient in whom an unfavorable phenotype (ie, disease likely to be poorly responsive or nonresponsive to RT) is found, a different treatment approach would be required, potentially involving either beyond total mesorectal excision surgery (without preoperative therapy) or neoadjuvant chemotherapy, which is the subject of the ongoing Prospect multicenter RCT (ClinicalTrials.gov NCT01515787). Furthermore, it is important to acknowledge that the clinical landscape is evolving in RC; the introduction of watch-and-wait strategies, given our currently limited understanding of the biology underpinning radiation response, is somewhat controversial. The ability to robustly predict response to RT based on tumor biology will be an essential stratifier to guide patient selection for organ preserving treatment. The final point to consider in terms of the value of radiation response characterization is the potential to modulate radiation response once the key molecular drivers are defined.

In conclusion, what has been demonstrated by this work is that the goal of a single biomarker remains both out of reach and unrealistic based on current evidence. However, as for the implications of radiation response biomarkers going forward, one can posit a number of clinical end points. Foremost, as debates on watch and wait and selection of optimal up-front therapy continues (chemoradiation vs. chemotherapy), the ability to accurately predict radiation response (and more importantly to be able to adequately survey the patient in a longitudinal fashion) would solidify the oncobiological credibility of these approaches. For practical purposes, this would obviously require a clinically translatable, bedside, or office-based test (rather than tissue biopsy), and work would need to be guided to establish how tissue-based biomarkers may be reflected in a more readily accessible liquid biopsy source, be that from stool, blood, or urine.

Acknowledgments

R.M. acknowledges funding and support from the National Institute for Health Research Imperial Biomedical Research Centre. L.P. acknowledges funding support from the Bowel Disease Research Foundation. D.G. and K.V. acknowledge the financial support of Wellcome Trust, Biotechnology and Biological Sciences Research Council, and Imperial College Stratified Medicine Training Programme in Systems Medicine and Spectroscopic Profiling (STRATiGRAD).

Disclosure

The authors have stated that they have no conflict of interest.

Supplemental Data

Supplemental material accompanying this article can be found in the online version at https://doi.org/10.1016/j.clcc.2019.01.004.

References

1. Heald RJ, Husband EM, Ryall RD. The mesorectum in rectal cancer surgery—the clue to pelvic recurrence? *Br J Surg* 1982; 69:613-6.
2. Kreis ME, Ruppert R, Ptok H, et al. Use of preoperative magnetic resonance imaging to select patients with rectal cancer for neoadjuvant chemoradiation—interim analysis of the German OCUM trial (NCT01325649). *J Gastrointest Surg* 2016; 20:25-32.
3. Sebag-Montefiore D, Stephens RJ, Steele R, et al. Preoperative radiotherapy versus selective postoperative chemoradiotherapy in patients with rectal cancer (MRC CR07 and NCIC-CTG C016): a multicentre, randomised trial. *Lancet* 2009; 373:811-20.
4. Erlandsson J, Holm T, Pettersson D, et al. Optimal fractionation of preoperative radiotherapy and timing to surgery for rectal cancer (Stockholm III): a multicentre, randomised, non-blinded, phase 3, non-inferiority trial. *Lancet Oncol* 2017; 18:336-46.
5. van Gijn W, Marijnen CA, Nagtegaal ID, et al. Preoperative radiotherapy combined with total mesorectal excision for resectable rectal cancer: 12-year follow-up of the multicentre, randomised controlled TME trial. *Lancet Oncol* 2011; 12:575-82.
6. Creavin B, Ryan E, Martin ST, et al. Organ preservation with local excision or active surveillance following chemoradiotherapy for rectal cancer. *Br J Cancer* 2016; 20:417.
7. Habr-Gama A, Sabbaga J, Gama-Rodrigues J, et al. Watch and wait approach following extended neoadjuvant chemoradiation for distal rectal cancer: are we getting closer to anal cancer management? *Dis Colon Rectum* 2013; 56:1109-17.
8. AmiGO 2. The Gene Ontology database, Available at: http://amigo.geneontology.org/amigo. Accessed April 10, 2016.
9. Hu P, Bader G, Wigle DA, et al. Computational prediction of cancer-gene function. *Nat Rev Cancer* 2007; 7:23-34.
10. Huh JW, Lee JH, Kim HR. Pretreatment expression of 13 molecular markers as a predictor of tumor responses after neoadjuvant chemoradiation in rectal cancer. *Ann Surg* 2014; 259:508-15.
11. Kobunai T, Watanabe T, Fukusato T. *REG4, NEIL2*, and *BIRC5* gene expression correlates with gamma-radiation sensitivity in patients with rectal cancer receiving radiotherapy. *Anticancer Res* 2011; 31:4147-53.
12. McDowell DT, Smith FM, Reynolds JV, et al. Increased spontaneous apoptosis, but not survivin expression, is associated with histomorphologic response to neoadjuvant chemoradiation in rectal cancer. *Int J Colorectal Dis* 2009; 24:1261-9.
13. Rodel F, Hoffmann J, Distel L, et al. Survivin as a radioresistance factor, and prognostic and therapeutic target for radiotherapy in rectal cancer. *Cancer Res* 2005; 65:4881-7.
14. Sim SH, Kang MH, Kim YJ, et al. P21 and CD166 as predictive markers of poor response and outcome after fluorouracil-based chemoradiotherapy for the patients with rectal cancer. *BMC Cancer* 2014; 14:241.
15. Bertolini F, Bengala C, Losi L, et al. Prognostic and predictive value of baseline and posttreatment molecular marker expression in locally advanced rectal cancer treated with neoadjuvant chemoradiotherapy. *Int J Radiat Oncol Biol Phys* 2007; 68:1455-61.
16. Edden Y, Wexner SD, Berho M. The use of molecular markers as a method to predict the response to neoadjuvant therapy for advanced stage rectal adenocarcinoma. *Colorectal Dis* 2012; 14:555-61.
17. Kobayashi H, Hashiguchi Y, Ueno H, et al. Absence of cyclooxygenase-2 protein expression is a predictor of tumor regression in rectal cancer treated with preoperative short-term chemoradiotherapy. *Dis Colon Rectum* 2007; 50:1354-62.
18. Brophy S, Sheehan KM, McNamara DA, et al. GLUT-1 expression and response to chemoradiotherapy in rectal cancer. *Int J Cancer* 2009; 125:2778-82.
19. Lopez-Crapez E, Bibeau F, Thezenas S, et al. p53 status and response to radiotherapy in rectal cancer: a prospective multilevel analysis. *Br J Cancer* 2005; 92:2114-21.
20. Gunther K, Dimmler A, Rodel F, et al. P27 does not predict histopathological response to radiochemotherapy in rectal cancer. *J Surg Res* 2003; 113:179-88.
21. Farkas R, Pozsgai E, Schally AV, et al. Possible predictors of histopathological response to neoadjuvant chemoradiotherapy for rectal cancer. *J Cancer Res Clin Oncol* 2012; 138:387-95.
22. Carlomagno C, Pepe S, D'Armiento FP, et al. Predictive factors of complete response to neoadjuvant chemoradiotherapy in patients with rectal cancer. *Oncology* 2010; 78:369-75.
23. Jakob C, Liersch T, Meyer W, et al. Predictive value of Ki67 and p53 in locally advanced rectal cancer: correlation with thymidylate synthase and histopathological tumor regression after neoadjuvant 5-FU–based chemoradiotherapy. *World J Gastroenterol* 2008; 14:1060-6.
24. Yan H, Wang R, Yu J, et al. Predictive value of Smac, VEGF and Ki-67 in rectal cancer treated with neoadjuvant therapy. *Oncol Lett* 2010; 1:641-7.
25. He HL, Lee YE, Shiue YL, et al. Overexpression of REG4 confers an independent negative prognosticator in rectal cancers receiving concurrent chemoradiotherapy. *J Surg Oncol* 2014; 110:1002-10.
26. Avoranta ST, Korkeila EA, Minn HRI, et al. Securin identifies a subgroup of patients with poor outcome in rectal cancer treated with long-course (chemo) radiotherapy. *Acta Oncol* 2011; 50:1158-66.
27. Senetta R, Duregon E, Sonetto C, et al. YKL-40/c-met expression in rectal cancer biopsies predicts tumor regression following neoadjuvant chemoradiotherapy: a multi-institutional study. *PLoS One* 2015; 10:e0123759.
28. Allal AS, Kahne T, Reverdin AK, et al. Radioresistance-related proteins in rectal cancer. *Proteomics* 2004; 4:2261-9.
29. Rodel C, Grabenbauer GG, Papadopoulos T, et al. Apoptosis as a cellular predictor for histopathologic response to neoadjuvant radiochemotherapy in patients with rectal cancer. *Int J Radiat Oncol Biol Phys* 2002; 52:294-303.
30. Chang HJ, Jung KH, Kim DY, et al. Bax, a predictive marker for therapeutic response to preoperative chemoradiotherapy in patients with rectal carcinoma. *Hum Pathol* 2005; 36:364-71.
31. Kelley ST, Coppola D, Yeatman T, et al. Tumor response to neoadjuvant chemoradiation therapy for rectal adenocarcinoma is mediated by p53-dependent and caspase 8-dependent apoptotic pathways. *Clin Colorectal Cancer* 2005; 5:114-8.
32. Qiu J, Yang G, Lin A, et al. Human phosphatidylethanolamine-binding protein 4 promoted the radioresistance of human rectal cancer by activating Akt in an ROS-dependent way. *PLoS One* 2014; 9:e90062.
33. Cao WT, Zhou ZY, Deng YH, et al. Clinical value of MR diffusion weighted imaging in prediction of pathological complete response of rectal cancer after neoadjuvant therapy. *Zhonghua Wei Chang Wai Ke Za Zhi* 2013; 16:1164-8.
34. Saigusa S, Inoue Y, Tanaka K, et al. Lack of M30 expression correlates with factors reflecting tumor progression in rectal cancer with preoperative chemoradiotherapy. *Mol Clin Oncol* 2014; 2:99-104.
35. Dou X, Wang RB, Meng XJ, et al. PDCD4 as a predictor of sensitivity to neoadjuvant chemoradiotherapy in locally advanced rectal cancer patients. *Asian Pac J Cancer Prev* 2014; 15:825-30.
36. Moussata D, Amara S, Siddeek B, et al. XIAP as a radioresistance factor and prognostic marker for radiotherapy in human rectal adenocarcinoma. *Am J Pathol* 2012; 181:1271-8.
37. Min BS, Choi YJ, Pyo HR, et al. Cyclooxygenase-2 expression in pretreatment biopsy as a predictor of tumor responses after preoperative chemoradiation in rectal cancer. *Arch Surg* 2008; 143:1091-7.
38. Yan H, Yu J, Wang R, et al. Prognostic value of Smac expression in rectal cancer patients treated with neoadjuvant therapy. *Med Oncol* 2012; 29:168-73.
39. Toiyama Y, Inoue Y, Saigusa S, et al. Gene expression profiles of epidermal growth factor receptor, vascular endothelial growth factor and hypoxia-inducible factor-1 with special reference to local responsiveness to neoadjuvant chemoradiotherapy and disease recurrence after rectal cancer surgery. *Clin Oncol* 2010; 22:272-80.
40. Guedj N, Bretagnol F, Rautou PE, et al. Predictors of tumor response after preoperative chemoradiotherapy for rectal adenocarcinomas. *Hum Pathol* 2011; 42:1702-9.
41. Li C, Wang Z, Liu F, et al. CXCL10 mRNA expression predicts response to neoadjuvant chemoradiotherapy in rectal cancer patients. *Tumour Biol* 2014; 35:9683-91.
42. Hur H, Kim NK, Min BS, et al. Can a biomarker-based scoring system predict pathologic complete response after preoperative chemoradiotherapy for rectal cancer? *Dis Colon Rectum* 2014; 57:592-601.
43. Kurt A, Yanar F, Asoglu O, et al. Low Mmp 9 and VEGF levels predict good oncologic outcome in mid and low rectal cancer patients with neoadjuvant chemoradiation. *BMC Clin Pathol* 2012; 12:27.
44. Zlobec I, Steele R, Compton CC. VEGF as a predictive marker of rectal tumor response to preoperative radiotherapy. *Cancer* 2005; 104:2517-21.
45. Garcia-Florez LJ, Gomez-Alvarez G, Frunza AM, et al. Predictive markers of response to neoadjuvant therapy in rectal cancer. *J Surg Res* 2015; 194:120-6.
46. Sheu MJ, Li CF, Lin CY, et al. Overexpression of ANXA1 confers independent negative prognostic impact in rectal cancers receiving concurrent chemoradiotherapy. *Tumour Biol* 2014; 35:7755-63.
47. He HL, Lee YE, Chen HP, et al. Overexpression of DNAJC12 predicts poor response to neoadjuvant concurrent chemoradiotherapy in patients with rectal cancer. *Exp Mol Pathol* 2015; 98:338-45.
48. Cecchin E, Agostini M, Pucciarelli S, et al. Tumor response is predicted by patient genetic profile in rectal cancer patients treated with neo-adjuvant chemoradiotherapy. *Pharmacogenomics J* 2011; 11:214-26.
49. Terrazzino S, Agostini M, Pucciarelli S, et al. A haplotype of the methylenetetrahydrofolate reductase gene predicts poor tumor response in rectal cancer patients receiving preoperative chemoradiation. *Pharmacogenet Genomics* 2006; 16:817-24.
50. Molinari C, Casadio V, Foca F, et al. Gene methylation in rectal cancer: predictive marker of response to chemoradiotherapy? *J Cell Physiol* 2013; 228:2343-9.
51. Berardi R, Maccaroni E, Mandolesi A, et al. Nuclear factor-kappaB predicts outcome in locally advanced rectal cancer patients receiving neoadjuvant radiochemotherapy. *Dig Liver Dis* 2012; 44:617-22.
52. Qin CJ, Song XM, Chen ZH, et al. XRCC2 as a predictive biomarker for radioresistance in locally advanced rectal cancer patients undergoing preoperative radiotherapy. *Oncotarget* 2015; 6:32193-204.
53. Yu Z, Zhang C, Chai R, et al. Prognostic significance and molecular mechanism of ATP-binding cassette subfamily C member 4 in resistance to neoadjuvant radiotherapy of locally advanced rectal carcinoma. *PLoS One* 2014; 9:e85446.
54. Drebber U, Madeja M, Odenthal M, et al. beta-Catenin and Her2/neu expression in rectal cancer: association with histomorphological response to neoadjuvant therapy and prognosis. *Int J Colorectal Dis* 2011; 26:1127-34.
55. Wang L, Zhang XM, Li Z, et al. Overexpression of nuclear beta-catenin in rectal adenocarcinoma is associated with radioresistance. *World J Gastroenterol* 2013; 19:6876-82.
56. Shinto E, Hashiguchi Y, Ueno H, et al. Pretreatment CD133 and cyclooxygenase-2 expression as the predictive markers of the pathological effect of

chemoradiotherapy in rectal cancer patients. *Dis Colon Rectum* 2011; 54:1098-106.
57. Meng X, Wang R, Huang Z, et al. Human epidermal growth factor receptor-2 expression in locally advanced rectal cancer: association with response to neoadjuvant therapy and prognosis. *Cancer Sci* 2014; 105:818-24.
58. Holck S, Nielsen HJ, Hammer E, et al. IQGAP1 in rectal adenocarcinomas: localization and protein expression before and after radiochemotherapy. *Cancer Lett* 2015; 356(2 pt B):556-60.
59. Yu Z, Zhang C, Wang H, et al. Multidrug resistance-associated protein 3 confers resistance to chemoradiotherapy for rectal cancer by regulating reactive oxygen species and caspase-3-dependent apoptotic pathway. *Cancer Lett* 2014; 353:182-93.
60. Huang MY, Lin CH, Huang CM, et al. Relationships between SMAD3 expression and preoperative fluoropyrimidine-based chemoradiotherapy response in locally advanced rectal cancer patients. *World J Surg* 2015; 39:1257-67.
61. Lin CY, Chang IW, Sheu MJ, et al. Low thrombospondin 2 expression is predictive of low tumor regression after neoadjuvant chemoradiotherapy in rectal cancer. *Am J Transl Res* 2015; 7:2423-32.
62. Smith FM, Reynolds JV, Kay EW, et al. COX-2 overexpression in pretreatment biopsies predicts response of rectal cancers to neoadjuvant radiochemotherapy. *Int J Radiat Oncol Biol Phys* 2006; 64:466-72.
63. Avoranta ST, Korkeila EA, Ristamaki RH, et al. ALDH1 expression indicates chemotherapy resistance and poor outcome in node-negative rectal cancer. *Hum Pathol* 2013; 44:966-74.
64. Yeo SG, Kim DY, Kim KH, et al. Hydroxymethylglutaryl-coenzyme a synthase 2 expression is associated with chemoradiotherapy responses in colorectal cancer. *Dis Colon Rectum* 2012; 55:686-94.
65. Lee YE, He HL, Shiue YL, et al. The prognostic impact of lipid biosynthesis-associated markers, HSD17B2 and HMGCS2, in rectal cancer treated with neoadjuvant concurrent chemoradiotherapy. *Tumour Biol* 2015; 36:7675-83.
66. Lee YY, Li CF, Lin CY, et al. Overexpression of CPS1 is an independent negative prognosticator in rectal cancers receiving concurrent chemoradiotherapy. *Tumour Biol* 2014; 35:11097-105.
67. Lin CY, Sheu MJ, Li CF, et al. Deficiency in asparagine synthetase expression in rectal cancers receiving concurrent chemoradiotherapy: negative prognostic impact and therapeutic relevance. *Tumour Biol* 2014; 35:6823-30.
68. Dou X, Wang R, Meng X, et al. The prognostic role of TCF4 expression in locally advanced rectal cancer patients treated with neoadjuvant chemo-radiotherapy. *Cancer Biomark* 2015; 15:187-94.
69. Jakob C, Liersch T, Meyer W, et al. Immunohistochemical analysis of thymidylate synthase, thymidine phosphorylase, and dihydropyrimidine dehydrogenase in rectal cancer (cUICC II/III): correlation with histopathologic tumor regression after 5-fluorouracil–based long-term neoadjuvant chemorodiotherapy. *Am J Surg Pathol* 2005; 29:1304-9.
70. Saw RPM, Morgan M, Koorey D, et al. p53, deleted in colorectal cancer gene, and thymidylate synthase as predictors of histopathologic response and survival in low, locally advanced rectal cancer treated with preoperative adjuvant therapy. *Dis Colon Rectum* 2003; 46:192-202.
71. Conradi LC, Bleckmann A, Schirmer M, et al. Thymidylate synthase as a prognostic biomarker for locally advanced rectal cancer after multimodal treatment. *Ann Surg Oncol* 2011; 18:2442-52.
72. Jakob C, Aust DE, Meyer W, et al. Thymidylate synthase, thymidine phosphorylase, dihydropyrimidine dehydrogenase expression, and histological tumour regression after 5-FU-based neo-adjuvant chemoradiotherapy in rectal cancer. *J Pathol* 2004; 204:562-8.
73. Perez RE, Shen H, Duan L, et al. Modeling the etiology of p53-mutated cancer cells. *J Biol Chem* 2016; 291:10131-47.
74. Carnero A, Paramio JM. The PTEN/PI3K/AKT pathway in vivo, cancer mouse models. *Front Oncol* 2014; 4:252.
75. Bishnupuri KS, Luo Q, Sainathan SK, et al. Reg IV regulates normal intestinal and colorectal cancer cell susceptibility to radiation-induced apoptosis. *Gastroenterology* 2010; 138:616-26.
76. Matsumoto S, Konishi H, Maeda R, et al. Expression analysis of the regenerating gene (Reg) family members Reg-IIIbeta and Reg-IIIgamma in the mouse during development. *J Comp Neurol* 2012; 520:479-94.
77. Rafa L, Dessein AF, Devisme L, et al. REG4 acts as a mitogenic, motility and pro-invasive factor for colon cancer cells. *Int J Oncol* 2010; 36:689-98.
78. Kim PJ, Plescia J, Clevers H, et al. Survivin and molecular pathogenesis of colorectal cancer. *Lancet* 2003; 362:205-9.
79. Shao R. YKL-40 acts as an angiogenic factor to promote tumor angiogenesis. *Front Physiol* 2013; 4:122.
80. Kawada M, Seno H, Kanda K, et al. Chitinase 3-like 1 promotes macrophage recruitment and angiogenesis in colorectal cancer. *Oncogene* 2012; 31:3111-23.
81. Shibahara K, Asano M, Ishida Y, et al. Isolation of a novel mouse gene *MA-3* that is induced upon programmed cell death. *Gene* 1995; 166:297-301.
82. He H, Liu D, Lin H, et al. Phosphatidylethanolamine binding protein 4 (PEBP4) is a secreted protein and has multiple functions. *Biochim Biophys Acta* 2016; 1863(7 pt A):1682-9.
83. Moeller BJ, Dewhirst MW. HIF-1 and tumour radiosensitivity. *Br J Cancer* 2006; 95:1-5.
84. Moeller BJ, Richardson RA, Dewhirst MW. Hypoxia and radiotherapy: opportunities for improved outcomes in cancer treatment. *Cancer Metastasis Rev* 2007; 26:241-8.
85. Karin N, Wildbaum G, Thelen M. Biased signaling pathways via CXCR3 control the development and function of $CD4^+$ T cell subsets. *J Leukoc Biol* 2016; 99:857-62.
86. Jameson KL, Mazur PK, Zehnder AM, et al. IQGAP1 scaffold-kinase interaction blockade selectively targets RAS-MAP kinase-driven tumors. *Nat Med* 2013; 19:626-30.
87. Munshi A, Ramesh R. Mitogen-activated protein kinases and their role in radiation response. *Genes Cancer* 2013; 4:401-8.
88. Taketo MM, Takaku K. Gastrointestinal tumorigenesis in Smad4 (Dpc4) mutant mice. *Hum Cell* 2000; 13:85-95.
89. Schally AV, Varga JL, Engel JB. Antagonists of growth-hormone-releasing hormone: an emerging new therapy for cancer. *Nat Clin Pract Endocrinol Metab* 2008; 4:33-43.
90. Rick FG, Seitz S, Schally AV, et al. GHRH antagonist when combined with cytotoxic agents induces S-phase arrest and additive growth inhibition of human colon cancer. *Cell Cycle* 2012; 11:4203-10.
91. Balasubramanian MN, Butterworth EA, Kilberg MS. Asparagine synthetase: regulation by cell stress and involvement in tumor biology. *Am J Physiol Endocrinol Metab* 2013; 304:E789-99.
92. Tuncer S, Banerjee S. Eicosanoid pathway in colorectal cancer: recent updates. *World J Gastroenterol* 2015; 21:11748-66.
93. Labrie F, Luu-The V, Lin S, et al. Role of 17beta-hydroxysteroid dehydrogenases in sex steroid formation in peripheral intracrine tissues. *Trends Endocrinol Metab* 2000; 11:421-7.
94. Liberti MV, Locasale JW. The Warburg effect: how does it benefit cancer cells? *Trends Biochem Sci* 2016; 41:211-8.
95. Yarla NS, Bishayee A, Sethi G, et al. Targeting arachidonic acid pathway by natural products for cancer prevention and therapy. *Semin Cancer Biol* 2016; 40-41:48-81.
96. Bouzourene H, Yan P, Sandmeier D, et al. The role of COX-2 in rectal cancer treated with preoperative radiotherapy. *Virchows Arch* 2008; 452:499-505.
97. Di Maggio FM, Minafra L, Forte GI, et al. Portrait of inflammatory response to ionizing radiation treatment. *J Inflamm (Lond)* 2015; 12:14.
98. Dworak O, Keilholz L, Hoffmann A. Pathological features of rectal cancer after preoperative radiochemotherapy. *Int J Colorectal Dis* 1997; 12:19-23.
99. Mandard AM, Dalibard F, Mandard JC, et al. Pathologic assessment of tumor regression after preoperative chemoradiotherapy of esophageal carcinoma. Clinicopathologic correlations. *Cancer* 1994; 73:2680-6.
100. Rodel C, Grabenbauer GG, Papadopoulos T, et al. Phase I/II trial of capecitabine, oxaliplatin, and radiation for rectal cancer. *J Clin Oncol* 2003; 21:3098-104.
101. Ryan R, Gibbons D, Hyland JM, et al. Pathological response following long-course neoadjuvant chemoradiotherapy for locally advanced rectal cancer. *Histopathology* 2005; 47:141-6.
102. Wheeler JM, Warren BF, Mortensen NJ, et al. Quantification of histologic regression of rectal cancer after irradiation: a proposal for a modified staging system. *Dis Colon Rectum* 2002; 45:1051-6.
103. Fokas E, Liersch T, Fietkau R, et al. Downstage migration after neoadjuvant chemoradiotherapy for rectal cancer: the reverse of the Will Rogers phenomenon? *Cancer* 2015; 121:1724-7.
104. Appelt AL, Vogelius IR, Ploen J, et al. Long-term results of a randomized trial in locally advanced rectal cancer: no benefit from adding a brachytherapy boost. *Int J Radiat Oncol Biol Phys* 2014; 90:110-8.

Copyright Clearance Center | RightsLink®

Home | Help | Email Support | Sign in | Create Account



Network Mapping of Molecular Biomarkers Influencing Radiation Response in Rectal Cancer

Author: Liam Poynter,Dieter Galea,Kirill Veselkov,Alexander Mirnezami,James Kinross,Jeremy Nicholson,Zoltán Takáts,Ara Darzi,Reza Mirnezami

Publication: Clinical Colorectal Cancer

Publisher: Elsevier

Date: June 2019

© 2019 The Authors. Published by Elsevier Inc.

Journal Author Rights

Please note that, as the author of this Elsevier article, you retain the right to include it in a thesis or dissertation, provided it is not published commercially. Permission is not required, but please ensure that you reference the journal as the original source. For more information on this and on your other retained rights, please visit: https://www.elsevier.com/about/our-business/policies/copyright#Author-rights

BACK | CLOSE WINDOW

© 2021 Copyright - All Rights Reserved | Copyright Clearance Center, Inc. | Privacy statement | Terms and Conditions

Comments? We would like to hear from you. E-mail us at customercare@copyright.com