

ASSESSING CIRCULATING CELL-FREE TUMOUR
DNA AS A POTENTIAL BIOMARKER FOR EARLY
DETECTION & CHEMOPREVENTION OF
COLORECTAL CANCER

Thesis submitted for the degree of
Doctor of Philosophy
at the University of Leicester

by

Ni Ni Moe Myint, BSc
Leicester Cancer Research Centre
University of Leicester

October 2017

Abstract

Assessing Circulating Cell-free Tumour DNA as a Potential Biomarker for Early Detection & Chemoprevention of Colorectal Cancer

Ni Ni Moe Myint

Colorectal cancer (CRC) is the second-leading cause of cancer-related deaths in the UK, and patient mortality is strongly correlated to the cancer stage at diagnosis. Circulating cell-free DNA (cfDNA) are short DNA fragments released into the blood from dying cells, and the use of cfDNA as a predictive and diagnostic biomarker for CRC patients has gained traction recently. The work undertaken in the thesis aimed to assess whether cfDNA could be a suitable biomarker for detection and monitoring of early colorectal lesions in preclinical and clinical settings.

Custom qPCR and ddPCR assays were designed and implemented for quantitative cfDNA analysis and for detection of tumour-derived mutations in plasma samples. Preclinical studies investigating the cfDNA dynamics during early tumourigenesis were conducted with the *Lgr5-EGFP-IRES-creER^{T2+/0};Apc^{fl/fl}* mouse model of CRC. In parallel, plasma cfDNA was also assessed in an adenoma patient cohort (n=76) against a 'polyp-free' control group (n=37).

Analysis of mice failed to show a correlation between total cfDNA levels and the adenoma development in three separate preclinical studies, whereas detection of a surrogate tumour-derived mutation in animal plasma samples showed sensitivities between 16 and 25%. Similarly, total plasma cfDNA concentrations between the adenoma patient and control group showed no significant difference (p=0.1012). Targeted detection of *BRAF* and *KRAS* mutations in the plasma was achievable, but hampered by low sensitivities (0-25%). Multi-regional targeted next-generation sequencing (NGS) was performed on selected patient cases. The data unmasked extensively and previously unappreciated intra-tumour heterogeneity in colorectal adenomas. Nonetheless, identification and plasma targeting of tumour truncal mutations did not improve the detection rate.

In conclusion, these results suggested that further methodological optimisation is necessary to achieve improved diagnostic sensitivity with plasma cfDNA liquid biopsy for the early detection of colorectal lesions.

Acknowledgement

This thesis is the culmination of both personal and professional challenges I have overcome with the help of many wonderful individuals whom I encountered throughout my journey. Firstly, I would like to express my deepest gratitude to my supervisors, Dr Alessandro Rufini and Professor Karen Brown, for all their guidance and help that have made this project a reality. Moreover, my great appreciation is extended to Dr Mafalda Pires Damaso, Dr Hong Cai and Dr Saif Al-Aqbi for their contributions to the successful completion of the preclinical studies. I am also thankful to Dr Peter Greaves, Dr Kevin West and Dr Ultan McDermott for their help with the pathological assessment of samples and sequencing analysis. Additionally, I would like to offer my appreciation to Dr Howard Pringle, Professor Jacqui Shaw, Dr Barbara Ottolini, Dr Ricky Trigg and Ms. Lindsay Primrose for their guidance on various aspects of my project. I would also like to mention the wonderful support I received from Ms. Nalini Foreman, Ms. Maria Szpek and Ms. Melanie Haberle for smooth running of my experiments. I am thankful to Dr Catherine Andreadi, Dr Lynne Howells, Professor Andreas Gescher and Ms. Karen Kulbicki for their help and encouragement throughout my project. I am also very grateful to all my wonderful colleagues – Ankur, Beth, Con, Gintare and Dhafer – for their constant help and friendship that have made this a memorable experience. Last but not least, I feel blessed to have the unwavering support and confidence from my family and friends to guide me through this PhD journey until the end. Finally, I would like to express my sincere gratitude to the funding bodies, MRC Doctoral Training Grant and Hope Against Cancer Foundation, for giving me this invaluable opportunity.

List of contents

Abstract	i
Acknowledgement	ii
List of contents	iii
Table of contents	iii
List of Figures	ix
List of Tables	xii
List of Abbreviations	xiv
List of Publications & Conference Abstracts	xix

Table of contents

Chapter 1 Introduction	1
1.1 Gastrointestinal Tract	1
1.2 Colorectal Cancer	4
1.2.1 Histology of Colorectal Lesions	6
1.2.2 Cancer Staging	7
1.2.3 Symptoms & Diagnosis	8
1.2.4 Treatment Options	9
1.3 Molecular Pathology of CRCs	10
1.3.1 WNT/ β -catenin Pathway	10
1.3.1.1 Pathway Overview	10
1.3.1.2 <i>APC</i> Tumour Suppressor Gene	13
1.3.1.3 <i>APC</i> Mutations & Implications in CRC	14
1.3.2 Mitogen-activated Protein Kinase Signalling Pathway	16
1.3.2.1 Pathway Overview	16
1.3.2.2 <i>RAS</i> Oncogenes	19
1.3.2.3 <i>RAF</i> Oncogenes	21
1.3.3 Pathways of Genomic Instability	24
1.3.3.1 CpG Island Methylator Phenotype	24
1.3.3.2 Microsatellite Instability	24
1.3.3.3 Chromosomal Instability	25
	iii

1.3.3.4	CRC Classification based on Genomic Alterations	26
1.3.4	Genetics, Genomics & CRC Pathways	27
1.4	Tumour Heterogeneity in CRC	29
1.5	Genetically Engineered Mouse Models of CRC	31
1.6	Methods to Reduce CRC Incidence & Mortality	35
1.6.1	Screening & Early Detection	35
1.6.1.1	The Bowel Cancer Screening Programme	35
1.6.2	Prevention of CRC	37
1.6.2.1	Lifestyle & Diet	37
1.6.2.2	Chemoprevention	38
1.7	Biomarkers for CRC	42
1.7.1	Overview of Biomarkers	42
1.7.2	RAS & RAF Mutation Statuses	43
1.7.3	Carcinoembryonic Antigen	43
1.7.4	Carbohydrate Antigen	44
1.7.5	Faecal Calprotectin	44
1.7.6	Circulating Cell-free DNA	45
1.7.6.1	Properties of cfDNA	45
1.7.6.2	Mechanisms of Release	46
1.7.6.3	Pre-analytical & Analytical Considerations	47
1.7.6.4	Techniques for cfDNA Detection	49
1.7.6.5	Prognostic & Predictive Values of cfDNA	50
1.7.6.6	cfDNA as a Biomarker for Early Detection	51
1.7.7	Summary	52
1.8	Technologies used for Targeted Mutation Detection	53
1.8.1	Real-time Quantitative Polymerase Chain Reaction	53
1.8.1.1	Allele-specific Mutation Detection Assays	55
1.8.1.2	Peptide Nucleic Acids	57
1.8.2	Droplet Digital Polymerase Chain Reaction	59
1.9	Hypothesis and Aims	61
	Chapter 2 Materials & Methods	63
2.1	Materials	63

2.1.1	Cell Culture	63
2.1.2	DNA Extraction	64
2.1.3	Traditional End-point PCR	66
2.1.4	qPCR & ddPCR	67
2.1.5	<i>In Vivo</i> Studies	69
2.1.6	Immunohistochemistry	70
2.1.7	Patients	71
2.2	Methods	72
2.2.1	Cell Culture	72
2.2.2	Protocols for DNA Extraction & Processing	72
2.2.2.1	DNA Extraction from Murine Samples	73
2.2.2.2	DNA Extraction from Human Samples	74
2.2.2.3	Processing of DNA Samples	76
2.2.3	Polymerase Chain Reaction (PCR)	76
2.2.3.1	Designing Primers & Probes	76
2.2.3.2	Traditional End-point PCR	77
2.2.3.3	qPCR	79
2.2.3.4	ddPCR	82
2.2.4	Multi-regional Targeted NGS of Patient Adenoma Samples	84
2.2.5	<i>In Vivo</i> Studies	85
2.2.5.1	Preparation of Tamoxifen & Vehicle Solutions	85
2.2.5.2	Diet Allocation	86
2.2.5.3	Health Checks	86
2.2.5.4	Study End-points	86
2.2.5.5	Blood Collection & Plasma Extraction	87
2.2.5.6	Tissue Collection	87
2.2.5.7	Collection of Faecal Samples	87
2.2.6	Histology	88
2.2.6.1	Haematoxylin & Eosin Staining	88
2.2.6.2	Immunohistochemical Analysis of β -catenin Expression in <i>Cre-Apc^{fl/fl}</i>	
	Mice	88
2.2.6.3	Image Capturing	89

2.2.7	Statistical Analysis	89
Chapter 3	Validation of Assays for Preclinical Studies	91
3.1	Introduction	91
3.2	Sequencing Upstream <i>loxP</i> in Intron 13 of <i>Apc^{fl/fl}</i> Allele	91
3.3	Sequencing Downstream <i>loxP</i> in Intron 14 of <i>Apc^{fl/fl}</i> Allele	93
3.4	Designing & Optimisation of Custom Assays for Detection of Recombined & Non-recombined Alleles	96
3.5	Summary	102
Chapter 4	Investigating the Potential of Plasma cfDNA as a Surrogate Biomarker for Preclinical Studies	103
4.1	Introduction	103
4.2	1 st <i>In Vivo</i> Study	104
4.2.1	Study Groups	104
4.2.2	Histopathological Analysis	105
4.2.3	Quantification of Total Plasma cfDNA Levels	108
4.2.4	Checking for Recombined Alleles in the Plasma	110
4.3	2 nd <i>In Vivo</i> Study	112
4.3.1	Study Groups	112
4.3.2	Histopathological Analysis	113
4.3.3	Degree of <i>Apc^{fl/fl}</i> Recombination in the Proximal SI	116
4.3.4	Quantification of Total Plasma cfDNA Levels	117
4.3.5	Analysis of Non-recombined & Recombined Alleles in the Plasma	119
4.4	3 rd <i>In Vivo</i> Study	121
4.4.1	Study Groups	121
4.4.2	Survival Analysis	123
4.4.3	Histopathological Analysis	124
4.4.4	Degree of <i>Apc^{fl/fl}</i> Recombination in the GI Tract	131
4.4.5	Analysis of Nuclear β -catenin Index	133
4.4.6	Quantification of Total Plasma cfDNA Levels	137
4.4.7	Analysis of Non-recombined & Recombined Alleles in the Plasma	139
4.4.8	Analysis of Recombined Alleles in the Faecal Samples	142
4.5	Summary	145

Chapter 5 Validation of Assays for the Analysis of Patient Samples	149
5.1 Introduction	149
5.2 Validation of <i>BRAF</i> & <i>KRAS</i> Mutation Detection Assays with PNA Clamping on the qPCR System	149
5.2.1 <i>BRAF</i> -V600E Assay	150
5.2.2 <i>KRAS</i> -121 Assay	151
5.2.3 <i>KRAS</i> -122 Assay	155
5.2.4 Efficiencies of Assays	156
5.3 Validation of <i>BRAF</i> & <i>KRAS</i> Mutation Detection Assays on the ddPCR System	158
5.3.1 <i>BRAF</i> -V600E Assay	158
5.3.2 <i>KRAS</i> -121, <i>KRAS</i> -122 & <i>KRAS</i> -G13D Assays	160
5.4 Validation of Patient-specific Mutation Detection Assays	164
5.5 Summary	171
Chapter 6 Analysis of Tissue & Plasma Samples from Patients with Colorectal Adenomas	173
6.1 Introduction	173
6.2 Histopathology of Patient Samples	174
6.3 Screening for <i>BRAF</i> & <i>KRAS</i> Mutations in FFPE Samples	176
6.4 Quantification of Total Plasma cfDNA Levels	179
6.5 Analysis of <i>BRAF</i> & <i>KRAS</i> Mutations in Plasma cfDNA Samples	181
6.5.1 Analysis of <i>BRAF</i> Mutation in the Plasma cfDNA Samples	181
6.5.2 Analysis of <i>KRAS</i> Mutations at base c.34G in Plasma cfDNA Samples	183
6.5.3 Analysis of <i>KRAS</i> Mutations at base c.35G in the Plasma cfDNA Samples	186
6.5.4 Analysis of <i>KRAS</i> (c.38G>A) Mutation in the Plasma cfDNA Samples	191
6.6 Multi-regional Targeted Sequencing of Patient Samples	196
6.7 Plasma Analysis with Patient-Specific Mutation Detection Assays	204
6.8 Summary	211
Chapter 7 Concluding Discussion	214
Chapter 8 Appendices	223
8.1 Accompanying Data for Quantification of Total Plasma cfDNA Levels in the Patient & Control Groups in Section 6.4	223

8.2	Details of the Custom Targeted Panel used in the Multi-regional Targeted NGS Analysis in Section 6.6	227
8.3	Mutation Profiles for Patient Cases from Multi-regional Targets NGS Analysis in Section 6.6	231
Chapter 9	References	234

List of Figures

Figure 1-1: Structure of the Gastrointestinal Tract	1
Figure 1-2: Structural Differences between the SI & Colon Epithelium	3
Figure 1-3: Statistics & Anatomical Distribution of CRC in UK	5
Figure 1-4: Comparison of CRC Classification Systems	7
Figure 1-5: Canonical WNT/Frizzled Signalling Pathway in ON & OFF States	11
Figure 1-6: Structural Domains & Functions of APC	13
Figure 1-7: Distribution of Somatic APC Mutations in the Intestine	15
Figure 1-8: ERK & PI3K Signalling Pathways	17
Figure 1-9: Structure of Wild-type & Mutant KRAS	20
Figure 1-10: Structure & Function of BRAF	22
Figure 1-11: Distribution of Somatic BRAF Mutations in the Intestine	23
Figure 1-12: Molecular Classification for CRC Subtypes	26
Figure 1-13: Genetic & Genomic Alterations in Colorectal Tumourigenesis	28
Figure 1-14: Branching Evolution of Intra-tumour Heterogeneity in CRC	30
Figure 1-15: Gene Manipulation using the Cre-lox System	32
Figure 1-16: Site-specific Sequence Deletion in the <i>Apc^{fl/fl}</i> Allele	33
Figure 1-17: Mechanisms of cfDNA Release into Peripheral Circulation	47
Figure 1-18: Methods used for cfDNA Analysis	49
Figure 1-19: Parameters for Fluorescence Data Collection in the qPCR System	54
Figure 1-20: Allele-specific Mutation Detection with TaqMan qPCR System	56
Figure 1-21: Watson-Crick Pairing between PNA & DNA Strands	57
Figure 1-22: Mutation Detection qPCR Assay with PNA Clamping	58
Figure 1-23: Workflow for a ddPCR Assay	60
Figure 3-1: Sequence details for the Upstream <i>loxP</i> site in the <i>Apc^{fl/fl}</i> Allele	92
Figure 3-2: PCR amplification of Downstream <i>loxP</i> sequence	94
Figure 3-3: Downstream <i>loxP</i> Sequence in Intron 14 of <i>Apc^{fl/fl}</i> Allele	95
Figure 3-4: Locations of <i>loxP</i> sites in the <i>Apc^{fl/fl}</i> allele	96
Figure 3-5: Apc-Rec Assay Design for Detection of the Recombined Allele	97
Figure 3-6: Determining Efficiency of the Apc-Rec qPCR Assay	98
Figure 3-7: Assay Design for Detection of the Non-recombined Alleles	99
Figure 3-8: Testing the Apc-Rec & Apc-NR Assays on ddPCR	100
Figure 4-1: Comparison of Histopathological Features in Control & Apc-deleted Mice	107
Figure 4-2: Comparing Total cfDNA Quantities in the Plasma of Control & Apc-deleted <i>Cre-Apc^{fl/fl}</i> Mice	109
Figure 4-3: Analysis of Plasma Samples to Detect Recombined Alleles	110
Figure 4-4: Large Polyp Growth in the Caecum of Apc-deleted Mouse	114

Figure 4-5: Comparison of Histopathological Features in Control & Apc-deleted Mice from 2 nd <i>In Vivo</i> Study	115
Figure 4-6: Assessment of <i>Apc^{fl/fl}</i> Recombination Levels in the Proximal SI	116
Figure 4-7: Comparing Total cfDNA Quantities in the Plasma of Control & Apc-deleted <i>Cre-Apc^{fl/fl}</i> Mice from the 2 nd <i>in vivo</i> Study	118
Figure 4-8: Analysis of Non-recombined & Recombined Alleles in Plasma Samples from Control & Apc-deleted Mice in 2 nd <i>in vivo</i> study	120
Figure 4-9: Survival Analysis of <i>Cre-Apc^{fl/fl}</i> Mice in the 3 rd Study	123
Figure 4-10: Comparison of Histopathological Features in Control & Apc-deleted Mice in the NFD Group	126
Figure 4-11: Comparison of Histopathological Features in Control & Apc-deleted Mice in the HFD Group	127
Figure 4-12: Comparing the Numbers & Areas of Polyps in the Proximal SI & Caecum of Apc-deleted Mice in the NFD & HFD Groups	128
Figure 4-13: Kozuha's Criteria used for Histological Grading in Table 4-8	129
Figure 4-14: Assessment of <i>Apc^{fl/fl}</i> Recombination Levels in the GI Tract	131
Figure 4-15: β -catenin Staining of Proximal SI Tissues from the NFD Mice	133
Figure 4-16: β -catenin Staining of Proximal SI Tissues from the HFD Mice	134
Figure 4-17: Comparing Nuclear β -catenin Indices in Crypts of Apc-deleted Mice on NFD & HFD	135
Figure 4-18: Comparison of Nuclear β -catenin Indices in Adenomas from Apc-deleted Mice on NFD & HFD	136
Figure 4-19: Total Plasma cfDNA Levels in <i>Cre-Apc^{fl/fl}</i> Mice from NFD & HFD Groups	138
Figure 4-20: 1D Amplitude Plots for Apc-Rec & Apc-NR Assays used to Detect Non-recombined & Recombined Alleles	139
Figure 4-21: Detection of Non-recombined & Recombined Alleles in the Plasma	140
Figure 4-22: Changes in the Fractional Abundance of Recombined Alleles in Faecal Samples with Time after Injection	144
Figure 5-1: Testing the BRAF-V600E qPCR Assay with PNA Clamping	150
Figure 5-2: Ineffectiveness of PNA Clamping in KRAS-121 qPCR reactions containing both WT and MU Probes	152
Figure 5-3: PNA Clamping was Necessary for Detection at Low MU Fractions	153
Figure 5-4: An Alternative Strategy for the KRAS-121 Assay	154
Figure 5-5: Testing the KRAS-122 qPCR with PNA Clamping	155
Figure 5-6: Assay Efficiencies for <i>BRAR</i> & <i>KRAS</i> qPCR Assays with PNA Clamping	157
Figure 5-7: Validation of BRAF-V600E Assay on ddPCR	159
Figure 5-8: Validation of KRAS-121 Assay on ddPCR	160

Figure 5-9: Validation of KRAS-122 Assay on ddPCR	161
Figure 5-10: Validation of KRAS-G13D on ddPCR	162
Figure 5-11: Validation of APC1338 Assay on ddPCR	165
Figure 5-12: Validation of APC1397 Assay on ddPCR	166
Figure 5-13: Validation of APC1429 Assay on ddPCR	167
Figure 5-14: Validation of TP53-273 Assay on ddPCR	168
Figure 5-15: Validation of PIK1047 Assay on ddPCR	169
Figure 6-1: Histopathological Information of Patient Samples	174
Figure 6-2: Representative Histological Subtypes in H&E Stained FFPE samples	175
Figure 6-3: Frequently Mutated Genes in Colon Adenomas from the COSMIC Database	177
Figure 6-4: Distribution of <i>BRAF</i> and <i>KRAS</i> Mutations in the Colon	178
Figure 6-5: Comparison of Total Plasma cfDNA Levels in the Patient & Control Groups	180
Figure 6-6: Testing Lyophilised Plasma cfDNA Samples using the BRAF-V600E Assay	182
Figure 6-7: Testing for <i>KRAS</i> c.34G Mutations in the Plasma cfDNA Samples using the KRAS-121 Assay	184
Figure 6-8: Testing for <i>KRAS</i> c.35G Mutations in the Plasma cfDNA Samples using the KRAS-122 Assay – 1D Plots	187
Figure 6-9: Testing for <i>KRAS</i> c.35G Mutations in the Plasma cfDNA Samples using the KRAS-122 Assay – Graphs for all Samples tested	188
Figure 6-10: Testing for <i>KRAS</i> c.38G Mutation in the Plasma cfDNA Samples using the KRAS-G13D Assay – 1D Plots	192
Figure 6-11: Testing for <i>KRAS</i> c.38G Mutation in the Plasma cfDNA Samples using the KRAS-G13D Assay – Graphs for all Samples tested	193
Figure 6-12: Summary of Multiregional Sequencing Data for Case H263	198
Figure 6-13: Summary of Multiregional Sequencing Data for Case H265	199
Figure 6-14: Summary of Multiregional Sequencing Data for Case H266	200
Figure 6-15: Comparing the PCR & NGS Data on the Percentage MU Abundances in Tissue Samples	202
Figure 6-16: Analysis of Patient Case H266 using the APC1338 Assay	205
Figure 6-17: Analysis of Patient Case H265 using the APC1397 Assay	206
Figure 6-18: Analysis of Patient Case H263 using the APC1429 Assay	207
Figure 6-19: Analysis of Patient Case H266 using the TP53-273 Assay	208
Figure 6-20: Analysis of Patient Case H263 using the PIK1047 Assay	209
Figure 8-1: Summary of Multiregional Sequencing Data for Case H149	231
Figure 8-2: Summary of Multiregional Sequencing Data for Case H154	232
Figure 8-3: Summary of Multiregional Sequencing Data for Case H264	233

List of Tables

Table 1-1: Summary of CRC Biomarkers & Potential Clinical Applications	53
Table 2-1: List of Media & Buffers used for Cell Culture	63
Table 2-2: Miscellaneous Buffers & Solutions used in DNA Extraction	64
Table 2-3: List of Buffers included in DNA Clean & Concentrator Kit	64
Table 2-4: Lists of Buffers included in Qiagen DNA Extraction Kits	65
Table 2-5: List of Reagents for Traditional End-point PCR Assays	66
Table 2-6: Primers for Traditional PCR Assays	66
Table 2-7: List of Reagents for qPCR & ddPCR Assays	67
Table 2-8: Primers & Probes for Analysis of <i>In Vivo</i> Samples	67
Table 2-9: Primers & Probes for Analysis of Patient Samples	68
Table 2-10: List of Reagents & Solutions for <i>In Vivo</i> Studies	69
Table 2-11: List of Buffers & Solutions for Immunohistochemistry	70
Table 2-12: Cell Lines used for extracting positive control DNA	72
Table 4-1: Genotype, Phenotype & Health Data for Mice used in 1 st <i>In Vivo</i> Study	105
Table 4-2: Polyp Counts in <i>Cre-Apc^{fl/fl}</i> Mice from 1 st Study	106
Table 4-3: Genotype, Phenotype & Health Data for Mice used in 2 nd <i>In Vivo</i> Study	112
Table 4-4: Macroscopic Polyp Counts in <i>Cre-Apc^{fl/fl}</i> Mice from the 2 nd Study	113
Table 4-5: Poisson Analysis for Positive Detection of Recombined Alleles	120
Table 4-6: Genotype, Phenotype & Health Data for Mice used in 3 rd <i>In Vivo</i> Study	122
Table 4-7: Microscopic Polyp Counts in <i>Cre-Apc^{fl/fl}</i> Mice from the 3 rd Study	125
Table 4-8: Histological Grading of Polyps in the Proximal SI & Caecum of Apc-deleted Mice in NFD & HFD Groups	129
Table 4-9: Health & Welfare Analysis Data for 3 Positive Samples	141
Table 4-10: Poisson Analysis for the Samples that were Positive for Recombined Alleles	141
Table 4-11: Total DNA Yields from Faecal Samples	143
Table 4-12: Statistical Analysis of Percentage Fractional Abundance of Recombined Alleles in Faecal Samples from Apc-deleted Mice (companion to Figure 4-22)	145
Table 5-1: LoDs for Patient-specific Mutation Detection Assays	170
Table 6-1: Poisson Distribution Analysis for Mutant-positive cases from Figure 6-7	185
Table 6-2: Histology Data & Mutation Frequencies for the Matched FFPE Cases Corresponding to Plasma cfDNA Samples Analysed in Figure 6-7	185
Table 6-3: Histology Information for the Matched FFPE Cases Corresponding to Plasma cfDNA Samples Analysed in Figure 6-9	189
Table 6-4: Poisson Distribution Analysis for MU-positive cases from Figure 6-9	190
Table 6-5: Histology Information for the Matched FFPE Cases Corresponding to Plasma cfDNA Samples Analysed in Figure 6-11	194

Table 6-6: Poisson Distribution Analysis for MU-positive cases from Figure 6-11	195
Table 6-7: Histology Details on the Sequenced Patient Cases	197

List of Abbreviations

7TMR	Seven transmembrane receptor
AJCC	American Joint Committee on Cancer
AKT/PKB	Ak strain transforming/ protein kinase B
Alu	Arthrobacter luteus
AML	Acute myelogenous leukaemia
AMPK	Adenosine monophosphate-activated protein kinase
APC	Adenomatous polyposis coli
ARAF	V-Raf murine sarcoma 3611 viral oncogene homolog 1
ASCL2	Achaete–Scute complex homologue 2
BAT25	Big adenine tract 25
BAT26	Big adenine tract 26
BCSP	Bowel Cancer Screening Programme
BEAMing	Beads, Emulsion, Amplification and Magnetics
BMP3	Bone morphogenetic protein 3
BRAF	V-Raf murine sarcoma viral oncogene homolog B
BUB	Budding uninhibited by benzimidazoles
BUBR1	Budding uninhibited by benzimidazoles 1 homolog beta
CA	Carbohydrate antigen
CACNA1G	Calcium voltage-gated channel subunit alpha1 G
CAPP	Concerted Action Polyposis Prevention
CaPP	Colorectal Adenoma/Carcinoma Prevention Programme
CAPP-Seq	CANcer Personalised Profiling by deep Sequencing
CBC	Crypt base columnar
CCND1	Cyclin D-1
CDKN2A	Cyclin-dependent kinase Inhibitor 2A
CEA	Carcinoembryonic antigen
cfDNA	Cell-free DNA
CGH	Comparative genomic hybridisation
CIMP	CpG island methylator phenotype
CIN	Chromosomal instability
CK1 α	Casein kinase 1 α
CKI	Cyclin-dependent kinase inhibitor
CL	Confidence level
COSMIC	Catalogue of somatic mutations in cancer
COX	Cyclo-oxygenase
CR	Conserved region
CRABP1	Cellular retinoic acid binding protein 1
CRAF	V-Raf-1 murine leukemia viral oncogene homolog 1
CRC	Colorectal cancer
CRD	Cysteine-rich domain
Cre	Cyclization recombination

CRUK	Cancer Research UK
Ct	Cycle threshold
ctDNA	Circulating-tumour DNA
CTNNB1	Catenin beta 1
DAB	Diaminobenzidine
ddPCR/dPCR	(Droplet) Digital Polymerase Chain Reaction
DPI ₃	Dihydrocyclopyrroloindole tripeptide
DVL	Dishevelled segment polarity
EB1	End-binding protein 1
EGF	Epidermal growth factor
EGFP	Enhanced green fluorescent protein
EGFR	Epidermal growth factor receptor
ELISA	Enzyme-linked immunosorbent assay
EMA	European Medicines Agency
ER	Estrogen receptor
ERK	Extracellular signal regulated kinases
FAP	Familial adenomatous polyposis
FCP	Faecal calprotectin
FDA	Food and Drug Administration
FFPE	Formalin-fixed, paraffin-embedded
FIT	Faecal immunochemical test
FOX	Forkhead box
FP	Forward primer
Fz	Frizzled
GAP	GTPase activating protein
GDP	Guanosine diphosphate
GEF	Guanine nucleotide exchange factor
GEMM	Genetically engineered mouse model
gFOBT	Guaiac faecal occult blood test
GI	Gastrointestinal
GNAS	Guanine nucleotide binding protein, alpha stimulating
GOF	Gain-of-function
GRB2	Growth factor receptor-bound protein 2
Grg	Gro-related gene
GSK3 β	Glycogen synthase kinase β
GTP	Guanosine triphosphate
H&E	Haematoxylin & Eosin
HDL	Human disc large
HEX	Hexachloro fluorescein
HFD	High fat diet
hgDNA	Human genomic DNA
hMLH1	Human MutL homolog 1
hMSH2	Human mutS homolog 2

HNPCC	Hereditary non-polyposis colorectal cancer
HP	Hyperplastic polyp
HRAS	Harvey rat sarcoma viral oncogene homolog
IGF1	Insulin-like growth factor 1
IGF2	Insulin-like growth factor 2
IPH	Inter-patient heterogeneity
IQGAP1	Ras GTPase-activating-like protein 1
IRES	Internal ribosomal entry sequence
ISC	Intestinal stem cell
ITH	Intra-tumour heterogeneity
IUPAC	International Union of Pure and Applied Chemistry
KRAS	Kirsten rat sarcoma viral oncogene homolog
KSR1	Kinase suppressor of RAS-1
LEF	Lymphoid enhancing factor
LGR5	Leucine-rich repeat-containing G-protein coupled receptor 5-expressing
LoD	Limit of detection
LOH	Loss of heterozygosity
loxP	locus of X(cross)-over in P1
LRPs	Lipoprotein receptor-related proteins
MAD	Mitotic arrest deficient
MAPK	Mitogen-activated protein kinase
MCR	Mutation cluster region
MEK	Mitogen-activated protein kinase kinase
MET	Metabolic equivalent task
mGapdh	Mouse glyceraldehyde 3-phosphate dehydrogenase
mgDNA	Mouse genomic DNA
MMR	Mismatch repair
MS	Microsatellite
MSI	Microsatellite instability
mTOR	Mammalian target of rapamycin
MU	Mutant
MUP	Mutant probe
MYC	V-Myc viral oncogene homolog
NDRG4	N-Myc downstream-regulated gene 4
Neo	Neomycin
NES	Nuclear export sequence
NET	Neutrophil extracellular trap
NEUROG1	Neurogenin 1
NFD	Normal fat diet
NF- κ B	Nuclear factor- κ B
NFQ	Non-fluorescent quencher
NGS	Next generation sequencing

NHS	National Health Service
NLS	Nuclear localisation sequence
NRAS	Neuroblastoma RAS viral oncogene homolog
NREC	NHS Research Ethics Committee
NSAID	Non-steroidal anti-inflammatory drug
NSCLC	Non-small cell lung cancer
NTC	Non template control
OIS	Oncogene-induced senescence
OLFM4	Olfactomedin 4
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PDK1	3-phosphoinositide-dependent kinase 1
PGK1	Phosphoglycerate kinase 1
PI3K	Phosphatidylinositol-4,5-bisphosphate 3-kinase
PIK3CA	Phosphatidylinositol-4,5-Bisphosphate 3-Kinase Catalytic Subunit Alpha
PIP ₂	Phosphatidylinositol 4,5-bisphosphate
PIP ₃	Phosphatidylinositol 3,4,5-trisphosphate
PNA	Peptide nucleic acid
qPCR	(Real-time) quantitative polymerase chain reaction
RAF	Rapidly accelerated fibrosarcoma oncogene homolog
RAS	Kirsten rat sarcoma viral oncogene homolog
RBD	RAS binding domain
RCT	Randomised clinical trials
RER	Replicative error
Rn	Normalised reporter
Rnf43	Ring finger protein 43
RP	Reverse primer
RTK	Receptor tyrosine kinase
RUNX3	Runt related transcription factor 3
SAMP	Serine-Alanine-Methionine-Proline repeats
SCF	Skp1–Cullin1–F-box protein
SH2	Src homology 2
SI	Small intestine
SMAD2	Sma- and mad-related protein 2
SMAD4	SMAD Family Member 4
SOCS1	Suppressor of cytokine signalling 1
SOS	Son-of-sevenless
SSA/P	Sessile serrated adenomas/polyp
TA	Tubular adenoma
TAC	Transit-amplifying cell
Tam-Seq	Tagged-amplicon deep Sequencing
TBE	Tris Borate, EDTA

TCF	T-cell factor
TE	Trypsin-EDTA
TFPI2	Tissue factor pathway inhibitor 2
TGF- β	Transforming growth factor β
T_m	Melting temperature
TNM	Tumour, Lymph Node, Metastasis
TP53	Tumour protein p53
TSA	Traditional serrated adenoma
TVA	Tubulovillous adenoma
UHL	University Hospital Leicester
VA	Villous adenoma
VEGF	Vascular endothelial growth factor
WBC	White blood cells
WES	Whole Exome Sequencing
WGS	Whole Genome Sequencing
WHO	World Health Organisation
WNT	Wingless/Integrin-1
WRE	Wnt-responsive element
WT	Wild-type
WTP	Wild-type probe
Znrf3	Zinc and ring finger 3
β -TrCP	β -transducin repeats-containing protein
ΔR_n	Baseline-corrected normalized reporter

List of Publications & Conference Abstracts

Abstracts

- Moe Myint, N.N., Pringle, J.H., Shaw, J., Trigg, R., Brown, K. & Rufini, A. Allele-specific Mutation Detection qPCR: A Sensitive Method for Detection of Early Colorectal Cancer Mutations Using Circulating Cell-free DNA. Poster presentation. 44th European Environmental Mutagen Society (EEMS) Annual Conference. Lancaster, United Kingdom, July 2014.
- Moe Myint, N.N., Pringle, J.H., Shaw, J., West, K., Verma, A.M., Baljit, S., Brown, K. & Rufini, A. Circulating Cell-free DNA: Hide-and-Seek in Blood to Catch Cancer Early. Poster presentation. 2nd European Association for Cancer Research (EACR) Special Conference: Cancer Genomics. Cambridge, United Kingdom, June 2015.
- Moe Myint, N.N., West, K., Verma, A.M., McDermott, U., Baljit, S., Pringle, J.H., Shaw, J., Brown, K. & Rufini, A. The Role of Circulating Tumour DNA as a Predictive Biomarker for Chemoprevention of Colorectal Cancer. Poster presentation. British Association for Cancer Research (BACR) & European Association for Cancer Research (EACR) Joint Meeting: Therapeutic Interventions for Cancer Prevention – the Way Forward. Bristol, United Kingdom, July 2016.

Chapter 1 Introduction

1.1 Gastrointestinal Tract

The gastrointestinal (GI) tract is a long, tubular organ made up of specialised structures and is responsible for digestion, absorption of nutrients and water, and efficient excretion of waste products [1, 2]. The GI tract is split into two sections. The upper section consists of oesophagus, stomach and duodenum (proximal small intestine (SI)), and most of the food digestion takes place here [1, 2]. The lower section consists of jejunum (medial SI), ileum (distal SI), caecum and large intestine (LI; also known as colon) and rectum (Figure 1-1-A, [3]), and nutrient absorption and formation of waste products occur here [1, 2].

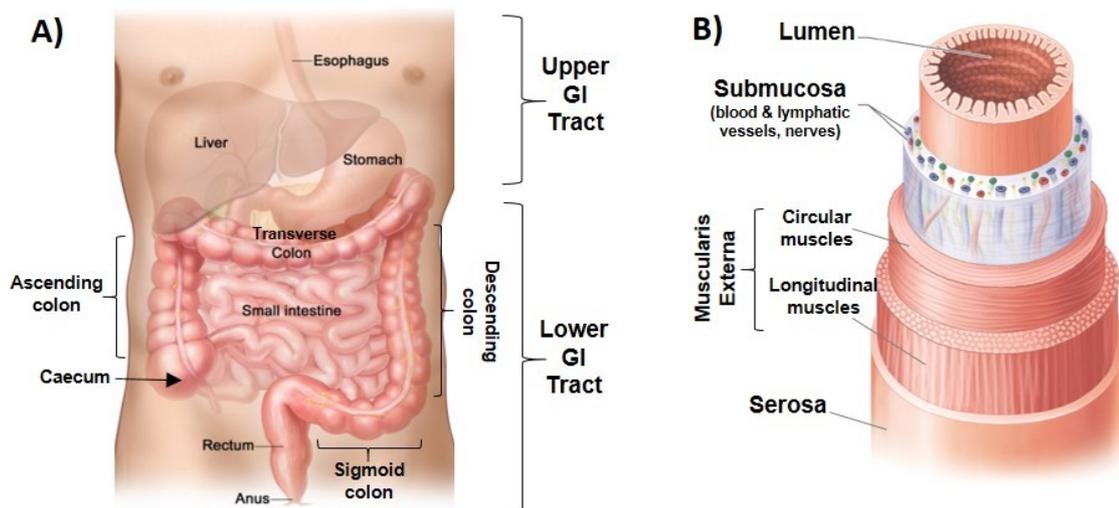


Figure 1-1: Structure of the Gastrointestinal Tract

A) Structural organisation of the GI tract that is split in to the upper and lower tracts. [2]

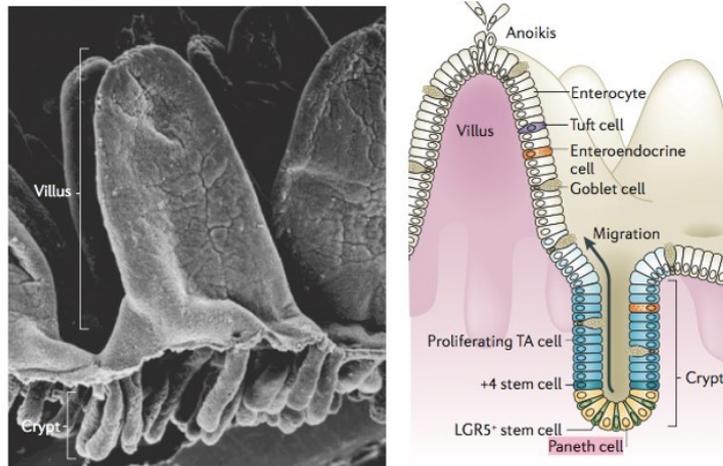
B) Organisation of tissue layers in the GI tract. [4]

The GI tract is layered into four different sections (Figure 1-1-B, [4]) [2]. The innermost mucosa opens into the lumen, and it is further split into an epithelial layer, lamina propria and a thin muscle layer called muscularis mucosae [1]. The mucosa is connected to the submucosa, which is a layer of connective tissues containing blood vessels, lymphatic vessels and nerves [1]. Next is the muscularis externa, consisting of an inner circular and an outer longitudinal smooth muscle layers that are responsible

for peristalsis [1]. The outermost serosa layer is mainly composed of connective tissues that helps with structural integrity of the GI tract and secretes serous fluid to reduce friction.

The epithelium is arranged into functional structures of crypts and villi [1]. Crypts are invaginations in the layer that contains a base of proliferating intestinal stem cells (ISCs), also known as crypt base columnar (CBC) stem cells, interspersed with few Paneth cells whose function is to provide immunity via secretion of anti-pathogenic enzymes (Figure 1-2) [5, 6]. Villi are cellular protrusions into the lumen, and they mostly consist of absorptive enterocytes and some goblet cells for mucus secretion (Figure 1-2) [5, 7]. There are also enteroendocrine cells that secrete hormones such as catecholamines for GI regulation (Figure 1-2) [5, 7]. CBCs at the crypt bottom continually divide and give rise to transit-amplifying cells (TAC), which migrate upwards to the villi where they differentiate and replace the dying cell at the tip of villi (Figure 1-2) [5, 7]. On average, intestinal epithelial cells have a half-life of 3-5 days and up to 10^{11} cells are lost daily [2, 7]. In the SI, each villus is surrounded by approximately 6 crypts, each with a stem cell compartment to help maintain this constant cell turnover [7]. Activity of CBC is regulated by complex signalling between neighbouring epithelial and stromal cells [8-10]. Various *in vivo* studies have identified molecular markers of CBCs, including achaete-Scute complex homologue 2 (*ASCL2*), olfactomedin 4 (*OLFM4*) and leucine-rich repeat-containing G-protein coupled receptor 5 (*LGR5*), which are all involved in the Wntless/Integrin-1 (WNT) pathway, the main signalling pathway responsible for the maintenance of the crypt stem-cell niche [7, 8, 11, 12]. Of these, *LGR5* is most commonly used as a stem-cell marker, and it has been estimated that between 3-16 *LGR5*⁺ CBCs are involved in maintaining the epithelial homeostasis [7-9, 13].

A) Small Intestine



B) Colon

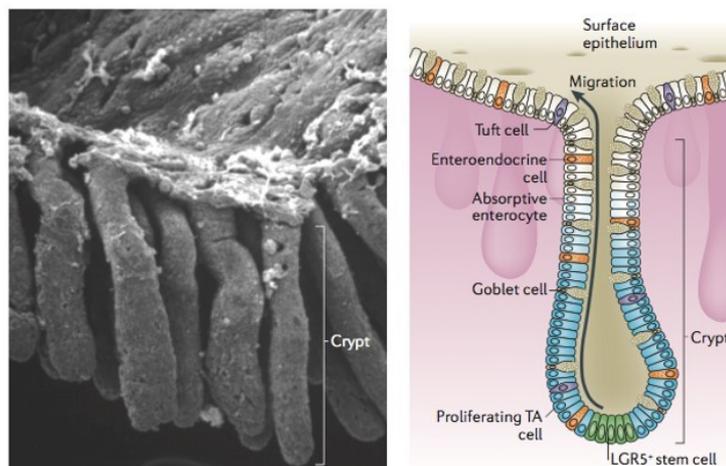


Figure 1-2: Structural Differences between the SI & Colon Epithelium

Electron micrograph (left) and cellular organisation (right) pictures of epithelium.

A) SI epithelium has long villi to aid nutrient absorption. The crypt base is occupied by *LGR5*⁺ CBC cells that continuously generate proliferating transit-amplifying cells (TA cells). The +4 stem cells act as reserves in case of loss or damage. [7]

B) Colon has a prominent, flat epithelial surface, but with longer crypts. Goblet cells are abundant, which secrete mucus to aid the passage of solid waste products. Paneth cells are absent in the colonic crypts. [7]

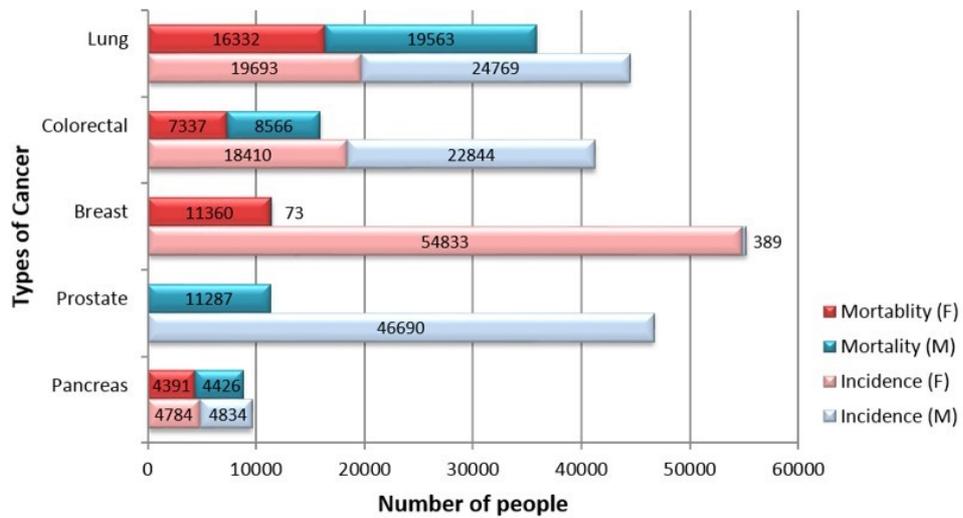
The structural organisation of gut epithelium changes distinctly as it transverses from the SI to LI, with a prominent decrease in villi length to almost lack of villus structure in the LI (Figure 1-2) [7, 14]. As food travels along the upper GI, complex macromolecules (e.g. carbohydrates) are digested into smaller units (e.g. monosaccharides) by the activity of digestive enzymes and gastric acid [1, 2]. Most nutrients are absorbed in the SI whose epithelial layer is arranged into long villi with a

high density of absorptive enterocytes (Figure 1-2) [7]. Residual ions and water contents are absorbed in the LI, where the solid, indigestible materials are formed into faeces until they reach the rectum [1, 2]. The colonic epithelium is structured with a high density of goblets cells that secrete mucus to aid with the passage of faecal matters (Figure 1-2) [7].

1.2 Colorectal Cancer

Colorectal cancer (CRC) is the fourth most prevalent cancer and the second leading cause of cancer-related deaths in United Kingdom (UK), where it accounts for approximately 12% of new cancer cases diagnosed each year (Figure 1-3-A, [15]) [15]. CRC is a heterogeneous disease and shows a predilection for the rectum and sigmoid regions of the lower GI tract in both genders (Figure 1-3-B, [16]) [16]. Since the 1990s, the incidence rate for CRC in the UK has seen a slight increase of approximately 4% [15, 17]. Despite the improvement in the overall 5-year survival rate of CRC, currently standing at approximately 59%, the rates of morbidity and mortality caused by CRC could be significantly reduced if patients are to be diagnosed at earlier stages [18]. Indeed, a drastically higher 5-year survival rate (>90%) is associated with diagnosis at stage I compared to <10% for those diagnosed at stage IV [18]. Currently, more than 50% of cases were diagnosed at a late stage across the UK [18, 19]. This undoubtedly highlights the need for clinically efficient means of early diagnosis and effective preventive strategies that can help reduce the cancer burden.

A) Top 5 Cancer Mortalities in UK



B)

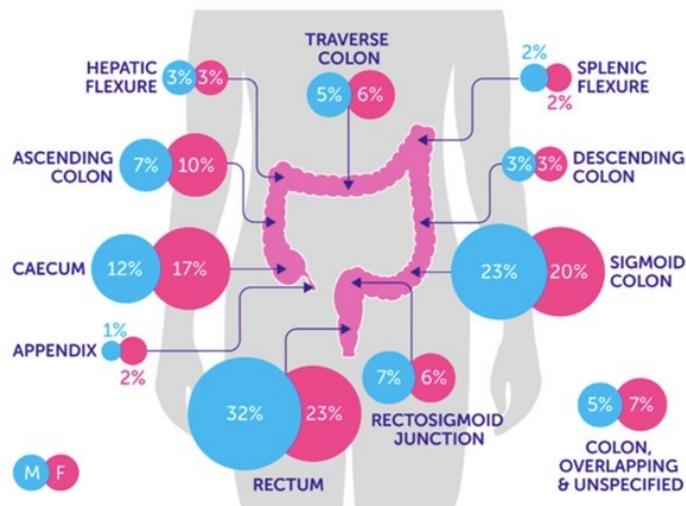


Figure 1-3: Statistics & Anatomical Distribution of CRC in UK

A) Graph comparing the gender (M = males; F = females) ratio of incidence and mortality for the five most common cancers in the UK. [15]

B) Anatomical distribution of CRC. Incidences are the most frequent in the rectum, followed by the sigmoid colon in both genders. [16]

1.2.1 Histology of Colorectal Lesions

Over 90% of CRCs have mucosa epithelial origin, with few cases arising from sources like neuroendocrine and squamous cells [20]. All CRCs start out as polyps, and macroscopically they are either sessile (i.e. raised, flat lesions on the mucosal layer) or pedunculated (stalked lesions extending from the mucosa) [20, 21]. There are two main categories of precancerous lesions: serrated and traditional adenomatous polyps.

The traditional adenomas are further classified as tubular adenomas (TA), tubulovillous adenomas (TVA) and villous adenomas (VA) subtypes, based on the percentage contribution of villous components [20]. TAs are the most common subtype (65-80%) and often seen as pedunculated polyps with <25% villous components [21, 22]. VAs are the least common subtype at 5-10% of cases, and they are characterised by >75% villous dysplasia and seen as stalk-like projections [20, 21]. TVAs contribute to 10-25% of cases and has features of both TAs and VAs [20, 21]. They also show common dysplastic features such as rounded nuclei, prominent nucleoli and depolarised epithelial cells [20].

Serrated polyps are also a heterogeneous group consisting of hyperplastic polyps (HP), sessile serrated adenomas/polyp (SSA/P) and traditional serrated adenoma (TSA) subtypes, and they feature characteristic epithelial serrations along the villi [20, 23]. HPs are the most common type (80-90%) with the propensity for left colon and rectum, and are small (<5mm diameter) with elongated crypts, prominent luminal serrations [20, 23]. SSA/Ps are the next common type (15-20%) that show predilection for the right colon [20, 23]. They are larger than HPs with >5mm in diameter, and have prominent crypt serrations, reduced stromal area and hyper-mucinous epithelium [20, 23]. TSAs are the least common type (1-5%) and frequently grow in the left colon [23]. They are mostly pedunculated and typically display low grade dysplasia [20, 23].

Additionally, the term 'mixed polyps' is also widely used in pathological analysis to refer to lesions that show combined features of the traditional and serrated polyps (e.g. TSA and SSA) with varying degrees of cytological dysplasia [23].

1.2.2 Cancer Staging

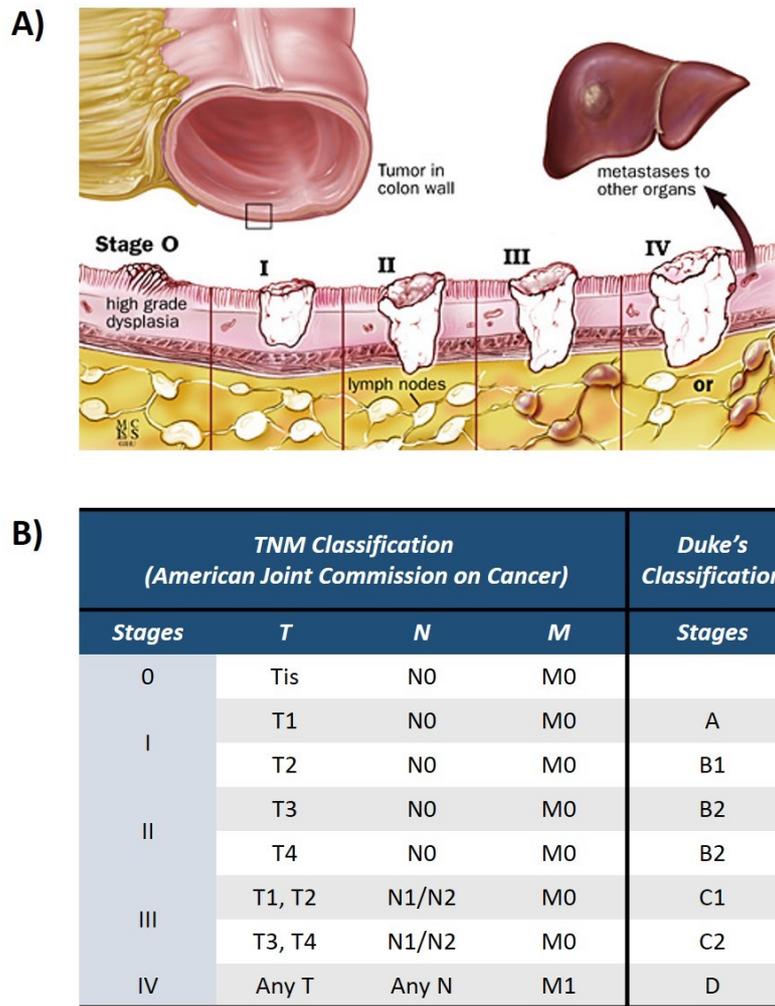


Figure 1-4: Comparison of CRC Classification Systems

A) The TNM (Tumour, Lymph Node, Metastasis) pathological staging system determines stages of CRC from benign polyps (stage 0) to carcinoma (stage IV). [20]

B) Comparison between the TNM and Dukes' staging systems - Duke's system includes Astler-Coller modification for stage B and C divisions. Tis = in situ carcinoma (low/high-grade), T1 = submucosa, T2 = muscularis propria, T3 = subserosa or perirectal tissues, T4 = other organs; N0 = no node metastasis, N1 = 1-3 nodes, N2 = >4 nodes; M0 = no distance metastasis, M1 = distant metastasis. [24]

As the benign colorectal lesions progress to carcinoma, cancer staging is carried out to support the treatment process and prediction of prognostic outcome in the clinics [20]. Cancer staging is the process of determining the severity of the disease based on the anatomical occupancy of the cancerous mass. The newer TNM staging (Figure 1-4), endorsed by American Joint Committee on Cancer (AJCC), has been widely used in the clinics for diagnostic assessment worldwide. It determines the cancer stage based on

the location and size of tumours (T), the degree of spread to the lymph nodes (N) and whether it has metastasised or not (M) (Figure 1-4; A & B, [24]). There is another system called Dukes' staging that implements less detailed classification criteria than the TNM system, but it remains applicable for epidemiological studies due to its widespread use over the years (Figure 1-4-B) [19, 20].

1.2.3 Symptoms & Diagnosis

Diagnosis of CRC can be challenging due to the lack of specific symptoms associated with the disease, and the commonality of symptoms with other non-neoplastic diseases such as irritable bowel syndrome [25, 26]. Several symptomatic presentations can indicate the underlying colorectal disease in an undiagnosed individual [25, 26]. A change in bowel habits, be it diarrhoea or constipation, persistent abdominal pain and rectal bleeding are the conspicuous abdominal symptoms that may prompt an affected individual to seek consultation with a general practitioner for a decisive diagnosis [25, 26]. Symptoms may also present as weight loss and fatigue, which are unlikely to be caused by either dietary or lifestyle choices [25, 26]. As reported by one study with 194 CRC patients, more than half of patients presented with at least one abdominal symptom with the rectal bleeding being the most common presentation in 58% of patients [27].

The diagnostic process from the first detection of symptoms to confirmation by clinical procedures, such as colonoscopy, is a lengthy process with the median duration of ≈ 4 months [25, 26]. Additionally, an individual at risk may also contribute to a delay in diagnosis due to the misinterpretation of the symptoms as being caused by a less serious malaise or an inevitable consequence of an old age [25, 26]. These factors contribute to more than half of CRC cases diagnosed at late stages with consequences of poorer prognosis and reduced survival rates [18, 19]. This highlights the need for less-invasive procedures that can shorten the diagnostic process for the benefit of patients.

1.2.4 Treatment Options

Treatment decisions are made based on the stage of CRC. Surgery is the standard treatment modality for the localised stage 0-II cases [28, 29]. For stage III & IV cases, chemotherapy may either precede or follow surgical procedures [28, 29]. For small superficial masses in stage 0 CRC, local polypectomy is sufficient [28]. However, surgical resection of part of colons, called colectomy, may be necessary for larger localised masses of stage I-II cases – colectomy may be partial when performed to remove selected sections of colon [28, 29]. Surgery for advanced cases may result in removal of the whole colon in a procedure called subtotal colectomy [28, 29]. Colectomy is either followed by anastomosis where the two remaining ends of colon are surgically reconnected, or by colostomy where the healthy colon is attached to the outer abdominal wall to create a stoma through which faecal matters can be excreted [28, 29]. For <T4 tumours that are smaller than 8cm, laparoscopic (keyhole) surgery is possible with the benefit of faster recovery [28, 29].

Chemotherapy can be provided as an adjuvant treatment to reduce the risk of recurrence following surgery, as a neoadjuvant treatment to minimise a tumour mass to enable surgical procedures and as a palliative option to improve the quality of life [28]. Oxaliplatin is a platinum containing cytotoxic agent that is usually combined with 5-fluorouracil (5-FU) and leucovorin, which is a folinic acid that increases efficacy of 5-FU, in the FOLFOX regimen or with capecitabine (oral form of 5-FU) in the XELOX regimen [28]. Irinotecan is another cytotoxic agent that can replace oxaliplatin in a combination treatment with 5-FU/leucovorin (FOLFIRI) or with capecitabine (XELIRI) [28]. For metastatic cases, targeted monoclonal antibody treatments such as bevacizumab, an anti-vascular endothelial growth factor, and cetuximab, an anti-epidermal growth factor receptor, may be given in conjunction with chemotherapy combination regimens to improve prognosis [28, 30]. For rectal cancers, neoadjuvant chemoradiotherapy is the standard to reduce a tumour mass prior to surgical resection [28, 31].

1.3 Molecular Pathology of CRCs

CRCs arise from either sporadic ($\approx 80\%$) or germline ($\approx 20\%$) causes, and disease progression from early colorectal polyps into malignant carcinoma is accompanied by gradual accumulation of genetic and genomic aberrations [32-35].

1.3.1 WNT/ β -catenin Pathway

1.3.1.1 Pathway Overview

The WNT/ β -catenin signalling pathway controls a plethora of physiological processes by transducing signals from cell surface to the nucleus (Figure 1-5) [36, 37]. *Wnt* is a proto-oncogene that encodes a 40kDa lipidated ligand, which binds and activates seven transmembrane receptors (7TMRs) called Frizzled (Fz) that are complexed to lipoprotein receptor-related proteins (LRPs) 5/6 for functional activity (Figure 1-5). Fz receptors contain a large extracellular cysteine-rich domain (CRD) with a hydrophobic groove for Wnt binding via its lipid moiety [38].

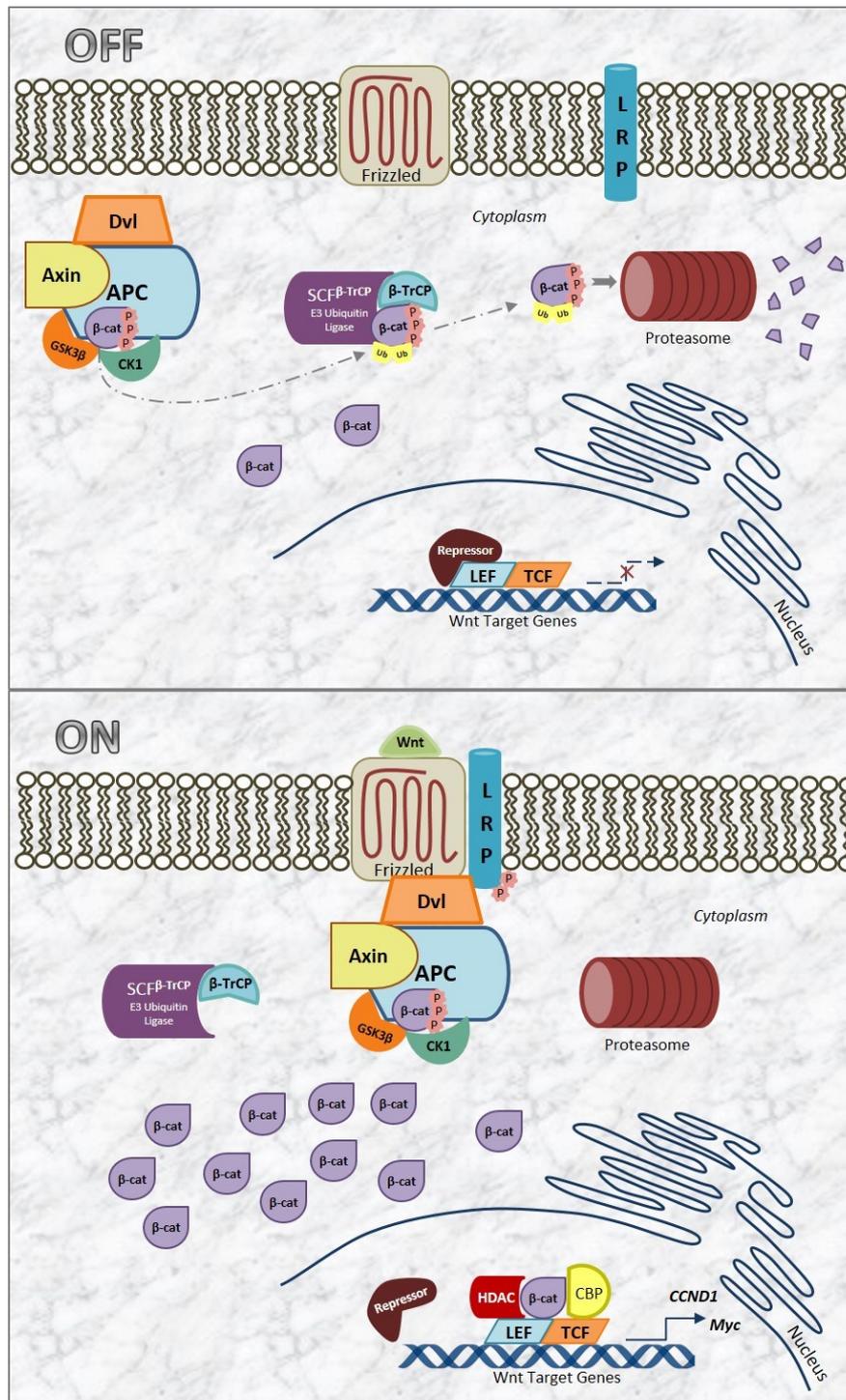


Figure 1-5: Canonical WNT/Frizzled Signalling Pathway in ON & OFF States

In an **OFF** state, Frizzled 7TMR and LRPs remain dissociated due to lack of Wnt ligand binding. This enables the 'cytoplasmic destruction' complex to form using APC as a scaffold: the axin protein tethered to the complex binds β -catenin that becomes phosphorylated by GSK3 β and CK1 kinases – the phosphorylated β -catenin is then destined for proteasomal degradation. [38]

In an **ON** state, the destruction complex on APC-scaffold is recruited to the Wnt-bound Frizzled 7TMR. This hinders the phosphorylation and subsequent ubiquitination of β -catenin, allowing cytoplasmic β -catenin to translocate to the nucleus and activate gene transcriptions. [38]

In the absence of WNT ligand (OFF stage; Figure 1-5), dishevelled segment polarity (DVL), adenomatous polyposis coli (APC), glycogen synthase kinase β (GSK3 β) and casein kinase 1 α (CK1 α) tether to the Axin scaffolding protein to form a cytoplasmic supramolecular destruction complex [37, 39]. This recruits and phosphorylates β -catenin at N-terminal serine (Ser) and threonine (Thr) residues: CK1 α phosphorylates at Ser45, which promotes sequential phosphorylation at Thr41, Ser37 and Ser33 of β -catenin by GSK3 β [37, 39]. Phosphorylated β -catenin is then ubiquitinated by the β -transducin repeats-containing protein (β -TrCP) subunit of the Skp1–Cullin1–F-box (SCF $^{\beta$ -TrCP) E3 ubiquitin ligase [37, 39]. This induces proteasomal degradation of β -catenin and prevents activation of β -catenin nuclear targets [37-39].

When WNT ligand binds to the Fz/LPR5-6 receptor (ON stage; Figure 1-5), it induces a conformational change, leading to phosphorylation of the cytoplasmic tail of LRP by the serine-threonine kinases GSK3 β and CK1 γ (Figure 1-5) [40]. The phosphorylated tail recruits Axin to the plasma membrane in a DVL-dependent manner, resulting in the inactivation of the destruction complex (Figure 1-5) [37, 40]. The newly synthesised, unphosphorylated β -catenin is then able to translocate to the nucleus. In the absence of nuclear β -catenin, the T-cell factor/lymphoid enhancing factor (TCF/LEF) family of transcription factors bind to the DNA at the consensus WNT-responsive elements (WRE) and complex with co-repressors (e.g. Groucho, Gro-related gene (Grg)) to block transcription of WNT-dependent genes [39, 40]. However, nuclear β -catenin binds to the N-terminus of the TCF/LEF and replaces co-repressors to promote recruitment of transcriptional co-activators and histone modifiers (e.g. CREB-binding protein), resulting in the active transcription of many WNT-regulated genes, including cyclin D-1 (*CCND1*), V-Myc viral oncogene homolog (*MYC*) and *LGR5* [37-39, 41].

The WNT pathway is crucial for developmental signalling as well as maintenance of various tissues in adults via regulation of cellular differentiation, proliferation and migration [42]. In the GI tract, WNT signals play an important role in maintaining the stem-cell niche at the crypt base where the activity is at its highest [2, 7, 8, 13, 42]. Wnt pathway also engages in a negative feedback loop via expression of ring finger protein 43 (Rnf43) and zinc and ring finger 3 (Znrf3) transmembrane E3 ligases, which promotes endocytosis and lysosomal destruction of Wnt receptors [43]. However, interaction of small secreted R-spondin (Rspo) proteins with Lgr5 leads to membrane clearance of

Rnf43 and Znf3, resulting in the potentiation of WNT signalling [43]. Due to its role in the homeostasis of CBC cells, majority of CRCs are driven by unregulated activation of WNT pathway [37, 44-46].

1.3.1.2 APC Tumour Suppressor Gene

APC is a tumour suppressor gene that encodes a large, multi-domain protein of 2843 amino acids (Figure 1-6) [36, 37]. In mammals, there are two APC genes: APC is expressed in foetal and epithelial cells, and APC2 homologue is expressed ubiquitously [36]. The APC gene is located on the chromosome arm 5q21 and consists of 8535 base pairs across 21 exons [36, 37].

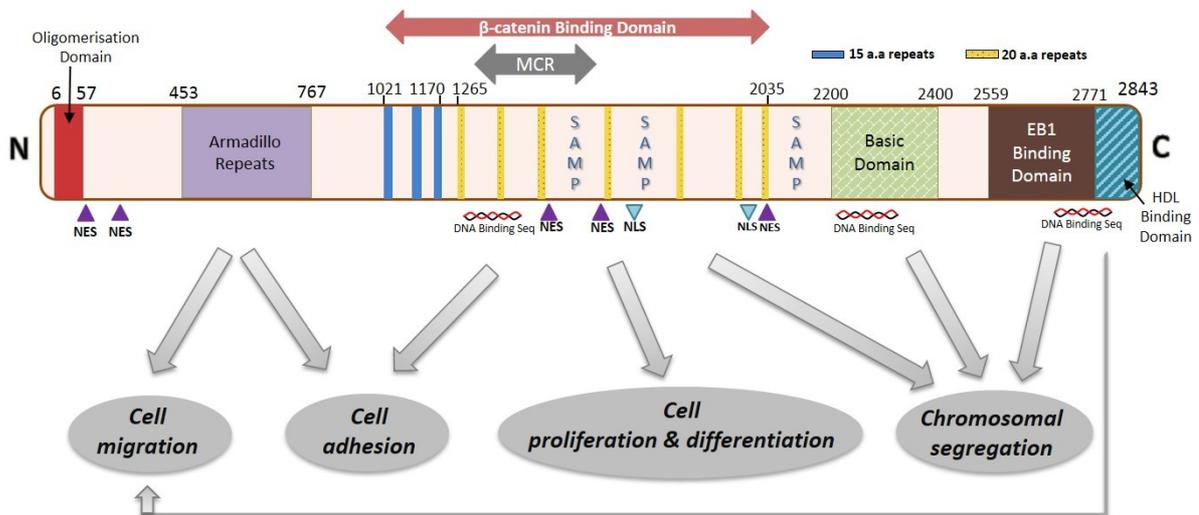


Figure 1-6: Structural Domains & Functions of APC

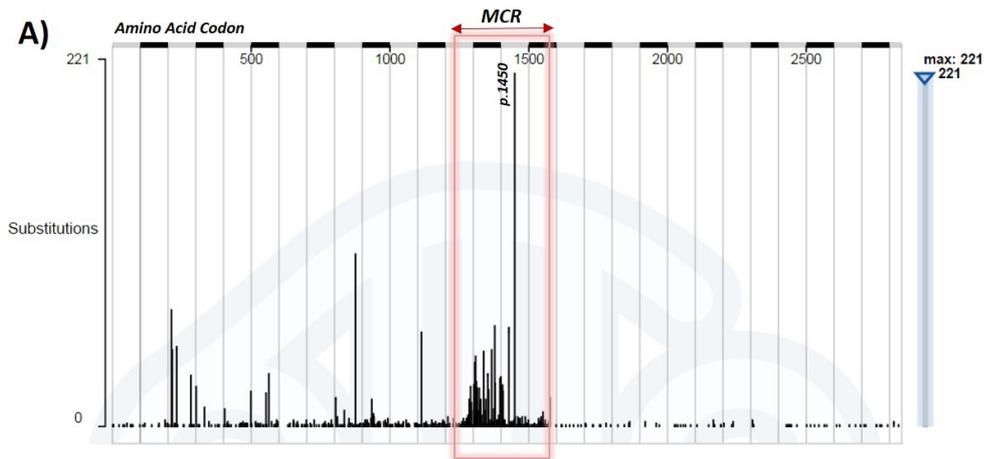
Structural domains linked to various functions of the APC protein. Mutation cluster region (MCR) is located within the β -catenin binding domain. SAMP = Serine-Alanine-Methionine-Proline repeats, EB1 = end-binding protein 1, HDL = Human Disc Large (HDL) binding domain, NES = nuclear export sequences, NLS = nuclear localisation sequences (NLS). [37]

At the N-term, there is an oligomerisation domain that facilitates homo-dimerisation between APC proteins, and armadillo repeats that play a central role in regulating cell migration and adhesion through interaction with other proteins such as Ras GTPase-activating-like protein 1 (IQGAP1) to control actin filaments and microtubules [36, 37]. There is a central β -catenin binding domain interspersed with 15

and 20 amino acid repeats, which acts as a scaffold that recruit and regulate β -catenin activity in the WNT pathway as previously described [36, 37]. The 20 amino acid and armadillo repeats also play a role in the ubiquitination and degradation of β -catenin [36, 47]. The SAMP repeats within the β -catenin binding domain also mediate binding of Axin and GSK3 β to promote phosphorylation of β -catenin [37]. The basic domain, rich in lysine and arginine residues, plays a role in chromosomal segregation and tubulin polymerisation [36, 37]. The EB1 and HDL binding domains are also involved in binding to microtubules for the maintenance of cytoskeleton integrity and regulating chromosomal segregation [36]. There are several NES and NLS involved in shuttling APC between cytoplasm and nucleus, and nuclear APC can regulate β -catenin-mediated transcription through binding to co-repressors via the S/TPXX repeat sequences at the C-terminus [48-51].

1.3.1.3 APC Mutations & Implications in CRC

Due to the role of *APC* as a negative regulator of the WNT signalling, mutations in *APC* are widely implicated in CRCs [33, 37, 50, 52]. *APC* mutations drive tumourigenesis through sustained WNT signalling and proliferation caused by disrupting the β -catenin destruction complex [36-39]. Loss of *APC* function is implicated in approximately 85% of all sporadic and germline CRCs [52, 53]. Most mutations are in the MCR within the exon 15 (Figure 1-6), which contains more than 75% of the coding sequence [37]. Additionally, the WNT pathway can also be deregulated by low frequency gain-of-function (GOF) mutations in the β -catenin-encoding *CTNNB1*, mostly in exon 3 phosphorylation sites [44, 45, 54]. Overall, WNT pathway deregulation is implicated in 93% of CRCs [55].



B)

Type of Mutation	No. of Samples Tested	Percentage*
Substitution (nonsense)	1800	50.32
Substitution (missense)	186	5.20
Substitution (synonymous)	37	1.03
Insertion (inframe)	1	0.03
Insertion (frameshift)	574	16.05
Deletion (inframe)	1	0.03
Deletion (frameshift)	1238	34.61
Complex	14	0.39
Other	140	3.91

Figure 1-7: Distribution of Somatic APC Mutations in the Intestine

A) Base substitutions affecting APC codons. MCR = mutation cluster region (red box). Highest peak at the hotspot codon p.R1450 where the most common alteration is the C>T substitution at c.4348, resulting in a stop codon at p.1450. [56]

B) Distribution of different types of APC alterations. Base substitutions are the most common. Note that percentages (*) were calculated based on the total number of unique samples tested for each type of mutations. [56]

Approximately 60% of sporadic CRCs are driven by APC mutations, and they usually give rise to the most common TA subtype [21, 22]. Somatic alterations are spread throughout the gene but 60% of them are clustered within the MCR (codons 1250-1550), where two hotspots are located at codons 1309 and 1450 (Figure 1-7, [56]) [37, 44-46]. APC mutations drive tumorigenesis in line with Knudson's two-hit hypothesis [57, 58]. Tumour suppressor genes functions to regulate cellular proliferation, and while an initial mutation, usually a point mutation, may cause a loss of function in one allele, affected cells still contain one wild-type (WT) copy to modulate

its tumour suppressing functions [57, 58]. As such, most first hits in tumour suppressor genes are recessive and give rise to a heterozygous state [57]. A second hit of additional loss-of-function mutations, such as genetic alterations or epigenetic inactivation through promoter methylation, fulfil the Knudson's model where 'loss of heterozygosity' (LOH) of a tumour suppressor gene is essential for malignant progression [33, 37, 57-59].

Genetic risk factors are implicated in about 10-20% of germline CRCs, and germline *APC* mutations are implicated in the inherited CRC predisposition syndrome called familial adenomatous polyposis (FAP) [33, 60]. FAP is an autosomal-dominant disease that shows an early onset (10s to 20s) with characteristic development of many colorectal polyps (>100), mostly in descending and sigmoid colon, which inevitably progress into cancerous lesions in later stages of life (over 40s) [60-64]. Two mutational hotspots at codons 1061 and 1309 account for >33% of the total germline mutations identified to date [37, 65, 66].

1.3.2 Mitogen-activated Protein Kinase Signalling Pathway

1.3.2.1 Pathway Overview

The mitogen-activated protein kinase (MAPK) signalling pathway relies on sequential phosphorylation of kinases to transduce downstream physiological effects in response to various signals (e.g. growth factors, hormones, cytokines) at the membrane level [67-69]. The RAS-RAF-MEK-ERK pathway, also known as the extracellular signal regulated kinase (ERK) pathway, is the most understood MAPK pathway in humans (Figure 1-8) [67-69]. Typical pathway activation occurs when a ligand binds to surface receptors called receptor tyrosine kinases (RTKs), a family of growth factor receptors (Figure 1-8) [70].

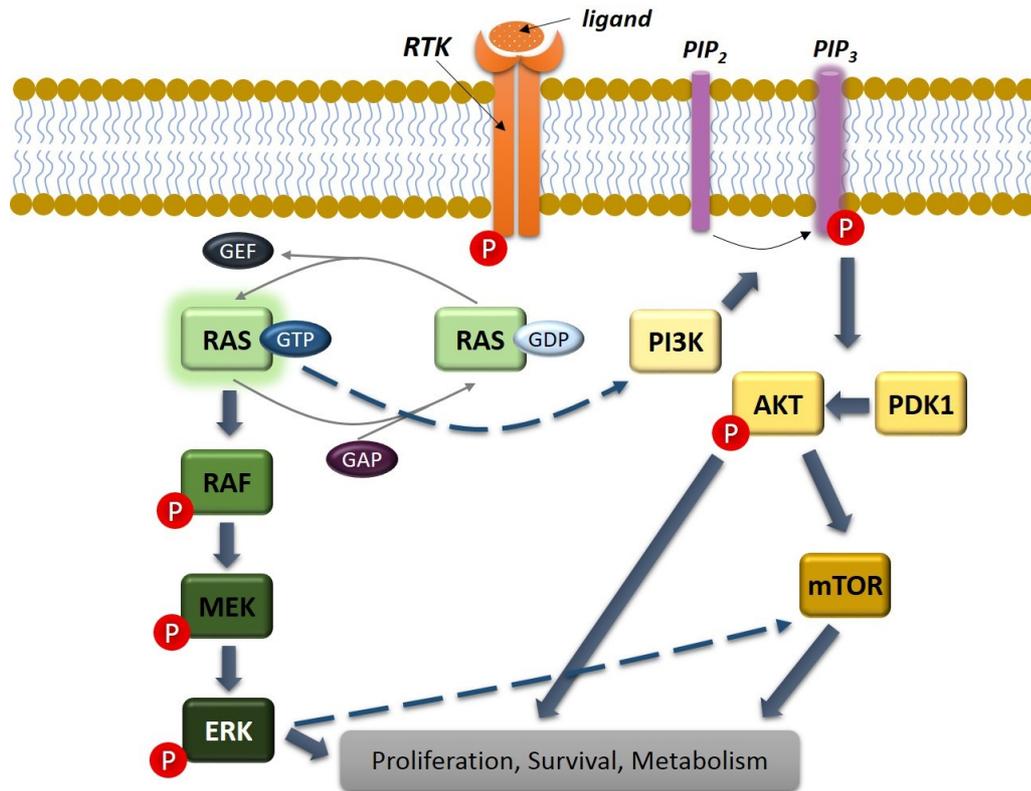


Figure 1-8: ERK & PI3K Signalling Pathways

A simple overview of the ERK and PI3K signalling pathways, and activating pathway cross-talks are indicated by dotted arrows. [61, 69-71]

P = phosphorylated residues, RTK = receptor tyrosine kinase, RAS = kirsten rat sarcoma viral oncogene homolog, GDP = guanosine diphosphate, GTP = guanosine triphosphate, GEF = guanine nucleotide exchange factor, GAP = GTPase activating protein, RAF = rapidly accelerated fibrosarcoma oncogene homolog, MEK = mitogen-activated protein kinase kinase, ERK = extracellular signal regulated kinase, PI3K = phosphatidylinositol-4,5-bisphosphate 3-kinase, PIP₂ = phosphatidylinositol 4,5-bisphosphate, PIP₃ = phosphatidylinositol 3,4,5-trisphosphate, AKT = Ak strain transforming, PDK1 = 3-phosphoinositide-dependent kinase 1, mTOR = mammalian target of rapamycin

For instance, a signal, such as epidermal growth-factor (EGF) binding to its RTK, an epidermal growth factor receptor (EGFR), induces dimerisation of the receptors and activates them via *trans*-autophosphorylation of multiple tyrosine residues (Figure 1-8) [70]. These phosphotyrosine residues recruit adaptor proteins, such as growth factor receptor-bound protein 2 (GRB2) and son-of-sevenless (SOS), via interaction with the src homology 2 (SH2) domain. This complex then interacts with the inactive GDP-bound kirsten rat sarcoma viral oncogene homolog (RAS), resulting in GDP-GTP exchange on RAS that is catalysed by guanine nucleotide exchange factors (GEFs) [70, 71]. Active RAS

interacts with V-Raf murine sarcoma viral oncogene homolog B (BRAF) via the N-term RAS-binding domain (RBD), triggering dimerization and auto-phosphorylation of key residues Thr599 and Ser602 in the activation segment of BRAF (Figure 1-10) [67, 72]. Activated BRAF, in turn, phosphorylates mitogen-activated protein kinases (MEK) at Ser218 and Ser222 in the activation loop and activates them [67]. Activated MEK then phosphorylates and activates ERK [67]. This sequential activation of RAF-MEK-ERK is supported by scaffolding proteins such as kinase suppressor of RAS-1 (KSR1) [67, 73]. Activated ERK has many cytoplasmic and nuclear targets, depending on the cell type (Figure 1-8) [67, 74].

The ERK pathway also exhibits cross-talking with the phosphatidylinositol-4,5-bisphosphate 3-kinase (PI3K) pathway, which also promotes cell cycle progression and cell survival (Figure 1-8) [75, 76]. GTP-bound RAS can directly bind, via its effector domain, and activate p110 α catalytic subunit of PI3K [76-78]. This results in the phosphorylation of membrane-bound phosphatidylinositol 4,5-bisphosphate (PIP₂) into phosphatidylinositol 3,4,5-trisphosphate (PIP₃), which then recruits Akt strain transforming/ protein kinase B (AKT/PKB) to the membrane where it is activated through phosphorylation by 3-phosphoinositide-dependent kinase 1 (PDK1) (Figure 1-8) [75, 76]. Active AKT/PKB has several downstream targets, including activation of the mammalian target of rapamycin (mTOR) serine/threonine kinase to promote protein synthesis, nucleotide biosynthesis and lipid metabolism amongst others [75, 76]. The small GTPase Rheb in the form of Rheb-GTP is also an upstream activator of mTOR, and active tuberous sclerosis complex (TSC) inhibits mTOR activity by promoting Rheb-GTP to Rheb-GDP hydrolysis. Sustained ERK signalling can also lead to pathway cross-activation at the level of mTOR, especially mTOR complex 1 (mTORC1), through ERK-mediated inhibitory phosphorylation of the TSC [76, 79]. PI3K pathway is also activated by various non-RAS-dependent pathways [80].

Additionally, there is a certain degree of convergence on downstream effectors between the ERK and PI3K pathways to induce pro-survival and proliferative consequences. The family of forkhead box O (FOXO) transcription factors regulate multiple tumour suppressor genes such as pro-apoptotic Bcl-2 interacting mediator of cell death (*BIM*) and Fas ligand (*FasL*) [81, 82]. Both AKT and ERK phosphorylates FOXO3a, leading to the murine double minute 2 (MDM2)-mediated degradation of

FOXO3a, which in turn promotes survival and tumourigenesis [81, 82]. Phosphorylation of FOXO transcription factors by AKT or ERK also leads to the cytosolic sequestration of FOXOs by association with phosphoserine-binding 14-3-3 proteins, which preclude the transcriptional activation of pro-apoptotic genes [81, 82]. Activated ERK also leads to transcriptional upregulation of cell cycle protein *CCND1* via MYC transcription factors, which results in the expression of cyclin D1 that regulates G1-S progression [67, 83].

1.3.2.2 RAS Oncogenes

RAS oncogenes encode membrane-associated small GTPases that are involved in various signal transduction pathways, including the ERK pathway as previously described [80, 84]. The *RAS* family contains 3 genes – Kirsten rat sarcoma viral oncogene homolog (*KRAS*), neuroblastoma RAS viral oncogene homolog (*NRAS*) and Harvey rat sarcoma viral oncogene homolog (*HRAS*) – with 82-90% sequence identity between them [80]. The *KRAS* gene contains 6 exons and is expressed as two functional splice variants, *KRAS4A* and *KRAS4B*, which are 188 and 189 amino acids long respectively; the *KRAS4B* is the dominant form in human cells [84, 85]. The low intrinsic GTPase activity of *RAS* proteins is regulated by *RAS*-binding proteins such as GEFs and GTPase activating proteins (GAPs) (Figure 1-8 & Figure 1-9) [80, 84]. GEFs facilitate binding of GTP to *RAS* and inducing a conformational change into its active state, whereas GAPs inactivates *RAS* activity by promoting hydrolysis of GTP to GDP [84]. *KRAS* interacts with GEF and GAP via the Switch I and II domains, and changes between active and inactive conformations during GDP-GTP cycling (Figure 1-9) [80, 86].

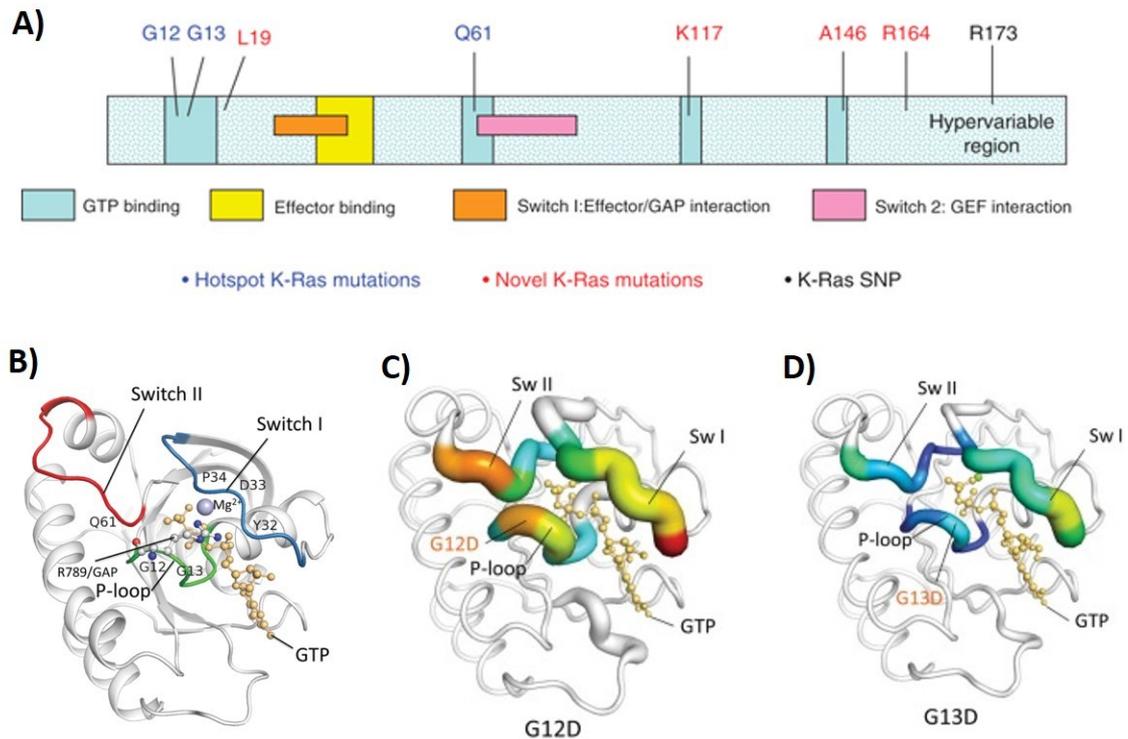


Figure 1-9: Structure of Wild-type & Mutant *KRAS*

A) Functional domains of *KRAS* and mutational hotspots. [87-88]

B) Computer-simulated structure of the GTP-bound WT *KRAS*. The hotspot codons G12 and G13 constitute the P-loop (phosphate-binding loop). [87-88]

C) The size of coloured tubes shows positional fluctuations of atoms: blue is lowest and red is highest. Aspartic acid (D) introduces steric hindrance and prevents GAP binding. [87-88]

D) Computer-simulated structure of G13D mutant; less atomic fluctuation was observed than the G12D mutant. [87-88]

RAS oncogenes, especially *KRAS*, are mutated in many cancers. *KRAS* mutations are implicated in approximately 45% of CRCs, whereas *NRAS* mutations are implicated at a lower frequency of 7.5% in CRCs [80]. There are two hotspot codons in exon 1 of *KRAS*: more than 70% of mutations affect codon 12, comprising of 6 possible missense mutations at bases c.34G and c.35G, and 20% affect codon 13 [80, 85]. These mutations introduce amino acids with larger side chains (e.g. aspartic acid, D) in the place of glycine (G) in the WT *KRAS*, causing steric hindrance that reduces *KRAS* binding affinity to GAP [80, 87]. This leads to reduced GTPase activity in a MU *KRAS* and resulting in a constitutively active, GTP-bound *RAS* [80]. Consequently, the signalling pathway remains active without the need for ligand-receptor interaction at the membrane level

[84, 86-88]. *KRAS* mutant cancers are clinically relevant as they are resistance to anti-EGFR therapies [84-86]. Mutant *RAS*-driven tumours are also linked to deregulations in the PI3K pathway, and these cases often present with overlapping *KRAS* and phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha (*PIK3CA*) mutations [80, 89]. *PIK3CA* encodes the p110 α catalytic subunit of class I PI3K, and *PIK3CA* mutations are implicated at 17% frequency in CRCs [56, 80].

1.3.2.3 *RAF* Oncogenes

Similarly, the *RAF* family of oncogenes encode serine/threonine kinases that are involved in MAPK signalling pathways [68, 69, 73]. There are 3 *RAF* proteins in humans, with the highest basal kinase activity found in BRAF that is predominantly implicated in cancers [68, 90]. BRAF has conserved regions (CR) 1 and 2, which encompass a RAS-binding domain (RBD) important for binding to RAS-GTP, as well as a kinase domain CR3, containing ATP and substrate binding sites (Figure 1-10) [73, 91]. Interaction with GTP-RAS triggers auto-phosphorylation of Thr599 and Ser602 in the CR3 activation segment for the full active conformation (Figure 1-10) [67, 68, 72, 73].

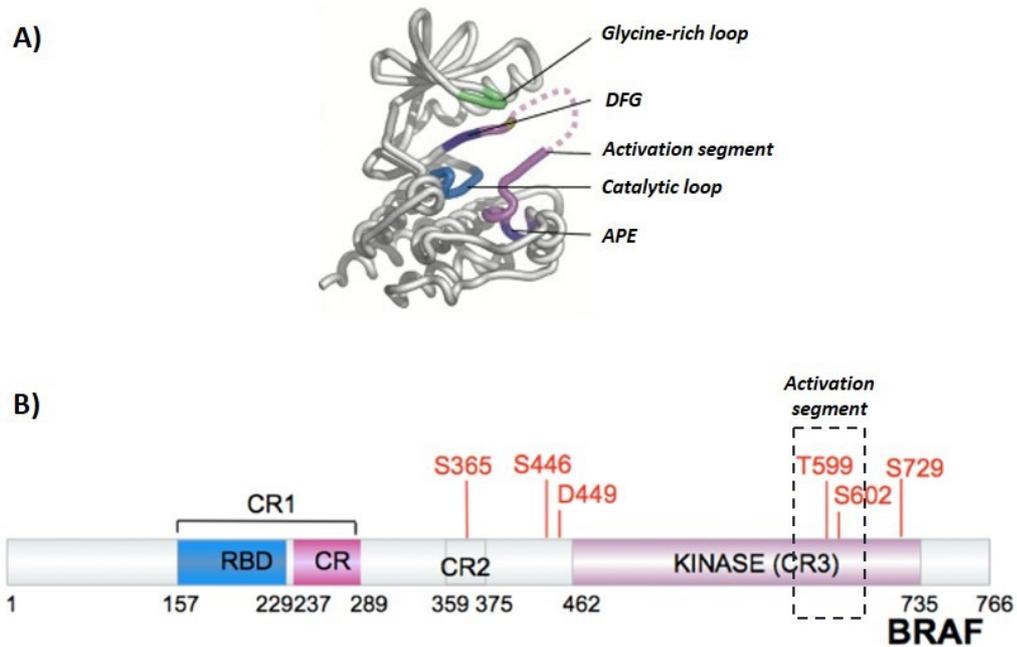
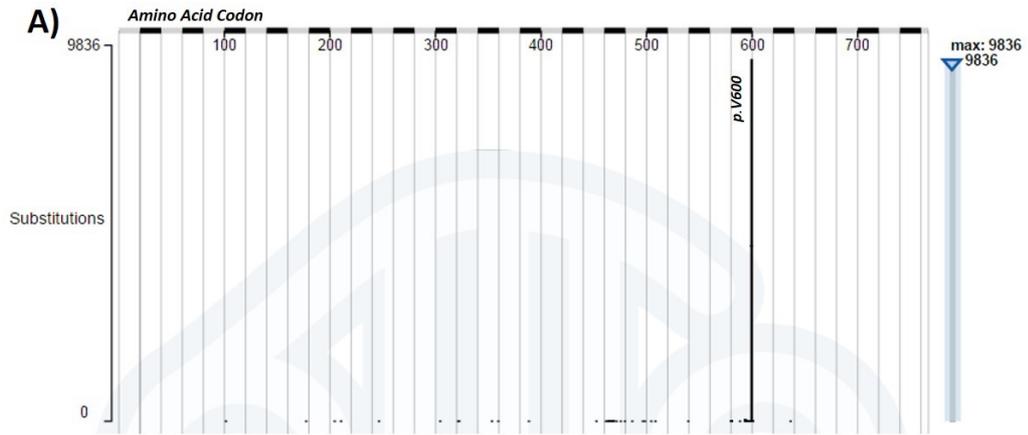


Figure 1-10: Structure & Function of *BRAF*

A) Crystal structure of BRAF with key structural regions involved in its catalytic activity. DFG = Asp-Phe-Gly motif, APE = Ala-Pro-Glu motif. [69, 92]

B) Functional domains and key residues. RBD = RAS-binding domain. CR = conserved regions. [69, 92]

In the absence of signalling, WT BRAF remains in an inactive conformation through interaction between the glycine-rich loop (464-469) and the activation segment (594-600) (Figure 1-10) [69, 92]. Additionally, Ser365 and Ser729 are phosphorylated and bound to the 14-3-3 protein, causing the CR3 domain to fold over to CR2 [91, 93]. In cancers, the predominant *BRAF* oncogenic mutation is the transversion of T to A at base c.1799, resulting in the switch from valine (V) to glutamic acid (E) at codon 600 of the BRAF protein ($BRAF^{V600E}$), located in the activation segment in the CR3 domain [86, 90]. This introduces a negative charge in the middle of the activation segment that mimics the active conformation (Figure 1-10), resulting in a constitutively active kinase with a 500-fold increase in activity compared to the WT counterpart, which leads to sustained, growth factor-independent activation of the downstream MAPK signalling pathway [68, 69, 73, 91]. This hotspot mutation (c.1799T>A/p.V600E) contributes to more than 95% of the BRAF mutations identified in CRC, and is implicated in 10-20% of sporadic CRC cases (Figure 1-11, [56])[90, 91, 94, 95].



B)

<i>Type of Mutation</i>	<i>No. of Samples Tested</i>	<i>Percentage*</i>
Substitution (nonsense)	2	0.02
Substitution (missense)	9917	99.86
Substitution (synonymous)	7	0.07
Insertion (inframe)	2	0.02
Insertion (frameshift)	2	0.02
Deletion (inframe)	0	0.00
Deletion (frameshift)	3	0.03
Complex	2	0.02
Other	0	0.00

Figure 1-11: Distribution of Somatic *BRAF* Mutations in the Intestine

A) Graph showing *BRAF* codons affected by base substitutions in intestinal tumours. There is a mutational hotspot at p.V600. [56]

B) Distribution of different alterations affecting *BRAF*. An overwhelming percentage of mutations are missense base substitutions. [56]

Overall, mutually exclusive mutations in *BRAF* and *KRAS* oncogenes lead to deregulation of MAPK signalling pathways, which are frequently observed as early events in CRCs [23, 44, 86, 92]. Moreover, deep sequencing analysis of matched primary and metastatic colorectal lesions showed that these hotspot genetic alternations are absolutely concordant over the course of tumour progression, indicating that these hotspot mutations serve as good targets for mutation detection and genetic profiling of CRCs [96].

1.3.3 Pathways of Genomic Instability

1.3.3.1 CpG Island Methylator Phenotype

CpG island methylator phenotype (CIMP) refers to methylation of cytosine residues in CpG island promoter regions and subsequent hetero-chromatinisation and transcriptional inactivation of genes [97]. CIMP is implicated in 20-30% of CRCs through aberrant inactivation of tumour suppressor genes, and it is one mechanism by which LOH of *APC* occur in the Knudson's two-hit model [58, 97]. Other prominent targets silenced by CIMP include the cyclin-dependent kinase Inhibitor 2A (*CDKN2A*) tumour suppressor gene, which encodes an anti-proliferative p16 protein, and human MutL homolog 1 (*hMLH1*) gene involved in the DNA mismatch repair (MMR) system [97-100]. There are 3 type of CIMP – CIMP-H (high), CIMP-L (low) and CIMP-Neg (no CIMP) – based on how many markers are positive for methylation [32, 99-101]. Currently, a panel of 4-8 CpG islands (*CACNA1G*, *CDKN2A*, *CRABP1*, *IGF2*, *MLH1*, *NEUROG1*, *RUNX3*, and *SOCS1*) is used to assess the CIMP status of tumours, where methylation at ≥ 5 out of 8 makers represents the CIMP-H status. However, it has been acknowledged that these markers are not ideal for accurately identifying CIMP-L or CIMP-Neg phenotypes [100, 102]. CIMP-H lesions frequently has *BRAF* mutations, and they show a predilection for proximal colon, has a poorly differentiated histology and poor prognosis [99, 100, 103].

1.3.3.2 Microsatellite Instability

Microsatellites (MS) are between 1 to 9 units of tandem nucleotide repeats, where each unit consists of 1-4 nucleotides, which exists within the genomic repetitive regions [104]. MS regions are susceptible to replicative error caused by stalling of DNA polymerase due to strand slippage, and functioning MMR system corrects this error to preserve genomic integrity [100, 105, 106]. Defective MMR system is caused by either genetic mutations affecting MMR genes, such as *hMLH1* and human mutS homolog 2 (*hMSH2*), or epigenetic silencing through CIMP, resulting in microsatellite instability (MSI) where MS vary in length [100, 105-108]. Affected cells are said to have the replicative error (RER) phenotype. MSI is classified into 3 types – MSI-H (high), MSI-L (low) and MSS (stable) [35, 100]. The Bethesda panel is one of many that can be used to assess the MSI status, and it consists of two mononucleotide MS loci (*BAT25* and *BAT26*)

and three dinucleotide MS loci (*D2S123*, *D5S346* and *D17S250*) that are used as references to check for variations in MS sequences [35, 100, 106]. Size alteration in ≥ 2 out of 5 markers is considered as MSI-H, whereas MSI-L shows alteration in one marker [35, 100, 106]. MSI Multiplex System is an alternative test that uses five mononucleotide MS loci (*BAT25*, *BAT26*, *NR21*, *NR24*, and *MONO27*), which has been reported to be more sensitive and specific [106, 109].

MSI contributes to a 100-fold increase in mutation rates and is present in approximately 15% of CRCs, of which 12% are sporadic cases [35, 97, 100]. MSI-H CRCs also frequently associate with CIMP-H and *BRAF* mutations, and show predilection for proximal colon as well as exhibiting histopathological features such as mucinous differentiation, high levels of tumour-infiltrating lymphocytes and tumour [35, 100, 110]. As for MSI-L and MSS phenotypes, their classification as distinct molecular phenotypes in CRCs has been argued as most tumours show some degree of MSI [35, 100, 111-113]. However, MSI-L tumours have been reported to have shorter survival compared to MSS cases at Dukes' C staging, suggesting their clinical relevance [35, 100, 111-113]. MSI is also implicated in more than 95% of hereditary non-polyposis colorectal cancer (HNPCC), also known as Lynch syndrome, which is a CRC predisposition syndrome that accounts for ~3% of all CRC cases [60, 97, 107]. It has a late onset (>40 years) and shows a predilection for proximal colon with histological characteristics such as poor differentiation and hyper-mucinous epithelium [114-118].

1.3.3.3 Chromosomal Instability

Chromosomal instability (CIN) is a common pathway in CRCs and it is observed in approximately 70% of sporadic cases [45]. CIN is characterised by widespread LOH and chromosomal abnormalities, such as aneuploidy [97]. CIN positivity can be measured in numerous ways, and the latest method is to use comparative genomic hybridisation (CGH) arrays that can identify amplifications and deletions with a high resolution [45]. Mutations in *APC* has 70% prevalence in CIN-positive tumours, and they are pivotal in early tumorigenesis as they drive aberrant WNT signalling as previously described [45, 97, 105]. Mutations in the mitotic checkpoint sensors, such as mitotic arrest deficient (*MAD*) and budding uninhibited by benzimidazoles (*BUB*), also promotes

CIN through defective chromosomal segregation [45, 97, 105]. LOH in tumour suppressor genes like tumour protein 53 (*TP53*) also plays a role in the CIN pathway through the loss of cell cycle control; *TP53* LOH also has an add-on effect on chromosomal segregation via downregulation of the checkpoint gene *BUBR1* [45, 119]. There is an approximately 70% concurrence between LOH in *TP53* and chromosome 18q, which contains tumour suppressors such as Sma- and mad-related proteins (*SMAD*) 2 and 4, and 18q loss has been suggested to feed into the CIN pathway through deregulation of the transforming growth factor- β (TGF- β) pathway that controls proliferation and apoptosis [45, 97, 105]. It has yet to be elucidated whether CIN is a cause or a consequence of aforementioned genetic and genomic alterations, but *APC* mutations have been proposed as initiating events for CIN due to their role in WNT signalling as well as cytoskeletal regulation (Figure 1-6) [45, 120, 121].

1.3.3.4 CRC Classification based on Genomic Alterations

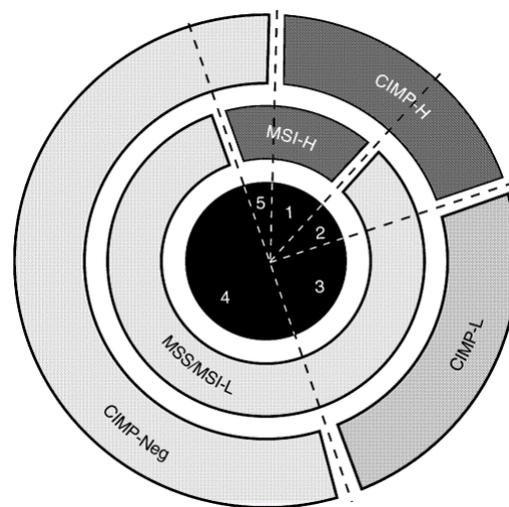


Figure 1-12: Molecular Classification for CRC Subtypes

Jass (2007) proposed 5 molecular subtypes of CRCs based on MSI and CIMP profiles in conjunction with the mutation profiles in prevalent genes implicated in CRC (e.g. *APC*, *BRAF*, *KRAS*). CIMP = CpG island methylator phenotype, MSS = microsatellite stable, MSI = microsatellite instable, L= low, H = high. [111]

CRCs can be classified into subtypes based on their CIMP and MSI profiles (Figure 1-12) [100, 111]. MSI-L/MSS and/or CIMP-L/CIMP-Neg CRCs (type 3 and 4) are the most frequent, contributing to 75-80% of all CRCs – this group is frequently associated with

APC, *KRAS* and *TP53* mutations, show predilection for the left colon [100, 111]. The MSI-H/CIMP-H tumours contribute to about 10% of cases (type 1), and these frequently show *BRAF* mutations, poor differentiation, lymphocytic infiltration, mucinous, serrated histology and have good prognosis [100, 111]. The MSI-H/CIMP-L or Neg tumours (type 5) contribute to approximately 5% of all adenoma cases and these may present with mutations in *APC* and *KRAS*, lymphocytic infiltration and mucinous differentiation [100, 111]. As for the MSI-L/MSS; CIMP-H tumours (type 2), they contribute to 5-10% of all CRCs with the predilection for the right colon and are associated with *BRAF* mutations; they show a mucinous, serrated morphology, lymphocytic infiltration and have poor prognosis [100, 111].

1.3.4 Genetics, Genomics & CRC Pathways

Colorectal tumourigenesis from a benign stage to the development of carcinomas is accompanied by accumulation of distinct genetic and genomic alterations (Figure 1-13) [122, 123]. The Cancer Genome Atlas (TCGA) project carried out whole genome sequencing and identified that about 84% of sporadic CRC cases are non-hypermuted with frequently alterations in these eight genes: *APC*, *TP53*, *KRAS*, *PIK3CA*, *FBXW7*, *SMAD4*, *TCF7L2* and *NRAS* [124]. As for the hypermutated sporadic cases, about 75% of them are associated with MSI brought about by CIMP and the silencing of *MLH1* gene and the remaining 25% bear mutations in DNA repair gene DNA polymerase epsilon (*POLE*) [124]. Concurrently, genomic aberrations accompany accumulation of somatic mutations during tumourigenesis. For instance, hypomethylation of interspersed and tandem repeat elements along the genome is an early event in CRC and promotes chromosomal rearrangements and contributes to tumourigenesis via CIN [123, 125]. Another early event is the gain or amplification of chromosome 20q observed in over half of sporadic cases and is implicated in deregulation of ERK and TP53 pathways as well as affecting transcriptional activation by MYC [123, 126]. Loss of chromosomal arms 17q and 18q are also common in over half of the sporadic CRC cases, which contributes to loss-of-heterozygosity in tumour suppressor genes *TP53* (chr 17q) and *SMAD4* (chr 18q) [124]. During late stages of tumourigenesis, loss of chromosome arm 8q affects the *MYC* transcription factor that is a downstream effector of WNT, ERK and PI3K pathways [124]. Additionally, recurrent

gene fusions of R-spondin small proteins, *RSPO2* and *RSPO3*, whose function is to potentiate WNT signalling via acting on the LGR5 receptors have been identified in 10% of sporadic CRCs [127]. These gene fusions are associated the MSS subtype and bore mutations in either *KRAS* or *BRAF* but found to be mutually exclusive with *APC* mutations, suggesting that they also promote colorectal tumourigenesis via the WNT pathway [127].

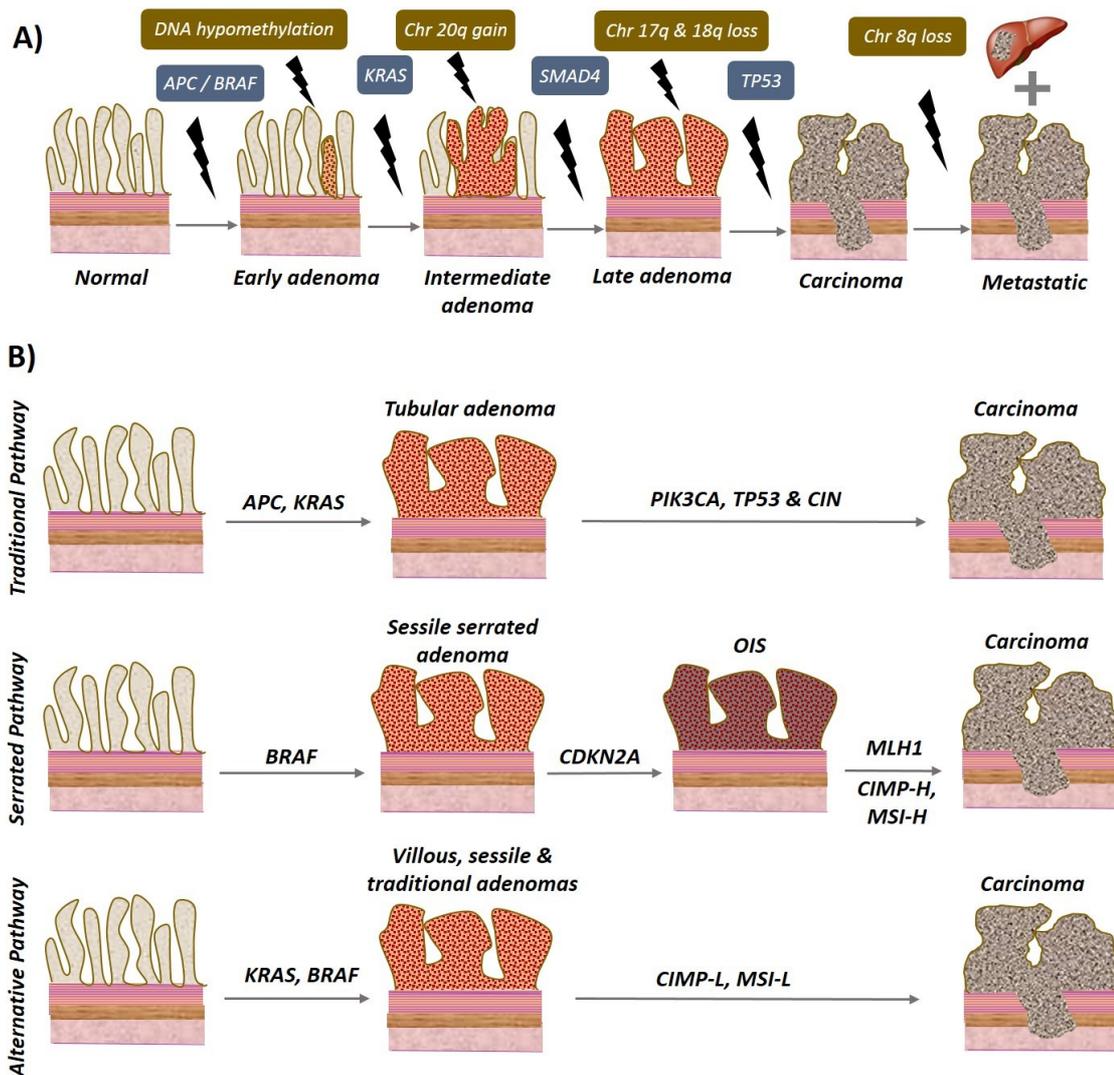


Figure 1-13: Genetic & Genomic Alterations in Colorectal Tumourigenesis

A) A modified Vogelgram highlighting the frequently implicated genetic and genomic alterations that accumulate over the course of colorectal tumourigenesis. [123]

B) Three proposed pathways of colorectal tumourigenesis: the traditional pathway, the serrated pathway with a possible oncogene-induced senescence (OIS) stage and the alternative pathway. [35]

Based on amalgamation of genetic and genomic profiles, colorectal tumourigenesis can also be divided into three pathways: traditional, serrated and alternative pathways (Figure 1-13) [35]. The traditional pathway is the most common and present in 50-70% of sporadic CRCs [35, 128]. Lesions with TA histology are characteristic of this pathway and they are usually driven by *APC* mutations with additional aberrations in *KRAS*, *PIK3CA* and *TP53* genes [35, 128]. This pathway is also associated with CIN and poor prognosis [98]. The serrated pathway accounts for 10-20% of sporadic cases, and is associated with *BRAF* mutations and the CIMP-H/MSI-H phenotypes [35, 111, 128]. Serrated adenomas may also exhibit oncogene-induced senescence (OIS) through activation of cyclin-dependent kinase inhibitors (CKIs) such as *CDKN2A*, and OIS exit requires additional hits such as CIMP-induced LOH of various tumour suppressor genes (e.g. *MLH1*) [95, 129, 130]. As for the alternative pathway, it is heterogeneous and mostly initiated by mutation in *KRAS*, and, to a lesser extent, *BRAF*, and this pathway is commonly associated with SSA and TSA histological subtypes [128]. CIMP-L and MSI-L are the characteristic genomic alterations observed in this subtype of lesions during malignant progression, and this pathway is usually associated with the worst prognostic outlook with poor response to chemotherapy [35, 128]. Nevertheless, these pathways are not exclusive and overlapping of genotypes can be observed occasionally [35].

1.4 Tumour Heterogeneity in CRC

Tumour heterogeneity refers to evolution of tumour clones with differing genetic and genomic profiles over the course of tumour progression [131-133]. There are different types of heterogeneity. Inter-patient heterogeneity (IPH) refers to presentation of variable tumour genetic profiles between patients with the same histological type of primary tumours, whereas intra-tumour heterogeneity (ITH) refers to genetic variations within the same lesion in one patient [131, 134, 135]. ITH can be further categorised into spatial (i.e. variations between tumour clones located in different anatomical parts) and temporal (i.e. variations between primary clone and successive sub-clones) heterogeneity [131, 132, 134, 135].

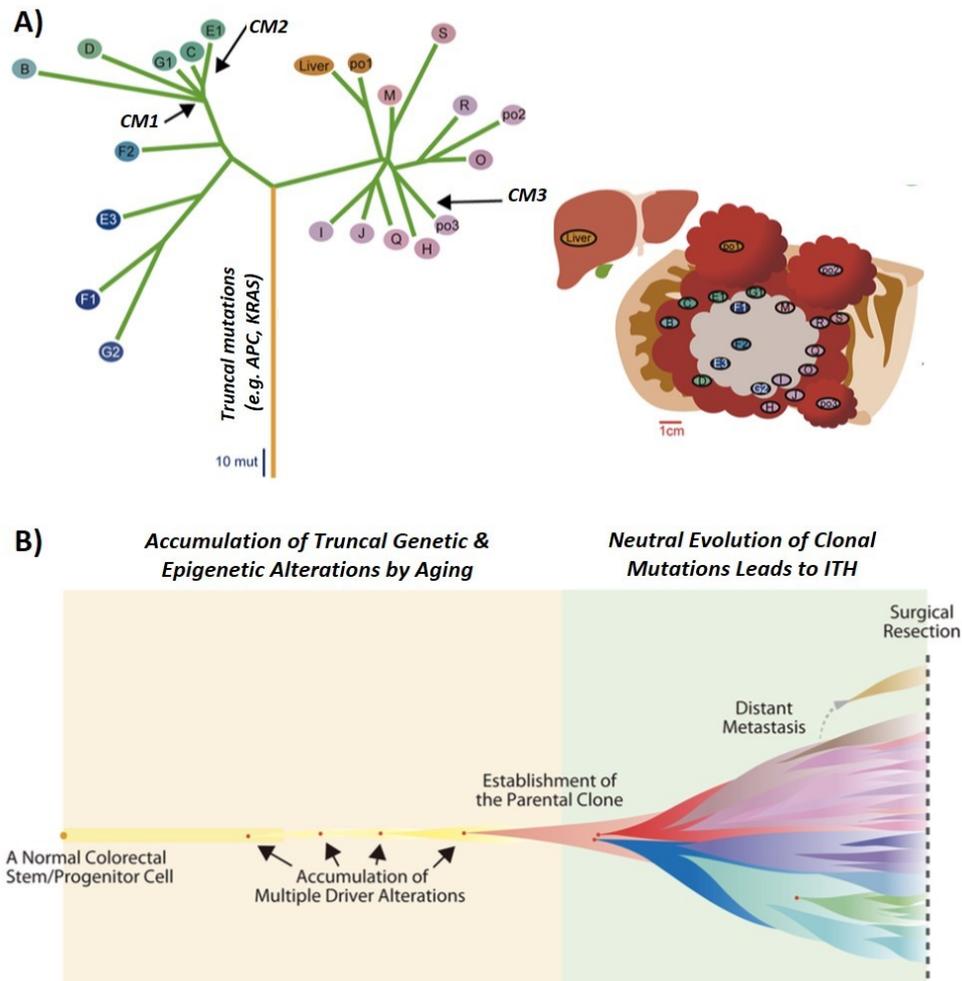


Figure 1-14: Branching Evolution of Intra-tumour Heterogeneity in CRC

- A)** An example case of a clonal evolutionary tree obtained from multi-regional analysis (n=21) of a stage IV CRC patient. Orange line represents multiple truncal mutations (e.g. *APC*, *KRAS*), and green represents multiple clonal mutations (CM). Coloured 'leaves' correspond to anatomical regions, and colour similarity shows similarity in mutation profiles. [136]
- B)** Proposed model of clonal evolution in CRC. Multiple truncal mutations accumulate as a consequence of aging, establishing parental clone population. Heterogeneous clonal mutations are acquired over time. [136]

A heterogeneous population of cancerous cells contains varying frequencies of truncal and clonal mutations (Figure 1-14) [134, 136]. Truncal mutations occur in early stages as cancer-initiating events and they are present in all parental tumour clones, which then branch into multiple sub-clones each containing various combination of clonal mutations (Figure 1-14) [134, 136]. This phenomenon is known as branching clonal evolution of tumour cells and has been observed in both early and late colorectal lesions [134-137]. Uchi's group performed multi-regional exome sequencing of spatially

separated advanced tumour samples from 9 CRC patients, and reported high genetic ITH in all cases (Figure 1-14) [134]. In CRCs, early driver mutations are usually acquired as a result of aging, whereas the clonal mutations may arise as a consequence of neutral evolution where they do not initially confer survival advantage [134, 136].

ITH has clinical relevance. A single tissue biopsy may not be representative of the whole tumour, and it could negatively impact treatment outcome if longitudinal treatment decisions are centred on the biopsied and archived primary specimens [131-133, 135]. A small, clonal population of cancerous cells may evade therapies targeted at truncal mutations, which can eventually lead to emergence of resistance clones in different evolutionary branches [132, 134]. One example is the positive selection of clones with *KRAS* mutations that confer treatment resistance to EGFR-targeted therapies (e.g. gefitinib) [138-141]. Treatment with anti-EGFR therapies in patients who are identified as low *KRAS* mutant phenotype (<10% frequency) will inadvertently lead to selection pressure on the clonal *KRAS* mutant population, resulting in the emergence of resistance clones [141]. However, in the case of solid tumours, consequential biopsies are not routine in the clinics to track the recurrence of localised primary tumour nor emergence of resistance clones [131]. Identification of evolving mutational branches as the cancer progresses can improve prognosis by adjusting treatment regimens: for instance, a combination therapy with MEK and EGFR inhibitors would be more efficacious in patients with secondary *KRAS* mutations than an EGFR-targeted therapy [138-141]. Advances in biomolecular diagnostics herald the era of personalised medicine, where platforms like next-generation sequencing (NGS) platforms (e.g. Illumina HiSeq series) can be used to capture comprehensive clonal evolution of tumours to dictate treatment decisions [131, 132, 134, 135, 141, 142].

1.5 Genetically Engineered Mouse Models of CRC

Many genetically engineered mouse models (GEMMs) have been developed to recapitulate and study CRC in various stages of progression [143]. For this project, the GEMM of interest is the *Lgr5-EGFP-IRES-creER^{T2+/0};Apc^{fl/fl}* model, thereafter referred to as *Cre-Apc^{fl/fl}*, where tumourigenesis is driven by loss of Apc function in the Lgr5-expressing stem cells [9, 143]. This mouse model can either have the Cre allele positive

(*Cre^{+ve}-Apc^{fl/fl}*) genotype that expresses the Cre protein or the negative (*Cre^{-ve}-Apc^{fl/fl}*) genotype that does not express it [8, 143, 144].

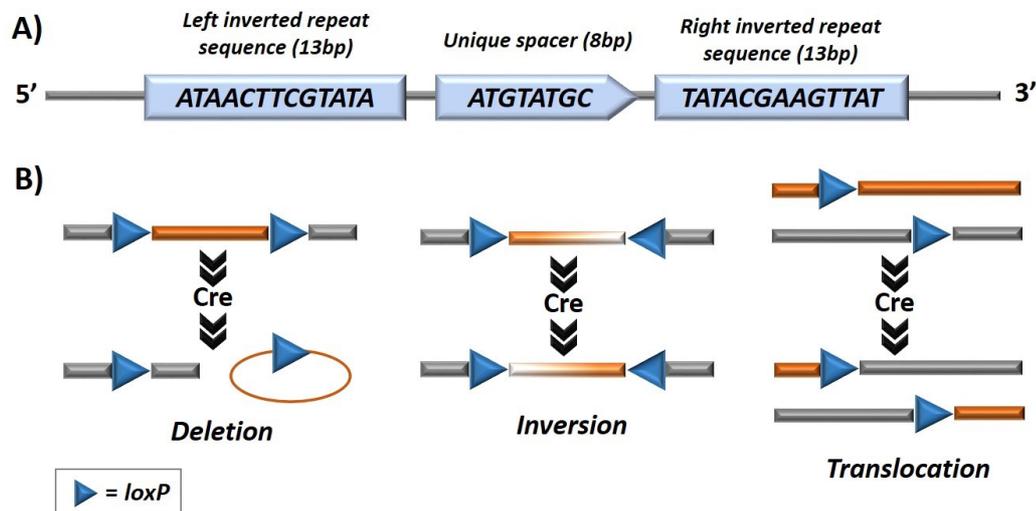


Figure 1-15: Gene Manipulation using the Cre-lox System

A) A consensus *loxP* sequence of 34 bp, where a unique 8bp spacer sequence is flanked by two 13bp palindromic repeats. The asymmetry in the spacer gives *loxP* directionality. [127]

B) Three possible outcomes of a Cre-*loxP* system depending on the directionality of the inserted *loxP* sequences. [126-127]

This *Cre-Apc^{fl/fl}* mouse model uses the Cre-lox technology (Figure 1-15) to spatially and temporally control the disease onset [9, 143]. The 34bp *loxP* (Locus of X(cross)-over in P1) is a component of the P1 bacteriophage, and it consists of two 13bp palindromic sequences separated by an 8-bp asymmetrical spacer (Figure 1-15-A) [145-147]. Cre (cyclization recombination) recombinase is a P1 bacteriophage enzyme that mediates site-specific recombination of inserted *loxP* sites [145-147]. Cre recognises the two 13bp repeat sequences of *loxP* with a high affinity and forms a dimer, which then forms a functioning tetramer when two *loxP* sites are aligned in parallel, resulting in reciprocal recombination of the *loxP* spacers [145-147]. The Cre-lox system can be used for sequence deletion, inversion or translocation depending on the directionality of inserted *loxP* sites (Figure 1-15) [145-147]. The *Cre* gene used in the *Cre-Apc^{fl/fl}* model has been modified by fusing the endogenous Cre to the human estrogen receptor (ER) ligand binding domain that contains three specific non-synonymous mutations G400V/M543A/L544A, producing the fusion *CreER^{T2}* gene [148, 149]. This *CreER^{T2}*

fusion protein (simply referred to as Cre thereafter) only becomes activated in the presence of 4-OH tamoxifen, a metabolite of tamoxifen [150-152]. Upon tamoxifen administration, cytoplasmic Cre localises to the nucleus where it mediates recombination of *loxP*-modified genes [145, 153].

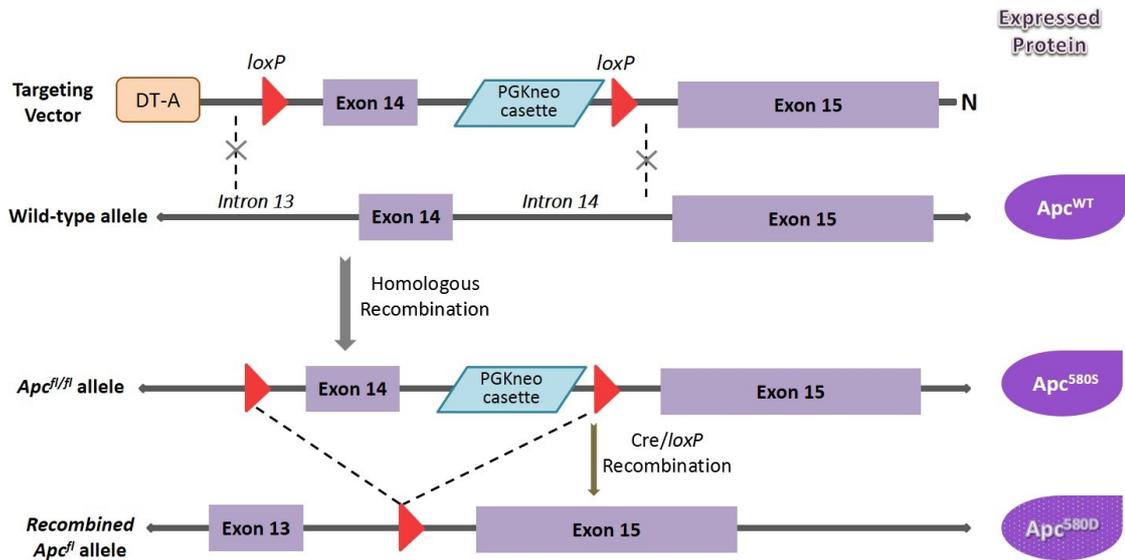


Figure 1-16: Site-specific Sequence Deletion in the *Apc^{fl/fl}* Allele

Homologous recombination between the targeting vector and the wild-type *Apc* allele generates the *Apc^{fl/fl}* allele, which expresses *Apc^{580S}* protein with a silent frameshift mutation at codon 580 with intact tumour suppressing functions of *Apc^{WT}*. In the presence of Cre, the exon 14 is excised out, resulting in a frame-shift mutation that stops at codon 580 – this expresses the truncated *Apc^{580D}* protein. DT-A is diphtheria toxin A used for gene selection during homologous recombination. [135]

Shibata's group developed this *Cre-Apc^{fl/fl}* model where an *Apc* targeting vector, in which the exon 14 of endogenous *Apc* is flanked by two *loxP* sequences in the same orientation, is inserted into introns 13 and 14 to generate a floxed (flanked by l*oxP* sites) *Apc* allele (*Apc^{fl/fl}*) [144]. Upon expression of Cre recombinase, exon 14 is deleted in *Apc^{fl/fl}* (Figure 1-16). It has been reported that *Apc^{fl/fl}* allele is hypomorphic, expressing ~30% less compared to the *Apc^{WT}* allele, suggesting that this *loxP* modification has a negative effect on gene expression [144]. Barker's group further modified this model to restrict the Cre expression to Lgr5-expressing stem cells by generating a knock-in allele where the first exon of *Lgr5* gene is replaced by the *EGFP-IRES-creER^{T2}* cassette [9]. The presence of an internal ribosomal entry sequence (IRES) allows for the simultaneous expression of the enhanced green fluorescent protein (EGFP) and Cre [9]. Inducible

models of *Cre-Apc^{fl/fl}* are heterozygous for the *Lgr5-EGFP-IRES-creER^{T2}* allele as the loss of *Lgr5* function causes lethality in newborn mice [9].

In the *Cre-Apc^{fl/fl}* model, in the absence of *Cre*, the targeted *Apc^{fl/fl}* allele expresses the *Apc^{580S}* protein with a silent frameshift mutation at codon 580 that is functionally synonymous to that of *Apc^{WT}* protein and showed no tumourigenic phenotype [144]. Upon tamoxifen injection, nuclear localisation of *Cre* occurs in the *Lgr5*-expressing CBC cells, resulting in the *Cre*-dependent exon 14 deletion to produce a recombined *Apc^{fl}* allele (Figure 1-16). This expresses a truncated *Apc^{580D}* protein with a frame-shift mutation at codon 580, and loss of *Apc* function initiates tumourigenesis through deregulated WNT signalling as previously described (section 1.3.1) [144]. Barker's group showed that *Cre-Apc^{fl/fl}* mice developed macroscopic adenomas within 3-5 weeks after tamoxifen administration, supporting the validity of this model for studies into the early stage colorectal tumourigenesis [7, 8, 144, 154, 155].

While the focus of this project is on understanding early tumourigenesis of CRC, there are several other GEMMs available that can emulate a more invasive and aggressive form of colorectal tumourigenesis [143]. Invasive models of GEMM require double mutations – one in the driver gene such as *Apc* to initiate adenoma development and to reduce latency period, and another in genes such as *Kras*, *Pten* or *Tp53* to promote invasiveness [143]. One such model is the *AhCre Apc^{fl/+}; Kras^{LSL-G12V/+}* that shows low frequency (17%) invasive carcinoma into smooth muscle when *Kras^{G12V}* is expressed following *Apc* deletion [156]. *AhCre Apc^{fl/+}; Kras^{LSL-G12D/+}* is another metastatic model that can faithfully recapitulate the adenoma-carcinoma-metastasis sequence of human CRCs, but the drawback of this model is the need for surgery to deliver the adeno-*Cre* for low-level stochastic deletion of *Apc* in the distal colon [143, 157]. Another excellent murine model of CRC is *AhCre Apc^{fl/+}; p53^{R172H/+}* that shows 100% invasive phenotype, which mirrors invasiveness associated with *TP53^{R175H}* mutation in human CRCs [143, 158]. One important consideration for invasive GEMMs is to ensure that the *Cre* induction is titrated so that the initial tumourigenic drive by *Apc* deletion is not too strong to cause accelerated adenoma development, which many inadvertently result in reaching ethical end-point before additional mutations can cause invasive lesions [143].

1.6 Methods to Reduce CRC Incidence & Mortality

1.6.1 Screening & Early Detection

Survival rates in CRC patients are strongly correlated to the cancer stage at the time of diagnosis. Current 5-year survival rate stands at 93.2% for those diagnosed at Dukes' A, which decreases to 77.0% at Dukes' B, 47.7% at Dukes' C and to 6.6% at Dukes' D [159]. This highlights the role of early detection in reducing CRC morbidity and mortality, as shown by the data from the Nottingham (UK) trial where 15% reduction in cumulative CRC mortality was observed in 45-74 year olds who underwent screening [160, 161]. Other randomised controlled trials conducted around the world have also shown up to 25% reduction in CRC mortality in individuals who participated in annual or biennial screening programmes [19, 162-166]. A retrospective analysis of the WHO CRC mortality database (1970-2011) for the European countries have also linked reductions in CRC mortality to increased participation in screening programmes and improved care [167]. These data undoubtedly showcase the importance of screening for early detection in reducing CRC mortality and incidence.

1.6.1.1 The Bowel Cancer Screening Programme

The NHS in the UK currently operates the nation-wide Bowel Cancer Screening Programme (BCSP) since 2006 [168, 169]. The screening invites eligible individuals (aged 60-74 years) to complete three non-hydrated guaiac faecal occult blood test (gFOBT) biennially [168, 169]. Each test kit contains 6 test windows that need to be smeared with faecal samples within a 10-day period [168, 169]. The gFOBT uses an α -guaiaconic acid soaked guaiac paper onto which a faecal sample can be smeared; if the sample contains occult (invisible) blood, the haem group of haemoglobin catalyses the oxidation of α -guaiaconic acid into a blue-coloured, highly conjugated *bis*-methylenequinone compound when hydrogen peroxide is applied [170, 171]. If a test is positive in ≥ 5 test windows, the individual is referred for colonoscopic examination [169]. If a test gives 1-4 positive windows, the individual is asked to re-do the test and is referred for colonoscopy if any windows show positive for occult blood in the 2nd test. They are only discharged if 2nd and 3rd tests showed all negatives [169]. Individuals whose 1st gFOBT

tests show all negatives are considered normal and would be re-invited for testing again in 2 years [169].

The 2011 BCSP data showed that the screening resulted in the diagnosis of early stage CRCs in 71.3% of those who underwent colonoscopy: 9.8% of cases are diagnosed with polyp cancer, 32.0% with Dukes' A, 29.5% with Dukes' B, 25.7% with Dukes' C and 3.0% with Dukes' D [169]. Interestingly, it also reported a significantly higher detection rate for left-sided lesions (77.3%) compared to the right-sided ones (14.3%), which may be indicative of the propensity of right-sided lesions in producing detectable occult blood only when they are in more advanced stages (Dukes' B or higher) [169, 172].

However, a limitation of gFOBT is its susceptibility to false positive and negative results [169, 173]. False negatives can arise from enzymatic degradation of the haem group in the upper GI tract, or inhibition of the α -guaiaconic acid oxidation due to ingestion of food or supplements containing vitamin C [174, 175]. In contrast, false positives can be caused by haem groups present in ingested red meats and other peroxidases from uncooked vegetables [170, 171]. Moreover, gFOBT has been shown to have low sensitivities (5-33%) in detecting adenomas [117, 175, 176]. Another limitation of gFOBT is poor compliance, especially in deprived areas: the 2006-2009 BCSP data showed 35% compliance in the most deprived areas compared to 61% in the least deprived areas [169, 177].

There is a revised version of FOBT called faecal immunochemical test (FIT), which targets the globin component of haemoglobin, making it more sensitive in detecting the occult blood [117, 174, 175]. FIT has advantages such as improved diagnostic accuracy, lower sample requirement and elimination of dietary and medication restrictions [117, 174, 175]. Whilst FIT has yet to be implemented as part of the UK BCSP screening, a 2015 review has highlighted the cost-effectiveness, increased compliance, better sensitivity (up to 92%) and specificity (up to 96%) of FIT compared to gFOBT [173, 178].

From 2013 onwards, the UK bowel screening programme also offers an additional one-off test called bowel scope screening, known as flexible sigmoidoscopy, to those aged over 55 years [168]. A multi-centre randomised controlled trial conducted across 14 UK centres (11 in England, 2 in Wales and 1 in Scotland) with eligible 55-64

year olds showed that this technique can detect lesions within distal colon and rectum with a sensitivity comparable to that of gFOBT test, and accountable for 33% and 40% reduction in incidence and mortality respectively [179]. Flexible sigmoidoscopy has the benefits of reduced time, cost and risk compared to a full colonoscopy, but the limitation is that it only screens for sigmoidal lesions [179].

1.6.2 Prevention of CRC

CRC prevention can be implemented in primary, secondary and tertiary stages to reduce cancer incidence and mortality [180]. Primary intervention aims at stopping the development of precancerous lesions by reducing the exposure to various known risk factors through education, informed lifestyle choices and providing therapeutic choices (e.g. aspirin) [180, 181]. Secondary intervention targets individuals who likely have the disease, and involves screening for early lesions and offering treatment options to reduce the risk of malignant progression: the UK BCSP screening is one such example [180, 181]. Primary and secondary interventions can overlap especially when individuals with germline conditions (e.g. Lynch syndrome sufferers) are involved [180, 181]. Tertiary prevention occurs after CRC has been diagnosed, and involves implementation of various treatment programmes aimed at reducing cancer progression or recurrence to improve prognosis [180].

1.6.2.1 Lifestyle & Diet

A recent meta-analysis has shown that physical activities (>21 metabolic equivalent task hours (MET)/week) can provide 20-30% risk reduction in both men and women compared to those who are less active (<2 MET hours/week) [182, 183]. Smoking is another lifestyle factor that is associated with 20% increase in the CRC risk for both active and former smokers who has smoked for 40+ years [182, 184]. Dietary choices like high consumption of red and processed meat (>160g/day) are also associated with up to 20% increase in 10-year CRC risk in 50+ year olds, compared those who consumed low amounts (<20g/day) [181, 182, 185]. One postulation is that red meat increases endogenous insulin secretion that leads to activation of Insulin-like growth factor 1 (IGF1) receptor, resulting in cell proliferation [182, 185]. Similarly, low

consumption of fish (all types; <10g/day) was associated with 40% increase in 10-year CRC risk, compared to those with high fish intake (>80g/day) [182, 185].

Lifestyle and diet choices underlie whether an individual present with obesity, which is one of the main CRC risk factors in the developed countries [186]. Body mass index (BMI) in kg/m² is the standard measure of an individual weight and a BMI of 30 is considered as obese. Studies have shown that a 5-unit increase in BMI in kg/m² is associated with 30% and 12% increase in the colon cancer risk for men and women respectively, whereas there is a lower 12% increase in the rectal cancer risk for obese men but no correlation was observed for female participants [186, 187]. Obesity, especially abdominal obesity, leads to insulin resistance where there is reduced cellular responsiveness to insulin, resulting in the need for higher insulin levels to maintain plasma glucose levels [188]. Consequently, this leads to hyperinsulinemia and an increase in free insulin-like growth factor 1(IGF-1) - IGF receptors are expressed in the colon and are activated by free IGF-1, resulting in the inhibition of colonocyte apoptosis and induction of proliferative responses [187-189]. Additionally, obesity can also promote chronic inflammation through increased production of lipids and high plasma glucose levels, which stimulate the expression of proinflammatory cytokine tumor necrosis factor α (TNF- α) that contributes to increased CRC risk [187-189]. Taken together, primary intervention can be implemented by raising public awareness that conscientious adjustment of lifestyle and diet choices can lead to lowering long-term CRC risk [181, 182, 185].

1.6.2.2 Chemoprevention

Chemoprevention refers to administration of natural, synthetic or biological agents to either lower the risk of CRC incidence, delay the disease onset or prevent malignant progression [190-192]. Chemoprevention overlaps as both primary and secondary intervention [191, 192]. In the last decade, this field has gathered momentum due to proven efficacies in solid cancers, including CRC [191-193].

1.6.2.2.1 Aspirin

Aspirin, or acetylsalicylic acid, is a non-steroidal anti-inflammatory drug (NSAID) that is commonly used as an analgesic, anti-inflammatory and anti-pyretic drug [194]. Its mechanism of action centres on inhibition of prostaglandins and thromboxane production by irreversibly inhibiting the action of cyclo-oxygenase (COX) enzymes [194].

Interest in aspirin as a chemopreventive agent stemmed from observations made in cardiovascular disease prevention trials that long-term aspirin intake (≥ 75 mg/day) is associated with lower CRC mortality, especially from proximal tumours [195-197]. Since then, epidemiological studies and clinical trials have provided substantial evidence for aspirin as a chemopreventive agent in prevention of sporadic CRCs [191, 195, 198-201]. The 24-year long landmark Nurses' Health Study, involving approximately 80,000 female volunteers, reported that low to moderate aspirin intake is associated with 24% reduction in CRC risk and 35% reduction in mortality from CRC effective from 10 years onwards [196, 197, 202-204]. Studies also showed that lower aspirin dosing (< 300 mg/day) takes longer (> 5 years) to be efficacious at reducing CRC mortality compared to higher doses (≥ 300 mg/day), although aspirin at any dose demonstrated risk reduction in adenoma recurrence within shorter intervals (< 3 years) [191, 196, 205].

Several randomised clinical trials (RCT), known as Colorectal Adenoma/Carcinoma Prevention Programme (CaPP), conducted over two decades have evaluated the long-term chemopreventive potential of aspirin in patients with HNPCC and FAP [195, 198-201]. The initial study, CAPP1, evaluated the chemopreventive effects of aspirin (600mg/day) in 201 FAP sufferers and reported smaller polyp sizes in participants who took aspirin for > 1 year, but observed no effect on polyp numbers after 17 months of aspirin intake [195, 198]. The second study, CAPP2, was conducted with 937 Lynch syndrome, as known as HNPCC, sufferers using the same aspirin dose as CAPP1, and showed that aspirin has a delayed chemopreventive effect in reducing CRC incidence by 63%, which only became efficacious from 4 years after dosing [195, 199, 201]. Currently, a double-blind, non-inferiority trial with different doses of aspirin (100mg, 300mg and 600mg) in Lynch syndrome sufferers called Cancer Prevention Programme (CaPP) 3 is running until 2021 and the results from this study will provide

more information on the optimal dose, adverse effects and efficacy of aspirin as a chemopreventive agent [195].

Interestingly, the clinical response to aspirin seems to depend on the genetic profiles as one study showed that the regular aspirin (325mg) intake of >14 tablets/week lowered the risk of CRC only in *BRAF^{WT}* cases (hazard ratios 0.43 vs. 1.03) [206]. This mutation-specific efficacy suggests that understanding individual genetic profiles may improve treatment efficacies.

1.6.2.2.2 Celecoxib

There is also an interest in substituting standard NSAIDs with selective COX2 inhibitors, such as celecoxib, for early chemoprevention due to their lower GI toxicity profiles [191, 207, 208]. An initial study with celecoxib taken at 400mg dose twice a day for 6 months (n=83) showed 31% reduction in adenoma burden in FAP patients, which led to accelerated FDA approval of celecoxib, marketed as Celebrex[®], as an adjunct treatment for FAP management [208]. Similarly, data from the Adenoma Prevention with Celecoxib trial showed that the same dose of celecoxib is associated ~45% reduction in adenoma recurrence for sporadic cases, but they reported an increase in cardiovascular risk [182, 209]. Subsequent trials with larger cohorts also reported up to 3-fold increase in cardiovascular risk [210, 211], leading to celecoxib withdrawal by Pfizer in 2011 due to unsatisfactory long-term (6+ months) clinical efficacy and safety profiles [212].

1.6.2.2.3 Curcumin

Curcumin (diferuloylmethane) is an active agent in turmeric commonly used in curries [213], and various *in vitro* studies showed that curcumin exerts its preventive effects through anti-proliferative, anti-angiogenic and pro-apoptotic properties [213-215]. Curcumin has been shown to downregulate transcription factor nuclear factor κ B (NF- κ B), resulting in decreased expression of cyclin D1 and COX2, as well as downregulating pro-angiogenic factors, such as vascular endothelial growth factor (VEGF) [213-215]. A small clinical trial involving 12 CRC patients have shown that

administration of curcumin at 3,600mg/day for a week resulted in anti-oxidant effect through reduction in oxidative DNA adducts [213, 216]. However, despite its low toxicity in humans, difficulty with administration of curcumin above 8,000mg/day combined with poor oral bioavailability – only trace serum levels were detected at $\geq 8,000$ mg/day – limit its use as a chemopreventive agent [213, 216-218]. However, preclinical studies with Meriva, a phospholipid formulation of curcumin, have shown improved bioavailability as well as chemopreventive effects, suggesting its potential as a curcumin substitute [219].

1.6.2.2.4 Resveratrol

Resveratrol (*trans*-3,5,4'-trihydroxystilbene) is a polyphenolic phytochemical found in the skin of red grapes, nuts and plants, and the preclinical and clinical studies conducted in our team have led to promising chemopreventive outcomes with resveratrol and its analogues [213, 220-222]. In one study with *Apc^{Min/+}* model of CRCs, resveratrol and its synthetic analogue (3,4,5,4'-tetramethoxystilbene), both administered at 0.2% in the diet (240mg/kg), hindered adenoma development by 27% and 24% respectively [222]. Daily resveratrol intake, even at dietary achievable concentrations of up to 1g/kg body weight, was shown to be associated with low adverse profile in study mice [216, 223, 224]. Interestingly, it has been observed that a low dose of resveratrol is correlated to a more potent chemopreventive effect in *Apc^{Min/+}* models – animals receiving 0.00007% resveratrol (0.07mg/kg) showed approximately 52% reduction in tumour burden compared to 25% reduction at higher dose of 0.0143% resveratrol (14mg/kg) [225].

Our team has also conducted a clinical study involving 20 CRC patients who were administered with either 0.5mg or 1.0mg/day resveratrol for 8 days prior to surgical resection [220, 221]. Post-surgical tissue analysis revealed that resveratrol and its metabolites were detected at metabolically relevant concentrations – for instance, resveratrol was detected at 94.1 ± 89.2 nmol/g and resveratrol sulfate glucuronide at 27.1 ± 21.6 nmol/g in tumour tissues from the right side of colon in patients taking 1.0mg/day resveratrol [220, 221]. The study found that both doses elicited ~10% reduction in tumour cell proliferation [220, 221]. However, whilst single doses of resveratrol (0.5-5.0g) have short-term low toxicity, more pronounced adverse effects

(e.g. diarrhoea, nausea) were observed in volunteers (28/40) when same doses were taken daily for 29 days, suggesting long-term toxicity even at low doses [220, 226, 227].

The mechanism of action of resveratrol is still unclear, but several pathways have been implicated. Efficacy of resveratrol is correlated to phosphorylation and activation of the adenosine monophosphate-activated protein kinase (AMPK), a crucial metabolic sensor, and is involved in insulin sensitisation, fatty acid oxidation, regulation of autophagy and induction of cellular senescence [220, 225, 228]. In particular, resveratrol may exert its chemopreventive effect through autophagy, a tumour-suppressing mechanism controlling proteins/organelles functions and energy homeostasis [225, 229]. *In vivo* data showed that resveratrol induced short-term, upregulation of autophagy through AMPK activation [220, 225, 228]. Some studies suggested that resveratrol also exerts its chemopreventive and anti-inflammatory effects through downregulation of transcription factor NF- κ B activity [220, 230, 231].

Overall, the chemopreventive effects of aforementioned compounds have been very promising and there has been considerable interest in developing these agents into the clinics for cancer prevention [213, 220-222]. However, the lack of suitable surrogate biomarkers has hampered their clinical translation.

1.7 Biomarkers for CRC

1.7.1 Overview of Biomarkers

A biomarker is an objectively measurable, biological molecule that exists in either bodily fluids or tissue, and is susceptible to physiological changes to represent either a normal or abnormal condition [232-234]. Good specificity, sensitivity, cost-effectiveness and ease of application define a good biomarker. Sensitivity is a measure of how many true positives are identified, and specificity is a measure of how many true negatives are identified [232-234]. Biomarkers can be predictive, diagnostic or prognostic [232-235]. Predictive biomarkers measure how an individual would respond to a therapeutic agent or treatment regimen [234, 235]. Diagnostic biomarkers are used to identify the genetic or genomic alterations and severity disease in individuals who have already been diagnosed [234, 235]. Prognostic biomarkers assess how a disease

may develop and progress in an individual irrespective of treatments [234, 235]. Biomarkers can also be surrogates if their measurements act as substitutes for more laborious, but accurate, outputs [234, 235].

1.7.2 RAS & RAF Mutation Statuses

Anti-EGFR therapies, such as monoclonal antibodies (Mabs) cetuximab and panitumumab, have been approved for treatment of various *KRAS* and *BRAF* MU cancers, including CRCs [138-141]. In metastatic CRCs, mutation status of *KRAS* and *BRAF*, upstream components of EGFR signalling, have been proven as predictive biomarkers for anti-EGFR therapies [141, 236]. A retrospective analysis of metastatic patients (n=649) treated with a combination of cetuximab with chemotherapy showed that *KRAS* WT patients had better response to therapy (35.8% vs. 6.7%) and better overall survival (50 weeks vs. 32 weeks) compared to *KRAS* MU patients [236]. Similarly, *BRAF* WT status was linked to better response (38.0% vs. 8.3%) as well as better overall survival (54 weeks vs. 26 weeks) compared to *BRAF* MU cases [236]. Another retrospective study with metastatic CRC patients (n=168) on either anti-EGFR Mab alone (cetuximab/panitumumab) or with chemotherapy also showed that *KRAS* MU status, even at low MU frequency (<10%), is associated with lower response (6.7% vs. 37.0%) and a 2-fold reduction in progression-free survival period [141]. Similarly, *BRAF* status serves as a predictive biomarker in metastatic CRC patients treated with vemurafenib, a selective inhibitor of MU BRAF [237, 238]. Vemurafenib as a single agent, despite showing high potency (80%) in melanoma, only elicited response from 5% of *BRAF* MU CRCs due to feedback upregulation of EGFR signalling; this feedback was not observed in melanoma cells due to low EGFR expression [237, 238]. Additionally, for stage II+ CRC cases, BRAF mutation status has prognostic value in conjunction with MSI status; the presence of *BRAF*^{V600E} mutation in MSS tumours is positively correlated with poor prognosis [239].

1.7.3 Carcinoembryonic Antigen

Carcinoembryonic antigen (CEA), a glycosylphosphatidylinositol (GPI)-linked membrane glycoprotein, is a widely-used biomarker for monitoring therapeutic

response in metastatic CRC patients, and it is also used as a prognostic biomarker in patients undergoing surgical resections [233, 234, 240]. High pre-operative CEA levels in serum (>5mg/ml) in primary CRCs may suggest poor prognosis, and more than 30ng/ml CEA levels are associated with lack of response to treatment and reduced survival in metastatic cases [233, 240]. However, CEA as a biomarker for early detection is ineffective as many studies have reported lack of CEA elevation in both precancerous and stage I patients, as well as having a low sensitivity (10%) in detecting benign lesions [233, 234, 241, 242].

1.7.4 Carbohydrate Antigen

Carbohydrate antigen (CA) 19-9, a glycoprotein, is another heavily investigated biomarker but there are insufficient data to confirm its clinical implementation [233, 240]. One study showed that prognostic sensitivity of CA 19-9 (threshold of >37U/ml) increases with CRC staging; 17% in Dukes' B, 34% in Dukes' C, 32% in Dukes' D1 and 69% in Dukes' D2, but the number was too low to properly assess the prognostic value of CA 19-9 for Dukes' A patients (n=4) [243]. Another study also reported that CA 19-9 has a low sensitivity (9.4%) in detecting early recurrences, but suggested that it might perform well as a prognostic biomarker when implemented in combination with other biomarkers [233, 240, 244].

1.7.5 Faecal Calprotectin

Another CRC biomarker that has been investigated is faecal calprotectin (FCP), a non-glycosylated calcium-binding protein present in the cytoplasm of neutrophils, which is released in response to tumour-induced inflammation or haemorrhage [245, 246]. A multi-centre study (n=453) has reported that a one-off enzyme-linked immunosorbent assay (ELISA)-based FCP testing of FCP in 50-100mg faeces can detect early Dukes' stages (A & B) with 84% sensitivity and 68% specificity [245, 246]. Similarly, another large study conducted in York (n=654) reported that FCP has a high sensitivity (91.9%) but a low specificity (36.4%) at detecting neoplasia (polyps \geq 10mm) [247]. Low specificity in detecting early lesions is a major limitation for FCP [245-247].

1.7.6 Circulating Cell-free DNA

Although the concept of circulating cell-free DNA (cfDNA) has been introduced since 1948 by Mandel and Metais, it was only in 1994 that the biomarker potential of cfDNA was first reported by two studies where *KRAS* and *NRAS* mutations were detected in matching plasma and tissue samples from pancreatic cancer and acute myelogenous leukaemia (AML) patients respectively [248-251]. This plasma DNA of tumour origin came to be defined as circulating-tumour DNA (ctDNA) or tumour-derived cfDNA [142, 252, 253]. Traditional single biopsies provide only a small representation of the overall cancer and underestimate the prevalence of mutations in heterogeneous tumours, increasing the risk of recurrence and resistance [254-258]. Therefore, the appeal of tumour-specific cfDNA lies in its representation of a whole tumour, and cfDNA analysis as *liquid biopsy* can hypothetically provide a comprehensive profile of genetic heterogeneity within a patient [254-258]. In 2016, the FDA and EMA have approved the tumour-specific cfDNA biomarker test called cobas® EGFR mutation test v2 (Roche) for plasma analysis of advanced lung cancer patients as a companion diagnostic test [259].

1.7.6.1 Properties of cfDNA

cfDNA are double-stranded, short DNA fragments (~150bp) of nuclear and mitochondrial origins [142, 253]. It has an average half-life of two hours, and there is evidence that it forms complexes with lipids and proteins to form cytosolic vesicles and may also be associated with extracellular vesicles called exosomes in circulation [142, 260, 261]. Healthy individuals have low cfDNA levels (1-30ng/ml), and cfDNA has a physiological role in immune response and blood coagulation through serving as an essential component of neutrophil extracellular traps (NETs) – fibrous traps of granule proteins and chromatin – that bind and degrade microbial pathogens [253, 262]. Elevated levels of cfDNA have been detected in individuals affected by a range of conditions including cancer, inflammatory bowel disease and autoimmune diseases [142, 261, 263-265]. Moreover, obesity, a CRC risk factor, has shown to be implicated in increasing plasma cfDNA levels in both humans and mice by promoting adipocyte degeneration and subsequent cfDNA release into circulation, which in turn induces inflammatory response due to macrophage accumulation via the activation of Toll-like

receptor 9 (TLR9) upon which cfDNA acts as a ligand [266, 267]. In advanced stages of cancer, circulating cfDNA levels can increase up to ~230ng/ml, although it can have a low range of 1-5ng/ml plasma; differences in quantitative methods used are likely the cause of this wide range of concentrations [268-270].

1.7.6.2 Mechanisms of Release

It has been postulated that cell lysis and apoptosis are the two main mechanisms of cfDNA release (Figure 1-17) [254, 271]. The basis for apoptotic cfDNA release came from the observation that gel electrophoresis of plasma or serum cfDNA shows a ladder pattern bands with sizes that are multiples of nucleosomal DNA length (~180-200bp), and nuclease-mediated fragmentation of nucleosomal DNA is a hallmark of late apoptosis [261, 271-273]. This notion is also fostered by an increase in nucleosome-complexed cfDNA levels after chemotherapy in cancer patients, suggesting the link between cfDNA and apoptosis [272, 273]. It has been noted that that cfDNA from cancer patients shows decreased DNA integrity (i.e. higher fragmentation), indicative of higher apoptotic rates, and given the positive correlation between increased proliferation and apoptosis in cancers, this further supports the link between apoptosis and cfDNA release [274-276]. Some studies hypothesised that necrotic cell death also contributes to peripheral cfDNA pool due to the presence of large DNA fragments (>10,000bp), which is a hallmark of necrotic cell death, seen as a smear on gel electrophoresis due to incomplete and non-specific digestion [271, 277, 278]. Another possibility is that cfDNA is actively released from cells, based on *in vitro* studies with labelled DNA where spontaneous release of DNA has been observed [263, 271, 279, 280]. However, the mechanism of this active processes of cfDNA release and, most importantly, its relevance *in vivo* remains to be investigated.

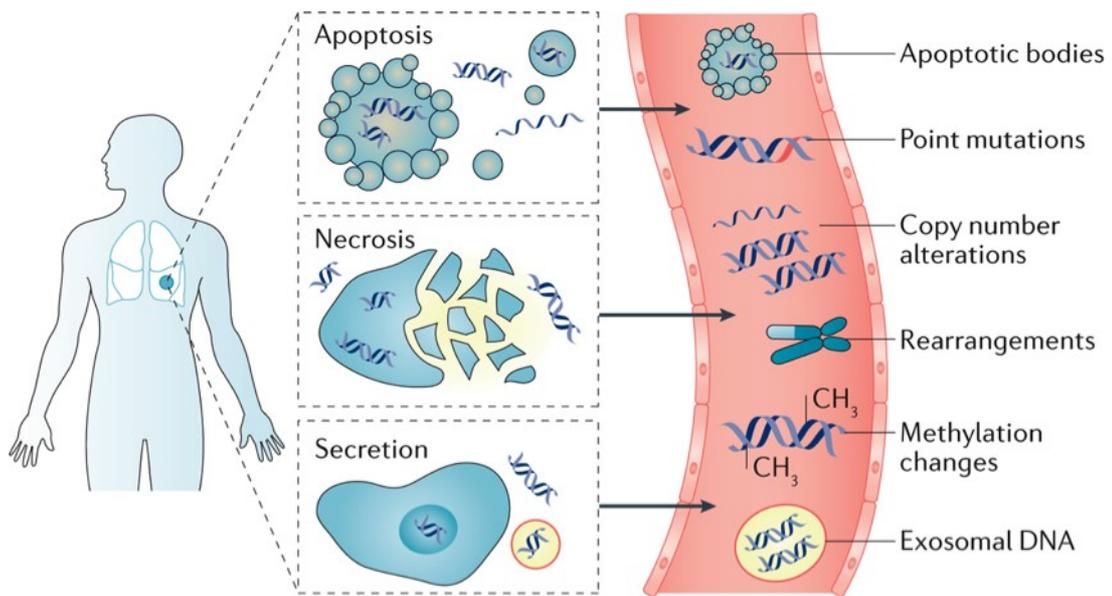


Figure 1-17: Mechanisms of cfDNA Release into Peripheral Circulation

Short fragments (~150bp) of cfDNA are released into circulation when cells, normal or cancerous, die from either apoptosis or necrosis. Increased cell death associated with tumour growth is one of the contributing factors for higher levels of cfDNA observed in cancer patients compared to healthy individuals. [271-273, 302]

In healthy individuals, the cfDNA levels are low because physiological rates of programmed cell death are relatively low and debris, such as naked cfDNA and aging neutrophils, are cleared away by macrophage phagocytosis to prevent manifestation of pro-thrombogenic or inflammatory effects [255, 261, 281]. However, in cancers, the increased rate of cell death presumably overwhelms the ability of infiltrating phagocytes to process cellular debris, leading to elevation in circulating cfDNA pool that originated from dead cancerous cells and their micro environment [248, 274].

1.7.6.3 Pre-analytical & Analytical Considerations

cfDNA can be isolated from either plasma or serum. Although it has been shown that the serum gives higher yields, it was due to contamination with genomic DNA from clotting and lysis of white blood cells (WBC), which increases the background DNA content and reduces the detectability of tumour-specific cfDNA [280, 282-284].

Therefore, plasma is preferentially used for cfDNA analysis as it gives better yields of tumour-derived cfDNA, even in cases of delayed sample processing [253, 285-287].

Blood samples are usually collected in EDTA-coated tubes to prevent coagulation and WBC lysis [280, 288, 289]. EDTA also indirectly inhibits endogenous DNases by chelating divalent ions and protects cfDNA from degradation [288]. Recommended sample processing follows sequential centrifugation steps at 1,000 g followed by 2,000g at 4°C for 10 minutes each to produce clean plasma [285, 290]. Samples can be stored at -20°C for short-term (≤ 3 months) without affecting cfDNA quality [291]. Freeze-thaw cycle should be kept to a minimum, as the cfDNA integrity shows a decrease after three cycles [282, 285]. Samples should be kept at -80°C for long-term storage, and although studies have shown that cfDNA levels and integrity decreases with storage time, it does not affect the detection of tumour-derived cfDNA even after >12 years [285, 290, 292-294].

Different studies used different cfDNA extraction techniques, but the most commonly used are the commercial kits (e.g. QIAamp Circulating Nucleic Acid kit from QIAGEN) that employed proprietary column-based extraction chemistry with proven high yields [252, 253, 290, 295]. Interestingly, Simon and colleagues also successfully performed direct plasma analysis for cfDNA quantification; although it is more economical, this could limit detection of tumour-specific cfDNA due to contaminants present in the plasma [296]. Plasma volume requirement varies between different techniques, and various studies have used volumes between 0.85-10ml of plasma for detection of tumour-specific alterations on PCR and sequencing platforms [276, 297, 298]. However, volumes as low as 0.5ml plasma have been successfully used for quantitative analysis, but with the caveat that a sensitive target is used; good reference genes are usually interspersed or tandem repeat elements such as repetitive retrotransposon *LINE-1* (long interspersed nuclear element-1) sequences, comprising 17-18% of human genome, and *ALU* elements or housekeeping genes such as glyceraldehyde 3-phosphate dehydrogenase (*GAPDH*) [299].

Different methodologies and sample volumes used between studies make it difficult to interpret and compare data across various studies, highlighting the need for

a standard operating procedure that can be widely implemented for cfDNA analysis [257, 300].

1.7.6.4 Techniques for cfDNA Detection

Whatever its origin, cfDNA is amenable to analysis by different technologies, although PCR and sequencing-based methods are widely used for detection and/or quantification of plasma cfDNA copies [248, 252, 297, 298, 301, 302]. Different methods have different limits of detection (LoD) and their suitability depends on the abundance of tumour-derived cfDNA in a sample (i.e. advanced vs. early stage cancers) (Figure 1-18) [252, 302].

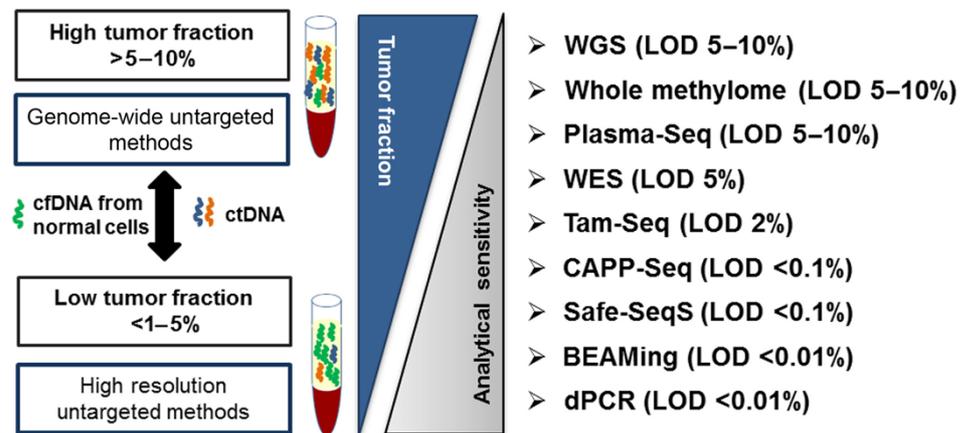


Figure 1-18: Methods used for cfDNA Analysis

Various sequencing and PCR-based methods available for plasma cfDNA analysis. [226]

LOD = limit of detection, WGS = whole genome sequencing, WES = whole exome sequencing, Tam-Seq = tagged-amplicon deep sequencing, CAPP-Seq = cancer personalised profiling by deep sequencing, BEAMing = Beads, Emulsion, Amplification and Magnetics, dPCR = digital PCR.

The genome-wide sequencing-based techniques (e.g. WGS, WES) have the advantage of providing comprehensive genetic and genomic profiles of cancers, eliminating the need for prior knowledge of patient-specific aberrations, usually obtained from analysis tissue biopsies [252, 302]. However, these methods are only effective at analysing samples with high tumour fractions (5-10%). For instance, Rosenfeld's group showed that exome sequencing (WES; Illumina HiSeq2500) of serial

plasma samples from breast, ovarian and lung cancers (n=6) identified emergence of resistance in response to therapy, but its applicability is limited to advanced cases with relatively high plasma MU cfDNA load (median burden = 5-10%) [303]. The use of sequencing-based analysis of cfDNA also has drawbacks such as labour-intensiveness, high costs and the necessary technical expertise [252, 295, 301, 303, 304].

In contrast, high resolution approaches, such as targeted sequencing (CAPP-seq) and PCR-based methods, may fall short on providing comprehensive tumour profiles, but they are able to detect tumour-derived cfDNA fragments that are present at the limit of detection (LoD) threshold [252, 302]. For example, CAPP-seq technique can detect tumour-derived mutations in cfDNA with 96% specificity at ~0.02% LoD in the plasma of non-small cell lung cancer (NSCLC) patients [301, 304]. Similarly, numerous studies have employed PCR-based targeted mutation detection assays to identify hotspot mutations in *KRAS*, *BRAF* and *PIK3CA* at or below 0.01% LoD in various cancers, including CRC [305-308]. For targeted approaches, the degree of cfDNA fragmentation may be a limitation in detecting tumour-derived MU fragments, but this can be solved by designing assays with short amplicon sizes (<100bp) [248, 255, 276, 302].

1.7.6.5 Prognostic & Predictive Values of cfDNA

Several studies have demonstrated the prognostic value of cfDNA in both primary and metastatic CRC patients. One study showed that elevated cfDNA levels (>1µg/ml blood) were correlated with reduced survival in advanced CRC patients (n=55) [270]. Another study with un-resectable metastatic CRC patients (n=29) showed that the presence of tumour-derived *KRAS* mutations in serum was associated with shorter progression free survival (5 months vs. 14 months) [309]. Similar findings were reported in metastatic CRC patients where high plasma *KRAS* MU fractions (>75% quartile) were associated with 0% disease control rate (i.e. neither respond to treatment nor have a stable disease), compared to the 42% control rate observed in patients with low *KRAS* MU fractions (<25% quartile) (range 50-180,000 allele/ml) [310].

Moreover, studies have shown that cfDNA has better predictive values in tracking clonal evolution, detecting early recurrence and monitoring emergence of

resistance in solid tumours compared to other biomarkers (e.g. CEA, CA) [142, 295, 297, 298, 302]. Andersen's group used patient-specific assays to test somatic aberrations in the serial plasma samples, and showed that relapse can be detected earlier (10 months on average) in post-operative CRC patients with 100% sensitivity and specificity compared to conventional methods (e.g. CT scanning) [298]. They also reported the same predictive specificity for serial monitoring by CEA biomarker, but with a much lower sensitivity of 67% and more delayed relapse detection (average of 3.5 months), suggesting that cfDNA out-performs CEA as a biomarker [298]. Another study with 230 resected stage II CRC patients showed that serial sequencing-based analysis of 10ml plasma cfDNA samples detected recurrence with 78.6% sensitivity and 92.1% specificity, performing better than the CEA biomarker [295]. cfDNA has also been shown to predict *KRAS* mutation-mediated resistance to anti-EGFR treatment in metastatic CRC patients up to 10 months before detectable radiologic evidence appears [138, 140, 311].

1.7.6.6 cfDNA as a Biomarker for Early Detection

The potential of cfDNA to serve as a biomarker for early detection has been explored in few studies [248, 299, 312]. An early study has highlighted issues with low sensitivity (39%) and specificity (42%) in detecting codon 12 *KRAS* mutations in plasma cfDNA from patient with colorectal neoplastic polyps (n=64) [313]. Similarly, a recent study also showed that detectable levels of tumour-derived cfDNA (at least 1 MU fragment/5ml plasma) were correlated to the stage in localised cancers: 47% of stage I patients were positive for tumour-derived cfDNA compared to 82% of stage IV patients [312]. A recent study by Gazouli's group also reported the low sensitivities in detecting *KRAS* codon 12 (60%, n=5) and *BRAF* V600E (66%, n=8) mutations in the plasma of patients with neoplastic polyps [314].

Certainly, low MU fractions (i.e. proportion of MU fragments to excess WT background) present in early stage plasma samples is one main limitation for using plasma cfDNA as a biomarker for early detection. In advanced stages, there is a larger population of tumour cells to contribute to a higher MU fraction in circulation and vice versa [312, 315]. Indeed, Vogelstein's group detected *APC* mutations at a higher mean fractional abundance in Dukes' D samples (11.1%) compared to Dukes' A samples

(0.04%) [260, 316, 317]. In addition, the truncal or clonal nature of targeted mutations further affects the MU fractions present in the plasma; for instance, Rosenfeld and Caldas' groups detected truncal mutations at higher fractions of 3.8-34.9% compared to clonal mutations at lower fractions of 2.5-19.1% [318].

One solution for this low detectability with early stage samples is to target repetitive sequences in the plasma cfDNA instead of tumour-specific cfDNA fragments in order to improve sensitivity. For example, Cree's group detected significantly higher total cfDNA levels in their early stage cohort (15.04ng/ml) compared to the controls (7.94ng/ml) by targeting repetitive LINE-1 retrotransposons [319]. Similarly, Watanabe's group reported elevated levels of cfDNA in their early stage CRC patients (9.8ng/ml) compared to the controls (7.7ng/ml) using the same LINE-1 targets [299]. However, the limitation of this method is that total cfDNA levels are affected by a range of physiological states and conditions unrelated to cancer [142, 261, 263-265].

Nonetheless, cfDNA liquid biopsy offers several advantages compared to the traditional tissue biopsies for the management of solid tumours [142, 256-258, 318, 320, 321]. It is a minimally invasive technique and is advantageous where tissue sampling is unavailable or impractical. Liquid biopsies are also suitable for serial collections to monitor disease progression efficiently and cost-effectively compared to tissue biopsies. Therefore, the early diagnostic potential of cfDNA certainly warrants thorough investigation in the hope of fulfilling this unmet clinical need.

1.7.7 Summary

The list of biomarkers (Table 1-1) for CRC that are currently recommended or being considered for clinical application continues to grow with an increase in our understanding of the mechanisms and pathology underlying these agents [234, 240]. The biomarkers discussed in this work are selected few chosen from a list of agents being investigated at present for their potential in screening, prognostic, predictive or monitoring colorectal tumours in the clinics [234, 240]. As the process of cancer management becomes increasingly personalised, biomarkers will undoubtedly take a

more prominent role in assisting the clinicians in tailoring the treatment options to achieve the best prognosis at different stages of CRC.

Table 1-1: Summary of CRC Biomarkers & Potential Clinical Applications

<i>Biomarker</i>	<i>Application</i>	<i>Sample Type</i>	<i>Recommendations for clinical application</i>
KRAS	Prognostic, Predictive	Tissue, Plasma	Recommended as a predictive biomarker for anti-EGFR therapies, but limited evidence for its prognostic value
BRAF	Prognostic, Predictive	Tissue, Plasma	Recommended as a predictive biomarker for anti-EGFR therapies, and has good prognostic value in conjunction with MSI status
CEA	Screening, Prognostic, Follow-up	Serum	Some prognostic value in metastatic cases with poorer performance than cfDNA in relapse detection, but ineffective in detecting early lesions
CA 19-9	Prognostic, Follow-up	Serum	Not recommended as prognostic sensitivity decreases for early cancer stages & low sensitivity in detecting recurrences
FCP	Screening	Faeces	Not recommended due to low specificity in detecting early lesions
cfDNA	Screening, Prognostic, Predictive, Follow-up	Plasma, Serum	Good evidence for prognostic and predictive values in metastatic cases, but poor sensitivity and specificity in detecting early neoplasms

CEA = carcinoembryonic antigen, CA 19-9 = carbohydrate antigen 19-9, FCP = faecal calprotectin

1.8 Technologies used for Targeted Mutation Detection

For this project, the following two PCR systems were primarily implemented due to their time- and cost-effectiveness in testing limited plasma cfDNA material.

1.8.1 Real-time Quantitative Polymerase Chain Reaction

Real-time quantitative PCR (qPCR) system is used for various purposes including gene expression analysis, copy number variation analysis, quantification of target genes and rare mutant detection amongst others [322]. This system enables the amplification data of target templates to be measured during the early exponential phase where the amplified products are proportional to the target templates theoretically, provided there are no PCR inhibitors present in the reaction mixture (assuming 100% amplification efficiency) [323]. The qPCR efficiency ($E = [10^{(-1/m)} - 1] * 100$, where m is the

slope of a standard curve) indicates how well targets are amplified in one cycle during the amplification phase [323, 324]. An assay with good primer and probe designs should have at least 90% efficiency, although presence of residual PCR inhibitors may give efficiency estimates of over 100%, with an acceptable efficiency range of 90-110% for any given assay run [323, 324].

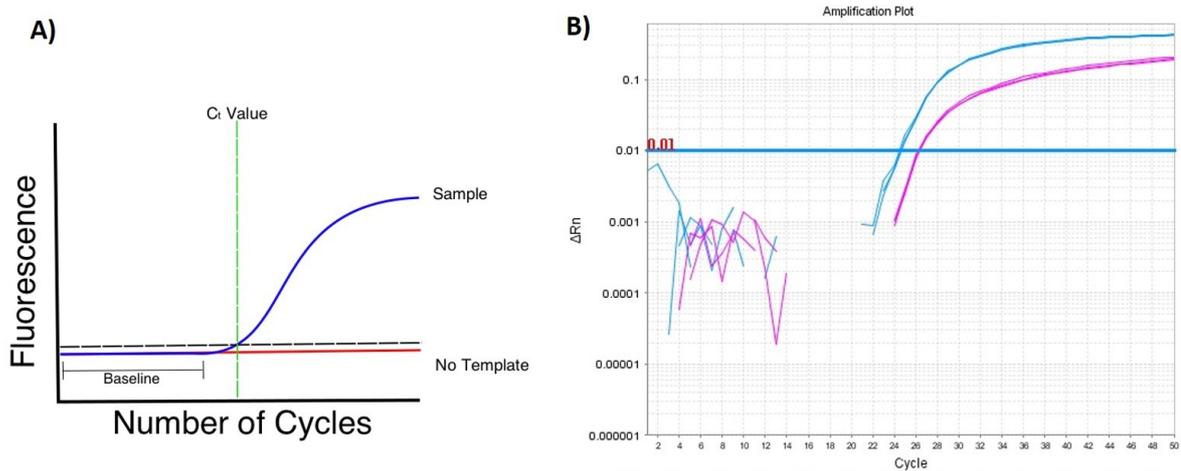


Figure 1-19: Parameters for Fluorescence Data Collection in the qPCR System

A) Graph showing how the baseline value is used to guide the fluorescence threshold (dotted line) to obtain the threshold cycle (C_t) value (green line), which is the least number of PCR cycles needed to detect actual fluorescence signal from the target (blue line). [326]

B) An example ΔR_n vs. cycle amplification plot with two different fluorescent signals. The threshold at $\Delta R_n = 0.01$ determines the C_t values: the blue lines have lower C_t value than the pink lines, indicating that the blue target has a higher quantity than the pink target. [325]

The StepOnePlus™ qPCR system was used for all qPCR analysis performed in this project. At each amplification stage of a PCR programme, real-time fluorescent data was obtained by exciting the fluorophores in reaction wells and collecting the emission wavelengths [323]. The pre-calibrated ROX™ dye is used as a passive reference to normalise fluctuations in reporter dyes, caused by differences in dye concentrations between reaction wells – this gives a normalised reporter (R_n) signal [323]. All data was analysed by the StepOnePlus software, which produces a ΔR_n vs. Cycle amplification plot where amplitudes of baseline-corrected normalized reporter (ΔR_n) [$\Delta R_n(\text{cycle}) = R_n(\text{cycle}) - R_n(\text{baseline})$] are plotted against the cycle numbers (Figure 1-19) [323]. This amplification plot is then used to obtain cycle threshold (C_t) values for each reaction well, and C_t is the PCR cycle number where the ΔR_n fluorescence is above the baseline

and within the exponential phase (Figure 1-19, [325]) [323]. C_t values inversely correlate to the sample quantity and are determined by the baseline: a good baseline is set during the initial cycling stages where there are few changes in the fluorescence signal [323].

The qPCR system has advantages over the traditional PCR. In a traditional PCR reaction, targets are analysed at the end-point via gel electrophoresis, resulting in lower precision due to size-based determination of PCR targets [326]. End-point detection is also subjected to sample and reagent bias because, at this point, a reaction is in a plateau phase and amplification has almost stopped due to consumption of reaction components, and PCR products also start to degrade [326]. In contrast, real-time qPCR collects data during the early exponential phase where the reaction components are in abundance and exact doubling of targets occurs, provided 100% efficiency, which gives a more precise data [323, 326].

1.8.1.1 Allele-specific Mutation Detection Assays

Allele-specific mutation detection assays on qPCR utilise TaqMan™ fluorescent probes (Figure 1-20). They are short probes (≈ 15 bp) that have been designed to anneal to a specific sequence, and they are tagged with fluorescence dyes at the 5' end for target detection [323, 326]. They also contain a 3'-end minor groove binder (MGB) moiety, dihydrocyclopyrroloindole tripeptide (DPI₃), whose isohelical confirmation binds to the minor groove of B-form DNA and is stabilised by van der Waals forces, providing these probes with higher melting temperature (T_m) and makes them more specific at discriminating single base mismatches [327]. TaqMan™ probes also contain a proprietary, non-fluorescent quencher (NFQ) complexed to the MGB moiety at 3' end, which quenches the reporter dye at the 5' end when a probe is intact [323]. During amplification, the 5'-3' exonuclease activity of the thermostable Taq (*Thermus aquaticus*) polymerase digests the probe from the 5' end, releasing the dye from the proximity of NFQ to enable fluorescence emission (Figure 1-20) [322, 328, 329]. A common pair of reporter dyes used are VIC™ (absorbance maximum= 538 nm; emission maximum= 554 nm), a proprietary product of Life Technologies, and 6-carboxyfluorescein™ (6-FAM™ or FAM™) (absorbance maximum= 492 nm; emission maximum= 517 nm) dyes as they show non-overlapping fluorescence profiles, enabling

two targets to be measured in parallel within one reaction [330, 331]. For all the custom assays developed as part of this project, VIC dye was assigned to WT alleles and FAM dye to MU alleles for mutation detection (Figure 1-20).

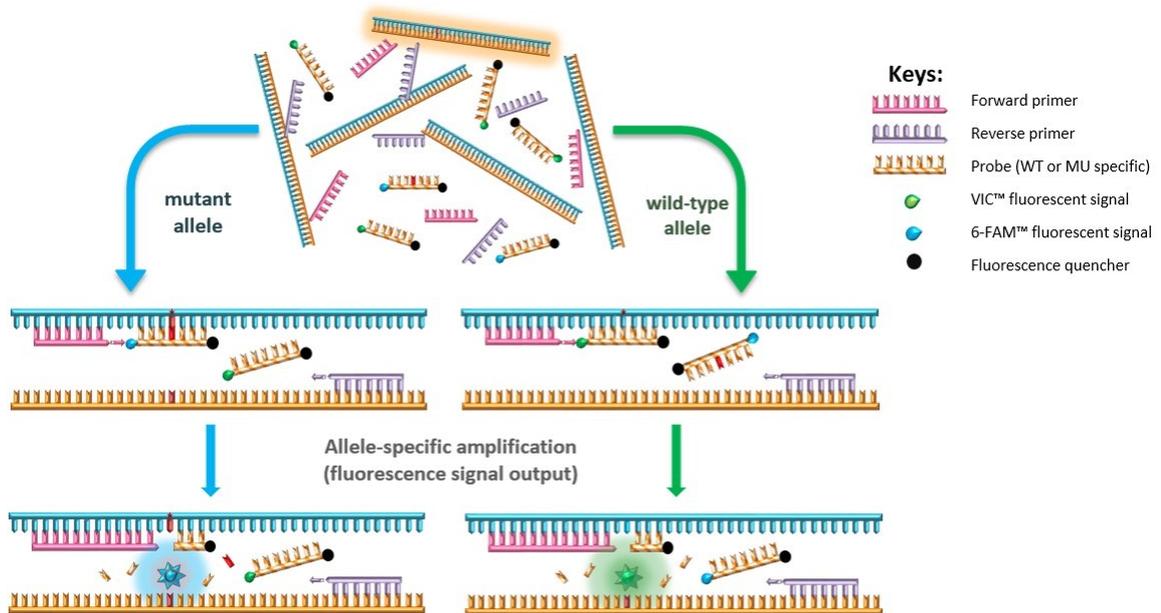


Figure 1-20: Allele-specific Mutation Detection with TaqMan qPCR System

An example assay set-up for detection of WT and MU alleles using sequence specific probes. The forward and reverse primers are shared between the two alleles. The allele-specific probes have a fluorescent reporter (VIC or FAM) at the 5' end and non-fluorescent quencher (NFQ) at the 3' end complexed to minor groove binder (MGB). During an extension phase, the fluorescent reporter is released from the quencher by the action of a 5'-3' exonuclease and fluoresces when hit by an excitation wavelength. [324]

The assays are designed so that WT and MU amplicons have the common forward and reverse primers. Allele-specific probes have sequence similarity with the exception of a mutant base, which is used as a focal point to achieve the best mismatch discrimination [332]. Hybridisation dynamics between primers, probes and templates are crucial to achieve efficient target detection. Both forward and reverse primers are designed to be 18-24 bp in length with an ideal T_m range of 58-60°C, and probes are designed to have higher T_m by a difference of 5-10°C; this prioritises probe-template hybridisation before primers anneal to ensure that probes are ready to be digested during the amplification phase [323].

1.8.1.2 Peptide Nucleic Acids

Plasma samples from precancerous patients contain very low MU fractions (<1%) in great excess of WT background, which can greatly hinder detection and amplification of tumour-derived cfDNA fragments [252, 302]. Therefore, allele-specific mutation detection assays (Figure 1-20) are run in the presence of peptide nucleic acids (PNAs) to facilitate detection of few MU copies in the plasma.

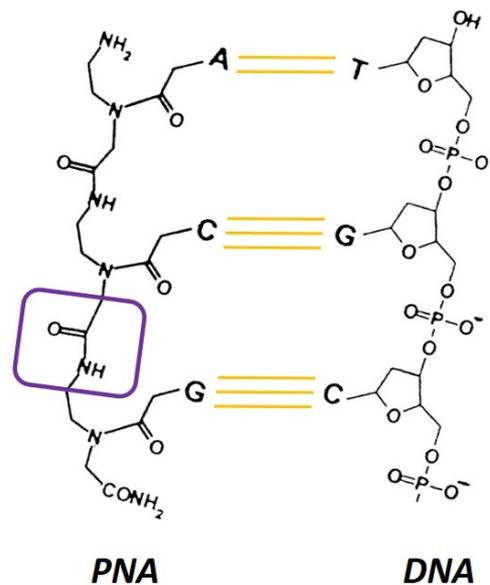


Figure 1-21: Watson-Crick Pairing between PNA & DNA Strands

The non-polar PNA backbone is made of 2-aminoethyl-glycine links (purple box), enabling it to pair with DNA via Watson-Crick complementarity (yellow line represents hydrogen bonds). [335]

PNAs are synthetic structural mimics of DNA strands, and they have a peptide backbone with N-(2-aminoethyl)-glycine linkages instead of the phosphodiester-bonded polar backbone of DNA (Figure 1-21) [333, 334]. The purine and pyrimidine bases are attached via methyl carbonyl bonds to produce amide-linked backbone polymers that are analogues of oligonucleotides [333, 334]. The non-polar peptide backbone grants PNA oligomers resistance to digestion by exo/endo-nucleases [333, 334]. PNAs as short oligomers (15-17bp) can bind to complementary DNA sequences via Watson-Crick pairing, and due to the lack of electrostatic repulsion between PNA-DNA duplexes, this hybridisation shows higher thermal stability (i.e. higher T_m) with reduced susceptibility

to changes in ionic concentrations [333, 334]. Additionally, PNA-DNA duplexes are more sensitive to single base mismatches – a single mismatch can lower the T_m of a 15-mer PNA-DNA duplex by 4°C more than that of a DNA-DNA duplex – and making them ideal for mutation detection assays [333, 334].

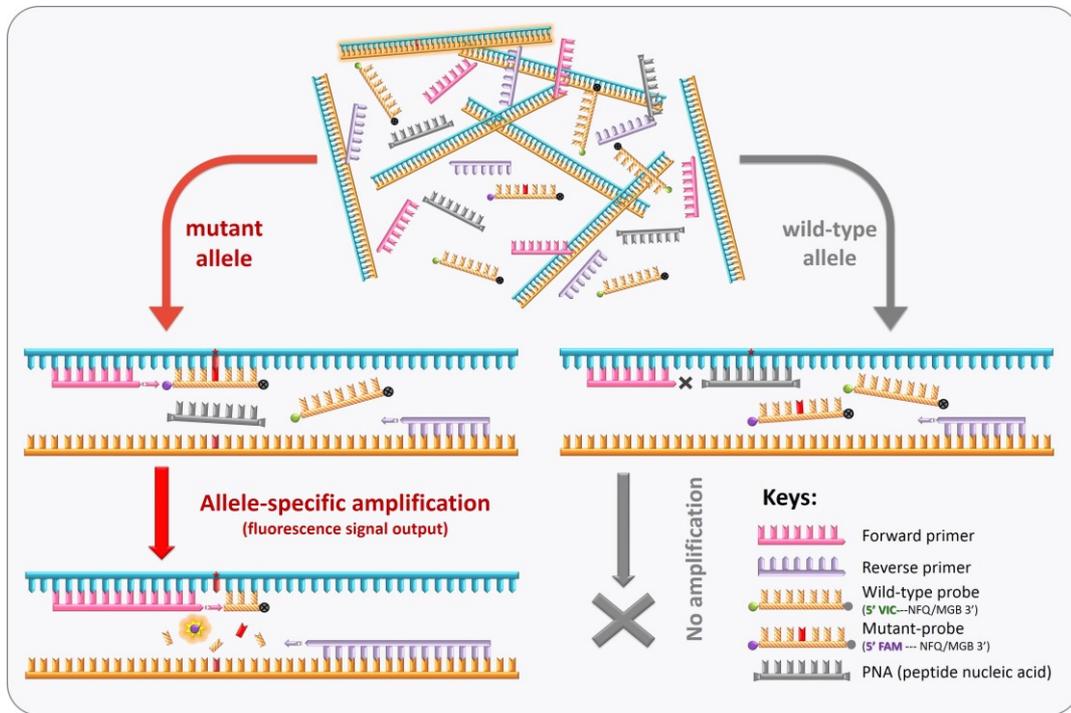


Figure 1-22: Mutation Detection qPCR Assay with PNA Clamping

An assay set-up for detection of low frequency mutations by selective PNA clamping. Primers are shared between the two alleles. Probes specific for WT and MU alleles are labelled with two distinct fluorophores (VIC/FAM) at 5' end for dual signal output. The PNA clamp anneals to the WT allele and inhibits its amplification, facilitating the preferential amplification of few MU alleles.

In this project, PNA oligomers were designed to anneal to the WT sequences that encompass mutated bases – they act as WT-specific blockers to enhance MU detection [332]. In the absence of a PNA blocker, excess WT background will be preferentially amplified, compromising the amplification of few MU alleles that become increasingly concealed by exponential accumulation of WT copies with each round of the PCR extension stage. When a PNA is present, it will bind to the WT allele – as PNA cannot be digested by polymerases, it prevents WT amplification whilst enabling preferential amplification of MU alleles that exist at a much lower fraction in a reaction well (Figure 1-22) [255, 333, 334]. By running reactions with and without a PNA blocker for each sample, MU to WT ratio for a given sample can be measured. This high-affinity

sequence-specific interaction of PNA was first proven to be sufficient in discriminating three different base mutations at a single locus by Stanley's group in 1993, and since then many other groups have utilised this PNA-based PCR applications to detect point mutations (e.g. *KRAS*) specifically and sensitively [255, 333, 335, 336].

1.8.2 Droplet Digital Polymerase Chain Reaction

Amongst currently available methods (Figure 1-18), the droplet digital PCR (ddPCR) system has been demonstrated to be an ideal system for high resolution analysis of samples that contain very low fractional abundance of targets (<1%) [252, 253, 302]. Vogelstein & Kinzler (1999) first used this ddPCR approach to detect *KRAS* mutations in the stool of CRC patients and many technological refinements since then have allowed this ddPCR system to detect MU alleles with remarkable LoDs down to 0.01% [305, 306, 308, 337, 338].

For this project, the QX200™ droplet reader system from BioRad was employed for much of the sample analysis. Using the QX200™ droplet generator, the reaction volume (20µl) is partitioned into evenly sized 15,000-20,000 emulsion droplets of approximately 1nl in volume using the proprietary microfluidics technology (Figure 1-23). To avoid signal saturation, the recommended maximum DNA quantity input is 110,000 copies/well, equivalent to ~360ng/well, at which 99.5% saturation of droplets is achieved [339, 340]. Each ddPCR droplet would either contain none, one or a few DNA copies – be it WT, MU or a mixture of both – and functions as an independent PCR reaction, essentially enabling many PCR reactions to run from a single reaction volume [341]. After droplet generation, a PCR is performed to the end-point using a conventional thermal cycler (Figure 1-23). Following this, fluorescence of individual droplets is read using the proprietary QX200™ droplet reader system with dual channel fluorescence detection: FAM or EvaGreen in one channel, and HEX or VIC in the other [339]. The properties of FAM and VIC are as previously described (section 1.8.1). HEX (hexachloro fluorescein) (absorbance maximum= 535 nm; emission maximum= 556 nm) is a reporter dye commonly used as a substitute for VIC as they share similar fluorescence profiles [342], and EvaGreen is a double-strand intercalating dye for non-probe-based assay designs (BioRad®) [339, 343].

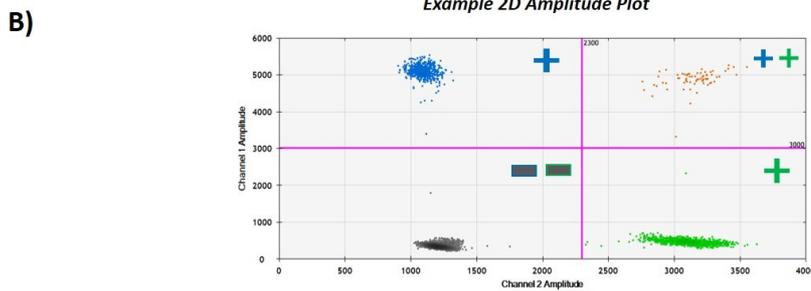
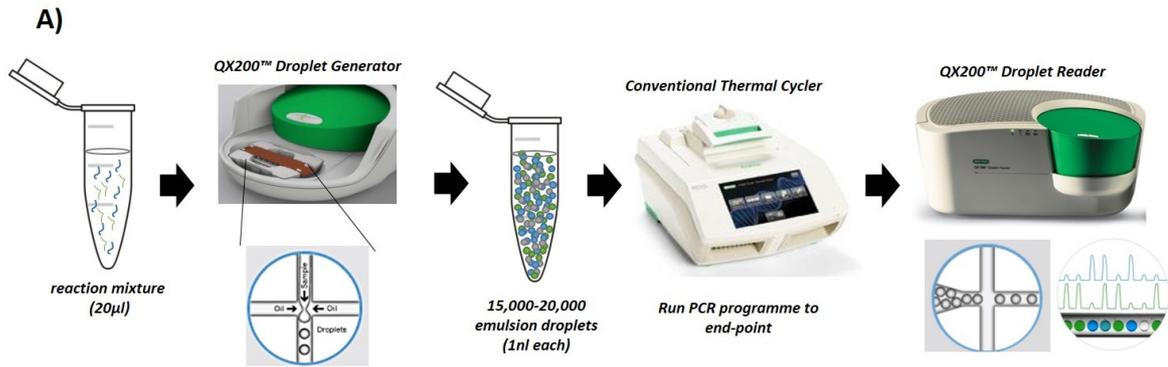


Figure 1-23: Workflow for a ddPCR Assay

A) A ddPCR workflow from sample preparation to fluorescence data collection. [310]

B) An example 2D amplitude plot of fluorescence output, showing 4 possible populations: channel 1 (e.g. FAM – blue plus sign) positive, channel 2 (e.g. VIC/HEX – green plus sign) positive, both positive (blue and green plus signs) and background (double negative, two minus signs). [310]

Much like the qPCR system, a ddPCR assay can be designed to detect two different targets (i.e. WT vs. MU) in parallel using TaqMan fluorescent probes. Reaction droplets with two reporter dyes have four possible outcomes: a droplet emitting fluorescence specific for target 1, a droplet emitting fluorescence specific for target 2, a droplet emitting both fluorescent wavelengths (i.e. both targets) and a droplet emitting background fluorescence (i.e. null) (Figure 1-23-B) [339, 343]. Either manual or automatic thresholding decides the positive or negative fluorescence readings within each well. Data analysis is then carried out by the proprietary software QuantaSoft (BioRad), which collects the fluorescence amplitudes from each droplet and fractionates the positive and negative droplets according to the Poisson distribution algorithm to account for droplets that contain more than one target copies [339, 341, 343]. For the

absolute quantification of targets, the software uses this formula – copies per droplet = $-\ln(1-p)$; where p = fraction of positive droplets [339].

The ddPCR offers several advantages over the qPCR system. Absolute quantification by ddPCR assays eliminates the need for standard curve analysis, unlike the qPCR system, and removes the bias introduced by quantitative references [339, 341, 344-346]. A qPCR quantitative analysis relies on constructing a standard curve using endogenous controls (e.g. *GAPDH*, *ALU*, *LINE-1*) as references, and therefore its accuracy depends on how precisely those controls are quantified prior to analysis [339, 341, 345]. These controls are usually quantified using UV spectrophotometry (e.g. NanoDrop spectrophotometer) and this can introduce bias as it cannot distinguish between single- and double-stranded DNA [339, 341, 345]. Moreover, partitioning ddPCR reaction volumes into droplets essentially removes the excess WT background in each droplet via distribution, which enhances the detectability of few MU alleles in a sample [339]. Droplet partitioning also removes the need for background (WT) blockers such as PNAs, making it more time- and cost-efficient for assay development [339, 341].

1.9 Hypothesis and Aims

Early detection has a profound effect on cancer mortality where early stage diagnosis of CRCs is associated with >90% 5-year survival rate compared to <10% when diagnosed at advanced stages [18]. Early detection of CRC also helps to facilitate chemopreventive interventions, which have the potential to significantly reduce mortality from CRC (section 1.6). However, preventive strategies are often hindered by the lack of surrogate biomarkers that can assess the efficacy of treatments. In this regard, non-invasive biomarkers, such as tumour-specific cfDNA, can potentially fulfil this unmet clinical need. For instance, plasma cfDNA analysis of driver mutations (e.g. *BRAF*) can be used to monitor treatment response in high-risk individuals who are undergoing chemopreventive treatments (e.g. regular doses of aspirin). Plasma cfDNA analysis could also be implemented as a complementary procedure to currently available diagnostic methods as a way of improving detection sensitivity and specificity. However, before cfDNA can be implemented as a biomarker for early detection, there

are several hurdles to overcome from standardisation of sample processing to technical optimisation to circumvent low MU fractions associated with early stage samples.

In this regard, this project was undertaken to fulfil the following three aims:

- To investigate cfDNA as a surrogate biomarker for disease progression in the *Cre-Apc^{f/f}* model of CRC, in which adenoma development is driven by loss of Apc function. This model allows the disease onset to be controlled to investigate the earliest detectable time-point for plasma cfDNA in relation to the adenoma development, and whether detection of cfDNA depends on the tumour burden. As it is a widely used pre-clinical model of CRC, a cfDNA surrogate biomarker, if validated, has potential uses beyond this project, especially for future preventive (or therapeutic) pre-clinical testing. (Chapters 3 & 4)
- To assess whether cfDNA is a suitable biomarker for identification of patients with precancerous lesions. For this, samples from a patient cohort (n=76) and a 'polyp-free' control cohort (n=37) have been chosen for analysis. Quantitative analysis of total cfDNA levels will be carried out to assess whether elevated cfDNA levels are associated with adenoma development. Optimised mutation detection assays on qPCR and ddPCR platforms will also be used to detect tumour-specific *KRAS* and *BRAF* mutations, chosen for their frequent implications in CRCs, in the matched tissue and plasma samples from adenoma patients. (Chapters 5 & 6)
- To assess the heterogeneity of early lesions and whether it affects the detectability of tumour-derived cfDNA in the plasma. Multi-regional sequencing will be performed for 6 patient cases with large adenomas to check for the presence of ITH in precancerous lesions. Additionally, to assess whether targeting mutations that were present at higher frequencies could improve the sensitivity of detection in the precancerous samples. (Chapters 5 & 6)

Chapter 2 Materials & Methods

2.1 Materials

2.1.1 Cell Culture

Table 2-1: List of Media & Buffers used for Cell Culture

<i>Name</i>	<i>Properties</i>	<i>Manufacturer</i>	<i>Purpose</i>
RPMI 1640	Liquid media with 300mg/l L-Glutamine and 5mg/l phenol red	Gibco, UK	Cell culture – A549 (lung carcinoma)
Dulbecco's Modified Eagle's Medium (DMEM)	Modified with 4500mg/L glucose, 0.584g/l L-glutamine, 110mg/l sodium pyruvate and 0.0159g/l phenol red	Sigma, UK	Cell culture – SW480 & HT29 (both colorectal adenocarcinoma)
McCoy's 5A (modified)	With 3000mg/l glucose, 600mg/l bacto-peptone, 219.2mg/l L-glutamine and phenol red	Gibco, UK	Cell culture – HCT116 (colorectal adenocarcinoma)
Foetal bovine serum (FBS)	Heat-inactivated, ≤10 EU/ml endotoxin level, ≤25 mg/dl haemoglobin level	Gibco, USA	Cell culture
GlutaMAX™	200 mM L-alanyl-L-glutamine dipeptide in 0.85% NaCl, pH range (4.7-6.0)	Gibco, UK	Cell culture
Trypsin-EDTA (TE)	0.5% 10x buffer with 8500mg/l NaCl, 2000mg/l EDTA and 5000mg/l Trypsin	Gibco, USA	Cell culture
Phosphate Buffered saline (PBS) Dulbecco A	Tablets form containing 8.0g/l NaCl, 0.2g/l KCl, 1.15g/l Na ₂ HPO ₄ 1.15g/l and 0.2g/l KH ₂ PO ₄ , pH 7.3±0.2	Oxoid, UK	Cell culture

2.1.2 DNA Extraction

Table 2-2: Miscellaneous Buffers & Solutions used in DNA Extraction

<i>Name</i>	<i>Properties</i>	<i>Manufacturer</i>	<i>Purpose</i>
Phosphate Buffered saline (PBS) Dulbecco A	Tablets form containing 8.0g/l NaCl, 0.2g/l KCl, 1.15g/l Na ₂ HPO ₄ 1.15g/l and 0.2g/l KH ₂ PO ₄ , pH 7.3±0.2	Oxoid, UK	DNA Extraction – general use
RNase A	100 mg/ml	Qiagen, Germany	DNA extraction – RNA degradation
Isopropanol	≥99.9 %, molecular biology grade	Fisher Scientific, UK	DNA extraction – only used with mouse ear tissues for genotyping
Absolute ethanol	≥99.9 %, molecular biology grade	Fisher Scientific, UK	DNA extraction – DNA precipitation
Xylene	≥99.0%, molecular biology grade	Fisher Scientific, UK	DNA extraction – dewaxing FFPE sections
Industrial methylated spirit (IMS)	99% denatured alcohol	Genta Medical, UK	DNA extraction – dewaxing FFPE sections

Table 2-3 listed the buffers included in DNA Clean & Concentrator™-5 kit (Zymo Research®, USA) that was used to purify PCR DNA products in section 2.2.3.2.4.

Table 2-3: List of Buffers included in DNA Clean & Concentrator Kit

<i>Name</i>	<i>Properties</i>
DNA binding buffer	<i>Proprietary of Manufacturer</i>
DNA wash buffer	<i>Proprietary of Manufacturer – contains ethanol</i>
DNA elution buffer	10mM Tris-HCL, 0.1mM EDTA, pH 8.5

Table 2-4 lists the buffers included in various DNA extraction kits. Number assignments are: 1 = Genra Puregene kit, 2 = DNeasy® Blood & Tissue Kit, 3 = QIAamp® DNA Blood Mini Kit, 4 = QIAamp® Fast DNA Stool Mini Kit, 5 = QIAamp® Circulating Nucleic Acid kit, 6 = GeneRead™ DNA FFPE kit. All kits were manufactured by Qiagen, Germany.

Table 2-4: Lists of Buffers included in Qiagen DNA Extraction Kits

Name	Properties	Kit No.	Purpose
Cell lysis solution	<i>Proprietary of Manufacturer</i> – an anionic detergent with DNA stabiliser	1	DNA extraction from mouse ear tissues for genotyping
Protein precipitation solution	<i>Proprietary of Manufacturer</i> – high salt buffer to reduce solubility of proteins	1	Isolation of proteins during DNA extraction
DNA hydration solution	1mM EDTA, 10mM Tris-Cl, pH 7.5	1	Resuspension of isolated DNA
Proteinase K	Subtilisin-type protease with high specificity, 20 mg/ml	2, 4, 5	Digestion of proteins during Various DNA extraction processes
Qiagen protease	Broad-spectrum serine protease	3	DNA extraction from murine plasma
Buffer AL	<i>Proprietary of Manufacturer</i> - guanidine-based lysis buffer	2, 3, 4, 6	Various DNA extraction processes
Buffer AW1	<i>Proprietary of Manufacturer</i> –ethanol-based wash solution	2, 3, 4, 6	Various DNA extraction processes
Buffer AW2	<i>Proprietary of Manufacturer</i> –Tris-based wash solution	2, 3, 4, 6	Various DNA extraction processes
Buffer AE	10mM Tris-HCl, 0.5mM EDTA, pH 9.0	2, 3	DNA extraction from cell pellets, murine plasma & tissues
Buffer ATL	<i>Proprietary of Manufacturer</i> - tissue lysis buffer with SDS	3, 6	DNA extraction from murine tissues & patient FFPE samples
InhibitEX® buffer	<i>Proprietary of Manufacturer</i> – lysis buffer to remove PCR inhibitors	4	DNA extraction from murine faeces
Buffer ATE	10mM Tris-HCl (pH 8.3), 0.1mM EDTA, 0.04% sodium azide (NaN ₃)	4, 6	DNA extraction from murine tissues & patient FFPE samples
Buffer ACL	<i>Proprietary of Manufacturer</i> –lysis buffer with guanidine salts	5	cfDNA extraction from patient plasma samples
Buffer ACB	<i>Proprietary of Manufacturer</i> – binding buffer with guanidine salts	5	cfDNA extraction from patient plasma samples
Buffer ACW1	<i>Proprietary of Manufacturer</i> – wash buffer with guanidine salts	5	cfDNA extraction from patient plasma samples
Buffer ACW2	<i>Proprietary of Manufacturer</i> – wash buffer for membranes	5	cfDNA extraction from patient plasma samples
Buffer AVE	<i>Proprietary of Manufacturer</i> – contains 0.04% sodium azide (NaN ₃)	5	cfDNA extraction from patient plasma samples

2.1.3 Traditional End-point PCR

Table 2-5: List of Reagents for Traditional End-point PCR Assays

<i>Name</i>	<i>Properties</i>	<i>Manufacturer</i>	<i>Purpose</i>
10x PCR Run Buffer (-MgCl ₂)	200mM Tris-HCl, pH 8.4, 500mM KCl	Invitrogen, USA	PCR mastermix
MgCl ₂	50mM solution in high-purity water	Invitrogen, USA	PCR mastermix
dNTP Mix (2.5mM)	2.5 mM of each dNTP (dATP, dCTP, dGTP, and dTTP) in highly purified water (pH 8.0)	Invitrogen, USA	PCR mastermix
<i>Taq</i> DNA Polymerase Recombinant	5 U/μl of enzyme (94 kDa): 5'→3' DNA polymerase activity & 5'→3' exonuclease activity	Invitrogen, USA	PCR mastermix
Dimethyl sulfoxide (DMSO)	≥99.5%, assay grade	Sigma, Japan	PCR mastermix
Agarose	DNase/RNase free, molecular biology grade	Appleton, UK	Gel electrophoresis
Tris Borate, EDTA (TBE) buffer	10x (89mM Tris, 89mM Borate, 2mM EDTA), pH 8.0-8.5	Thermo Scientific, USA	Gel electrophoresis
UltraPure™ Ethidium Bromide	10mg/ml solution in high-purity water	Invitrogen, New Zealand	Gel electrophoresis
Gel loading dye (blue)	6x (2.5% Ficoll®-400, 11mM EDTA, 3.3mM Tris-HCL, 0.017% SDS, 0.015% bromophenol blue, pH 8.0)	New England Biolabs, UK	Gel electrophoresis
TrackIt™ 100bp DNA Ladder	0.10μg/μl ladder in 10 mM Tris-HCl (pH 7.5), 10 mM EDTA (pH 8.0), 0.06% XCFE, 0.6% tartrazine, and 5% glycerol	Invitrogen, USA	Gel electrophoresis

Table 2-6 lists primers used in the traditional PCR assays (2.2.3.2). All primers were manufactured by Sigma, UK.

Table 2-6: Primers for Traditional PCR Assays

<i>Name</i>	<i>Type</i>	<i>Sequence (5'-3')</i>	<i>Amplicon size (bp)</i>	<i>Assay</i>	<i>Target</i>
Apc-P3	FP	GTTCTGTATCATGGAAAGATAGGTGGTC	308 (<i>Apc^{fl/fl}</i>); 226 (<i>Apc^{WT}</i>)	Apc-lox1	Upstream (1 st) <i>loxP</i> region in the <i>Apc^{fl/fl}</i> allele
Apc-P4	RP	CACTCAAACGCTTTTGAGGGTTG			
PGK-F	FP	CATTCTGCACGCTTCAAAG	1,596	Apc-lox2	Downstream (2 nd) <i>loxP</i> region in the <i>Apc^{fl/fl}</i> allele
P5-2	RP	GAGTACGGGGTCTCTGTCTCAG			
P5-3F	FP	TCGCCTTCTTGACGAGTTCT	736		
P5-3R	RP	GAGTACGGGGTCTCTGTCTC			

2.1.4 qPCR & ddPCR

Table 2-7: List of Reagents for qPCR & ddPCR Assays

Name	Properties	Manufacturer	Purpose
Mouse genomic DNA (mgDNA)	100µg in 10mM Tris-HCl (pH 8.0), 1mM EDTA	Promega, USA	Control for PCR assays
Human genomic DNA (hgDNA)	100µg in 10mM Tris-HCl (pH 8.0), 1mM EDTA	Roche, Germany	Control for PCR assays
Fast SYBR® Green Master Mix	SYBR® Green 1 Dye, AmpliTaq Gold® Fast DNA Polymerase LD, dNTPs (dUTP/dTTP blend), Passive Reference 1 and optimized buffer components	Applied Biosystems, Lithuania	qPCR master mix – testing primers
TaqMan Genotyping PCR Master Mix	2x AmpliTaq Gold® DNA Polymerase UP (Ultra Pure), dNTPs without dUTP, Passive Reference ROX dye and optimized mix components	Applied Biosystems, USA	qPCR master mix - DNA quantification & mutation detection
ddPCR™ Supermix for Probes (no dUTP)	<i>Proprietary of Manufacturer</i>	BioRad, USA	ddPCR master mix
ddPCR™ Droplet Generation Oil for Probes	<i>Proprietary of Manufacturer</i>	BioRad, USA	ddPCR – droplet generation
ddPCR™ Droplet Reader Oil	<i>Proprietary of Manufacturer</i>	BioRad, USA	ddPCR – detection of fluorescence signals

For Table 2-8 and Table 2-9, all primers were manufactured by Sigma (UK) and all TaqMan™ probes were from Applied Biosystems (UK). 5' reporter dyes for probes are shown in brackets. PNA blockers were from Eurogentec, Belgium. KRAS-G13D assay mixes were from BioRad, USA. Abbreviations: FP= forward primer, RP= reverse primer, WTP= wild-type probe, MUP= mutant probe, PNA= peptide nucleic acid, asterisk (*) = STOP codon, D (degenerate base) = A/G/T (IUPAC code).

Table 2-8: Primers & Probes for Analysis of *In Vivo* Samples

Name	Type	Sequence (5'-3')	Amplicon size (bp)	Assay	Target
mGapdh-F	FP	GTCGTGGATCTGACGTGCC	70	mGapdh qPCR	30 <i>mGapdh</i> pseudogenes for DNA quantification
mGapdh-R	RP	CCTGCTTACCACCTTCTTGA			
mGapdh-P	WTP	ACCTGCCAAGTATGATGA (VIC)			
Apcf1-102NF1	FP	TAAGGGCTAACAGTCAATATAATGC	102	Apc-Rec qPCR & ddPCR	Recombined <i>loxP</i> site in <i>Cre-Apc^{fl/fl}</i> mice
Apcf1-102NR2	RP	CATTAGTTTAATCCTGTGTTGATC			
Apcf1-P1	MUP	CTAGTGGATCCGATAACTTCG (FAM)			
Apcf1-NRF1	FP	CGATGATCTCGTCGTGACCC	75	Apc-NR ddPCR	Non-recombined <i>Apc^{fl/fl}</i> allele in <i>Cre-Apc^{fl/fl}</i> mice
Apcf1-NRR1	RP	AATCCAGAAAAGCGGCCATT			
Apcf1-NRP1	WTP	CTTGCCGAATATCA (VIC)			

Table 2-9: Primers & Probes for Analysis of Patient Samples

Name	Type	Sequence (5'-3')	Amplicon size (bp)	Assay	Purpose
hALU-F	FP	GACCATCCCGGCTAAAACG	69	hALU69 qPCR	Quantification of patient FFPE samples
hALU-69R	RP	CCACTACGCCCGGCTAATTT			
hALU-P	WTP	CCCCGTCTCTACTAAA (VIC)			
hGAPDH-F	FP	GGCTAGCTGGCCCCGATTT	95	hGAPDH qPCR	Quantification of patient plasma samples
hGAPDH-R	RP	GGACACAAGAGGACCTCCATAAA			
hGAPDH-P	WTP	ATGCTTTTCTAGATTATTC (FAM)			
BRAF-600F	FP	TCATGAAGACCTCACAGTAAAAATAGGT	94	BRAF-V600E qPCR/ddPCR	Detects missense <i>BRAF</i> mutation at c.1799T>A
BRAF-600R	RP	ATCCAGACAAGTGTCAAAGTATG			
BRAF-WTP	WTP	CTAGCTACAGTGAAATC (VIC)			
BRAF-MUP	MUP	TAGCTACAGAGAAATC (FAM)			
BRAF-PNA	PNA	TAGCTACAGTGAAATC			
KRAS-78F	FP	AGGCCTGCTGAAAATGACTGA	78	KRAS-121 or KRAS-122 qPCR/ddPCR	Detects missense <i>KRAS</i> mutations: KRAS-121 (c.34G>A/T/C) & KRAS-122 (c.35G>A/T/C)
KRAS-78R	RP	TGTATCGTCAAGGCACTCTTGC			
KRAS-WTP	WTP	CTACGCCACCAGCTC (VIC)			
KRAS-121P	MUP	TACGCCACDAGCTC (FAM)			
KRAS-122P	MUP	TACGCCADCAGCTC (FAM)			
KRAS-PNA	PNA	TACGCCACCAGCTCC			
KRAS-G13D WT mix	N/A	Proprietary of Manufacturer – assay ID dHsaCP2000014	90	KRAS-G13D ddPCR	Detects missense <i>KRAS</i> mutation at c.38G>A
KRAS-G13D MU mix	N/A	Proprietary of Manufacturer – assay ID dHsaCP2000013			
APC1338-F2	FP	TCACAGCACCTAGAACCAAATC	68	APC1338 ddPCR	Detects nonsense <i>APC</i> mutation at c.4012C>T/p.Q1338*
APC1338-R2	RP	TGCCTGGCTGATTCTGAAGATAA			
APC1338-WTP	WTP	AGACTGCAGGGTTC (VIC)			
APC1338-MUP	MUP	CAGACTGTAGGGTTCT (FAM)			
APC1397-F1	FP	CATGTTTAGCAGATGTACTTCTGTCACT	86	APC1397 ddPCR	Detects nonsense <i>APC</i> mutation at c.4189G>T/p.E1397*
APC1397-R2	RP	TGGTTCACTCTGAACGGAGCT			
APC1397-WTP	WTP	ATAGTTTTGAGAGTCGTTT (VIC)			
APC1397-MUP	MUP	TGATAGTTTTAGAGTCGTTT (FAM)			
APC1429-F1	FP	CCCAGTGATCTTCCAGAT	61	APC1429 ddPCR	Detects nonsense <i>APC</i> mutation at c.4285C>T/p.Q1429*
APC1429-R1	RP	AGGTGTTTTACTTCTGCTTGGT			
APC1429-WTP	WTP	CCTGGACAAACCAT (VIC)			
APC1429-MUP	MUP	CCCTGGATAAACCAT (FAM)			
TP53-273F2	FP	CAGTGGAATCTACTGGGACGGA	67	TP53-273 ddPCR	Detects missense <i>TP53</i> mutation at c.817C>T/p.R273C
TP53-273R1	RP	CCGGTCTCTCCAGGACA			
TP53-273WTP	WTP	TGAGGTGCGTGTTT (VIC)			
TP53-273MUP	MUP	TGAGGTGTGTGTTTGT (FAM)			
PIK1047-F2	FP	GCTTTGGAGTATTTTCATGAAACAAAT	77	PIK1047 ddPCR	Detects missense <i>PIK3CA</i> mutation at c.3140A>G/p.H1047R
PIK1047-R2	RP	TGGAAGATCCAATCCATTTTTGT			
PIK1047-WTP	WTP	TGATGCACATCATGGT (VIC)			
PIK1047-MUP	MUP	TGATGCACGTCATGGT (FAM)			

2.1.5 *In Vivo* Studies

Table 2-10: List of Reagents & Solutions for *In Vivo* Studies

<i>Name</i>	<i>Properties</i>	<i>Manufacturer</i>	<i>Purpose</i>
Tamoxifen	≥99%, in powder form	Sigma, China	To induce <i>loxP</i> recombination in <i>Cre-Apc^{f/f}</i> mice
Absolute ethanol	≥99.9 %, molecular biology grade	Fisher Scientific, UK	Preparing tamoxifen solution
Sunflower seed oil	oil from <i>Helianthus annuus</i> , 0.92 g/ml	Sigma, China	Vehicle control to tamoxifen injection in <i>Cre-Apc^{f/f}</i> mice
Formal saline	Formaldehyde: 10% (3.8-4% v/v), NaCl	Atom Scientific, UK	Tissue fixative
Isoflurane	Liquid form, 100% w/w inhalation vapour	Zoetis, USA	Terminal anaesthesia
EURodent Diet 14% (5LF2*)	Calories provided: 16.79% protein, 6.61% fat (ether extract) & 76.6% carbohydrates	LabDiet, USA	Standard mouse diet used in the 1 st & 2 nd <i>in vivo</i> studies
AIN-93G Growth Purified Diet (#5801-G)	Calories provided: 18.8% protein, 16.4% fat (ether extract) & 65.1% carbohydrates	TestDiet, USA	Provided as normal fat diet (NFD) to mice in the 3 rd study
Modified AIN-93G Diet (5SGM)	Calories provided: 13.6% protein, 59.7% fat (ether extract – coconut oil) & 26.7% carbohydrates	TestDiet, USA	Provided as normal fat diet (HFD) to mice in the 3 rd study

2.1.6 Immunohistochemistry

Table 2-11: List of Buffers & Solutions for Immunohistochemistry

<i>Name</i>	<i>Properties</i>	<i>Manufacturer</i>	<i>Purpose</i>
Phosphate Buffered saline (PBS) Dulbecco A	Tablets form containing 8.0g/l NaCl, 0.2g/l KCl, 1.15g/l Na ₂ HPO ₄ 1.15g/l and 0.2g/l KH ₂ PO ₄ , pH 7.3±0.2	Oxoid, UK	Washing steps, diluent for antibodies
Xylene	≥99.0%, molecular biology grade	Fisher Scientific, UK	Dewaxing & rehydration of FFPE sections
Industrial methylated spirit (IMS)	99% denatured alcohol	Genta Medical, UK	Dewaxing & rehydration of FFPE sections
Purified Mouse Anti-β-Catenin	Monoclonal mouse IgG1, immunogen mouse β-catenin (aa. 571-781): 250 µg/ml in BSA, glycerol, and ≤0.09% sodium azide	BD Biosciences, UK	Primary antibody for β-catenin staining
Mouse IgG1 X0931	Isotype – IgG1, kappa, in 0.05 mol/l Tris-HCL (pH 7.2), 15 mmol/l sodium azide	Dako, Denmark	Negative control for β-catenin staining
Citrate monophosphate	≥98%, reagent grade, powder form	Sigma, Austria	Antigen retrieval
Avidin blocking solution	0.1% avidin in Tris-HCl buffered saline, 0.015 mol/L sodium azide	Dako, USA	Inhibit non-specific, endogenous biotin activity
Biotin blocking solution	0.01% unconjugated biotin in Tris-HCl buffer, 0.015 mol/L sodium azide	Dako, USA	Inhibit non-specific, endogenous biotin activity
Hydrogen peroxide	Supplied as 30 % (w/w) in H ₂ O, contains stabiliser	Sigma, USA	Inhibit endogenous peroxidase activity
Bovine Serum Albumin (BSA)	≥96%, gel electrophoresis grade, lyophilised powder	Sigma, USA	Added to primary antibody solution to reduce background
Triton™ X-100 (BioXtra)	Molecular biology grade	Sigma, USA	Used as a non-ionic surfactant in primary antibody solution
Biotinylated link	Biotin labelled, affinity isolated goat anti-mouse/rabbit Ig in PBS, 15 mmol/l sodium azide	Dako, USA	Binds primary antibody to strengthen staining
Streptavidin-HRP	Streptavidin conjugated to horseradish peroxidase in PBS, stabilising protein & anti-microbial agents	Dako, USA	Binds biotinylated link for specific staining
DAB chromogen	1.74% w/v 3,3'-diaminobenzidine	Novolink, UK	Antibody staining
DAB Substrate Buffer (polymer)	Buffered solution containing ≤0.1% hydrogen peroxide and preservative	Novolink, UK	Antibody staining
Haematoxylin	<0.1% haematoxylin	Novolink, UK	Counterstaining
DPX	A mixture of toluene, xylene and dibutyl phthalate	Sigma, Spain	Mountant for FFPE sections

2.1.7 Patients

All patient samples were collected as part of the national Bowel Cancer Screening Programme (BCSP). Colonoscopy, surgical resection and blood sample collection were carried out at the Glenfield Hospital, Leicester, UK. All patients consented for their biological samples to be used in the study 'Biomarkers for Bowel Disease (2011)' (UHL study 11005; NREC 10/H0408/116). For patients who underwent surgical resection, tissue samples were collected for the formalin-fixed paraffin-embedded (FFPE) histology archive. Venous blood samples (15-20ml) were collected 1-4 hours prior to colonoscopy or surgical resection, and processed within 2 hours of collection by the trained staff at Glenfield Hospital. Blood was fractionated via centrifugation and samples were stored at -80°C until use.

Out of 131 patient cases, 76 patient cases (58 males, 18 females) with precancerous lesions were selected for analysis based on sample availability. The age range was 57-87 years (mean = 67.5 years). Most samples are small (≤ 20 mm diameter) and low-grade, and almost three-quarter of cases were from left-sided colon (sigmoid & rectal regions) – details will be discussed in Chapter 6 (section 6.2). For cases with multiple FFPE blocks, all available samples were included for analysis. Cases with mutation-positive FFPE samples were subsequently selected for plasma analysis.

The control group consisted of 37 individuals who participated in the programme but showed no clinically observable polyps following false-positive FOBT results. As such, no FFPE tissue samples were available, and only blood samples were collected. 22 were males and 15 were females with the age range of 60-74 years (average= 64.8 years). Additional diagnosis was noted for 8 individuals in the control group: 4 had diverticulosis, 3 had ulcerative colitis and 1 had angioectasia (appendix 8.1).

2.2 Methods

2.2.1 Cell Culture

DNA extracted from various cell lines (Table 2-12) were used as positive controls for mutation detection assays.

Table 2-12: Cell Lines used for extracting positive control DNA

Name	Cancer Type	Growth Conditions	Mutations
A549	Lung carcinoma	RPMI 1640, 10% FCS, 1% GlutaMAX	<i>KRAS</i> (c.34G>A/p.G12S)
SW480	Colorectal adenocarcinoma	DMEM, 10% FCS, 1% GlutaMAX	<i>KRAS</i> (c.35G>T/p.G12V)
HCT116	Colorectal adenocarcinoma	McCoy's 5A, 10% FCS	<i>KRAS</i> (c.38G>A/p.G13D)
HT29	Colorectal adenocarcinoma	DMEM, 15% FCS, 1% GlutaMAX	<i>BRAF</i> (c.1799T>A/p.V600E)
SK-MEL-28	Malignant melanoma	N/A (<i>not cultured, only used the DNA extracts</i>)	<i>BRAF</i> (c.1799T>A/p.V600E)

All culturing was performed in a sterilised Class II Microbiology Safety Cabinet. Media and solutions used were listed in Table 2-1. Cell aliquots were stored in liquid nitrogen prior to resuscitation by snap-thawing in a 37°C water bath, and re-suspended in appropriate media. Cells were then pelleted by centrifugation at 1,200 rpm for 6 minutes, re-suspended in media and plated. After plating, cells were incubated until >70% confluency. For passaging, cells were washed with PBS before 10% TE digestion. Once cells have detached, TE was deactivated by adding FCS-supplemented media. Cells were pelleted by centrifugation at 1,200rpm for 6 minutes and re-suspended in media. Cells were either passaged further or stored as pellets at -20°C for DNA extraction.

2.2.2 Protocols for DNA Extraction & Processing

The following protocols were implemented to process different types of murine and patient samples (fresh tissue, FFPE and plasma) for DNA extraction. All the reagents and kits used were listed in Table 2-2 and Table 2-4. All sample aliquots were either stored at 4°C (short-term) or at -20°C (long-term).

2.2.2.1 DNA Extraction from Murine Samples

2.2.2.1.1 DNA Extraction from Murine Ear Tissues

Ear tissues were lysed in 300µl of cell lysis solution by vortexing, and then 1.5µl of Proteinase K solution was added to digest proteins by incubating at 55°C overnight. Following this, 100µl of protein precipitation solution was added and protein pellets were precipitated by centrifugation at 13,000rpm for 3 minutes. Supernatant was recovered and 300µl 100% Isopropanol was added before centrifugation at 13,000rpm for 3 minutes. Supernatant was decanted and 300µl 70% ethanol was added to wash DNA pellets by mixing gently, and centrifuged again at 13,000rpm for 3 minutes. Supernatant was removed and DNA pellets were air-dried for 10-15 minutes. Afterwards, 50µl DNA Hydration solution was added to sample pellets and incubated for 1 hour at 65°C.

2.2.2.1.2 DNA Extraction from Murine Blood Plasma

QIAmp® DNA Blood Mini Kit (Qiagen®) was used for this protocol. Blood collection and processing was described in 2.2.5.5. If a plasma volume was <200µl, PBS was added to reach 200µl. Proteins were digested by adding Qiagen protease (10µl per 100µl plasma) and vortexed. Buffer AL was added in 1:1 ratio and incubated at 56°C for 20 minutes. After incubation, 100% ethanol was added in 1:1 ratio and vortexed. The solution mixture was passed through the QIAmp silica-gel membrane columns by centrifugation at 6,000g for 1 minute. Columns were washed with 500µl Buffer AW1 by centrifuging at 6,000g for 1 minute to remove non-specific binding. This wash step was repeated with 500µl of Buffer AW2 by centrifuging at 20,000g for 3 minutes. Columns were incubated in 70µl Buffer AE and DNA was eluted by centrifuging at 6,000g for 1 minute. The step with Buffer AE was repeated once more to increase the yield.

2.2.2.1.3 DNA Extraction from Murine Fresh & Frozen tissues

DNA was extracted using the DNeasy® Blood & Tissue Kit (Qiagen®). Frozen tissues stored at either -20°C or -80°C were left to equilibrate to room temperature, and

were cut into small pieces (maximum weight = 25mg). 180µl buffer ATL and 20µl Proteinase K were added to the sample and incubated at 56°C overnight, or longer if necessary, to achieve full digestion. Samples were left to cool and 4µl 100mg/ml RNase A was added. Following this, 400µl each of buffer AL and 100% ethanol were added. Samples were loaded onto the DNeasy membrane columns. Column washing and eluting steps were as described in section 2.2.2.1.2.

2.2.2.1.4 DNA Extraction from Murine Faecal Samples

Murine faecal samples collection was as described in section 2.2.5.7. DNA was extracted using the QIAamp® Fast DNA Stool Mini Kit (Qiagen®). Faecal samples were weighed and mashed using a small spatula. 1ml InhibitEX® buffer was added and incubated at 70°C for 10 minutes to remove DNA-degrading substances and PCR inhibitors. Insoluble faecal pieces were pelleted by centrifugation at 13,000rpm for 1 minute. The supernatant was mixed with 25µl proteinase K and vortexed. Then, 600µl Buffer AL was added and the mixture was incubated at 70°C for 10 minutes. Afterwards, 600µl of 100% ethanol was added. The solution was filtered through the QIAamp® spin column by centrifuging the columns at 13,000rpm for 1 minute. The spin column was washed with 500µl Buffer AW1 via centrifugation at 13,000rpm for 1 minute, and with 500µl Buffer AW2 via centrifugation at 13,000rpm for 3 minutes. The column membrane was dried by centrifugation at 13,000rpm for 3 minutes, and incubated in 100µl elution Buffer ATE at room temperature for 5 minutes. DNA was eluted via centrifugation at 13,000rpm for 1 minute.

2.2.2.2 DNA Extraction from Human Samples

2.2.2.2.1 DNA Extraction from Cell Pellets

DNeasy® Blood & Tissue kit was used for this protocol. Frozen cell pellets were equilibrated to room temperature for 30 minutes before being re-suspended in 200µl PBS. Then, 20µl Proteinase K and 200µl buffer AL were added and incubated at 56°C for 10 minutes. Mixture was cooled down and 200µl absolute ethanol was added. It was passed through a silica-based DNeasy membrane column by centrifugation at 8,000 rpm

for 1 minute. Column wash steps with buffers AW1 and AW2 were as described in section 2.2.2.1.2. The DNA was eluted from the membrane by incubating the columns in 200µl buffer AE at room temperature for 5 minutes and centrifuging at 8,000rpm for 1 minute.

2.2.2.2.2 cfDNA Extraction from Frozen Patient Plasma Samples

QIAamp® Circulating Nucleic Acid kit was used for this protocol. Plasma samples were stored at -80°C and thawed at room temperature for approximately 1 hour. 2.5-3 ml plasma was used for extraction. The plasma was centrifuged at 1,000g for 5 minutes and the supernatant was mixed with 300µl Proteinase K and 2.4ml Buffer ACL. The mixture was then incubated at 60°C for 30 minutes before mixing with 5.4ml buffer ACB – ACB ensures that DNA fragments were primed for optimal binding to the silica membrane column. Samples were then incubated on ice for 5 minutes. Using a QIAvac 24 Plus vacuum manifold (Qiagen®), samples were poured into the QIAamp mini columns extended with Vac connectors and 20ml capacity funnels. Vacuum was applied to pass the samples through the silica membrane. The membranes were washed with 600µl Buffer ACW1, 750µl Buffer ACW2 and 750µl absolute ethanol sequentially by applying vacuum. Columns were centrifuged at 14,000rpm for 3 minutes before incubating at 56°C for 10 minutes to dry the membranes. Then, 150µl Buffer AVE was added and incubated at room temperature for 3 minutes. DNA was eluted by centrifugation at 14,000rpm for 1 minute into 1.5ml DNA LoBind micro-centrifuge tubes (Eppendorf®).

2.2.2.2.3 DNA extraction from FFPE samples

GeneRead™ DNA FFPE tissue kit was used for this protocol. A minimum of three 4µm FFPE sections were used per sample. Sections were heated at 65°C for 20 minutes before de-waxing in xylene for 3 minutes twice. Sections were re-hydrated by serial immersions in 99% IMS twice followed by 95% IMS once for 1 minute each. Tissues were scraped off the glass slides into an Eppendorf. For complete removal of xylene, 1ml absolute ethanol was added and centrifuged at 13,000rpm for 2 minutes. Supernatant was decanted and tubes were incubated at 37°C for 10 minutes with the lid open to

remove the residual ethanol. Pellets were suspended in 180µl buffer ATL and 20µl proteinase K, and incubated at 56°C overnight or until complete digestion. Then, samples were incubated at 90°C for 1 hour to partially reverse the formaldehyde cross-links. Once cooled down, 4µl 100mg/ml RNase A was added at room temperature for 2 minutes. Following this, 200µl buffer AL and 200µl absolute ethanol were added and the mixture was passed through the QIAamp® MinElute column. The columns were washed in 500µl buffer AW1 and 500µl AW2 sequentially by centrifugation at 8,000 rpm for 1 minute each. Then the membranes were dried by centrifugation at 13,000 rpm for 3 minutes and then incubated in 150µl buffer ATE for 5 minutes at room temperature before eluting.

2.2.2.3 Processing of DNA Samples

2.2.2.3.1 Lyophilisation of Plasma cfDNA Samples

Lyophilisation of plasma cfDNA samples was performed using the VirTis BenchTop Pro Freeze Dryer with Omnitronics™ (SP Scientific). Condenser compartment of the freeze-dryer was cooled down to -65°C for 30 minutes prior to lyophilisation. Starting sample volumes ranged from 40-70µl, and all available sample aliquots after quantification were lyophilised to obtain the highest concentration possible for each sample. In preparation, samples were transferred to 1.5ml Eppendorf® DNA LoBind micro-centrifuge tubes (Sigma®) and the lids were pierced with a needle. Samples were then flash-frozen in liquid nitrogen and lyophilised for approximately 3 hours. Samples were reconstituted in 5-25µl dH₂O and left to stand for 2+ hours at 4°C to ensure complete dissolution.

2.2.3 Polymerase Chain Reaction (PCR)

2.2.3.1 Designing Primers & Probes

The Primer Express (v3.0) software (Applied Biosystems™) was used to design custom primers and probes for various assays. Primers were designed to be 18-26bp in length, 40-60% GC content and within the melting temperature (T_m) range of 58-62°C. The length of the amplicon was limited to <100 base-pairs to achieve efficient detection

of cfDNA fragments. Tentative primer pairs were then analysed with the UCSC *in silico* PCR platform [<https://genome.ucsc.edu/cgi-bin/hgPcr>], using the *GRCh38/hg38* genome build. They were further validated using SYBR Green qPCR (section 2.2.3.3.1).

Custom TaqMan™ probes were designed to be approximately 10°C higher in T_m compared to their primer counterparts, to ensure that probes anneal to target alleles before primer-target hybridisation occurs. Probe sequences were 15-20bp in length with 40-60% GC content, and the last 5 nucleotides at 3' end contained no more than 2 G/C nucleotides to avoid the 'GC clamping' effect that promotes non-specific priming due to strong covalent bonding between G-C bases. For mutation-specific probes, the mutant bases were placed in the middle of the sequence to achieve optimal mismatch discrimination [332].

2.2.3.2 Traditional End-point PCR

2.2.3.2.1 Apc-lox1 PCR to amplify upstream *loxP* region in the *Apc^{fl/fl}* Allele

Apc-lox1 PCR amplifies the upstream *loxP* region in the *Apc^{fl/fl}* allele from *Cre-Apc^{fl/fl}* mice. Each 10µl PCR reaction contained 5.0µl 10xPCR Run Buffer (-MgCl₂), 2.5mM MgCl₂, 20µM dNTP Mix, 0.2µl Taq DNA Polymerase, 400nM of Apc-P3 FP, 400nM of Apc-P4 RP and 39.0µl dH₂O (Table 2-5 & Table 2-6). Each reaction was run with 2.5µl of DNA aliquots extracted from ear tissues (2.2.2.1). The assay was run on the GeneAmp® PCR System 9700 machine using the following programme: 95°C for 3 minutes, 40 cycles of 95°C for 30 seconds, 60°C for 30 seconds and 72°C for 1 minute, and 72°C for 5 minutes. Presence of *Apc^{fl/fl}* allele was visualised as 308bp bands on the gel (2.2.3.2.3).

2.2.3.2.2 Apc-lox2 PCR to amplify downstream *loxP* region in the *Apc^{fl/fl}* Allele

Semi-nested Apc-lox2 PCR was run to amplify the downstream *loxP* site and the adjacent PGKneo cassette region in the *Apc^{fl/fl}* allele from *Cre-Apc^{fl/fl}* mice [144]. Each 10µl PCR reaction contained 5.0µl 10xPCR Run Buffer (-MgCl₂), 2.5mM MgCl₂, 20µM dNTP Mix, 0.2µl Taq DNA Polymerase Recombinant, 0.5µl DMSO, 400nM PGK-F FP, 400nM P5-2 RP and 38.5µl of dH₂O (Table 2-5 & Table 2-6). The assay was run using the

following programme: 95°C for 3 minutes, 40 cycles of 95°C for 30 seconds, 58°C for 45 seconds and 72°C for 150 seconds, and 72°C for 10 minutes. Amplified DNA (1,596bp) was visualised on agarose gel (2.2.3.2.3). The PCR amplicon was retrieved by pricking the gel and the next semi-nested PCR reaction was performed using this as a template.

The subsequent semi-nested step produces a shorter amplicon (<1000bp) for more efficient amplification and sequencing. The composition of PCR reaction was kept the same as before, except it contained a new primer pair of 400nM P5-3F FP and 400nM P5-3R RP. Gel electrophoresis and UV visualisation was carried out as described in section 2.2.3.2.3. Positive results appeared as 736bp bands. This amplified DNA as retrieved from gel and used as a template to repeat the PCR. Subsequently, products were purified and sequenced as described in section 2.2.3.2.3.

2.2.3.2.3 Gel Electrophoresis

The PCR products were analysed by gel electrophoresis in 1-2% Agarose gel in TBE buffer. 5µl Ethidium Bromide was added per 100ml gel for DNA visualisation under UV light. For loading, 2µl Gel Loading Dye was mixed with 10µl sample. The gel was run in TBE buffer at 100V for 40 minutes using Labnet ENDURO™ GEL XL machine. 5µl TrackIt™ 100bp DNA Ladder was used as the DNA size marker and the gel was visualised using the SynGene™ UV Transilluminator and the GeneSnap™ software (version 7.12).

2.2.3.2.4 Purification & Sequencing of PCR products

PCR DNA products from Apc-lox1 and Apc-lox2 assays were purified using the DNA Clean & Concentrator™-5 kit (Table 2-3). A pipette tip was used to prick the bands in the gel to collect the DNA products. The tip was swirled in 10µl dH₂O to dissolve the DNA and 50µl DNA binding buffer was added in 5:1 ratio of buffer to sample. The mixture was passed through the Zymo-Spin™ column by centrifugation at 8,000rpm for 1 minute. The columns were washed twice with 200µl DNA wash buffer by centrifugation at 8,000rpm for 1 minute. The column matrix was incubated in 50µl DNA elution buffer at room temperature for 5 minutes. DNA was eluted by centrifugation at 8,000 rpm for 1 minute. DNA aliquots were quantified using NanoDrop™ Spectrophotometer (ND-1000). Sequencing of the PCR products was carried out by the PNAFL (Protein Nucleic

Acid Chemistry Laboratory, The Centre for Core Biotechnology Services, University of Leicester, UK) facility using an automated Applied Biosystems™ 3730 Sanger sequencer.

2.2.3.3 qPCR

qPCR assays were used for DNA quantification and mutation detection analysis. All assays were run on the StepOnePlus™ Real-Time PCR System Thermal Cycling Block (ABI®). PCR steps were performed at the standard mode ramp rate of 1.6°C/second, except where stated otherwise. All data were analysed using the StepOnePlus™ Software v2.3 (ABI®). All reagents, primers and probes used were listed in Table 2-7, Table 2-8 and Table 2-9.

2.2.3.3.1 Testing Primer Pairs

Potential primer pairs were tested using SYBR® Green qPCR. Each reaction volume contained 5.0 µl fast SYBR® Green master mix, 400nM of each FP and RP. 10ng of commercially available hgDNA or mgDNA was used as controls for human and mouse targets, respectively. The tests were run using the VeriFlex™ programme to test a range of annealing temperatures: 95°C for 10 minutes and 40 cycles of 95°C for 15 seconds and 58°C/60°C/62°C for 1 minute. This was followed by melt curve analysis at 95°C for 15 seconds, 60°C for 1 minute and 95°C for 15 seconds at 0.3% ramp rate, to check for primer dimers caused by non-specific annealing.

2.2.3.3.2 DNA Quantification of Murine Samples

The mGapdh qPCR assay, designed by Dr R. Trigg (Dept. of Cancer Studies, University of Leicester, UK), was used to quantify murine DNA samples. The assay targets 30 *mGapdh* pseudogenes. 10ng of mgDNA control was diluted 1:2 ratio in dH₂O for 7-8 dilutions to construct a standard curve. Each dilution was run in either duplicates or triplicates. Reactions were run in 10µl volumes containing 5.0µl 2x TaqMan® Genotyping PCR Master Mix, 600nM of mGapdh-F FP, 600nM of mGapdh-R RP, 100nM mGapdh-P probe and 3.0µl of sample. The amplicon size was 70bp. The qPCR was run using the

following programme: 50°C for 2 minutes, 95°C for 10 minutes and 50 cycles of 95°C for 15 seconds followed by 60°C for 1 minute.

2.2.3.3.3 Analysis of Recombined Alleles in Murine Samples

The Apc-Rec qPCR assay was used for detection of recombined *Apc^{fl}* alleles in *Cre-Apc^{fl/fl}* mice. Reactions were run in a 10µl volume containing 5.0µl 2x TaqMan® Genotyping PCR Master Mix, 600nM Apcf1-102NF1 FP, 600nM of Apcf1-102NR2 RP, 100nM of Apcf1-P1 MUP and 3.0µl of sample. The amplicon size was 102bp. The assay was run using the following programme: 50°C for 2 minutes, 95°C for 10 minutes and 50 cycles of 95°C for 15 seconds followed by 60°C for 1 minute.

2.2.3.3.4 DNA Quantification of Patient FFPE Samples

Patient FFPE tissue DNA samples were quantified using the hALU69 assay designed by Dr J. Pringle (Dept. of Cancer Studies, University of Leicester, UK). It targets repetitive *Alu* retrotransposon elements as a reference. 10ng of hgDNA control was diluted 1:2 ratio in dH₂O for 7-8 dilutions to construct a standard curve. Each dilution was run in triplicates. Reactions were run in 10µl volumes containing 5.0µl 2x TaqMan® Genotyping PCR Master Mix, 600nM of hALU-F FP, 600nM of hALU-69R RP, 100nM hGAPDH-P probe and 3.0µl of DNA sample. The amplicon size was 69bp. The assay was run using the programme: 50°C for 2 minutes, 95°C for 10 minutes and 50 cycles of 95°C for 15 seconds followed by 60°C for 1 minute.

2.2.3.3.5 DNA Quantification of Patient Plasma Samples

Patient plasma DNA samples were quantified using the hGAPDH qPCR assay designed by Prof J. Shaw's team (Dept. of Cancer Studies, University of Leicester, UK). This assay targets an intronic sequence of the housekeeping gene *hGAPDH* as a reference. 10ng of hgDNA control was diluted 1:2 ratio in dH₂O for 7-8 dilutions to construct a standard curve. Each dilution was run in triplicates. The assay was run in a 10µl reaction volume containing 5.0µl 2x TaqMan® Genotyping PCR Master Mix, 600nM

of hGAPDH-F FP, 600nM of hGAPDH-R RP, 100nM of hGAPDH-P probe and 3.0µl of DNA sample. The amplicon size was 95bp. The qPCR programme was as described in section 2.2.3.3.4.

2.2.3.3.6 BRAF & KRAS Mutation Detection qPCR Assays with PNA Clamping

The following 3 assays, BRAF-V600E, KRAS-121 and KRAS-122, were used for detection of *BRAF* and *KRAS* hotspot mutations in patient FFPE samples. These assays were designed by Dr J Pringle (Dept. of Cancer Studies, University of Leicester, UK) [333, 334]. Validation of these assays will be described in Chapter 5. All assays were run in 10µl reaction volumes containing 5.0µl of 2x TaqMan® Genotyping PCR Master Mix with the respective primers, probes and PNA blocker as described below. Each sample was run in triplicates. The assays were run using the following programme: holding stage at 50°C for 2 minutes followed by 95°C at 10 minutes, then 50 cycles of 95°C for 15 seconds and 60°C for 1 minute.

BRAF-V600E qPCR

This assay detects the hotspot *BRAF* mutation at c.1799T>A/p.V600E. Each 10µl reaction contained 600nM BRAF-600F, 600nM BRAF-600R, 100nM BRAF-WTP, 100nM BRAF-MUP, 150nM BRAF-PNA and 3.0µl of sample. The amplicon size was 94bp.

KRAS-121 qPCR

This assay detects the 3 possible missense *KRAS* mutations at base c.34G – c.34G>A/p.G12S, c.34G>C/p.G12R and c.34G>T/p.G12C – by using a MU probe KRAS-121P with a degenerate base D (A/T/G) at base c.34 [347]. One reaction was run with WT probe only without PNA clamping – it contained 600nM KRAS-78F, 600nM of KRAS-78R, 100nM KRAS-WTP and 3.0µl of sample. For the same sample, duplicate reactions were run with the MU probe and the PNA blocker – each reaction contained 600nM KRAS-78F, 600nM KRAS-78R, 100nM KRAS-121P, 60nM KRAS-PNA and 3.0µl of sample aliquot. The amplicon size was 78bp.

KRAS-122 qPCR

This assay detects the 3 possible missense *KRAS* mutations at base c.35G – c.35G>A/p.G12D, c.35G>C/p.G12A, c.35G>T/G12V – by using a MU probe KRAS-122P with a degenerate base D (A/T/G) at base c.35 [347]. Each 10µl reaction contained 600nM KRAS-78F, 600nM KRAS-78R, 100nM KRAS-WTP, 100nM KRAS-121P, 60nM KRAS-PNA and 3.0µl sample aliquot. The amplicon size was 78bp.

2.2.3.4 ddPCR

2.2.3.4.1 Assay Set-up

All ddPCR assays were run on the QX200™ ddPCR system (section 1.8.2). Each sample was run in 20µl volumes containing 10µl ddPCR Supermix for Probes (no dUTP). DG8™ cartridges and gaskets (BioRad) were used for droplet generation – 70µl droplet generation oil was used per 20µl sample. Droplets were transferred to a 96-well, semi-skirted Eppendorf® twin-tec PCR plates (Sigma) and sealed with a pierceable foil heat seal on a PX1™ PCR plate sealer (BioRad). Within one hour of droplet generation, plates were run on the C1000 Touch™ Thermal Cycler with 96-Deep Well Reaction Module. All PCRs were run at a ramp rate of 2°C/sec. Droplets were read using the QX200™ Droplet Reader. Data capture and analysis was done by the proprietary software QuantaSoft™ software (version 1.6.6.0320, BioRad®). All reagents, primers and probes used were listed in Table 2-7, Table 2-8 and Table 2-9.

2.2.3.4.2 Analysis of Recombined & Non-Recombine *Apc^{fl/fl}* Alleles in Murine Samples

The Apc-Rec assay targets the recombined *loxP* region in the *Apc^{fl}* alleles that had undergone Cre-mediated recombination. The Apc-NR assay targets the inserted PGK-neo cassette that is only present in the non-recombined *Apc^{fl/fl}* alleles. The Apc-Rec and Apc-NR assays were run in parallel to detect and analyse the fractional abundance of recombined alleles. The Apc-Rec assay was run with 450nM Apcfl-102NF1, 450nM Apcfl-102NR2 and 250nM Apcfl-P1. The Apc-NR assay was run with 450nM Apcfl-NRF1, 450nM Apcfl-NRR2 and 250nM Apcfl-NRP1. 3.0µl of sample was tested per 20µl reaction volume. The following run programme was used: enzyme activation at 95°C for 10

minutes, 40 cycles of extension at 94°C for 30 seconds and 60°C for 1 minutes, and enzyme deactivation at 98°C for 10 minutes with holding at 4°C.

2.2.3.4.3 Analysis of *BRAF* & *KRAS* mutations in Patient Samples

The three mutation assays, *BRAF*-V600E, *KRAS*-121 and *KRAS*-122, which were initially implemented using qPCR (section 2.2.3.3.6), were also adopted for the QX200 ddPCR system for analysis of patient plasma samples. They were run using the same primers and probes as described in section 2.2.3.3.6, with the omission of PNA clamps. For all assays, samples were tested in 20µl reaction volumes containing respective FP and RP at 450nM, and WTP and MUP at 250nM (Table 2-9). The assay was run using the following programme: enzyme activation at 95°C for 10 minutes, 40 cycles of denaturation and extension at 94°C for 30 seconds and 59°C for 1 minutes respectively, enzyme deactivation at 98°C for 10 minutes with holding at 4°C.

An additional *KRAS* assay – *KRAS*-G13D – was also used to detect the c.38G>A/p.G13D *KRAS* mutation in patient tissue and plasma samples. This assay was only run on the ddPCR system using the commercial mixes (dHsaCP2000013-14; BioRad®). Each sample was tested in a 20µl reaction volume containing 10.0µl ddPCR Supermix for Probes, 1.0µl *KRAS*-G13D WT mix and 1.0µl *KRAS*-G13D MU mix (Table 2-9). The assay was run using the following programme: enzyme activation at 95°C for 10 minutes, 40 cycles of denaturation and extension at 94°C for 30 seconds and 55°C for 1 minutes respectively, enzyme deactivation at 98°C for 10 minutes with holding at 4°C.

2.2.3.4.4 Patient-specific Mutation Detection Assays

Multi-regional, targeted next generation sequencing (NGS) data (section 2.2.4) from patients with large adenomas were used to design five patient-specific mutation detection assays (chapter 5 & 6). Five mutations – three in the MCR of *APC*, one in *TP53* DNA-binding domain and one in *PIK3CA* kinase domain – were chosen as targets:

1. APC1338 assay for *APC* mutation at codon 1338 (c.4012C>T/p.Q1338*)
2. APC1397 assay for *APC* mutation at codon 1397 (c.4189G>T/p.E1397*)

3. APC1429 assay for *APC* mutation at codon 1429 (c.4285C>T/p.Q1429*)
4. TP53-273 assay for *TP53* mutation at codon 273 (c.817C>T/p.R273C)
5. PIK1047 assay for *PIK3CA* mutation at codon 1047 (c.3140A>G/p.H1047R)

Validation of these assays will be discussed in Chapter 5, and analysis of patient samples will be discussed in Chapter 6. For all assays, each 20µl reaction was run with 450nM of respective FP and RP, and 250nM of respective WTP and MUP (Table 2-9). The assays were run using the following programme: enzyme activation at 95°C for 10 minutes, 40 cycles of denaturation and extension at 94°C for 30 seconds and 57°C (except at 60°C for the APC1429 assay) for 1 minutes respectively, enzyme deactivation and droplet hardening at 98°C for 10 minutes with holding at 4°C.

2.2.3.4.5 Defining the Droplet Threshold

For mutation detection ddPCR assays, limit of detection (LoD) is defined by the minimum target (MU) concentration that can be discriminated from the non-target (WT) background [339, 348]. The QuantaSoft software makes a positive calling if a droplet has fluorescence that is above the threshold, which can be set manually [339, 348]. Assuming an assay has a false-positive rate of 0 droplet, for single-well assay runs at a LoD of 0.01%, where 1 MU target exists in a background of 10,000 WT copies, at least 3 MU-positive copies need to be detected above the fluorescence threshold to fulfil the Rule of Three where positive MU target calling can be made with 95% statistical confidence interval [339, 348].

2.2.4 Multi-regional Targeted NGS of Patient Adenoma Samples

A total of 30 FFPE samples from 6 patient cases (H149, H154, H263, H264, H265 and H266) were selected for NGS analysis in collaboration with Wellcome Trust Sanger Institute, Cambridge, UK. These samples were pathologically assessed by Dr Kevin West (Leicester Royal Infirmary, UK) for tumour abundance and macroscopic enrichment (Table 6-7). DNA extraction was as described in section 2.2.2.2.3, and quantified as described in section 2.2.3.3.4 – 500ng of each sample was shipped for sequencing.

Sequencing was carried out using a custom, proprietary CRC panel consisting of 116 mutations, 22 gene amplification/deletion, 23 recurrent focal amplification/deletion, 121 MSI regions and 2 gene fusions (appendix 8.2). Custom baits (SureSelect, Agilent, UK) were designed by the Cancer Genome Project staff at the Sanger Institute, and they target all coding exons of 116 genes implicated in CRC; 95 were identified as cancer genes from systematic analysis of 229 whole exome sequences (WES) and 21 were recurrently mutated genes in MSI tumours. Prior to sequencing, DNA is fragmented to an average insert size of 150bp and subjected to Illumina DNA sequencing library preparation using Bravo automated liquid handling platform. Sequencing was performed on an Illumina HiSeq2000 machine using the 75-bp paired-end protocol with the target of 1Gb sequence per sample. Data quality has been confirmed with 95% target coverage at 100x. Mutation analysis was performed by bioinformaticians at Sangers Institute using their in-house algorithm. NGS data will be discussed in Chapter 6 with additional details reported in appendix 8.3.

2.2.5 *In Vivo* Studies

All *in vivo* studies with *Cre-Apc^{fl/fl}* mice were carried out under the project licence (PPL) no. 60/4370 and the personal licence (PIL) no. I48C35887. Animal studies were conducted at the Central Research Facility, University of Leicester, UK. Mice were housed in a climate-controlled environment with a 12 hour day/night cycle.

2.2.5.1 Preparation of Tamoxifen & Vehicle Solutions

Intraperitoneal injection of tamoxifen (10mg/ml tamoxifen ($\geq 99\%$) in sunflower oil; Table 2-10) was used to induce adenoma formation in *Cre-Apc^{fl/fl}* mice. Tamoxifen solution was prepared by mixing 0.1g of tamoxifen powder in 1ml ethanol, and dissolved by vortexing with intermittent sonication to accelerate the process. Then, it was mixed with 9ml sunflower seed oil (Table 2-10) and vortexed until the solution was homogeneous. The mixture was protected from light by covering with foil and aliquoted into 1ml amber glass vials for storage at -20°C . The vehicle control was prepared by mixing sunflower seed oil and absolute ethanol in 9:1 ratio at room temperature.

All injections were performed by trained staff at the Central Research Facility, University of Leicester, UK. Study mice were either injected with 3mg tamoxifen or 300µl vehicle for the first study. For the second and third studies, tamoxifen/vehicle volumes injected were matched to the starting body weights of each mouse at 100µl/10g.

2.2.5.2 Diet Allocation

All mice were fed EURodent Diet 14% (5LF2*; Table 2-10) as a maintenance diet before studies commenced. In the 1st and 2nd studies, mice were given the same maintenance diet for the study duration. For the 3rd study, in which effects of the dietary fat content on tumourigenesis was investigated, mice were allocated to either normal fat diet (NFD) and high fat diet (HFD) groups two weeks prior to tamoxifen induction. Pellets of AIN-93G growth purified diet (#5801-G; Table 2-10) were provided as NFD, and pellets of modified AIN-93G (60% Kcal from Fat as Coconut Oil; Table 2-10) were provided as HFD. Diet pellets were stored at -20°C and intake was observed and replaced weekly, or at earlier intervals if necessary. All diet and water were consumed *ad libitum*.

2.2.5.3 Health Checks

All the study mice were monitored by experienced staff at the Central Research Facility, University of Leicester, UK. Health checks and weight measurements were performed weekly, and this was increased to twice a week once the mice had started to exhibit signs of sickness that include, and not limited to, anaemia observed as pale feet, sudden weight loss, droopy eyes and ears, raised fur and melancholy behaviour.

2.2.5.4 Study End-points

For the 1st study, study end-point was designated at 8 weeks post-injection provided that the mice had not lost 10% of highest body weights measured after day 0 (i.e. date of injection). If any mice exhibited more than two signs of sickness or lost more than 10% body weight, they were euthanised. For the 2nd study, study end-point was increased to 15% loss of highest body weights measured at day 0 (i.e. date of injection)

or showing more than two signs of sickness, or a mixture of both. For the 3rd study, the end-point was further increased to 20% weight loss; other parameters were kept the same. All mice were euthanised via cardiac puncture under Isoflurane-induced terminal anaesthesia (Table 2-10) to collect blood and tissue samples.

2.2.5.5 Blood Collection & Plasma Extraction

All blood collection was done via cardiac puncture. Blood was collected in 1.3 ml EDTA-coated tubes (K3E, TEKLAB™) to prevent clotting. Tubes were kept on ice and plasma was isolated within two hours by sequential centrifugation at 1,000g for 10 minutes and then at 2,000g for 15 minutes to remove any residual cellular components. Plasma volumes ranged from 150-900µl and samples were stored at -80°C until use.

2.2.5.6 Tissue Collection

Small intestine, colon and caecum were harvested and fixed in 10% formal saline (Table 2-10) over night before being processed for paraffin embedment by Core Biotechnological Services, University of Leicester, UK. Small intestine was divided into 3 or 6 sections and swiss-rolled for embedding. Swiss-rolling was also performed for colon samples. Small pieces of tissue (approximately 3x3 mm) were also collected from each tissue section and stored at -80°C for DNA extraction. Other tissues such as spleen were also processed if they showed abnormal phenotypes such as enlargement.

2.2.5.7 Collection of Faecal Samples

Faecal samples were collected from the mice in the 3rd study. Collection of faecal samples started on day 0 when diets were first allocated. Samples were collected on a weekly basis, and at end-points. For collection, each mouse was transferred to a sterile cage briefly and monitored until they had excreted faecal samples. Quantities of faecal samples collected ranged from 9-234mg (median= 39mg) depending on diet consumption and general health conditions of each mouse. Samples were immediately snap-frozen in liquid nitrogen and stored at -80°C for further analysis.

2.2.6 Histology

2.2.6.1 Haematoxylin & Eosin Staining

Haematoxylin and eosin (H&E) staining of FFPE samples was performed to examine any pathological changes in the tissues. Staining was done using ~4µm thick sections of FFPE samples. All H&E staining was performed by Core Biotechnological Services at University of Leicester, UK.

2.2.6.2 Immunohistochemical Analysis of β -catenin Expression in *Cre-Apc^{fl/fl}* Mice

All solutions in this protocol are listed in Table 2-11. FFPE sections (~4µm) were mounted onto Menzel-Gläser polysine[®] glass slides (Thermo Scientific[®]) and incubated at 65°C for 30 minutes. The sections were dewaxed via serial immersions in xylene, 99% IMS and 95% IMS for 3 minutes each; all steps were done twice. The sections were washed in running tap water for 5 minutes. Antigen retrieval was performed by microwaving the slides in 10mM citrate monophosphate buffer (pH 6.0) for 20 minutes at 900W. The sections were left to cool at room temperature in the hot buffer for ~45 minutes before washing in PBS for 5 minutes.

The staining protocol used the DakoCytomation Biotin blocking system designed to prevent non-specific staining. Sections were incubated in avidin blocking solution for 10 minutes to block endogenous biotin, and then washed in PBS for 5 minutes. Then, sections were incubated with biotin blocking solution for 10 minutes to block any residual non-specific biotin activity, and then washed in PBS for 5 minutes. Afterwards, sections were incubated in 3% (w/w) hydrogen peroxide in dH₂O for 20 minutes to inhibit endogenous peroxidase activity, and then washed in PBS for 5 minutes. Sections were incubated in the primary mouse anti- β -catenin primary antibody (#610154; 1:1000 or 2µg/ml) in 3% BSA in PBS with 0.1% Triton X-100 at 4°C for ~20 hours. For negative controls, sections were incubated in 1.0 µg/ml Mouse IgG1 X0931 (1:1000) in the same diluent. Following this, sections were washed in PBS for 5 minutes three times.

Dako LSAB2 System-HRP kit (#K0675, Agilent Technologies[®]) was used for antigen visualisation. Sections were incubated in biotinylated link for 10 minutes at

room temperature, and then washed in PBS for 5 minutes. Next, sections were incubated in streptavidin-HRP for 10 minutes at room temperature, and washed in PBS for 5 minutes. Peroxidase activity was developed by incubating in DAB working solution, a 1:20 mixture of DAB chromogen in DAB substrate buffer (polymer), for 5 minutes in the dark. Sections were then washed in PBS for 5 minutes and counterstained in Haematoxylin for 5 minutes. Sections were washed in PBS for 5 minutes and dehydrated by serial immersions in 95% IMS for 3 minutes twice, 99% IMS for 3 minutes twice and xylene for 3 minutes twice. Sections were mounted with DPX and left to dry overnight.

2.2.6.3 Image Capturing

Histological images from the 1st preclinical study were acquired using a Leica DM2500 light microscope with the Leica DFC420 (C-Mount 0.55x) camera. All other images were captured using the NanoZoomer-XR digital slide scanner C12000 (Hamamatsu Photonics®) and analysed using the NDP viewer software v.2.4.26 (Hamamatsu Photonics®).

2.2.7 Statistical Analysis

Total plasma cfDNA concentrations between the control and experimental groups for *in vivo* and patient samples were analysed using a T-test and Mann Whitney test. For the 3rd *in vivo* study, the effect of dietary fat content on the survival function was analysed using the log rank test. Changes in the percentage fractional abundance of recombined *Apc^{fl}* alleles in relation to consumption of diets with different fat content and the GI tract locations were analysed using a 2-way ANOVA with Sidak's multiple comparison test. Multiple comparisons of nuclear β -catenin indices in the crypts of control and experimental 'Apc-deleted' mice in the NFD and HFD groups were made using 2-way ANOVA and Sidak's correction. Nuclear β -catenin indices in the adenomas from NFD and HFD groups were analysed using Mann Whitney test. Moreover, changes in the percentage fractional abundance of recombined *Apc^{fl}* alleles in faecal samples at various time-points between NFD and HFD groups were analysed using 2-way ANOVA with Tukey's multiple comparison to allow for unequal sample sizes.

Additionally, efficiencies of qPCR and ddPCR assays were calculated using linear regression analysis. For all mutation detection assays performed ddPCR, Poisson distribution analysis was used to predict the probability of independent and random distribution of targets into droplets at 95% confidence level.

Chapter 3 Validation of Assays for Preclinical Studies

3.1 Introduction

As described in the Introduction (section 1.5), tamoxifen-mediated activation of Cre nuclear localisation triggers exon 14 deletion in the Lgr5-expressing CBC cells and initiates adenoma formation in *Cre-Apc^{fl/fl}* mice. In other words, the recombination status of the *Apc^{fl}* allele reflects the tumour-specific alteration in this model. Therefore, to fulfil the first aim of this project, which was to use the *Cre-Apc^{fl/fl}* model to investigate whether cfDNA is detectable during early stages of cancer, I designed PCR-based assays to analyse the plasma levels of recombined *Apc^{fl}* alleles to assess how early these could be detected during the initial stages of disease development. In addition, I sought out to investigate whether the quantities of recombined *Apc^{fl}* fragments in plasma are informative of disease burden.

To accurately design primers and probes for custom PCR assays for *in vivo* analysis, it was first necessary to sequence the genomic regions encompassing the two *loxP* sites in the *Apc^{fl/fl}* allele (Figure 1-16). This chapter describes the targeted sequencing of these *loxP* sites, and subsequent designing and validation of custom PCR assays used to analyse the levels of non-recombined and recombined alleles in the mouse tissue and plasma samples.

3.2 Sequencing Upstream *loxP* in Intron 13 of *Apc^{fl/fl}* Allele

Apc-lox1 PCR (section 2.2.3.2.1) was used to amplify the sequence containing the upstream (1st) *loxP* region in intron 13 of the *Apc^{fl/fl}* allele (Figure 3-1-A). DNA extracted from ear-snip tissues of three *Cre^{+ve}-Apc^{fl/fl}* mice were used as templates for this assay, and positive 308bp bands were successfully detected on the agarose gel for all samples (Figure 3-1-B).

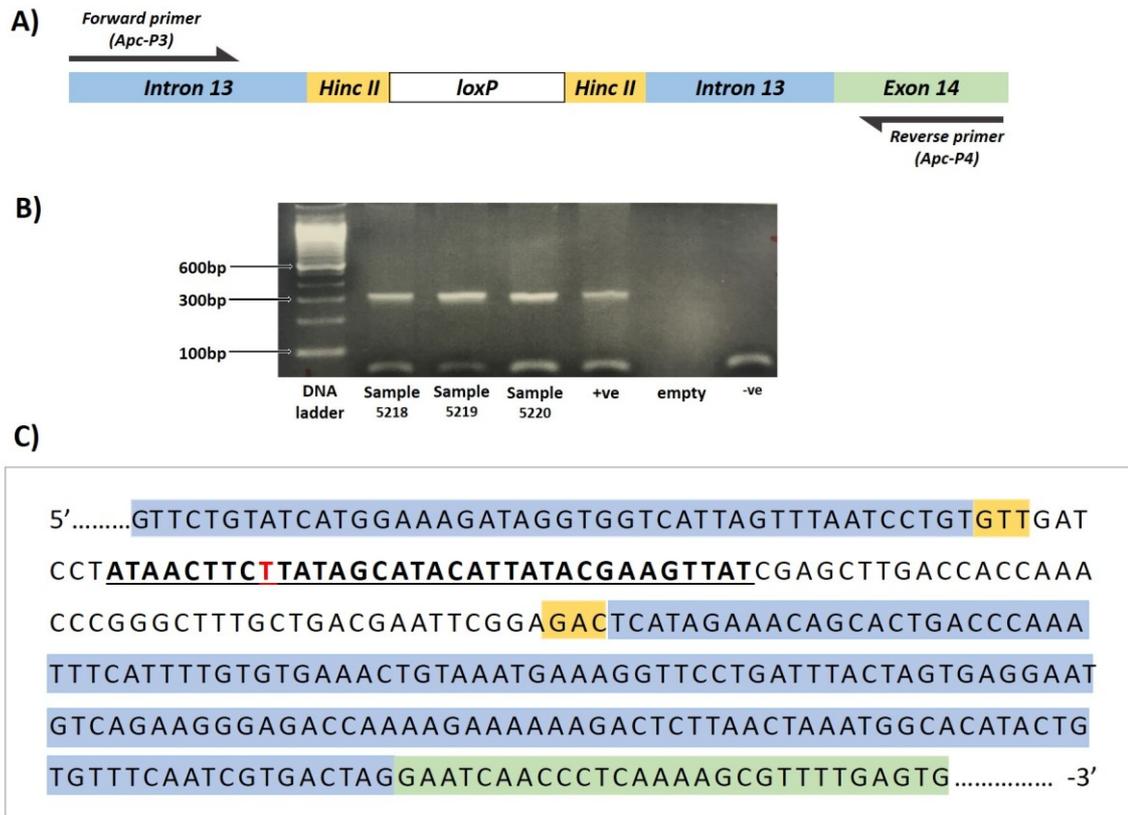


Figure 3-1: Sequence details for the Upstream *loxP* site in the *Apc^{fl/fl}* Allele

A) Diagram showing the upstream *loxP* and surrounding sequences. The locations of the *Apc-lox1* PCR primers, *Apc-P3* and *Apc-P4*, were also indicated. *Hinc II* was the restriction enzyme sites used for *loxP* insertion. [348]

B) Positive bands appeared at 308bp on the gel for all three samples (ID 5218, 5219, 5220). Positive (+ve) control was the DNA sample from a *Cre^{+ve}-Apc^{fl/fl}* mouse (ID 5181). Negative (-ve) control was water. Bands at the bottom (<100bp) were primer dimers. TrackIt™ 100bp DNA ladder was used a size marker.

C) The sequencing data for the upstream *loxP* region. The base (T) in red was the observed mismatch in the upstream *loxP* compared to the published sequence. [146] **Blue** = intron 13, **green** = exon 14, **yellow** = *Hinc II* restriction sites, **underlined** = upstream *loxP* region.

Details on the purification and sequencing of the amplified DNA were described in section 2.2.3.2.4. Two sequencing data sets were obtained for each sample; one from each primer. No discordance in the sequences were observed from the 3 mice, and both strand sequences concurred as well. Additionally, the endogenous mouse sequences surrounding the *loxP* were confirmed using the BLAST programme (blastn suite). The final sequence was reported in Figure 3-1-C. The intron 13 of the endogenous *Apc* gene was shown in blue highlights and the *Hinc II* restriction sites (GTx/xAC) used for *loxP* insertion were shown in yellow highlights [349]. Endogenous exon 14 sequence was

shown in green. Interestingly, the *loxP* sequence obtained showed a guanosine (G) to thymidine (T) change compared to the previously published *loxP* sequence [144, 146]:

5' ATAACCTTC**T**TATAGCATACATTATACGAAGTTAT 3' (in-house *loxP* sequence)

5' ATAACCTTC**G**TATAGCATACATTATACGAAGTTAT 3' (published *loxP* sequence)

The reason for this discrepancy was not known but, despite this change, successful *loxP* recombination was observed in all tamoxifen-treated *Cre^{+ve}-Apc^{fl/fl}* animals (Chapter 4). However, I cannot rule out that this base change might affect the efficiency of Apc exon 14 deletion.

3.3 Sequencing Downstream *loxP* in Intron 14 of *Apc^{fl/fl}* Allele

Sequencing of the downstream (2nd) *loxP* site in intron 14 proved to be more challenging as a 4,575bp PGKneo cassette [350], which contained the phosphoglycerate kinase 1 (PGK1) promoter and the neomycin (neo) resistance gene, was inserted upstream of this *loxP* sequence (Figure 3-2-A). To sequence an amplicon of over 1,000bp, a semi-nested Apc-*lox2* PCR was used (2.2.3.2.2). The first PCR was run using the PGK-F FP, which was used by Addgene for plasmid verification [350, 351], and the P5-2 RP to produce a 1,596bp amplicon (Figure 3-2: A & B). Ear-snip tissue DNA from mouse 5215 was used as a template, and the reaction was run in duplicate to obtain enough DNA for sequencing. However, the full sequence of the amplicon was not obtained, which was likely caused by the long amplicon size.

Therefore, using the partial sequence obtained from the first PCR amplicon, a second primer pair was designed to amplify a shorter amplicon of 736bp in a semi-nested PCR approach. This second PCR reaction used a P5-3F, which annealed downstream of PGK-P5 primer, and a P5-3R, which was a shortened version of P5-2 primer for optimised annealing (Figure 3-2-A). Ear-snip DNA samples from four *Cre^{+ve}-Apc^{fl/fl}* mice (5215, 5217, 5220 and 5222) were used to run this semi-nested reaction, and the amplification proved successful in all cases as shown by 736bp positive bands on gel (Figure 3-2-C). PCR DNA products were purified and sequencing as described in section 2.2.3.2.4.

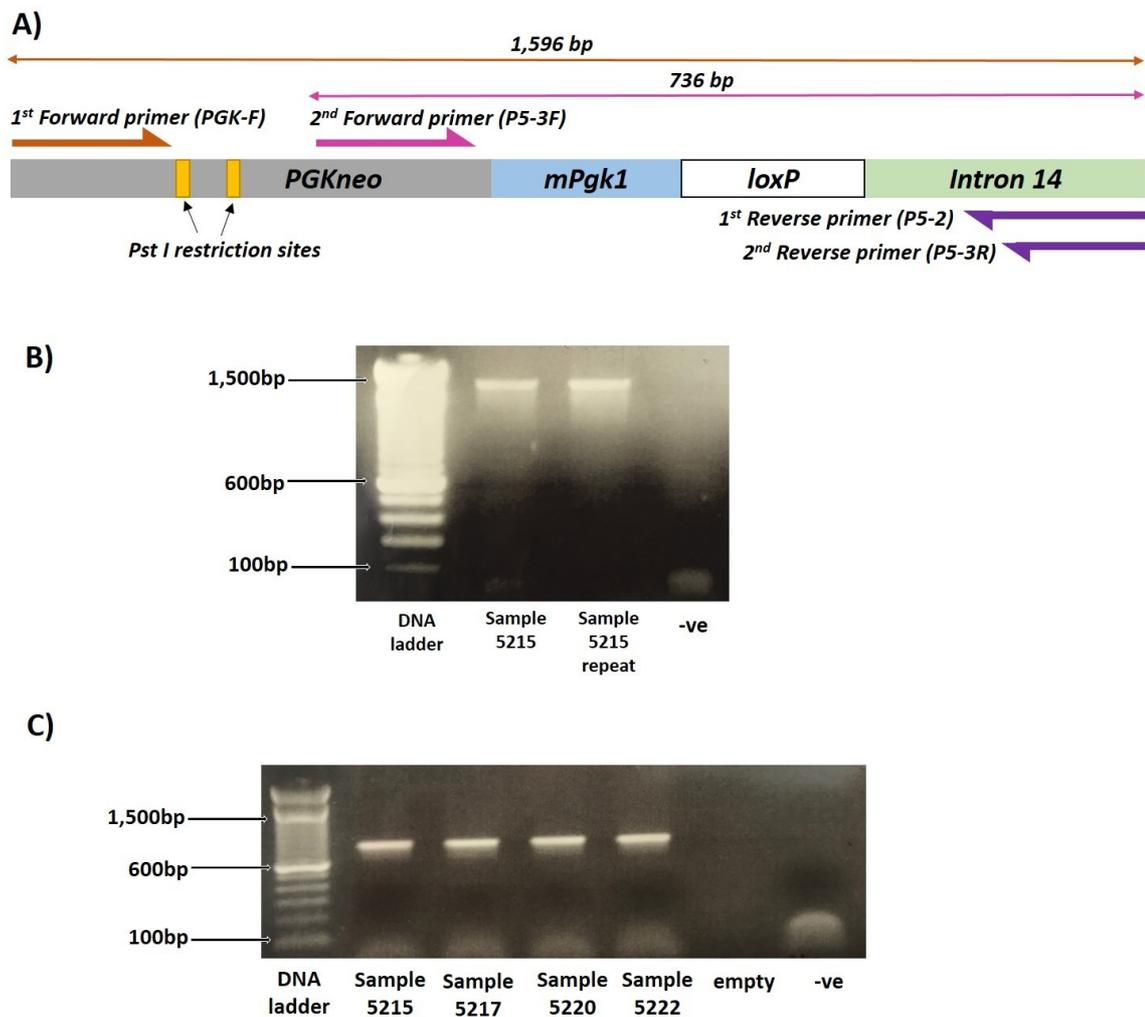


Figure 3-2: PCR amplification of Downstream *loxP* sequence

A) Schematic of the *loxP* site and PGKneo cassette in intron 14 of the *Apc* gene. The PGK-F and P5-2 pair gave a 1,596bp amplicon, and the P5-3F and P5-3R pair gave a 736bp amplicon.

B) PCR agarose gel for the 1st reaction. The DNA template was from mouse 5215. Negative (-ve) control was water. TrackIt™ 100bp DNA ladder was used as a size marker.

C) PCR agarose gel for the 2nd semi-nested PCR reaction. DNA samples from four mice (ID 5215, 5217, 5220 and 5222) were used as positive templates. Negative (-ve) control was water. TrackIt™ 100bp DNA ladder was used as a size marker.

5' **CATTCTGCACGCTTCAAAG**CGCACGTCTGCCGCGCTGTTCTCCTC
TTCCTCATCTCCGGCCTTTCGAC**CTGCAG**CCAATATGGGATCGGCCATTGAA
CAAGATGGATTGCACGCAGGTTCTCCGGCCGCTTGGGTGGAGAGGCTATTTCG
GCTATGACTGGGCACAACAGACAATCGGCTGCTCTGATGCCGCCGTGTTCCGG
CTGTCAGCGCAGGGGCGCCCGTCTTTTTGTCAAGACCGACCTGTCCGGTG
CCCTGAATGAA**CTGCAG**GACGAGGCAGCGCGGCTATCGTGGCTGGCCACGAC
GGGCGTTCCTTGCAGCTGTGCTCGACGTTGTCACTGAAGCGGGAAGGGAC
TGGCTGCTATTGGGCGAAGTGCCGGGGCAGGATCTCCTGTCATCTCACCTTGC
TCCTGCCGAGAAAGTATCCATCATGGCTGATGCAATGCGGCGGCTGCATACGC
TTGATCCGGCTACCTGCCATTCGACCACCAAGCGAAACATCGCATCGAGCGA
GCACGTA CTGGATGGAAGCCGGTCTTGTGATCAGGATGATCTGGACGAAG
AGCATCAGGGGCTCGCGCCAGCCGAACTGTTCCGCCAGGCTCAAGGCGCGCAT
GCCCGACGGCGATGATCTCGTCGTGACCCATGGCGATGCCTGCTTGCCGAATA
TCATGGTGGAAAATGGCCGCTTTTCTGGATTCATCGACTGTGGCCGGCTGGGT
GTGGCGGACCGCTATCAGGACATAGCGTTGGCTACCCGTGATATTGCTGAAGA
GCTTGGCGGCGAATGGGCTGACCGCTTCCTCGTGCTTTACGGTATCGCCGCTC
CCGATTCGCAGCGCATCGCCTTCTA**TCGCCTTCTTGACGAGTTCT**TCTGAGGG
GATC**CGCTGTAAGTCTGCAGAAATTGATGATCTATTAACAATAAAGATGTCCA**
CTAAAATGGAAGTTTTCTGTCATACTTTGTTAAGAAGGGTGAGAACAGAGT
ACCTACATTTTGAATGGAAGGATTGGAGCTACGGGGGTGGGGGTGGGGTGG
GATTAGATAAATGCCTGCTCTTTACTGAAGGCTCTTTACTATTGCTTTATGATAA
TGTTTCATAGTTGGATATCATAATTTAAACAAGCAAACCAAATTAAGGGCCAG
CTCATTCTCCCACTCATGATCTATAGATCTATAGATCTCTCGTGGGATCATTGT
TTTTCTCTTGATTCCCACCTTGTGGTTCTAAGTACTGTGGTTTCCAAATGTGTC
AGTTTCATAGCCTGAAGAACGAGATCAGCAGCCTCTGTTCCACATACTTCA
TTCTCAGTATTGTTTTGCCAAGTTCTAATTCATCAGAAGCTGGTCGATCGAAT
TCCTGCAGCCCGGGGGATCCTATAACTTCTTATAGCATACTTATACGAAGTTA
TCGGATCCACTAGTTCTAGCATTATATTGACTGTTAGCCCTTATATTATATGCTT
TTTGATTTTTTAAAACCTATACTTCATGTTATTTCTTAAAATAATGCTTATATA
CAGTCTGCCAAAGTGTGCTTTGGACTTGGTGTCTTCACTGAGACAGAGACCC
CGTACTC.....3'

Figure 3-3: Downstream loxP Sequence in Intron 14 of Apc^{fl/fl} Allele

The sequencing data for downstream loxP sequence (1,596bp). As with the upstream loxP sequence, a base change of G>T (in red) was observed at the same location in the palindromic region of loxP [146]. The underlined nucleotide 'C' at the start of intron 14 (green) represents the 5'-sticky end of the Sac I digestion site where 2nd loxP sequence was inserted [351].

Orange = PGK-F primer, **grey** = 2691-3582bp region in PGKneobpA plasmid [350], **yellow** = PstI restriction sites, **pink** = P5-3F primer, **blue** = partial alignment with plasmid vector β-lactamase gene (pHM2/3), **underlined** = loxP, **green** = intron 14, **purple** = P5-3R primer.

Using the PCR-amplified products from the two PCR reactions (Figure 3-2), 1,596bp sequence of the downstream *loxP* and surrounding regions was obtained (Figure 3-3). The sequencing data was confirmed using BLAST (blastn suite: Align) to position the *loxP* site respective to the intronic and exonic regions of endogenous *Apc* gene (Figure 3-3). The first 892 bases of sequenced data showed alignment with 2,691-3,582 bp region of PGKneobpA plasmid (grey), and it also contained the *Pst1* restriction sites (yellow) [350, 351]. The following 467 bases (blue) showed partial alignment with plasmid vector β -lactamase gene (pHM2/3), which is likely a residual gene region from the original construct [144]. The *loxP* was inserted at the sticky end 'C' (5'-GAGCT,C-3'; underlined & green) produced by *Sac I* digestion at the start of intron 14 (green) [352]. Alignment with the intron 14 of *Apc* was confirmed using the GRCm38.p4 assembly of C57BL/6J strain (blastn suite). Interestingly, like the 1st *loxP* site, this 2nd *loxP* sequence (underlined) also has the G>T base change (in red) from the reported consensus WT *loxP* site sequence [144, 146]. This was confirmed by paired-end sequencing. Both upstream and downstream *loxP* sites were accompanied by unique sequences that were probably derived from the original DNA construct used for somatic recombination [144, 146].

3.4 Designing & Optimisation of Custom Assays for Detection of Recombined & Non-recombined Alleles

Using the sequencing data obtained in sections 3.2 and 3.3 (Figure 3-4), custom primers and probes were designed for the detection of recombined *Apc^{fl}* alleles as a tumour biomarker in *Cre-Apc^{fl/fl}* mice, with an additional assay designed to detect the non-recombined alleles as a reference.



Figure 3-4: Locations of *loxP* sites in the *Apc^{fl/fl}* allele

Schematic representation of the two *loxP* sites in the *Apc^{fl/fl}* allele.

As described in section 1.5, Cre-mediated recombination led to deletion of *Apc* exon 14, and this resulted in bringing the introns 13 and 14 together (Figure 3.5). Therefore, the custom Apc-Rec assay was designed to target and amplify a 102bp amplicon encompassing this recombined *loxP* site (2.2.3.3.3). The Apcf1-102NF1 FP was designed to partially anneal with the unique sequence in the upstream of *loxP* site and the intron 13 to improve target specificity, and the Apcf1-102NR2 RP annealed to the intron 14 sequence adjacent to the recombined *loxP* (Figure 3-5). Sequence-specific TaqMan probe (Apcf1-P1) tagged with a FAM reporter dye was targeted to the palindromic region of the *loxP* site (Figure 3-5). Hypothetically, this assay could also amplify the non-recombined locus as the Apcf1-P1 probe recognises any *loxP* site, but this is a highly unlikely event, as it necessitates producing an amplicon of more than 5000bp that includes the PGK-neo cassette sequence (Figure 3-4).

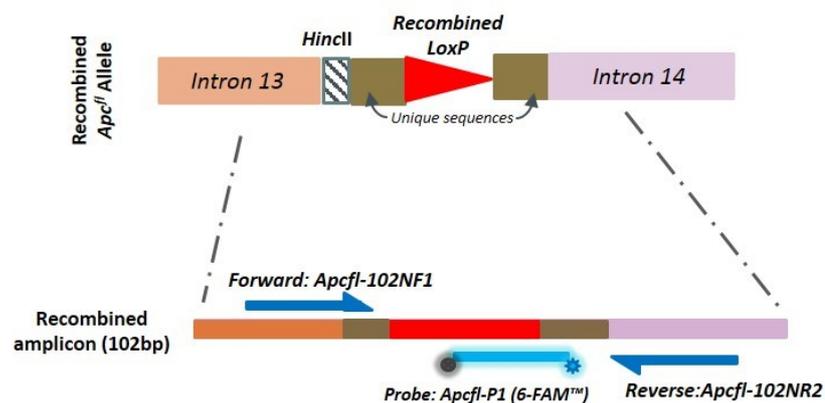


Figure 3-5: Apc-Rec Assay Design for Detection of the Recombined Allele

Schematic representation of the recombined *Apc^{fl}* allele. Annealing sites for the primers and probe were indicated. See Table 2-8 for sequence details.

Initially, the Apc-Rec assay was validated on the StepOnePlus qPCR system. The assay efficiency was tested by standard curve analysis using serial dilutions of DNA isolated from the SI tissue of a tamoxifen-treated *Cre^{+ve}-Apc^{fl/fl}* mouse (5284) (Figure 3-6-A). The recombined status, thereafter referred to as ‘Apc-deleted’, of this mouse was confirmed by sequencing of its PCR product as described in section 2.2.3.2.4 (data not shown). This DNA template aliquot contained a mixture of recombined and non-recombined alleles as DNA was extracted from a piece of proximal SI tissue rather than from isolated Lgr5-expressing CBC cells.

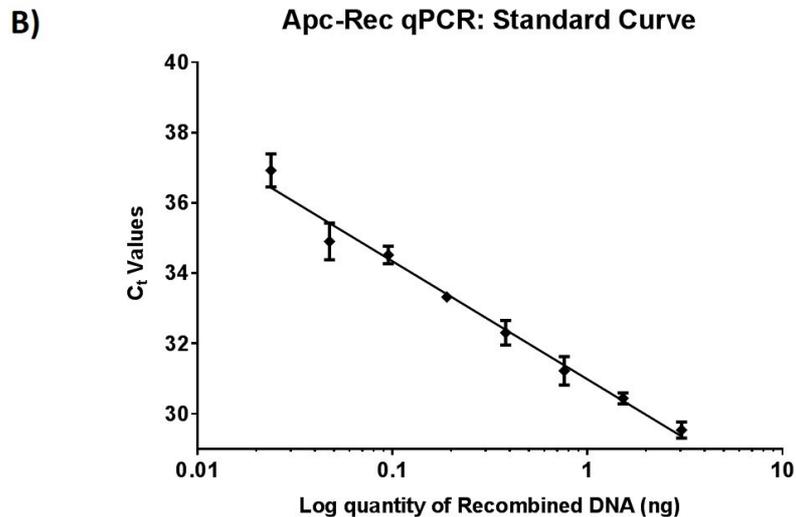
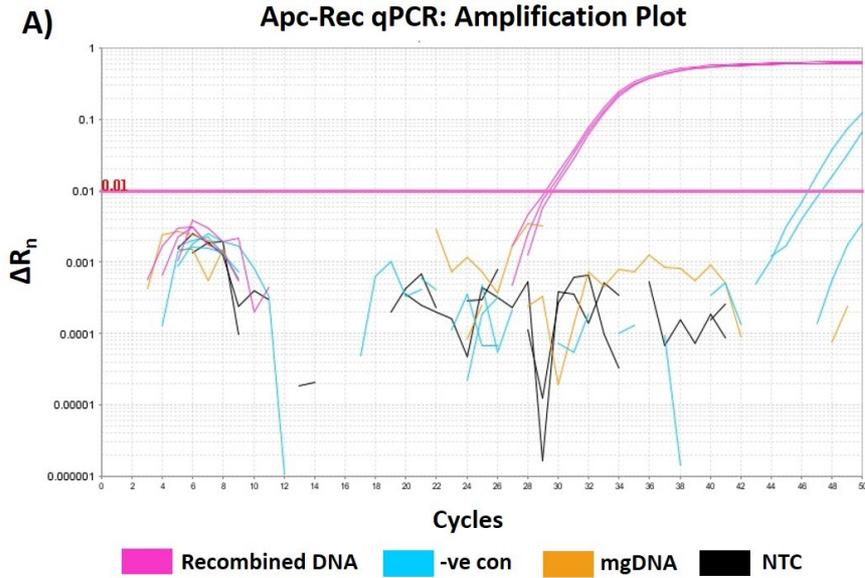


Figure 3-6: Determining Efficiency of the Apc-Rec qPCR Assay

A) The Apc-Rec assay was run as described in section 2.2.3.3.3. The positive (recombined DNA) control was SI tissue DNA (10ng) extracted from an Apc-deleted mouse (5284). The negative (-ve con) control was ear tissue DNA (10ng) from the same mouse. 10ng mgDNA was included as an additional negative control. DNA concentrations were determined using Nanodrop 1000 Spectrophotometer. NTC = no template control. Threshold manually set at 0.01.

B) A standard curve of recombined DNA serial dilutions in triplicates. The starting dilution was 10ng/well, and this was diluted 1:2 ratio for 8 series. The curve has a slope of -3.359, y-intercept at 30.98 and R^2 value of 0.9724. Efficiency was 98.47%.

The Apc-Rec assay amplified the recombined alleles with good specificity and efficiency (98.47%; Figure 3-6). The recombined DNA (positive control) wells showed good amplification signals (pink lines, Figure 3-6-A). Trace fluorescence signals (blue

lines, Figure 3-6-A) were detected for the negative (-ve) control, which was the ear tissue DNA from the same mouse, but these signals had very high C_t cycles (>46), indicating amplification of negligible amounts of DNA that was likely caused by cross-contamination. No amplification was observed for the additional negative control mgDNA well (10ng/well) (orange lines, Figure 3-6-A), validating the target specificity of this assay. This Apc-Rec assay on qPCR was used for analysis of samples from the 1st preclinical study (Chapter 5).

Following this, an Apc-NR assay (section 2.2.3.4.2) was designed and validated for the detection of non-recombined $Apc^{fl/fl}$ alleles, which could be interpreted as background reference DNA. This assay amplified a 75bp region within the PGKneo cassette that was only present in the non-recombined alleles, as this region was excised out during Cre-mediated recombination (Figure 3-4). The use of a sequence-specific TaqMan probe (ApcfI-NRP1) tagged with VIC reporter dye enabled the Apc-Rec and Apc-NR assays to be run in parallel (Figure 3-7). By doing so, the fractional abundance of recombined Apc^{fl} alleles could be calculated using the background non-recombined alleles as a reference.

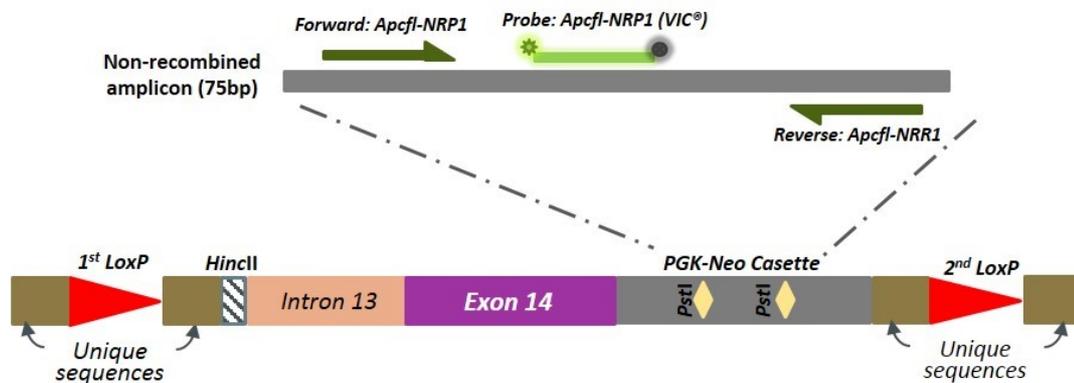


Figure 3-7: Assay Design for Detection of the Non-recombined Alleles

Schematic representation of the non-recombined $Apc^{fl/fl}$ allele. The annealing sites for primers and probe were indicated. See Table 2-8 for sequence details.

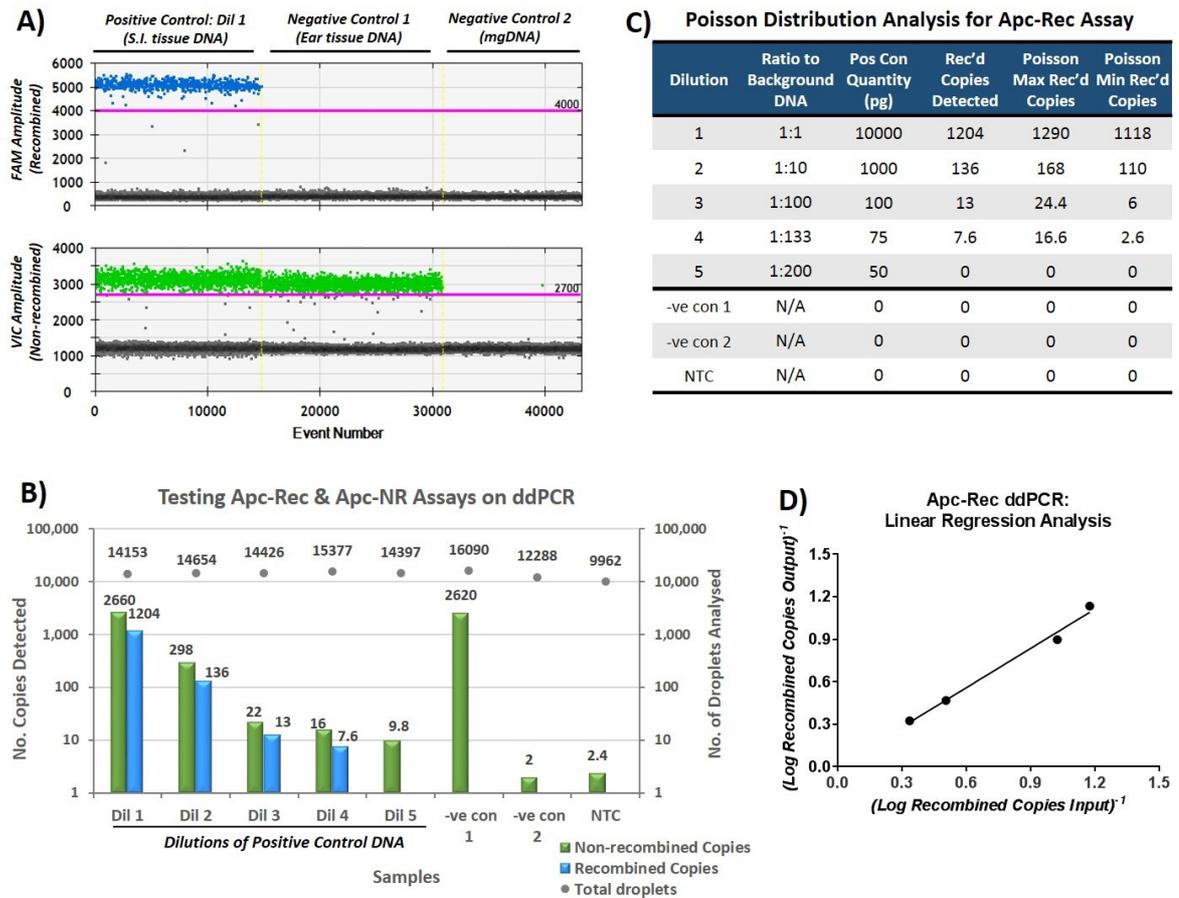


Figure 3-8: Testing the Apc-Rec & Apc-NR Assays on ddPCR

A) Representative amplitude plots for positive and negative controls. 10ng of DNA was tested in each reaction. Each dot represents one droplet. Pink line are manual thresholds.

B) Graph showing the numbers of recombined (FAM – blue) or non-recombined copies (VIC - green) detected on the left y-axis, and the number of droplets analysed (grey dots) on the right y-axis. Positive control was SI tissue DNA from the Apc-deleted mouse (5284). Total quantity was 10ng for all five dilutions (Dil 1-5) tested. Negative control 1 (-ve con 1) was 10ng of ear tissue DNA from the same mouse (5284). Negative control 2 (-ve con 2) was 10ng mgDNA. NTC = no template control.

C) Poisson distribution analysis for the Apc-Rec assay, based on the maximum and minimum error bars of droplet distribution at 95% confidence level (CL).

D) Linear regression analysis - graph plotted as log-log values of recombined copies input (x-axis) and copies detected (y-axis), as a way of evaluating the assay efficiency. The line has the equation $Y = 0.9255X + 0.002226$ with the R^2 value of 0.9882 ($p = 0.0059$).

As the BioRad QX200™ ddPCR system was acquired during the assay development, the validation process of the Apc-Rec and Apc-NR assays was transferred to the ddPCR platform as it provided additional advantages over the qPCR system (Figure

3-8). For validation, SI tissue DNA from the Apc-deleted mouse (5284) was used as a positive control, and was tested at different dilutions in excess of background mgDNA to assess the LoD for the recombined alleles. For all dilutions, total amount of DNA was kept constant at 10ng per reaction. The Apc-Rec assay showed specific detection of the recombined alleles as FAM signal (blue) was apparent only in the reaction wells containing the positive control (Dil 1 to 4; Figure 3-8-B). As before, the ear tissue DNA from the same animal (5284) were used as a negative control (-ve con 1; Figure 3-8-B) for the Apc-Rec assay, and the lack of FAM signal (blue) confirmed the target specificity of this assay. Additionally, mgDNA was included as a negative control (-ve con 2) for both Apc-Rec and Apc-NR assays. Few signals for non-recombined amplification (VIC, green) were observed in the mgDNA (-ve con 2) and NTC reaction wells (Figure 3-8-B), which were likely caused by cross-contamination either during sample preparation or during the droplet transfer process. It could also suggest false positives brought about by the background noise, and the stringency of positive calling might be improved by implementing open-access bioinformatics tools such as 'ddpcRquant' for better threshold determination and 'definetherain' for more accurate calling of low target copies [353, 354]. However, both mgDNA and NTC wells were negative for the recombined amplification (FAM, blue). Indeed, the risk of accidental contamination was a drawback associated with highly sensitive techniques such as ddPCR, despite preparing the reaction in a UV-sterilised PCR station to minimise cross-contamination.

The data Figure 3-8 also informed the LoD of the Apc-Rec assay; LoD is defined as the lowest target concentration (i.e. recombined copies) that can be distinguished from negative (non-target) controls [348]. In the above experiment, the lowest number of recombined copies detected was 7.6 copies in a background of approximately 3030 haploid genome equivalents in dilution 4 (Figure 3-8-B), assuming that one haploid human genome is 3.3pg of DNA [348]. Therefore, the number of copies detected was above the 3.0 copies threshold, indicating that the dilution 4 represented the LoD for the Apc-Rec assay, which was calculated to be $[(7.6/3030)*100]$ 0.25%. At 0.25% LoD, the Apc-Rec assay can potentially detect 1 recombined copy in a background of 400 copies or $(400*3.3\text{pg})$ 1,320pg DNA, assuming there is no sample loss during droplet generation and no PCR inhibitors are present in the reaction well [348].

3.5 Summary

The first aim of my project was to investigate the biomarker potential of tumour-derived cfDNA for detection and monitoring of early disease in the *Cre-Apc^{fl/fl}* model of CRC. To this end, the first objective was to obtain the sequence details for the upstream and downstream *loxP* sites flanking exon 14 in the *Apc^{fl/fl}* allele [144]. Using the Apc-lox1 assay, the upstream *loxP* site (section 3.2) was amplified and sequenced. Sequencing of the downstream *loxP* site proved to be challenging as this site was adjacent to the inserted 4,575bp PGKneo cassette (section 3.3). However, using a semi-nested Apc-lox2 PCR, a 1,596bp long sequence data was retrieved containing a portion of the inserted PGKneobpA plasmid, the vector β -lactamase, downstream *loxP* region and the intron 14 of *Apc* (Figure 3-3). Interestingly, both sequenced *loxP* regions contained a G>T base change when compared with the consensus *loxP* sequence, but no significant hindrance of recombination was observed due to this change [144, 146]. Most importantly, these sequencing data identified the unique sequences adjacent to *loxP* sites, and this proved essential for designing of PCR-based assays described in section 3.4.

Using the sequencing data for the two *loxP* sites, I designed and tested custom PCR-based assays to detect the recombined and non-recombined alleles. The Apc-Rec assay was designed to detect the recombined alleles as a surrogate cfDNA biomarker, as it was the best approximation of tumour-derived cfDNA in this mouse model. This assay was first designed and optimised for the qPCR system (section 3.4), and used for analysis of samples from the 1st *in vivo* study as discussed in the next chapter. Subsequent acquisition of the QX200 ddPCR system prompted transfer of the optimisation process to the ddPCR system as it offered advantages over the qPCR system (section 1.8.2). Using dilution series experiments, the Apc-Rec assay was shown to be specific for the recombined target with the LoD of 0.25% (Figure 3-8). Moreover, the Apc-NR assay, designed as a complementary test to the Apc-Rec assay, was used to detect non-recombined alleles as a reference for fractional quantification of the recombined copies. This was made possible by using TaqMan probes tagged with different fluorophores (VIC for Apc-NR, FAM for Apc-Rec), so that these assays could be run in parallel. This facilitated a quick and simple workflow, and allowed direct quantification of fractional abundances of the recombined alleles for the analysis of preclinical samples, which will be discussed in the next chapter.

Chapter 4 Investigating the Potential of Plasma cfDNA as a Surrogate Biomarker for Preclinical Studies

4.1 Introduction

Several studies with xenograft models have shown that elevation of cfDNA was correlated to tumour burden, and that the cfDNA from tumour cells represented more than half of total cfDNA in these xenograft models [287, 307, 355]. However, preclinical studies have focussed on xenograft models, which do not emulate pathophysiological environment of early tumourigenesis. Therefore, this project aimed to use the *Cre-Apc^{fl/fl}* model of CRC GEMM to understand cfDNA dynamics during an early disease stage. In this model, Cre-mediated recombination of the *Apc^{fl/fl}* alleles was targeted to the Lgr5-expressing CBC cells to initiate adenoma growth (section 1.5), and recombined alleles were present in subsequent developing adenomas. The hypothesis was that eventual identification of these alleles in the plasma could occur if cells from precancerous tissues shed DNA into the circulation, and an increase in disease burden would be reflected in the elevation of recombined DNA levels in the plasma. Therefore, work in this chapter aimed to assess whether the recombined alleles could be detected in the plasma of adenoma-bearing mice, and if so, how early they could be detected, and whether cfDNA levels correlated with tumour burden. In parallel, total plasma cfDNA levels were quantitatively assessed to see if they were elevated in the presence of lesions, similar to what has been reported in humans [299, 312, 319].

To this end, three *in vivo* studies were conducted and data from each experiment was used to refine the design of the following investigations. The last *in vivo* study also involved analysis on whether dietary fat content has an influence on cfDNA dynamics during early stages of tumourigenesis. Studies have indicated that high-fat content promotes tumorigenesis in preclinical models [225, 356, 357]. In humans, obesity is the 2nd highest risk factor for cancer in the UK, and is implicated in 11% of CRC cases in Europe [182, 183, 358]. Therefore, I reasoned that HFD could aggravate the tumour burden in the *Cre-Apc^{fl/fl}* mice and influence the detection of recombined alleles in plasma. The data from this study could help us understand the influence of different dietary fat contents on cfDNA dynamics, which has yet to be properly investigated. Moreover, as the detection of recombined alleles proved to be arduous in the plasma

of adenoma-bearing animals in the first two studies, this prompted an investigation of whether recombined alleles were present at detectable levels in the faecal matters. Studies with CRC patients have shown that genetic/epigenetic changes are detectable in faecal DNA samples [317, 359-362]. Therefore, analysis of stool samples from the 3rd study could inform its potential as an alternative procedure to test tumour-related mutations in preclinical models.

4.2 1st *In Vivo* Study

4.2.1 Study Groups

This study involved 13 *Cre-Apc^{fl/fl}* mice (12 females, 1 male) aged 7-9 months with the starting weights of 24-37g (Table 4-1). 6 *Cre^{+ve}-Apc^{fl/fl}* were allocated to the 'Apc-deleted' group, where they were injected intraperitoneally with 3mg tamoxifen to induce adenoma development. The control group consisted of 4 *Cre^{-ve}-Apc^{fl/fl}* mice, which were injected with 3mg tamoxifen to eliminate the possibility of tamoxifen-independent basal Cre activity, and 3 *Cre^{+ve}-Apc^{fl/fl}*, which were injected with vehicle.

Table 4-1: Genotype, Phenotype & Health Data for Mice used in 1st *In Vivo* Study

	Mouse ID	Lgr5-Cre Genotype (+ve / -ve)	Tamoxifen (T) or Vehicle (V) Injection	Starting Weight (g)	End Weight (g)	No. of Days in study	Reason for Culling
Controls	5173	-ve	T	34	35	57	End of study
	5180	-ve	T	36	36	57	End of study
	5193	-ve	T	36	34	57	End of study
	5195	-ve	T	28	28	57	End of study
	5174	+ve	V	26	27	57	End of study
	5177	+ve	V	30	30	57	End of study
	5181	+ve	V	29	30	34	Showed >2 signs of sickness*
Apc-deleted	5178	+ve	T	24	23	29	Stroke – emergency culling
	5182	+ve	T	28	28	57	End of study
	5183	+ve	T	32	31	57	End of study
	5185	+ve	T	25	23	50	Showed >2 signs of sickness*
	5179	+ve	T	31	31	57	End of study
	5207	+ve	T	31	29	57	End of study

**Signs of sickness include, and not limited to, hunched posture, drastic weight loss, pale feet due to blood loss, raised fur and lethargic movement.*

Animals were euthanised for tissue and plasma collection at 8-week post tamoxifen-induction (section 2.2.5.4), unless they had to be euthanised early due to sickness (Table 4-1). The 8-week end-point was established based on previous findings in our laboratory and the data from literature [9], which suggested that adenomas are detectable at this stage. One distressed mouse (5178), probably affected by a stroke, had to be culled at day 33. Two animals were culled at day 36 (5181) and at day 54 (5185) as they showed signs of distress and poor health status (Table 4-1). Mouse 5181 was a vehicle-injected $Cre^{+ve}-Apc^{fl/fl}$ mouse, and no tumours were observed when tissues were inspected post-mortem, so its health deterioration was independent of adenoma development (Table 4-1).

4.2.2 Histopathological Analysis

Macroscopic polyps were observed only in mouse 5185 during the tissue collection and assessment of H&E histological sections identified micro-adenomas in 3 additional animals.

Table 4-2: Polyp Counts in *Cre-Apc^{fl/fl}* Mice from 1st Study

	Mouse ID	<i>Lgr5-Cre</i> Genotype (+ve/-ve)	Tamoxifen (T)/ Vehicle (V) Injection	Number of Polyps		
				S.I	Colon	Caecum
Controls	5173	-ve	T	N/A	N/A	N/A
	5180	-ve	T	0	0	0
	5193	-ve	T	0	0	0
	5195	-ve	T	0	0	0
	5174	+ve	V	0	0	0
	5177	+ve	V	0	0	0
	5181	+ve	V	0	0	0
Apc-deleted	5178	+ve	T	2	0	1
	5182	+ve	T	1	0	3
	5183	+ve	T	0	0	0
	5185	+ve	T	13	2	1*
	5179	+ve	T	0	0	0
	5207	+ve	T	0	0	3

* Polyp in the caecum of 5185 showed high grade dysplasia.

Control mice did not present any adenoma growth as expected (Table 4-2). Tissue samples were unavailable for mouse 5173. With the exception of mouse 5185, the numbers of SI polyps observed were very low in the Apc-deleted mice, and no lesions were observed in two animals (5183 and 5179) (Table 4-2). The relative low tumour burden observed in this study was in contrast to what was reported by Clevers' group – they observed multiple macroscopic adenomas at 24 days post-induction in their *Cre-Apc^{fl/fl}* cohort [8]. Based on this, an 8-week end-point was set as it was expected that the Apc-deleted group would show significant disease burden at this stage. However, this difference between the two studies could be explained by the physiological differences between our animal cohorts – my *Cre-Apc^{fl/fl}* mice were significantly older at 7-9 months and weighed more when compared to 6-8 weeks old mice used by Clevers' group [8]. Age-related metabolic and physiological differences might have therefore contributed to the milder tumour development observed in this experiment, and one study has shown that older mice exhibited less aggressive tumour development in response to tumour cell implants compared to younger mice in a physiological process hypothesised as age-dependent cancer resistance [363].

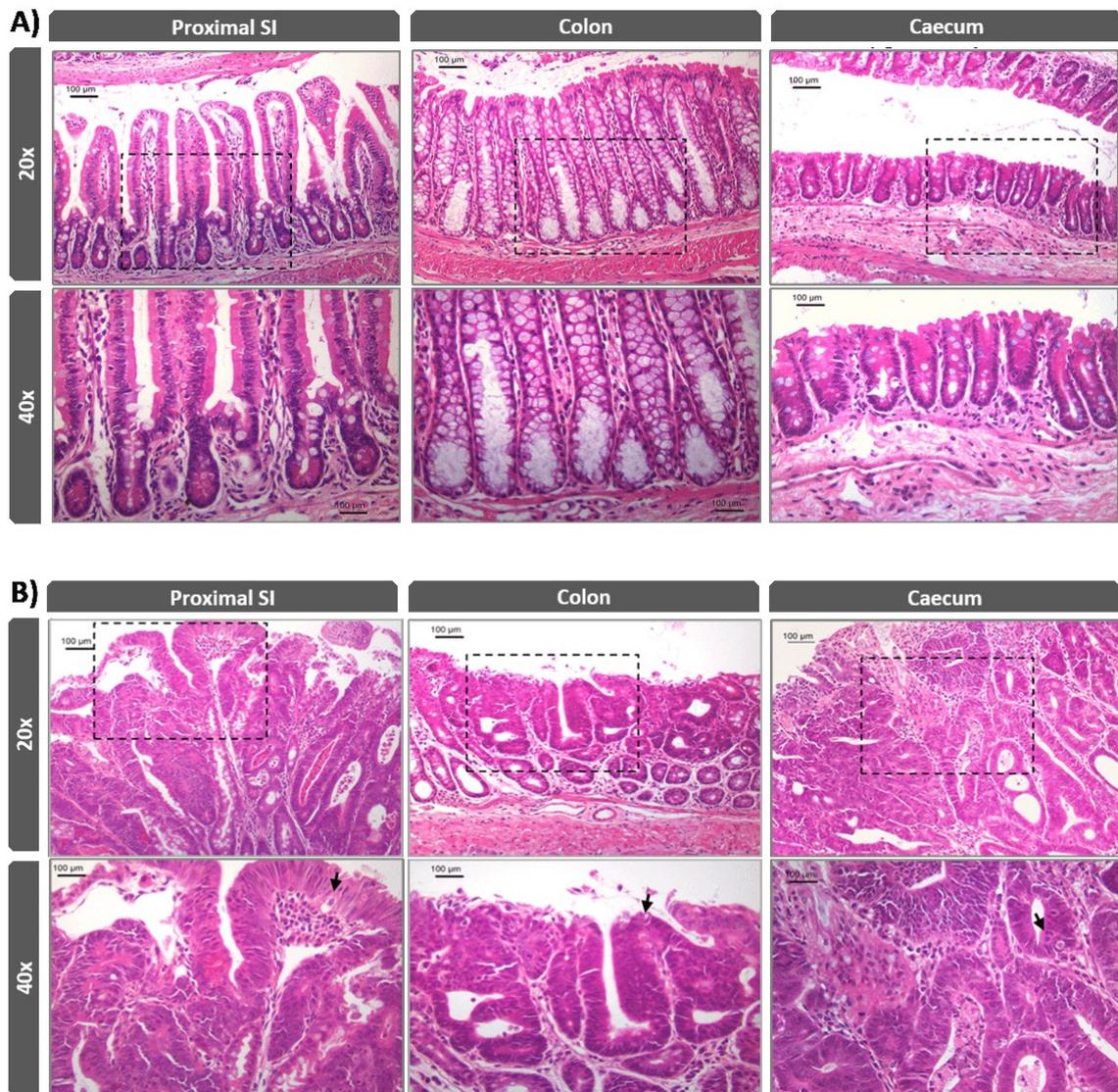


Figure 4-1: Comparison of Histopathological Features in Control & Apc-deleted Mice

Photos of H&E stained sections acquired using Leica DM2500 light microscope. Insets in the upper panels (20x) marked the regions magnified in the lower panels (40x).

A) Samples from a control mouse (5193). Intestinal epithelium showed normal morphology with intact submucosa and muscularis propria layers.

B) Samples from an Apc-deleted mouse (5185). Dysplasia was evident in the adenomatous regions with loss of basal-apical polarity, loss of mucinous cells and narrowing of the intraglandular stroma. Presence of mitotic cells (black arrows) indicated sustained proliferation.

H&E analysis confirmed the presence of adenomatous lesions in Apc-deleted mice, which showed structural abnormalities such as crowding of cells with elongated, stratified nuclei along the epithelial lining, and loss of apical-basal polarity [364] (Figure 4-1-B). Lack of goblet cells was evident, as well as the presence of mitotic cells (black

arrows; Figure 4-1-B) suggesting sustained proliferation. Similar characteristics have been observed in APC-mutant human polyps, confirming the validity of this experimental mouse model [111]. Interestingly, growth of adenomas was observed in the caecum, which was not reported in the initial descriptions of the *Cre-Apc^{fl/fl}* model [9, 155]. However, Vermeulen's group reported the presence of large caecal tumours in a similar mouse model with the genotype *Lgr5-Cre^{ER}.Bcl^{fl/fl}; Apc^{fl/fl}*, which prompted culling of these mice earlier than expected due to a high disease burden [365]. Caecal lesions are particularly vicious as they obstruct the passage of food, causing rapid weight loss and frequently results in culling of animals within shorter intervals after tamoxifen treatment [365].

4.2.3 Quantification of Total Plasma cfDNA Levels

Based on the studies that reported elevated total cfDNA levels in correlation to disease burden [287, 299, 319], a repeat experiment was performed with the *Cre-Apc^{fl/fl}* mice to assess if the same observation could be made. Plasma samples were processed (section 2.2.2.1.2) and quantified using the mGapdh qPCR assay (section 2.2.3.3.2).

Despite a trend towards increased cfDNA levels in *Apc*-deleted animals (75.42 ± 68.31 ng/ml) compared to the controls (35.84 ± 26.68 ng/ml), no statistically significant difference was observed between the two groups ($p=0.2949$) (Figure 4-2-C). However, the presence of an outlier mouse (5178, Figure 4-2-B) with 204.80 ng/ml plasma cfDNA level, probably biased the analysis and contributed to the higher mean value of total cfDNA levels in the *Apc*-deleted group. Interestingly, this was the mouse that had to be culled as an emergency due to a suspected stroke (Table 4-1), and this underlying condition might have contributed to an increase in cfDNA levels – it has been reported that up to a 5-fold increase in total cfDNA can be observed in patients who died from an acute stroke [366]. Another possibility was that the high cfDNA content came from leukocytic contamination during sample collection – occasionally, there were few instances where it was difficult to draw blood via cardiac puncture and that unfortunately resulted in clotting. Overall, irrespective of this outlier, this initial study

failed to identify an increase in total cfDNA content in the Apc-deleted mice with early lesions compared to the controls.

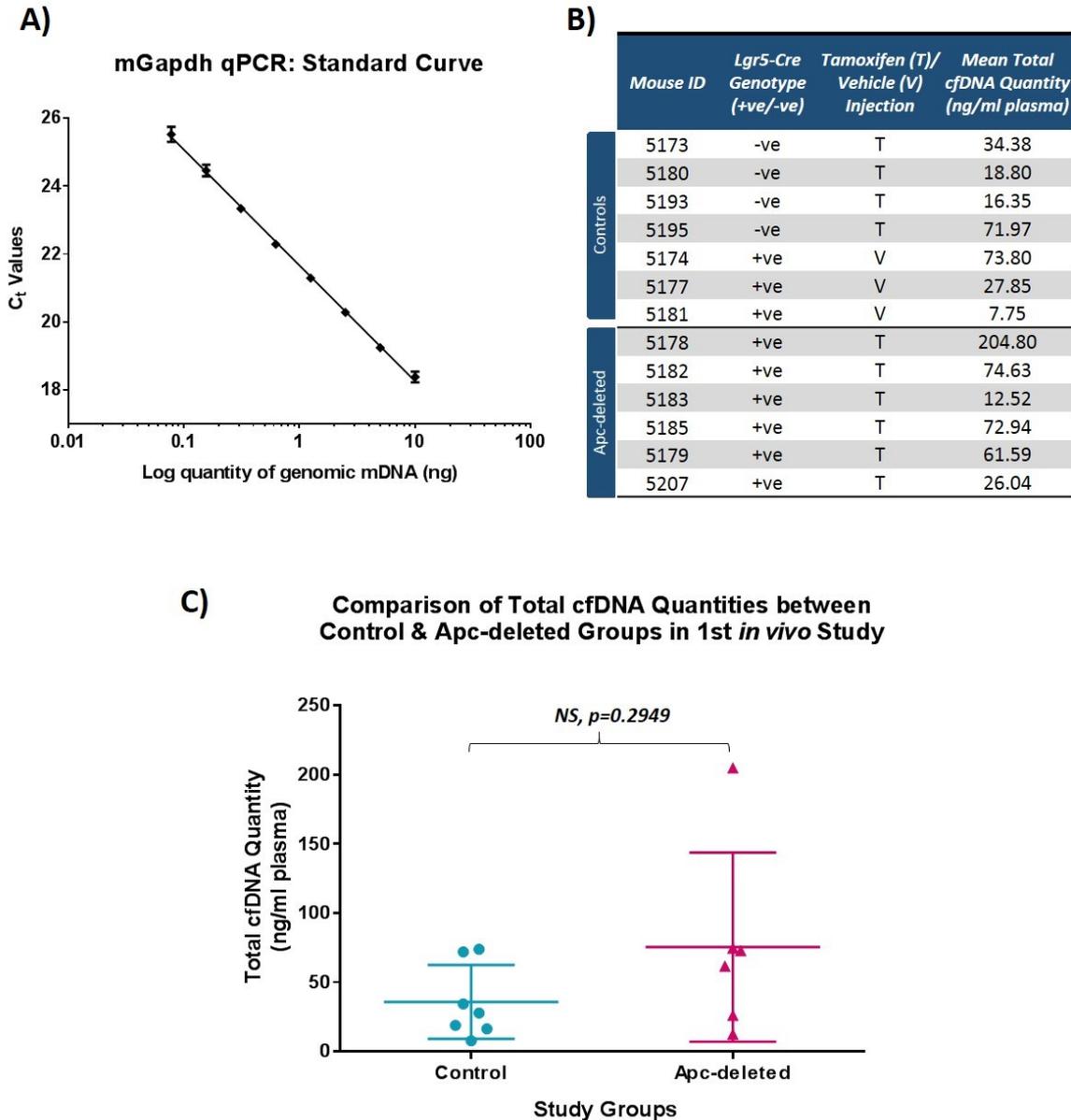


Figure 4-2: Comparing Total cfDNA Quantities in the Plasma of Control & Apc-deleted *Cre-Apc^{fl/fl}* Mice

The mGapdh assay was run twice, and each sample was run in duplicate per assay run. 3.0µl plasma sample was tested per reaction.

A) Standard Curve of C_t values (y-axis) against serial dilutions of the mgDNA standard. It has the slope of -3.408, y-intercept at 21.67 and R^2 value of 0.9966. Assay efficiency was 96.53%.

B) Table listing the details of study mice and the mean total cfDNA quantities in plasma.

C) Graph comparing total cfDNA quantities between two study groups. Each dot represents one mouse, and horizontal lines are mean \pm SD. Mean total cfDNA quantities are 35.84 (\pm 26.68) ng/ml for the control group and 75.42 (\pm 68.31) ng/ml for the Apc-deleted group ($p=0.2949$ at 95% CL; Mann-Whitney test, non-parametric, unpaired).

4.2.4 Checking for Recombined Alleles in the Plasma

Plasma samples were also analysed using the Apc-Rec qPCR assay (section 2.2.3.3.3) in the attempt to detect the presence of recombined alleles (Figure 4-3-A). To increase the probability of detection, all the sample aliquots available after the total cfDNA quantification were lyophilised (section 2.2.2.3.1) and reconstituted in 8µl dH₂O, which was sufficient to run duplicate reactions (3.0µl plasma/ reaction).

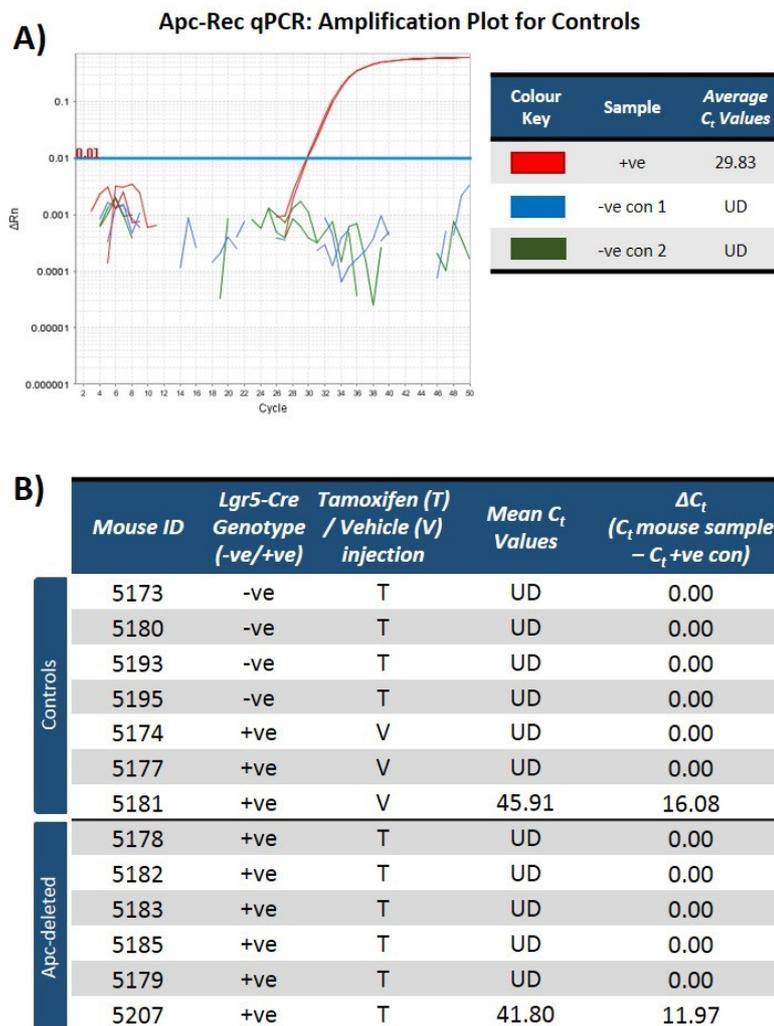


Figure 4-3: Analysis of Plasma Samples to Detect Recombined Alleles

A) Amplification plots. Positive control (+ve, red) was 10ng of SI tissue DNA from the Apc-deleted mouse (5284). Negative control (-ve con 1, blue) was 10ng of ear tissue DNA from the same mouse. 10ng of mgDNA (-ve con 2, green) was an additional negative control. Blue line was the manual threshold.

B) Table listing the average C_t values for amplification of the recombined alleles in control and Apc-deleted mice. Delta C_t (ΔC_t) values represent the difference in C_t cycles between the sample and the reference (+ve con). UD= undetermined.

The Apc-Rec assay showed target amplification in only 1 (5207) out of 6 Apc-deleted mice, and this occurred at a late C_t cycle of 41.80 (Figure 4-3-B). Notably, no amplification of recombined alleles was detected in the plasma of mouse 5185, which showed the highest number of polyps (Table 4-2) and was therefore expected to carry the highest amount of circulating tumour DNA. As C_t values are inversely proportional to the target quantity, a high C_t cycle of 41.80 indicated that the amount of recombined DNA copies present in the plasma was very low – this positive C_t value was only detected in only one of the duplicate wells, which further indicated that the target quantity was likely to be too low for detection. Moreover, the ΔC_t value can be used for comparative quantification of the sample against the positive control reference [367]. For sample 5207, ΔC_t of 11.97 indicated that it contained ($2^{11.97}$) 4012-fold lower recombined DNA quantity compared to the reference, which was a very negligible amount to confidently call it a positive detection. The possibility that this positive detection was artefactual due, for example, to very weak contamination of the original sample cannot be ruled out. In fact, one control mouse (5181) showed trace amplification at the C_t cycle of 45.91 ($\Delta C_t = 16.08$, Figure 4-3-B), which was very likely to be an experimental artefact. In support of this, no polyps were observed in this control mouse when its tissue samples were histologically assessed.

The lack of detection of recombined alleles in the Apc-deleted mice was unlikely to be caused by the plasma DNA extraction process, as cfDNA samples had been successfully amplified using mGapdh qPCR assay in section 4.2.3 (Figure 4-2). Therefore, one deduction was that the disease burden at the time of blood collection (i.e. 8-week post-injection) was yet too low to contribute significantly to the plasma cfDNA pool, a possibility suggested by the low number of adenomas identified during histological assessment of intestinal tissues (Table 4-2). Therefore, this eventuality was taken into consideration for the refinement of study protocols for subsequent *in vivo* studies.

4.3 2nd *In Vivo* Study

4.3.1 Study Groups

The 2nd study was conducted using a refined protocol to increase the disease burden in *Apc*-deleted mice. A total of 17 *Cre-Apc^{fl/fl}* mice (10 females, 7 males), aged 9-12 months and weighed 25-35 grams, were involved (Table 4-3). The *Apc*-deleted group consisted of 8 tamoxifen-injected (T) *Cre^{+ve}-Apc^{fl/fl}* mice (Table 4-3). The control group consisted of 6 tamoxifen-injected (T) *Cre^{-ve}-Apc^{fl/fl}* mice and 3 vehicle-injected (V) *Cre^{+ve}-Apc^{fl/fl}* mice (Table 4-3). The volume of injected tamoxifen was matched to body weight - 1mg injected per 10g body weight –to reduce variability in adenoma growth caused by different degrees of tamoxifen-mediated recombination.

Table 4-3: Genotype, Phenotype & Health Data for Mice used in 2nd *In Vivo* Study

	Mouse ID	Lgr5-Cre Genotype (+ve / -ve)	Tamoxifen (T) /Vehicle (V) Injection	Starting Weight (g)	End Weight (g)	No. of Days Survived	Reason for Culling
Controls	5228	+ve	V	33.5	34.6	81	End of study
	5246	+ve	V	24.1	24.4	58	>2 signs of sickness
	5247	+ve	V	24.2	23.0	34	>2 signs of sickness
	5254	-ve	T	34.3	34.4	81	End of study
	5255	-ve	T	34.0	33.9	81	End of study
	5256	-ve	T	30.0	31.3	45	>2 signs of sickness
	5257	-ve	T	26.4	28.8	40	>2 signs of sickness
	5258	-ve	T	33.2	32.1	81	End of study
	5260	-ve	T	33.5	33.7	81	End of study
Apc-deleted	5225	+ve	T	33.2	32.5	53	>2 signs of sickness
	5226	+ve	T	36.1	36.4	81	>2 signs of sickness
	5227	+ve	T	33.9	30.6	40	>2 signs of sickness
	5242	+ve	T	32.1	31.0	34	>2 signs of sickness
	5243	+ve	T	26.3	25.6	58	>2 signs of sickness
	5244	+ve	T	24.5	24.6	55	>2 signs of sickness
	5245	+ve	T	28.8	28.4	69	>2 signs of sickness
	5248	+ve	T	33.2	32.1	53	>2 signs of sickness

**Signs of sickness include, and not limited to, hunched posture, drastic weight loss, pale feet due to blood loss, raised fur and lethargic movement.*

The end-point for this study was set when the body weight loss reached 15% of the maximum body weight or if animals showed signs of distress (Table 4-3). The end of study was determined when the last Apc-deleted mouse was euthanised when it reached the appointed severity limit: this study ended at 81 days post-induction (Table 4-3). Interestingly, four control mice –two vehicle-injected *Cre^{+ve}-Apc^{fl/fl}* mice and two tamoxifen-injected *Cre^{-ve}-Apc^{fl/fl}* mice – were culled earlier than 81 days due to showing signs of sickness, but no macroscopic tumours were observed when histologically analysed (Table 4-3).

4.3.2 Histopathological Analysis

The modified protocol did produce the intended effect as Apc-deleted mice in this study developed greater numbers of macroscopic polyps (Table 4-4) compared to the previous study (Table 4-2).

Table 4-4: Macroscopic Polyp Counts in *Cre-Apc^{fl/fl}* Mice from the 2nd Study

	Mouse ID	Lgr5-Cre Genotype (+ve / -ve)	Tamoxifen (T) /Vehicle (V) Injection	Number of Polyps		
				SI	Colon	Caecum
Controls	5228	+ve	V	0	0	0
	5246	+ve	V	0	0	0
	5247	+ve	V	0	0	0
	5254	-ve	T	0	0	0
	5255	-ve	T	0	0	0
	5256	-ve	T	0	0	0
	5257	-ve	T	0	0	0
	5258	-ve	T	0	0	0
	5260	-ve	T	0	0	0
Apc-deleted	5225	+ve	T	32	0	0
	5226	+ve	T	5	0	0
	5227	+ve	T	N/A	N/A	N/A
	5242	+ve	T	N/A	N/A	N/A
	5243	+ve	T	34	0	1
	5244	+ve	T	6	0	0
	5245	+ve	T	4	0	0
	5248	+ve	T	4	0	1

Once again, a few Apc-deleted mice (5243 and 5248) developed large caecal lesions that obstructed passage of food through the gut and caused rapid loss of body weight and sickness (Figure 4-4). Unfortunately, macroscopic polyp count data was missing for 2 Apc-deleted mice (5227 and 5242).

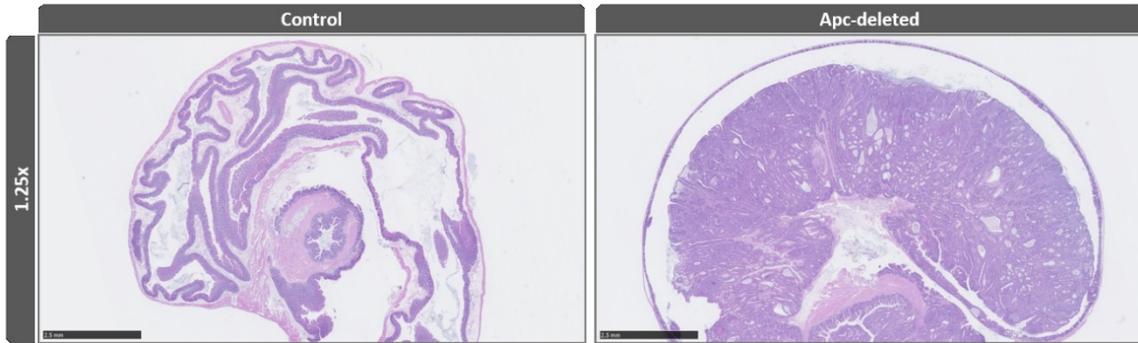


Figure 4-4: Large Polyp Growth in the Caecum of Apc-deleted Mouse

Comparing H&E-stained caecal tissues from the control and Apc-deleted mice. The control (left) was a vehicle-injected $Cre^{+ve}-Apc^{fl/fl}$ mouse (5247) (Table 4-3), showing a polyp-free caecum. The Apc-deleted mouse (right, 5243) was culled at 58 days post-injection due to sickness, and a large polyp was observed in the caecum. Image was captured using the NanoZoomer-XR digital slide scanner C12000 at 1.25x magnification (scale bar= 2.5mm).

Despite the titration of tamoxifen to animal body weight, there was heterogeneity in the lesion numbers. Two Apc-deleted mice (5225 and 5243, Table 4-4) developed more than 30 macroscopic lesions in the proximal SI, whereas the rest of Apc-deleted developed less than 10 lesions in the same region (Table 4-4). Histopathological features of adenomatous growth were observed along the GI tract of an Apc-deleted mouse (Figure 4-5-B). There were hyperplastic areas with epithelial cells that started show loss of basal-apical polarity (Figure 4-5-B). Mitotic features (black arrows) were present, indicating that these cells were proliferating, with the presence of apoptotic debris (red arrows) in the vicinity (40x, Figure 4-5-B). All control animals were free of pathological alterations (Figure 4-5-A). Overall, histopathological analysis suggested that the modified protocol led to substantial adenoma growth in Apc-deleted mice.

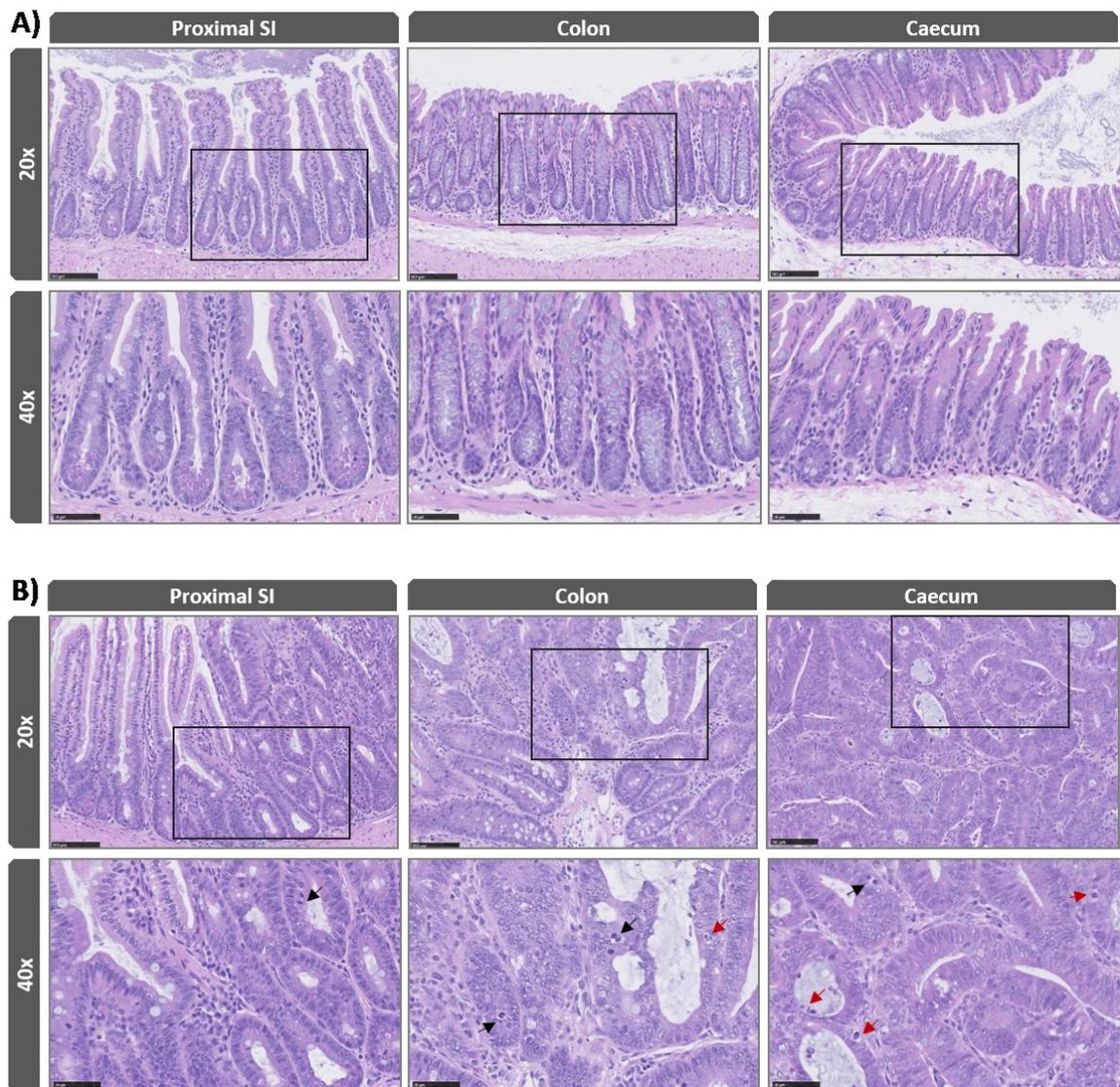


Figure 4-5: Comparison of Histopathological Features in Control & Apc-deleted Mice from 2nd *In Vivo* Study

H&E stained images of proximal SI, colon and caecum. Insets in the upper panels (20x, scale bar= 100 μ m) marked the regions magnified in the lower panels (40x, scale bar= 50 μ m). All images were acquired with NanoZoomer-XR digital slide scanner C12000.

A) Control mouse (5247) showed a normal epithelium with organised crypts and villi. Goblet cells were also present.

B) Apc-deleted mouse (5243) had several adenomas along the GI tract. Presence of mitotic cells (black arrows, 40x) indicated proliferative nature of these lesions. Apoptotic bodies (red arrows) were also observed.

4.3.3 Degree of *Apc^{fl/fl}* Recombination in the Proximal SI

The fractional abundance of the recombined alleles in proximal SI tissues was measured to investigate whether matching tamoxifen injection to body weights had influenced the efficiency of Cre-mediated recombination. Proximal SI region was chosen as it was where most adenoma development was observed. DNA extraction was detailed in section 2.2.2.1.3. Aliquots were quantified using NanoDrop™ 1000 Spectrophotometer and 10ng of each sample was tested using the *Apc-Rec* and *Apc-NR* assays on the ddPCR system.

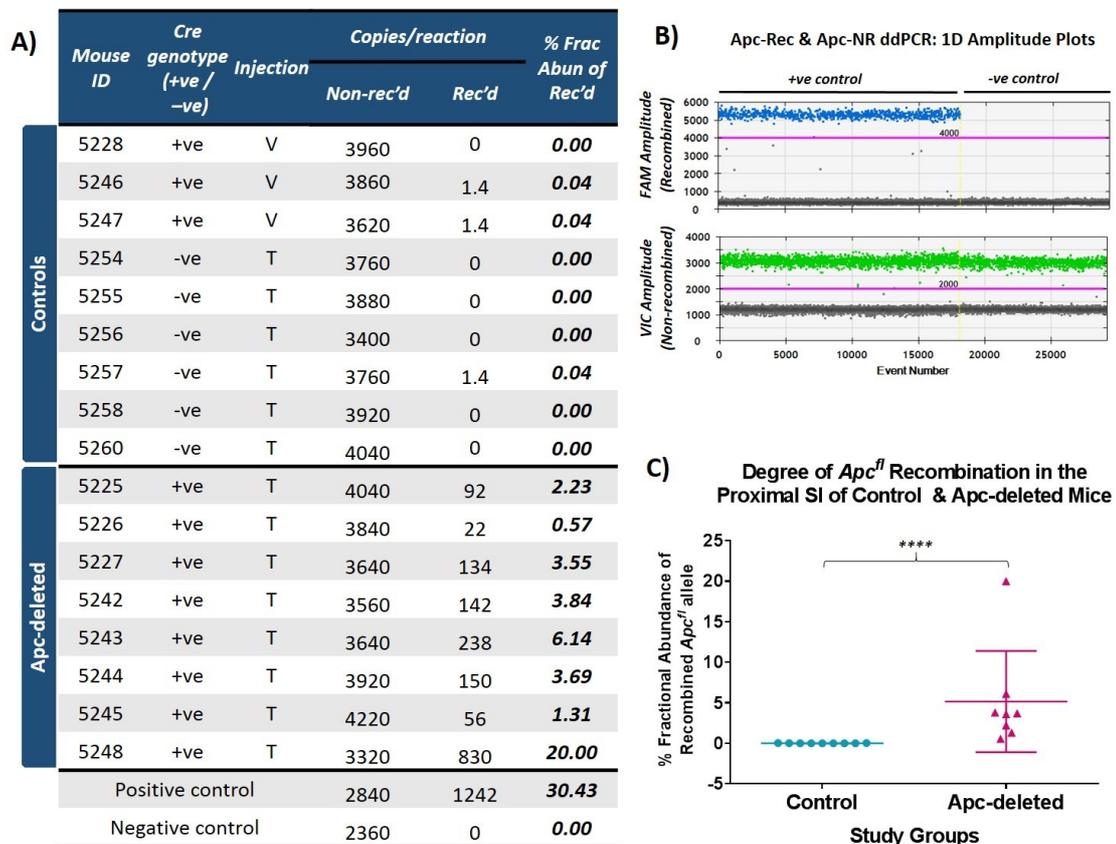


Figure 4-6: Assessment of *Apc^{fl/fl}* Recombination Levels in the Proximal SI

The *Apc-Rec* and *Apc-NR* ddPCR assays were run as described in section 2.2.3.4.2.

A) Table listing the samples analysed with corresponding genotypes, and numbers of non-recombined and recombined allele copies. Positive (+ve) control was SI tissue DNA from the *Apc*-deleted mouse (5284), and negative (-ve) control was ear tissue DNA from the same mouse.

B) 1D amplification plots showing manual threshold gating (pink lines) for the controls.

C) Percentage fractional abundances of recombined alleles in the control (blue) and *Apc*-deleted (pink) mice. Each dot represents one mouse; error bars represent mean \pm SD. Control mice had 0% fractional abundance. *Apc*-deleted mice showed a mean fractional abundance of 5.16 \pm 6.24% ($p < 0.0001$ at 95% CL; Mann-Whitney test, non-parametric, unpaired).

As expected, the control group (blue; Figure 4-6-C) showed very little (<2.0 recombined copies) to no *Apc^{f/f}* recombination – few recombined copies detected were likely due to the background fluorescence. All *Apc*-deleted mice were confirmed to have undergone *Apc^{f/f}* allele recombination in the proximal SI, and that the degree of recombination was fairly consistent between the animals (mean = $5.16 \pm 6.24\%$) (Figure 4-6-C). One outlier mouse (5248) did show 20% fractional abundance, but this did not correlate to the disease burden as the mice only harboured few polyps (four in the proximal SI and one in the caecum; Table 4-4). One possibility was the 5248 tissue used for DNA extraction contained higher fraction of cancerous cells, thus contributing to increased quantity recombined DNA. Nonetheless, this data confirmed that the Cre-mediated *Apc* deletion was induced by tamoxifen injection and highlighted that some variability between study animals was inevitable.

Although the mean recombination efficiency of 5.16 (± 6.24) % might appear low, the DNA aliquots were extracted from whole tissues, including stroma and muscle, which likely contributed to diluting the percentage fractional abundances. Therefore, this data served as an estimation of *Apc^{f/f}* recombination between animals, rather than an absolute measurement. Moreover, it has been reported that the *Lgr5-GFP-IRE5-CreER^{T2}* allele showed mosaic expression of the Cre gene [9, 368], an independent factor that could have also influenced the recombination efficiency. Clevers' team also reported that *Lgr5* stem cells divide symmetrically (i.e. a dividing stem cell generates two daughter stem cells or two daughter transit-amplifying cells) and this leads to each crypt becoming progressively “monoclonal” [13], but Winton's group demonstrated that, despite loss of *Apc* might partially favour stem cells in this drift, WT stem cells remain able to out-compete the *Apc* MU cells in most circumstances [369]. Hence, it was also possible that the low levels of recombination observed in *Apc*-deleted mice might be due to losing recombination-positive stem cells via the neutral drift competition.

4.3.4 Quantification of Total Plasma cfDNA Levels

As with the previous study, total cfDNA levels in the plasma samples were quantified to test whether cfDNA levels were elevated in correlation to adenoma development in *Apc*-deleted mice. Plasma samples were collected and processed as

described in section 4.2.3, and the aliquots were analysed using the mGapdh qPCR assay (section 2.2.3.3.2).

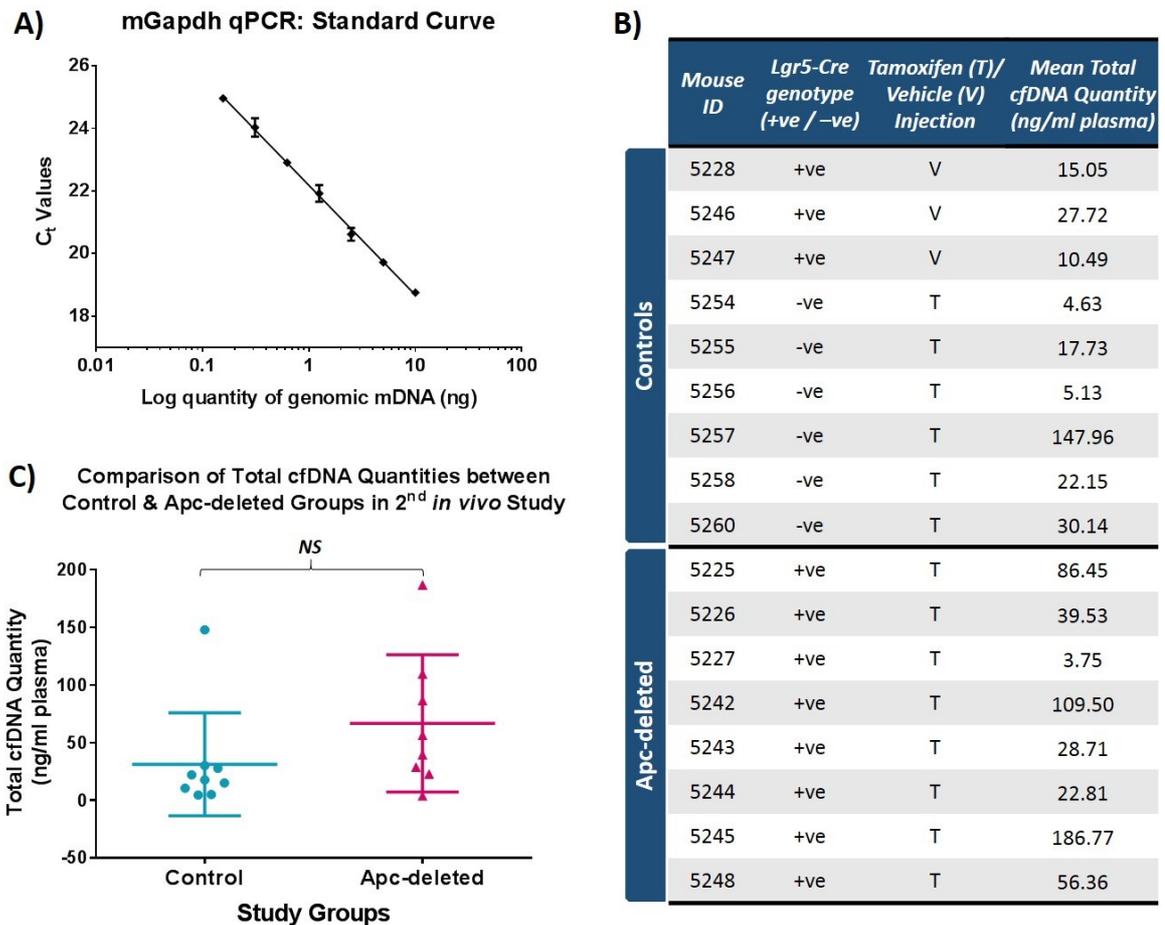


Figure 4-7: Comparing Total cfDNA Quantities in the Plasma of Control & Apc-deleted *Cre-Apc^{fl/fl}* Mice from the 2nd *in vivo* Study

The plasma cfDNA load was quantified using the mGapdh assay. Each sample was run in duplicates and 3.0µl sample was tested per reaction.

A) Standard Curve analysis of C_t values against serial dilutions of mgDNA as a reference. It has the slope of -3.503, y-intercept at 22.18 and R^2 value of 0.9949. Assay efficiency was 92.96%.

B) Table listing the details of study mice and mean total plasma cfDNA quantities.

C) Graph comparing mean total cfDNA loads between two study groups. Each dot represents one mouse, and horizontal lines are mean \pm SD. Mean total cfDNA concentrations were 66.74 ± 59.54 ng/ml for the Apc-deleted group and 31.22 ± 44.70 ng/ml for the control group ($p=0.0927$, 95% CL; Mann-Whitney test, non-parametric and unpaired).

Similar to what was observed in the previous study (Figure 4-2), plasma samples from the Apc-deleted mice showed a trend for elevated levels of total plasma cfDNA (66.74 ± 59.54 ng/ml), compared to the controls (31.22 ± 44.70 ng/ml), but this difference

did not reach statistical significance ($p=0.0927$; Figure 4-7). There were also outliers in each group: control mouse 5257 had 147.96ng/ml and Apc-deleted mouse 5245 had 186.77ng/ml. When the statistical analysis was repeated, but excluding these two samples, the p value (0.0541) approached statistical significance (graph not shown). It remains unclear why some mice show such a spike in cfDNA levels. In the previous study, the outlier mouse was suspected to have suffered a stroke, a condition that could have contributed to the higher cfDNA load (Table 4-1, Figure 4-2) [366]. However, for this analysis, neither mice suffered from sudden debilitating conditions (Table 4-3). It is possible that contamination with leukocyte DNA might have contributed to the spikes in plasma cfDNA levels in these samples [284].

4.3.5 Analysis of Non-recombined & Recombined Alleles in the Plasma

Plasma samples were also analysed using the Apc-Rec and Apc-NR ddPCR assays (section 2.2.3.4.2) to check whether non-recombined and recombined alleles could be detected. As recombination only occurred in the Apc-deleted mice (Figure 4-6), any recombined alleles detected in the plasma would have originated from Lgr5-expressing cells. Prior to analysis, all available aliquots were lyophilised and re-suspended in 10 μ l volumes, and 3.0 μ l of each sample was tested per ddPCR reaction.

Different amounts of non-recombined copies of the *Apc^{fl}* allele were detected across the samples (Figure 4-8), mirroring the variation in total cfDNA loads quantified by the mGapdh assay (Figure 4-7). No copies of the recombined *Apc^{fl}* allele were detected in the control group. Only 2 out of 8 plasma samples (25%) from the Apc-deleted animals showed positive recombined alleles – 1.4 recombined copies were detected for 5227 (0.30% fractional abundance) and 8.0 copies for 5245 (0.15% fractional abundance) (Figure 4-8). The Poisson analysis calculated that, at the minimum level (PoissonMin), 0.0 and 2.8 recombined copies will be detected for 5227 and 5245 respectively. Statistically, positive detection of recombined copies for 5227 may not occur in independent runs and the possibility that 1.4 copies detected might have been due to the background (Figure 4-6) could not be excluded, which suggested that only 5245 could be called as a true positive (Table 4-5).

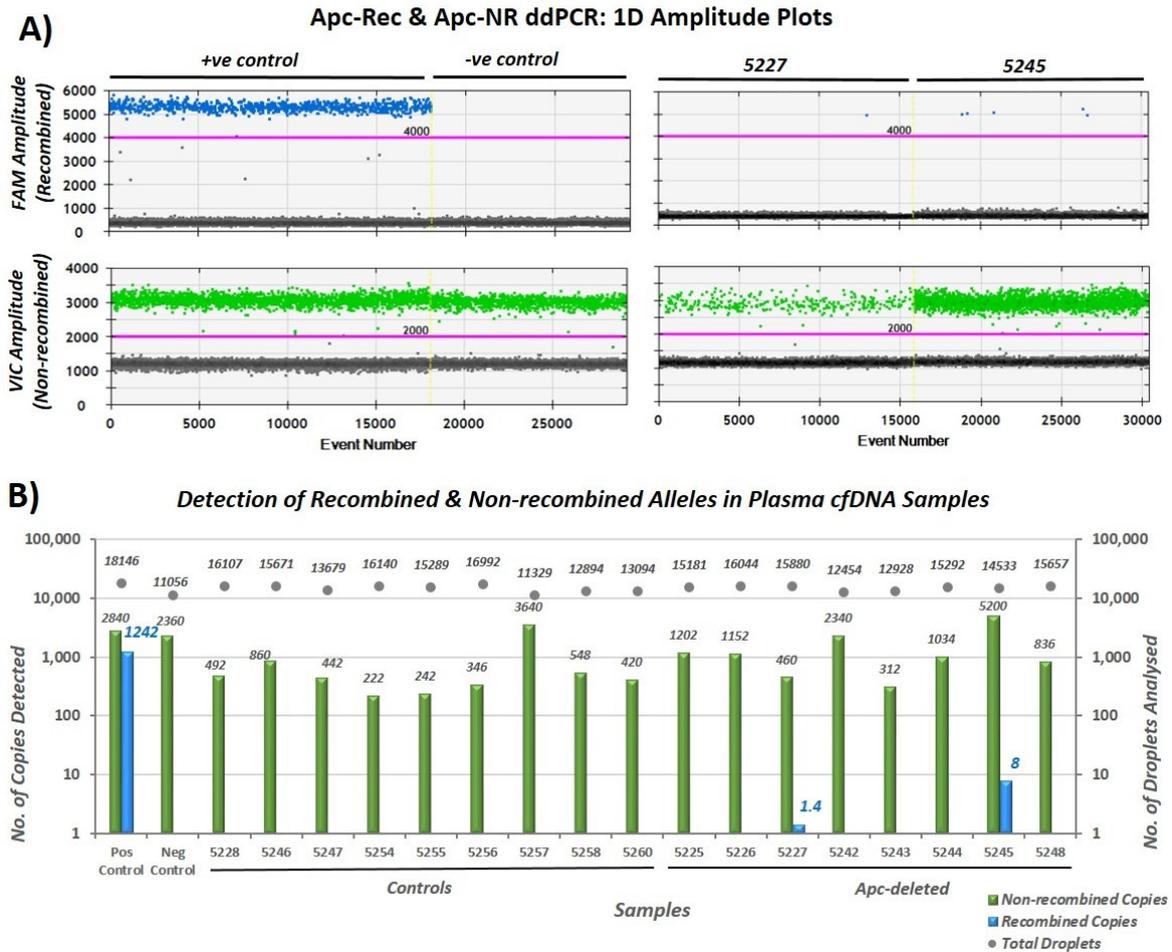


Figure 4-8: Analysis of Non-recombined & Recombined Alleles in Plasma Samples from Control & Apc-deleted Mice in 2nd *in vivo* study

Plasma cfDNA samples were analysed using the Apc-Rec and Apc-NR assays in parallel.

A) 1D amplification plots for controls (left) and samples that were positive for recombined alleles (right). Pink lines are manual thresholds. Each dot represents one droplet. Positive (+ve) control was SI tissue DNA from the Apc-deleted mouse (5284) and negative (-ve) control was ear tissue DNA from the same mouse.

B) Graph showing the numbers of non-recombined (green bars) and recombined (blue bars) copies detected in the control and Apc-deleted mice (left y-axis). Number of droplets analysed (grey dots) for each samples was shown on right y-axis.

Table 4-5: Poisson Analysis for Positive Detection of Recombined Alleles

ID	Concentration (copies/ μ l)	Copies/20 μ l Reaction	PoissonMax (Conc)	PoissonMax (no. of copies)	PoissonMin (Conc)	PoissonMin (no. of copies)
5227	0.07	1.4	0.35	7.0	0.00	0.0
5245	0.4	8.0	0.88	17.6	0.14	2.8

Overall, the 2nd study has shown that detection of recombined alleles in the plasma of Apc-deleted mice was achievable, albeit at a low sensitivity, and the detection of recombined cfDNA occurred at low quantities despite the robustness of disease burden observed in Apc-deleted mice. Furthermore, no correlation between the total cfDNA levels and the disease burden was observed: for instance, mouse 5245 had the highest total cfDNA concentration of 186.77ng/ml despite the relatively low disease burden of 4 polyps in the SI, and in contrast mouse 5243 that showed a high disease burden of 34 SI polyps and one caecal polyp only had a low total cfDNA concentration of 28.71ng/ml (Table 4-4 & Figure 4-7). Overall, this study data suggested that further modification was necessary to improve the detection of recombined cfDNA in the plasma.

4.4 3rd *In Vivo* Study

4.4.1 Study Groups

This study protocol was modified to include allocation of mice into two diet groups – NFD (16.4% Kcal from fat) and HFD (59.7% Kcal from fat) (Table 2-10). The purpose was to stimulate the growth of adenomas in Apc-deleted mice, as studies have reported that high fat intake promotes colorectal tumorigenesis in mice [225, 357]. Due to the introduction of a diet factor, the NFD used in this study was obtained from the same source as the HFD to remove the composition variability. As such, this NFD had a higher calorie content from fat (16.4%) than the maintenance diet (EURodent Diet 5LF2*, 6.61% Kcal from fat) that was provided in the previous two studies (Table 2-10). Mice were allowed to adapt to the new diets for two weeks before tamoxifen injection. The severity limit for humane endpoint was also increased to 20% body weight loss. These changes were implemented to enhance the disease burden to augment DNA release into the circulation. Additionally, as detection of the recombined alleles in plasma had proven to be arduous for the last two studies, faecal samples were also collected to assess whether recombined alleles could be detected in the faeces.

Table 4-6: Genotype, Phenotype & Health Data for Mice used in 3rd *In Vivo* Study

	Mouse ID	Lgr5-Cre Genotype (+ve / -ve)	HFD/NFD	Starting Weight (g)	End Weight (g)	No. of Days Survived	Reason for Culling
Apc-deleted	6363	+ve	HFD	30.2	27.7	65	>2 signs of sickness
	6355	+ve	HFD	28.3	26.9	44	>2 signs of sickness
	6349	+ve	HFD	26.4	23.7	44	>2 signs of sickness
	6368	+ve	HFD	26.6	23.8	64	>2 signs of sickness
	6370	+ve	HFD	29.6	27.1	92	>2 signs of sickness
	6232	+ve	HFD	26.5	25.1	51	>2 signs of sickness
	6150	+ve	HFD	28.3	25.5	30	>2 signs of sickness
	6155	+ve	HFD	26.0	22.1	43	>2 signs of sickness
	6156	+ve	HFD	27.6	23.4	35	>2 signs of sickness
Controls	6167	-ve	HFD	27.6	30.8	92	End of study
	6214	-ve	HFD	26.0	26.8	92	End of study
	6100	-ve	HFD	31.7	36.7	92	End of study
	6242	-ve	HFD	22.3	24.8	92	End of study
Apc-deleted	6356	+ve	NFD	29.4	24.4	58	>2 signs of sickness
	6357	+ve	NFD	28.8	23.5	50	>2 signs of sickness
	6154	+ve	NFD	24.1	21.2	37	>2 signs of sickness
	6236	+ve	NFD	26.0	21.6	57	>2 signs of sickness
	6364	+ve	NFD	27.6	23.0	71	>2 signs of sickness
	6365	+ve	NFD	31.1	26.0	85	>2 signs of sickness
	6228	+ve	NFD	32.6	26.0	35	>2 signs of sickness
	6341	+ve	NFD	27.6	20.2	64	>2 signs of sickness
	6342	+ve	NFD	27.9	24.1	49	>2 signs of sickness
	6348	+ve	NFD	25.8	23.3	72	>2 signs of sickness
Controls	6144	-ve	NFD	26.0	28.2	86	End of study
	6186	-ve	NFD	27.2	26.7	86	End of study
	6172	-ve	NFD	25.8	25.7	86	End of study
	6223	-ve	NFD	31.3	32.9	86	End of study

**Signs of sickness include, and not limited to, hunched posture, drastic weight loss, pale feet due to blood loss, raised fur and lethargic movement.*

A total of 27 mice, aged 6-10 months, were involved in this study (Table 4-6). They were allocated as follow: 9 *Cre*^{+ve}-*Apc*^{fl/fl} and 4 *Cre*^{-ve}-*Apc*^{fl/fl} in HFD, and 10 *Cre*^{+ve}-

$Apc^{fl/fl}$ and 4 $Cre^{ve}-Apc^{fl/fl}$ in NFD. All mice were induced with tamoxifen (1mg per 10g body weight), two weeks after diet allocation. All Apc-deleted mice developed polyps and were culled when they reached the appointed severity limit. Control mice in each group were culled when the last experimental mouse reached the end-point. Control mice did not show any visible sign of sickness during the study and no lesions were observed in post-mortem autopsies (Table 4-6).

4.4.2 Survival Analysis

The Kaplan-Meier survival analysis showed that the Apc-deleted mice in the HFD group had a shorter median survival compared to those in the NFD group – 44.0 days vs. 57.5 days – but the log-rank analysis gave no statistical difference in the survival rates between the two groups ($p=0.7190$; Figure 4-9). The lack of statistical significance might have been partly contributed by the small numbers of study mice used for this analysis.

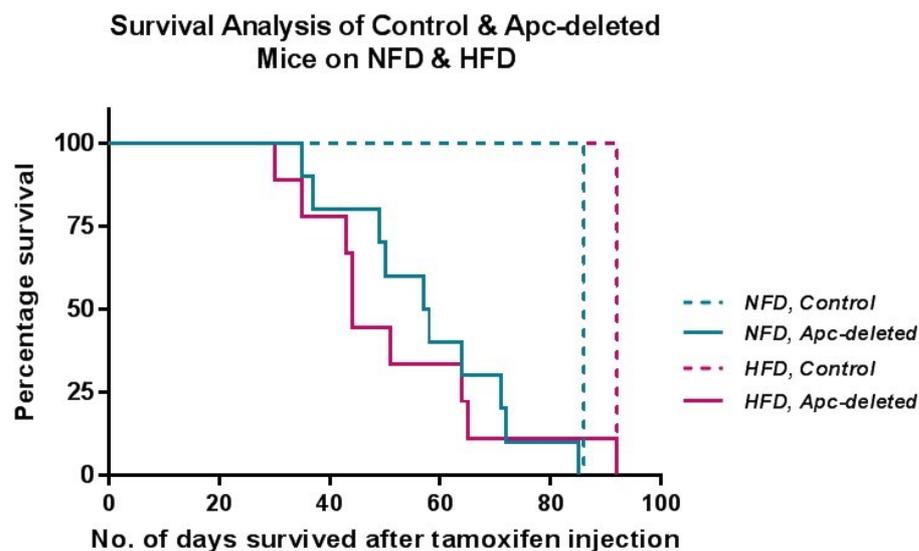


Figure 4-9: Survival Analysis of $Cre-Apc^{fl/fl}$ Mice in the 3rd Study

Days survived were counted from the time of tamoxifen injection to end-point. Control mice in both NFD (dotted blue line) and HFD (dotted pink line) groups were culled when all Apc-deleted mice (solid lines) had reached the end-point for each diet group. Apc-deleted mice on NFD (solid blue line) had median survival of 57.5 days compared to 44.0 days for the HFD mice (solid pink line) ($p = 0.7190$; log-rank test). NFD to HFD hazard ratio is 0.8363.

4.4.3 Histopathological Analysis

Tissue samples were analysed to investigate if the HFD exacerbated tumorigenesis in the Apc-deleted animals. Manual counting was deemed impractical due to the extensive growth of micro-adenomas that were invisible to the naked eye. Instead, H&E stained slides were used to assess the extent of disease burden in these mice. Cross-sections of swiss-rolled tissue samples were used to manually count the number of polyps in the SI, colon and caecum to give an approximation of the disease burden in these mice (Table 4-7). Additional analysis was also performed by Dr P. Greaves (University of Leicester, UK) in the form of histological grading (Table 4-8), using Kozuha's diagnostic criteria (Figure 4-13), to assess the disease burden between the two diet groups [370].

Table 4-7: Microscopic Polyp Counts in *Cre-Apc^{fl/fl}* Mice from the 3rd Study

	Mouse ID	Lgr5-Cre Genotype (+ve / -ve)	HFD/NFD	Number of Polyps		
				SI	Colon	Caecum
Apc-deleted	6363	+ve	HFD	110	0	1
	6355	+ve	HFD	14	0	2
	6349	+ve	HFD	80	0	4
	6368	+ve	HFD	81	0	8
	6370	+ve	HFD	58	0	1
	6232	+ve	HFD	50	0	8
	6150	+ve	HFD	58	0	1
	6155	+ve	HFD	38	0	3
	6156	+ve	HFD	47	0	4
Controls	6167	-ve	HFD	0	0	0
	6214	-ve	HFD	0	0	0
	6100	-ve	HFD	0	0	0
	6242	-ve	HFD	0	0	0
Apc-deleted	6356	+ve	NFD	57	0	4
	6357	+ve	NFD	25	0	6
	6154	+ve	NFD	45	0	4
	6236	+ve	NFD	45	0	9
	6364	+ve	NFD	96	0	8
	6365	+ve	NFD	28	0	10
	6228	+ve	NFD	12	0	6
	6341	+ve	NFD	3	0	3
	6342	+ve	NFD	9	0	3
	6348	+ve	NFD	13	0	5
Controls	6144	-ve	NFD	0	0	0
	6186	-ve	NFD	0	0	0
	6172	-ve	NFD	0	0	0
	6223	-ve	NFD	0	0	0

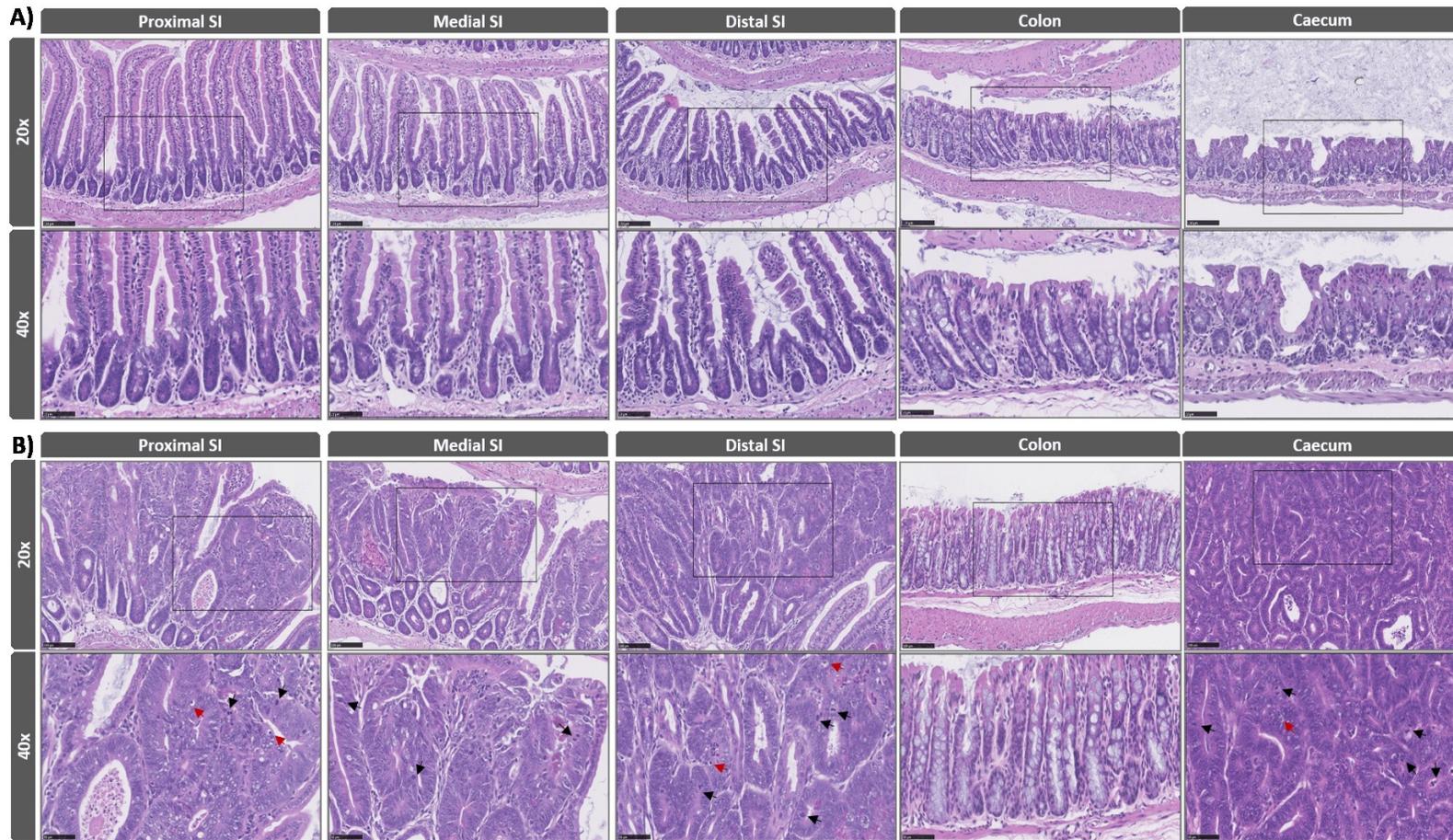


Figure 4-10: Comparison of Histopathological Features in Control & Apc-deleted Mice in the NFD Group

Photos of H&E stained tissues from GI tract. Black box insets in the upper panels (20x, scale bar= 100µm) marked the regions magnified in the lower panels (40x, scale bar= 50µm). Images were acquired using NanoZoomer-XR digital slide scanner and the NDP.view 2 software.

A) Samples from a control mouse (6172). Intestinal epithelium showed normal morphology with polarised nuclei and scattering of goblet cells.

B) Samples from an Apc-deleted mouse (6364). Polyp growth was observed in all but colon. Epithelial dysplasia was evident as shown by the loss of basal-apical polarity, loss of mucinous cells and narrowing of the intra-glandular stroma. Presence of mitotic cells (black arrows) indicated sustained proliferation. Apoptotic bodies (red arrows) were also present.

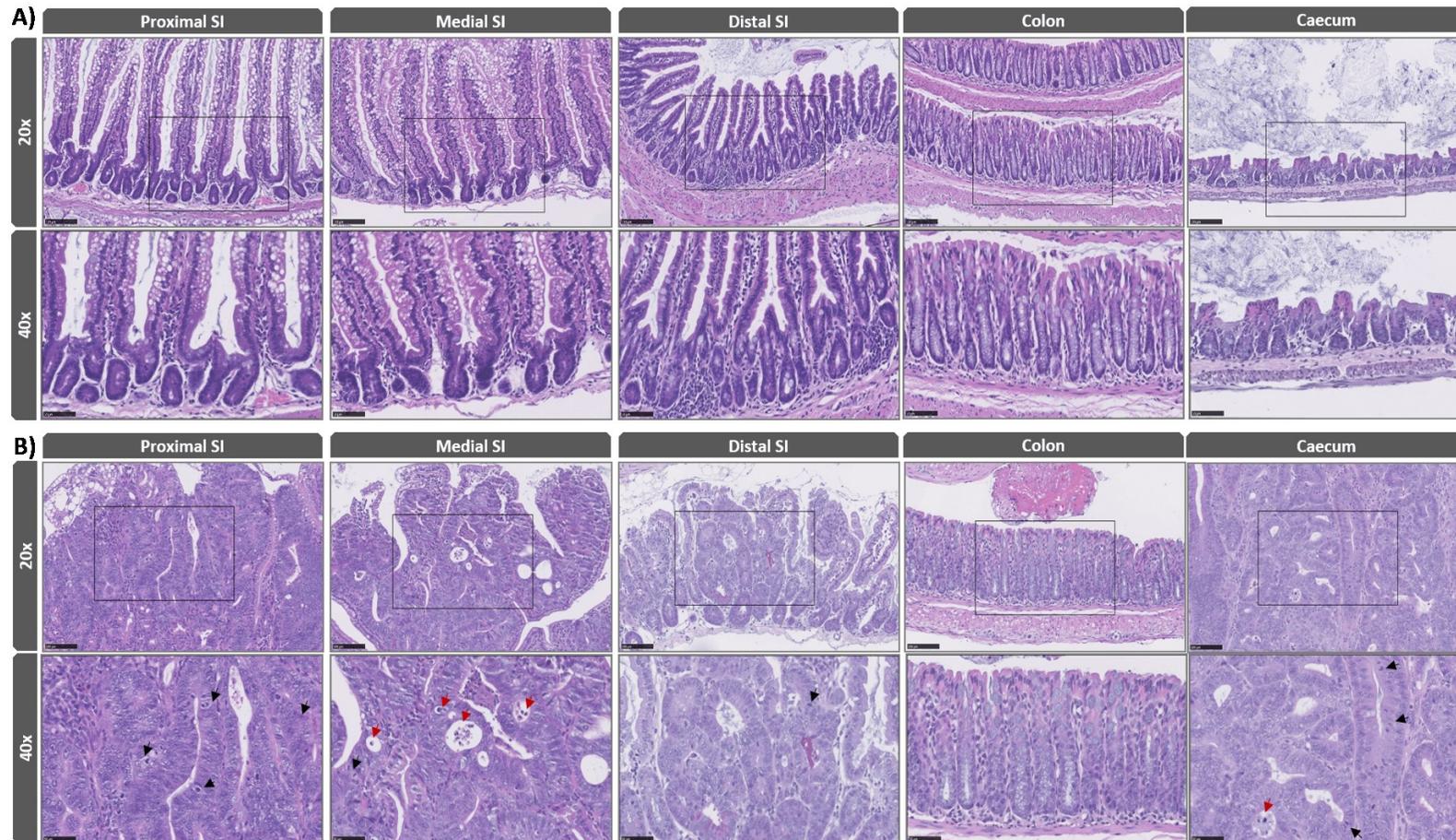


Figure 4-11: Comparison of Histopathological Features in Control & Apc-deleted Mice in the HFD Group

Photos of H&E stained tissues from GI tract. Black box insets in the upper panels (20x, scale bar= 100µm) marked the regions magnified in the lower panels (40x, scale bar= 50µm). Images were acquired using NanoZoomer-XR digital slide scanner and the NDP.view 2 software.

A) Samples from a control mouse (6167) showed normal intestinal epithelium with abundant fat globules, consistent with high fat intake.

B) Samples from an Apc-deleted mouse (6349). Lesions were present in all but colon. Dysplastic epithelium was evident with depolarised nuclei and narrowing of the intra-glandular stroma. Mitotic cells (black arrows) and apoptotic bodies (red arrows) were indicative of proliferation.

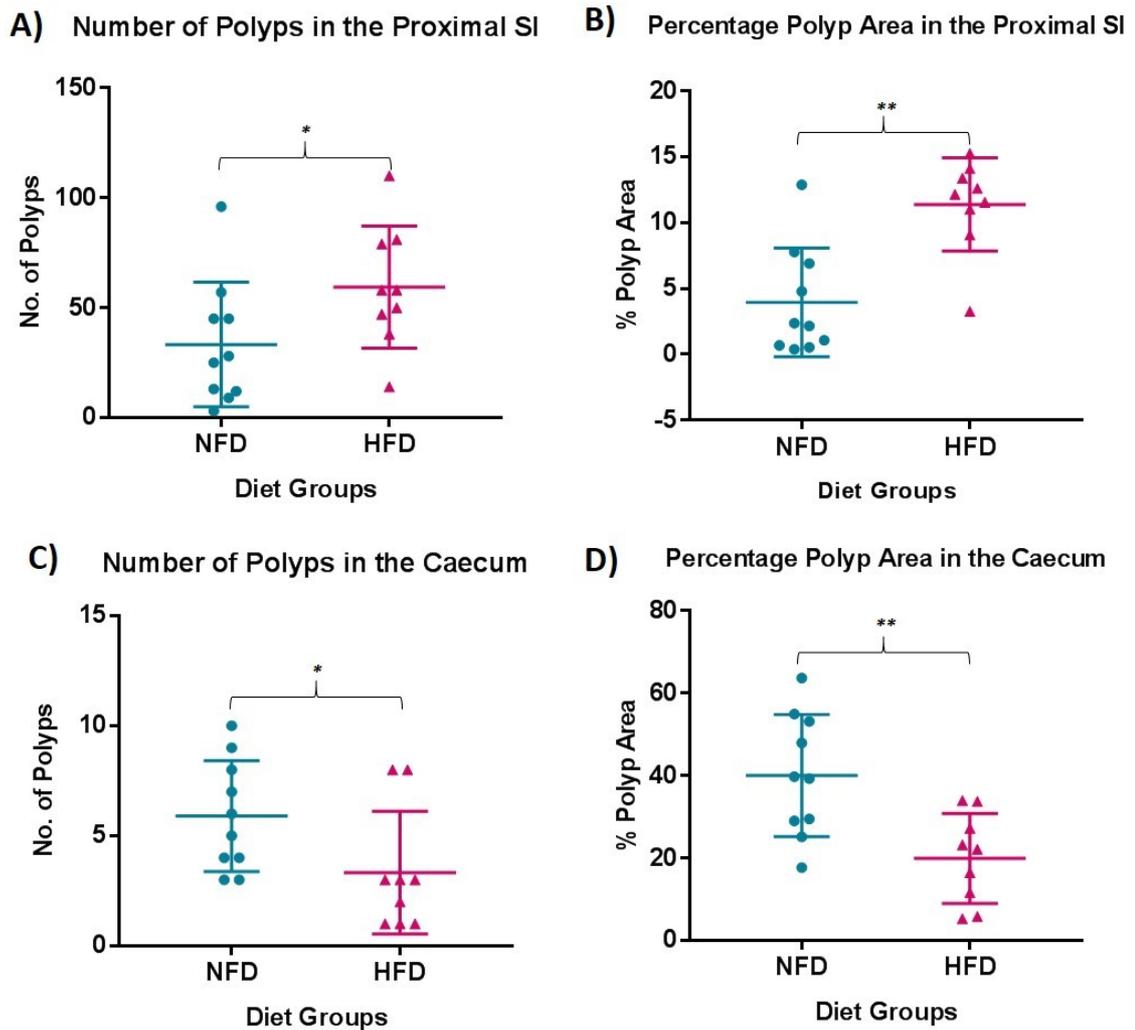


Figure 4-12: Comparing the Numbers & Areas of Polyps in the Proximal SI & Caecum of Apc-deleted Mice in the NFD & HFD Groups

Only Apc-deleted mice were included in this analysis since no tumour growth was observed in control mice. Image analysis was done using the NDP.view 2 software (Hamamatsu®). Number of polyps were counted in H&E stained cross-sectional images of swiss-rolls. Percentage polyp area was calculated as a ratio of adenoma-to-total tissue areas. All statistical analysis was performed using the two-tailed, non-parametric, Mann-Whitney test.

A) Average polyp numbers in the proximal SI were 59.44 ± 27.79 polyps for the HFD group and 33.3 ± 28.36 polyps for the NFD group ($p=0.0262$).

B) Average percentage polyp areas in the proximal SI were $11.38 \pm 3.54\%$ for the HFD group and $3.95 \pm 4.12\%$ for the NFD group ($p=0.0021$).

C) Average polyp numbers in the caecum were 3.33 ± 2.78 polyps for the HFD group and 5.90 ± 2.51 polyps for the NFD group ($p=0.0248$).

D) Average percentage polyp areas in the caecum were $19.96 \pm 10.9\%$ for the HFD group and $40.03 \pm 14.8\%$ for the NFD group ($p=0.0057$).

Table 4-8: Histological Grading of Polyps in the Proximal SI & Caecum of Apc-deleted Mice in NFD & HFD Groups

	Mouse ID	Proximal SI		Caecum	
		Cytology grade	Branching	Cytology grade	Branching
NFD ; Apc-deleted	6356	IV	2	I	0
	6357	III	1	IV	2
	6154	III	2	IV	2
	6236	III	2	IV	2
	6364	III	1	III	2
	6365	III	1	IV	secondary
	6228	III	0	IV	2
	6341	III	2	III	2
	6342	I	0	IV	secondary
	6348	III	1	IV	2
HFD ; Apc-deleted	6363	III	2	II	1
	6355	III	1	III	1
	6349	III	2	III	2
	6368	IV	2	III	2
	6370	III	1	III	2
	6232	IV	2	III	2
	6150	III	1	III	1
	6155	III	2	III	2
	6156	IV	2	IV	2

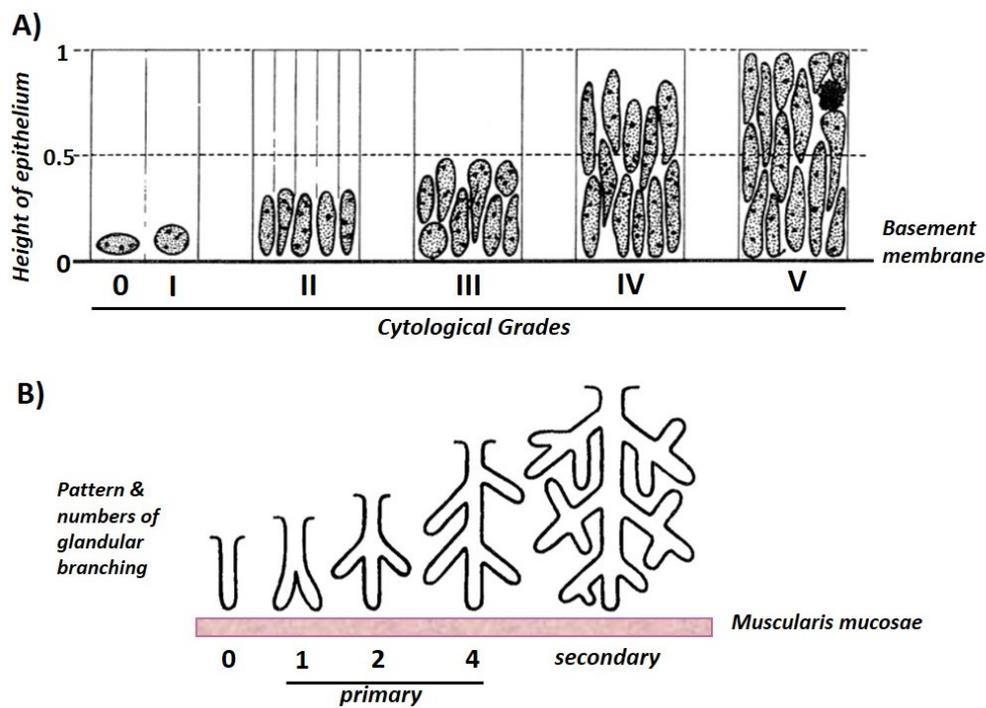


Figure 4-13: Kozuha's Criteria used for Histological Grading in Table 4-8

A) Epithelial pseudo-stratification criteria for cytological grading. Grade 0 is normal. Grade I is near-normal with epithelial elongation, but flattened/cuboidal nuclei remain basal. Grade II shows crowding of cells with elongated nuclei. Grade III shows apparent pseudo-stratification with basal nuclei. Grade IV and V are characterised by severe pseudo-stratification, and grade V also shows scattering of nuclei throughout epithelium. [336]

B) Illustration of pattern and number of glandular branching. Grade 0 is normal (i.e. non-branching). The grade of primary branching is determined by the number of additional branches. Secondary branching is when additional branches stemmed from primary branches. [336]

Polyp growth was concentrated in the proximal SI and caecum, irrespective of the diet group (Figure 4-10 & Figure 4-11), confirming what was observed in the previous studies. Few polyps were also identified in the medial and distal SI regions in both diet groups (Figure 4-11-B), which was not observed in the previous studies, and this was likely due to the dietary changes and the increased severity limit. In the two previous studies, all mice were fed standard diet with lower fat content (6.608% fat) even when compared to the NFD (16.4% fat) in this study (Table 2-10). No polyp growth was observed in the colon (Figure 4-10 & Figure 4-11).

HFD-fed mice showed statistically significant higher number of polyps in the proximal SI compared to the NFD-fed mice (59.44 ± 27.79 polyps vs. 33.3 ± 28.36 polyps, $p=0.0262$; Figure 4-12-A). Most of these polyps were micro-adenomas and thus only become apparent under microscopic analysis. HFD mice also had wider total polyp area compared to the NFD mice ($11.38 \pm 3.54\%$ vs. $3.95 \pm 4.12\%$, $p=0.0021$; Figure 4-12-B) in the proximal SI. This was supported by historical grading where polyps from HFD mice showed higher cytological grading and glandular branching, on average, compared to the NFD mice (Table 4-8 & Figure 4-13). In contrast, when the caecal samples were analysed, HFD mice showed statistically lower polyp count compared to the NFD mice (3.33 ± 2.78 polyps vs. 5.9 ± 2.51 polyps, $p=0.0248$; Figure 4-12-C). Similarly, HFD mice showed significantly reduced polyp area compared to the NFD mice ($19.96 \pm 10.9\%$ vs. $40.03 \pm 14.8\%$, $p=0.0057$; Figure 4-12-D). This data was supported by histological grading where caecal polyps from NFD mice showed increased epithelial pseudo-stratification and glandular branching, with two NFD mice (6365 & 6342) even showing secondary branching, compared to the HFD mice (Table 4-8 & Figure 4-13).

The reason for this contrasting behaviour in tumour development is unclear. Since caecal lesions were debilitating, it was intriguing to speculate that the higher number of lesions observed in the caecum of the NFD mice might have worsened the fitness of these animals, which could have contributed to the comparable survival rates observed between the two groups (Figure 4-9). Nevertheless, this data proved the successful implementation of the modified protocol to increase the disease burden in Apc-deleted mice, as well as confirming the previous observations from our group and others that HFD promotes adenoma growth in the SI of CRC mouse models [225, 356].

4.4.4 Degree of $Apc^{fl/fl}$ Recombination in the GI Tract

Fresh-frozen tissue samples were tested for the fractional abundances of recombined alleles, as a way of assessing the degree of Cre-mediated recombination. DNA extraction was as described in section 2.2.2.1.3, and 10ng of each sample was analysed using the Apc-Rec and Apc-NR ddPCR assays (section 2.2.3.4.2).

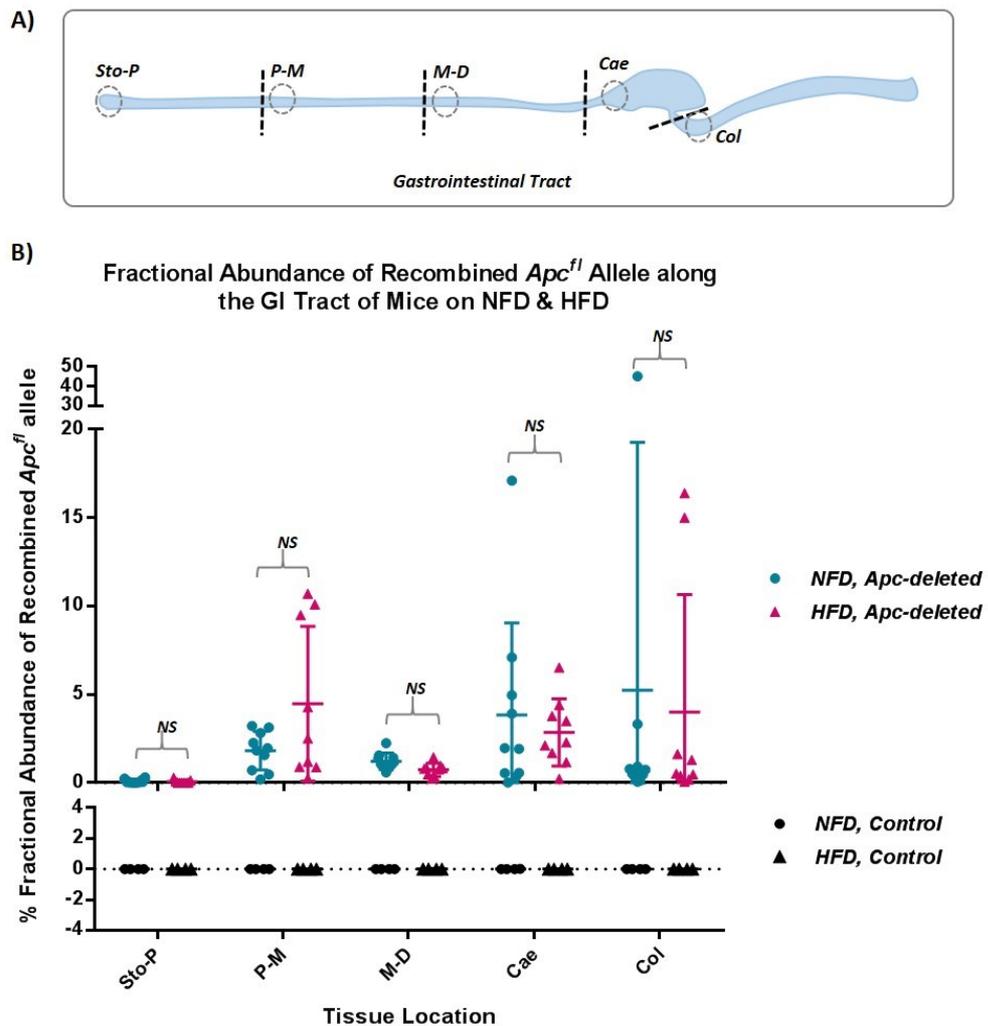


Figure 4-14: Assessment of $Apc^{fl/fl}$ Recombination Levels in the GI Tract

A) Diagram of GI tract with circles indicating where samples were collected. Sto-P = stomach-proximal SI, P-M = proximal-medial SI, M-D = medial-distal SI, Cae = caecum, Col = colon.

B) All control mice showed 0% abundance (i.e. no recombination). Apc -deleted mice showed variable fractional abundances throughout the GI tract. At the Sto-P region, NFD= $0.08 \pm 0.11\%$ vs. HFD= $0.06 \pm 0.10\%$ ($p > 0.9999$). At the P-M region, NFD= $1.82 \pm 1.09\%$ vs. HFD= $4.48 \pm 4.38\%$ ($p = 0.8259$). At the M-D region: NFD= $1.22 \pm 0.46\%$ vs. HFD= $0.74 \pm 0.37\%$ ($p > 0.9999$). At the Cae region: NFD= $3.85 \pm 5.20\%$ vs. HFD= $2.86 \pm 1.91\%$ ($p = 0.9975$). At the Col region: NFD= $5.24 \pm 14.04\%$ vs. HFD= $4.00 \pm 6.66\%$ ($p = 0.9927$). Several outliers were present in the Cae and Col regions. All p values were calculated using 2-way ANOVA with Sidak's multiple comparison test.

Unlike the previous analysis where only the proximal SI tissues were tested, different GI tissue regions were analysed this time to check if recombination levels show spatial variation (Figure 4-14-A). All control mice showed no recombination as expected. The recombined allele was present throughout the whole intestine in *Apc*-deleted mice, but more abundantly in the samples collected at the junction between proximal and medial SI (*P-M*) and in the caecum (*Cae*) (Figure 4-14: A & B) where tumour development was concentrated (Figure 4-12). At the *P-M* junction, the *Apc*-deleted mice in the HFD group showed higher fractional abundances of the recombined locus compared to the NFD group – $4.48 \pm 4.38\%$ vs. $1.82 \pm 1.09\%$ respectively – but this was not statistically significant ($p=0.8259$; Figure 4-14-B). These fractional abundances were lower than what was observed in the 2nd *in vivo* study ($5.16 \pm 6.24\%$; Figure 4-6-C). In contrast, caecal tissues showed that the NFD *Apc*-deleted mice had slightly higher fractional abundance of recombined alleles compared to the HFD counterparts – $3.85 \pm 5.20\%$ vs. $2.86 \pm 1.91\%$ respectively – but they showed no statistical significance ($p=0.9974$). Interestingly, the varying fractional abundances of recombined alleles in the *P-M* and *Cae* regions between two groups seemed to parallel the previous analysis where the proximal SI of HFD mice had the propensity for increased number of lesions compared to the NFD mice (Figure 4-12 and Table 4-8).

In general, the remaining GI tissue regions (*Sto-P*, *M-D* and *Col*) had very low fractional abundances of recombined alleles (Figure 4-14-B), which correlated with the lack of polyp growth observed in these locations. Few animals showed a high degree of *Apc* recombination in the colon, although this was not supported by the post-mortem analysis as no substantial colonic lesions were observed (Figure 4-11). As the colon samples were collected at the caecal junction (Figure 4-14-A), it was possible that tissue crossover at the time of collection might be responsible for these outliers. Indeed, exclusion of outliers from analysis gave low fractional abundances of $0.81 \pm 0.99\%$ for the NFD mice and $0.66 \pm 0.58\%$ for the HFD mice in the colon. Nonetheless, the variables described previously with regards to dilution of the recombined signal through stromal DNA, mosaicism of Cre expression and the neutral drift competition of stem cells should all be taken into consideration here as well [9, 13, 368]. Notwithstanding, this analysis provided an estimation of the degree of Cre-mediated recombination along the GI tract, and indicated that recombination events were more frequent in the proximal SI and

caecum, in line with their susceptibility for adenoma development, but this effect was independent of differences in the dietary fat content.

4.4.5 Analysis of Nuclear β -catenin Index

Tissue samples were stained with anti- β -catenin antibody (Table 2-11) to measure the nuclear β -catenin index, which can be interpreted as a readout of Wnt/ β -catenin pathway activation. Staining protocol was described in section 2.2.6.2. This analysis was restricted to the proximal SI, as this region showed the highest levels of recombination as well as being the preferential location of adenoma development.

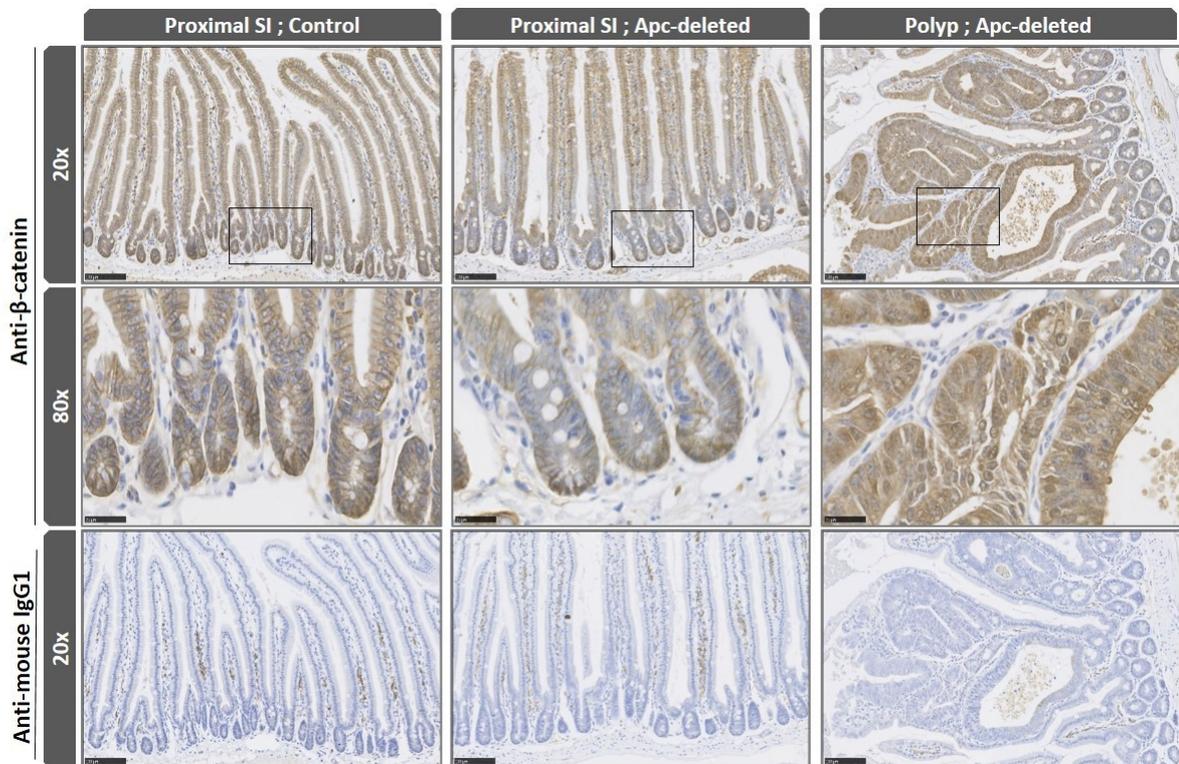


Figure 4-15: β -catenin Staining of Proximal SI Tissues from the NFD Mice

Control mouse was 6144 and Apc-deleted mouse was 6364 (Table 4-6). The images were captured at using NanoZoomer-XR digital slide scanner C12000. Black box insets in images taken at 20x magnification (top row, scale bar= 100 μ m) were further magnified at 80x (middle row, scale bar= 25 μ m). Negative controls were shown on bottom row. CBC cells at the base of crypts were positive for nuclear β -catenin staining in both control (left column) and Apc-deleted (middle column) mice. In the polyp region from an Apc-deleted mouse (right column), there were abundant cells with positive nuclear staining.

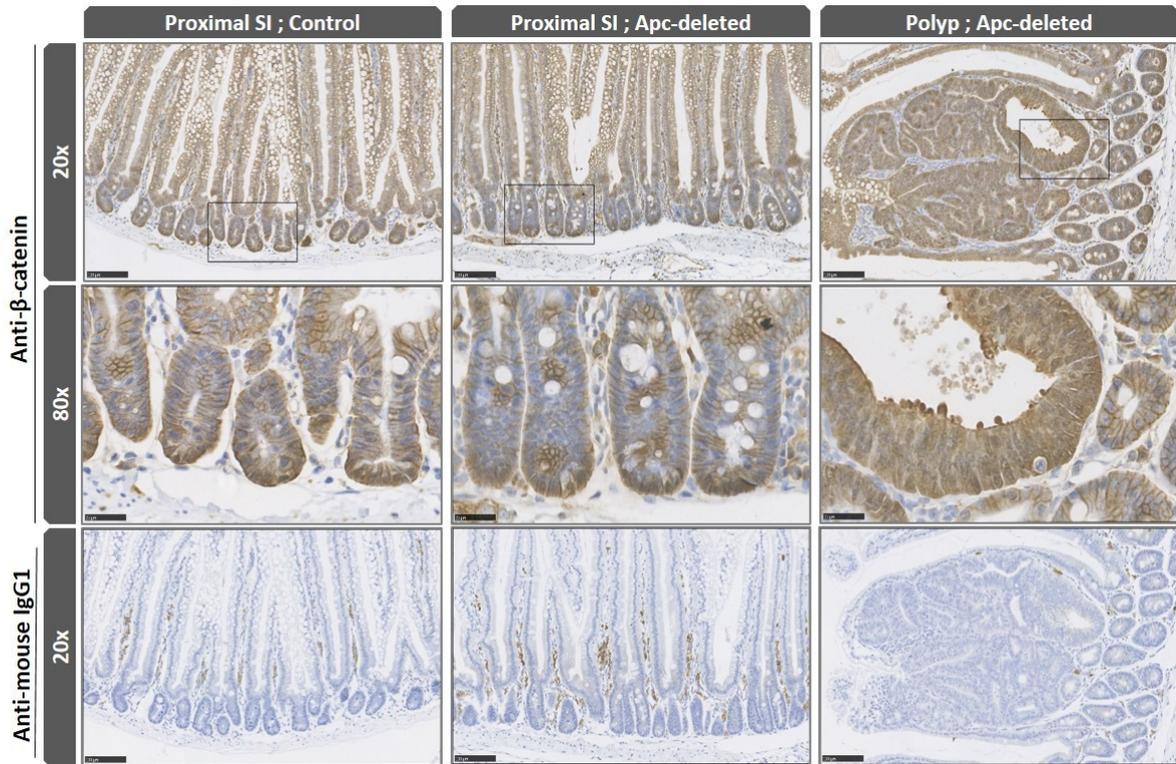


Figure 4-16: β -catenin Staining of Proximal SI Tissues from the HFD Mice

Control mouse was 6167 and Apc-deleted mouse was 6349 (Table 4-6). The images were captured at using NanoZoomer-XR digital slide scanner. Black box insets in images taken at 20x magnification (top row, scale bar= 100 μ m) were further magnified at 80x (middle row, scale bar= 25 μ m). Negative controls were shown on the bottom row. CBC cells at the base of crypts showed nuclear β -catenin staining in both control (left column) and Apc-deleted (middle column) mice. Abundant lipid globules were scattered throughout the epithelium, which was likely caused by the intake of high dietary fat content. Intense nuclear staining was observed for the cells in the polyp region from an Apc-deleted mouse (right column).

β -catenin nuclear localisation, seen as intense brown spots in images, was evident in the intestinal crypts in both Apc-deleted and control mice, with higher number of positive staining observed for the adenomatous regions in Apc-deleted mice (Figure 4-15 and Figure 4-16). These stained sections were used to score the nuclear β -catenin index by counting the numbers of β -catenin positive and negative crypt cells (50 crypts/mouse).

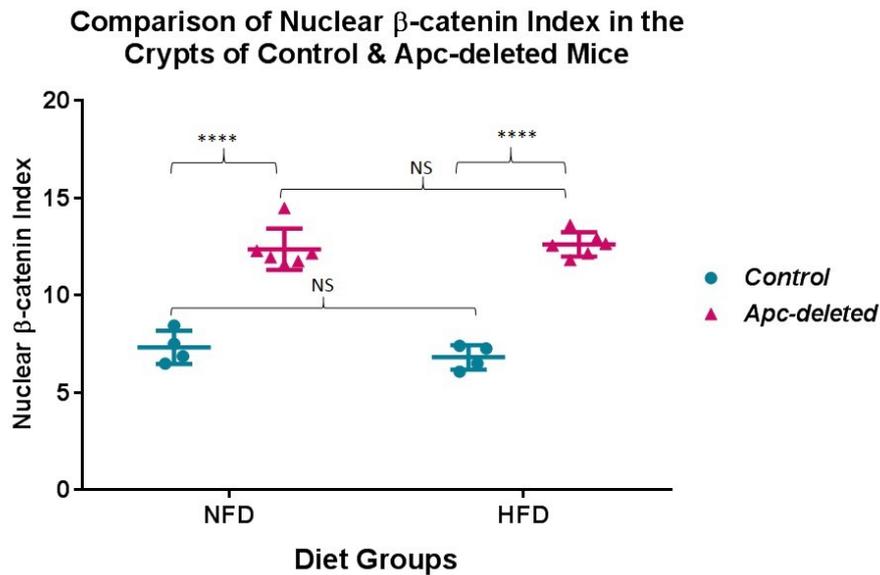


Figure 4-17: Comparing Nuclear β -catenin Indices in Crypts of Apc-deleted Mice on NFD & HFD

For each mouse, 50 crypts within the proximal SI were scored for nuclear β -catenin indices, calculated as the percentage of cells with nuclear β -catenin positivity. Each dot represents one mouse, and horizontal bars are mean \pm SD. The NFD controls have 7.33 ± 0.86 and the HFD control have 6.82 ± 0.64 indices ($p = 0.9508$). The Apc-deleted NFD mice have 12.38 ± 1.07 and HFD mice have 12.63 ± 0.62 indices ($p=0.9966$). However, for each diet group, p values are <0.0001 between Apc-deleted mice and controls. All p values were calculated at 95% CL, and 2-way ANOVA with Sidak's multiple comparisons test was used for statistical analysis.

As expected, an increase in nuclear β -catenin index was recorded in the crypts of Apc-deleted mice compared to the control mice in both diet groups (both $p < 0.0001$; Figure 4-17), consequent to the tamoxifen-induced recombination of *Apc^{fl/fl}* alleles. Sabatini and Yilmaz's group reported higher nuclear β -catenin activities in their HFD cohort [357], but I did not observe this difference in my HFD Apc-deleted mice as they showed comparable nuclear β -catenin indices to the NFD Apc-deleted mice (NFD= 12.38 ± 1.07 vs. HFD= 12.63 ± 0.62 ; $p=0.9966$), which suggested that dietary fat content did not have a significant influence on the canonical WNT pathway, at least in this study.

Additionally, nuclear β -catenin indices in the adenomas were assessed (Figure 4-18). For each mouse, five field-of-view snapshots of adenomatous regions were captured at 40x magnification and all nuclear β -catenin positive and negative cells were counted.

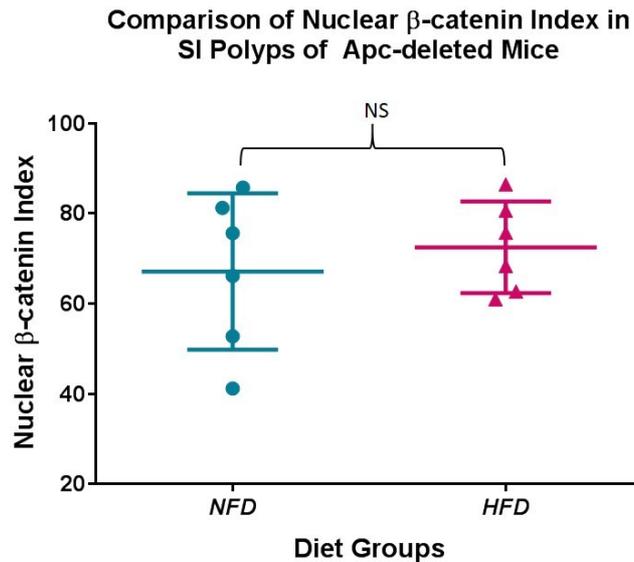


Figure 4-18: Comparison of Nuclear β -catenin Indices in Adenomas from *Apc*-deleted Mice on NFD & HFD

For each mouse, five random, field-of-view snapshots (40x magnification) of polyp regions from β -catenin-stained proximal SI sections used for scoring. Six *Apc*-deleted mice were counted for each diet group. Each dot represents one mouse, and horizontal bars represent mean and \pm SD. The HFD group has 72.52 ± 10.16 and the NFD group has 67.16 ± 0.62 indices ($p=0.6991$, 95% CL; Mann-Whitney test).

Irrespective of the diets, nuclear β -catenin indices in adenomas showed a substantial increase when compared to crypts: there was a 4.43-fold increase in the mean index value in the *Apc*-deleted, NFD mice (crypts = 12.38 ± 1.07 , adenomas = 67.16 ± 0.62) and a 4.74-fold increase in the *Apc*-deleted, HFD mice (crypt = 12.63 ± 0.62 , adenoma = 72.52 ± 10.16) (Figure 4-17 & Figure 4-18). However, despite the higher disease burden seen in the proximal SI of HFD mice (Figure 4-12), the nuclear β -catenin indices in adenomatous regions between the diet groups showed no statistical difference ($p=0.6991$; Figure 4-18). Although this data indicated there was a robust Wnt activity in the lesions present in *Apc*-deleted animals, it also suggested that an increase in dietary fat intake did not have a significant effect on the nuclear β -catenin localisation upon loss of *Apc* function. It should be noted that these mice were culled at different end-points depending on when they reached the severity limit (Table 4-6), but, considering that the survival between two groups was comparable (Figure 4-9), it was unlikely that this might have masked any effect of HFD.

4.4.6 Quantification of Total Plasma cfDNA Levels

Plasma samples were collected and processed for cfDNA extraction as previously described in section 4.2.3. Quantification of total cfDNA was carried out using the mGapdh qPCR assay (section 2.2.3.3.2) – 3.0µl of each sample was tested in triplicates. Irrespective of the study diets, significantly reduced total cfDNA load was observed in the Apc-deleted mice compared to the controls (Figure 4-19). For the NFD group, the mean cfDNA quantity was 14.65 (\pm 11.52) ng/ml in the Apc-deleted mice compared to 110.61 (\pm 105.03) ng/ml in the controls ($p=0.0384$). Similarly, for the HFD group, the mean cfDNA quantity was 32.43 (\pm 28.18) ng/ml in the Apc-deleted mice compared to 226.26 (\pm 102.89) ng/ml in the controls ($p<0.0001$) (Figure 4-19-B). As for the Apc-deleted mice, no statistically significant difference was observed between the two diet groups ($p=0.9000$; Figure 4-19-B). This data contradicted previous results from 1st and 2nd studies where higher cfDNA levels were associated with Apc-deleted mice (Figure 4-2 and Figure 4-7). Moreover, the Apc-deleted mice in this study showed considerably lower mean cfDNA levels compared to the previous two studies: the Apc-deleted mice from the 1st study had mean cfDNA levels of 75.42 (\pm 68.31) ng/ml and the 2nd study had 66.74 (\pm 59.54) ng/ml.

One possible explanation for such striking differences in the mean cfDNA quantities between the controls and Apc-deleted mice could lie in the dynamics of cfDNA in circulation. In this study, the Apc-deleted mice exhibited higher disease burden (section 4.4.3), which led to exacerbating the intestinal bleeding consequent to the development of adenomas, and mice became very sick from severe anaemia. This could have resulted in shedding considerable amount of DNA through faeces, rather than into the peripheral circulation. To check this possibility, faecal samples were analysed and results will be discussed in section 4.4.8. Severe anaemia also triggered a compensatory extramedullary haematopoiesis in the spleen, which was indeed enlarged in sick animals (not shown). Whether these alterations in blood homeostasis and haematopoiesis might have influenced cfDNA yield is yet to be confirmed. Of note, all Apc-deleted mice were culled and blood was collected when each mouse reached an ethical end-point, and the controls for each group were euthanised on the same day as when the last Apc-deleted mouse had reached the end-point. Therefore, it was possible that the blood samples

from controls were contaminated with DNA from lymphocytes during sample processing, leading to outliers with markedly elevated total plasma cfDNA levels.

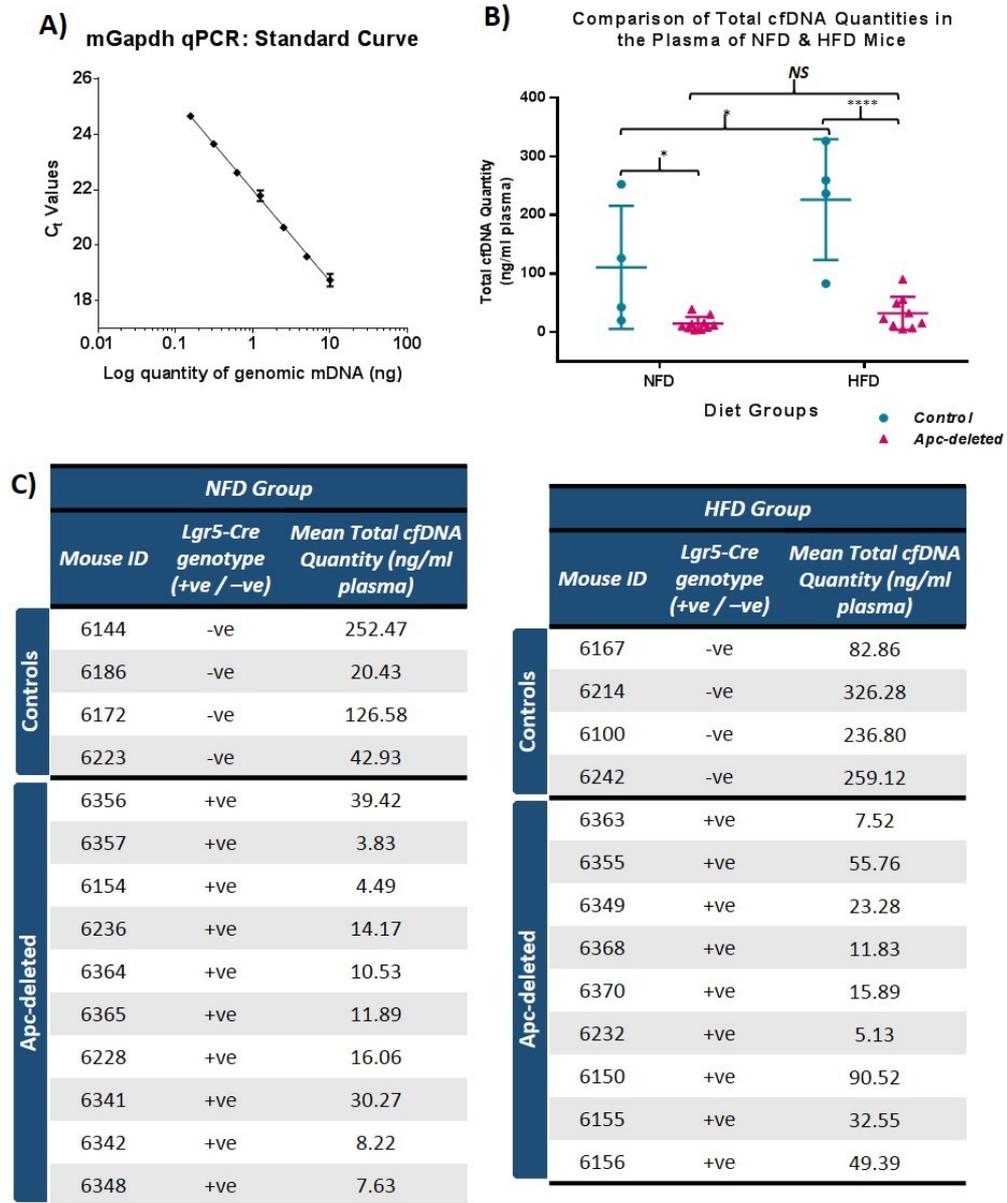


Figure 4-19: Total Plasma cfDNA Levels in *Cre-Apc^{fl/fl}* Mice from NFD & HFD Groups

A) Standard Curve analysis of C_t values against serial dilutions of mgDNA as a standard. It has the slope of -3.295, y-intercept at 21.99 and R^2 value of 0.9952. Assay efficiency was 101.14%.

B) Graph comparing mean plasma cfDNA levels. Each dot represents one mouse, and horizontal lines are mean \pm SD. Mean quantities are 226.3 ± 102.89 ng/ml for the HFD controls and 110.6 ± 105.03 ng/ml for the NFD controls ($p=0.0364$). Mean quantities are 32.43 ± 28.18 ng/ml for the HFD Apc-deleted mice and 14.65 ± 11.52 ng/ml for the NFD Apc-deleted mice ($p=0.9000$). For differences in the mean quantities between the controls and Apc-deleted mice, the NFD group has $p=0.0384$ and the HFD group has $p<0.0001$. All p values were calculated at 95% CL, and 2-way ANOVA with Tukey's multiple comparison was used for statistical analysis.

C) Tables listing the mouse IDs and the average total cfDNA quantities.

4.4.7 Analysis of Non-recombined & Recombined Alleles in the Plasma

Plasma samples were also analysed using the Apc-Rec and Apc-NR ddPCR assays (section 2.2.3.4.2) in the attempt to detect the recombined alleles. All available plasma samples were lyophilised (section 2.2.2.3.1), and 3µl of the reconstituted samples were tested per reaction.

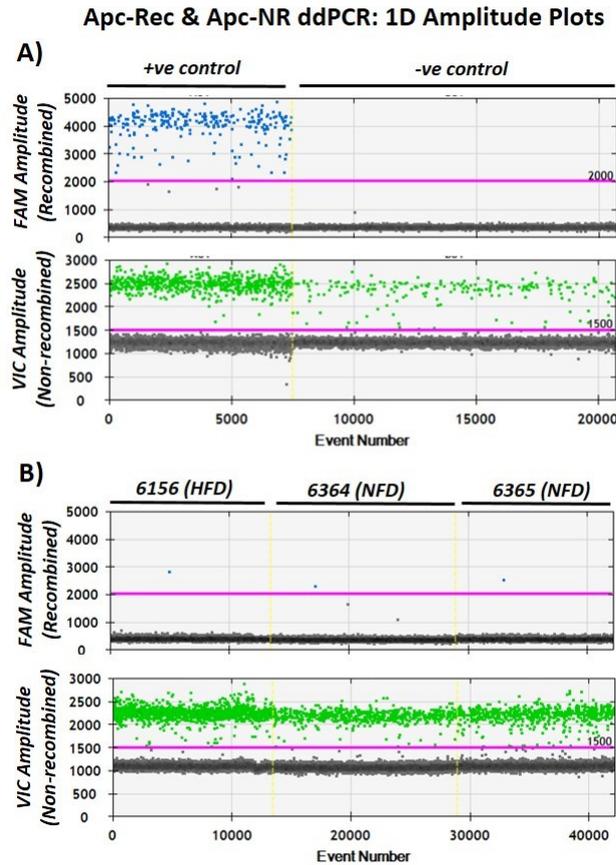


Figure 4-20: 1D Amplitude Plots for Apc-Rec & Apc-NR Assays used to Detect Non-recombined & Recombined Alleles

These plots are companion figures to the graphs in Figure 4-21.

A) 1D ddPCR plots showing manual threshold gating (pink lines) for the Apc-Rec (FAM – recombined) and Apc-NR (VIC – non-recombined) assays for the positive (+ve) and negative (-ve) controls. Positive control was 10ng of SI tissue DNA from the Apc-deleted mouse (5284). Negative control was 1ng of ear tissue DNA from the same mouse.

B) 1D plots for positive detection of recombined copies in the plasma from three Apc-deleted mice: one HFD mouse (6165) and two NFD mice (6364 & 6365).

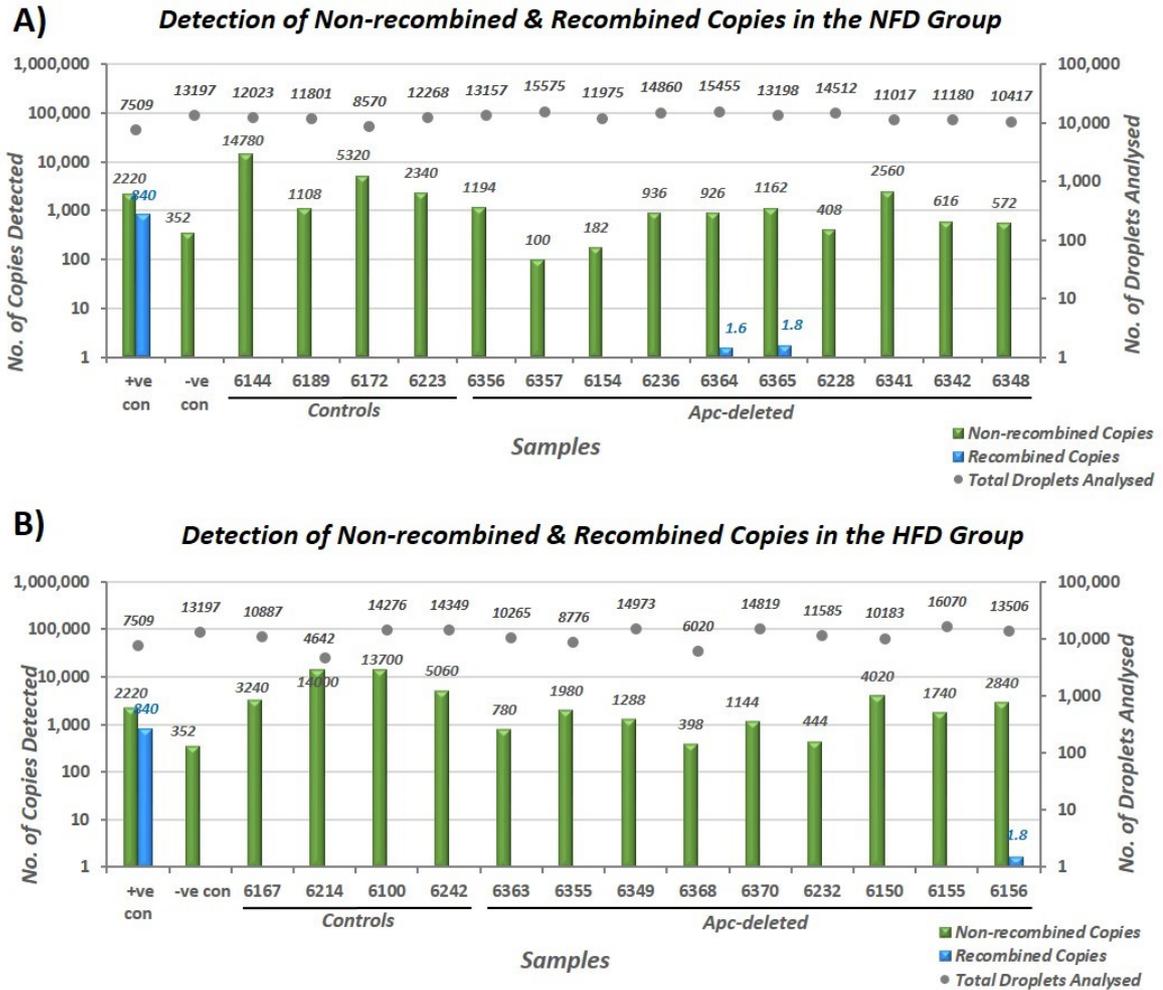


Figure 4-21: Detection of Non-recombined & Recombined Alleles in the Plasma

1D plots and manual threshold gating were described in Figure 4-20. Each graph has the numbers of non-recombined (green bars) and recombined (blue bars) copies detected on left y-axis, and the number of droplets analysed (grey dots) on right y-axis.

A) NFD group. Non-recombined copies were detected at varying quantities in all samples. Controls were free of recombined copies. Plasma samples from two Apc-deleted mice (6364 and 6365) showed positive amplification of recombined alleles, albeit at very low quantities.

B) HFD group. Varying numbers of non-recombined copies were detected across the samples. Positive recombined alleles were detected in only one Apc-deleted mouse (6156).

Similar to the previous studies, non-recombined alleles were promptly detected in plasma samples, though with varying quantities across the study groups (Figure 4-21). Interestingly, control sample 6214 showed an overwhelming number of non-recombined copies (14,000) despite the low droplet count (4,642). More copies of non-recombined alleles were detected in the controls compared to the Apc-deleted mice in both diet groups, reflecting higher total cfDNA yields observed in the controls with the mGapdh assay (Figure 4-19). As for the detection of recombined copies, only 3 out of 19

(15.79%) Apc-deleted mice were positive – two were NFD mice (6364 and 6365) and one was HFD mice (6156) – and the recombined copies detected were very low (<2 copies/well; Figure 4-21). The positive detection of recombined copies in these 3 mice did not seem to be correlated to their total cfDNA levels, which were all comparable to the averages of each group. The NFD mice, 6364 and 6356, had 10.53ng/ml and 11.89ng/ml respectively, which were just below the average concentration (14.65 ±11.52ng/ml) of the group (Figure 4-19-C). The HFD mouse 6156 had 49.39ng/ml total cfDNA compared to the group’s average of 32.43 ±28.18ng/ml (Figure 4-19-C).

Table 4-9: Health & Welfare Analysis Data for 3 Positive Samples

<i>ID</i>	<i>Gender</i>	<i>Diet</i>	<i>Weight loss (%)</i>	<i>Time survived (days)</i>	<i>No. of Non-recombined Copies</i>	<i>No. of Recombined Copies</i>	<i>% Fractional Abundance</i>
6156	F	HFD	15.22	35	2840	1.8	0.0352
6364	M	NFD	16.67	71	926	1.6	0.1725
6365	M	NFD	16.40	85	1162	1.8	0.1547

Moreover, the welfare and health data (Table 4-9) of these mice did not provide any indication as to why recombined copies were detected in these samples and not in the other 16 Apc-deleted mice. Whilst all 3 mice showed similar percentage of weight loss at the time of culling, their survival rates varied. Also, the percentage fractional abundances of recombined copies detected were low in all three samples (Table 4-9).

Table 4-10: Poisson Analysis for the Samples that were Positive for Recombined Alleles

<i>ID</i>	<i>Diet</i>	<i>Concentration (copies/μl)</i>	<i>Copies/ 20μl Reaction</i>	<i>PoissonMax (Conc)</i>	<i>PoissonMax (no. of copies)</i>	<i>PoissonMin (Conc)</i>	<i>PoissonMin (no. of copies)</i>
6364	NFD	0.08	1.6	0.36	7.2	0.00	0.0
6365	NFD	0.09	1.8	0.43	8.6	0.00	0.0
6156	HFD	0.09	1.8	0.42	8.4	0.00	0.0

In addition, Poisson analysis predicted the maximum number of copies detected (PoissonMax) to be between 7.2-8.6 copies and the minimum number of copies (PoissonMin) to be 0.0 for all 3 samples (Table 4-10). As the number of recombined copies detected were very few (<2.0 copies), there was a possibility that these may be caused by high background fluorescence noise. Poisson minimum values were an indication that there was a possibility of not detecting any recombined copies in

independent assay runs, suggesting the detection of recombined copies in the above 3 samples could not be confidently called as true positives.

Overall, the data presented here suggest that, despite the development of multiple adenomatous lesions (Figure 4-11 & Figure 4-12), the levels of tumour-derived, recombined alleles were below the detection threshold in the majority of Apc-deleted mice. The difficulty with detecting recombined copies in these samples was likely exacerbated by the unexpectedly low cfDNA yields in the Apc-deleted mice, irrespective of their diets (Figure 4-19). This analysis was further compounded by the technical problem with droplet generation where number of droplets generated and analysed (grey dots) were below the standard (>10,000/reaction) for samples 6355 and 6388 (Figure 4-21); however due to insufficient aliquots, it was not possible to repeat the experiment.

4.4.8 Analysis of Recombined Alleles in the Faecal Samples

Despite raising the severity limit, the data so far suggested that the quantity of recombined alleles in plasma remained below the detection threshold for most Apc-deleted mice (Figure 4-21). Therefore, faecal samples were analysed to assess whether detection of recombined alleles in faeces could be a viable alternative to the plasma testing, as studies with CRC patients that have successfully used faecal testing for detection of tumour-specific aberrations [359-362]. Faecal DNA samples were quantified using the mGapdh assay (section 2.2.3.3.2) and 10ng DNA per sample was tested using the Apc-Rec and Apc-NR assays on ddPCR (section 2.2.3.4.2).

Total DNA yields from faecal samples showed great fluctuations in the control and Apc-deleted mice from both diet groups across all time-points (Table 4-11), suggesting DNA is shed into faeces in a stochastic manner, irrespective of physical conditions. Even though the Apc-deleted mice showed increasing signs of sickness in response to adenoma development as more time elapsed post-induction, this was not reflected in the total faecal DNA yield over the course of study.

Table 4-11: Total DNA Yields from Faecal Samples

Time after injection (weeks)	NFD				HFD			
	Control		Apc-deleted		Control		Apc-deleted	
	No. of Mice	Yield \pm SD (ng/mg)	No. of Mice	Yield \pm SD (ng/mg)	No. of Mice	Yield \pm SD (ng/mg)	No. of Mice	Yield \pm SD (ng/mg)
0	4	45.18 \pm 29.32	10	21.11 \pm 13.37	4	54.99 \pm 41.26	9	55.66 \pm 37.28
2	4	54.98 \pm 42.58	10	23.25 \pm 21.93	4	46.49 \pm 50.18	9	35.56 \pm 18.32
4	4	19.34 \pm 13.05	10	12.83 \pm 4.67	4	33.43 \pm 20.44	9	24.53 \pm 15.63
6	4	14.41 \pm 10.03	9	14.01 \pm 5.96	4	30.77 \pm 19.76	7	29.42 \pm 26.97
7	4	17.53 \pm 21.98	8	24.23 \pm 10.47	4	13.94 \pm 12.96	7	24.87 \pm 11.02
8	4	25.60 \pm 35.89	7	44.68 \pm 30.01	4	12.52 \pm 7.26	4	24.36 \pm 9.33
9	4	29.29 \pm 27.56	6	23.67 \pm 17.10	4	13.77 \pm 5.70	3	22.85 \pm 12.00
10	4	20.74 \pm 4.43	3	35.01 \pm 45.66	4	26.25 \pm 21.14	3	12.33 \pm 7.43
11	4	10.89 \pm 8.19	3	29.24 \pm 20.49	4	22.62 \pm 6.71	1	15.66 \pm N/A
12	4	28.04 \pm 19.88	1	2.13 \pm N/A	4	22.09 \pm 12.18	1	77.06 \pm N/A
End-point	4	10.89 \pm 5.98	10	13.27 \pm 8.83	4	21.71 \pm 10.43	10	9.56 \pm 5.21

The recombined alleles were successfully detected in the faecal samples from Apc-deleted mice in both diet groups, with detectable fractional abundances rising above 1% for both diet groups from 4-week post-injection (Figure 4-22 & Table 4-12). In contrast to the total faecal DNA yields that showed great fluctuations across all time-points (Table 4-11), there was a trend towards increased abundances of recombined alleles in correlation to the elapsed time, with the HFD group showing a lesser increase (Figure 4-22 & Table 4-12). The mean fractional abundances of recombined alleles detected was consistently lower for the weekly samples from the HFD mice throughout the study, but these differences never reached statistical significance (Table 4-12).

Additionally, faecal samples were also collected at the time of culling (end-point) and analysed in the same manner (Figure 4-22). Irrespective of diet groups, the average percentage fractional abundance of recombined alleles was higher in these end-point samples than the weekly ones (Table 4-12), suggesting that likely increased blood loss at the end-point, when the disease burden was highest, contributed to more recombined alleles shed into faeces. For these samples, the mean fractional abundance was statistically lower in the HFD group compared to the NFD group (4.16% \pm 2.41 vs 7.44% \pm 2.92, $p = 0.0099$; Table 4-12).

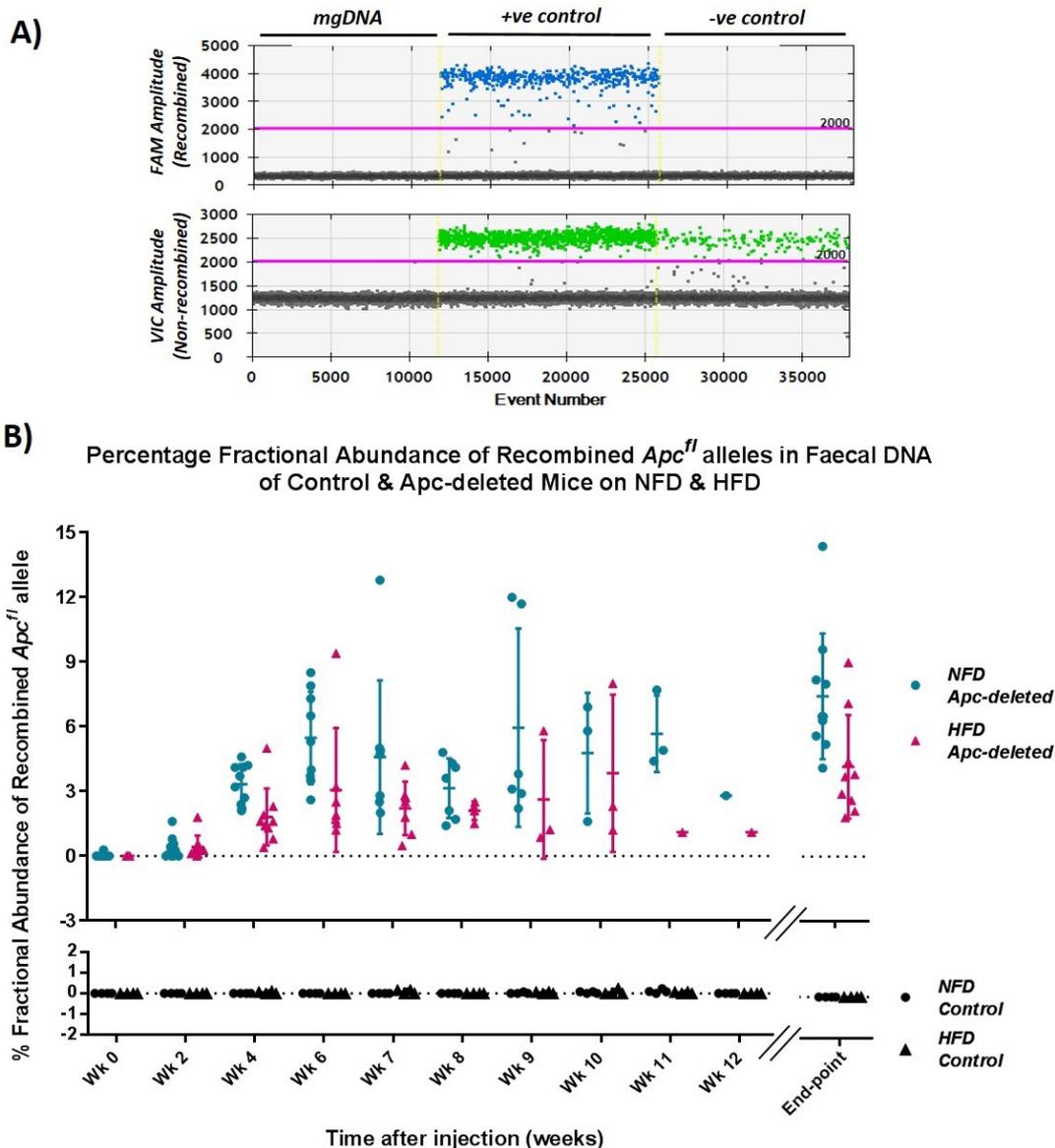


Figure 4-22: Changes in the Fractional Abundance of Recombined Alleles in Faecal Samples with Time after Injection

A) 1D amplitude plots with manual threshold gating (pink lines). Positive (+ve) control was 10ng SI tissue DNA from the Apc -deleted mouse (5284). Negative (-ve) control was 1ng ear tissue DNA from the same mouse. 10ng mgDNA was an additional negative control.

B) Each dot represents one mouse. All control mice (black dots) were negative for recombined alleles. The number of samples decreased with time elapsed due to sacrificing of sick animals as they reached end-point. End-points varied between study mice. Percentage fraction abundances of recombined alleles for Apc -deleted mice in the NFD (blue) and HFD (pink) groups and accompanying statistical analysis data are listed in Table 4-12.

Table 4-12: Statistical Analysis of Percentage Fractional Abundance of Recombined Alleles in Faecal Samples from Apc-deleted Mice (companion to Figure 4-22)

Time after injection (weeks)	NFD		HFD		Adjusted P Values	Significance at 95% CL
	No. of Mice	% Frac Abun \pm SD	No. of Mice	% Frac Abun \pm SD		
0	10	0.03 \pm 0.09	9	0.00 \pm 0.00	>0.9999	NS
2	10	0.38 \pm 0.51	9	0.41 \pm 0.54	>0.9999	NS
4	10	3.33 \pm 0.93	9	1.81 \pm 1.32	0.7478	NS
6	9	5.48 \pm 2.15	7	3.06 \pm 2.88	0.2337	NS
7	8	4.59 \pm 3.56	7	2.21 \pm 1.24	0.2905	NS
8	7	3.14 \pm 1.38	4	2.10 \pm 0.43	0.9979	NS
9	6	5.95 \pm 4.60	3	2.63 \pm 2.75	0.2603	NS
10	3	4.77 \pm 2.80	3	3.83 \pm 3.65	>0.9999	NS
11	3	5.67 \pm 1.78	1	1.10 \pm N/A	0.5056	NS
12	1	2.80 \pm N/A	1	1.10 \pm N/A	>0.9999	NS
End-point	10	7.44 \pm 2.92	10	4.16 \pm 2.41	0.0099	S

Statistical analysis was performed using 2-way ANOVA with Sidak correction for multiple comparisons. All p values were calculated at 95% CL.

Overall, this analysis did show that the recombined alleles were readily detected in faeces, and that significant DNA loads were shed into faeces regardless of phenotypic differences (Table 4-11). As hypothesised earlier, the presence of DNA in faecal material might have contributed to the difficulty with detecting recombined copies in plasma samples from the Apc-deleted mice (Figure 4-19), as these mice suffered intestinal blood loss as the disease progressed, which likely resulted in losing substantial DNA loads through faeces. Current data showed an increase in fractional abundances of recombined alleles with time elapsed after injection (Table 4-12), and with disease progression by proxy, which suggested an intriguing possibility of using the faecal DNA to monitor surrogate biomarker levels in relation to disease burden. Nevertheless, this warrants further investigation before any definitive conclusions can be made.

4.5 Summary

In this chapter, three *in vivo* studies were carried out to investigate whether the detection of recombined alleles as a surrogate biomarker could be used to inform and monitor early adenoma development in the *Cre-Apc^{fl/fl}* model of CRC.

In the 1st *in vivo* study, the disease burden at 8 weeks post-injection was lower than expected, with only 4 out of 6 animals showing at least one macroscopic polyp (Table 4-2). Quantitative analysis showed that although higher total plasma cfDNA concentrations were observed in the Apc-deleted mice compared to the controls (75.42 ±68.31 ng/ml vs. 35.84 ±26.68 ng/ml), the difference was not statistically significant (p=0.2949) (Figure 4-2). Moreover, recombined alleles were detected in only 1 (5207) out of 6 Apc-deleted mice and detection happened at a high C_t cycle of 41.80 with the ΔC_t of 11.97 (relative to the positive control), suggesting that it might be a false positive due to cross-well contamination from the positive control (Figure 4-3-A) or a background noise in the system.

This led to implementation of the 2nd study using a modified protocol that resulted in boosting adenoma growth in the Apc-deleted mice (Table 4-4). As before, elevated total plasma cfDNA levels were detected in Apc-deleted mice compared to the controls (66.74 ±59.54 ng/ml vs. 31.22 ±44.70 ng/ml), but this again was not statistically significant (p=0.0927; Figure 4-7). Additionally, recombined copies were detected in the plasma from only 2 (5227 & 5242) out of 7 Apc-deleted mice (Figure 4-8), with Poisson analysis predicting positive calling for only 1 out of these 2 positive samples at the minimum level of detection (Table 4-5). Therefore, despite the protocol refinement, the extent of disease burden did not prove to be sufficient to cause significant shedding of cfDNA into the circulation to reach the limit of detection.

Consequently, the protocol was further refined for the final *in vivo* study, which involved a bigger sample size of 27 mice, split into two diet groups of NFD (16.4% calories from fat) and HFD (59.7% calories from fat) (Table 2-10). HFD was introduced to promote adenoma development, based on the data reported in the literature [225, 356, 371]. Additionally, the severity limit was raised, which likely contributed to the extensive adenoma growth observed in the Apc-deleted mice from both diet groups (Figure 4-11). However, diets had no effect on the survival between the two groups (HFD= 44.0 days vs. NFD= 57.5 days, p=0.7190; Figure 4-9). Microscopic polyp counts and histological grading (Table 4-8) indicated that the HFD-fed mice had propensity for higher disease burden in the proximal SI compared to the NFD-fed mice (59.44 ±27.49 polyps vs. 33.3 ±28.36 polyps, p=0.0262; Figure 4-12-A), whilst the NFD-fed mice showed increased

burden in the caecum compared to the HFD-fed mice (5.98 ± 2.51 polyps vs. 3.33 ± 2.78 polyps, $p=0.0248$; Figure 4-12-C). Moreover, the intestinal crypts of Apc-deleted animals showed increased β -catenin nuclear localisation compared to the controls, but the nuclear β -catenin indices were not affected by the diets (section 4.4.5). Overall, at least for this study, HFD had insignificant effect on the development of lesions compared to previous reports [225, 357].

Moreover, quantitative analysis revealed a surprising trend where lower plasma cfDNA levels were observed in the Apc-deleted mice (NFD = 14.65 ± 11.52 ng/ml, HFD = 32.43 ± 28.18 ng/ml) compared to the healthy controls (NFD = 110.6 ± 105.03 ng/ml, HFD = 226.3 ± 102.89 ng/ml), in both diet groups (Figure 4-19). This was in contrary to the data from the previous studies where elevated plasma cfDNA levels were detected in Apc-deleted mice compared to the controls (Figure 4-2 and Figure 4-7). However, the control groups from the 3rd study showed great variations in their plasma cfDNA quantities (NFD = 110.61 ± 105.03 ng/ml, HFD = 226.26 ± 102.89 ng/ml), which could have been caused by contamination with DNA from lysed WBC; although it was also possible that they likely represented genuine, though striking, physiological differences. Furthermore, despite the increased disease burden, only 3 out of 19 Apc-deleted mice showed positivity for recombined alleles in the plasma, and the fractional abundances of recombined copies detected was very low (0.0352%-0.1725%; Figure 4-21, Table 4-9). Poisson analysis also predicted minimum detection at 0.0 recombined copies for all 3 samples, suggesting that these samples could not be confidently called as true positives (Table 4-10). This difficulty with detection of recombined alleles in the plasma had led me to perform analysis of faecal samples as an alternative method. Quantification analysis showed that whilst faeces contained considerable DNA loads, regardless of their phenotypes, the DNA yields were greatly variable (Table 4-11). However, faecal analysis did detect recombined alleles more efficiently compared to plasma samples (Figure 4-22 & Table 4-12). The data also showed a trend where the Apc-deleted mice on NFD had higher fractional abundances of recombined alleles compared to the HFD mice at weekly time-points (Table 4-12), and this difference became significant when faecal samples collected at the end-point were compared (NFD $7.44 \pm 2.92\%$ vs. HFD $4.16 \pm 2.41\%$, $p=0.0099$) (Table 4-12). However, it remained to be elucidated whether the differences in the quantities of recombined alleles in faecal samples between the groups were affected by high fat intake.

Of note, in all the studies, the development of caecal lesions in the Apc-deleted mice had debilitating consequences and shortened their participation in the studies, but it has yet to be determined whether this had affected cfDNA release into the circulation. Interestingly, this has not been widely reported, except by Vermeulen's group where they observed caecal polyps in some of their study mice with the *Lgr5-CreER^{T2}* allele [365].

Overall, the data from these *in vivo* studies questioned whether the release of cfDNA was associated with early disease progression, as it failed to confirm findings from similar studies conducted in human patients, which showed a correlation between cfDNA levels and early lesions [299, 312, 319]. Moreover, analysis of plasma samples gave very low sensitivities in detecting recombined alleles in Apc-deleted mice, despite the presence of substantial disease burden. However, one limitation of this project was the use of a single GEMM and this *Cre-Apc^{fl/fl}* model emulates the loss of Apc function which drives an aggressive development of multiple polyps with short latency, inevitably causing the study mice to be sacrificed due to a high disease burden without the lesions becoming invasive, and this could have contributed to the reduced representation of tumour-derived cfDNA in the plasma in this GEMM [143]. One future improvement would be to include a more aggressive GEMM of CRC such as the double mutant model *AhCre Apc^{fl/+}; Kras^{LSL-G12V/+}* to improve the detectability of tumour-derived cfDNA in plasma [143]. Nevertheless, based on the *in vivo* data obtained for this project, the use of plasma cfDNA for detection of early lesions in preclinical studies is not currently recommended. However, faecal analysis data suggested that faeces could provide an alternative mean to the plasma analysis for detection of tumour-derived surrogate biomarkers in preclinical studies, but further investigation is necessary.

Chapter 5 Validation of Assays for the Analysis of Patient Samples

5.1 Introduction

Many studies have shown that tumour-specific mutations found in cfDNA can be used to tailor treatment options, monitor responses to therapy and identify residual diseases [298, 302, 372]. However, most studies involved patients with advanced, malignant CRC, whereas less information is available with regards to pre-malignant lesions. Furthermore, attempts to detect tumour-specific cfDNA in patients with the early stage disease has been challenging due to low frequencies of MU alleles [299, 312, 318, 319]. Therefore, several PCR-based assays were designed and optimised for analysis of patient samples for this project.

For mutation detection assay targets, *BRAF* and *KRAS* were chosen for their frequent implications in CRCs and their clinical relevance as discussed in the introduction (section 1.3.2). Hotspot mutations in these two oncogenes are mutually exclusive and occur in all stages of tumourigenesis, including premalignant lesions, thus giving a good coverage when screening for mutations in the patient samples [373, 374]. Furthermore, as will be discussed in the following chapter, six patient cases (H149, H154, H263, H264, H265 and H266), each with large and/or multiple colorectal adenomas, were selected for multiregional targeted NGS to identify mutations that could be suitable targets for patient-specific ddPCR cfDNA analysis. This approach was chosen based on reports that higher frequency mutations likely contribute to higher fractional abundances in the plasma, which would make them easier to be detected [318]. Although the details of the NGS experiments will be discussed in the following chapter, here the development and validation of patient-specific mutation detection assays will be described.

5.2 Validation of *BRAF* & *KRAS* Mutation Detection Assays with PNA Clamping on the qPCR System

The PNA-clamped qPCR assay system has been widely used for detection of rare MU copies in cancer patient samples, such as faeces and blood, where excess background of WT DNA proves to be a technical challenge [255, 375, 376].

5.2.1 BRAF-V600E Assay

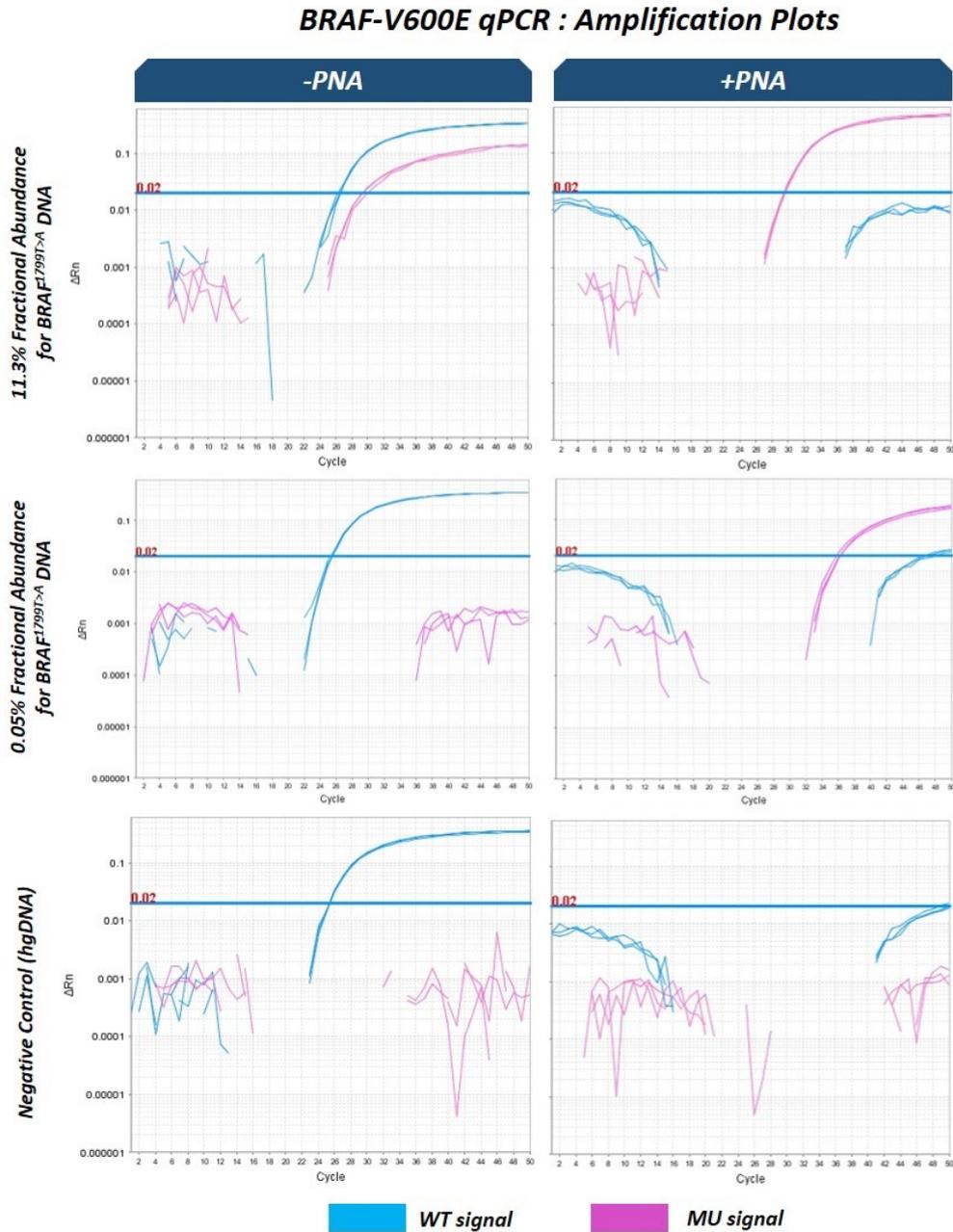


Figure 5-1: Testing the BRAF-V600E qPCR Assay with PNA Clamping

Amplification plots of ΔR_n (baseline-corrected relative normalised reporter dye) vs. cycles. Experimental conditions were as described in 2.2.3.3.6. Each reaction contained 10ng DNA. Each DNA dilution was tested with and without PNA, in triplicates. DNA from SK-MEL-28 (*BRAF* c.1799T>A) was used as a positive control. Positive control DNA was diluted in hgDNA as a WT background, and % fractional abundances of MU DNA were indicated on the y-axis. Negative control was 10ng hgDNA. Each line represents amplification signal from one reaction. Threshold was set at 0.02 ΔR_n (blue horizontal line). PNA clamping resulted in suppression of the WT (blue) signals, and allowed mutant copies (MU) (pink) to be amplified at a low MU abundance.

The BRAF-V600E assay targets the c.1799T>A/p.V600E mutation in exon 15 of the *BRAF* gene, which is implicated in 10%-15% of all CRCs [90, 373, 377]. The assay was tested with and without PNA to assess how effective PNA clamping was in suppressing the background WT signal to improve MU allele detection (Figure 5-1). DNA isolated from the SK-MEL-28 BRAF-MU cell line was used as a positive control and was diluted in hgDNA as a background, to obtain sample aliquots with different % MU fractional abundances to test the LoD. The commercial hgDNA effectively served as a negative control. The results showed that PNA clamping was very effective in suppressive signals from WT BRAF alleles (blue lines), which was essential in extending the LoD for amplification of few MU alleles (pink lines) at the low fractional abundance of 0.05% (Figure 5-1).

5.2.2 KRAS-121 Assay

Mutations in codons 12 and 13 of *KRAS* are implicated in approximately 30-50% of cases [80, 378]. This KRAS-121 assay was designed to detect three possible missense mutations in codon 12 at base c.34G: c.34G>A/p.G12S, c.34G>C/p.G12R and c.34G>T/p.G12C. Targeting multiple base substitutions at one location was made possible by using a degenerate TaqMan™ probe with the 'D' base (D=G/A/T) in the complementary position to the mutated base in the target allele (section 2.2.3.3.6).

During the testing of the KRAS-121 assay, I encountered a technical issue where the PNA clamping effect was rendered ineffective at blocking WT amplification (top row; Figure 5-2). It was unlikely to have been an intrinsic problem with the PNA sequence design as the clamping was successful at dampening the WT signal in the negative control (bottom row, Figure 5-2). Re-designing was not possible because the sequence designs of probes and PNAs are restricted to the mutated base as a central point to achieve an optimal hybridisation chemistry [332].

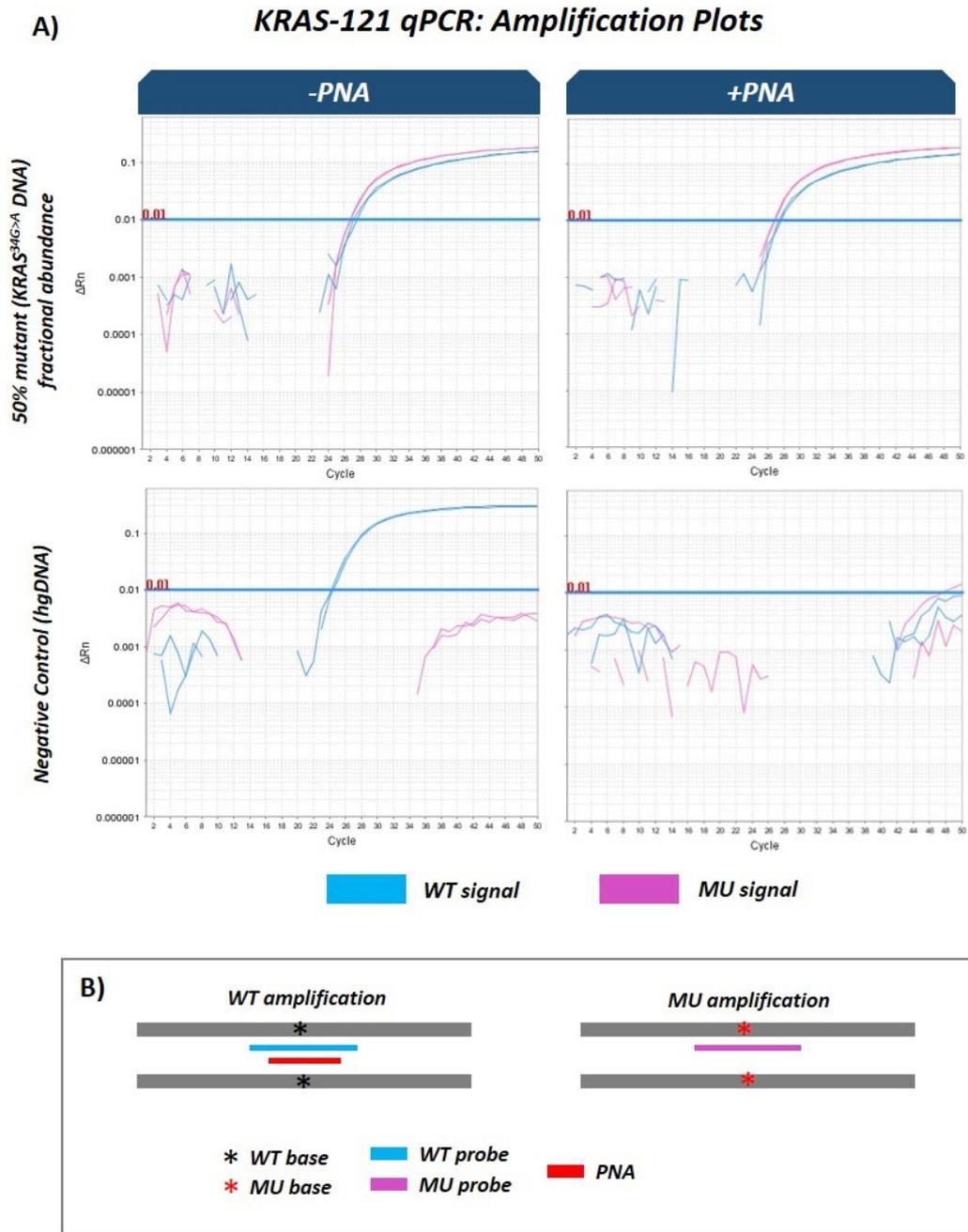


Figure 5-2: Ineffectiveness of PNA Clamping in KRAS-121 qPCR reactions containing both WT and MU Probes

A) Experimental conditions were as described in section 2.2.3.3.6. DNA from A549 (*KRAS* c.34G>A) cell line was used as a positive control, and diluted in hgDNA as a WT background. 10ng hgDNA was used as a negative control. Blue lines represent WT signals and pink lines represent MU signals. Manual threshold was set at $\Delta R_n=0.01$.

B) Simple illustration to show where probes and PNA anneal. Both probes and the PNA have overlapping hybridisation regions.

Therefore, different combinations of reagents were tested in an attempt to solve this unexpected problem with the PNA clamping. Surprisingly, I found that the PNA was effective at suppressing WT amplification when tested in the absence of the MU probe (Figure 5-3). Using a homozygous cell line (A549, *KRAS* G34>A) DNA, I confirmed that PNA clamping had no effect on MU amplification by running the reactions with PNA (pink) and without PNA (blue) (Figure 5-3-A). I also showed that PNA clamping was necessary to dampen the WT signal to detect MU alleles at a low MU fractional abundance (0.14%; Figure 5-3-B). Even though this did not provide an explanation for the unexpected ineffectiveness of this PNA sequence, it indicated that WT background signal could be suppressed if the assay was run in the absence of WT probe.

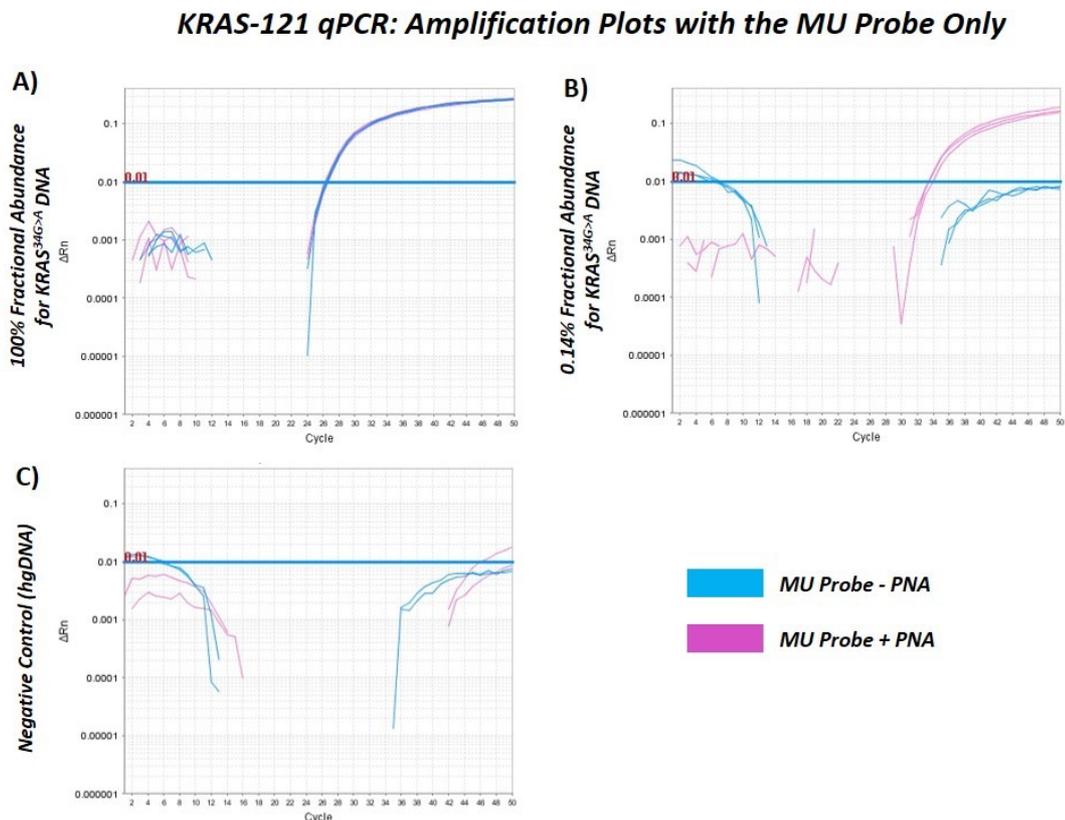


Figure 5-3: PNA Clamping was Necessary for Detection at Low MU Fractions

All reactions were run with the MU probe only in triplicates, with PNA (blue) and without PNA (pink). 10ng DNA was tested per reaction. DNA from cell line A549 (*KRAS* c.34G>A) was used as a positive control. It was diluted hgDNA to obtain aliquots with varying MU frequencies as indicated on the y-axis. Manual threshold was set at $\Delta Rn=0.01$.

A) At 100% MU fraction, PNA had no effect on MU amplification.

B) At a low (0.14%) MU fraction, MU signal was only detected in the presence of PNA (pink lines).

C) Negative control was 10ng hgDNA. No MU amplification was detected, confirming the target specificity for the MU alleles.

Therefore, an alternative strategy was adopted to run the KRAS-121 assay as shown in Figure 5-4: one reaction was run with the WT probe only without PNA (blue lines), and duplicate reactions were run with the MU probe only in the presence of PNA (pink lines). In this way, the WT signal provided the background signal for estimation of the fractional abundance of MU DNA. This alternative strategy was successful at detecting MU alleles at fractional abundances as low as 0.10% (Figure 5-4-B).

KRAS-121 qPCR: Alternative Strategy for Mutation Detection

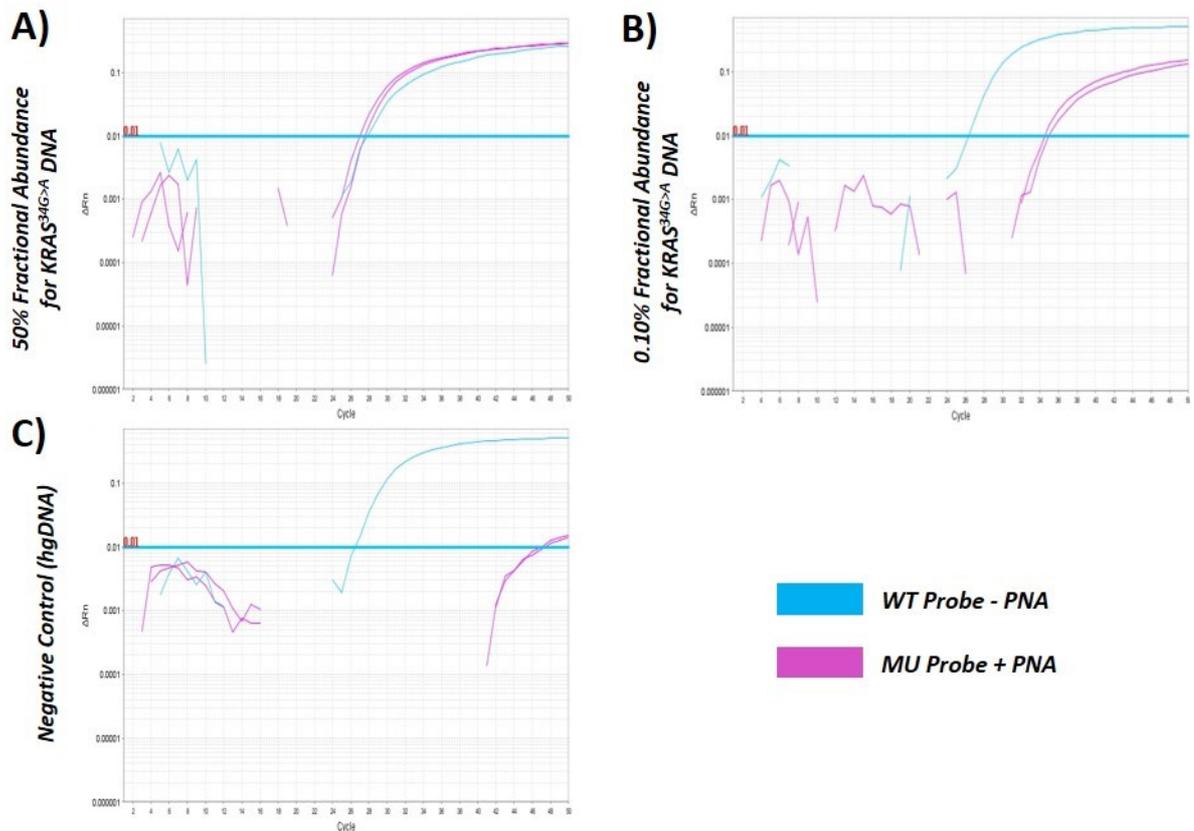


Figure 5-4: An Alternative Strategy for the KRAS-121 Assay

A modified strategy for the KRAS-121 assay; conditions were described in section 2.2.3.3.6. For each sample, one reaction was run with the WT probe only in the absence of PNA (blue), and duplicate reactions were run with the MU probe only in the presence of PNA (pink). DNA from cell line A549 (*KRAS* c.34G>A) was used as a positive control, and this was diluted in excess hgDNA. 10ng of DNA was tested per reaction. hgDNA was used as a negative control.

- A) At 1 WT: 1 MU fractions, WT and MU amplification profiles were superimposable.
- B) At a low MU fractional abundance (0.10%), PNA clamping facilitated amplification of few MU copies (pink lines).
- C) Negative controls showed only the WT amplification, confirming target specificity.

5.2.3 KRAS-122 Assay

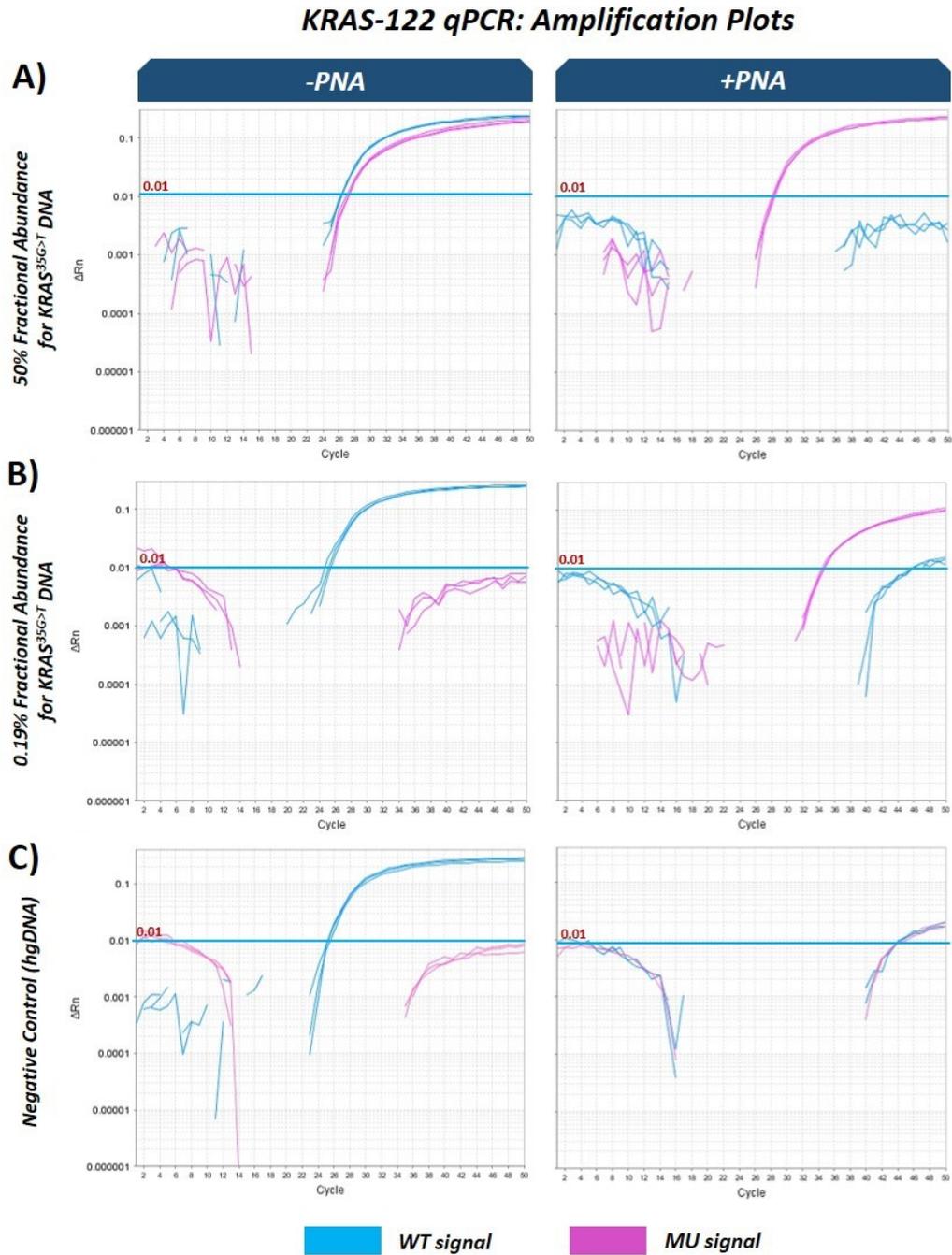


Figure 5-5: Testing the KRAS-122 qPCR with PNA Clamping

Experimental conditions were as described in section 2.2.3.3.6. Reactions were run in triplicates, with and without PNA. Each line represents one reaction. Threshold was manually set at ΔR_n 0.01. DNA from the cell line SW480 (*KRAS* c.35G>T) was used as a positive control, and was diluted in excess of hgDNA at various ratios. 10ng of total DNA was tested per reaction. PNA was effective at blocking WT signals (blue) and enabled detection of mutant (MU) copies (pink) at both high (50%) and low (0.19%) fractional abundances of MU DNA.

Similarly, the KRAS-122 assay was designed to detect the three possible missense mutations in codon 12 at base c.35G: c.35G>A/p.G12D, c.35G>C/p.G12A and c.35G>T/G12V. Like the KRAS-121 assay, this assay made use of the degenerate TaqMan™ probe with the 'D' base (D=G/A/T) at the position c.35G.

The effectiveness of the PNA clamping was similarly validated for the KRAS-122 assay (Figure 5-5). While PNA clamping did not play an essential role in the detection of MU alleles when the reaction contained abundant MU alleles (top row, Figure 5-5), clamping became necessary for MU detection at low fractional abundances (middle row, Figure 5-5). When compared to the BRAF-V600E assay, the KRAS assays had higher LoDs (KRAS-121=0.10%, KRAS-122=0.19%) and performed poorly below these MU fractional abundances despite DNA clamping (not shown). This was likely caused by the use of degenerate MU probes for simultaneous detection of different base substitutions, as studies have reported that degenerate bases, in either primers or probes, can negatively affect sensitivity [379, 380]. Nonetheless, although it was possible to design specific probes and separate assays for each base mutations on codon 12, we decided to persist with the use of degenerate probes to maintain the cost-effectiveness of our approach, and to allow analysis of multiple mutations in a single PCR reaction, which was an advantage in dealing with low target quantities in plasma samples.

5.2.4 Efficiencies of Assays

Amplification efficiencies of the qPCR assays were calculated by performing standard curve analysis as described in the introduction (section 1.8.1). Respective positive control DNAs were diluted in an increasing background of WT DNA in 1:2 ratio, whilst keeping the total DNA quantity the same at 10ng. Each dilution was then run in triplicates. The data was then analysed using the linear regression (semi-log) analysis by plotting the MU DNA quantity (x-axis) against the C_t cycles for each dilution (y-axis) (Figure 5-6).

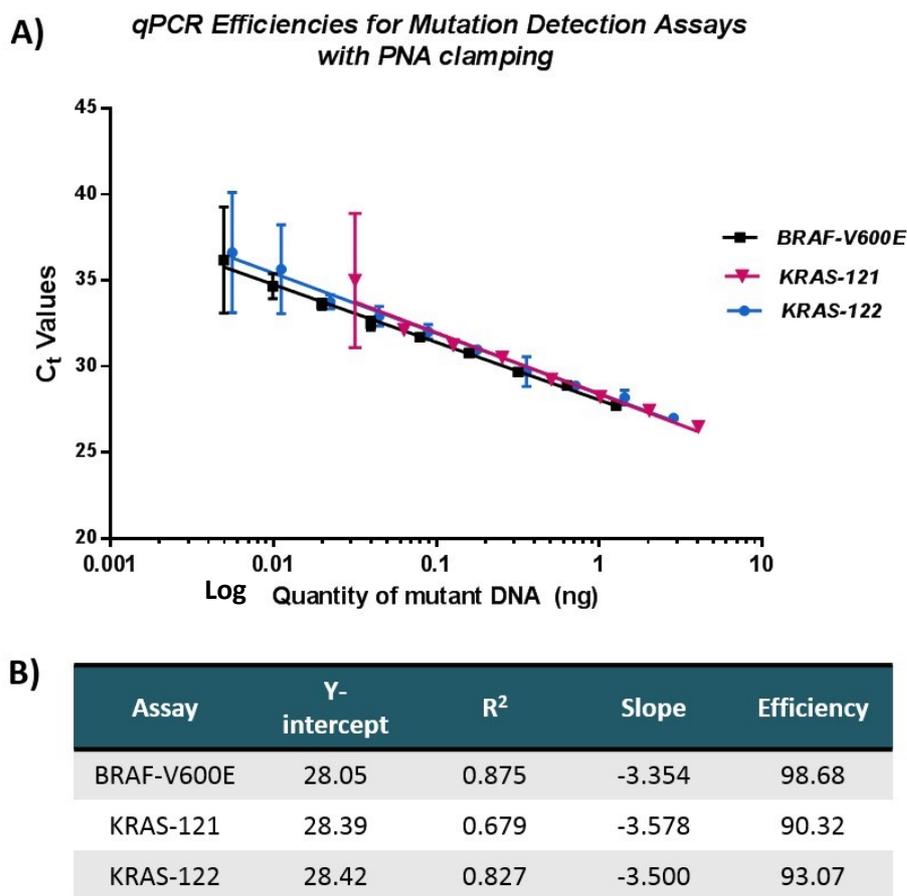


Figure 5-6: Assay Efficiencies for *BRAF* & *KRAS* qPCR Assays with PNA Clamping

A) Standard Curves were plotted using serial dilutions of positive controls, in excess of hgDNA. Assay conditions were described in section 2.2.3.3.6. All reactions were run with PNA clamping. C_t values for MU amplification (FAM) were determined by manual gating of thresholds at 0.02 ΔR_n for BRAF-V600E, and at 0.01 ΔR_n for KRAS-121 and KRAS-122 assays. Each dot represents one dilution; the values were averaged from three assays runs, and each dilution was run in triplicate per assay run. Bars represent ±SD.

B) Tabulated data of assay efficiencies.

As discussed in Introduction, 100% efficiency means there is an exact doubling of the DNA template at every PCR cycle, and qPCR assays have a tolerable range of 90-110% [323, 324]. BRAF-V600E showed the best amplification efficiency at 98.68%, whereas the KRAS-121 and KRAS-122 assays showed slightly lower efficiencies at 90.32% and 93.07%, respectively (Figure 5-6). The lower efficiencies for the KRAS assays were likely caused by the use of degenerate probes to detect all three possible base substitutions at bases c.34G (KRAS-121) and c.35G (KRAS-122).

5.3 Validation of *BRAF* & *KRAS* Mutation Detection Assays on the ddPCR System

Due to the acquisition of the QX200™ ddPCR system (BioRad) half way through this project, I proceeded to validate the *BRAF* and *KRAS* assays on this system. Additionally, a commercial assay for detection of the *KRAS* mutation c.38G>A/p.G13D in codon 13 (*KRAS* G13D assay) was validated exclusively on the ddPCR platform. The performance and target specificity of assays were tested through serial dilutions of MU DNA isolated from cancer cell lines in excess of hgDNA. For the *BRAF*-V600E, *KRAS*-121 and *KRAS*-122 assays, they were tested using the same primers and probe as the qPCR counterpart, but with the exclusion of PNA clamps. The ddPCR system has made PNA clamping obsolete as enrichment of MU DNA amplification was achieved through multipartite droplet generation. The WT and MU targets were detected in parallel using different fluorophore conjugations (VIC/HEX for WT and FAM for MU). Experimental conditions were described in section 2.2.3.4.3. All reactions were run with 10ng total DNA input. Data analysis was done using the QuantaSoft software (version 1.6.6.0320, BioRad®).

5.3.1 *BRAF*-V600E Assay

On the ddPCR system, the *BRAF*-V600E assay showed specific and sensitive detection of MU alleles in an excess of hgDNA background (Figure 5-7). At the lowest dilution (Dil 6 – 1:1000), 2.8 MU copies were detected in a background of 4660 WT copies (MU fractional abundance at 0.060%; Figure 5-7, B & C). However, Poisson analysis for this dilution predicted that, at the minimum value, the number of MU copies detected will be at 0.4 copy (Figure 5-7-C), which was less than the minimum of 1.0 copy needed for positive calling. Therefore, the effective LoD lies above 0.060% MU fractional abundance for the *BRAF*-V600E assay. Linear regression (log-log) analysis showed a good correlation between the input MU copies and no. of MU copies detected ($R^2=0.9831$; Figure 5-7-D).

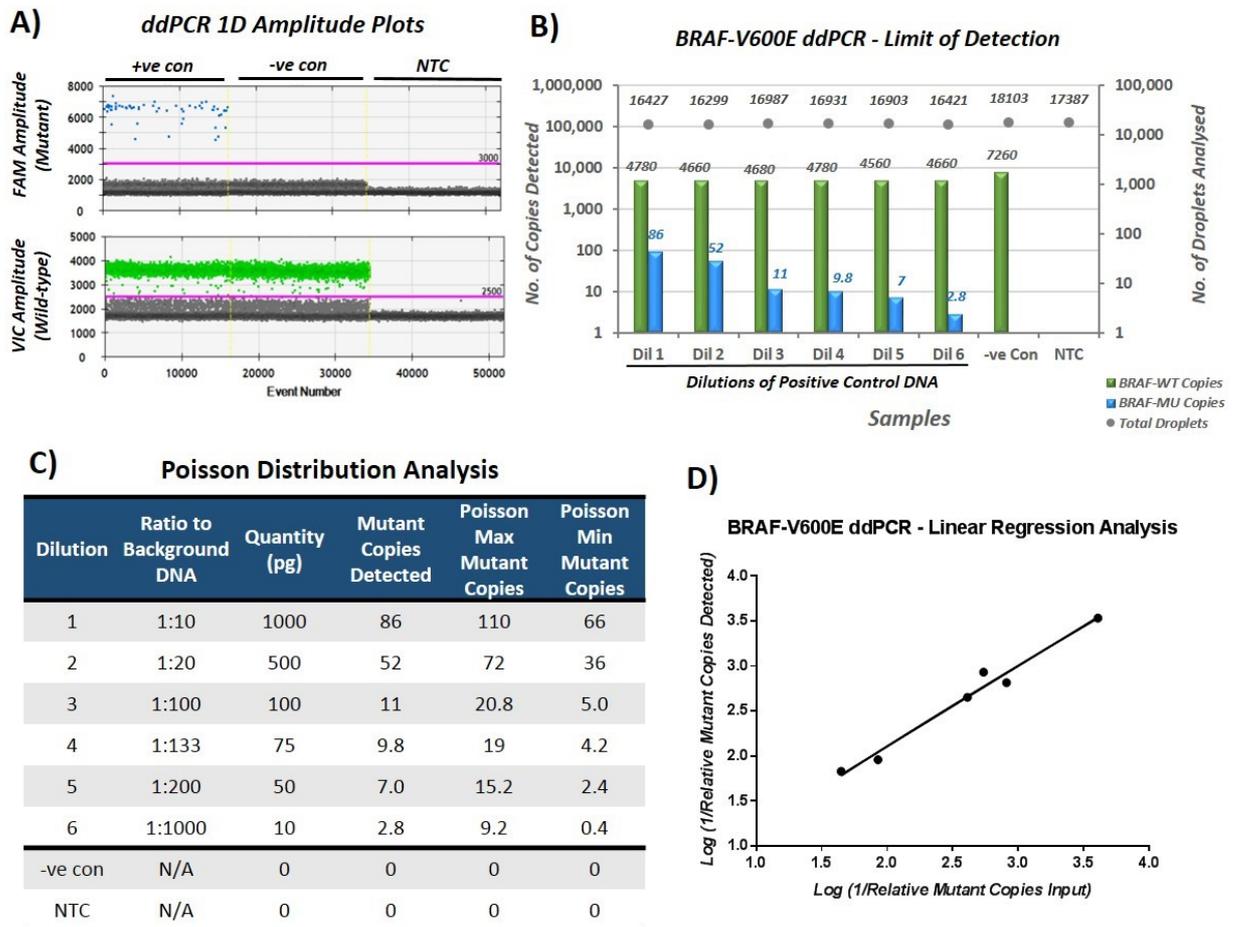


Figure 5-7: Validation of BRAF-V600E Assay on ddPCR

A) 1D ddPCR plots showing manual threshold gating (pink lines) for the MU (FAM) and WT (VIC) targets. The positive control (+ve con) was DNA from HT-29 (*BRAF* c.1799T>A) cell line diluted 1:10 in hgDNA (Dil 1). Negative control (-ve con) was hgDNA. NTC = no template control.

B) Dilutions (Dil) 1 to 6 represent reactions with decreasing mutant DNA quantity with increasing WT background (see table C). BRAF WT copies (green bars) and MU copies (blue bars) detected were plotted on the left y-axis. Right y-axis showed the number of droplets analysed (grey dots).

C) Poisson analysis for the MU detection.

D) Linear regression analysis of MU copies input (x-axis) vs. MU copies detected (y-axis). The line has the equation $Y=0.9133X+0.2909$ with the R^2 value of 0.9831 ($p=0.0009$).

5.3.2 KRAS-121, KRAS-122 & KRAS-G13D Assays

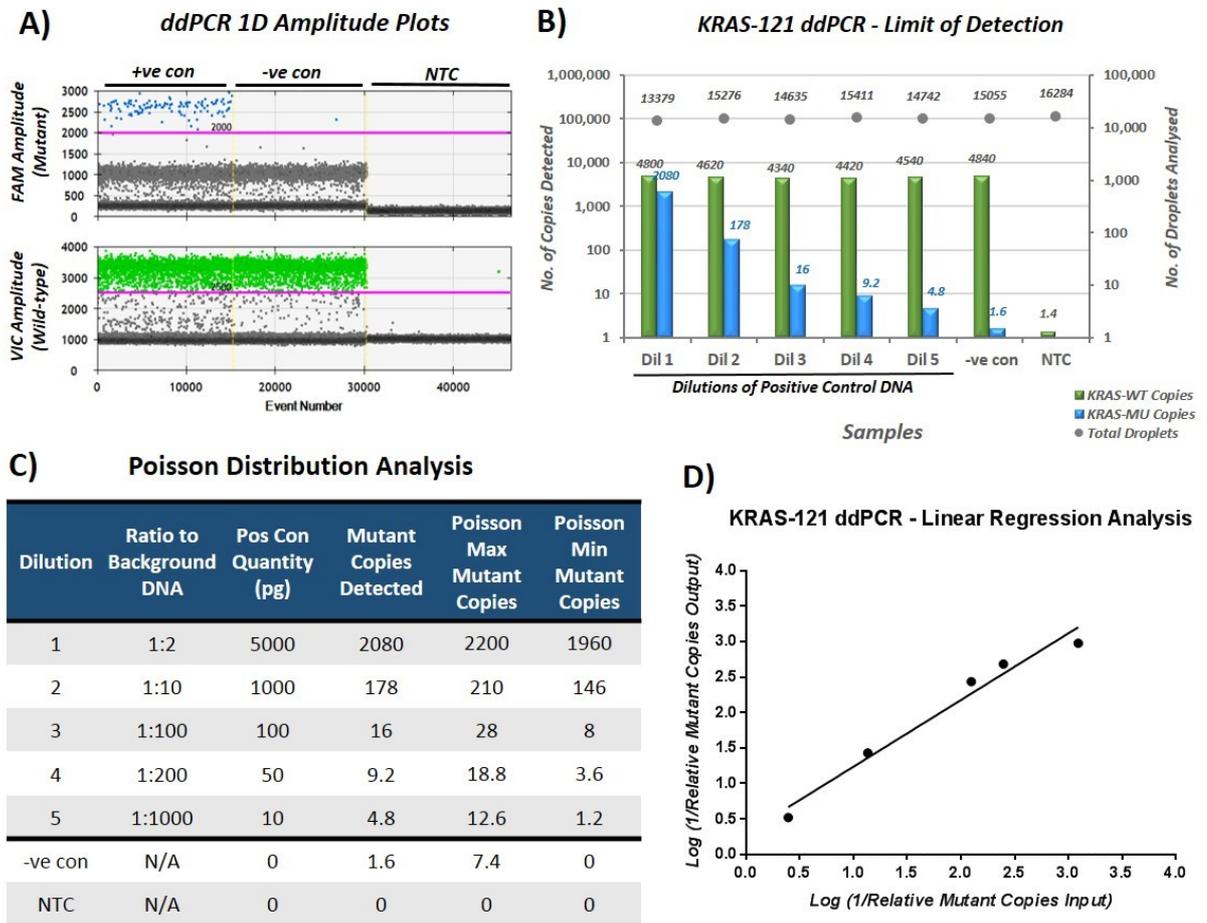


Figure 5-8: Validation of KRAS-121 Assay on ddPCR

A) 1D ddPCR plots showing manual threshold gating (pink lines) for the MU (FAM) and WT (VIC) targets. The positive control (+ve con) was DNA from A549 (*KRAS* c.34G>A) cell line diluted 1:10 in hgDNA (Dil 2). Negative control (-ve con) was hgDNA. NTC = no template control.

B) Dilutions (Dil) 1 to 5 represent reactions with decreasing MU DNA quantity with increasing WT background (see table C). KRAS WT copies (green bars) and MU copies (blue bars) detected were plotted on the left y-axis. Right y-axis showed the number of droplets analysed (grey dots).

C) Poisson distribution analysis for the MU target detection.

D) Linear regression analysis of MU copies input (x-axis) vs. MU copies detected (y-axis). The line has the equation $Y=0.9392X+0.2982$ with the R^2 value of 0.9699 ($p=0.0022$).

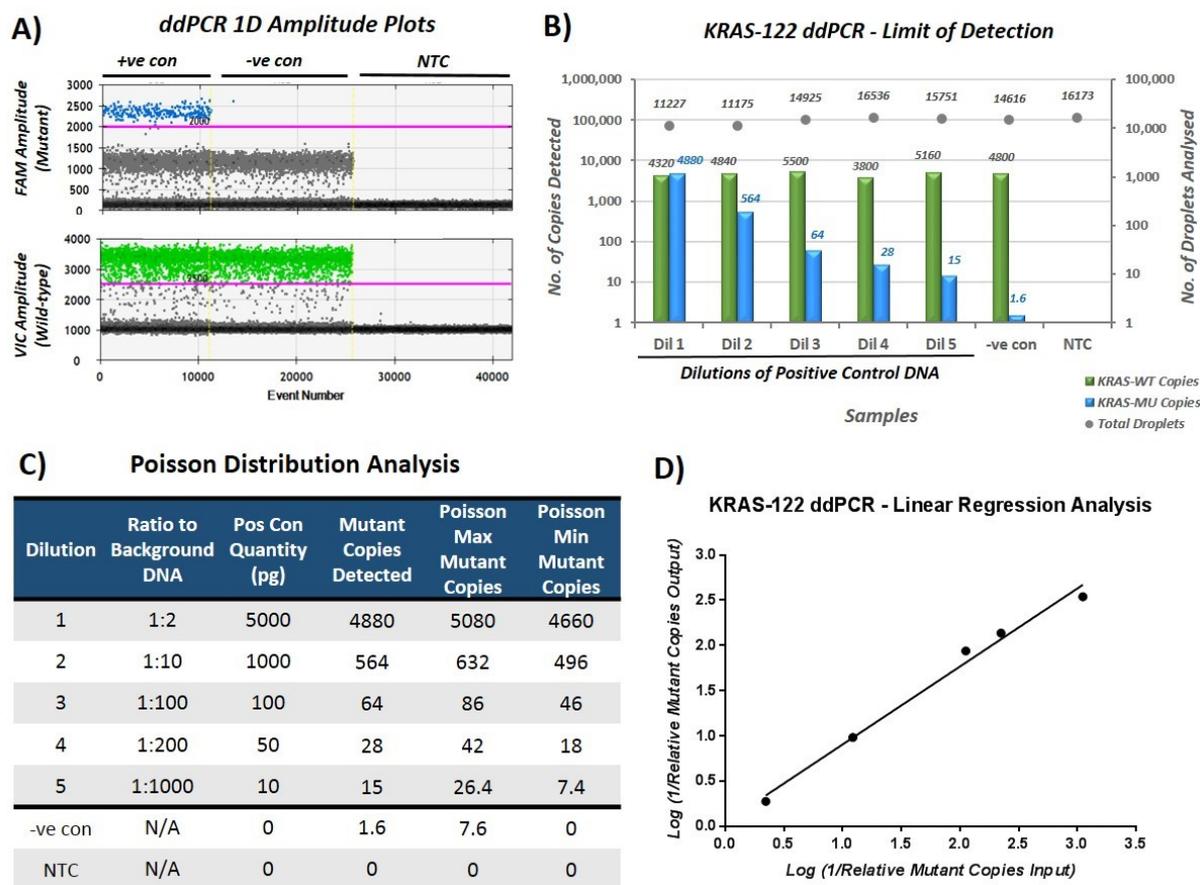


Figure 5-9: Validation of KRAS-122 Assay on ddPCR

A) 1D ddPCR plots showing manual threshold gating (pink lines) for the MU (FAM) and WT (VIC) targets. The positive control (+ve con) was DNA from SW480 (*KRAS* c.35G>T) cell line diluted 1:10 in hgDNA (Dil 2). Negative control (-ve con) was hgDNA. NTC = no template control.

B) Dilutions (Dil) 1 to 5 represent reactions with decreasing MU DNA quantity with increasing WT background (see table C). KRAS WT copies (green bars) and MU copies (blue bars) detected were plotted on the left y-axis. Right y-axis showed the number of droplets analysed (grey dots).

C) Poisson distribution analysis for the MU target detection.

D) Linear regression analysis of MU copies input (x-axis) vs. MU copies detected (y-axis). The line has the equation $Y=0.8625X+0.04289$ with the R^2 value of 0.9874 ($p=0.0006$).

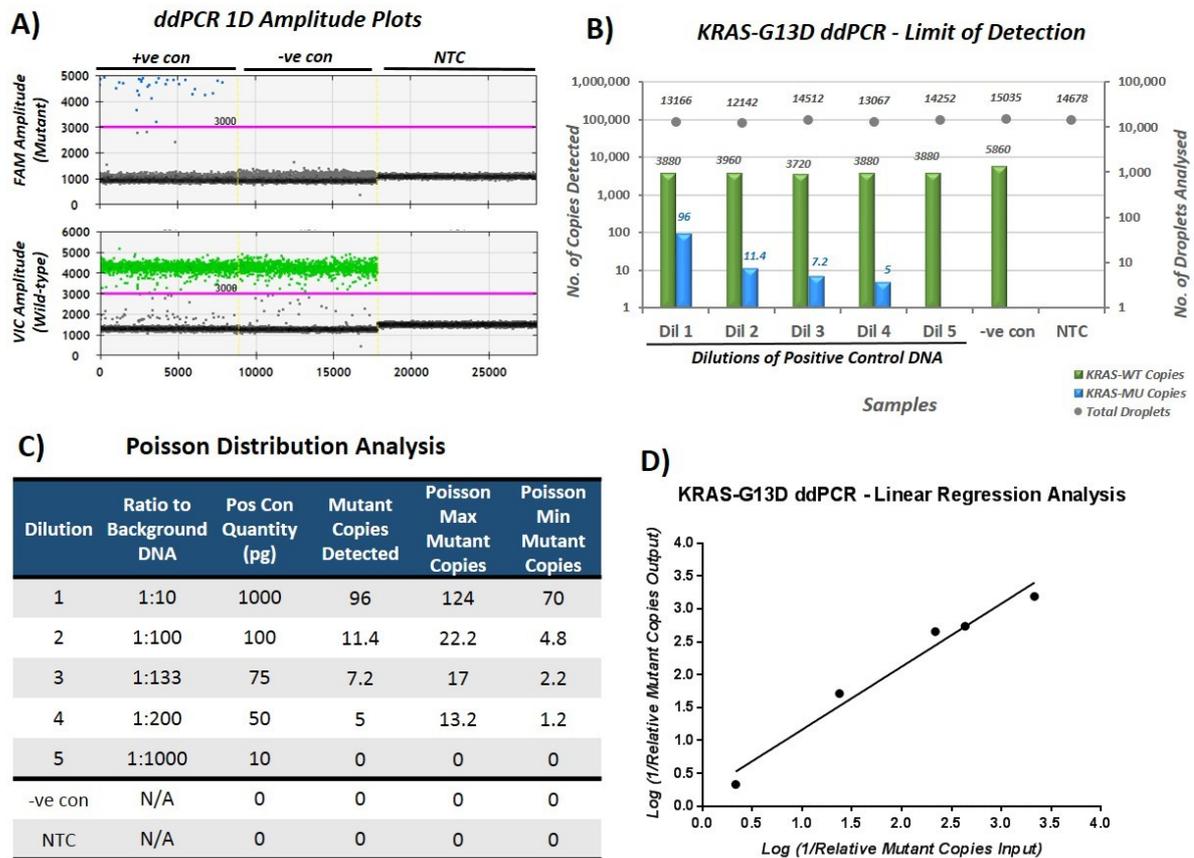


Figure 5-10: Validation of KRAS-G13D on ddPCR

A) 1D ddPCR plots showing manual threshold gating (pink lines) for the MU (FAM) and WT (VIC) targets. The positive control (+ve con) was DNA from HCT116 (*KRAS* c.38G>A) cell line diluted 1:10 in hgDNA (Dil 1). Negative control (-ve con) was hgDNA. NTC = no template control.

B) Dilutions (Dil) 1 to 5 represent reactions with decreasing MU DNA quantity with increasing WT background (see table C). KRAS WT copies (green bars) and MU copies (blue bars) detected were plotted on the left y-axis. Right y-axis showed the number of droplets analysed (grey dots).

C) Poisson distribution analysis for the MU detection.

D) Linear regression analysis of MU copies input (x-axis) vs. MU copies detected (y-axis). The line has the equation $Y=0.9572X+0.2085$ with the R^2 value of 0.9688 ($p=0.0024$).

The KRAS-121, KRAS-122 and KRAS G13D assays performed well with some variability in LoDs for MU alleles (Figure 5-8, Figure 5-9 and Figure 5-10). The KRAS-121 assay on ddPCR system was comparable to its performance on qPCR, showing a LoD at 0.11% MU fractional abundance (4.8 MU copies detected in 4,540 WT background in dilution 5), and this LoD was supported by the Poisson analysis that predicted positive detection of 1.2 MU copies at a minimum level (Figure 5-8, B & C). This assay also showed good correlation between the amount of mutant DNA input and the mutant copies detected ($R^2=0.9699$; Figure 5-8-C).

As for the KRAS-122 assay, its LoD on ddPCR was 0.29% MU fractional abundance (15 MU copies detected in 5,160 WT background in dilution 5; Figure 5-9-B), which was higher than its qPCR counterpart. Interestingly, the Poisson minimum value for MU copies was calculated at 7.4 copies (Figure 5-9-C), which was quite high, suggesting that the LoD might be lower than 0.29%. Similarly, this assay showed good correlation ($R^2=0.9874$; Figure 5-9-C) between the input mutant copies and the number of mutant copies detected. Weak WT and MU signals were observed in the controls for both KRAS-121 and KRAS-122 assays, which were likely caused by cross-contamination during sample preparation or non-specific amplification, a possibility exacerbated by the sensitive nature of these ddPCR assays (Figure 5-8 & Figure 5-9).

As for the commercial KRAS-G13D assay, it also showed a LoD of 0.13% MU fractional abundance (5 mutant copies detected in 3880 WT background copies in dilution 4; Figure 5-10-B), which was comparable to the KRAS-121 assay. This LoD was supported by the Poisson analysis that predicted detection of 1.2 MU copies at the minimum level (Figure 5-10-C). This assay also showed a good correlation ($R^2=0.9688$; Figure 5-10-D) between input and detected copies of MU DNA.

Overall, this validation process confirmed the suitability of these assays for the analysis of matched FFPE and plasma samples from patients with precancerous lesions, and the results will be discussed in the next chapter.

5.4 Validation of Patient-specific Mutation Detection Assays

Five custom ddPCR assays were designed and validated for the detection of patient-specific mutations based on the NGS data. Although the details of the NGS experiments are discussed in the following chapter, work in this chapter described the validation of these custom assays. To this end, the following five targets were selected:

- 3 *APC* non-sense mutations (denoted as *) located in the MCR of the *APC* gene:
 - APC1338 assay – c.4012C>T/p.Q1338*
 - APC1397 assay – c.4189G>T/p.E1397*
 - APC1429 assay – c.4285C>T/p.Q1429*
- 1 *TP53* missense mutation in the DNA binding region:
 - TP53-273 assay – c.817C>T/p.R273C
- 1 *PIK3CA* missense mutation in the catalytic subunit:
 - PIK1047 assay – c.3140A>G/p.H1047R

All assay designs were restricted to <100bp amplicon sizes to facilitate the detection of fragmented cfDNA. For each assay, the WT and MU targets were detected in parallel using different fluorophore conjugations (VIC for WT and FAM for MU). All reactions were run with 10ng total DNA input. Sequences and amplicon sizes were listed in Table 2-9, and experimental conditions were described in section 2.2.3.4.4. Data analysis was done using the QuantaSoft software (version 1.6.6.0320, BioRad®).

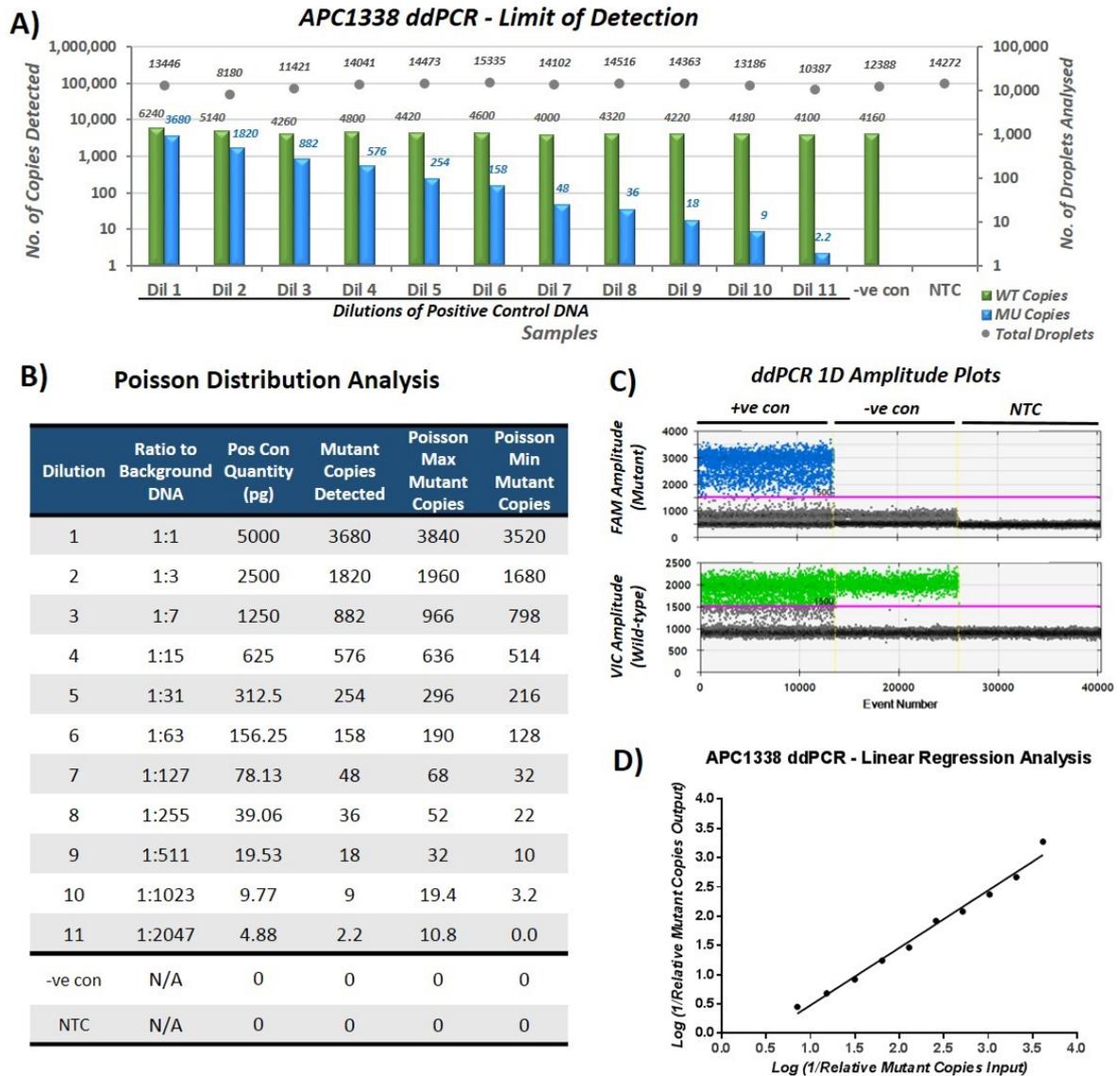


Figure 5-11: Validation of APC1338 Assay on ddPCR

A) DNA extracted from the sequenced FFPE sample (H266) was used as a positive control that has an average % allele frequency of 45.5 for the target MU. Positive control was diluted in 1:2 ratio serially in excess of hgDNA as background. Dilutions (Dil) 1 to 11 represent reactions with decreasing MU DNA quantity. WT (green bars) and MU copies (blue bars) detected were plotted on the left y-axis. Right y-axis showed the number of droplets analysed (grey dots). Negative control (-ve con) was hgDNA. NTC = no template control.

B) Poisson distribution analysis for the MU target detection.

C) 1D ddPCR plots showing manual threshold gating (pink lines) for the MU (FAM) and WT (VIC) targets. The positive control (+ve con) was Dil 1.

D) Linear regression analysis of MU copies input (x-axis) vs. MU copies detected (y-axis). The line has the equation $Y = 0.9796X - 0.4992$ with the R^2 value of 0.9864 ($p < 0.0001$).



Figure 5-12: Validation of APC1397 Assay on ddPCR

A) DNA extracted from the sequenced FFPE sample (H265) was used as a positive control that has an average % allele frequency of 43.0 for the target MU. Positive control was diluted in 1:2 ratio serially in excess of hgDNA. Dilutions (Dil) 1 to 11 represent reactions with decreasing MU DNA quantity. WT (green bars) and MU copies (blue bars) detected were plotted on the left y-axis. Right y-axis showed the number of droplets analysed (grey dots). Negative control (-ve con) was hgDNA. NTC = no template control.

B) Poisson distribution analysis for MU target detection.

C) 1D ddPCR plots showing manual threshold gating (pink lines) for the MU (FAM) and WT (VIC) targets. The positive control (+ve con) was Dil 1.

D) Linear regression analysis of MU copies input (x-axis) vs. MU copies detected (y-axis). The line has the equation $Y=0.9035X-0.4723$ with the R^2 value of 0.9965 ($p<0.0001$).

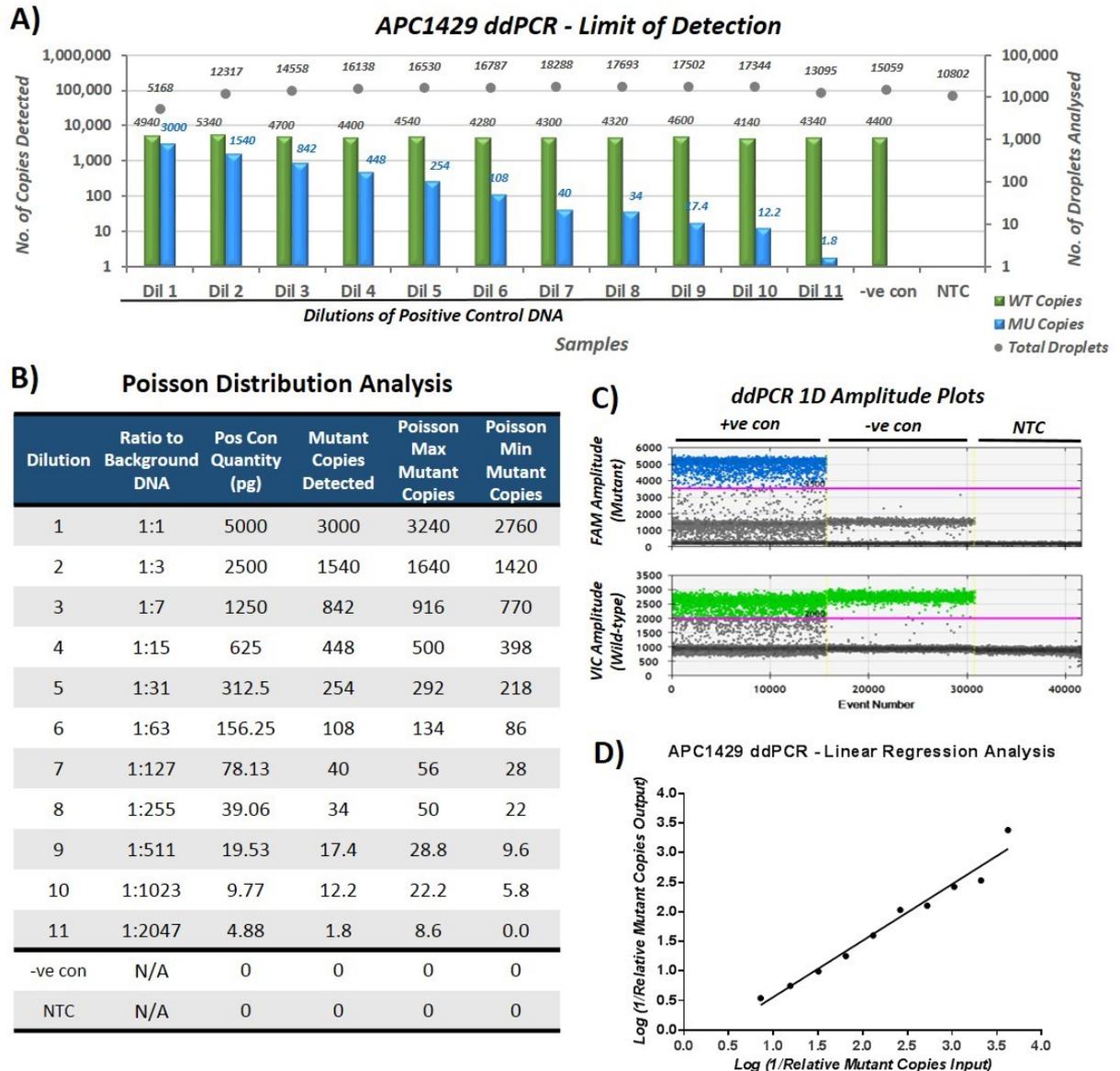


Figure 5-13: Validation of APC1429 Assay on ddPCR

A) DNA extracted from the sequenced FFPE sample (H263) was used as a positive control that has an average % allele frequency of 44.4 for the target MU. Positive control was diluted in 1:2 ratio serially in excess of hgDNA. Dilutions (Dil) 1 to 11 represent reactions with decreasing MU DNA quantity. WT (green bars) and MU copies (blue bars) detected were plotted on the left y-axis. Right y-axis showed the number of droplets analysed (grey dots) for each reaction. Negative control (-ve con) was hgDNA. NTC = no template control.

B) Poisson distribution analysis for the mutant target detection.

C) 1D ddPCR plots showing manual threshold gating (pink lines) for the MU (FAM) and WT (VIC) targets. The positive control (+ve con) was Dil 1.

D) Linear regression analysis of MU copies input (x-axis) vs. MU copies detected (y-axis). The line has the equation $Y=0.9548X-0.3964$ with the R^2 value of 0.9706 ($p<0.0001$).

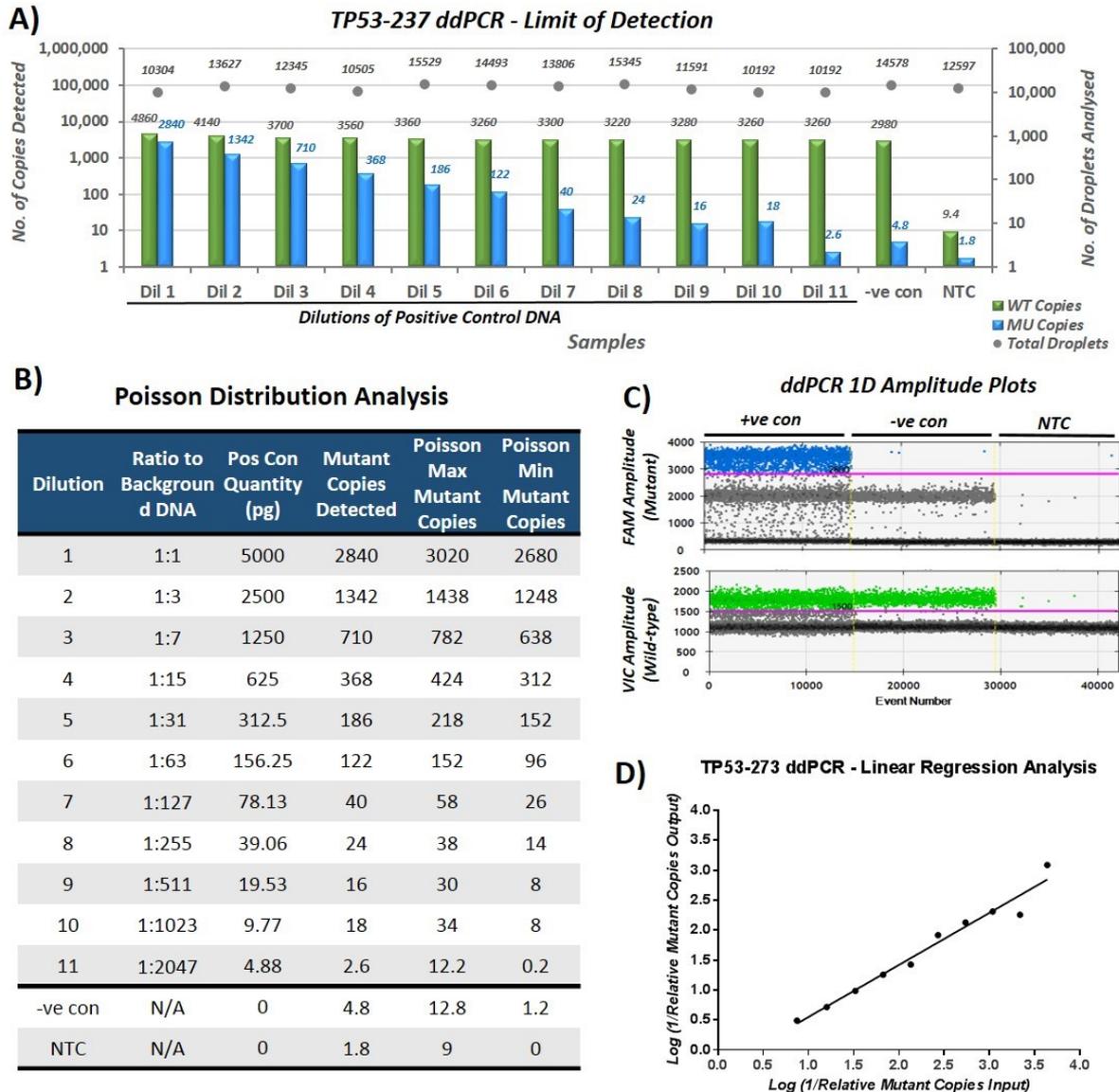


Figure 5-14: Validation of TP53-273 Assay on ddPCR

A) DNA extracted from sequenced FFPE sample (H266) was used as a positive control that has an average % allele frequency of 52.5 for the target MU. Positive control was diluted in 1:2 ratio serially in excess of hgDNA as background. Dilutions (Dil) 1 to 11 represent reactions with decreasing MU DNA quantity. WT (green bars) and MU copies (blue bars) detected were plotted on the left y-axis. Right y-axis showed the number of droplets analysed (grey dots) for each reaction. Negative control (-ve con) was hgDNA. NTC = no template control. Both the negative control and NTC showed contamination.

B) Poisson distribution analysis for the mutant target detection.

C) 1D ddPCR plots showing manual threshold gating (pink lines) for the MU (FAM) and WT (VIC) targets. The positive control (+ve con) was Dil 1.

D) Linear regression analysis of MU copies input (x-axis) vs. MU copies detected (y-axis). The line has the equation $Y=0.8704X-0.3201$ with the R^2 value of 0.9672 ($p<0.0001$).

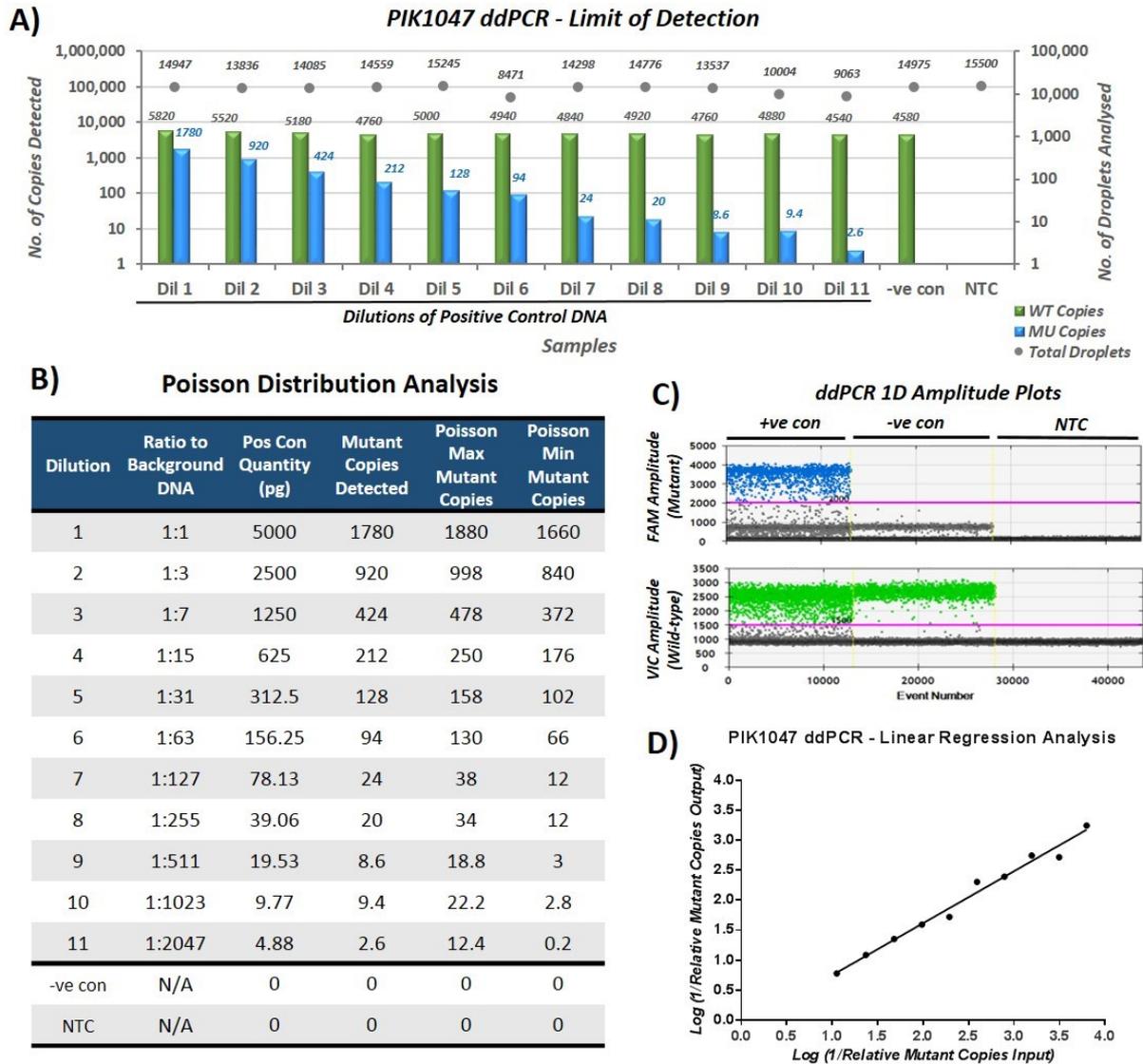


Figure 5-15: Validation of PIK1047 Assay on ddPCR

A) DNA extracted from sequenced FFPE sample (H263) was used as a positive control that has an average % allele frequency of 9.6 for the target MU. Positive control was diluted in 1:2 ratio serially in hgDNA as background. Dilutions (Dil) 1 to 11 represent reactions with decreasing MU DNA quantity. WT (green bars) and MU copies (blue bars) detected were plotted on the left y-axis. Right y-axis showed number of droplets analysed (grey dots) for each reaction. Negative control (-ve con) was hgDNA. NTC = no template control.

B) Poisson distribution analysis for the mutant target detection.

C) 1D ddPCR plots showing manual threshold gating (pink lines) for the MU (FAM) and WT (VIC) targets. The positive control (+ve con) was Dil 1.

D) Linear regression analysis of MU copies input (x-axis) vs. MU copies detected (y-axis). The line has the equation $Y=0.8671X-0.1187$ with the R^2 value of 0.982 ($p<0.0001$).

Table 5-1: LoDs for Patient-specific Mutation Detection Assays

<i>Assay Name</i>	<i>Figure Reference</i>	<i>MU Copies Detected at Lowest Dilution</i>	<i>WT Copies Detected at Lowest Dilution</i>	<i>MU Fractional Abundance (%)</i>
<i>APC1338</i>	Figure 5-11	9.0	4180	0.215
<i>APC1397</i>	Figure 5-12	6.6	4300	0.153
<i>APC1429</i>	Figure 5-13	5.8	4140	0.294
<i>TP53-273</i>	Figure 5-14	18.0	3260	0.549
<i>PIK1047</i>	Figure 5-15	9.4	4880	0.192

Whilst all five custom assays had good correlations (R^2 values approaching 1.0) between the input MU copies and the number of MU copies detected, they showed varying LoDs for respective MU targets (Table 5-1). The assays targeting the 3 *APC* mutations (*APC1338*, *APC1397*, *APC1429*) and the *PIK1047* all have low LoDs at MU fractional abundances ranging from 0.153% to 0.294% (Table 5-1). However, for the *TP53-273* assay, the LoD was relatively high at 0.549% (18 MU copies in a background of 3260 WT copies in dilution 10; Table 5-1). The next dilution (Dil 11 at 1:2047 ratio) for the *TP53-273* assay did detect 2.6 MU copies (Figure 5-14-A), but the Poisson analysis gave 0.2 MU copies at the minimum level of detection (Figure 5-14-B), suggesting that the assay might be unable to consistently call positive MU targets at such a low MU fraction, indicating that the effective LoD for *TP53-273* assay was indeed at dilution 10. Moreover, the *TP53-273* assay was compounded by non-specific amplifications present in both the negative control and NTC wells (Figure 5-14-A). One possible cause for these was cross-contamination with positive controls during the droplet transfer process. Another possibility was that the negative control (hgDNA), which was a commercial sample consisting of genomic DNA from various sources, contained a high MU baseline – this observation has been reported by Diehn’s group as they observed a high baseline for *TP53* mutations in their healthy cohort [301]. Nevertheless, this was the best LoD that could be achieved for the *TP53-273* assay after the trial and error of testing various combinations of primer and probe sequences.

Overall, this validation process ensured that these assays were performing optimally for the analysis of patient tissue and plasma samples, and the results will be discussed in the next chapter.

5.5 Summary

This chapter described the validation of various PCR-based assays for the analysis of patient samples. The qPCR-based mutation detection assays with PNA clamping had been validated to detect the *BRAF* and *KRAS* mutations specifically and sensitively, with the LoDs at 0.05% for BRAF-V600E (Figure 5-1), 0.10% for KRAS-121 (Figure 5-4) and 0.19% for KRAS-122 assays (Figure 5-5). The higher LoDs for the KRAS assays were likely due to the use of degenerate probes to detect all three possible base substitutions at bases c.34G (KRAS-121) and c.35G (KRAS-122) in codon 12. During the validation process, an unexpected problem arose with the KRAS-121 assay where PNA clamping was ineffective at blocking WT signals when both MU and WT targets were tested in parallel (Figure 5-2) – whilst the cause of this was not elucidated, it might have been due to cross-interactions between two probes and the PNA. However, as PNA clamping was necessary to enhance MU detection at low fractional abundances, an alternative strategy where the WT and MU targets were detected in separate reactions was used for the KRAS-121 assay runs (Figure 5-4). All three assays showed good efficiencies: BRAF-V600E at 98.68%, KRAS-121 at 90.32% and KRAS-122 at 93.07% (Figure 5-6).

Half-way through this project, our laboratory acquired the QX200™ ddPCR system, prompting the transfer of validation processes to this platform. As the ddPCR system eliminated the need for PNA clamping due to enriching of MU detection through droplet partitioning, the assay development process was made more cost-effective and time-efficient. To this end, the BRAF-V600E, KRAS-121 and KRAS-122 assays were re-validated on the ddPCR. The BRAF-V600E assay showed a LoD at above 0.06% MU fractional abundance (Figure 5-7), comparable to its qPCR counterpart. The same was true for the KRAS-121 assay, which showed a comparable LoD at 0.11% (Figure 5-8). Interestingly, the KRAS-122 assay had a higher LoD at 0.29% (Figure 5-9), but it was possible that the lowest dilution tested (dilution 5) did not reach the LoD threshold. In addition to these assays, a commercial KRAS-G13D assay was tested for MU detection at codon 13 (c.38G>A/p.G13D), and this was shown to have a LoD of 0.13% (Figure 5-10).

Moreover, five patient-specific mutation detection assays were designed based on the NGS data obtained from selected patient cases with large adenomas (Chapter 6). The mutation targets were 3 nonsense substitutions in the MCR of *APC* gene (APC1338,

APC1397 and APC1429 assays), 1 missense mutation in the DNA-binding domain of *TP53* (TP53-273 assay) and 1 missense mutation in the catalytic subunit of *PIK3CA* (PIK1047 assay). All these five custom assays were able to efficiently detect their respective MU targets with a variable range of LoDs (Figure 5-11 to Figure 5-15), with lowest LoD observed for the APC1397 assay at 0.153% (Table 5-1). As for the TP53-273 assay, despite testing more than eight combinations of primer-probe pairs, the lowest LoD achieved was 0.549%. The validation of this assay did prove to be arduous with the persistent issue of non-specific amplification in the controls (Figure 5-14-A), despite strict care being taken during the assay preparation to minimise cross-contamination as much as possible from using a sterilised PCR station to UV-treating consumables prior to each assay run. Another possibility was that these signals were off-target, low-efficiency amplification, which could be filtered out by increasing the T_m during the extension phase to increase stringency of complementary hybridisation. Notwithstanding, these custom assays did exhibit sensitive detection of respective MU targets, warranting their use in the analysis of patient plasma samples, and the results will be discussed in the next chapter.

Chapter 6 Analysis of Tissue & Plasma Samples from Patients with Colorectal Adenomas

6.1 Introduction

Many studies on cfDNA have reported its potential clinical use as a predictive and prognostic biomarker for advanced CRC cases [306, 310, 312, 372, 381]. However, studies regarding the role of cfDNA in precancerous lesions have been few in comparison due to its challenging nature, although several groups have reported a correlation between elevated cfDNA levels and development of early lesions [299, 312]. Nonetheless, further research is necessary to investigate the practicality of using cfDNA as a biomarker for early disease detection and monitoring preventive interventions, and this project aimed to fulfil this gap.

Based on the studies that reported an increase in plasma cfDNA levels in cancer patients, even in early stages of CRC [299, 319], a similar analysis was performed to assess whether the same correlation can be observed in my patient cohort. However, as it is known that total cfDNA levels are influenced by factors other than tumour development (e.g. trauma, autoimmune diseases) [142, 261, 263-265], I also performed mutation-specific detection assays were performed to assess whether targeting the tumour-derived cfDNA was a feasible undertaking in these early stage samples by implementing the qPCR and ddPCR assays validated in Chapter 5. Subsequently, matched plasma and tissue samples from mutation-positive patient cases were analysed. Furthermore, six patient cases with large and/or multiple adenomas were selected for multiregional targeted NGS (see appendix 8.2 for details) in collaboration with the Wellcome Trust Sanger Institute, Cambridge, UK. This was undertaken to identify mutation targets that appeared at high frequencies for subsequent development of five patient-specific mutation detection assays, which have been discussed in section 5.4. These assays were used to detect patient-specific mutations in the matched plasma samples, the results from which will be discussed in this chapter.

6.2 Histopathology of Patient Samples

Samples analysed for this project were obtained from the patients who participated in the national BCSP (section 2.1.7). Out of 131 patient cases with early colorectal lesions, 76 were available for research as described below; the remaining samples were excluded due to limited tissue availability.

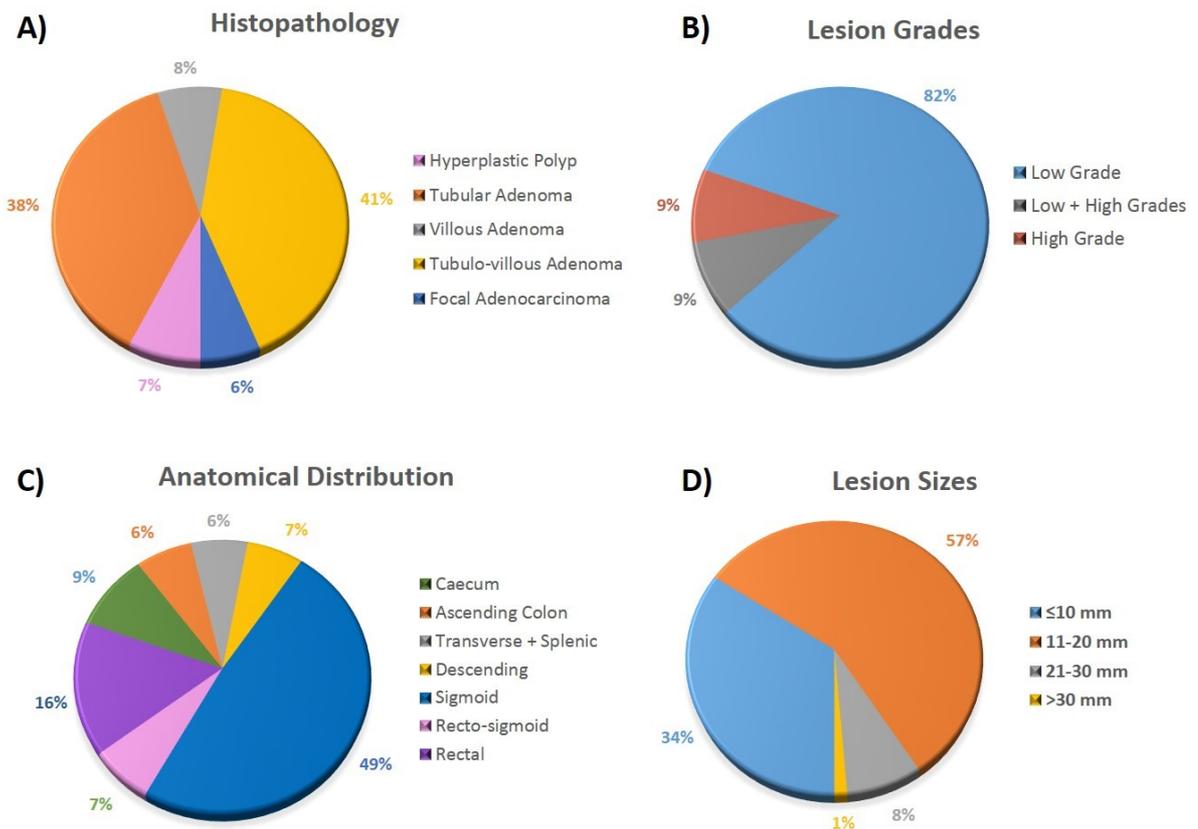


Figure 6-1: Histopathological Information of Patient Samples

All data were obtained from histology reports for the 76 patient cases.

A) Histopathology. Distribution of different histological subtypes. A small portion of lesions (6%) showed focal adenocarcinoma (AC) within adenomatous regions.

B) Lesion grades. Pathology assessment showed that most were low-grade lesions (82%).

C) Anatomical distribution. Half of the lesions were found in the sigmoid colon (49%), with a significant portion located in the rectum (16%). The rest were distributed throughout the colon.

D) Lesion sizes. Longest measurements of diameter listed on histology report were used; measurements of cut FFPE sections were used as substitutes when a report did not state any measurements. Only one patient case had a lesion with >30mm in diameter.

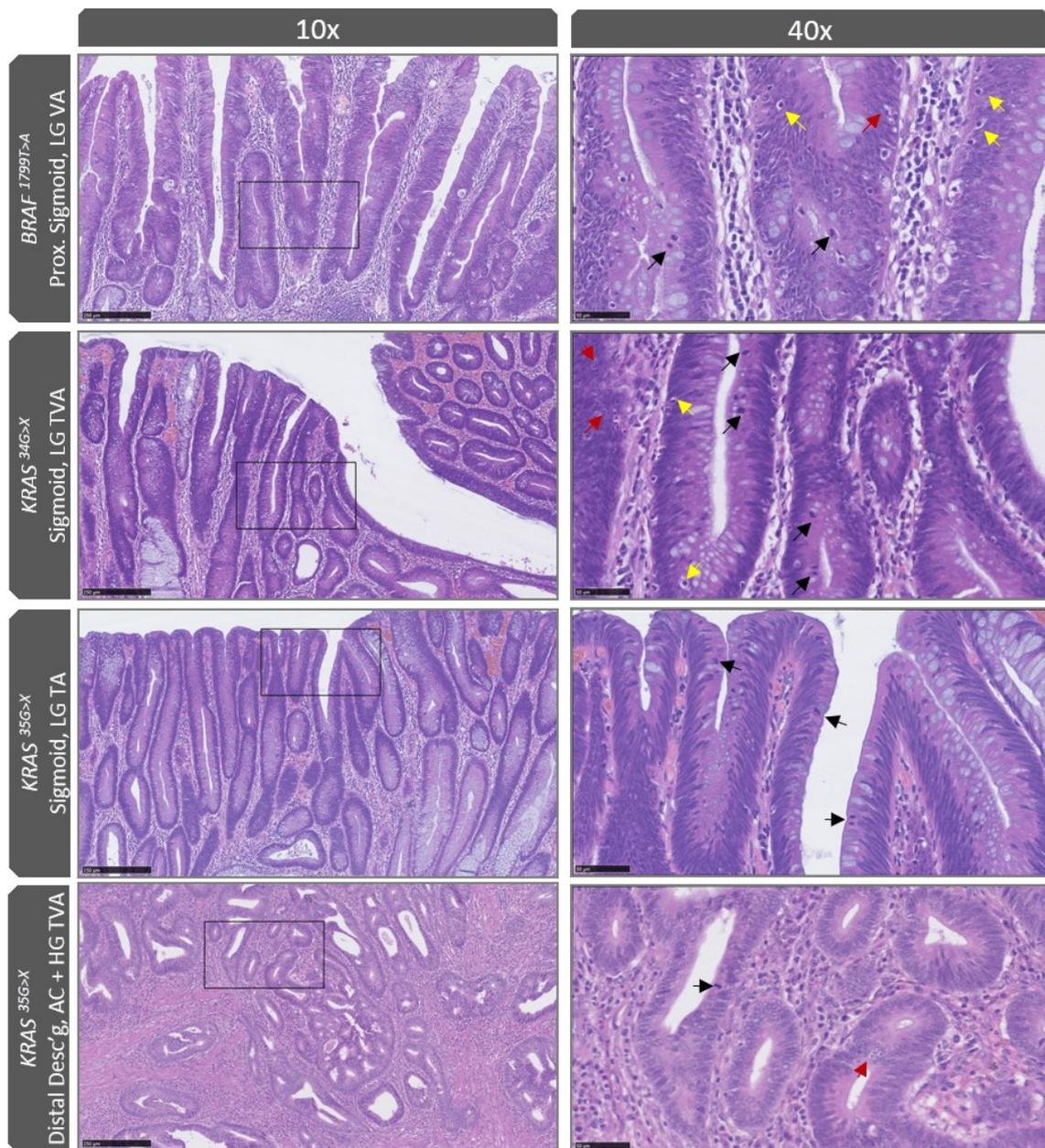


Figure 6-2: Representative Histological Subtypes in H&E Stained FFPE samples

H&E staining was as described in section 2.2.6.1. Images were captured using the NanoZoomer-XR scanner. Black boxes in the images on the left column (10x, scale bar= 250µm) were magnified in the right column (40x, scale bar= 50µm). Proliferative nature of these regions was indicated by the presence of mitotic features (black arrows). Apoptotic bodies (red arrows) and tumour-infiltrating lymphocytes (yellow arrows) were also present. Mutations, location of polyps and grades were listed on the tabs. LG = low grade, HG = high grade, VA = villous adenoma, TA = tubular adenoma, TVA = tubulo-villous adenoma, AC = adenocarcinoma, 'X' base change = A/C/T.

The pie charts in Figure 6-1 illustrated the histopathology information for the 76 cases used in this project with regards to the histological subtypes, grades, location and size of the lesions. The predominant histological subtypes were TVA (41%) and TA (38%), with 82% of all cases classified as low grade lesions by the clinical pathologists. There were few cases of advanced adenomas with focal adenocarcinomas (6%) (Figure 6-2). The majority of lesions were surgically excised from the sigmoidal (49%) and rectal (16%) regions, with the remaining tissue samples more or less evenly distributed throughout the colon, which was a similar pattern to the distribution observed amongst CRC patients in the UK population (Figure 1-3). More than 90% of lesions tested were small; 34% were ≤ 10 mm in diameter and 57% were 11-20mm in diameter (Figure 6-1).

6.3 Screening for *BRAF* & *KRAS* Mutations in FFPE Samples

A query of the COSMIC database showed that mutations in *BRAF* and *KRAS* oncogenes occurred at significant frequencies, 9% and 37% respectively, in adenoma samples, with other notable mutations present in *APC* (51%), *TP53* (27%), *GNAS* (23%) and *CTNNB1* (13%) (Figure 6-3) [56]. Similarly, comprehensive molecular profiling by TCGA network also identified *BRAF* and *KRAS* mutations at 7.36% and 36.1% respectively in sporadic CRCs, irrespective of the hypermutation status [124]. Using the previously validated *BRAF* and *KRAS* assays (Chapter 5), patient FFPE tissue DNA samples were screened for the hotspot *BRAF* and *KRAS* mutations. FFPE DNA extraction was detailed in section 2.2.2.2.3; as most were small lesions (Figure 6-1-D), whole FFPE tissue sections were processed for DNA extraction without enriching for adenomatous areas. DNA sample aliquots were quantified using the hALU69 qPCR assay (section 2.2.3.3.4), and 10ng DNA was tested per sample. For each sample, four assays were run to check for the *BRAF* and *KRAS* mutations – the *BRAF*-V600E, *KRAS*-121 and *KRAS*-122 assays were run on the StepOnePlus qPCR system, whereas the *KRAS*-G13D assay was run on the QX200 ddPCR system. All experimental conditions were detailed in Chapter 2. For the qPCR results, positive mutation detection was determined as $<10 \Delta C_t$ cycles for the MU signals using the WT reference [382]. For the ddPCR results, positive detection was determined at $\geq 1\%$ fractional abundance of MU copies [348]. Amplification signals detected within these cut-off values represent detection of at least 10 MU copies in a

background of 990 WT copies, in a reaction containing 1000 detectable copies [348, 382].

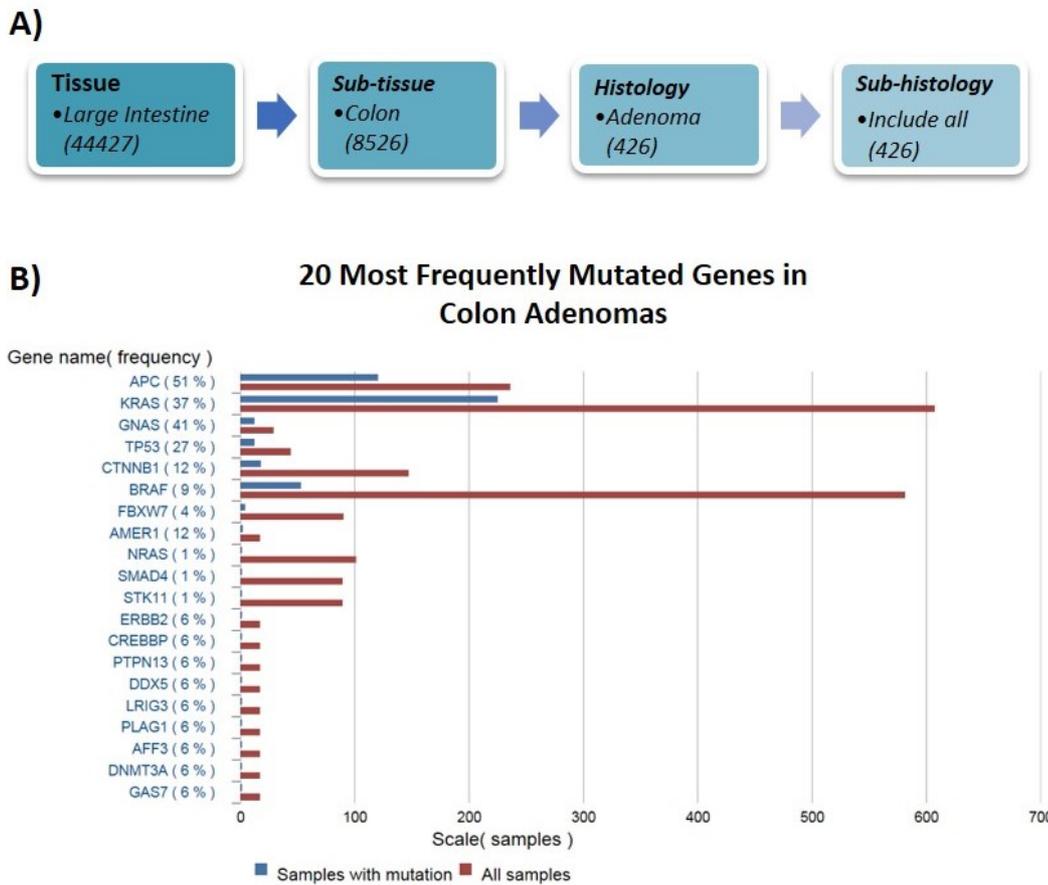


Figure 6-3: Frequently Mutated Genes in Colon Adenomas from the COSMIC Database

A) The search criteria used for querying the COSMIC database. [50]

B) 20 most frequent mutations in colon adenomas. For each mutation, the blue bars represent the number of positive samples, and the red bars represent the total number tested. Mutation frequencies were calculated based on the ratio of positive-to-total samples for each gene entry.

The *BRAF* hotspot mutation at V600E was identified at a frequency of 5.26% (4 out of 76 cases) in this patient cohort, slightly lower than what has been generally reported in the TCGA network and COSMIC database [90, 91, 94, 95, 124] (Figure 6-3). However, Chang's group identified a similar frequency of *BRAF* mutations (4.2%) in their adenoma samples (n=191) using exome sequencing [383]. All four *BRAF*-positive cases were from the recto-sigmoid region, with 3 of them classified as HPs and 1 as a low-grade VA. All 4 cases measured <10mm in their largest diameter. This study cohort showed a higher frequency of *BRAF* mutations compared to Velculescu's group who reported 0% *BRAF* mutation frequency in their adenoma samples (n=34) with ≤10mm

diameter [374]. It was likely that the differences in sample preparation and the use of a more targeted approach – a qPCR assay with PNA clamping against Velculescu’s exome sequencing – contributed to the identification of few positive *BRAF*-MU cases in my patient cohort [374].

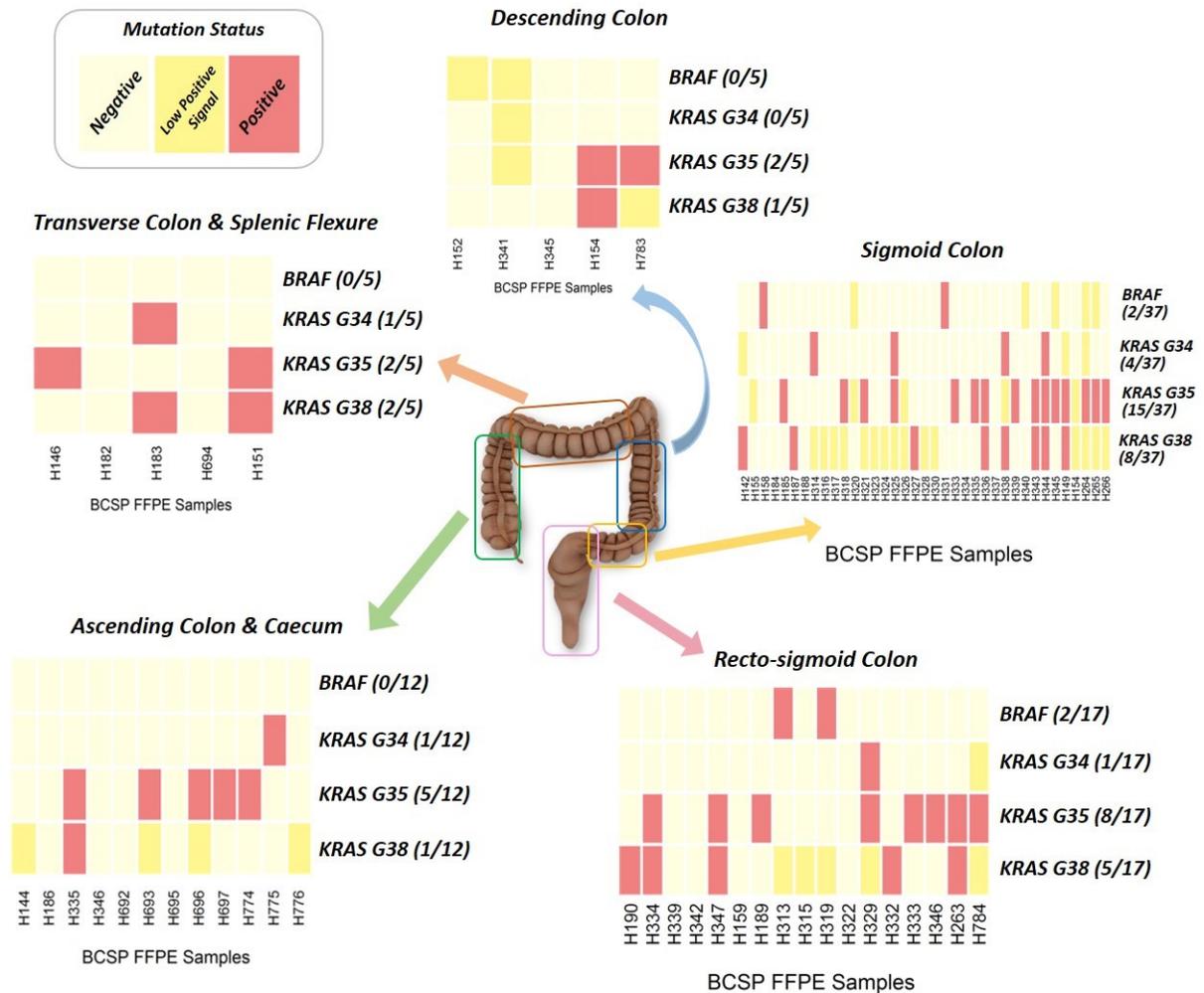


Figure 6-4: Distribution of *BRAF* and *KRAS* Mutations in the Colon

Data from the four mutation detection assays (*BRAF*-V600E, *KRAS*-121, *KRAS*-122 and *KRAS*-G13D) were compiled to show the *BRAF* and *KRAS* mutation status of samples in relation to their anatomical locations. Low positive signal was determined as either >10 C_t cycle difference between WT (VIC) and MU (FAM) signals for *BRAF*-V600E, *KRAS*-121 (c.34G) and *KRAS*-122 (c.35G) qPCR assays, or <1% fractional abundance of mutant copies for *KRAS*-G13D (c.38G>A) ddPCR assay. 4 out of 76 cases (5.26%) were positive for the *BRAF*-V600E assay. As for *KRAS* mutations, base substitutions at c.34G were found in 7 cases (9.21%), at c.35G in 32 cases (42.1%) and c.38G in 17 cases (22.4%). Overall, 44 out of 76 cases (57.9%) have at least one *KRAS* mutation, and 14 cases (out of 44) showed two or more positive signals for mutations affecting different bases for *KRAS* oncogene.

In this patient cohort, *KRAS* mutations were identified in 44 out of 76 (57.9%) patient cases (Figure 6-4), a frequency higher than the COSMIC dataset (Figure 6-3). Out of these 44 *KRAS*-positive cases, mutations at the base c.35G was the most common with 32 cases (42.1%) – of these, 16 cases were TVAs, 10 were TAs, 5 were VAs and 1 was a benign HP. Mutations at the base c.38G were found in 17 cases (22.4%) – of these, 12 cases were TVAs and 5 were TAs. Mutations at the base c.34G was found in 7 cases (9.21%) – of these, 3 cases were TVAs, 2 were TAs, 1 was VA and 1 was an unknown subtype with focal AC. *KRAS* mutations did not show a predilection for a GI region in this cohort. However, *KRAS* positivity was more common in the TVA subtype than the TA, despite screening similar numbers of TVA and TA cases (41% vs. 38% of 76 cases respectively; Figure 6-2-A). Due to the highly sensitive nature of these assays, sometimes trace amounts of MU signal (yellow boxes - Figure 6-4) could be detected; however, at such a low MU signal, it was difficult to confidently identify these samples as *KRAS* mutant. Interestingly, 14 out of 76 (18.4%) cases showed double *KRAS* mutations, with 12 of those cases (15.8%) affecting codon 12 and 13 simultaneously. This phenomenon has also been reported in the literature where double mutations affecting *KRAS* codons 12 and 13 were observed at a frequency of 27% in CRC patients [384, 385]. The simultaneous presence of multiple mutations in *KRAS* suggested a robust, and underappreciated, degree of tumour heterogeneity in early adenomas.

Overall, the above data identified a total of 48 patient cases bearing mutations in either *KRAS* or *BRAF*, which served as a basis for the subsequent detection of these tumour-derived mutations in matched plasma samples.

6.4 Quantification of Total Plasma cfDNA Levels

For this experiment, 48 patient cases (i.e. the 'patient' group) that were positive for at least one *BRAF* or *KRAS* mutation were selected for quantitative analysis of total plasma cfDNA levels to test the hypothesis that elevated cfDNA levels are associated with the presence of early lesions [299, 319]. The 'control' group consisted of 37 individuals that had a negative colonoscopy outcome following referral for a false-positive FOBT result (section 2.1.7). Plasma cfDNA extraction was described in section

2.2.2.2, and 3µl plasma aliquot was tested per reaction using the hGAPDH assay (section 2.2.3.3.5). Each sample was run in triplicates.

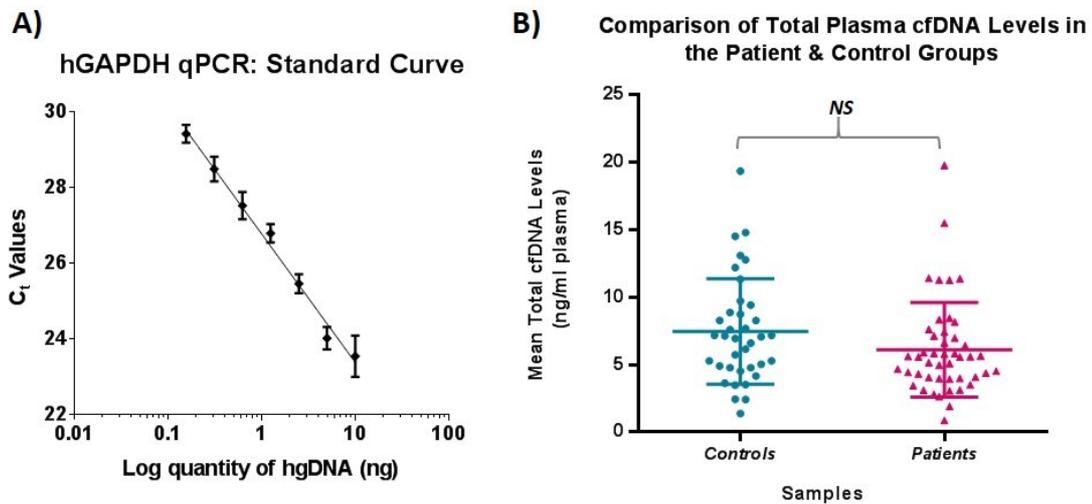


Figure 6-5: Comparison of Total Plasma cfDNA Levels in the Patient & Control Groups

Comparing total plasma cfDNA levels in the patient (n=48) and the control (n=37) groups.

A) Standard Curve (semi-log line) plotted for C_t values (y-axis) against serial dilutions of hgDNA as a standard. The line has the slope of -3.397, y-intercept at 26.78 and R² value of 0.9678. The assay efficiency was 96.96%.

B) Each dot represents one sample. Horizontal bars represent mean ±SD. The mean total cfDNA levels were 7.456 ± 0.642 ng/ml for the control group and 6.106 ± 0.5167 ng/ml for the patient group (p=0.1012 at 95% CL; parametric, unpaired t-test).

The mean total cfDNA concentration was slightly higher in the control group (7.456 ±0.642 ng/ml) compared to the patient group (6.106 ±0.5167 ng/ml), but this difference was not statistically significant (p=0.1012; Figure 6-5). Although some polyp-free individuals in the control group suffered from underlying physiological conditions (e.g. colitis, diverticulosis; appendix 8.1), this was unlikely to have biased the analysis, as my data did not show any overt deviation from the quantification data of healthy cohorts reported in other studies [299, 319]. For instance, Cree's group measured a mean cfDNA concentration of 7.96ng/ml for their control group, and Watanabe's group also reported a comparable mean cfDNA level of 7.7ng/ml in their healthy controls; both groups used the repetitive *LINE-1* retrotransposon locus as a reference for quantification [299, 319]. Notwithstanding the use of different reference standards (i.e. *LINE1* vs. *hGAPDH*), their data supported the results obtained for the control group. In comparison, the cfDNA levels in my patient group were lower than the published reports: Watanabe's group quantified a mean cfDNA concentration of 9.8ng/ml in their

cohort (n=57) and Cree's group reported a markedly higher cfDNA level of 15.04ng/ml in their cohort (n=26). However, the main difference was that their patient groups contained more advanced cases: Watanabe's group included both Stage I and II samples, and Cree's group included large high-grade polyps (mean size 56.1mm) in their sample cohorts. The difference in the sample stages between our patient cohorts might explain the differences in plasma cfDNA levels observed between our experiments. It was also possible that the small sizes of polyps in this patient group also diminished the extent of cfDNA release into circulation. This lack of significant differences in total plasma cfDNA levels between patients with premalignant lesions and healthy individuals has also been reported by Pilotti's group [386].

Overall, this experiment indicated that no discernible alteration in total plasma cfDNA concentrations were observed in patients with precancerous lesions compared to the healthy controls, at least for this patient cohort. Nevertheless, the lack of an increase in plasma cfDNA levels did not necessarily exclude the possibility that tumour-specific mutations could be detected in the plasma. Therefore, the next stage of analysis was to investigate whether *BRAF* and *KRAS* mutations identified in the FFPE samples could also be detected in the matched plasma samples.

6.5 Analysis of *BRAF* & *KRAS* Mutations in Plasma cfDNA Samples

Matched plasma samples to the 48 mutation-positive cases were lyophilised (section 2.2.2.3.1) and reconstituted in 25µl dH₂O prior to mutation detection. This was done to increase the input cfDNA concentrations to maximise the probability of detecting tumour-derived cfDNA. Assay conditions were described in section 2.2.3.4.3. All data were analysed using the QuantaSoft software (version 1.6.6.0320, BioRad®).

6.5.1 Analysis of *BRAF* Mutation in the Plasma cfDNA Samples

The plasma samples, corresponding to the 4 *BRAF*-MU FFPE cases identified in section 6.3, were analysed using the *BRAF*-V600E assay on ddPCR (section 5.3.1). For each sample, 3.0µl lyophilised plasma cfDNA aliquot was tested per reaction well. The assay was run twice.

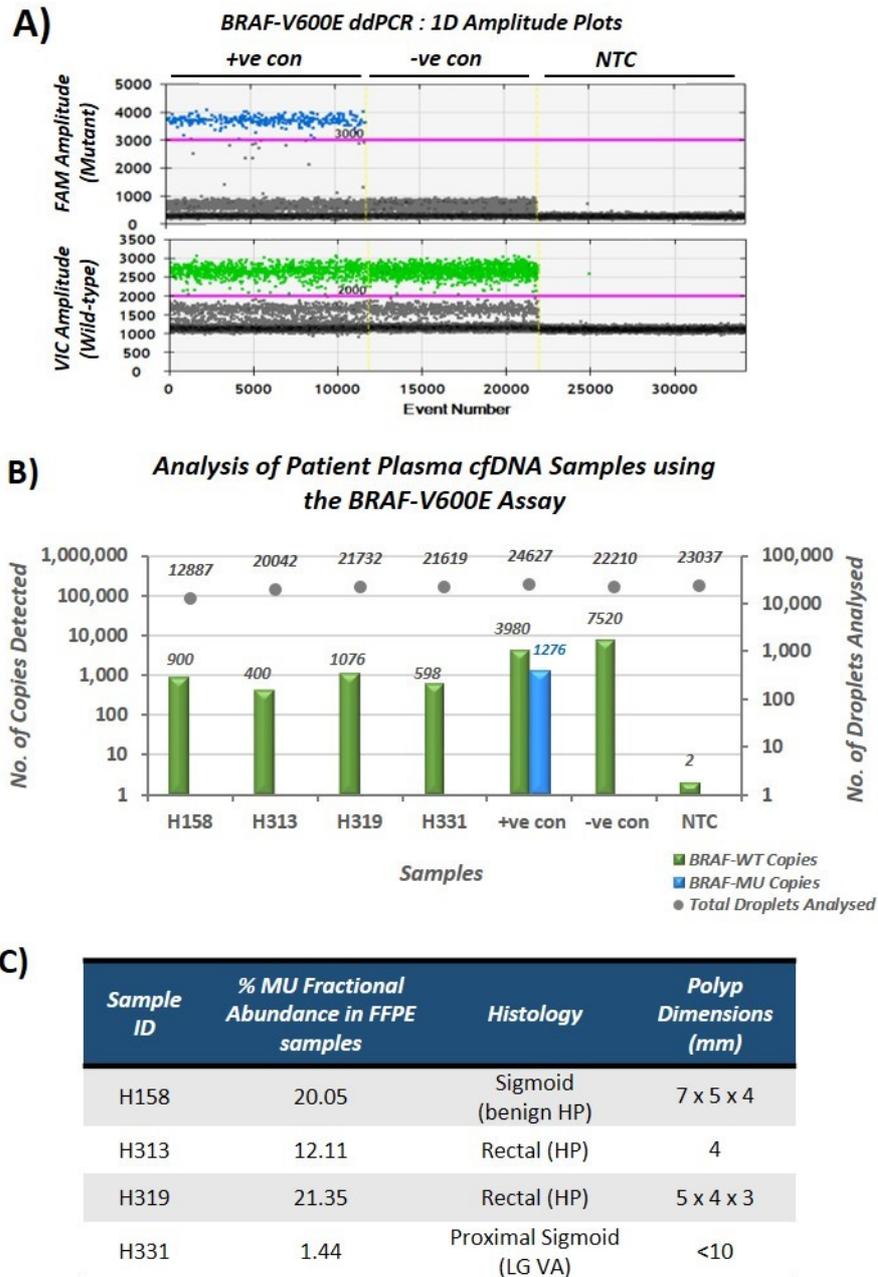


Figure 6-6: Testing Lyophilised Plasma cfDNA Samples using the BRAF-V600E Assay

A) 1D ddPCR plots showing manual threshold gating (pink lines) for the WT (VIC) and MU (FAM) amplifications. Positive control (+ve con) was 10ng of HT-29 (c.1799T>A/p.V600E) DNA. Negative control was 10ng of hgDNA. NTC = no template control.

B) Merged data from two runs of the BRAF-V600E assay. WT (green bars) and MU (blue bars) copies detected were shown on left y-axis. Number of droplets analysed (grey dots) were shown on right y-axis. NTC well showed a possible cross-contaminated WT signal.

C) Tabulated data of patient samples tested. Histology data and polyp dimensions were obtained from histology reports. HP = hyperplastic polyp, LG = low-grade, VA = villous adenoma.

Unfortunately, no *BRAF*-MU cfDNA copies were detected in any of the four plasma cfDNA samples analysed (Figure 6-6-A). One possible cause was that all *BRAF*-MU adenomas were very small (diameter ≤ 10 mm; Figure 6-6-B), and presented with low MU fractional abundances (1.44-21.35%) in matching tissue samples (Figure 6-6-B), suggesting that their representation in the total cfDNA was also likely to be weak. The total cfDNA quantification data obtained in section 6.4 (also see appendix 8.1) also showed that these patient samples had low plasma cfDNA levels (0.88-4.01 ng/ml) compared to the average concentrations reported in the literature, which ranged from 9.8-15.04 ng/ml [299, 319]. This indicated that these samples were likely to have very low MU fractions, which could have contributed to the lack of MU detection in these plasma samples. The difficulty with MU detection was likely exacerbated by low droplet counts, especially for sample H158 where a total of 12,887 droplets were analysed between two assay runs – this was below the threshold of $>10,000$ droplets/reaction.

6.5.2 Analysis of *KRAS* Mutations at base c.34G in Plasma cfDNA Samples

Next, a total of 13 matched plasma samples from *KRAS* (c.34G>X) positive cases, including those showing borderline *KRAS* positivity (Figure 6-4), were tested using the *KRAS*-121 assay on the ddPCR system (section 5.3.2). The cfDNA samples were lyophilised and processed as previously described, and 3.0 μ l lyophilised aliquots were tested per reaction well. The assay was run twice.

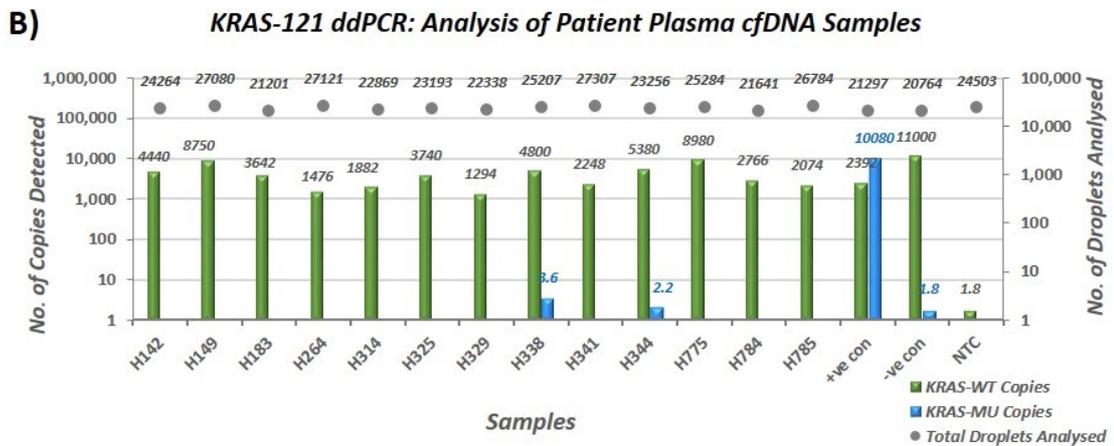
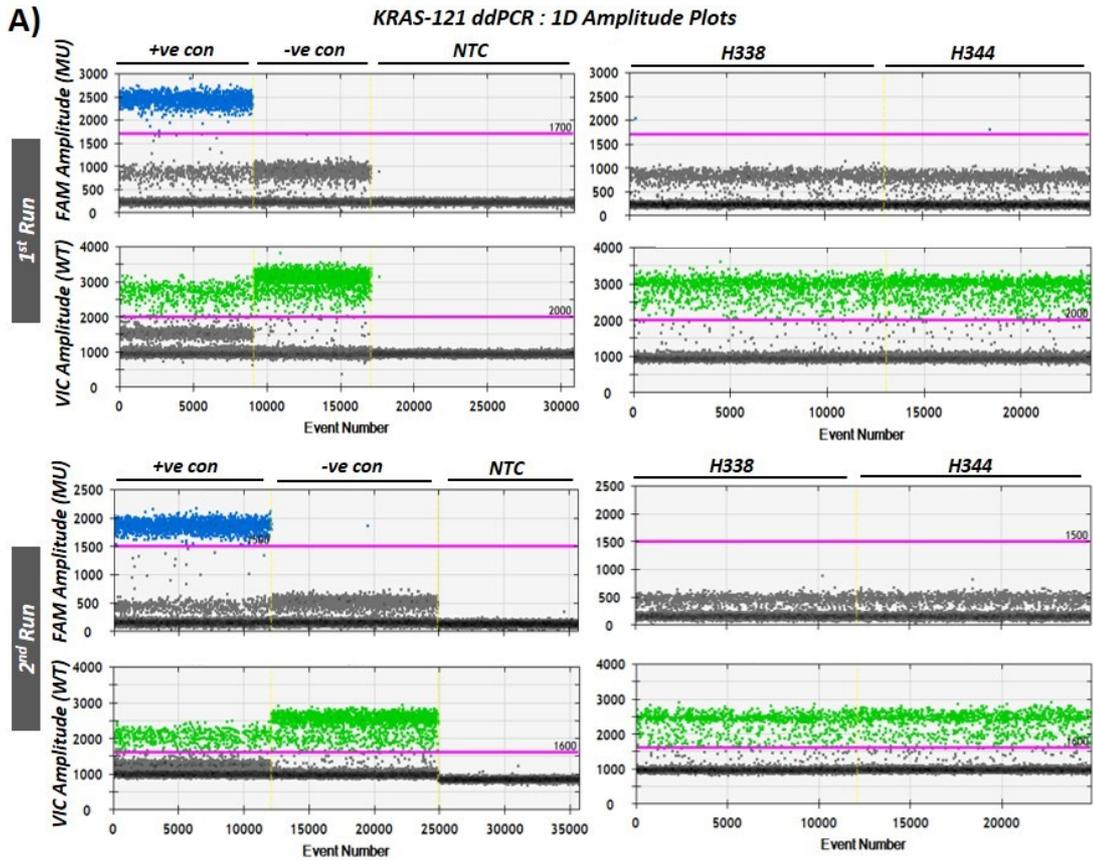


Figure 6-7: Testing for *KRAS* c.34G Mutations in the Plasma cfDNA Samples using the *KRAS*-121 Assay

A) 1D plots with manual threshold gating (pink lines) for the WT (VIC) and MU (FAM) amplification. Positive control (+ve con) was 10ng of A549 (c.34G>A/p.G12S) DNA. Negative control was 10ng of hgDNA. NTC = no template control. The 1D plots for the two samples (H338 and H344) that showed positive MU (FAM) amplification were also shown on the right.

B) Merged data from two assay runs. WT (green bars) and MU (blue bars) copies detected were shown on left y-axis. Number of droplets analysed (grey dots) were shown on right y-axis.

Table 6-2: Poisson Distribution Analysis for Mutant-positive cases from Figure 6-7

Sample	Total MU copies	Total WT copies	% MU fractional abundance	Assay no.	MU Copies/ Well	PoissonMax (no. of MU copies)	PoissonMin (no. of MU copies)
H338	3.6	4800	0.075	1 st	3.6	11.6	0.6
				2 nd	0.0	0.0	0.0
H344	2.2	5380	0.041	1 st	2.2	10.6	0.0
				2 nd	0.0	0.0	0.0
-ve con	1.8	11000	0.016	1 st	0.0	0.0	0.0
				2 nd	1.8	8.8	0.0

Table 6-1: Histology Data & Mutation Frequencies for the Matched FFPE Cases Corresponding to Plasma cfDNA Samples Analysed in Figure 6-7

Sample ID	No. of FFPE Blocks	% Mutant Fractional Abundance in FFPE samples	Histology	Polyp Dimensions (mm)
H142	1	0.0895	Sigmoid (LG TVA)	17 x 10 x 15
H149	8	0.0000 – 0.0229	Sigmoid (LG TVA)	15 x 15 x 15 – 25 x 20 x 25
H183	1	0.9667	Mid Transverse (LG TA)	10 x 6
H264	7	0.0000 – 0.2559	Sigmoid (LG + HG TVA)	5 x 5 x 3 – 10 x 10 x 6
H314	1	2.7618	Distal Sigmoid (LG TA)	16 x 7 x 5
H325	1	1.0923	Colonic (LG TVA)	20 x 14 x 8
H329	2	0.0006 – 5.3010	Rectal (LG VA)	11 x 9 x 9 – 14 x 9 x 7
H338	2	0.7336 – 3.5509	Sigmoid (LG TVA)	16 x 15 x 10
H341	3	0.0000 – 0.1422	Descending (LG TVA)	9 x 8 x 7 – 20 x 15 x 16
H344	2	0.0000 – 1.9799	Colonic (LG TVA)	20 x 20 x 10
H775	1	11.093	Colonic (MD AC, Duke's C1)	N/A
H784	2	0.0000 – 0.1084	Colonic (Polyp – HG VA, AC – invasive, pT1)	N/A
H785	1	9.0330	Recto-sigmoid (sessile HG TVA, focal pT1 AC)	16 x 11 x 9

Histology and polyp dimensions were obtained from the histology reports. For cases where more than one FFPE block were tested, lowest and highest mutation frequencies were reported.

LG = low-grade, HG = high-grade, TA = tubular adenoma, VA = villous adenoma, TVA = tubulovillous adenoma, MD = mild differentiated, AC = adenocarcinoma.

Out of 13 samples analysed, few MU cfDNA copies were detected in only 2 cases: H338 and H344 (Figure 6-7). Few positive copies were detected in the negative control and NTC wells, which were likely caused by minor cross-contamination (Figure 6-7). For the two positive samples, MU fractional abundances in the plasma was calculated at 0.075% for H338 (3.6 MU copies in 4800 WT copies) and 0.041% for H344 (2.2 MU copies in 5380 WT copies) (Table 6-2). Poisson analysis for these two samples predicted that, at the minimum level of droplet distribution, there is a possibility of detecting 0.0-0.6 MU copies (i.e. negative) for both samples (Table 6-2), suggesting inconsistency with positive calling for these samples. Moreover, the fact that positive MU copies were only detected in one out of the two independent ddPCR runs for both samples agreed with the Poisson analysis (Figure 6-7-A, Table 6-2). The difficulty with MU detection was also exacerbated by the fact that most samples have very low MU fractions in the matched tissue samples (Table 6-1). Although stromal DNA might have contributed to diluting the MU fractions in the tissues, the data in Table 6-2 nevertheless suggested these *KRAS* mutations might be clonal for some cases. For instance, two FFPE samples were analysed for case H329, and one sample showed a very low MU fraction of 0.0006% whilst the other showed a higher fraction of 5.3010%, which was highly suggestive of *KRAS* clonality. This would have made it more difficult to detect tumour-derived mutations in the plasma [318].

6.5.3 Analysis of *KRAS* Mutations at base c.35G in the Plasma cfDNA Samples

Similar to the previous analysis, the *KRAS*-122 ddPCR assay (section 5.3.2) was used to test 34 plasma samples matching the *KRAS* (c.35G>X) positive cases, including those that showed borderline MU positivity. The cfDNA samples were lyophilised and processed as previously described, and 3.0µl lyophilised aliquots were tested per reaction. The assay was run twice. However, duplicate testing of several samples was not possible – IDs noted in the Figure 6-9 – because they had been identified with more than one mutation, which necessitated their inclusion in two or more assays that led to complete exhaustion of small sample volumes.

KRAS-122 ddPCR : 1D Amplitude Plots

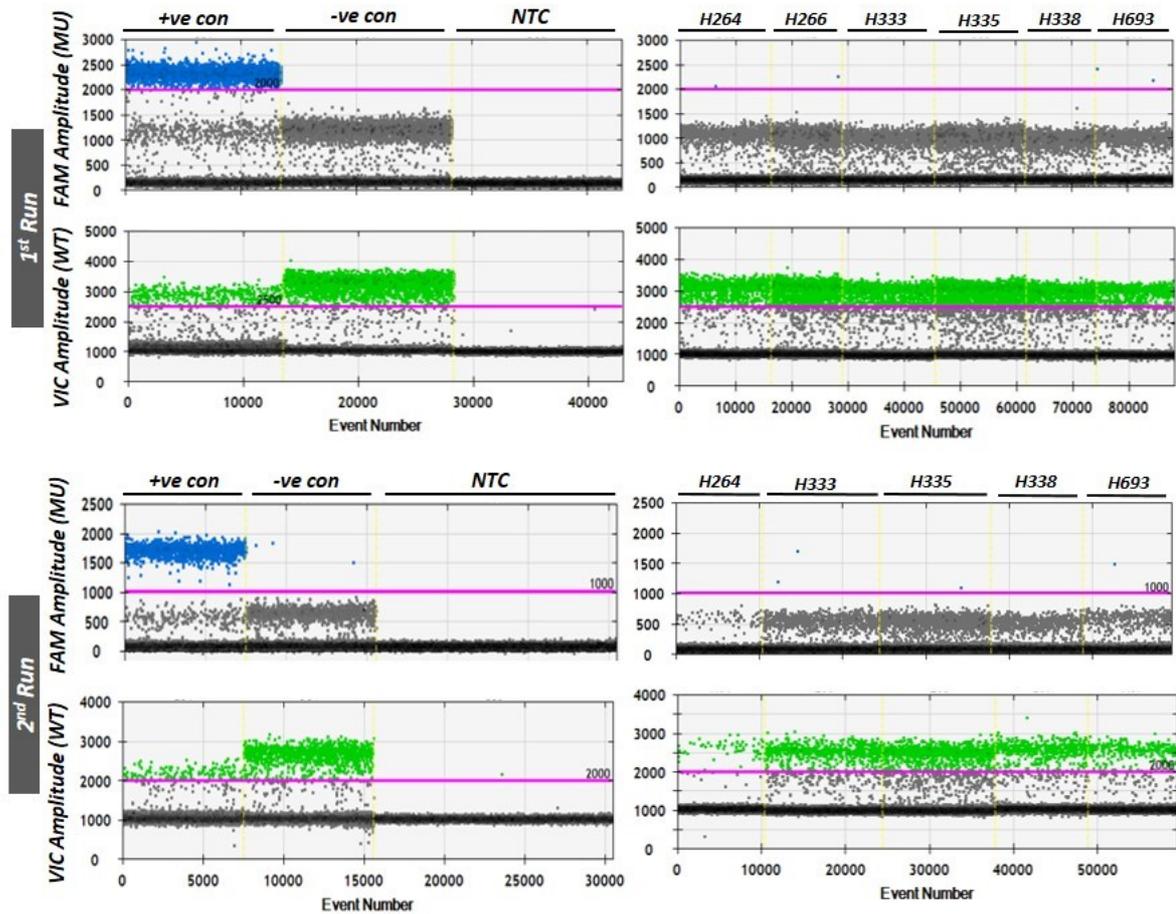
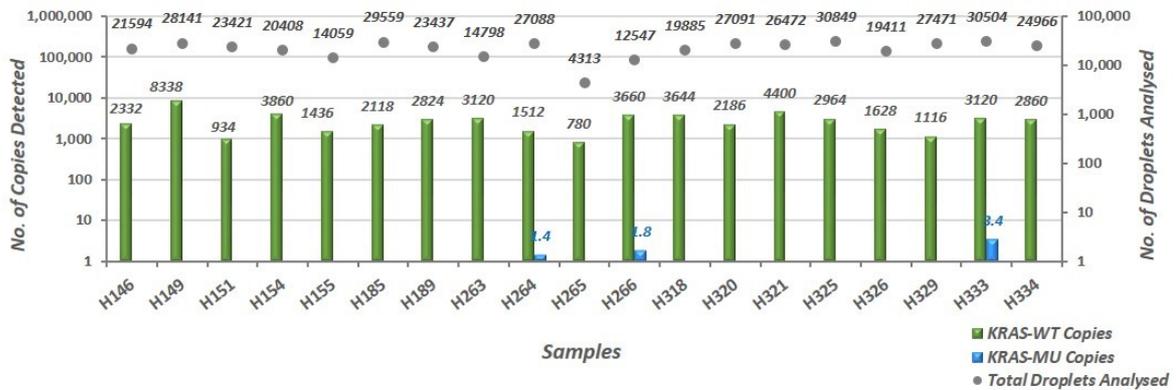


Figure 6-8: Testing for *KRAS* c.35G Mutations in the Plasma cfDNA Samples using the *KRAS*-122 Assay – 1D Plots

These 1D ddPCR plots were accompanying figures to the graphs in Figure 6-9. Each plot showed manual threshold gating (pink lines) for the WT (VIC) and MU (FAM) amplification. Positive control (+ve con) was 10ng of SW480 (c.35G>T/p.G12V) DNA. Negative control was 10ng hgDNA. NTC = no template control.

The 1D plots for the six samples (H264, H266, H333, H335, H338 and H693) that showed positive MU (FAM) amplification were shown on the right. Due to the lack of sample aliquot, H266 was only tested once (1st run). H264 and H266 were positive for MU amplification only in the 1st assay run. H333, H335 and H338 were positive for MU amplification only in the 2nd assay run. Only H693 tested positive for *KRAS* mutation in both assay runs.

KRAS-122 ddPCR: Analysis of Patient Plasma cfDNA Samples (Graph 1 of 2)



KRAS-122 ddPCR: Analysis of Patient Plasma cfDNA Samples (Graph 2 of 2)

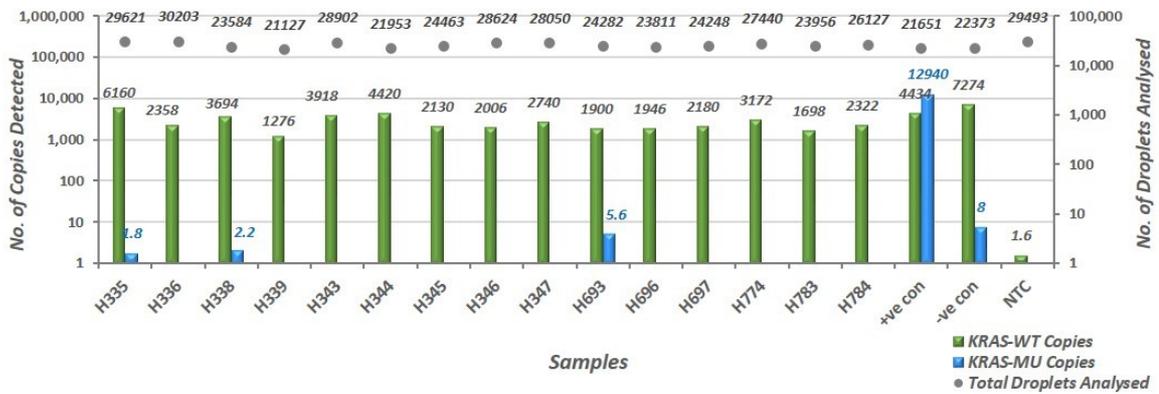


Figure 6-9: Testing for KRAS c.35G Mutations in the Plasma cfDNA Samples using the KRAS-122 Assay – Graphs for all Samples tested

Merged data from two runs of the KRAS-122 assay. WT (green bars) and MU (blue bars) copies detected were shown on left y-axis. Number of droplets analysed (grey dots) were shown on right y-axis. Controls were as described in Figure 6-8. Despite the repeated runs, some reactions (H155, H265 and H326) failed to meet the threshold for the accepted number of droplets (>10,000 droplets/reaction). 3 samples (H263, H265 and H266) were only tested once due to insufficient sample volume as these were also used for the analysis described in section 6.7.

Table 6-3: Histology Information for the Matched FFPE Cases Corresponding to Plasma cfDNA Samples Analysed in Figure 6-9

Sample ID	No. of FFPE Blocks	% Mutant Fractional Abundance in FFPE samples	Histology	Polyp Dimensions (mm)
H146	1	9.4001	Transverse (benign HP)	8 x 5 x 4
H149	8	0.0000 – 0.2459	Sigmoid (LG TVA)	15 x 15 x 15 – 25 x 20 x 25
H151	1	3.1402	Splenic Flexure (LG TA)	5 x 4
H154	4	0.0000 – 10.8343	Descending-sigmoid (LG TA/TVA)	17 x 15 x 14
H155	1	0.0733	Sigmoid (LG + HG TA)	10 x 8 x 5
H185	1	1.2587	Sigmoid (LG TVA)	10 x 8 x 17
H189	1	18.612	Rectal (LG TA)	12
H263	11	0.0000 – 6.2256	Rectal (LG + HG TVA, focal AC)	17 x 30 x 10 – 30 x 25 x 22
H264	7	0.0000 – 0.2563	Sigmoid (LG + HG TVA)	5 x 5 x 3 – 10 x 10 x 6
H265	13	19.5678 – 37.4255	Distal Sigmoid (LG + HG VA)	55 x 35 x 20
H266	9	12.9668 – 26.7434	Sigmoid (LG VA)	N/A
H318	1	0.4275	Sigmoid (LG TVA)	15
H320	1	0.0621	Sigmoid (LG TVA)	11
H321	1	0.9613	Colonic (LG + HG TVA)	15 x 7
H325	1	1.0923	Colonic (LG TVA)	20 x 14 x 8
H326	1	0.0969	Colonic (LG TA)	7 x 5 x 6
H329	2	4.5123 – 9.0246	Rectal (LG VA)	11 x 9 x 9 – 14 x 9 x 7
H333	2	6.2192 – 8.3315	Recto-Sigmoid (LG TA)	19 x 16 x 10
H334	2	0.0000 – 1.5870	Recto-Sigmoid (LG TA + TVA)	7 x 9 x 4 – 12 x 10 x 4
H335	2	2.8364 – 8.9973	Sigmoid-Caecal (LG TVA)	10 - 12
H336	2	5.3522 – 11.8328	Sigmoid (LG TVA)	14
H338	2	0.0000 – 0.0188	Sigmoid (LG TVA)	16 x 15 x 10
H339	2	0.0005 – 1.2086	Recto-sigmoid (LG TA)	6 x 5 x 5 – 12 x 8 x 8
H343	2	0.7317 – 1.3000	Sigmoid (LG TVA)	15 x 10 x 15
H344	2	2.1852 – 2.8104	Colonic (LG TVA)	20 x 20 x 10
H345	2	0.0016 – 1.5049	Descending-sigmoid (LG TA)	7 x 5 x 4 – 11 x 9 x 7
H346	2	0.0003 – 2.5250	Ascending-Rectal (LG TA)	12-15
H347	2	0.7243 – 1.8001	Recto-sigmoid (LG TVA)	21 x 15 x 10
H693	2	0.3259 – 1.6247	Caecal (LG TA)	N/A
H696	2	7.0348 – 7.4390	Ascending (LG TVA)	N/A
H697	1	5.4522	Caecal (LG TA)	10 x 4 x 2
H774	2	9.3524 – 18.4350	Caecal (AC Duke's B)	N/A
H783	2	13.9502 – 14.6877	Distal descending (HG TVA with focal AC)	N/A
H784	2	18.9137 – 23.7891	Colonic (Polyp – HG VA, AC – invasive, pT1)	N/A

For cases where more than one FFPE block were tested, lowest and highest mutation frequencies were reported.

HP = hyperplastic polyp, LG = low-grade, HG = high-grade, TA = tubular adenoma, VA = villous adenoma, TVA = tubulovillous adenoma, MD = mild

Table 6-4: Poisson Distribution Analysis for MU-positive cases from Figure 6-9

Sample	Total MU copies	Total WT copies	% MU fractional abundance	Assay no.	MU Copies/ Well	PoissonMax (no. of MU copies)	PoissonMin (no. of MU copies)
H264	1.4	1512	0.093	1 st	1.4	6.8	0.0
				2 nd	0.0	0.0	0.0
H266	1.8	3660	0.049	1 st	1.8	9.0	0.0
				2 nd	N/A	N/A	N/A
H333	3.4	3120	0.109	1 st	0.0	0.0	0.0
				2 nd	3.4	10.8	0.6
H335	1.8	6160	0.029	1 st	0.0	0.0	0.0
				2 nd	1.8	8.4	0.0
H338	2.2	3694	0.060	1 st	0.0	0.0	0.0
				2 nd	2.2	10.2	0.0
H693	5.6	1900	0.294	1 st	3.4	11.0	0.6
				2 nd	2.2	10.6	0.0
-ve con	8.0	7274	0.110	1 st	0.0	0.0	0.0
				2 nd	8.0	24	2.0

Of 34 plasma samples tested, only 6 samples (17.6%) were positive for *KRAS*-MU copies in the plasma (Figure 6-9) and they were detected over a range of MU fractional abundances – the lowest was 0.029% for H335 and the highest at 0.294% for H693 (Table 6-4). Sample H693 was the only case that showed MU positivity in both assay runs (Table 6-4). For all positive cases, Poisson analysis predicted detection of 0.0-0.6 MU copies at the minimum threshold – as they were <1.0 MU copies, this strongly suggested the possibility of obtaining false negatives in independent runs (Table 6-4). Indeed, with the exception of H266 that was only tested once, positive MU calling was inconsistent in the two independent assays run for 4 out of 5 samples (Table 6-4). Additionally, the results showed no correlation between the matched FFPE and plasma MU fractions (Table 6-3 and Table 6-4); although it can be said that the calculations were subjected to bias depending on how much stromal DNA tissue had diluted the MU signal. As an example, for the case H693, its FFPE tissues showed MU fractions of 0.3259-1.6247%, yet the matched plasma has the highest MU fraction (0.294%) amongst the positive plasma samples tested (Table 6-3; Table 6-4). In contrast, FFPE sections from case H333 were tested to have high MU fraction (6.2192-8.3315%), yet the matched plasma has a relatively low MU fraction of 0.109% (Table 6-3; Table 6-4). Surprisingly, case H265, whose FFPE analysis gave a high MU fraction in the tissues, failed to show any MU

positivity in the plasma, although it was very likely due to poor droplet generation (4,313 droplets/reaction) (Table 6-3; Figure 6-9). Nonetheless, I was unable to re-test the H265 sample as the rest of the aliquot was used up in the analysis described in section 6.7.

One possible limitation of this analysis could have been the likely clonal nature of *KRAS* mutations in these patient cases. For instance, case H263 showed highly variable *KRAS* MU fractions (0.0000-6.2250%) amongst the 11 FFPE tissue samples tested, which was strongly suggestive of the *KRAS* clonality in this patient (Table 6-3). Although stromal DNA could have diluted the percentage abundance of *KRAS* MU copies in the FFPE aliquots, this alone was unlikely to be the cause for wide variations in the percentage MU abundances within samples from one patient (e.g. H263).

Overall, this analysis demonstrated that scarce copies of *KRAS* mutations could be detected in plasma samples from patients with early lesions, albeit at a low sensitivity (i.e. 6 in 34 samples tested), and successful detection was subjected to the robustness of ddPCR technology as well as the clonality of mutation targets.

6.5.4 Analysis of *KRAS* (c.38G>A) Mutation in the Plasma cfDNA Samples

For this analysis, matched plasma samples from 24 patient cases, which had been identified as positive for the *KRAS* c.38G>A/p.G13D mutation in codon 13, were tested using the *KRAS*-G13D ddPCR assay (section 5.3.2). This included samples that showed borderline MU positivity. However, sample H263 was excluded from this analysis as all available aliquot had been used up. All plasma samples were lyophilised and processed as previously described, and 3.0µl of each sample was tested per reaction. The assay was run twice.

KRAS-G13D ddPCR : 1D Amplitude Plots

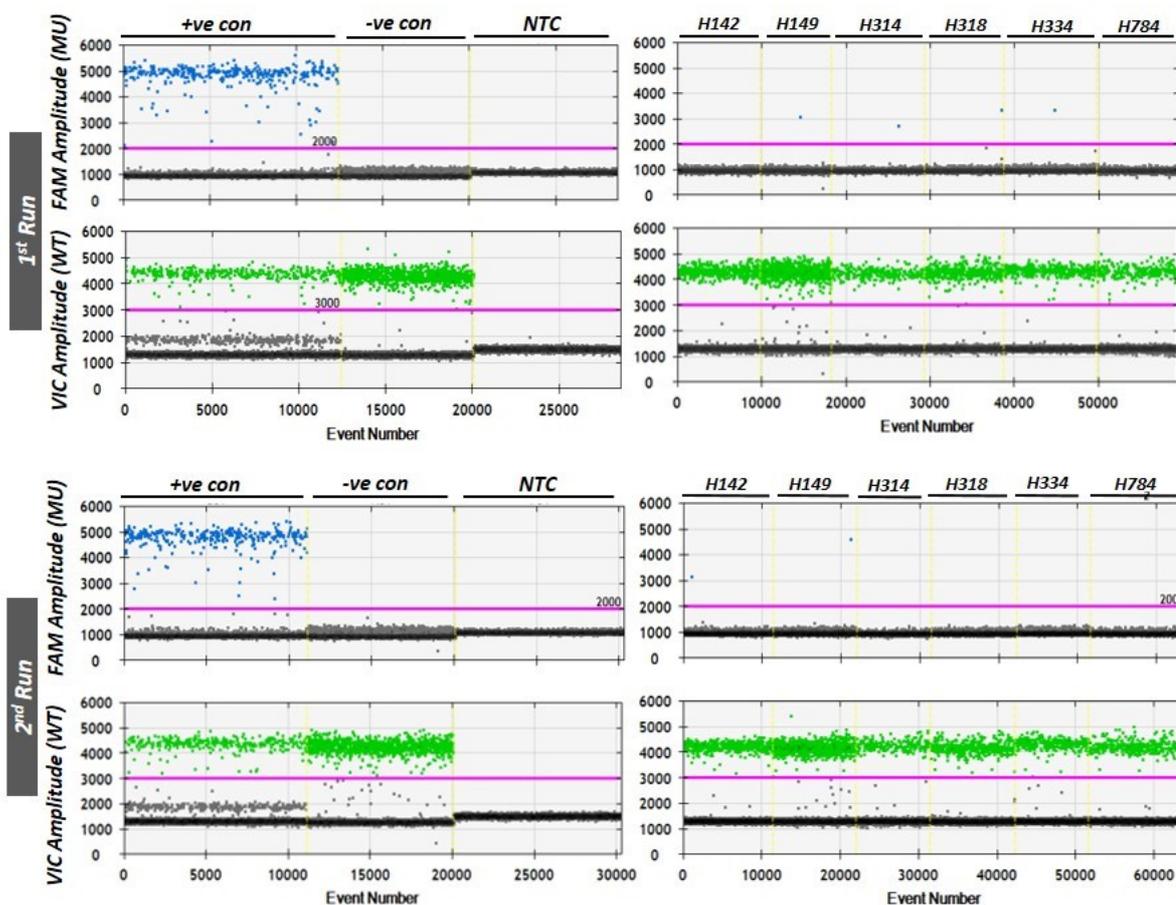
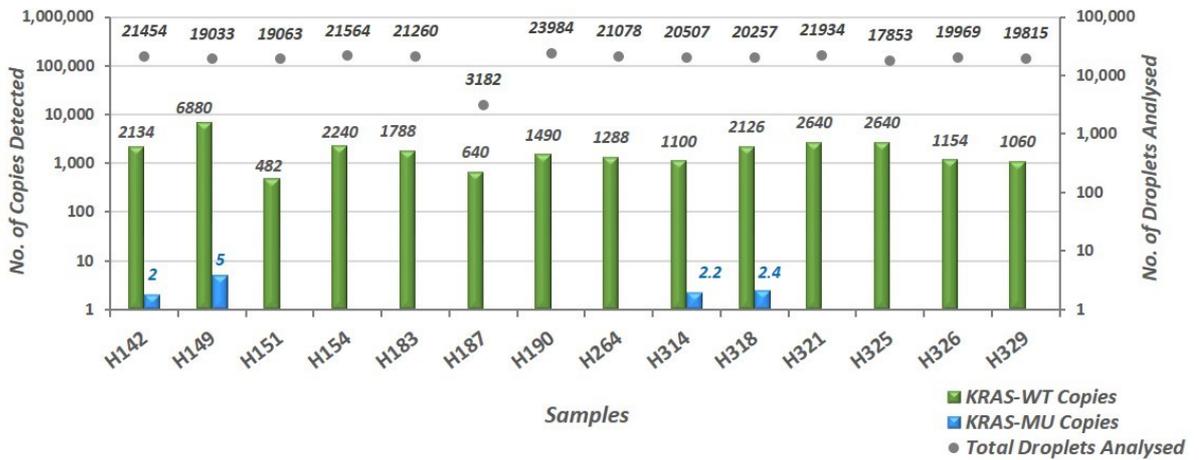


Figure 6-10: Testing for *KRAS* c.38G Mutation in the Plasma cfDNA Samples using the *KRAS*-G13D Assay – 1D Plots

These 1D ddPCR plots were accompanying figures to the graphs in Figure 6-11. Pink lines are manual threshold gating for the WT (HEX) and MU (FAM) amplification. The plots on the left were the control wells. Positive control (+ve con) was 10ng of HCT116 (c.38G>A/p.G13D) DNA. Negative control was 10ng hgDNA. NTC = no template control.

The 1D plots for the six samples (H142, H149, H314, H318, H334 and H784) that showed positive MU (FAM) amplification were shown on the right. H314, H318 and H334 were positive for MU amplification only in the 1st assay run. H142 and H784 were positive for MU amplification only in the 2nd assay run. Only H149 tested positive for the *KRAS* mutation in both assay runs.

KRAS-G13D ddPCR: Analysis of Patient Plasma cfDNA Samples (Graph 1 of 2)



KRAS-G13D ddPCR: Analysis of Patient Plasma cfDNA Samples (Graph 2 of 2)

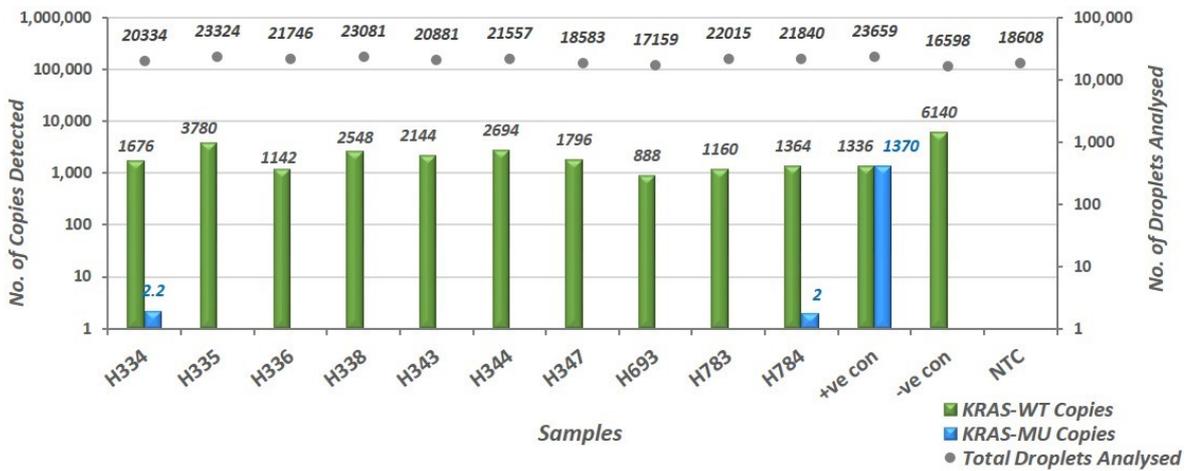


Figure 6-11: Testing for KRAS c.38G Mutation in the Plasma cfDNA Samples using the KRAS-G13D Assay – Graphs for all Samples tested

Merged data from two assay runs. WT (green bars) and MU (blue bars) copies detected were shown on left y-axis. Number of droplets analysed (grey dots) were shown on right y-axis. Controls were as described in Figure 6-10. Despite repeated runs, the numbers of droplets analysed were below the standard (>10,000 droplets/reaction) for these 8 samples: H149, H151, H187, H325, H326, H329, H347 and H187. In particular, sample H187 had very poor droplet generation.

Table 6-5: Histology Information for the Matched FFPE Cases Corresponding to Plasma cfDNA Samples Analysed in Figure 6-11

Sample ID	No. of FFPE Blocks	% MU Fractional Abundance in FFPE samples	Histology	Polyp Dimensions (mm)
H142	1	16.3347	Sigmoid (LG TVA)	17 x 10 x 15
H149	8	0.0000 – 3.8205	Sigmoid (LG TVA)	15 x 15 x 15 – 25 x 20 x 25
H151	1	5.0351	Splenic Flexure (LG TA)	5 x 4
H154	4	0.0361 – 9.4767	Descending-sigmoid (LG TA/TVA)	17 x 15 x 14
H183	1	6.2698	Mid Transverse (LG TA)	10 x 6
H187	1	0.2664	Sigmoid (LG TA)	13 x 13 x 11
H190	1	30.4348	Recto-sigmoid (LG TVA)	14
H264	7	0.0000 – 0.1928	Sigmoid (LG + HG TVA)	5 x 5 x 3 – 10 x 10 x 6
H314	1	0.0845	Distal Sigmoid (LG TA)	16 x 7 x 5
H318	1	0.2430	Sigmoid (LG TVA)	15
H321	1	0.2206	Colonic (LG + HG TVA)	15 x 7
H325	1	0.1070	Colonic (LG TVA)	20 x 14 x 8
H326	1	0.1671	Colonic (LG TA)	7 x 5 x 6
H329	2	0.0000 – 0.0482	Rectal (LG VA)	11 x 9 x 9 – 14 x 9 x 7
H334	2	0.0000 – 4.6363	Recto-Sigmoid (LG TA + TVA)	7 x 9 x 4 – 12 x 10 x 4
H335	2	0.0565 – 1.4967	Sigmoid-Caecal (LG TVA)	10 - 12
H336	2	2.3732 – 7.1057	Sigmoid (LG TVA)	14
H338	2	0.0000 – 0.3967	Sigmoid (LG TVA)	16 x 15 x 10
H343	2	0.2017 – 0.2258	Sigmoid (LG TVA)	15 x 10 x 15
H344	2	0.0364 – 0.4505	Colonic (LG TVA)	20 x 20 x 10
H347	2	0.0500 – 2.1894	Recto-sigmoid (LG TVA)	21 x 15 x 10
H693	2	0.0000 – 0.0449	Caecal (LG TA)	N/A
H783	2	0.0631 – 0.1532	Distal descending (HG TVA with focal AC)	15 x 15 x 17
H784	2	0.0000 – 0.1562	Colonic (Polyp – HG VA, AC – invasive, pT1)	N/A

For cases where more than one FFPE block were tested, lowest and highest mutation frequencies were reported.

LG = low-grade, HG = high-grade, TA = tubular adenoma, VA = villous adenoma, TVA = tubule-villous adenoma, MD = mild differentiated, AC = adenocarcinoma.

Table 6-6: Poisson Distribution Analysis for MU-positive cases from Figure 6-11

Sample	Total MU copies	Total WT copies	% MU fractional abundance	Assay no.	Copies/ 20µl Reaction	PoissonMax (no. of copies)	PoissonMin (no. of copies)
H142	2.0	2134	0.094	1 st	0.0	0.0	0.0
				2 nd	2.0	9.8	0.0
H149	5.0	6880	0.073	1 st	2.8	13.4	0.2
				2 nd	2.2	10.6	0.0
H314	2.2	1100	0.200	1 st	2.2	10.2	0.0
				2 nd	0.0	0.0	0.0
H318	2.4	2126	0.113	1 st	2.4	11.8	0.2
				2 nd	0.0	0.0	0.0
H334	2.2	1676	0.131	1 st	2.2	10.2	0.0
				2 nd	0.0	0.0	0.0
H784	2.0	1364	0.146	1 st	0.0	0.0	0.0
				2 nd	2.0	9.2	0.0

Out of 24 cases tested, MU copies were detected in 6 cases (25.0%), and only 1 (H149) of them showed positive MU copies in two independent assay runs (Figure 6-11; Table 6-6). This problem with being able to reproduce positive results independently was a major limitation for this experiment. This issue was reflected in the Poisson distribution analysis that predicted detection of 0.0-0.2 MU copies at the minimum level for all 6 samples (Table 6-6), suggesting the high possibility of false negatives. Indeed, 5 out of 6 samples failed to show consistent MU positivity in two independent assay runs. Furthermore, this difficulty with MU detection in the plasma was exacerbated by the low MU fractions in the matched tissue samples (Table 6-6). For patient cases where multiple FFPE samples were tested, the percentage MU fractions varied greatly between samples within a patient, suggesting the possible clonality of *KRAS* mutations (Table 6-6). Additionally, no correlation was observed between MU fractions quantified in the FFPE and plasma samples from a patient. For example, the FFPE sample from patient H142 showed 16.3347% MU fractions and yet the matched plasma only had 0.094% MU fraction (Table 6-5 & Table 6-6). In contrast, a higher MU fraction (0.200%) was detected in the plasma of patient H314 whose FFPE sample only has 0.0845% MU fraction (Table 6-5 & Table 6-6). However, it should be noted that the FFPE analysis was likely affected by stromal DNA contribution, and therefore might not fully reflect the extent of *KRAS* MU fractions in the lesions. Nevertheless, results obtained from multiple tissue analysis were strongly suggestive of heterogeneity: good examples were H334 and H336, where

KRAS MU positivity was identified in only one out of two tissue samples tested for each patient case (Table 6-5). In this case, we could then expect that only a minor portion of the whole neoplastic tissue to contribute to the plasma cfDNA pool, making the detection of tumour-derived cfDNA in the plasma even more arduous.

Moreover, there was a frequent technical issue with droplet generation in the aforementioned experiments. For example, a mere total of 3,182 droplets were analysed in two combined reactions for sample H187, which was below the acceptable threshold (Figure 6-11). In addition, 7 other samples (H149, H151, H187, H325, H326, H329 and H347) also had droplet counts below the standard (<10,000/reaction). One possible cause might lie in the technical problem with the QX200 Droplet Generator itself. Other possibilities were that the samples contained contaminants that caused blockage of cartridge capillaries during droplet generation, or random sample wells were affected by air bubbles interfering with droplet generation despite the care taken during pipetting to prevent generation of bubbles. Generally, this stochastic failure in droplet generation was a technical limitation for these experiments.

Independently of the assays used or samples analysed, any detection of MU cfDNA in the plasma approached the LoD on the ddPCR system, which offered the lowest threshold for mutation detection among the currently available methods (Figure 1-18). This clearly raised a concern whether this approach has potential for clinical implementation, as it will be difficult to identify MU positive patients with certainty and might also lead to a substantial number of false negative results. Therefore, although these results were a proof-of-concept that tumour-derived cfDNA can be identified in the plasma of adenoma-bearing patients, the data so far suggested that further investigation is necessary to ascertain the clinical applicability of these PCR-based methods for detection of precancerous lesions.

6.6 Multi-regional Targeted Sequencing of Patient Samples

Experiments with the *BRAF* and *KRAS* mutation detection assays had shown low sensitivities with detecting tumour-derived alterations in the plasma. Data from FFPE tissue analysis had suggested that *BRAF* and *KRAS* mutations were clonal, rather than

truncal, in most of the samples, and this was the intrinsic sample limitation that contributed to the difficulty of mutation detection in the plasma. Therefore, to assess the heterogeneity of early lesions and to identify mutation targets that were present at higher frequencies, six patient cases (H149, H154, H263, H264, H265 and H266) were selected for multi-regional targeted sequencing in collaboration with the Wellcome Trust Sanger Institute, Cambridge, UK. These cases were chosen as they harboured large adenomas that had been split into multiple FFPE blocks, enabling different polyp regions to be sequenced. Sequencing was carried out using an in-house and proprietary custom targeted panel, which targets the most common alterations in CRC including 116 point mutations, 22 gene amplification/deletion, 23 recurrent focal amplification/deletion, 121 MSI regions and 2 gene fusions (see section 2.2.4 & appendix 8.2 for the details).

Prior to sequencing, the chosen FFPE blocks were analysed by a cancer pathologist (Dr Kevin West, Leicester Royal infirmary, UK) to identify the tumour area for macroscopic enrichment (Table 6-7). All six adenoma cases originated in the left-sided colon. DNA aliquots were quantified using the hALU69 assay (section 2.2.3.3.4) and approximately 500ng of purified DNA per sample was shipped for sequencing. The sequencing data was then used to design the patient-specific mutation detection assays described in section 5.4. Assay targets were selected from these three patient cases (H263, H265 & H266), whose sequencing data were reported as followed.

Table 6-7: Histology Details on the Sequenced Patient Cases

<i>Case No.</i>	<i>FFPE Blocks Tested</i>	<i>% Tumour Area</i>	<i>Histology</i>	<i>Details of Lesions</i>
H149	5, 6, 7, 8, 11, 12	50-100	Sigmoid, low-grade, TVA	Polyp 1 (FFPE No. 5, 6, 7, 8) – 25 x 20 x 25 Polyp 2 (FFPE No. 11, 12) – 15 x 15 x 15
H154	3, 4	80	Descending colon, low-grade, TVA	17 x 15 x 14
H263	3, 4, 6, 8, 10-1	100	Rectum, low & high-grade TVA with focal AC	3 large polypoid pieces: Piece 1 (FFPE No. 3, 4, 6) – 30 x 25 x 22 Piece 2 (FFPE No. 8) – 17 x 30 x 10 Piece 3 (FFPE No. 10-1) - 30 x 15 x 10
H264	1, 2, 5, 8	80-100	Sigmoid, low & high-grade TVA	5 x 5 x 3 – 10 x 10 x 16
H265	2, 3, 4-1, 5-1, 5-2, 6-1, 7-1	100	Distal sigmoid, low & high-grade VA	20 x 35 x 55 (multiple fragments from one large polypoid)
H266	1, 2, 3, 5, 7-1, 7-2	100	Sigmoid, low-grade VA	60 (no further measurements reported, multiple polypoid tissues)

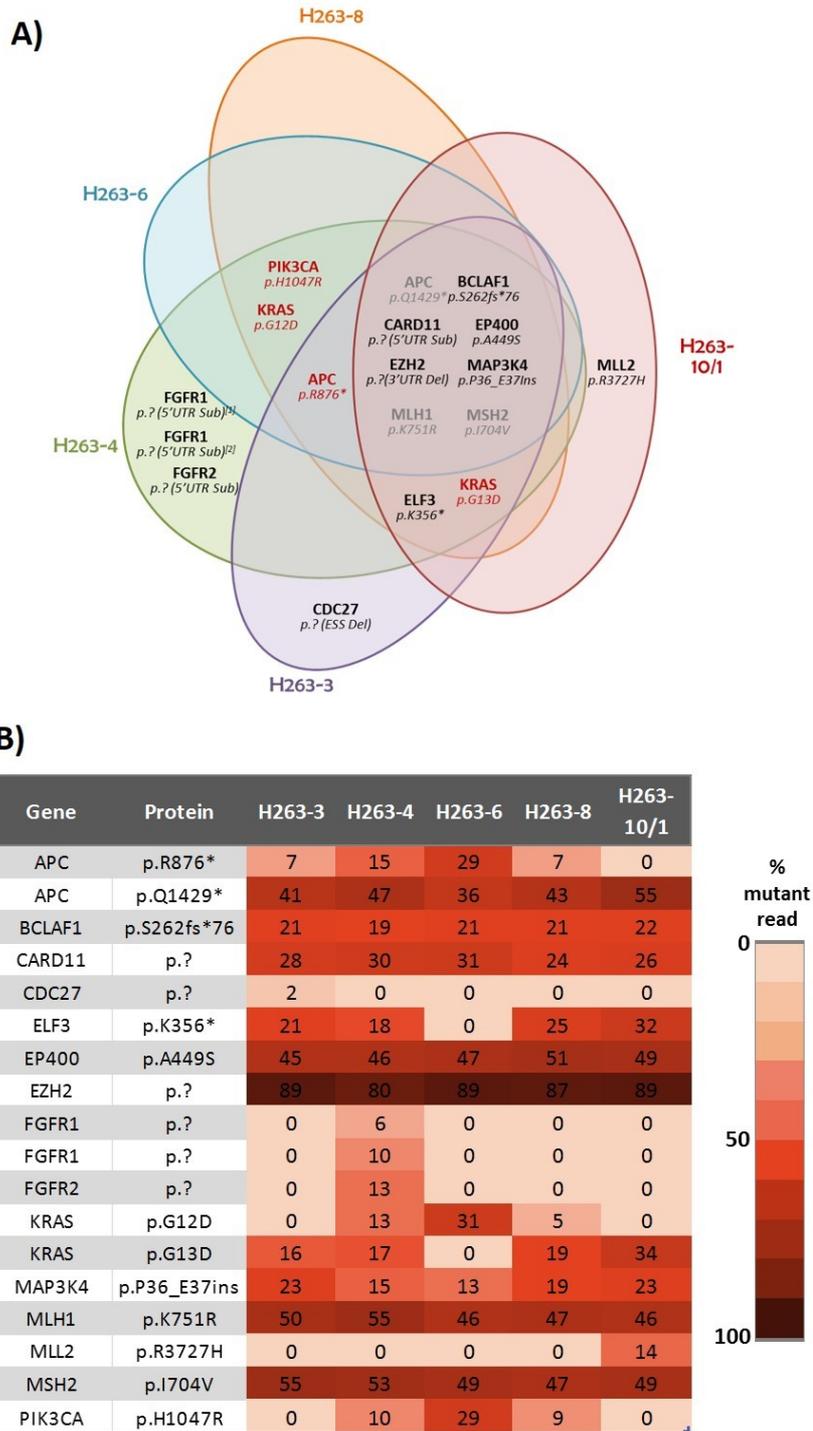


Figure 6-12: Summary of Multiregional Sequencing Data for Case H263

The patient had a rectal TVA with both low- and high-grade regions, as well as a focal invasive adenocarcinoma in the submucosa. Five FFPE samples that originated from three polypoid pieces (Table 6-7) were sequenced. Sequencing details were described in section 2.2.4.

A) Venn diagram illustrating the distributions of mutations in the different sequenced regions. Each coloured oval represents one sequenced region. Mutations in red are driver mutations as determined by the in-house calling algorithm.

B) A heat map showing the percentage of mutant reads for each mutation.

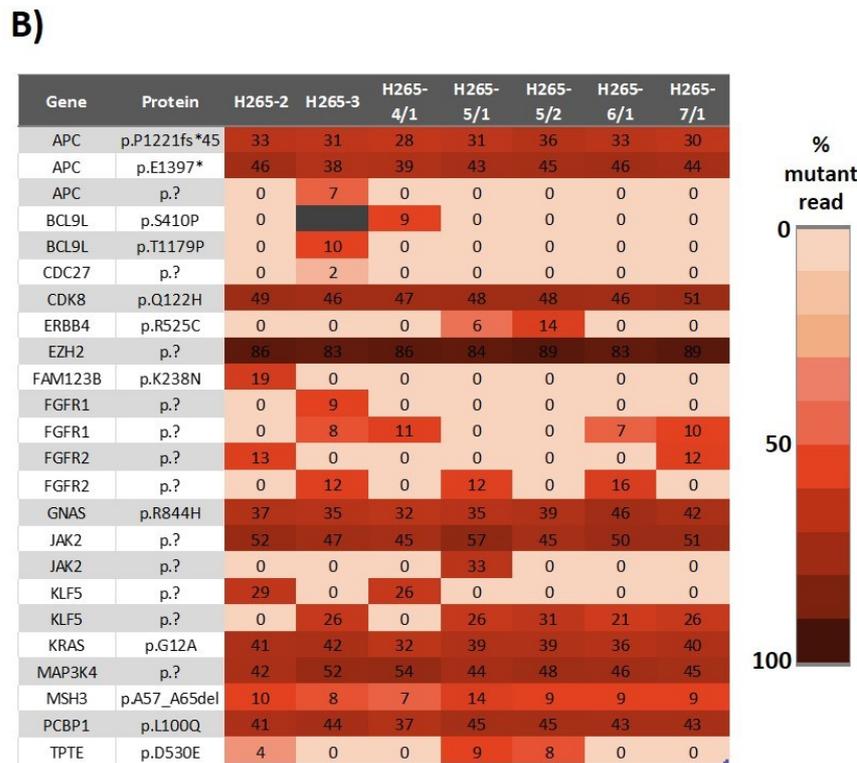
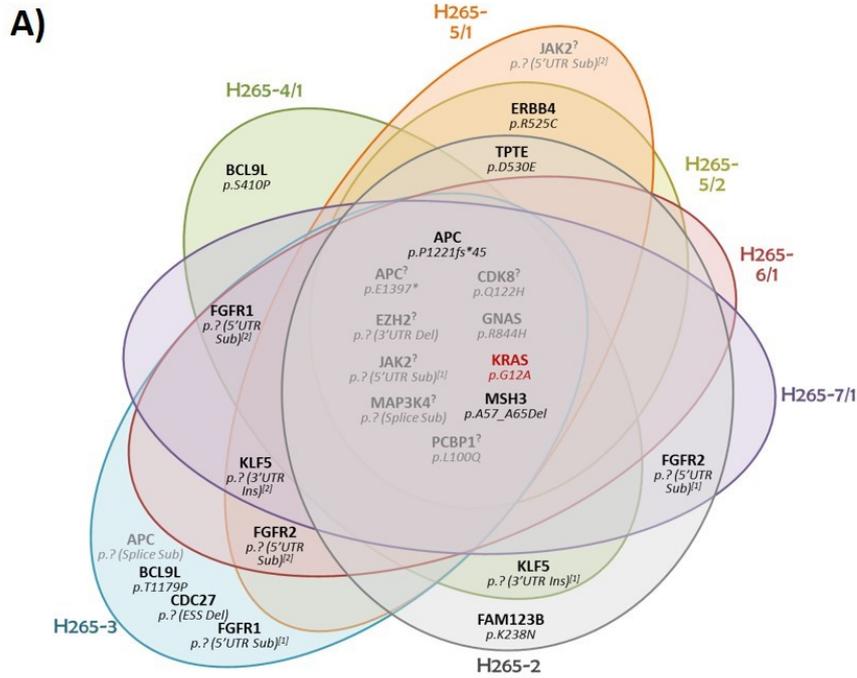


Figure 6-13: Summary of Multiregional Sequencing Data for Case H265

This patient had low- and high-grade VA in the distal sigmoid. All 7 FFPE regions were from one large polyp (Table 6-7). Sequencing details were described in section 2.2.4.

A) Venn diagram illustrating the distribution of mutations between the samples. Each coloured oval represents one FFPE region. Mutations in red are driver mutations as determined by the in-house calling algorithm.

B) A heat map showing the percentage of mutant reads for each mutation.

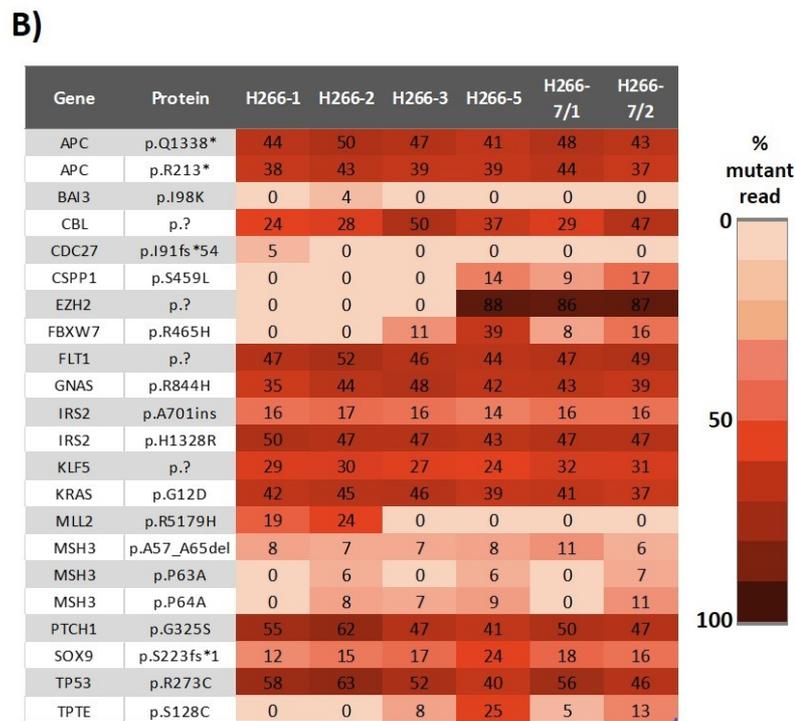
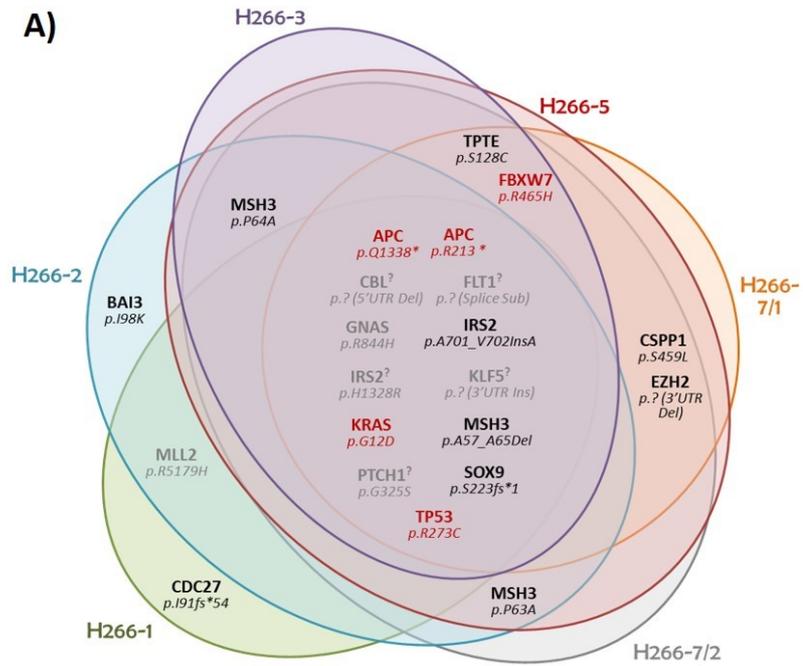


Figure 6-14: Summary of Multiregional Sequencing Data for Case H266

This patient had a low-grade VA in the sigmoid, which was excised as multiple polypoid fragments. All 6 FFPE regions originated from one lesion (Table 6-7) were sequenced. Sequencing details were described in section 2.2.4.

A) Venn diagram illustrating the distributions of mutations between the sequenced FFPE regions. Each coloured oval represents one FFPE region. Mutations in red are driver mutations as determined by the in-house calling algorithm.

B) A heat map showing the percentage of mutant reads for each mutation.

The sequencing results for the cases H263, H265 and H266 were reported in Figure 6-12, Figure 6-13 and Figure 6-14, respectively. The other three cases (H149, H154 and H264) were reported in appendix 8.3. Based on the sequencing data for patient H263 (Figure 6-12), two mutation targets were chosen: *APC* (c.4285C>T/p.Q1429*) and *PIK3CA* (c.3140A>G/p.H1047R). The *APC* mutation at codon 1429 – a non-sense substitution, resulting in a STOP codon (denoted as *) – was identified at high frequencies in all 5 sequenced samples, suggesting that this was a truncal mutation (Figure 6-12: A & B). The *PIK3CA* mutation at c.3140A>G/p.H1047R was detected in 3 out of 5 FFPE samples with frequencies ranging from 9-29% (Figure 6-12-B); this target was chosen based on its implication in both initiation and progression of CRC [387]. From the sequencing data of H265 (Figure 6-13), another *APC* mutation (c.4189G>T/p.E1397*) was chosen as the target: this mutation appeared at high frequencies (38-46%) in all 7 samples tested and it was confirmed as a somatic alteration in CRCs [56]. The remaining two targets were chosen from the sequencing data of case H266: *APC* (c.4012C>T/p.Q1338*) and *TP53* (c.817C>T/p.R273C). Both targets were detected at high MU frequencies (41-50% for *APC*^{Q1338*} and 40-63% for *TP53*^{R273C}) all 6 sequenced regions (Figure 6-14-B). The reason for primarily choosing *APC* mutation as targets was because of their strong implications as common drivers of colorectal tumourigenesis. Whilst the sequencing data for these six cases also identified several other mutation targets, such as *GNAS* in H265 (Figure 6-13), the custom assay development was limited to five targets as not to exhaust the resources available. The validation of these assays has been described in Chapter 5.

A) Patient Case H263

FFPE Block No.	% MU Abundance (KRAS 35G)		% MU Abundance (KRAS 38G)	
	qPCR	NGS	ddPCR	NGS
2	1.28	N/A	20.49	N/A
3	0.02	0	16.71	16
4	1.82	13	22.02	17
5	3.06	N/A	11.99	N/A
6	4.89	31	3.71	0
7	6.23	N/A	0.55	N/A
8	0.01	5	18.88	19
9	0.00	N/A	23.21	N/A
10-1	0.00	0	28.71	34
10-2	0.00	N/A	26.37	N/A
11	0.00	N/A	39.11	N/A

B) Patient Case H265

FFPE Block No.	% MU Abundance (KRAS 35G)	
	qPCR	NGS
2	26.09	41
3	29.39	42
4-1	26.85	32
4-2	28.96	N/A
5-1	26.12	39
5-2	25.47	39
5-3	21.99	N/A
6-1	32.87	36
6-2	25.77	N/A
7-1	19.57	40
7-2	37.43	N/A
8	23.93	N/A
9	22.60	N/A

C) Patient Case H266

FFPE Block No.	% MU Abundance (KRAS 35G)	
	qPCR	NGS
1	20.13	42
2	19.50	45
3	22.25	46
4	12.97	N/A
5	26.74	39
6	25.80	N/A
7-1	21.62	41
7-2	20.22	37
7-3	21.64	N/A

Figure 6-15: Comparing the PCR & NGS Data on the Percentage MU Abundances in Tissue Samples

The *KRAS* c.35G data was obtained from the *KRAS*-122 qPCR analysis (Table 6-3), and the *KRAS* c.38G data was obtained from the *KRAS*-G13D ddPCR analysis (Table 6-5). NGS analysis was as described in section 2.2.4. FFPE samples excluded from NGS analysis were denoted as N/A.

A) Patient case H263 was identified with double *KRAS* mutations at based c.35G and c.38G. Both the PCR and NGS data indicated that these were clonal mutations.

B) Patient case H265 was positive for the *KRAS* c.35G mutation. Both the PCR and NGS data indicated that this was a truncal alteration as it was present in all the samples.

C) Patient case H266 was also positive for the *KRAS* c.35G mutation. Both PCR and NGS data showed that this was a truncal mutation as it was present in all the samples.

Furthermore, as previously mentioned, the disparity in the *KRAS* MU fractions between multiple FFPE samples from individual patients suggested that these mutations might be clonal (Table 6-2, Table 6-3 and Table 6-5). Indeed, at least for case H263, the sequencing data confirmed this deduction and showed evidence of ITH in precancerous lesions (Figure 6-12). The patient case H263 contained two *KRAS* mutations: 3 out of 5 sequenced regions were positive for the c.35G>A/G12D mutation at 5-31% MU frequencies, and 4 out of 5 regions were positive for the c.38G>A/p.G13D mutation at 16-34% MU frequencies (Figure 6-12-B). This data was confirmed by the previous PCR-based analysis, which identified that same *KRAS* mutations, but at much lower MU frequencies, which was likely caused by stromal DNA dilution (Figure 6-15). Interestingly, when the percentage MU data obtained from PCR-based analyses and NGS were compared, it appeared that qPCR analysis was less accurate at calling MU copies, and consistently called lower percentages, compared to the NGS data (Figure 6-15), which might be explained by the lower qPCR efficiency as a result of using degenerate probes. In contrast, the *KRAS*-G13D ddPCR analysis showed comparable, or slightly better, accuracy than the NGS data; in fact, for FFPE block no. 6 from case H263, ddPCR analysis identified *KRAS* c.38G mutation at 3.71% abundance whereas the NGS failed to call any MU positivity (Figure 6-15-A), indicating that the limit of sensitivity was higher in NGS.

On the other hand, the *KRAS* mutation at c.35G>C/p.G12A was present in all 7 regions tested for patient H265, suggesting that this might be a truncal variant (Figure 6-13). This supported the *KRAS*-122 qPCR analysis data where the same mutation was detected at substantial fractions in the FFPE samples (Table 6-3). Unfortunately, despite the truncal nature of this mutation, no MU copies were detected in the matched plasma due to very poor droplet generation (Figure 6-9). Also for the case H266, the sequencing data showed high mutant reads (37-46%) for the *KRAS* c.35G>A/p.G12D mutation in all 6 FFPE samples tested (Figure 6-14), and this confirmed the high MU fractions (~20%; Table 6-3) obtained from the *KRAS*-122 qPCR analysis of the same FFPE samples. Furthermore, the matched plasma sample from this patient did test positive for the MU cfDNA, albeit very few copies were detected in the plasma (Table 6-4).

Overall, the NGS analysis provided the mutation profiles of these patient cases and guided the development of five custom assays described in section 5.4, which were

used to test the matched plasma samples to detect tumour-derived mutations in the plasma.

6.7 Plasma Analysis with Patient-Specific Mutation Detection Assays

Prior to testing, all samples were lyophilised (section 2.2.2.3.1) and 3.0µl sample was tested per reaction. Unfortunately, duplicate runs were not possible due to limited availability of plasma samples as these were also used in the previous *BRAF* and *KRAS* mutation analysis. Cases H263 and H266 harboured two mutation targets each (H263 - *APC*^{Q1429*} and *PIK3CA*^{H1047R}; H266 - *APC*^{Q1338*} and *TP53*^{R273C}), which further exhausted the availability of sample aliquots. For these experiments, the leukocytic DNA extracted from the matched buffy coat layers was also analysed to eliminate the possibility of germline polymorphic variants. Experimental conditions were described in section 2.2.3.4.4 and all data were analysed using the QuantaSoft software (version 1.6.6.0320, BioRad®).

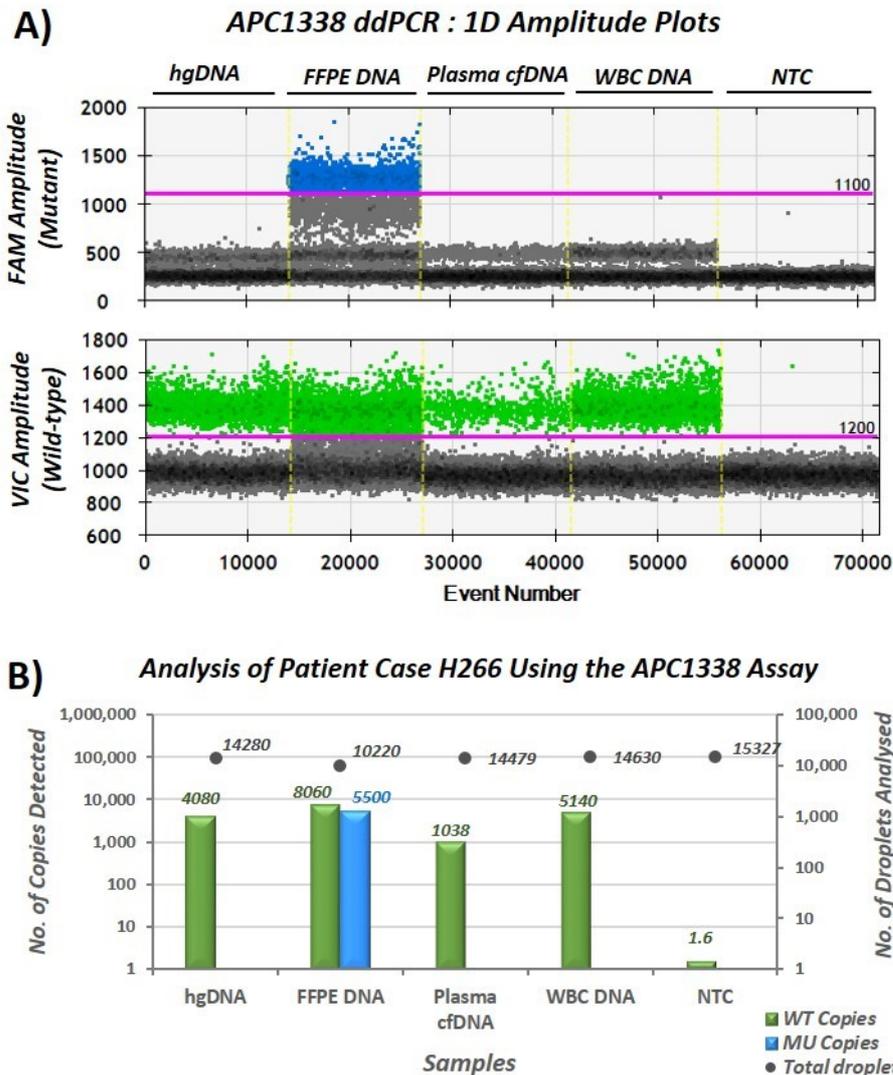


Figure 6-16: Analysis of Patient Case H266 using the APC1338 Assay

The APC1338 ddPCR assay targets the c.4012C>T/p.Q1338* mutation. 10ng hgDNA was used as a negative control. 10ng of FFPE DNA from H266 was used as a positive control. 10ng of WBC DNA was included as a control for polymorphic variants. NTC = no template control.

A) 1D amplitude plots showing the manual threshold gating (pink line) for VIC (green) and FAM (blue) amplitudes.

B) Graph plotting the number of WT copies (green bars) and MU copies (blue bars) detected on the left y-axis. Number of droplets analysed (grey dots) were shown on the right y-axis.

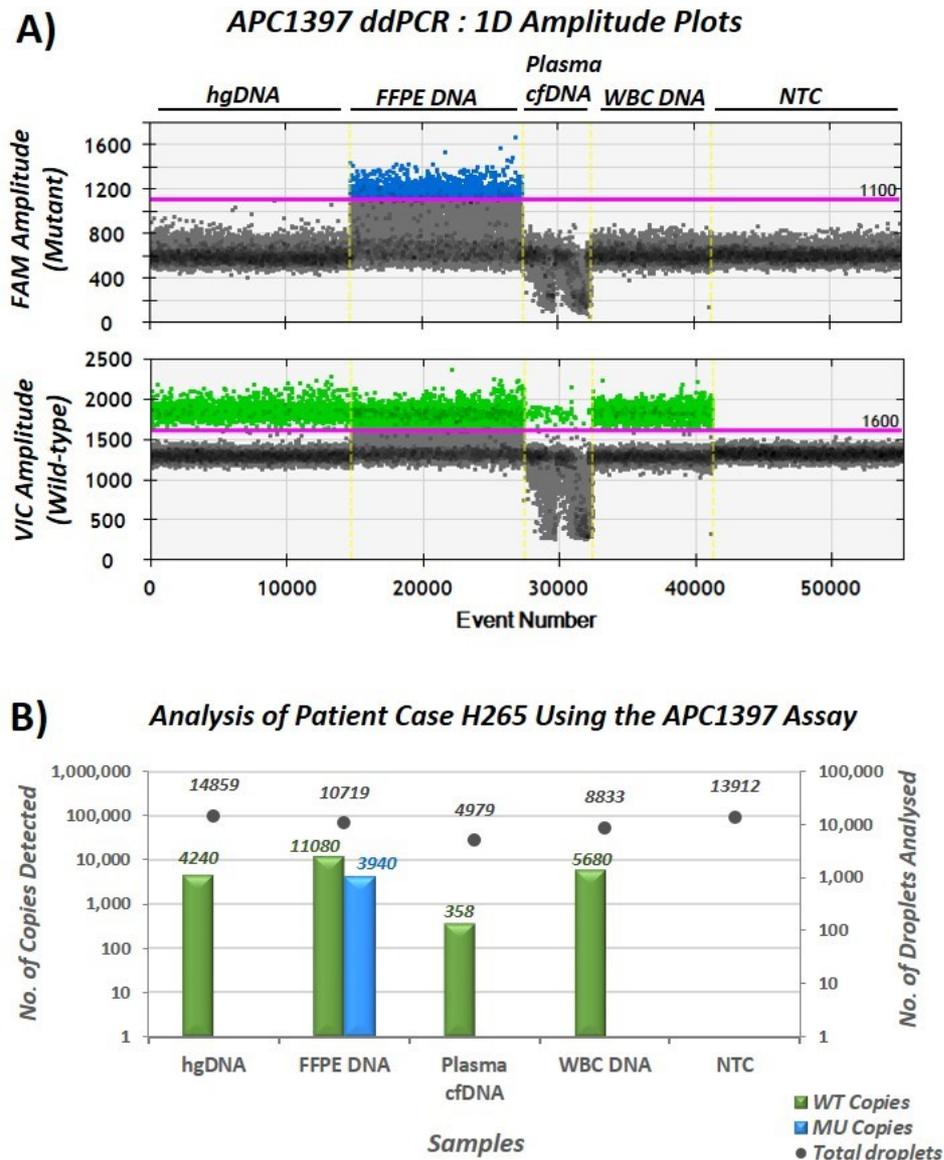


Figure 6-17: Analysis of Patient Case H265 using the APC1397 Assay

The APC1397 ddPCR assay targets the c.4189G>T/p.E1397* mutation. 10ng hgDNA was used as a negative control. 10ng of FFPE DNA from H265 was used as a positive control. 10ng of WBC DNA was included as a control for polymorphic variants. NTC = no template control.

A) 1D amplitude plots showing the manual threshold gating (pink line) for VIC (green) and FAM (blue) amplitudes. There was an issue with droplet generation in the plasma cfDNA well as seen by the reduction in background fluorescence (grey dots), indicating these droplets contained less reaction mixture.

B) Graph plotting the number of WT copies (green bars) and MU copies (blue bars) detected on the left y-axis. Number of droplets analysed (grey dots) were shown on the right y-axis. Problem with droplet generation for the plasma cfDNA well resulted in only 4,979 droplets analysed.

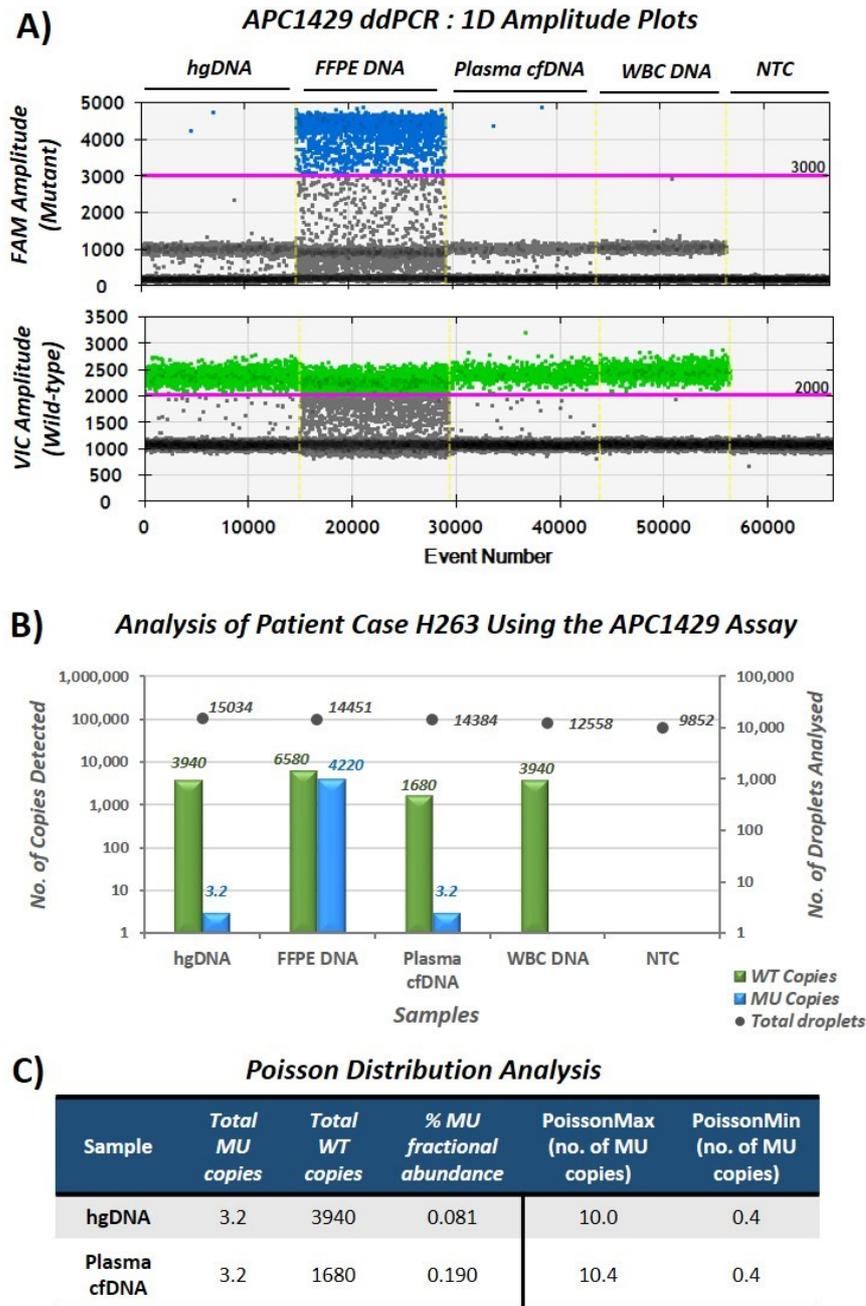


Figure 6-18: Analysis of Patient Case H263 using the APC1429 Assay

The APC1429 ddPCR assay targets the c.4285C>T/p.Q1429* mutation. 10ng hgDNA was used as a negative control. 10ng of FFPE DNA from H263 was used as a positive control. 10ng of WBC DNA was included as a control for polymorphic variants. NTC = no template control.

A) 1D amplitude plots showing the manual threshold gating (pink line) for VIC (green) and FAM (blue) amplitudes.

B) Graph plotting the number of WT copies (green bars) and MU copies (blue bars) detected on the left y-axis. Number of droplets analysed (grey dots) were shown on the right y-axis.

C) Poisson distribution analysis for MU detection.

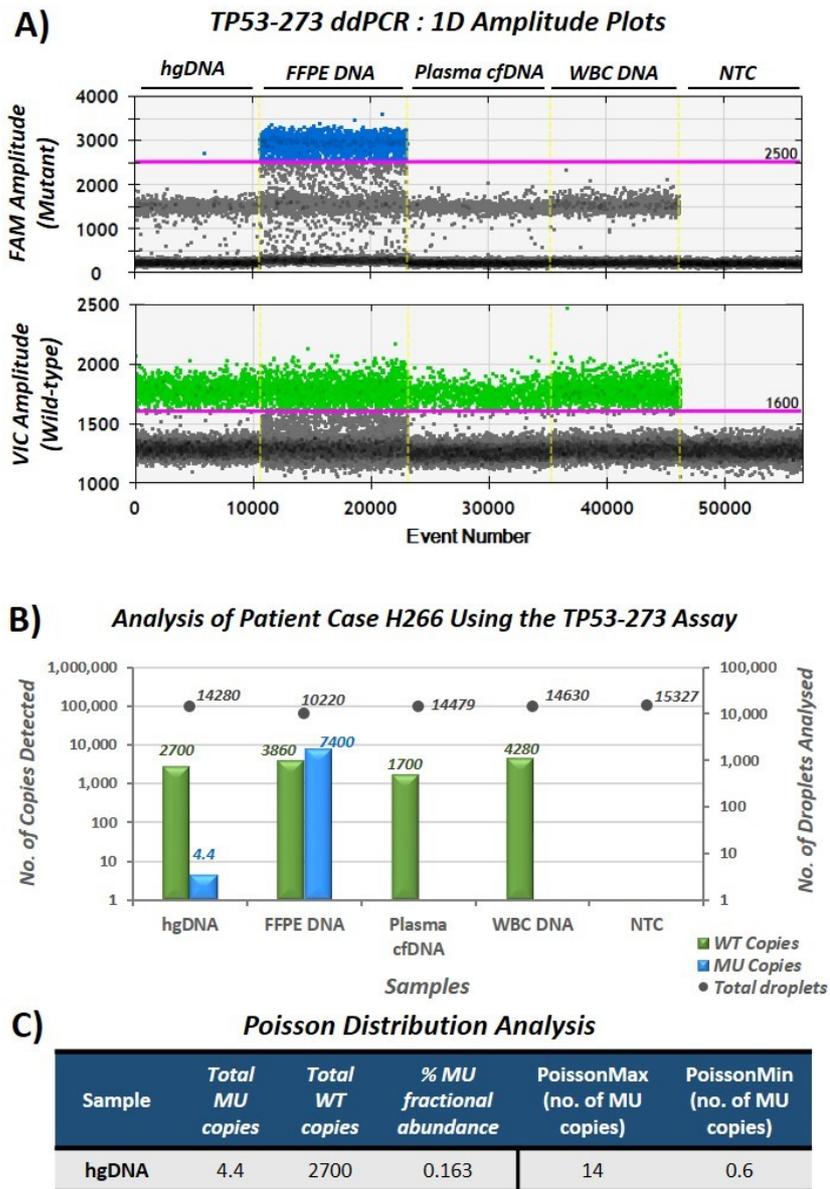


Figure 6-19: Analysis of Patient Case H266 using the TP53-273 Assay

The TP53-273 assay targets the c.817C>T/p.R273C mutation. 10ng hgDNA was used as a negative control. 10ng FFPE DNA from H266, used for sequencing, was used as a positive control. 10ng WBC DNA was included as a control for polymorphic variants. NTC = no template control.

A) 1D amplitude plots showing the manual threshold gating (pink line) for VIC (green) and FAM (blue) amplitudes.

B) Graph plotting the number of WT copies (green bars) and MU copies (blue bars) detected on the left y-axis. Number of droplets analysed (grey dots) were shown on the right y-axis.

C) Poisson distribution analysis for MU detection.

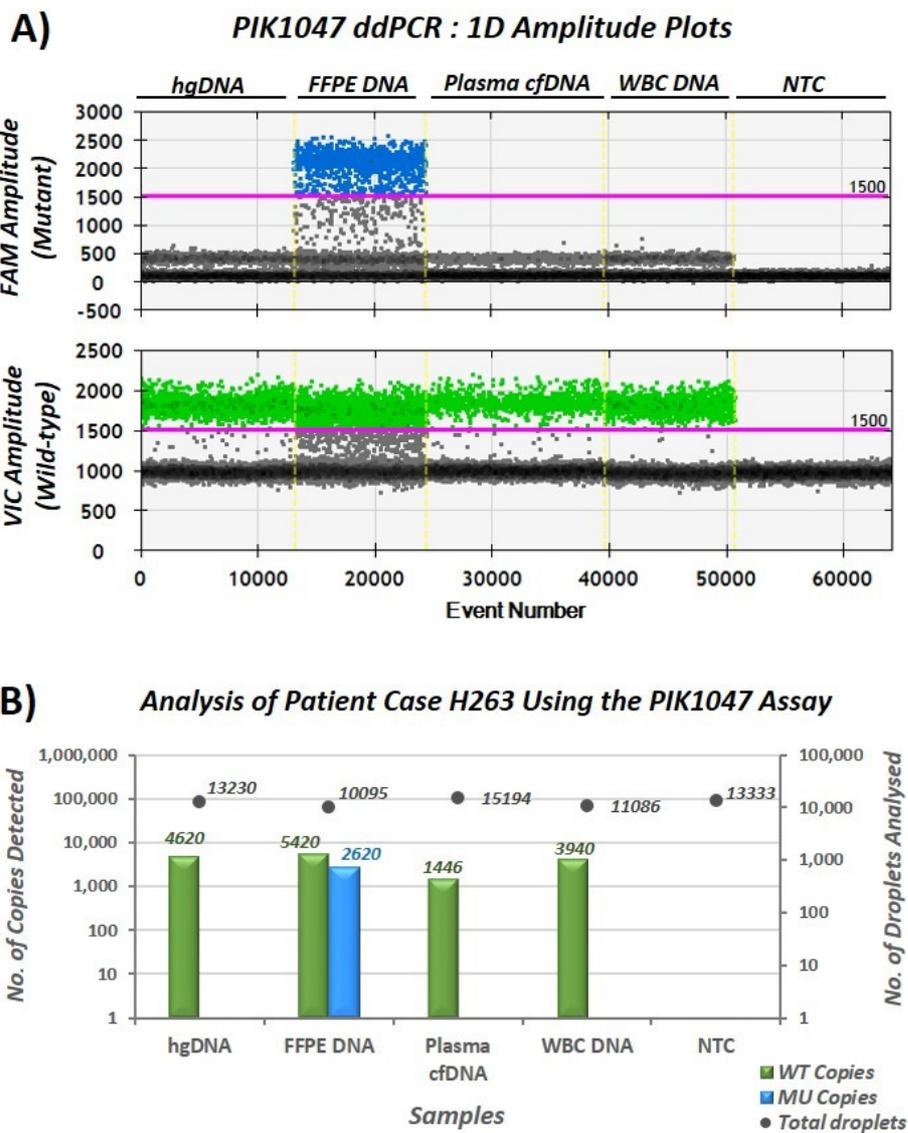


Figure 6-20: Analysis of Patient Case H263 using the PIK1047 Assay

The PIK1047 assay targets the c.3140A>G/p.H1047R mutation. 10ng hgDNA was used as a negative control. 10ng FFPE DNA from H263, used for sequencing, was used as a positive control. 10ng WBC DNA was included as a control for polymorphic variants. NTC = no template control.

A) 1D amplitude plots showing the manual threshold gating (pink line) for VIC (green) and FAM (blue) amplitudes.

B) Graph plotting the number of WT copies (green bars) and MU copies (blue bars) detected on the left y-axis. Number of droplets analysed (grey dots) were shown on the right y-axis.

Out of the 5 custom assays, only the APC1429 assay proved successful for mutation detection in the plasma sample from patient H263 – 3.2 MU copies were detected in the plasma in a background of 1680 WT copies (0.19% MU fractional abundance; Figure 6-18 Figure 6-16-B). The WBC DNA reaction well was negative for MU copies (Figure 6-18-B), indicating any MU cfDNA copies detected were genuine somatic alterations. However, Poisson analysis predicted a minimum threshold of 0.4 MU copies for the plasma cfDNA reaction at the low error bar of droplet distribution, suggesting the possibility of false negatives in independent runs. Additionally, a few MU copies were detected in the hgDNA (-ve con) control, likely caused by cross-contamination with the positive control well or they were background noise (Figure 6-18-B).

Unfortunately, the other 4 assays failed to detect any MU copy in the plasma. Nevertheless, these assays confirmed that all mutations were cancer-related somatic variants, as WBC DNA controls showed no alterations of the corresponding genes (Figure 6-16, Figure 6-17, Figure 6-19 and Figure 6-20). Of interest, for the TP53-273 assay run, few *TP53* MU copies were detected in the hgDNA (-ve con) well, which could likely have been caused by cross-contamination from the positive control (Figure 6-19). However, an unexpectedly high background of *TP53* mutations at codon 175 has been previously reported by Diehn's group in their healthy plasma controls [301], leading to the speculation that an unusual *TP53* MU baseline might also explain the MU positivity in my negative control well. Moreover, there was an issue with the H265 plasma samples where very few droplets (<5,000) were generated; this problem was already noted in the previous *KRAS*-122 ddPCR analysis (Figure 6-9). Sub-optimal droplet generation (<10,000) did occur randomly during ddPCR experiments, however this problem with H265 plasma sample seemed to lie with the aliquot itself. Although the reason remained unclear, one possibility was that there was a problem with cfDNA extraction that resulted in the elution of contaminants, which might have interfered with droplet generation.

Overall, despite assay optimisation and targeting high frequency mutations, positive MU detection occurred in only 1 out of 5 assays tested. Sample limitation and poor droplet generation also partly contributed to the difficulty with the plasma mutation detection analysis. Nevertheless, within the parameters of this project, these experiments have suggested that the use of plasma cfDNA for identification of tumour-

specific mutations is not currently feasible for early stages of CRC, as any positive copies were likely to be detected at the detection threshold with a high probability of false negatives.

6.8 Summary

This chapter detailed the results obtained from investigation into the potential of using plasma cfDNA as a biomarker for the detection of precancerous colorectal lesions in patients. To this end, our objectives were investigated in a cohort of 76 adenoma patients (section 2.1.7).

Firstly, FFPE samples from the 76 patient cases were screened for the *BRAF* and *KRAS* hotspot mutations using these four PCR-based assays: *BRAF*-V600E, *KRAS*-121, *KRAS*-122 and *KRAS*-G13D. This analysis identified 44 out of 76 (57.9%) cases as having at least one *KRAS* mutation, either in codon 12 or 13, with 14 of them positive for two simultaneous *KRAS* mutations (Figure 6-4). Additionally, this screening identified 4 (5.26%) *BRAF*-MU patient cases (Figure 6-4). These mutation-positive patient cases were then selected for the subsequent analysis of matched plasma samples.

Several studies have reported a correlation between increased levels of cfDNA and the presence of early stage disease in patients [299, 319], and to test this hypothesis, quantitative analysis of total plasma cfDNA levels was performed in the patient and 'polyp-free' control cohorts using the hGAPDH assay (section 2.1.7). The results showed that total plasma cfDNA levels were comparable between my patient and control groups: 6.106 ± 0.5167 ng/ml vs. 7.456 ± 0.642 ng/ml, respectively ($p=0.1057$; Figure 6-5). Although the mean plasma cfDNA concentration in the control group was comparable to what have been reported in the literature, my patient group showed markedly lower plasma cfDNA levels, on average, when compared to other groups [299, 319]. One possible explanation of this difference might lie in our study cohorts: both Cree's and Watanabe's group had included larger and advanced lesions, including Stage I/II cancers, in their patient group compared to the smaller and less advanced lesions that made up my patient cohort (Figure 6-1). It was also possible that technical differences in sample collection, storage, processing and analysis might be responsible

for the different observations. However, the data presented here echoed what had been reported by Pilotti's group where they also failed to observe discernible differences in total plasma cfDNA levels between patients with premalignant lesions and healthy controls [386]. Nevertheless, the results presented here suggested that total cfDNA quantification, at least for early stage cancer patients, is an unreliable method to identify individuals with early colorectal lesions.

Following this, experiments were performed to investigate if mutations in *KRAS* and *BRAF* oncogenes, which have been shown to be readily detectable in plasma of CRC patients [255, 260, 312, 337, 388], could also be identified in the plasma of individuals with benign lesions. Towards this end, matched plasma samples were compared to the FFPE cases that were positively identified with at least one *BRAF* or *KRAS* mutation (Figure 6-4). Using custom ddPCR assays, this analysis resulted in a marginal but successful identification of tumour-derived cfDNA in the plasma samples, but this was limited to *KRAS* mutations: 2 out of 13 samples (15.4%) were tested positive for *KRAS* c.34G mutations (Figure 6-7), 6 out of 34 samples (17.6%) were positive for *KRAS* c.35G mutations (Figure 6-9) and 6 out of 24 samples (25%) had few plasma copies of *KRAS* c.38G>A mutation (Figure 6-11). Plasma analysis for the *BRAF* mutation failed to produce any positive detection in all four samples analysed (Figure 6-6). These experiments highlighted the issue with low sensitivities in detecting mutations in the plasma of adenoma patients, and this issue has also been reported in the literature [312, 386]. However, one limitation of this analysis could lie in the low MU fractions of these *KRAS* and *BRAF* mutations in the tissue samples, suggesting that these alterations might be clonal for most patient cases (Table 6-2, Table 6-3 and Table 6-5), which would exacerbate the difficulty in detecting these mutations in the plasma due to their lesser presentation in the plasma cfDNA pool.

Therefore, for the next stage of analysis, six patient cases (H149, H154, H263, H264, H265 and H266) with large adenomas were selected for multiregional targeted sequencing (section 6.6, appendices 8.2 & 8.3) to assess heterogeneity in early lesions and to identify mutations that occur at high frequencies as potential assay targets. The resulting NGS data did confirm the presence of ITH in early stages of colorectal tumorigenesis, in agreement with what has been reported in the literature [389, 390]. Moreover, from the sequencing data of three patient cases (H263, H265 and H266), five

mutation targets – 3 *APC* mutations, 1 *TP53* mutation and 1 *PIK3CA* mutation – were selected for patient-specific mutation detection assays (section 5.4) to analyse the matched plasma samples (section 6.7). Disappointingly, only 1 out of 5 assays resulted in positive MU detection in the plasma. This positive result was obtained using the APC1429 assay where 3.2 MU copies (0.190% fractional abundance) were detected in the plasma from patient H263 (Figure 6-16-B). Due to limited sample availability, I was unable to repeat the test for independent confirmation. While performing these experiments, some reaction wells failed to reach the droplet generation threshold (>10,000/reaction), and this was a persistent issue especially with the H265 plasma sample where low droplet counts were observed with both KRAS-122 (Figure 6-9) and APC1397 assays (Figure 6-20-B). This was either an intrinsic problem with the sample quality or an extrinsic technical issue. Nonetheless, this problem with droplet generation was a technical limitation that likely contributed to the difficulty with mutation detection in these plasma samples.

Overall, the data presented in this chapter had shown that, using PCR-based methods, it was possible to detect tumour-derived cfDNA in the plasma from patients with early lesions. However, even using the ddPCR system, which currently offers the best resolution for targeted mutation detection (Figure 1-18), all experiments showed low sensitivities for tumour-derived cfDNA detection in the plasma. For most samples, MU detection occurred at the limit of the ddPCR technique, with a high probability of false negative results. Whilst it might be possible to increase the detection capability by performing multi-regional sequencing of early lesions prior to targeted mutation detection in the matching plasma samples, this would undoubtedly be cost- and labour-intensive, which made this process impractical for clinical implementation. Based on the data collected so far, advancements in methodologies and sample processing workflows might be the key to improve detection sensitivities sufficiently before plasma cfDNA liquid biopsy can be used for early detection of colorectal lesions.

Chapter 7 Concluding Discussion

CRC is a heterogeneous disease and it is the second leading cause of cancer deaths in the UK [15]. Morbidity and mortality of CRC patients are strongly correlated with the cancer stage at diagnosis: current 5-year relative survival rate is >90% for those diagnosed at stage I cancer compared to <10% for those diagnosed at stage IV [18]. Certainly, numerous epidemiological studies have identified the link between improving CRC survival rates and the detection of early lesions [19, 162-167]. Ever since the detection of tumour-derived cfDNA in the plasma of pancreatic cancer patients was reported more than 2 decades ago [250], numerous studies have shown that plasma cfDNA can be implemented as a prognostic and predictive biomarker in monitoring treatment responses and detecting early emergence of resistance for various cancers [142, 295, 297, 298]. Although several studies have also investigated the potential of cfDNA biomarker for early detection, their findings have yet to find success in clinical translation [299, 314, 319, 386, 391]. Therefore, this project aimed to investigate whether cfDNA could be used to detect and monitor progression of precancerous colorectal lesions, and whether cfDNA could be implemented as a surrogate biomarker for preclinical studies. To fulfil these objectives, I carried out a combination of preclinical studies, involving *Cre-Apc^{fl/fl}* mouse model of CRC, and analysis of adenoma patient samples to investigate the dynamics of plasma cfDNA, using qPCR, ddPCR and NGS as experimental tools.

Firstly, I performed quantitative analysis of total plasma cfDNA levels in both preclinical and patient samples to test the hypothesis that an increase in plasma cfDNA levels are associated with the presence of lesions, even with early, localised stages of CRC, based on studies that reported this correlation [299, 319, 392]. The quantitative data obtained from the three preclinical studies revealed an interesting correlation. Results from the first preclinical study showed that whilst the plasma cfDNA levels were higher, on average, in the *Apc*-deleted mice (75.42 ± 68.31 ng/ml, $n=6$) compared to the controls (35.84 ± 26.68 ng/ml, $n=7$), this difference was not statistically significant ($p=0.2949$, Figure 4-2). Similar findings were made in the second preclinical study where the average plasma cfDNA levels were higher in the *Apc*-deleted group (66.74 ± 59.54

ng/ml, n= 8) compared to the control group (31.22 ±44.70 ng/ml, n=9), but this again was not statistically significant (p=0.0927, Figure 4-7). Interestingly, the p value was lower for the second data set, which might have been a consequence of higher adenoma counts observed in the Apc-deleted mice from the second study (Table 4-4).

Therefore, for the third study, a diet factor was introduced and mice were allocated into the normal (NFD) and high fat diet (HFD) groups to induce higher disease burden, as our group had shown that HFD promoted adenoma growth in *Apc^{Min}* models of CRC [225]. Whilst the allocation of new study diets did not significantly affect the survival rates for the Apc-deleted mice (NFD=57.5 days vs. HFD=44.0 days, p=0.7190; Figure 4-9), mice in this study did develop a substantial number of adenomas throughout the SI, even in the medial and distal regions, and in the caecum (Figure 4-4), irrespective of the diet groups. Despite this, analysis of the total plasma cfDNA levels in these mice showed unexpectedly higher levels in the controls compared to the Apc-deleted mice, in both diet groups (Figure 4-19). In the NFD group, mean cfDNA level was 110.6 ±105.03 ng/ml (n=4) for the controls and 14.65 ±11.52 ng/ml (n=10) for the Apc-deleted mice (p=0.0384). In the HFD group, it was 226.3 ±102.80 ng/ml (n=4) for the controls and 32.43 ±28.18 (n=9) for the Apc-deleted mice (p<0.0001). One possible explanation for these surprising results maybe the acute intestinal bleeding prevalent in these APC-deleted mice resulted in shedding DNA into the faeces preferentially, which indirectly diminished the extent of cfDNA release into the circulation. We also noted that the severe anaemia associated with intestinal haemorrhages caused a substantial extramedullary haematopoiesis in the spleen, whether this physiological adaptation has any impact on cfDNA dynamics needs to be established, but cannot be excluded. The increased cfDNA yields in control animals compared to the previous two studies remains puzzling, but it was possible that that these elevated levels were caused by contamination with lysed WBC cells during the sample collection. Based on this, for further work, one improvement would be to develop a PCR-based assay to detect recombined immunoglobulin loci to identify contaminated samples [393].

In parallel, quantitative analysis of human plasma samples also gave statistically insignificant (p=0.1012, Figure 6-5) levels of total plasma cfDNA concentrations between the patient cohort (6.106 ±0.5167 ng/ml, n=48) and the controls (7.456 ±0.642 ng/ml, n=37) (section 2.1.7). As other groups reported comparable levels of plasma cfDNA in

their control groups, despite using different reference genes for quantification [299, 319], this led to the speculation that this patient cohort showed relatively low plasma cfDNA levels, on average. For comparison, Cree's and Watanabe's groups reported mean cfDNA levels of 15.04ng/ml (n=26) and 9.8ng/ml (n=57), respectively, for their patient cohorts [299, 319]. It was likely that the inclusion of larger and advanced polyps, even stage I and II samples, contributed to the quantification of elevated cfDNA levels in their studies [299, 319]. Interestingly, the data from this experiment supported the findings from Pilotti's group where they also failed to observe statistically relevant elevation of plasma cfDNA levels between patients with precancerous lesions and 'polyp-free' controls [386]. Overall, the data showed that there was no statistically significant correlation between the total plasma cfDNA levels and the presence of adenomas, at least for this patient cohort.

Quantitative analysis of total plasma cfDNA is an appealing concept due to its ease of use – for instance, a master assay can be designed to test samples irrespective of cancer types. However, its inability to discriminate cfDNA fragments from normal and cancerous origins is a major limitation of its applicability, as studies have shown that plasma cfDNA levels can be influenced by various non-cancer-related conditions (e.g. autoimmune diseases) [142, 263-265]. Therefore, although limited observations have been made that the presence of adenomas might contribute to raising plasma cfDNA levels in the preclinical samples, these findings were not consistent nor statistically significant. Moreover, analysis of patient samples failed to support this correlation. Overall, this study has suggested that the use of quantitative analysis of plasma cfDNA is not suitable for the detection of precancerous colorectal lesions.

Therefore, this leads to the question of whether plasma analysis can be refined to measure tumour-derived cfDNA during early tumourigenesis in CRC. As before, both preclinical and patient samples were used to test if tumour-derived cfDNA exists at a detectable level in plasma when disease burden is low, and whether it can be quantitatively assessed to inform disease progression.

With regards to the analysis of preclinical samples, as mentioned previously, recombined *Apc^{fl}* alleles were designated as a surrogate for the tumour-derived alteration, as it represents the consequence of *Apc* deletion that drives adenoma growth

in the *Cre-Apc^{fl/fl}* model. However, analysis of preclinical plasma samples from the three independent studies yielded very few positive samples. In the first study cohort, recombined alleles were detected in only 1 out of 6 samples using targeted Apc-Rec qPCR assay – this positive detection occurred at a very late C_t cycle (41.80) and with a ΔC_t of 11.97 against the positive control, which represented amplification of a very negligible amount of DNA, casting doubt over the reliability of this result (Figure 4-3). However, low disease burden in these mice exacerbated the difficulty in detecting recombined DNA in the plasma. Nevertheless, even raising the severity limit to increase disease burden for the two subsequent *in vivo* studies and transferring analysis onto the ddPCR platform had little effect on the detection rate. In the second study, I was only able to detect few recombined copies in 2 out of 8 Apc-deleted mice (25%) at low target fractions of 0.30% and 0.15% (Figure 4-8 & Table 4-5). In the third study, low fractions (<0.2%) of recombined DNA was detected in a total of 3 out of 19 (16%) Apc-deleted mice from both NFD and HFD groups (Table 4-9 & Figure 4-21).

Taken together, these data indicated that even though it was possible to detect recombined DNA as a surrogate biomarker in the plasma of adenoma-bearing mice, the sensitivity of this analysis was too low to be implemented in preclinical studies with the *Cre-Apc^{fl/fl}* model of CRC. It could be argued that this analysis was limited by the generally low fractions of target cfDNA, leaving the possibility that further optimisation and/or a re-design of the assay might give better sensitivities. It should be noted that these Apc-deleted mice, especially in the third study, had multiple lesions and presented a high disease burden, more so than what can be expected in humans. Although the hypothesis was that tumour-derived DNA would be readily detected in the plasma due to this high disease burden, the findings here contradicted this, which suggested that precancerous lesions might have a reduced propensity to release cfDNA in this mouse model. In this regard, one possible refinement for future studies could be the use of a more aggressive GEMM of CRCs, such as the *AhCre Apc^{fl/+} Kras^{G12V}* model that can harbour invasive carcinomas, to induce greater cfDNA release into peripheral circulation as a way of increasing the probability of detecting surrogate cfDNA [143].

Moreover, patient samples were analysed using custom ddPCR assays to test whether tumour-derived cfDNA can be detected in the plasma. FFPE samples from the patient cohort were screened for the hotspot mutations in the frequently implicated

BRAF and *KRAS* oncogenes (Figure 6-3), and matched plasma samples were then tested for tumour-derived cfDNA. No *BRAF*-MU copies were detected in the plasma from positive patient cases (n=4), but as all cases presented with very small lesions (Figure 6-6), this likely meant that very few tumour-derived cfDNA fragments were present in the plasma cfDNA pool. However, analysis of *KRAS* mutations in plasma gave more encouraging results. Analysis of 3 possible missense mutations at *KRAS* c.34G resulted in 2 positive samples in a total of 13 patient cases (15.4%) (Figure 6-7). A similar analysis for mutations at *KRAS* c.35G yielded 6 positive plasma samples out of 34 positive patient cases (17.6%) (Figure 6-9). Moreover, 6 out of 24 plasma samples (25.0%) were positively identified for the *KRAS* codon 13 mutation (c.38G) (Figure 6-11). Whilst these data showed that tumour-derived cfDNA fragments can be detected in the plasma of adenoma patients, sensitivities of these targeted ddPCR assays were low; however, these sensitivities were at a comparable range to the gFOBT analysis that has 5-33% sensitivity in identifying adenomas [117, 175, 176]. In addition, there was no correlation between the MU fractional abundances in the FFPE samples and MU detection in the matched plasma samples. Moreover, all MU detection in the plasma occurred at the MU fractions lower than 0.3%, highlighting the challenges of detecting disease at this early stage (Table 6-1, Table 6-4 & Table 6-6). For these adenoma cases, analysis of tissue samples showed that *BRAF* and *KRAS* mutation frequencies were low with the majority of cases having <30% MU allele frequencies (Table 6-2, Table 6-3 and Table 6-5). Even if the stromal contribution was taken into account as a dilution factor, this nevertheless highlighted the weak representation of target allele frequencies in pre-neoplastic cases compared to cancer (stage I+) samples, which could explain for low MU fractions in the plasma. Indeed, Vogelstein's group showed that MU fractions in the plasma were correlated to cancer stages as they observed only 0.02% MU fraction for one of their adenoma patients, compared to 0.01-0.12% MU fractions for stage I and up to 27% MU fractions in stage IV patients [260, 316, 317]. These low MU fractions likely serve as an intrinsic limitation for tumour-derived cfDNA detection in the plasma from adenoma patients. These findings agreed with other studies that have also reported low detection rates for *KRAS* mutations in the plasma from patients with neoplastic polyps [314, 391].

Interestingly, analysis of FFPE samples suggested the possibility that *KRAS* and *BRAF* mutations might be clonal alternations for the majority of patient cases, based on wide discrepancies in MU fractions between FFPE samples collected from individual

patients (Table 6-2, Table 6-3 & Table 6-5). One example was patient H364 who had one FFPE region with 0.0003% *KRAS* MU fraction and the other 2.5250% (Table 6-3). Subsequent multi-regional NGS analysis did confirm that ITH was present in precancerous lesions (section 6.6), especially in patient H263 that harboured simultaneous *KRAS* mutations in codons 12 and 13 (Figure 6-15-A). The *KRAS* mutation at codon 12 (c.35G) was confirmed as clonal by both the qPCR and NGS data (Figure 6-15-A). On the other hand, the clonality of *KRAS* codon 13 (c.38G) mutation was debated as the ddPCR analysis identified positive MU fractions in all FFPE pieces analysed, despite NGS calling *KRAS* c.38G mutations in only 4 out of 5 FFPE regions sequenced (Figure 6-15-A).

Nonetheless, the individual genetic profiles provided by the NGS analysis had guided the development of patient-specific mutation detection assays that targeted high frequencies mutations, to improve the detection of tumour-derived cfDNA in the plasma [252, 318]. To this end, five patient-specific mutation targets were selected: 3 *APC*, 1 *TP53* and 1 *PIK3CA* (section 5.4). The *APC* and *TP53* mutations were all identified at high frequencies in the FFPE samples, and were deduced to be truncal, and although the *PIK3CA* was a clonal alteration, it was included due to its implication in both early and late stages of CRC (section 6.6) [387]. Unfortunately, only 1 out of 5 assays was successful at detecting tumour-derived cfDNA in the plasma (section 6.6); the *APC*1429 ddPCR assay detected plasma cfDNA copies of the *APC* c.4285C>T/p.Q1429* mutation at 0.19% MU fractional abundance (Figure 6-18-B), although the accompanying Poisson analysis predicted independent detections at below the 1.0 copies threshold at the minimum level of droplet distribution (Figure 6-18-C). Even though the detectable levels of MU fraction in this adenoma patient was higher than what was reported by Vogelstein's group [316], this study has suggested that even targeting truncal mutations in plasma likely comes with a caveat of false calling. Nonetheless, this patient-specific mutation analysis was a small experiment with few samples, and further investigation is necessary to form any definitive conclusion on the diagnostic sensitivity of this approach.

During this project, a pilot experiment was performed with the preclinical samples from the third study to assess if faecal samples could be used for DNA biomarker detection as an alternative method to the plasma analysis. This was based on

numerous reports that genetic and epigenetic alterations are detectable in faecal samples from human patients [317, 359-362]. Although quantitative analysis of faecal samples did not show any correlation between total faecal DNA yields and the adenoma development (Table 4-11), recombined DNA was successfully detected in the faeces from adenoma-bearing mice (Figure 4-22). Moreover, the fractions of recombined DNA in faeces showed an overall upward trend with the disease progression, measured as time elapsed since tamoxifen induction – the NFD mice showed higher average recombined fractions compared to the HFD counterparts, but this trend was not statistically significant between the two diet groups at weekly time-points (Figure 4-22, Table 4-12). However, analysis of faecal samples collected at the end-point revealed an overall increase in the mean recombined fractions compared to the weekly samples, and the NFD mice showed statistically higher recombined DNA fractions compared to the HFD mice ($p=0.0099$, Table 4-12). Taken together, there is a potential in using faecal DNA as a biomarker to detect the presence of adenomas, and possibly monitor disease progression, in preclinical models, but further work is necessary to better understand the dynamics of faecal DNA in relation to diets and the disease burden.

It must be acknowledged that all plasma analysis performed in this project was based on the prior knowledge of mutation profiles obtained from the matched FFPE tissues, and this is a limitation in implementing the plasma-based mutation analysis for cases where tissue biopsies are not available. For instance, plasma-based mutation detection will not be applicable for preventive measures such as population-wide screening in healthy individuals. However, this approach may be feasible as a complementary diagnostic test for monitoring recurrence in patients after a primary resection [140, 295]. Nonetheless, this also means that any clinical implementation of plasma cfDNA analysis necessitates prior knowledge of patient-specific mutation profiles, most likely obtained through targeted NGS, to enable the development of custom assays for longitudinal follow-up throughout the treatment pipeline, which will be costly in long-term. However, in recent years, there is a growing interest in optimising targeted NGS platforms for use with plasma cfDNA. One such example is the OncoPrint™ Colon cfDNA assay (Thermo Fisher®), which covers >240 hotspots of frequently implicated genes, on the Ion Torrent™ system, and the manufacturer has suggested up to 0.1% LoD with 20ng input DNA. Therefore, one future direction of this investigation into the potential of plasma cfDNA biomarker for early detection maybe to assess if

using targeted NGS panels can serve as an alternative method, with better coverage, to ddPCR-based assays.

Technical limitations also needed to be considered when using early stage samples that containing low MU fractions (<1%) [252]. The ddPCR system is cost-effective and offers high-resolution, targeted analysis with LoDs as low as 0.01% (Figure 1-18). Paradoxically, the high resolution of ddPCR technique could also lead to artefacts and trace contaminants being detected in control wells despite using a UV-sterilised station for assay preparation (sections 6.5 & 6.7). Consequently, this could make it difficult to call true positives for plasma samples where positive detection occurs at the detection threshold. Furthermore, frequent issues with droplet generation were encountered where the number of droplets 'accepted' (i.e. analysed) were below 10,000 droplets threshold per 20µl reaction – this issue affected random reaction wells within an 8-well cartridge, which further exacerbated the difficulty in detecting few MU copies when wells containing plasma samples were affected.

One other limitation with using plasma cfDNA for early detection lies in the plasma samples themselves. Early plasma samples contain very low MU fractions (<1%) [252, 260, 316], as proven throughout this project, and this contributed to the difficulty with cfDNA biomarker detection in the plasma from both preclinical and patient cases, despite using a very sensitive technique like ddPCR. One possible solution for future studies is to increase the volume of blood collected to obtain higher cfDNA yields as a way of improving the probability of cfDNA biomarker detection in early plasma samples. Indeed, other studies have used up to 10ml of plasma for their analysis [276, 297, 298].

Overall, though a statistically significant correlation between total plasma cfDNA levels and the presence of adenomas was not observed, it was indeed possible to detect tumour-derived cfDNA in plasma samples from both mice and patients with precancerous lesions, albeit at low sensitivities. Moreover, pre-clinical studies in animals suggested that faecal material is a promising reservoir of surrogate DNA biomarkers for monitoring the disease onset and progression in future preclinical studies. Although results here argued against the clinical implementation of cfDNA biomarker for early detection, further studies are necessary to ascertain the earliest detection threshold and to assess the extent of ITH in premalignant lesions, and to understand how these factors

influence plasma cfDNA dynamics between adenoma patients. To this end, future improvements in sample processing and technical advancement in the PCR or NGS systems may be crucial in realising the potential of cfDNA biomarker for the early detection of CRC.

Chapter 8 Appendices

8.1 Accompanying Data for Quantification of Total Plasma cfDNA Levels in the Patient & Control Groups in Section 6.4

Samples with <i>BRAF</i> c.1799T>A Mutations	
<i>Sample ID</i>	<i>Total cfDNA Concentration (ng/ml plasma)</i>
H158	3.0509
H313	0.8838
H319	4.0662
H331	2.6403

Samples with <i>KRAS</i> c.34G>D Mutations	
<i>Sample ID</i>	<i>Total cfDNA Concentration (ng/ml plasma)</i>
H142	7.4209
H149	19.7400
H183	4.9616
H264	5.8920
H314	3.0947
H325	11.2470
H329	3.9262
H338	8.3362
H341	3.9831
H344	11.2680
H775	15.4672
H784	4.5181
H785	4.3156

Samples with <i>KRAS</i> c.35G>D Mutations	
<i>Sample ID</i>	<i>Total cfDNA Concentration (ng/ml plasma)</i>
H146	5.0917
H149	19.7400
H151	1.9108
H154	5.5731
H155	4.0437
H185	5.5847
H189	3.9646
H263	11.3956
H264	5.8920
H265	3.1237
H266	8.1693
H318	7.1245
H320	3.4461
H321	6.3984
H325	11.2470
H326	4.6661
H329	3.9262
H333	7.5999
H334	5.1554
H335	11.3602
H336	4.3811
H338	8.3362
H339	2.7978
H343	8.4432
H344	11.2680
H345	5.7988
H346	4.4443
H347	5.6335
H693	3.5021
H696	5.6065
H697	5.8280
H774	6.9757
H783	5.6107
H784	4.5181

Samples with <i>KRAS</i> c.38G>A Mutations	
<i>Sample ID</i>	<i>Total cfDNA Concentration (ng/ml plasma)</i>
H142	7.4209
H149	19.7400
H151	1.9108
H154	5.5731
H183	4.9616
H187	6.6138
H190	5.8279
H264	5.8920
H314	3.0947
H318	7.1245
H321	6.3984
H325	11.2470
H326	4.6661
H329	3.9262
H334	5.1554
H335	11.3602
H336	4.3811
H338	8.3362
H343	8.4432
H344	11.2680
H347	5.6335
H693	3.5021
H783	5.6107
H784	4.5181

Controls (without polyps)		
<i>Sample ID</i>	<i>Total cfDNA Concentration (ng/ml plasma)</i>	<i>Additional Diagnosis</i>
H2/13	7.1735	None
H10/13	7.6800	Diverticulosis
H41/13	6.1433	None
H42/13	2.4503	None
H51/13	5.2750	None
H52/13	4.7810	None
H277/12	12.7436	Diverticulosis, previous polyps
H287/12	7.1609	None
H290/12	5.7342	None
H291/12	13.0749	None
H292/12	8.2572	None
H293/12	8.2650	None
H295/12	6.5900	None
H388/12	19.3187	None
H390/12	3.4802	None
H396/12	4.8907	None
H398/12	14.4831	None
H403/12	8.8570	None
H958/12	5.0244	None
H959/12	12.1742	None
H967/12	4.1597	None
H972/12	2.4152	None
H973/12	4.5233	None
H978/12	1.3920	Colitis (non-specific)
H985/12	7.0666	None
H986/12	3.5270	None
H992/12	3.6339	None
H993/12	7.5910	None
H999/12	7.1137	Angioectasia, benign stricture
H1311/12	9.7025	None
H1315/12	6.9291	None
H1320/12	5.2795	None
H1323/12	14.7685	Distal quiescent ulcerative colitis
H1336/12	9.4004	Diverticulosis, previous polyps
H1339/12	11.3066	None
H1341/12	8.7326	None
H968/12	4.7776	Diverticulosis, previous polyps

*Sample IDs listed for control cases followed a different labelling system as they do not have corresponding histology number for FFPE cases.

8.2 Details of the Custom Targeted Panel used in the Multi-regional Targeted NGS Analysis in Section 6.6

<i>List of Genes Checked for Single Base Mutations</i>		
ACVR1B	EZH2	NRAS
ACVR2A	FAM123B	PCBP1
APC	FBXW7	PDGFRA
ARFGEF1	FGFR1	PIK3CA
ARID1A	FGFR2	PIK3R1
ARID2	FLT1	PMS2
ARID4A	GATA3	POLE
ASXL1	GNAS	PPP1R3A
ATM	H3F3A	PTCH1
ATR	H3F3B	PTEN
ATRX	HNF4A	PTPN11
AXIN2	IGF2	RB1
BAI3	IKBKB	RET
BCL9L	IRS2	RNF43
BCLAF1	JAK2	RSPO2
BCOR	KAT6A	RSPO3
BRAF	KDM3B	SETD2
BRCA2	KDM6A	SMAD2
CARD11	KDR	SMAD3
CBL	KIT	SMAD4
CDC27	KLF5	SOX9
CDC73	KRAS	STAG2
CDH1	LIG1	STK11
CDK12	MAP2K4	TBX3
CDK8	MAP3K4	TCF7L2
CDKN2A	MARK1	TGFBR2
CLSPN	MET	TOP2B
CREBBP	MGA	TP53
CSPP1	MLH1	TP53BP1
CTNNB1	MLH3	TPTE
DKK2	MLL2	TRIM23
DUSP16	MLL3	TRRAP
EGFR	MSH2	TSHR
ELF3	MSH3	VEGFA
EP300	MSH6	VHL
EP400	MTOR	VTI1A
ERBB2	MYC	WT1
ERBB3	NF1	ZC3H13
ERBB4	NF2	(n/a)

<i>Recurrent Focal amp/deletion Regions</i>				
<i>Chromosome</i>	<i>Start</i>	<i>Stop</i>	<i>Known Drivers</i>	<i>Recurrent Gain or Loss</i>
2	77578574	78173445	n/a	Loss
3	138179356	138620106	n/a	Gain
4	1792651	1797902	FGFR3	Loss
4	15985254	15987573	n/a	Loss
4	88077504	88083039	n/a	Loss
4	134598864	138487116	n/a	Loss
4	143080258	143083016	n/a	Loss
4	178910263	178912845	n/a	Loss
5	99023205	114756546	APC	Loss
6	10816126	10817328	n/a	Loss
6	27092680	27095422	n/a	Loss
6	102247162	102249888	n/a	Loss
7	42266969	42268947	n/a	Loss
7	63112265	64847086	n/a	Gain
7	111761657	120704534	MET	Gain
8	28220844	28227238	n/a	Loss
8	41203037	42933372	n/a	Gain
8	101260500	103876943	UBR5	Gain
8	127307422	129461559	MYC	Gain
9	21882049	22287560	CDKN2A	Loss
9	125154417	125154723	n/a	Loss
9	137722407	137733151	n/a	Loss
10	53196243	53419759	n/a	Loss
10	89225082	90666002	PTEN	Loss
10	95405434	95407142	n/a	Loss
10	114572158	114581916	n/a	Loss
11	456120	2115325	HRAS	Loss
11	9870401	9872248	n/a	Loss
11	48146614	48149266	n/a	Loss
11	95569426	95569912	n/a	Loss
11	100222555	100227422	n/a	Loss
13	26893617	29601545	FLT3	Gain
13	73494500	74402782	n/a	Gain
13	105667270	114987458	n/a	Gain
15	27571787	27573178	n/a	Loss
15	67307237	67495109	n/a	Loss
16	29874277	31019371	n/a	Gain
16	46507665	54369484	CYLD	Gain
16	54567302	55881172	n/a	Gain
17	11050665	12089643	MAP2K4	Loss
17	36642500	36848051	n/a	Gain
17	37809357	38020228	ERBB2	Gain
17	70068506	70532966	n/a	Loss
18	48509092	48544565	n/a	Loss
18	66378166	66380421	n/a	Loss
19	11166176	11173230	SMARCA4	Loss
19	21714129	21721990	n/a	Loss
20	24525630	26197899	n/a	Gain
21	15347621	31500240	n/a	Loss
22	42337638	49331012	n/a	Loss
23	29041185	29871762	n/a	Loss

List of MSI Regions Checked (Chr_start_stop)

3_15084347-15084466	11_27741126-27741245	15_24927833-24927952
16_19073889-19074008	4_70160574-70160693	4_177250143-177250262
2_165551236-165551355	12_389465-389584	11_118969083-118969202
11_94212872-94212991	2_48030581-48030700	4_83785506-83785625
5_79970856-79970975	6_10756805-10756924	20_61536633-61536752
4_153832044-153832163	6_131481216-131481335	17_16454361-16454480
14_50930736-50930855	1_204228352-204228471	4_68481819-68481938
16_10867143-10867262	2_211297807-211297926	4_13570395-13570514
16_8876734-8876853	6_111587301-111587420	12_8926622-8926741
10_29760056-29760175	6_33263905-33264024	17_46629677-46629796
5_140049043-140049162	6_29641299-29641418	14_60761375-60761494
1_113067412-113067531	2_48603021-48603140	11_85395501-85395620
6_656021-656140	20_50400923-50401042	8_38175221-38175340
3_114057944-114058063	5_77334847-77334966	11_69466215-69466334
20_46286262-46286381	2_148683627-148683746	2_203420071-203420190
12_122242599-122242718	1_233807609-233807728	3_30691813-30691932
6_31323303-31323422	11_89925004-89925123	6_80751838-80751957
17_48433907-48434026	12_80169651-80169770	14_59992040-59992159
11_31812258-31812377	11_32637461-32637580	8_23289622-23289741
8_103289290-103289409	14_93707972-93708091	6_43586640-43586759
21_35122507-35122626	7_131175922-131176041	1_32673016-32673135
4_3430339-3430458	3_5260482-5260601	8_98740046-98740165
12_83444673-83444792	9_96422553-96422672	6_158507950-158508069
14_74205713-74205832	9_69421845-69421964	18_18533620-18533739
3_129155489-129155608	3_72424491-72424610	5_32786407-32786526
16_85682230-85682349	6_80751837-80751956	18_211971-212090
8_125325159-125325278	15_34157483-34157602	17_56435102-56435221
12_40805816-40805935	2_27717457-27717576	19_55814976-55815095
18_74593370-74593489	16_4862169-4862288	4_164446802-164446921
3_51417545-51417664	11_72549591-72549710	7_44920772-44920891
4_83785506-83785625	19_38375975-38376094	1_6257725-6257844
16_15802627-15802746	19_50157982-50158101	7_77423401-77423520
11_89444552-89444671	1_115469074-115469193	4_48428105-48428224
10_64159454-64159573	2_210561607-210561726	8_95686551-95686670
5_140482885-140483004	10_93043874-93043993	11_3660920-3661039
15_79750526-79750645	5_175800368-175800487	6_129898673-129898792
13_47325599-47325718	1_204923974-204924093	2_239974051-239974170
X_54470987-54471106	17_47005643-47005762	2_196544882-196545001
11_4703416-4703535	11_124844989-124845108	4_107156445-107156564
15_55912833-55912952	14_24040377-24040496	8_48701495-48701614
2_234394178-234394297	4_38140023-38140142	X_152140937-152141056

**List of Genes Checked for
Copy Number
Amplification/Deletions**

APC
CDK8
CDKN2A
ERBB2
FGFR1
HNF4A
IGF2
IKBKB
IRS2
KAT6A
KLF5
MAP2K4
MET
MSH2
MYC
PTEN
SMAD3
SMAD4
SOX9
TCF7L2
VEGFA
VT11A

**List of Genes Checked for
Fusions**

RSPO2
RSPO3

8.3 Mutation Profiles for Patient Cases from Multi-regional Targets NGS Analysis in Section 6.6

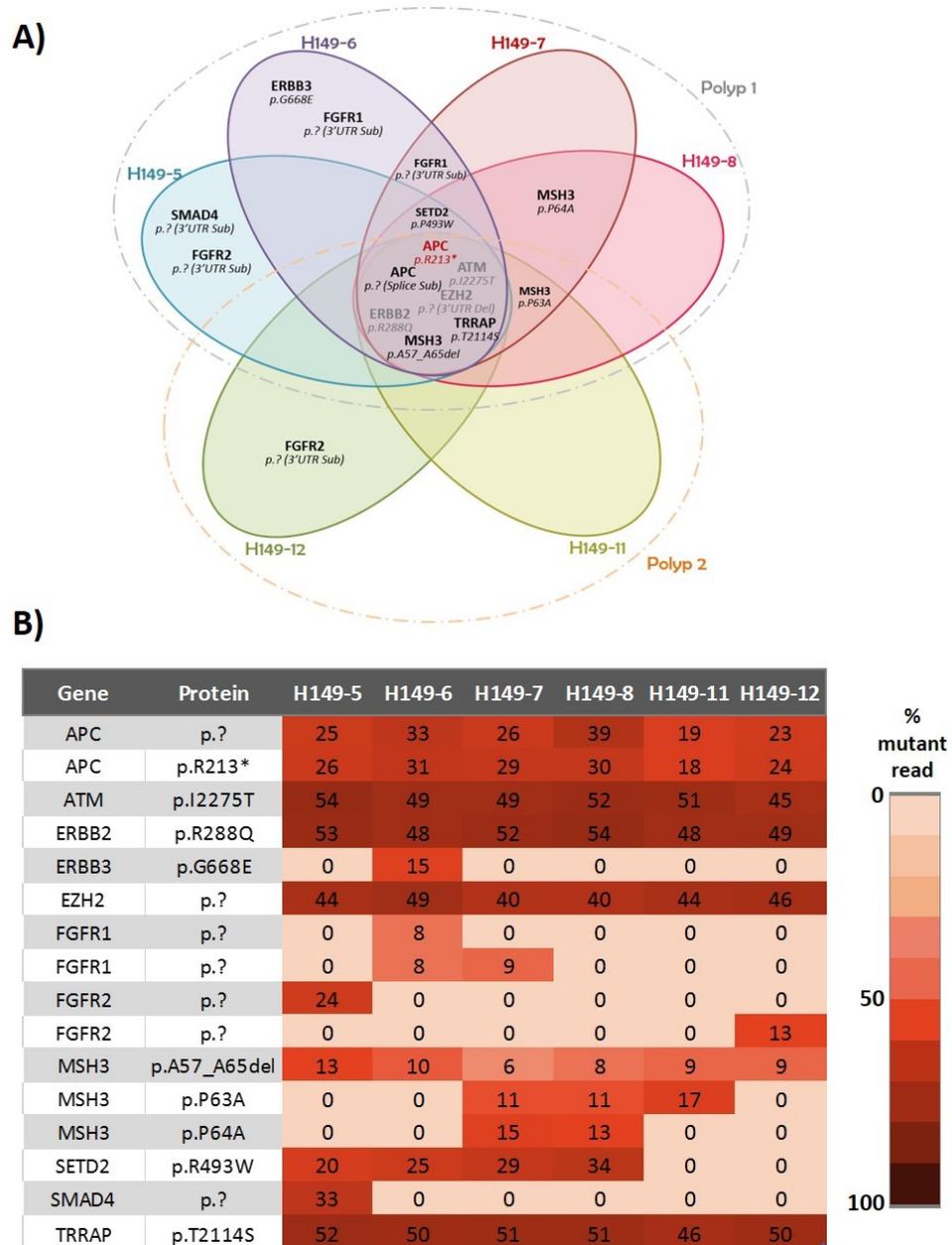


Figure 8-1: Summary of Multiregional Sequencing Data for Case H149

The patient had two low-grade TVAs in the sigmoid. FFPE samples 5-8 originated from one polyp, and samples 11 and 12 originated from the other polyp.

A) Venn diagram illustrating the distributions of mutations in the different sequenced regions. Each coloured oval represents one sequenced region. Mutations in red are driver mutations as determined by the in-house calling algorithm.

B) A heat map showing the percentage MU reads for each mutation.

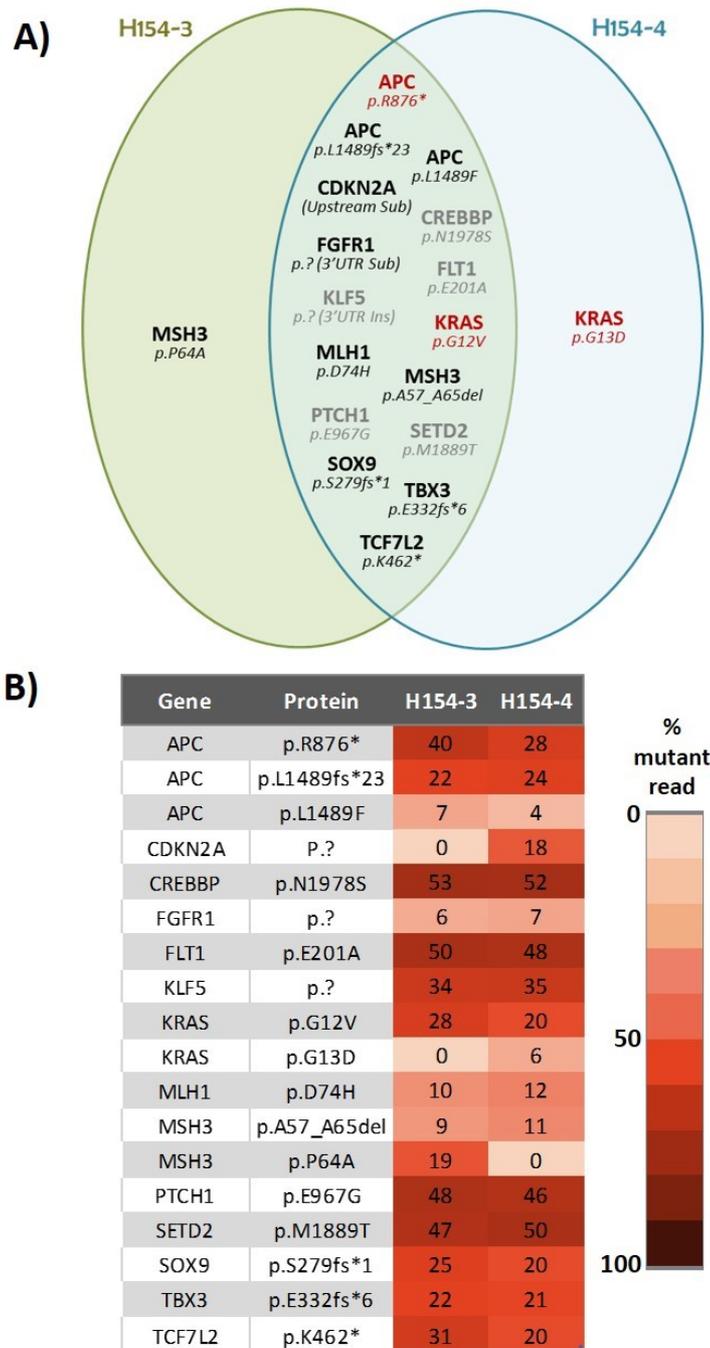


Figure 8-2: Summary of Multiregional Sequencing Data for Case H154

For this patient, the histology report does not specify the number allocation of polyp biopsies for FFPE archiving. The assumption was that the samples originated from a low-grade TVA in the descending colon.

A) Venn diagram illustrating the distributions of mutations in the different sequenced regions. Each coloured oval represents one sequenced region. Mutations in red are driver mutations as determined by the in-house calling algorithm.

B) A heat map showing the percentage MU reads for each mutation.

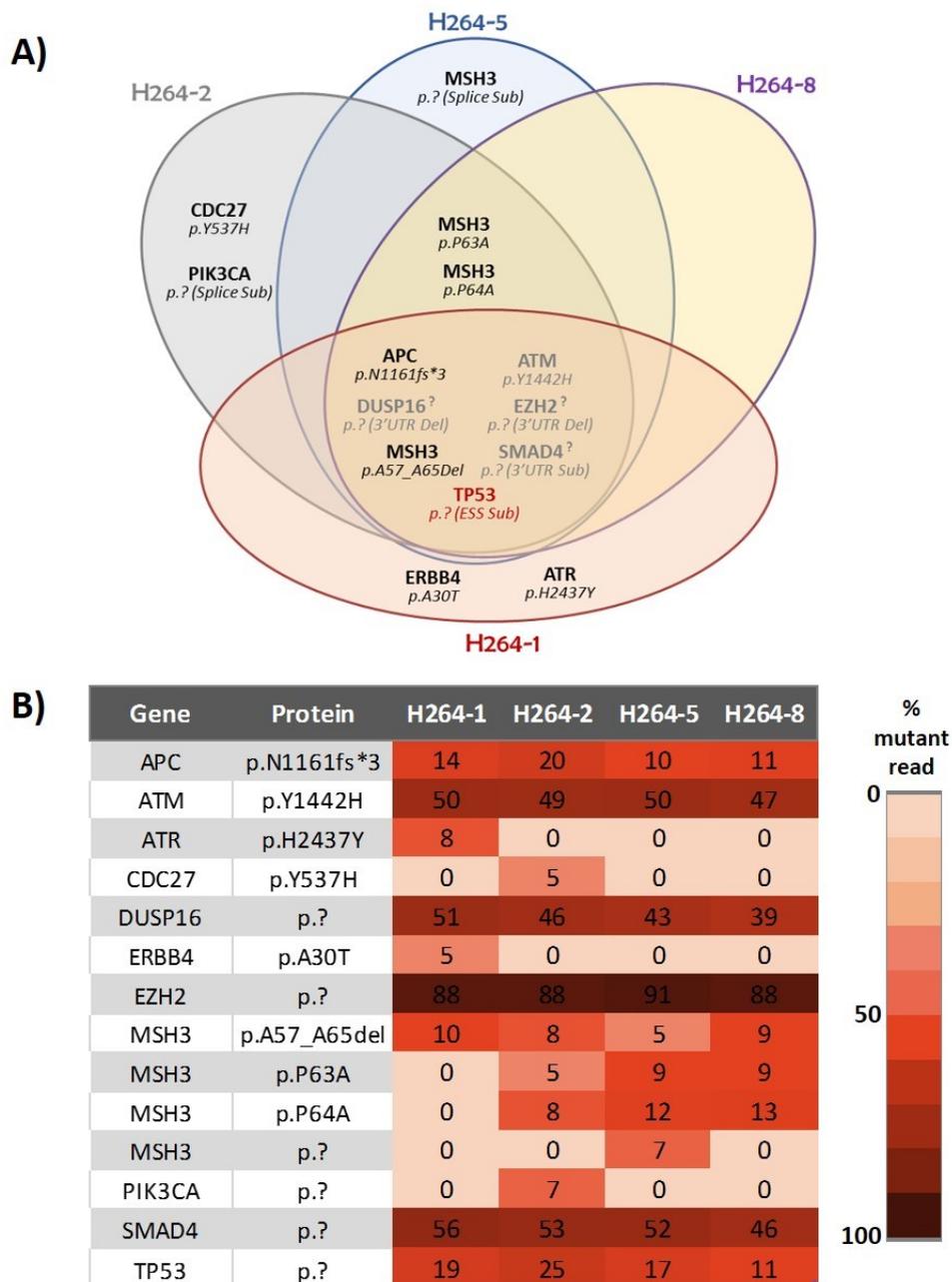


Figure 8-3: Summary of Multiregional Sequencing Data for Case H264

For this patient, all the sequenced FFPE regions originated from one large, sigmoidal TVA with low- and high-grade dysplasia.

A) Venn diagram illustrating the distributions of mutations in the different sequenced regions. Each coloured oval represents one sequenced region. Mutations in red are driver mutations as determined by the in-house calling algorithm.

B) A heat map showing the percentage MU reads for each mutation.

Chapter 9 References

1. Semrin, M.G., et al., *Anatomy, Histology, Embryology, and Developmental Anomalies of the Stomach and Duodenum*, in *Sleisenger and Fordtran's Gastrointestinal and Liver Disease (9th edition)*. 2010, Saunders, Philadelphia, U.S.A. . p. 773-785.
2. Shen, L., *Functional morphology of the gastrointestinal tract*. *Curr Top Microbiol Immunol*, 2009. **337**: p. 1-35.
3. *Gastrointestinal Tract (GI Tract)*. NIH - National Institute of Diabetes and Digestive and Kidney Diseases [cited 2017 05/06]; Available from: <https://www.ncbi.nlm.nih.gov/pubmedhealth/PMHT0022855/>.
4. *Biology of Humans*. [cited 2017 05/06]; Available from: <http://schoolbag.info/biology/humans/20.html>.
5. Crosnier, C., D. Stamatakis, and J. Lewis, *Organizing cell renewal in the intestine: stem cells, signals and combinatorial control*. *Nat Rev Genet*, 2006. **7**(5): p. 349-59.
6. Sato, T., et al., *Paneth cells constitute the niche for Lgr5 stem cells in intestinal crypts*. *Nature*, 2011. **469**(7330): p. 415-+.
7. Barker, N., *Adult intestinal stem cells: critical drivers of epithelial homeostasis and regeneration*. *Nat Rev Mol Cell Biol*, 2014. **15**(1): p. 19-33.
8. Barker, N., et al., *Crypt stem cells as the cells-of-origin of intestinal cancer*. *Nature*, 2009. **457**(7229): p. 608-11.
9. Barker, N., et al., *Identification of stem cells in small intestine and colon by marker gene Lgr5*. *Nature*, 2007. **449**(7165): p. 1003-7.
10. Krausova, M. and V. Korinek, *Wnt signaling in adult intestinal stem cells and cancer*. *Cell Signal*, 2014. **26**(3): p. 570-9.
11. van der Flier, L.G., et al., *OLFM4 is a robust marker for stem cells in human intestine and marks a subset of colorectal cancer cells*. *Gastroenterology*, 2009. **137**(1): p. 15-7.
12. van der Flier, L.G., et al., *Transcription factor achaete scute-like 2 controls intestinal stem cell fate*. *Cell*, 2009. **136**(5): p. 903-12.
13. Snippert, H.J., et al., *Intestinal crypt homeostasis results from neutral competition between symmetrically dividing Lgr5 stem cells*. *Cell*, 2010. **143**(1): p. 134-44.
14. San Roman, A.K. and R.A. Shivdasani, *Boundaries, junctions and transitions in the gastrointestinal tract*. *Experimental Cell Research*, 2011. **317**(19): p. 2711-2718.
15. *Cancer incidence for common cancers*. [cited 2017 05/06]; Available from: <http://www.cancerresearchuk.org/health-professional/cancer-statistics/incidence/common-cancers-compared#heading-Three>.
16. *Bowel cancer incidence statistics*. 17/12/2015 [cited 2017 05/06]; Available from: <http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/bowel-cancer/incidence#heading-Four>.
17. Center, M.M., A. Jemal, and E. Ward, *International trends in colorectal cancer incidence rates*. *Cancer Epidemiol Biomarkers Prev*, 2009. **18**(6): p. 1688-94.
18. *Bowel cancer survival statistics*. [cited 2017 05/06]; Available from: <http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/bowel-cancer/survival#heading-Zero>.
19. McClements, P.L., et al., *Impact of the UK colorectal cancer screening pilot studies on incidence, stage distribution and mortality trends*. *Cancer Epidemiol*, 2012. **36**(4): p. e232-42.
20. Fleming, M., et al., *Colorectal carcinoma: Pathologic aspects*. *J Gastrointest Oncol*, 2012. **3**(3): p. 153-73.
21. Shussman, N. and S.D. Wexner, *Colorectal polyps and polyposis syndromes*. *Gastroenterol Rep (Oxf)*, 2014. **2**(1): p. 1-15.

22. Goldstein, N.S., *Serrated pathway and APC (conventional)-type colorectal polyps: molecular-morphologic correlations, genetic pathways, and implications for classification*. Am J Clin Pathol, 2006. **125**(1): p. 146-53.
23. Aust, D.E., G.B. Baretton, and G.I.P.o.t.G.S.o.P. Members of the Working Group, *Serrated polyps of the colon and rectum (hyperplastic polyps, sessile serrated adenomas, traditional serrated adenomas, and mixed polyps)-proposal for diagnostic criteria*. Virchows Arch, 2010. **457**(3): p. 291-7.
24. *Staging of Colorectal Cancer*. [cited 2017 05/06]; Available from: http://www.hopkinscoloncancercenter.org/CMS/CMS_Page.aspx?CurrentUDV=59&CMS_Page_ID=EEA2CD91-3276-4123-BEEB-BAF1984D20C7.
25. John, S.K., et al., *Symptoms and signs in patients with colorectal cancer*. Colorectal Dis, 2011. **13**(1): p. 17-25.
26. Vega, P., F. Valentin, and J. Cubiella, *Colorectal cancer diagnosis: Pitfalls and opportunities*. World J Gastrointest Oncol, 2015. **7**(12): p. 422-33.
27. Majumdar, S.R., R.H. Fletcher, and A.T. Evans, *How does colorectal cancer present? Symptoms, duration, and clues to location*. Am J Gastroenterol, 1999. **94**(10): p. 3039-45.
28. Health, P., *Colon Cancer Treatment (PDQ®)*. 2017(Health Professional Version).
29. Xynos, E., et al., *Clinical practice guidelines for the surgical management of colon cancer: a consensus statement of the Hellenic and Cypriot Colorectal Cancer Study Group by the HeSMO*. Ann Gastroenterol, 2016. **29**(1): p. 3-17.
30. Braun, M.S. and M.T. Seymour, *Balancing the efficacy and toxicity of chemotherapy in colorectal cancer*. Ther Adv Med Oncol, 2011. **3**(1): p. 43-52.
31. Li, Y., et al., *A Review of Neoadjuvant Chemoradiotherapy for Locally Advanced Rectal Cancer*. Int J Biol Sci, 2016. **12**(8): p. 1022-31.
32. Carethers, J.M. and B.H. Jung, *Genetics and Genetic Biomarkers in Sporadic Colorectal Cancer*. Gastroenterology, 2015. **149**(5): p. 1177-1190 e3.
33. Fearhead, N.S., J.L. Wilding, and W.F. Bodmer, *Genetics of colorectal cancer: hereditary aspects and overview of colorectal tumorigenesis*. Br Med Bull, 2002. **64**: p. 27-43.
34. Worthley, D.L. and B.A. Leggett, *Colorectal cancer: molecular features and clinical opportunities*. Clin Biochem Rev, 2010. **31**(2): p. 31-8.
35. Yamagishi, H., et al., *Molecular pathogenesis of sporadic colorectal cancers*. Chin J Cancer, 2016. **35**: p. 4.
36. Aoki, K. and M.M. Taketo, *Adenomatous polyposis coli (APC): a multi-functional tumor suppressor gene*. J Cell Sci, 2007. **120**(Pt 19): p. 3327-35.
37. Fearhead, N.S., M.P. Britton, and W.F. Bodmer, *The ABC of APC*. Hum Mol Genet, 2001. **10**(7): p. 721-33.
38. Clevers, H. and R. Nusse, *Wnt/beta-catenin signaling and disease*. Cell, 2012. **149**(6): p. 1192-205.
39. MacDonald, B.T., K. Tamai, and X. He, *Wnt/beta-catenin signaling: components, mechanisms, and diseases*. Dev Cell, 2009. **17**(1): p. 9-26.
40. Wu, G., et al., *Inhibition of GSK3 phosphorylation of beta-catenin via phosphorylated PPPSPXS motifs of Wnt coreceptor LRP6*. PLoS One, 2009. **4**(3): p. e4926.
41. Novak, A. and S. Dedhar, *Signaling through beta-catenin and Lef/Tcf*. Cellular and Molecular Life Sciences, 1999. **56**(5-6): p. 523-537.
42. Schepers, A. and H. Clevers, *Wnt signaling, stem cells, and cancer of the gastrointestinal tract*. Cold Spring Harb Perspect Biol, 2012. **4**(4): p. a007989.
43. de Lau, W., et al., *The R-spondin/Lgr5/Rnf43 module: regulator of Wnt signal strength*. Genes Dev, 2014. **28**(4): p. 305-16.

44. Jass, J.R., et al., *Advanced colorectal polyps with the molecular and morphological features of serrated polyps and adenomas: concept of a 'fusion' pathway to colorectal cancer*. *Histopathology*, 2006. **49**(2): p. 121-31.
45. Pino, M.S. and D.C. Chung, *The Chromosomal Instability Pathway in Colon Cancer*. *Gastroenterology*, 2010. **138**(6): p. 2059-2072.
46. Rowan, A.J., et al., *APC mutations in sporadic colorectal tumors: A mutational "hotspot" and interdependence of the "two hits"*. *Proceedings of the National Academy of Sciences of the United States of America*, 2000. **97**(7): p. 3352-3357.
47. Yang, J., et al., *Adenomatous polyposis coli (APC) differentially regulates beta-catenin phosphorylation and ubiquitination in colon cancer cells*. *J Biol Chem*, 2006. **281**(26): p. 17751-7.
48. Bienz, M., *The subcellular destinations of APC proteins*. *Nat Rev Mol Cell Biol*, 2002. **3**(5): p. 328-38.
49. Neufeld, K.L., *Nuclear APC*. *Adv Exp Med Biol*, 2009. **656**: p. 13-29.
50. Sierra, J., et al., *The APC tumor suppressor counteracts beta-catenin activation and H3K4 methylation at Wnt target genes*. *Genes Dev*, 2006. **20**(5): p. 586-600.
51. Willert, K. and K.A. Jones, *Wnt signaling: is the party in the nucleus?* *Genes & Development*, 2006. **20**(11): p. 1394-1404.
52. Bienz, M. and H. Clevers, *Linking colorectal cancer to Wnt signaling*. *Cell*, 2000. **103**(2): p. 311-20.
53. Kinzler, K.W. and B. Vogelstein, *Lessons from hereditary colorectal cancer*. *Cell*, 1996. **87**(2): p. 159-70.
54. Luchtenborg, M., et al., *Mutations in APC, CTNNB1 and K-ras genes and expression of hMLH1 in sporadic colorectal carcinomas from the Netherlands Cohort Study*. *BMC Cancer*, 2005. **5**: p. 160.
55. *Comprehensive molecular characterization of human colon and rectal cancer*. *Nature*, 2012. **487**(7407): p. 330-7.
56. v81, C. *Catalogue of Somatic Mutations In Cancer*. 2014 [cited 2017 05/06]; Available from: <http://cancer.sanger.ac.uk/cosmic/browse/tissue>.
57. Chial, H., *Tumor Suppressor (TS) Genes and the Two-Hit Hypothesis*. *Nature Education*, 2008. **1**(1): p. 177.
58. Knudson, A.G., *Two genetic hits (more or less) to cancer*. *Nat Rev Cancer*, 2001. **1**(2): p. 157-62.
59. Fearon, E.R., *Molecular genetics of colorectal cancer*. *Annu Rev Pathol*, 2011. **6**: p. 479-507.
60. Dionigi, G., et al., *Genetic alteration in hereditary colorectal cancer*. *Surg Oncol*, 2007. **16 Suppl 1**: p. S11-5.
61. Bodmer, W.F., et al., *Localization of the gene for familial adenomatous polyposis on chromosome 5*. *Nature*, 1987. **328**(6131): p. 614-6.
62. Macpherson, A.J., I. Bjarnason, and I.C. Forgacs, *Discovery of the gene for familial adenomatous polyposis*. *BMJ*, 1992. **304**(6831): p. 858-9.
63. Nieuwenhuis, M.H. and H.F.A. Vasen, *Correlations between mutation site in APC and phenotype of familial adenomatous polyposis (FAP): A review of the literature*. *Critical Reviews in Oncology Hematology*, 2007. **61**(2): p. 153-161.
64. Wennstro.J, E.R. Pierce, and V.A. Mckusick, *Hereditary Benign and Malignant Lesions of Large Bowel*. *Cancer*, 1974. **34**(3): p. 850-857.
65. Beroud, C. and T. Soussi, *APC gene: database of germline and somatic mutations in human tumors and cell lines*. *Nucleic Acids Res*, 1996. **24**(1): p. 121-4.
66. Ficari, F., et al., *APC gene mutations and colorectal adenomatosis in familial adenomatous polyposis*. *Br J Cancer*, 2000. **82**(2): p. 348-53.

67. Dhillon, A.S., et al., *MAP kinase signalling pathways in cancer*. *Oncogene*, 2007. **26**(22): p. 3279-90.
68. Garnett, M.J. and R. Marais, *Guilty as charged: B-RAF is a human oncogene*. *Cancer Cell*, 2004. **6**(4): p. 313-319.
69. Mercer, K., et al., *Expression of endogenous oncogenic V600E-raf induces proliferation and developmental defects in mice and transformation of primary fibroblasts*. *Cancer Res*, 2005. **65**(24): p. 11493-500.
70. Hubbard, S.R. and W.T. Miller, *Receptor tyrosine kinases: mechanisms of activation and signaling*. *Curr Opin Cell Biol*, 2007. **19**(2): p. 117-23.
71. Margolis, B. and E.Y. Skolnik, *Activation of Ras by receptor tyrosine kinases*. *J Am Soc Nephrol*, 1994. **5**(6): p. 1288-99.
72. Kohler, M., et al., *Activation loop phosphorylation regulates B-Raf in vivo and transformation by B-Raf mutants*. *EMBO J*, 2016. **35**(2): p. 143-61.
73. Wan, P.T.C., et al., *Mechanism of activation of the RAF-ERK signaling pathway by oncogenic mutations of B-RAF*. *Cell*, 2004. **116**(6): p. 855-867.
74. Lu, Z. and S. Xu, *ERK1/2 MAP kinases in cell survival and apoptosis*. *IUBMB Life*, 2006. **58**(11): p. 621-31.
75. Aksamitiene, E., A. Kiyatkin, and B.N. Kholodenko, *Cross-talk between mitogenic Ras/MAPK and survival PI3K/Akt pathways: a fine balance*. *Biochem Soc Trans*, 2012. **40**(1): p. 139-46.
76. Mendoza, M.C., E.E. Er, and J. Blenis, *The Ras-ERK and PI3K-mTOR pathways: cross-talk and compensation*. *Trends Biochem Sci*, 2011. **36**(6): p. 320-8.
77. Kodaki, T., et al., *The activation of phosphatidylinositol 3-kinase by Ras*. *Curr Biol*, 1994. **4**(9): p. 798-806.
78. Rodriguez-Viciana, P., et al., *Phosphatidylinositol 3' kinase: one of the effectors of Ras*. *Philos Trans R Soc Lond B Biol Sci*, 1996. **351**(1336): p. 225-31; discussion 231-2.
79. Sengupta, S., T.R. Peterson, and D.M. Sabatini, *Regulation of the mTOR complex 1 pathway by nutrients, growth factors, and stress*. *Mol Cell*, 2010. **40**(2): p. 310-22.
80. Cox, A.D., et al., *Drugging the undruggable RAS: Mission possible?* *Nat Rev Drug Discov*, 2014. **13**(11): p. 828-51.
81. Hu, M.C., et al., *IkappaB kinase promotes tumorigenesis through inhibition of forkhead FOXO3a*. *Cell*, 2004. **117**(2): p. 225-37.
82. Yang, J.Y., et al., *ERK promotes tumorigenesis by inhibiting FOXO3a via MDM2-mediated degradation*. *Nat Cell Biol*, 2008. **10**(2): p. 138-48.
83. Chambard, J.C., et al., *ERK implication in cell cycle regulation*. *Biochim Biophys Acta*, 2007. **1773**(8): p. 1299-310.
84. Ellis, C.A. and G. Clark, *The importance of being K-Ras*. *Cell Signal*, 2000. **12**(7): p. 425-34.
85. Chetty, R. and D. Govender, *Gene of the month: KRAS*. *J Clin Pathol*, 2013. **66**(7): p. 548-50.
86. Smith, G., et al., *Activating K-Ras mutations outwith 'hotspot' codons in sporadic colorectal tumours - implications for personalised cancer medicine*. *Br J Cancer*, 2010. **102**(4): p. 693-703.
87. Chen, C.C., et al., *Computational analysis of KRAS mutations: implications for different effects on the KRAS p.G12D and p.G13D mutations*. *PLoS One*, 2013. **8**(2): p. e55793.
88. Malumbres, M. and M. Barbacid, *RAS oncogenes: the first 30 years*. *Nat Rev Cancer*, 2003. **3**(6): p. 459-65.
89. Zhang, J., et al., *Molecular spectrum of KRAS, NRAS, BRAF and PIK3CA mutations in Chinese colorectal cancer patients: analysis of 1,110 cases*. *Sci Rep*, 2015. **5**: p. 18678.
90. Davies, H., et al., *Mutations of the BRAF gene in human cancer*. *Nature*, 2002. **417**(6892): p. 949-54.

91. Mercer, K.E. and C.A. Pritchard, *Raf proteins and cancer: B-Raf is identified as a mutational target*. Biochim Biophys Acta, 2003. **1653**(1): p. 25-40.
92. Pritchard, C., et al., *Mouse models for BRAF-induced cancers*. Biochemical Society Transactions, 2007. **35**: p. 1329-1333.
93. Matallanas, D., et al., *Raf family kinases: old dogs have learned new tricks*. Genes Cancer, 2011. **2**(3): p. 232-60.
94. Barras, D., *BRAF Mutation in Colorectal Cancer: An Update*. Biomark Cancer, 2015. **7**(Suppl 1): p. 9-12.
95. Carragher, L.A., et al., *V600EBraf induces gastrointestinal crypt senescence and promotes tumour progression through enhanced CpG methylation of p16INK4a*. EMBO Mol Med, 2010. **2**(11): p. 458-71.
96. Brannon, A.R., et al., *Comparative sequencing analysis reveals high genomic concordance between matched primary and metastatic colorectal cancer lesions*. Genome Biol, 2014. **15**(8): p. 454.
97. Colussi, D., et al., *Molecular pathways involved in colorectal cancer: implications for disease behavior and prevention*. Int J Mol Sci, 2013. **14**(8): p. 16365-85.
98. Burrell, R.A., et al., *Replication stress links structural and numerical cancer chromosomal instability*. Nature, 2013. **494**(7438): p. 492-6.
99. Nazemalhosseini Mojarad, E., et al., *The CpG island methylator phenotype (CIMP) in colorectal cancer*. Gastroenterol Hepatol Bed Bench, 2013. **6**(3): p. 120-8.
100. Ogino, S. and A. Goel, *Molecular classification and correlates in colorectal cancer*. J Mol Diagn, 2008. **10**(1): p. 13-27.
101. Weisenberger, D.J., et al., *CpG island methylator phenotype underlies sporadic microsatellite instability and is tightly associated with BRAF mutation in colorectal cancer*. Nat Genet, 2006. **38**(7): p. 787-93.
102. Ogino, S., et al., *Evaluation of markers for CpG island methylator phenotype (CIMP) in colorectal cancer by a large population-based sample*. J Mol Diagn, 2007. **9**(3): p. 305-14.
103. Ogino, S., et al., *CpG island methylator phenotype, microsatellite instability, BRAF mutation and clinical outcome in colon cancer*. Gut, 2009. **58**(1): p. 90-6.
104. Gemayel, R., et al., *Beyond junk-variable tandem repeats as facilitators of rapid evolution of regulatory and coding sequences*. Genes (Basel), 2012. **3**(3): p. 461-80.
105. Armaghany, T., et al., *Genetic alterations in colorectal cancer*. Gastrointest Cancer Res, 2012. **5**(1): p. 19-27.
106. Murphy, K.M., et al., *Comparison of the microsatellite instability analysis system and the Bethesda panel for the determination of microsatellite instability in colorectal cancers*. J Mol Diagn, 2006. **8**(3): p. 305-11.
107. Boland, C.R. and A. Goel, *Microsatellite instability in colorectal cancer*. Gastroenterology, 2010. **138**(6): p. 2073-2087 e3.
108. Kheirelseid, E.A., et al., *Mismatch repair protein expression in colorectal cancer*. J Gastrointest Oncol, 2013. **4**(4): p. 397-408.
109. Bacher, J.W., et al., *Development of a fluorescent multiplex assay for detection of MSI-High tumors*. Dis Markers, 2004. **20**(4-5): p. 237-50.
110. Miyakura, Y., et al., *Extensive methylation of hMLH1 promoter region predominates in proximal colon cancer with microsatellite instability*. Gastroenterology, 2001. **121**(6): p. 1300-1309.
111. Jass, J.R., *Classification of colorectal cancer based on correlation of clinical, morphological and molecular features*. Histopathology, 2007. **50**(1): p. 113-30.
112. Kohonen-Corish, M.R., et al., *Low microsatellite instability is associated with poor prognosis in stage C colon cancer*. J Clin Oncol, 2005. **23**(10): p. 2318-24.
113. Tomlinson, I., et al., *Does MSI-low exist?* J Pathol, 2002. **197**(1): p. 6-13.

114. Lanza, G., et al., *Medullary-type poorly differentiated adenocarcinoma of the large bowel: a distinct clinicopathologic entity characterized by microsatellite instability and improved survival*. J Clin Oncol, 1999. **17**(8): p. 2429-38.
115. Liu, B., et al., *Analysis of mismatch repair genes in hereditary non-polyposis colorectal cancer patients*. Nat Med, 1996. **2**(2): p. 169-74.
116. Lynch, H.T., et al., *Genetics, natural history, tumor spectrum, and pathology of hereditary nonpolyposis colorectal cancer: an updated review*. Gastroenterology, 1993. **104**(5): p. 1535-49.
117. Soares-Weiser, K., et al. *Diagnostic Accuracy and Cost-effectiveness of Faecal Occult Blood Tests (FOBT) used in screening for colorectal cancer: A Systematic Review*. 2007 [cited 2014 20 January]; Available from: https://www.york.ac.uk/inst/crd/CRD_Reports/crdreport36.pdf
118. Velayos, F.S., et al., *The mechanism of microsatellite instability is different in synchronous and metachronous colorectal cancer*. Journal of Gastrointestinal Surgery, 2005. **9**(3): p. 329-335.
119. Oikawa, T., et al., *Transcriptional control of BubR1 by p53 and suppression of centrosome amplification by BubR1*. Mol Cell Biol, 2005. **25**(10): p. 4046-61.
120. Aoki, K., et al., *Chromosomal instability by beta-catenin/TCF transcription in APC or beta-catenin mutant cells*. Oncogene, 2007. **26**(24): p. 3511-20.
121. Hadjihannas, M.V., et al., *Aberrant Wnt/beta-catenin signaling can induce chromosomal instability in colon cancer*. Proc Natl Acad Sci U S A, 2006. **103**(28): p. 10747-52.
122. Fearon, E.R. and B. Vogelstein, *A genetic model for colorectal tumorigenesis*. Cell, 1990. **61**(5): p. 759-67.
123. Rao, C.V. and H.Y. Yamada, *Genomic instability and colon carcinogenesis: from the perspective of genes*. Front Oncol, 2013. **3**: p. 130.
124. Cancer Genome Atlas, N., *Comprehensive molecular characterization of human colon and rectal cancer*. Nature, 2012. **487**(7407): p. 330-7.
125. Ehrlich, M., *DNA hypomethylation in cancer cells*. Epigenomics, 2009. **1**(2): p. 239-59.
126. Tabach, Y., et al., *Amplification of the 20q chromosomal arm occurs early in tumorigenic transformation and may initiate cancer*. PLoS One, 2011. **6**(1): p. e14632.
127. Seshagiri, S., et al., *Recurrent R-spondin fusions in colon cancer*. Nature, 2012. **488**(7413): p. 660-4.
128. Pancione, M., A. Remo, and V. Colantuoni, *Genetic and epigenetic events generate multiple pathways in colorectal cancer progression*. Patholog Res Int, 2012. **2012**: p. 509348.
129. Cisowski, J., et al., *Oncogene-induced senescence underlies the mutual exclusive nature of oncogenic KRAS and BRAF*. Oncogene, 2016. **35**(10): p. 1328-33.
130. Goel, A., et al., *Characterization of sporadic colon cancer by patterns of genomic instability*. Cancer Research, 2003. **63**(7): p. 1608-1614.
131. Bedard, P.L., et al., *Tumour heterogeneity in the clinic*. Nature, 2013. **501**(7467): p. 355-64.
132. Fidler, I.J. and I.R. Hart, *Biological diversity in metastatic neoplasms: origins and implications*. Science, 1982. **217**(4564): p. 998-1003.
133. Nowell, P.C., *The clonal evolution of tumor cell populations*. Science, 1976. **194**(4260): p. 23-8.
134. Uchi, R., et al., *Integrated Multiregional Analysis Proposing a New Model of Colorectal Cancer Evolution*. PLoS Genet, 2016. **12**(2): p. e1005778.
135. Gerlinger, M., et al., *Genomic architecture and evolution of clear cell renal cell carcinomas defined by multiregion sequencing*. Nat Genet, 2014. **46**(3): p. 225-33.
136. Sottoriva, A., et al., *A Big Bang model of human colorectal tumor growth*. Nat Genet, 2015. **47**(3): p. 209-16.

137. Thirlwell, C., et al., *Clonality assessment and clonal ordering of individual neoplastic crypts shows polyclonality of colorectal adenomas*. *Gastroenterology*, 2010. **138**(4): p. 1441-54, 1454 e1-7.
138. Bardelli, A. and S. Siena, *Molecular mechanisms of resistance to cetuximab and panitumumab in colorectal cancer*. *J Clin Oncol*, 2010. **28**(7): p. 1254-61.
139. Ciardiello, F. and G. Tortora, *EGFR antagonists in cancer treatment*. *N Engl J Med*, 2008. **358**(11): p. 1160-74.
140. Misale, S., et al., *Emergence of KRAS mutations and acquired resistance to anti-EGFR therapy in colorectal cancer*. *Nature*, 2012. **486**(7404): p. 532-6.
141. Tougeron, D., et al., *Effect of low-frequency KRAS mutations on the response to anti-EGFR therapy in metastatic colorectal cancer*. *Ann Oncol*, 2013. **24**(5): p. 1267-73.
142. Perdignes, N. and M. Murtaza, *Capturing tumor heterogeneity and clonal evolution in solid cancers using circulating tumor DNA analysis*. *Pharmacol Ther*, 2017. **174**: p. 22-26.
143. Jackstadt, R. and O.J. Sansom, *Mouse models of intestinal cancer*. *J Pathol*, 2016. **238**(2): p. 141-51.
144. Shibata, H., et al., *Rapid colorectal adenoma formation initiated by conditional targeting of the Apc gene*. *Science*, 1997. **278**(5335): p. 120-3.
145. *Cre-lox Plasmid*. Special Collections [cited 2017 05/06]; Available from: <https://www.addgene.org/cre-lox/>.
146. Missirlis, P.I., D.E. Smailus, and R.A. Holt, *A high-throughput screen identifying sequence and promiscuity characteristics of the loxP spacer region in Cre-mediated recombination*. *BMC Genomics*, 2006. **7**: p. 73.
147. Sauer, B., *Inducible gene targeting in mice using the Cre/lox system*. *Methods*, 1998. **14**(4): p. 381-92.
148. Abremski, K. and R. Hoess, *Bacteriophage P1 site-specific recombination. Purification and properties of the Cre recombinase protein*. *J Biol Chem*, 1984. **259**(3): p. 1509-14.
149. Database, J.M. 2013 [cited 2013 1 December]; Available from: <http://jaxmice.jax.org/strain/008875.html>.
150. Feil, S., N. Valtcheva, and R. Feil, *Inducible Cre mice*. *Methods Mol Biol*, 2009. **530**: p. 343-63.
151. Hayashi, S. and A.P. McMahon, *Efficient recombination in diverse tissues by a tamoxifen-inducible form of Cre: a tool for temporally regulated gene activation/inactivation in the mouse*. *Dev Biol*, 2002. **244**(2): p. 305-18.
152. Metzger, D. and P. Chambon, *Site- and time-specific gene targeting in the mouse*. *Methods*, 2001. **24**(1): p. 71-80.
153. Wunderlich, F.T., *Generation of inducible Cre systems for conditional gene inactivation in mice, Doctoral Thesis, in Mathematics & Science Faculty*. 2004, University of Cologne: University of Cologne, Frangenheimstraße 4, 50931 Köln, Germany.
154. Su, L.K., et al., *Multiple Intestinal Neoplasia Caused by a Mutation in the Murine Homolog of the Apc Gene*. *Science*, 1992. **256**(5057): p. 668-670.
155. Young, M., L. Ordonez, and A.R. Clarke, *What are the best routes to effectively model human colorectal cancer?* *Molecular Oncology*, 2013. **7**(2): p. 178-189.
156. Sansom, O.J., et al., *Loss of Apc allows phenotypic manifestation of the transforming properties of an endogenous K-ras oncogene in vivo*. *Proc Natl Acad Sci U S A*, 2006. **103**(38): p. 14122-7.
157. Hung, K.E., et al., *Development of a mouse model for sporadic and metastatic colon tumors and its use in assessing drug treatment*. *Proc Natl Acad Sci U S A*, 2010. **107**(4): p. 1565-70.
158. Muller, P.A., et al., *Mutant p53 drives invasion by promoting integrin recycling*. *Cell*, 2009. **139**(7): p. 1327-41.

159. *Colorectal Cancer Survival by Stage*. [cited 2017 05/06]; Available from: http://www.ncin.org.uk/publications/data_briefings/colorectal_cancer_survival_by_age.
160. Hardcastle, J.D., et al., *Randomised controlled trial of faecal-occult-blood screening for colorectal cancer*. *The Lancet*, 1996. **348**(9040): p. 1472-1477.
161. Scholefield, J.H., et al., *Nottingham trial of faecal occult blood testing for colorectal cancer: a 20-year follow-up*. *Gut*, 2012. **61**(7): p. 1036-40.
162. Hewitson, P., et al., *Cochrane systematic review of colorectal cancer screening using the fecal occult blood test (hemoccult): an update*. *Am J Gastroenterol*, 2008. **103**(6): p. 1541-9.
163. Kronborg, O., et al., *Randomised study of screening for colorectal cancer with faecal-occult-blood test*. *The Lancet*, 1996. **348**(9040): p. 1467-1471.
164. Lindholm, E., H. Brevinge, and E. Haglund, *Survival benefit in a randomized clinical trial of faecal occult blood screening for colorectal cancer*. *Br J Surg*, 2008. **95**(8): p. 1029-36.
165. Mandel, J.S., et al., *Reducing mortality from colorectal cancer by screening for fecal occult blood. Minnesota Colon Cancer Control Study*. *N Engl J Med*, 1993. **328**(19): p. 1365-71.
166. Towler, B., et al., *A systematic review of the effects of screening for colorectal cancer using the faecal occult blood test, hemoccult*. *BMJ*, 1998. **317**(7158): p. 559-65.
167. Ait Ouakrim, D., et al., *Trends in colorectal cancer mortality in Europe: retrospective analysis of the WHO mortality database*. *BMJ*, 2015. **351**: p. h4970.
168. *Bowel cancer screening* 09/02/2015 [cited 2017 06/06]; Available from: <http://www.nhs.uk/Conditions/bowel-cancer-screening/Pages/Introduction.aspx>.
169. Logan, R.F.A., et al., *Outcomes of the Bowel Cancer Screening Programme (BCSP) in England after the first 1 million tests*. *Gut*, 2012. **61**(10): p. 1439.
170. Guthlein, W., et al., *Guaiaconic acid A from guaiac resin*. 1981, Google Patents.
171. Kratochvil, J.F., et al., *Isolation and characterization of α -guaiaconic acid and the nature of guaiacum blue*. *Phytochemistry*, 1971. **10**(10): p. 2529-2531.
172. Lee, T.J., et al., *High yield of colorectal neoplasia detected by colonoscopy following a positive faecal occult blood test in the NHS Bowel Cancer Screening Programme*. *J Med Screen*, 2011. **18**(2): p. 82-6.
173. Mackie, A. *Moving from guaiac faecal occult blood test (gFOBT) to a faecal immunochemical test for haemoglobin (FIT) in the bowel screening programme: A consultation*. 2015. **Legacy Screening Portal**.
174. Daly, J.M., C.P. Bay, and B.T. Levy, *Evaluation of fecal immunochemical tests for colorectal cancer screening*. *J Prim Care Community Health*, 2013. **4**(4): p. 245-50.
175. Levi, Z., et al., *A higher detection rate for colorectal cancer and advanced adenomatous polyp for screening with immunochemical fecal occult blood test than guaiac fecal occult blood test, despite lower compliance rate. A prospective, controlled, feasibility study*. *Int J Cancer*, 2011. **128**(10): p. 2415-24.
176. Brenner, H., et al., *Diagnostic performance of guaiac-based fecal occult blood test in routine screening: state-wide analysis from Bavaria, Germany*. *Am J Gastroenterol*, 2014. **109**(3): p. 427-35.
177. von Wagner, C., et al., *Inequalities in participation in an organized national colorectal cancer screening programme: results from the first 2.6 million invitations in England*. *Int J Epidemiol*, 2011. **40**(3): p. 712-8.
178. Moss, S., et al., *OC-002 Increased participation in colorectal cancer screening during a pilot of faecal immunochemical test for haemoglobin (FIT) in England*. *Gut*, 2015. **64**(Suppl 1): p. A1.2-A1.
179. Atkin, W.S., et al., *Once-only flexible sigmoidoscopy screening in prevention of colorectal cancer: a multicentre randomised controlled trial*. *Lancet*, 2010. **375**(9726): p. 1624-33.

180. Davis, J.S. and X. Wu, *Current state and future challenges of chemoprevention*. *Discov Med*, 2012. **13**(72): p. 385-90.
181. Tárraga López, P.J., J.S. Albero, and J.A. Rodríguez-Montes, *Primary and Secondary Prevention of Colorectal Cancer*. *Clinical Medicine Insights. Gastroenterology*, 2014. **7**: p. 33-46.
182. Chan, A.T. and E.L. Giovannucci, *Primary prevention of colorectal cancer*. *Gastroenterology*, 2010. **138**(6): p. 2029-2043 e10.
183. Wolin, K.Y., et al., *Physical activity and colon cancer prevention: a meta-analysis*. *Br J Cancer*, 2009. **100**(4): p. 611-6.
184. Liang, P.S., T.Y. Chen, and E. Giovannucci, *Cigarette smoking and colorectal cancer incidence and mortality: systematic review and meta-analysis*. *Int J Cancer*, 2009. **124**(10): p. 2406-15.
185. Norat, T., et al., *Meat, fish, and colorectal cancer risk: the European Prospective Investigation into cancer and nutrition*. *J Natl Cancer Inst*, 2005. **97**(12): p. 906-16.
186. Larsson, S.C. and A. Wolk, *Obesity and colon and rectal cancer risk: a meta-analysis of prospective studies*. *Am J Clin Nutr*, 2007. **86**(3): p. 556-65.
187. Ma, Y., et al., *Obesity and risk of colorectal cancer: a systematic review of prospective studies*. *PLoS One*, 2013. **8**(1): p. e53916.
188. Renehan, A.G., J. Frystyk, and A. Flyvbjerg, *Obesity and cancer risk: the role of the insulin-IGF axis*. *Trends Endocrinol Metab*, 2006. **17**(8): p. 328-36.
189. Gunter, M.J. and M.F. Leitzmann, *Obesity and colorectal cancer: epidemiology, mechanisms and candidate genes*. *J Nutr Biochem*, 2006. **17**(3): p. 145-56.
190. Sporn, M.B., *Approaches to prevention of epithelial cancer during the preneoplastic period*. *Cancer Res*, 1976. **36**(7 PT 2): p. 2699-702.
191. Steward, W.P. and K. Brown, *Cancer chemoprevention: a rapidly evolving field*. *Br J Cancer*, 2013. **109**(1): p. 1-7.
192. Tsao, A.S., E.S. Kim, and W.K. Hong, *Chemoprevention of cancer*. *CA Cancer J Clin*, 2004. **54**(3): p. 150-80.
193. Wu, X., S. Patterson, and E. Hawk, *Chemoprevention--history and general principles*. *Best Pract Res Clin Gastroenterol*, 2011. **25**(4-5): p. 445-59.
194. Fuster, V. and J.M. Sweeny, *Aspirin: a historical and contemporary therapeutic overview*. *Circulation*, 2011. **123**(7): p. 768-78.
195. Burn, J., J.C. Mathers, and D.T. Bishop, *Chemoprevention in Lynch syndrome*. *Fam Cancer*, 2013. **12**(4): p. 707-18.
196. Rothwell, P.M., et al., *Short-term effects of daily aspirin on cancer incidence, mortality, and non-vascular death: analysis of the time course of risks and benefits in 51 randomised controlled trials*. *The Lancet*, 2012. **379**(9826): p. 1602-1612.
197. Rothwell, P.M., et al., *Long-term effect of aspirin on colorectal cancer incidence and mortality: 20-year follow-up of five randomised trials*. *The Lancet*, 2010. **376**(9754): p. 1741-1750.
198. Burn, J., et al., *A randomized placebo-controlled prevention trial of aspirin and/or resistant starch in young people with familial adenomatous polyposis*. *Cancer Prev Res (Phila)*, 2011. **4**(5): p. 655-65.
199. Burn, J., et al., *Effect of aspirin or resistant starch on colorectal neoplasia in the Lynch syndrome*. *N Engl J Med*, 2008. **359**(24): p. 2567-78.
200. Burn, J., et al., *Diet and cancer prevention: the concerted action polyp prevention (CAPP) studies*. *Proc Nutr Soc*, 1998. **57**(2): p. 183-6.
201. Burn, J., et al., *Long-term effect of aspirin on cancer risk in carriers of hereditary colorectal cancer: an analysis from the CAPP2 randomised controlled trial*. *Lancet*, 2011. **378**(9809): p. 2081-7.

202. Chan, A.T., et al., *Long-term aspirin use and mortality in women*. Archives of Internal Medicine, 2007. **167**(6): p. 562-572.
203. Flossmann, E. and P.M. Rothwell, *Effect of aspirin on long-term risk of colorectal cancer: consistent evidence from randomised and observational studies*. The Lancet, 2007. **369**(9573): p. 1603-1613.
204. Garcia-Albeniz, X. and A.T. Chan, *Aspirin for the prevention of colorectal cancer*. Best Practice & Research in Clinical Gastroenterology, 2011. **25**(4-5): p. 461-472.
205. Logan, R.F., et al., *Aspirin and folic acid for the prevention of recurrent colorectal adenomas*. Gastroenterology, 2008. **134**(1): p. 29-38.
206. Nishihara, R., et al., *Aspirin use and risk of colorectal cancer according to BRAF mutation status*. JAMA, 2013. **309**(24): p. 2563-71.
207. Chell, S., et al., *Prospects in NSAID-derived chemoprevention of colorectal cancer*. Biochem Soc Trans, 2005. **33**(Pt 4): p. 667-71.
208. Steinbach, G., et al., *The effect of celecoxib, a cyclooxygenase-2 inhibitor, in familial adenomatous polyposis*. N Engl J Med, 2000. **342**(26): p. 1946-52.
209. Bertagnolli, M.M., et al., *Celecoxib for the prevention of sporadic colorectal adenomas*. N Engl J Med, 2006. **355**(9): p. 873-84.
210. Solomon, S.D., et al., *Cardiovascular risk associated with celecoxib in a clinical trial for colorectal adenoma prevention*. N Engl J Med, 2005. **352**(11): p. 1071-80.
211. Solomon, S.D., et al., *Effect of celecoxib on cardiovascular events and blood pressure in two trials for the prevention of colorectal adenomas*. Circulation, 2006. **114**(10): p. 1028-35.
212. Pfizer, Inc.; *Withdrawal of Approval of Familial Adenomatous Polyposis Indication for CELEBREX*, F.a.D. Administration, Editor. 2012.
213. Patel, V.B., et al., *Colorectal cancer: chemopreventive role of curcumin and resveratrol*. Nutr Cancer, 2010. **62**(7): p. 958-67.
214. Gururaj, A.E., et al., *Molecular mechanisms of anti-angiogenic effect of curcumin*. Biochem Biophys Res Commun, 2002. **297**(4): p. 934-42.
215. Shishodia, S., et al., *Curcumin (diferuloylmethane) inhibits constitutive NF-kappaB activation, induces G1/S arrest, suppresses proliferation, and induces apoptosis in mantle cell lymphoma*. Biochem Pharmacol, 2005. **70**(5): p. 700-13.
216. Garcea, G., et al., *Consumption of the putative chemopreventive agent curcumin by cancer patients: assessment of curcumin levels in the colorectum and their pharmacodynamic consequences*. Cancer Epidemiol Biomarkers Prev, 2005. **14**(1): p. 120-5.
217. Cheng, A.L., et al., *Phase I clinical trial of curcumin, a chemopreventive agent, in patients with high-risk or pre-malignant lesions*. Anticancer Res, 2001. **21**(4B): p. 2895-900.
218. Lao, C.D., et al., *Dose escalation of a curcuminoid formulation*. BMC Complement Altern Med, 2006. **6**: p. 10.
219. Mahale, J., *Preclinical Evaluation of Meriva (a formulation of curcumin) as a putative chemopreventive agent for Lung Cancer*, in *Department of Cancer Studies*. 2016, University of Leicester.
220. Gescher, A., W.P. Steward, and K. Brown, *Resveratrol in the management of human cancer: how strong is the clinical evidence?* Resveratrol and Health, 2013. **1290**: p. 12-20.
221. Patel, K.R., et al., *Clinical pharmacology of resveratrol and its metabolites in colorectal cancer patients*. Cancer Res, 2010. **70**(19): p. 7392-9.
222. Sale, S., et al., *Comparison of the effects of the chemopreventive agent resveratrol and its synthetic analog trans 3,4,5,4'-tetramethoxystilbene (DMU-212) on adenoma development in the Apc(Min+) mouse and cyclooxygenase-2 in human-derived colon cancer cells*. Int J Cancer, 2005. **115**(2): p. 194-201.

223. Baur, J.A. and D.A. Sinclair, *Therapeutic potential of resveratrol: the in vivo evidence*. Nat Rev Drug Discov, 2006. **5**(6): p. 493-506.
224. Crowell, J.A., et al., *Resveratrol-associated renal toxicity*. Toxicol Sci, 2004. **82**(2): p. 614-9.
225. Scott, E., et al., *Less is more for cancer chemoprevention: evidence of a non-linear dose response for the protective effects of resveratrol in humans and mice*. Science translational medicine, 2015. **7**(298): p. 298ra117-298ra117.
226. Boocock, D.J., et al., *Phase I dose escalation pharmacokinetic study in healthy volunteers of resveratrol, a potential cancer chemopreventive agent*. Cancer Epidemiol Biomarkers Prev, 2007. **16**(6): p. 1246-52.
227. Brown, V.A., et al., *Repeat dose study of the cancer chemopreventive agent resveratrol in healthy volunteers: safety, pharmacokinetics, and effect on the insulin-like growth factor axis*. Cancer Res, 2010. **70**(22): p. 9003-11.
228. Baur, J.A., et al., *Resveratrol improves health and survival of mice on a high-calorie diet*. Nature, 2006. **444**(7117): p. 337-42.
229. Kimmelman, A.C., *The dynamic nature of autophagy in cancer*. Genes Dev, 2011. **25**(19): p. 1999-2010.
230. Pezzuto, J.M., *Resveratrol as an Inhibitor of Carcinogenesis*. Pharmaceutical Biology, 2008. **46**(7-8): p. 443-573.
231. Ren, Z., et al., *Resveratrol inhibits NF-kB signaling through suppression of p65 and IkkappaB kinase activities*. Pharmazie, 2013. **68**(8): p. 689-94.
232. De Gruttola, V.G., et al., *Considerations in the evaluation of surrogate endpoints in clinical trials. summary of a National Institutes of Health workshop*. Control Clin Trials, 2001. **22**(5): p. 485-502.
233. Langan, R.C., et al., *Colorectal cancer biomarkers and the potential role of cancer stem cells*. J Cancer, 2013. **4**(3): p. 241-50.
234. Lech, G., et al., *Colorectal cancer tumour markers and biomarkers: Recent therapeutic advances*. World J Gastroenterol, 2016. **22**(5): p. 1745-55.
235. Shaw, A., et al., *Tumour biomarkers: diagnostic, prognostic, and predictive*. BMJ, 2015. **351**: p. h3449.
236. De Roock, W., et al., *Effects of KRAS, BRAF, NRAS, and PIK3CA mutations on the efficacy of cetuximab plus chemotherapy in chemotherapy-refractory metastatic colorectal cancer: a retrospective consortium analysis*. Lancet Oncol, 2010. **11**(8): p. 753-62.
237. Kopetz, S., et al., *Phase II Pilot Study of Vemurafenib in Patients With Metastatic BRAF-Mutated Colorectal Cancer*. J Clin Oncol, 2015. **33**(34): p. 4032-8.
238. Prahallad, A., et al., *Unresponsiveness of colon cancer to BRAF(V600E) inhibition through feedback activation of EGFR*. Nature, 2012. **483**(7387): p. 100-3.
239. Samowitz, W.S., et al., *Poor survival associated with the BRAF V600E mutation in microsatellite-stable colon cancers*. Cancer Res, 2005. **65**(14): p. 6063-9.
240. Locker, G.Y., et al., *ASCO 2006 update of recommendations for the use of tumor markers in gastrointestinal cancer*. J Clin Oncol, 2006. **24**(33): p. 5313-27.
241. Blake, K.E., et al., *Clinical significance of the preoperative plasma carcinoembryonic antigen (CEA) level in patients with carcinoma of the large bowel*. Dis Colon Rectum, 1982. **25**(1): p. 24-32.
242. Doos, W.G., et al., *CEA levels in patients with colorectal polyps*. Cancer, 1975. **36**(6): p. 1996-2003.
243. Filella, X., et al., *Prognostic value of CA 19.9 levels in colorectal cancer*. Annals of Surgery, 1992. **216**(1): p. 55-59.
244. Nicolini, A., et al., *Intensive risk-adjusted follow-up with the cea, tpa, ca19.9, and ca72.4 tumor marker panel and abdominal ultrasonography to diagnose operable colorectal cancer recurrences: Effect on survival*. Archives of Surgery, 2010. **145**(12): p. 1177-1183.

245. B. Johne, O.K.H.I.T.J., *A New Fecal Calprotectin Test for Colorectal Neoplasia: Clinical Results and Comparison with Previous Method*. Scandinavian Journal of Gastroenterology, 2009. **36**(3): p. 291-296.
246. Davies, R.J., R. Miller, and N. Coleman, *Colorectal cancer screening: prospects for molecular stool analysis*. Nature Reviews Cancer, 2005. **5**(3): p. 199-209.
247. Turvill, J., et al., *Faecal calprotectin in patients with suspected colorectal cancer: a diagnostic accuracy study*. Br J Gen Pract, 2016. **66**(648): p. e499-506.
248. Crowley, E., et al., *Liquid biopsy: monitoring cancer-genetics in the blood*. Nat Rev Clin Oncol, 2013. **10**(8): p. 472-84.
249. Mandel, P. and P. Metais, *Les acides nucleiques du plasma sanguine chez l'homme [in French]*. C R Seances Soc Biol Fil, 1948. **142**: p. 241-243.
250. Sorenson, G.D., et al., *Soluble normal and mutated DNA sequences from single-copy genes in human blood*. Cancer Epidemiol Biomarkers Prev, 1994. **3**(1): p. 67-71.
251. Vasioukhin, V., et al., *Point mutations of the N-ras gene in the blood plasma DNA of patients with myelodysplastic syndrome or acute myelogenous leukaemia*. Br J Haematol, 1994. **86**(4): p. 774-9.
252. Perakis, S., et al., *Advances in Circulating Tumor DNA Analysis*. Adv Clin Chem, 2017. **80**: p. 73-153.
253. Volik, S., et al., *Cell-free DNA (cfDNA): Clinical Significance and Utility in Cancer Shaped By Emerging Technologies*. Mol Cancer Res, 2016. **14**(10): p. 898-908.
254. Anker, P. and M. Stroun, *Circulating DNA in plasma or serum*. Medicina (B Aires), 2000. **60**(5 Pt 2): p. 699-702.
255. Diaz, L.A., Jr. and A. Bardelli, *Liquid biopsies: genotyping circulating tumor DNA*. J Clin Oncol, 2014. **32**(6): p. 579-86.
256. Gormally, E., et al., *Circulating free DNA in plasma or serum as biomarker of carcinogenesis: Practical aspects and biological significance*. Mutation Research-Reviews in Mutation Research, 2007. **635**(2-3): p. 105-117.
257. Hodgson, D.R., et al., *Circulating tumour-derived predictive biomarkers in oncology*. Drug Discovery Today, 2010. **15**(3-4): p. 98-101.
258. Pantel, K. and C. Alix-Panabieres, *Real-time Liquid Biopsy in Cancer Patients: Fact or Fiction?* Cancer Research, 2013. **73**(21): p. 6384-6388.
259. *FDA approves first blood test to detect gene mutation associated with non-small cell lung cancer*. 2016 01/06/2016 [cited 2017 06/06]; Available from: <https://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm504488.htm>.
260. Diehl, F., et al., *Circulating mutant DNA to assess tumor dynamics*. Nat Med, 2008. **14**(9): p. 985-90.
261. Giacona, M.B., et al., *Cell-free DNA in human blood plasma: length measurements in patients with pancreatic cancer and healthy controls*. Pancreas, 1998. **17**(1): p. 89-97.
262. Brinkmann, V., et al., *Neutrophil extracellular traps kill bacteria*. Science, 2004. **303**(5663): p. 1532-5.
263. Breitbach, S., S. Tug, and P. Simon, *Circulating cell-free DNA: an up-coming molecular marker in exercise physiology*. Sports Med, 2012. **42**(7): p. 565-86.
264. Tug, S., et al., *Exercise-induced increases in cell free DNA in human plasma originate predominantly from cells of the haematopoietic lineage*. Exerc Immunol Rev, 2015. **21**: p. 164-73.
265. Tug, S., et al., *Correlation between cell free DNA levels and medical evaluation of disease progression in systemic lupus erythematosus patients*. Cell Immunol, 2014. **292**(1-2): p. 32-9.
266. Haghiaç, M., et al., *Increased death of adipose cells, a path to release cell-free DNA into systemic circulation of obese women*. Obesity (Silver Spring), 2012. **20**(11): p. 2213-9.

267. Nishimoto, S., et al., *Obesity-induced DNA released from adipocytes stimulates chronic adipose tissue inflammation and insulin resistance*. *Sci Adv*, 2016. **2**(3): p. e1501332.
268. Mouliere, F. and A.R. Thierry, *The importance of examining the proportion of circulating DNA originating from tumor, microenvironment and normal cells in colorectal cancer patients*. *Expert Opin Biol Ther*, 2012. **12 Suppl 1**: p. S209-15.
269. Perkins, G., et al., *Multi-Purpose Utility of Circulating Plasma DNA Testing in Patients with Advanced Cancers*. *Plos One*, 2012. **7**(11).
270. Schwarzenbach, H., et al., *Detection and monitoring of cell-free DNA in blood of patients with colorectal cancer*. *Ann N Y Acad Sci*, 2008. **1137**: p. 190-6.
271. Stroun, M., et al., *About the possible origin and mechanism of circulating DNA - Apoptosis and active DNA release*. *Clinica Chimica Acta*, 2001. **313**(1-2): p. 139-142.
272. Schwarzenbach, H., D.S. Hoon, and K. Pantel, *Cell-free nucleic acids as biomarkers in cancer patients*. *Nat Rev Cancer*, 2011. **11**(6): p. 426-37.
273. Ward, T.H., et al., *Biomarkers of apoptosis*. *Br J Cancer*, 2008. **99**(6): p. 841-6.
274. Evan, G.I. and K.H. Vousden, *Proliferation, cell cycle and apoptosis in cancer*. *Nature*, 2001. **411**(6835): p. 342-8.
275. Madhavan, D., et al., *Plasma DNA integrity as a biomarker for primary and metastatic breast cancer and potential marker for early diagnosis*. *Breast Cancer Res Treat*, 2014. **146**(1): p. 163-74.
276. Mouliere, F., et al., *High fragmentation characterizes tumour-derived circulating DNA*. *PLoS One*, 2011. **6**(9): p. e23418.
277. Elshimali, Y.I., et al., *The clinical utilization of circulating cell free DNA (CCFDNA) in blood of cancer patients*. *Int J Mol Sci*, 2013. **14**(9): p. 18925-58.
278. Jahr, S., et al., *DNA fragments in the blood plasma of cancer patients: quantitations and evidence for their origin from apoptotic and necrotic cells*. *Cancer Res*, 2001. **61**(4): p. 1659-65.
279. Anker, P., M. Stroun, and P.A. Maurice, *Spontaneous release of DNA by human blood lymphocytes as shown in an in vitro system*. *Cancer Res*, 1975. **35**(9): p. 2375-82.
280. Bronkhorst, A.J., J. Aucamp, and P.J. Pretorius, *Cell-free DNA: Preanalytical variables*. *Clin Chim Acta*, 2015. **450**: p. 243-53.
281. Savill, J.S., et al., *Macrophage phagocytosis of aging neutrophils in inflammation. Programmed cell death in the neutrophil leads to its recognition by macrophages*. *J Clin Invest*, 1989. **83**(3): p. 865-75.
282. Chan, K.C., et al., *Effects of preanalytical factors on the molecular size of cell-free DNA in blood*. *Clin Chem*, 2005. **51**(4): p. 781-4.
283. Lee, T.H., et al., *Quantitation of genomic DNA in plasma and serum samples: higher concentrations of genomic DNA found in serum than in plasma*. *Transfusion*, 2001. **41**(2): p. 276-82.
284. Xue, X., et al., *Optimizing the yield and utility of circulating cell-free DNA from plasma and serum*. *Clin Chim Acta*, 2009. **404**(2): p. 100-4.
285. El Messaoudi, S., et al., *Circulating cell free DNA: Preanalytical considerations*. *Clin Chim Acta*, 2013. **424**: p. 222-30.
286. Jung, M., et al., *Changes in concentration of DNA in serum and plasma during storage of blood samples*. *Clin Chem*, 2003. **49**(6 Pt 1): p. 1028-9.
287. Thierry, A.R., et al., *Origin and quantification of circulating DNA in mice with human colorectal cancer xenografts*. *Nucleic Acids Res*, 2010. **38**(18): p. 6159-75.
288. Barra, G.B., et al., *EDTA-mediated inhibition of DNases protects circulating cell-free DNA from ex vivo degradation in blood samples*. *Clin Biochem*, 2015.
289. Lam, N.Y., et al., *EDTA is a better anticoagulant than heparin or citrate for delayed blood processing for plasma DNA analysis*. *Clin Chem*, 2004. **50**(1): p. 256-7.

290. Page, K., et al., *Influence of plasma processing on recovery and analysis of circulating nucleic acids*. PLoS One, 2013. **8**(10): p. e77963.
291. Frattini, M., et al., *Reproducibility of a semiquantitative measurement of circulating DNA in plasma from neoplastic patients*. J Clin Oncol, 2005. **23**(13): p. 3163-4; author reply 3164-5.
292. Kopreski, M.S., et al., *Detection of mutant K-ras DNA in plasma or serum of patients with colorectal cancer*. Br J Cancer, 1997. **76**(10): p. 1293-9.
293. Page, K., et al., *The importance of careful blood processing in isolation of cell-free DNA*. Ann N Y Acad Sci, 2006. **1075**: p. 313-7.
294. Sozzi, G., et al., *Effects of prolonged storage of whole plasma or isolated plasma DNA on the results of circulating DNA quantification assays*. J Natl Cancer Inst, 2005. **97**(24): p. 1848-50.
295. Tie, J., et al., *Circulating tumor DNA analysis detects minimal residual disease and predicts recurrence in patients with stage II colon cancer*. Sci Transl Med, 2016. **8**(346): p. 346ra92.
296. Breitbach, S., et al., *Direct quantification of cell-free, circulating DNA from unpurified plasma*. PLoS One, 2014. **9**(3): p. e87838.
297. Forshew, T., et al., *Noninvasive identification and monitoring of cancer mutations by targeted deep sequencing of plasma DNA*. Sci Transl Med, 2012. **4**(136): p. 136ra68.
298. Reinert, T., et al., *Analysis of circulating tumour DNA to monitor disease burden following colorectal cancer surgery*. Gut, 2016. **65**(4): p. 625-34.
299. Nagai, Y., et al., *LINE-1 hypomethylation status of circulating cell-free DNA in plasma as a biomarker for colorectal cancer*. Oncotarget, 2017. **8**(7): p. 11906-11916.
300. *Cell-free DNA Consensus Meeting: Meeting Report*, in *Experimental Cancer Medicine Centres (ECMC) Network and National Cancer Research Institute (NCRI) Biomarkers & Imaging Clinical Studies Group*. 2014.
301. Newman, A.M., et al., *An ultrasensitive method for quantitating circulating tumor DNA with broad patient coverage*. Nat Med, 2014. **20**(5): p. 548-54.
302. Wan, J.C., et al., *Liquid biopsies come of age: towards implementation of circulating tumour DNA*. Nat Rev Cancer, 2017. **17**(4): p. 223-238.
303. Murtaza, M., et al., *Non-invasive analysis of acquired resistance to cancer therapy by sequencing of plasma DNA*. Nature, 2013. **497**(7447): p. 108-12.
304. Benesova, L., et al., *Mutation-based detection and monitoring of cell-free tumor DNA in peripheral blood of cancer patients*. Anal Biochem, 2013. **433**(2): p. 227-34.
305. Beaver, J.A., et al., *Detection of cancer DNA in plasma of patients with early-stage breast cancer*. Clin Cancer Res, 2014. **20**(10): p. 2643-50.
306. Chen, D., et al., *High-sensitivity PCR method for detecting BRAF V600E mutations in metastatic colorectal cancer using LNA/DNA chimeras to block wild-type alleles*. Anal Bioanal Chem, 2014. **406**(9-10): p. 2477-87.
307. Moulriere, F., et al., *Circulating Cell-Free DNA from Colorectal Cancer Patients May Reveal High KRAS or BRAF Mutation Load*. Translational Oncology, 2013. **6**(3): p. 319-328.
308. Sanmamed, M.F., et al., *Quantitative cell-free circulating BRAFV600E mutation analysis by use of droplet digital PCR in the follow-up of patients with melanoma being treated with BRAF inhibitors*. Clin Chem, 2015. **61**(1): p. 297-304.
309. Lefebure, B., et al., *Prognostic value of circulating mutant DNA in unresectable metastatic colorectal cancer*. Ann Surg, 2010. **251**(2): p. 275-80.
310. Spindler, K.L., et al., *Quantitative cell-free DNA, KRAS, and BRAF mutations in plasma from patients with metastatic colorectal cancer during treatment with cetuximab and irinotecan*. Clin Cancer Res, 2012. **18**(4): p. 1177-85.

311. Newton, K.F., W. Newman, and J. Hill, *Review of biomarkers in colorectal cancer*. *Colorectal Dis*, 2012. **14**(1): p. 3-17.
312. Bettegowda, C., et al., *Detection of circulating tumor DNA in early- and late-stage human malignancies*. *Sci Transl Med*, 2014. **6**(224): p. 224ra24.
313. Kopreski, M.S., et al., *Somatic mutation screening: identification of individuals harboring K-ras mutations with the use of plasma DNA*. *J Natl Cancer Inst*, 2000. **92**(11): p. 918-23.
314. Galanopoulos, M., et al., *Comparative Study of Mutations in Single Nucleotide Polymorphism Loci of KRAS and BRAF Genes in Patients Who Underwent Screening Colonoscopy, With and Without Premalignant Intestinal Polyps*. *Anticancer Res*, 2017. **37**(2): p. 651-657.
315. Berger, A.W., et al., *Treatment monitoring in metastatic colorectal cancer patients by quantification and KRAS genotyping of circulating cell-free DNA*. *PLoS One*, 2017. **12**(3): p. e0174308.
316. Diehl, F., et al., *Detection and quantification of mutations in the plasma of patients with colorectal tumors*. *Proc Natl Acad Sci U S A*, 2005. **102**(45): p. 16368-73.
317. Diehl, F., et al., *Analysis of mutations in DNA isolated from plasma and stool of colorectal cancer patients*. *Gastroenterology*, 2008. **135**(2): p. 489-98.
318. Murtaza, M., et al., *Multifocal clonal evolution characterized using circulating tumour DNA in a case of metastatic breast cancer*. *Nat Commun*, 2015. **6**: p. 8760.
319. Mead, R., et al., *Circulating tumour markers can define patients with normal colons, benign polyps, and cancers*. *Br J Cancer*, 2011. **105**(2): p. 239-45.
320. Park, J.L., et al., *Quantitative analysis of cell-free DNA in the plasma of gastric cancer patients*. *Oncol Lett*, 2012. **3**(4): p. 921-926.
321. Shapiro, B., et al., *Determination of circulating DNA levels in patients with benign or malignant gastrointestinal disease*. *Cancer*, 1983. **51**(11): p. 2116-20.
322. Wilhelm, J. and A. Pingoud, *Real-time polymerase chain reaction*. *Chembiochem*, 2003. **4**(11): p. 1120-8.
323. *Applied Biosystems StepOne™ and StepOnePlus™ Real-Time PCR Systems - Reagent Guide*. 2010, Applied Biosystems.
324. Svec, D., et al., *How good is a PCR efficiency estimate: Recommendations for precise and robust qPCR efficiency assessments*. *Biomol Detect Quantif*, 2015. **3**: p. 9-16.
325. Porterfield, A., *What Is a Ct Value?* BitesizeBio.
326. *Real-Time PCR Vs. Traditional PCR*.
327. Kutyavin, I.V., et al., *3'-minor groove binder-DNA probes increase sequence specificity at PCR extension temperatures*. *Nucleic Acids Res*, 2000. **28**(2): p. 655-61.
328. Holland, P.M., et al., *Detection of specific polymerase chain reaction product by utilizing the 5'----3' exonuclease activity of *Thermus aquaticus* DNA polymerase*. *Proceedings of the National Academy of Sciences of the United States of America*, 1991. **88**(16): p. 7276-7280.
329. Livak, K.J., et al., *Oligonucleotides with fluorescent dyes at opposite ends provide a quenched probe system useful for detecting PCR product and nucleic acid hybridization*. *PCR Methods Appl*, 1995. **4**(6): p. 357-62.
330. *VIC fluorescent dye (ABI)*. [cited 2017 06/06]; Available from: http://www.genelink.com/newsite/products/mod_detail.asp?modid=63.
331. *FAM (6-fluorescein amidite (6-FAM))*. [cited 2017 06/06]; Available from: http://www.genelink.com/newsite/products/mod_detail.asp?modid=33.
332. Cheng, J., Y. Zhang, and Q. Li, *Real-time PCR genotyping using displacing probes*. *Nucleic Acids Research*, 2004. **32**(7): p. e61-e61.
333. Paulasova, P. and F. Pellestor, *The peptide nucleic acids (PNAs): a new generation of probes for genetic and cytogenetic analyses*. *Ann Genet*, 2004. **47**(4): p. 349-58.

334. Ray, A. and B. Norden, *Peptide nucleic acid (PNA): its medical and biotechnical applications and promise for the future*. *Faseb Journal*, 2000. **14**(9): p. 1041-1060.
335. Orum, H., et al., *Single base pair mutation analysis by PNA directed PCR clamping*. *Nucleic Acids Res*, 1993. **21**(23): p. 5332-6.
336. Thiede, C., et al., *Simple and sensitive detection of mutations in the ras proto-oncogenes using PNA-mediated PCR clamping*. *Nucleic Acids Research*, 1996. **24**(5): p. 983-984.
337. Diehl, F. and L.A. Diaz, Jr., *Digital quantification of mutant DNA in cancer patients*. *Curr Opin Oncol*, 2007. **19**(1): p. 36-42.
338. Zhang, B.O., et al., *Comparison of droplet digital PCR and conventional quantitative PCR for measuring EGFR gene mutation*. *Exp Ther Med*, 2015. **9**(4): p. 1383-1388.
339. *Droplet Digital PCR Applications Guide*. BioRad.
340. Pinheiro, L.B., et al., *Evaluation of a droplet digital polymerase chain reaction format for DNA copy number quantification*. *Anal Chem*, 2012. **84**(2): p. 1003-11.
341. Sanders, R., et al., *Evaluation of digital PCR for absolute DNA quantification*. *Anal Chem*, 2011. **83**(17): p. 6474-84.
342. HEX (Hexachloro-Fluorescein). [cited 2017 06/06]; Available from: http://www.genelink.com/newsite/products/mod_detail.asp?modid=34.
343. Baker, M., *Digital PCR hits its stride*. *Nature Methods*, 2012. **9**(6): p. 541-544.
344. Hayden, R.T., et al., *Comparison of droplet digital PCR to real-time PCR for quantitative detection of cytomegalovirus*. *J Clin Microbiol*, 2013. **51**(2): p. 540-6.
345. Hindson, C.M., et al., *Absolute quantification by droplet digital PCR versus analog real-time PCR*. *Nat Methods*, 2013. **10**(10): p. 1003-5.
346. Pohl, G. and M. Shih le, *Principle and applications of digital PCR*. *Expert Rev Mol Diagn*, 2004. **4**(1): p. 41-7.
347. Iserte, J.A., et al., *Family-specific degenerate primer design: a tool to design consensus degenerated oligonucleotides*. *Biotechnol Res Int*, 2013. **2013**: p. 383646.
348. *Rare Mutation Detection Best Practices Guidelines*. BioRad.
349. BioLabs®. *Products: Hincll*. 2014 [cited 2014 17 January]; Available from: <https://www.neb.com/products/r0103-hincii>.
350. Addgene. *Addgene: Plasmid 13442*. [cited 2014 17 January]; Available from: <http://www.addgene.org/13442/>.
351. Soriano, P., G. Friedrich, and P. Lawinger, *Promoter interactions in retrovirus vectors introduced into fibroblasts and embryonic stem cells*. *J Virol*, 1991. **65**(5): p. 2314-9.
352. BioLabs®. *Products: SacI*. 2014 [cited 2014 17 January]; Available from: <https://www.neb.com/products/r0156-saci>
353. Trypsteen, W., et al., *ddpcRquant: threshold determination for single channel droplet digital PCR experiments*. *Anal Bioanal Chem*, 2015. **407**(19): p. 5827-34.
354. Jones, M., et al., *Low copy target detection by Droplet Digital PCR through application of a novel open access bioinformatic pipeline, 'definetherain'*. *J Virol Methods*, 2014. **202**: p. 46-53.
355. Czeiger, D., et al., *Measurement of circulating cell-free DNA levels by a new simple fluorescent test in patients with primary colorectal cancer*. *Am J Clin Pathol*, 2011. **135**(2): p. 264-70.
356. Baltgalvis, K.A., et al., *The interaction of a high-fat diet and regular moderate intensity exercise on intestinal polyp development in Apc Min/+ mice*. *Cancer Prev Res (Phila)*, 2009. **2**(7): p. 641-9.
357. Beyaz, S., et al., *High-fat diet enhances stemness and tumorigenicity of intestinal progenitors*. *Nature*, 2016. **531**(7592): p. 53-8.
358. Martinez-Useros, J. and J. Garcia-Foncillas, *Obesity and colorectal cancer: molecular features of adipose tissue*. *J Transl Med*, 2016. **14**: p. 21.

359. Ahlquist, D.A., et al., *Stool DNA and occult blood testing for screen detection of colorectal neoplasia*. *Ann Intern Med*, 2008. **149**(7): p. 441-50, W81.
360. Ahlquist, D.A., et al., *Next-generation stool DNA test accurately detects colorectal cancer and large adenomas*. *Gastroenterology*, 2012. **142**(2): p. 248-56; quiz e25-6.
361. Dhaliwal, A., et al., *Fecal DNA testing for colorectal cancer screening: Molecular targets and perspectives*. *World J Gastrointest Oncol*, 2015. **7**(10): p. 178-83.
362. Kadiyska, T. and A. Nossikoff, *Stool DNA methylation assays in colorectal cancer screening*. *World J Gastroenterol*, 2015. **21**(35): p. 10057-61.
363. Pettan-Brewer, C., et al., *B16 melanoma tumor growth is delayed in mice in an age-dependent manner*. *Pathobiol Aging Age Relat Dis*, 2012. **2**.
364. Xiao, S., et al., *Pathology of Small-Bowel Tumours*, in *Cancer of the Upper Gastrointestinal Tract*. 2002, American Cancer Society, BC Decker Inc, Hamilton, Ontario, USA. . p. 343-358.
365. van der Heijden, M., et al., *Bcl-2 is a critical mediator of intestinal transformation*. *Nat Commun*, 2016. **7**: p. 10916.
366. van der Vaart, M. and P.J. Pretorius, *Circulating DNA. Its origin and fluctuation*. *Ann N Y Acad Sci*, 2008. **1137**: p. 18-26.
367. *Introduction To Real-Time Quantitative PCR (qPCR)*. [cited 2017 08/06]; Available from: <http://sabiosciences.com/manuals/IntrotoqPCR.pdf>.
368. Kim, T.H., S. Escudero, and R.A. Shivdasani, *Intact function of Lgr5 receptor-expressing intestinal stem cells in the absence of Paneth cells*. *Proc Natl Acad Sci U S A*, 2012. **109**(10): p. 3932-7.
369. Lopez-Garcia, C., et al., *Intestinal stem cell replacement follows a pattern of neutral drift*. *Science*, 2010. **330**(6005): p. 822-5.
370. Kozuka, S., et al., *Premalignancy of the mucosal polyp in the large intestine: II. Estimation of the periods required for malignant transformation of mucosal polyps*. *Dis Colon Rectum*, 1975. **18**(6): p. 494-500.
371. DeClercq, V., D.N. McMurray, and R.S. Chapkin, *Obesity promotes colonic stem cell expansion during cancer initiation*. *Cancer Lett*, 2015. **369**(2): p. 336-43.
372. Tie, J., et al., *Circulating tumor DNA as an early marker of therapeutic response in patients with metastatic colorectal cancer*. *Ann Oncol*, 2015. **26**(8): p. 1715-22.
373. Clarke, C.N. and E.S. Kopetz, *BRAF mutant colorectal cancer as a distinct subset of colorectal cancer: clinical characteristics, clinical behavior, and response to targeted therapies*. *J Gastrointest Oncol*, 2015. **6**(6): p. 660-7.
374. Rajagopalan, H., et al., *Tumorigenesis: RAF/RAS oncogenes and mismatch-repair status*. *Nature*, 2002. **418**(6901): p. 934.
375. Lee, Y.W., *Peptide nucleic acid clamp polymerase chain reaction reveals a deletion mutation of the BRAF gene in papillary thyroid carcinoma: A case report*. *Exp Ther Med*, 2013. **6**(6): p. 1550-1552.
376. Luo, J.D., et al., *Detection of rare mutant K-ras DNA in a single-tube reaction using peptide nucleic acid as both PCR clamp and sensor probe*. *Nucleic Acids Res*, 2006. **34**(2): p. e12.
377. Fransen, K., et al., *Mutation analysis of the BRAF, ARAF and RAF-1 genes in human colorectal adenocarcinomas*. *Carcinogenesis*, 2004. **25**(4): p. 527-33.
378. Imamura, Y., et al., *Specific mutations in KRAS codons 12 and 13, and patient prognosis in 1075 BRAF wild-type colorectal cancers*. *Clin Cancer Res*, 2012. **18**(17): p. 4753-63.
379. Gardner, S.N., et al., *Limitations of TaqMan PCR for Detecting Divergent Viral Pathogens Illustrated by Hepatitis A, B, C, and E Viruses and Human Immunodeficiency Virus*. *Journal of Clinical Microbiology*, 2003. **41**(6): p. 2417-2427.

380. Muradrasoli, S., et al., *Broadly targeted multiprobe QPCR for detection of coronaviruses: Coronavirus is common among mallard ducks (Anas platyrhynchos)*. J Virol Methods, 2009. **159**(2): p. 277-87.
381. Spindler, K.L., et al., *Circulating free DNA as biomarker and source for mutation detection in metastatic colorectal cancer*. PLoS One, 2015. **10**(4): p. e0108247.
382. Lang, A.H., et al., *Optimized allele-specific real-time PCR assays for the detection of common mutations in KRAS and BRAF*. J Mol Diagn, 2011. **13**(1): p. 23-8.
383. Lin, J.K., et al., *Clinical relevance of alterations in quantity and quality of plasma DNA in colorectal cancer patients: based on the mutation spectra detected in primary tumors*. Ann Surg Oncol, 2014. **21 Suppl 4**: p. S680-6.
384. Hostein, I., et al., *The detection of double mutations in KRAS depends on the mutation-detection assay used*. Clin Chem, 2011. **57**(7): p. 1077-9.
385. Macedo, M.P., et al., *Multiple mutations in the Kras gene in colorectal cancer: review of the literature with two case reports*. Int J Colorectal Dis, 2011. **26**(10): p. 1241-8.
386. Perrone, F., et al., *Circulating free DNA in a screening program for early colorectal cancer detection*. Tumori, 2014. **100**(2): p. 115-21.
387. Yueh, A.E., et al., *Colon Cancer Tumorigenesis Initiated by the H1047R Mutant PI3K*. PLoS One, 2016. **11**(2): p. e0148730.
388. Diaz, L.A., Jr., et al., *The molecular evolution of acquired resistance to targeted EGFR blockade in colorectal cancers*. Nature, 2012. **486**(7404): p. 537-40.
389. Giaretti, W., et al., *Intratumor heterogeneity of k-ras and p53 mutations among human colorectal adenomas containing early cancer*. Anal Cell Pathol, 2000. **21**(2): p. 49-57.
390. Sylvester, B.E. and E. Vakiani, *Tumor evolution and intratumor heterogeneity in colorectal carcinoma: insights from comparative genomic profiling of primary tumors and matched metastases*. J Gastrointest Oncol, 2015. **6**(6): p. 668-75.
391. Yan, Z.H., et al., *Comparative study of mutations in SNP loci of K-RAS, hMLH1 and hMSH2 genes in neoplastic intestinal polyps and colorectal cancer*. World J Gastroenterol, 2014. **20**(48): p. 18338-45.
392. Frattini, M., et al., *Quantitative and qualitative characterization of plasma DNA identifies primary and recurrent colorectal cancer*. Cancer Lett, 2008. **263**(2): p. 170-81.
393. Pallisgaard, N., et al., *Controls to validate plasma samples for cell free DNA quantification*. Clin Chim Acta, 2015. **446**: p. 141-6.