

The biology of colorectal peritoneal malignancy and its prognostic significance.

By

Sally Hallam

A thesis submitted to the University of Birmingham for the degree Doctor of Philosophy

The Institute of Cancer and Genomic Sciences

College of Medical and dental Sciences

The University of Birmingham

May 2020

UNIVERSITY OF
BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

This unpublished thesis/dissertation is copyright of the author and/or third parties. The intellectual property rights of the author or third parties in respect of this work are as defined by The Copyright Designs and Patents Act 1988 or as modified by any successor legislation.

Any use made of information contained in this thesis/dissertation must be in accordance with that legislation and must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the permission of the copyright holder.

Abstract (200 words)

Peritoneal metastasis, (CPM) develop in 15% of colorectal cancers. Cytoreductive surgery and heated intraperitoneal chemotherapy (CRS & HIPEC) aims to achieve macroscopic tumour resection and ablation of microscopic disease. However, 5-year survival varies widely and morbidity and mortality are relatively high. There is a need to improve patient selection and a paucity of research concerning CPM disease biology. Prognostic classifiers have been developed in a number of cancers, though none in CPM. I aimed to identify a multi-dimensional classifier to improve patient selection for CRS & HIPEC and to further the understanding of CPM disease biology.

Firstly, clinical prognostic factors for patients with CPM undergoing CRS & HIPEC were identified by systematic review. Secondly, comprehensive molecular profiling of the transcriptome, epigenome and genome of CPM was performed to develop a prognostic multidimensional classifier. Finally, a potential biomarker of tumour recurrence following CRS & HIPEC was examined using circulating tumour DNA.

This study has identified clinical and molecular features capable of predicting poor prognosis CPM and has potential applications in improved patient selection for treatment and the development of novel personalised treatments.

Acknowledgements

This work would not have been possible without the participation of a group of patients and their loved ones who enthusiastically and generously gave their time and samples in the pursuit of improved future treatments for patients with CPM. I am honoured to have met you and grateful for your participation.

The completion of this thesis would not have been possible without the extensive and patient help and support of my supervisors Mr Andrew Beggs and Mr Haney Youssef.

I am particularly grateful to the team at Good Hope hospital; nurse specialist Joanne Williams, the theatre team, Mr Raju, Mr Fallis, Dr Asif Ali, Leah Murphy, Cathy Padmore, Julie Tyler and Julian Bomba, and all of the clinic nurses. Without all of your help, patient recruitment and sample collection would not have been possible.

To Joanne Stockton, Valerie Pestinger, Celina Whalley, Claire Bryer, João D. Barros-Silva, Rachel Hoare and Konstantinos Kagiopoulos. You collectively patiently taught me the laboratory techniques needed to make this work possible. As well as providing much fun along the way. To Tom, Agata, Ollie, Rob, Kasun, Tojo, Yash, Helen and Helen without whom this would not have been such an enjoyable process.

To my parents who worked tirelessly to allow me the privilege of going to University to study medicine and have continued to support me throughout my surgical training. To Alex for all his love and support.

Finally, I am grateful to the University of Hospitals Birmingham and to the Royal College of Surgeons for their generous funding awards.

Chapter 1 Introduction	1
1.1 The epidemiology of colorectal cancer	2
1.2 The aetiology of colorectal cancer	3
1.3 Tumorigenesis	4
1.3.1 Mismatch repair inactivation and microsatellite instability	5
1.3.2 The Wnt signalling pathway	7
1.3.3 Chromosomal instability (CIN)	9
1.3.4 p53.....	9
1.3.5 Transforming growth factor Beta.....	10
1.3.6 The MAPK pathway	10
1.3.7 The role of aberrant methylation in colorectal cancer	13
1.3.7.1 Transcriptional silencing of tumour suppressor and DNA repair genes	13
1.3.7.2 The CpG island methylator phenotype.....	14
1.3.7.3 Mutability of methylcytosines	14
1.3.7.4 Methylation inhibition.....	14
1.3.8 Tumour immune recognition and evasion.....	15
1.3.9 The adenoma-carcinoma sequence	15
1.3.10 Copy number alteration	18
1.4 The molecular classification of colorectal cancer.....	19
1.4.1 Consensus molecular subtypes	19
1.4.2 Colorectal intrinsic subtypes	21
1.4.3 Molecular prognostic classifiers	21
1.5 The presentation of colorectal cancer.....	23
1.6 The diagnosis of colorectal cancer	23
1.7 The staging and grading of colorectal cancer	23
1.8 Mechanism and routes of metastasis in CRC.....	25

1.9 Peritoneal metastasis	25
1.10 The treatment of peritoneal metastasis	27
1.10.1 Palliative chemotherapy.....	27
1.10.2 Cytoreductive surgery and heated intraperitoneal chemotherapy, history	28
1.10.3 Cytoreductive surgery and heated intraperitoneal chemotherapy, technique	29
1.10.4 Heated intraperitoneal chemotherapy, technique	32
1.10.5 Adjuvant intraperitoneal chemotherapy	33
1.10.6 Cytoreductive surgery and heated intraperitoneal chemotherapy, oncological outcomes	33
1.10.7 Cytoreductive surgery and heated intraperitoneal chemotherapy, morbidity and mortality.....	34
1.11 Biomarkers of colorectal peritoneal metastasis, circulating tumour DNA	34
1.12 The biology of colorectal peritoneal metastasis	35
1.12.1 MAPK pathway mutations.....	35
1.12.2 PI3K pathway mutations	36
1.11.3 Metastasis-associated colon cancer 1 gene (MACC1)	36
1.12.4 Histone acetyltransferase (Tip60)	36
1.12.5 Neurone glial-related cell adhesion molecule (Nr-CAM)	36
1.12.6 Collagen Triple Helix Repeat Containing 1 (CTHRC1)	37
1.12.7 Increased TOPO1, Cox2, ERCC1 and reduced RRM1 protein expression in CPM.....	37
1.13 Conclusion.....	39
1.14 Hypotheses	40
1.15 Aims.....	40
Chapter 2 Materials and Methods	41
2.1 List of reagents and suppliers	42
2.1.1 General consumables and equipment common to all techniques.....	42
2.1.2 Tissue collection and storage	43

2.1.3 Extraction of RNA and DNA.....	43
2.1.4 Quantification of DNA and RNA.....	44
2.1.5 RNA library preparation	45
2.1.6 DNA sequencing.....	45
2.1.7 Methylation arrays	46
2.1.8 Exome sequencing	47
2.1.9 DNA panel sequencing	47
2.2 Methods.....	48
2.2.1 Ethics.....	48
2.2.2 Clinical sample preparation	48
2.2.3 Nucleic acid extraction.....	49
2.2.3.1 DNA & RNA extraction from retrospective FFPE tissue samples	49
2.2.3.2 DNA & RNA extraction from prospective fresh tissue preserved in RNA later.....	51
2.2.3.3 DNA extraction from prospective matched normal blood samples	52
2.2.3.4 Cell-free DNA extraction from plasma samples	52
2.2.4 Fluorometric Quantification of RNA and DNA.....	53
2.2.4.1 Qubit™ quantification.....	53
2.2.5 DNA & RNA electrophoresis.....	53
2.2.5.1 TapeStation.....	54
2.2.5.2 Bio-Analyser.....	56
2.2.6 Polymerase chain reaction, an introduction	56
2.2.7 RNA sequencing.....	56
2.2.7.1 Contemporary RNA sequencing techniques	59
2.2.7.2 RNA library preparation, 3' mRNA-seq method	64
2.2.7.3 RNAseq bioinformatics.....	66
2.2.8 DNA sequencing.....	66
2.2.8.1 DNA sequencing method	69
2.2.9 Epigenetics	69
2.2.9.1 Histone modification.....	70

2.2.9.2 Non-coding RNA associated gene silencing	70
2.2.9.3 DNA methylation	70
2.2.9.4 Methylation analysis.....	73
2.2.9.5 Methylation analysis techniques	76
2.2.9.6 Next-generation sequencing	80
2.2.9.7 Methylation analysis Infinium MethylationEPIC array, methods	80
2.2.9.8 Methylation bioinformatics.....	84
2.2.10 Exome sequencing	85
2.2.10.1 Exome sequencing, method.....	88
2.2.10.2 Exome sequencing bioinformatics	91
2.2.11 DNA panel sequencing	93
2.2.11.1 DNA panel sequencing method	96
2.2.12 Biosigner classifier development	98
2.2.13 Circulating tumour DNA sequencing techniques.....	99
2.2.13.1 Circulating tumour DNA, bioinformatics	99

Chapter 3 A meta-analysis of prognostic factors for patients with colorectal peritoneal metastasis undergoing cytoreductive surgery and heated intraperitoneal chemotherapy..... 101

3.1 Introduction 102

3.2 Method..... 103

3.2.1 Literature search.....	103
3.2.2 Study selection	104
3.2.3 Assessment of risk of bias	104
3.2.4 Data Extraction	104
3.2.5 Prognostic factor selection.....	105
3.2.6 Statistical analysis	105

3.3 Results 105

3.3.1 Prognostic factors	115
3.3.1.1 Patient factors	115
3.3.1.2 Tumour factors.....	115

3.3.1.3 Treatment factors.....	122
3.4 Discussion	128
 Chapter 4 : Describing the biology of colorectal peritoneal malignancy - an investigation of the gene expression, methylation and somatic mutation alterations associated with the development of colorectal peritoneal malignancy and with poor prognosis.	
4.1 Introduction	132
4.2 Results	135
4.2.1 Retrospective patient cohort	135
4.2.2 RNA sequencing, differential gene expression	138
4.2.2.1 Differential gene expression, primary CRC vs. matched metachronous CPM	138
4.2.2.2 Differential gene expression profile Poor vs. Good prognosis CPM.....	148
4.2.2.3 Correlation between metastasis development, prognosis and CMS and CRIS subtypes	155
4.2.3 Methylation	162
4.2.3.1 Methylation Microarray Analysis, primary CRC vs. matched metachronous CPM	162
4.2.3.2 Differential methylation at the probe level, primary CRC vs. matched metachronous CPM.....	162
4.2.3.3 Differential methylated regions, primary CRC vs. matched metachronous CPM	166
4.2.3.4 Copy number alteration, primary CRC vs. matched metachronous CPM	169
4.2.3.5 Methylation Microarray Analysis, Poor vs. Good prognosis CPM	171
4.2.3.6 Differential methylation at the probe level, Poor vs. Good prognosis CPM	171
4.2.3.7 Differential methylated regions, Poor vs. Good prognosis CPM	174

4.2.3.8 Copy number alteration, Poor vs. Good prognosis CPM.....	176
4.2.4 Exome sequencing	178
4.2.4.1 Sequencing quality.....	178
4.2.4.2 Primary CRC, somatic variations	178
4.2.4.3 Primary CRC, candidate driver mutations	180
4.2.4.4 Primary CRC, pathways.....	183
4.2.4.5 CPM, somatic variations	185
4.2.4.6 CPM, candidate driver mutations	188
4.2.4.7 CPM, pathways.....	194
4.2.5 Prognostic groups.....	196
4.2.5.1 Candidate driver mutations for poor prognosis CPM	199
4.3 Discussion	200
 Chapter 5 Developing a bio-molecular classifier to predict response to treatment with cytoreductive surgery and heated intraperitoneal chemotherapy in patients with colorectal peritoneal malignancy.....	 204
5.1 Introduction	205
5.2 Results	209
5.2.1 Prospective patient cohort	209
5.2.2 RNAseq classifier	212
5.2.2.1 Analysis of significant classifier features in the prospective cohort.....	215
5.2.3 Methylation classifier	217
5.3 Discussion	220
 Chapter 6 Investigating use of circulating tumour DNA to predict tumour recurrence and determine the mutational burden in patients treated with curative intent CRS & HIPEC for CPM	 223
6.1 Introduction	224
6.2 Results	227
6.2.1 Prospective patient cohort	227

6.2.2 Hypermutant tumour	229
6.2.3 ctDNA mutation number significantly higher than tumour	229
6.2.4 ctDNA remained detectable following complete CRS & HIPEC.....	234
6.2.5 The use of ctDNA as a biomarker of tumour recurrence.....	234
6.2.6 Mutations detected in ctDNA	237
6.3 Discussion	244
Chapter 7 Discussion.....	248
7.1 Interpretation and implications	249
7.1.1 Prognostic factors for patients with colorectal peritoneal metastasis undergoing Cytoreductive Surgery and Heated Intraperitoneal chemotherapy	249
7.1.2 Describing the biology of colorectal peritoneal malignancy	249
7.1.3 Developing a bio-molecular classifier to predict prognosis for patients with CPM undergoing CRS & HIPEC.....	251
7.1.4 The use of circulating tumour DNA to predict tumour recurrence and determine mutational burden in patients with CPM	252
7.2 Strengths and limitations	253
7.3 Future directions	254
7.3.1 Validation	254
7.3.2 Future research	255
7.4 Conclusion.....	259

Table of Figures

Figure 1-1 The DNA mismatch repair pathway.....	6
Figure 1-2 The Wnt signalling pathway	8
Figure 1-3 The RAS/RAF/MEK pathway and the PI3K/AKT/mTOR pathways.	12
Figure 1-4 The adenoma-carcinoma sequence.....	17
Figure 1-5 Routes of CRC metastasis.....	26
Figure 1-6 The peritoneal carcinomatosis index (PCI)	30
Figure 2-1 Tape station electropherogram	55
Figure 2-2 Transcription and translation.....	58
Figure 2-3 Microarray technology.....	60
Figure 2-4 3'Prime RNA sequencing	63
Figure 2-5 Next generation sequencing	68
Figure 2-6 DNA methylation	72
Figure 2-7 Bisulfite conversion of genomic DNA	75
Figure 2-8 The Infinium methylation array	79
Figure 2-9 Exon capture methods	87
Figure 2-10 QIAseq™ targeted DNA panel.....	95
Figure 3-1 PRISMA flow diagram	110
Figure 3-2 Prognostic factor effect on overall survival, (Hazard ratio and 95% confidence interval)	114
Figure 4-1 Kaplan-Meier disease-free survival analysis, good vs. poor prognosis patient cohorts.....	137
Figure 4-2 Heatmap, differential gene expression CPM vs. primary samples	146
Figure 4-3 Principal component analysis, differential gene expression CPM vs. primary samples	147
Figure 4-4 Sankey diagram of the CMS transition from primary to CPM.....	156
Figure 4-5 Sankey diagram of the CRIS transition from primary to CPM	158
Figure 4-6 Heatmap differential gene expression good vs. poor CPM	160
Figure 4-7 Principal component analysis, good vs. poor prognosis CPM.....	161
Figure 4-8 Box plot of Epiplakin expression CPM vs. primary CRC	165
Figure 4-9 Mutations exclusive to and shared between primary CRC and matched CPM	187

Figure 4-10 Mutations exclusive to and shared between good and poor prognosis CPM	197
Figure 5-1 Description of the Biosigner algorithm for feature selection	208
Figure 5-2 Kaplan-Meier disease-free survival analysis, good vs. poor prognosis patient cohorts.....	211
Figure 5-3 RNAseq classifier, differential expression of the features selected by the random forest classifier in the good and poor prognosis cohorts	214
Figure 5-4 Individual boxplots for the features selected by the retrospective random forest RNA classifier in the prospective cohort.....	216
Figure 5-5 Methylation classifier, differential expression of the features selected by the RF and PLSA-DA classifier features in the good and poor prognosis cohorts...	219
Figure 6-1 DFS according to ctDNA detection status	228
Figure 6-2 Sample ctDNA tape station electropherogram	231
Figure 6-3 Correlation between mutation number in CPM tumour and baseline ctDNA samples	232
Figure 6-4 Correlation between mutation number in CPM tumour and baseline ctDNA samples.....	233
Figure 6-5 Mutation number tracking from CRS & HIPEC to final follow up for individual patients.....	236
Figure 6-6 Lollipop plots showing the location and distribution of mutations in	239
Figure 6-7 Lollipop plots showing the location and distribution of mutations in	240
Figure 6-8 Lollipop plots showing the location and distribution of mutations in	241
Figure 6-9 The proportion of gene mutations detected in baseline ctDNA samples before CRS & HIPEC in patients with good and poor prognosis	243
Figure 7-1 Novel therapeutic targets identified	258

Table of Tables

Table 1-1 The consensus molecular subtypes of colorectal cancer	20
Table 1-2 TNM staging colorectal cancer	24
Table 1-3 The completeness of cytoreduction, (CC) score.	31
Table 2-1 Genes targeted by the QIAseq™ targeted panel CDHS-14542Z-1197	94
Table 3-1 Demographics	107
Table 3-2 Adjustment factors used in multivariate analysis.....	109
Table 3-3 Risk of bias assessment results for each study using the Quality in Prognostic Studies (QUIPS) tool	111
Table 3-4 Survival outcomes	112
Table 3-5 Tumour factors	117
Table 3-6 Treatment factors	124
Table 4-1 Comparison of good and poor prognosis patient cohorts	136
Table 4-2 KEGG pathway enrichment analysis, primary CRC vs. matched metachronous CPM.....	139
Table 4-3 The top 10 genes with significantly reduced expression (FDR<0.1) in CRS samples compared with primary CRC samples	142
Table 4-4 The top 10 genes with significantly increased expression (FDR<0.1 in CPM samples compared with primary CRC samples	145
Table 4-5 KEGG pathway enrichment analysis, poor vs. good prognosis CPM, p= <0.05	149
Table 4-6 The top 10 genes with significantly increased expression (FDR<0.1) in poor prognosis CPM.....	152
Table 4-7 The five genes with significantly reduced expression (FDR<0.1) in poor prognosis CPM.....	154
Table 4-8 CMS classification primary vs. CPM.....	156
Table 4-9 CRIS classification primary vs. CPM	158
Table 4-10 CMS classification good vs. poor prognosis CPM	159
Table 4-11 CRIS classification good vs. poor prognosis CPM	159
Table 4-12 Differentially methylated CpG sites, CPM vs. primary CRC	164
Table 4-13 The top 10 differentially methylated regions, (DMRs), CPM vs. primary CRC	168

Table 4-14 The top 10 CNAs, CPM vs. primary CRC.....	170
Table 4-15 Differentially methylated CpG sites, good vs. poor prognosis CPM	173
Table 4-16 Differentially methylated regions, (DMRs) good vs. poor prognosis CPM	175
Table 4-17 Known CRC driver mutations identified in the primary CRC cohort.....	179
Table 4-18 The top 10 genes with a bias towards a high functional level of mutation in primary CRC	182
Table 4-19 The top 10 KEGG pathways with accumulated mutations of high functional impact in primary CRC	184
Table 4-20 Known CRC driver gene mutations identified in the CPM cohort	186
Table 4-21 The top 10 genes with a bias towards high functional levels of mutations in CPM	193
Table 4-22 The top 10 KEGG pathways with accumulated mutations of high functional impact in CPM.....	195
Table 4-23 Potential candidate variants, poor prognosis CPM.....	198
Table 5-1 Comparison of good and poor prognosis patient cohorts	210
Table 5-2 RNAseq classifier confusion table and resulting sensitivity, specificity and error rate.....	213
Table 5-3 Methylation classifier signature	218
Table 6-1 Gene panel mutation proportions detected in baseline ctDNA in all patients and by prognostic group	238

List of supplementary tables and figures

Figure S 1 Forest and funnel plots for each prognostic factor	302
Figure S 2 Methylation array, quality control checks	309
Table S 1 Methylation bioinformatics R code	300
Table S 2 Biosigner classifier R code	300
Table S 3 Significantly (FDR<0.1) downregulated genes in CPM samples compared with primary CRC samples	312
Table S 4 Significantly (FDR<0.1) upregulated genes in CPM samples compared with primary CRC samples	313
Table S 5 Significantly (FDR<0.1) upregulated genes in poor prognosis CRS samples compared with good prognosis CRS samples	314
Table S 6 Methylation array, CpGs satisfying the detection p-value of 0.05 and 0.01	317
Table S 7 Differentially Methylated Regions (DMRs), in CPM samples compared with primary CRC samples	319
Table S 8 Copy number alterations, primary CRC and CPM, p value < $\times 10^{-07}$	319
Table S 9 Copy number alteration, poor and good prognosis CPM, p value < $\times 10^{-10}$	320
Table S 10 Exome sequencing quality metrics	323
Table S 11 Genes with a bias towards high functional levels of mutations in the primary CRC cohort, potential drivers	325
Table S 12 The KEGG pathways with accumulated mutations of high functional impact in primary CRC	327
Table S 13 Genes with a bias towards high functional levels of mutations in the CPM CRC cohort, potential drivers	328
Table S 14 The KEGG pathways with accumulated mutations of high functional impact in CPM CRC	337

List of abbreviations

5-fu	5-Fluorouracil, Fluoropyrimidine chemotherapy
5Mc	5-methylcytosine
AFA	Adaptive Focussed Acoustics
APC	Adenomatous Polyposis Coli
APC/C	Anaphase promoting complex
BF	Bayes factor
bp	Base pair
°C	Celsius
CC score	Completeness of cytoreduction score
cDNA	Complementary Deoxyribonucleic acid
ccfDNA	Circulating cell-free DNA
CHASM	Cancer-specific high throughput annotation of somatic mutations
CH3	Methyl group
Chr	Chromosome
c.i.	Confidence interval
CIN	Chromosomal Instability
CIMP	CpG island methylation phenotype
CMS	Consensus Molecular Subtype
CNA	Copy number alteration
COBRA	Combined bisulfite restriction analysis
CPM	Colorectal peritoneal metastasis

CpG	Cytosine followed by guanine in a 5' to 3' direction
CRC	Colorectal cancer
CRS	Cytoreductive surgery
CRCSC	Colorectal cancer Subtyping Consortium
CT	Computed tomography
ctDNA	Circulating tumour DNA
ddNTP	Dideoxynucleotide
DEG	Differentially expressed gene
DNA	Deoxynucleic acid
DNMT	DNA methyltransferases enzymes
DMR	Differentially methylated regions
DMP	Differentially methylated probes
dNTP	Deoxyribonucleotide
dsDNA	Double-stranded DNA
DFS	Disease-free survival
DNA	Deoxyribonucleic acid
EB	Elution buffer
EDTA	Ethylenediaminetetraacetic acid
EGAPP	Evaluation of genomic appliances in practice and prevention
EGFR	Epithelial growth factor receptor
EMT	Epithelial-mesenchymal transition
EPIC	European prospective investigation into cancer
EPIC	Early postoperative intraperitoneal chemotherapy

EPPK1	Epiplakin
EST	Expression sequence tags
FAP	Familial adenomatous polyposis
FFPE	Formalin Fixed Paraffin Embedded
FOBT	Faecal occult blood test
FOLFOX	Chemotherapy combination, 5-fu/Leucovorin/Oxaliplatin
FOLFIRI	Chemotherapy combination, 5-Fu/Leucovorin/ Irinotecan
FZD	Frizzled receptor protein
GDP	Guanosine diphosphate
GEF	Genome nucleotide exchange factors
GEO	Gene expression omnibus
HCC	Hepatocellular carcinoma
HDU	High dependency unit
HBRC	Human biomaterials Resource Centre
HIPEC	Heated Intra-Peritoneal Chemotherapy
HNPCC	Hereditary non-polyposis colorectal cancer
HR	Hazard ratio
HS	High sensitivity
JPS	Juvenile polyposis syndrome
kb	Kilobase
KEGG	Kyoto Encyclopedia of Genes and Genomes
ITU	Intensive care unit
LINE	Long interspersed nuclear elements

lncRNA	Long non-coding RNA
MACC1	Metastasis-associated colon cancer 1 gene
MAP	MUTYH-associated polyposis
MAPK	Mitogen-activated protein kinase
MHC	Major histocompatibility complex
ml	Millilitre
MMR	Mismatch repair
mRNA	Messenger Ribonucleic Acid
MS-DGGE	Methylation-specific denaturing gradient gel electrophoresis
MS-DHPLC	Methylation-specific denaturing high-performance liquid chromatography
MSI	Microsatellite instability
MSRE	Methylation sensitive restriction enzymes
MS-PCR	Methylation-specific PCR
MS-SnuPE	Methylation-sensitive single-nucleotide primer extension
MSS	Microsatellite stability
MUC2	Mucin 2
miRNA	MicroRNA
mRNA	Messenger RNA
NaOH	Sodium Hydroxide
NCBI	National Center for Biotechnology Information
Ng	Nanograms
NGS	Next-generation sequencing
NF- κ B	Nuclear factor kappa-light-chain-enhancer of activated B cells

NF-κBIA	NK-Inhibitor Alpha
Nr-CAM	Neurone glial-related cell adhesion molecule
OS	Overall survival
P53	Tumour suppressor gene
PB1	Reagent used to prepare BeadChips for hybridization
PCR	Polymerase chain reaction
PCI	Peritoneal carcinomatosis index
Phix	Bacteriophage genomic library
PJS	Peutz-Jeghers syndrome
PI3KCA	Phosphatidylinositol 3-kinase gene
piRNA	Piwi-interacting RNA
PMP	Paramagnetic Particles
Poly(A)	Polyadenylated
PRISMA	Preferred reporting items for systematic reviews and meta-analysis
PSDSS	Peritoneal Surface Disease Severity Score
PTEN	Phosphate and tensin homolog tumour suppressor gene
RAS	Oncogene Family, including KRAS
RAF	Oncogene family, including BRAF
RCT	Randomised controlled trial
RECIST	Response evaluation criteria in solid tumours
RELB	Proto-Oncogene
RF	Random forest
RFS	Relapse free survival

RNA	Ribonucleic Acid
RNAi	Interference RNA
rRNA	Ribosomal RNA
RLT	Lysis buffer
RSB	Resuspension Buffer
RTK	Receptor tyrosine kinase
SD	Standard deviation
siRNA	Small interfering RNA
SNP	Single nuclear polymorphisms
SNV	Single nucleotide variation
snoRNA	Small nucleolar RNAs
snRNA	Small nuclear RNA
SMB	Streptavidin magnetic beads
SPB	Sample Purification Beads
SPIC	Sequential postoperative chemotherapy
SSV	Somatic shared variants
TCGA	The cancer genome atlas
TGF- β	Transforming Growth factor beta
TGFR1/2	Transforming Growth factor receptor
Tip60	Histone acetyltransferase gene
TNM	Tumour node metastasis system
TRIS	Tris(hydroxymethyl)aminomethane
tRNA	transfer RNA

UMI	Unique molecular indices
UV	Ultraviolet
μl	Microlitre
VEGF	Vascular endothelial growth factor
% v/v	Volume concentration
xg	Relative centrifugal force
Xist	X-inactive specific transcript
WES	Whole-exome sequencing
WGS	Whole-genome sequencing
Wnt	Wingless/Integrated pathway

Chapter 1 Introduction

1.1 The epidemiology of colorectal cancer

Colorectal cancer (CRC) is the fourth most common cancer in the UK with 41,804 new cases in 2015 and 16,384 deaths in 2016 (1, 2). Rates of CRC are higher in males than females. The incidence of CRC has increased by 4% since the 1990s (1, 2). This may be attributable to an increased prevalence of risk factors in previous years and improved detection with the introduction of bowel cancer screening in 2006 (3).

There is wide variation in incidence globally with high rates in Australia, New Zealand, Europe and America compared to Africa and Asia. This global variation is likely to represent the effects of environmental exposures including red meat, dietary fat and reduced fibre. Migrants from low to high-risk areas adopt the incidence of the country they migrate to implicating environmental and lifestyle risk factors in the development of CRC (4). Low socioeconomic status is associated with a 14% increase in CRC (3). Several studies have established environmental risk factors, which may explain the socioeconomic disparity in the development of CRC. The European Prospective Investigation into Cancer (EPIC) was a large multicentre prospective cohort, which recruited over half a million participants and followed them for 15 years to investigate the relationship between diet, nutritional status, lifestyle, environmental factors and the incidence of disease. EPIC established a protective effect of a high fibre and fish diet for CRC and an increased risk associated with a high intake of processed and red meat (5, 6). The United States nurses' health study was a multicentre prospective cohort study of over 280,000 participants followed for 24 years investigating risk factors for several diseases. The study established risk factors associated with CRC: obesity, low levels of exercise, high alcohol intake and smoking (7).

1.2 The aetiology of colorectal cancer

The aetiology of CRC is multifactorial with a complex interplay between environmental factors, heritable components and spontaneous genetic mutations.

The majority of CRC is sporadic. There is a wide variation in cancer incidence according to the tissue of origin from 4.82% for cancers of the large intestine to 0.003% for cancers of the brain (8). Some of this variation can be explained by environmental factors known to increase the risk of CRC: obesity, intake of red and processed meat, Tobacco, alcohol intake (5-7), inflammatory bowel disease (9, 10) and abdominal irradiation (11). Variation in tissue-specific cancer incidence persists despite adjustment for environmental exposure, meaning environmental factors account for only a small proportion of the cases of CRC. In 2015 Vogelstein et al established a correlation between the number of stem cell divisions and the risk of cancers in a given tissue, suggesting that random mutations which occur during deoxyribonucleic acid (DNA) replication of normal tissues are a significant causative factor in the development of CRC (8).

Familial association studies suggest up to 30% of CRC cases have a heritable component. Only 5%, however, are the result of well-defined CRC syndromes with known genetic mutations. Hereditary non-polyposis colorectal cancer (HNPCC) also known as Lynch syndrome, is an autosomal dominant condition associated with an 80% lifetime risk of CRC. HNPCC results from an autosomal dominant germline mutation of the mismatch repair genes (MMR) *MLH1*, *MSH2*, *PMS2* or *MSH6*. Familial adenomatous polyposis (FAP) is characterised by numerous adenomatous polyps in the colon and rectum resulting in CRC if untreated. FAP results from a dominantly inherited germline mutation of the adenomatous polyposis coli (*APC*) gene. HNPCC and FAP account for the majority of inherited CRC. Other less common CRC syndromes associated with known heritable mutations include; *MUTYH*-associated polyposis (MAP), an autosomal recessive condition resulting from mutation of the *MUTYH* gene. Peutz-Jeghers syndrome (PJS), an autosomal dominant polyposis syndrome resulting from mutation of the *STK11* gene. Juvenile polyposis syndrome (JPS), an autosomal dominant syndrome resulting from a mutation in either the

BMPR1A or *SMAD4* genes (12), and autosomal dominant proofreading polyposis syndromes, characterised by a mutation in the *POLE1* or *POLD1* genes respectively.

1.3 Tumorigenesis

In CRC arising outside of such inherited syndromes, multiple sequential genetic events accumulate leading to the inactivation of tumour suppressor genes, the activation of oncogene pathways and tumorigenesis.

A gene mutation is a permanent DNA sequence alteration. Mutations range in size from a single nucleotide base pair (SNV) to the insertion or deletion of a sequence of nucleotides (indel), to structural variants greater than 1kb pairs such as inversions or translocations. The size, type and location of the mutation dictates the effect on its gene protein product. This can range from no effect (synonymous) to alterations in protein structure, function or production (non-synonymous). Germline DNA mutations inherited from a parent are present in every cell in the body. Acquired or somatic mutations occur at a time point after conception, result from environmental factors or errors in cell replication and are only present in certain cells. Somatic mutations directly affecting the potential of the cell to become cancerous, to progress and to metastasise are cancer driver mutations. These mutations confer a clonal growth advantage and are selected as the tumour cell population increases. Passenger mutations are those that do not confer a growth advantage and do not contribute to tumorigenesis. Cancer driver mutations are uncommon in the general population, other genetic mutations occur more frequently. Genetic changes common to more than 1% of the population are known as polymorphisms, these are considered normal DNA variants.

Tumour suppressor genes suppress inappropriate cell growth and division and stimulate cell death as well as maintaining DNA repair pathways. These functions prevent mutation accumulation and abnormal tissue growth. Loss of function mutations in tumour suppressor genes is commonly recessive. For cancer to develop both gene alleles must acquire mutations. This is known as the “two-hit” hypothesis and was first proposed by Alfred Knudson in 1971 (13). The common genetic events leading to CRC tumorigenesis are discussed below.

1.3.1 Mismatch repair inactivation and microsatellite instability

The fidelity of DNA replication is a very reliable process, incorrect base incorporation or the failure of the DNA polymerases proofreading function result in an error rate of around 10^{-7} base pairs per genome (14). The DNA mismatch repair pathway (MMR) ensures stable cell replication by targeting base substitution mismatches and insertion-deletion mismatches, which have escaped proofreading by DNA polymerases (Figure 1.1). Following DNA replication, the mismatch proofreading system is signalled to the daughter strand. A series of mismatch repair proteins (MutS, MutH and MutL) recognise and repair errors in the daughter strand. MutS recognises the mismatched base and binds to it; MutL binds and activates MutH, which nicks the daughter strand. A DNA Helicase separates the DNA strand and allows MutSHL complex to excise it. The strand is digested by an exonuclease. The gap is repaired by a DNA polymerase using the other strand as a template (15).

The MMR pathway contributes up to 1000-fold to the fidelity of cell replication. Inactivation of the MMR pathway, therefore, results in a high rate of spontaneous mutations. The hallmark of MMR-deficient cells is microsatellite instability (MSI) characterised by single-base mismatches and small insertion and deletion loops of unpaired bases, between 12–17 % of colorectal cancers are MSI (16), around 4-5% of metastatic colorectal cancers are MSI (17).

Kim et al examined the association of MSI status with recurrence patterns, DFS, OS in 2940 patients with CRC. MSI high (MSI-H) tumours recurred more frequently with local or peritoneal recurrence and less frequently in the lung or liver, additionally patients with recurrent MSI tumours had reduced OS compared with patients with microsatellite stable (MSS) tumours (18).

MSI-H tumours have a high tumour mutational burden (TMB) ≥ 20 mut/Mb, This results in an increased neo-antigen burden and potential response to immune checkpoint inhibition (19).

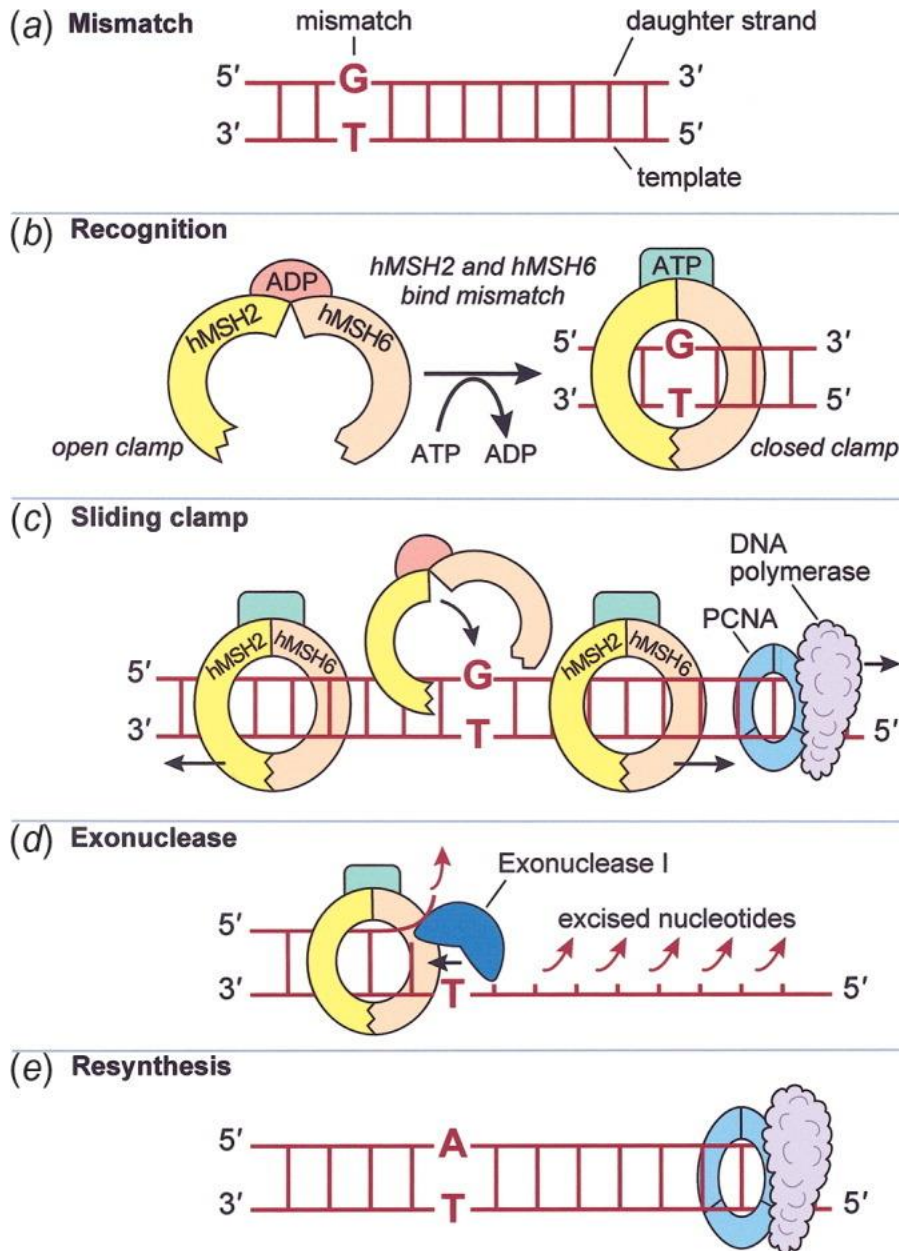


Figure 1-1 The DNA mismatch repair pathway.

A) A single nucleotide mismatch is formed during mitosis. B) Mismatch repair proteins identify the error in the newly synthesised DNA and clamp onto its location. C) The sliding clamp can migrate in either direction of the mismatch, recruitment of MutSa clamps ensure movement in the 5' to 3' direction encountering the PCNA-DNA polymerase complex. D) Exo1 excises the daughter strand back to the site of the mismatch. E) The error is corrected by resynthesis This figure is reprinted from Jascur et al, 2006, figure 1 (20)

1.3.2 The Wnt signalling pathway

The Wnt signalling pathway regulates intestinal cell proliferation, differentiation and cell fate decisions (Figure 1.2). Wild type Wnt is a signalling molecule which binds to cell receptor complexes (Frizzled, *FZD*) initiating the recruitment of scaffold proteins and disruption of the β -catenin destruction complex (*AXIN*, *APC*, *CK1*, *GSK3 β*) inhibiting its nuclear localisation. In the absence of the destruction complex, β -catenin accumulates and translocates into the nucleus where it associates with TCF transcription factors to drive the transcription of target genes. Wnt activation maintains crypt stem cell compartments in the healthy colon by promoting proliferation and stem-cell-like phenotypes. When persistently activated by mutation it leads to the development of CRC (21, 22). The Wnt signalling pathway is activated in over 90% of CRC due to mutations in either the β -catenin gene (*CTNNB1*) or *APC*. Inactivation of the *APC* tumour suppressor gene results in inappropriate and consistent activation of the Wnt signalling pathway and cell clonal expansion. *APC* is mutated in the majority of sporadic CRC as well as in FAP. In both FAP and sporadic mutations of the *APC* gene, a bi-allelic or 'two-hit' mutation must occur for tumorigenesis to occur. In a small group with wild type *APC*, mutations of the β -catenin gene (*CTNNB1*) located on exon-3, inappropriately activate the Wnt signalling pathway driving the transcription of Wnt regulated genes (23).

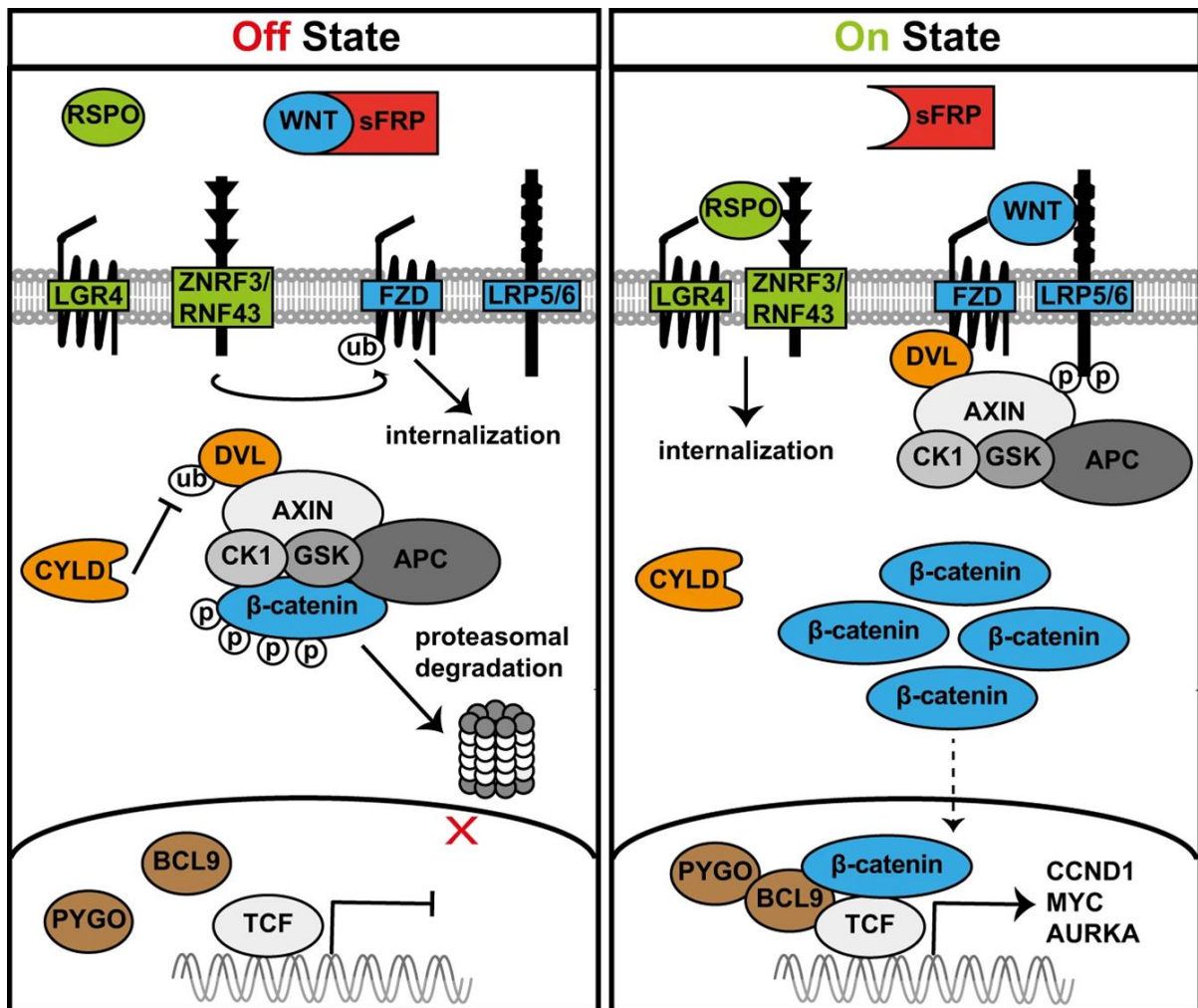


Figure 1-2 The Wnt signalling pathway

The Wnt signalling pathway. Off state: in the absence of Wnt ligands, β -catenin is phosphorylated by a destruction complex which marks it for proteasomal degradation. Wnt signalling is inhibited by multiple negative Wnt regulators (LGR4, ZNRF3, RNF43). On state: Binding of a Wnt ligand to its receptor Frizzled induces phosphorylation of co-receptor LRP5/6, forming a docking site for AXIN which disrupts the destruction complex allowing stabilisation and nuclear translocation of β -catenin. This results in transcription of Wnt target genes. This figure is reprinted from Van Andel et al, 2019, figure 1 (24).

1.3.3 Chromosomal instability (CIN)

Chromosomal instability (CIN) is characterised by widespread imbalances in chromosome number (aneuploidy), loss of heterozygosity or the formation of structurally abnormal chromosomes. CIN can result from several aberrant processes: The spindle assembly or mitotic checkpoint ensures fidelity of chromosome segregation by delaying anaphase until all pairs of chromatids are properly aligned. Defects here lead to mis-segregation and aneuploidy. Similarly, extra centrosomes lead to multiple spindle poles and the unequal distribution of chromosomes. Telomeres are DNA-protein complexes, which protect the chromosome from fusing or breaking during replication. When the telomere becomes shortened, apoptosis is triggered. Where this is compromised, chromosome ends enter breakage, fusion, bridge cycles leading to genome reorganisation and tumour formation. DNA damage response machinery protects cells from genotoxic stress by initiating cell cycle arrest and repair or apoptosis. p53 is a DNA repair protein implicated in CRC and predisposing to CIN (25).

Loss of heterozygosity is a hallmark of CIN; 25–75% of alleles are lost in tumours. This is thought to be a consequence of interchromosomal recombination and deletions. CIN is present in 70–85% of CRC and is associated with drug resistance and poor outcomes with a reduction in both disease-free survival (DFS) and overall survival (OS) (26).

1.3.4 p53

Wild type *TP53* encodes a transcription factor inducing cell cycle arrest, senescence, and apoptosis in response to cellular stress or DNA damage. p53 binds to DNA, stimulating the production of p21 protein, which interacts with a cell division stimulating protein cdk2. The p21/cdk2 complex inhibits cell cycle transition resulting in cell cycle arrest or senescence. p53 induces the expression of pro-apoptotic proteins (BCL-2, BAX) triggering apoptosis (27). p53 levels are tightly regulated by interaction with the E3 ubiquitin ligase MDM2, itself induced by p53, which targets p53 for ubiquitination and subsequent degradation (22, 28).

Mutant *TP53* is unable to bind to DNA, the stop signal is switched off and cells continue to divide unchecked. Autosomal dominant mutations of *TP53* result in Li-Fraumeni syndrome, characterised by the early onset of multiple cancers of various types. *p53* mutations occur in up to 70% of CRC and are associated with lymphatic and vascular invasion, chemotherapy resistance and poor prognosis (12, 27).

1.3.5 Transforming growth factor Beta

TGF- β is a multifunctional cytokine, which regulates proliferation, cell death, differentiation, adhesions, motility and epithelial to mesenchymal transition (EMT) (29). TGF- β ligands are activated by extracellular proteases (matrix metalloproteinase) and bind to its receptors (TGFR1 and TGFR2). Activated TGFR1 phosphorylates downstream SMAD transcription factors. The resulting complex translocates to the nucleus initiating the transcription of TGF- β responsive genes, which suppress cell proliferation or induce apoptosis (29).

TGF- β mutation can occur at all points of the pathway (TGF- β ligands, TGFR1/2, SMAD). TGF- β signalling inactivation is not able to induce tumorigenesis alone but is dependent on convergent signalling pathway dysregulation. A TGF- β activated microenvironment suppresses the differentiation and activity of lymphocytes excluding T-cells and allowing tumours to escape from immune surveillance. This immunosuppressive effect is associated with poor outcomes (30). Alterations in the TGF- β pathway affect 40%–50% of all CRCs and are associated with a high risk of disease recurrence (29).

1.3.6 The MAPK pathway

The mitogen-activated protein kinase (MAPK) cascade regulates cell proliferation and differentiation. It receives divergent stimuli from metabolic stress, DNA damage, altered protein concentrations and communications from other cells (22, 31). The MAPK pathway is triggered by the binding of a ligand (growth factor, cytokine or hormone) to a receptor tyrosine kinase (RTK) a transmembrane protein. Activated RTK is bound by cytoplasmic proteins. Guanine–nucleotide exchange factors (GEFs) then activate RAS, a GTPase protein at the plasma membrane. RAS is inactive when bound to Guanosine diphosphate (GDP), GEF displaces GDP allowing GTP to bind.

RAS then activates BRAF (MAPK). Activated BRAF facilitates the phosphorylation of a MEK (MAPK). MEK, in turn, activates the final MAPK in the cascade ERK. ERK translocates to the nucleus, activating transcription factors that result in gene expression (Figure 1.3) (22).

Several mutations resulting in upregulation of the MAPK pathway have been identified. Epithelial growth factor receptor (EGFR) is a receptor tyrosine kinase transmembrane protein mutated or over-amplified in up to 77% of CRC (31). Mutation of the *KRAS* signal transducer (mutated in 50% of CRC) activates *BRAF* kinases (mutated in 5-15% CRC), which in turn upregulates transcription factors and drives the MAPK signalling cascade (22, 31). *KRAS* mutation in CRC is associated with an increased risk of recurrence and death (32). *KRAS* mutation leads to the continued activation of the MAPK pathway in the absence of its upstream regulators. It is therefore associated with poor response to anti-EGFR targeted therapy commonly used in CRC such as Cetuximab and Panitumumab (33). *BRAF* mutations result in continuous activation of MEK and ERK and the transcription of factors involved in cell proliferation and survival. They are associated with reduced OS in both the early and the metastatic setting (34). In malignant melanoma treatment with the BRAF inhibitor Vemurafenib is associated with improved response and progression-free survival. In CRC however, treatment with BRAF inhibitors is associated with no significant survival gains (35). In CRC single-agent BRAF inhibition provides inadequate suppression of the MAPK pathway and can trigger an increase in upstream EGFR activation, downstream CRAF activation and signaling through the MAPK or PI3K pathways (36, 37).

Up to one-third of CRC bears a phosphatidylinositol 3-kinase gene (*PI3KCA*) mutation or loss of the Phosphate and tensin homolog tumour suppressor gene (*PTEN*), both of which increase PI3K pathway signalling which in turn affects cell growth, differentiation and proliferation (22). Mutations in *PI3KCA* are associated with resistance to anti EGFR therapies in the absence of *EGFR* mutation suggesting they may drive aberrant EGFR signalling (38).

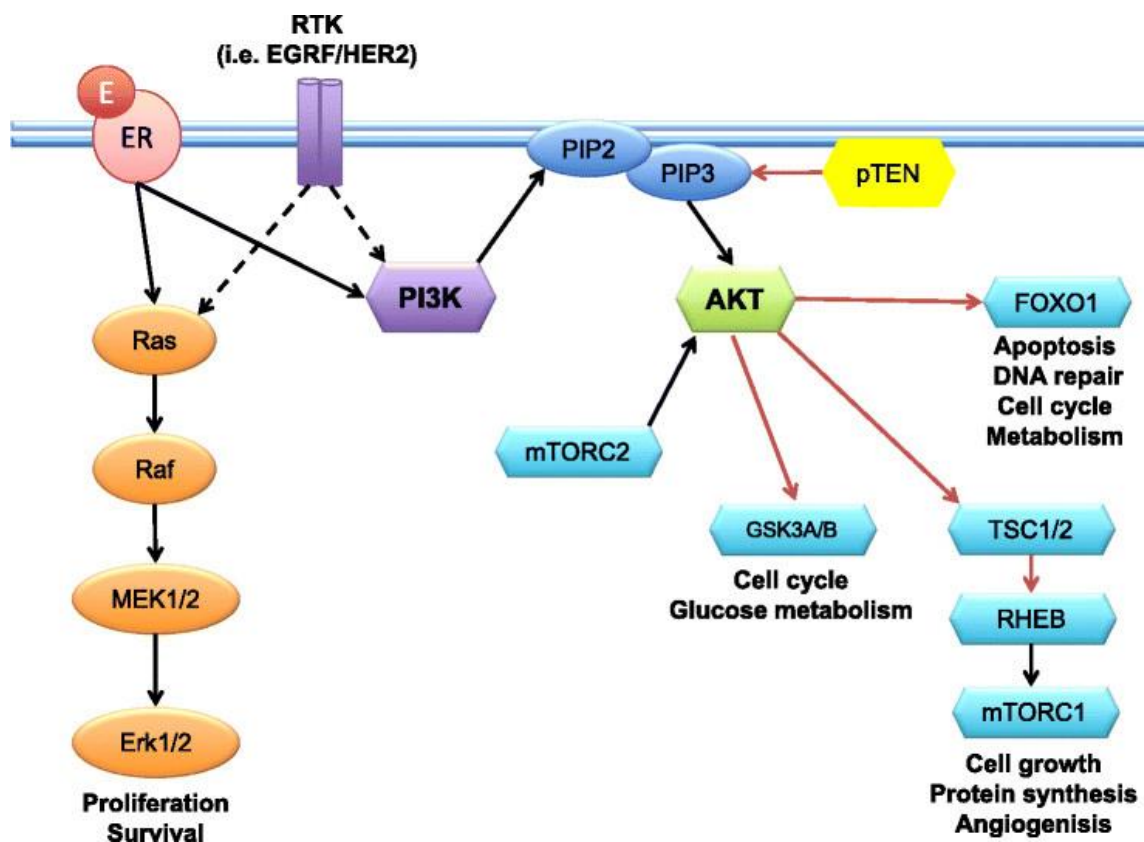


Figure 1-3 The RAS/RAF/MEK pathway and the PI3K/AKT/mTOR pathways.

Ligand binding activates the receptor tyrosine kinase (RTK) transmembrane protein, which is bound by cytoplasmic proteins. GTPase proteins and Guanine–nucleotide exchange factors (GEFs) then activate the RAS oncogene. RAS then activates BRAF, a mitogen-activated protein kinase (MAPK). Activated BRAF facilitates the phosphorylation of a MEK (MAPK). MEK, in turn, activates the final MAPK in the cascade, ERK. ERK translocates to the nucleus, activating transcription factors that result in gene expression responsible for cell growth, protein synthesis and angiogenesis (39). The RAS/RAF/MEK/MAPK pathway converges with the PI3K/AKT pathway, Phosphoinositide 3-kinase (PI3K) is recruited to the membrane by activated growth factor receptors. PI3K phosphorylates phosphoinositides producing phosphatidylinositol-3,4,5-triphosphate (PIP3) which signals to AKT through enzymes, kinases and transcription factors. The activated PI3K/AKT pathway promotes metabolism, proliferation, cell survival, growth and angiogenesis. This figure is reprinted from Toss et al, 2015, figure 1 (39).

1.3.7 The role of aberrant methylation in colorectal cancer

In health DNA methylation allows cells to acquire and maintain a specialised state, suppresses the expression of viral and non-host DNA and facilitates the cell response to environmental stimuli and ageing. Aberrant DNA methylation (hyper or hypo) is implicated in many disease processes including ageing, neurological disorders, metabolic disorders and cancer.

1.3.7.1 Transcriptional silencing of tumour suppressor and DNA repair genes

In health, the majority (70-80%) of CpG sites are methylated resulting in epigenetic silencing. CpG islands conversely, located in promoter regions of genes are usually un-methylated (40). In cancer cells, there is a loss of global methylation and a gain of methylation at the promoters of selected CpG islands. This results in the transcriptional silencing of specific genes including tumour suppressor genes (41). Methylation inhibits transcription directly by inhibiting transcription factor binding and indirectly by recruiting methyl CpG binding proteins and their associated repressive chromatin remodelling (42).

Up to 50% of genes mutated in hereditary forms of cancer undergo methylation associated transcriptional silencing in sporadic cancers (43). Loss of gene function arising from aberrant promoter methylation can provide a survival advantage to neoplastic cells similar to that resulting from mutations (43). Genes regulating the cell cycle, tumour cell invasion, DNA repair, chromatin remodelling, cell signalling, transcription and apoptosis become aberrantly methylated and silenced in nearly every tumour type (42).

The impact of global hypomethylation of non-promotor regions and structural elements such as centromeric DNA in cancer is less well understood but may result in genomic instability such as those resulting from germline mutations of *DNMT3* which results in a loss of methylation at centromeric regions and profound chromosomal structural changes (43). Embryonic stem cells deficient in DNMTs are viable but die when induced to differentiate (42).

Whilst a modest increase in DNMT activity and protein expression has been noted in some cancers, this has not been able to explain the patterns of aberrant methylation in CRC (41).

1.3.7.2 The CpG island methylator phenotype

A group of cancers display a 3-5 fold elevation in methylation compared to age-matched controls, these are known as CpG Island Methylator Phenotype (CIMP) (41).

Up to 70-80% of sporadic MSI colorectal cancers display CIMP and specifically are attributable to the methylation associated silencing of the *MLH1* mismatch repair gene (44).

CIMP positive tumours commonly occur in older, female patients and the proximal colon. On a molecular level, CIMP tumours display high levels of *KRAS* and *BRAF* mutation and low levels of *TP53* mutation (44). Histologically, CIMP tumours are poorly differentiated. CIMP status has additional prognostic implications; CIMP+MSI+ tumours are associated with good prognosis whereas CIMP+MSI- tumours a poor prognosis (44).

1.3.7.3 Mutability of methylcytosines

Cytosine methylation can influence tumorigenicity. Methylated cytosines are themselves mutagenic and can undergo spontaneous hydrolytic deamination causing C–T transitions. This enhanced mutagenesis is seen in mutations of the *TP53* tumour suppressor gene, up to 50% of inactivating point mutations in the coding region of *TP53* occur at methylated cytosines (43).

1.3.7.4 Methylation inhibition

DNA methylation inhibitors such as 5-azacytidine can restore the function of genes transcriptionally silenced by methylation (43). These drugs have been trialled with promise *in vitro* and animal studies. Clinical trials of DNA methylation inhibitors in patients with metastatic CRC have demonstrated that they are generally well tolerated at low doses and are capable of reversing hypermethylation. They are, however, associated with only a minimal objective response in the majority of patients (45-48).

DNA methylation inhibitors result in genome-wide methylation changes and this global effect may reduce their efficacy. Whilst reversal of methylation at one gene site may result in drug sensitisation at another it may reduce drug sensitivity. Clinical effectiveness of methylation inhibitors is limited by low bioavailability and dependence on nucleoside transporters for cellular uptake (48). Ongoing and future clinical trials focus on the use of methylation inhibitors as a priming agent to minimise resistance preceding further cytotoxic or immunotherapy (49).

1.3.8 Tumour immune recognition and evasion

Genetic and epigenetic alterations allow the immune system to distinguish tumour cells from host tissue. Tumour cells display antigenic peptides (epitopes) on major histocompatibility complex (MHC) molecules on the cell surface. These are recognised as 'non-self' by CD8 T-lymphocytes triggering their destruction.

Tumours, however, can evade immune recognition and elimination via several mechanisms. Constant tumour cell division produces new tumour cell variants. Loss of MHC antigen expression, loss of surface antigen expression and the secretion of immunosuppressive cytokines such as VEGF and TGF- β result in decreased immune visibility (50).

Tumour antigen epitopes can be shared between different tumours or encoded by unique somatic mutations private to an individual tumour (neoepitopes). Immunotherapies can be targeted towards shared epitopes or personalised to an individual tumours neoepitope (51).

1.3.9 The adenoma-carcinoma sequence

Tumorigenesis is a multistep process arising from a combination of mutations or epigenetic changes (Figure 1.4). Vogelstein et al hypothesised a genetic model for CRC in which the sequential accumulation of known genetic mutations in *APC*, *KRAS* and *TP53* drive the transition from healthy colonic epithelium to dysplasia to CRC (52). It is accepted that tumorigenesis is a sequential process arising from a combination of mutations. Sequencing of a large cohort of CRC, however, demonstrated only a small proportion of tumours contained mutations in all genes implicated in the adenoma-

carcinoma sequence and more than one-third of tumours contained mutations in only one of *APC*, *KRAS* or *P53* (53). The model proposed by Vogelstein is an oversimplification, whilst tumorigenesis occurs as a sequential process, the genetic mutations driving tumorigenesis can vary in number, type and timing in individual tumours (54).

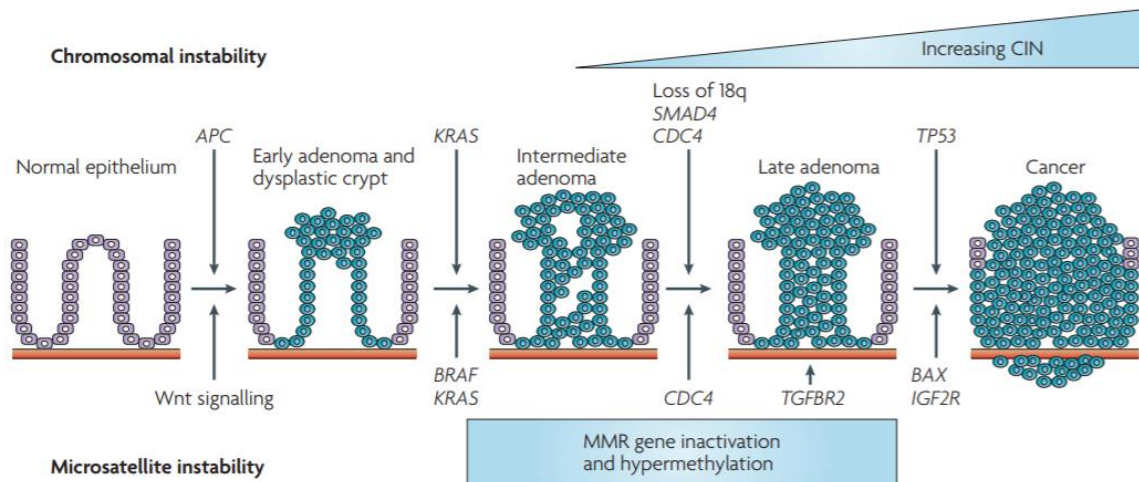


Figure 1-4 The adenoma-carcinoma sequence

The adenoma-carcinoma sequence describes the sequential accumulation of known genetic mutations, associated with recognisable stages of tumorigenesis, which drive the transition from healthy colonic epithelium to dysplasia to colorectal cancer (CRC (55)). Loss of adenomatous polyposis coli (APC) is the first step transforming normal colorectal epithelium to adenoma. Larger adenomas and early carcinomas acquire mutations in the small GTPase KRAS, followed by loss of chromosome 18q with SMAD4, which is downstream of transforming growth factor- β (TGF β), and mutations in TP53 in frank carcinoma. The accumulation of sequential mutations lead to chromosomal instability (CIN), mismatch repair gene inactivation leads to microsatellite instability. This figure is reprinted from Walther et al, 2009, figure 1 (55).

1.3.10 Copy number alteration

Copy number alterations (CNAs) are somatic changes resulting in gain or loss of DNA sections from the genome (56). CNAs contribute to the onset and the progression of cancer by the activation of proto-oncogenes and the inactivation of tumour suppressor genes (56). The initiation and progression of CRC involve the stepwise accumulation of CNAs. Adenomas have comparable levels of CNAs to carcinomas, metastatic CRC has increased CNAs when compared with invasive CRC (57). CNAs frequently identified in CRC are gains at chromosomes 1q, 7p, 7q, 8p, 8q, 12q, 13q, 19q, 20p, 20q and losses at chromosomes 8p, 17p, 18p and 18q (58). These common CNAs identify several oncogenes commonly amplified or deleted in CRC. Commonly amplified oncogenes include *EGFR*, *ERBB2*, *CND1*, *IGF2*, *NET* and *MYC* (58). Amplification of these genes through CNA results in tumour developments and progression and confers resistance to some drugs such as Cetuximab. Commonly deleted tumour suppressor genes include *TP53*, *APC*, *PTEN* and *SMAD4* (58). Deletion of these genes through CNA results in cell growth with no regulatory mechanisms such as apoptosis or cell cycle arrest. These CNA result in dysregulation in several well-defined CRC pathways including *WNT*, *MAPK*, *PI3K*, *TGF- β* and *p53* (58).

1.4 The molecular classification of colorectal cancer

1.4.1 Consensus molecular subtypes

Gene expression profiles in use in CRC include the Consensus Molecular Subtypes (CMS) and the colorectal cancer intrinsic subtypes (CRIS). The colorectal cancer subtype consortium (CRCSC) used gene expression profiles of approximately 4000 largely early stage primary CRCs (8% stage IV) to define four biologically distinct Consensus Molecular Subtypes (CMS) (Table 1.1). CMS1 comprises most tumours with MSI and is characterised by the infiltration of activated immune cells, CMS2 and CMS3 show epithelial characteristics with oncogene amplification, high Wnt and Myc signalling in CMS2 and metabolic reprogramming in CMS3. CMS4 comprises mesenchymal-like cancers with high stromal infiltration and poor prognosis (59). The CMS classifiers have a number of clinical associations. CMS1 tumours are common in females with right-sided CRC and high histopathological grade, they are associated with poor survival following relapse (59). CMS2 tumours are commonly left-sided and associated with high survival following relapse (59). CMS4 tumours are commonly diagnosed at an advanced stage and display worse OS and relapse-free survival, (RFS) despite adjustment for clinicopathological variables, MSI status, BRAF and KRAS mutations (59). CMS classifiers can be used to deliver targeted therapy, for example in CMS4 there is evidence that stromal cells mediate resistance to chemotherapies and targeted therapy (60). *TGF β* signalling inhibitors have been trialled with success in pre-clinical models to block cross-talk between the tumour microenvironment and halt disease progression of stromal rich CMS4 CRC (61). *TGF β* receptor inhibitors are currently being trialled in combination with chemotherapy in this patient subgroup (60).

CMS are classified according to the gene expression of all heterogeneous components which make up a tumour, cancer cells, vessels, fibroblasts and immune cells (62). Non-neoplastic cells within the extracellular matrix send extrinsic signals which dictate the prognosis of tumours such as the CMS4 mesenchymal subtype (62). Stromal traits in CRC influence prognosis, which is a weakness of the CMS classifier as it is a stroma specific classifier. The transcription profile of heterogeneous tumours with a high

proportion of non-cancerous cells may mask the subtle but biologically relevant gene expression profiles of neoplastic cells (62).

Table 1-1 The consensus molecular subtypes of colorectal cancer

CMS1 MSI immune	CMS2 Canonical	CMS3 Metabolic	CMS4 Mesenchymal
MSI, CIMP high, hypermutation	SCNA high	Mixed MSI status, SCNA low, CIMP low	SCNA high
BRAF mutation		KRAS mutations	
Immune infiltration and activation	WNT and MYC activation	Metabolic deregulation	Stromal infiltration, TGF- β activation, angiogenesis
Worse survival after relapse			Worse relapse-free and overall survival

CIMP, CpG island methylator phenotype; MSI, microsatellite instability; SCNA, somatic copy number alterations. This table is adapted from Guinney et al, 2015, figure 5 (59).

1.4.2 Colorectal intrinsic subtypes

An alternative classification system, the colorectal cancer intrinsic subtypes (CRIS) aims to classify tumours according to the gene expression of neoplastic cells alone excluding the heterogeneous stroma (62). CRIS categorises CRC into five novel transcriptional classes. CRIS-A tumours are commonly BRAF-mutated, MSI, KRAS-mutated and MSS, a group for which no targeted treatment is currently available (62). CRIS-A tumours exhibit strong glycolytic/hypoxic signatures suggesting they may respond to anti-metabolic therapy (62). CRIS-B tumours are invasive, associated with poor prognosis and high *TGF β* signalling (62). CRIS-B tumours may respond to treatments targeting *TGF β* stromal signals (62). CRIS-C tumours are dependent on *EGFR* signals and sensitive to treatment with EGFR antibodies (62). CRIS-D tumours display high expression of IGF2 associated with resistance to *EGFR* blockade in *KRAS* wild type tumours (62). The final subtype, CRIS-E includes *KRAS*-mutated, paneth cell-like CIN tumours with high WNT pathway activity which are commonly resistant to treatment with anti-EGFR antibodies. Wnt inhibitors may be a useful treatment option in this group (62).

The CRIS subtypes can be used in conjunction with the extent of stromal infiltration to predict prognosis and treatment response (62). Patients with high levels of cancer-associated fibroblasts in CRIS-B tumours have a poorer prognosis than those with low cancer-associated fibroblasts and are likely to benefit from adjuvant therapy (62).

1.4.3 Molecular prognostic classifiers

Only a small fraction of features identified through next generation sequencing (NGS) techniques have established clinical effects which can be used to predict clinical outcomes. A group of such features can be defined as a biomolecular classifier. The development of NGS classifiers has some inherent problems, sample sizes are commonly small, datasets have high dimensionality, the number of features is orders of magnitude greater than the number of samples. Genes are connected via pathways resulting in multicollinearity of features (63). Extracting the features capable of predicting a phenotype in subsequent patient cohorts from such complex data is mathematically challenging. Many mathematical methods exist to extract the

significant features from high-dimensional molecular data, three models are discussed below. Partial least squares discriminatory analysis (PLS-DA) models, find features which maximise sample co-variance for a given phenotype (64). PLS-DA is well suited to the dimensionality and multicollinearity of NGS data (64). PLS-DA is a linear model which requires normally distributed data. Support vector machine (SVM) models find a linear or non-linear hyperplane which maximises the separation between phenotypes (64). Random forest (RF) models are formed of an ensemble of classification trees, the dataset is repeatedly sampled and the trees aggregated into a predictive ensemble (64). Although these models achieve good prediction accuracies they do however tend to produce an excess of features increasing the risk of overfitting and prediction variability (63). To make clinical predictions these models require a restricted list of features before entering the verification phase of classifier development. The analysis of the thousands to millions of combinations of features identified by NGS is not computationally feasible for large datasets. Several statistical methods of feature selection have been described to extract a restricted set of variables reducing the dimensionality of the dataset. One strategy is to filter the features before building the classifier. In filter feature selection features are ranked according to a univariate or multivariate metric (63). Filter feature selection is fast, however as it is performed before model construction features may not be optimal for classifier performance, additionally, the resulting classifier is large (63). A second method combines feature selection and model construction into a single step by including a penalisation constraint into the building of the classifier, the embedded approach limits the number of features with non-zero coefficients in the model (63). Embedded methods are fast, but the resulting classifier is large and unstable on repetition. Embedded feature selection uses one type of classifier. Several studies suggest optimal classifier performance results from the application of distinct classifier models depending on the dataset structure (63). Wrapper feature selection can be applied to any classifier and iteratively selects features based upon the classifier performance. With wrapper feature selection each new subset is used to train the model which is tested on a holdout dataset (63).

1.5 The presentation of colorectal cancer

CRC may be asymptomatic, with 10% of cases detected by routine screening in the UK (3). The introduction of a UK National screening program for CRC has led to a shift in the mode of presentation from symptomatic to screening, with resulting earlier stage at diagnosis and a reduction in emergency and metastatic presentations (65). The UK CRC screening program began in 2006 and targets men and women aged between 60 and 74 years. Screening comprises biennial faecal occult blood testing (FOBT) plus or minus colonoscopy or CT colonoscopy as appropriate for those with positive FOBT (66).

Symptomatic CRC commonly presents with bleeding per rectum, abdominal pain, iron deficiency anaemia and or change in bowel habit. Emergency presentations resulting from advanced disease include bowel obstruction or perforation (67).

1.6 The diagnosis of colorectal cancer

A diagnosis of CRC is made by clinical history and examination. Diagnosis is confirmed with colonoscopy to visualise the tumour extent and its anatomical location. At the time of colonoscopy, the tumour is marked with tattoo dye to aid in its localisation during subsequent surgery. Tumour biopsies allow histological confirmation and categorisation.

1.7 The staging and grading of colorectal cancer

Computed tomography (CT) imaging of the chest, abdomen and pelvis is used to estimate the stage of disease pre-operatively (68). The pathological grade from 1–3 is established with endoscopic biopsy. Three pathological staging systems are used in the UK to describe the spread and extent of CRC. The Dukes system (69), the tumour node metastasis system (TNM) (70), and a numbered staging system from I-IV. Staging systems describe the extent of local tumour invasion, the presence, location and number of nodal metastasis and the presence and location of metastasis (Table 1.2).

Table 1-2 TNM staging colorectal cancer

Stage	TNM						Dukes
	Tumour		Node		Metastasis		
Stage I	Tis	Carcinoma in situ	N0		M0		Dukes A
	T1	Invades submucosa					
	T2	Invades muscularis propria					
Stage II	T3	Invades through muscularis propria into pericorectal tissues	N0		M0		Dukes B
	T4a	Penetrates surface of visceral peritoneum					
	T4b	Directly invades, adherent to other organs			M0		
Stage III	T1-4		N1	1-3 regional nodes	M0		Dukes C
			N2	4 + regional nodes			
Stage IV	T1-4		N1-2		M1a	One organ or site	Dukes D
					M1b	More than one site, organ	
					M1c	Peritoneum with or without organ	

Table adapted from Brierley et al, 2017 (70).

1.8 Mechanism and routes of metastasis in CRC

Up to 26% of CRC patients present with metastasis, 5-year survival for such patients is just 7% (3). Metastasis is the spread, survival and growth of tumour cells in organs distant to their primary tissue of origin. CRC commonly metastasises in order of frequency: to the liver, regional lymph nodes, peritoneum and lungs (71).

Paget in 1889 described the 'seed and soil' hypothesis of metastasis. Neoplasms are biologically heterogeneous, diverse cell populations (72). Metastasis selects for cells 'seeds', able to succeed in invasion, embolization, survival and multiplication (72). Different organ microenvironments 'soil' are biologically unique, expressing different cell receptors and growth factors which influence the phenotype of tumours which grow there (72). Metastasis depends on cross-talk between the metastasising cell and micro-environment in which they grow.

CRC metastasis occurs via several mechanisms: Haematogenous, tumour cells detach and invade the circulation, cell arrest occurs via attachment to endothelial cells or decreased vessel luminal size. Extravasation and transmigration are followed by growth or dormancy (73). Lymphatic, tumour cell invasion of lymphatic channels results in regional lymph node metastasis (73). Direct invasion, tumour growth breaches the bowel's serosal surface and further growth is contiguous through adjacent organs and peritoneum (74). Cell seeding, tumour rupture or capsular breach (iatrogenic or spontaneous) allows tumour cells to seed and grow throughout the abdominal cavity (Figure 1.5) (74).

1.9 Peritoneal metastasis

Colorectal peritoneal metastasis (CPM) develop in up to 15% of CRC. CPM occurs in 10% of CRC at diagnosis (synchronous) and in 25% of patients with recurrent disease (metachronous) (75-77). CPM is increased in patients with advanced tumour stage, nodal metastasis, vascular invasion, tumour obstruction or perforation (75-77). CPM is isolated in 3-10% of all cases of CRC, presenting in the absence of distant metastasis (75, 77).

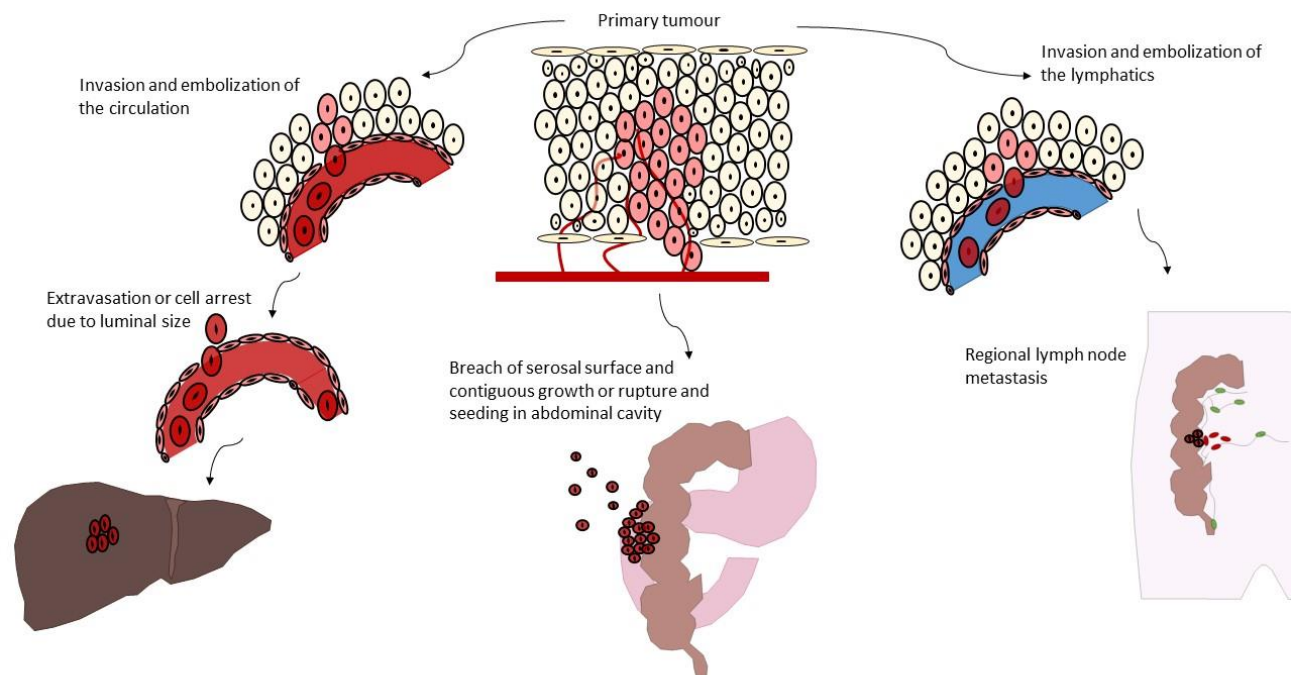


Figure 1-5 Routes of CRC metastasis

Colorectal cancer metastasis occurs via common mechanisms. Haematogenous, lymphatic, direct invasion or cell seeding by tumour rupture or capsular breach. This figure is adapted from Francia et al, 2011, figure 1 (78).

1.10 The treatment of peritoneal metastasis

1.10.1 Palliative chemotherapy

With palliation survival with CPM is limited to just 6 months (79). Traditionally, CPM was considered a terminal stage of CRC. Treatment was focused on palliation and included palliative surgery (tumour de-bulking or a de-functioning stoma) and palliative Fluoropyrimidine based chemotherapy. Overall survival with palliative treatment ranged from between 8–12 months (80, 81). In truth, outcomes were likely to be improved for isolated CPM as study populations were mixed to include patients with distant metastasis (81). This treatment rationale was challenged by a surgeon from the USA, Dr Paul Sugarbaker who defined CPM as locoregional disease suitable for aggressive treatment (82).

Current cytotoxic treatment of metastatic CRC remains Fluoropyrimidine based intravenous 5-fluorouracil (5-FU) or its oral equivalent Capecitabine. Combination therapy, when tolerated, produces improved response rates, progression-free and overall survival (83). Two combinations with similar activity are commonly used, 5-Fu/Leucovorin/Oxaliplatin (FOLFOX) or 5-Fu/Leucovorin/ Irinotecan (FOLFIRI) (83). Patients progressing on primary treatment can be switched to secondary chemotherapy, progression on FOLFOX would be followed by FOLFIRI and vice versa (83).

Biologically targeted agents can be used in combination with intravenous chemotherapy. Biologics include monoclonal antibodies (Bevacizumab) active against vascular endothelial growth factor receptor (VEGF) or Cetuximab, active against epidermal growth factor receptor (EGFR).

Cytotoxic combination treatments have improved survival for CPM. A retrospective analysis in patients with metachronous CPM treated with biological targeted agents and systemic chemotherapy revealed significant improvements in OS compared with systemic chemotherapy alone (13 months vs. 20.3 months) (84).

Despite optimal cytotoxic and biological treatments the survival of patients with CPM is reduced when compared with patients with other sites of distant metastasis suggesting a relative resistance to systemic chemotherapy in the peritoneum. Franko

et al performed a subgroup analysis of two, phase 3 chemotherapy trials. Whilst overall survival was improved for patients with CPM with Oxaliplatin based chemotherapy as compared to Irinotecan, (15.7 months vs. 12.7 months) their survival was significantly reduced compared with patients with metastatic CRC and no PM (15.7 months vs. 20.9) (85). The CAIRO II trial added Bevacizumab (a monoclonal antibody agent active against vascular endothelial growth factor (VEGF)) and or Cetuximab (a monoclonal antibody agent active against EGFR) to Capecitabine and Oxaliplatin based chemotherapy. The addition of monoclonal antibodies improved OS from 10.4 to 15.2 months for patients with CPM. Survival for patients without CPM and with other sites of distant metastasis, however, was significantly greater (20.9 months vs. 15.2) (86). Whilst systemic cytotoxic and biologic therapies improve outcomes for patients with CPM, survival with peritoneal metastasis remains significantly lower than for patients with other sites of metastasis. Systemic cytotoxic and biologic therapy remains the standard of care in some cases: to downstage disease potentially converting unresectable to resectable, to prolong survival, improve symptoms and maintain quality of life for patients with unresectable disease (83).

1.10.2 Cytoreductive surgery and heated intraperitoneal chemotherapy, history

Surgical tumour de-bulking was first described in the 1930s for patients with advanced ovarian cancer and subsequently pseudomyxoma peritonei. In the 1970s de-bulking surgery became more radical with the aim of removal of all tumour deposits greater than 1cm in diameter. This was termed cytoreductive surgery (CRS) (87, 88).

In the 1980s several trials demonstrated the pharmacokinetic advantages of intraperitoneal vs. systemic chemotherapy for peritoneal disease allowing the administration of higher dose delivery locally with reduced systemic side effects (89). Ovarian cancer patients undergoing IP chemotherapy for low volume peritoneal disease (<2cm deposits) showed significantly improved OS when compared with high volume disease (>2cm deposits), (49 months vs. 8 months) (90). Hyperthermia works synergistically with cytotoxic drugs increasing their efficacy via increased tissue perfusion and drug delivery as well as causing direct tumour cell damage (91). Intraperitoneal chemotherapy is therefore heated to improve its efficacy, heated intraperitoneal chemotherapy (HIPEC).

CRS & HIPEC was first applied to gastrointestinal cancers and specifically CRC by Dr Paul Sugarbaker in the 1980s (92). Several phase I and II studies followed which showed good tolerability of intraperitoneal chemotherapy, with encouraging improvements in survival but relatively high post-operative morbidity (93, 94).

1.10.3 Cytoreductive surgery and heated intraperitoneal chemotherapy, technique

The technique for CRS & HIPEC includes a methodical laparotomy, peritonectomy, omentectomy and multi-visceral resections, as required depending on the spread of peritoneal metastasis (95).

The extent of peritoneal tumour burden at surgery is described by the peritoneal carcinomatosis index (PCI) (Figure 1.6) (96). This is an additive score from 0-39. The peritoneal cavity is divided into 13 regions and in each region, the largest tumour nodule is measured. Tumour deposits > 5cm score 3, tumour deposits 0.5-5cm score 2, tumours <0.5cm score 1 and where no tumour is visible the score is 0. The scores from each region are added with a final score ranging from 0-39. Increased PCI is a poor prognostic indicator for CPM (97). Other scoring systems in use internationally describing the extent of tumour burden include the Gilly peritoneal carcinomatosis staging and Dutch simplified peritoneal cancer index. The Peritoneal Surface Disease Severity Score (PSDSS) incorporates PCI, clinical symptoms and histological grade to predict response to CRS & HIPEC (98).

The completeness of cytoreduction score (CC score) describes the degree of macroscopic tumour remaining after CRS and the likelihood of benefit from intraperitoneal chemotherapy (99). Patients with no residual tumour score CC0, residual tumour of <0.25cm CC1, residual tumour 0.25-2.5cm CC2 and >2.5cm CC3. CC scores of CC0 and CC1 are termed complete cytoreduction and gain the greatest benefit from HIPEC which can penetrate the peritoneum to a depth of between 3-5mm (100) (Table 1.3).

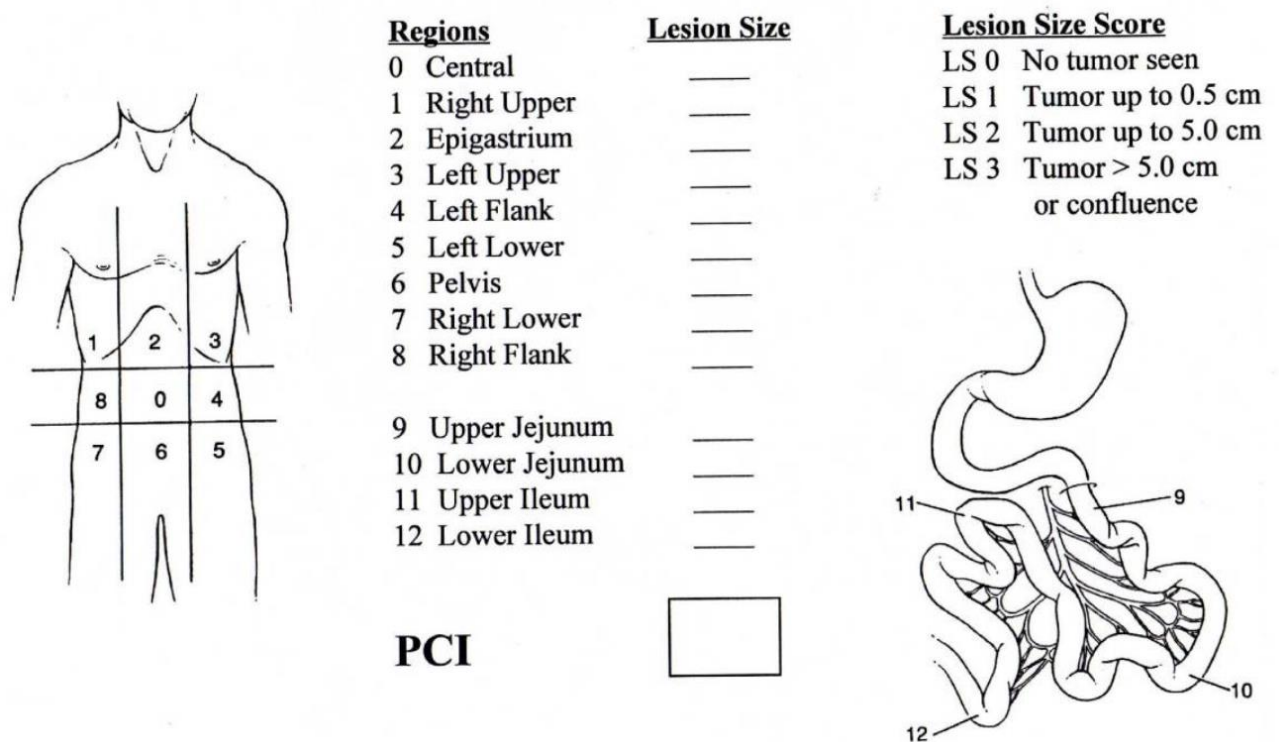


Figure 1-6 The peritoneal carcinomatosis index (PCI)

The extent of peritoneal metastasis is described by the peritoneal carcinomatosis index (PCI). The peritoneal cavity is divided into 13 regions and in each region, the largest tumour nodule is measured. Tumour deposits > 5cm score 3, tumour deposits 0.5-5cm score 2, tumours <0.5cm score 1 and where no tumour is visible tumour the score is 0. The scores from each region are added with a final score ranging from 0-39 (101). This figure is reprinted from Harmon et al, 2005, figure 5 (101).

Table 1-3 The completeness of cytoreduction, (CC) score.

Residual peritoneal disease after CRS	CC score
No visible disease	CC0
Residual disease <0.25cm	CC1
Residual disease 0.25-2.5cm	CC2
Residual disease >2.5cm	CC3

The completeness of cytoreduction score (CC score) describes the degree of macroscopic tumour remaining after CRS and the likelihood of benefit from IP chemotherapy (99). This table is adapted from Sugarbaker et al, 1999, figure 5 (99).

1.10.4 Heated intraperitoneal chemotherapy, technique

HIPEC is delivered via a variety of techniques. The open or coliseum technique suspends the skin edges from a Thompson® or Omni-tract® retractor using a running suture. A plastic sheet is sewn between the skin edges. Fixed-flow rate pumps placed inside the abdomen maintain a flow rate of 1L/minute, a heat exchanger and temperature probes ensures the perfusate is maintained at between 41-43°C (102). An alternative closed technique involves the same pumps and probes with the closure of the laparotomy giving a closed perfusion circuit. Following both techniques the abdomen is opened, perfusate washed away anastomoses formed and the abdomen closed.

The coliseum technique allows improved access to the abdomen for the operating surgeon, uniform exposure of all abdominal areas to heat and chemotherapy and protection to theatre staff from direct chemotherapy contact. The closed technique rapidly achieves and maintains hyperthermia, there is reduced risk of chemotherapy contact for theatre staff, both as direct and aerosolized. The distribution of chemotherapy and heat is however less uniform, pooling of heat and chemotherapy may result in undertreatment of some areas and increased risks of morbidity such as ileus, perforation and fistula formation (103). No randomised controlled trials exist to compare the two techniques, and the method used varies according to departmental preference and experience (104).

Chemotherapy drugs used in HIPEC are varied. Desirable features of HIPEC drugs include slow clearance from the peritoneal cavity, synergy with hyperthermia as well as rapid hepatic metabolism and renal clearance to avoid systemic effects (105). Commonly used agents include Mitomycin C (an alkylating antibiotic) or Oxaliplatin (a platinum derivative) which is given simultaneously with intravenous 5-fluorouracil and Leucovorin. No prospective or randomised studies exist comparing HIPEC chemotherapy drugs in CRC, superiority has not been demonstrated in retrospective studies and the drugs used vary according to local preference (106, 107).

1.10.5 Adjuvant intraperitoneal chemotherapy

CRS & HIPEC may be followed by adjuvant intraperitoneal chemotherapy. Early postoperative intraperitoneal chemotherapy (EPIC) is given in the immediate postoperative period (days 1-5) via abdominal drains. Retrospective cohorts including patients undergoing a variety of intraperitoneal chemotherapy, CRS & HIPEC, CRS & EPIC or CRS & HIPEC & EPIC found no significant difference in oncological outcome despite the addition of EPIC (108-110). Sequential postoperative intraperitoneal chemotherapy (SPIC) is given over a prolonged postoperative period, in six, 6-day cycles. Two retrospective cohorts comparing CRS & HIPEC and CRS & SPIC, both demonstrated improved oncological outcomes with HIPEC (111, 112). One randomised trial, (ICARuS), comparing CRS & HIPEC with CRS & EPIC is ongoing and due to complete in 2019 (113).

1.10.6 Cytoreductive surgery and heated intraperitoneal chemotherapy, oncological outcomes

To date, one randomised controlled trial (RCT) exists comparing CRS & HIPEC to systemic intravenous chemotherapy for CPM (114). CRS & HIPEC was associated with significant improvements in DFS (22.2 months vs. 12.6 months) and a 5-year OS of 45% for those patients with complete cytoreduction. (115) One further RCT was attempted in 2016 comparing CRS & HIPEC to systemic chemotherapy, however, it terminated with just 24 patients in each arm due to recruitment difficulties, median OS was 25 months with CRS & HIPEC vs. 18 months with systemic chemotherapy (116). Four case-controlled studies compared CRS & HIPEC to systemic IV chemotherapy. CRS & HIPEC resulted in a median OS of 34–62.7 months compared to 16.8–23.9 months for systemic IV chemotherapy (111, 112, 117, 118). The remainder of studies concerning the use of CRS & HIPEC for CPM are case series reporting OS ranging from 3.7-35 months and 5-year OS of between 0-55% (108-110, 119-139).

Following CRS & HIPEC OS is improved for patients with a low PCI, in whom complete cytoreduction is achieved. None the less the oncological outcomes for CRS & HIPEC vary widely across the literature with limited consensus on patient suitability for treatment (140).

1.10.7 Cytoreductive surgery and heated intraperitoneal chemotherapy, morbidity and mortality

CRS & HIPEC can last up to 12 hours in duration and necessitates a high dependency or intensive care stay and up to a 2-week length of stay. CRS & HIPEC is associated with high postoperative mortality of 0-12% and morbidity of 7–63% (110-112, 120, 121, 123-131, 133, 135, 138, 141-145). Common postoperative complications include fistula formation 2-10% (108, 110, 125-130, 142), abscess formation 1.8-21% (110, 121, 125, 128, 130, 142), and anastomotic leak in 5-10.4% (124-126, 130). Re-operation, as a result, is necessary in between 4–20.8% of cases (111, 121, 125, 128-130).

1.11 Biomarkers of colorectal peritoneal metastasis, circulating tumour DNA

A cancer biomarker is secreted by a tumour or by the body as a reaction to the presence of a tumour. Biomarkers can be used in cancer diagnosis, to guide treatment or as markers of tumour recurrence.

Fragments of circulating tumour DNA (ctDNA) circulate in the bloodstream. The mechanisms by which these fragments enter the circulation remains unclear but three potential origins have been proposed, apoptosis or necrosis of tumour cells, release from viable tumour cells and from circulating tumour cells (146). ctDNA represents a proportion of ccfDNA normally present at low levels in healthy individuals (146). In health, ccfDNA is thought to originate from haematopoietic cells through apoptosis, necrosis, secretion or inefficient phagocytosis (146). ccfDNA increases in response to several conditions including myocardial infarction, serious infections, inflammation and pregnancy (146).

ctDNA is an attractive option as a tumour marker. As it is composed of tumour DNA it has potential as a non-invasive, liquid biopsy, which can determine genomic and epigenetic markers allowing the delivery of targeted personalised treatments without the need for surgical resection or biopsy. Although ctDNA fragments are short (150 – 250 bp), evidence suggests they provide a comprehensive view of the tumour genome

as they are released from multiple tumour regions or foci, in some cases, they have been able to detect mutations missed in corresponding tumour samples (147).

1.12 The biology of colorectal peritoneal metastasis

There is a paucity of research regarding the biology of CPM and in particular the biology of the subset of patients with isolated and limited CPM suitable for treatment with CRS & HIPEC.

Sequencing of a variety of tumours has determined high intratumor heterogeneity in primary tumours with gene expression signatures of good and poor prognosis detected within the same tumour (148). It is hypothesised that metastasis develop from heterogenous primary tumour clones selected through competitive selection and late clonal expansion (148). Clonal cells may acquire differential organ specific metastatic traits (148). Comparison of genetic variations by WGS reveal disparity between primary tumour and metastasis suggesting additional mutations are acquired after metastasis (148).

Several retrospective cohort studies have examined the biology of advanced or recurrent primary CRC. In these studies, genes commonly mutated in CRC are correlated with clinicopathological features such as advanced tumour size, the presence of lymph node, distant or peritoneal metastasis and survival outcomes.

1.12.1 MAPK pathway mutations

Four retrospective cohort studies examined the prevalence of *KRAS* mutation in patients with advanced and recurrent CRC and reported the proportion with synchronous or metachronous PM. Three studies found no increased prevalence of PM in patients with *KRAS* mutation (149-151). One study found an increased frequency of PM in patients with *KRAS* mutation, 11.1% vs. 6% wild type (152).

Four retrospective cohort studies examined the prevalence of mutation of the Raf-kinase *BRAF* in advanced or recurrent CRC and found high rates of synchronous or metachronous PM in patients with a *BRAF* mutation of between 30-60% vs. 7-18% in *BRAF* wild type tumours (149-151, 153). *BRAF* mutation was a marker of reduced OS on multivariate analysis alongside poor performance status, increased age, mucinous

or signet ring histology and greater than three sites of distant metastasis (149-151, 153).

1.12.2 PI3K pathway mutations

Three retrospective cohorts examined mutation of the PI3K pathway in patients with advanced or recurrent CRC and found no relationship with PM (150-152).

1.11.3 Metastasis-associated colon cancer 1 gene (MACC1)

MACC1 is a key regulator of hepatocyte growth factor. *MACC1* is involved in cellular growth, epithelial-mesenchymal transition, angiogenesis, cell motility, invasiveness, and metastasis. Shirahata et al examined the expression of *MACC1* in a retrospective cohort of patients with advanced and recurrent primary CRCs (154). They found *MACC1* was upregulated in patients with CPM and advanced tumour stage (154). These were primary CPM making it difficult to determine whether *MACC1* was a marker of advanced stage, PM or both.

1.12.4 Histone acetyltransferase (Tip60)

The *Tip60* gene regulates DNA repair and apoptosis. Sakuraba et al examined the expression of Tip60 in a cohort of advanced and recurrent primary CRCs (155). They demonstrated that Tip60 was downregulated in patients with large tumour size, poor differentiation and the presence of peritoneal or distant metastasis making it difficult to determine whether Tip60 was a marker of advanced stage, PM or both.

1.12.5 Neurone glial-related cell adhesion molecule (Nr-CAM)

Nr-CAM is involved in the development and function of the nervous system, it is also overexpressed in CRC and has been shown to enhance cell motility and invasiveness. Chan et al examined the expression of *Nr-CAM* in a retrospective cohort of advanced and recurrent primary CRCs (156). *Nr-CAM* expression was increased in patients with lymph node and distant metastasis including PM, vascular invasion and *p53* expression, *Nr-CAM* expression was not related to tumour size (156). Patients with over-expression of *Nr-CAM* had a poor response to 5-fluorouracil-based

chemotherapy, however, given the retrospective study design, this may be a result of selection bias (156).

1.12.6 Collagen Triple Helix Repeat Containing 1 (CTHRC1)

CTHRC1 is a protein-coding gene associated with digestive tract tumours which may contribute to cell invasion and metastasis. Tan et al examined the expression of CTHRC1 in patients with advanced and recurrent CRC. Patients with CPM had increased expression of CTHRC1 which was associated with reduced OS, 31.36 +/- 5.63 months vs. 57.84 +/- 3.69 months (157). However, no information was provided about the surgical or systemic treatment of either the primary CRC or the PM meaning that the difference in OS may be due to un-reported treatment differences.

1.12.7 Increased *TOPO1*, *Cox2*, *ERCC1* and reduced *RRM1* protein expression in CPM

El-Deiry et al performed next-generation sequencing using a panel of 45 known biomarkers with therapeutic targets. This large cohort included 6892 metastatic CRC of which 465 were CPM with unmatched controls (158). The *TOPO1* gene was upregulated in CPM. *TOPO1* is the target of Irinotecan based chemotherapy and it has been hypothesised increased expression predicts increased benefit from Irinotecan (158). *ERCC1* expression was upregulated in CPM. *ERCC1* gene expression is predictive of resistance to Oxalaplatin based chemotherapy (158). *RRM1* expression was reduced in CPM. Reduced expression of *RRM1* suggests an improved benefit to gemcitabine-based chemotherapy, an agent not commonly in use in CPM (158).

Whilst these studies demonstrate the prevalence of known mutations in subgroups of patients with CPM, some problems exist which mean we cannot extrapolate the findings to a population with isolated CPM. All retrospective patient cohorts included were undergoing palliative chemotherapy for advanced or recurrent CRC, there was poor reporting of treatment detail (surgical or adjuvant) or it's intent as palliative or curative. Mutation analysis was performed most commonly on primary CRC resection specimens and assumed to be identical in synchronous or metachronous CPM.

The cohorts included large proportions of patients with PM in addition to numerous distant sites of metastasis. Whilst this suggests mutation status may be a risk factor for advanced disease or metastasis, analysis were not specific to isolated CPM.

1.13 Conclusion

CPM occurs in a small but significant number of patients with CRC. With palliation few patients survive greater than 6 months, systemic chemotherapy can increase survival to between 12–20 months depending upon patient selection, tumour biology and chemotherapy choice, however peritoneal metastasis show relative resistance to systemic chemotherapy when compared with other metastasis.

Selected patients can be treated with curative intent using CRS & HIPEC (117). Following NICE guidance on CRS & HIPEC for peritoneal carcinomatosis, the technique has been commissioned for CPM in England since 2013. Despite this, ongoing controversy exists regarding the role of CRS & HIPEC for CPM.

There is a paucity of research regarding the biology of CPM and in particular the biology of the subset of patients with isolated and limited CPM suitable for treatment with CRS & HIPEC. Given the significant morbidity and mortality associated with CRS & HIPEC and wide variation in oncological outcomes reported in the available literature, there is significant room for improvement in the identification of prognostic factors for CPM and patient selection for CRS & HIPEC.

1.14 Hypotheses

1. Metachronous CPM is will acquire biological changes distinct from matched primary CRC in keeping with evolutionary adaptation, disease progression and the development of metastatic disease.
2. A clinical and molecular classifier can be developed to predict prognosis and response to the best available treatment for CPM, CRS & HIPEC.
3. Circulating tumour DNA can be used as a liquid biopsy prior to surgery, guiding treatment, and following surgery to determine residual disease and the need for further treatment.

1.15 Aims

1. For patients with CPM undergoing treatment with CRS & HIPEC, to define the significant clinical prognostic factors via systematic review and meta-analysis of the existing literature.
2. To determine the gene expression methylation and somatic mutation profile associated with the development of CPM and conferring 'poor prognosis' in patients with CPM following CRS & HIPEC (DFS equivalent to palliative chemotherapy treatment, <12 months).
3. To develop a biomolecular classifier (gene expression and methylation) for use with known clinical factors, capable of predicting prognosis for patients with CPM.
4. To determine whether ctDNA mutations correlate with those found in CPM specimens and can be used to predict poor prognosis and recurrence following CRS & HIPEC.

Chapter 2 Materials and Methods

2.1 List of reagents and suppliers

2.1.1 General consumables and equipment common to all techniques

DNA LoBind tube, 1.5ml, Eppendorf AG, Hamburg, Germany

Conical tube, 15ml, Corning, New York, USA

FALCON conical tube, 15ml, Corning, New York, USA

8-Strip 'Non-Flex' PCR Tubes with Individual Attached Caps, 0.2ml, Starlab, Milton Keynes, UK

Cryotube, 1.8ml, Thermo Fisher Scientific, Northumberland, UK

Hard-Shell PCR Plates, 96-well 0.8mL, BIO-RAD, Vienna, Austria

Abgene™ Polypropylene Deepwell Storage Plate, 96-well 0.8mL,

Thermo Fisher Scientific, Northumberland, UK

Adhesive PCR Plate Foils, Thermo Fisher Scientific, Northumberland, UK

Nunc™ 96-Well Cap Mats, Thermo Fisher Scientific, Northumberland, UK

Magnetic stand- 96, Ambion, Northumberland, UK

MagnaRack™, Invitrogen, California, United States

Trough, Thermo Fisher Scientific, Northumberland, UK

Sodium hydroxide pellets, Sigma-Aldrich, Missouri, United States

Nuclease free water, Omega-Bio-tek, Georgia, USA

Isopropanol, Thermo Fisher Scientific, Northumberland, UK

Ethanol absolute, Thermo Fisher Scientific, Northumberland, UK

Tris(hydroxymethyl)aminomethane, Sigma-Aldrich, Missouri, United States

C1000 Touch Thermal Cycler, Bio-Rad, California, USA

Plate centrifuge, Greiner bio-one, Kremsmünster, Austria

Hereaus Megafuge 16R Centrifuge, Thermo Fisher, Northumberland, UK

2.1.2 Tissue collection and storage

PAXgene Blood ccfDNA tube, Qiagen, Manchester UK

Vacutainer™ Hemogard Closure Plastic K2-EDTA Tube, Thermo Fisher Scientific, Northumberland, UK

RNA later stabilisation solution, Thermo Fisher Scientific, Northumberland, UK

2.1.3 Extraction of RNA and DNA

TruXTRAC FFPE total NA Kit, Covaris, Brighton, UK

E220 evolution focused-ultrasonicator, Covaris, Brighton, UK

TURBO DNA-free kit, Thermo Fisher Scientific, Northumberland, UK

RNA Clean & Concentrator – 5, Zymo research, California, USA

QIAamp DNA FFPE kit, Qiagen, Manchester, UK

AllPrep DNA/RNA Mini Kit, Qiagen, Manchester, UK

TissueLyser II, Qiagen, Manchester, UK

TissueLyser Adapter Set 2 x 24, Qiagen, Manchester, UK

Stainless steel beads, 5mm, Qiagen, Manchester, UK

Maxwell® RSC instrument, Promega, Wisconsin, USA

Maxwell® RSC Buffy Coat DNA Kit, Promega, Wisconsin, USA

Cell3™ Xtract, Nonacus, Birmingham, UK

2.1.4 Quantification of DNA and RNA

Qubit™ 3.0 Fluorometer, Thermo Fisher Scientific, Northumberland, UK

Qubit™ dsDNA HS (high sensitivity assay kit) and Qubit™ RNA HS (high sensitivity) assay kit, Thermo Fisher Scientific, Northumberland, UK

2200 TapeStation Nucleic Acid System and the 2200 TapeStation Software A.01.05, Agilent, Santa Clara, USA

Genomic DNA ScreenTape and the Genomic DNA Reagents, Agilent, Santa Clara, USA

D1000 DNA ScreenTape and the D1000 reagents, Agilent, Santa Clara, USA

High Sensitivity RNA ScreenTape and the RNA High sensitivity reagents, Agilent, Santa Clara, USA

Loading tips, Agilent, Santa Clara, USA

Optical Tube 8x Strip, Agilent, Santa Clara, USA

2100 Bio-analyser, Agilent, Santa Clara, USA

High sensitivity DNA chips, Agilent, Santa Clara, USA

2.1.5 RNA library preparation

Quant Seq 3' mRNA-Seq Library Prep kit, Lexogen, Vienna, Austria

TruSeq stranded total RNA kit, Illumina, California, USA

SuperScript II Reverse Transcriptase, Thermo Fisher, Northumberland, UK

Agencourt RNAClean XP beads, Beckman Coulter, California, USA

2.1.6 DNA sequencing

NextSeq500, Illumina, San Diego, California, United States

75-cycle High Output flow cell, Illumina, San Diego, California, United States

150-cycle High Output flow cell, Illumina, San Diego, California, United States

PhiX, Illumina, San Diego, California, United States

Partek® flow® genomics suite software package, Partek, St Louis, Missouri, United States

2.1.7 Methylation arrays

EZ DNA Methylation Kit, Zymo, California, USA

ZR-96 DNA Clean & Concentrator-5 Kit (2 x 96 Preps), Zymo, California, USA

Infinium HD FFPE Restore, Illumina, California, USA

Infinium MethylationEPIC BeadChip Kit (32 samples), Illumina, California, USA

Multi-Sample BeadChip Alignment Fixture, Illumina, California, USA

Infinium BeadChip Expansion Package, Illumina, California, USA

Infinium BeadChip Wash rack, Illumina, California, USA

Infinium BeadChip Glass tray, Illumina, California, USA

Te-Flow 48 position rack chamber, Tecan, Männedorf, Switzerland

Infinium Hybridization Chamber, Illumina, California, USA

Infinium Hybridization Chamber Inserts (8), Illumina, California, USA

ALPS 25 Manual Heat Sealer, Thermo Fisher, Northumberland, UK

Illumina Hybridization Oven (220V), Illumina, California, USA

High-Speed Microplate Shaker, Illumina, California, USA

Star-Vitrum-Desiccator Horizontal, borosilicate glass 3.3, SICCO, Grünsfeld, Germany

Hybex Micro sample Incubator, SciGene, California, USA,

I-Scan®, 110 V/220 V, Illumina, California, USA

GenomeStudio 2.0.4, Illumina, California, USA

RStudio desktop (2015), Massachusetts, USA

2.1.8 Exome sequencing

8-microTUBE-50 AFA Fibre Strip V2 (12), Covaris, Brighton, UK

E220 evolution focused-ultrasonicator, Covaris, Brighton, UK

TruSeq® Exome Kit (96 Samples), Illumina, California, USA

2.1.9 DNA panel sequencing

QIAseq™ Targeted DNA Panel CDHS-14542Z-1197, Qiagen, Hilden, Germany

2.2 Methods

2.2.1 Ethics

This study obtained ethical approval from the North West Haydock Research Ethics Committee (15/NW/0079), project ID (171283). Participants gave informed consent.

2.2.2 Clinical sample preparation

Retrospective patients were recruited from an internally held database of all patients undergoing CRS & HIPEC at Good Hope hospital from 2011–2017. Patients with CPM (adenocarcinoma), no extra-abdominal metastasis, a CC0 resection and a PCI of < 12 were eligible for inclusion. Patients were divided into two groups. With palliative chemotherapy, DFS is 11-13 months and therefore patients, post-treatment (CRS & HIPEC) with DFS <12 months were defined as the “poor prognosis” cohort (114, 115). Patients undergoing therapy with DFS > 12 months were defined as a “good prognosis” cohort. For each patient tissue was sourced from the colorectal primary tumour, the metachronous colorectal peritoneal metastasis and normal tumour free tissue. Formalin-fixed, paraffin-embedded (FFPE) tumour blocks were sectioned into triplicate and further divided into 5x 8-micron scrolls. Tissue was sourced through the Human Biomaterials Resource Centre (HBRC) at the University of Birmingham. Histologically confirmed, tumour free tissue samples for paired analysis were taken from the cytoreductive specimen.

Prospective patients were recruited from the peritoneal malignancy clinic, given an information sheet to take away and were consented on the day of surgery. Samples were sectioned immediately, preserved in RNAlater solution (Thermo Fisher, AM7021), refrigerated overnight and frozen at -20 °C.

Demographics, tumour and treatment details were collated from the patient electronic record and anonymised via the HBRC. The cohorts were compared, for continuous variables, the student's T-test was applied to normally distributed data and Mann Whitney-U to non-normally distributed data. Categorical variables were compared with the Chi-squared test or Fisher's exact test, a *P*-value of < 0.5 was considered statistically significant. DFS survival between the good and poor prognosis cohorts was

compared using the Kaplan Meier method in IBM SPSS Statistics for Windows, Version 24.0 (159).

At the induction of anaesthesia matched normal whole blood samples were collected in 5 ml EDTA tubes (Thermo Fisher, 367525), centrifuged at 900 xg for 10 minutes, the lymphocyte rich buffy coat collected and stored at -20 °C. Blood samples for circulating tumour DNA (ctDNA) were collected into 10ml PAXgene tubes (Qiagen, 768115) at the induction of anaesthesia, postoperatively and at 3, 6, 9 and 12 months follow up appointments. The PAXgene tube was centrifuged for 15 minutes at 1900xg, plasma pipetted into a 15 ml falcon tube and centrifuged for a further 10 minutes at 1900xg. Plasma was stored in Cryotubes (Thermo Fisher, 374500TS) at -20 °C.

2.2.3 Nucleic acid extraction

2.2.3.1 DNA & RNA extraction from retrospective FFPE tissue samples

DNA and RNA were extracted from FFPE scrolls using Adaptive Focused Acoustics (AFA), the truTRAC® FFPE total NA Kit (Covaris, 520220) and the E220 evolution focused-ultrasonicator (Covaris, 500429) as per the manufacturer's instructions. Briefly, 2x 5 µm FFPE scrolls were loaded into a Covaris microtube with 120 µl of tissue lysis buffer and proteinase-K. The focused-ultrasonicator dissociates the paraffin and re-hydrates the sample. Tissue lysis is achieved by incubation at 56 °C for 15 minutes and centrifuged at 5,000 xg for 15 minutes.

To reverse formalin crosslinking the RNA containing supernatant was collected and heated to 80 °C for 15 minutes. RNA was precipitated and purified using a silica spin column and collection tube, isopropanol, buffers and series of centrifuge and wash steps as per the manufacturer's instructions. In-solution DNase digestion was performed using the TURBO DNA-free kit to remove DNA contaminants (Thermo Fisher, AM1907). RNA columns were incubated with a TURBO DNase mixture for 25 minutes, purified with wash buffer steps and centrifuged as per the manufacturer's instructions. RNA was eluted into a DNA LoBind tube using 30 µl of elution buffer (EB) and stored at -80 °C.

Tissue lysis buffer and proteinase-K mix were added to the DNA pellet, tissues lysis and homogenisation were performed in the focused-ultrasonicator. Proteinase-K

digestion was performed at 56 °C for one hour, followed by incubation at 80 °C for one hour to achieve formalin de-crosslinking. Samples were processed once again in the focussed-ultrasonicator. DNA was precipitated and purified using a silica spin column, collection tube, ethanol and buffers and a series of centrifuge and wash steps as per the manufacturer's instructions. DNA was eluted in buffer warmed to 70 °C and stored at -20°C until further processing.

Samples resulting in low concentrations of RNA were extracted again using the second set of scrolls. The two samples were combined and concentrated using the RNA Clean & Concentrator-5 kit (Zymo, R1015). Samples were combined with RNA binding buffer and a 1x volume of 95% (v/v) ethanol, transferred to a Zymo-Spin Column and centrifuged at 11,000xg for 30 seconds. Two further wash and centrifuge steps were performed with RNA Prep buffer and RNA wash buffer as per the manufacturer's instructions. RNA was eluted into a DNA LoBind tube in 15 µl of EB and stored at -80 °C.

Samples resulting in low concentrations of DNA were extracted again using the second set of scrolls. The two samples were combined and concentrated using the Genomic DNA Clean & Concentrator kit (Zymo, D4066). Briefly, samples were combined with DNA binding buffer, transferred to a Zymo-Spin Column, and centrifuged at 11,000xg for 30 seconds. Two further wash and centrifuge steps were performed with DNA wash buffer as per the manufacturer's instructions. Finally, DNA was eluted into a clean DNA LoBind tube using 50 µl of EB and stored at -20 °C.

For some samples, this did not provide adequate concentrations of DNA for both exome sequencing and methylation arrays. For these, a repeat extraction was performed using an alternative method allowing a higher input of DNA material, the QIAamp DNA FFPE kit (Qiagen, 56404). In brief, up to 8 scrolls were added to a 1.5 ml Eppendorf with 1 ml of xylene and vortexed for 10 seconds to dissociate the paraffin. The samples were centrifuged at full speed and the supernatant discarded. The pellet was lysed by incubation in lysis buffer ATL and proteinase K for one hour at 56 °C. Further incubation at 90 °C for one-hour reverses formalin crosslinking. DNA was precipitated using a QIAamp MinElute silica spin column, collection tube, buffer AL, ethanol and centrifuged as per the manufacturer's instructions. DNA was purified using

a series of buffer wash and centrifuge steps as per the manufacturer's instructions. DNA was eluted from the column using 50 µl of buffer ATE and stored at -20 °C.

2.2.3.2 DNA & RNA extraction from prospective fresh tissue preserved in RNA later

DNA and RNA were extracted from fresh tissue which had been preserved in RNA later solution (Thermo Fisher, AM7021) using the Qiagen AllPrep DNA/RNA mini kit (Qiagen, 80204). In brief, tissues were removed from RNA later, to a 2 ml Eppendorf no more than 30 mg of tissue was added with 600 µl of Lysis buffer (RLT) / B-mercaptoethanol (1 ml / 10 µl). One 5 mm stainless steel bead was added (Qiagen, 69989) and tissue homogenization performed with the Qiagen TissueLyser II (Qiagen, 85300). Following homogenization, the lysate was centrifuged for 3 minutes at maximum speed. The supernatant was transferred to an AllPrep DNA spin column and centrifuged for 30 seconds at $\geq 8000 \times g$. The dry DNA spin column was placed into a new collection tube until further processing. The flow-through continued to RNA purification.

2.2.3.2.1 RNA purification

One volume of 70% (v/v) ethanol was added to the RNA containing flow-through, mixed by pipette, transferred to an RNeasy silica spin column and centrifuged $\geq 8000 \times g$ for 15 seconds. A series of three wash and centrifuge steps of the RNeasy spin column was performed to purify the RNA using RW1 and RPE wash buffers as per manufacturer's instructions. RNA was eluted into a DNA LoBind tube in 50 µl of NFW by centrifuge and stored at -80 °C.

2.2.3.2.2 DNA purification

A series of two wash and centrifuge steps of the AllPrep DNA spin column was performed to purify the DNA using AW1 and AW2 wash buffers as per manufacturer's instructions, centrifuged at $\geq 8000 \times g$ and the flow-through discarded. DNA was eluted from the spin column with 50 µl of EB by centrifuge. To achieve a higher DNA concentration, this step was repeated once. DNA samples were stored at -20 °C.

2.2.3.3 DNA extraction from prospective matched normal blood samples

DNA was extracted from 250 µl of the buffy coat using the Maxwell® RSC Buffy Coat DNA Kit, a cartridge-based magnetic bead extraction method (Promega, AS1540) and the Maxwell® RSCI instrument (Promega, AS4500). In brief, DNA extraction occurs within the pre-filled cartridge, which is split into wells containing lysis buffer, MagneSil™ Paramagnetic Particles (PMPs) and wash buffers. DNA is bound to and moved between wells by the PMPs and the application of a magnetic force to disposable plungers. Cells are lysed in a guanidine-based lysis buffer and mechanical lysis from the disposable plungers. DNA is moved through a series of wash wells to remove impurities. Purified DNA is then eluted from the beads in a tris(hydroxymethyl)aminomethane (TRIS), Ethylenediaminetetraacetic acid (EDTA) based lysis buffer heated to 56 °C.

2.2.3.4 Cell-free DNA extraction from plasma samples

A variety of methods are available for the collection of ctDNA. The proportion of ctDNA in total circulating cell free DNA (ccfDNA) is greater in the plasma than serum, higher ctDNA levels in serum are due to leakage from leucocytes as clotting occurs, therefore plasma separation is recommended as soon as possible following blood collection (160).

DNA was extracted from the full volume of plasma taken from each patient at each time point (3-7 ml), using the Cell3™Xtract kit (Nonacus, C3048LK). Plasma was thawed and mixed with the 5x digestion buffer (250 µl) and Proteinase-K (100 µl) per ml of plasma. Samples were incubated for 30 minutes at 55 °C. Two volumes of DNA binding buffer were added to the digested sample, placed into a silica spin column assembly and centrifuged at 1,000 xg for 2 minutes. DNA Equilibrium buffer (400 µl) was added to the column and centrifuged at >10,000 xg for 1 minute. The DNA was purified using two wash buffer and centrifuge steps. DNA was eluted from the spin column in 50 µl of EB warmed to 70 °C.

2.2.4 Fluorometric Quantification of RNA and DNA

The quantification of nucleic acid concentration allows standardisation or normalisation in experiments. Fluorometric quantification uses fluorescent dyes, specific to DNA, RNA or protein. The dyes have low fluorescence until they bind to their target. Once bound they become intensely fluorescent. At a given volume of dye, the fluorescence signal, read by a Fluorometer is directly proportional to the concentration of target in solution. The concentration is calculated by comparison to known standards of DNA concentration.

2.2.4.1 Qubit™ quantification

DNA and RNA concentrations were quantified using the Qubit™ 3.0 Fluorometer (Invitrogen, Q33216). The Qubit™ measures the fluorescence intensity of fluorescent dye bound to nucleic acids. A Qubit™ double-stranded DNA (dsDNA), HS (high sensitivity) assay kit and Qubit™ RNA HS assay kit were used with the Qubit™ 3.0 Fluorometer. Standards were prepared with 10 µl of each standard and 190 µl of Qubit™ working solution. Samples were prepared by adding 1 µl of sample to 199 µl of Qubit™ working solution and read using the Fluorometer. The nucleic acid concentration was recorded in ng/µl.

2.2.5 DNA & RNA electrophoresis

Electrophoresis separates charged macromolecules such as DNA, RNA and proteins based on size according to their response to an electric current. An electrical current is placed across a gel and the molecules move towards the opposite charge with the smallest fragments moving more rapidly than the largest. To visualise the fragments the gel is pre-stained with a dye, which binds the molecule and fluoresces when exposed to ultraviolet (UV) light.

Electrophoresis was performed using either the Agilent 2200 TapeStation Nucleic Acid System (Agilent, G2965AA), Agilent 2200 TapeStation Software A.01.05 or the Agilent 2100 Bio-analyser (Agilent, G2939BA).

The TapeStation uses a ready-made three-layer polymer gel (ScreenTape) containing 16-sample channels, buffer chambers and an electrode layer. The gel is loaded by a

mini-robot directly from a 96-well plate or strip tubes. Up to 16 samples are run simultaneously and imaged. Dye within the chip allows the migrating fragments to be detected. A computer records fragment sizes and concentrations.

The Bio-analyser uses a ready-made chip with 16 sample channels etched into the chip. Sample and ladder are mixed with a dye-containing buffer and loaded onto a 1% (v/v) agarose precast gel. Separation of samples through these channels when an electric voltage is applied occurs according to fragment mass. Dye within the chip allows the migrating fragments to be detected a connected computer records fragment sizes and concentrations.

2.2.5.1 TapeStation

Genomic DNA was analysed using the Genomic DNA ScreenTape (Agilent, 5067–5365) and the Genomic DNA Reagents (Agilent, 5067–5366) with samples drawn from sample tube strips. RNA was analysed using the High Sensitivity RNA ScreenTape (Agilent, 5067-5579) and the RNA High sensitivity reagents (Agilent, 5067-5580, 5067-5581) with samples drawn from sample tube strips. dsDNA was analysed using the D1000 High Sensitivity ScreenTape (Agilent, 5067-5584) and the D1000 High sensitivity reagents (Agilent, 5067-5585) (Figure 2.1).

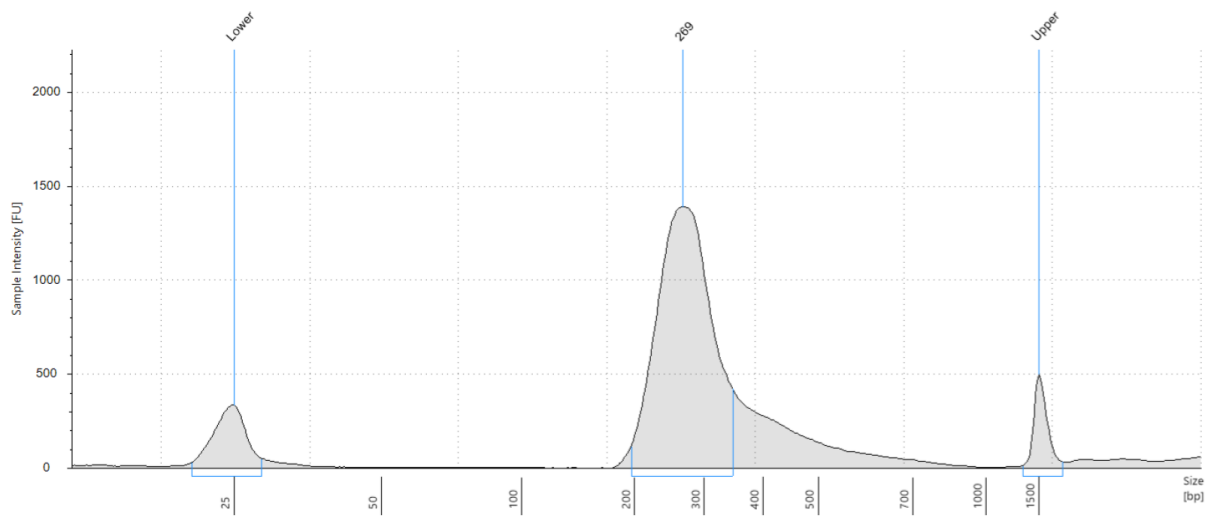


Figure 2-1 Tape station electropherogram

A representative tape station electropherogram. Markers indicate lower marker, upper marker and the average size of the nucleic acid sample in base pairs. This figure has been taken from my own results.

2.2.5.2 Bio-Analyser

During RNA library preparation amplification products were checked on the Agilent 2100 Bio-analyser (Agilent, G2939BA). Briefly, a gel dye mix was prepared by mixing the high sensitivity DNA dye concentrate and gel matrix (Agilent, 5067-4627). Gel-dye mix (9 μ l) was then loaded into the High sensitivity DNA chip to cast a gel (Agilent, 5067-4626). DNA marker (5 μ l) and 1 μ l of ladder or sample were loaded into each well, the chip incubated for 1 minute at room temperature and read using the Agilent 2100 Bio-analyser.

2.2.6 Polymerase chain reaction, an introduction

Polymerase chain reaction (PCR) is an *in-vitro* technique commonly used to amplify specific DNA regions of interest, identified by a primer. PCR uses thermal cycling, repeated cycles of heating and cooling to allow temperature-dependent reactions. High temperatures first separate the dsDNA (94-96 °C). PCR primers anneal to the region of interest at 50-65 °C. The temperature is increased to the optimal temperature of a DNA polymerase (75-80 °C) and primers are extended with the addition of deoxyribonucleoside triphosphates (dNTP). As each PCR cycle progresses DNA is exponentially doubled. Amplified sections of DNA produced via PCR are described as amplicons.

2.2.7 RNA sequencing

The genome comprises all the genetic material present in an organism. Expression of the genome relies on transcription, messenger Ribonucleic Acid (mRNA) acts as a transient copy or transcript of the genomic DNA, which is translated into encoded proteins.

Transcription occurs in the following steps: RNA polymerase and transcription factors bind to promoter regions of DNA. RNA polymerase separates ds-DNA and complementary RNA nucleotides bind to a single strand of complementary DNA (cDNA). The precursor messenger RNA (mRNA) strand is released, further processing in the nucleus transforms it from precursor mRNA to mature mRNA. Polyadenylation describes the addition of a polyadenylated poly(A) tail of adenosine repeats to mRNA.

Capping describes the addition of a five-prime cap, a specially altered nucleotide at the 5' ends of mRNA. Splicing describes the removal of introns and the joining of exons before translation (Figure 2.2).

The sum of RNA transcripts is known as the transcriptome. Measuring gene expression in different tissues or conditions creates a snapshot of which cellular processes are active and which are dormant in a cell at a particular point in time.

The transcriptome is composed of many species of RNA; in addition to protein-coding mRNAs, the genome encodes various non-coding RNAs. Some RNA species play an essential role in protein synthesis. Transfer RNAs (tRNA) deliver amino acids to the ribosome, where ribosomal RNA (rRNA) links amino acids to form proteins (161). Small nuclear RNAs are less than 200 nucleotides in length. They comprise several non-coding RNA molecules that often function to silence RNA. MicroRNA (miRNA) or small interfering RNA (siRNA) are both types of small RNA able to degrade mRNA and repress translation (161). Other species of small RNA included piwi-interacting RNA (piRNA) which also participates in RNA silencing. Small nucleolar RNAs (snoRNA) participate in post-transcriptional modification of gene expression levels known as RNA interference (161). In RNA interference miRNAs attenuate gene expression by promoting the degradation of mRNA or repressing translation (161). Small nuclear RNAs (snRNA), participate in RNA splicing, forming a large protein complex known as the spliceosome which recognises splice sites in the precursor mRNA (161). Long non-coding RNAs (lncRNA) act to regulate gene expression through many diverse functions including the modification of chromatin, direct transcriptional regulation, the regulation of precursor RNA processing, modification of miRNA, the production of siRNA and the regulation of genomic imprinting (161). Here, we primarily address the sequencing of protein-coding mRNA.

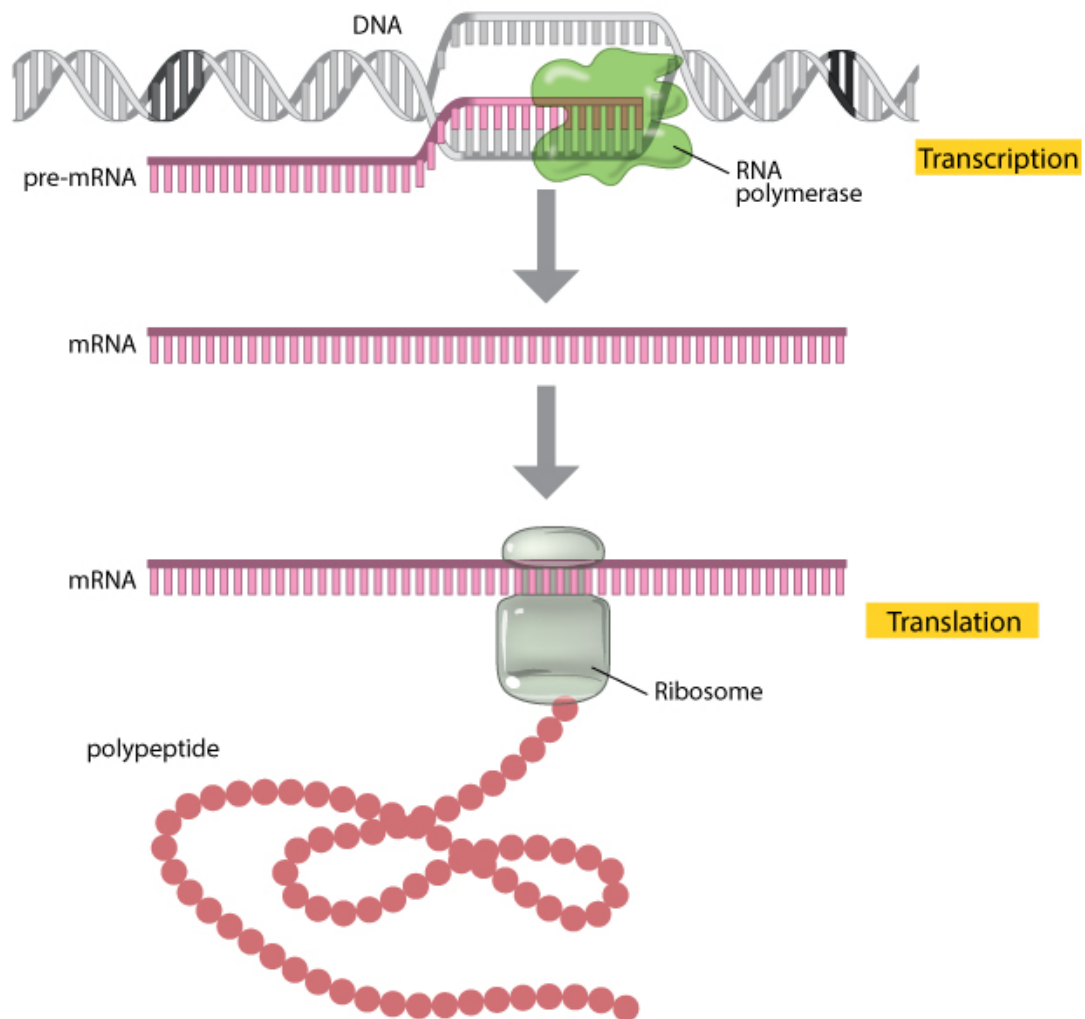


Figure 2-2 Transcription and translation

Gene expression is the manufacture of its protein product. DNA serves as a template for complementary base pairing, a polymerase enzyme catalyses the formation of an mRNA precursor which is further processed by polyadenylation, capping and splicing before translation into a protein molecule. This figure is reprinted from Clancy et al, 2008, Figure 1 (162).

2.2.7.1 Contemporary RNA sequencing techniques

2.2.7.1.1 Microarray

Microarrays quantify a set of predetermined RNA transcripts. The array consists of short nucleotide probes designed to match known transcripts, arrayed on a solid substrate (Figure 2.3) (163).

The transcriptome is prepared from an experimental condition, cDNA reverse transcripts are prepared and labelled with a red fluorescent dye (163). A control library is constructed from a non-experimental condition and labelled with a green fluorescent dye (163). Experimental and control libraries are hybridised to the microarray probes (163). A dual-channel laser excites the dye and the fluorescence indicates the degree of hybridisation (163). Gene expression is measured as the ratio of the two fluorescence wavelengths (163). Increased expression is visualised as red and decreased expression as green, constitutive expression shows as black (163). Transcript abundance is determined by the fluorescent intensity at each probe location (163).

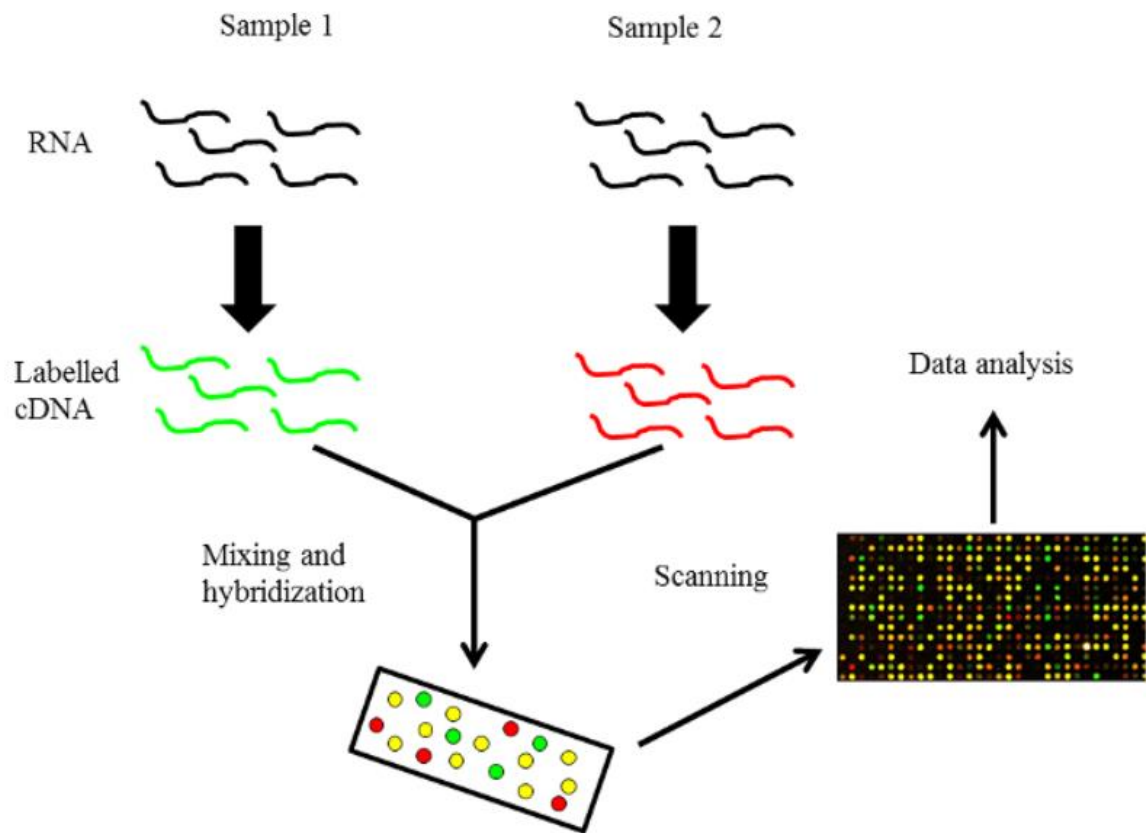


Figure 2-3 Microarray technology

The transcriptome is prepared from an experimental condition, cDNA reverse transcripts are prepared and labelled with a red fluorescent dye. A control library is constructed from a non-experimental condition and labelled with a green fluorescent dye. Experimental and control libraries are hybridised to the microarray probes. A dual-channel laser excites the dye and the fluorescence indicates the degree of hybridisation. Gene expression is measured as the ratio of the two fluorescence wavelengths. Increased expression is visualised as red and decreased expression as green. The constitutive expression is shown as black. The figure is reprinted from Afzal et al, 2015, Figure 1 (164).

Microarrays have several limitations. Arrays are limited to interrogating for the genes for which probes are designed (165). Background hybridisation describes the non-specific attachment of the probe to the membrane background (165). Cross hybridisation described signal produced by transcripts with non-perfect but significant sequence similarity with the probe (165). Because of background and cross hybridisation microarrays have difficulty detecting transcripts in low abundance (165). Conversely, microarrays are prone to hybridisation saturation for highly abundant genes meaning they cannot give reliable measures of subtle changes of genes with high expression levels (165). Microarray probe intensity depends upon the affinity to the probe under given hybridisation conditions. In microarray multiple probes interrogate different regions of the same gene, expression level changes are not always consistent across all probe sets for the same gene (165). Exon specific probes mean splice variants of the same gene can be missed (165). These factors can lead to incorrect conclusions regarding expression (165).

2.2.7.1.2 RNA-seq

RNA-seq combines high-throughput sequencing with bioinformatics to quantify gene expression. Reads of many short fragments are computationally reconstructed and aligned to a reference genome allowing deep sequencing of the entire transcriptome (166).

RNA-seq avoids hybridisation based limitations such as background noise or with probe issues such as the limited detection range of individual probes, cross-hybridisation and non-specific hybridisation (165). RNA-seq can detect novel transcripts, discover sequence variations and splice variants and accurately quantify the expression of low and high abundance genes (165). Input RNA requirements are low allowing examination of small samples down to the single-cell level. The high sensitivity of RNA-seq allows quantification of gene expression within infrequent cell populations from a sample (166).

RNA-seq involves enrichment of the transcriptome of interest (via enrichment for poly-adenylated mRNA), fragmentation, amplification followed by single or paired-end sequencing (166). Enrichment of the RNA species of interest increases sequencing

sensitivity: Oligonucleotide probes separate mRNA molecules by binding to their poly(A) tails, alternatively, ribo-depletion removes abundant ribosomal RNAs (167). mRNAs are longer than typical sequencing read lengths so transcripts are fragmented before sequencing (167). The cDNA transcript copy is then amplified, this allows input of very-low amounts of RNA with enrichment of fragments containing the expected 5' and 3' adaptor sequences (167). The addition of spike-in controls provides quality control of library preparation and sequencing in terms of fragment length and bias due to the fragment position in a transcript (168).

Sequencing of the transcript in one direction (single-end) is rapid, cheap and sufficient for quantification of gene expression levels. Sequencing in both directions (paired-end) improves alignment and transcript isoform discovery (166, 167).

RNA-seq sensitivity depends on the number of reads per sample. Large numbers of reads are needed to ensure sufficient coverage of the transcriptome and detection of low abundance transcripts (167). The current benchmark is 70-fold coverage for standard RNA-seq and 500-fold coverage to detect rare transcripts or isoforms (167).

Analysis of gene expression relies on counting sequencing read alignments to the reference genome while compensating for transcript size variation. Longer genes therefore potentially have higher counts at the same expression level (168). A solution to this is limiting the sequencing read to the 3' region of the transcript, counts are then based on the same region of each gene. Other benefits of 3' sequencing include a reduction in the depth of sequencing required and allowing for a higher degree of sample degradation such as that found in FFPE samples.

In 3' sequencing normalised total RNA is added to an oligodT primer, which binds to the poly(A) tail of mRNA. Following cDNA first-strand synthesis the RNA is removed and second-strand synthesis initiated by binding of a DNA polymerase. The library is cleaned using a magnetic bead-based purification step, indexed by the addition of indices and amplified using PCR before single-end sequencing (Figure 2.4).

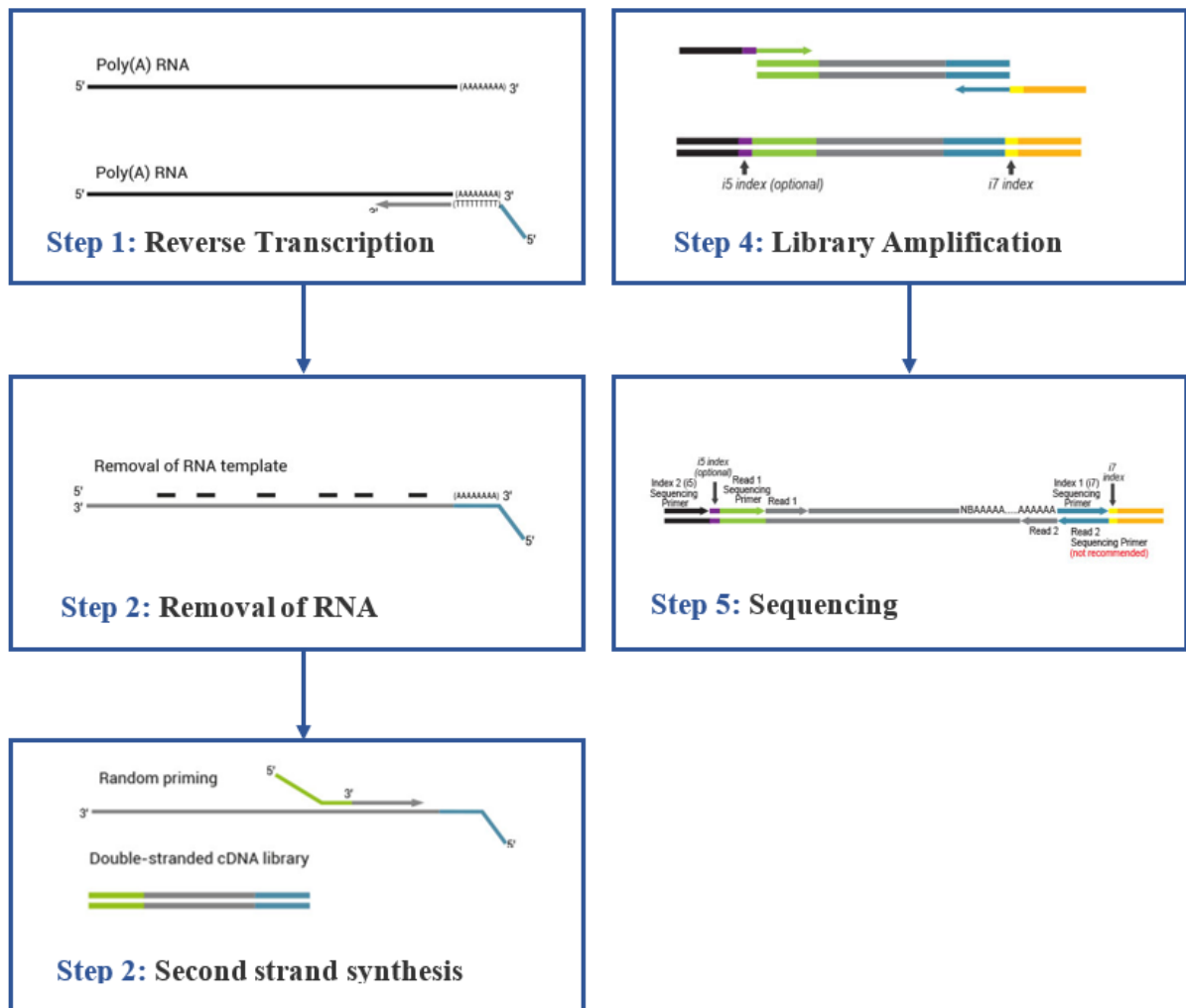


Figure 2-4 3'Prime RNA sequencing

An oligodT primer binds to the polyA tail of mRNA. Following cDNA first-strand synthesis the RNA template is removed, DNA polymerase binding initiates second strand synthesis. Magnetic bead-based purification is followed by indexing and PCR amplification before sequencing. This figure is reprinted from the Lexogen Quantseq workflow, 2019 (169).

2.2.7.2 RNA library preparation, 3' mRNA-seq method

Library preparation was performed using the Quant Seq 3' mRNA-Seq Library Prep kit (Lexogen, SKU:015.96). RNA samples were defrosted on ice, normalised to 2 ng in 5 µl of NFW in a 96 well plate. Following each thermocycler step samples were pulse centrifuged. Following the addition of each reagent, samples were pipette mixed and pulse centrifuged.

2.2.7.2.1 First-strand synthesis:

An oligo-dT primer selectively binds the poly(A) tail of mRNA, the cleaved RNA fragments are then reverse transcribed into the first-strand cDNA. A master mix at 42 °C, (0.95 µl first-strand cDNA synthesis mix 2 / 0.5 µl enzyme mix 1) was mixed with the normalised samples. The plate was incubated on the thermal cycler for 1 hour at 42 °C (Bio-Rad, 184-1100).

2.2.7.2.2 RNA removal:

RNA removal solution (5 µl) was added and incubated at 95 °C for one hour, then cooled to 25 °C on the thermal cycler to degrade the RNA template, (Bio-Rad, 184-1100).

2.2.7.2.3 Second strand synthesis:

This process synthesises a replacement strand to generate ds-cDNA. Second strand synthesis one (10 µl) was added, incubated at 98 °C on the thermal cycler, (Bio-Rad, 184-1100) then cooled slowly to 25 °C for 30 minutes. A second master mix, (4 µl second strand synthesis two / 1 µl enzyme 2) was added and incubated for a further 15 minutes.

2.2.7.2.4 Purification:

Purification beads, (16 µl) were mixed with each sample, incubated for 5 minutes, placed onto a magnetic stand, (Thermo Fisher, AM10027) until the liquid cleared. The supernatant discarded and the pellet resuspended in EB (40 µL).

Purification solution (48 µL) was added, incubated at room temperature for 5 minutes, placed onto a magnetic stand (Thermo Fisher, AM10027) until the liquid cleared and the supernatant discarded. Two 120 µL ethanol washes were performed and the pellet air-dried for up to 10 minutes. The pellet was resuspended in 20 µl of EB, incubated for 2 minutes, placed on a magnetic stand (Thermo Fisher, AM10027) and 17 µl of supernatant transferred to a new midi plate (Abgene, AB0859).

2.2.7.2.5 Library amplification and indexing:

The dsDNA library was amplified to add the complete adaptor sequences required for cluster generation and provide sufficient material for QC and sequencing. Seven microliters of a third master-mix (7 µl PCR mix / 1 µl Enzyme mix 3) was added with 5 µl of the respective index in a 96 well plate. Twenty-six cycles of PCR were performed on the thermal cycler, (Bio-Rad, 184-1100) (initial denaturation 98 °C for 30 seconds, 26 cycles of PCR at 98 °C for 10x seconds, 65 °C for 20x seconds, 72 °C for 30x seconds, 72 °C for 1 minute and hold at 10 °C. Amplification products were checked on the bio-analyser.

2.2.7.2.6 Purification:

A purification step was repeated as before with 27 µl of SPB, 30 µl of EB, 30 µl of purification solution, 2x ethanol washes and resuspension in 20 µl of EB resulting in a final volume of 17 µl of the sample in a new 8-strip PCR tube (Starlab, I1402-3700).

2.2.7.2.7 Quantification:

The DNA concentration was quantified using the Qubit™ 3.0 Fluorometer using methods described previously. DNA Library size was checked by electrophoresis using the 2200 TapeStation Nucleic Acid System (Agilent, G2965AA), 2200 TapeStation Software A.01.05 and D1000 High Sensitivity ScreenTape (Agilent, 5067-5584) and the D1000 High sensitivity reagents (Agilent, 5067-5585). Libraries were frozen at -20 °C before sequencing on the Nextseq500 with a 75-cycle high Output flow cell.

2.2.7.3 RNAseq bioinformatics

Unaligned RNA-seq data was uploaded into the Partek® Flow® genomics suite software package (Partek, St Louis, MI, USA). Pre-alignment quality assurance/quality control (QA/QC) was performed to check the total number of reads per sample, read length and quality. Trimming was performed to remove any low-quality reads and the data aligned to the reference genome (NCBI build 37, hg19). Following alignment, a further QA/QC step removed any unaligned reads.

Gene Specific Analysis (GSA) Modelling was performed using the Partek® flow®. GSA identifies the optimum statistical model for a specific transcript and then uses that model to test for differential expression. GSA produces *P*-values, fold change estimates and other measures of importance for the factor of interest. A false discovery rate (FDR) filter of <0.1 was applied to the data meaning no more than 10% of differential gene expression is type 1 error.

CMS and CRIS classifications were performed using 'CMScaller' (v0.99.1) in the R/Bioconductor statistics package, (by Dr Andrew Beggs). With CMScaller, prediction confidence is estimated from gene resampling ($n = 1000$) and samples with false discovery rate adjusted, *P*-value > 0.05 were "not assigned" (NA) (170). Fisher's exact test was used to compare contingency between primary and CPM or good and poor prognosis CPM test in IBM SPSS Statistics for Windows, Version 24.0 (159). A *P*-value of < 0.05 was considered significant.

2.2.8 DNA sequencing

Next-generation sequencing is based on the chain/dye-terminator method described by Sanger et al in the 1970s (171). Sequencing is based on the selective incorporation of fluorescently labelled nucleotides by DNA polymerase in a PCR reaction (Figure 2.5) (172).

Samples are prepared, fragmented and bound to adaptors, complementary to oligonucleotides bound to the flow cell, as well as indices which identify the sample (172). When applied to the flow cell, each fragment binds to an oligonucleotide sequence complementary to its adaptor (172). PCR creates a complementary strand

of DNA. The dsDNA is denatured and the template washed away (172). Fragments are amplified via bridge amplification, the strand bends and the opposite adaptor binds to a second complementary adaptor, further PCR reaction creates an identical strand of DNA (172). This reaction is performed in parallel for each fragment of DNA resulting in clonal amplification of fragments and clusters of identical DNA (172).

During each cycle, a mixture of all four fluorescently tagged nucleotides and 3'-blocked dNTPs are added and compete for addition to the DNA primer (172). After the addition of each nucleotide, the unbound dNTPs are removed, the clusters are excited by a light source, a fluorescent signal is emitted specific to the base type and the surface imaged to identify which dNTP was incorporated (172). The fluorophore and blocking group are removed and a new cycle can begin (172). Each cluster and DNA fragment is sequenced in parallel. The number of cycles dictates the length of the transcript read (172).

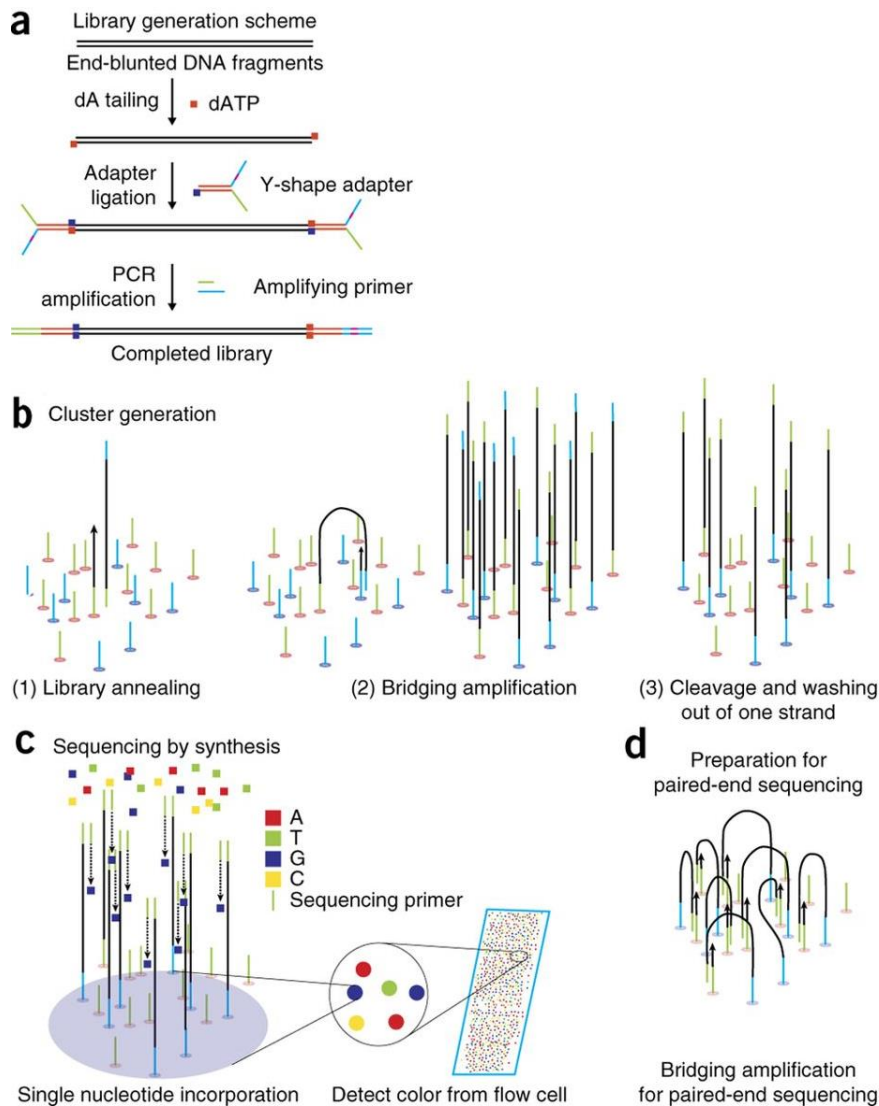


Figure 2-5 Next generation sequencing

(a) A library is composed of fragments of double-stranded DNA flanked by adapter sequences and indices, correctly ligated libraries are amplified. (b) Cluster generation on the surface of the flow cell. (1) Fragments bind complementary oligonucleotides. (2) Bridging amplification generates clusters. (3) One strand from the double-strand DNA is cleaved and washed away (c) In each sequencing cycle, 3'-blocked deoxynucleotides (dNTPs) and fluorescently labelled A, T, G and C bases are applied. After the addition of each nucleotide, the sequencing reaction is stopped, and the image is taken. (d) In paired-end sequencing when the first-strand sequencing reaction is finished, the synthesized strand is removed and the process is repeated for the opposite strand. This figure is reprinted from Shin et al 2014, figure 5 (173).

2.2.8.1 DNA sequencing method

The 75, 150 or 300-cycle mid or high Output flow cell and reagents (Illumina, FC-404-2005, FC-404-2002, 20024905) were thawed and allowed to equilibrate to room temperature.

PhiX (Illumina, FC-110-3001) is a concentrated bacteriophage genomic library with an average size of 500bp, this acts as a biological control for the sequencing run. DNA libraries were normalised to a 0.5 nm concentration and pooled. The DNA library and PhiX control were denatured by combining with equal volumes of a 0.2M sodium hydroxide (NaOH) solution (SigmaAldrich, 221465-500G) vortexed briefly, centrifuged at 280 xg for 1 minute and incubated at room temperature for 5 minutes. An equal volume of Tris-HCL was added, (Sigma-Aldrich, 77-86-1) vortexed briefly and centrifuged at 280 xg for 1 minute. The denatured library and Illumina PhiX sequencing control were then diluted to a 20 pm concentration with the addition of HT1 hybridisation buffer to a final loading concentration of 1.6 pm in 1.3 ml before addition to the Illumina reagent cartridge.

The reagent cartridge buffer cartridge and flow cell were loaded into the NextSeq500. Indices were assigned to sample IDs in base space and run parameter assigned including single or paired-end read and the cycle number and the sequencing run commenced.

2.2.9 Epigenetics

Epigenetics refers to several bio-molecular mechanisms that modify gene expression and the resulting phenotype, without alteration of the DNA sequence or genotype (40). In health, DNA methylation allows cells to acquire and maintain a specialised state and function, silence the expression of non-host genomic DNA and facilitates the cellular response to environmental stimuli (40).

An example of epigenetics in health is the random inactivation of one X-chromosome in female mammals by the non-coding RNA gene, X-inactive specific transcript (Xist)

(40). As males only possess one copy of the X-chromosome, x-linked disorders have a classical phenotype, heterozygous females, however, have increased phenotypic variation depending on which chromosome is silenced (40). Three mechanisms comprise the main drivers of epigenetic modification;

2.2.9.1 Histone modification

Histones are alkaline proteins which DNA bases are wound around allowing them to be packaged and organised into Chromatin. A variety of post-translational histone modifications are possible, resulting in a multiplicity of possible combinations and modifications (40). Histone acetylation marks transcriptionally active regions whereas hypo-acetylation marks transcriptionally inactive regions (40). In contrast, histone methylation can be a marker for both active and inactive regions (40). Other post-transcriptional modifications include phosphorylation and ubiquitination (40).

2.2.9.2 Non-coding RNA associated gene silencing

Non-coding RNA such as anti-sense transcripts, non-coding RNA or the RNA interference pathway (RNAi) can lead to enzymatic degradation of mRNA, the formation of heterochromatin (tightly packaged, inactive DNA) and mitotically heritable transcriptional silencing.

2.2.9.3 DNA methylation

Methyl groups (CH₃) can be added to cytosine or adenine DNA bases by DNA methyltransferases enzymes (DNMTs) to give 5-methylcytosine (5Mc) (Figure 2.6) (40). There are three types of DNMT. DNMT1 is the most abundant, it binds to hemimethylated DNA at CpG sites, after replication, it methylates cytosine in the new strand maintaining the CpG methylation pattern through mitosis (174). DNMT3A and DNMT3B are responsible for de-novo methylation and show an affinity for hemimethylated and none methylated DNA, they are required for genome-wide methylation of DNA occurring after embryo implantation.

CpG sites are DNA regions where a cytosine nucleotide is followed by a guanine nucleotide in a 5' to 3' direction. Cytosines in CpG dinucleotide sequences can be methylated to form 5-methylcytosine (40). In mammals, 70-80% of CpG sites are

methyated which results in epigenetic silencing (40). CpG islands are short regions of between 0.5-4 kbp with a GC content greater than 55%. They are located in promoter regions of genes, usually un-methylated and conserved during evolution (40). CpG island methylation causes stable heritable transcriptional silencing. Aberrant de novo methylation of CpG islands is a hallmark of human cancer found early in carcinogenesis. DNA methylation allows cells to acquire and maintain a specialised state, suppresses the expression of viral and non-host DNA and facilitates the cell response to environmental stimuli and ageing. Aberrant DNA methylation (hyper or hypo) is implicated in many disease processes including ageing, neurological disorders, metabolic disorders and cancer.

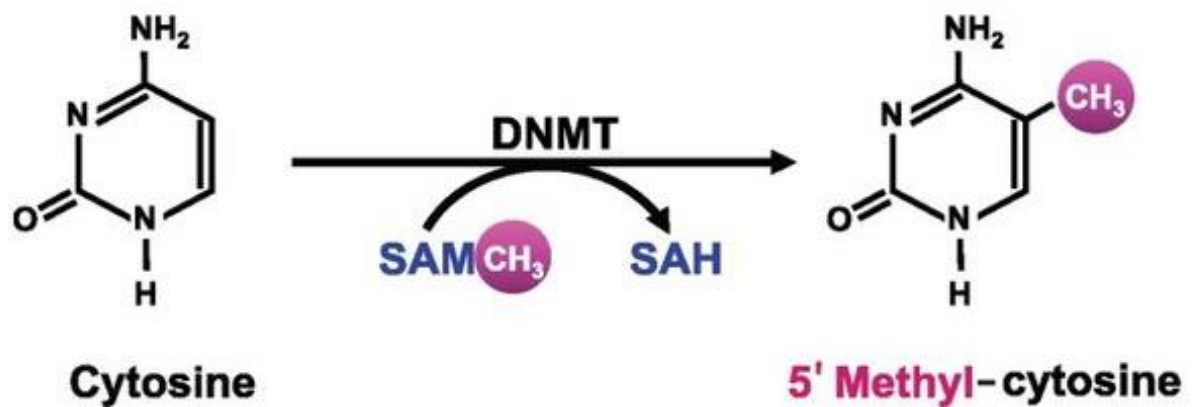


Figure 2-6 DNA methylation

The addition of methyl groups (CH_3) to cytosine DNA bases by DNA methyltransferase (DNMT) to give 5-methylcytosine (5Mc). S-adenosylmethionine (SAM) is a co-substrate in methyl transfers, S-adenosylhomocysteine (SAH) is formed by the demethylation SAM. This figure is reprinted from Zakhari et al, 2013, Figure 1 (175).

2.2.9.4 Methylation analysis

2.2.9.4.1 Methylation dependent DNA treatment

5MC is not distinguishable from unmethylated cytosine by hybridisation and as DNMTs are not present during PCR this information is erased during amplification. Therefore, all sequence-specific DNA methylation analysis relies on a type of methylation-dependent DNA treatment before amplification or hybridisation. Several treatments exist and are discussed below.

2.2.9.4.2 Endonuclease digestion

Isoschizomers of bacterial restriction endonucleases with different sensitivities to 5MC methylation-sensitive restriction enzymes (MSREs) are used to determine the methylation status of cytosine at specific sites (176). One enzyme of isoschizomer pair is insensitive to 5MC, the other will not cut DNA if the cleavage site contains 5MC (176).

Endonucleases provide information only about CpGs within the cleavage sites of the specific enzymes. In addition, this technique is prone to false-positive results due to incomplete endonuclease digestion, they should therefore not be used for the detection of low-abundance methylated molecules (176).

2.2.9.4.3 Affinity enrichment

Affinity enrichment isolates methylated DNA from the rest by antibody immunoprecipitation with methyl CpG binding proteins (176). Antibodies specific for 5MC are combined with denatured genomic DNA allowing the enrichment of methylated regions (176).

Affinity based methods do not reveal information of individual CpG dinucleotides and require significant bioinformatics adjustment for varying CpG density at different regions of the genome (176).

2.2.9.4.4 Bisulphite conversion

The treatment of genomic DNA with sodium bisulphite deaminates unmethylated cytosine residues to uracil much more rapidly than methylated cytosines (176). Preferential deamination of cytosine by bisulfite occurs by sequential sulfonation, hydrolytic deamination and alkaline desulfonation reactions (176). The DNA double helix is converted into two single strands (176). DNA is fragmented by incubation at high temperature, sodium bisulphite deaminates cytosine and incubation at high pH removes the sulphite generating uracil (176). 5MC is not susceptible to bisulfite conversion and remains intact (Figure 2.7) (176). The bisulphate induced sequence differences in DNA template, (conversion of cytosine to thymine, by uracil) form the basis for discriminating between methylated and unmethylated DNA (176). Methylation of any CpG in the genome can be determined – for total cell population and individual DNA molecules (176).

Bisulphite conversion is prone to several issues. Bisulphite conversion relies on the conversion of every unmethylated cytosine to uracil. Incomplete conversion may if dsDNA is not completely denatured or if it is not maintained in single-stranded conformation during the process due to variations in parameters such as temperature or salt concentration (176). In PCR amplification, the primers may not amplify the methylated and unmethylated sequences with equal efficacy (176).

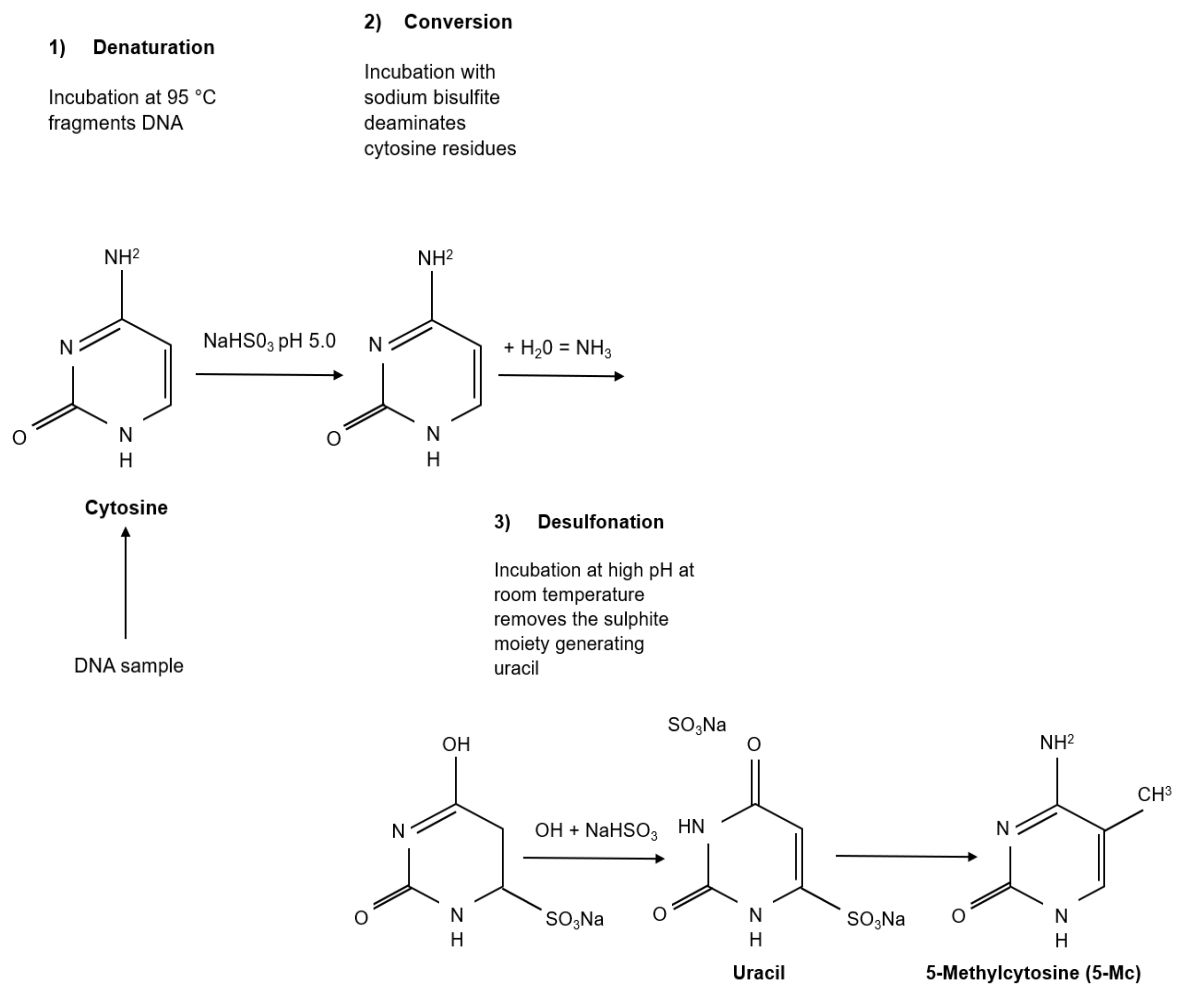


Figure 2-7 Bisulfite conversion of genomic DNA

1) DNA is denatured by incubation at high temperature 2) Incubation with sodium bisulphite deaminates cytosine 3) Incubation at high pH removes the sulphite generating uracil. 5-methylcytosine is not susceptible to bisulphite conversion and remain intact. (This figure was drawn in PowerPoint, Microsoft office)

2.2.9.5 Methylation analysis techniques

A variety of methods exist to analyse methylation. These can be divided into those that determine the methylation status of a single CpG locus of a short sequence and those that determine methylation at a global or genome-wide level.

Single CpG loci or short sequence methylation analysis techniques

2.2.9.5.1 Methylation-specific PCR (MS-PCR) / quantitative MS-PCR

Bisulphite treatment of DNA converts unmethylated cytosine to uracil whilst 5MC is preserved (177). During PCR uracil is amplified as thymidine which 5Mc is unchanged (177). Primers for both methylated and unmethylated DNA are designed to determine the methylation status of a locus of interest and gel electrophoresis is used to discriminate between the two (177). Methylated specific PCR is qualitative and can be used as an initial screen for methylation of specific genes but requires subsequent validation using a quantitative method (177).

Quantitative MS-PCR is possible with technology such as MethyLight. Bisulphite converted DNA is amplified using PCR with 3x oligonucleotides, a dual labelled fluorescent probe flanked by a forward and reverse primer (177). After annealing to the PCR product, the probe is cleaved by a Taq DNA polymerase triggering a fluorescent signal (177). Fluorescence is measured in each cycle and is proportional to the PCR product generated (177).

2.2.9.5.2 Methylation-sensitive single-nucleotide primer extension (Ms-SnuPE)

Methylation-sensitive single-nucleotide primer extension (Ms-SnuPE) uses sodium bisulphite treated DNA amplified with PCR to give a template of the target sequence (177). An oligonucleotide primer anneals to a PCR generated template immediately 5' to the nucleotide of interest (177). The primer is then extended in the presence of DNA polymerase and a labelled deoxyribonuclease (177). The relative amount of incorporated label is proportional to the relative amount of that base at that site in the DNA template (177). Ms-SnuPE is quantitative and allows the analysis of small amounts of DNA (177).

2.2.9.5.3 Bisulphite pyrosequencing

Bisulphite pyrosequencing quantifies methylation in bisulphite treated DNA via the detection and quantification of pyrophosphate released from incorporated nucleotides by DNA polymerase during amplification. Pyrophosphate undergoes an enzymatic process and emits a light signal which is recorded by a camera. This method does not rely on primer sequencing for specificity and can quantify multiple CpG sites in a single run. It is however limited to a sequence length of 100-150bp.

2.2.9.5.4 Combined bisulphite restriction analysis (COBRA)

Bisulphite treated DNA is amplified using PCR primers which do not discriminate between methylated and unmethylated sites (177). PCR products are a mixed population which have retained or lost their CpG restriction enzyme sites (177). Products are treated with a restriction enzyme which cleaves sites which were originally methylated (177). Digested fragments are separated by electrophoresis, large bands correspond to undigested fragments and small bands to digested methylated fragments (177). The bands can be imaged quantifying the methylation percentage of the sample. COBRA is quantitative and allows the analysis of small amounts of DNA (177).

2.2.9.5.5 High resolution melting analysis

Bisulphite treated DNA is amplified using PCR. Precise warming of the PCR product results in melting of the DNA strands (177). This is monitored in real-time using a fluorescent dye which fluoresces when bound to double-stranded DNA and at a low level when the DNA denatures (177). This is captured by a camera and plotted in a melting curve showing the fluorescence vs. the temperature (177). Methylation-specific denaturing high-performance liquid chromatography (MS-DHPLC) is an in-tube high resolution melting analysis assay (177). Methylation-specific denaturing gradient gel electrophoresis (MS-DGGE) separates differentially methylated melted molecules in denaturing gel electrophoresis (177).

2.2.9.5.6 Methylation-specific microarray

Bisulphite treated DNA is amplified using non-discriminatory PCR with fluorescently labelled primers and hybridised to a glass slide with oligonucleotides which discriminate between methylated and unmethylated cytosines (177). This method allows the analysis of multiple genes on one array, however closely spaced CpGs may not be accurately analysed if the gene is heterogeneously methylated (177).

Genome-wide methylation analysis techniques

High throughput methods make it possible to simultaneously analyse the status of thousands of CpG islands across the genome. Some of the technologies in use currently are discussed below.

2.2.9.5.7 Affinity enrichment Chromatin immunoprecipitation

This technique uses an antibody or protein with an affinity for 5MC to isolate methylated DNA (176). The enriched and input DNA are labelled with fluorescent dyes and hybridised to a microarray with probes specific for CpG islands or promotor regions (176). Bioinformatics analysis is performed to reveal methylation using the relative signal intensities of the enriched and input DNA (176). Affinity based methods allow genome-wide assessment of DNA methylation but are not able to yield information on individual CpG dinucleotides (176).

2.2.9.5.8 Beadchip Microarray

This technique from Illumina analyses methylation sites quantitatively across the genome at single-nucleotide resolution allowing genome-wide profiling of human DNA methylation (176). Following bisulfite conversion DNA is fragmented and hybridised to methylation-specific oligomers linked to individual bead types (176). Each bead corresponds to a specific DNA CpG site and methylation state (176). The current assay interrogates 850,000 methylation sites (176). The Illumina BeadChip microarray has relatively low DNA input requirements, is adapted to allow the use of degraded FFPE DNA and is relatively low cost (Figure 2.8).

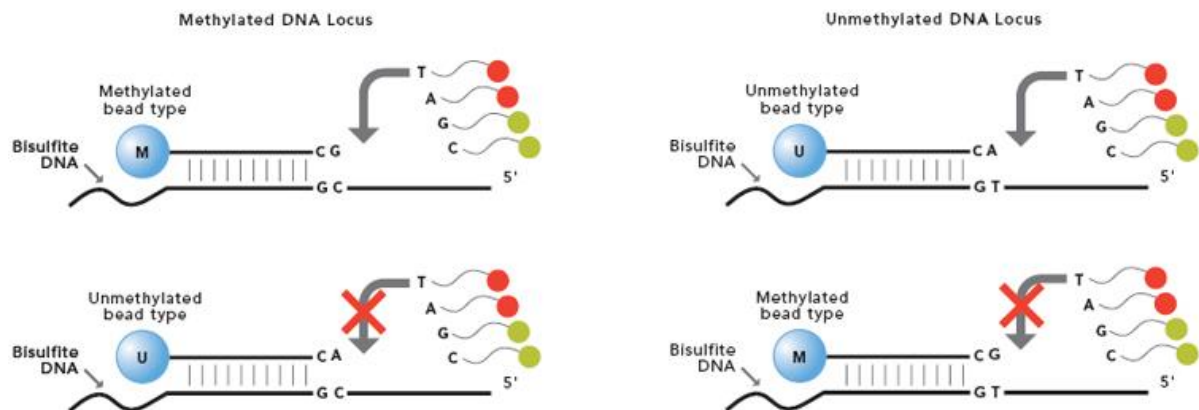


Figure 2-8 The Infinium methylation array

The U bead type matches the unmethylated CpG site; the M bead type matches the methylated site. In the top figure, the unmethylated CpG target site matches with the U probe, enabling single-base extension and detection. It has a single-base mismatch to the M probe, which inhibits extension. If the CpG locus of interest is methylated (bottom figure), the reverse occurs. This figure is reprinted from the Infinium methylation assay product literature, 2019 (178).

2.2.9.6 Next-generation sequencing

Sequence-based analysis allows allele-specific methylation analysis, does not require a designed microarray and can cover a larger proportion of the genome with less DNA input. Bisulphite treated DNA is fragmented, ligated to adaptors and sequenced using next-generation sequencing allowing visualisation of differentially methylated sites at single-nucleotide resolution. Next-generation sequencing is relatively expensive and complex in terms of bioinformatics analysis.

Methylation analysis methods used in this thesis were selected to allow genome-wide differential methylation analysis, at an affordable cost per sample, using techniques which have been successfully performed within our laboratory. As the Illumina Infinium EPIC array is essentially an SNP array this method of methylation analysis allow the analysis of CNA within the genomic regions targeted by the methylation probes with similar sensitivity to SNP platforms (71%) (179).

2.2.9.7 Methylation analysis Infinium MethylationEPIC array, methods

Following the addition of each reagent, samples were mixed by pipette or vortex and pulse centrifuged. Following each thermal cycler step samples were pulse centrifuged.

2.2.9.7.1 Bisulphite treatment of FFPE DNA

Bisulphite conversion was performed using the Zymo EZ-DNA methylation kit (D5004). CT conversion reagent is prepared by mixing with 7.5 ml of NFW and 1.85 ml of dilution buffer. In a midi plate (Abgene, AB0859) 350 ng of normalised DNA in 42.5 µl of NFW (Omega BIO-TEK PD092) was mixed with 1 µl of dilution buffer. The plate was then incubated at 42 °C for 30 minutes on a thermal cycler (Bio-Rad, 184-1100). CT conversion reagent (97.5 µl) was mixed with each sample and the plate incubated, in the dark, on the thermal cycler (Bio-Rad, 184-1100) for 16 cycles of (95 °C for 30 seconds, then 50 °C for 60 minutes).

Binding buffer (400 µl) was combined with samples in a Zymo-Spin™ I-96 plate. A wash was performed with wash buffer, centrifuged at > 3,000 xg for 5 minutes and the flow-through discarded. Desulfonation buffer was added (200 µl) and incubated at

room temperature for 20 minutes, centrifuged at >3,000xg for 5 minutes and the flow-through discarded. Two further washes were performed with 500 µl of wash buffer and centrifuged at >3,000 xg for 5 minutes then 10 minutes. DNA was eluted in a final volume of 8 µl of EB.

2.2.9.7.2 FFPE DNA restoration

Bisulphite converted, degraded FFPE was restored using the Infinium HD FFPE restore kit (Illumina, WQ-901-2004). Bisulphite treated DNA was added to a midi plate and denatured with 0.1N (4 µl) sodium hydroxide. Primer restore reagent (34 µl) and amplification mix restore reagent (38 µl) were added, sealed, mixed by inversion and incubated at 37 °C for one hour.

DNA was cleaned using the ZR-96 DNA clean and concentrator-5 kit (Zymo, D4023). Samples were mixed with binding buffer (560 µl), added to a Zymo-spin column, centrifuged at 2250 xg for 2 minutes and the flow-through discarded. A wash step was performed using Zymo-wash buffer, (600 µl) centrifuged at 2250 xg for 2 minutes and the flow through discarded. DNA was eluted in a final volume of 10 µl of EB.

DNA was incubated at 95 °C for 2 minutes and placed on ice. Convert master mix reagent (10 µl) was added, then incubated at 37 °C for 1 hour. DNA was cleaned again using the ZR-96 DNA clean and concentrator-5 kit, (Zymo, D4023) as described previously. DNA was eluted in 10 µl of deionised water. Eluted DNA was used as input for the Infinium MethylationEPIC BeadChip Kit (Illumina, WG-317-1002).

2.2.9.7.3 Amplify bisulphite converted DNA

DNA (8 µl) was added to a midi plate, (Abgene, AB0859) with multi-sample amplification mix 1 (20 µl) and 4 µl of 0.1N sodium hydroxide to denature and neutralise the samples. The plate was sealed with a cap mat (Thermo Fisher, 276000) and incubated at room temperature for 10 minutes. Random primer mix (68 µl) and multi-sample amplification master mix (75 µl) were mixed, the plate sealed and incubated in the Illumina Hybridisation Oven for 20-24 hours at 37 °C, (Illumina, SE-901-1002).

2.2.9.7.4 Fragment DNA

Fragmentation solution (50 µl) was mixed with each sample containing well of the midi plate (Abgene, AB0859) sealed with a cap mat and incubated for 1 hour at 37 °C.

2.2.9.7.5 Precipitate DNA

Precipitation solution (100 µl) was mixed with each sample, sealed with a cap mat and incubated at 37 °C for 5 minutes. Isopropanol 100% (v/v) (300 µl) (Thermo Fisher, 10674732) was added to each sample, sealed and inverted 10x to mix. The plate as then incubated at 4 °C for 30 minutes and centrifuged at 3000 xg, 4 °C for 20 minutes. The plate was inverted on to an absorbent pad, smacked down and tapped firmly for 1 minute until all wells were dry. The plate was left for 1 hour to air dry the DNA pellet.

2.2.9.7.6 Re-suspend DNA

RSB (46 µl) was added to each sample. A foil seal was applied and firmly sealed using a heat sealer (Thermo Fisher, 11590314) The plate was then incubated in the Illumina Hybridisation oven for 1 hour at 48 °C.

2.2.9.7.8 Hybridize to bead chip

Samples were placed on a heat block at 95 °C for 20 minutes to denature and allowed to cool to room temperature. The hybridisation chambers were assembled (Illumina, BD-60-402, WG-15-301). Humidifying buffer (400 µl) was added to the reservoirs. Each BeadChip was placed into its Hybridisation chamber insert. The DNA samples were loaded onto the BeadChip (26 µl) and the BeadChips placed into the hybridisation chamber. The chamber was transferred into the Hybridisation oven at 48 °C for 16 hours.

2.2.9.7.9 Wash bead chip

The hybridisation chamber was removed from the oven and cooled to room temperature for 30 minutes. Wash dishes (Illumina, BD-60-460) were filled with a reagent used to prepare BeadChips for hybridization (PB1) (200 ml). The BeadChips were removed from their inserts, the covers removed then placed into wash racks (Illumina, BD-60-450). BeadChips were washed by repeated submersion in the PB1

for 1 minute. This was repeated once in fresh PB1. Flow-through chambers were assembled (Illumina, WG-10-201).

2.2.9.7.10 Extend and stain bead chip

The chamber rack (TECAN, 3002329) was warmed to 44 °C and each flow-through chamber assembly placed into the chamber rack. Into the reservoir of each flow-through chamber, the following reagents were dispensed:

- Wash solution (150 µl) incubated for 30 seconds and repeated 5x.
- XStain BeadChip solution 1 (450 µl) incubated for 10 minutes.
- XStain BeadChip solution 2 (450 µl) incubated for 10 minutes.
- Two-Color Extension Master Mix (200 µl) incubated for 15 minutes.
- 95% (v/v) formamide 1 mM EDTA incubated for 1 minute and repeated once.
- Incubate for 5 minutes then ramp the chamber rack temperature to 39.2 °C.
- XStain BeadChip solution 3 (450 µl) incubate for 1 minute and repeated once.

Into each reservoir of the flow-through chamber, the following were dispensed:

- Superior Two-Color Master Mix (250 µl), incubated for 10 minutes, XStain BeadChip solution 3, incubated for 1-minute repeated once, wait for 5 minutes.
- Anti-Stain Two-Colour Master Mix (250 µl) and incubate for 10 minutes, XStain BeadChip solution 3, incubate for 1 minute and repeat once, wait for 5 minutes.
- Superior Two-Colour Master Mix (250 µl) and incubate for 10 minutes, XStain BeadChip solution 3, incubate for 1 minute and repeat once, wait for 5 minutes.
- Anti-Stain Two-Colour Master Mix (250 µl) and incubate for 10 minutes, XStain BeadChip solution 3, incubate for 1 minute and repeat once, wait for 5 minutes.
- Superior Two-Colour Master Mix (250 µl) and incubate for 10 minutes, XStain BeadChip solution 3, incubate for 1 minute and repeat once, wait for 5 minutes.

The flow-through chambers were removed from the chamber rack. A wash was repeated with 310 ml of PB1 in a wash rack for 5 minutes. A final wash dish was prepared with XStain BeadChip solution 4 (310 ml). The BeadChip containing rack was

submerged 10x in the dish and allowed to soak for 5 minutes. BeadChips were removed from the rack and dried in a vacuum desiccator for 50-55 minutes at 675 mm Hg (SICCO, V 1821-07).

2.2.9.7.11 Image bead chip

Bead chips were imaged using the iScan system (Illumina, SY-101-1001). Data was exported as raw image files and proprietary iDAT files.

2.2.9.8 Methylation bioinformatics

Raw data in the form of IDAT files were loaded into the RStudio version 3.5.0 software using the *minifi* package and Bioinformatics analysis performed using the Chip Analysis Methylation Pipeline (ChAMP) R package, version 2.10.2 (180, 181). In brief, red/green intensity levels were captured from the Illumina IDAT files, background corrected and SWAN-normalised to produce beta-values. The beta-value is an estimate of methylation level using the raw intensities between methylated and unmethylated alleles. Beta-values are between 0-1, where 0 is unmethylated and 1 is fully methylated. Raw methylation data were filtered to exclude: inconsistent probes which fail to detect a signal with high confidence, (a detection value > 0.01), failed probes with <3 beads, non-CpG probes, SNP related probes that would give genetic rather than epigenetic observations (182), multi-hit probes and probes on chromosome X and Y chromosome to exclude gender bias. To check the dataset was suitable for analysis a quality check was performed. A shift in the distribution of methylation beta-values results from the different chemistry of array probes across the Infinium assays. This requires a correction strategy of normalisation using the BMIQ strategy (183, 184). To identify unwanted types of variation not of biological origins, such as batch effects a singular value decomposition (SVD) was performed (185). Batch effects may arise due to an experiment being performed on different days. As batch effects were not identified, a normalisation method to correct for this was not required.

Differential methylation was analysed via identification of differentially methylated probes (DMPs). DMP implements the limma package to calculate the *P*-value for differential methylation between groups using a linear regression model (186). DMPs are mapped to the genome and genes containing at least one methylation probe were

extracted. The methylation fold change was obtained by calculating the difference between the mean beta-value for each probe. The threshold for statistical significance was set at a Bayes factor (BF) of 4.

Differentially methylated regions (DMRs) are extended segments of the genome which show a qualitative alteration in DNA methylation levels between two groups. DMR analysis was performed using the Bumphunter algorithm (187). Bumphunter groups all probes into small clusters then applies random permutation methods to estimate candidate DMRs. Bumphunter output includes all detected DMRs, their length, clusters and the number of CpGs annotated. The threshold for statistical significance was set at a *P*-value of <0.05. Copy number alteration calling was performed using the CHAMP CNA function, a *P*-value of < $\times 10^{-10}$ was considered to be significant (179). The R code used to complete the methylation bioinformatics is displayed in appendices, table S1.

2.2.10 Exome sequencing

A more comprehensive analysis of the genome is required to identify the somatic mutations associated with the development of CPM and conferring a poor prognosis. Whole-genome sequencing, (WGS) determines the complete DNA sequence of the cells within a sample. WGS was first performed in humans in 2007, it was initially prohibitively expensive. The cost of WGS reduced substantially in the following decade. The price per sample, however, still sits at between £1500-2000. This in addition to associated start-up costs for sequencers capable of providing the depth and breadth of coverage required. An alternative is to sequence a targeted area of the genome. The exome comprises the protein-coding portion of the genome. Although exons comprise just 1% of the genome they are thought to harbour up to 85% of mutations associated with disease (188). In WES DNA is fragmented before library preparation, this results in a reduced uniformity of coverage compared to WGS. WES, therefore, requires a higher depth to give adequate coverage, (100X vs. 30X). The probes used for target enrichment vary in length, sequence and overlap, this can lead to variability in target specificity between WES kits (the Illumina TruSeq exome kits are designed to give $\geq 80\%$ on target reads). As reads are focussed on a small subset of the genome it requires less breadth of sequencing, leading to reduced costs and a

smaller data output which is faster and easier to analyse than WGS. It is difficult to perform WGS on FFPE samples due to sample degradation and reduced DNA yields, WES, however, is feasible due to an in-house protocol developed in our laboratory.

To allow exome sequencing the DNA is fragmented and the exon sequences captured using a variety of methods. Solid-phase hybridisation use probes complementary to the exon sequences fixed to a microarray. DNA is applied to the microarray and the exons hybridise to the probes, non-targeted fragments are washed away and enriched DNA eluted for sequencing. Liquid-phase hybridisation uses liquid phase biotinylated probes bound to magnetic streptavidin beads. The exons hybridise to the probes and the non-targeted fragments are washed away on a magnetic plate, enriched DNA is eluted for sequencing (Figure 2.9).

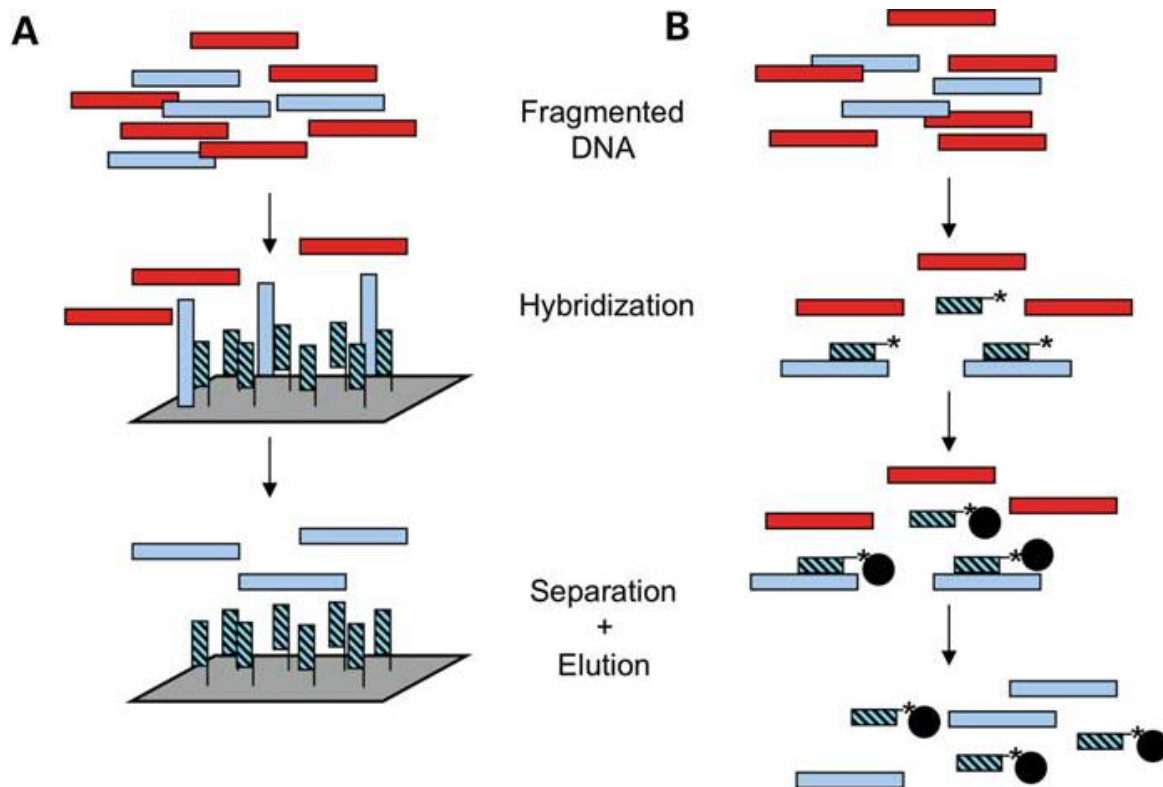


Figure 2-9 Exon capture methods

Blue bars represent the desired exon sequence, red the unwanted intron sequence. A) Solid-phase hybridisation. Exons bind to complementary probes on the microarray surface, the array is washed and DNA eluted. B) Liquid phase capture. Biotinylated probes are mixed with fragmented DNA, the exons hybridise to the probes in solution. The bead-DNA complex is washed and the desired DNA eluted. This figure is reprinted from Teer et al, 2010, figure 1 (189).

2.2.10.1 Exome sequencing, method

Following the addition of each reagent, samples were pipette mixed and pulse centrifuged. Following each thermal cycler step samples were pulse centrifuged.

2.2.10.1.1 Fragment DNA

In an 8-strip PCR tube (Starlab, I1402-3700) 300ng of normalised DNA in 49 μ L of RSB was added to 1 μ L of 50 mM EDTA. Samples were transferred to microTUBEs (Covaris, 520174) and fragmented to 150 bp using the E220 evolution focused-ultrasonicator (Covaris, 500429) with the following settings: Duty factor 10%, peak power 175, cycles/burst 200, duration 150 seconds, temperature 7 °C, water level 5. Fragment length was checked by electrophoresis using the 2200 TapeStation Nucleic Acid System (Agilent, G2965AA), Agilent 2200 TapeStation Software A.01.05 and D1000 High Sensitivity ScreenTape (Agilent, 5067-5584) and the D1000 High sensitivity reagents (Agilent, 5067-5585).

2.2.10.1.2 Clean-up fragmented DNA

In a midi plate (Abgene, AB0859) DNA samples (50 μ L) were mixed with sample purification beads (SPB) (100 μ L), incubated at room temperature for 5 minutes, then placed on a magnetic stand (Thermo Fisher, AM10027) until the liquid cleared. Two 80% (v/v) ethanol washes were performed (200 μ L) of 80% (v/v) ethanol on a magnetic stand (Thermo Fisher, AM10027). DNA was then re-suspended in RSB (62.5 μ L) incubated at room temperature for 2 minutes, placed on a magnetic stand (Thermo Fisher, AM10027) until the liquid cleared, then 60 μ L of sample transferred to a new 8-strip PCR tube (Starlab, I1402-3700).

2.2.10.1.3 Repair ends and select library size

End repair mix (40 μ L) was mixed with each sample then placed on the thermal cycler, (Bio-Rad, 184-1100) for 30 minutes at 30 °C. A clean-up step was repeated as before with 250 μ L of SPB, 2x 80% (v/v) ethanol washes and resuspension in 20 μ L of RSB resulting in a final volume of 17.5 μ L of the sample.

2.2.10.1.4 Adenylate 3' Ends

A-Tailing mix (12.5 µl) was added to each sample and incubated on a thermal cycler, thermal cycler, (Bio-Rad, 184-1100) at 37 °C for 30 minutes, 70 °C for 5 minutes and held at 4 °C.

2.2.10.1.5 Ligate adaptors

The following reagents (2.5 µl) were added to each tube, RSB, ligation mix, DNA index adaptors, incubated on the thermal cycler, (Bio-Rad, 184-1100) at 30 °C for 10 minutes then held at 4 °C. The reaction was stopped by the addition of stop ligation buffer (5 µl).

2.2.10.1.6 Clean up Ligated fragments

A clean-up step was repeated as before with 42.5 µl of SPB, 2x 80% (v/v) ethanol washes and resuspension in 52.5 µl of RSB resulting in a final volume of 50 µl of the sample in a new 8-strip PCR tube (Starlab, I1402-3700).

2.2.10.1.7 Enrich DNA fragments

PCR primer cocktail (5 µl) and enhanced PCR mix (20 µl) was added to each sample. Samples were placed on the thermal cycler (Bio-Rad, 184-1100) for 95 °C for 3 minutes, 12x cycles of 98 °C for 20 seconds, 60 °C for 15 seconds, 72 °C for 5 minutes and held at 4 °C.

2.2.10.1.8 Clean up amplified DNA

SPB (40 µl) was mixed with each sample, transferred to a midi plate, (Abgene, AB0859) incubated for 5 minutes at room temperature, placed onto a magnetic stand until the liquid cleared and 82 µl of supernatant transferred to a new midi plate, (Abgene, AB0859).

A clean-up step was repeated as before with 82 µl of SPB, 2x ethanol washes and resuspension in 17.5 µl of RSB resulting in a final volume of 15 µl of the sample in a new 8-strip PCR tube (Starlab, I1402-3700).

2.2.10.1.9 Validate and quantify libraries

The DNA concentration was quantified using the Qubit™ 3.0 Fluorometer using methods described previously. DNA Library size was checked by electrophoresis using the Agilent 2200 TapeStation Nucleic Acid System (Agilent, G2965AA), Agilent 2200 TapeStation Software A.01.05 and D1000 High Sensitivity ScreenTape (Agilent, 5067-5584) and the D1000 High sensitivity reagents (Agilent, 5067-5585).

2.2.10.1.10 Hybridise probes

Libraries were pooled in groups of three using 500ng of each library and a final volume of 40 µl of RSB. To each library pool capture target buffer, (50 µl) and coding exome oligo's (10 µl) were mixed and placed on the thermal cycler (Bio-Rad, 184-1100) for 95 °C for 10 minutes, 10x cycles of 1 minute at 94 °C, decreasing by 2 °C per cycle and held at 58 °C for at least 90 minutes.

2.2.10.1.11 Capture hybridised Probes

Streptavidin magnetic beads, SMB (250 µl) were mixed with the sample in a 1.5ml DNALoBind tube, incubated for 25 minutes, placed onto a magnetic stand until the liquid cleared and the supernatant discarded. Streptavidin wash solution (200 µl) was mixed with each sample, placed on a heat block at 50 °C for 30 minutes then onto the magnetic stand until the liquid clears and the supernatant discarded. This wash was repeated as above once.

Elution pre-mix (23 µl) was mixed with the SMB to re-suspend the sample, incubated at room temperature for 2 minutes and placed on a magnetic stand until the liquid cleared. The supernatant (21 µl) was transferred to a new 8-strip PCR tube (Starlab, I1402-3700).

2.2.10.1.12 Perform second hybridisation

To each library pool capture target buffer (50 µl) and coding exome oligo's (10 µl) were mixed, and placed on the thermal cycler, (Bio-Rad, 184-1100) for 95 °C for 10 minutes, 10x cycles of 1 minute at 94 °C, decreasing by 2 °C per cycle and held at 58 °C for at least 14.5 hours.

2.2.10.1.13 Perform the second capture

An identical 'Capture hybridised probes' step was performed as before. The resulting supernatant (21 µl) was transferred to a new 8-strip PCR tube (Starlab, I1402-3700).

2.2.10.1.14 Clean up captured library

A clean-up step was repeated as before with 45 µl of SPB, 2x ethanol washes and resuspension in 27.5 µl of RSB resulting in a final volume of 25 µl of the sample in a new 8-strip PCR tube (Starlab, I1402-3700).

2.2.10.1.15 Amplify enriched library

PCR primer cocktail (5 µl) and enrichment amplification mix (20 µl) were mixed with each sample and placed on the thermal cycler (Bio-Rad, 184-1100) at 98 °C for 30 seconds, 10x cycles of 98 °C for 10 seconds, 60 °C for 30 seconds, 72 °C for 5 minutes and held at 4 °C.

2.2.10.1.16 Clean up the amplified enriched library

A clean-up step was repeated as before with 45 µl of SPB, 2x ethanol washes and resuspension in 22 µl of RSB resulting in a final volume of 20 µl of the sample in a new 8-strip PCR tube (Starlab, I1402-3700).

2.2.10.1.17 Validate enriched library

The DNA concentration was quantified using the Qubit™ 3.0 Fluorometer using methods described previously. DNA Library size was checked by electrophoresis using the Agilent 2200 TapeStation Nucleic Acid System (G2965AA), Agilent 2200 TapeStation Software A.01.05 and D1000 High Sensitivity ScreenTape (Agilent t, 5067-5584) and the D1000 High sensitivity reagents (Agilent, 5067-5585). Libraries were frozen at -20 °C until sequenced before sequencing on the Next seq 500 with a 150-cycle high Output flow cell (in batches of 12 samples).

2.2.10.2 Exome sequencing bioinformatics

Sequencing reads were assessed using FastQC. The Phred quality scores are logarithmically related to the probability of errors in base-calling. Sequences with a

Phred score of <30 were removed giving a base call accuracy of 99.9%. In accordance with Illumina sequencing coverage recommendations, 12 samples were multiplexed on a high output flow cell to give 50X coverage and a minimum of 33.3 million reads. Sequence reads were aligned to the human reference genome, (NCBI build 37 (hg19)) using the Burrows-Wheeler Aligner (BWA) package (190). SAMTools was used to generate chromosomal coordinate-sorted BAM files and Picard was used to remove sequence duplicates which suggest PCR duplication (191). Somatic variants were called from matched tumour-normal samples using Strelka2 in tumour/normal mode (192). Dr Andrew Beggs performed the initial QC, alignment and somatic variant calling discussed above. The subsequent mutation filtering, annotation and analysis were performed by me.

Somatic variants were viewed, filtered and annotated in Genomics Workbench (193). Variants which are present with greater than 1% of the population are unlikely to result in disease. Therefore mutations with a MAF of >1% in known variant databases, (dbSNP and 100,000 genomes) were filtered. Variants were compared to known variant databases, (dbSNP and 100,000 genomes) to annotate experimental variants with information from known variants and to identify novel rare variants. Significant somatic mutations are likely to be clonally selected for and conserved during evolution. Therefore, to add information to novel variants the conservation score was added, (PhastCons). The PhastCons ranges from 0-1 with a score of 1 indicating the mutation is highly conserved. Mutations were further annotated to include genomic information, exon number and functional consequences such as amino acid change. The prognostic CPM groups were compared in genomics workbench using Fischer exact test, (applied to the number of occurrences of each allele of the variant in the case and control cohorts) to identify potential candidate driver mutations for poor prognosis CPM.

Somatic mutations for each tumour cohort, (primary minus normal, CPM minus normal and CPM minus primary) were entered into the IntOGen platform for further analysis (194). The IntOGen-mutation platform incorporates several pipelines to determine the functional consequences of mutations and to identify cancer driver mutations (194).

The OncodriveFM pipeline receives as input the list of synonymous, nonsynonymous and frameshift-indel mutations. It adds to this three scores which describe the functional impact of mutations on amino acid structure and function, (Sorting Intolerant From Tolerant, (SIFT) (195), PolyPhen2 (196) and Mutation Assessor scores) (194, 197). OncodriveFM then assesses whether any gene shows a trend towards the accumulation of mutations with a high functional impact, as compared to the background distribution of these functional impact scores in all mutations detected across the cohort of tumour samples (194). The resulting *P*-values are combined using Fisher's approach to produce an integrated *P*-value, (to allow for possible non-dependence of the 3 values an FDR of <0.05 is applied) and a Q-value, (multiple testing corrected *P*-value) (194). These scores are transformed using transFIC to compensate for the tolerance of different proteins to functional variants, to assess the likelihood that a specific somatic mutation is a cancer driver (194). The OncodriveCLUST pipeline receives nonsynonymous and silent mutations as input. It then assesses the clustering of protein affecting mutations to identify relevant activated pathways (194).

Tumour mutational burden was calculated by dividing the number of non-synonymous mutations by the exome panel size (45 Mb).

Microsatellite instability was determined using the Microsatellite instability classifier in R studio (by Dr Andrew Beggs) (198, 199).

2.2.11 DNA panel sequencing

Target enrichment allows the targeting of specific DNA regions of interest, for example, a panel of selected genes. This allows increased sequencing depth whilst minimising cost. The Qiagen QIAseq™ panel integrates unique molecular indices or barcodes (UMIs) which are integrated into a single gene-specific, primer based targeted enrichment process. Each unique DNA molecule is barcoded before amplification, sequence reads with different barcodes represent original molecules whilst sequences with the same barcode are results of PCR duplication. The kit contains 1197 primers in a single tube targeting 30 CRC specific genes (Table 2.1), (Figure 2.10).

Table 2-1 Genes targeted by the QIAseq™ targeted panel CDHS-14542Z-1197

ACVR2A	APC	ARID1A	ATM	B2M	BCL9L	BMPR2	BRAF
FBXW7	GNAS	KRAS	MLH1	MSH2	MSH6	NRAS	PIK3CA
PTEN	RNF43	RPL22	SMAD2	SMAD4	SOX9	TCF7L2	TGIF1
ZFP36L2	POLE	ELF3	TP53	POLD1	CTNNB1		

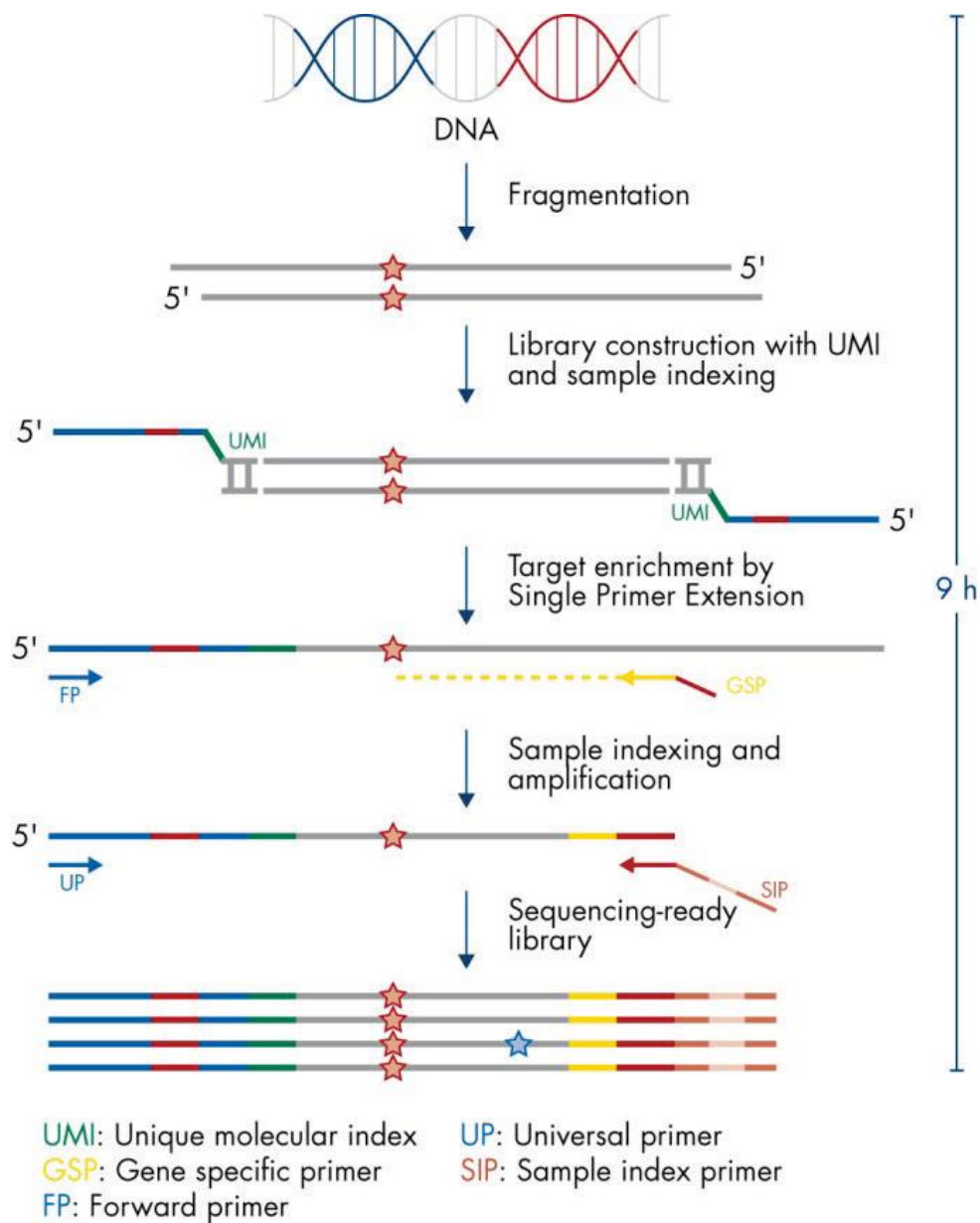


Figure 2-10 QIAseq™ targeted DNA panel

Schematic representation of the QIAseq™ targeted DNA panel workflow. The DNA sample undergoes enzymatic fragmentation, adaptors are ligated including unique molecular indices and sample indices, target enrichment is performed with primer extension, sample indexing and amplification prior to sequencing. Figure reprinted from the Qiagen targeted DNA panel handbook, Figure 2 (200).

2.2.11.1 DNA panel sequencing method

Following the addition of each reagent, samples were pipette mixed and pulse centrifuged. Following each thermal cycler step samples were pulse centrifuged. Text in italics indicates the modifications made for the sequencing of ctDNA.

2.2.11.1.1 Enzymatic fragmentation

In an 8-strip PCR tube (Starlab, I1402-3700) 40 ng of normalised DNA in 16.75 µL of NFW was added to 2.5 µl of fragmentation buffer, 0.75 µl FERA solution *and 1.25 µl of FG solution*. On ice 5 µl of Fragmentation enzyme was added and the reaction incubated on a thermal cycler, (Bio-Rad, 184-1100) for 32 °C for 24 minutes, 72 °C for 30 minutes and held at 4 °C.

2.2.11.1.2 Adaptor ligation

Ligation buffer (10 µl), an IL-N7 index (2.8 µl, *0.5 µl*) DNA ligase 5 µl, ligation solution 7.2 µl *and 2.3 µl of NFW* was added to the reaction and incubated at 20 °C on a thermal cycler for 15 minutes, (Bio-Rad, 184-1100).

2.2.11.1.3 Clean-up of adaptor-ligated DNA

NFW was added to bring the sample to a volume of 100 µl, (*80 µl*) and transferred to a midi plate (Abgene, AB0859). The samples were incubated with 100 µl, (*112 µl*) of QIAseq magnetic beads for 5 minutes, placed onto a magnetic stand until the liquid cleared and the supernatant discarded. The bead pellet was washed with 2x 200 µl 80% (v/v) ethanol washes and resuspended in 52 µl of NFW, 50 µl was transferred to a new midi plate

A clean-up step was repeated as before with 50 µl, (*70 µl*) of QIAseq beads, 2x 200 µl 80% (v/v) ethanol washes and resuspension in 12 µl of NFW resulting in a final volume of 9.4 µl of the sample in a new 8-strip PCR tube (Starlab, I1402-3700).

2.2.11.1.5 Target enrichment

TEPCR buffer (4 µl), the QIAseq DNA panel (5 µl), Illumina forward primer (0.8 µl) and HotStarTaq DNA polymerase (0.8 µl) was added to the sample and incubated on a

thermal cycler (Bio-Rad, 184-1100) at 95 °C for 13 minutes, 98 °C for 2 minutes, 8x cycles of 98 °C for 15 seconds, 68 °C for 10 minutes, 72 °C for 5 minutes and held at 4 °C.

2.2.11.1.6 Clean-up of target enrichment

NFW was added to result in a volume of 100 µl, (90 µl) and transferred to a midi plate, (Abgene, AB0859). The samples were mixed with 100 µl, (108 µl) of QIAseq magnetic beads and incubated for 5 minutes at room temperature, placed onto a magnetic stand until the liquid cleared and the supernatant discarded. The bead pellet was washed with 2x 200 µl 80% (v/v) ethanol washes and re-suspended in 16 µl of NFW, 13.4 µl was transferred to a new 8-strip PCR tube (Starlab, I1402-3700).

2.2.11.1.7 Universal PCR

UPCR buffer (4 µl), HotStarTaq DNA polymerase (1 µl) and NFW (1.6 µl) were added to the sample and mixed with a dry unique index, IL-S502-S511. The reaction was incubated on a thermal cycler (Bio-Rad, 184-1100) with the following settings. (95 °C for 13 minutes, 98 °C for 2 minutes, 8x cycles of 98 °C for 15 seconds, 68 °C for 10 minutes, 72 °C for 5 minutes and held at 4 °C).

2.2.11.1.8 Clean-up of universal PCR

NFW was added to result in a volume of 100 µl, (90 µl) and transferred to a midi plate, (Abgene, AB0859). The samples were mixed with 100 µl, (108 µl) of QIAseq magnetic beads and incubated for 5 minutes at room temperature, placed onto a magnetic stand until the liquid cleared and the supernatant discarded. The bead pellet was washed with 2x 200 µl 80% (v/v) ethanol washes and re-suspended in 30 µl of NFW, 28 µl was transferred to a new 8-strip PCR tube (Starlab, I1402-3700).

2.2.11.1.9 Validate enriched library

The DNA concentration was quantified using the Qubit™ 3.0 Fluorometer using methods described previously. DNA Library size was checked by electrophoresis using the 2200 TapeStation Nucleic Acid System (Agilent, G2965AA), Agilent 2200 TapeStation Software A.01.05 and D1000 High Sensitivity ScreenTape (Agilent, 5067-

5584) and the D1000 High sensitivity reagents (Agilent, 5067-5585). Libraries were frozen at -20 °C until sequenced on the Nextseq 500 with a 300-cycle mid Output flow cell.

2.2.12 Biosigner classifier development

For both the retrospective and prospective patient cohorts unaligned RNA-seq data was uploaded into the Partek® Flow® genomics suite software package (Partek, St Louis, MI, USA). Pre-alignment quality assurance/quality control (QA/QC) was performed to check the total number of reads per sample, read length and quality. Trimming was performed to remove any low-quality reads and the data aligned to the reference genome hg19. Following alignment, a further QA/QC step removed any unaligned reads. Gene counts for both the retrospective and prospective cohorts were downloaded and labelled with Entrez numerical gene IDs. Normalised methylation beta values for both the retrospective and prospective cohorts were downloaded.

Gene counts or beta values were loaded into RStudio version 3.5.0 software and Bioinformatics analysis performed using the Biosigner R package, version 2.10.2 (63). The Biosigner package generates test subsets of variables by random bootstrapping. As the dataset has a small number of samples the recommended number of bootstraps of 50 was doubled to 100 to increase the stability of the resulting classifier. Biosigner selects variables, which significantly contribute to the prediction performance of a binary classifier, in this case, prognosis. The procedure is repeated by applying the algorithm to the dataset restricted to the significant variables until for a given round all candidate variables are found significant if there is no variable left to be tested. The Biosigner algorithm returns stable signature (S), in addition to several tiers from A – E corresponding to variables discarded during previous iterations. Signatures from three distinct classifiers are run in parallel (PLS-DA, RF and SVM) as the performance of each classifier varies depending upon the dataset structure. The Biosigner classifier R code is displayed in the appendices, supplementary table S2.

2.2.13 Circulating tumour DNA sequencing techniques

ctDNA analysis can focus on several targets, simple quantification of DNA levels, analysis of single or panels of mutations, methylation and clonality. A variety of methods are available for the analysis of ctDNA. Digital PCR methods detect ctDNA mutations by partitioning DNA fragments in an emulsion (201). BEAMing is a digital PCR method which utilises beads, emulsion, amplification, and magnetics (201). Water droplets in an oil emulsion act as reaction vessels containing a mixture of template, primers, PCR reagents, and magnetic beads (201). Fluorescently labelled dideoxynucleotide terminators discriminate wild type and mutant DNA sequences and are analysed by flow cytometry (201). Droplet-digital PCR is based on water-oil emulsion droplet technology by the distributes the DNA sample into thousands to millions of droplets containing a single mutated or wild type strand that can be distinguished by flow cytometry using fluorescent TaqMan-based probes (201). Real-time PCR is a rapid, cheap method to detect ctDNA mutations at an allele frequency of 10-20% (201). Several modifications improve sensitivity. Most of these are based on either using a blocking oligo at the 3'-end to block the amplification of the normal allele and allowing the amplification of the mutant allele or they make use of a modification step in the PCR protocol that enriches variant alleles from a mixture of wild-type and mutation-containing DNA (201). NGS analyses millions of short ctDNA sequences in parallel, these are then aligned to the reference genome (201). NGS in ctDNA analysis can focus on specific target regions of known mutation panels or analyse novel mutations via WES or WGS (201).

2.2.13.1 Circulating tumour DNA, bioinformatics

ctDNA levels in ng/ μ L were examined at each time point. The DFS for patients with undetectable ctDNA post-CRS & HIPEC was compared to those with detectable ctDNA using the Kaplan Meier method. Median mutation number and levels of ctDNA in ng/ μ L were compared to that detected in the tumour and at subsequent time points using Mann-Whitney U. The correlation between tumour and baseline ctDNA mutation number was examined using Pearson's coefficient correlation. All statistical tests were performed in IBM SPSS Statistics for Windows, Version 24.0 (159).

Sequencing reads in FastQC format were uploaded to the Qiagen Biomedical Genomics Workbench. Reads were trimmed to remove adaptors and primer-dimer artefacts (<40bp). Sequence reads were aligned to the human reference genome, (NCBI build 37 (hg19)) using the Burrows-Wheeler Aligner (BWA-MEM). Poorly mapped reads were filtered, and gene-specific primer sequences trimmed and annotated with UMI. Variant calling was performed using smCounter2. SmCounter2 determines whether the allele frequency of the variant is significantly above the background error rate.

Lollipop plots were produced to determine the distribution and frequency of mutations as well as the associated amino acid changes for the most frequent variants. VCF files were converted to mutation annotation format using vcf2maf and then merged into a single file stratified by sample identifier (202). The MAF file was then loaded into R/Bioconductor and processed using MAFtools to generate lollipop plots.

Chapter 3 A meta-analysis of prognostic factors for patients with colorectal peritoneal metastasis undergoing cytoreductive surgery and heated intraperitoneal chemotherapy.

Some of the work in this chapter has been previously published in Hallam *et al*/2019 (203). In this publication I was responsible for the writing of the text, collection of the data and also production any of the figures used in this thesis.

3.1 Introduction

CPM occurs in up to 15 per cent of cases of CRC (75-77). The prognosis for CPM is poor, untreated median OS is 6 just months (79). With systemic chemotherapy, the median OS is improved to up to 20 months (84-86). The standard for CPM is CRS & HIPEC which may improve median OS by 20 – 63 months (108, 109, 112, 115, 117, 118, 124, 130, 137).

CRS & HIPEC is a long, high-cost operation associated with a protracted inpatient and high dependency (HDU) or intensive care (ITU) stay, an associated mortality rate of 1-12 per cent and morbidity rate of 7-63 per cent (112, 121, 123, 124, 127, 130, 135, 138, 144, 145). Improving patient selection is therefore crucial to maximising patient outcomes whilst minimising morbidity and mortality.

Variation in outcomes for CRS & HIPEC can be explained in part by patient selection, for example necessitating the ability to achieve complete cytoreduction at CRS, the exclusion of patients with extensive CPM as assessed by Sugarbaker's peritoneal carcinomatosis index (PCI), and selection of patients with minimal co-morbidity and good performance status.

Several clinicopathological variables which impact on survival have been identified in the literature including lymph node status, tumour differentiation and histological findings, the completeness of cytoreduction (CC score) and PCI. There is wide variation however in the variables reported by studies and selection criteria for centres performing CRS & HIPEC worldwide and in turn in their outcomes.

Studies reporting prognostic factors in CRS & HIPEC are cohort in design with small samples. Each examines numerous and varying prognostic factors on different scales of measurement. No consensus exists as to which prognostic factors contraindicate CRS & HIPEC or predict a good outcome. Comprehensive evidence synthesis is therefore called for to determine relevant prognostic factors.

The aim of this systematic review and meta-analysis was to analyse all prognostic factors affecting OS in patients with CPM undergoing CRS & HIPEC.

3.2 Method

3.2.1 Literature search

A comprehensive literature search was conducted in accordance with the Preferred Reporting Items for Systematic Review and Meta-Analysis (PRISMA) (204). MEDLINE, Embase and Cochrane Library electronic databases, registers of clinical trials (ClinicalTrials.gov and the WHO International Clinical Trials Registry) and the Conference Proceedings Citation Index, Zetoc, were searched from inception to the present. The search strategy captured terms for colorectal cancer, peritoneal metastasis and CRS & HIPEC techniques, separated by the Boolean operator 'AND'. Searches were supplemented by a hand search of selected journals and the reference lists of all included studies.

Example Search Strategy:

-
- 1 exp Colorectal Neoplasms/
 - 2 ((colon* or colorectal) adj2 (cancer or carcinoma or tumo\$r or malignan*)).ti,ab.
 - 3 exp Peritoneal Neoplasms/
 - 4 ((peritoneal or peritoneum) adj2 (cancer or carcinoma* or malignan\$ or spread* or neoplasm*)).ti,ab.
 - 5 1 or 2
 - 6 3 or 4
 - 7 cytoreductive surgery.ti,ab.
 - 8 ((intraperitoneal or intra-peritoneal) adj2 chemotherapy).ti,ab.
 - 9 CRS.ti,ab.
 - 10 HIPEC.ti,ab.
 - 11 7 or 8 or 9 or 10
 - 12 5 and 6
 - 13 11 and 12

3.2.2 Study selection

English-language studies were eligible for inclusion if they reported on the impact of prognostic factors on OS for patients with CPM undergoing CRS & HIPEC. Where multiple studies describe the same cohort of patients, the largest and most complete dataset was included. Review articles, case reports or case series of fewer than 10 patients were excluded. Additional exclusion criteria included studies involving patients with primary tumours other than CRC and studies where a proportion of the cohort did not receive both CRS & HIPEC.

After screening the titles and abstracts, articles fulfilling the eligibility criteria were identified and their full-text publications reviewed. Literature search and study selection were done independently by two researchers, and any disagreements were resolved by discussion with senior reviewers. After a qualitative assessment, articles were screened to ensure they presented adequate statistical information to be included in the meta-analysis: hazard ratios (HRs) with confidence intervals or Kaplan–Meier curves with the number of events and patients at risk.

3.2.3 Assessment of risk of bias

The quality and risk of bias of individual studies were assessed using the Quality in Prognosis Studies tool (QUIPS) tool (205). This tool reviews each study in six criteria: study participation, attrition, prognostic factor measurement, outcome measurement, confounding factors and the statistical analysis and reporting. Two authors scored all articles independently (SH & RT).

3.2.4 Data Extraction

Data were independently extracted by two reviewers (SH & RT) using a piloted data extraction form. The number of patients, study design, patient demographics, tumour characteristics, use of adjuvant and neoadjuvant regimens, CRS & HIPEC techniques, survival and prognostic factors were recorded.

The unadjusted hazard ratio (HR) and its 95% confidence interval (c.i.) and *P*-value were extracted. When adjusted HRs (with confidence intervals and *P*-values) were reported, these were extracted along with the set of adjustment factors. If HRs, confidence intervals or *P*-values were not provided directly, the methods of Tierney et

al were used to estimate them indirectly from Kaplan Meier curves, when presented in adequate detail with the numbers of events and patients at risk (206, 207).

Prognostic factors reported on a continuous scale were extracted. If results were categorised into three or more categories, results for each comparison were extracted and when clinically relevant these were grouped to form a binary comparison.

3.2.5 Prognostic factor selection

All prognostic factors adequately described and reported by two or more independent studies were included.

3.2.6 Statistical analysis

Owing to clinical and methodological heterogeneity, a random-effects meta-analysis was used (on the log(HR) scale) using the method of DerSimonian and Laird (208). The combined effect size was described by the pooled HR, its confidence interval and *P*-value, with *P* < 0.050 considered significant. For prognostic factors reported by more than two studies, a 95 per cent prediction interval (a measure of the variation in treatment effects) is presented (209). Heterogeneity was also described by the *I*² statistic (210). All analyses were performed in STATA® version 15 (StataCorp, College Station, Texas, USA).

3.3 Results

The final literature search was performed on the 3rd of April 2018. Literature searches (after removal of duplicates) identified 1052 records. Titles and abstract screening identified 158 abstracts for full-text review. Of these, fifty-one studies met the inclusion criteria. Twenty-four unique studies reporting on 3128 patients with CPM presented adequate data to be included in the meta-analysis (112, 123, 135, 139, 141, 211-229) (Figure 3.1). Of the 24 cohort studies six were prospective (123, 214, 217, 219, 223, 226), and the remaining 18 retrospective (112, 135, 139, 141, 211-213, 215, 216, 218, 220-222, 224, 225, 227-229) (Table 3.1).

Several studies presented both an unadjusted and adjusted hazard ratio. As adjustment factors varied widely between studies (table 3.2) meta-analysis was stratified according to whether the hazard ratio was unadjusted or adjusted.

There was a low risk of bias from study participation. The moderate risk of bias due to study attrition reflects the poor reporting of loss to follow-up. There was a low risk of bias due to prognostic factor measurement. Prognostic factors were objective, clearly defined and clinically relevant. Reporting of all prognostic factors and treatment variations was incomplete in the majority of studies, resulting in a moderate risk of bias. The presentation of results and analytical strategy was sufficient in the majority of studies resulting in a low risk of bias (Table 3.3).

Pooled median overall survival across all studies was 32 months (12.2 – 51 range) with a pooled median follow up of 28.1 months (13.3 - 62.4 range) (Table 3.4)

Table 3-1 Demographics

Author (year)	Country of origin	Study design	Number patients	Age, years (range)	Gender (M:F)	Follow up (months, median (range)	Comorbidity (range)	Performance status
Baratti (2014)	Italy	RC	101	60 (29-79)	40:61	44.9 (24.1- 65.7 95%CI)	Charlston score 4 (2-7)	ECOG0 54(53) ECOG1 42(42) ECOG2 5(5)
Beniziri (2012)	France	RC	49	52.7 \bar{x} +/-SD 11	19:30	27 \bar{x} +/-8	NS	ECOG<2 (100)
Cashin (2012)	Sweden	RC	126	55 \bar{x} (14-79)	76:75	49 (0.5-200)	NS	WHO <2 (100)
Da Silva (2006)	USA	RC	70	45.5 \bar{x} (18-71)	27:43	46.5 \bar{x} (6-241)	NS	NS
Duraj (2013)	Sweden	PC	11 LM+PM 22 PM	60 LM+PM 57 PM	2:9 LM+PM 7:15 PM	57 LM+PM, 45 PM	NS	NS
Elias (2010)	International centre (n23)	RC	523	54 (16-88)	227:296	45 (23-79 IQR)	NS	WHO0 387(76) WHO1 102(20) WHO2 20(4)
Elias (2013)	France	RC	114	50 +/-SD 10	49:65	46.8 (0-140.4)	NS	NS
Elias (2014)	France	PC	139	SB 49.1 NSB 49.7	SB 37:63 NSB 14:25	62.4 (55.6- 77.6)	NS	NS
Faron (2016)	France	RC	173	48.9 (20-74)	71:102	48 (41.2-56.3)	NS	NS
Franko (2008)	USA	RC	65	51.6+/- SD 13.2	18:47	NS	NS	NS
Froysnes (2016)	Norway	RC	119	58 (22-77)	42:77	45	NS	ECOG 0-1 (100)
Huang (2014)	China	PC	60	< 60 46 (77) > 60 14 (23)	26:34	NS	NS	NS
Huang (2016)	Australia	RC	168	534 \bar{x} SD 14.1	51:104	20.1 (0.2-112)	NS	NS

Ihemelandu (2017)	USA	RC	318	50.6 \bar{x} (18-86)	171:147	15 (3-264)	NS	NS
Lorimer (2017)	France	RC	22 PM+LM 36 PM	60 (21-77) median NS 0.82	29:29	29.2 (5-125)	NS	ECOG <2 (100)
Maggiori (2013)	France	PC	37 PM+LM 61 PM	PM+LM 49 (27-67) PM: 50 (25-63)	PM+LM: 11:26 PM: 16:45	36+/-28	NS	ECOG <2 (100)
Ng (2016)	Singapore	RC	50	50 (14-71)	18:32	13.3 (0.8 – 87.1)	n31 yes	ECOG0 41 (82) ECOG1 9(18)
Shen (2004)	USA	PC	77	54 \bar{x}	45:32	15	NS	ECOG0 18(23) ECOG1 46(60) ECOG2 7(9) ECOG3 2(3) ECOG4 4(5)
Simkins, Razenberg (2016)	Netherlands	RC	445	69.3+/-SD 11.6 AC 68.5+/-SD 12.1 MC 65.3+/-SD 14.2 SR	NS	NS	NS	NS
Simkens, Van Oudheusden (2016)	Netherlands	PC	58 colon 29 rectal	59.4+/-SD 11.2 62.3 +/-SD 9.4	14:15 rectal 28:30 colon	21.8 (3.2-71.3) rectal 21.3 (2.5-96.6) colon	NS	Colon:rectum ASA 1 4:5 ASA 2 22:45 ASA 3 3:5
Sluiter (2016)	Netherlands	RC	65	62 (31-78)	25:40	21 (2-50)	NS	NS
Teo (2015)	Singapore	RC	35	51 (14-71)	11:24	24.7 (0.6-81.8)	NS	ECOG \leq 1 35(100)
Ung (2013)	Australia	RC	125	58.9 (NS)	50:75	23.3 (1-156)	NS	NS
Winer (2014)	USA	RC	30	54.9 \bar{x} +/-SD 14.5	22:8	52 (IQR 1.04-11.5)	NS	NS

(NS not specified, RC retrospective cohort, PC prospective cohort, LM liver metastasis, PM peritoneal metastasis, AC adenocarcinoma, MC mucinous, SR signet ring, ECOG Eastern Cooperative Oncology Group performance status score, ASA American Society of anaesthesiologists score, SB small bowel, NSB not small bowel, NS not specified, \bar{x} mean, IQR inter-quartile range, CI confidence interval, SD standard deviation, number followed by value in parenthesis, percentage).

Table 3-2 Adjustment factors used in multivariate analysis

Study ID	Age	Gender	Colon Rectum	ECOG	Lymph node	Adverse features	Hepatic metastasis	Signet ring Mucinous	Timing PM	PCI	CC score	GI anastomosis	MVR	R Status	Neo-adjuvant	Adjuvant	HIPEC EPIC SPIC	Morbidity	PSS
Baratti (2014)				X			X		X	X	X							X	
Benziri (2012)										X	X								
Cashin (2012)		X								X					X	X	X		X
Da Silva (2006)	X				X	X				X									
Duraj (2013)	X	X	X		X		X			X		X		X			X		X
Elias (2010)					X					X	X					X			
Elias (2013)	X	X	X				X			X									
Elias (2014)	X	X			X		X			X									
Faron (2016)					X					X									
Franko (2008)												X	X						
Froysnes (2016)	X	X	X		X					X									
Huang (2014)										X	X								
Huang (2016)	X				X			X		X									
Ihemalindu (2017)	X	X						X			X								x
Lorimer (2017)			X		X		X	X		X	X	X	X					X	
Maggiore (2013)	X	X	X		X		X			X						X	X		
Ng (2016)										X									X
Shen (2004)	X	X		X	X	X	X					X		X					
Sluiter (2016)	X				X					X									
Simkins, Razenberg (2016)								X											
Simkens, Van Oudheusden (2016)			X																
Ung (2013)	X				X					X									
Winer (2014)	X	X									X							X	

(ECOG, PM peritoneal metastasis, PCI Peritoneal carcinomatosis index, CC score completeness of cytoreduction, GI gastrointestinal, MVR multi-visceral resection, R status resection status, HIPEC heated intraperitoneal chemotherapy, EPIC Early post-operative chemotherapy, SPIC Sequential postoperative intraperitoneal chemotherapy ASA American society of anaesthesiologists score VEGF, Vascular endothelial growth factor, VCAN Versican, LN lymph node, WBC white blood cell).

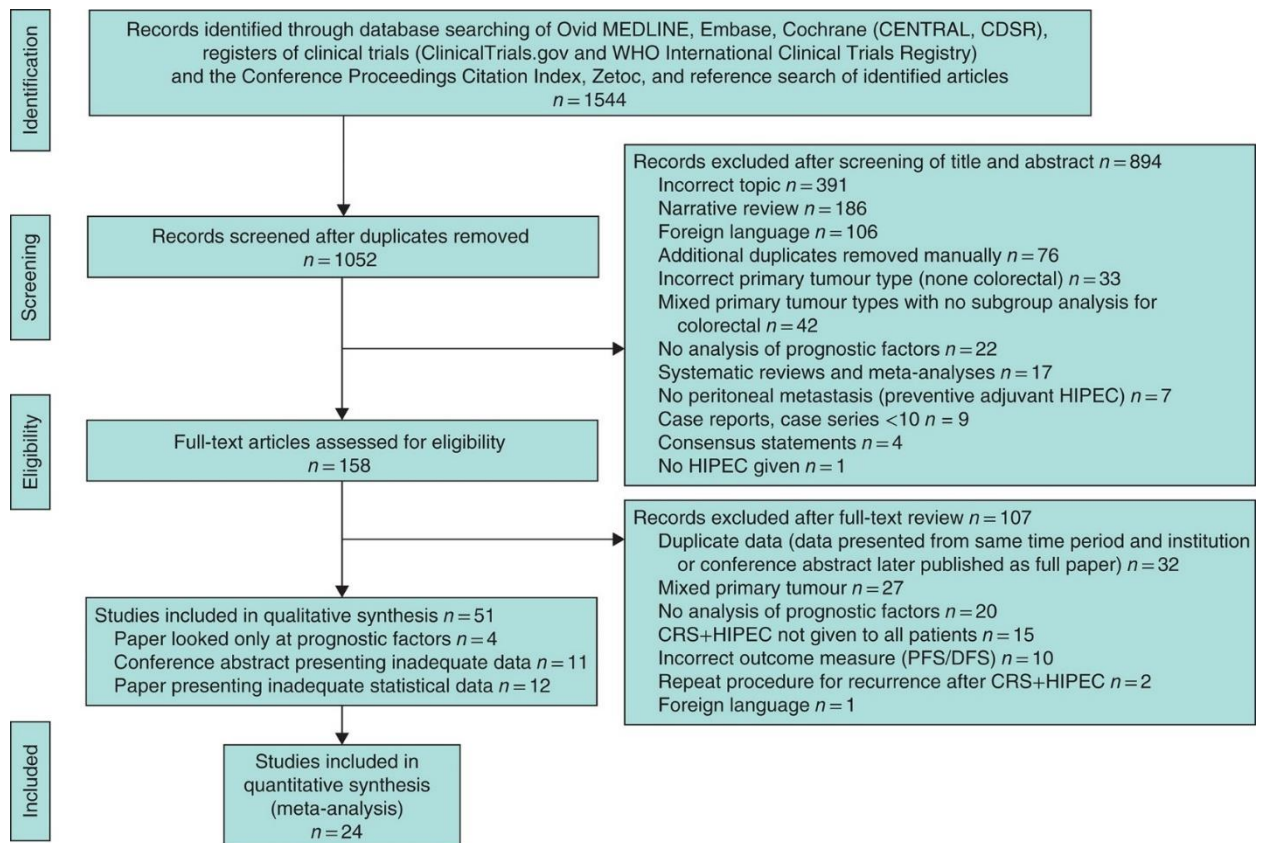


Figure 3-1 PRISMA flow diagram

PRISMA flow diagram detailing the records identified, included and excluded and the reasons for this. This figure is reprinted from Hallam et al, 2019, figure 1 (203).

Table 3-3 Risk of bias assessment results for each study using the Quality in Prognostic Studies (QUIPS) tool

Study ID	Study participation	Study attrition	Prognostic factor measurement	Outcome measurement	Study confounding	Statistical analysis and reporting
Baratti (2014)	Moderate	Moderate	Low	Low	Moderate	Low
Benziri (2012)	Low	Moderate	Low	Moderate	High	Low
Cashin (2012)	Low	Moderate	Low	Low	Moderate	Moderate
Da Silva (2006)	Moderate	High	High	Moderate	Moderate	Low
Duraj (2013)	Low	Moderate	Low	Low	Moderate	Low
Elias (2010)	Low	Moderate	Low	Moderate	Moderate	Low
Elias (2013)	Moderate	Low	High	Low	Moderate	Low
Elias (2014)	Low	Moderate	Low	Moderate	High	Low
Faron (2016)	Low	Moderate	Low	Low	Moderate	Moderate
Franko (2008)	Moderate	Moderate	Low	Moderate	Moderate	Low
Froysnes (2016)	Low	Moderate	Low	Low	Moderate	Low
Huang (2014)	Low	Low	Low	Low	Moderate	Low
Huang (2016)	Low	Moderate	Low	Moderate	Moderate	Low
Ihemalindu (2017)	Low	Moderate	Low	Moderate	Moderate	Low
Lorimer (2017)	Low	Moderate	Low	Low	Moderate	Low
Maggiori (2013)	Low	Low	Low	Low	Moderate	Low
Ng (2016)	High	High	Low	Moderate	Moderate	Low
Shen (2004)	Low	Moderate	Low	Moderate	Moderate	Low
Sluiter (2016)	Low	Moderate	Moderate	Moderate	High	Low
Simkins, Razenberg (2016)	High	High	Moderate	Moderate	Moderate	Low
Simkens, Van Oudheusden (2016)	Low	Moderate	Moderate	Moderate	Moderate	Low
Teo (2015)	Low	Low	Low	Low	Moderate	Low
Ung (2013)	Low	Moderate	Low	Moderate	Moderate	Low
Winer (2014)	Moderate	Moderate	Low	Mow	Moderate	Low

Table 3-4 Survival outcomes

Author (year)	Deaths	Median OS (95% CI)	5-year OS (95% CI)
Baratti (2014)	n=37 (31 disease related)	32 (16.2-55.9)	11.7% (9.5-24.5)
Beniziri (2012)	NS	51 (NS)	NS
Cashin (2012)	NS	34 (2-77) CRS&HIPEC 25 (2-188) CRS&SPIC, $P=0.047$	40% CRS&HIPEC 18% CRS&SPIC, $P=0.047$
Da Silva (2006)	NS	33 (NS)	32% (NS)
Duraj (2013)	NS	15 (6-46) PM+LM 34 (19-37) PM, $P=0.2$	NS
Elias (2010)	NS	30.1 (NS)	27% (21-33%)
Elias (2013)	n=47 (41)	NS	NS
Elias (2014)		SB 30.9 \bar{x} (0.23-113) NSB 45.5 \bar{x} (0.79-144)	39% (NS)
Faron (2016)	NS	41 (32-50)	42% (NS)
Franko (2008)	NS	15.3 (NS) 20.2 R0/1 MVR 32.8 (NS) / Controls 20 = 0.787	NS
Froysnes (2016)	NS	47 (42-52)	NS
Huang (2014)	NS	16 (12.2-19.8)	22%
Huang (2016)	NS	42.1 (33.7-50.4)	34% (NS)
Ihemelandu (2017)	NS	21.5 (NS)	25% (NS)
Lorimer (2017)	NS	PM 25.2 (14.8-82.6) PM+LM 36.1 (19.6-113.7), $P=0.3$	PM: 40.7% (23.4-77.3) PM+LM: 42.1% (19.7-63)
Maggiori (2013)	NS	NS	PM+LM: 26% (18-35) PM: 43% (35-51)
Ng (2016)	n=18 (36)	\bar{x} 28.8 (18.3-39.1)	NS
Shen (2004)	NS	16 (10 – 26)	17% (NS)
Simkins, Razenberg (2016)	NS	32.8 (27.8-37.8)	NS

Simkens, Van Oudheusden (2016)	NS	26 (22.2-29.9) rectal 35.1 (22.8-47.3) colon	32% (NS) rectal 24% (NS) colon
Sluiter (2016)	NS	34.4 (29.4-39.2)	NS
Teo (2015)	n=13 (37)	27.1 (15.3-39.1)	19.1 (NS)
Ung (2013)	NS	37.1 (NS) colon 29.6 (NS) rectum	33% (NS) colon 20% (NS) rectum
Winer (2014)	n=25 (83)	12.2 (7.5-17.2)	NS

(NS not specified, NK not known, \bar{x} median with (range), PM: Peritoneal metastasis, LM liver metastasis, SB small bowel, NSB none small bowel, CRS cytoreductive surgery, HIPEC heated intraperitoneal chemotherapy, EPIC Early postoperative intraperitoneal chemotherapy, SPIC sequential postoperative intraperitoneal chemotherapy, P= P-value, n=number followed by value in parenthesis, percentage).

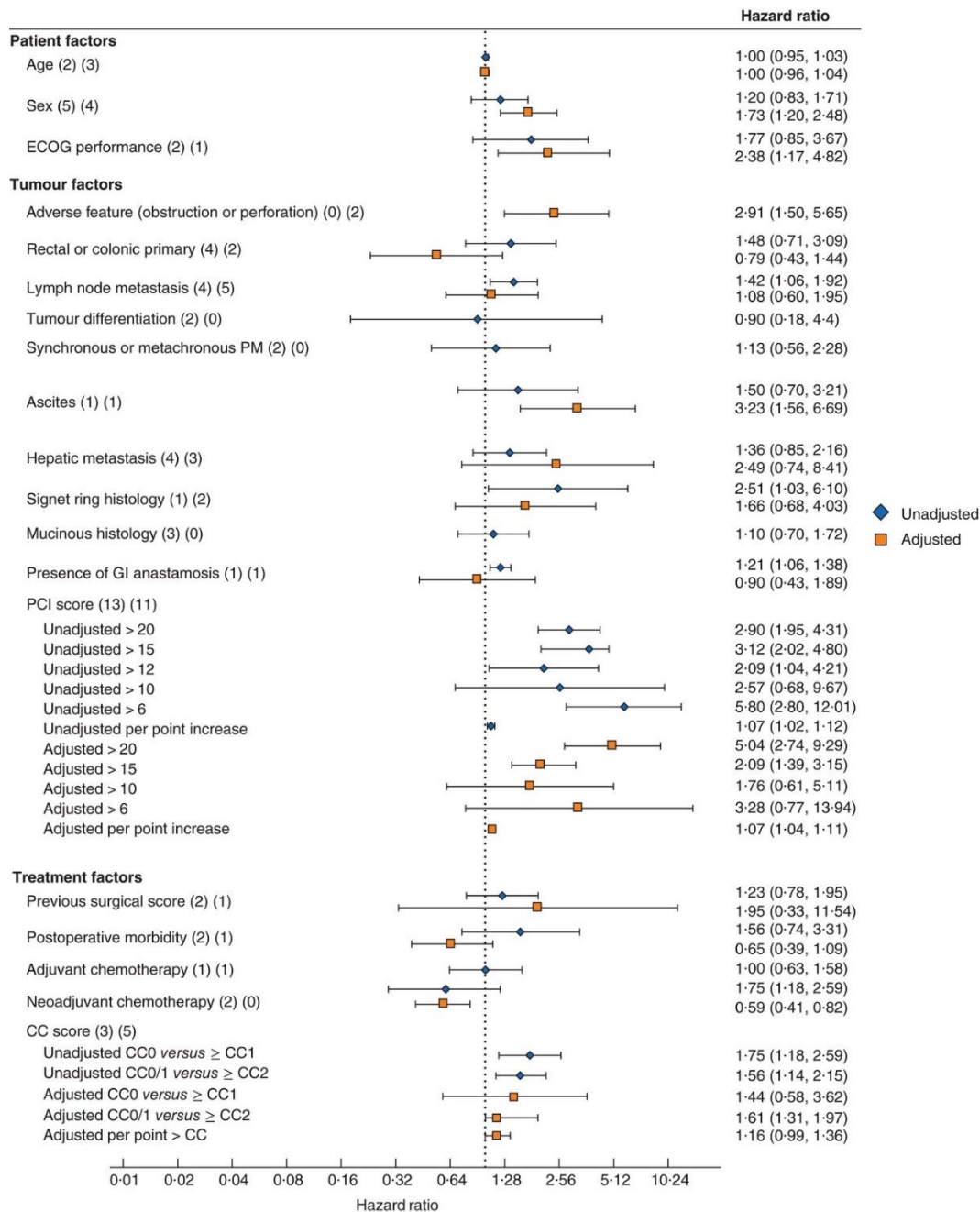


Figure 3-2 Prognostic factor effect on overall survival, (Hazard ratio and 95% confidence interval)

Hazard ratios are shown with 95 per cent confidence intervals. Values in parentheses after each factor indicate the numbers of studies providing unadjusted and adjusted hazard ratios respectively. ECOG, Eastern Cooperative Oncology Group; PM, peritoneal metastasis; GI, gastrointestinal; PCI, Peritoneal Carcinomatosis Index; CC, completeness of cytoreduction. *This figure is reprinted from Hallam et al, 2019, figure 2 (203).*

3.3.1 Prognostic factors

Individual forest plots and funnel plots for each prognostic factor are presented in appendices, figure S1.

3.3.1.1 Patient factors

3.3.1.1.1 Age

The pooled median age of patients was 54 years (range 45.5–69.3). Eleven studies reported on age as a prognostic factor (123, 126, 135, 139, 213, 214, 217, 219, 221, 227, 228). Ten studies reported adequate data to be included in the meta-analysis. Five studies categorised age into binary outcomes which we were unable to meaningfully combine (123, 139, 213, 219, 228). The pooled unadjusted HR was 1 (0.98 – 1.03), $I^2 = 38.5\%$, and adjusted HR 1 (0.96 – 1.04), $I^2 = 69.3\%$, for an increase in age of one year. These data provide no evidence that age is a useful predictor of OS (Figure 3.2).

3.3.1.1.2 Sex

Nine studies reported the effect of sex on OS (123, 126, 135, 214, 216, 217, 219, 221, 229). Eight studies presented adequate data to be included in the meta-analysis. The pooled unadjusted HR was 1.20 (0.83 – 1.71), $I^2 = 33.8\%$ and adjusted HR 1.73 (1.20 – 2.48), $I^2 = 11.6\%$. It is unclear therefore whether sex is a useful predictor of OS.

3.3.1.1.3 ECOG performance status

Five studies reported the influence of ECOG on OS (123, 126, 211, 222, 230). Three studies presented adequate data to be included in the meta-analysis. The pooled unadjusted HR for an ECOG score of at least 2 was 0.78 (0.57 – 1.07), $I^2 = 0\%$. These data provide no evidence that ECOG is a useful predictor of OS.

3.3.1.2 Tumour factors

3.3.1.2.1 Adverse primary tumour features

Three studies reported the influence of adverse features of the primary tumour (obstruction/perforation) on OS (123, 126, 213). Two studies presented adequate data to be included in the meta-analysis. The pooled adjusted HR was 2.91 (1.5 – 5.69) $I^2 = 0\%$. These data indicate adverse primary features are a useful predictor of OS.

Table 3-5 Tumour factors

Author, (year)	Synch: Metach	DFI (month)	Colonic: Rectal	T stage	N stage	Distant metastasis	Differentiation Well:Mod:Poor	Histological type	PCI (range)	CC score 0:1:2:3
Baratti, (2014)	48:54	NS	92:7 RS n=1 NS n=1	NS	NS	LM n=8	8:68:25	NS	10 (1-39)	87:12:2:0
Beniziri, (2012)	NS	20 \bar{x} +/- SD 9	8:41	NS	N+ n=22 N- n=27	NS	6:35:8	NS	10 \bar{x} +/- SD 6.2	37:12:0:0
Cashin, (2012)	102:48	NS	135:15	NS	NS	Excluded	NS	MC n=89 SR n=18	1-10 n=49 11-20 n=45 21-39 n=56	97:0:0:0 0:≥54
Chua, (2010)	22:34	NS	NS	NS	N1/2 n=22	NS	15:22:8	SR 8	<10 n=23 10-20 n=28 >20 n=5	0/1 n=56
Da Silva, (2006)	NS	<12 n=35 >12 n=27	64:6	NS	N0 n=14 N1/2 n=53 NS n=3	NS	11:50:9	MC n=36 SR n=5	<20 n=60 >20 n=10	0/1 n=70
Duraj, (2013)	8:3 PM+LM 15:7 PM	NS	9:2 PM+HM 18:4 PM	NS	PM+HM N+ n=8, N0 n=3 PM N+ n=14, N0 n=6, NS n=2	LM	0:10:0 PM+HM, 2:13:7 PM	MC 5/11, 11/22	NS	R1 n=10 R2 n=1 PM+HM R1 n=20 R2 n=2 PM
Elias, (2010)	161:300	NS	433:36 RS n=8 NS n=73	NS	LN+ n=325 N0 n=158	LM 77	NS	NS	1-6 n=181 (37) 7- 12 n=132 (28) 13- 19 n=96 (21) >19 n=69 (14)	439:53:22:0
Elias, (2013)	NS	NS	94:23	NS	NS	LM n=24 Ovarian n=46	NS	NS	9.2 \bar{x} (0-27)	0/1 n=114

Elias, (2014)	NS	NS	NS	NS	NS	LM SB n=16, NSB n=11	SB 50:0:18 NSB 19:0:9	SB MC n=28, SR n=10 NSB MC n=10, SR n=3	SB 12x(3- 36) NSB 4 (2-11) $P=<0.001$	0/1 n=139
Faron, (2016)	NS	NS	135:38	NS	NS	LM n=52	NS	NS	10.2 \bar{x} +/-6.8	0/1 n=173
Franko, (2008)	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS
Froysnes, (2016)	73:46	15x(7- 67)	109:10	T1 n=1 T3 n=54 T4 n=63 NS n=1	N0 n=34 N1 n=41 N2 n=42 N2 n=2	Excluded	8:67:33 NS n=11	SR n=14 MC n=33	9x(0- 28) 0- 10 n=74 11-20 n=35 >20 n=10	113:5:0:0 NS n=1
Huang, (2014)	24:36	NS	35:25	NS	NS	NS	AC, well / mod diff. n=26 AC, poor diff. / MC / SR n=34		<20 n=28 > 20 n=32 21x(1- 39)	17:15: \geq 28
Huang, (2016)	Metach n=168 (100)	NS	NS	NS	RPLN n=102	Excluded	3:97:27 Benign n=22	SR n=11	9.5 \bar{x} +/- SD 6.6	0/1 n=168
Huo, (2016)	NS	NS	NS	NS	NS	NS	NS	NS	NS	0/1 n=164
	**Specified for patients with high and low tumour marker levels not for the whole group									
Ihemelandu (2017)	NS	NS	NS	NS	NS	NS	NS	AC n=150 SR n=11	15.2 \bar{x} +/-SD 11.2	\leq 1 188:37:80
Lorimer, (2017)	n=58 (100) synch	NA	54:4	T>2 n=55	N+ n=43	LM n=22	NS	MC n=16 (28)	13x(1- 39)	44:9:5:0
Maggiori, (2013)	n=98 (100) synch	NA	50:11 PM 32:5 PM+LM	PM T/2 n=1 T3/4 n= 60 PM+L M T3/4 n=37	PM N+ n=54, N- n=7 PM+LM N+ n=31 N- n=6	LM n=37	NS	NS	PM: 9 (2-26) PM+LM : 11 (1- 26)	0/1 n=98
Ng, (2016)	NS	NS	49:1	NS	NS	NS	0:27:13	MC (26) SR (2)	10x(1- 27)	49:1:0:0
Shen, (2004)	21:56	14 (3- 85)	74:3	NS	NS	LM n=10 None n=65 NS n=2	NS	MC n=43 (56) SR n=7 (9) NS n=27 (35)	NS	R0/1 n=37 >R1 n=40

Simkins, Razenberg (2016)	n=445 (100) Synch	NA	NS	NS	NS	NS	NS	AC n=246 (55) MC n=156 (35) SR n=43(10)	NS	NS
Simkens, Van Oudheusden (2016)	19:10 rectal 31:27 colon	NS	58:29	Rectal: Colon ≤T3 19:38 ≥T4 9:20	Rectal:col on N0 7:17 N1 9:17 N2 12:24	LM Rectal n=3 Colon n=7	Rectal:col on Good-mod 19:34 Poor 2:8	Rectal: colon MC 9:5 SC 6:3	Rectum 9.3 +/- 4.7 Colon 9 +/- 5.1	R1 (100)
Sluiter, (2016)	35:30	NS	49:5 n=11 RS	T1 n=1 T2 n=2 T3 n=27 T4 n=35	Positive n=46 Negative n=17 NS n=2	NS	3:22:11 NS n=29	AC n=39 MC n=22 SC n=4	*SPCI <2 n=1 2-4 n=42 5 n=14 >5 n=8	R1 n=37 R2a n=22 R2b n=6
Teo, (2015)	NS	15 (1.7-95.9)	NS n=15 (43) Right n=20 (57) Left	T1 n=1 T2 n=0 T3 n=9 T4 n=22	N0 n=10 N1 n=10 N2 n=10	Excluded	NS	AC n=23 MC n=12	12 (1-27)	33:1:1:0
Ung, (2013)	NS	NS	109:16	NS	N+ n=84 N0 n=41	LM n=51 (41)	8:85:32	MC n=56 (45)	1-6 n=43 7-12 n=46 13-19 n=24 >20 n=12	120:5:0:0
Winer, (2014)	21:9	NS	NS	NS	NS	NS	NS	SR n=30(100)	11.5 (1-19)	14:9:2:2 NS n=3

(Synch synchronous, metach metachronous, RS rectosigmoid, RPLN retroperitoneal lymph node, DFI disease-free interval, NS not specified, NA not applicable, LM liver metastasis, PM peritoneal metastasis, AC adenocarcinoma, MC mucinous, SR signet ring, PCI peritoneal carcinomatosis index, sPCI short peritoneal carcinomatosis index, LM liver metastasis, CC completeness of cytoreduction score, R0-2 completeness of resection, CRS cytoreductive surgery, HIPEC heated intraperitoneal chemotherapy, SB small bowel, NSB not small bowel, NS not specified, \bar{x} mean, x median, CI confidence interval, SD standard deviation, P= P-value, n=number followed by value in parenthesis, percentage.

3.3.1.2.2 Rectal or colonic primary

Eight studies reported the influence of a rectal or colonic primary on OS (135, 214, 216, 219, 222, 226, 230, 231). Six studies presented adequate data to be included in the meta-analysis. The pooled unadjusted HR was 1.48 (0.71 – 3.09), $I^2 = 68.5\%$. These data provide no evidence that a rectal primary is a useful predictor of OS.

3.3.1.2.3 Lymph node metastasis

Eleven studies reported on the effect of lymph node status on OS (112, 123, 139, 213-215, 217, 227, 228, 230, 231). Nine studies presented adequate data to be included in the meta-analysis. The pooled unadjusted HR was 1.42 (1.06 – 1.92) $I^2 = 0\%$ and adjusted HR 1.08 (0.6 – 1.95) $I^2 = 68.6\%$. It is unclear therefore whether lymph nodes are a useful predictor of OS.

3.3.1.2.4 Tumour differentiation

Two studies reported the effect of primary tumour differentiation on OS (214, 219). The pooled unadjusted HR was 0.9 (0.18–4.40), $I^2 = 72.4\%$. These data provide no evidence that primary tumour differentiation is a useful predictor of OS.

3.3.1.2.5 Timing of CPM (Synchronous or metachronous)

Three studies reported on the effect of the timing of CPM (synchronous or metachronous) on OS (112, 219, 231). Two presented adequate data to be included in the meta-analysis. The pooled unadjusted HR was 1.13 (0.56 – 2.28), $I^2 = 57.1\%$, indicating that the timing of CPM is not a useful predictor of OS.

3.3.1.2.6 Ascites

Two studies reported on the effect of malignant ascites on OS (123, 219). One paper presented an unadjusted HR (1.50, 95 per cent c.i. 0.70 to 3.21) (219) and one an adjusted HR (3.23, 1.56 to 6.69) (123) meaning it was not possible to provide a pooled effect estimate.

3.3.1.2.7 Hepatic metastasis

Eleven studies included patients with hepatic metastasis (HM) (123, 211, 214-218, 222, 223, 226, 228), two studies did not specify that HM were resected (123, 218) (Table 3.5). Eight studies reported on the effect of surgically treated hepatic metastasis

on OS (123, 126, 211, 214, 216, 217, 222, 223). Seven presented adequate data to be included in the meta-analysis. The pooled unadjusted HR was 1.36 (0.85 – 2.16) $I^2 = 43.6\%$ and the adjusted HR 2.49 (0.74 – 8.41) $I^2 = 85.3\%$. These data provide no evidence that the presence of surgically treated hepatic metastasis are a useful predictor of OS.

3.3.1.2.8 Signet ring histology

Three studies reported on the effect of signet ring histology on OS (112, 221, 225). Two included adequate data to be included in the meta-analysis. The pooled adjusted HR was 1.65 (0.68–4.03), $I^2 = 0\%$ ($p=0.926$) suggesting signet ring histology is not a useful predictor of OS.

3.3.1.2.9 Mucinous histology

Four studies reported in the influence of mucinous histology on OS (112, 139, 222, 225). Three studies presented adequate data to be included in the meta-analysis. The pooled unadjusted HR was 1.10 (0.70–1.72), $I^2 = 0\%$ ($p=0.456$) suggesting mucinous histology is not a useful predictor of OS.

3.3.1.2.10 Gastrointestinal anastomosis

Two studies reported on the effect of one or more gastrointestinal (GI) anastomosis on OS (123, 222). One paper presented an unadjusted HR and one an adjusted meaning it was not possible to provide a pooled effect estimate.

3.3.1.2.11 PCI

Eighteen studies reported on the effect of the extent of CPM as described by the PCI (112, 135, 139, 211-220, 222, 224, 228, 231, 232). PCI was reported as a continuous variable or condensed into categorical variables. Sixteen studies presented adequate data to be included in the meta-analysis.

A PCI of greater than the following levels was predictive of reduced OS: >20 , unadjusted HR 2.9 (1.95–4.31), $I^2 = 0$, adjusted HR 5.04 (2.74– 9.29), $I^2 = 0$, > 15 , unadjusted HR 3.12 (2.02-4.80), $I^2 = 0\%$ ($p=0.515$), adjusted HR 2.09 (1.39 – 3.15), $I^2 = 0\%$ ($p=0.570$). PCI as a continuous variable was predictive of reduced OS: pooled unadjusted HR 1.07 (1.02 –1.12), $I^2 = 46.1\%$, pooled adjusted HR 1.07 (1.04

– 1.11), $I^2 = 67.1\%$. Lower PCI levels were not predictive of OS: PCI > 10, unadjusted pooled HR 2.57 (0.69–9.67), $I^2 = 90\%$. PCI > 6, adjusted pooled HR 3.28 (0.77 – 13.94), $I^2 = 92\%$.

3.3.1.3 Treatment factors

3.3.1.3.1 Prior surgical score

Three studies reported on the influence of the prior surgical score (PSS) on OS (112, 214, 221). The pooled unadjusted HR was 1.23 (0.78 – 1.95) $I^2 = 0\%$. These data provide no evidence that PSS is a useful predictor of OS.

3.3.1.3.2 Post-operative morbidity

Three studies reported the association between postoperative morbidity and OS (219, 222, 228). The pooled unadjusted HR was 1.56 (0.74–3.31) $I^2 = 48.7\%$. These data provide no evidence that postoperative morbidity is a useful predictor of OS.

3.3.1.3.3 Neo-adjuvant and adjuvant chemotherapy

Neo-adjuvant and adjuvant treatment of the primary tumour was poorly reported (Table 3.6). None of the studies reported response to treatment or whether it was completed as planned.

Two studies reported the effect of neoadjuvant chemotherapy prior to CRS & HIPEC on OS (112, 222). The pooled unadjusted HR was 1.00 (0.63 – 1.58), $I^2 = 0\%$ ($p=0.425$). These data provide no evidence that neoadjuvant chemotherapy is a useful predictor of OS.

Four studies reported on the effect of adjuvant chemotherapy, following CRS & HIPEC on OS (112, 215, 222, 228). The pooled unadjusted HR suggested that there is no evidence that the use of adjuvant chemotherapy is predictive of OS 0.6 (0.29–1.21), $I^2 = 68.1\%$.

3.3.1.3.4 Completeness of cytoreduction

Eight studies reported on the effect of the completeness of cytoreduction on OS (112, 212, 215, 219, 221, 222, 229, 232). Seven studies presented adequate data to be included in the meta-analysis. A CC score of >0 and ≥ 2 was predictive of a reduction

in OS, unadjusted HR 1.75 (1.18-2.59), $I^2 = 79.5\%$ and adjusted HR 1.61 (1.31–1.97), $I^2 = 0\%$ respectively.

Table 3-6 Treatment factors

Da Silva, (2006)	Chua, (2010)	Cashin, (2012)	Beniziri, (2012)	Baratti, (2014)	Author, (year)	
					No (%)	Neo- adjuvant therapy
NS	NS	n=47 (37)	n=37 (72.5)	n=87 (86)		
NS	NS	NS	NS	5-FU/CAP n=24, 5-FU + OXA n=12, 5FU/CAP + OXA n=26, 5FU +OXA + bevacizumab or Cetuximab n=11, 5-FU based n=6, Other n3	Type	
n=34 (after 1997) MMCC 10mg/m ² F/12mg/m ² M + 5x days 5-FU 650mg/m ²	MMC 10- 20mg/m ²	HIPEC n=67, MMC 30m/m ² (n=2), OXA 460mg/m ² /+ IV 5-FU 400mg/mg + calcium folinate 60mg/m ² (n=44), OXA 360mg/m ² + IRI 360mg/m ² + IV 5-FU 450- 500mg/m ² + calcium folinate 60mg/m ² (n=23)	MMC 12.5mg/m ² M10mg/m ² F	CIS 25ml/m ² + MMC 3.3mg/m ²	Agent	HIPEC
90	90	90	90	60	Duration (min)	
41-42C	42		42	42.5	Temp (°C)	
n=36 EPIC alone MMC 10mg/m ² F/12mg/m ² M + IV 5-FU 650mg/m ² 23 hrs + 6x days EPIC 5- FUAdjuvant 6/12	n=33 (56+EPIC (Details NS)	SPIC*, 5FU 500-600mg/m ² + IV LV 60mg/m ² daily 7 days, 8x cycles in 6 months	NA	NA		EPIC / SPIC
NS	NS	NA	NS	n=8 liver resection, n=4 RFA liver		Surgery for metastasi
NS	NS	n=27 (21)	NS	n=70 (69)	No (%)	Adjuvant
NS	NS	NS	NS	NS	Type	

Franko, (2008)	Faron, (2016)	Elias, (2014)	Elias, (2013)	Elias, (2010)	Duraj, (2013)
n=64 (98)	2 months	n=139 (100)	n=114 (100)	n=370 (70)	Y:N 10:1 PM/HM
Oxa n=37, Iri n=16, bevacizumab n=21	NS	NS	NS	NS	NS
MMC 40mg	300mg/m ² OXA + 200mg/m ² IRI IV 400mg/m ² 5FU + LV 20mg/m ²	OXA 460-mg/m ² OR OXA 300mg/m ² + IRI 200mg/m ² IV 5-FU 400mg/m ² + LV 20mg/m ²	OXA 460mg/m ² OR OXA 300mg/m ² + IRI 200mg/m ² + IV 5-FU 400mg/m ² / LV 20mg/m ²	MMC 30-50mg/m ² +/- CIS 50-100mg/m ² OR, OXA 360-460mg/m ² +/- IRI 200mg/m ² + IV 5-FU + LV,	n=13 OXA 460mg/m ² IP + 5-FU 400mg/m ² /LV 60mg/m ² IV, n=7 OXA 360mg/m ² + IRI 360mg/m ² + IV 5-FU 400mg/m ² /LV 60mg/m ²
100	30	30	30	60-120	30
NS	42-43	42	43	41	41-42
NA	NA	NA	NA	MMC 10mg/m2 D1, 5-FU 600mg/m2 days 2-5	n=8 + EPIC, 5-FU 550mg/m ² IP + LV 60mg/m2, n=12 +SPIC 5-FU 500-600mg/m2 + LV 60mg/m2 6 days, 8x cycles over 6 months
NS	30.2% liver resection	Synchronous liver n=37	Synchronous liver n=24	Synchronous liver resection n=77	LM n=2 major n=12 Segmental
NS	6 months total including neo-adjuvant	84%	84%	n=232 (47%) objective response to neo-adjuvant OR poor prognostic factors CC1-2, LN+, liver mets	n=4(36) PM/HM, n=7(12) PM
NS	NS	NS	NS	NS	NS

Maggioli, (2013)	Lorimer, (2017)	Ihemelandu, (2017)	Huang, (2016)	Huang, (2014)	Froysnes, (2016)
2-3/12 pre op – no progression on tx	2-3 months with response	Recommend 3-4 cycles	NS	cycles 8x (2-18), <6 n=15, >6 n=45	n=81 (68)
NS	NS	5FU + Oxalaplatin +/- bevacizumab	NS	NS	NS
460mg/m ² OXA +IV 5-FU 400mg/m ² + LV 20mg/m ²	2002 – OXA 460mg/m ² <2003 – MMC 30mg/m ² + CIS 100mg/m ²	<1995 – MMC 15mg / m ² >1995 – MMC 15mg / m ² >2010 MMC + DOX 15mg/m ²	OXA 350mg/m ² OR MMC 12.5mg/m ²	120mg CIS +30mg MMC	MMC 35mg/m ²
30	30	90	30	90	90
43-45	42-45	Cold	42	43+/-0.5C	41.4
MMC+5FU	NA	5-FU 400 (F), 600 (M) for 5 days post-op for C1/2 patients, After 1995 + LV 20mg/m ² , (Not if massive cytoeduction, or long term systemic chemo before CRS	NS	Once postoperative physical condition recovered' DOC 75mg/m ² + CBDCA day 1 every 3 weeks With adjuvant or alternate cycles Median 4 cycles (1-10)	n=16 additional HIPEC usually within 1x week CRS
Minor liver n=25, Major liver n=12 , +/- RFA n7	Minor liver n=17, Major liver n=5	NS	NA	NS	NA
PM n=48, PM+LM n=30	PM n=20 (90.9), PM+LM n=30 (80.3)	NS	NS	FOLFOX or FOLFIRI	Not routinely given
NS	NS	NS	NS	NA	NA

Winer, (2014)	Ung, (2013)	Teo, (2015)	Sluiter, (2016)	Simkens, Van Oudheusden (2016)	Simkins, Razenberg, (2016)	Shen, (2004)	Ng, (2016)
n=25 (83)	n=76 (61)	NS	n=17 (26)	n=14 (48.3) rectal, n=7 (12.1) colon	NS	n=58 (75)	NS
NS	NS	NS	NS	NS	NS	NS	NS
MMC 40mg	MMC 12.5 mg/m ²	MMC dose NS	MMC C 35mg/m ²	MMC C 25mg/m ² OXA 460mg/m ² + IV 4FU, LV 20mg/m ²	MMC C 35mg/m ²	MMC C 40mg *dose reduction 30mg – elderly, extensive prior chemo, poor performance status extensive peritonectomy	MMC OR OXA + LV + 5-FU (dose NS)
40	NS	60	90	90	90	120	60-90
42	42	42	39-41	NS	41-42	39.5-43	39-43
NA	5-FU 5 days 650mg/m ²	AI + EPIC 5-FU 5 days, dose NS	NA	NS	NS	NA	NA
NS	n=51 synchronous	NA (excluded)	NS	n=5 pre CRS&IPEC, n=4 synchronous, *1 not resected	NS	n=10 synchronous liver	NS
NS	n=99 (79)	'most'	n=41 (63)	n=21 (72) rectal, n=40 (69) colon	NS	some	n=11 yes, n=26 no, n=13 NK
NS	NS	NA	NA	NS	NS	NS	NS

(MMC Mitomycin C, OXA Oxalaplatin, LV Leucovorin, 5-FU 5 Fluorouracil, CIS Cisplatin, DOX Doxorubicin, DOC docetaxel, CBDCA carboplatin, IRI Irinotecan, CAP Capecitabine, mg milligrams, m² metres squared, Min minute, C centigrade, EPIC Early postoperative intraperitoneal chemotherapy, SPIC sequential postoperative intraperitoneal chemotherapy, CRS cytoreductive surgery, HIPEC heated

intraperitoneal chemotherapy, NA not applicable, NS not specified, deviation, n=number followed by value in parenthesis, percentage)

3.4 Discussion

This systematic review and meta-analysis examined the effect of prognostic factors on OS following CRS & HIPEC for CPM. Emergency presentation with obstruction or perforation of the primary tumour as well as the extent of CPM and the completeness of resection, as described by the PCI and the CC score respectively, were the only significant prognostic factors.

An emergency presentation of the primary tumour with obstruction or perforation was predictive of reduced OS (for both synchronous and metachronous CPM). Several factors may contribute to this. In the primary setting, colorectal cancers presenting with obstruction or perforation are associated with decreased cancer-specific survival and increased postoperative mortality (233). Obstructed or perforated colorectal cancer is, by definition, advanced in stage and increases the risk of metastasis; additionally, emergency presentation limits the possibility of neoadjuvant treatment and may delay adjuvant treatment owing to postoperative morbidity. The extent of peritoneal metastasis as described by the PCI was predictive of reduced OS as a continuous variable and when the PCI score was 15 or above. A complete cytoreduction (CC0) was predictive of improved OS. Improving patient selection is therefore reliant on the ability to predict accurately the extent of peritoneal metastasis and the ability to resect it completely. Specialist radiologists have demonstrated good concordance between radiological and surgical PCI estimations, particularly when combining modalities; however, these tend to be most accurate in patients with high PCI scores (234, 235). In some centers, this is used in combination with diagnostic laparoscopy before CRS & HIPEC. This is feasible in the majority of patients and may help to reduce the laparotomy rate in patients for whom CRS & HIPEC may not be possible (236).

Included patients were relatively young at 54 (range 45–69) years compared with the incident age of colorectal cancer (80-90 years) (1). In addition, performance status was not reported by the majority of studies in the review and limited to an ECOG <2 by a further nine. Within these limits, no other patient or tumour factor was predictive of OS.

Details of neoadjuvant and adjuvant treatments were reported poorly in the included studies. Within these limits, the use of neoadjuvant or adjuvant chemotherapy was not predictive of OS after CRS & HIPEC. One meta-analysis, by Kwakman and colleagues from 2016 (237), examined the effect of clinicopathological variables only on OS following CRS & HIPEC. Significant prognostic factors identified in the present review (adverse features of the primary tumour, PCI, CC score) were not comparable with those from Kwakman *et al* as they were not examined (237). In contrast to the present study, Kwakman and co-workers found performance status, the presence of lymph node or hepatic metastasis, tumour differentiation, signet ring histology, a rectal primary and the use of neoadjuvant chemotherapy to be predictive prognostic factors (237). A number of differences may explain this variation in findings: the exclusion of 73 papers described as unavailable in full text may introduce a potential selection bias (237). In addition, a number of studies included by Kwakman and colleagues were excluded in the present study for the following reasons: presentation of inadequate data to estimate the HR accurately (118, 126, 231, 238), the inclusion of mixed primary tumours (227, 239, 240) and the inclusion of patients having repeat CRS & HIPEC procedures. Finally, the meta-analysis combined all studies regardless of the presentation of unadjusted or adjusted HR (237). The strengths of the present study include the comprehensive and systematic literature search including 3128 patients with CPM undergoing CRS & HIPEC, all potential prognostic factors were included, and a consistent association was found between predictive prognostic factors across different studies. The low heterogeneity associated with these factors adds to the strength and generalizability of the findings. The application of strict inclusion criteria limits the potential impact of factors such as mixed primary tumour origin. The present analysis takes into account the adjustment factors used in primary studies to ensure meta-analysis of time to event data was performed only when data were comparable.

Some limitations must, however, be acknowledged. This review is limited by the quality of primary studies and the heterogeneity of the population. Prognostic factor systematic reviews, by their nature, represent one of the most difficult categories due to their retrospective and observational nature. As CRS & HIPEC is performed at tertiary centres, data concerning the primary tumour and its treatment may not have been captured or reported fully. Additionally, the statistical analysis and presentation

of data necessary for accurate meta-analysis varied widely across primary studies, which may introduce a degree of inclusion bias.

There was a lack of molecular and genetic data concerning patients with CPM undergoing CRS & HIPEC in comparison with studies of primary colorectal cancer, and this is an area for future research. Discordance between primary genetic mutations and those in peritoneal metastases may identify novel therapeutic targets and prognostic markers in this metastatic group. Recent research by Schneider and colleagues found that *RAS/RAF* mutations impair survival after CRS/HIPEC (241), although this is the only study that has considered this prognostic factor and validation is required. Further large-scale research is needed to account for these factors; this would require collaboration between CRS & HIPEC centres, standardization of the analysis and presentation of prognostic factor time to event data.

Chapter 4 : Describing the biology of colorectal peritoneal malignancy - an investigation of the gene expression, methylation and somatic mutation alterations associated with the development of colorectal peritoneal malignancy and with poor prognosis.

Some of the work in this chapter has been previously published in Hallam *et al* 2020 (242). In this publication I was responsible for the writing of the text, collection of the data and also production any of the figures used in this thesis.

4.1 Introduction

In the preceding chapters, we established the known clinical prognostic factors for patients with CPM. Obstruction or perforation of the primary tumour, the extent of PM as described by the PCI score and the completeness of cytoreduction as described by the CC score. Despite optimal selection for CRS & HIPEC using these known prognostic factors, there is still significant variation in survival for these patients. Little is known about the biology of isolated CPM. Understanding tumour biology may identify which patients with primary CRC are at risk of developing CPM, and which are suitable for treatment with CRS & HIPEC or alternative existing or novel treatment strategies.

Gene expression profiling studies in primary CRC have identified genetic signatures associated with the development of lymph node metastasis (243), systemic metastasis (243), and predicting prognosis (243). One small study of CPM identified a gene expression profile predictive of reduced OS following CRS & HIPEC (244). This study, however, combined small numbers of CPM with a larger cohort of appendix adenocarcinoma, a biologically distinct tumour with a significantly improved prognosis.

Methylation profiling studies in primary CRC have identified changes associated with the development of metastasis and predicting prognosis. The methylation of markers of CIMP appear to be stable between matched primary CRC and hepatic metastasis suggesting the global hypermethylation of CIMP is established before the development of metastasis (245, 246). In support of this Murata *et al* examined the methylation levels of long interspersed nuclear elements (LINEs). LINEs comprise 17% of the genome and act as a surrogate marker of global DNA methylation (247). They found equivalent LINE-1 methylation in primary CRC and matched liver metastasis (247). Zhang *et al* found CIMP was associated with an unfavourable response to chemotherapy and a reduced response to anti-EGFR antibody therapy amongst *KRAS* wild type tumours

(248). This may relate to a higher mutation frequency of *EGFR* related genes such as *BRAF*, *NRAS* and *PIK3CA*, (74.1%) (248). Given the 70-80% correlation between CIMP and MSI high tumours, it is difficult to know whether the reduced chemotherapy response is a reflection of MSI high or CIMP status (18, 44). A number of studies examined the methylation status of individual genes of known importance in CRC and found hypermethylation of genes including *KRAS*, the *miR-34* and *CDKN2A* tumour suppressors and hypomethylation of the *MYC* oncogene in metastatic CRC which was associated with reduced OS (249-252). Hypermethylation of Wnt modulators and the resulting Wnt pathway dysregulation are associated with recurrence in primary CRC and reduced OS (252, 253).

Copy number alterations result in the activation of oncogene or deactivation of proto-oncogenes. To our knowledge, no group have analysed the difference in CNA primary CRC vs. CPM. Lee et al compared CNA in primary CRC and matched liver metastasis (254). Forty-seven per cent of patients showed significant differences in CNA between primary and metastasis (254).

The sequentially accumulated somatic mutations driving the development of primary CRC tumorigenesis have been described in detail in the literature. The mutations associated with the development of metastasis are less clearly defined. Whilst a number of studies have examined the association between mutations, (known CRC drivers or potential) and the development of metastasis their results have been contradictory. Goryca et al performed WES exome sequencing in 7 primary CRC and matched liver metastasis. Using the Cancer-Specific High-throughput Annotation of Somatic Mutations (CHASM) method, they identified potential drivers which were exclusive to metastatic sites, (*BAHD1*, *PLXND1*) (255). Mlecnik et al examined primary CRC samples from 212 patients, (with and without distant metastasis) for a panel of 48 commonly mutated CRC genes and found similar patterns of genomic alterations in both groups (256). Huang et al performed a meta-analysis to investigate the association between key CRC driver mutations and the development of metastasis (257). They found *KRAS*, *p53*, *SMAD4* and *BRAF* mutations were associated with lymph node and distant metastasis (257). In all studies samples were taken from retrospective primary CRC specimens, it may be that later mutations preceding the

development of metastasis were not captured, additionally, the site of metastasis was not specified and included lymph node which is an earlier stage of metastasis than distant CPM.

Few studies have examined the somatic mutations found in CPM and their prognostic implications. These studies, however, are limited to individual or small panels of mutations routinely tested for in clinical practice, there is limited evidence to suggest which genes should be included in panel sequencing in CPM. Schneider et al examined the *KRAS* and *BRAF* mutation status of patients with CPM who underwent CRS & HIPEC (241). They found mutations of *RAS/RAF* were associated with reduced OS and that this association was independent of the use of targeted anti-EGFR treatment (241). Sasaki et al examined the *KRAS*, *BRAF* and *PIKCA* mutation status of 526 patients with metastatic CRC, with or without CPM (150). They found the incidence of *BRAF* mutation was significantly associated with the presence of CPM but not with prognosis (150).

There is a need to improve the outcomes for patients with CPM and significant variation in survival despite optimal patient selection for treatment using currently known prognostic factors. There is a paucity of knowledge concerning CPM tumour biology. Understanding the transcriptome, epigenome and genome of CPM will form the basis to identify which patients with primary CRC are at risk of developing CPM, those suitable for treatment with CRS & HIPEC or alternative existing or novel treatment strategies.

Aim: to determine the gene expression methylation and somatic mutation profile associated with the development of CPM and conferring 'poor prognosis' in patients with CPM following CRS & HIPEC.

4.2 Results

4.2.1 Retrospective patient cohort

From 2011-2017 a total of 161 patients underwent CRS & HIPEC at University Hospitals Birmingham, 88 patients for metachronous CPM.

Patients were excluded for the following reasons: rectal primary, 5, incomplete CC2 resection, 8, PCI of ≥ 12 20, follow up ≤ 12 months 27 leaving 28 patients. Complete information regarding the primary CRC pathology and treatment was available for 26 patients who form the basis of this study. Each patient had matched normal, primary CRC and CPM samples.

Thirteen patients had a DFS of 24 months (15–72 range) following CRS & HIPEC and formed the 'good prognosis' cohort, thirteen patients had a DFS of 6 months (2-11 range) and formed the poor prognosis cohort (Figure 4.1). There were no significant differences in patient demographics, primary CRC or CPM tumour, treatment or follow up between cohorts, (table 4.1).

Following nucleic acid extraction all, 26 patients had adequate CPM RNA for RNAseq (13 good, 13 poor prognosis), 25 had matched primary CRC samples. For methylation array 24 patients (12 good, 12 poor prognosis) had adequate DNA. As the Infinium methylation array comes in a 32-prep kit, 4 good and 4 poor prognosis primary tumours were matched to these. For exome sequencing 24 patients (12 good, 12 poor prognosis) had adequate DNA from both the primary and CPM samples, extraction of DNA from normal tissue resulted in 18 samples (9 good, 12 poor prognosis).

Table 4-1 Comparison of good and poor prognosis patient cohorts

		Good Prognosis	Poor prognosis	P-value
Age, mean +/- SD		58 +/-13	58 +/-9	0.97
Gender, male		n=7 (54)	n=7 (54)	0.68
Tumour location	Right	n=9 (69)	n=6 (46)	0.33
	Transverse	n=1 (8)	n=0 (0)	
	Left	n=3 (23)	n=7 (54)	
T stage primary	3	n=3 (23)	n=3 (23)	0.66
	4a	n=5 (38.5)	n=7 (54)	
	4b	n=5 (38.5)	n=3 (23)	
N stage primary	0	n=4 (31)	n=1 (8)	0.86
	1	n=7 (54)	n=5 (38)	
	2	n=2 (15)	n=7 (54)	
DFI months, mean +/- SD		25 +/-9	24 +/-12	0.83
PCI score, median (range)		5 (3-12)	8 (2-12)	0.019
CC score	CC0	n=13 (100)	n=13 (100)	1
	CC1	n=0 (0)	n=0 (0)	
	CC2	n=0 (0)	n=0 (0)	
Follow up, months, median (range)		29 (19-72)	16 (5-55)	0.11
Adjuvant treatment	Yes	n=11 (85)	n=12 (92)	0.38
	No	n=2 (15)	n=1 (8)	
DFS, median (range)		24 (15-72)	6 (2-11)	<0.0001
OS, median (range)		29 (19-72)	16 (5-55)	0.12

(n= number, value in parenthesis, percentage, DFI, disease free interval, PCI, peritoneal carcinomatosis index, CC score, completeness of cytoreduction, DFS, disease free survival, OS, overall survival, SD standard deviation)

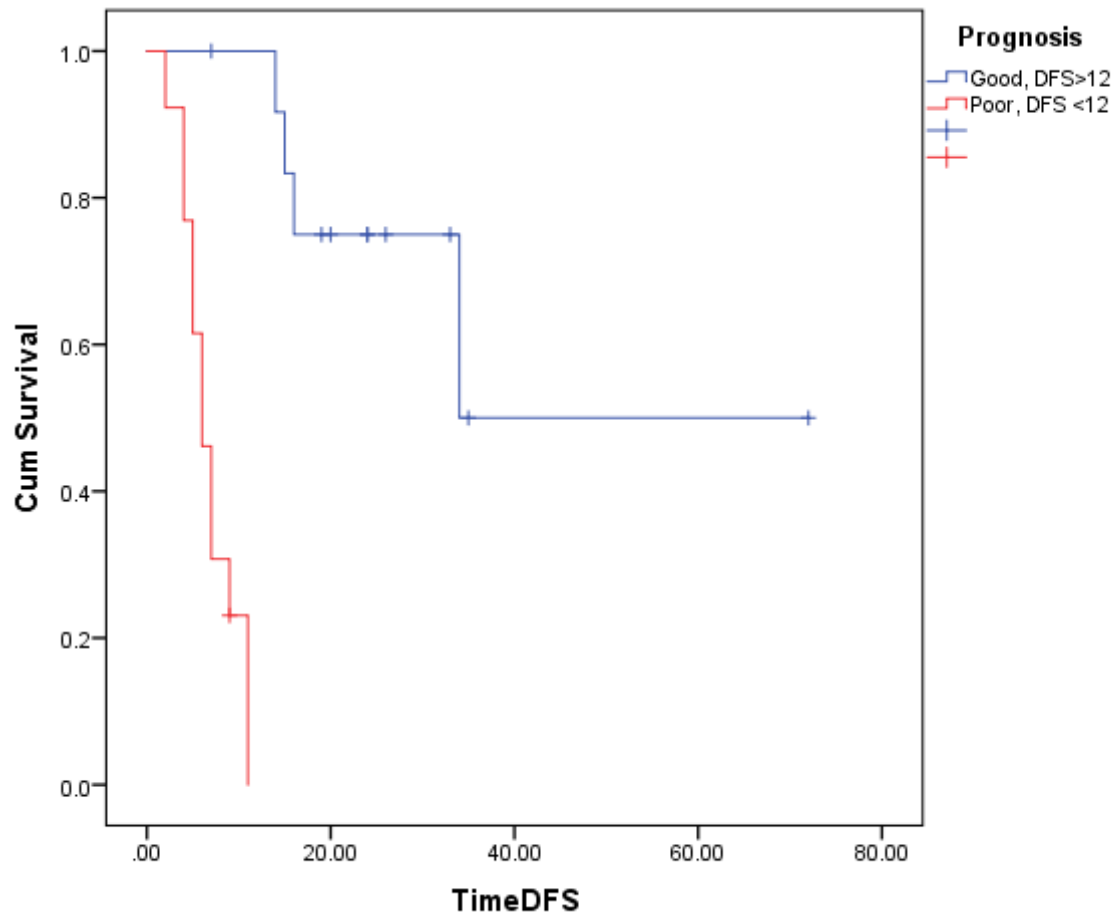


Figure 4-1 Kaplan-Meier disease-free survival analysis, good vs. poor prognosis patient cohorts

Kaplan-Meier curve showing disease free survival in patients in the retrospective patient cohort with good and poor prognosis following CRS & HIPEC for CPM. Log-rank $p < 0.0001$. This figure was taken from my own results and produced in SPSS.

4.2.2 RNA sequencing, differential gene expression

4.2.2.1 Differential gene expression, primary CRC vs. matched metachronous CPM

A Kyoto Encyclopaedia of Genes and Genomes, (KEGG) pathway analysis was performed to identify the enriched biological functions among the differentially expressed genes. Differentially expressed genes participated in a variety of functions. Pathways identified are detailed in table 4.2. The most enriched pathways include those involved in translation, cell motility, cancer-specific pathways and immune pathways.

Table 4-2 KEGG pathway enrichment analysis, primary CRC vs. matched metachronous CPM

Gene set	Class	KEGG pathway	Enrichment score	P-value
hsa04810	Cellular Processes; Cell motility	Regulation of actin cytoskeleton	4.19	0.02
hsa05206	Human Diseases; Cancers: Overview	MicroRNAs in cancer	3.44	0.03
hsa05222	'Human Diseases; Cancers: Specific types	Small cell lung cancer	3.43	0.03
hsa04672	Organismal Systems; Immune system	Intestinal immune network for IgA production	3.30	0.04
hsa04670	Organismal Systems; Immune system	Leukocyte transendothelial migration	3.22	0.04

P = <0.05

Primary CRC and matched CPM showed differential expression of 65 genes with an FDR <0.1 (Figure 4.2 and Figure 4.3). Sixteen genes showed decreased expression in CPM compared with primary CRC. The full list can be seen in appendices, table S3. The 10 genes with the highest fold change of known significance are discussed below (Table 4.3).

In the literature *FABP6*, *DEFA6*, *DMBT1*, *OLFM4* and *MUC2* are highly expressed in adenoma or early CRC with reduced expression in metastasis suggesting a role in early tumorigenesis which is not required for metastasis. The expression of *FABP6*, an intercellular bile acid transporter, was decreased 34.30-fold in CPM. Elevated bile acid concentrations associated with high-fat diets are detrimental to the colonic epithelial architecture through mechanisms of oxidative DNA damage, inflammation and increased cell proliferation (258). Omachi et al found high expression of *FABP6* in primary CRC compared with low levels in lymph node metastasis (259).

DEFA6 expression, part of a cytotoxic peptide family involved in host defence of the bowel and a downstream Wnt target, was decreased 8.15-fold in CPM. Radeva et al found high expression of *DEFA6* in colorectal adenomas compared to CRC (260). Decreased *DEFA6* expression in CPM suggests the development of Wnt pathway dysregulation and immune system evasion in the transition from primary to CPM (261). *DMBT1* expression was decreased 6.06-fold in CPM. *DMBT1* controls the interaction between tumour cells and the immune system as well as epithelial differentiation (261). No studies have described *DMBT1* expression in colorectal metastasis, however, Du et al demonstrated high expression in benign prostatic tissue, weak expression in early primary prostatic cancer and no expression in advanced prostatic cancer. Given the role of *DMBT1* in immune surveillance of tumour cells its downregulation may reflect tumour cell immune evasion in metastasis (262). *IGHA1* is an immune receptor (261), its expression was reduced 3.66-fold in CPM. Fehlker et al found reduced expression of *IGHA1* along with other immune response genes in metastatic CRC (263). Taken together decreased expression of these immune regulators suggests an immune evasive phenotype in CPM.

OLFM4 is a target of the Wnt/ β -catenin pathway and marker of colorectal stem cells (264), its expression was reduced 3.77-fold in CPM. *OLFM4* downregulates β -catenin signalling by binding to Frizzled receptors in place of Wnt ligands, *OLFM4* deletion promotes colon tumorigenesis (265). High expression of *OLFM4* is associated with early tumour stage, and improved survival and is downregulated in poorly differentiated, advanced primary tumours and metastasis resulting in unchecked Wnt pathway activation (265, 266).

Mucin 2, (*MUC2*) encodes a mucinous coating which forms a protective intestinal barrier, expression was reduced 2.34-fold in CPM (261). Reduced expression of *MUC2* is associated with progression and metastasis of CRC (267). *In vitro* *MUC2* suppresses the migration of CRC cells, its downregulation results in upregulation of interleukins and E-cadherins promoting EMT, invasion and metastasis (267).

The expression of *CES2* was reduced 3.2-fold in CPM, *CES2* encodes an intestinal enzyme involved in the intestinal clearance of drugs and activation of ester-containing prodrugs such as capecitabine and irinotecan, commonly used chemotherapy agents in CRC (261). The decreased expression may suggest resistance in this cohort of patients, Shaojun et al demonstrated the expression of *CES2* predicts Irinotecan sensitivity in metastatic CRC (268).

Table 4-3 The top 10 genes with significantly reduced expression (FDR<0.1) in CRS samples compared with primary CRC samples

Rank	Gene name	Fold change	FDR <i>P</i>-value
1	<i>FABP6</i>	-34.30	1.74 x 10 ⁻⁰⁶
2	<i>DEFA6</i>	-8.15	8.55 x 10 ⁻⁰⁶
3	<i>DMBT1</i>	-6.06	2.43 x 10 ⁻⁰⁴
4	<i>TTC38</i>	-4.56	5.80 x 10 ⁻⁰⁵
5	<i>OLFM4</i>	-3.77	1.01 x 10 ⁻⁰⁴
6	<i>IGHA1</i>	-3.66	4.23 x 10 ⁻⁰⁵
7	<i>CES2</i>	-3.20	6.84 x 10 ⁻⁰⁵
8	<i>NDUFS6</i>	-2.70	7.74 x 10 ⁻⁰⁵
9	<i>P2RY11</i>	-2.53	6.37 x 10 ⁻⁰⁴
10	<i>MUC2</i>	-2.34	7.22 x 10 ⁻⁰⁴

Forty-nine genes showed increased expression in CPM compared with primary CRC. The full list can be seen in appendices, table S4. The 10 genes with the highest fold change of known significance are discussed below (Table 4.4).

CD53 a tetraspanin and *TSC22D3* an anti-inflammatory protein glucocorticoid (GC)-induced leucine zipper both have a mechanistic role in cell survival and the inhibition of apoptosis, a cellular process essential for metastasis (261, 269, 270). Their expression was increased 7.29 and 3.36-fold respectively suggesting downregulation of apoptotic pathways in CPM.

Increased expression of *CYR61*, *CXCL12*, *CTGF* and *CSTB* suggest an invasive, epithelial-mesenchymal phenotype. *CYR61* promotes migration, invasion a loss of polarity and adhesion (261, 271), increased expression correlates with EMT, metastasis and poor prognosis (271, 272). *CXCL12* triggers proliferation, migration, angiogenesis, invasion and metastasis *in vitro* (273) and correlates with distant recurrence and poor OS (274). *CTGF* encodes a matricellular protein with a role in fibrosis, inflammation and connective tissue remodelling, high expression is noted in the CMS4 subtype suggesting a role in EMT (261, 275). *CSTB* encodes a cystatin which participates in the dissolution and remodelling of connective tissue and basement membrane in tumour invasion and metastasis, high expression is found in metastatic CRC and correlates with decreased OS (276). Their expression of *CYR61*, *CXCL12*, *CTGF* and *CSTB* were increased 4.24, 3.64 and 3.49 and 3.41-fold respectively suggesting EMT in the development of CPM.

NR2F1 expression was increased 3.53-fold in CPM (261). *NR2F1* is upregulated in dormant cancer cells, blocking *NR2F1 in vivo* results in a rapid switch from dormancy to proliferation and systemic recurrence (277). Cancer cell dormancy may be an important mechanism in the recurrence of CPM from dormant cell populations following CRS & HIPEC.

DCN can initiate several signalling pathways including *ERK* and *EGFR* leading to growth suppression, its expression was increased 3.3-fold in CPM, this was unexpected and in contrast to the literature (278). *In vitro* reduced *DCN* expression stimulates tumorigenesis, growth and angiogenesis, increased expression reduces cell

motility, proliferation and invasion (278). Clinically reduced expression of *DCN* is associated with the development of metastasis in both breast and lung cancer (278). Phosphatase And Tensin Homolog (*PTEN*) is a tumour suppressor gene capable of inactivating the PI3K-dependent signalling pathway (261). *PTEN* expression was increased 3.25-fold in CPM, this was unexpected and in contrast to the literature. There is a gradual decline in the levels of *PTEN* in the sequence of transition from normal to polyp to carcinoma to distant metastasis. *PTEN* levels are inversely correlated with clinical severity (TNM stage, reduced survival) in CRC (279).

NF- κ B Inhibitor Alpha, (*NF- κ BIA*) encodes a protein which inhibits the *NF- κ B* transcription factor by masking nuclear localisation signals keeping *NF- κ B* sequestered in the cytoplasm and blocking transcription factor binding (261). *NF- κ BIA* expression was increased 3.24-fold in CPM, its upregulation may reflect increased *NF- κ B* activity in the development of CPM. *NF- κ B* activity in cancer cells can be thought of as a malignant reflection of its normal behaviour in protecting the organism from danger, promoting cell survival, proliferation and metastasis (280).

Table 4-4 The top 10 genes with significantly increased expression (FDR<0.1 in CPM samples compared with primary CRC samples

Rank	Gene ID	Fold change	FDR <i>P</i>-value
1	CD53	7.29	5.87 x 10 ⁻⁰⁵
2	CYR61	4.24	3.12 x 10 ⁻⁰⁴
3	CXCL12	3.64	9.25 x 10 ⁻⁰⁴
4	NR2F1	3.53	7.09 x 10 ⁻⁰⁴
5	CTGF	3.49	1.55 x 10 ⁻⁰⁴
6	CSTB	3.41	6.13 x 10 ⁻⁰⁴
7	TSC22D3	3.36	3.94 x 10 ⁻⁰⁴
8	DCN	3.30	6.19 x 10 ⁻⁰⁵
9	PTEN	3.25	9.28 x 10 ⁻⁰⁴
10	NFKBIA	3.24	1.06 x 10 ⁻⁰⁴

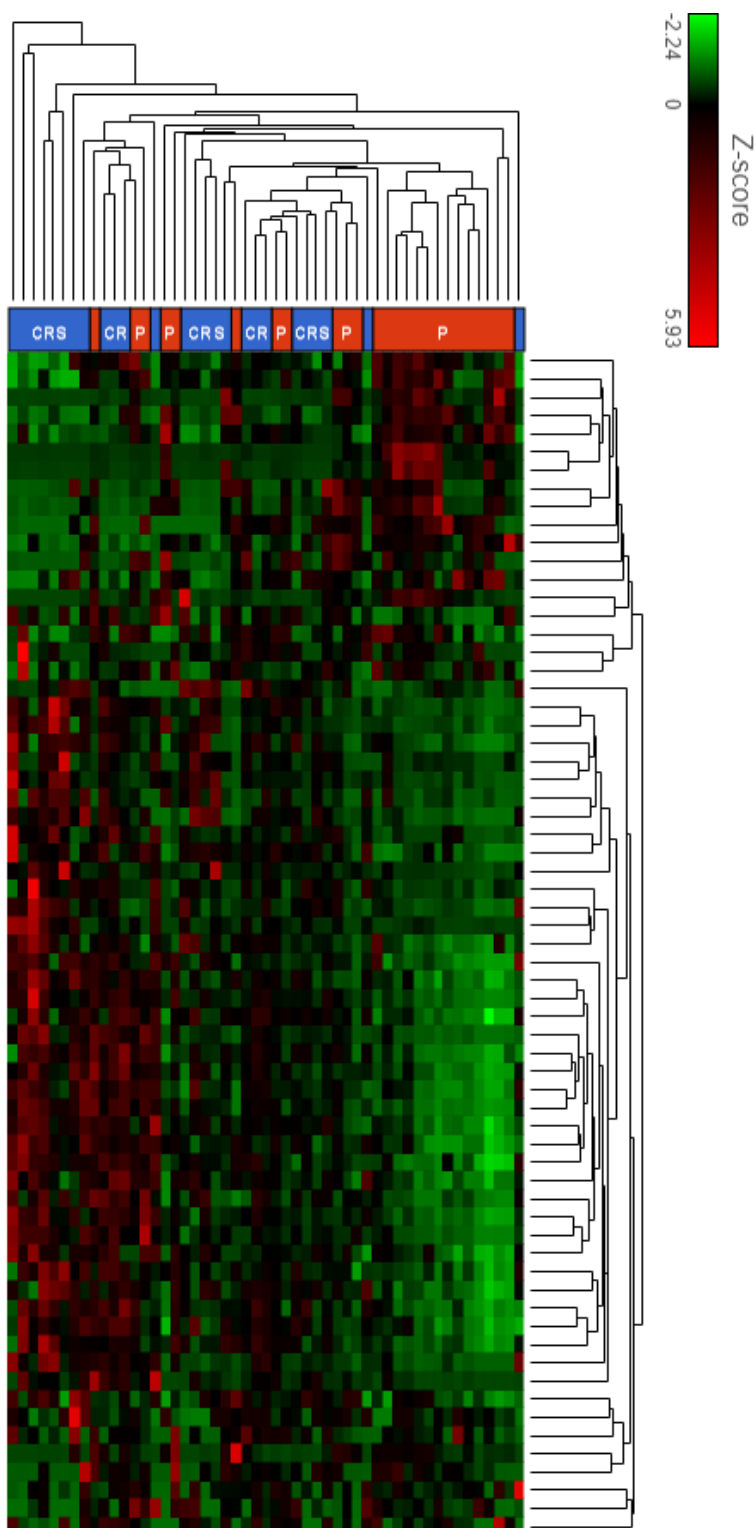


Figure 4-2 Heatmap, differential gene expression CPM vs. primary samples

Heatmap depicting differential genes expression in CPM and matched primary CRC. Sample type is indicated at the upper border of the heatmap, CRS blue = CPM, P red = primary. This figure is reprinted from Hallam et al, 2020, figure 2 (242).

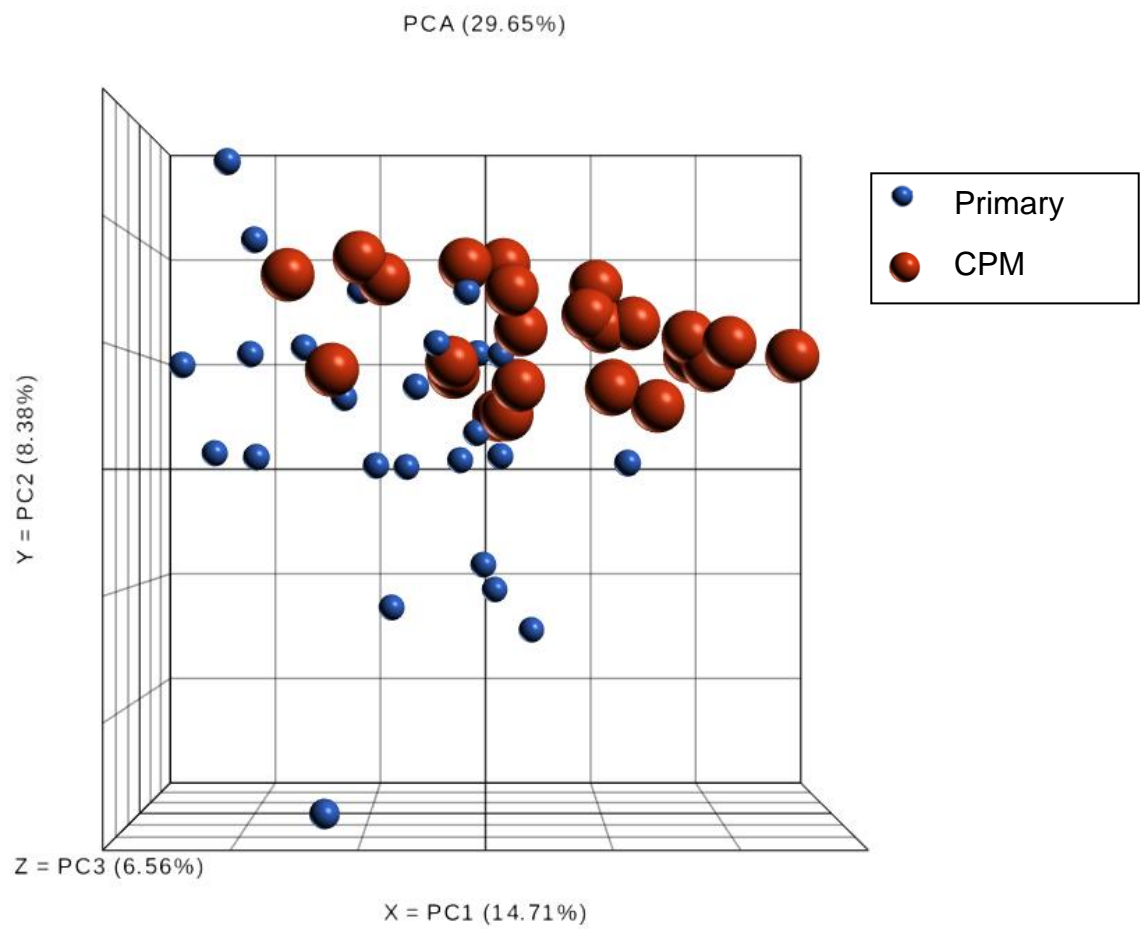


Figure 4-3 Principal component analysis, differential gene expression CPM vs. primary samples

Principle component analysis depicting differential genes expression in CPM and matched primary CRC.

4.2.2.2 Differential gene expression profile Poor vs. Good prognosis CPM

KEGG pathway analysis was performed to identify the enriched biological functions among the differentially expressed genes. Differentially expressed genes participated in a variety of functions. Pathways identified are detailed in table 4.5. The most enriched pathways include those involved in endocytosis, metabolism, phagocytosis, cell movement and architecture, bacterial and viral cell infection, transcription and the expression of genes controlling apoptosis, cell cycle, oxidative stress resistance and longevity.

Table 4-5 KEGG pathway enrichment analysis, poor vs. good prognosis CPM, $p < 0.05$

Gene set	Class	KEGG pathway	Enrichment score	P-value
has04144	Cellular Processes; Transport and catabolism	Endocytosis	11.67	8.56 x10-6
hsa00130	Metabolism; Metabolism of cofactors and vitamins	Ubiquinone and other terpenoid-quinone biosynthesis	5.07	6.28 x10-3
hsa05100	Human Diseases; Infectious diseases: Bacterial	Bacterial invasion of epithelial cells	4.76	8.56 x10-3
hsa05110	Human Diseases; Infectious diseases: Bacterial	Vibrio cholerae infection	4.14	0.02
hsa04145	Cellular Processes; Transport and catabolism	Phagosome	3.92	0.02
hsa05165	Human Diseases; Infectious diseases: Viral	Human papillomavirus infection	3.85	0.02
hsa00561	Metabolism; Lipid metabolism	Glycerolipid metabolism	3.64	0.03
hsa04966	Organismal Systems; Excretory system	Collecting duct acid secretion	3.63	0.03
hsa00230	Metabolism; Nucleotide metabolism	Purine metabolism	3.37	0.03
hsa05120	Human Diseases; Infectious diseases: Bacterial	Epithelial cell signalling in Helicobacter pylori infection	3.31	0.04
hsa00020	Metabolism; Carbohydrate metabolism	Citrate cycle (TCA cycle)	3.20	0.04
hsa01230		Biosynthesis of amino acids	3.20	0.04
hsa04520	Cellular Processes; Cellular community – eukaryotes	Adherens junction	3.13	0.04
hsa03020	Genetic Information Processing; Transcription	RNA polymerase	3.08	0.05
hsa04068	Environmental Information	FoxO signalling pathway	3.00	0.05

One hundred and fifty-nine genes showed increased expression in poor prognosis CPM compared with good prognosis (Figure 4.6 and Figure 4.7). The full list can be seen in appendices, table S5. The 10 genes with the highest fold change of known significance are discussed below, (table 4.6).

The expression of *CEACAM1*, a member of the carcinoembryonic antigen (*CEA*) immunoglobulin family was increased 8.27-fold in poor prognosis CPM (261). Cellular activities of *CEACAM1* include adhesion, the inhibition of differentiation, apoptosis, disruption of tissue architecture and the inhibition of NK cell immune-mediated death (281). *CEACAM1* expression correlates with metastasis and reduced survival in CRC (281). CEA levels are not routinely measured in patients with CPM, and it would be interesting to see whether this is reflected in serum CEA levels of patients with colorectal cancer. Novel therapies in the form of CEA TCB IgG-based T-cell bispecific antibodies result in immune-mediated tumour lysis in tumours overexpressing CEA, a potential therapeutic target in patients with CPM given the high expression of *CEACAM1* (282).

Genes regulating cell survival and inhibiting apoptosis were upregulated in poor prognosis CPM suggesting rapid and unchecked tumour cell growth. *BCYRN1*, a small non-messenger RNA and *RELB* Proto-Oncogene, (*RELB*) promote cell proliferation, survival, inhibit apoptosis and correlate with tumour stage and size (280, 283). Their expression was increased 7.96 and 6.17-fold respectively in poor prognosis CPM. *DCAF12* is an apoptosis regulator (261), expression was increased 6.17-fold in poor prognosis CPM (284).

Genes regulating transcription, cell growth and angiogenesis were upregulated in poor prognosis CPM suggesting a rapid growth. *VEZF1* is an endothelial cell-specific transcription factor involved in angiogenesis and lymph-angiogenesis (261, 285). *VEZF1* expression was increased 10.4-fold in poor prognosis CPM suggesting a rapid rate of angiogenesis, lymph-angiogenesis in keeping with metastatic growth in the poor prognosis cohort. *POLR3GL* encodes an RNA polymerase (261), its expression was increased 6.05-fold in poor prognosis CPM suggesting an increased rate of transcriptional activity. *FEM1C* expression was increased 5.69-fold in poor prognosis CPM. *FEM1C* proteins regulate stem-loop binding proteins during the progression of

the cell cycle from G2 – S phase, perhaps indicating an increased rate of cell growth (286). *KHDRBS3* activities include pre-mRNA splicing and cell cycle control. *KHDRBS3* expression was increased 5.42-fold in poor prognosis CPM. Its function has only been studied *in vitro* where it is downregulated as cancer cell lines become immortalised.

AXIN1 encodes a cytoplasmic protein which forms the β -Catenin destruction complex, a negative regulator of the WNT signalling pathway (23). *AXIN1* expression was increased 5.42-fold in poor prognosis CPM which suggests increased activity of the Wnt signalling pathway (287).

Table 4-6 The top 10 genes with significantly increased expression (FDR<0.1) in poor prognosis CPM

Rank	Gene name	Fold change	FDR <i>P</i> -value
1	<i>VEZF1</i>	-10.40	3.76 x10 ⁻⁰⁵
2	<i>CEACAM1</i>	-8.27	1.83 x10 ⁻⁰³
3	<i>BCYRN1</i>	-7.96	7.50 x10 ⁻⁰⁵
4	<i>RP11-192H23.5</i>	-6.22	2.38 x10 ⁻⁰⁴
5	<i>RELB</i>	-6.17	9.50 x10 ⁻⁰⁴
6	<i>DCAF12</i>	-6.17	1.36 x10 ⁻⁰⁵
7	<i>POLR3GL</i>	-6.05	1.48 x10 ⁻⁰³
8	<i>FEM1C</i>	-5.69	1.87 x10 ⁻⁰³
9	<i>AXIN1</i>	-5.42	1.67 x10 ⁻⁰⁴
10	<i>KHDRBS3</i>	-5.01	1.23 x10 ⁻⁰³

Five genes showed decreased expression in poor prognosis CPM compared with good prognosis, however, none had a fold change ≥ 1.5 suggesting a minimal difference in expression between the good and poor prognosis cohorts (table 4.7).

Table 4-7 The five genes with significantly reduced expression (FDR<0.1) in poor prognosis CPM

Rank	Gene name	Fold change	FDR <i>P</i>-value
1	<i>KIAA0319L</i>	1.30	2.89 x10-03
2	<i>CORO7</i>	1.16	7.36 x10-04
3	<i>DIP2C</i>	1.13	1.90 x10-04
4	<i>STK38</i>	1.05	6.38 x10-05
5	<i>CLINT1</i>	1.03	2.73 x10-03

4.2.2.3 Correlation between metastasis development, prognosis and CMS and CRIS subtypes

Amongst the 51 primary CRC and CPM samples, 35 were representative of each CMS subtype, the remaining 16 samples did not have a consistent pattern within one of the four CMS subtypes, (Table 4.8, Figure 4.4). Comparison of the CMS subtypes in primary or CPM groups and prognostic groups within CPM revealed the following patterns. Firstly, there appeared to be a transition from primary CRC between CMS subtypes. No primary CRC samples were classified as CMS4 (mesenchymal subtype characterized by prominent transforming growth factor activation, stromal invasion and angiogenesis) compared to 31% of CPM, *P*-value 0.085. Secondly, patients with poor prognosis CPM were more commonly CMS4, 46% vs. 15%, *P*-value 0.005, (table 4.10). Whilst CMS were identified from a largely early stage cohort of primary CRC (59), they have subsequently been validated as prognostic in the setting of metastatic CRC (288, 289).

Table 4-8 CMS classification primary vs. CPM

	Non-consensus	CMS1	CMS2	CMS3	CMS4	Total
Primary	10 (40)	4 (16)	6 (24)	5 (20)	0 (0)	25
CPM	12 (46)	1 (4)	3 (12)	2 (7)	8 (31)	26

Fishers exact P-value 0.085, values in parenthesis percentages

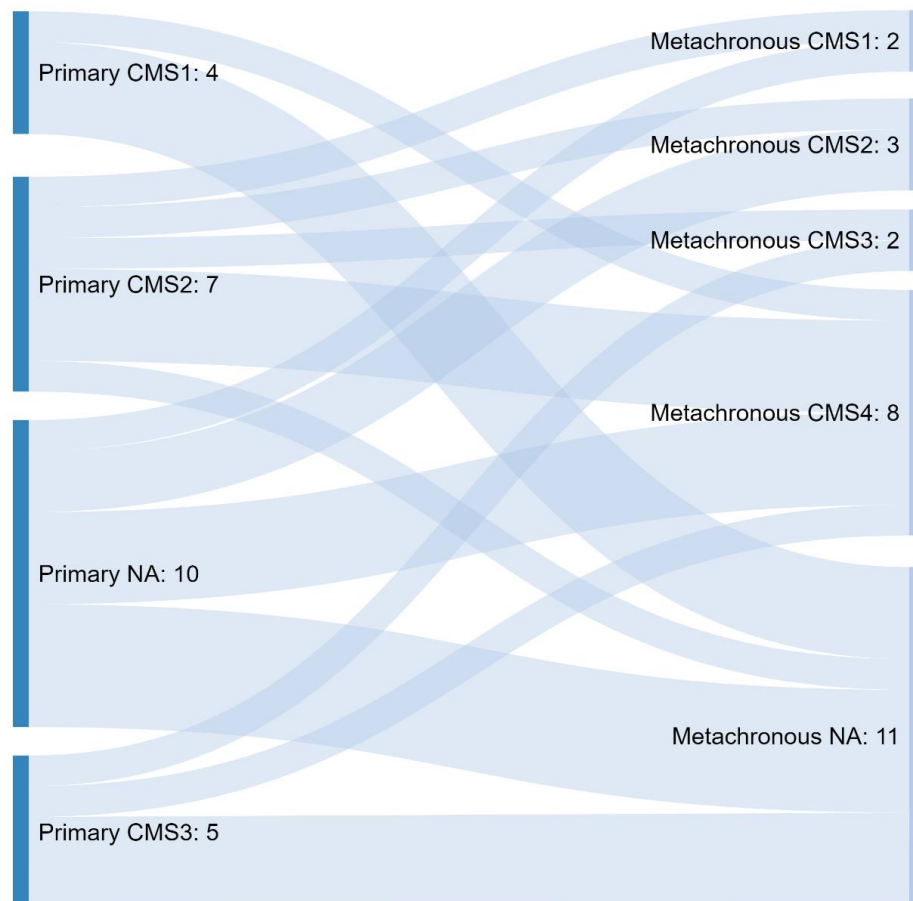


Figure 4-4 Sankey diagram of the CMS transition from primary to CPM

Sankey diagram showing the transition in CMS type from primary CRC to matched CPM. This figure is reprinted from Hallam et al, 2020, figure 4 (242).

Amongst the 51 primary CRC and CPM samples, 42 were representative of each CRIS subtype, the remaining nine samples did not have a consistent pattern within one of the four CMS subtypes, (Table 4.9, Figure 4.5). There was no significant difference in CRIS classification between primary and CPM, *P*-value 0.365, or between good and poor prognosis CPM, *P*-value 0.148, (table 4.11). The non-concordance with CMS subtypes may be a reflection of the increased stromal mesenchymal tissue present in patients with CPM. As CRIS classification is based only on tumour cells such changes in intratumoral heterogeneity are not accounted for.

Table 4-9 CRIS classification primary vs. CPM

	Non-consensus	CRISA	CRISB	CRISC	CRISD	CRISE	Total
Primary	4 (16)	6 (24)	4 (16)	8 (32)	0 (0)	3 (12)	25
CPM	5 (19)	4 (16)	7 (30)	5 (19)	3 (12)	1 (4)	26

Fishers exact P-value 0.365, values in parenthesis percentages

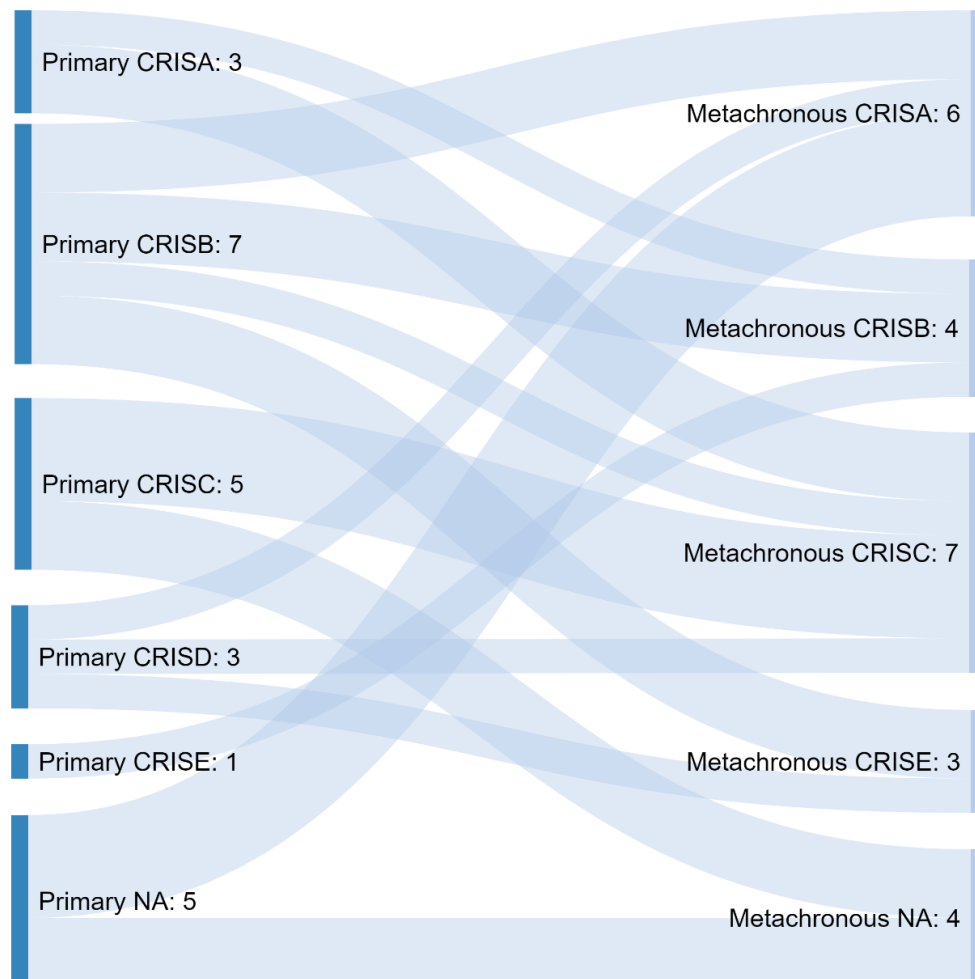


Figure 4-5 Sankey diagram of the CRIS transition from primary to CPM

Sankey diagram showing the transition in CRIS type from primary CRC to matched CPM.

Table 4-10 CMS classification good vs. poor prognosis CPM

	Non-consensus	CMS1	CMS2	CMS3	CMS4	Total
Good prognosis CPM	10 (77)	0 (0)	0 (0)	1 (8)	2 (15)	13
Poor prognosis CPM	2 (15)	1 (8)	3 (23)	1 (8)	6 (46)	13

Fishers exact P-value 0.005, values in parenthesis percentages

Table 4-11 CRIS classification good vs. poor prognosis CPM

	Non-consensus	CRISA	CRISB	CRISC	CRISD	CRISE	Total
Good prognosis CPM	5 (39)	2 (15)	2 (15)	3 (23)	0 (0)	1 (8)	13
Poor prognosis CPM	1 (8)	2 (15)	5 (39)	2 (15)	3 (23)	0 (0)	13

Fishers exact P-value 0.148, values in parenthesis percentages

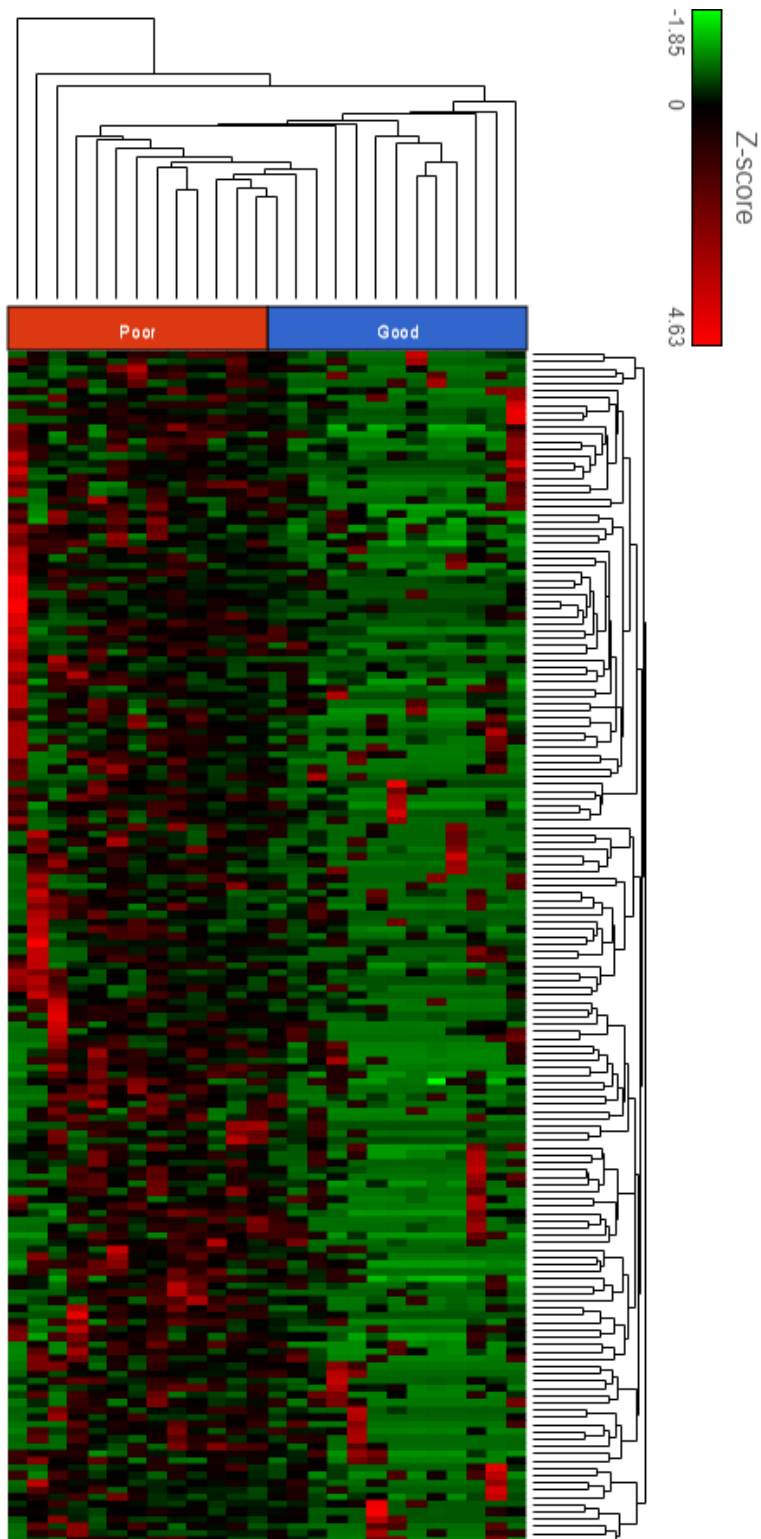


Figure 4-6 Heatmap differential gene expression good vs. poor CPM
Heatmap depicting differential genes expression in good and poor prognosis CPM. Sample type is indicated at the upper border of the heatmap, CRS blue = CPM, P red = primary. This figure is reprinted from Hallam et al, 2020, figure 2 (242).

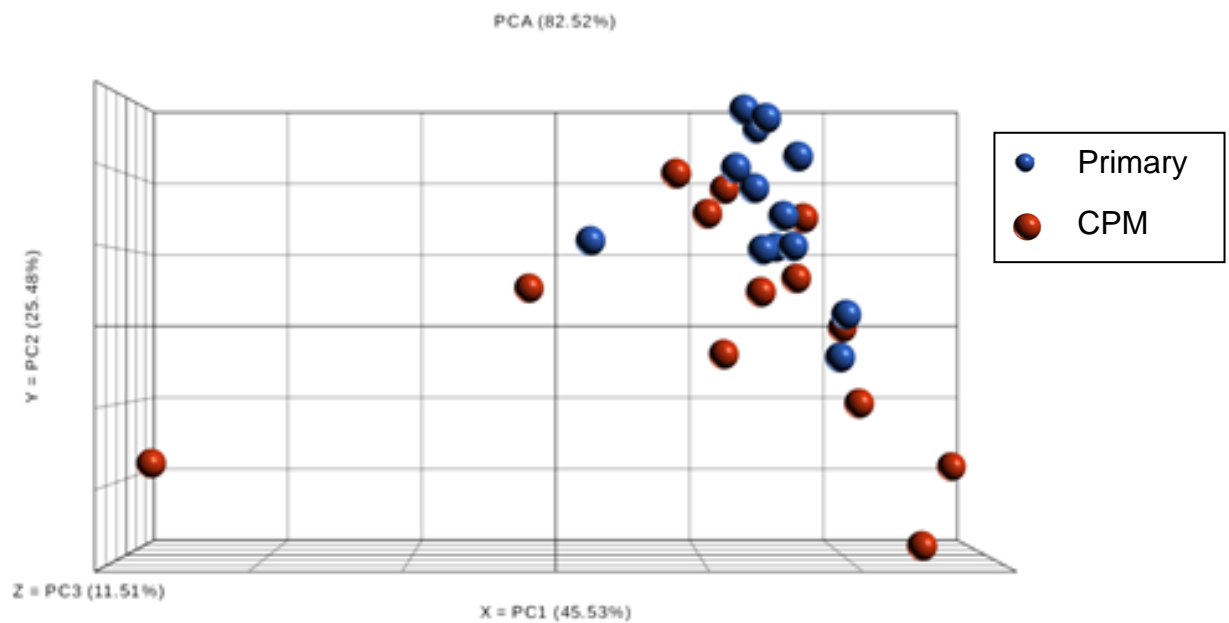


Figure 4-7 Principal component analysis, good vs. poor prognosis CPM

Principle component analysis depicting differential genes expression in good and poor prognosis CPM.

4.2.3 Methylation

4.2.3.1 Methylation Microarray Analysis, primary CRC vs. matched metachronous CPM

Assay performance was assessed initially using the Illumina GenomeStudio Methylation Module. Thirty-two samples in total were hybridised successfully to the Illumina HumanMethylation EPIC microarrays. All arrays passed the manufacturers quality control metrics for DNP, Biotin staining, extension, hybridisation, target removal, specificity, negative, non-polymorphic controls and bisulphite conversion.

Data pre-processing and quality control was carried out using the ChAMP pipeline in R studio. All probes were detected as expected in the dataset, 411 control, 865918 methylated and 865918 unmethylated probes. The following probes were removed during filtering: probes with a detection *P*-value above 0.01, (20139) probes with a bead count of <3 in at least 5% of samples, (8424), non-CpG probes, (2788) SNP probes, (77224), multi-hit probes (49), and probes located on the X and Y chromosome, (16650). The final number of probes included in the analysis was 740644. Quality control can be seen in appendices figure S1 and table S6.

4.2.3.2 Differential methylation at the probe level, primary CRC vs. matched metachronous CPM

DMPs were called between the primary CRC and CPM with a BF of 4, (Table 4.12). Two probes were significantly differentially methylated. The top-ranked differentially methylated probe was cg04146982 (chr8:144,943,810-144,943,810, hg19 coordinates), which tags a CpG dinucleotide 3651 bp upstream of the transcription start site of gene Epiplakin 1 (*EPPK1*) (290). Review of RNAseq data revealed the expression of *EPPK1* was decreased in CPM vs. matched primary CRC (Figure 4.8). *EPPK1* is part of the Plakin family an important component of the cell cytoskeleton. Plakins are critical in the regulation of cell adhesion, migration and proliferation as well as playing a key role in mediating signalling transduction such as Wnt and ERK1/2 signalling (261). The expression of *EPPK1* is increased in CRC (291). Whilst plakins have no therapeutic agents which specifically target them several anticancer drugs are common ligands for plakins suggesting they may be a potential therapeutic target (291).

The other DMP was cg12209861 (chr4:37,459,078-37,459,078, hg19 coordinates), this CpG dinucleotide 3526 bp upstream of the transcription start site of gene Chromosome 4 Open Reading Frame 19, (*C4orf19*). *C4orf19* is a protein-coding gene of unknown function, no research exists analysing its expression in CRC (261).

Table 4-12 Differentially methylated CpG sites, CPM vs. primary CRC

Probe ID	<i>Beta value</i>	<i>BF</i>	Adj. P	Gene name
cg04146982	0.79	34.5	5.67 x 10 ⁻¹⁶	EPPK1
cg12209861	0.13	7.1	0.059	C4orf19

Beta value, the proportion of methylation at a given locus ranges from 0 for unmethylated to 1 fully methylated. BF Bayes factor, adjusted P, the P-value corrected for multiple testing.

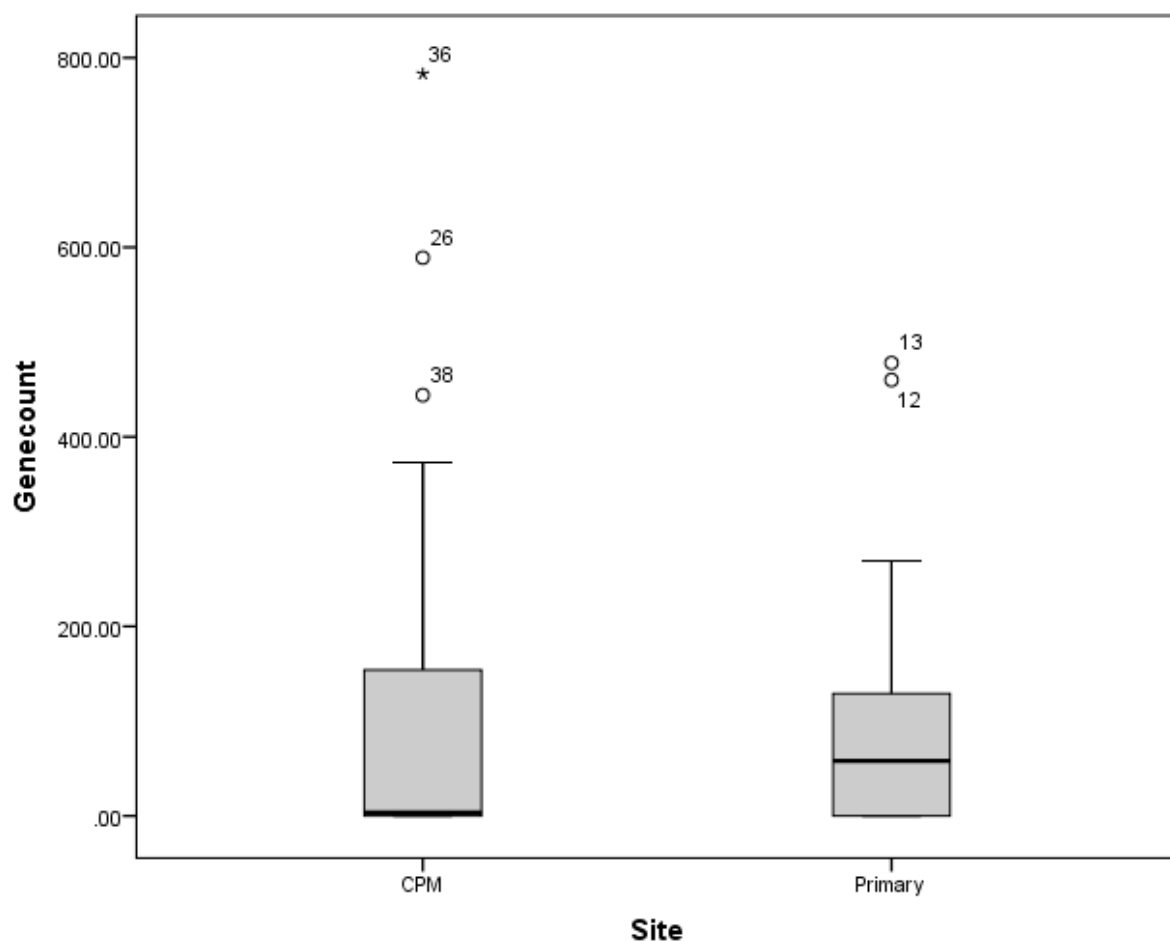


Figure 4-8 Box plot of Epiplakin expression CPM vs. primary CRC

Gene expression of EPPK1 in the retrospective patient cohort. The median expression of EPPK1 was 3.5 (IQR 156.25) in CPM and 58 (IQR 135) in primary CRC (Wilcoxon rank-sum P -value = 0.000, z = -4.878).

4.2.3.3 Differential methylated regions, primary CRC vs. matched metachronous CPM

DMRs were called between primary CRC and CPM via the dmrLasso function of the CHAMP pipeline (Table 4.13). The top 10 most significant DMRs were located at chr11:2160904-2161586, which tags an intron of gene *IGF2*. Chr19:37157318-37157945 which tags an exon of *ZNF461*. Chr15:79383167-79383980 which tags an exon of *RASGRF1*. Chr18:35145983-35147090 which tags an exon of *CELF4*. Chr19:58609269- 58609987 which tags an exon of *ZSCAN18*. Chr13:78493229-78493712 which tags an exon of gene *EDNRB*. Chr6:28602543-28603027 located 18,554 bp before gene *ZBED9*. Chr5:135415693-135416613 which tags an exon of gene *VTRNA2-1*. Chr19:58458572-58459358 which tags an exon of *ZNF256* and chr5:38258007- 38259243 which tags an exon of *EGFLAM*. The complete list of DMR between primary CRC and CPM can be seen in appendices table S7.

Insulin-Like Growth Factor 2 (*IGF2*) is a growth factor and commonly used marker of the CIMP phenotype (261). Methylation of *IGF2* is associated with reduced OS in patients with metastatic CRC (292). The insulin/IGF system plays a key role in the development and progression of CRC. Hyperactivation of the insulin/IGF pathway establishes mitogenic and pro-angiogenic signals favouring neoplastic transformation of colonic epithelium. The insulin/IGF system contributes to resistance to chemotherapy and anti-EGFR antibodies by increasing PI3K/Akt signalling which hinders apoptotic signals desensitising CRC cells to the effects of these treatments (293).

ZNF461, *ZSCAN18* and *ZNF256* are transcriptional regulators commonly methylated in CRC and overexpressed in metastatic CRC (261, 294). *ZNF461* is associated with the progression of non-small cell lung cancer, its role in the development of colorectal metastasis is unknown but given their role in transcription regulation, this suggests an increased rate of cell growth in CPM (295). *RASGRF1* is commonly methylated in CRC, its hypermethylation correlates with tumour stage, LN metastasis and reduced OS (296). *RASGRF1* is a key component of the MAPK pathway its hypermethylation

suggests dysregulation of the MAPK pathway and resulting increased cell proliferation and differentiation in CPM (22).

EDNRB encodes a G protein-coupled receptor which regulates several processes including angiogenesis and the stimulation of cell growth and division (261). Hypermethylation of *EDNRB* and its decreased expression results in increased cell proliferation, angiogenesis and metastasises (297). *EDNRB* expression is reduced in several metastatic cancers though its exact role in the process has not been described (298).

VTRNA2-1 is a tumour suppressing lncRNA, its hypermethylation abolishes its expression and tumour suppressor function (299).

Table 4-13 The top 10 differentially methylated regions, (DMRs), CPM vs. primary CRC

DMR	Chr	Start (bp)	End (bp)	Length (bp)	DMR <i>P</i>-value	Located Gene	CpG island
DMR1	chr11	2160904	2161586	682	0.002	IGF2	302
DMR2	chr19	37157318	37157945	627	0.003	ZNF461	39
DMR3	chr15	79383167	79383980	813	0.003	RASGRF1	195
DMR4	chr18	35145983	35147090	1107	0.003	CELF4	196
DMR5	chr19	58609269	58609987	718	0.003	ZSCAN18	61
DMR6	chr13	78493229	78493712	483	0.003	EDNRB	70
DMR7	chr6	28602543	28603027	484	0.004	ZBED9	32
DMR8	chr5	135415693	135416613	920	0.005	VTRNA2-1	24
DMR9	chr19	58458572	58459358	786	0.005	ZNF256	52
DMR10	chr5	38258007	38259243	1236	0.008	EGFLAM	108

4.2.3.4 Copy number alteration, primary CRC vs. matched metachronous CPM

Comparison of CNA between primary and CPM did not identify any significant differences at a stringent P -value of $< \times 10^{-10}$ however a number of CNA were identified at a lower significance threshold, $p=2.78 \times 10^{-07}$ suggesting increase or decrease in the relevant genes activity, appendices table S8. The top 10 CNAs can be seen in table 4.14.

Genes showing CNA gains of known significance in patients with CPM included; *TRIM3*, 5, 6, 21 and 22, TRIM proteins are part of the E3 ubiquitin ligase family responsible of the post-translational modification of oncoproteins and tumour suppressor proteins (261). They have diverse roles in transcriptional regulation, cell growth, apoptosis, development and tumorigenesis (261, 300).

MT1A, 2A, 3, 4 encode proteins of the metallothionein family, MT are oxidative stress-induced proteins capable of transcription factor regulation. In oesophageal and CRC MT protein expression correlates with LN and distant metastasis, chemoresistance and reduced survival (301, 302).

MMP26 encodes a matrix metalloproteinase a protein known to promote cell invasion and cancer progression in several cancer types (261). *ILK* is overexpressed in a diverse array of cancers including colorectal. *ILK* expression correlates with cell migration, invasion, EMT, metastasis and chemoresistance (303).

POLR2C, *APBB1* and *TAF10* are transcriptional regulators (261). *TAF10*, *APBB1*, *RRM1*, *CCKBR* and *UBQLN3* regulate cell cycle progression, proliferation and cellular differentiation (261).

Table 4-14 The top 10 CNAs, CPM vs. primary CRC

Chr	Start	End	Type	Count	Score	GW_P	Gene
chr11	Gain	3817692	3818253	22	0.91	2.78 X10-07	NUP98
chr11	Gain	3825449	4629195	22	0.91	2.78 X10-07	RRM1 TRIM21
chr11	Gain	4629529	5704417	22	0.91	2.78 X10-07	MMP26 UBQLN3 TRIM5, 6
chr11	Gain	5705995	5705995	22	0.91	2.78 X10-07	TRIM5
chr11	Gain	5712109	6518263	22	0.91	2.78 X10-07	TRIM3, 5, 22 FAM160A2 CCKBR SMPD1 APBB1 ARFIP2
chr11	Gain	6518907	6650480	22	0.91	2.78 X10-07	DNHD1 ILK TPP1 TAF10 RRP8
chr16	Gain	56635680	56645731	22	0.91	2.78 X10-07	MT1A, 2A, 3, 4
chr16	Gain	57333722	57335017	22	0.91	2.78 X10-07	TRNALeu
chr16	Gain	57495650	57503371	22	0.91	2.78 X10-07	POLR2C
chr16	Gain	71842906	71843647	22	0.91	2.78 X10-07	AP1G1

Located on chromosome 11 and 16, P-value of 2.78×10^{-07} were gains located on chromosome 11 and 16.

4.2.3.5 Methylation Microarray Analysis, Poor vs. Good prognosis CPM

Twenty-four samples in total were hybridised successfully to the Illumina Human Methylation EPIC microarrays. All arrays passed the manufacturers quality control metrics for DNP, Biotin staining, extension, hybridisation, target removal, specificity, negative, non-polymorphic controls and bisulphite conversion.

Data pre-processing and quality control was carried out with the ChAMP pipeline in R studio. All probes were detected as expected in the dataset, 411 control, 865918 methylated and 865918 unmethylated probes. The following probes were removed during filtering: probes with a detection *P*-value above 0.01, (18586) probes with a bead count of <3 in at least 5% of samples, (6550) , non-CpG probes, (2806) SNP probes, (77592), multi-hit probes (49), and probes located on the X and Y chromosome, (16801). The final number of probes included in the analysis was 743534. Quality control can be seen in appendices figure S1 and table S6.

4.2.3.6 Differential methylation at the probe level, Poor vs. Good prognosis CPM

DMPs were called between the good and poor prognosis CPM with a BF of 4, (Table 4.15). Three probes were significantly differentially methylated. The top-ranked differentially methylated probe was cg07951355 (cr1:40123717) which tags an intergenic region 1076 bp before gene *NT5C1A*. Cg25909064 (chr11:120081487-120082345) which tags an intron of gene *OAF* and cg12977942 (chr5:92839309-92839309) which tags an intron of gene *NR2F1-AS1* (290).

NT5C1A helps to regulate adenosine in conditions of ischemia and hypoxia which are typical in a rapidly growing aggressive tumour such as poor prognosis CPM (290). *NR2F1-AS1* is a lncRNA which regulates tumour cell growth and quiescence (304). *NR2F1-AS1* is upregulated in oxaliplatin resistant hepatocellular carcinoma (HCC) *in vitro* knockdown of *NR2F1-AS1* decreased invasion, migration and drug resistance of

HCC cells (305). Oxaliplatin is a commonly used chemotherapy agent in primary CRC and CPM, both systemically and in HIPEC, if NR2F1-AS1 is capable of mediating resistance in CRC this would contribute to the poor prognosis in this cohort of CPM. *OAF* encodes a poorly described protein whose overall function or prognostic significance in colorectal cancer is unknown (290).

Table 4-15 Differentially methylated CpG sites, good vs. poor prognosis CPM

Probe ID	<i>T</i>	<i>BF</i>	Adj. P	Gene name
cg07951355	-6.68	6	0.47	NT5C1A
cg25909064	-5.72	4	0.47	OAF
cg12977942	-5.63	4	0.47	NR2F1-AS1

T: the *t* value (the size of the difference relative to the variation in the sample) *P* the raw *P*-value, not corrected for multiple testing; adjusted *P*, the *P*-value corrected for multiple testing.

4.2.3.7 Differential methylated regions, Poor vs. Good prognosis CPM

DMRs were called between primary CRC and CPM via the dmrLasso function of the CHAMP pipeline (Table 4.16). Six significant DMRs were located at chr10:134600295-134600919, 758 bp following gene *NKX6-2*. Chr12:133464419-133464892 which tags an exon of gene *CHFR*. Chr10:8096305-8096991 which tags an exon of gene *GATA3*. Chr16:54972078-54973579 located 3683 bp following gene *IRX5*. Chr20:30639816-30640256 tags an exon of gene *HCK* and chr11:2397201-2397655 which tags an intron of gene *BC019904*.

NKX6-2, *GATA3* and *IRX5* encode transcription regulators, their hypermethylation suggests an increased rate of growth in poor prognosis CPM (261). *GATA3* regulates tumour angiogenesis, *in vitro* it represses the expression of chemokine *CXCL1* allowing the extravasation of proangiogenic monocytes (306). *IRX5* is found in many cancers and regulates genes controlling apoptosis, it is also thought to reduce the sensitivity to the cytostatic effect of *TGF- β* (307). The expression of *IRX5* is correlated with its neighbouring lncRNA *CRNDE-h* suggesting that the expression of both genes is under a level of coordinate control. In CRC high expression of *CRNDE-h* is associated with reduced OS (307).

CHFR regulates cell cycle progression and entry into mitosis and functions as a negative regulator of NF- κ B (308). *CHFR* is epigenetically silenced by hypermethylation resulting in angiogenesis, cell migration, cytokine production and cell survival *in vitro*, suggesting a loss of *CHFR* tumour suppressor function in CPM. Clinically in CRC hypermethylation of *CHFR* is associated with LN metastasis and reduced OS (309).

Proto-Oncogene *HCK* is a member of the Src family of tyrosine kinases found in hematopoietic cells (261). *HCK* plays an important role in the regulation of innate immune responses (261). Pho et al found that high *HCK* levels correlate with reduced survival of patients with CRC and that in a mouse model *HCK* promotes the growth of CRC (310). Its hypermethylation in CPM suggests dysregulation of innate immunity

and *HCK* oncogene activity. BC019904 encodes a poorly described gene, its significance is unknown in CRC (261).

Table 4-16 Differentially methylated regions, (DMRs) good vs. poor prognosis CPM

DMR	Chr	Start (bp)	End (bp)	Length (bp)	DMR <i>P</i>-value	Located Gene	CpG island
DMR_1	chr10	134600357	134600919	562	0.007	NKX6-2	569
DMR_2	chr12	133464496	133464737	473	0.003	CHFR	91
DMR_3	chr10	8096311	8096991	680	0.006	NA	509
DMR_4	chr20	30639816	30640256	440	0.009	HCK	71
DMR_5	chr16	54972078	54973579	1501	0.015	NA	179
DMR_6	chr11	2397201	2397655	454	0.016	BC019904	NA

4.2.3.8 Copy number alteration, Poor vs. Good prognosis CPM

Comparison of CNA between the CPM prognostic groups identified 34 gene losses at chromosomes 3, 4, 14, 15, 17 and 19, these are detailed in appendices table S9.

CNA losses clustered in the RAS-MAPK-ERK signalling pathway suggesting dysregulation in poor prognosis CPM. The RAS oncogenes *RAB8A* and *RAB34* are small GTPases capable of activating the expression of genes involved in cell growth, differentiation and survival (261). *FGF5* regulates cell proliferation and differentiation, *FGF5* interacts with ERK, FGF/ERK signalling is capable of maintaining pluripotency and driving differentiation (261, 311). *BMP3* hypermethylation is associated with the CIMP+ phenotype. Epigenetic inactivation of *BMP3* is associated with the activation of the RAS-RAF-MEK-ERK signalling pathway (312).

PTPN21 and *VASH1* regulate cell growth and angiogenesis (261), *PTPN21* and *FAM32* regulate the cell cycle and apoptosis, *ZKSCAN7* regulates transcription and *IGFBP7* cell growth. CNA loss suggests increased rates of unchecked growth in poor prognosis CPM (261).

RFTN1 regulates T and B-cell antigen receptor-mediated signalling, antibody production and cytokine secretion (261). CNA loss of *RFTN1* suggests immune evasion in poor prognosis CPM.

BMP3 is a ligand of the *TGF-β* protein family (261), CNA loss of a *TGF-β* ligand may suggest dysregulation of the *TGF-β* pathway in poor prognosis CPM.

Comparison of CNA between the CPM prognostic groups identified 19 gene gains at chromosomes 9, 10 and 11 these are detailed in appendices table S9.

Genes showing CNA gains in patients with poor prognosis CPM included; *SIT1* a negative regulator of T-cell antigen mediated signalling (261). *RNF38* a gene involved in oncogenesis and signal transduction (261). *MELK* a gene regulating the cell cycle, proliferation and carcinogenesis (261). *PAX5* a transcription factor regulating tissue differentiation (261). *SHB* a gene regulating signal transduction and angiogenesis (261). *ZEB1* a gene involved in EMT by the recruitment of *SMARCA4/BRG1* and in promoting tumorigenicity by repressing stemness inhibiting microRNAs (261). *DEAF1* a transcription regulator (261). *ANTXR* a transmembrane protein regulating cell

adhesion (261). *EPS8L2* a gene regulating actin cytoskeleton remodelling (261).
PIDD1 an effector of *p53* dependent apoptosis (261).

4.2.4 Exome sequencing

4.2.4.1 Sequencing quality

In Accordance with Illumina sequencing coverage recommendations, 12x samples were multiplexed on a high output flow cell to give 50X coverage and a minimum of 33.3 million reads. Across all six sequencing runs, we obtained a median of 59.6X coverage (2.3-166.4) with a median uniformity of 88.1% (70.67-89.32). The median Q30 score obtained was 91.73% (85.23-93.68). Sequences with a Phred score of <30 were removed giving a base call accuracy of 99.9%, appendices, table S10.

4.2.4.2 Primary CRC, somatic variations

In the primary CRC, there were 112,420 somatic SNV's, (primary, normal subtraction). Variants present at >1% in the dbSNP common or 1000-Genomes databases were filtered leaving 38,366 variants. Of these SNVs, 7267 were classified as non-synonymous. Eighty variants were found in the dbSNP or 1000 genome database, the remainder were novel. Of the novel variants, 3657 had a PhastCons score of 1 suggesting they are highly conserved and likely functionally important mutations. Four samples have a TMB \geq 20 mut/Mb, median 69.79 (26.62 – 203.53) (313). Additionally, 12,727 indels were identified. Variants present at >1% in the dbSNP common or 1000 Genomes databases were filtered leaving 7956 variants. Of these 736 variants were classified as non-synonymous. Sixty-seven variants were found in the dbSNP or 1000 genome database, the remainder were novel, of these 228 variants had a PhastCons score of 1 suggesting they are highly conserved and likely functionally important mutations. The Somatic SNV and indel mutations in primary CRC were compared to known CRC driver mutations reported on the IntOGen database. Mutations were present in 64 of 95 known CRC driver genes (Table 4.17) (314).

Table 4-17 Known CRC driver mutations identified in the primary CRC cohort

BRAF	MED24	KRAS	ASPM	TRIO	CHD4	CNOT1	TAF1
MECOM	MED12	ARID1A	BMPR2	PIK3R1	NUP107	NF1	PTPRU
EGFR	DIS3	TP53	FN1	APC	CEP290	AXIN2	DNMT3A
NRAS	PTGS1	CASP8	TGFBR2	RAD21	SOS2	SOX9	CAD
CUL1	AKAP9	MAP3K4	ACSL6	ATRX	CHD9	BPTF	PPP2R1A
NUP98	PTPN11	BRWD1	FXR1	PTEN	TP53BP1	GNAS	SF3B1
NTN4	SMAD4	CLSPN	POLR2B	TCF7L2	TCF12	ITSN1	CTNNB1
TTBX3	MSR1	CDC73	FBXW7	ATM	IDH2	BCOR	PIK3CA

4.2.4.3 Primary CRC, candidate driver mutations

Somatic mutations were entered into the IntOGen pipeline to determine their function and to identify candidate driver mutations. OncodriveFM identified 46 candidate driver genes with high levels of functional mutation, (Q-value<0.05) in the primary CRC cohort, appendices, table S11. The top 10 potential driver genes are discussed below (Table 4.18). *KRAS* mutation was present in 69% of primary CRC samples. Mutation of the *KRAS* oncogene activates *BRAF* kinases, driving the MAPK signalling pathway leading to tumour proliferation and growth (31). *KRAS* mutation is associated with a high risk of recurrence in keeping with a cohort of primary CRC which have developed CPM (32). The high proportion of *KRAS* mutation in this cohort meant they would have had a poor response to anti-EGFR therapy (33). Tumour suppressor mutations of *p53* and *ARHGEF12* were present in 62% and 23% of samples respectively. *p53* induces cell cycle arrest, senescence and apoptosis in response to cellular stress of DNA damage. The loss of *p53* through mutation allows tumour growth to establish, loss of checkpoint control enables continued cycling and the acquisition of additional mutations enabling cells to breach the basement membrane, acquire invasive and migratory potential and undergo EMT ultimately resulting in CPM (315). Wild type *ARHGEF12* results in reduced cell proliferation and migration *in vitro* suggesting a tumour suppressor role in CRC which has most likely been lost in this cohort of primary CRC (261, 316). *CAD* mutation was present in 40% of primary CRC samples. *CAD* controls pyrimidine nucleotide biosynthesis, an essential step in cell proliferation. *CAD* levels are regulated by the MAPK cascade (261). Wild type *CAD* is a negative regulator of the WNT/ β -catenin signalling pathway in CRC inhibiting cell migration (317). Loss of *CAD* through mutation is likely to result in dysregulation of the MAPK and WNT/ β -catenin pathways with unchecked growth and metastasis.

NME1-NME2 mutation was present in 30% of primary CRC, the function of *NME1-NME2* is poorly described though mutations are commonly found in highly metastatic cells such as neuroblastoma (261). *NMD3* and *SDHA* are ribosome biogenesis factors present in 31% of primary CRC, and components of the mitochondrial respiratory chain present in 15% of primary CRC respectively, their mutation may reflect the increased rate of cell growth and proliferation in primary CRC (261). *CDC42BPA* mutation was

present in 23% of primary CRC. *CDC42BPA* is a member of the serine/threonine-protein kinase family, a downstream effector of *CDC42* which plays a role in cytoskeleton reorganisation, cell migration and cell cycle regulation (261) mechanisms essential for the growth and metastasis of primary CRC. *AP1S1* mutation was present in 23% of samples, *AP1S1* encodes part of the clathrin coat assembly complex. *RCN3* mutation was present in 23% of primary CRC, *RCN3* encodes a protein essential for biosynthesis and transport. The significance of these mutations is unknown (261). Of note in the remaining potential driver genes identified by the IntOGen pipeline, *APC* mutations were identified in 38% of primary CRC samples. Mutational inactivation of the *APC* tumour suppressor gene results in inappropriate and consistent activation of the Wnt signalling pathway.

Table 4-18 The top 10 genes with a bias towards a high functional level of mutation in primary CRC

	P-value	Q-value	PPH2 P-value	MA P-value	SIFT P-value
<i>KRAS</i>	3.66E-15	1.03E-11	0.044	0.000	0.000
<i>TP53</i>	3.90E-12	5.46E-09	0.000	0.000	0.002
<i>CAD</i>	1.36E-11	1.27E-08	0.001	0.000	0.005
<i>NME1-NME2</i>	1.35E-08	9.47E-06	0.081	0.000	0.054
<i>NMD3</i>	1.46E-06	0.001	0.001	0.001	0.011
<i>RCN3</i>	3.43E-05	0.014	0.010	0.002	0.011
<i>CDC42BPA</i>	3.38E-05	0.014	0.004	0.007	0.011
<i>SDHA</i>	5.17E-05	0.017	0.020	0.000	0.054
<i>AP1S1</i>	5.71E-05	0.017	0.029	0.000	0.054
<i>ARHGEF12</i>	6.22E-05	0.017	0.003	0.015	0.011

Identified using OncodriveFM. Intogen, Q-value <0.05, potential drivers in primary CRC cohort. (PPH2, Polyphen2, MA Mutation Assessor, SIFT Sorting Intolerant from Tolerant).

4.2.4.4 Primary CRC, pathways

Somatic mutations were entered into the IntOGen pipeline to determine the KEGG pathways with multiple accumulated mutations. Ten cancer KEGG pathways were identified including the CRC pathway, 12 metabolic pathways and 4 associated with DNA replication and repair (Table 4.19). The complete list of KEGG pathways identified can be found in appendices, table S12.

Table 4-19 The top 10 KEGG pathways with accumulated mutations of high functional impact in primary CRC

Kegg ID	Name	Class	Q-value
hsa00280	Valine, leucine and isoleucine degradation	Metabolism	7.54E-06
hsa05217	Basal cell carcinoma	Cancers: Specific types	2.13E-05
hsa05210	Colorectal cancer	Cancers: Specific types	0.000
hsa04726	Serotonergic synapse	Nervous system	0.000
hsa03008	Ribosome biogenesis in eukaryotes	Genetic; Translation	0.000
hsa00240	Pyrimidine metabolism	Metabolism	0.000
hsa05216	Thyroid cancer	Cancers: Specific types	0.000
hsa04730	Long-term depression	Nervous system	0.000
hsa03430	Mismatch repair	Genetic Replication and repair	0.001
hsa01100	Metabolic pathways	Metabolism	0.001

4.2.4.5 CPM, somatic variations

In the CPM cohort 244,531 somatic SNV's were identified, (CPM, primary subtraction). Variants present at >1% in the dbSNP common or 1000 Genomes databases were filtered leaving 33,652 variants. Six thousand and twenty-three variants were classified as non-synonymous, 28 mutations were present in three samples, 286 were present in 2 samples, the remainder were present in 1 sample. Ninety-three variants were found in the dbSNP or the 1000 genome database, the remainder were novel, of these 3241 variants had a PhastCons score of 1 suggesting they are highly conserved and likely functionally important mutations. Seven samples had a TMB \geq 20 mut/mb, median 198.76 (162.5-321.99) (313). Additionally, 19,979 indels were identified. Variants present at >1% in the dbSNP common or 1000 Genomes databases were filtered leaving 4812 variants. Of these 313 variants were classified as non-synonymous, 2 mutations were present in three samples, 21 were present in 2 samples, the remainder were present in 1 sample. Sixty-nine variants were found in the dbSNP or the 1000 genome database, the remainder were novel, of these 61 variants had a PhastCons score of one suggesting they are highly conserved and likely functionally important mutations. The Somatic SNV and indel mutations in CPM were compared to known CRC driver mutations reported in the IntOGen database. Mutations were identified in 69 of 95 known CRC driver genes, 51 were shared between the primary and CPM, 13 were novel, (table 4.20) (314). Of the somatic variants identified in CPM, 58,958, (29%) were present in the primary CRC, 205,552 variants occurred exclusively in the CPM suggesting a significant accumulation of mutations in the transition to CPM (Figure 4.9).

Table 4-20 Known CRC driver gene mutations identified in the CPM cohort

BRAF	MED24	KRAS	ASPM	TRIO	CHD4	CNOT1	TAF1
MED12	ARID1A	NUP107	NF1	PTPRU	DIS3	TP53	FN1
APC	DNMT3A	NRAS	PTGS1	CASP8	RAD21	SOX9	CAD
CUL1	AKAP9	MAP3K4	ACSL6	ATRX	CHD9	BPTF	NUP98
PTPN11	BRWD1	FXR1	TP53BP1	GNAS	SF3B1	NTN4	POLR2B
TCF7L2	CTNNB1	CDC73	ATM	BCOR	PIK3CA	TBX3	NR4A2
FOXP1	WIPF1	ITSN1	SRGAP3	SOS2	WT1	GATA3	MGA
GOLGA5	SMAD2	ELF3	CREBBP	SMC1A	ACO1	ZC3H11A	PPP2R1A
RBM10	AXIN2	PIK3R1					

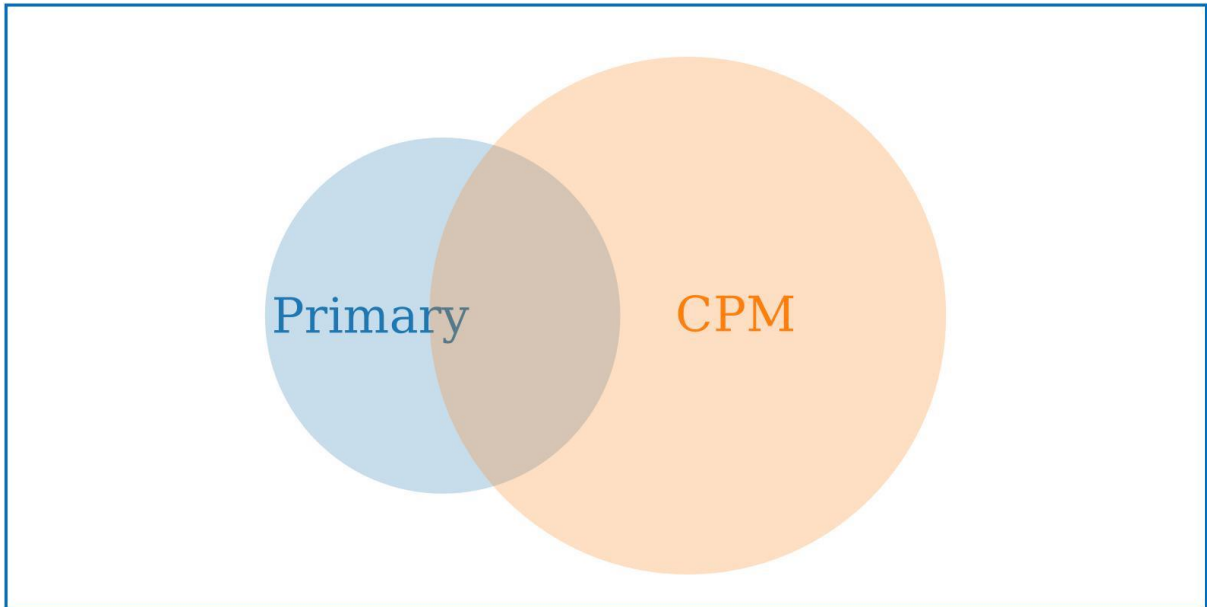


Figure 4-9 Mutations exclusive to and shared between primary CRC and matched CPM

Venn diagram showing the proportion of mutations exclusive to and shared between primary CRC and matched CPM. This figure is adapted from Hallam et al, 2020, figure 1 (242).

4.2.4.6 CPM, candidate driver mutations

Somatic mutations were entered into the IntOGen pipeline to determine their functional importance and to identify potential driver mutations. OncodriveFM identified 265 potential driver genes with high levels of functional mutation, (Q-value<0.05) in the CPM cohort. The top 10 potential driver genes are discussed below (Table 4.21). The full list can be seen in appendices, table S13.

FLNB, *SPTB*, *PPL*, and *SVEP1* regulate cell adhesion, invasion and the induction of EMT, their mutation in CPM explain in part the process of metastasis from primary to CPM through the loss of cell adhesion and polarity typical of EMT. *FLNB* mutations were present in 46% of CPM samples. *FLNB* is an actin-binding protein which anchors the cell onto plasma adhesion receptors such as integrins regulating the dynamic changes of the actin cytoskeleton in response to extracellular signals (261). *FLNB* deficiency in a xenograft mouse model increases angiogenesis and promotes tumour growth and metastasis (318). Alternate splicing of *FLNB* is strongly associated with EMT signatures in breast cancer (319). *SPTB* mutations were present in 46% of CPM. *SPTB* encodes a spectrin protein involved in cell shape and membrane integrity, mutations likely reflect changes in cell adherence, a hallmark of metastatic cancer (261). *PPL* mutations were present in 38% of CPM samples. *PPL* encodes periplakin, *AKT1/PKB* interacts with *PPL* suggesting a role in cell growth and survival (261). The expression of wild type *PPL* is reduced in CRC vs. normal tissue. Knockdown of *PPL* *in vitro* promotes cell proliferation, migration, invasion and EMT transition (320). *SVEP1* mutations were present in 54% of CPM samples. *SVEP1* is a cell adhesion molecule. *SVEP1* expression is stimulated by *TNF α* a pro-inflammatory cytokine able to affect cell adhesion, migration and induce EMT (261, 321). These mutations suggest a mechanism for the development of metastasis and the CMS4 mesenchymal phenotype seen more commonly in the CPM cohort.

p53 tumour suppressor mutation was present in 38% of CPM, the loss of *p53* through mutation allows tumours to establish, invade and metastasis as discussed in more detail in section 4.4.4.2.

PDE4DIP, *RIOK2* and *CDC16* are cell cycle regulators, their mutation suggests increased rates of cell growth and unchecked proliferation in CPM. *PDE4DIP* was mutated in 46% of CPM samples. *PDE4DIP* functions as an anchor, sequestering components of the cAMP-dependant pathway to Golgi and or centrosomes (261). This is a crucial process in cell migration, mitosis and cell-cycle progression (261). *RIOK2* mutations were present in 31% of CPM. *RIOK2* is a serine/threonine biogenesis-associated protein kinase regulating cell cycle progression (261). *RIOK2* overexpression promotes cell proliferation and loss to cell cycle exit and apoptosis in glioblastoma cells (322). *CDC16* encodes a ubiquitin ligase component of the anaphase-promoting complex, *APC/C* a cyclin degradation system governing the exit from mitosis by the degradation of G1 and mitotic checkpoint regulators. *In vitro* overexpression of *CDC16* suppresses invasion and metastasis of lung cancer cells (323).

NUP98 and *CDC16* are mitotic checkpoint regulators. Mutation of mitotic checkpoint regulators may result in the accumulation of additional mutations in CPM. *NUP98* mutations were present in 31% of CPM samples. Nuclear pore complexes (NPCs) regulate the transport of macromolecules between the nucleus and cytoplasm. *NUP98* is known to form a complex with *RAE1* preventing aneuploidy through the inhibition of premature securing degradation during mitosis (261). *RAE1* and *NUP98* play a critical role in regulating chromosomal segregation through the inhibition of securing ubiquitination mediated by the anaphase-promoting complex, (*APC/C*) (261). *CDC16* mutations were present in 23% of CPM samples. *CDC16* encodes a ubiquitin ligase component of the anaphase-promoting complex, *APC/C* a cyclin degradation system governing the exit from mitosis by the degradation of G1 and mitotic checkpoint regulators. *In vitro* overexpression of *CDC16* suppresses invasion and metastasis of lung cancer cells suggesting a potential role in CRC metastasis (323). *RRP7A* mutations were present in 38% of CPM. *RRP7A* encodes a protein involved in ribosome biogenesis and rRNA processing, its significance in CRC is unknown (261).

Of note in the remaining candidate driver genes identified by the IntOGen pipeline, the following known CRC driver mutations were identified.

HLA-A mutations were present in 54% of CPM samples. *HLA-A* is one of 6 major histocompatibility complex antigen alleles responsible for presenting antigens to T-cells, which triggers the apoptosis of non-self-cells (261). T-cell infiltration of CRC is inversely correlated with *HLA-A* expression and improved DFS. HLA antigen loss is common in MSI-H CRC where immune selective pressure leads to cell proliferation with defects in antigen presentation (324).

MSH6 mutations were present in 38% of CPM samples. *MSH6* encodes a mismatch repair protein, *MutS* involved in the repair of both single-base mismatches and insertion-deletion loops (261). Mutations of *MSH6* are associated with hereditary non-polyposis colon cancer, colorectal and endometrial cancer (261). No patient in the prospective cohort had a family history of CRC, which would be expected for an autosomal dominant germline mutation resulting in an up to 70% lifetime risk of CRC, suggesting these mutations were somatic. Such mutations may result in microsatellite instability and have been associated with the presence of isolated peritoneal metastasis (18). Microsatellite instability was found in 46% of CPM samples, all had associated high TMB median 195.59 (162-321.99).

APC mutations were identified in 38% of CPM samples, the same proportion mutated in primary CRC, section 4.4.4.2.

TET2 mutations were identified in 23% of CPM samples. *TET2* is a dioxygenase that catalyses the conversion of 5MC into oxidative products, resulting in active DNA demethylation (261). *TET2* has emerged as an important tumour suppressor in cancer is one of the most frequently mutated genes in a wide range of cancer types. The regulation of DNA methylation by TET enzymes protects malignant transformation (325). *TET2* is expressed in CRC it does, however, show reduced nuclear localisation, particularly in patients with distant metastasis suggesting a loss of function (326). DNA hypermethylation of tumour suppressor genes may reflect a loss of active DNA

methylation, reactivation of active DNA methylation may, therefore, result in re-expression of tumour suppressors (326). *In vitro*-nuclear export inhibitor, Leptomycin-B can increase levels of 5HMC and is a potential therapeutic option for these patients in addition to inhibitors of DNA methyltransferases (326). *CHD7* mutations were identified in 15% of CPM samples, *CHD7* is a transcriptional regulator and regulator of chromatin (261). Amongst tumours with *CIMP* and resulting global hypermethylation *CDH7* is commonly mutated suggesting aberrant chromatin regulation resulting from *CDH7* mutation may contribute to the development of *CIMP* CRC (327).

Mutation of *RNF43*, *FAM123B* and *TSC1* key components of the Wnt pathway suggest Wnt dysregulation in CPM. *RNF43* mutations were present in 23% of CPM samples. *RNF43* encodes an E3 ubiquitin-protein ligase, a negative regulator of Wnt signalling which promotes the turnover of *frizzled* and *LRP6* (261). *In vivo* inhibition of *RNF43* enhances Wnt/ β -catenin signalling in contrast to R-spondin, which enhances Wnt signalling by inhibiting *RNF43* (328). *RNF43* mutations exist upstream of the Wnt pathway making them an interesting therapeutic target as they would inhibit the Wnt pathway despite the presence of downstream pathway mutations (328). *FAM123B* mutations were identified in 15% of CPM samples. *FAM123B* is a regulator of the Wnt signalling pathway interacting with the beta-catenin destruction complex. *FAM123B* can act as a positive Wnt regulator by promoting *LRP6* phosphorylation or as a negative regulator acting as a scaffold protein for the beta-catenin destruction complex (329). *FAM123B* is commonly mutated in CRC resulting in its inactivation, tumours with *FAM123B* mutations exhibit a mesenchymal phenotype characterised by inhibition of the Wnt pathway (329). *TSC1* mutations were identified in 23% of CPM samples. *TSC1* and *TSC2* encode hamartin and tuberlin, which form a functional complex affecting cell growth, differentiation and proliferation. *TSC1* is a negative regulator of the *mTOR* signalling pathway, a promotor of cell growth (261). Mutation of *TSC1* results in reduced survival for patients with CRC (330). Wild type *TSC1* negatively regulates β -catenin stability and activity by participating in the β -catenin degradation complex (331).

CCND2 mutations were identified in 15% of CPM samples. *CCND2* encodes a cyclin which forms a complex with *CDK4* or *CDK6* regulating cell cycle transition at G1/S (261). *CCND2* expression correlates with proliferation and increased metastatic potential in CRC cells (332).

Table 4-21The top 10 genes with a bias towards high functional levels of mutations in CPM

	<i>P</i> -value	Q-value	PPH2 <i>P</i> -value	MA <i>P</i> -value	SIFT <i>P</i> -value
<i>FLNB</i>	1.38E-13	7.53E-10	1.00E-08	0.0001	0.000
<i>TP53</i>	1.10E-12	3.00E-09	0.000	1.00E-08	0.002
<i>RIOK2</i>	2.33E-12	3.73E-09	0.000	1.00E-08	0.001
<i>PDE4DIP</i>	3.42E-12	3.73E-09	0.001	1.00E-08	0.001
<i>RRP7A</i>	2.88E-12	3.73E-09	0.001	1.00E-08	0.001
<i>SPTB</i>	1.35E-11	1.22E-08	0.000	1.00E-08	0.009
<i>PPL</i>	1.92E-10	1.49E-07	1.00E-08	0.007	0.007
<i>NUP98</i>	2.49E-10	1.70E-07	1.00E-08	0.008	0.008
<i>SVEP1</i>	2.16E-09	1.31E-06	0.000	0.000	0.000
<i>CDC16</i>	1.75E-08	7.35E-06	0.000	0.000	0.003

Identified using OncodriveFM, IntOGen, Q-value <0.05, potential drivers in CPM CRC cohort. (PPH2, Polyphen2, MA Mutation Assessor, SIFT Sorting Intolerant from Tolerant).

4.2.4.7 CPM, pathways

Somatic mutations were entered into the IntOGen pipeline to determine the KEGG pathways with multiple accumulated mutations. The top ten KEGG pathways included 6 concerning the immune system, one signalling, one metabolic and one cancer, (table 4.22). The complete list of KEGG pathways identified can be found in appendices, table S14.

Table 4-22 The top 10 KEGG pathways with accumulated mutations of high functional impact in CPM

KEGG	Name	Class	Q-value
hsa04512	ECM-receptor interaction	Signalling molecules and interaction	1.90E-07
hsa04940	Type I diabetes mellitus	Endocrine and metabolic disease	1.90E-07
hsa05330	Allograft rejection	Immune diseases	2.09E-07
hsa05332	Graft-versus-host disease	Immune diseases	3.06E-07
hsa05145	Toxoplasmosis	Infectious diseases	6.45E-07
hsa05310	Asthma	Immune diseases	6.45E-07
hsa05320	Autoimmune thyroid disease	Immune diseases	9.86E-07
hsa04672	Intestinal immune network for IgA production	Immune system	1.42E-05
hsa00410	beta-Alanine metabolism	Metabolism	4.32E-05
hsa05217	Basal cell carcinoma	Cancers: Specific types	0.000

4.2.5 Prognostic groups

Patients with poor prognosis CPM had a higher proportion of somatic mutations than patients with good prognosis, 55% of all mutations in CPM cohort vs. 45%. Of the somatic mutations identified in poor prognosis CPM, 115,860 (45%) were present in patients with good prognosis CPM, 145,089 (55%) variants occurred exclusively in patients with poor prognosis CPM (Figure 4.10). Patients with poor prognosis CPM were more commonly MSI-H with a high TMB 57%, of these 25% were *BRAF* mutant.

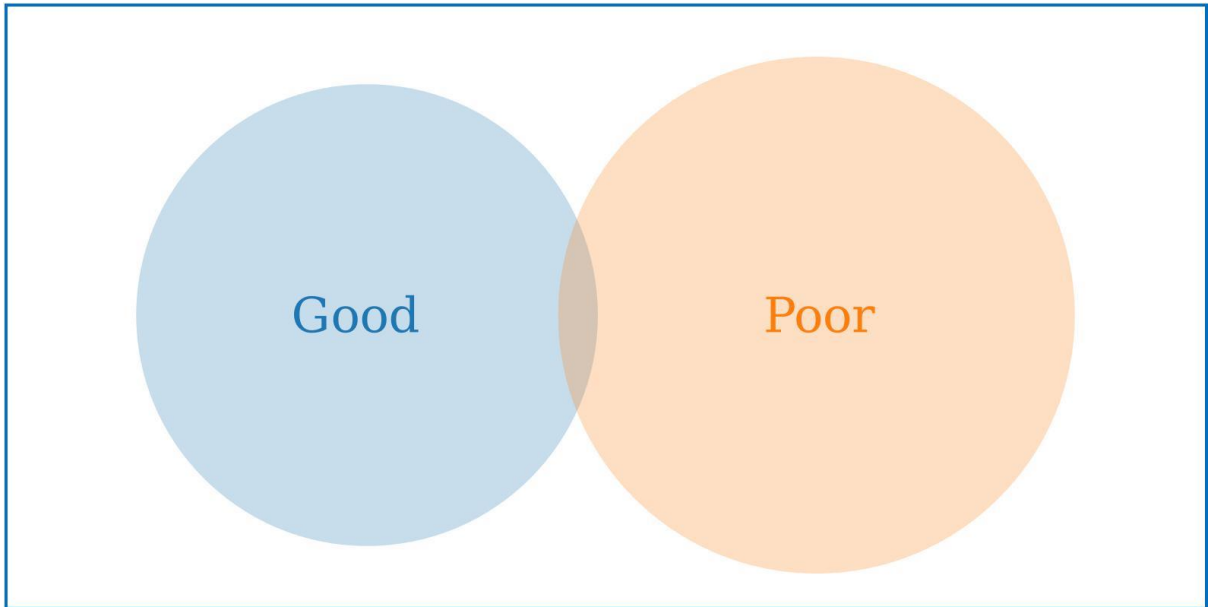


Figure 4-10 Mutations exclusive to and shared between good and poor prognosis CPM

Venn diagram showing the proportion of mutations exclusive to and shared between good and poor prognosis CPM. This figure is adapted from Hallam et al, 2020, figure 1 (242).

Table 4-23 Potential candidate variants, poor prognosis CPM

Chr	Reference	Allele	P-value	FDR	Sample frequency (case)	Sample frequency (control)	Gene ID
4	C	G	0.007	0.53	62.5	0	FAM13A
18	G	C	0.023	0.53	50	0	PIEZO2

CPM identified through Fisher exact test, genomics workbench (Chr, chromosome, FDR, false discovery rate)

4.2.5.1 Candidate driver mutations for poor prognosis CPM

Comparison of somatic mutations, (Fisher exact test, genomics workbench) in patients with good and poor prognosis identified two potential candidate genes, we must, however, note the high FDR of 0.53 indicating that these results are at high risk of being false positives (Table 4.23).

FAM13A is a protein-coding gene it is known to interact with several pathways important in CRC including Wnt/ β catenin and *TGF β* . *FAM13A* can activate the Wnt/ β catenin pathway by increasing the stability of β catenin, which accumulates and translocates to the nucleus driving gene transcription (333, 334). Knockdown of *FAM13A* reduces Wnt signalling in human lung cancer cells. *FAM13A* is inversely correlated with *TGF β* , *TGF β* can act as a tumour suppressor and tumour promoter (335). In early-stage cancers it inhibits cell growth and proliferation, in advanced cancers it contributes to disease progression and the induction of cell migration and metastasis (335). *In vitro*, *TGF β* induces apoptosis in lung cancer cell lines and downregulates both the RNA and protein levels of *FAM13A* (335). Silencing *FAM13A* with *FAM13* siRNA results in a downregulation of *FAM13A*, promoting Rho protein activity which results in a decrease in cell number and proliferation but an increase in random cell migration (335). *FAM13A* mutation in poor prognosis CPM suggests dysregulation of the Wnt and *TGF β* pathways.

PIEZO2 encodes a transmembrane mechanically activated ion channel protein (336). The *PIEZO2* generated calcium signal activates the downstream *RhoA* pathway necessary for the regulation of the actin cytoskeleton (336). In breast cancer brain metastasis, *PIEZO2* plays a key role in the ability of cells to sense and transduce mechanical cues from the environment conferring an advantage in the cells ability to proliferate, invade and migrate within their secondary environment (336). In glioblastoma cells, *PIEZO2* knockdown results in tumour growth inhibition, reduced vascular density and vascular hyperpermeability (337). This effect was thought to be mediated through the reduced secretion of Wnt/ β catenin protein *Wnt11* in *PIEZO2* knockdown cells (337). *PIEZO2* mutation may have a role in the process of invasion, migration and metastasis in poor prognosis CPM.

4.3 Discussion

This study determined the gene expression, CNA, methylation and somatic mutation profile of primary CRC and matched metachronous CPM to determine whether there was a change associated with the development of CPM or predicting prognosis for patients with CPM. To our knowledge, this is the first such analysis in a cohort of patients with isolated CPM suitable for treatment with CRS & HIPEC.

Comparison of patients with primary CRC and metachronous CPM identified biological changes associated with the transition from primary CRC to CPM. There was downregulation of markers of early CRC tumorigenesis in CPM (*FABP6*, *DEFA6*, *DMBT1*, *OLFM4* and *MUC2*). The expression of genes promoting cell survival and inhibiting apoptosis was upregulated (*CD53*, *TSC72D3* and *NF-κB*). Genes regulating cell growth, proliferation, transcription and angiogenesis were hypermethylated or altered via CNA or somatic mutation (*IGF2*, *EDNRB*, *ZNF461*, *ZSCAN256*, *ZSCAN18*, *POLR2C*, *APBB1*, *TAF10*, MT proteins, *PDE4DIP*, *RIOK2*, *CDC16*, *NUP98* and *CCND2*). Hypermethylation, CNA and somatic mutation resulted in the inactivation of tumour suppressors and oncogene activation in CPM (*TP53*, *VTRA2-1*, *TRIM* proteins). These changes suggest a rapid rate of tumour growth unchecked by tumour suppressor or apoptotic mechanisms in CPM.

Increased MAPK and Wnt/β-catenin pathway activation were noted in CPM. Gene expression of markers of colon stem cells and negative regulators of the Wnt pathway was reduced (*OLFM4*, *DEFA6*), negative Wnt regulators contained somatic mutations (*APC*, *RNF43*, *FAM123B* and *TSC1*). and hypermethylation of MAPK pathway marker (*RASFGFR1*) suggesting persistent activation of MAPK and Wnt pathways in CPM. Multiple mutations of negative Wnt signalling regulators make this an attractive therapeutic target. Numerous agents have been developed targeting the Wnt pathway the majority of which are still at the clinical trial stage. Porcupine inhibitors mediate the palmitoylation of Wnt ligands blocking Wnt signalling. The porcupine inhibitor LGK974 inhibits the upstream negative Wnt regulator mutant *RNF43* and would be a potential therapeutic target in the CPM cohort (338).

CPM contained a high proportion of *MSH6* somatic mutations which suggests a deficiency in the mismatch repair pathway, 46% of CPM were MSI-H with associated

high TMB median 195.59 (162-321.99). Of patients with poor prognosis 57% were MSI-H TMB-H. Of these 15% were *BRAF* mutant, it is thought that *BRAF* mutation drives the poor prognosis in MSI-H tumours. TMB-H tumours display an increased neo-antigen load and dense immune cell infiltration, they respond well to treatment with immune checkpoint inhibitors such as nivolumab and pembrolizumab (339-341).

The expression of genes regulating innate immunity was downregulated (*DEFA6*, *DMBT1*, *MUC2*) or altered via somatic mutations, (HLA-A antigen) suggesting immune evasion in the transition to CPM.

The expression of genes suppressing invasion, migration and EMT were downregulated or hypermethylated (*MUC2*, *MMP26*, *ILK*, *FLNB*, *SPTB*, *PPL*, and *SVEP1*) and those triggering these processes upregulated (*CYR61*, *CXCL12*, *CTGF*, and *CSTB*). These changes suggest a mechanism by which CPM cells metastasise from the primary CRC. In keeping with changes in EMT regulators there appeared to be a transition in CMS subtypes towards CMS4 from primary CRC to CPM. The CMS4 subtype is an interesting therapeutic target, *TGF β* signalling inhibitors and targeted immunotherapies have been trialled with success in pre-clinical models to block cross-talk between the tumour microenvironment and halt disease progression of stromal rich CMS4 CRC (61, 342).

Methylation appeared to be dysregulated in CPM with somatic mutation of the *TET2* tumour suppressor and *CDH7* chromatin regulator. Active DNA demethylation by TET enzymes is an important tumour suppressor mechanism in a range of cancer types. The *CDH7* gene is commonly associated with global hypermethylation and the CIMP phenotype. The presence of *TET2* mutations in the CPM cohort in addition to mutations of the *CDH7* gene suggests potential benefit from methylation inhibitors in addition to other cytotoxic or immunotherapy (43, 49, 327).

There was downregulation of *CES2*, a gene known to activate the prodrug irinotecan, a commonly used chemotherapy in the adjuvant treatment of primary CRC and CPM, (*CES2*). Resistance to the treatment of primary CRC may in part explain the development of CPM in this cohort.

Biological changes were noted in poor prognosis CPM distinct to those found in good prognosis CPM which may confer poor prognosis in this cohort. The expression of genes promoting transcription, cell growth, cycle, angiogenesis and inhibition of apoptosis were upregulated, (*VEZF1*, *POLR3GL*, *FEM1C*, *KHDRBS3*, *BCYRN1* and *DCAF12*). There was increased expression and hypermethylation of oncogenes, (*RELB*, *HCK*). There was hypermethylation and CNA of regulators of transcription, cell growth, cycle and angiogenesis, (*NKX6-2*, *GATA3*, *IRX5*, *ZKSCAN7*, *PAX5*, *DEAF1A*, *CHFR*, *PTPN21*, *VASH1*, *IGFBP7*, *SHB*, *PTPN21*, *FAM32* and *MELK*) These changes suggest a rapid rate of unchecked tumour growth in poor prognosis CPM.

The expression of genes regulating the innate immune evasion was upregulated, (*CEACAM1*) along with hypermethylation of genes regulating the innate immune response, (*RFTN1*, *SIT1*) suggesting immune evasion in poor prognosis CPM. Novel therapies in the form of CEA TCB IgG-based T-cell bispecific antibodies may, therefore, be of benefit in this patient cohort, they result in immune-mediated tumour lysis in tumours overexpressing *CEA* (282).

There was downregulation of gene expression of negative regulators of the Wnt pathway (*AXIN1*) and somatic mutations of key Wnt regulators (*FAM13A*). There was hypermethylation of MAPK and TGF- β pathway markers (*RAB8A*, *RAB34*, *FGF5* and *BMP3*) suggesting persistent activation of MAPK, TGF- β and Wnt in poor prognosis CPM.

There were hypermethylation and somatic mutation of genes regulating invasion, migration and EMT (*NR2F1-AS1*, *ZEB1*, *EPS8L2* and *PIEZO*) suggesting an increased invasive and migratory capacity and the development of EMT in poor prognosis CPM. There was a correlation between poor prognosis and the CMS4 mesenchymal subtype which can be targeted by several novel therapies including TGF β signalling inhibitors and targeted immunotherapies (61, 342).

There was hypermethylation of a gene known to mediate resistance to oxaliplatin, a commonly used chemotherapy in the treatment of primary CRC and CPM (*NR2F1-AS1*). Resistance to the treatment may in part explain the poor prognosis CPM.

A weakness of this study is the lack of validation of the findings. Publicly available genomic datasets were searched for comparable patient cohorts. The Cancer Genome Atlas (TCGA) contains one dataset from Yaeger et al which used MK-IMPACT, a hybridisation capture-based NGS assay targeting a limited panel of 486 genes, the dataset contains 1134 patients with metastatic CRC, 52 patients had peritoneal metastasis, the metastasis sample sequenced was peritoneal in 11, 5 of these were primary CRC and 6 metachronous (343-345). Of the mutations, Yaeger et al identified, three were shared with this CPM cohort, *TP53*, *APC* and *RNF43*, given the difference in methods used these results are not directly comparable (343-345). The Gene Expression Omnibus (GEO) contained no suitable datasets with isolated peritoneal metastasis for comparison (346). To confirm the findings, results will need to be validated on a prospective patient cohort. The transcriptome is a snapshot of cellular activity at a moment in time. Pathway activity may not be identified if it were not active at the time the tumour was removed and sectioned. It is encouraging however that the biological changes identified in the transcriptome appear to be reflected by those found in methylation and somatic mutation profiles. We must acknowledge that this is a small cohort of patients, the biological changes identified here form a starting point in identifying the tumour biology associated with the development of CPM and predicting poor prognosis disease.

In conclusion, we have identified biological changes in the transcriptome, epigenome and genome associated with the progression from primary CRC to CPM and with poor prognosis CPM. When validated and used with known clinical prognostic markers this can be used to predict patient outcome and to define new therapeutic targets for patients with CPM.

Chapter 5 Developing a bio-molecular classifier to predict response to treatment with cytoreductive surgery and heated intraperitoneal chemotherapy in patients with colorectal peritoneal malignancy

5.1 Introduction

The preceding chapter has identified changes in the transcriptome, genome and epigenome associated with the transition from primary to CPM and associated with poor prognosis following CRS & HIPEC using NGS. NGS techniques identify thousands to millions of differentially expressed features, only a small number of these have established clinical significance which can be used to predict a particular phenotype or outcome such as survival or disease recurrence, a group of such features can be defined as a biomolecular classifier.

Due to the costs of NGS, sample sizes are commonly small. As a result, the number of features identified is much greater than the number of samples. Datasets are complex and features connected via pathways (63). Mathematical methods such as machine learning have been developed to overcome these challenges and extract the significant features from NGS datasets. Biomolecular classifiers have been developed and are in clinical use for prognostication and therapeutic stratification in breast cancer (347, 348).

Several studies have used the Cox proportional hazards regression models to identify biomolecular features associated with a phenotype of interest such as poor survival or disease recurrence in primary CRC (349, 350). Due to the high dimensionality of NGS data, the Cox proportional hazards model requires large training cohorts which would be difficult to achieve given the small number of patients suitable with isolated CPM which have been resected allowing sampling. Additionally, many NGS techniques are relatively high cost meaning large cohorts require significant funding. These papers, therefore, used large datasets from The Cancer Genome Atlas (TCGA) or gene expression omnibus (GEO).

Machine learning classification models have been used for a variety of purposes in CRC, these have been applied to and validated using datasets taken from GEO. Diagnostic classifiers have aimed to differentiate with greater certainty CRC from benign or other types of cancer. Long et al applied a random forest model to GEO differentially expressed gene (DEG) datasets comparing CRC to adenoma and normal tissue to improve the diagnostic accuracy of CRC (351). Sun et al applied an SVM model to a filtered list of DEG taken from GEO datasets to differentiate CRC from other

types of cancer (352). Prognostic classifiers have aimed to identify patients with CRC at high risk of recurrence and in need of further adjuvant treatments. Zhi et al applied an SVM model to a filtered list of differentially expressed genes taken from GEO datasets in metastatic and primary CRC to identify patients at high risk of metastasis (353). Xu et al applied an SVM model to a filtered list of differentially expressed genes to identify a prognostic signature (354). The model was developed and validated using GEO and TCGA datasets, it resulted in an 15 gene signature with 77% accuracy for the prediction of disease recurrence (354). Yang applied a random forest classifier to GEO DEG datasets in patients with and without recurrence to extract a 4x gene signature capable of predicting recurrence in stage II CRC, on validation its AUC was 0.943 (355). Zhi et al applied SVM methods to five DEG GEO datasets to develop a classifier 40 gene classifier capable of predicting the development of metastasis in CRC. Predictive classifiers have aimed to identify those patients likely to respond to a particular treatment. Gan et al applied a random forest classifier to GEO datasets in patients with and without FOLFOX treatment response to predict treatment response to FOLFOX in patients with advanced CRC (356).

Public datasets containing patients with CPM are however limited with no datasets containing patients with isolated CPM in the absence of another site of metastasis. There is merit therefore in applying a machine learning classifier to the RNA and methylation datasets from chapter 4 to identify a prognostic signature capable of predicting response to CRS & HIPEC for patients with CPM.

The Biosigner package in R uses wrapper feature selection to restrict the dataset reducing the risk of overfitting and prediction variability (63). The Biosigner package combines three binary classifiers to optimise the classifier performance according to the dataset structure (PLS-DA, SVM and RF) (63). The dataset is split into training and testing subsets by bootstrapping (63). The model is trained on the training set and its accuracy determined on the holdout test set (63). Features are ranked according to their significance in the model (63). The dataset is restricted to the selected features and steps 1 – 4 are repeated until the selected list of features are stable (Figure 5.1) (63). As the training and testing subsets are not separated, (resampling between selection rounds) the performance may be over-optimistic. External validation on a new dataset is therefore required to refine the estimation of model performance (63).

Aim: To develop a biomolecular classifier (gene expression and methylation) capable of predicting prognosis for patients with CPM.

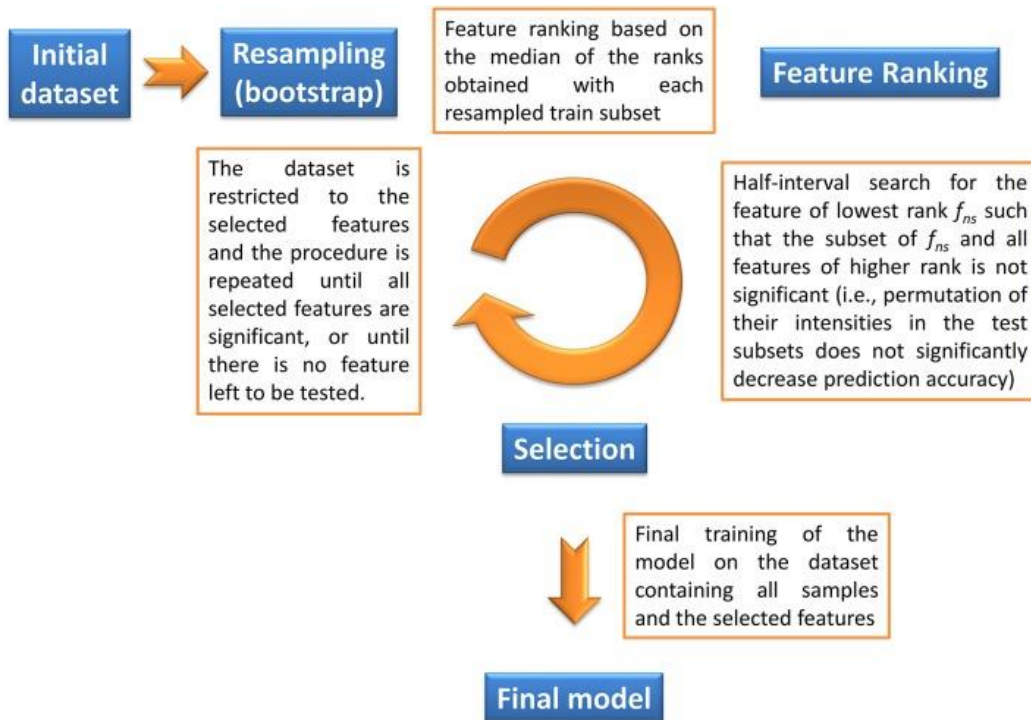


Figure 5-1 Description of the Biosigner algorithm for feature selection

Figure showing the workflow of the Biosigner algorithm. The dataset is split into training and testing subsets by bootstrapping, features are ranked according their importance, the dataset is restricted to selected features and the process repeated until the list of features remains stable. Figure reprinted from Rinaudo et al, 2016, Figure 1 (63).

5.2 Results

5.2.1 Prospective patient cohort

Prospective patients undergoing CRS & HIPEC for metachronous CPM were recruited from the peritoneal malignancy clinic between the 1st of July 2017 and the 1st of April 2018. Patients were eligible for inclusion if they had CPM (adenocarcinoma), no extra-abdominal metastasis, a CC0 resection and a PCI of < 12. Twenty-one patients were recruited.

Thirteen patients had a DFS of 14 months (13–14 range) following CRS & HIPEC and formed the 'good prognosis' cohort, eight patients had a DFS of 5 months (3-12 range) and formed the poor prognosis cohort (Figure 5.2). Whilst all patients were a minimum of 12 months post-CRS & HIPEC it is likely that the follow up for the prospective cohort is incomplete, sometimes it takes several months for data regarding recurrence or death to feedback to the tertiary CRS & HIPEC centre meaning patients classified as good prognosis may have recurred. The poor prognosis cohort had significantly more men 63% vs. 54% $p=0.018$, right-sided primary tumours 50% vs. 23% $p=0.018$, fewer N2 tumours 38% vs. 46% $p=0.042$, a higher PCI score 9 vs. 5, $p=0.04$ and patients with a CC1 resection 25% vs 0% $p=0.001$ (Table 5-1).

Table 5-1 Comparison of good and poor prognosis patient cohorts

		Good Prognosis	Poor prognosis	P-value
Age		61 +/-11.52	61 +/-11	0.85
Gender (male)		n=7 (54)	n=5 (63)	0.018
Tumour location	Right Transverse Left	n=3 (23) n=1 (8) n=9 (69)	n=4 (50) n=0 n=4 (50)	0.014
T stage primary	3 4a 4b	n=4 (31) n=5 (38) n=4 (31)	n=1 (12) n=3 (38) n=4 (50)	0.05
N stage primary	0 1 2	n=1 (8) n=6 (46) n=6 (46)	n=1 (12) n=4 (50) n=3 (38)	0.042
DFI, mean +/- SD		22 +/-10	19 +/-12	0.95
PCI, mean +/- SD		5 +/-4	9 +/-4	0.04
CC score	CC0 CC1 CC1	n=13 (100) n=0 n=0	n=6 (75) n=2 (25) n=0	0.001
Follow up, months, median (range)		8 (4-15)	8 (4-15)	1
Adjuvant treatment	Yes No	11 2	10 1	0.38
DFS median (range)		14 (13-14)	5 (3-12)	0.000
OS median (range)		15 (15-15)	10 (7-13)	0.68

(DFI, disease-free interval, PCI, peritoneal carcinomatosis index, CC score, completeness of cytoreduction, DFS, disease-free survival, OS, overall survival, SD standard deviation)

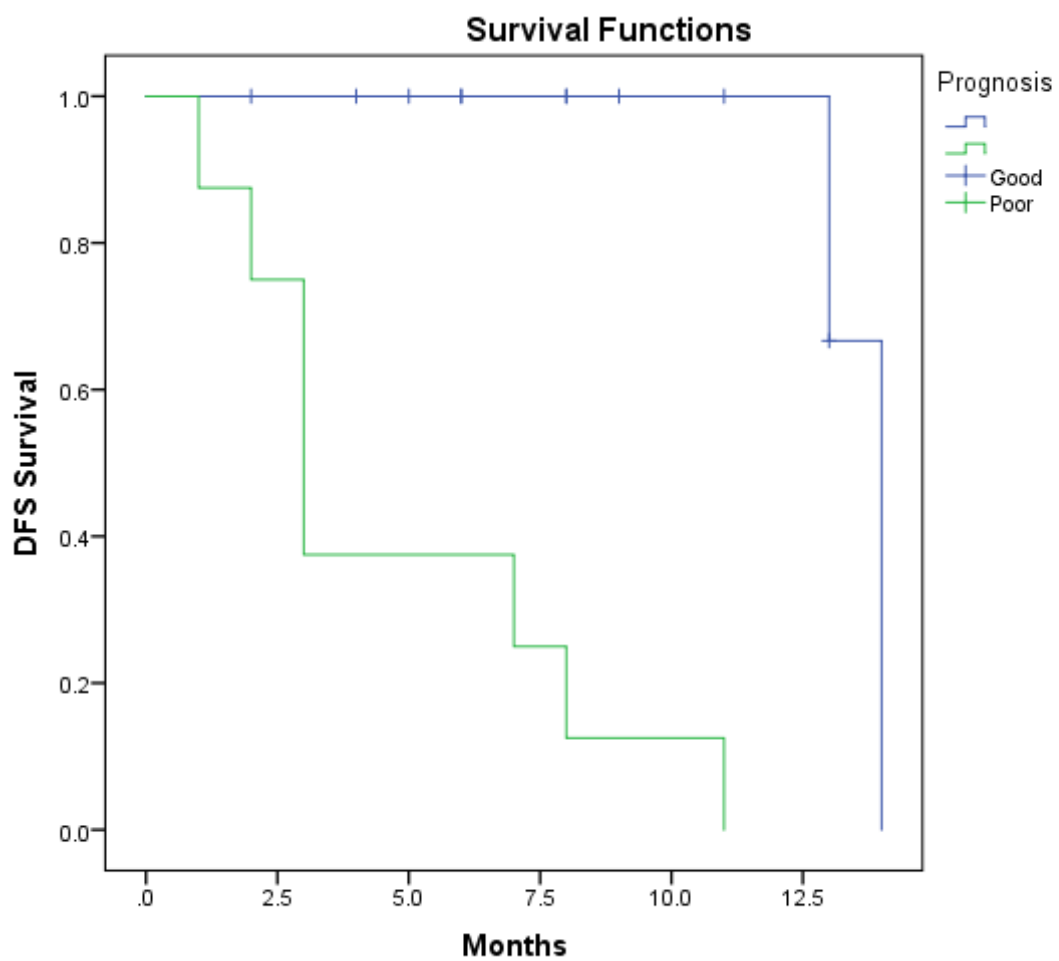


Figure 5-2 Kaplan-Meier disease-free survival analysis, good vs. poor prognosis patient cohorts

Kaplan-Meier curve showing disease free survival in patients in the prospective patient cohort with good and poor prognosis following CRS & HIPEC for CPM. Log-rank $p < 0.0001$. This figure was taken from my own results and produced in SPSS.

5.2.2 RNAseq classifier

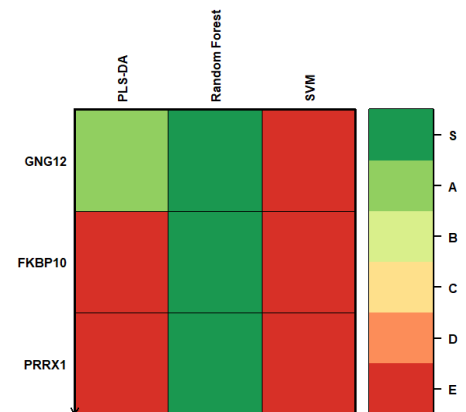
A differential gene expression prognostic signature was identified using the Biosigner method in the retrospective cohort. The dataset was split into a 60% training and 40% testing subset. The signature was unstable on repeat with the recommended bootstrapping of 50x. Bootstrapping was therefore increased in 50x increments and the procedure repeated until the classifier remained stable upon repeat at a level of 400x (357).

The significant features identified by the classifier included genes *GNG12*, *FKB10* and *PRRX1*. The resulting classifier had an area under the curve of 0.997 on the training subset. The performance of the resulting classifier on the hold out test set resulted in a classifier accuracy of 0.607 (Table 5.2 and Figure 5.3).

GNG12 is a G-protein with activity in the PI3K-AKT, MAPK and RAS pathways (261) suggesting increased tumour proliferation and growth in the poor prognosis cohort. *FKBP10* is a PPlase responsible for the acceleration of protein folding during protein synthesis. Its function in CRC has not been described in the literature. Ge et al found increased expression of *FKBP10* in Renal cell carcinoma, knockdown *in vitro* resulted in a cell-cycle arrest at G0/G1, reduced cell proliferation, invasion and migration suggesting a role in tumour proliferation and the process of metastasis (358). *PRRX1* encodes a transcription co-activator, required for the induction of genes by growth and differentiation factors (261). *PRRX1* is capable of inducing EMT in CRC cells *in vitro*, the expression of *PRRX1* in CRC is significantly associated with metastasis and poor prognosis (359). This supports the preceding RNAseq findings of an increased proportion of mesenchymal CMS4 tumours in poor prognosis CPM and may in part explain the process of EMT in this cohort.

Table 5-2 RNAseq classifier confusion table and resulting sensitivity, specificity and error rate

Gene ID	PLSA-DA	Random forest	SVM
GNG12	A	S	E
FKBP10	A	S	E
PRRX1	E	S	E
Accuracy			
Full	0.463	0.726	0.469
AS	0.901	0.996	NA
S (final)	NA	0.997	NA



	Predicted negative	Predicted positive	Total
Actual Negative	TN= 8	FP = 0	8
Actual Positive	FN = 0	TP = 7	7
Total	8	7	15
Error rate (FP + FN / total) = 0+0 / 15 = 0%			
Sensitivity, (TP / actual yes) = 7/7 = 100%			
Specificity, (TN / actual no) = 8/8 = 100%			

Classifiers PLSA-DA, partial least square discriminant analysis, RF, random forest, SVM, support vector machines. Full is the initial classifier, AS the intermediate and S the final model containing features found to be significant in all selection steps. The selected classifiers are plotted as tiers from S – E by decreasing relevance. TN, true negative, FN false negative, FP false positive, TP, true positive. Figure produced by the R Biosigner pipeline and taken from own results.

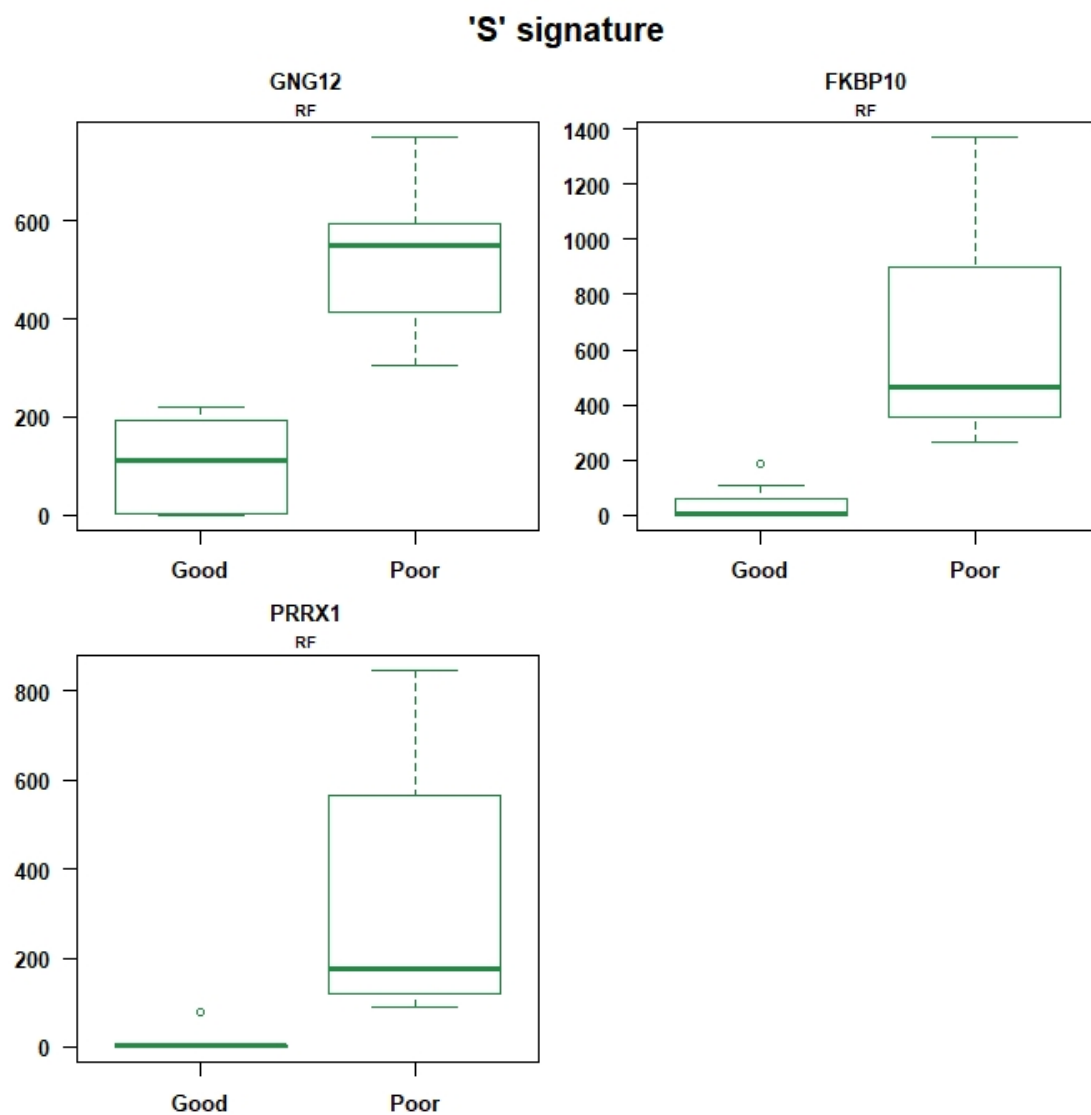


Figure 5-3 RNAseq classifier, differential expression of the features selected by the random forest classifier in the good and poor prognosis cohorts

Differential gene expression of the features selected by the random forest classifier in the retrospective good and poor prognosis cohorts. Figure produced by the R Biosigner pipeline and taken from own results

5.2.2.1 Analysis of significant classifier features in the prospective cohort

The prospective cohort was analysed for differential gene expression for the significant features identified by the retrospective classifier in patients with good and poor prognosis. The median gene count of *GNG12* was not significantly different between the good and poor prognosis cohorts at 35.96 (9.10-80.72) and 34.51 (12.2-63.28) respectively, *P*-value 0.772. The median gene count of *FKBP10* and *PRRX1* was higher in the poor prognosis cohort reflecting the retrospective classifier, however, the difference was not statistically significant. *FKBP10* was 10.99 (1.99-35.64) in the good prognosis cohort and 20.69 (4.15-34.87) in the poor prognosis cohort *P*-value = 0.11. *PRRX1* was 18.26 (3.24-75.94) in the good prognosis cohort and 33.75 (14.21-151.11) in the poor prognosis cohort *P*-value 0.09 (Figure 5.4).

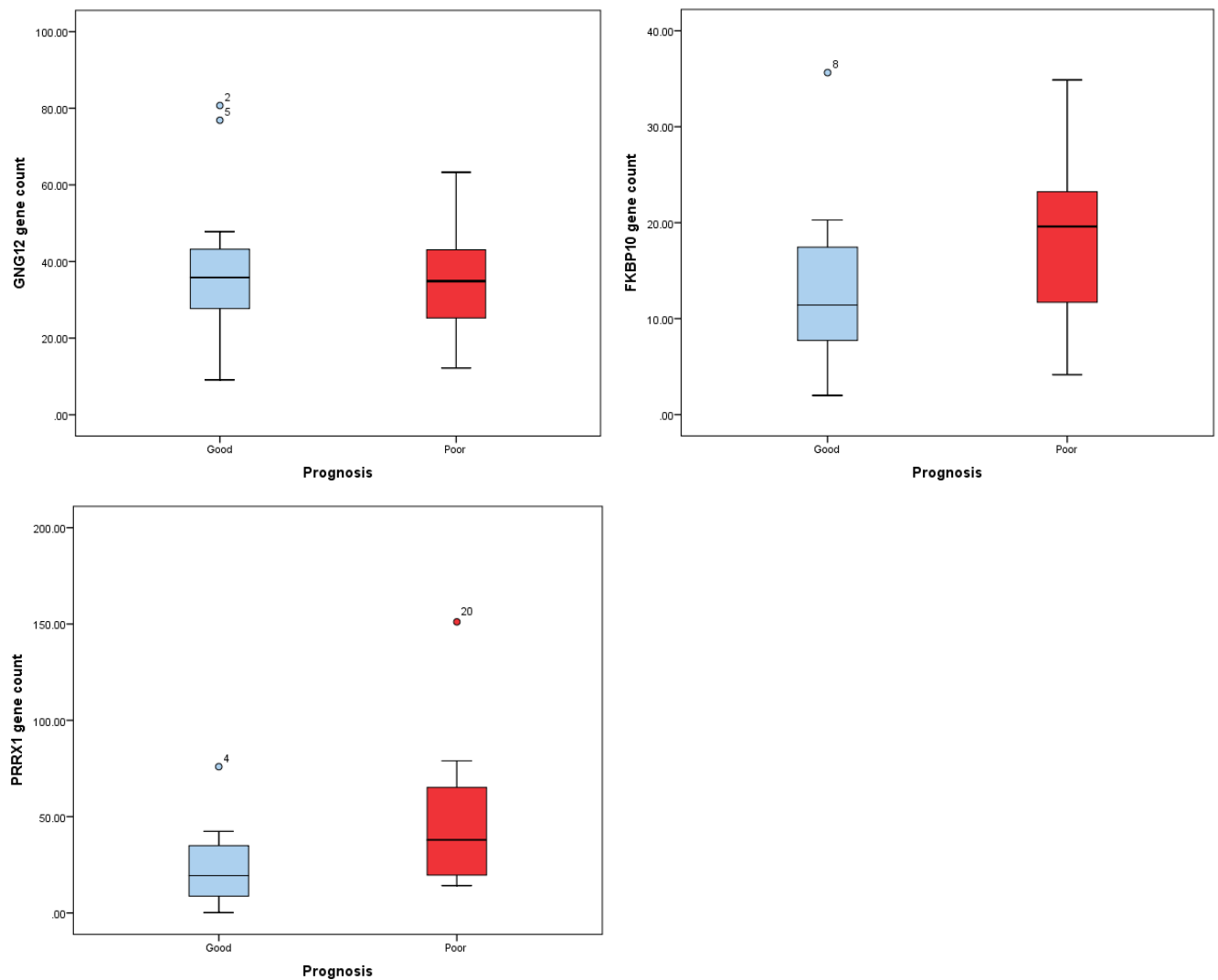


Figure 5-4 Individual boxplots for the features selected by the retrospective random forest RNA classifier in the prospective cohort

Differential gene expression of the features selected by the random forest classifier in the prospective good and poor prognosis patient cohorts. Figure produced in SPSS and taken from own results

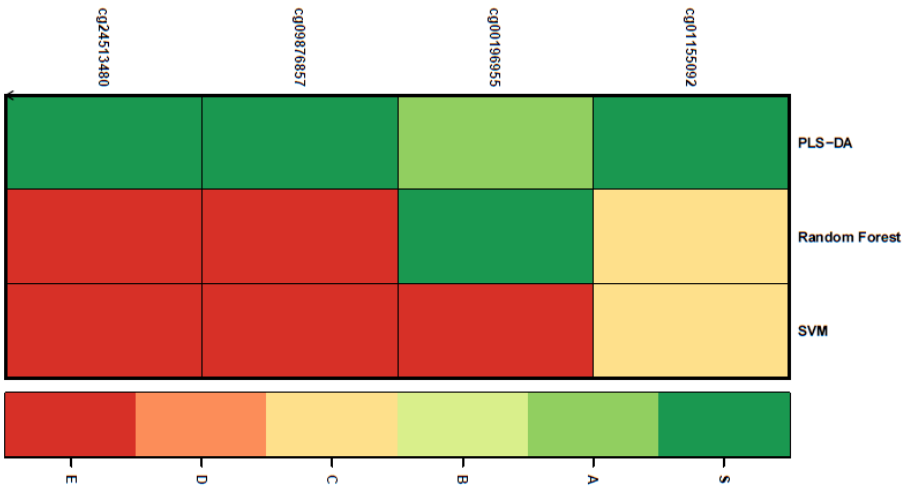
5.2.3 Methylation classifier

A differential methylation prognostic signature was identified using the Biosigner method in the retrospective cohort. The significant features identified by the classifier included CpG sites relating to the following genes *TMEM100*, *RP11-22P4.1*, *POLR2M*, *GCOM1* and *SOX2*. The resulting classifier had an accuracy of 0.939 at the training stage (Table 5.4 and Figure 5.5). This will require training, testing and validation on a prospective cohort.

TMEM100 encodes a gene essential for angiogenesis. Its function in CRC has not been described in the literature. *TMEM100* has been described as a potential tumour suppressor in lung and hepatocellular carcinoma, overexpression inhibits invasion, migration and proliferation *in vitro* (360, 361), methylation of *TMEM100* in poor prognosis CRC suggests a loss of its tumour suppressor role in this patient cohort. *METTL15* encodes an S-adenosyl-L-methionine-dependent methyltransferase, a poorly described gene whose function in CRC has not been described (261). *POLR2M* and *GCOM1* are gene paralogs encoding an RNA polymerase which negatively regulates transcriptional activation (261), methylation in poor prognosis CRC suggests an increase in transcriptional activation. *SOX2* encodes a transcription factor which induces a stem-cell state in CRC cells (362) (363). High expression of *SOX2* is associated with reduced relapse-free survival in CRC, the induction of EMT and reduced Wnt activity (362-364). The silencing of *SOX2 in vitro* has been linked with the induction of EMT in CRC. This finding again supports the increased proportion of mesenchymal CMS4 tumours in poor prognosis CPM and may in part explain the process of EMT in this cohort.

Table 5-3 Methylation classifier signature

Chr	Position		Gene ID	EPIC Cg site	PLSA-DA	Random forest	SVM
17	53810494		<i>TMEM100</i>	cg01155092	S	C	C
11	28373379		<i>METTL15</i>	cg00196955	A	S	E
15	57999170		<i>POLR2M, GCOM1</i>	cg09876857	S	E	E
3	181428046		<i>SOX2</i>	cg24513480	S	E	E
Accuracy							
				Full	0.659	0.726	0.469
				AS	0.979	0.996	NA
				S (final)	0.939	0.997	NA



PLSA-DA, partial least square discriminant analysis, RF, random forest, SVM, support vector machines. Full is the initial classifier, AS the intermediate and S the final model containing features found to be significant in all selection steps. The selected classifiers are plotted as tiers from S – E by decreasing relevance. Figure produced by the R Biosigner pipeline and taken from own results.

'S' signature

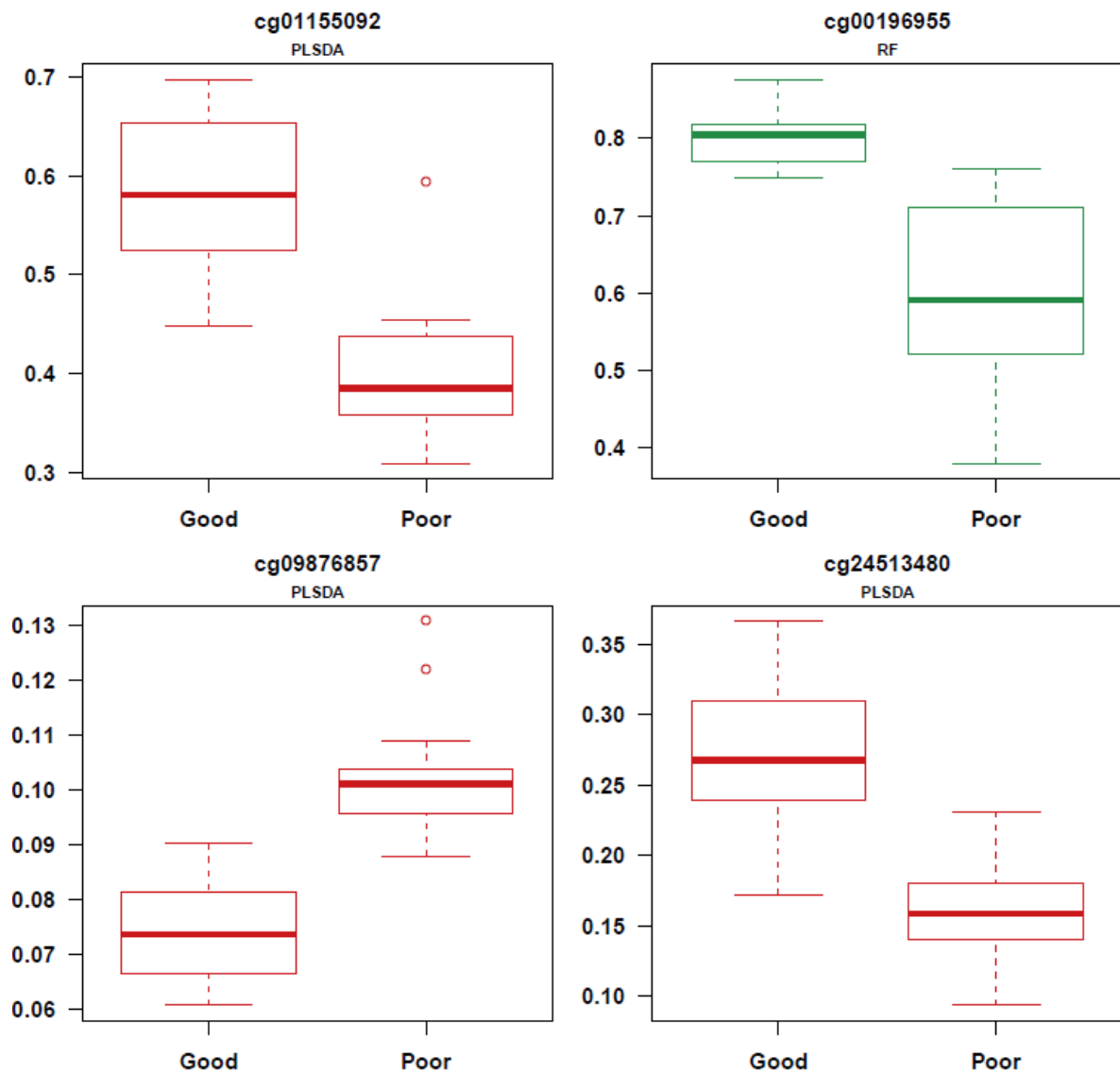


Figure 5-5 Methylation classifier, differential expression of the features selected by the RF and PLSDA classifier features in the good and poor prognosis cohorts

Differential gene expression of the features selected by the partial least squares and random forest classifiers in the retrospective good and poor prognosis cohorts. Figure produced by the R Biosigner pipeline and taken from own results

5.3 Discussion

In the present study, a machine learning algorithm was applied to determine differential gene expression and methylation prognostic classifier for patients with CPM treated with CRS & HIPEC.

The Biosigner pipeline combines three classifiers to extract and restrict meaningful signatures from complex NGS data sets. To ensure a robust classifier a high bootstrap number of 400x was used to increase feature stability. Bootstrap sampling increases the stability of the classifier and reduces the risk of overfitting. The RNAseq data was partitioned into a training and holdout testing subset. Due to the increased size and computational demands of the methylation dataset, it was not possible to do this for the methylation signature.

The RNA signature identified three gene features capable of predicting prognosis in the Random forest classifier with an accuracy of 0.607. *GNG12* a G-protein active in the PI3K-AKT, MAPK and RAS pathways, *FKBP10* a PPlase controlling cell proliferation, invasion and migration and *PRRX1* a transcription co-factor known to induce EMT. The expression of these genes was examined in a prospective cohort of patients. The expression of *GG12* was equivalent between those with good and poor prognosis. The expression of *FKBP10* and *PRRX1* was increased in the poor prognosis cohort, this difference, however, was not statistically significant. This may be a true failure of the signature to validate, this may mean there is no true gene expression signature or, more likely that the sample size was too small to establish one. It may also reflect the incomplete follow up of the prospective cohort and the inability to confidently classify the good prognosis cohort as such.

The methylation signature identified two classifiers, the random forest contained one CpG feature and the PLSDA three CpG features capable of predicting prognosis with an accuracy of 9.939 and 0.997 respectively at the model training stage. This will require validation on a second dataset. *TMEM100* is a gene with a putative tumour suppressor function controlling angiogenesis, *METTL 15* encodes a methyltransferase,

POLR2M and *GCOM1* encode RNA polymerases and *SOX2* encodes a gene capable of inducing stemness and EMT as well as downregulating Wnt in CRC.

mRNA classifiers such as Oncotype DX® and MammaPrint® are currently used clinically for prognostic classification and therapeutic stratification of adjuvant treatments in breast cancer (347, 348). Such classifiers have not yet been developed for patients with CPM. They do exist for primary CRC in the form of the CMS and CRIS classification systems though they are not in widespread clinical use as the results are not yet clinically actionable (59, 62). We know that genomic heterogeneity increases in metastatic disease, basing treatment decisions on the mutations of the primary CRC may lead to treatment failure due to tumour evolutionary adaptation. The analysis of a large cohort of patients with CPM is necessary therefore to determine the tumour heterogeneity and clonal events which occur during tumour progression and the development of CPM.

To our knowledge, this is the first gene expression and methylation prognostic classifier in patients with CPM from CRC. This study, however, has several limitations. This a small cohort of patients, the results of the classifier vary from that in the prospective cohort if this a true observation confirmed by increased follow up it suggests a high level of variance in the learning algorithm which would be corrected by increasing the size of both the training and testing cohorts. Were the follow up complete in the prospective group it may have been possible to merge the retrospective and prospective cohorts, however with potentially incomplete prospective follow up there is a risk that patients have been classified incorrectly as good prognosis, this would result in incorrect training information for the classifier. Small sample size, however, reflects the limited numbers of patients suitable for and undergoing CRS & HIPEC in the UK, Good Hope Hospital performed an average of 37 procedures per year over the 5 years the service has been established. The other UK commissioned centres include Manchester, Basingstoke and Dublin. Increasing sample size would benefit from collaboration between CRS & HIPEC centres which could be achieved as part of a clinical trial. As discussed above the follow up of the prospective cohort is not complete which limits the validation of the RNAseq classifier. Both will, therefore, require subsequent validation. This classifier captures gene expression and methylation feature signatures. This is however not the whole picture

the CPM tumour is inherently more complex. These features must be considered synergistically with the exome sequencing and clinical prognostic data. Other moderators of gene expression not captured here include post-transcriptional modifications from other species of RNA, histone modification and post-translational modifications of protein function such as phosphorylation and ubiquitination.

Future directions for this study include increasing the duration of follow up to determine the true prognosis of this cohort. Both the gene expression and methylation classifiers require validation in a larger cohort of patients to increase its stability and robustness. RNA sequencing captures a snapshot of time within a tumour's evolution, additionally, RNA expression levels vary by cell type, cell cycle and cell viability. The addition of a somatic variant classifier would increase the applicability of a prognostic classifier in CPM. This, however, is not a simple task. It can be difficult to predict the functional impact of variants at a molecular level. Cancer driver mutations co-exist with harmless passenger mutations and somatic variants are sparsely distributed in the coding regions of tumour samples (365). Somatic mutations can be used as binary data e.g. the presence or absence of a mutation, this is, however, an over-simplification as mutations and their outcomes are not identical. The development of a somatic classifier is beyond the scope of this work but is an interesting future direction.

**Chapter 6 Investigating use of circulating tumour
DNA to predict tumour recurrence and determine the
mutational burden in patients treated with curative
intent CRS & HIPEC for CPM**

6.1 Introduction

CRS & HIPEC is a treatment used with curative intent in selected patients with CPM with limited resectable disease (366). A significant number of patients, however, develop tumour recurrence following treatment. In a European cohort of 274 patients who underwent complete CRS & HIPEC for CPM, 69.3% developed a recurrence within 5-years (367). Thirty-one per cent were isolated peritoneal recurrences, 35% distant. Iterative treatment with curative intent was performed in 39.5% of patients, in this group subsequent 5-year OS was 46% (367). Iterative treatment with CRS & HIPEC is, therefore, a viable option in selected patients with isolated intraperitoneal recurrences. This in addition to other locoregional treatments for isolated distant metastasis and systemic treatments for more extensive recurrence. It is imperative therefore to detect recurrences at an early, treatable stage.

The follow up of patients following CRS & HIPEC for CPM is not currently standardised by national or international guidelines. Despite this, there is an international consensus. An international survey of nineteen surgeons performing CRS & HIPEC at the 2018 Peritoneal Surface Oncology Group International meeting found agreement in follow up schedules amongst the majority of surgeons surveyed (366). Survey respondents recommended > 2 follow up visits in the first 2 years with a physical examination, CEA measurement and abdominal and thoracic CT at least once per year for the first five years following CRS & HIPEC (366).

Contrast-enhanced CT is the imaging modality of choice in the follow up of patients post-CRS & HIPEC. The sensitivity for detecting peritoneal metastasis with CT varies widely from 25-100% with a specificity of 78-100% (368). Sensitivity is particularly reduced for deposits of less than 5 mm and those in certain anatomical locations including the root of the mesentery, lesser omentum, left hemidiaphragm and the small bowel serosa (368). Specialist radiologists in a limited number of centres can predict CPM tumour volume however there is low radiological sensitivity for patients with a low PCI score (234, 235). Recurrences, therefore, may not be detected by conventional imaging until they become widespread or of a large enough volume to preclude curative treatment. Additionally, CPM is entirely asymptomatic in the early stages,

symptoms including nausea, diarrhoea, distension, abdominal pain or bowel obstruction are late signs associated with widespread, advanced disease. Patients are therefore unlikely to present with symptoms triggering additional imaging or further investigation as a result of tumour recurrence.

One tumour marker is in clinical use for CRC follow-up following primary treatment, CEA. CEA is a cell surface glycoprotein, detectable in the serum at low concentrations in healthy adults and high concentrations in a large proportion of CRCs (281). CEA levels are elevated in numerous carcinomas including lung, breast, gastrointestinal and gynaecological (281). In the absence of systemic disease, levels normalise within weeks following primary resection (281). The sensitivity and specificity of CEA for the detection of distant metastasis can exceed 60% (281). Sensitivity, however, varies according to the site of recurrence with >70% for liver and <50% for peritoneal disease (281). There is a need, therefore, to identify an effective tumour marker for the monitoring of patients following treatment with CRS & HIPEC for CPM.

Fragments of ctDNA circulate in the bloodstream. ctDNA is an attractive option as a tumour marker. As it is composed of tumour DNA it has potential as a non-invasive, liquid biopsy, which can determine genomic and epigenetic markers allowing the delivery of targeted personalised treatments without the need for surgical resection or biopsy. Although ctDNA fragments are short (150 – 250 bp), evidence suggests they provide a comprehensive view of the tumour genome as they are released from multiple tumour regions or foci, in some cases, they have been able to detect mutations missed in corresponding tumour samples (147). To be adopted as a tumour marker ctDNA must fill several criteria defined by the Evaluation of Genomic Applications in Practice and Prevention (EGAPP) Initiative before adoption into clinical practice (147). Analytical validity, (accurate and reproducible) clinical validity, (to be able to identify patients with distinct clinical outcomes) and clinical utility, (as a result of being tested patients have improved outcomes) (147).

ctDNA levels vary according to tumour type, with higher levels in patients with CRC vs. glioma and with higher levels in metastatic than primary tumour (146). ctDNA has been used with success in the identification of minimal residual disease, response to

adjuvant therapy and recurrent disease in CRC (160, 369). This information can be used to guide closer surveillance or identify the need for additional further adjuvant therapy (160). ctDNA has been used as a prognostic biomarker in CRC, patients with high ctDNA levels at diagnosis have significantly shorter OS (160). ctDNA can be used to detect both primary and acquired resistance to systemic therapies (160). Acquired resistance to anti-EGFR therapy occurs via the acquisition of KRAS mutations which can be detected as drug-resistant sub-clones up to 10 months earlier than radiographic disease progression (160). Stopping therapy can lead to a decline in mutant KRAS clones allowing potential re-challenge in the future (160). Primary resistance to anti-EGFR therapy has been detected in KRAS wild type patients with ERBB2 amplification (160).

Aim: To determine whether ctDNA mutations correlate with those found in CPM specimens and can be used to predict poor prognosis and recurrence following CRS & HIPEC.

6.2 Results

6.2.1 Prospective patient cohort

The prospective patient cohort is described in detail in chapter 5, section 5.2.1. The same patient cohort remains the focus of this chapter. In brief, twenty-three patients were recruited, patients P1 and P2 had a macroscopic complete pathological response to neoadjuvant chemotherapy and so have no tumour specimens. Patient P10 had a prior liver transplant and so had CRS & HIPEC at the hepatobiliary surgical unit, a tumour specimen was not possible in this case. All patients have a pre-operative ctDNA sample, nineteen patients have a post-operative ctDNA sample, sixteen patients have a 3-month ctDNA sample, thirteen a six- or nine-month sample and two a nine- or twelve-month sample. Difficulties in collecting follow up ctDNA samples were particularly noted in patients who were unwell with post-operative morbidity or early tumour recurrence. As patients come from a wide geographical area in these situations, they were either an inpatient at their local hospital or were too unwell to travel to the clinic for review.

Following tumour resection with CRS & HIPEC ctDNA levels remained detectable for the majority of patients (90%). The absence of detectable ctDNA post CRS & HIPEC (4) was not predictive of improved DFS, median DFS 14 months (95% CI 9.84 – 18.16) vs. 10.20 months (95% CI 13.15 – 13) $p=0.88$ (Figure 6.1).

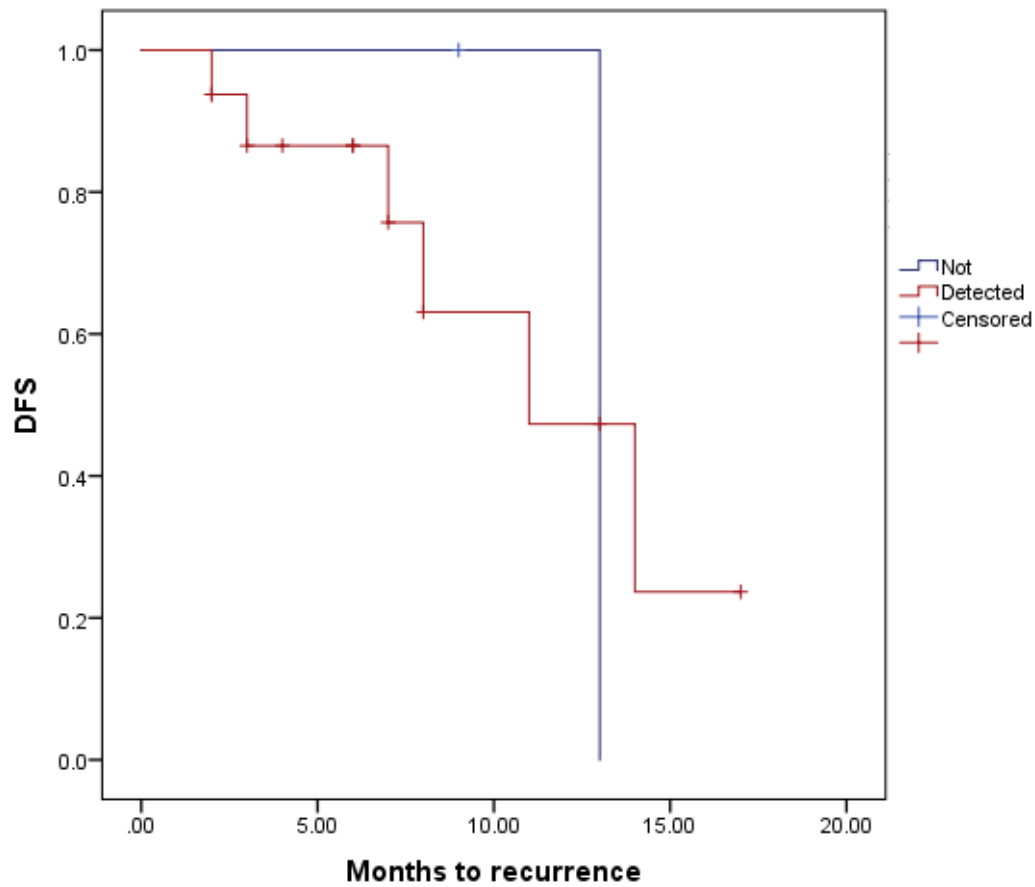


Figure 6-1 DFS according to ctDNA detection status

Kaplan-Meier curve showing disease free survival for patients with ctDNA detected post CRS & HIPEC, (10.20 months (95% CI 13.15 – 13)), and not detected, (14 months (95% CI 9.84 – 18.16)), p-value 0.88. Figure produced in SPSS and taken from my own results.

6.2.2 Hypermutant tumour

One patient had significantly higher levels of mutation than others (P1). This patient had a metachronous CPM at the site of their previous sigmoid tumour, they had an apparent complete pathological response to neoadjuvant single-agent capecitabine chemotherapy with no tumour specimen at CRS & HIPEC. Despite complete pathological response baseline ctDNA levels at the time of CRS & HIPEC showed 1964 non-synonymous variants, this may suggest high TMB (this would need to be confirmed on WES) and eligibility for PD-1 monoclonal antibody blockade. At CRS & HIPEC the anastomosis was resected with a pelvic peritonectomy and omentectomy. The complete pathological response was confirmed on histology. Following CRS & HIPEC the number of non-synonymous variants fell to 26 at 4 weeks, adjuvant chemotherapy was completed with 6x cycles of FOLFOX. Asymptomatic recurrence was detected 7 months post-CRS & HIPEC on CT and the chemotherapy changed to FOLFIRI. Preceding the radiological diagnosis of recurrence, the number of non-synonymous variants increased to 245 at 6 months. The high levels of non-synonymous mutation despite apparent complete pathological response has not been described in the literature but may suggest a hypermutant subclone that expands.

This patient has a hypermutated CRC. Hypermutation is highly correlated with MSI. Three per cent of MSI tumours are associated with Lynch syndrome and 12% result from sporadic mutations of the DNA mismatch repair system (370). MSI tumours express various immunological checkpoint inhibitors such as PD-1 and may respond well to immune checkpoint inhibitors such as pembrolizumab and nivolumab (370). In addition to MMR dysfunction, mutations in DNA polymerases POLD1 and POLE genes are known to result in hypermutation (370). Analysis of variant consequences for P1 identified potentially damaging mutations of high functional consequence in ARID1A, MSH1, MSH2 and MSH6 suggesting dysfunction of the MMR pathway resulting in hypermutation (371). No germline variants were detected in any Lynch syndrome associated genes in this patient.

6.2.3 ctDNA mutation number significantly higher than tumour

There was a weak negative correlation between tumour and ctDNA mutation number, Pearson's correlation -0.302 P-value = 0.22 (Figure 6.3). In the whole cohort the

mutation number detected in baseline ctDNA was significantly higher than that detected in the tumour for each gene sequenced with a median for mutations across all genes of 47 (3-1963 range) vs. 10 (1-27 range) P -value = 0.0001 (Figure 6.4). The difference in mutation detection cannot be explained by the sufficiency of coverage. Tumour samples had a higher input of DNA at 40ng vs. 2ng for ctDNA and a resulting higher mean coverage at 1685x (293x-6941x) vs. 103x (10x-1206x). False-positive mutation calls increase with decreased sample input amounts, the input of DNA for the ctDNA samples was lower than matched tumour however several features of the technique used mitigate for this. DNA fragments are bound to unique molecular indices before PCR meaning only those which correctly bind to a UMI are amplified by PCR. In addition, during variant filtering, only alleles with a frequency above the background error rate and high confidence are selected and false calls filtered.

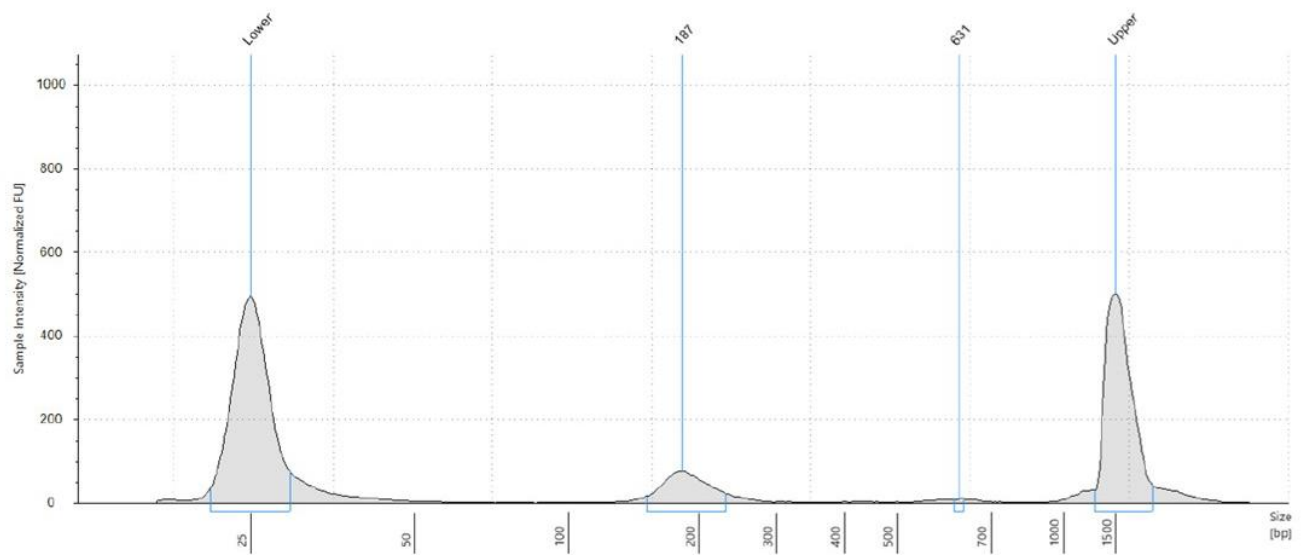


Figure 6-2 Sample ctDNA tape station electropherogram

Electropherogram showing the expected ctDNA fragment size of 150 – 250 bp. Markers indicate lower marker, upper marker and the average size of the nucleic acid sample in base pairs. This figure has been taken from my own results.

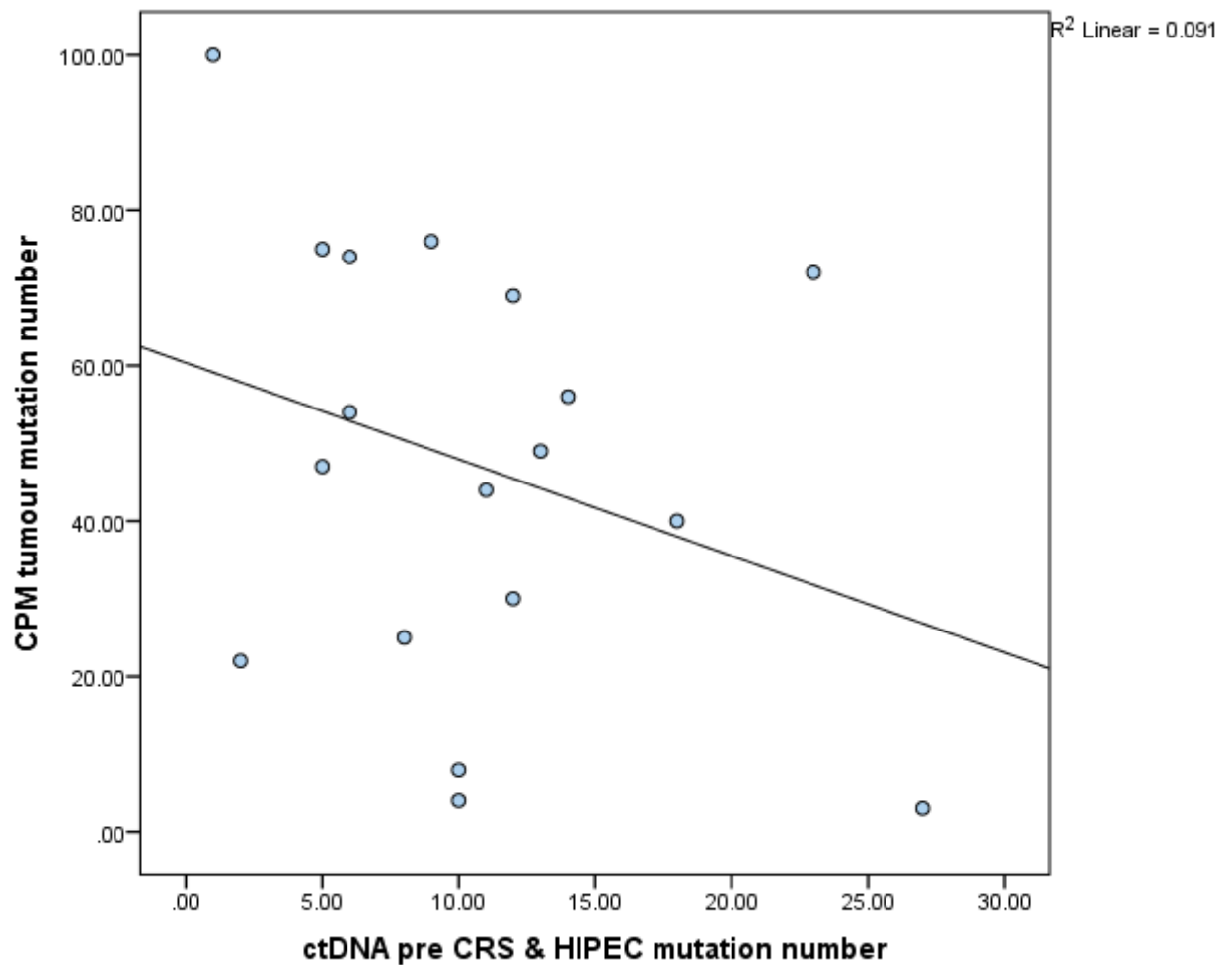


Figure 6-3 Correlation between mutation number in CPM tumour and baseline ctDNA samples

Pearson's correlation showing a weak negative correlation in mutation number between CPM tumour and matched ctDNA samples. Pearson's correlation $-.302$, P -value = 0.22 . This figure has been taken from my own results.

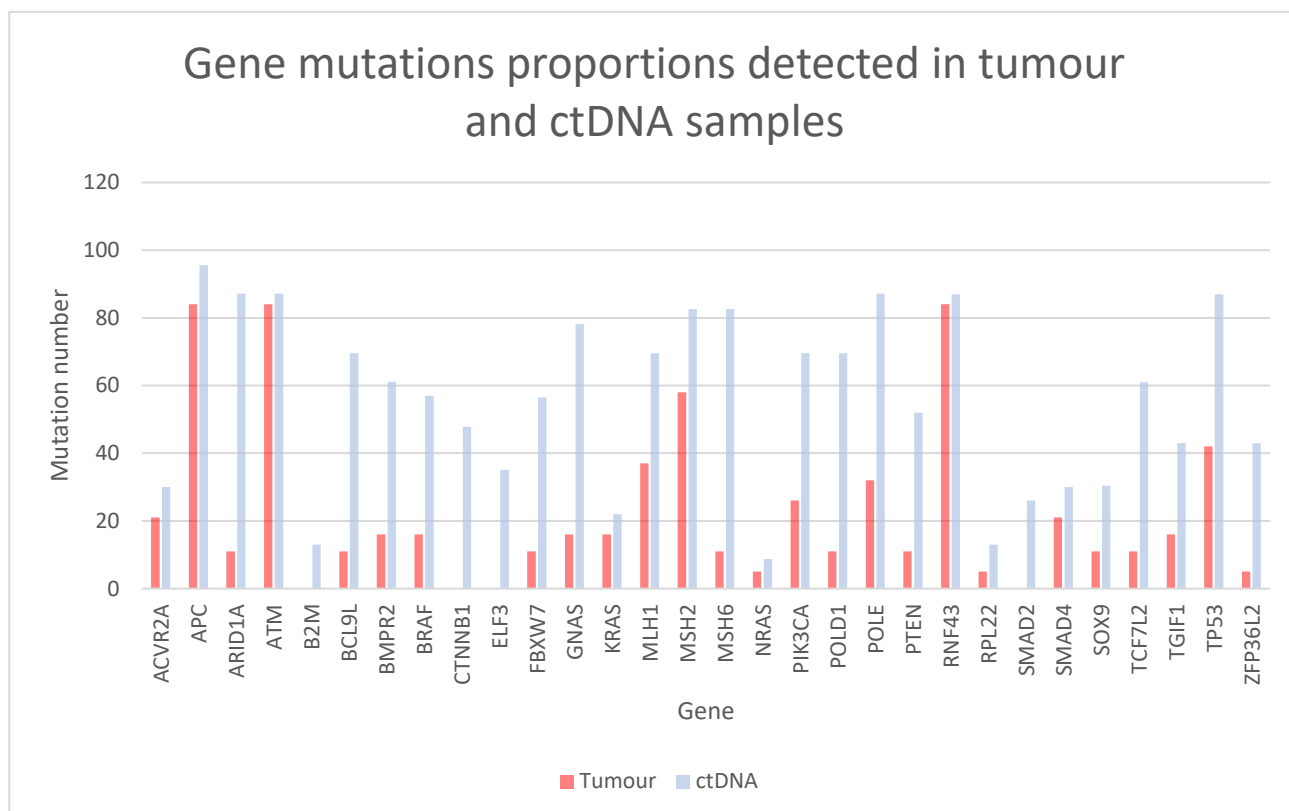


Figure 6-4 Correlation between mutation number in CPM tumour and baseline ctDNA samples

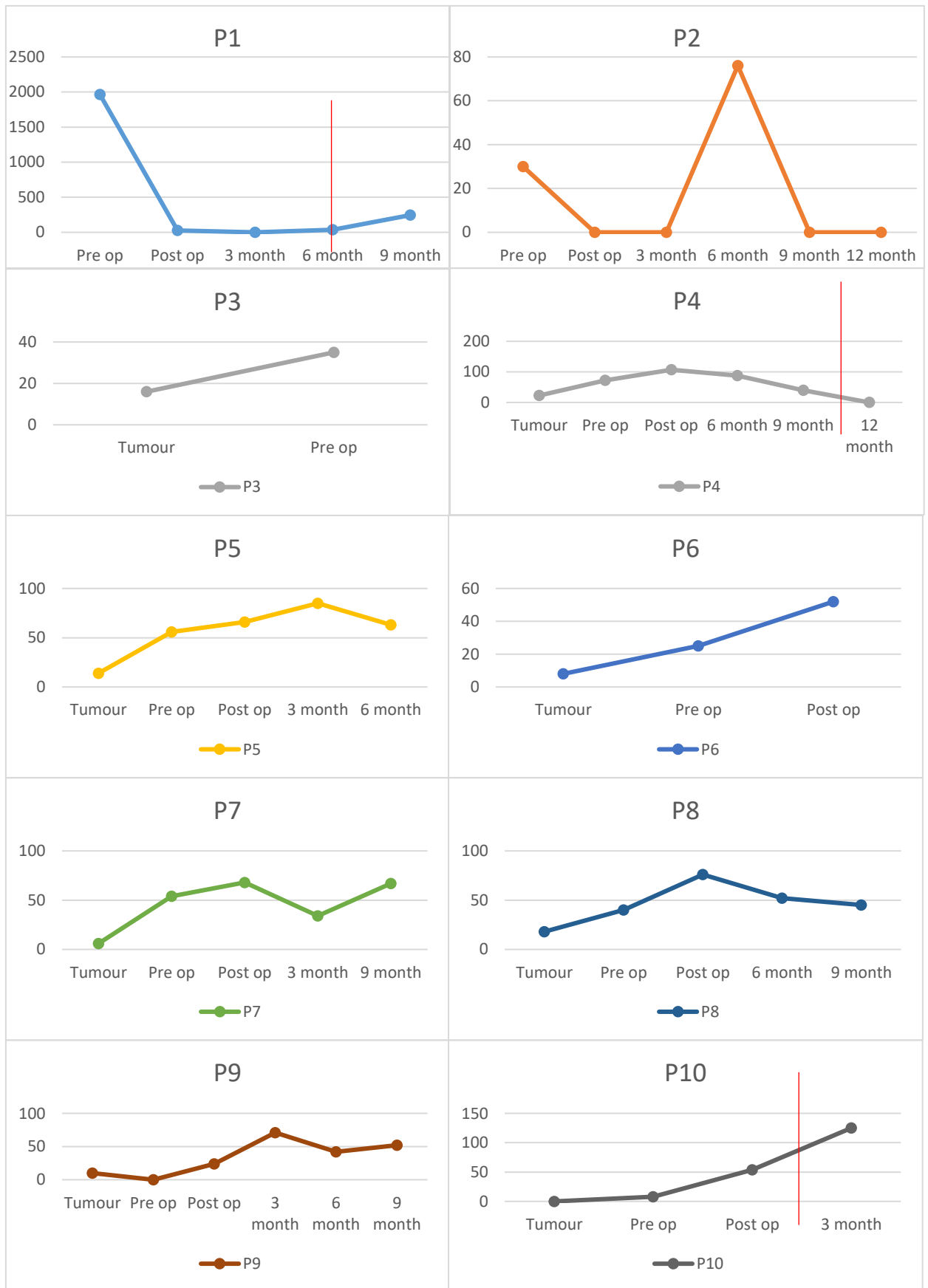
Number of non-synonymymous mutations in CPM tumour sample (red) and paired ctDNA sample (blue) for each gene sequenced in the QIAseq™ gene panel.

6.2.4 ctDNA remained detectable following complete CRS & HIPEC

The half-life of ctDNA is less than 2 hours (372). ctDNA was collected following CRS & HIPEC at a minimum of 5 days post-resection, median 38 days (5-84). All patients had a complete cytoreduction, 90% CC0 meaning no residual disease following CRS and 10% CC1 meaning <2.5mm residual disease following CRS amenable to treatment with HIPEC. Despite this ctDNA levels remained detectable following CRS & HIPEC and were marginally higher than pre-CRS & HIPEC levels, 0.93 ng/ μ L (0-1.04) vs. 0.54ng/ μ L (0.128-5.34), *P*-value 0.08. This trend was reflected in the mutation number 52 (0-107), vs. 47 (3 – 1963), *P*-value 0.39.

6.2.5 The use of ctDNA as a biomarker of tumour recurrence

Several limitations mean it was not possible to determine whether increased ctDNA levels or mutation number can be used to predict tumour recurrence. Firstly, there was a limited number of ctDNA samples per patient due to loss to follow up, this was particularly noted in patients with post-operative or adjuvant treatment morbidity or recurrence. Secondly, this was a small cohort of patients with a limited number of recurrences (3) during the period of ctDNA sampling. Finally, the follow up is incomplete, whilst for all patients, it is a minimum of 12 months following CRS & HIPEC, the median follow up with clinic contact is just 8 months (4-15). As CRS & HIPEC is performed at a tertiary referral centre it may take several months for data regarding recurrence or death captured by the patient's local hospital to be fed back. Of the three patients with recurrences during the period of ctDNA sampling 2 (P1 and P10) showed increased levels of ctDNA preceding the diagnosis of tumour recurrence (Figure 6.5). These numbers are too small to draw any conclusions.



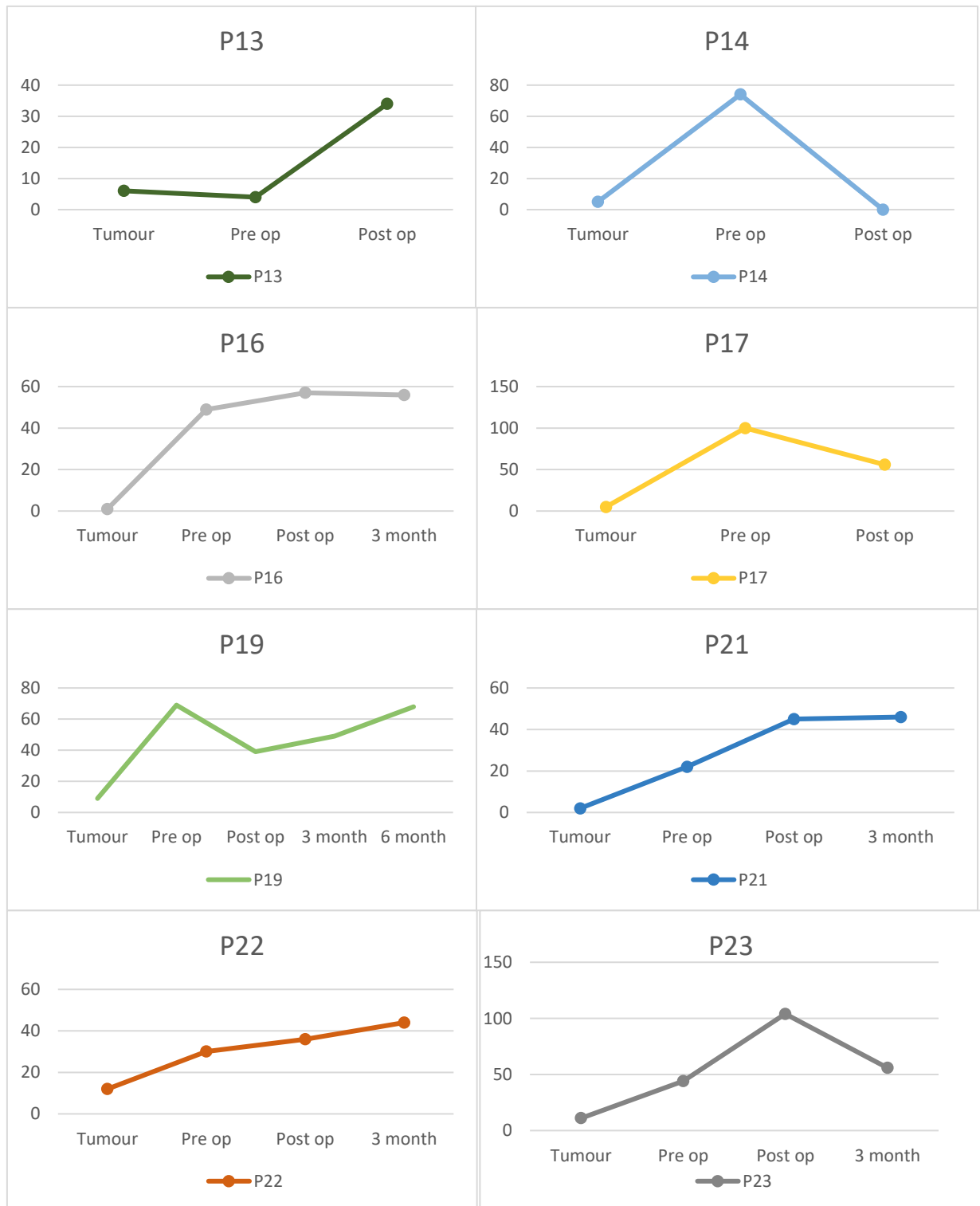


Figure 6-5 Mutation number tracking from CRS & HIPEC to final follow up for individual patients

Mutation number at each ctDNA timepoint sampled. All patients with two or more samples are shown, re-occurrences within the ctDNA sampling timeframe are indicated with a vertical red line

6.2.6 Mutations detected in ctDNA

The QIAseq™ gene panel in use in our laboratory targets 30 genes of known significance in CRC. The mutations commonly detected in ctDNA the CPM cohort are discussed below (Table 6.1). The location and distribution of key CRC gene mutations (*APC*, *RNF43*, *POLE*, *POLD*, *BRAF* and *KRAS*) are displayed in lollipop plots, Figures 6.6-6.8).

High levels of mutations were found in key genes of the Wnt pathway. *APC* mutations were present in 96% of patients, Inactivation of the *APC* tumour suppressor gene results in inappropriate and consistent activation of the Wnt signalling pathway. *APC* is mutated in the majority of sporadic CRC (23). *BCL9L* and *TCF7L2* mutations were present in 70% and 61% of patients. *BCL9L* and *TCF7L2* are beta-catenin co-factors which form a transcriptional activation complex and Wnt activation. *RNF43* mutations were present in 87% of patients, *RNF43* is an upstream negative Wnt regulator which can be targeted by porcupine inhibitors (261). *PTEN* and *PI3K* mutations were present in 52% and 70% of patients respectively. *PTEN* is a negative regulator of *PI3K*, *PI3K* inhibits the activity of *GSK-3B* which results in B-catenin accumulation and Wnt activation. Mutations in *PI3K* are associated with resistance to anti EGFR therapies in the absence of *EGFR* mutation suggesting they may drive aberrant *EGFR* signalling (38).

ARID1, *ATM*, *MLH1*, *MSH2*, *MSH6* contained mutations in 87%, 87%, 70%, 83% and 83% of patients respectively. *MLH1*, *MSH2* and *MSH6* are important DNA mismatch repair genes, *ARID1* recruits *MSH2* to chromatin during DNA replication promoting MMR (261). *ATM* coordinates DNA repair by activating enzymes to fix broken strands (261). Such high levels of mutations in the MMR pathway suggest MSI. *POLE* and *POLD1* mutations were present in 87% and 70% of mutations respectively. *POLE* and *POLD1* DNA polymerases are associated with hypermutant CRCs (370).

Table 6-1 Gene panel mutation proportions detected in baseline ctDNA in all patients and by prognostic group

	All patients (%)	Poor prognosis (%)	Good prognosis (%)	Fishers exact Good vs. Poor P-value
ACVR2A	30	6 (75)	5 (33)	0.006*
APC	96	7 (88)	15 (100)	0.35
ARID1A	87	6 (75)	14 (93)	0.29
ATM	87	6 (75)	14 (93)	0.29
B2M	13	1 (13)	4 (27)	1
BCL9L	70	4 (50)	12 (80)	0.18
BMPR2	61	3 (38)	12 (80)	0.07
BRAF	57	4 (50)	10 (67)	0.40
CTNNB1	48	4 (50)	8 (53)	1
ELF3	35	2 (25)	7 (47)	0.40
FBXW7	57	5 (63)	9 (60)	1
GNAS	78	4 (50)	13 (87)	0.13
KRAS	22	1 (13)	5 (33)	0.37
MLH1	70	5 (63)	9 (60)	1
MSH2	83	6 (75)	13 (87)	0.59
MSH6	83	7 (88)	12 (80)	1
NRAS	9	1 (13)	1 (7)	1
PIK3CA	70	4 (50)	11 (73)	0.37
POLD1	70	5 (63)	11 (73)	0.66
POLE	87	8 (100)	13 (87)	0.53
PTEN	52	5 (63)	7 (47)	0.67
RNF43	87	7 (88)	13 (87)	1
RPL22	13	2 (25)	2 (13)	0.59
SMAD2	26	3 (38)	4 (27)	0.62
SMAD4	30	3 (38)	5 (33)	0.62
SOX9	30	3 (38)	5 (33)	1
TCF7L2	61	4 (50)	11 (73)	1
TGIF1	43	3 (38)	7 (47)	1
TP53	87	8 (100)	12 (80)	0.53
ZFP3	43	4 (500)	6 (40)	0.67

Comparison with Fisher's exact test. Values in parenthesis are percentages, *significant P-value.

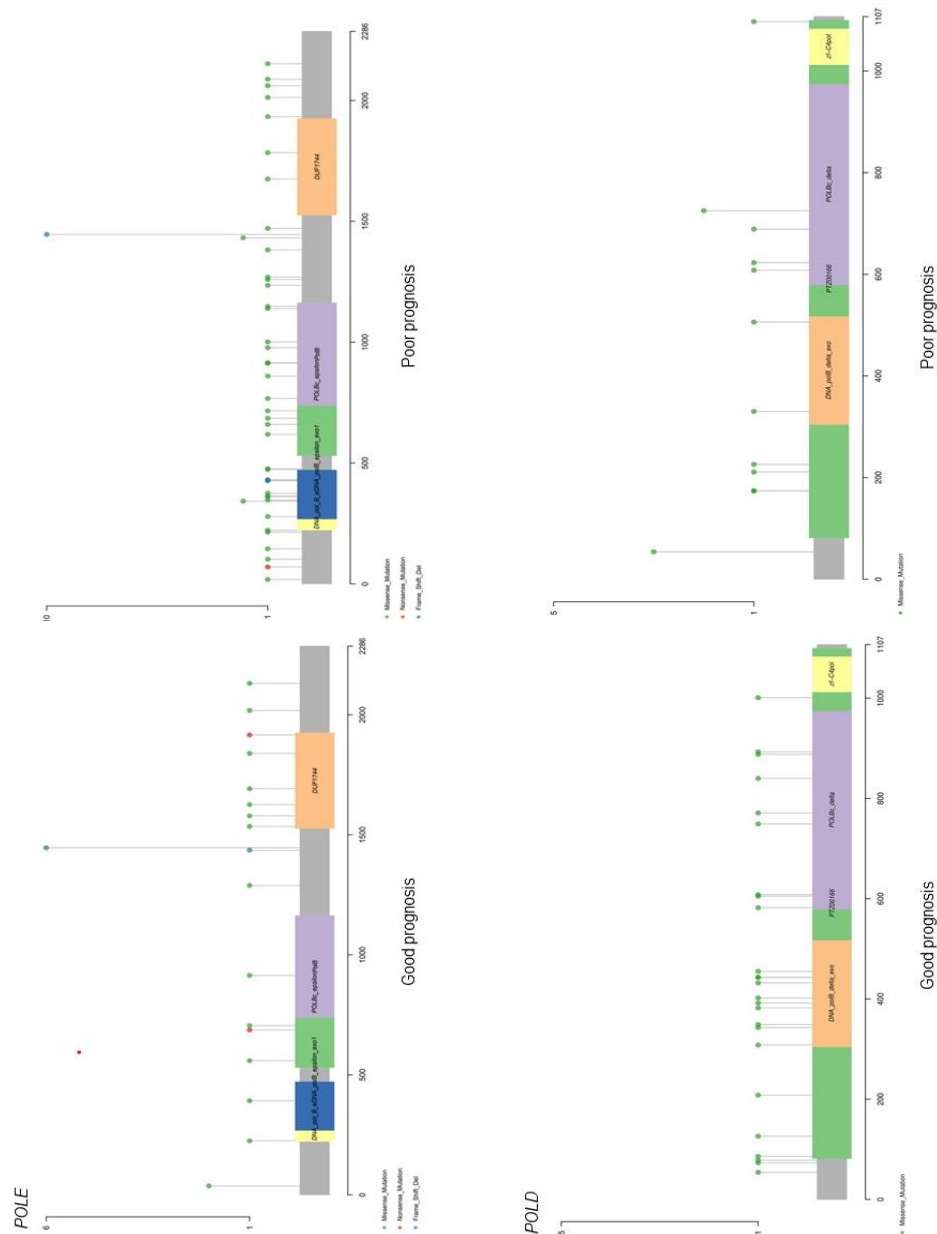


Figure 6-7 Lollipop plots showing the location and distribution of mutations in the *POLE* and *POLD* genes in good and poor prognosis ctDNA samples

Variants relative to a schematic representation of the gene. Any position with a mutation obtains a circle, the length of the line depends on the number of mutations detected at that codon. The grey bar represents the entire protein with the different amino acid positions.



Figure 6-8 Lollipop plots showing the location and distribution of mutations in the *BRAF* and *KRAS* genes in good and poor prognosis ctDNA samples

Variants relative to a schematic representation of the gene. Any position with a mutation obtains a circle, the length of the line depends on the number of mutations detected at that codon. The grey bar represents the entire protein with the different amino acid positions.

BRAF mutations were present in 57% of patients, *BRAF* is a serine/threonine kinase downstream of *KRAS* in the MAPK signalling pathway (22). *BRAF* mutation is associated with hypermutation, CIMP and MSI (22). *BMPT2* was mutated in 61% of CRC. *BMPT2* mutation is thought to contribute to CRC development by the inactivation of BMP/ SMAD signalling (373). *FBXW7* and *p53* were mutated in 57% and 87% of patients. Both *FBXW7* and *p53* are important tumour suppressors in CRC. *GNAS* mutations were present in 78% of patients. *GNAS* can dysregulate several pathways of importance in CRC including Wnt, hedgehog and TGF- β (261).

The numbers of non-synonymous mutations in each gene were compared in the good and poor prognosis groups using the Fishers exact test. The proportion of patients with and without gene panel mutations was similar across the prognostic groups with significantly more frequent mutations of *ACVR2A* only in the poor prognosis cohort (Figure 6.9).

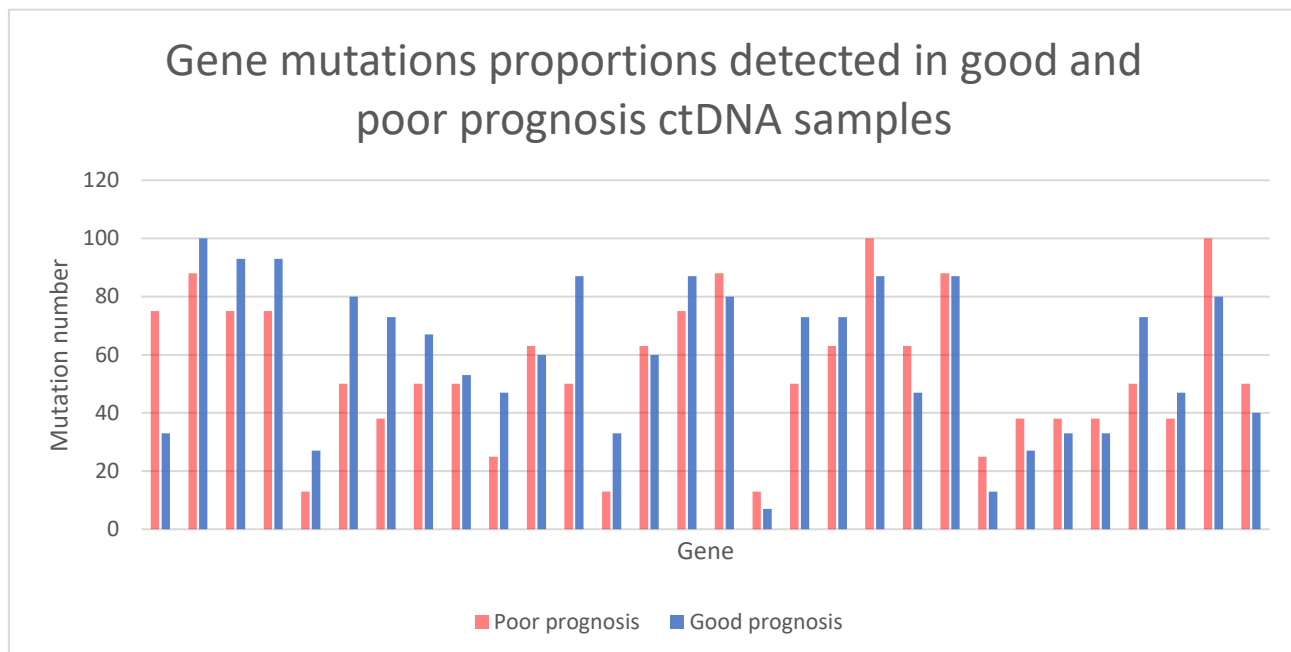


Figure 6-9 The proportion of gene mutations detected in baseline ctDNA samples before CRS & HIPEC in patients with good and poor prognosis

Number of non-synonymous mutations in good prognosis (blue) and poor prognosis (red) CPM tumour samples for each gene sequenced in the QIAseq™ gene panel.

6.3 Discussion

In this study, ctDNA samples had significantly higher mutational burden when compared with matched CPM tumour samples. This may reflect the origin of ctDNA fragments from multiple tumour sites which have not been fully captured by the tumour sample. Analysis of tumour samples gives a small sample of cells at a given time which gives a low temporal and spatial resolution of the tumour as a whole. ctDNA may, therefore, have several advantages over tumour sample or biopsy allowing the assessment of inter tumour heterogeneity and subclonal mutations which in the literature have been used to predict treatment response, the development of resistance to therapies necessitating a change in treatment and even allow re-treatment when resistant subclones decline (374-377).

We were unable to draw any conclusion with regards to the utility of ctDNA as a biomarker of tumour recurrence. This may reflect the small volume of residual disease left following CRS & HIPEC. The literature, however, suggests ctDNA can be used as a marker of tumour recurrence following the treatment of both primary and metastatic CRC. Reinert et al examined the temporal presence of somatic structural variants, (SSV) in a cohort of patients post-primary CRC resection, they found SSV in ctDNA were capable of predicting relapse in 100% of patients at a mean of 10 months before the detection of metastatic recurrence with conventional follow up (369). Zhou et al compared the utility of ctDNA variant allele frequency to detect CRC disease relapse compared to CEA, cancer antigen 19-9 and imaging (378). They found ctDNA mutations decreased following resection or adjuvant treatment and increased 13-fold upon disease recurrence despite normal CEA and CA19-9 (378). Thierry et al examined the concordance between *KRAS*, *NRAS* and *BRAF* mutations between tumour and ctDNA samples in mCRC (379). They found ctDNA was able to detect a higher proportion of mutations than tumour biopsy and was quicker to process than tumour suggesting it could replace tissue biopsy to determine clinically actionable mutations in patients with mCRC (379). Kidess et al examined temporal ctDNA mutations in patients with CRC liver metastasis undergoing resection (380). They found ctDNA samples captured additional mutations not detected in tumour tissue and

that ctDNA anticipated tumour recurrence earlier than CEA and imaging (380). Ng et al examined temporal ctDNA following the primary resection of CRC (381), they found ctDNA detection correlated with the early detection of metastasis (381).

Following complete CRS & HIPEC levels of ctDNA increased marginally. The post-CRS & HIPEC sample was taken in the clinic at between 4-8 weeks post-resection, at a median of 57 days (IQR 46-75). This rise in ctDNA occurred in patients with samples taken in the first 7 days post-resection as well as those taken later suggesting this is not a reflection of early disease recurrence. The impact of CRS & HIPEC on ctDNA has not been reported in the literature for patients with CPM. One hypothesis is that this represents a transient rise in novel mutations as a reaction to de novo treatment with Mitomycin which is not used in the systemic treatment of CRC (382). An alternative hypothesis is that ctDNA originates from multiple tumour sites and may be able to predict a more comprehensive picture of intra-tumour clonal heterogeneity than a single site tumour sample (147). Once apoptosis is induced fragmented cellular contents are packaged and efficiently cleared by macrophages, in the case of chemotherapy for cancer the high volume of apoptosis may engulf the capacity of local phagocytes resulting in the release of fragmented DNA into the circulation (383). No studies have analysed the effects of topical chemotherapy on ctDNA levels, it is likely that HIPEC results in tumour cell apoptosis and shedding over several days which may explain a peak in levels following CRS & HIPEC. A study by Tie et al found in cases of incomplete surgical resection or tumour debulking there followed a rise in ctDNA levels which may reflect surgically induced tissue injury, this is also a potential in cases of CRS where small areas of tumour are left behind (384).

The absence of detectable ctDNA following resection was not predictive of improved DFS in this study. This is not in keeping with the literature and is likely a reflection of a small cohort of patients, of which 4 had undetectable ctDNA post-CRS & HIPEC and just one of these recurred within the timeframe of ctDNA collection.

Within this targeted gene panel, patients with CPM had frequent mutations of Wnt pathway genes, *APC*, beta-catenin co-factors *BCL9* and *TCF7L2*, upstream Wnt regulator *RNF43*, *PTEN* and *PI3K*. Mismatch repair pathway genes including *ARID1*, *ATM*, *MLH1*, *MSH2* and *MSH6* were commonly mutated suggesting microsatellite

instability. DNA polymerases *POLE* and *POLD1* were commonly mutated which in addition to MMR defects can result in hypermutant CRCs. Other common mutations included *BRAF*, *BMPT2*, *FBXW7*, *TOP53* and *GNAS*.

There were significantly more frequent *ACVR2A* mutations in the poor prognosis cohort. The prognostic cohorts had otherwise similar mutation frequency across the gene panel. *ACVR2A* is part of the TGF- β signalling pathway. Zhou et al examined a cohort of 497 patients and found reduced *ACVR2A* expression correlated with advanced stage, metastasis, reduced survival and increased cell migration *in vitro* (385).

This study is the first to our knowledge to examine the use of ctDNA as a biomarker of tumour recurrence in patients who have undergone CRS & HIPEC for CPM. This study has several limitations. This is a small patient cohort with significant loss to follow up in cases of tumour recurrence and post-operative morbidity limiting patient's clinic attendance. This means we are unable to draw any conclusion with regards to the utility of ctDNA as a biomarker of tumour recurrence. The follow up of this cohort is incomplete, a number of patients likely have asymptomatic recurrence which will be reported at a later date meaning the distinction between good and poor prognosis cohorts is unclear. Finally, the study contains no normal controls. The panel used looks for mutations of known importance in CRC that are unlikely to be found in circulating cell-free DNA of non-tumour origin. In extracting ctDNA the blood sample was centrifuged twice to separate the plasma from the serum within 24 hours of collection limiting the contamination of white blood cell cfDNA, we know however that there is a high correlation between background mutations found in white blood cells and those found in cfDNA meaning this can be a confounding factor (386).

Future directions for this study include increasing the duration of follow up to determine the true prognosis of this cohort and capture any further patients with poor prognosis and recurrence occurring before 12 months post-CRS & HIPEC. To determine the utility of ctDNA as a biomarker of tumour recurrence for patients with CPM this study would need to be repeated on a larger cohort with a longer follow up period and a group of normal controls. A reduction in the loss to follow up may be achieved by local ctDNA collection through the patient's local oncology or surgical clinic. Though this

may prove logistically challenging unless part of a larger clinical trial with onsite research support.

Chapter 7 Discussion

7.1 Interpretation and implications

7.1.1 Prognostic factors for patients with colorectal peritoneal metastasis undergoing Cytoreductive Surgery and Heated Intraperitoneal chemotherapy

In this thesis, I firstly identified current prognostic factors for patients with CPM suitable for treatment with CRS & HIPEC using systematic review and meta-analysis. These factors were obstruction or perforation of the primary tumour (HR 2.91, 95 % CI 1.5 to 5.65), the extent of peritoneal metastasis as described by the PCI (per increase of 1 PCI point: (HR 1.07, 95 % CI 1.02 to 1.12) and the completeness of resection as described by the CC score (CC score above zero: HR 1.75, 95 % CI 1.18 to 2.59). These prognostic factors are commonly used in patient selection for CRS & HIPEC. Despite this survival still varies widely and can be difficult to predict.

7.1.2 Describing the biology of colorectal peritoneal malignancy

I therefore secondly sought to describe the biology of CPM, to identify the biological changes in gene expression, CNA, methylation and somatic mutation associated with the transition from primary CRC to isolated CPM and conferring a poor prognosis in response to the best available treatment CRS & HIPEC. Each CPM sample was matched to a germline normal and primary CRC sample to identify somatic variants and those changes unique to metachronous CPM.

Biological changes identified in primary CRC were subtracted from matched CPM samples to identify changes unique to and associated with the progression from primary CRC to isolated CPM. A rapid rate of unchecked tumour growth was suggested by upregulation, hypermethylation, CNA or somatic mutation of genes promoting cell survival, the inhibition of apoptosis, regulating cell growth, proliferation, transcription and angiogenesis (*CD53*, *TSC72D3*, *NF-Kb*, *IGF2*, *EDNRB*, *ZNF461*, *CDC16*, *NUP98*, *CCND2*, *TP53*, *VTRA2-1* and TRIM proteins). The MAPK and Wnt/ β -catenin pathways were persistently activated in CPM with decreased expression and somatic mutation of Wnt regulators (*OLFM4*, *DEAFA6*, *APC*, *RNF43*, *FAM123B*, and *TSC1*) and hypermethylation of MAPK marker (*RASFGFR1*). Negative Wnt regulators such as *RNF43* are an attractive therapeutic target for treatments such as the porcupine inhibitor LGK974 in the CPM cohort (338). The CPM cohort had a high proportion of MSI-H, mismatch repair mutations and TMB. Of the patients with poor prognosis 57%

were MSI-H, TMB-H, of these 25% were *BRAF* mutant. Of the patients with good prognosis 38% were MSI-H, TMB-H, of these 33% were *BRAF* mutant. High TMB is a surrogate marker for increased neoantigen expression, a biomarker for response to immune checkpoint inhibitors such as anti-PD-1/L1 therapy (339-341).

The CPM cohort had downregulation or somatic mutations of regulators of innate immunity (*DEFA6*, *DMBT1*, *MUC2*, *HLA-A* antigen) suggesting an immune evasive phenotype. As expected for a metastasis CPM samples had downregulation of genes suppressing invasion, migration and EMT (*MUC2*, *MMP26*, *ILK*, *FLNB*, *SPTB*, *PPL* and *SVEP1*) and those triggering these were upregulated (*CYR61*, *CXCL12*, *CTGF* and *CSTB*). CPM tumour were more commonly of the mesenchymal subtype suggesting a potential response to *TGF β* signalling inhibitors and targeted immunotherapies (61, 342). Methylation was dysregulated in CPM with somatic mutation of the *TET2* enzyme and *CDH7* chromatin regulator suggesting uncontrolled hyper-methylation and a theoretical benefit from methylation inhibitors (43, 49, 327). There was downregulation of *CES2*, a gene known to activate the prodrug irinotecan, a commonly used chemotherapy in the adjuvant treatment of primary CRC and CPM (*CES2*). Resistance to the treatment of primary CRC may in part explain the development of CPM in this cohort.

Comparison of biological changes in good and poor prognosis cohorts identified changes exclusive to and potentially conferring poor prognosis in CPM. Oncogenes, genes regulating cell growth, cycle, angiogenesis, transcription and the inhibition of apoptosis were upregulated (*RELB*, *HCKVEZF1*, *POLR3GL*, *FEM1C*, *KHDRBS3*, *BCYRN1* and *DCAF12*) or altered via hypermethylation or CNA (*NKX6-2*, *GATA3*, *IRX5*, *ZKSCAN7*, *PAX5*, *DEAF1A*, *CHFR*, *PTPN21*, *VASH1*, *IGFBP7*, *SHB*, *PTPN21*, *FAM32* and *MELK*) suggesting a rapid rate of unchecked tumour growth. Increased expression of genes regulating immune invasion (*CEACAM1*) and hypermethylation of regulators of the immune response (*RFTN1*, *SIT1*) suggest an immune evasive phenotype in poor prognosis CPM. This is an interesting therapeutic target, CEA TCB IgG-based T-cell bispecific antibodies result in immune-mediated tumour lysis in tumour cells with high expression of CEA (282). MAPK, Wnt/ β -catenin and TGF- β pathway regulators (*AXIN1*, *FAM13A*, *RAB8A*, *RAB34*, *FGF5* and *BMP3* were downregulated or altered by somatic mutation suggesting persistent activation of these

pathways in poor prognosis CPM. Poor prognosis CPM had hypermethylation and somatic mutation of regulators of invasion, migration and EMT (*NR2F1-AS1*, *ZEB1*, *EPS8L2* and *PIEZO*). Poor prognosis CPM was more commonly of the mesenchymal subtype which can be targeted by several novel therapies including *TGF- β* signalling inhibitors and targeted immunotherapies (61, 342). There was hypermethylation of a gene known to mediate resistance to oxaliplatin, a commonly used chemotherapy in the treatment of primary CRC and CPM (*NR2F1-AS1*). Resistance to the treatment may in part explain the poor prognosis CPM.

7.1.3 Developing a bio-molecular classifier to predict prognosis for patients with CPM undergoing CRS & HIPEC

I thirdly sought to identify from the biological changes discussed above a biomolecular classifier capable of predicting prognosis in response to the best currently available treatment for CPM, CRS & HIPEC. A machine-learning algorithm was applied to the RNA and methylation datasets to identify a prognostic classifier for patients with CPM. The RNA signature contained three genes capable of predicting prognosis with an accuracy of 99% at the training stage and 61% on the final hold outset, (*GNG12*, *FKBP10* and *PRRX1*). The expression of these genes was examined in the prospective cohort, the expression of *GNG12* was equivalent, *FKBP10* and *PRRX1* expression were increased non-significantly in the poor prognosis cohort. This may be a true failure of the signature to validate, this may mean there is no true gene expression signature or that the sample size was too small to establish one. It may also reflect the incomplete follow up of the prospective cohort and the inability to confidently classify the good prognosis cohort as such. The methylation signature identified two classifiers, the random forest contained one CpG feature and the PLSDA three CpG features capable of predicting prognosis with an accuracy of 0.939 and 0.997 respectively at the training stage, (*TMEM100*, *METT-15*, *POLR2M*, *GCOM1* and *SOX2*). Due to the increased size of the methylation dataset, it was not possible to partition into a training and holdout subset and so this will need to be validated. Classifiers are in clinical use for prognostic classification and therapeutic stratification of adjuvant treatments in breast cancer (347, 348). Whilst we have not identified from this work a prognostic classifier for patients with CPM, we have identified a technique to do so. It is hoped that applying this to a larger cohort of patients with a longer

duration of follow up and validation set will result in a clinically relevant classifier for CPM.

7.1.4 The use of circulating tumour DNA to predict tumour recurrence and determine mutational burden in patients with CPM

Finally, I aimed to determine the use of ctDNA as a liquid biopsy before surgery and as a biomarker of residual disease and recurrence following CRS & HIPEC. Matched ctDNA samples had a significantly higher tumour mutational burden when compared with CPM tumour samples. This may reflect the ability of ctDNA to better capture the temporal and spatial resolution of the tumour. This improved assessment of tumour heterogeneity and subclonal mutations has been used in the literature to predict treatment response, resistance to therapies and to allow re-treatment when resistant sub-clones decline (374-377). Of those patients with recurrence during the period of ctDNA sampling, 75% showed increasing levels of ctDNA preceding tumour recurrence, these numbers are too small to draw any conclusions regarding the utility of ctDNA as a marker of tumour recurrence. Due to loss to follow up, there were a limited number of ctDNA samples per patient, this was particularly pronounced in patients with tumour recurrence. This was a small cohort of patients with just 3 recurrences during the period of ctDNA sampling and limited duration of follow up. Following complete resection with CRS & HIPEC ctDNA levels increased marginally and remained detectable in 90% of cases. This may simply reflect lysis of microscopic tumour deposits in response to the chemotherapy and heat in the HIPEC treatment. One hypothesis is that it may represent a transient rise in novel mutations in response to de novo treatment with Mitomycin C which is not used in the treatment of CRC (382). The gene panel identified frequent mutations of Wnt pathways genes (*APC*, beta-catenin cofactors *BCL9* and *TCF7L2*, *RNF43*, *PTEN* and *PI3K*), mismatch repair pathway genes (*ARID1*, *ATM*, *MLH1*, *MSH2* and *MSH6*) and DNA polymerases (*POLE* and *POLD1*). Comparison of mutations between the good and poor prognosis cohorts identified a mutation *ACVR2A* which was significantly more frequent in patients with poor prognosis and associated with advanced stage, metastasis and reduced survival in the literature (385).

7.2 Strengths and limitations

This is the first study to comprehensively profile isolated colorectal peritoneal metastasis and to determine the biological changes associated with the development of peritoneal metastasis and associated with poor prognosis disease.

I must acknowledge that this is a small cohort of patients. Whilst we identified significant biological changes associated with the development of CPM and associated with poor prognosis disease these must form a starting point in identifying the tumour biology of CPM. Increased sample size will reduce variance in the machine learning algorithm used to develop the classifier, increasing the confidence that the features identified reflect the population the cohort is taken from. Small sample size additionally meant there were few recurrences during the period of ctDNA sampling meaning we were unable to assess the utility of ctDNA as a marker of tumour recurrence. Small sample size reflects the limited number of patients with isolated CPM suitable for and undergoing CRS & HIPEC at the limited number of centres performing CRS & HIPEC in the UK.

We were unable to validate the findings from chapter 4 as no suitable public dataset of isolated CPM exists in the literature or TCGA, GEO datasets to compare to. One dataset from Yaeger et al contains 52 patients with CPM, however in just 6 of these the sample sequenced was CPM (343-345). As we have demonstrated there is a significant change in tumour biology from primary CRC to CPM development which will not be captured in the analysis of the primary CRC or other sites of metastasis. We recruited a prospective cohort of patients to validate the retrospective findings. Whilst all patients in the prospective cohort had a minimum of 12 months follow up following CRS & HIPEC information regarding tumour recurrence and death was fed back from referring centres, this process is not immediate and means that it is likely that follow up is incomplete meaning patients classified as good prognosis may have recurred. This means we cannot be sure if the comparison between retrospective and prospective cohorts is accurate.

In the prospective cohort, we had a significant loss to follow up. This was particularly apparent in patients with recurrence following either CRS & HIPEC or adjuvant treatments. This is understandable as patients travel long distances to the CRS &

HIPEC centre and are less likely to do so when unwell. This meant we were unable to accurately determine the use of ctDNA as a biomarker of tumour recurrence.

The machine learning algorithms used in chapter 5 can develop classifiers for a set of features, in this case, we applied the methods to gene expression and methylation datasets. Tumour biology is more complex than this, genes connect by pathways and their expression and function are modulated via several mechanisms such as post-transcriptional modification by methylation, other RNA species, histone modification and post-translational modifications of protein function such as phosphorylation and ubiquitination. A single classifier will capture only one dimension of tumour biology. These features must be considered synergistically with the other biological and clinical prognostic data.

In chapter 6 we examined the use of ctDNA as a liquid biopsy and biomarker. In extracting ctDNA the blood sample was centrifuged twice to separate plasma from serum within 24 hours of collection to minimise sample contamination with white blood cell cfDNA. Additionally, we used a panel of mutations specific to CRC which should not be prevalent in germline samples. We know however that there is a high correlation between germline mutations in white blood cells and those found in cfDNA which can be a confounding factor (386). The use of normal control blood samples could be used as a comparator however clearly these would not be matched meaning germline mutations may not be comparable.

7.3 Future directions

7.3.1 Validation

In NGS the biological changes identified are often orders of magnitude greater than the sample size, they, therefore, require validation in an additional cohort (387). This is a relatively small cohort and there would be a benefit in the validation of these findings in a larger cohort of patients with isolated CPM. This would be challenging as the numbers of patients suitable for CRS & HIPEC (having their tumours sampled) are small and the number of centres performing this procedure is limited. Collaboration between CRS & HIPEC centres would be required to recruit more patients. This would allow validation of RNA, methylation and exome sequencing data in a second larger

cohort. Experimental validation is then required using an additional technique. For RNA seq, real-time quantitative reverse transcription PCR, (qRT-PCR) can be used for absolute and relative qualification of gene expression. For methylation, the degree of methylation can be quantitatively measured and validated using pyrosequencing. Quantitative bisulfite pyrosequencing determines DNA methylation levels by analysing artificial C/T SNPs at CpG sites within specific pyrosequencing assays. Candidate driver mutations identified through exome sequencing can be validated using a gene panel targeted towards the mutations of interest, this would allow an additional benefit of increased read depth.

Circulating tumour DNA has proven useful as a liquid biopsy and marker of tumour recurrence and treatment response in several cancers (374-377). Our results require validation in a larger cohort of patients. The loss to follow up was a problem in this study, strategies to improve this could include blood sample collection at the local hospital where patients attend for their imaging and adjuvant treatments or for this to form part of a clinical trial with a clearly defined follow up schedule.

7.3.2 Future research

Irinotecan is a chemotherapy agent used in the treatment of metastatic colorectal cancer. Carboxylesterase 2 (*CES2*) converts Irinotecan to its active metabolite SN-38. The expression of *CES2* was reduced 3.2-fold in CPM, this is a mechanism of resistance in metastatic CRC (268). Oxaliplatin is a chemotherapy agent used in the treatment of metastatic colorectal cancer. ABC transporters export drugs into the extracellular space reducing their cytotoxic effect (388). LncRNA *NR2F1-AS1* results in increased expression of ABC transporters and resistance to oxaliplatin chemotherapy *in vitro* (305). Resistance to such a commonly used chemotherapy in the CPM cohort may explain in part the poor prognosis in this group. The functional impact of differences in gene expression and methylation must be confirmed by analysis of *CES2*, *NR2F1-AS1* and ABC protein expression. In a prospective cohort of patients treated with irinotecan or oxaliplatin the expression of these markers could be determined using immunohistochemistry, treatment response and survival for patients

with low and high expression of these markers could then be compared to investigate whether high expression results in chemotherapy resistance in patients with CPM. The identification of a group of patients with resistance to individual chemotherapy drugs would allow alternative treatment strategies and outcomes in these patients.

Novel therapeutic targets were identified in patients with CPM which have been targeted with good effect either alone or in combination with other treatments (Figure 7.1). RNF43 a negative Wnt regulator was mutated in 23% of CPM patients. As an upstream Wnt regulator, this can be targeted by porcupine inhibitors which inhibit the Wnt pathway irrespective of the presence of downstream mutations (328, 338, 389). CPM had frequent mutations of mismatch repair proteins and high tumour mutational burden suggesting microsatellite instability, this suggests a good response to immune checkpoint inhibitors such as pembrolizumab (339). *TET2* mutations were noted in 23% of CPM samples suggesting a potential benefit from methylation inhibitors (43, 49, 327). There appeared to be EMT and a mesenchymal subtype in CPM and particularly poor prognosis CPM, this can be targeted by several novel therapies including TGF β signalling inhibitors and targeted immunotherapies (61, 342). The expression of genes regulating innate immunity was downregulated or altered by hypermethylation or CNA (*DEFA6*, *DMBT1*, *MUC2*, *RTFN1*, *SIT1*). Genes regulating immune evasion were upregulated (*CEACAM1*). Novel therapies in the form of immune checkpoint inhibitors or CEA TCB IgG-based T-cell bispecific antibodies may, therefore, be of benefit in this patient cohort, they result in immune-mediated tumour lysis in tumours overexpressing CEA (282). Before use in clinical studies, drugs are initially tested *in-vitro*. Traditionally this was in a two-dimensional cell culture model. Two-dimensional cell cultures, however, are not representative of the *in vivo* situation. Heterogeneity, three-dimensional organ structure, cell to cell relationships are lost and signalling networks altered. It is unsurprising then that the success of drugs tested in 2D *in vitro* models commonly fails to translate into clinical trials (390). A type of three-dimensional cell culture model provides an alternative. An organoid is a clonal derivative of primary epithelial stem cells grown without mesenchyme or epithelial-mesenchymal co-culture (391). Organoids can be established in a short period and maintain tumour morphological and genetic features on expansion. They can,

therefore, be used as *in vitro* models to test drugs targeted towards patients individual mutations (392, 393). We have successfully cultured two organoid lines from patient's CPM samples. Expansion of this biobank and comparison with the tumours from which they are derived would be required prior to use of the model of *in vitro* drug testing. This would require histology to confirm tumour morphology, immunohistochemistry to confirm the expression of key CRC markers and a comparison of genetic mutation concordance using genome sequencing of paired tumour and organoid samples.

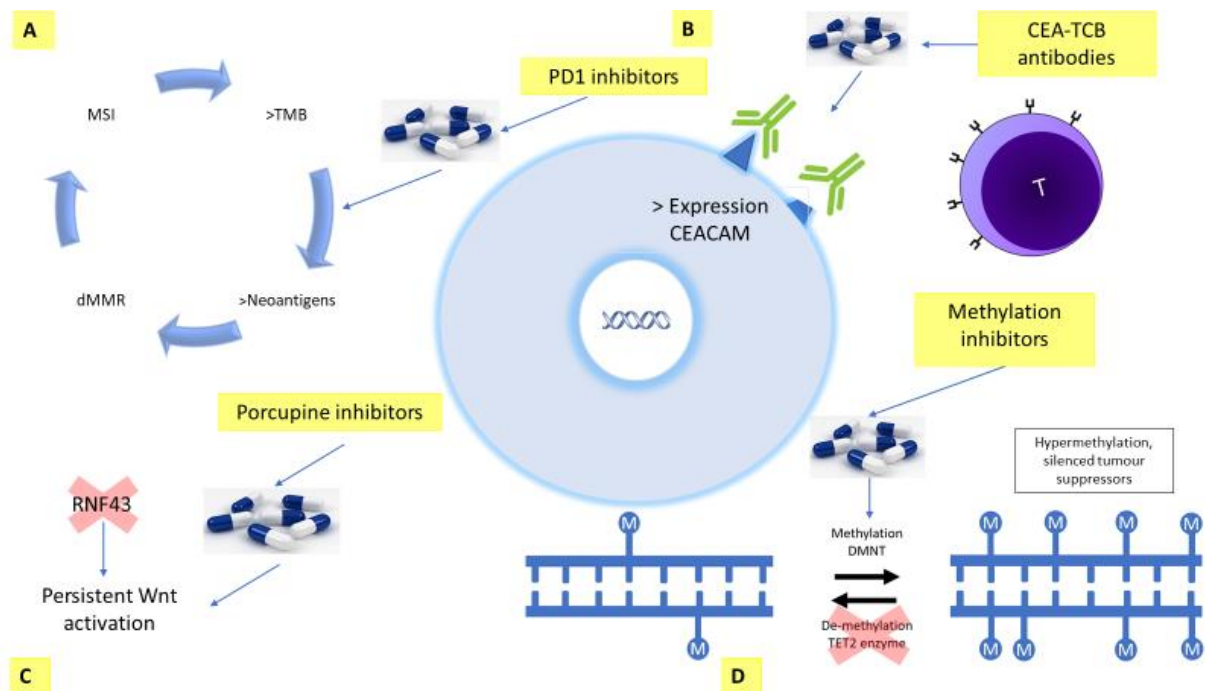


Figure 7-1 Novel therapeutic targets identified

A) Microsatellite instability and increased tumour mutational burden was noted in CPM and particularly in poor prognosis CPM, this results in an increased neo-antigen burden and suggests response to PD1 checkpoint inhibitors. B) Increased CEACAM antigen related cell adhesion molecule expression was noted in poor prognosis CPM, this can be targeted by novel CEA-TCB bi-specific antibodies. C) RNF43 an upstream negative Wnt regulator was commonly mutated in CPM, this results in persistent Wnt activation, this can be targeted by porcupine inhibitor LGK974. D) TET2 mutations were noted in CPM, this results in a low of active demethylation and hypermethylation and silencing of tumour suppressor genes, this is a potential target for methylation inhibitors such as 5-azacytidine. This figure was produced by myself in Microsoft powerpoint.

Successful treatments at the *in vitro* stage could move forward into a clinical trial. CEA TCB antibodies, Methylation inhibitors and porcupine inhibitors are already being trialled in some phase 1 studies (394-397). Pembrolizumab and Nivolumab have been trialled with success in patients with MSI mismatch repair deficient metastatic CRC (339-341). No trials exist in patients with isolated peritoneal metastasis.

Given the number of targetable mutations identified within patients with CPM an appropriate phase II trial design would be an umbrella trial with multiple treatment arms according to patients individual mutations. Patients with isolated metachronous CPM resected by CRS & HIPEC would undergo molecular profiling to determine their targetable mutations (*RNF43*, *TET2*, *CEACAM*, and MSI). Personalised targeted adjuvant treatments could then be given based on the individual tumours molecular features. Patients would be followed up clinically and with CT imaging for 24 months. Treatment response would be assessed according to the response evaluation criteria in solid tumours (RECIST) (398), in addition to a further analysis of long term survival outcomes at 5-years.

A clinical trial could be combined with the repetition of the ctDNA experiment performed in chapter 6 which aimed to determine whether ctDNA can be used as a liquid biopsy before surgery and as a biomarker of residual disease and recurrence following CRS & HIPEC. The enhanced follow up demanded by a clinical trial is likely to reduce the loss to follow up we encountered.

7.4 Conclusion

Patients with CPM secondary to CRC have limited survival with the best available treatments. Despite selection for treatment using known prognostic factors survival varies widely and can be difficult to predict. This thesis has identified biological changes associated with the development of isolated CPM and conferring 'poor prognosis' in response to the best currently available treatment in a small cohort of patients. These results require validation. Further investigation in a larger cohort is needed to develop a prognostic biomolecular classifier and to determine whether ctDNA can be used as a marker of tumour recurrence. Therapeutically targetable alterations identified here warrant further investigation with *in vitro* and clinical trials

References

1. Office for National Statistics. Cancer Registration Statistics, England: 2014. UK2016.
2. Office for National Statistics. Cancer survival in England. UK2016.
3. Cancer Research UK. Cancer incidence statistics [Available from: <http://www.cancerresearchuk.org/health-professional/cancer-statistics/incidence>.
4. Parkin D. International variation. *Oncogene*. 2004;23:6329–40.
5. Murphy N, Norat T, Ferrari P, Jenab M, Bueno-de-Mesquita B, Skeie G, et al. Dietary fibre intake and risks of cancers of the colon and rectum in the European prospective investigation into cancer and nutrition (EPIC). *PLoS One*. 2012;7(6):e39361.
6. Norat T, Bingham S, Ferrari P, Slimani N, Jenab M, Mazuir M, et al. Meat, fish, and colorectal cancer risk: the European Prospective Investigation into cancer and nutrition. *Journal of the National Cancer Institute*. 2005;97(12):906-16.
7. Song M, Giovannucci E. Preventable Incidence and Mortality of Carcinoma Associated With Lifestyle Factors Among White Adults in the United States. *JAMA Oncol*. 2016;2(9):1154-61.
8. Tomasetti C, Vogelstein B. Cancer etiology. Variation in cancer risk among tissues can be explained by the number of stem cell divisions. *Science*. 2015;347(6217):78-81.
9. von Roon AC, Reese G, Teare J, Constantinides V, Darzi AW, Tekkis PP. The risk of cancer in patients with Crohn's disease. *Dis Colon Rectum*. 2007;50(6):839-55.
10. Eaden JA, Abrams KR, Mayberry JF. The risk of colorectal cancer in ulcerative colitis: a meta-analysis. *Gut*. 2001;48(4):526-35.
11. Henderson TO, Oeffinger KC, Whitton J, Leisenring W, Neglia J, Meadows A, et al. Secondary gastrointestinal cancer in childhood cancer survivors: a cohort study. *Ann Intern Med*. 2012;156(11):757-66, w-260.
12. Jasperson K, Tuohy T, Neklason D, Burt R. Hereditary and Familial Colon Cancer. *Gastroenterology*. 2010;138(6):2044-58.

13. Knudson AG. Mutation and cancer: statistical study of retinoblastoma. *Proc Natl Acad Sci U S A*. 1971;68(4):820-3.
14. Kunkel TA, Bebenek K. DNA replication fidelity. *Annu Rev Biochem*. 2000;69:497-529.
15. Iyer RR, Pluciennik A, Burdett V, Modrich PL. DNA mismatch repair: functions and mechanisms. *Chem Rev*. 2006;106(2):302-23.
16. Boland C, Goel A. Microsatellite Instability in Colorectal Cancer. *Gastroenterology*. 2011;138(6):2073-87.
17. Koopman M, Kortman GA, Mekenkamp L, Ligtenberg MJ, Hoogerbrugge N, Antonini NF, et al. Deficient mismatch repair system in patients with sporadic advanced colorectal cancer. *British journal of cancer*. 2009;100(2):266-73.
18. Kim CG, Ahn JB, Jung M, Beom SH, Kim C, Kim JH, et al. Effects of microsatellite instability on recurrence patterns and outcomes in colorectal cancers. *British journal of cancer*. 2016;115(1):25-33.
19. Fabrizio DA, George TJ, Dunne RF, Frampton G, Sun J, Gowen K, et al. Beyond microsatellite testing: assessment of tumor mutational burden identifies subsets of colorectal cancer who may respond to immune checkpoint inhibition. *J Gastrointest Oncol*. 2018;9(4):610-7.
20. Jascur T, Boland CR. Structure and function of the components of the human DNA mismatch repair system. *International journal of cancer Journal international du cancer*. 2006;119(9):2030-5.
21. Zhan T, Rindtorff N, Boutros M. Wnt signalling in cancer. *Oncogene*. 2017;36:1461-73.
22. Markowitz SD, Bertagnolli MM. Molecular origins of cancer: Molecular basis of colorectal cancer. *N Engl J Med*. 2009;361(25):2449-60.
23. Morin P, Sparks A, Korinek V, Barker N, Clevers H, Vogelstein B, et al. Activation of β -Catenin-Tcf Signaling in Colon Cancer by Mutations in β -Catenin or APC. *Science*. 1997;275(5307):1787-90.
24. van Andel H, Kocemba KA, Spaargaren M, Pals ST. Aberrant Wnt signaling in multiple myeloma: molecular mechanisms and targeting options. *Leukemia*. 2019;33(5):1063-75.

25. Pino M, Chung D. The chromosomal instability pathway in colon cancer. *Gastroenterology*. 2010;138(6):2059-72.
26. McGranahan N, Burrell RA, Endesfelder D, Novelli MR, Swanton C. Cancer chromosomal instability: therapeutic and diagnostic challenges. *EMBO Rep*. 2012;13(6):528-38.
27. Li XL, Zhou J, Chen ZR, Chng WJ. P53 mutations in colorectal cancer - molecular pathogenesis and pharmacological reactivation. *World J Gastroenterol*. 2015;21(1):84-93.
28. Baker SJ, Preisinger AC, Jessup JM, Paraskeva C, Markowitz S, Willson JK, et al. p53 gene mutations occur in combination with 17p allelic deletions as late events in colorectal tumorigenesis. *Cancer Res*. 1990;50(23):7717-22.
29. Calon A, Espinet E, Palomo-Ponce S, Tauriello DV, Iglesias M, Céspedes MV, et al. Dependency of colorectal cancer on a TGF- β -driven program in stromal cells for metastasis initiation. *Cancer Cell*. 2012;22(5):571-84.
30. Tauriello D, Palomo-Ponce S, Stork D, Berenguer-Llergo A, Badia-Ramentol J, Iglesias M, et al. TGF β drives immune evasion in genetically reconstituted colon cancer metastasis. *Nature*. 2018;554(7693):538-43.
31. Roberts PJ, Der CJ. Targeting the Raf-MEK-ERK mitogen-activated protein kinase cascade for the treatment of cancer. *Oncogene*. 2007;26(22):3291-310.
32. Andreyev HJ, Norman AR, Cunningham D, Oates JR, Clarke PA. Kirsten ras mutations in patients with colorectal cancer: the multicenter "RASCAL" study. *Journal of the National Cancer Institute*. 1998;90(9):675-84.
33. Lièvre A, Bachet JB, Le Corre D, Boige V, Landi B, Emile JF, et al. KRAS mutation status is predictive of response to cetuximab therapy in colorectal cancer. *Cancer Res*. 2006;66(8):3992-5.
34. Roth AD, Tejpar S, Delorenzi M, Yan P, Fiocca R, Klingbiel D, et al. Prognostic role of KRAS and BRAF in stage II and III resected colon cancer: results of the translational study on the PETACC-3, EORTC 40993, SAKK 60-00 trial. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. 2010;28(3):466-74.
35. Kopetz S, Desai J, Chan E, Hecht JR, O'Dwyer PJ, Maru D, et al. Phase II Pilot Study of Vemurafenib in Patients With Metastatic BRAF-Mutated Colorectal

Cancer. *Journal of clinical oncology* : official journal of the American Society of Clinical Oncology. 2015;33(34):4032-8.

36. Prahallad A, Sun C, Huang S, Di Nicolantonio F, Salazar R, Zecchin D, et al. Unresponsiveness of colon cancer to BRAF(V600E) inhibition through feedback activation of EGFR. *Nature*. 2012;483(7387):100-3.

37. Dankner M, Rose AAN, Rajkumar S, Siegel PM, Watson IR. Classifying BRAF alterations in cancer: new rational therapeutic strategies for actionable mutations. *Oncogene*. 2018;37(24):3183-99.

38. Jhaver M, Goel S, Wilson AJ, Montagna C, Ling YH, Byun DS, et al. PIK3CA mutation/PTEN expression status predicts response of colon cancer cells to the epidermal growth factor receptor inhibitor cetuximab. *Cancer Res*. 2008;68(6):1953-61.

39. Toss A, Cristofanilli M. Molecular characterization and targeted therapeutic approaches in breast cancer. *Breast Cancer Res*. 2015;17:60.

40. Egger G, Liang G, Aparicio A, Jones PA. Epigenetics in human disease and prospects for epigenetic therapy. *Nature*. 2004;429(6990):457-63.

41. Issa J. CpG island methylator phenotype in cancer. *Nature Reviews Cancer*. 2004;4:988-93.

42. Robertson KD. DNA methylation and human disease. *Nat Rev Genet*. 2005;6(8):597-610.

43. Jones PA, Baylin SB. The fundamental role of epigenetic events in cancer. *Nat Rev Genet*. 2002;3(6):415-28.

44. Weisenberger DJ, Siegmund KD, Campan M, Young J, Long TI, Faasse MA, 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):787-93.

45. Azad NS, El-Khoueiry A, Yin J, Oberg AL, Flynn P, Adkins D, et al. Combination epigenetic therapy in metastatic colorectal cancer (mCRC) with subcutaneous 5-azacitidine and entinostat: a phase 2 consortium/stand up 2 cancer study. *Oncotarget*. 2017;8(21):35326-38.

46. Lin J, Gilbert J, Rudek MA, Zwiebel JA, Gore S, Jiemjit A, et al. A phase I dose-finding study of 5-azacytidine in combination with sodium phenylbutyrate in

patients with refractory solid tumors. *Clinical cancer research : an official journal of the American Association for Cancer Research*. 2009;15(19):6241-9.

47. Overman MJ, Morris V, Moinova H, Manyam G, Ensor J, Lee MS, et al. Phase I/II study of azacitidine and capecitabine/oxaliplatin (CAPOX) in refractory CIMP-high metastatic colorectal cancer: evaluation of circulating methylated vimentin. *Oncotarget*. 2016;7(41):67495-506.

48. Agrawal K, Das V, Vyas P, Hajdúch M. Nucleosidic DNA demethylating epigenetic drugs - A comprehensive review from discovery to clinic. *Pharmacol Ther*. 2018;188:45-79.

49. Carter CA, Oronsky BT, Roswarski J, Oronsky AL, Oronsky N, Scicinski J, et al. No patient left behind: The promise of immune priming with epigenetic agents. *Oncoimmunology*. 2017;6(10):e1315486.

50. Khong HT, Restifo NP. Natural selection of tumor variants in the generation of "tumor escape" phenotypes. *Nat Immunol*. 2002;3(11):999-1005.

51. Brennick CA, George MM, Corwin WL, Srivastava PK, Ebrahimi-Nik H. Neoepitopes as cancer immunotherapy targets: key challenges and opportunities. *Immunotherapy*. 2017;9(4):361-71.

52. Vogelstein B, Fearon ER, Hamilton SR, Kern SE, Preisinger AC, Leppert M, et al. Genetic alterations during colorectal-tumor development. *N Engl J Med*. 1988;319(9):525-32.

53. Smith G, Carey FA, Beattie J, Wilkie MJ, Lightfoot TJ, Coxhead J, et al. Mutations in APC, Kirsten-ras, and p53--alternative genetic pathways to colorectal cancer. *Proc Natl Acad Sci U S A*. 2002;99(14):9433-8.

54. Thirlwell C, Will OC, Domingo E, Graham TA, McDonald SA, Oukrif D, et al. Clonality assessment and clonal ordering of individual neoplastic crypts shows polyclonality of colorectal adenomas. *Gastroenterology*. 2010;138(4):1441-54, 54.e1-7.

55. Walther A, Johnstone E, Swanton C, Midgley R, Tomlinson I, Kerr D. Genetic prognostic and predictive markers in colorectal cancer. *Nat Rev Cancer*. 2009;9(7):489-99.

56. Oliveira D, Santamaria G, Laudanna C, Migliozi S, Zoppoli P, Quist M, et al. Identification of copy number alterations in colon cancer from

- analysis of amplicon-based next generation sequencing data. *Oncotarget*. 2018;9(29):20409-25.
57. Wang H, Liang L, Fang JY, Xu J. Somatic gene copy number alterations in colorectal cancer: new quest for cancer drivers and biomarkers. *Oncogene*. 2016;35(16):2011-9.
 58. Xie T, D' Ario G, Lamb JR, Martin E, Wang K, Tejpar S, et al. A comprehensive characterization of genome-wide copy number aberrations in colorectal cancer reveals novel oncogenes and patterns of alterations. *PLoS One*. 2012;7(7):e42001.
 59. Guinney J, Dienstmann R, Wang X, de Reyniès A, Schlicker A, Soneson C, et al. The consensus molecular subtypes of colorectal cancer. *Nat Med*. 2015;21(11):1350-6.
 60. Dienstmann R, Vermeulen L, Guinney J, Kopetz S, Tejpar S, Tabernero J. Consensus molecular subtypes and the evolution of precision medicine in colorectal cancer. *Nat Rev Cancer*. 2017;17(4):268.
 61. Calon A, Lonardo E, Berenguer-Llargo A, Espinet E, Hernando-Momblona X, Iglesias M, et al. Stromal gene expression defines poor-prognosis subtypes in colorectal cancer. *Nat Genet*. 2015;47(4):320-9.
 62. Isella C, Brundu F, Bellomo SE, Galimi F, Zanella E, Porporato R, et al. Selective analysis of cancer-cell intrinsic transcriptional traits defines novel clinically relevant subtypes of colorectal cancer. *Nat Commun*. 2017;8:15107.
 63. Rinaudo P, Boudah S, Junot C, Thévenot EA. biosigner: A New Method for the Discovery of Significant Molecular Signatures from Omics Data. *Front Mol Biosci*. 2016;3:26.
 64. Trainor PJ, DeFilippis AP, Rai SN. Evaluation of Classifier Performance for Multiclass Phenotype Discrimination in Untargeted Metabolomics. *Metabolites*. 2017;7(2).
 65. Mansouri D, McMillan DC, Crearie C, Morrison DS, Crichton EM, Horgan PG. Temporal trends in mode, site and stage of presentation with the introduction of colorectal cancer screening: a decade of experience from the West of Scotland. *British journal of cancer*. 2015;113(3):556-61.

66. Logan RF, Patnick J, Nickerson C, Coleman L, Rutter MD, von Wagner C, et al. Outcomes of the Bowel Cancer Screening Programme (BCSP) in England after the first 1 million tests. *Gut*. 2012;61(10):1439-46.
67. Moreno CC, Mittal PK, Sullivan PS, Rutherford R, Staley CA, Cardona K, et al. Colorectal Cancer Initial Diagnosis: Screening Colonoscopy, Diagnostic Colonoscopy, or Emergent Surgery, and Tumor Stage and Size at Initial Presentation. *Clin Colorectal Cancer*. 2016;15(1):67-73.
68. NICE. Colorectal cancer : the diagnosis and management of colorectal cancer. Cardiff: National Collaborating Centre for Cancer; 2011.
69. Dukes C, Bussey H. The Spread of Rectal Cancer and its Effect on Prognosis. *British Journal of Cancer*. 1932;12:309-20.
70. Brierley J, Gospodarowicz M, Wittekind C. *TNM Classification of Malignant Tumours* (8th edition). Oxford, UK: Wiley-Blackwell; 2017.
71. Hess K, Gauri R, Varadhachary M, Taylor S, Wei W, Raber M, et al. Metastatic Patterns in Adenocarcinoma. *Cancer*. 2005;106(7):1624-33.
72. Fidler I. The pathogenesis of cancer metastasis: the 'seed and soil' hypothesis revisited. *Nature reviews cancer*. 2002;3(6):453-8.
73. Sleeman JP, Nazarenko I, Thiele W. Do all roads lead to Rome? Routes to metastasis development. *Int J Cancer*. 2011;128(11):2511-26.
74. Sugarbaker P. *Peritoneal Carcinomatosis: Principles of Management*. Sugarbaker P, editor. USA: Kluwer academic publishers; 2006.
75. Jayne DG, Fook S, Loi C, Seow-Choen F. Peritoneal carcinomatosis from colorectal cancer. *Br J Surg*. 2002;89(12):1545-50.
76. Segelman J, Granath F, Holm T, Machado M, Mahteme H, Martling A. Incidence, prevalence and risk factors for peritoneal carcinomatosis from colorectal cancer. *Br J Surg*. 2012;99(5):699-705.
77. Quere P, Facy O, Manfredi S, Jooste V, Faivre J, Lepage C, et al. Epidemiology, Management, and Survival of Peritoneal Carcinomatosis from Colorectal Cancer: A Population-Based Study. *Dis Colon Rectum*. 2015;58(8):743-52.

78. Francia G, Cruz-Munoz W, Man S, Xu P, Kerbel RS. Mouse models of advanced spontaneous metastasis for experimental therapeutics. *Nat Rev Cancer*. 2011;11(2):135-41.
79. Chu DZ, Lang NP, Thompson C, Osteen PK, Westbrook KC. Peritoneal carcinomatosis in nongynecologic malignancy. A prospective study of prognostic factors. *Cancer*. 1989;63(2):364-7.
80. Sadeghi B, Arvieux C, Glehen O, Beaujard AC, Rivoire M, Baulieux J, et al. Peritoneal carcinomatosis from non-gynecologic malignancies: results of the EVOCAPE 1 multicentric prospective study. *Cancer*. 2000;88(2):358-63.
81. Kohne CH, Cunningham D, Di Costanzo F, Glimelius B, Blijham G, Aranda E, et al. Clinical determinants of survival in patients with 5-fluorouracil-based treatment for metastatic colorectal cancer: results of a multivariate analysis of 3825 patients. *Ann Oncol*. 2002;13(2):308-17.
82. Sugarbaker PH. Observations concerning cancer spread within the peritoneal cavity and concepts supporting an ordered pathophysiology. *Cancer Treat Res*. 1996;82:79-100.
83. Van Cutsem E, Cervantes A, Nordlinger B, Arnold D. Metastatic colorectal cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol*. 2014;25 Suppl 3:iii1-9.
84. van Oudheusden TR, Razenberg LG, van Gestel YR, Creemers GJ, Lemmens VE, de Hingh IH. Systemic treatment of patients with metachronous peritoneal carcinomatosis of colorectal origin. *Sci Rep*. 2015;5:18632.
85. Franko J, Shi Q, Goldman CD, Pockaj BA, Nelson GD, Goldberg RM, et al. Treatment of colorectal peritoneal carcinomatosis with systemic chemotherapy: a pooled analysis of north central cancer treatment group phase III trials N9741 and N9841. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. 2012;30(3):263-7.
86. Klaver YL, Simkens LH, Lemmens VE, Koopman M, Teerenstra S, Bleichrodt RP, et al. Outcomes of colorectal cancer patients with peritoneal carcinomatosis treated with chemotherapy with and without targeted therapy. *Eur J Surg Oncol*. 2012;38(7):617-23.

87. Griffiths CT, Parker LM, Fuller AF, Jr. Role of cytoreductive surgical treatment in the management of advanced ovarian cancer. *Cancer Treat Rep.* 1979;63(2):235-40.
88. Long RT, Spratt JS, Jr., Dowling E. Pseudomyxoma peritonei. New concepts in management with a report of seventeen patients. *Am J Surg.* 1969;117(2):162-9.
89. Zimm S, Cleary SM, Lucas WE, Weiss RJ, Markman M, Andrews PA, et al. Phase I/pharmacokinetic study of intraperitoneal cisplatin and etoposide. *Cancer Res.* 1987;47(6):1712-6.
90. Howell SB, Zimm S, Markman M, Abramson IS, Cleary S, Lucas WE, et al. Long-term survival of advanced refractory ovarian carcinoma patients with small-volume disease treated with intraperitoneal chemotherapy. *J Clin Oncol.* 1987;5(10):1607-12.
91. Cavaliere R, Ciocatto EC, Giovanella BC, Heidelberger C, Johnson RO, Margottini M, et al. Selective heat sensitivity of cancer cells. Biochemical and clinical studies. *Cancer.* 1967;20(9):1351-81.
92. Sugarbaker PH. Surgical management of peritoneal carcinosis: diagnosis, prevention and treatment. *Langenbecks Arch Chir.* 1988;373(3):189-96.
93. Sugarbaker PH, Gianola FJ, Speyer JL, Wesley R, Barofsky I, Myers CE. Prospective randomized trial of intravenous v intraperitoneal 5-FU in patients with advanced primary colon or rectal cancer. *Semin Oncol.* 1985;12(3 Suppl 4):101-11.
94. Gilly FN, Beaujard A, Glehen O, Grandclement E, Caillot JL, Francois Y, et al. Peritonectomy combined with intraperitoneal chemohyperthermia in abdominal cancer with peritoneal carcinomatosis: phase I-II study. *Anticancer Res.* 1999;19(3b):2317-21.
95. Sugarbaker PH. Peritonectomy procedures. *Ann Surg.* 1995;221(1):29-42.
96. Esquivel J, Farinetti A, Sugarbaker PH. [Elective surgery in recurrent colon cancer with peritoneal seeding: when to and when not to proceed]. *G Chir.* 1999;20(3):81-6.
97. Elias D, Blot F, El Otmany A, Antoun S, Lasser P, Boige V, et al. Curative treatment of peritoneal carcinomatosis arising from colorectal cancer by complete resection and intraperitoneal chemotherapy. *Cancer.* 2001;92(1):71-6.

98. Esquivel J, Lowy AM, Markman M, Chua T, Pelz J, Baratti D, et al. The American Society of Peritoneal Surface Malignancies (ASPSM) Multiinstitution Evaluation of the Peritoneal Surface Disease Severity Score (PSDSS) in 1,013 Patients with Colorectal Cancer with Peritoneal Carcinomatosis. *Annals of surgical oncology*. 2014;21(13):4195-201.
99. Sugarbaker PH. Successful management of microscopic residual disease in large bowel cancer. *Cancer Chemotherapy and Pharmacology*. 1999;43:S15-S25.
100. Los G, Mutsaers PH, van der Vijgh WJ, Baldew GS, de Graaf PW, McVie JG. Direct diffusion of cis-diamminedichloroplatinum(II) in intraperitoneal rat tumors after intraperitoneal chemotherapy: a comparison with systemic chemotherapy. *Cancer Res*. 1989;49(12):3380-4.
101. Harmon RL, Sugarbaker PH. Prognostic indicators in peritoneal carcinomatosis from gastrointestinal cancer. *Int Semin Surg Oncol*. 2005;2(1):3.
102. Stephens AD, Alderman R, Chang D, Edwards GD, Esquivel J, Sebbag G, et al. Morbidity and mortality analysis of 200 treatments with cytoreductive surgery and hyperthermic intraoperative intraperitoneal chemotherapy using the coliseum technique. *Annals of surgical oncology*. 1999;6(8):790-6.
103. Elias D, Antoun S, Goharin A, Otmany AE, Puizillout JM, Lasser P. Research on the best chemohyperthermia technique of treatment of peritoneal carcinomatosis after complete resection. *Int J Surg Investig*. 2000;1(5):431-9.
104. Glehen O, Cotte E, Kusamura S, Deraco M, Baratti D, Passot G, et al. Hyperthermic intraperitoneal chemotherapy: nomenclature and modalities of perfusion. *J Surg Oncol*. 2008;98(4):242-6.
105. de Bree E. Optimal drugs for HIPEC in different tumors. *J buon*. 2015;20 Suppl 1:S40-6.
106. Hompes D, D'Hoore A, Wolthuis A, Fieuws S, Mirck B, Bruin S, et al. The use of Oxaliplatin or Mitomycin C in HIPEC treatment for peritoneal carcinomatosis from colorectal cancer: a comparative study. *Journal of Surgical Oncology*. 2014;109(6):527-32.
107. Prada-Villaverde A, Esquivel J, Lowy AM, Markman M, Chua T, Pelz J, et al. The American Society of Peritoneal Surface Malignancies evaluation of HIPEC with

- Mitomycin C versus Oxaliplatin in 539 patients with colon cancer undergoing a complete cytoreductive surgery. *Journal of Surgical Oncology*. 2014;110(7):779-85.
108. Glehen O, Kwiatkowski F, Sugarbaker PH, Elias D, Levine EA, De Simone M, et al. Cytoreductive surgery combined with perioperative intraperitoneal chemotherapy for the management of peritoneal carcinomatosis from colorectal cancer: a multi-institutional study. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. 2004;22(16):3284-92.
109. Glehen O, Gilly FN, Boutitie F, Bereder JM, Quenet F, Sideris L, et al. Toward curative treatment of peritoneal carcinomatosis from nonovarian origin by cytoreductive surgery combined with perioperative intraperitoneal chemotherapy: a multi-institutional study of 1,290 patients. *Cancer*. 2010;116(24):5608-18.
110. Klaver YL, Chua TC, de Hingh IH, Morris DL. Outcomes of elderly patients undergoing cytoreductive surgery and perioperative intraperitoneal chemotherapy for colorectal cancer peritoneal carcinomatosis. *J Surg Oncol*. 2012;105(2):113-8.
111. Cashin PH, Graf W, Nygren P, Mahteme H. Intraoperative hyperthermic versus postoperative normothermic intraperitoneal chemotherapy for colonic peritoneal carcinomatosis: a case-control study. *Annals of Oncology*. 2012;23(3):647-52.
112. Cashin PH, Graf W, Nygren P, Mahteme H. Cytoreductive surgery and intraperitoneal chemotherapy for colorectal peritoneal carcinomatosis: prognosis and treatment of recurrences in a cohort study. *European Journal of Surgical Oncology*. 2012;38(6):509-15.
113. Nash G. ICARuS Post-operative Intraperitoneal Chemotherapy (EPIC) and Hyperthermic Intraperitoneal Chemotherapy (HIPEC) After Optimal Cytoreductive Surgery (CRS) for Neoplasms of the Appendix, Colon or Rectum With Isolated Peritoneal Metastasis (ICARuS). NCT01815359: ClinicalTrials.gov.
114. Verwaal VJ, van Ruth S, de Bree E, van Sloothen GW, van Tinteren H, Boot H, et al. Randomized trial of cytoreduction and hyperthermic intraperitoneal chemotherapy versus systemic chemotherapy and palliative surgery in patients with peritoneal carcinomatosis of colorectal cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. 2003;21(20):3737-43.

115. Verwaal VJ, Bruin S, Boot H, van Slooten G, van Tinteren H. 8-year follow-up of randomized trial: cytoreduction and hyperthermic intraperitoneal chemotherapy versus systemic chemotherapy in patients with peritoneal carcinomatosis of colorectal cancer. *Annals of surgical oncology*. 2008;15(9):2426-32.
116. Cashin PH, Mahteme H, Spång N, Syk I, Frödin JE, Torkzad M, et al. Cytoreductive surgery and intraperitoneal chemotherapy versus systemic chemotherapy for colorectal peritoneal metastases: A randomised trial. *European journal of cancer (Oxford, England : 1990)*. 2016;53:155-62.
117. Elias D, Lefevre JH, Chevalier J, Brouquet A, Marchal F, Classe JM, et al. Complete cytoreductive surgery plus intraperitoneal chemohyperthermia with oxaliplatin for peritoneal carcinomatosis of colorectal origin. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. 2009;27(5):681-5.
118. Franko J, Ibrahim Z, Gusani NJ, Holtzman MP, Bartlett DL, Zeh HJ. Cytoreductive surgery and hyperthermic intraperitoneal chemoperfusion versus systemic chemotherapy alone for colorectal peritoneal carcinomatosis. *Cancer*. 2010;116(16):3756-62.
119. Portilla AG, Sugarbaker PH, Chang D. Second-look surgery after cytoreduction and intraperitoneal chemotherapy for peritoneal carcinomatosis from colorectal cancer: analysis of prognostic features. *World J Surg*. 1999;23(1):23-9.
120. Pilati P, Mocellin S, Rossi CR, Foletto M, Campana L, Nitti D, et al. Cytoreductive surgery combined with hyperthermic intraperitoneal intraoperative chemotherapy for peritoneal carcinomatosis arising from colon adenocarcinoma. *Annals of surgical oncology*. 2003;10(5):508-13.
121. Glehen O, Cotte E, Schreiber V, Sayag-Beaujard AC, Vignal J, Gilly FN. Intraperitoneal chemohyperthermia and attempted cytoreductive surgery in patients with peritoneal carcinomatosis of colorectal origin. *Br J Surg*. 2004;91(6):747-54.
122. Mahteme H, Hansson J, Berglund A, Pahlman L, Glimelius B, Nygren P, et al. Improved survival in patients with peritoneal metastases from colorectal cancer: a preliminary study. *British journal of cancer*. 2004;90(2):403-7.
123. Shen P, Hawksworth J, Lovato J, Loggie BW, Geisinger KR, Fleming RA, et al. Cytoreductive surgery and intraperitoneal hyperthermic chemotherapy with

- mitomycin C for peritoneal carcinomatosis from nonappendiceal colorectal carcinoma. *Annals of surgical oncology*. 2004;11(2):178-86.
124. Cavaliere F, Valle M, De Simone M, Deraco M, Rossi CR, Di Filippo F, et al. 120 peritoneal carcinomatoses from colorectal cancer treated with peritonectomy and intra-abdominal chemohyperthermia: a S.I.T.I.L.O. multicentric study. *In Vivo*. 2006;20(6a):747-50.
125. Gusani NJ, Cho SW, Colovos C, Seo S, Franko J, Richard SD, et al. Aggressive surgical management of peritoneal carcinomatosis with low mortality in a high-volume tertiary cancer center. *Ann Surg Oncol*. 2008;15(3):754-63.
126. Varban O, Levine EA, Stewart JH, McCoy TP, Shen P. Outcomes associated with cytoreductive surgery and intraperitoneal hyperthermic chemotherapy in colorectal cancer patients with peritoneal surface disease and hepatic metastases. *Cancer*. 2009;115(15):3427-36.
127. Vaira M, Cioppa T, D'Amico S, de Marco G, D'Alessandro M, Fiorentini G, et al. Treatment of peritoneal carcinomatosis from colonic cancer by cytoreduction, peritonectomy and hyperthermic intraperitoneal chemotherapy (HIPEC). Experience of ten years. *In Vivo*. 2010;24(1):79-84.
128. Quenet F, Goere D, Mehta SS, Roca L, Dumont F, Hessissen M, et al. Results of two bi-institutional prospective studies using intraperitoneal oxaliplatin with or without irinotecan during HIPEC after cytoreductive surgery for colorectal carcinomatosis. *Ann Surg*. 2011;254(2):294-301.
129. Passot G, Vaudoyer D, Cotte E, You B, Isaac S, Noel Gilly F, et al. Progression following neoadjuvant systemic chemotherapy may not be a contraindication to a curative approach for colorectal carcinomatosis. *Annals of Surgery*. 2012;256(1):125-9.
130. Hompes D, D'Hoore A, Van Cutsem E, Fieuws S, Ceelen W, Peeters M, et al. The treatment of peritoneal carcinomatosis of colorectal cancer with complete cytoreductive surgery and hyperthermic intraperitoneal peroperative chemotherapy (HIPEC) with oxaliplatin: a Belgian multicentre prospective phase II clinical study. *Annals of surgical oncology*. 2012;19(7):2186-94.

131. Turrini O, Lambaudie E, Faucher M, Viret F, Blache JL, Houvenaeghel G, et al. Initial experience with hyperthermic intraperitoneal chemotherapy. *Arch Surg*. 2012;147(10):919-23.
132. Haslinger M, Francescutti V, Attwood K, McCart JA, Fakih M, Kane JM, 3rd, et al. A contemporary analysis of morbidity and outcomes in cytoreduction/hyperthermic intraperitoneal chemoperfusion. *Cancer Med*. 2013;2(3):334-42.
133. Yonemura Y, Canbay E, Ishibashi H. Prognostic factors of peritoneal metastases from colorectal cancer following cytoreductive surgery and perioperative chemotherapy. *ScientificWorldJournal*. 2013;2013:978394.
134. Lin EK, Hsieh MC, Chen CH, Lu YJ, Wu SY. Outcomes of cytoreductive surgery and hyperthermic intraperitoneal chemotherapy for colorectal cancer with peritoneal metastasis. *Medicine (Baltimore)*. 2016;95(52):e5522.
135. Froysnes IS, Larsen SG, Spasojevic M, Dueland S, Flatmark K. Complete cytoreductive surgery and hyperthermic intraperitoneal chemotherapy for colorectal peritoneal metastasis in Norway: Prognostic factors and oncologic outcome in a national patient cohort. *Journal of Surgical Oncology*. 2016;114(2):222-7.
136. Huang Y, Alzahrani NA, Chua TC, Huo YR, Morris DL. Clinical Outcomes of Patients with Extensive Peritoneal Carcinomatosis Following Cytoreductive Surgery and Perioperative Intraperitoneal Chemotherapy. *Anticancer Research*. 2016;36(3):1033-40.
137. Moran B, Cecil T, Chandrakumaran K, Arnold S, Mohamed F, Venkatasubramaniam A. The results of cytoreductive surgery and hyperthermic intraperitoneal chemotherapy in 1200 patients with peritoneal malignancy. *Colorectal Dis*. 2015;17(9):772-8.
138. Desantis M, Bernard JL, Casanova V, Cegarra-Escolano M, Benizri E, Rahili AM, et al. Morbidity, mortality, and oncological outcomes of 401 consecutive cytoreductive procedures with hyperthermic intraperitoneal chemotherapy (HIPEC). *Langenbecks Arch Surg*. 2015;400(1):37-48.
139. Teo MC, Ching Tan GH, Lim C, Chia CS, Tham CK, Soo KC. Colorectal peritoneal carcinomatosis treated with cytoreductive surgery and hyperthermic intraperitoneal chemotherapy: the experience of a tertiary Asian center. *Asian J Surg*. 2015;38(2):65-73.

140. Esquivel J, Sticca R, Sugarbaker P, Levine E, Yan T, Alexander R, et al. Cytoreductive surgery and hyperthermic intraperitoneal chemotherapy in the management of peritoneal surface malignancies of colonic origin: a consensus statement. 2007. p. 128–33.
141. Franko J, Gusani NJ, Holtzman MP, Ahrendt SA, Jones HL, Zeh HJ, 3rd, et al. Multivisceral resection does not affect morbidity and survival after cytoreductive surgery and chemoperfusion for carcinomatosis from colorectal cancer. *Annals of surgical oncology*. 2008;15(11):3065-72.
142. Yan TD, Morris DL. Cytoreductive surgery and perioperative intraperitoneal chemotherapy for isolated colorectal peritoneal carcinomatosis: experimental therapy or standard of care? *Annals of Surgery*. 2008;248(5):829-35.
143. Mirnezami R, Moran BJ, Harvey K, Cecil T, Chandrakumaran K, Carr N, et al. Cytoreductive surgery and intraperitoneal chemotherapy for colorectal peritoneal metastases. *World J Gastroenterol*. 2014;20(38):14018-32.
144. Simkens GA, van Oudheusden TR, Luyer MD, Nienhuijs SW, Nieuwenhuijzen GA, Rutten HJ, et al. Predictors of Severe Morbidity After Cytoreductive Surgery and Hyperthermic Intraperitoneal Chemotherapy for Patients With Colorectal Peritoneal Carcinomatosis. *Annals of surgical oncology*. 2016;23(3):833-41.
145. Simkens GA, van Oudheusden TR, Braam HJ, Luyer MD, Wiezer MJ, van Ramshorst B, et al. Treatment-Related Mortality After Cytoreductive Surgery and HIPEC in Patients with Colorectal Peritoneal Carcinomatosis is Underestimated by Conventional Parameters. *Annals of surgical oncology*. 2016;23(1):99-105.
146. Cree IA, Uttley L, Buckley Woods H, Kikuchi H, Reiman A, Harnan S, et al. The evidence base for circulating tumour DNA blood-based biomarkers for the early detection of cancer: a systematic mapping review. *BMC Cancer*. 2017;17(1):697.
147. Heitzer E, Haque I, Roberts C, Speicher M. Current and future perspectives of liquid biopsies in genomics- driven oncology *Nature Reviews Genetics*. 2019;20:71 – 88.
148. Wan L, Pantel K, Kang Y. Tumor metastasis: moving new biological insights into the clinic. *Nat Med*. 2013;19(11):1450-64.

149. Yokota T, Ura T, Shibata N, Takahari D, Shitara K, Nomura M, et al. BRAF mutation is a powerful prognostic factor in advanced and recurrent colorectal cancer. *British journal of cancer*. 2011;104(5):856-62.
150. Sasaki Y, Hamaguchi T, Yamada Y, Takahashi N, Shoji H, Honma Y, et al. Value of KRAS, BRAF, and PIK3CA Mutations and Survival Benefit from Systemic Chemotherapy in Colorectal Peritoneal Carcinomatosis. *Asian Pac J Cancer Prev*. 2016;17(2):539-43.
151. Kawazoe A, Shitara K, Fukuoka S, Kuboki Y, Bando H, Okamoto W, et al. A retrospective observational study of clinicopathological features of KRAS, NRAS, BRAF and PIK3CA mutations in Japanese patients with metastatic colorectal cancer. *BMC Cancer*. 2015;15:258.
152. Lan YT, Jen-Kou L, Lin CH, Yang SH, Lin CC, Wang HS, et al. Mutations in the RAS and PI3K pathways are associated with metastatic location in colorectal cancers. *J Surg Oncol*. 2015;111(7):905-10.
153. Wang X, Wei Q, Gao J, Li J, Gong J, Li Y, et al. Clinicopathologic features and treatment efficacy of Chinese patients with BRAF-mutated metastatic colorectal cancer: a retrospective observational study. *Chin J Cancer*. 2017;36(1):81.
154. Shirahata A, Shinmura K, Kitamura Y, Sakuraba K, Yokomizo K, Goto T, et al. MACC1 as a marker for advanced colorectal carcinoma. *Anticancer Res*. 2010;30(7):2689-92.
155. Sakuraba K, Yasuda T, Sakata M, Kitamura YH, Shirahata A, Goto T, et al. Down-regulation of Tip60 gene as a potential marker for the malignancy of colorectal cancer. *Anticancer Res*. 2009;29(10):3953-5.
156. Chan JY, Ong CW, Salto-Tellez M. Overexpression of neurone glial-related cell adhesion molecule is an independent predictor of poor prognosis in advanced colorectal cancer. *Cancer Sci*. 2011;102(10):1855-61.
157. Tan F, Liu F, Liu H, Hu Y, Liu D, Li G. CTHRC1 is associated with peritoneal carcinomatosis in colorectal cancer: a new predictor for prognosis. *Med Oncol*. 2013;30(1):473.
158. El-Deiry WS, Vijayvergia N, Xiu J, Scicchitano A, Lim B, Yee NS, et al. Molecular profiling of 6,892 colorectal cancer samples suggests different possible

treatment options specific to metastatic sites. *Cancer Biol Ther.* 2015;16(12):1726-37.

159. Corp I. IBM SPSS Statistics for Windows, Version 24.0 Armonk, NY IBM Corp; Released 2016 [

160. Khakoo S, Georgiou A, Gerlinger M, Cunningham D, Starling N. Circulating tumour DNA, a promising biomarker for the management of colorectal cancer. *Crit Rev Oncol Hematol.* 2018;122:72-82.

161. Necsulea A, Kaessmann H. Evolutionary dynamics of coding and non-coding transcriptomes. *Nat Rev Genet.* 2014;15(11):734-48.

162. Clancy S, Brown W. Translation: DNA to mRNA to Protein. *Nature Education.* 2008;1(1):101.

163. Gibson G, Muse S. *A Primer of Genome Science.* USA: Sinauer; 2002.

164. Afzal M, Manzoor I, Kuipers OP. A Fast and Reliable Pipeline for Bacterial Transcriptome Analysis Case study: Serine-dependent Gene Regulation in *Streptococcus pneumoniae*. *J Vis Exp.* 2015(98).

165. Zhao S, Fung-Leung WP, Bittner A, Ngo K, Liu X. Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells. *PLoS One.* 2014;9(1):e78644.

166. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet.* 2009;10(1):57-63.

167. Lowe R, Shirley N, Bleackley M, Dolan S, Shafee T. Transcriptomics technologies. *PLoS Comput Biol.* 2017;13(5):e1005457.

168. Oshlack A, Robinson MD, Young MD. From RNA-seq reads to differential expression results. *Genome Biol.* 2010;11(12):220.

169. Lexogen. quantseq-workflow 2019 [Available from: <https://www.lexogen.com/quantseq-workflow/>].

170. Eide PW, Bruun J, Lothe RA, Sveen A. CMScaller: an R package for consensus molecular subtyping of colorectal cancer pre-clinical models. *Scientific reports.* 2017;7(1):16618.

171. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A.* 1977;74(12):5463-7.

172. Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet.* 2016;17(6):333-51.
173. Shin J, Ming GL, Song H. Decoding neural transcriptomes and epigenomes via high-throughput sequencing. *Nat Neurosci.* 2014;17(11):1463-75.
174. Robertson KD, Uzvolgyi E, Liang G, Talmadge C, Sumegi J, Gonzales FA, et al. The human DNA methyltransferases (DNMTs) 1, 3a and 3b: coordinate mRNA expression in normal tissues and overexpression in tumors. *Nucleic Acids Res.* 1999;27(11):2291-8.
175. Zakhari S. Alcohol metabolism and epigenetics changes. *Alcohol Res.* 2013;35(1):6-16.
176. Laird PW. Principles and challenges of genomewide DNA methylation analysis. *Nat Rev Genet.* 2010;11(3):191-203.
177. Dahl C, Guldborg P. DNA methylation analysis techniques. *Biogerontology.* 2003;4(4):233-50.
178. Illumina. Infinium-methylation assay, beadarray technology [Available from: <https://emea.illumina.com/science/technology/beadarray-technology/infinium-methylation-assay.html>].
179. Feber A, Guilhamon P, Lechner M, Fenton T, Wilson GA, Thirlwell C, et al. Using high-density DNA methylation arrays to profile copy number alterations. *Genome Biol.* 2014;15(2):R30.
180. Morris TJ, Butcher LM, Feber A, Teschendorff AE, Chakravarthy AR, Wojdacz TK, et al. ChAMP: 450k Chip Analysis Methylation Pipeline. *Bioinformatics.* 2014;30(3):428-30.
181. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics.* 2014;30(10):1363-9.
182. Zhou W, Laird PW, Shen H. Comprehensive characterization, annotation and innovative use of Infinium DNA methylation BeadChip probes. *Nucleic Acids Res.* 2017;45(4):e22.
183. Fortin JP, Triche TJ, Hansen KD. Preprocessing, normalization and integration of the Illumina HumanMethylationEPIC array with minfi. *Bioinformatics.* 2017;33(4):558-60.

184. Teschendorff AE, Marabita F, Lechner M, Bartlett T, Tegner J, Gomez-Cabrero D, et al. A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. *Bioinformatics*. 2013;29(2):189-96.
185. Teschendorff AE, Menon U, Gentry-Maharaj A, Ramus SJ, Gayther SA, Apostolidou S, et al. An epigenetic signature in peripheral blood predicts active ovarian cancer. *PLoS One*. 2009;4(12):e8274.
186. Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol*. 2004;3:Article3.
187. Jaffe AE, Murakami P, Lee H, Leek JT, Fallin MD, Feinberg AP, et al. Bump hunting to identify differentially methylated regions in epigenetic epidemiology studies. *Int J Epidemiol*. 2012;41(1):200-9.
188. Choi M, Scholl UI, Ji W, Liu T, Tikhonova IR, Zumbo P, et al. Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proc Natl Acad Sci U S A*. 2009;106(45):19096-101.
189. Teer J, Mulikin J. Exome sequencing: The sweet spot before whole genomes. *Human Molecular Genetics*. 2010;19:R145-51.
190. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2010;26(5):589-95.
191. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009;25(16):2078-9.
192. Saunders CT, Wong WS, Swamy S, Becq J, Murray LJ, Cheetham RK. Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinformatics*. 2012;28(14):1811-7.
193. Qiagen. **CLC Genomics Workbench 11.0** [Available from: <https://www.qiagenbioinformatics.com>].
194. Gonzalez-Perez A, Perez-Llamas C, Deu-Pons J, Tamborero D, Schroeder MP, Jene-Sanz A, et al. IntOGen-mutations identifies cancer drivers across tumor types. *Nat Methods*. 2013;10(11):1081-2.
195. Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res*. 2003;31(13):3812-4.

196. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet*. 2013;Chapter 7:Unit7.20.
197. Reva B, Antipin Y, Sander C. Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic Acids Res*. 2011;39(17):e118.
198. Huang MN, McPherson JR, Cutcutache I, Teh BT, Tan P, Rozen SG. MSIseq: Software for Assessing Microsatellite Instability from Catalogs of Somatic Mutations. *Scientific reports*. 2015;5:13321.
199. Nakken S. MSI classification from somatic mutation profiles 2018 [Available from: https://rstudio-pubs-static.s3.amazonaws.com/387512_bb83ba45ab7446fe940464695c384f2d.html.
200. Qiagen. QIAseq Targeted DNA Panel Handbook 2019 [Available from: <https://www.qiagen.com/gb/resources/resourcedetail?id=8907edbe-a462-4883-ae1b-2759657e7fd0&lang=en>.
201. Han X, Wang J, Sun Y. Circulating Tumor DNA as Biomarkers for Cancer Detection. *Genomics Proteomics Bioinformatics*. 2017;15(2):59-72.
202. Mayakonda A, Lin DC, Assenov Y, Plass C, Koeffler HP. Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome Res*. 2018;28(11):1747-56.
203. Hallam S, Tyler R, Price M, Beggs A, Youssef H. Meta-analysis of prognostic factors for patients with colorectal peritoneal metastasis undergoing cytoreductive surgery and heated intraperitoneal chemotherapy. *BJS Open*. 2019;3(5):585-94.
204. Moher D, Liberati A, Tetzlaff J, Altman DG, Group P. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *J Clin Epidemiol*. 2009;62(10):1006-12.
205. Hayden JA, van der Windt DA, Cartwright JL, Côté P, Bombardier C. Assessing bias in studies of prognostic factors. *Ann Intern Med*. 2013;158(4):280-6.
206. Tierney JF, Stewart LA, Gherzi D, Burdett S, Sydes MR. Practical methods for incorporating summary time-to-event data into meta-analysis. *Trials*. 2007;8:16.
207. Guyot P, Ades A, Ouwers M, Welton N. Enhanced secondary analysis of survival data: reconstructing the data from published Kaplan-Meier survival curves. *BMC Medical Research Methodology*. 2012;12(9):1471-2288.

208. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials*. 1986;7(3):177-88.
209. IntHout J, Ioannidis JP, Rovers MM, Goeman JJ. Plea for routinely presenting prediction intervals in meta-analysis. *BMJ Open*. 2016;6(7):e010247.
210. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ*. 2003;327(7414):557-60.
211. Baratti D, Kusamura S, Iusco D, Bonomi S, Grassi A, Virzi S, et al. Postoperative complications after cytoreductive surgery and hyperthermic intraperitoneal chemotherapy affect long-term outcome of patients with peritoneal metastases from colorectal cancer: a two-center study of 101 patients. *Dis Colon Rectum*. 2014;57(7):858-68.
212. Benizri EI, Bernard JL, Rahili A, Benchimol D, Bereder JM. Small bowel involvement is a prognostic factor in colorectal carcinomatosis treated with complete cytoreductive surgery plus hyperthermic intraperitoneal chemotherapy. *World J Surg Oncol*. 2012;10:56.
213. da Silva RG, Sugarbaker PH. Analysis of prognostic factors in seventy patients having a complete cytoreduction plus perioperative intraperitoneal chemotherapy for carcinomatosis from colorectal cancer. *Journal of the American College of Surgeons*. 2006;203(6):878-86.
214. Duraj FF, Cashin PH. Cytoreductive surgery and intraperitoneal chemotherapy for colorectal peritoneal and hepatic metastases: a case-control study. *J Gastrointest Oncol*. 2013;4(4):388-96.
215. Elias D, Gilly F, Boutitie F, Quenet F, Bereder JM, Mansvelt B, et al. Peritoneal colorectal carcinomatosis treated with surgery and perioperative intraperitoneal chemotherapy: retrospective analysis of 523 patients from a multicentric French study. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. 2010;28(1):63-8.
216. Elias D, Borget I, Farron M, Dromain C, Ducreux M, Goere D, et al. Prognostic significance of visible cardiophrenic angle lymph nodes in the presence of peritoneal metastases from colorectal cancers. *European Journal of Surgical Oncology*. 2013;39(11):1214-8.

217. Elias D, Mariani A, Cloutier AS, Blot F, Goere D, Dumont F, et al. Modified selection criteria for complete cytoreductive surgery plus HIPEC based on peritoneal cancer index and small bowel involvement for peritoneal carcinomatosis of colorectal origin. *European Journal of Surgical Oncology*. 2014;40(11):1467-73.
218. Faron M, Macovei R, Goere D, Honore C, Benhaim L, Elias D. Linear Relationship of Peritoneal Cancer Index and Survival in Patients with Peritoneal Metastases from Colorectal Cancer. *Annals of surgical oncology*. 2016;23(1):114-9.
219. Huang CQ, Yang XJ, Yu Y, Wu HT, Liu Y, Yonemura Y, et al. Cytoreductive surgery plus hyperthermic intraperitoneal chemotherapy improves survival for patients with peritoneal carcinomatosis from colorectal cancer: a phase II study from a Chinese center. *PLoS One*. 2014;9(9):e108509.
220. Huang Y, Alzahrani NA, Chua TC, Liauw W, Morris DL. Impacts of peritoneal cancer index on the survival outcomes of patients with colorectal peritoneal carcinomatosis. *Int J Surg*. 2016;32:65-70.
221. Ihemelandu C, Sugarbaker PH. Management for Peritoneal Metastasis of Colonic Origin: Role of Cytoreductive Surgery and Perioperative Intraperitoneal Chemotherapy: A Single Institution's Experience During Two Decades. *Annals of surgical oncology*. 2017;24(4):898-905.
222. Lorimier G, Linot B, Paillocher N, Dupoirion D, Verri  le V, Wernert R, et al. Curative cytoreductive surgery followed by hyperthermic intraperitoneal chemotherapy in patients with peritoneal carcinomatosis and synchronous resectable liver metastases arising from colorectal cancer. *Eur J Surg Oncol*. 2017;43(1):150-8.
223. Maggiori L, Go  r   D, Viana B, Tzanis D, Dumont F, Honor   C, et al. Should patients with peritoneal carcinomatosis of colorectal origin with synchronous liver metastases be treated with a curative intent? A case-control study. *Ann Surg*. 2013;258(1):116-21.
224. Ng JL, Ong WS, Chia CS, Tan GH, Soo KC, Teo MC. Prognostic Relevance of the Peritoneal Surface Disease Severity Score Compared to the Peritoneal Cancer Index for Colorectal Peritoneal Carcinomatosis. *Int J Surg Oncol*. 2016;2016:2495131.
225. Simkens GA, Razenberg LG, Lemmens VE, Rutten HJ, Creemers GJ, de Hingh IH. Histological subtype and systemic metastases strongly influence treatment

and survival in patients with synchronous colorectal peritoneal metastases. *European Journal of Surgical Oncology*. 2016;42(6):794-800.

226. Simkens GA, van Oudheusden TR, Braam HJ, Wiezer MJ, Nienhuijs SW, Rutten HJ, et al. Cytoreductive surgery and HIPEC offers similar outcomes in patients with rectal peritoneal metastases compared to colon cancer patients: a matched case control study. *J Surg Oncol*. 2016;113(5):548-53.

227. Sluiter NR, de Cuba EM, Kwakman R, Meijerink WJ, Delis-van Diemen PM, Coupe VM, et al. Versican and vascular endothelial growth factor expression levels in peritoneal metastases from colorectal cancer are associated with survival after cytoreductive surgery and hyperthermic intraperitoneal chemotherapy. *Clin Exp Metastasis*. 2016;33(4):297-307.

228. Ung L, Chua TC, David L M. Peritoneal metastases of lower gastrointestinal tract origin:a comparative study of patient outcomes following cytoreduction and intraperitoneal chemotherapy. *J Cancer Res Clin Oncol*. 2013;139(11):1899-908.

229. Winer J, Zenati M, Ramalingam L, Jones H, Zureikat A, Holtzman M, et al. Impact of aggressive histology and location of primary tumor on the efficacy of surgical therapy for peritoneal carcinomatosis of colorectal origin. *Annals of surgical oncology*. 2014;21(5):1456-62.

230. Votanopoulos KI, Swett K, Blackham AU, Ihemelandu C, Shen P, Stewart JH, et al. Cytoreductive surgery with hyperthermic intraperitoneal chemotherapy in peritoneal carcinomatosis from rectal cancer. *Annals of surgical oncology*. 2013;20(4):1088-92.

231. Rivard JD, McConnell YJ, Temple WJ, Mack LA. Cytoreduction and heated intraperitoneal chemotherapy for colorectal cancer: are we excluding patients who may benefit? *J Surg Oncol*. 2014;109(2):104-9.

232. Füzün M, Sökmen S, Terzi C, Canda AE. Cytoreductive approach to peritoneal carcinomatosis originated from colorectal cancer: Turkish experience. *Acta Chir Iugosl*. 2006;53(2):17-21.

233. Chen HS, Sheen-Chen SM. Obstruction and perforation in colorectal adenocarcinoma: an analysis of prognosis and current trends. *Surgery*. 2000;127(4):370-6.

234. Dohan A, Hoeffel C, Soyer P, Jannot AS, Valette PJ, Thivolet A, et al. Evaluation of the peritoneal carcinomatosis index with CT and MRI. *Br J Surg*. 2017;104(9):1244-9.
235. Flicek K, Ashfaq A, Johnson CD, Menias C, Bagaria S, Wasif N. Correlation of Radiologic with Surgical Peritoneal Cancer Index Scores in Patients with Pseudomyxoma Peritonei and Peritoneal Carcinomatosis: How Well Can We Predict Resectability? *J Gastrointest Surg*. 2016;20(2):307-12.
236. Iversen LH, Rasmussen PC, Laurberg S. Value of laparoscopy before cytoreductive surgery and hyperthermic intraperitoneal chemotherapy for peritoneal carcinomatosis. *Br J Surg*. 2013;100(2):285-92.
237. Kwakman R, Schrama AM, van Olmen JP, Otten RH, de Lange-de Klerk ES, de Cuba EM, et al. Clinicopathological Parameters in Patient Selection for Cytoreductive Surgery and Hyperthermic Intraperitoneal Chemotherapy for Colorectal Cancer Metastases: A Meta-analysis. *Ann Surg*. 2016;263(6):1102-11.
238. Chua TC, Yan TD, Ng KM, Zhao J, Morris DL. Significance of lymph node metastasis in patients with colorectal cancer peritoneal carcinomatosis. *World Journal of Surgery*. 2009;33(7):1488-94.
239. Elias D, Glehen O, Pocard M, Quenet F, Goéré D, Arvieux C, et al. A comparative study of complete cytoreductive surgery plus intraperitoneal chemotherapy to treat peritoneal dissemination from colon, rectum, small bowel, and nonpseudomyxoma appendix. *Ann Surg*. 2010;251(5):896-901.
240. Baumgartner JM, Tobin L, Heavey SF, Kelly KJ, Roeland EJ, Lowy AM. Predictors of progression in high-grade appendiceal or colorectal peritoneal carcinomatosis after cytoreductive surgery and hyperthermic intraperitoneal chemotherapy. *Annals of surgical oncology*. 2015;22(5):1716-21.
241. Schneider MA, Eden J, Pache B, Laminger F, Lopez-Lopez V, Steffen T, et al. Mutations of RAS/RAF Proto-oncogenes Impair Survival After Cytoreductive Surgery and HIPEC for Peritoneal Metastasis of Colorectal Origin. *Ann Surg*. 2018;268(5):845-53.
242. Hallam S, Stockton J, Bryer C, Whalley C, Pestinger V, Youssef H, et al. The transition from primary colorectal cancer to isolated peritoneal malignancy is

associated with an increased tumour mutational burden. *Scientific reports*. 2020;10(1):18900.

243. Friederichs J, Rosenberg R, Mages J, Janssen KP, Maeckl C, Nekarda H, et al. Gene expression profiles of different clinical stages of colorectal carcinoma: toward a molecular genetic understanding of tumor progression. *Int J Colorectal Dis*. 2005;20(5):391-402.
244. Levine EA, Blazer DG, 3rd, Kim MK, Shen P, Stewart JHt, Guy C, et al. Gene expression profiling of peritoneal metastases from appendiceal and colon cancer demonstrates unique biologic signatures and predicts patient outcomes. *Journal of the American College of Surgeons*. 2012;214(4):599-606; discussion -7.
245. Ju HX, An B, Okamoto Y, Shinjo K, Kanemitsu Y, Komori K, et al. Distinct profiles of epigenetic evolution between colorectal cancers with and without metastasis. *Am J Pathol*. 2011;178(4):1835-46.
246. Konishi K, Watanabe Y, Shen L, Guo Y, Castoro RJ, Kondo K, et al. DNA methylation profiles of primary colorectal carcinoma and matched liver metastasis. *PLoS One*. 2011;6(11):e27889.
247. Murata A, Baba Y, Watanabe M, Shigaki H, Miyake K, Ishimoto T, et al. Methylation levels of LINE-1 in primary lesion and matched metastatic lesions of colorectal cancer. *British journal of cancer*. 2013;109(2):408-15.
248. Zhang X, Shimodaira H, Soeda H, Komine K, Takahashi H, Ouchi K, et al. CpG island methylator phenotype is associated with the efficacy of sequential oxaliplatin- and irinotecan-based chemotherapy and EGFR-related gene mutation in Japanese patients with metastatic colorectal cancer. *Int J Clin Oncol*. 2016;21(6):1091-101.
249. Miranda E, Destro A, Malesci A, Balladore E, Bianchi P, Baryshnikova E, et al. Genetic and epigenetic changes in primary metastatic and nonmetastatic colorectal cancer. *British journal of cancer*. 2006;95(8):1101-7.
250. Sharrard RM, Royds JA, Rogers S, Shorthouse AJ. Patterns of methylation of the c-myc gene in human colorectal cancer progression. *British journal of cancer*. 1992;65(5):667-72.
251. Siemens H, Neumann J, Jackstadt R, Mansmann U, Horst D, Kirchner T, et al. Detection of miR-34a promoter methylation in combination with elevated expression

- of c-Met and β -catenin predicts distant metastasis of colon cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research*. 2013;19(3):710-20.
252. Tang D, Liu J, Wang DR, Yu HF, Li YK, Zhang JQ. Diagnostic and prognostic value of the methylation status of secreted frizzled-related protein 2 in colorectal cancer. *Clin Invest Med*. 2011;34(2):E88-95.
253. Kim SH, Park KH, Shin SJ, Lee KY, Kim TI, Kim NK, et al. CpG Island Methylator Phenotype and Methylation of Wnt Pathway Genes Together Predict Survival in Patients with Colorectal Cancer. *Yonsei Med J*. 2018;59(5):588-94.
254. Lee SY, Haq F, Kim D, Jun C, Jo HJ, Ahn SM, et al. Comparative genomic analysis of primary and synchronous metastatic colorectal cancers. *PLoS One*. 2014;9(3):e90459.
255. Goryca K, Kulecka M, Paziewska A, Dabrowska M, Grzelak M, Skrzypczak M, et al. Exome scale map of genetic alterations promoting metastasis in colorectal cancer. *BMC Genet*. 2018;19(1):85.
256. Mlecnik B, Bindea G, Kirilovsky A, Angell HK, Obenauf AC, Tosolini M, et al. The tumor microenvironment and Immunoscore are critical determinants of dissemination to distant metastasis. *Sci Transl Med*. 2016;8(327):327ra26.
257. Huang D, Sun W, Zhou Y, Li P, Chen F, Chen H, et al. Mutations of key driver genes in colorectal cancer progression and metastasis. *Cancer Metastasis Rev*. 2018;37(1):173-87.
258. Bernstein C, Bernstein H, Garewal H, Dinning P, Jabi R, Sampliner RE, et al. A bile acid-induced apoptosis assay for colon cancer risk and associated quality control studies. *Cancer Res*. 1999;59(10):2353-7.
259. Ohmachi T, Inoue H, Mimori K, Tanaka F, Sasaki A, Kanda T, et al. Fatty acid binding protein 6 is overexpressed in colorectal cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research*. 2006;12(17):5090-5.
260. Radeva MY, Jahns F, Wilhelm A, Gleis M, Settmacher U, Greulich KO, et al. Defensin alpha 6 (DEFA 6) overexpression threshold of over 60 fold can distinguish between adenoma and fully blown colon carcinoma in individual patients. *BMC Cancer*. 2010;10:588.

261. Safran M, Dalah I, Alexander J, Rosen N, Iny Stein T, Shmoish M, et al. GeneCards Version 3: the human gene integrator. Database (Oxford). 2010;2010:baq020.
262. Du J, Guan M, Fan J, Jiang H. Loss of DMBT1 expression in human prostate cancer and its correlation with clinical progressive features. Urology. 2011;77(2):509.e9-13.
263. Fehlker M, Huska MR, Jöns T, Andrade-Navarro MA, Kemmner W. Concerted down-regulation of immune-system related genes predicts metastasis in colorectal carcinoma. BMC Cancer. 2014;14:64.
264. van der Flier LG, Haegebarth A, Stange DE, van de Wetering M, Clevers H. OLFM4 is a robust marker for stem cells in human intestine and marks a subset of colorectal cancer cells. Gastroenterology. 2009;137(1):15-7.
265. Liu W, Li H, Hong SH, Piszczek GP, Chen W, Rodgers GP. Olfactomedin 4 deletion induces colon adenocarcinoma in Apc. Oncogene. 2016;35(40):5237-47.
266. Seko N, Oue N, Noguchi T, Sentani K, Sakamoto N, Hinoi T, et al. Olfactomedin 4 (GW112, hGC-1) is an independent prognostic marker for survival in patients with colorectal cancer. Exp Ther Med. 2010;1(1):73-8.
267. Hsu HP, Lai MD, Lee JC, Yen MC, Weng TY, Chen WC, et al. Mucin 2 silencing promotes colon cancer metastasis through interleukin-6 signaling. Scientific reports. 2017;7(1):5823.
268. Shaojun C, Li H, Haixin H, Guisheng L. Expression of Topoisomerase 1 and carboxylesterase 2 correlates with irinotecan treatment response in metastatic colorectal cancer. Cancer Biol Ther. 2018;19(3):153-9.
269. Espinasse MA, Pépin A, Virault-Rocroy P, Szely N, Chollet-Martin S, Pallardy M, et al. Glucocorticoid-Induced Leucine Zipper Is Expressed in Human Neutrophils and Promotes Apoptosis through Mcl-1 Down-Regulation. J Innate Immun. 2016;8(1):81-96.
270. Yunta M, Lazo PA. Apoptosis protection and survival signal by the CD53 tetraspanin antigen. Oncogene. 2003;22(8):1219-24.
271. Huang X, Xiang L, Li Y, Zhao Y, Zhu H, Xiao Y, et al. Snail/FOXK1/Cyr61 Signaling Axis Regulates the Epithelial-Mesenchymal Transition and Metastasis in Colorectal Cancer. Cell Physiol Biochem. 2018;47(2):590-603.

272. Song YF, Xu ZB, Zhu XJ, Tao X, Liu JL, Gao FL, et al. Serum Cyr61 as a potential biomarker for diagnosis of colorectal cancer. *Clin Transl Oncol*. 2017;19(4):519-24.
273. Ma JC, Sun XW, Su H, Chen Q, Guo TK, Li Y, et al. Fibroblast-derived CXCL12/SDF-1 α promotes CXCL6 secretion and co-operatively enhances metastatic potential through the PI3K/Akt/mTOR pathway in colon cancer. *World J Gastroenterol*. 2017;23(28):5167-78.
274. Saigusa S, Toiyama Y, Tanaka K, Yokoe T, Okugawa Y, Kawamoto A, et al. Stromal CXCR4 and CXCL12 expression is associated with distant recurrence and poor prognosis in rectal cancer after chemoradiotherapy. *Annals of surgical oncology*. 2010;17(8):2051-8.
275. Ubink I, Verhaar ER, Kranenburg O, Goldschmeding R. A potential role for CCN2/CTGF in aggressive colorectal cancer. *J Cell Commun Signal*. 2016;10(3):223-7.
276. Kos J, Krasovec M, Cimerman N, Nielsen HJ, Christensen IJ, Brünner N. Cysteine proteinase inhibitors stefin A, stefin B, and cystatin C in sera from patients with colorectal cancer: relation to prognosis. *Clinical cancer research : an official journal of the American Association for Cancer Research*. 2000;6(2):505-11.
277. Sosa MS, Parikh F, Maia AG, Estrada Y, Bosch A, Bragado P, et al. NR2F1 controls tumour cell dormancy via SOX9- and RAR β -driven quiescence programmes. *Nat Commun*. 2015;6:6170.
278. Sofeu Feugaing DD, Götte M, Viola M. More than matrix: the multifaceted role of decorin in cancer. *Eur J Cell Biol*. 2013;92(1):1-11.
279. Copija A, Waniczek D, Walkiewicz k, Głogowski L, Augustyniak H, Nowakowska-Zajdel E. PTEN – clinical significance in colorectal cancer. *OncoReview*. 2016;6(2):A86-90.
280. Perkins N. The diverse and complex roles of NF- κ B subunits in cancer. *Nature Reviews Cancer*. 2012;12(2):121-32.
281. Beauchemin N, Arabzadeh A. Carcinoembryonic antigen-related cell adhesion molecules (CEACAMs) in cancer progression and metastasis. *Cancer Metastasis Rev*. 2013;32(3-4):643-71.

282. Bacac M, Fauti T, Sam J, Colombetti S, Weinzierl T, Ouaret D, et al. A Novel Carcinoembryonic Antigen T-Cell Bispecific Antibody (CEA TCB) for the Treatment of Solid Tumors. *Clinical cancer research : an official journal of the American Association for Cancer Research*. 2016;22(13):3286-97.
283. Gu L, Lu L, Zhou D, Liu Z. Long Noncoding RNA BCYRN1 Promotes the Proliferation of Colorectal Cancer Cells via Up-Regulating NPR3 Expression. *Cell Physiol Biochem*. 2018;48(6):2337-49.
284. Hwangbo DS, Biteau B, Rath S, Kim J, Jasper H. Control of apoptosis by *Drosophila* DCAF12. *Dev Biol*. 2016;413(1):50-9.
285. Gerald D, Adini I, Shechter S, Perruzzi C, Varnau J, Hopkins B, et al. RhoB controls coordination of adult angiogenesis and lymphangiogenesis following injury by regulating VEZF1-mediated transcription. *Nat Commun*. 2013;4:2824.
286. Dankert JF, Pagan JK, Starostina NG, Kipreos ET, Pagano M. FEM1 proteins are ancient regulators of SLBP degradation. *Cell Cycle*. 2017;16(6):556-64.
287. Zhang L, Guo L, Peng Y, Chen B. Expression of T-STAR gene is associated with regulation of telomerase activity in human colon cancer cell line HCT-116. *World J Gastroenterol*. 2006;12(25):4056-60.
288. Mooi JK, Wirapati P, Asher R, Lee CK, Savas P, Price TJ, et al. The prognostic impact of consensus molecular subtypes (CMS) and its predictive effects for bevacizumab benefit in metastatic colorectal cancer: molecular analysis of the AGITG MAX clinical trial. *Annals of oncology : official journal of the European Society for Medical Oncology / ESMO*. 2018;29(11):2240-6.
289. Lenz HJ, Ou FS, Venook AP, Hochster HS, Niedzwiecki D, Goldberg RM, et al. Impact of Consensus Molecular Subtype on Survival in Patients With Metastatic Colorectal Cancer: Results From CALGB/SWOG 80405 (Alliance). *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. 2019;37(22):1876-85.
290. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, et al. The human genome browser at UCSC. *Genome Res*. 2002;12(6):996-1006.
291. Hu L, Huang Z, Wu Z, Ali A, Qian A. Mammalian Plakins, Giant Cytolinkers: Versatile Biological Functions and Roles in Cancer. *Int J Mol Sci*. 2018;19(4).

292. Cha Y, Kim KJ, Han SW, Rhee YY, Bae JM, Wen X, et al. Adverse prognostic impact of the CpG island methylator phenotype in metastatic colorectal cancer. *British journal of cancer*. 2016;115(2):164-71.
293. Vigneri PG, Tirrò E, Pennisi MS, Massimino M, Stella S, Romano C, et al. The Insulin/IGF System in Colorectal Cancer Development and Resistance to Therapy. *Front Oncol*. 2015;5:230.
294. Mitchell SM, Ross JP, Drew HR, Ho T, Brown GS, Saunders NF, et al. A panel of genes methylated with high frequency in colorectal cancer. *BMC Cancer*. 2014;14:54.
295. Li J, Bi L, Shi Z, Sun Y, Lin Y, Shao H, et al. RNA-Seq analysis of non-small cell lung cancer in female never-smokers reveals candidate cancer-associated long non-coding RNAs. *Pathol Res Pract*. 2016;212(6):549-54.
296. Chen H, Xu Z, Yang B, Zhou X, Kong H. RASGRF1 Hypermethylation, a Putative Biomarker of Colorectal Cancer. *Ann Clin Lab Sci*. 2018;48(1):3-10.
297. Mousavi Ardehaie R, Hashemzadeh S, Behrouz Sharif S, Ghojazadeh M, Teimoori-Toolabi L, Sakhinia E. Aberrant methylated EDNRB can act as a potential diagnostic biomarker in sporadic colorectal cancer while KISS1 is controversial. *Bioengineered*. 2017;8(5):555-64.
298. Smith SL, Damato BE, Scholes AG, Nunn J, Field JK, Heighway J. Decreased endothelin receptor B expression in large primary uveal melanomas is associated with early clinical metastasis and short survival. *British journal of cancer*. 2002;87(11):1308-13.
299. Romanelli V, Nakabayashi K, Vizoso M, Moran S, Iglesias-Platas I, Sugahara N, et al. Variable maternal methylation overlapping the nc886/vtRNA2-1 locus is locked between hypermethylated repeats and is frequently altered in cancer. *Epigenetics*. 2014;9(5):783-90.
300. Hatakeyama S. TRIM proteins and cancer. *Nat Rev Cancer*. 2011;11(11):792-804.
301. Hishikawa Y, Koji T, Dhar DK, Kinugasa S, Yamaguchi M, Nagasue N. Metallothionein expression correlates with metastatic and proliferative potential in squamous cell carcinoma of the oesophagus. *British journal of cancer*. 1999;81(4):712-20.

302. Hishikawa Y, Kohno H, Ueda S, Kimoto T, Dhar DK, Kubota H, et al. Expression of metallothionein in colorectal cancers and synchronous liver metastases. *Oncology*. 2001;61(2):162-7.
303. Yan Z, Yin H, Wang R, Wu D, Sun W, Liu B, et al. Overexpression of integrin-linked kinase (ILK) promotes migration and invasion of colorectal cancer cells by inducing epithelial-mesenchymal transition via NF- κ B signaling. *Acta Histochem*. 2014;116(3):527-33.
304. Gao XL, Zhang M, Tang YL, Liang XH. Cancer cell dormancy: mechanisms and implications of cancer recurrence and metastasis. *Onco Targets Ther*. 2017;10:5219-28.
305. Huang H, Chen J, Ding CM, Jin X, Jia ZM, Peng J. LncRNA NR2F1-AS1 regulates hepatocellular carcinoma oxaliplatin resistance by targeting ABCC1 via miR-363. *J Cell Mol Med*. 2018;22(6):3238-45.
306. Sidibe A, Ropraz P, Jemelin S, Emre Y, Poittevin M, Pocard M, et al. Angiogenic factor-driven inflammation promotes extravasation of human proangiogenic monocytes to tumours. *Nat Commun*. 2018;9(1):355.
307. Liu T, Zhang X, Yang YM, Du LT, Wang CX. Increased expression of the long noncoding RNA CRNDE-h indicates a poor prognosis in colorectal cancer, and is positively correlated with IRX5 mRNA expression. *Onco Targets Ther*. 2016;9:1437-48.
308. Kashima L, Toyota M, Mita H, Suzuki H, Idogawa M, Ogi K, et al. CHFR, a potential tumor suppressor, downregulates interleukin-8 through the inhibition of NF- κ B. *Oncogene*. 2009;28(29):2643-53.
309. Sun Z, Liu J, Jing H, Dong SX, Wu J. The diagnostic and prognostic value of CHFR hypermethylation in colorectal cancer, a meta-analysis and literature review. *Oncotarget*. 2017;8(51):89142-8.
310. Poh AR, Love CG, Masson F, Preaudet A, Tsui C, Whitehead L, et al. Inhibition of Hematopoietic Cell Kinase Activity Suppresses Myeloid Cell-Mediated Colon Cancer Progression. *Cancer Cell*. 2017;31(4):563-75.e5.
311. Lanner F, Rossant J. The role of FGF/Erk signaling in pluripotent cells. *Development*. 2010;137(20):3351-60.

312. Hinoue T, Weisenberger DJ, Pan F, Campan M, Kim M, Young J, et al. Analysis of the association between CIMP and BRAF in colorectal cancer by DNA methylation profiling. *PLoS One*. 2009;4(12):e8357.
313. Goodman AM, Kato S, Bazhenova L, Patel SP, Frampton GM, Miller V, et al. Tumor Mutational Burden as an Independent Predictor of Response to Immunotherapy in Diverse Cancers. *Mol Cancer Ther*. 2017;16(11):2598-608.
314. IntOGen. IntOGen Colorectal cancer driver genes [Available from: <https://www.intogen.org/search>.
315. Powell E, Piwnica-Worms D, Piwnica-Worms H. Contribution of p53 to metastasis. *Cancer Discov*. 2014;4(4):405-14.
316. Ong D, Ho Y, Rudduck C, Chin K, Kuo W, Lie D, et al. LARG at chromosome 11q23 has functional characteristics of a tumor suppressor in human breast and colorectal cancer. *Oncogene*. 2010;26(28):4189–200.
317. Ding C, He J, Liao W, Chen J, Li J, He G, et al. Regulation of WNT/ β -catenin signaling by carbamoyl-phosphate synthetase 2, aspartate transcarbamylase and dihydroorotase (CAD) in colorectal cancer cell. *International Journal of Clinical and Experimental Medicine*. 2017;10(12):16243-53.
318. Bandaru S, Zhou AX, Rouhi P, Zhang Y, Bergo MO, Cao Y, et al. Targeting filamin B induces tumor growth and metastasis via enhanced activity of matrix metalloproteinase-9 and secretion of VEGF-A. *Oncogenesis*. 2014;3:e119.
319. Li J, Choi PS, Chaffer CL, Labella K, Hwang JH, Giacomelli AO, et al. An alternative splicing switch in FLNB promotes the mesenchymal cell state in human breast cancer. *Elife*. 2018;7.
320. Li X, Zhang G, Wang Y, Elgehama A, Sun Y, Li L, et al. Loss of periplakin expression is associated with the tumorigenesis of colorectal carcinoma. *Biomed Pharmacother*. 2017;87:366-74.
321. Bonomi S, Gallo S, Catillo M, Pignataro D, Biamonti G, Ghigna C. Oncogenic alternative splicing switches: role in cancer progression and prospects for therapy. *Int J Cell Biol*. 2013;2013:962038.
322. Read RD, Fenton TR, Gomez GG, Wykosky J, Vandenberg SR, Babic I, et al. A kinome-wide RNAi screen in *Drosophila* Glia reveals that the RIO kinases mediate

cell proliferation and survival through TORC2-Akt signaling in glioblastoma. *PLoS Genet.* 2013;9(2):e1003253.

323. Kim EJ, Park MK, Byun HJ, Kang GJ, Yu L, Kim HJ, et al. YdjC chitooligosaccharide deacetylase homolog induces keratin reorganization in lung cancer cells: involvement of interaction between YDJC and CDC16. *Oncotarget.* 2018;9(33):22915-28.

324. Kloor M, Becker C, Benner A, Woerner SM, Gebert J, Ferrone S, et al. Immunoselective pressure and human leukocyte antigen class I antigen machinery defects in microsatellite unstable colorectal cancers. *Cancer Res.* 2005;65(14):6418-24.

325. Rasmussen K, Helin K. Role of TET enzymes in DNA methylation, development, and cancer. *Genes and development.* 2016;30(7):733-50.

326. Huang Y, Wang G, Liang Z, Yang Y, Cui L, Liu CY. Loss of nuclear localization of TET2 in colorectal cancer. *Clin Epigenetics.* 2016;8:9.

327. Tahara T, Yamamoto E, Madireddi P, Suzuki H, Maruyama R, Chung W, et al. Colorectal carcinomas with CpG island methylator phenotype 1 frequently contain mutations in chromatin regulators. *Gastroenterology.* 2014;146(2):530-38.e5.

328. Hao HX, Xie Y, Zhang Y, Charlat O, Oster E, Avello M, et al. ZNRF3 promotes Wnt receptor turnover in an R-spondin-sensitive manner. *Nature.* 2012;485(7397):195-200.

329. Sanz-Pamplona R, Lopez-Doriga A, Paré-Brunet L, Lázaro K, Bellido F, Alonso MH, et al. Exome Sequencing Reveals AMER1 as a Frequently Mutated Gene in Colorectal Cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research.* 2015;21(20):4709-18.

330. Lee SJ, Kang BW, Chae YS, Kim HJ, Park SY, Park JS, et al. Genetic variations in STK11, PRKAA1, and TSC1 associated with prognosis for patients with colorectal cancer. *Annals of surgical oncology.* 2014;21 Suppl 4:S634-9.

331. Mak BC, Takemaru K, Kenerson HL, Moon RT, Yeung RS. The tuberin-hamartin complex negatively regulates beta-catenin signaling activity. *J Biol Chem.* 2003;278(8):5947-51.

332. Wang Y, Xue J, Kuang H, Zhou X, Liao L, Yin F. microRNA-1297 Inhibits the Growth and Metastasis of Colorectal Cancer by Suppressing Cyclin D2 Expression. *DNA Cell Biol.* 2017;36(11):991-9.
333. Jin Z, Chung JW, Mei W, Strack S, He C, Lau GW, et al. Regulation of nuclear-cytoplasmic shuttling and function of Family with sequence similarity 13, member A (Fam13a), by B56-containing PP2As and Akt. *Mol Biol Cell.* 2015;26(6):1160-73.
334. Jiang Z, Lao T, Qiu W, Polverino F, Gupta K, Guo F, et al. A Chronic Obstructive Pulmonary Disease Susceptibility Gene, FAM13A, Regulates Protein Stability of β -Catenin. *Am J Respir Crit Care Med.* 2016;194(2):185-97.
335. Eisenhut F, Heim L, Trump S, Mittler S, Sopel N, Andreev K, et al. FAM13A is associated with non-small cell lung cancer (NSCLC) progression and controls tumor cell proliferation and survival. *Oncoimmunology.* 2017;6(1):e1256526.
336. Pardo-Pastor C, Rubio-Moscardo F, Vogel-González M, Serra SA, Afthinos A, Mrkonjic S, et al. Piezo2 channel regulates RhoA and actin cytoskeleton to promote cell mechanobiological responses. *Proc Natl Acad Sci U S A.* 2018;115(8):1925-30.
337. Yang H, Liu C, Zhou RM, Yao J, Li XM, Shen Y, et al. Piezo2 protein: A novel regulator of tumor angiogenesis and hyperpermeability. *Oncotarget.* 2016;7(28):44630-43.
338. Bahrami A, Amerizadeh F, ShahidSales S, Khazaei M, Ghayour-Mobarhan M, Sadeghnia HR, et al. Therapeutic Potential of Targeting Wnt/ β -Catenin Pathway in Treatment of Colorectal Cancer: Rational and Progress. *J Cell Biochem.* 2017;118(8):1979-83.
339. Schrock AB, Ouyang C, Sandhu J, Sokol E, Jin D, Ross JS, et al. Tumor mutational burden is predictive of response to immune checkpoint inhibitors in MSI-high metastatic colorectal cancer. *Annals of oncology : official journal of the European Society for Medical Oncology / ESMO.* 2019.
340. Overman MJ, McDermott R, Leach JL, Lonardi S, Lenz HJ, Morse MA, et al. Nivolumab in patients with metastatic DNA mismatch repair-deficient or microsatellite instability-high colorectal cancer (CheckMate 142): an open-label, multicentre, phase 2 study. *Lancet Oncol.* 2017;18(9):1182-91.

341. Brahmer JR, Drake CG, Wollner I, Powderly JD, Picus J, Sharfman WH, et al. Phase I study of single-agent anti-programmed death-1 (MDX-1106) in refractory solid tumors: safety, clinical activity, pharmacodynamics, and immunologic correlates. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. 2010;28(19):3167-75.
342. Becht E, de Reyniès A, Giraldo NA, Pilati C, Buttard B, Lacroix L, et al. Immune and Stromal Classification of Colorectal Cancer Is Associated with Molecular Subtypes and Relevant for Precision Immunotherapy. *Clinical cancer research : an official journal of the American Association for Cancer Research*. 2016;22(16):4057-66.
343. Yaeger R, Chatila WK, Lipsyc MD, Hechtman JF, Cercek A, Sanchez-Vega F, et al. Clinical Sequencing Defines the Genomic Landscape of Metastatic Colorectal Cancer. *Cancer Cell*. 2018;33(1):125-36.e3.
344. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov*. 2012;2(5):401-4.
345. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal*. 2013;6(269):pl1.
346. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res*. 2002;30(1):207-10.
347. Cardoso F, van't Veer LJ, Bogaerts J, Slaets L, Viale G, Delaloge S, et al. 70-Gene Signature as an Aid to Treatment Decisions in Early-Stage Breast Cancer. *N Engl J Med*. 2016;375(8):717-29.
348. van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AA, Voskuil DW, et al. A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med*. 2002;347(25):1999-2009.
349. Chen L, Lu D, Sun K, Xu Y, Hu P, Li X, et al. Identification of biomarkers associated with diagnosis and prognosis of colorectal cancer patients based on integrated bioinformatics analysis. *Gene*. 2019;692:119-25.

350. Martinez-Romero J, Bueno-Fortes S, Martín-Merino M, Ramirez de Molina A, De Las Rivas J. Survival marker genes of colorectal cancer derived from consistent transcriptomic profiling. *BMC Genomics*. 2018;19(Suppl 8):857.
351. Long NP, Park S, Anh NH, Nghi TD, Yoon SJ, Park JH, et al. High-Throughput Omics and Statistical Learning Integration for the Discovery and Validation of Novel Diagnostic Signatures in Colorectal Cancer. *Int J Mol Sci*. 2019;20(2).
352. Sun L, Zhang X, Xu J, Wang W, Liu R. A Gene selection approach based on the fisher linear discriminant and the neighborhood rough set. *Bioengineered*. 2018;9(1):144-51.
353. Zhi J, Sun J, Wang Z, Ding W. Support vector machine classifier for prediction of the metastasis of colorectal cancer. *Int J Mol Med*. 2018;41(3):1419-26.
354. Xu G, Zhang M, Zhu H, Xu J. A 15-gene signature for prediction of colon cancer recurrence and prognosis based on SVM. *Gene*. 2017;604:33-40.
355. Yang WJ, Wang HB, Wang WD, Bai PY, Lu HX, Sun CH, et al. A network-based predictive gene expression signature for recurrence risks in stage II colorectal cancer. *Cancer medicine*. 2019.
356. Gan Z, Zou Q, Lin Y, Xu Z, Huang Z, Chen Z, et al. Identification of a 13-gene-based classifier as a potential biomarker to predict the effects of fluorouracil-based chemotherapy in colorectal cancer. *Oncol Lett*. 2019;17(6):5057-63.
357. Van Landeghem S, Abeel T, Saeys Y, Van de Peer Y. Discriminative and informative features for biomolecular text mining with ensemble feature selection. *Bioinformatics*. 2010;26(18):i554-60.
358. Ge Y, Xu A, Zhang M, Xiong H, Fang L, Zhang X, et al. FK506 Binding Protein 10 Is Overexpressed and Promotes Renal Cell Carcinoma. *Urol Int*. 2017;98(2):169-76.
359. Takahashi Y, Sawada G, Kurashige J, Uchi R, Matsumura T, Ueo H, et al. Paired related homoeobox 1, a new EMT inducer, is involved in metastasis and poor prognosis in colorectal cancer. *British journal of cancer*. 2013;109(2):307-11.
360. Han Z, Wang T, Han S, Chen Y, Chen T, Jia Q, et al. Low-expression of TMEM100 is associated with poor prognosis in non-small-cell lung cancer. *Am J Transl Res*. 2017;9(5):2567-78.

361. Ou D, Yang H, Hua D, Xiao S, Yang L. Novel roles of TMEM100: inhibition metastasis and proliferation of hepatocellular carcinoma. *Oncotarget*. 2015;6(19):17379-90.
362. Takeda K, Mizushima T, Yokoyama Y, Hirose H, Wu X, Qian Y, et al. Sox2 is associated with cancer stem-like properties in colorectal cancer. *Scientific reports*. 2018;8(1):17639.
363. Lundberg IV, Edin S, Eklöf V, Öberg Å, Palmqvist R, Wikberg ML. SOX2 expression is associated with a cancer stem cell state and down-regulation of CDX2 in colorectal cancer. *BMC Cancer*. 2016;16:471.
364. Han X, Fang X, Lou X, Hua D, Ding W, Foltz G, et al. Silencing SOX2 induced mesenchymal-epithelial transition and its expression predicts liver and lymph node metastasis of CRC patients. *PLoS One*. 2012;7(8):e41335.
365. Kuijjer ML, Paulson JN, Salzman P, Ding W, Quackenbush J. Cancer subtype identification using somatic mutation data. *British journal of cancer*. 2018;118(11):1492-501.
366. Bushati M, Rovers KP, Sommariva A, Sugarbaker PH, Morris DL, Yonemura Y, et al. The current practice of cytoreductive surgery and HIPEC for colorectal peritoneal metastases: Results of a worldwide web-based survey of the Peritoneal Surface Oncology Group International (PSOGI). *Eur J Surg Oncol*. 2018;44(12):1942-8.
367. Gelli M, Huguenin JFL, de Baere T, Benhaim L, Mariani A, Boige V, et al. Peritoneal and extraperitoneal relapse after previous curative treatment of peritoneal metastases from colorectal cancer: What survival can we expect? *European journal of cancer (Oxford, England : 1990)*. 2018;100:94-103.
368. Patel CM, Sahdev A, Reznick RH. CT, MRI and PET imaging in peritoneal malignancy. *Cancer Imaging*. 2011;11:123-39.
369. Reinert T, Schøler LV, Thomsen R, Tobiasen H, Vang S, Nordentoft I, et al. Analysis of circulating tumour DNA to monitor disease burden following colorectal cancer surgery. *Gut*. 2016;65(4):625-34.
370. Yuza K, Nagahashi M, Watanabe S, Takabe K, Wakai T. Hypermethylation and microsatellite instability in gastrointestinal cancers. *Oncotarget*. 2017;8(67):112103-15.

371. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, et al. The Ensembl Variant Effect Predictor. *Genome Biol.* 2016;17(1):122.
372. Diehl F, Schmidt K, Choti MA, Romans K, Goodman S, Li M, et al. Circulating mutant DNA to assess tumor dynamics. *Nat Med.* 2008;14(9):985-90.
373. Park SW, Hur SY, Yoo NJ, Lee SH. Somatic frameshift mutations of bone morphogenic protein receptor 2 gene in gastric and colorectal cancers with microsatellite instability. *APMIS.* 2010;118(11):824-9.
374. De Mattos-Arruda L, Mayor R, Ng CK, Weigelt B, Martínez-Ricarte F, Torrejon D, et al. Cerebrospinal fluid-derived circulating tumour DNA better represents the genomic alterations of brain tumours than plasma. *Nat Commun.* 2015;6:8839.
375. Oxnard GR, Thress KS, Alden RS, Lawrance R, Paweletz CP, Cantarini M, et al. Association Between Plasma Genotyping and Outcomes of Treatment With Osimertinib (AZD9291) in Advanced Non-Small-Cell Lung Cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology.* 2016;34(28):3375-82.
376. Siravegna G, Mussolin B, Buscarino M, Corti G, Cassingena A, Crisafulli G, et al. Clonal evolution and resistance to EGFR blockade in the blood of colorectal cancer patients. *Nat Med.* 2015;21(7):827.
377. Morelli MP, Overman MJ, Dasari A, Kazmi SM, Mazard T, Vilar E, et al. Characterizing the patterns of clonal selection in circulating tumor DNA from patients with colorectal cancer refractory to anti-EGFR treatment. *Annals of oncology : official journal of the European Society for Medical Oncology / ESMO.* 2015;26(4):731-6.
378. Zhou J, Chang L, Guan Y, Yang L, Xia X, Cui L, et al. Application of Circulating Tumor DNA as a Non-Invasive Tool for Monitoring the Progression of Colorectal Cancer. *PLoS One.* 2016;11(7):e0159708.
379. Thierry AR, El Messaoudi S, Mollevi C, Raoul JL, Guimbaud R, Pezet D, et al. Clinical utility of circulating DNA analysis for rapid detection of actionable mutations to select metastatic colorectal patients for anti-EGFR treatment. *Annals of oncology : official journal of the European Society for Medical Oncology / ESMO.* 2017;28(9):2149-59.
380. Kidess E, Heirich K, Wiggin M, Vysotskaia V, Visser BC, Marziali A, et al. Mutation profiling of tumor DNA from plasma and tumor tissue of colorectal cancer

patients with a novel, high-sensitivity multiplexed mutation detection platform. *Oncotarget*. 2015;6(4):2549-61.

381. Ng SB, Chua C, Ng M, Gan A, Poon PS, Teo M, et al. Individualised multiplexed circulating tumour DNA assays for monitoring of tumour presence in patients after colorectal cancer surgery. *Scientific reports*. 2017;7:40737.

382. Kato K, Uchida J, Kukita Y, Kumagai T, Nishino K, Inoue T, et al. Transient appearance of circulating tumor DNA associated with de novo treatment. *Scientific reports*. 2016;6:38639.

383. Poon IK, Lucas CD, Rossi AG, Ravichandran KS. Apoptotic cell clearance: basic biology and therapeutic potential. *Nat Rev Immunol*. 2014;14(3):166-80.

384. Tie J, Kinde I, Wang Y, Wong HL, Roebert J, Christie M, et al. Circulating tumor DNA as an early marker of therapeutic response in patients with metastatic colorectal cancer. *Annals of oncology : official journal of the European Society for Medical Oncology / ESMO*. 2015;26(8):1715-22.

385. Zhuo C, Hu D, Li J, Yu H, Lin X, Chen Y, et al. Downregulation of Activin A Receptor Type 2A Is Associated with Metastatic Potential and Poor Prognosis of Colon Cancer. *J Cancer*. 2018;9(19):3626-33.

386. Liu J, Chen X, Wang J, Zhou S, Wang CL, Ye MZ, et al. Biological background of the genomic variations of cf-DNA in healthy individuals. *Annals of oncology : official journal of the European Society for Medical Oncology / ESMO*. 2018.

387. Simon R. Development and evaluation of therapeutically relevant predictive classifiers using gene expression profiling. *Journal of the National Cancer Institute*. 2006;98(17):1169-71.

388. Kunická T, Souček P. Importance of ABCC1 for cancer therapy and prognosis. *Drug Metab Rev*. 2014;46(3):325-42.

389. Liu J, Pan S, Hsieh MH, Ng N, Sun F, Wang T, et al. Targeting Wnt-driven cancer through the inhibition of Porcupine by LGK974. *Proc Natl Acad Sci U S A*. 2013;110(50):20224-9.

390. Horvath P, Aulner N, Bickle M, Davies AM, Nery ED, Ebner D, et al. Screening out irrelevant cell-based models of disease. *Nat Rev Drug Discov*. 2016;15(11):751-69.

391. Shamir ER, Ewald AJ. Three-dimensional organotypic culture: experimental models of mammalian biology and disease. *Nat Rev Mol Cell Biol.* 2014;15(10):647-64.
392. Kim M, Mun H, Sung CO, Cho EJ, Jeon HJ, Chun SM, et al. Patient-derived lung cancer organoids as in vitro cancer models for therapeutic screening. *Nat Commun.* 2019;10(1):3991.
393. van de Wetering M, Francies HE, Francis JM, Bounova G, Iorio F, Pronk A, et al. Prospective derivation of a living organoid biobank of colorectal cancer patients. *Cell.* 2015;161(4):933-45.
394. Novartis. A study of LGK974 in Patients with malignancies dependant on Wnt Ligands Clinical trials.gov [Available from: <https://clinicaltrials.gov/ct2/show/NCT01351103>.
395. BioPharma. A. Study of WNT974 in Combination With LGX818 and Cetuximab in Patients With BRAF-mutant Metastatic Colorectal Cancer (mCRC) and Wnt Pathway Mutations Clinical trials.gov [Available from: <https://clinicaltrials.gov/ct2/show/NCT02278133>.
396. University. I. Trial of Decitabine as a sensitizer to carboplatin in platinum resistant recurrent ovarian cancer [Available from: <https://clinicaltrials.gov/ct2/show/NCT00477386?term=decitabine&rslt=With&phase=0&draw=4&rank=5>.
397. Tabernero JM, I. Wileke, R. Argiles, G. Marabelle, A. Rodriguez-Ruiz, ME. Albanell, J. Calvo, E. Moreno, V. Cleary, JM. Eder JP. Karanikas, V. Bouseida, S. Sandoval, F. Sabanes, D. Sreckovic, S. Hurwitz, H. Paz-Ares, LG. Saro-Suarez, JM. <p class="MsoNormal" style="margin-bottom:0cm, margin-bottom:.0001pt, line-height:1, normal, mso-layout-grid-align:none, text-autospace:none">, et al. Phase Ia and Ib studies of the novel carcinoembryonic antigen (CEA) T-cell bispecific (CEA CD3 TCB) antibody as a single agent and in combination with atezolizumab: Preliminary efficacy and safety in patients with metastatic colorectal cancer (mCRC). *Journal of clinical oncology.* 2017;35:3002-398. Schwartz L, Litière S, de Vries E, Ford R, Gwyther S, Mandrekar S, et al. RECIST 1.1 – Update and Clarification: From the RECIST Committee. *European journal of cancer.* 2016;62:132-7.

Appendix 1

Table S 1 Methylation bioinformatics R code

```
library("ChAMP")
read.idat():
bgxfile = dir(path = "/dummy_data/Image Data/Manifest/" ,pattern = ".bgx")
myLoad <- champ.load(arraytype="EPIC")
champ.QC()
myNorm <- champ.norm()
QC.GUI(myNorm)
write.csv(myNorm, file="Mynorm.csv")
champ.SVD()
myDMP <- champ.DMP(adjPVal = 0.1)
write.csv(myDMP,file="results.csv")
myDMR <- champ.DMR()
write.csv(myDMR,file="dmr.csv")
myCNA <- champ.CNA()
myCNA <- champ.CNA(intensity = myLoad$intensity, controlGroup = "T",
sampleCNA = FALSE, groupFreqPlots = TRUE, Rplot = FALSE, arraytype =
"EPIC")
write.table(myCNA, file="CNA.csv")
```

Table S 2 Biosigner classifier R code

```
Library(biosigner)

a<-read.table("sallycombinedrna_genecounts.txt",header=T,row.names=1,as.is=T)
pheno <- read.csv(file="latestpheno.csv",row.names=1,as.is=TRUE)

crsmatrix <- as.matrix(a)
```

```

pheno$Prognosis <- as.factor(pheno$Prognosis)

crstranspose <- t(crsmatrix)

training <- 1:floor(0.60 * nrow(crstranspose))

testing <- setdiff(1:nrow(crstranspose), training)

sigTrain <- biosign(crstranspose[training, ], pheno[training,"Prognosis"],
predicted=predFc,bootl=50)

dev.copy2pdf(file="rnasig-latest.pdf")

dev.off()

plot(sigTrain, typeC = "boxplot")

dev.copy2pdf(file="rnasig-boxplots.pdf")

dev.off()

plot(sigTrain, typeC = "boxplot")

diaFitDF <- predict(sigTrain)

lapply(diaFitDF, function(predFc) table(actual = pheno[training,"Prognosis"],
predicted = predFc))

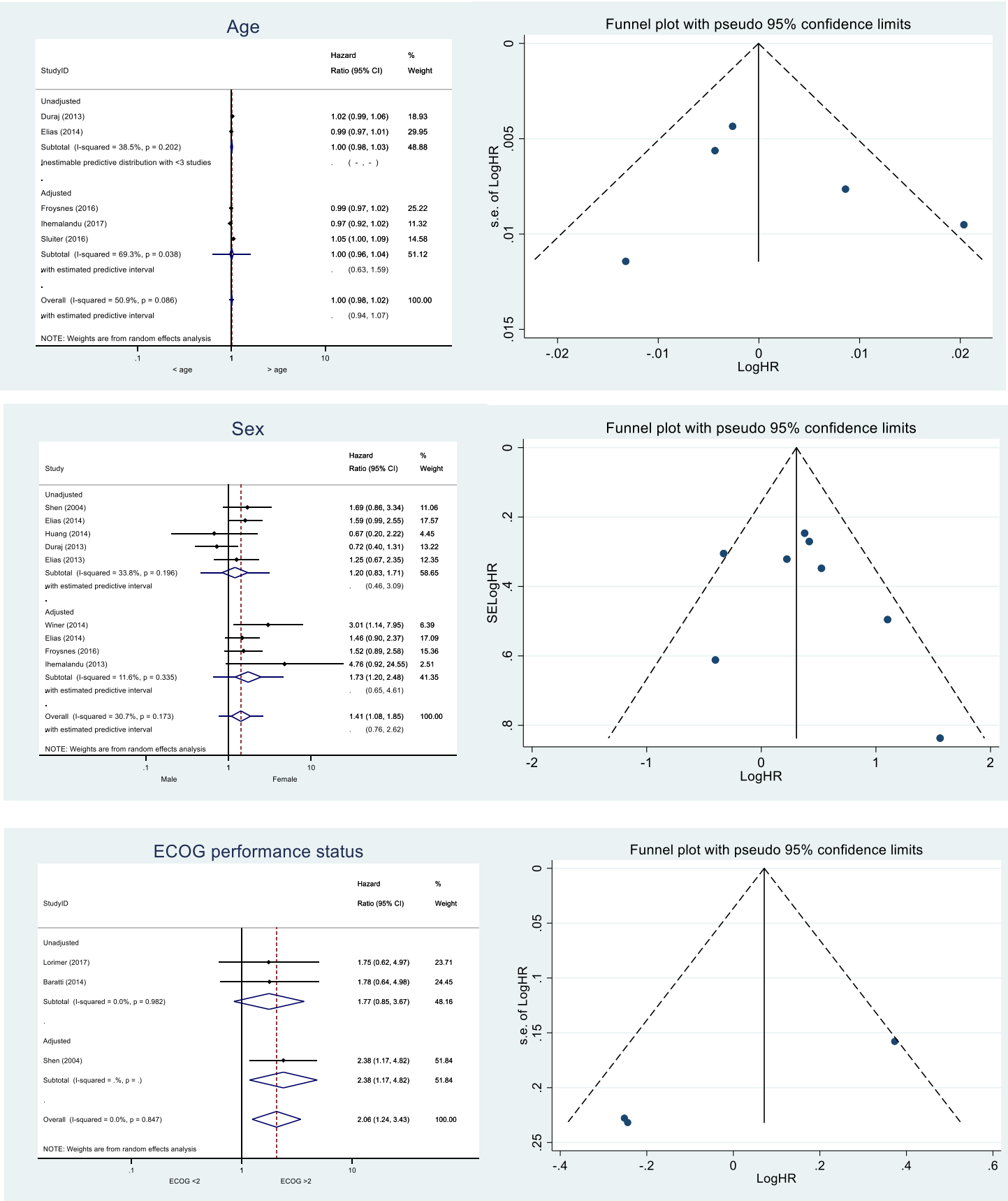
sapply(diaFitDF, function(predFc) {
conf <- table(pheno[training, "Prognosis"], predFc)
conf <- sweep(conf, 1, rowSums(conf), "/")
round(mean(diag(conf)), 3)
})

diaTestDF <- predict(sigTrain, newdata = crstranspose[testing, ])

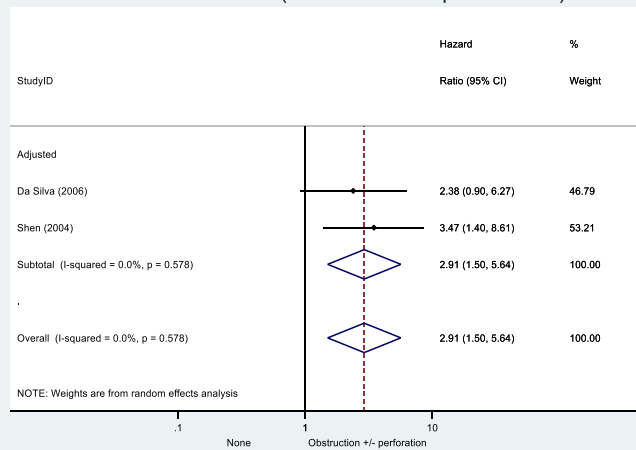
sapply(diaTestDF, function(predFc) {
conf <- table(pheno[testing, "Prognosis"], predFc)
conf <- sweep(conf, 1, rowSums(conf), "/")
round(mean(diag(conf)), 3)
})

```

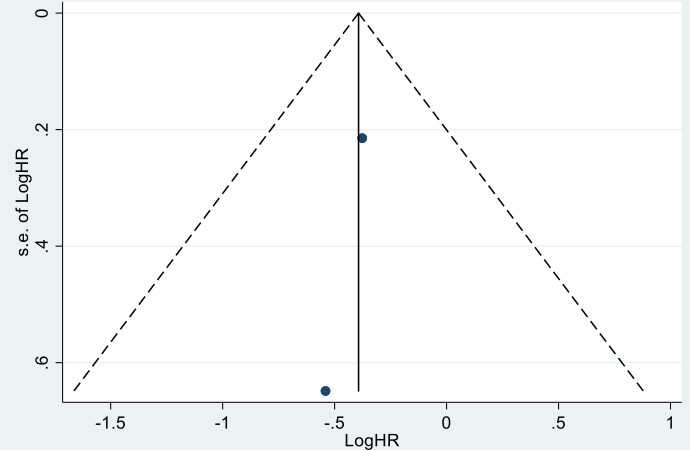

Figure S 1 Forest and funnel plots for each prognostic factor



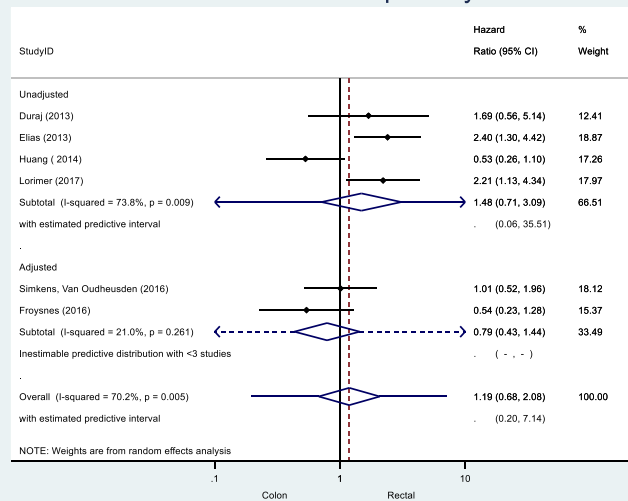
Adverse features (obstruction or perforation)



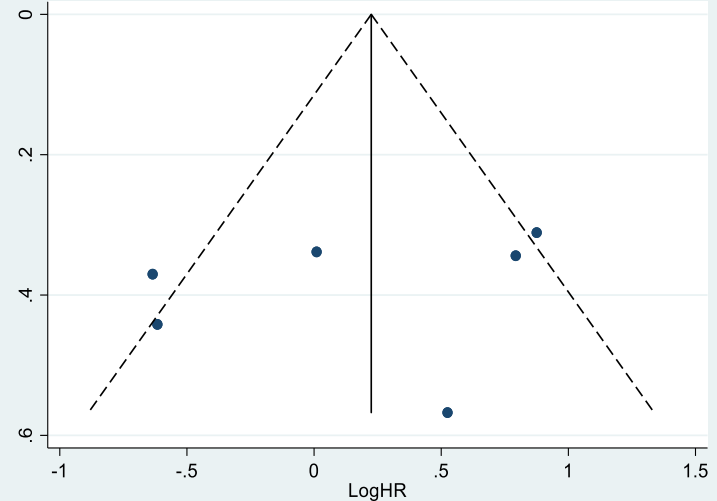
Funnel plot with pseudo 95% confidence limits



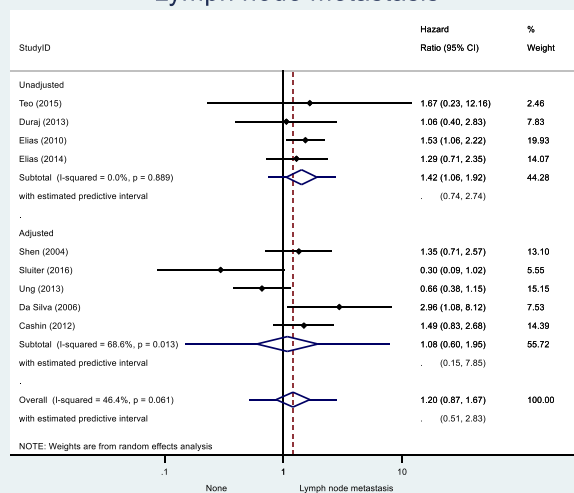
Rectal or colonic primary



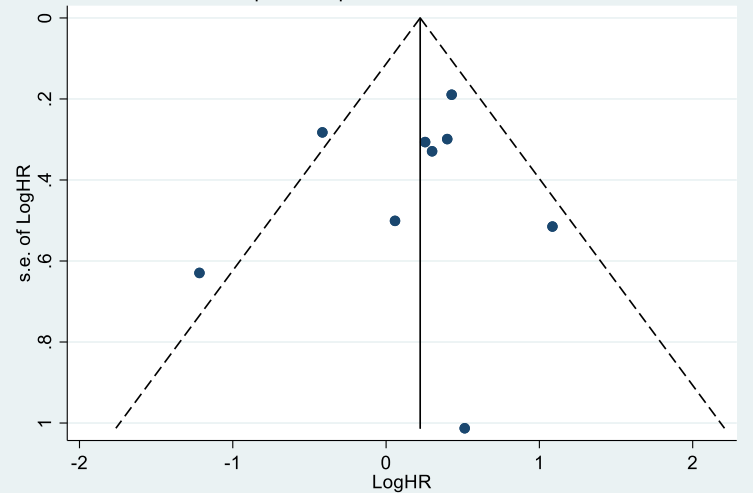
Funnel plot with pseudo 95% confidence limits



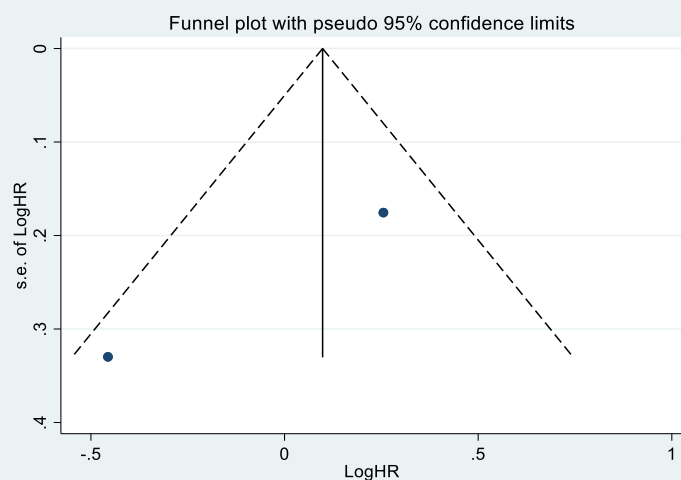
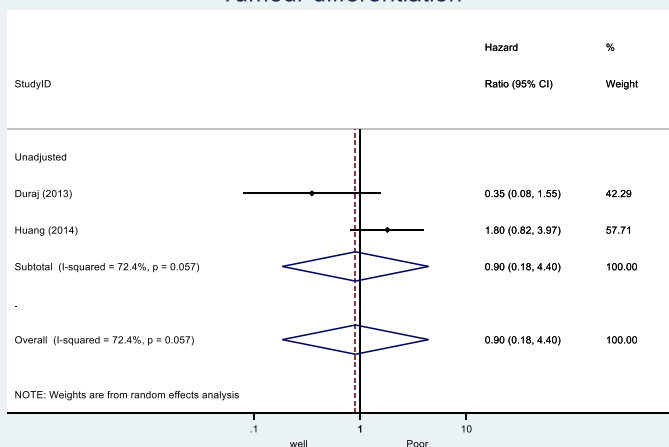
Lymph node metastasis



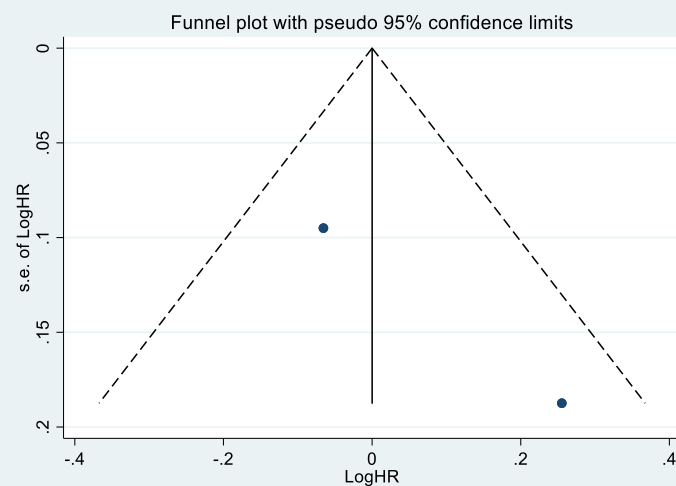
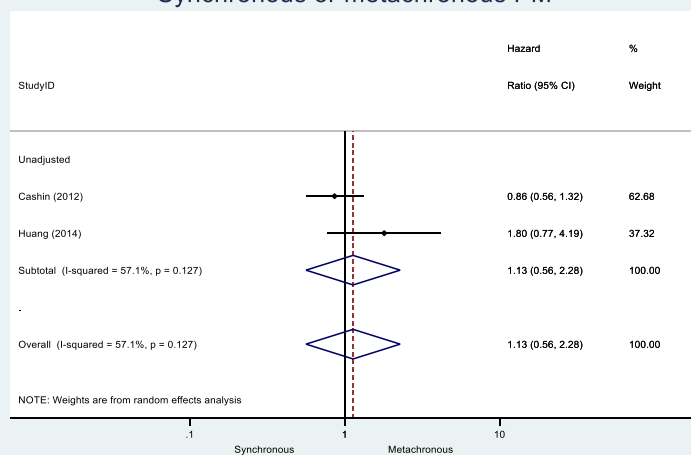
Funnel plot with pseudo 95% confidence limits



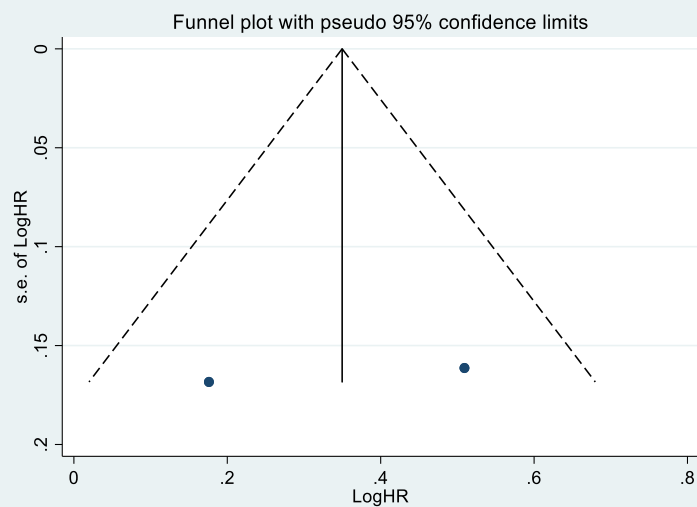
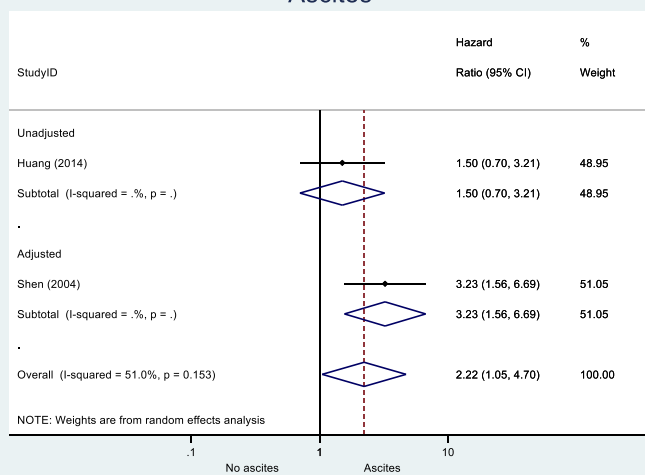
Tumour differentiation



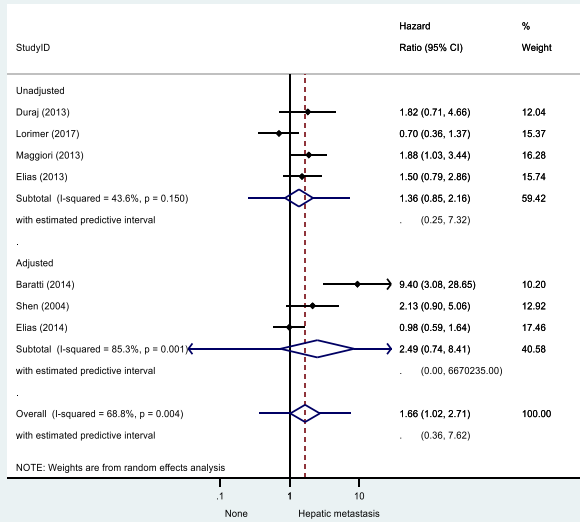
Synchronous or metachronous PM



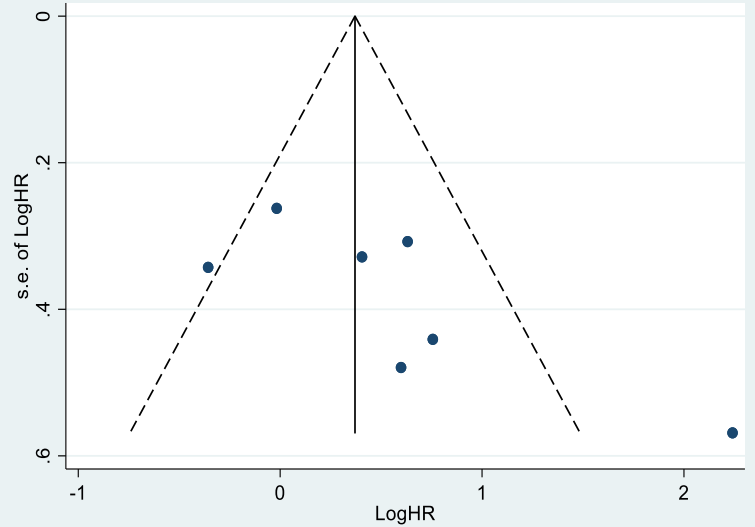
Ascites



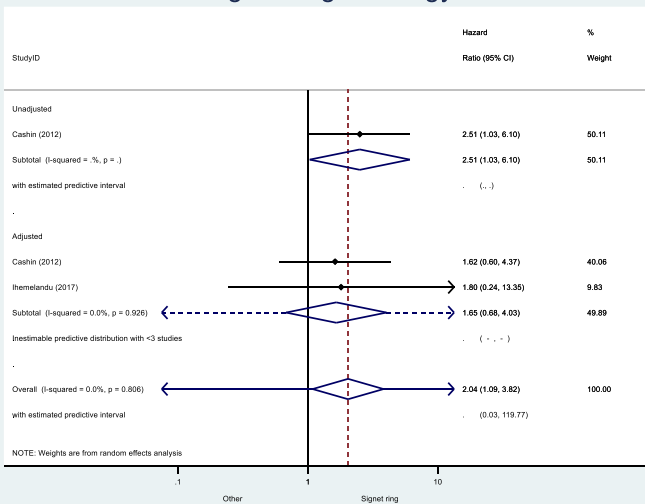
Hepatic metastasis



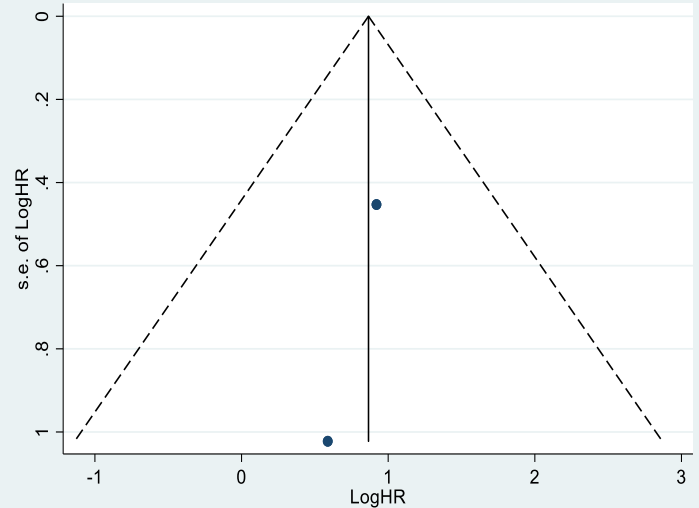
Funnel plot with pseudo 95% confidence limits



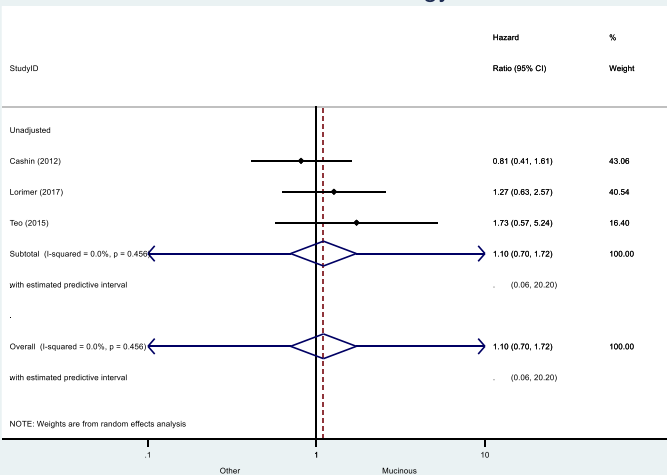
Signet ring histology



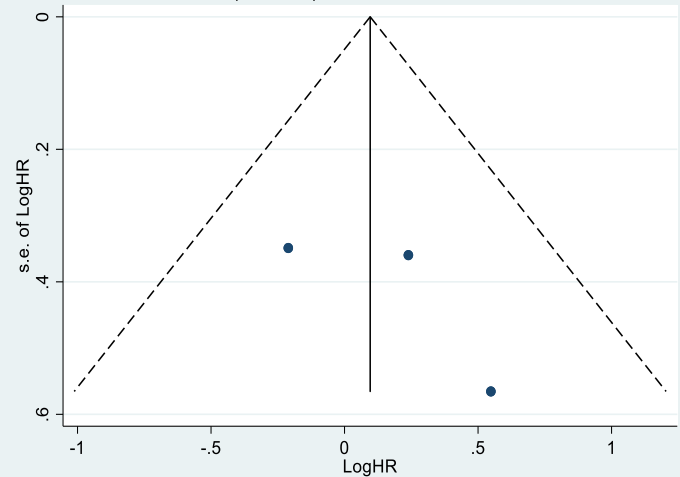
Funnel plot with pseudo 95% confidence limits



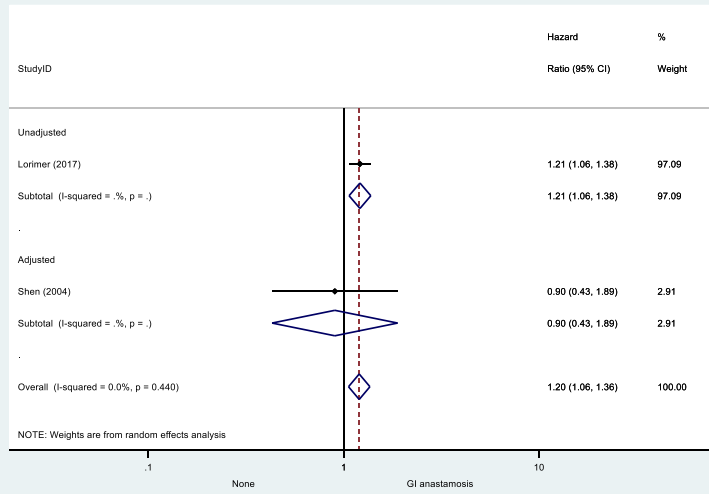
Mucinous histology



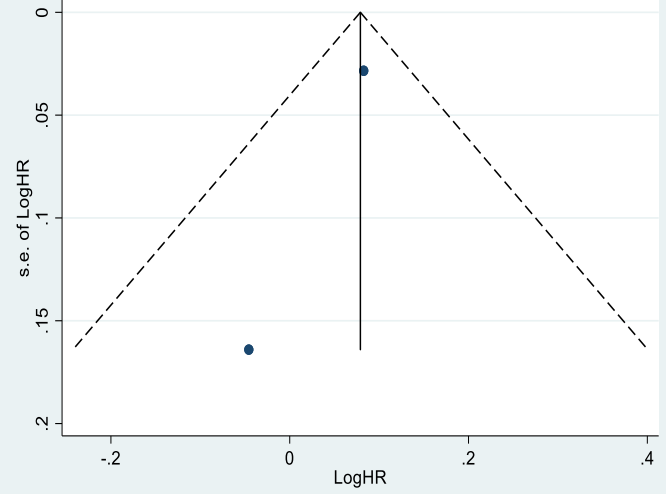
Funnel plot with pseudo 95% confidence limits



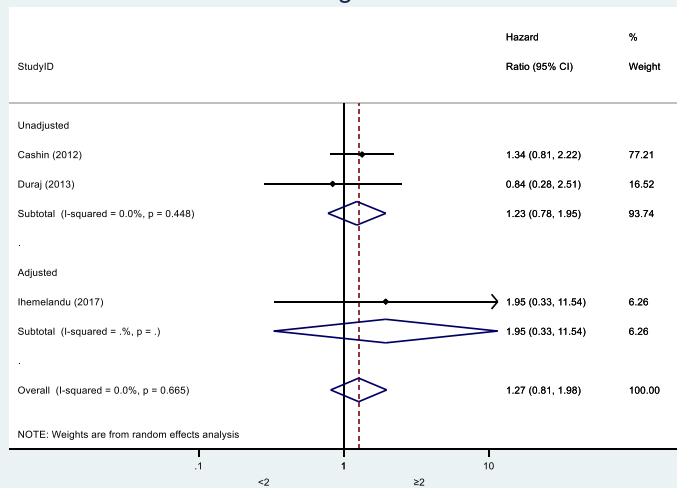
Presence of GI anastomosis



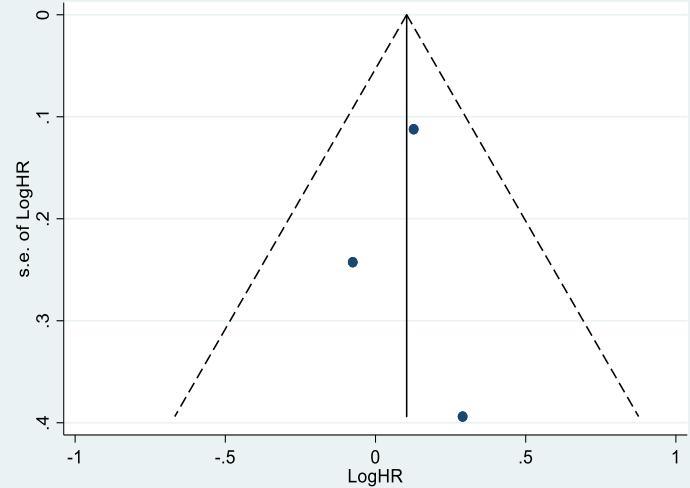
Funnel plot with pseudo 95% confidence limits



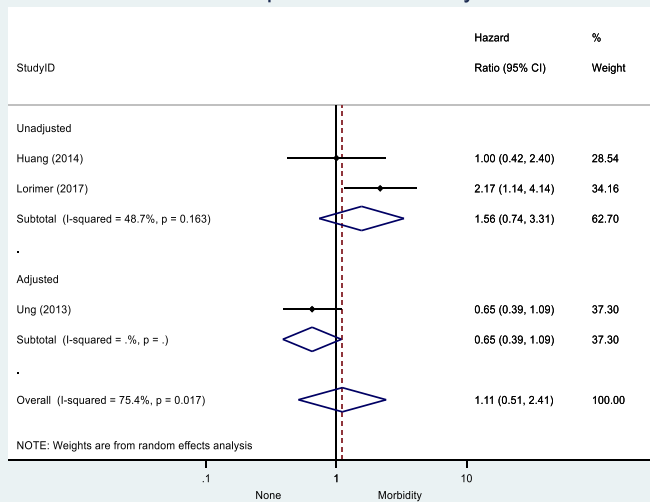
Prior surgical score



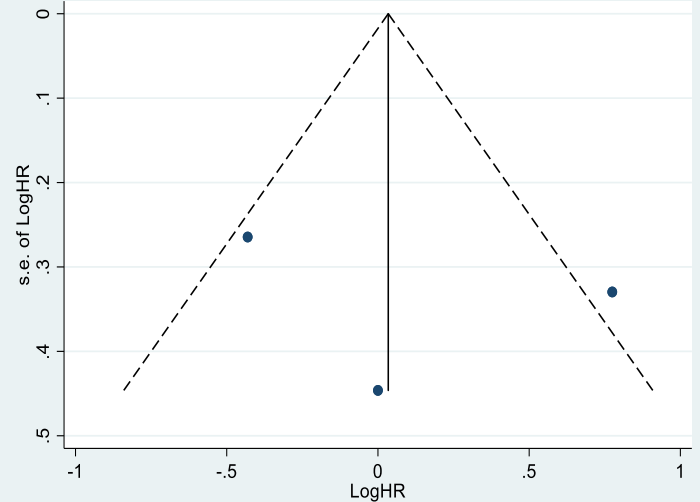
Funnel plot with pseudo 95% confidence limits



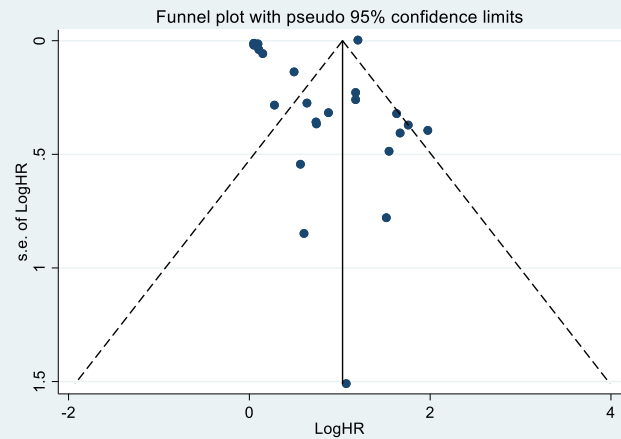
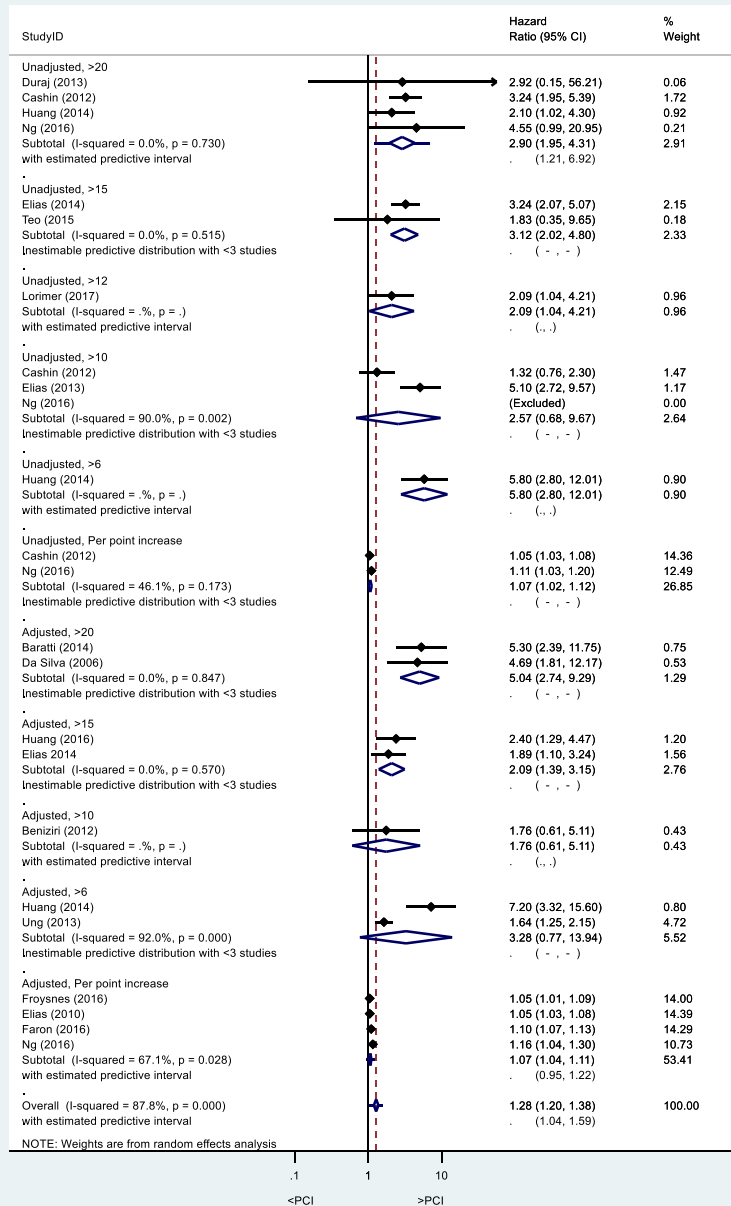
Post operative morbidity



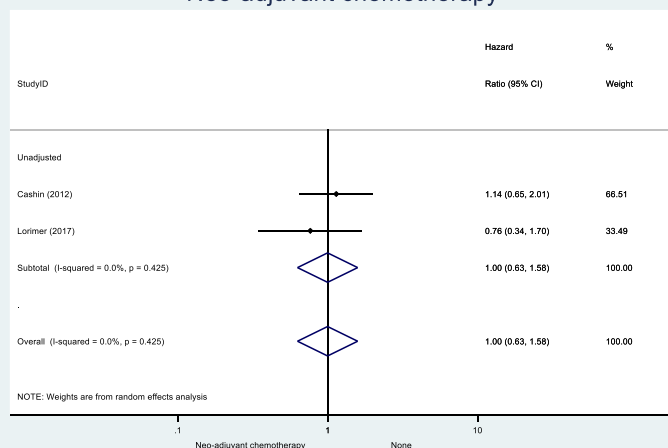
Funnel plot with pseudo 95% confidence limits



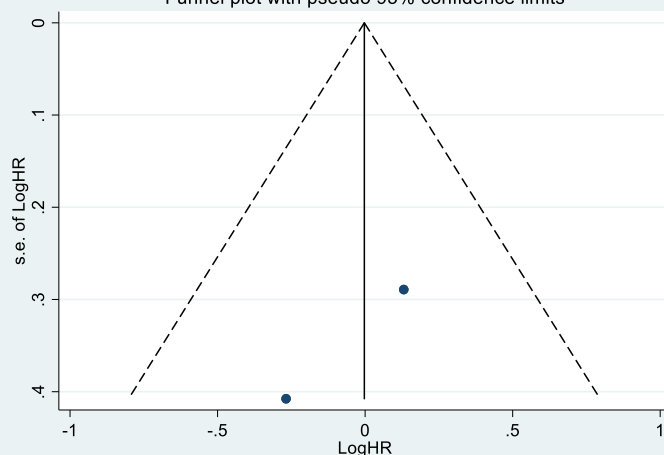
PCI score



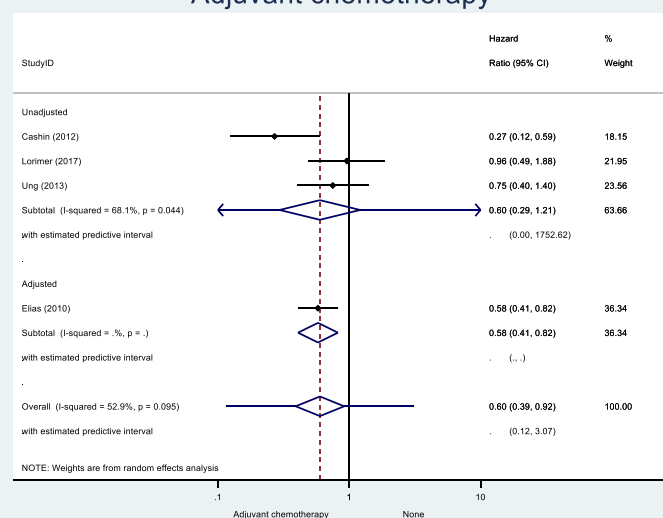
Neo-adjuvant chemotherapy



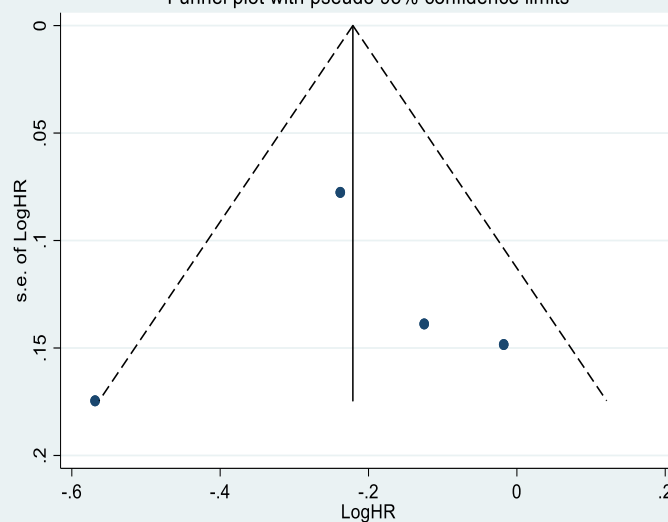
Funnel plot with pseudo 95% confidence limits



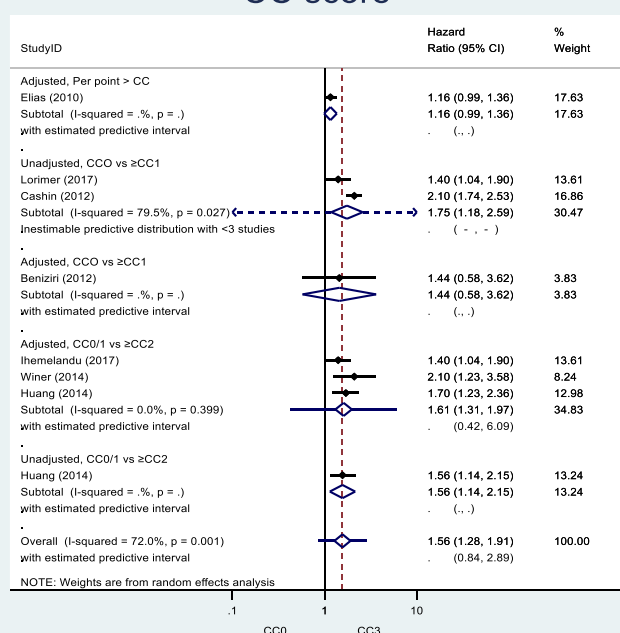
Adjuvant chemotherapy



Funnel plot with pseudo 95% confidence limits



CC score



Funnel plot with pseudo 95% confidence limits

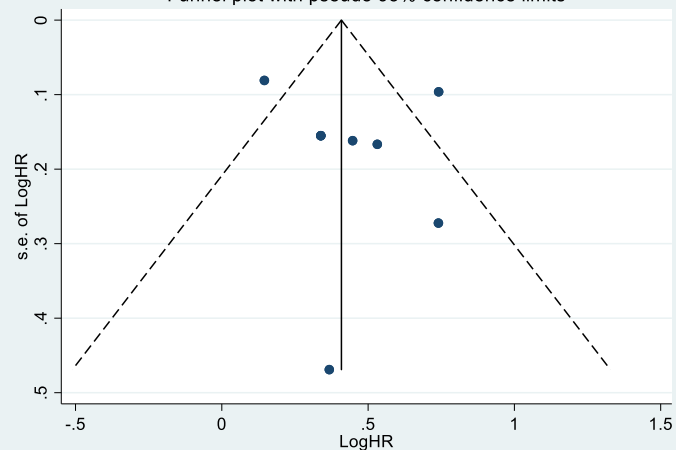
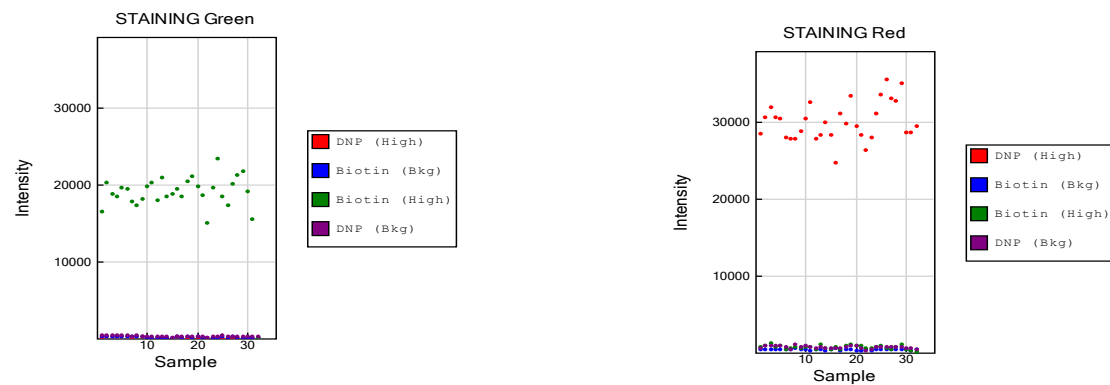
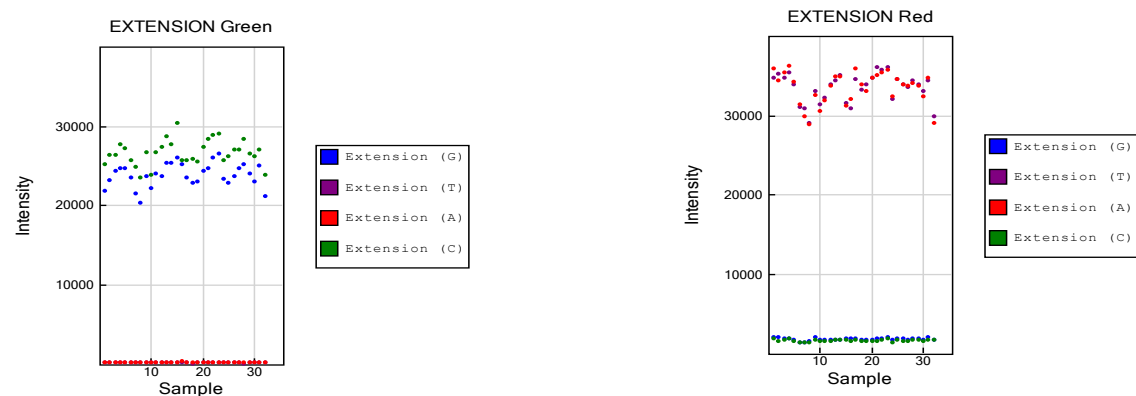


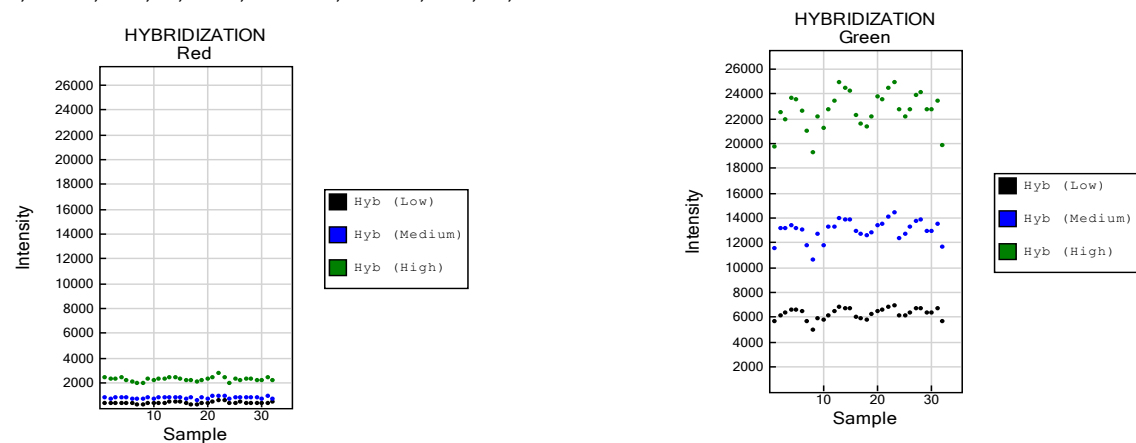
Figure S 2 Methylation array, quality control checks



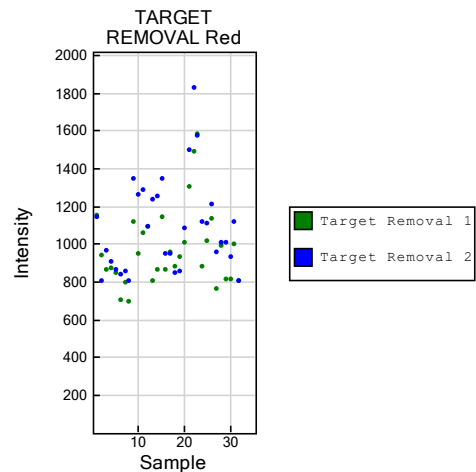
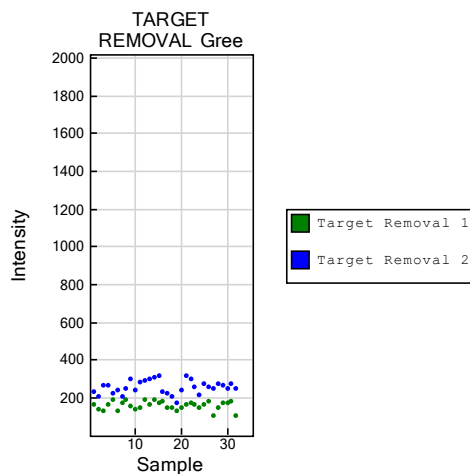
Staining control - examines the efficiency of the staining in both red and green channels. Ideal Green: Biotin High~ 20,000, Red: DNP~High 40,000



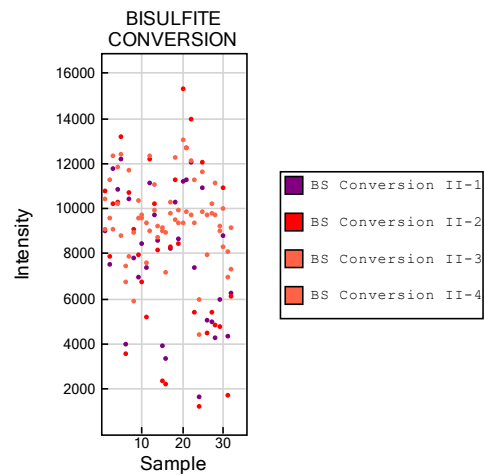
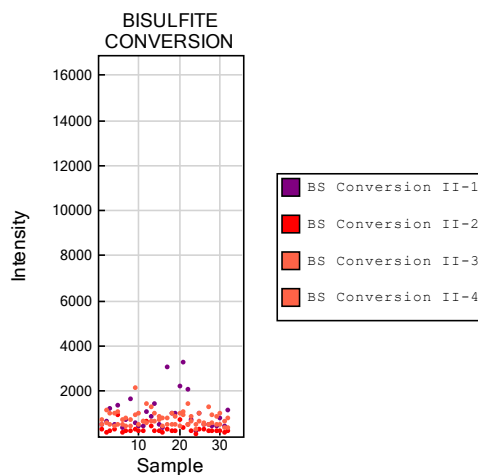
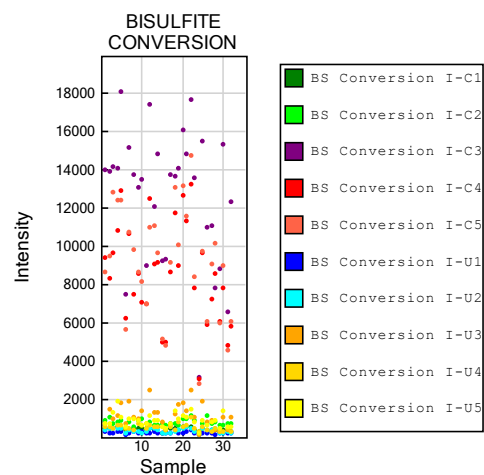
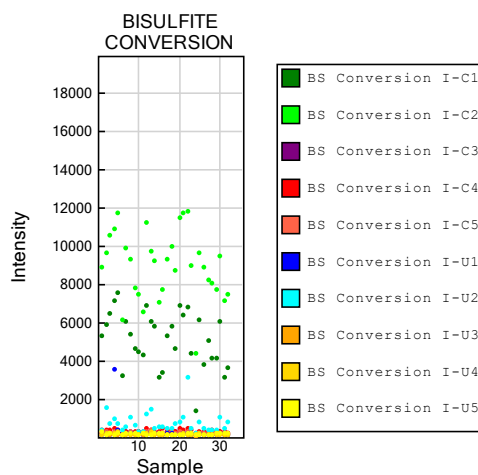
Extension control - examines the efficiency of A, T, C and G nucleotide extension Ideal Green: C,G~30,000; A,T ~0, Red: A,T ~ 40,000; G,C ~ 0



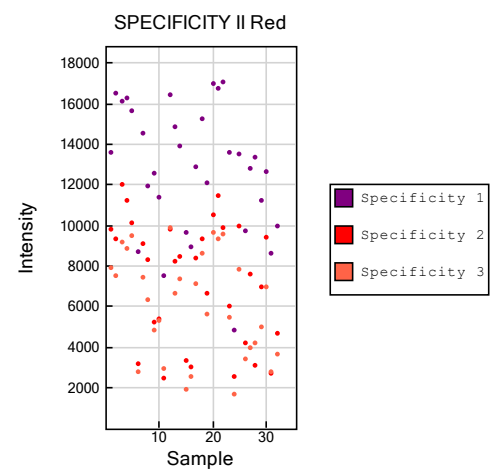
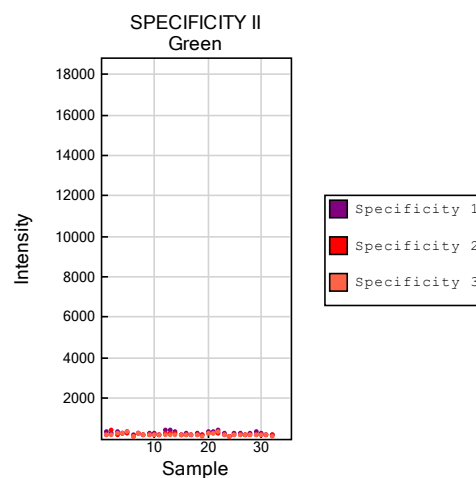
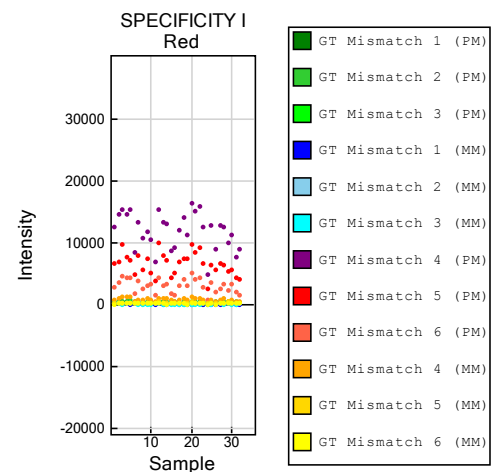
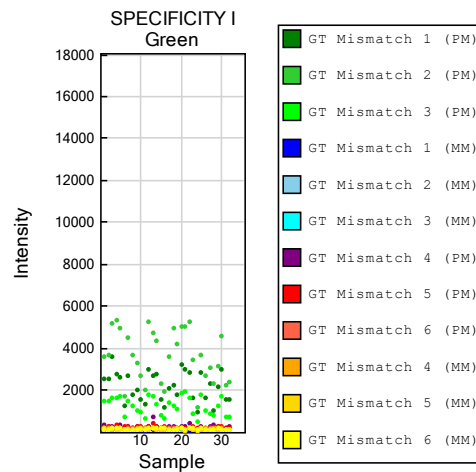
Hybridisation control – synthetic targets at thee concentrntations (high, medium, low) tests the overall performance of the assay. Only monitor green, Ideal Hyb (low) ~ 10,000, Hyb (medium) ~ 15,000, Hyb (high) ~ 30,000.



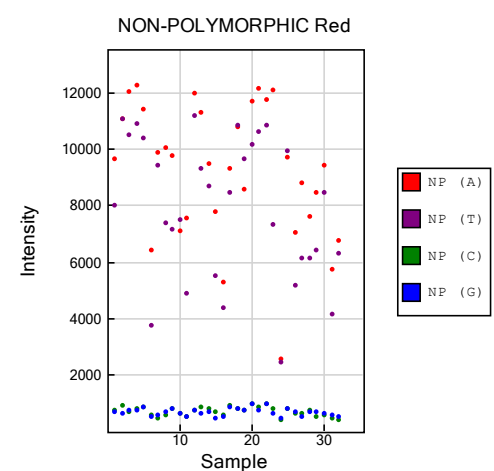
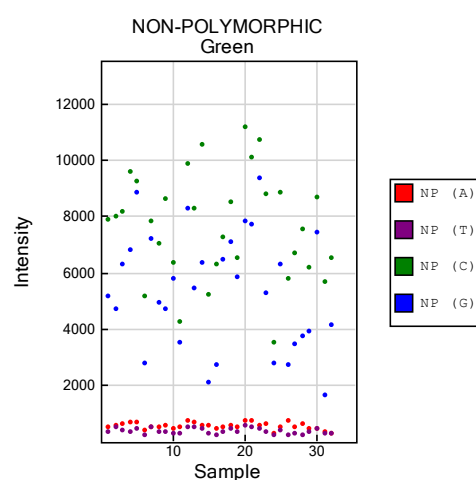
Target removal control – highlight efficiency of the removal of probe sequences post extension. Target removals should exhibit a low signal compared to hybridisation controls. Only monitor green, Ideal low background ~ 10,000



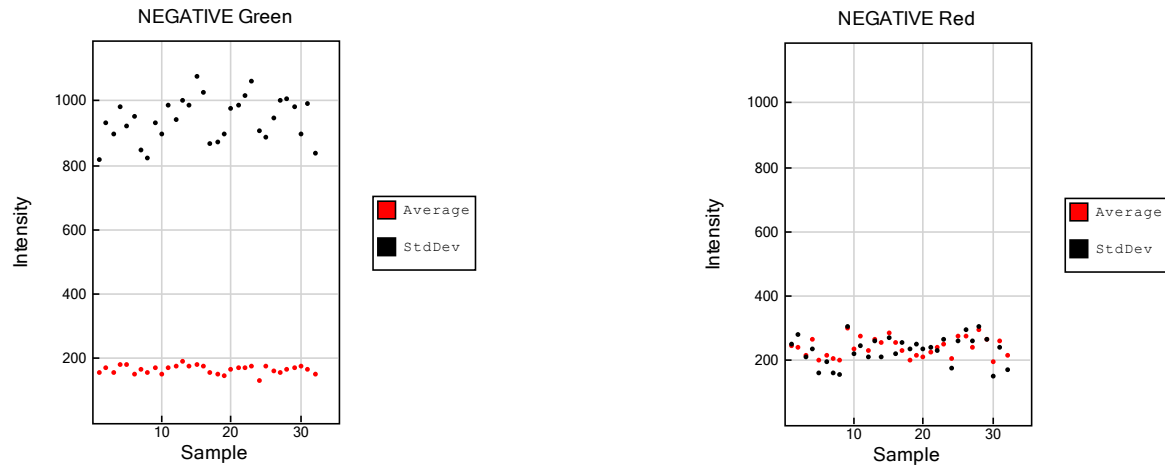
Bisulphite conversion control – if successful the converted C probes match the converted sequence and are extended, unconverted DNA - U probes are extended. Ideal Green: 1,2,3,4 ~ 0, Green: C1, C2, C3 ~ 15,000; U1, U2, U3 ~ 0, Red: C4, C5, C6 ~ 20,000; U4, U5, U6 ~ 0, Red: 1,2,3,4 ~ 20,000



Specificity control – check for non-specific signal against unmethylated background, Ideal Green 1,2,3 perfect match ~ 10,000, mismatch ~ 0, Red 1,2,3 ~ 20,000



Non-polymorphic control – query a base in a non-polymorphic region of genome, allows comparison of assay performance between samples. Ideal, Green A,T ~ 10,000, Red C,G ~ 20,000



Negative controls – bisulfite converted sequences not containing CpG dinucleotides are targeted in 600 negative controls on the array, defines the mean signal defines the system background <1000

Table S 3 Significantly (FDR<0.1) downregulated genes in CPM samples compared with primary CRC samples

Rank	Gene name	Fold change	FDR p value
1	FABP6	-34.30	1.74 x 10 ⁻⁰⁶
2	DEFA6	-8.15	8.55 x 10 ⁻⁰⁶
3	DMBT1	-6.06	2.43 x 10 ⁻⁰⁴
4	TTC38	-4.56	5.80 x 10 ⁻⁰⁵
5	OLFM4	-3.77	1.01 x 10 ⁻⁰⁴
6	IGHA1	-3.66	4.23 x 10 ⁻⁰⁵
7	CES2	-3.20	6.84 x 10 ⁻⁰⁵
8	NDUFS6	-2.70	7.74 x 10 ⁻⁰⁵
9	P2RY11	-2.53	6.37 x 10 ⁻⁰⁴
10	MUC2	-2.34	7.22 x 10 ⁻⁰⁴
11	DES	-2.32	4.68 x 10 ⁻⁰⁴
12	ECHS1	-2.25	9.80 x 10 ⁻⁰⁴
13	PIGR	-2.16	1.04 x 10 ⁻⁰³
14	RXRA	-2.06	5.13 x 10 ⁻⁰⁴
15	MTATP6P1	-1.78	1.10 x 10 ⁻⁰³
16	PDCD6	-1.68	1.11 x 10 ⁻⁰³

Table S 4 Significantly (FDR<0.1) upregulated genes in CPM samples compared with primary CRC samples

Rank	Gene ID	Fold change	FDR p value
1	CD53	7.29	5.87 x 10 ⁻⁰⁵
2	CYR61	4.24	3.12 x 10 ⁻⁰⁴
3	CXCL12	3.64	9.25 x 10 ⁻⁰⁴
4	NR2F1	3.53	7.09 x 10 ⁻⁰⁴
5	CTGF	3.49	1.55 x 10 ⁻⁰⁴
6	CSTB	3.41	6.13 x 10 ⁻⁰⁴
7	TSC22D3	3.36	3.94 x 10 ⁻⁰⁴
8	DCN	3.30	6.19 x 10 ⁻⁰⁵
9	PTEN	3.25	9.28 x 10 ⁻⁰⁴
10	NFKBIA	3.24	1.06 x 10 ⁻⁰⁴
11	PFKFB3	3.13	8.49 x 10 ⁻⁰⁴
12	SPARC	2.77	3.91 x 10 ⁻⁰⁵
13	AC120194.1	2.48	4.87 x 10 ⁻⁰⁴
14	TXNIP	2.43	6.33 x 10 ⁻⁰⁴
15	VIM	2.35	7.83 x 10 ⁻⁰⁴
16	MCL1	2.18	4.33 x 10 ⁻⁰⁴
17	EEF1A1	2.13	3.28 x 10 ⁻⁰⁸
18	RPS21	2.12	2.11 x 10 ⁻⁰⁵
19	RPS29	2.09	4.23 x 10 ⁻⁰⁶
20	PRPF8	2.08	6.48 x 10 ⁻⁰⁴
21	ITM2B	2.08	1.58 x 10 ⁻⁰⁵
22	MYL12A	2.02	9.83 x 10 ⁻⁰⁴
23	TMSB4X	1.99	2.21 x 10 ⁻⁰⁵
24	TOMM7	1.93	2.87 x 10 ⁻⁰⁴
25	RPL34	1.92	3.29 x 10 ⁻⁰⁵
26	RPL7A	1.88	6.17 x 10 ⁻⁰⁵
27	RHOA	1.85	1.05 x 10 ⁻⁰⁴
28	TPT1	1.84	3.81 x 10 ⁻⁰⁶
29	TMSB10	1.80	8.82 x 10 ⁻⁰⁵
30	FTH1	1.79	2.40 x 10 ⁻⁰⁵
31	SERF2	1.79	2.87 x 10 ⁻⁰⁴
32	RPL30	1.77	1.06 x 10 ⁻⁰³
33	RPL15	1.72	1.14 x 10 ⁻⁰⁵
34	RPLP1	1.71	1.06 x 10 ⁻⁰⁵
35	PFN1	1.69	6.59 x 10 ⁻⁰⁴
36	RPL39	1.69	1.64 x 10 ⁻⁰⁴
37	RPL23	1.65	1.22 x 10 ⁻⁰⁴
38	RPS3A	1.64	1.80 x 10 ⁻⁰⁴
39	CD63	1.63	1.79 x 10 ⁻⁰⁴
40	RPS6	1.62	2.29 x 10 ⁻⁰⁴

41	RPL41	1.62	3.65 x 10 ⁻⁰⁴
42	RPL31	1.61	5.26 x 10 ⁻⁰⁴
43	RPL32	1.60	5.04 x 10 ⁻⁰⁵
44	RPL11	1.59	4.69 x 10 ⁻⁰⁴
45	B2M	1.58	4.39 x 10 ⁻⁰⁴
46	RPS24	1.55	4.14 x 10 ⁻⁰⁴
47	RPS16	1.54	3.44 x 10 ⁻⁰⁴
48	RPLP0	1.54	7.90 x 10 ⁻⁰⁴
49	RPS12	1.49	5.87 x 10 ⁻⁰⁵

Table S 5 Significantly (FDR<0.1) upregulated genes in poor prognosis CRS samples compared with good prognosis CRS samples

Rank	Gene ID	Fold change	FDR p value
1	VEZF1	-10.40	3.76x 10 ⁻⁰⁵
2	CEACAM1	-8.27	1.83x 10 ⁻⁰³
3	BCYRN1	-7.96	7.50x 10 ⁻⁰⁵
4	RP11-192H23.5	-6.22	2.38x 10 ⁻⁰⁴
5	RELB	-6.17	9.50x 10 ⁻⁰⁴
6	DCAF12	-6.17	1.36x 10 ⁻⁰⁵
7	POLR3GL	-6.05	1.48x 10 ⁻⁰³
8	FEM1C	-5.69	1.87x 10 ⁻⁰³
9	AXIN1	-5.42	1.67x 10 ⁻⁰⁴
10	KHDRBS3	-5.01	1.23x 10 ⁻⁰³
11	TMEM167A	-4.68	1.30x 10 ⁻⁰⁴
12	FHOD1	-4.60	4.54x 10 ⁻⁰⁴
13	ATG9A	-4.59	1.28x 10 ⁻⁰³
14	DENND4B	-4.43	2.50x 10 ⁻⁰⁴
15	PDCD5	-4.34	7.63x 10 ⁻⁰⁴
16	RAB22A	-4.21	1.20x 10 ⁻⁰³
17	SNX3	-4.16	3.06x 10 ⁻⁰³
18	SEC61B	-4.13	2.13x 10 ⁻⁰³
19	PAM	-4.10	9.87x 10 ⁻⁰⁵
20	NKIRAS2	-4.10	2.77x 10 ⁻⁰³
21	PPP1R11	-4.10	1.95x 10 ⁻⁰³
22	IGSF8	-4.07	8.12x 10 ⁻⁰⁴
23	FOXK1	-3.95	2.90x 10 ⁻⁰³
24	SEPHS1	-3.76	1.87x 10 ⁻⁰³
25	NUBP2	-3.74	4.24x 10 ⁻⁰⁴
26	ITFG3	-3.64	1.31x 10 ⁻⁰³

27	DCP1A	-3.59	2.34x 10-03
28	HSPBP1	-3.41	8.96x 10-04
29	VPS26B	-3.38	1.54x 10-03
30	C14orf119	-3.38	1.34x 10-04
31	IDH3A	-3.37	2.57x 10-04
32	KLHL12	-3.34	1.45x 10-03
33	HSD3B7	-3.25	4.56x 10-04
34	RNF216	-3.22	6.91x 10-04
35	PES1	-3.20	8.67x 10-04
36	TTC39B	-3.17	1.90x 10-03
37	FMNL2	-3.11	1.00x 10-03
38	TMEM214	-3.09	9.61x 10-05
39	CDH17	-3.08	1.72x 10-03
40	POLR2H	-3.06	1.27x 10-04
41	CHMP4A	-3.04	9.91x 10-04
42	GGA1	-3.01	1.00x 10-03
43	MKLN1	-3.00	1.30x 10-03
44	VEGFB	-2.99	1.56x 10-03
45	NUDT3	-2.95	1.90x 10-04
46	ATP6V0A1	-2.93	1.04x 10-03
47	ANKH	-2.93	1.95x 10-03
48	CTNND1	-2.86	2.94x 10-03
49	RP11-345J4.5	-2.84	7.67x 10-04
50	SERTAD1	-2.82	4.57x 10-04
51	HBEGF	-2.82	5.48x 10-04
52	COX20	-2.78	8.51x 10-04
53	VKORC1	-2.77	2.29x 10-03
54	APLP2	-2.70	1.42x 10-03
55	SLC6A8	-2.69	1.54x 10-03
56	MGMT	-2.67	1.88x 10-03
57	LTBP1	-2.66	1.27x 10-03
58	WASL	-2.66	7.97x 10-04
59	TRAF7	-2.63	3.68x 10-04
60	KLF10	-2.63	1.97x 10-03
61	SHMT2	-2.63	9.78x 10-04
62	EP300	-2.62	3.30x 10-05
63	NQO1	-2.61	2.89x 10-03
64	ACLY	-2.58	3.07x 10-03
65	MAP1A	-2.58	1.23x 10-03
66	IFNAR1	-2.57	1.73x 10-03
67	PSMA1	-2.56	2.37x 10-03
68	RAD1	-2.56	5.06x 10-04
69	CD37	-2.55	1.21x 10-03
70	MIEN1	-2.52	6.78x 10-04
71	YARS	-2.48	1.14x 10-03
72	TRAK1	-2.48	2.74x 10-03

73	LDLR	-2.43	1.42x 10-03
74	C7orf73	-2.43	2.58x 10-03
75	TARS	-2.42	6.92x 10-04
76	ARFIP1	-2.40	2.34x 10-03
77	HNRNPH2	-2.39	6.49x 10-04
78	CCDC90B	-2.36	1.30x 10-03
79	DERL1	-2.36	2.52x 10-04
80	NUDT5	-2.35	1.27x 10-03
81	KIAA0368	-2.32	2.52x 10-03
82	TMEM165	-2.32	1.86x 10-03
83	SLC2A8	-2.32	1.84x 10-03
84	PRDX4	-2.30	1.73x 10-04
85	ANKRD52	-2.25	1.01x 10-03
86	ATP5C1	-2.21	1.13x 10-03
87	RBBP6	-2.21	3.57x 10-04
88	EHD2	-2.15	3.96x 10-04
89	TRPC4AP	-2.14	1.37x 10-03
90	SOWAHC	-2.11	1.62x 10-03
91	NOP10	-2.10	1.24x 10-04
92	PXN	-2.05	2.99x 10-03
93	PPAN	-2.01	1.67x 10-03
94	ALDH7A1	-1.99	1.79x 10-03
95	RNF40	-1.98	1.57x 10-03
96	EMP3	-1.98	2.73x 10-03
97	NDST1	-1.95	1.49x 10-03
98	NME7	-1.95	2.67x 10-03
99	MARCH2	-1.94	1.80x 10-04
100	PTPN14	-1.94	1.24x 10-03
101	NSUN5	-1.93	2.69x 10-03
102	CTA-363E6.7	-1.91	2.24x 10-03
103	TRIAP1	-1.90	1.80x 10-03
104	RNF149	-1.89	2.91x 10-03
105	STOM	-1.88	1.30x 10-03
106	KDM2B	-1.87	9.12x 10-04
107	PRPS1	-1.86	1.11x 10-03
108	MPLKIP	-1.85	1.87x 10-04
109	NCKAP1	-1.85	2.05x 10-03
110	LSM14A	-1.85	1.21x 10-04
111	LEPREL4	-1.83	6.29x 10-04
112	DGKA	-1.82	1.60x 10-03
113	CERCAM	-1.82	1.12x 10-03
114	RNASE6	-1.80	9.57x 10-04
115	ZHX1	-1.77	6.17x 10-04
116	PLRG1	-1.75	1.78x 10-03
117	PRKAG1	-1.74	1.37x 10-04
118	CIZ1	-1.74	3.19x 10-04

119	BRF2	-1.72	1.84x 10-03
120	SIGMAR1	-1.72	4.24x 10-04
121	AGPAT1	-1.69	1.95x 10-03
122	JOSD1	-1.69	3.70x 10-04
123	IFITM3	-1.67	1.58x 10-03
124	ARL2	-1.66	2.59x 10-04
125	TMEM230	-1.64	1.72x 10-03
126	ADIRF	-1.63	1.58x 10-03
127	UBR4	-1.62	3.48x 10-04
128	KCTD10	-1.60	1.00x 10-03
129	WIPF1	-1.60	2.53x 10-03
130	GSE1	-1.59	2.69x 10-03
131	MRPL15	-1.58	2.60x 10-03
132	TFRC	-1.57	6.55x 10-04
133	GBAS	-1.56	2.00x 10-03
134	PARVA	-1.56	3.00x 10-03
135	MXD1	-1.52	2.68x 10-04
136	ZEB1	-1.50	2.94x 10-03
137	SAFB2	-1.50	3.10x 10-03
138	CYTH2	-1.46	3.14x 10-03
139	CAPN2	-1.46	2.56x 10-03
140	PLOD3	-1.46	1.68x 10-03
141	TMEM164	-1.46	3.14x 10-03
142	MRPL18	-1.46	1.21x 10-03
143	PIK3R2	-1.45	1.26x 10-03
144	RAPGEFL1	-1.40	2.33x 10-03
145	NUDT22	-1.39	1.35x 10-03
146	VPS37C	-1.34	2.40x 10-03
147	STX12	-1.32	3.15x 10-03
148	ATP6V1B2	-1.32	2.76x 10-03
149	FAF2	-1.27	2.61x 10-04

Table S 6 Methylation array, CpGs satisfying the detection p-value of 0.05 and 0.01

Sample ID	Sentrix Barcode	CpG detection p-value (0.01)	%	CpG detection p-value (0.05)	%
CRS1	202784980004	843445	97.29494	861852	99.41827
CRS2	202784980004	830464	95.79753	860247	99.23313
CRS3	202784980004	859235	99.11639	864703	99.74714

CRS4	202784980004	855345	98.66766	864308	99.70158
CRS5	202784980004	861459	99.37293	864954	99.7761
CRS6	202784980004	822886	94.92338	859315	99.12561
CRS7	202784980004	861596	99.38874	865016	99.78325
CRS8	202784980004	861981	99.43315	864697	99.74645
CRS9	202784980045	804554	92.8087	854392	98.55773
CRS10	202784980045	838711	96.74886	860838	99.3013
CRS12	202784980045	794049	91.59691	854553	98.5763
CRS13	202784980045	859991	99.20359	864680	99.74449
CRS14	202784980045	849612	98.00633	862239	99.46291
CRS15	202784980045	848152	97.83792	863304	99.58576
CRS16	202784980045	767508	88.53529	845664	97.55091
CRS17	202784980045	781328	90.12949	838259	96.69672
CRS18	202784980029	833549	96.1534	861277	99.35194
CRS20	202784980029	851528	98.22735	863489	99.6071
CRS21	202784980029	834896	96.30878	860209	99.22874
CRS22	202784980029	850487	98.10727	864086	99.67597
CRS23	202784980029	852377	98.32529	864248	99.69466
CRS24	202784980029	856861	98.84254	864473	99.72061
CRS25	202784980029	822612	94.89177	857486	98.91463
CRS26	202784980029	734406	84.71683	822035	94.82521
P1	202784980001	843428	97.29298	862174	99.45541
P2	202784980001	768280	88.62434	846622	97.66142
P3	202784980001	826195	95.30508	857273	98.89006
P4	202784980001	791419	91.29352	846301	97.6244
P14	202784980001	818861	94.45908	854076	98.52127
P15	202784980001	861106	99.33221	865045	99.78659
P16	202784980001	773329	89.20677	844170	97.37858
P17	202784980001	852069	98.28976	863648	99.62544

Table S 7 Differentially Methylated Regions (DMRs), in CPM samples compared with primary CRC samples

DMR	Chr	Start (bp)	End (bp)	Length (bp)	DMR p-value	Located Gene	CpG island
DMR1	chr11	2160904	2161586	682	0.002	IGF2	302
DMR2	chr19	37157318	37157945	627	0.003	ZNF461	39
DMR3	chr15	79383167	79383980	813	0.003	RASGRF1	195
DMR4	chr18	35145983	35147090	1107	0.003	CELF4	196
DMR5	chr19	58609269	58609987	718	0.003	ZSCAN18	61
DMR6	chr13	78493229	78493712	483	0.003	EDNRB	70
DMR7	chr6	28602543	28603027	484	0.004	ZBED9	2
DMR8	chr5	135415693	135416613	920	0.005	VTRNA2-1	24
DMR9	chr19	58458572	58459358	786	0.005	ZNF256	52
DMR10	chr5	38258007	38259243	1236	0.008	EGFLAM	108
DMR11	chr13	96296844	96297281	437	0.013	DZIP1 EPB4 1L3	91
DMR12	chr18	5543271	5544237	966	0.010	EPB41L3	82
DMR13	chr8	65492528	65492936	408	0.016	LOC401463	143
DMR14	chr19	37328842	37329517	675	0.013	ZNF790	49
DMR15	chr5	146257347	146258427	1080	0.013	PPP2R2B	97
DMR16	chr19	38182773	38183418	645	0.016	ZNF781	47

Table S 8 Copy number alterations, primary CRC and CPM, p value < x10-07

Chr	Start	End	Type	Count	Score	P value
chr11	Gain	3817692	3818253	22	0.91	2.78 X10-07
chr11	Gain	3825449	4629195	22	0.91	2.78 X10-07
chr11	Gain	4629529	5704417	22	0.91	2.78 X10-07
chr11	Gain	5705995	5705995	22	0.91	2.78 X10-07
chr11	Gain	5712109	6518263	22	0.91	2.78 X10-07
chr11	Gain	6518907	6650480	22	0.91	2.78 X10-07
chr16	Gain	56635680	56645731	22	0.91	2.78 X10-07

chr16	Gain	57333722	57335017	22	0.91	2.78 X10-07
chr16	Gain	57495650	57503371	22	0.91	2.78 X10-07
chr16	Gain	71842906	71843647	22	0.91	2.78 X10-07
chr16	Gain	71844102	72096368	22	0.91	2.78 X10-07
chr16	Gain	72096449	72108059	22	0.91	2.78 X10-07
chr16	Gain	77246048	77246057	22	0.91	2.78 X10-07
chr16	Gain	77247012	77247012	22	0.91	2.78 X10-07
chr16	Gain	88877767	88878414	22	0.91	2.78 X10-07
chr5	Loss	32711355	32711431	21	0.87	6.73 X10-07
chr5	Loss	32711517	32713290	21	0.87	6.73 X10-07
chr5	Loss	32713464	32714010	21	0.87	6.73 X10-07
chr5	Loss	33892083	33892429	21	0.87	6.73 X10-07
chr5	Loss	33892457	33892457	21	0.87	6.73 X10-07
chr5	Loss	33936232	33936322	21	0.87	6.73 X10-07
chr5	Loss	33936402	33937673	21	0.87	6.73 X10-07
chr5	Loss	33937876	33945086	21	0.87	6.73 X10-07
chr5	Loss	38257541	38258007	21	0.87	6.73 X10-07
chr5	Loss	38258028	38258670	21	0.87	6.73 X10-07
chr5	Loss	38258853	38259214	21	0.87	6.73 X10-07

Table S 9 Copy number alteration, poor and good prognosis CPM, p value < x10-10

Chr	Start	End	Type	Count	Score	
chr3	6902689	6905031	Loss	12	1	2.22 X10-16
chr3	16554466	16554910	Loss	12	1	2.22 X10-16
chr3	36985697	36986697	Loss	12	1	2.22 X10-16
chr3	43020703	43022351	Loss	12	1	2.22 X10-16
chr3	44596684	44596846	Loss	12	1	2.22 X10-16
chr3	46939978	46940439	Loss	12	1	2.22 X10-16

chr4	25657365	25657628	Loss	12	1	2.22 X10-16
chr4	57975793	57976944	Loss	12	1	2.22 X10-16
chr4	81187125	81188051	Loss	12	1	2.22 X10-16
chr4	81951953	81952024	Loss	12	1	2.22 X10-16
chr4	84206068	84206074	Loss	12	1	2.22 X10-16
chr4	141489859	141490428	Loss	12	1	2.22 X10-16
chr14	76839844	76843917	Loss	12	1	2.22 X10-16
chr14	77227545	77228510	Loss	12	1	2.22 X10-16
chr14	77228690	77228753	Loss	12	1	2.22 X10-16
chr14	79744991	79745154	Loss	12	1	2.22 X10-16
chr14	89017776	89018372	Loss	12	1	2.22 X10-16
chr14	92788602	92790285	Loss	12	1	2.22 X10-16
chr15	84115811	84116151	Loss	12	1	2.22 X10-16
chr15	84116920	84116920	Loss	12	1	2.22 X10-16
chr17	27044852	27045113	Loss	12	1	2.22 X10-16
chr17	27045164	27046568	Loss	12	1	2.22 X10-16
chr17	30681379	31617566	Loss	12	1	2.22 X10-16
chr17	31618409	31620378	Loss	12	1	2.22 X10-16
chr17	31620500	32255252	Loss	12	1	2.22 X10-16
chr17	66596275	66597500	Loss	12	1	2.22 X10-16
chr19	16151557	16163163	Loss	12	1	2.22 X10-16
chr19	16223151	16235032	Loss	12	1	2.22 X10-16
chr19	16294776	16294776	Loss	12	1	2.22 X10-16
chr19	17378645	17392161	Loss	12	1	2.22 X10-16
chr19	17392298	17392923	Loss	12	1	2.22 X10-16
chr19	17393354	17415223	Loss	12	1	2.22 X10-16
chr19	17580131	17580131	Loss	12	1	2.22 X10-16
chr19	17581430	17581572	Loss	12	1	2.22 X10-16
chr9	35658919	36400252	Gain	12	1	2.22 X10-16
chr9	36401304	36598003	Gain	12	1	2.22 X10-16
chr9	36603035	37033489	Gain	12	1	2.22 X10-16

chr9	37033750	37800138	Gain	12	1	2.22 X10-16
chr9	37800654	70971005	Gain	12	1	2.22 X10-16
chr9	72435576	72436057	Gain	12	1	2.22 X10-16
chr9	74061323	74061323	Gain	12	1	2.22 X10-16
chr10	23461301	23489850	Gain	12	1	2.22 X10-16
chr10	23492470	23978755	Gain	12	1	2.22 X10-16
chr10	23982350	23984977	Gain	12	1	2.22 X10-16
chr10	31435733	31531682	Gain	12	1	2.22 X10-16
chr10	32240872	32266546	Gain	12	1	2.22 X10-16
chr11	625894	625894	Gain	12	1	2.22 X10-16
chr11	626307	636460	Gain	12	1	2.22 X10-16
chr11	637885	693717	Gain	12	1	2.22 X10-16
chr11	694295	703918	Gain	12	1	2.22 X10-16
chr11	704672	804954	Gain	12	1	2.22 X10-16
chr11	805666	809911	Gain	12	1	2.22 X10-16
chr11	1968310	1968980	Gain	12	1	2.22 X10-16

Chr, chromosomes, Type CNA, Start, start position of the CNA, End, end position of the CNA, count, number of samples carrying this type of event at this location, score, proportion of samples carrying this type of event at this location, GW_P, poisson-binomial test for overlap of event across individuals at this location given proportion chromosome affected, Gene – gene encoded by this portion of the genome.

Table S 10 Exome sequencing quality metrics

Sample	Percent aligned reads	Percent unique aligned reads	Unique base enrichment	Padded unique base enrichment	Percent targeted unique overlapping bases	Percent Q30 bases
N12	90.16	76.13	68.85	83.53	15.87	91.73
p26	87.34	84.52	73.88	81.72	25.01	90.75
N6	90.37	87.35	67.6	81.95	14.27	89.21
P11	86.91	82.14	73.81	82.74	21.92	85.23
C5	87.26	80.88	74.36	87.54	17.01	89.84
C16	91.37	88.86	64.7	78.33	14.65	92.25
N5	91.75	87.45	65.66	79.13	15.84	90.2
N10	91.29	85.74	65.22	85.16	9.7	92.11
P1	91.51	87.54	62.46	83.15	6.62	87.77
P9	90.02	86.39	64.94	82.9	8.09	87.02
C8	90.19	88.26	63.88	81.82	10.54	92.69
p23	92.87	90.52	66.54	83.69	9.97	93.24
N11	90.57	86.19	63.42	82.5	9.87	91.09
p16	92.34	90.74	71.72	85.25	13.79	92.68
C17	91.15	88.85	65.52	80.87	10.73	92.07
N7	91.76	88.42	60.04	78.95	8.31	90.3
C9	90.79	89.14	66.15	82.49	10.23	92
C12	92.91	90.82	67.58	82.33	13.18	93.18
P6	90.56	86.13	65.94	83.77	9.12	86.92
N17	93.42	90.57	69.84	85.47	13.26	93.63
CRS26	90.77	88.24	62.55	72.92	20.47	91.19
CRS23	92.23	90.58	60.76	78.92	8.45	90.9
p25	91.41	89.89	68.17	81.17	13.98	92.27
p18	92.25	91.14	69.37	84.55	11.59	93.09
P2	91.49	88.37	67.59	83.76	9.43	86.39
N22	93.22	90.92	65.01	84.95	7.66	92.71

C13	89.5	87.77	61.81	81.47	6.67	92.5
N21	91.14	88.15	68.59	85.36	10.97	91.98
C21	91.29	89.34	63.03	80.85	7.61	92.36
p24	92.72	90.86	68.41	83.99	10.96	93.17
p19	92.48	90.26	68.51	85.26	10.54	93.19
N20	92.93	88.48	63.86	83.46	9.86	93.24
C18	92.14	90.48	63.88	80.45	8.68	92.68
CRS25	92.46	90.05	59.04	75.25	10.62	90.59
p17	93.06	91.9	66.51	82.8	9.96	93.29
N2	92.02	90.25	55.33	73.31	8.66	90.43
C7	89.42	86.26	65.44	85.32	6.66	90.76
C14	92.47	90.71	64.93	82.07	8.48	93.25
C2	89.44	85.44	66.18	85.59	7.8	90.8
C4	89.71	86.42	66.36	86.11	7.16	90.86
N23	91.79	87.76	61.92	82.47	7.47	92.23
C10	93.14	92.04	61.86	79.23	9.48	93.48
N14	91.19	87.34	64.36	85.07	7.22	92.04
P10	89.83	84.64	66.29	81.58	10.57	87.4
p20	90.01	88.33	65.04	82.49	9.65	91.63
P13	89.04	84.25	64.16	75.85	15.02	87.26
CRS21	92.65	91.34	64.49	81.2	9.12	91.34
P4	92.48	88.84	65.92	81.81	10.75	88.78
N3	92.51	90.5	61.06	79.23	8.85	90.55
N1	91.88	89.74	59.95	79.65	7.16	90.53
P12	90.12	85.24	61.07	78.1	8.82	86.86
P3	91.19	85.89	57.35	73.64	10.08	87.1
CRS22	93.02	91.97	60.58	80.48	6.72	90.71
N26	91.22	87.77	67.9	84.52	10.22	92.59
p21	92.9	91.27	61.96	79.8	8.39	93.34
N15	91.2	88.71	65.99	82.95	10.04	92.38
p22	92.39	91.15	64.64	79.07	10.55	93.13
C15	92.36	90.85	63.28	80.66	8.15	93.24

C20	91.67	90.4	58.69	76.7	7.99	92.61
C1	90.73	87.93	67.01	85.72	7.8	91.73
N13	90.62	87.25	65.38	82.66	9.53	92.24
C3	88.3	83.18	65.91	84.36	7.37	91.17
p15	92.51	91.56	65.52	84.2	6.48	93.68
P14	89.13	84.02	59.35	74.32	9.95	87.29
C6	88.65	82.58	64.66	84.84	6.72	91.02
CRS24	92.69	89.55	61.48	78.33	10.04	91.06
P8	91.1	88.23	60.02	77.12	7.91	87.9

Table S 11 Genes with a bias towards high functional levels of mutations in the primary CRC cohort, potential drivers

	P-value	Q-value	PPH2 P-value	MA P-value	SIFT P-value
KRAS	3.66E-15	1.03E-11	0.044	0.000	0.000
TP53	3.90E-12	5.46E-09	0.000	0.000	0.002
CAD	1.36E-11	1.27E-08	0.001	0.000	0.005
NME1-NME2	1.35E-08	9.47E-06	0.081	0.000	0.054
NMD3	1.46E-06	0.001	0.001	0.001	0.011
RCN3	3.43E-05	0.014	0.010	0.002	0.011
CDC42BPA	3.38E-05	0.014	0.004	0.007	0.011
SDHA	5.17E-05	0.017	0.020	0.000	0.054
AP1S1	5.71E-05	0.017	0.029	0.000	0.054
ARHGEF12	6.22E-05	0.017	0.003	0.015	0.011
EFEMP1	7.35E-05	0.019	0.003	0.018	0.011
ZNF397	8.30E-05	0.019	0.006	0.002	0.054
MCCC2	9.68E-05	0.021	0.023	0.001	0.054
ATL2	0.000	0.021	0.006	0.003	0.054
CDH1	0.000	0.026	0.017	0.002	0.054
TTLL1	0.000	0.026	0.006	0.009	0.054

CRNKL1	0.000	0.026	0.165	0.001	0.011
ERICH1	0.000	0.026	0.006	0.009	0.054
HIBADH	0.000	0.026	0.006	0.009	0.054
GTPBP4	0.000	0.026	0.006	0.009	0.054
MSH6	0.000	0.026	0.003	0.040	0.016
ZNF410	0.000	0.026	0.006	0.009	0.054
POLR1B	0.000	0.026	0.087	0.002	0.011
OSER1	0.000	0.026	0.006	0.009	0.054
CEP350	0.000	0.026	0.013	0.089	0.002
EI24	0.000	0.026	0.006	0.009	0.054
AHCYL1	0.000	0.026	0.059	0.001	0.054
HLA-F	0.000	0.026	0.006	0.009	0.054
MPV17L2	0.000	0.026	0.006	0.009	0.054
EML2	0.000	0.028	0.006	0.011	0.054
RICTOR	0.000	0.028	0.013	0.014	0.019
PDE4DIP	0.000	0.028	0.017	0.010	0.021
CDKAL1	0.000	0.029	0.157	0.000	0.079
SPAG9	0.000	0.031	0.006	0.013	0.054
CNTRL	0.000	0.032	0.027	0.019	0.009
ZFYVE1	0.000	0.032	0.007	0.024	0.028
ELP2	0.000	0.032	0.010	0.009	0.054
ELP3	0.000	0.033	0.022	0.004	0.054
SMARCAL1	0.000	0.035	0.018	0.006	0.053
CHPF	0.001	0.036	0.006	0.018	0.054
DCHS1	0.001	0.036	0.006	0.011	0.096
APC	0.001	0.038	0.014	0.009	0.055
NAA15	0.001	0.038	0.011	0.011	0.054
SLC36A1	0.001	0.040	0.036	0.004	0.054
WDR19	0.001	0.040	0.014	0.011	0.054
HPS5	0.001	0.042	0.004	0.057	0.035

Identified using OncodriveFM, Intogen, Q-value <0.05, (PPH2, Polyphen2, MA Mutation Assesor, SIFT Sorting Intolerant from Tolerant).

Table S 12 The KEGG pathways with accumulated mutations of high functional impact in primary CRC

Kegg ID	Name	Class	Q-value
hsa00280	Valine, leucine and isoleucine degradation	Metabolism	7.54E-06
hsa05217	Basal cell carcinoma	Cancers: Specific types	2.13E-05
hsa05210	Colorectal cancer	Cancers: Specific types	0.000
hsa04726	Serotonergic synapse	Nervous system	0.000
hsa03008	Ribosome biogenesis in eukaryotes	Genetic; Translation	0.000
hsa00240	Pyrimidine metabolism	Metabolism	0.000
hsa05216	Thyroid cancer	Cancers: Specific types	0.000
hsa04730	Long-term depression	Nervous system	0.000
hsa03430	Mismatch repair	Genetic Replication and repair	0.001
hsa01100	Metabolic pathways	Metabolism	0.001
hsa05016	Huntington disease	Neurodegenerative disease	0.001
hsa03440	Homologous recombination	Genetic Replication and repair	0.001
hsa03030	DNA replication	Genetic Replication and repair	0.001
hsa03410	Base excision repair	Genetic Replication and repair	0.002
hsa05219	Bladder cancer	Cancers: Specific types	0.002
hsa00900	Terpenoid backbone biosynthesis	Metabolism	0.005
hsa00650	Butanoate metabolism	Metabolism	0.014

hsa05222	Small cell lung cancer	Cancers: Specific types	0.018
hsa00410	beta-Alanine metabolism	Metabolism	0.020
hsa05220	Chronic myeloid leukemia	Cancers: Specific types	0.020
hsa05212	Pancreatic cancer	Cancers: Specific types	0.020
hsa05213	Endometrial cancer	Cancers: Specific types	0.022
hsa04920	Adipocytokine signaling pathway	Endocrine system	0.024
hsa05030	Cocaine addiction	Human Diseases; Substance dependence	0.026
hsa00010	Glycolysis / Gluconeogenesis	Metabolism	0.032
hsa00640	Propanoate metabolism	Metabolism	0.032
hsa04340	Hedgehog signaling pathway	Signal transduction	0.032
hsa00071	Fatty acid degradation	Metabolism	0.032
hsa00340	Histidine metabolism	Metabolism	0.040
hsa00330	Arginine and proline metabolism	Metabolism	0.041
hsa00511	Other glycan degradation	Metabolism	0.054

Table S 13 Genes with a bias towards high functional levels of mutations in the CPM CRC cohort, potential drivers

	P-value	Q-value	PPH2 P-value	MA P-value	SIFT P-value
FLNB	1.38E-13	7.53E-10	1.00E-08	0.0001	0.000
TP53	1.10E-12	3.00E-09	0.000	1.00E-08	0.002
RIOK2	2.33E-12	3.73E-09	0.000	1.00E-08	0.001
PDE4DIP	3.42E-12	3.73E-09	0.001	1.00E-08	0.001
RRP7A	2.88E-12	3.73E-09	0.001	1.00E-08	0.001

SPTB	1.35E-11	1.22E-08	0.000	1.00E-08	0.009
PPL	1.92E-10	1.49E-07	1.00E-08	0.007	0.007
NUP98	2.49E-10	1.70E-07	1.00E-08	0.008	0.008
SVEP1	2.16E-09	1.31E-06	0.000	0.000	0.000
CDC16	1.75E-08	7.35E-06	0.000	0.000	0.003
HIPK1	1.75E-08	7.35E-06	0.000	0.000	0.003
TSC1	1.75E-08	7.35E-06	0.000	0.000	0.003
THAP6	1.75E-08	7.35E-06	0.000	0.000	0.003
CCHCR1	2.08E-08	8.12E-06	0.000	0.000	0.004
MED14	2.54E-08	9.25E-06	0.000	0.000	0.003
AHNAK2	8.26E-08	2.81E-05	0.001	0.000	0.005
MUC4	2.67E-07	8.55E-05	1.00E-08	0.147	0.775
ZNF462	5.26E-07	0.000	0.001	0.030	0.000
SPG7	6.63E-07	0.000	0.001	0.001	0.003
CENPJ	8.55E-07	0.000	0.001	0.002	0.002
AIFM1	9.91E-07	0.000	0.002	0.001	0.003
ZNF880	1.06E-06	0.000	0.000	0.004	0.004
ZMYM6	2.07E-06	0.000	0.001	0.003	0.007
TMEM2	2.17E-06	0.000	0.003	0.001	0.007
HSPG2	4.03E-06	0.001	0.010	0.002	0.001
FRG1BP	4.66E-06	0.001	0.000		0.003
KNL1	6.33E-06	0.001	0.001	0.006	0.011
AMACR	6.20E-06	0.001	0.018	0.000	0.011
POMT1	8.07E-06	0.002	0.008	0.004	0.002
TTL	9.06E-06	0.002	0.002	0.005	0.006
CNDP2	9.44E-06	0.002	0.012	0.000	0.050
GGT6	1.01E-05	0.002	0.003	0.004	0.005
NHS	9.84E-06	0.002	0.025	0.025	0.000
ALMS1	1.31E-05	0.002	0.000	0.040	0.007
BRWD1	1.31E-05	0.002	0.001	0.007	0.019
MYCBP2	1.76E-05	0.002	0.001	0.043	0.004

SH3YL1	3.45E-05	0.002	0.003	0.004	0.023
VCAN	2.04E-05	0.002	0.002	0.020	0.004
NEDD4L	3.45E-05	0.002	0.003	0.004	0.023
PIK3CB	3.45E-05	0.002	0.003	0.004	0.023
SYNE2	2.12E-05	0.002	0.008	0.001	0.021
GPATCH1	3.45E-05	0.002	0.003	0.004	0.023
RAB10	3.45E-05	0.002	0.003	0.004	0.023
DNMT3B	3.45E-05	0.002	0.003	0.004	0.023
GRAMD1A	3.45E-05	0.002	0.003	0.004	0.023
CRNKL1	3.45E-05	0.002	0.003	0.004	0.023
USP14	3.45E-05	0.002	0.003	0.004	0.023
ZNF419	3.45E-05	0.002	0.003	0.004	0.023
NCAPG	3.45E-05	0.002	0.003	0.004	0.023
SMAP1	2.19E-05	0.002	0.009	0.000	0.084
ITGA4	3.45E-05	0.002	0.003	0.004	0.023
DNAJC16	3.45E-05	0.002	0.003	0.004	0.023
MPL	3.45E-05	0.002	0.003	0.004	0.023
CCND2	3.45E-05	0.002	0.003	0.004	0.023
TMEM214	3.14E-05	0.002	0.001	0.019	0.011
EEF1E1	3.45E-05	0.002	0.003	0.004	0.023
ERGIC3	3.45E-05	0.002	0.003	0.004	0.023
LOXL1	2.42E-05	0.002	0.002	0.032	0.003
XPO7	3.06E-05	0.002	0.014	0.001	0.013
PTPRA	3.45E-05	0.002	0.003	0.004	0.023
ELP3	3.45E-05	0.002	0.003	0.004	0.023
CCNH	2.37E-05	0.002	0.016	0.004	0.003
MRPS9	3.45E-05	0.002	0.003	0.004	0.023
NDUFB5	3.45E-05	0.002	0.003	0.004	0.023
NAAA	3.45E-05	0.002	0.003	0.004	0.023
INTS7	3.45E-05	0.002	0.003	0.004	0.023
ZNF773	3.45E-05	0.002	0.003	0.004	0.023

ZKSCAN2	2.05E-05	0.002	0.001	0.010	0.011
TSPAN7	3.45E-05	0.002	0.003	0.004	0.023
ZNF276	2.72E-05	0.002	0.002	0.265	0.000
EDIL3	3.45E-05	0.002	0.003	0.004	0.023
CYB5A	3.45E-05	0.002	0.003	0.004	0.023
TNXB	1.72E-05	0.002	0.096	0.000	0.012
NPNT	3.29E-05	0.002	0.002	0.008	0.018
ADAL	3.45E-05	0.002	0.003	0.004	0.023
CSNK1G1	3.45E-05	0.002	0.003	0.004	0.023
CHCHD7	3.45E-05	0.002	0.003	0.004	0.023
SNCG	3.45E-05	0.002	0.003	0.004	0.023
RNF213	2.90E-05	0.002	0.002	0.024	0.005
LMLN	3.45E-05	0.002	0.003	0.004	0.023
CLN3	1.94E-05	0.002	0.001	0.079	0.003
FBF1	2.23E-05	0.002	0.020	0.020	0.000
ZNF720	3.45E-05	0.002	0.003	0.004	0.023
TSEN15	2.10E-05	0.002	0.001	0.037	0.003
MOB3C	3.55E-05	0.002	0.024	0.004	0.003
CUL7	3.70E-05	0.002	0.008	0.038	0.001
EML2	3.91E-05	0.002	0.003	0.005	0.023
LIG1	4.21E-05	0.003	0.007	0.005	0.011
MKI67	4.61E-05	0.003	0.000	0.061	0.020
HLA-DPB1	4.95E-05	0.003	0.277	0.001	0.002
SLC26A6	4.93E-05	0.003	0.016	0.001	0.021
WWC1	5.16E-05	0.003	0.001	0.019	0.027
CRYBG1	5.70E-05	0.003	0.004	0.009	0.013
ALDH3A2	5.94E-05	0.003	0.017	0.003	0.011
LAMC2	6.02E-05	0.003	0.055	0.006	0.002
PSPH	6.10E-05	0.003	0.022	0.001	0.023
ITGA6	6.24E-05	0.004	0.008	0.006	0.011
PHKB	7.06E-05	0.004	0.004	0.008	0.018

VRK2	7.19E-05	0.004	0.007	0.007	0.012
FAT4	7.47E-05	0.004	0.001	0.014	0.032
	8.11E-05	0.004	0.026	0.010	0.003
LIPE	9.58E-05	0.005	0.002	0.048	0.008
BOD1	9.63E-05	0.005	0.016	0.019	0.003
ATP6V0A2	9.66E-05	0.005	0.008	0.010	0.011
FNIP1	9.71E-05	0.005	0.007	0.012	0.011
TLR5	9.91E-05	0.005	0.009	0.004	0.026
NUMBL	0.000	0.005	0.003	0.014	0.023
MAP1A	0.000	0.006	0.000	0.010	1.000
FRAS1	0.000	0.006	0.030	0.001	0.024
CLEC2D	0.000	0.006	0.017	0.011	0.006
EGFLAM	0.000	0.006	0.012	0.010	0.008
FCHSD2	0.000	0.006	0.003	0.016	0.023
TET2	0.000	0.006	0.003	0.133	0.003
FKTN	0.000	0.006	0.003	0.025	0.016
PLCE1	0.000	0.006	0.009	0.098	0.001
PRUNE2	0.000	0.007	0.007	0.018	0.010
DPYD	0.000	0.007	0.007	0.011	0.018
GEMIN4	0.000	0.007	0.018	0.007	0.012
TSEN54	0.000	0.007	0.015	0.029	0.003
PTGES3L- AARSD1	0.000	0.007	0.004	0.133	0.003
TMED5	0.000	0.007	0.023	0.004	0.018
PPFIA1	0.000	0.007	0.004	0.038	0.009
WDR19	0.000	0.007	0.043	0.001	0.031
AKT1S1	0.000	0.007	0.004	0.074	0.006
EPHX2	0.000	0.007	0.005	0.019	0.018
TMEM63A	0.000	0.007	0.021	0.005	0.016
LPL	0.000	0.007	0.140	0.003	0.004
HLA-A	0.000	0.008	0.019	0.000	1.000

TCOF1	0.000	0.009	0.022	0.026	0.004
TCP11L1	0.000	0.009	0.021	0.003	0.030
CD320	0.000	0.009	0.003	0.219	0.003
KNSTRN	0.000	0.009	0.006		0.003
ZNF511	0.000	0.010	0.005	0.161	0.003
ASMTL	0.000	0.010	0.003	0.027	0.033
CAPN12	0.000	0.010	0.006	0.019	0.023
SMARCA1	0.000	0.010	0.009	0.008	0.036
ARHGEF3	0.000	0.010	0.013	0.023	0.008
SPG11	0.000	0.010	0.007	0.012	0.034
PDE3B	0.000	0.010	0.026	0.005	0.023
CLTCL1	0.000	0.010	0.010	0.011	0.025
SELL	0.000	0.011	0.009	0.010	0.034
TYK2	0.000	0.011	0.019	0.010	0.016
C16orf70	0.000	0.011	0.008	0.018	0.023
DDX49	0.000	0.012	0.014	0.010	0.023
ZSCAN32	0.000	0.012	0.002		0.013
XRRA1	0.000	0.012	0.005	0.020	0.039
APC	0.000	0.012	0.038	0.011	0.009
ZHX3	0.000	0.012	0.007	0.035	0.015
DNHD1	0.000	0.012	0.010	0.071	0.005
ZNF415	0.000	0.012	0.004	0.284	0.004
TBC1D5	0.000	0.012	0.005	0.083	0.008
PGM1	0.000	0.012	0.002	0.028	0.072
ACAP3	0.000	0.012	0.024	0.007	0.023
HLA-DRB5	0.000	0.012	0.311	0.001	0.012
UHRF1BP1	0.000	0.013	0.003	0.100	0.012
C20orf96	0.000	0.014	0.005	0.026	0.039
GGT1	0.000	0.014	0.015	0.019	0.016
KDEL2	0.000	0.015	0.050	0.004	0.023
HEATR1	0.000	0.015	0.019	0.009	0.030

KIF15	0.000	0.015	0.003	0.030	0.059
CYB5D1	0.000	0.016	0.016	0.115	0.003
CTTNBP2	0.000	0.016	0.006	0.024	0.039
CEP120	0.000	0.016	0.021	0.010	0.027
PALLD	0.001	0.017	0.033	0.013	0.013
TXNDC17	0.001	0.017	0.009	0.029	0.023
	0.000	0.017	0.004	0.042	0.039
ZNF626	0.001	0.017	0.003	0.026	0.078
CALCOCO1	0.001	0.017	0.007	0.020	0.040
HLA-DQB1	0.001	0.018	0.001	0.136	0.034
VWF	0.001	0.018	0.048	0.003	0.044
EPRS	0.001	0.018	0.098	0.005	0.016
MOK	0.001	0.019	0.022	0.006	0.053
FBRSL1	0.001	0.019	0.007	0.115	0.009
PKN2	0.001	0.020	0.009	0.010	0.084
INPP4B	0.001	0.020	0.009	0.010	0.084
RASA2	0.001	0.020	0.009	0.010	0.084
ACSF2	0.001	0.020	0.009	0.010	0.084
VEGFB	0.001	0.020	0.009	0.010	0.084
MAVS	0.001	0.020	0.006	0.054	0.025
MTRR	0.001	0.020	0.017	0.016	0.029
ABCB6	0.001	0.020	0.051	0.004	0.039
IQCB1	0.001	0.021	0.012	0.040	0.018
PTGR1	0.001	0.021	0.120	0.002	0.039
GBP3	0.001	0.022	0.050	0.005	0.035
SMYD4	0.001	0.022	0.017	0.025	0.023
TMEM116	0.001	0.023	0.023	0.007	0.059
ADAMTS5	0.001	0.023	0.004	0.025	0.090
AK9	0.001	0.023	0.012	0.187	0.005
TLR4	0.001	0.023	0.023	0.052	0.008
DOCK8	0.001	0.023	0.001	0.202	0.046

SMARCAL1	0.001	0.023	0.010	0.026	0.039
AMER1	0.001	0.023	0.007	0.063	0.023
SH2D4A	0.001	0.023	0.048	0.011	0.020
COPG1	0.001	0.024	0.279	0.003	0.013
SLC39A4	0.001	0.024	0.037	0.012	0.025
TP53BP1	0.001	0.024	0.008	0.139	0.010
UIMC1	0.001	0.025	0.002	0.068	0.115
TMEM176A	0.001	0.025	0.005	0.049	0.050
FAM72B	0.001	0.025	0.005	0.050	0.050
MAP1S	0.001	0.026	0.021	0.039	0.016
SP100	0.001	0.028	0.056	0.018	0.014
SALL2	0.001	0.028	0.004	0.148	0.023
KIAA1549	0.001	0.028	0.008	0.065	0.028
IDH3G	0.001	0.028	0.030	0.012	0.039
CDK5RAP1	0.001	0.028	0.010	0.059	0.023
RREB1	0.001	0.028	0.006	0.051	0.044
RNF43	0.001	0.029	0.017	0.161	0.006
RAPGEF3	0.001	0.029	0.006	0.144	0.018
WDR72	0.001	0.029	0.092	0.035	0.005
COX10	0.001	0.030	0.023	0.028	0.025
NAV2	0.001	0.031	0.003	0.029	0.201
WWOX	0.001	0.031	0.077	0.056	0.004
LGALS8	0.001	0.031	0.025	0.013	0.050
NARFL	0.001	0.031	0.028	0.026	0.023
ZP3	0.001	0.031	0.047	0.018	0.020
PACS1	0.001	0.031	0.014	0.021	0.058
ESYT2	0.001	0.031	0.005	0.072	0.050
SFI1	0.001	0.031	0.003	0.142	0.044
ZNF17	0.001	0.031	0.002	0.317	0.025
MTHFD1L	0.001	0.031	0.016	0.016	0.071
SKIV2L	0.001	0.032	0.186	0.034	0.003

NEK4	0.001	0.032	0.038	0.099	0.005
MRPL9	0.001	0.032	0.355	0.007	0.008
VAV3	0.001	0.033	0.121	0.006	0.025
MTMR2	0.001	0.033	0.292	0.005	0.012
ZNF777	0.001	0.033	0.012	0.052	0.032
COG5	0.001	0.034	0.064	0.023	0.014
MROH6	0.001	0.035	0.005	0.149	0.029
AK1	0.001	0.036	0.061	0.004	0.082
C1orf112	0.002	0.036	0.022	0.043	0.023
CD2AP	0.002	0.036	0.013	0.071	0.023
C3orf52	0.002	0.036	0.008	0.058	0.050
MICB	0.002	0.037	0.022	0.012	0.087
LAMA5	0.002	0.037	0.017	0.024	0.058
FAM83H	0.002	0.037	0.015	0.132	0.012
SZT2	0.002	0.038	0.019	0.154	0.008
OLFM2	0.002	0.038	0.016	0.188	0.008
MCAM	0.002	0.039	0.021	0.019	0.063
ZNF28	0.002	0.039	0.040	0.014	0.046
IGFBP3	0.002	0.039	0.017	0.025	0.059
SIPA1	0.002	0.039	0.009	0.087	0.032
MSH6	0.002	0.039	0.000	0.326	0.266
AARS2	0.002	0.039	0.021	0.021	0.061
UTP23	0.002	0.039	0.023	0.135	0.008
RASAL1	0.002	0.040	0.008	0.070	0.050
DOCK4	0.002	0.040	0.016	0.020	0.084
STEAP4	0.002	0.040	0.086	0.014	0.023
SUN1	0.002	0.040	0.037	0.133	0.006
IFT88	0.002	0.041	0.140	0.036	0.006
ZNF268	0.002	0.041	0.058	0.148	0.003
TMEM63B	0.002	0.042	0.027	0.040	0.027
ADGRL4	0.002	0.042	0.013	0.025	0.091

KCNN4	0.002	0.043	0.051	0.025	0.023
TIMM44	0.002	0.043	0.059	0.012	0.043
MCPH1	0.002	0.043	0.010	0.041	0.074
ERGIC1	0.002	0.043	0.034	0.039	0.023
RPP40	0.002	0.043	0.804	0.003	0.012
BDP1	0.002	0.043	0.062	0.040	0.012
QSOX2	0.002	0.043	0.024	0.012	0.106
LAMB3	0.002	0.043	0.113	0.005	0.053
ZNF700	0.002	0.043	0.009	0.056	0.061
E4F1	0.002	0.043	0.020	0.067	0.023
VPS37A	0.002	0.043	0.098	0.004	0.084
WDR90	0.002	0.044	0.090	0.041	0.009
KLHL25	0.002	0.044	0.029	0.051	0.023

Identified using OncodriveFM, Intogen, Q-value <0.05, (PPH2, Polyphen2, MA Mutation Assesor, SIFT Sorting Intolerant from Tolerant).

Table S 14 The KEGG pathways with accumulated mutations of high functional impact in CPM CRC

KEGG	Name	Class	Q-value
hsa04512	ECM-receptor interaction	Signaling molecules and interaction	1.90E-07
hsa04940	Type I diabetes mellitus	Endocrine and metabolic disease	1.90E-07
hsa05330	Allograft rejection	Immune diseases	2.09E-07
hsa05332	Graft-versus-host disease	Immune diseases	3.06E-07
hsa05145	Toxoplasmosis	Infectious diseases	6.45E-07
hsa05310	Asthma	Immune diseases	6.45E-07
hsa05320	Autoimmune thyroid disease	Immune diseases	9.86E-07

hsa04672	Intestinal immune network for IgA production	Immune system	1.42E-05
hsa00410	beta-Alanine metabolism	Metabolism	4.32E-05
hsa05217	Basal cell carcinoma	Cancers: Specific types	0.000
hsa05222	Small cell lung cancer	Cancers: Specific types	0.000
hsa05416	Viral myocarditis	Human Diseases; Cardiovascular diseases	0.00
hsa00053	Ascorbate and aldarate metabolism	Metabolism	0.00
hsa04514	Cell adhesion molecules	Signaling molecules and interaction	0.00
hsa05168	Herpes simplex infection	Infectious diseases	0.01
hsa03008	Ribosome biogenesis in eukaryotes	Genetic; Translation	0.01
hsa00340	Histidine metabolism	Metabolism	0.01
hsa00120	Primary bile acid biosynthesis	Metabolism	0.02
hsa00770	Pantothenate and CoA biosynthesis	Metabolism	0.03
hsa03430	Mismatch repair	Genetic; Replication and repair	0.03
hsa04612	Antigen processing and presentation	Immune system	0.03
hsa05150	Staphylococcus aureus infection	Infectious diseases	0.03
hsa05132	Influenza A	Infectious diseases	0.03
hsa05146	Amoebiasis	Infectious diseases	0.04
hsa00670	One carbon pool by folate	Metabolism	0.04

hsa04620	Toll-like receptor signaling pathway	Immune system	0.04
hsa05164	Influenza A	Infectious diseases	0.04

Appendix 2: Published works arising from this thesis

Meta-analysis of prognostic factors for patients with colorectal peritoneal metastasis undergoing cytoreductive surgery and heated intraperitoneal chemotherapy

S. Hallam¹, R. Tyler¹, M. Price², A. Beggs¹ and H. Youssef³

¹Institute of Cancer and Genomic Sciences and ²Institute of Applied Health Research, University of Birmingham, and ³Colorectal Surgery, Good Hope Hospital, University Hospitals Birmingham NHS Foundation Trust, Birmingham, UK

Correspondence to: Ms S. Hallam, Institute of Cancer and Genomic Sciences, University of Birmingham, Edgbaston, Birmingham B15 2SY, UK (e-mail: sally84hallam@gmail.com)

Background: Up to 15 per cent of colorectal cancers present with peritoneal metastases (CPM). Cyto-reductive surgery and heated intraperitoneal chemotherapy (CRS + HIPEC) aims to achieve macroscopic tumour resection combined with HIPEC to destroy microscopic disease. CRS + HIPEC is a major operation with significant morbidity and effects on quality of life (QoL). Improving patient selection is crucial to maximize patient outcomes while minimizing morbidity and mortality. The aim of this study was to identify prognostic factors for patients with CPM undergoing CRS + HIPEC.

Methods: A systematic search of MEDLINE, Embase and Cochrane Library electronic databases was performed using terms for colorectal cancer, peritoneal metastasis and CRS + HIPEC. Included studies focused on the impact of prognostic factors on overall survival following CRS + HIPEC in patients with CPM.

Results: Twenty-four studies described 3128 patients. Obstruction or perforation of the primary tumour (hazard ratio (HR) 2.91, 95 per cent c.i. 1.5 to 5.65), extent of peritoneal metastasis as described by the Peritoneal Carcinomatosis Index (PCI) (per increase of 1 PCI point: HR 1.07, 1.02 to 1.12) and the completeness of cytoreduction (CC score above zero: HR 1.75, 1.18 to 2.59) were associated with reduced overall survival after CRS + HIPEC.

Conclusion: Primary tumour obstruction or perforation, PCI score and CC score are valuable prognostic factors in the selection of patients with CPM for CRS + HIPEC.

Funding information

No funding

Paper accepted 3 April 2019

Published online 27 June 2019 in Wiley Online Library (www.bjsopen.com). DOI: 10.1002/bjs.5.50179

Introduction

Colorectal peritoneal metastasis (CPM) occurs in up to 15 per cent of patients with colorectal cancer^{1–3}. The prognosis of patients with CPM is poor: untreated median overall survival (OS) is just 6 months⁴. With systemic chemotherapy, median OS is improved to up to 20 months^{5–7}. The standard for CPM is cytoreductive surgery and heated intraperitoneal chemotherapy (CRS + HIPEC), which may improve median OS by 20–63 months^{8–16}.

CRS + HIPEC is a long, high-cost operation associated with a protracted inpatient and high-dependency or intensive care unit stay, and an associated mortality rate of 1–12 per cent and morbidity rate of 7–63 per cent^{11,14,16}.

²³. Improving patient selection is therefore

crucial to maximize patient outcomes whilst minimizing morbidity and mortality.

Variation in outcomes for CRS + HIPEC can be explained in part by patient selection, for example necessitating the ability to achieve complete cytoreduction at CRS, the exclusion of patients with extensive CPM as assessed by Sugarbaker's Peritoneal Carcinomatosis Index (PCI), and selection of patients with minimal co-morbidity and good performance status.

Several clinicopathological variables that impact on survival have been identified in the literature, including lymph node (LN) status, tumour differentiation and histological findings, the completeness of cytoreduction (CC score) and PCI. There is wide variation, however, in the variables reported by studies and selection criteria for

centres performing CRS + HIPEC worldwide, and in turn in their outcomes.

Studies reporting prognostic factors for CRS + HIPEC are cohort in design with small samples, and each examines numerous and varying prognostic factors on differing scales of measurement. No consensus exists as to which prognostic factors contraindicate CRS + HIPEC, or predict a good outcome. A comprehensive evidence synthesis is therefore called for to determine relevant prognostic factors.

The aim of this systematic review and meta-analysis was to analyse all prognostic factors affecting OS in patients with CPM undergoing CRS + HIPEC.

Methods

A comprehensive literature search was conducted in accordance with the PRISMA guidelines²⁴. MEDLINE, Embase and Cochrane Library electronic databases, registers of clinical trials (ClinicalTrials.gov and the WHO International Clinical Trials Registry) and the Conference Proceedings Citation Index, Zetoc, were searched from inception to the present. The search strategy captured terms for colorectal cancer, peritoneal metastasis and CRS + HIPEC techniques, separated by the Boolean operator 'AND'. For an example search strategy, see *Appendix S1* (supporting information). Searches were supplemented by a hand search of selected journals and the reference lists of all included studies.

Study selection

English-language articles were eligible for inclusion if they reported on the impact of prognostic factors on OS in patients with CPM undergoing CRS + HIPEC. Where multiple studies described the same cohort of patients, the largest and most complete data set was included. Review articles, case reports and case series of fewer than ten patients were excluded. Additional exclusion criteria included studies involving patients with a primary tumour other than colorectal cancer and studies in which a proportion of the cohort did not receive combined CRS + HIPEC.

After screening the titles and abstracts, articles fulfilling the eligibility criteria were identified and their full-text publications reviewed. Literature search and study selection were done independently by two researchers, and any disagreements were resolved by discussion with senior reviewers. After qualitative assessment, articles were screened to ensure they presented adequate statistical information to be included in the meta-analysis: hazard ratios (HRs) with confidence intervals or Kaplan–Meier curves with the number of events and patients at risk. When

HRs, confidence intervals or *P* values were not provided directly, the methods of Tierney *et al.*²⁵ were used to estimate them indirectly from Kaplan–Meier curves, when presented in adequate detail with the numbers of events and patients at risk²⁶.

Assessment of risk of bias

The quality and risk of bias of individual studies was assessed using the Quality in Prognosis Studies (QUIPS) tool²⁷. This tool reviews each study according to six criteria: study participation, attrition, prognostic factor measurement, outcome measurement, confounding factors, and statistical analysis and reporting. Two authors scored all articles independently.

Data extraction

Data were extracted independently by two reviewers using a dedicated and piloted data extraction form. The number of patients, study design, patient demographics, tumour characteristics, use of adjuvant and neoadjuvant regimens, CRS + HIPEC techniques, survival and prognostic factors were recorded.

The unadjusted HR and its 95 per cent c.i. and *P* value were extracted. When adjusted HRs (with confidence intervals and *P* values) were reported, these were extracted along with the set of adjustment factors used. If HRs, confidence intervals or *P* values were not provided directly, the methods of Tierney and colleagues²⁵ were used to estimate them indirectly from Kaplan–Meier curves, when presented in adequate detail with the numbers of events and patients at risk²⁶.

Prognostic factors reported on a continuous scale were extracted. If results were categorized into three or more categories, results for each comparison were extracted and, when clinically relevant, were grouped to form a binary comparison.

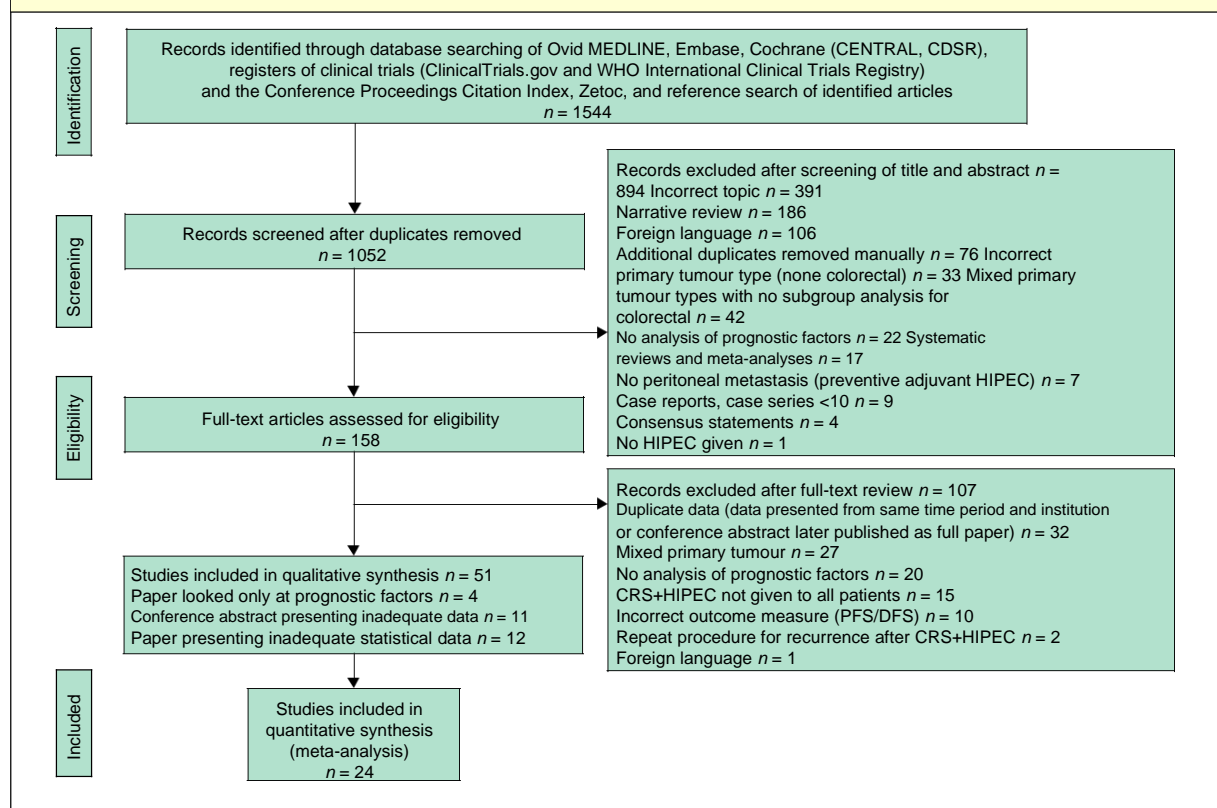
Prognostic factor selection

All prognostic factors described adequately and reported by two or more independent studies were included.

Statistical analysis

Owing to clinical and methodological heterogeneity, a random-effects meta-analysis was used (on the log(HR) scale) using the method of DerSimonian and Laird²⁸. The combined effect size was described by the pooled HR, its confidence interval and *P* value, with *P* < 0.050 considered significant. For prognostic factors reported by more than

Fig. 1 PRISMA diagram for the review



CENTRAL, Cochrane Central Register of Controlled Trials; CDSR, Cochrane Database of Systematic Reviews; DARE, Database of Abstracts of Reviews of Effects; HIPEC, heated intraperitoneal chemotherapy; CRS, cytoreductive surgery; PFS, progression-free survival; DFS, disease-free survival.

tudies, a 95 per cent prediction interval (a measure of the variation in treatment effects) is presented²⁹. Heterogeneity was also described by the I^2 statistic³⁰. All analyses were performed in STATA[®] version 15 (StataCorp, College Station, Texas, USA).

two s

Results

The final literature search was performed on 3 April 2018. Literature searches (after removal of duplicates) identified 1052 records. Titles and abstract screening identified 158 full-text articles for review. Of these, 51 studies met the inclusion criteria. Twenty-four unique studies^{11,18,20,31–51} reporting on 3128 patients with CPM presented adequate data to be included in the meta-analysis (Fig. 1). Of the 24 cohort studies, were prospective and the remaining 18^{11,20,31–33,35,36,38,39,41–43,45,46,48–51} were retrospective (Table S1, supporting Pooled median OS across all studies was 32 (range 12–51) months, with a pooled median follow-up of

information). A number of studies presented both an unadjusted and adjusted HR. As adjustment factors varied widely between studies (Table S2, supporting information), meta-analysis was stratified according to whether the HR was unadjusted or adjusted.

Quality of research

There was a low risk of bias from study participation. The moderate risk of bias due to study attrition reflects the poor reporting of loss to follow-up. There was a low risk of bias due to prognostic factor measurement. Prognostic factors were objective, clearly defined and clinically relevant.

Reporting of all prognostic factors and treatment variations was incomplete in the majority of studies, resulting in a moderate risk of bias. The presentation of results and analytical strategy was sufficient in the majority of studies resulting in a low risk of bias (Table S3, supporting information).

28.1 (13.3–62.4) months (Table S4, supporting information).

Patient factors

Age

The pooled median age of patients included was 54 (range 45.5–69.3) years. Eleven studies^{18,20,33,34,37,40,42,48–50,52} reported on age as a prognostic factor and ten presented adequate data to be included in the meta-analysis. Five studies categorized age into binary outcomes that could not be combined meaningfully. The pooled unadjusted HR was 1.00 (95 per cent c.i. 0.98 to 1.03) ($I^2 = 38.5$ per cent) and adjusted HR was 1.00 (0.96 to 1.04) ($I^2 = 69.3$ per cent), for an increase in age of 1 year (Fig. 2; Fig. S1, supporting information). These data provide no evidence that age is a useful predictor of OS.

Sex

Nine studies^{18,20,34,36,37,40,42,51,52} reported on the effect of sex on OS and eight presented adequate data to be included in the meta-analysis. The pooled unadjusted HR for male sex was 1.20 (95 per cent c.i. 0.83 to 1.71) ($I^2 = 33.8$ per cent) and adjusted HR was 1.73 (1.20 to 2.48) ($I^2 = 11.6$ per cent) (Fig. 2; Fig. S1, supporting information). It is therefore unclear whether sex is a useful predictor of OS.

Eastern Cooperative Oncology Group performance status

Five studies^{18,31,43,52,53} reported on the influence of Eastern Cooperative Oncology Group (ECOG) status on OS, and three presented adequate data to be included in the meta-analysis. The pooled unadjusted HR for an ECOG score of at least 2 was 1.77 (95 per cent c.i. 0.85 to 3.67) ($I^2 = 0$ per cent) (Fig. 2; Fig. S1, supporting information). These data provide no evidence that ECOG is a useful predictor of OS.

Tumour factors

Adverse primary tumour features

Three studies^{18,33,52} reported on the influence of adverse features of the primary tumour (obstruction or perforation) on OS; two presented adequate data to be included in the meta-analysis. The pooled adjusted HR for adverse features of the primary tumour was 2.91 (95 per cent c.i. 1.5 to 5.64) ($I^2 = 0$ per cent) (Fig. 2; Fig. S1, supporting information). These data indicate that adverse primary features

are a useful predictor of decreased OS (Table S5, supporting information).

Rectal or colonic primary

Eight studies^{20,34,36,40,43,47,53,54} reported on the influence of a rectal or colonic primary on OS and six presented adequate data to be included in the meta-analysis. The pooled

unadjusted HR for rectal primary (compared with colonic primary) was 1.48 (95 per cent c.i. 0.71 to 3.09) ($I^2 = 73.8$ per cent) (Fig. 2; Fig. S1, supporting information). These data provide no evidence that a rectal primary is a useful predictor of OS.

Lymph node metastasis

Eleven studies^{11,18,33–35,37,48–50,53,54} reported on the effect of LN status on OS; nine presented adequate data to be included in the meta-analysis. The pooled unadjusted HR for positive LNs (compared with negative LNs) was 1.42 (95 per cent c.i. 1.06 to 1.92) ($I^2 = 0$ per cent) and adjusted HR was 1.08 (0.60 to 1.95) ($I^2 = 68.6$ per cent) (Fig. 2; Fig. S1, supporting information). It is therefore unclear whether lymph nodes are a useful predictor of OS.

Tumour differentiation

Two studies^{34,40} reported on the effect of primary tumour differentiation on OS. The pooled unadjusted HR for well differentiated tumours was 0.90 (95 per cent c.i. 0.18 to 4.40) ($I^2 = 72.4$ per cent) (Fig. 2; Fig. S1, supporting information). These data provide no evidence that primary tumour differentiation is a useful predictor of OS.

Timing of colorectal peritoneal metastasis (synchronous or metachronous)

Three studies^{11,40,54} reported on the effect of the timing of CPM (synchronous or metachronous) on OS. Two presented adequate data to be included in the meta-analysis. The pooled unadjusted HR for synchronous CPM was 1.13 (95 per cent c.i. 0.56 to 2.28) ($I^2 = 57.1$ per cent) (Fig. 2; Fig. S1, supporting information), indicating that the timing of CPM is not a useful predictor of OS.

Ascites

Two studies^{18,40} reported on the effect of malignant ascites on OS. One paper⁴⁰ presented an unadjusted HR (1.50, 95 per cent c.i. 0.70 to 3.21) and one¹⁸ an adjusted HR (3.23, 1.56 to 6.69) (Fig. 2; Fig. S1, supporting information), so it was not possible to provide a pooled effect estimate.

Hepatic metastasis

Eleven studies^{18,31,34–38,43,44,47,50} reported on the effect of surgically treated hepatic metastasis on OS. Seven pre-sented adequate data to be included in the meta-analysis. The pooled unadjusted HR for surgically treated hepatic metastasis was 1.36 (95 per cent c.i. 0.85 to 2.16) ($I^2 = 43.6$ per cent) and the **adjusted HR was 2.49** (0.74 to 8.41) ($I^2 = 85.3$ per cent) (Fig. 2; Fig. S1, supporting information). These data provide no evidence that the presence of surgically treated hepatic metastasis is a useful predictor of OS.

Treatment factors

Previous surgical score

Three studies^{11,34,42} reported on the influence of the previous surgical score (PSS) on OS. The pooled unadjusted HR for a PSS of at least 2 was 1.23 (95 per cent c.i. 0.78 to 1.95) ($P = 0$ per cent) (Fig. 2; Fig. S1, supporting information). These data provide no evidence that PSS is a useful predictor of OS.

Postoperative morbidity

Three studies^{40,43,50} reported on the association between postoperative morbidity and OS. The pooled unadjusted HR for a Clavien–Dindo complication of grade III or above was 1.56 (95 per cent c.i. 0.74 to 3.31) ($P = 48.7$ per cent) (Fig. 2; Fig. S1, supporting information). These data provide no evidence that postoperative morbidity is a useful predictor of OS.

Neoadjuvant and adjuvant chemotherapy

Neoadjuvant and adjuvant treatment of the primary tumour was reported poorly (Table S6, supporting information). No study reported response to treatment, or whether it was completed as planned.

Two studies^{11,43} reported on the effect of neoadjuvant chemotherapy before CRS + HIPEC on OS. The pooled unadjusted HR for the use of neoadjuvant chemotherapy was 1.00 (95 per cent c.i. 0.63 to 1.58) ($P = 0$ per cent) (Fig. 2; Fig. S1,

supporting information). These data provide no evidence that neoadjuvant chemotherapy is a useful predictor of OS.

Adjuvant chemotherapy

Four studies^{11,35,43,50} reported on the effect of adjuvant chemotherapy after CRS + HIPEC on OS. The pooled unadjusted HR for adjuvant chemotherapy was 0.60 (95 per cent c.i. 0.29 to 1.21) ($P = 68.1$ per cent) (Fig. 2; Fig. S1, supporting information), which suggests there is no evidence that use of adjuvant chemotherapy is predictive of OS.

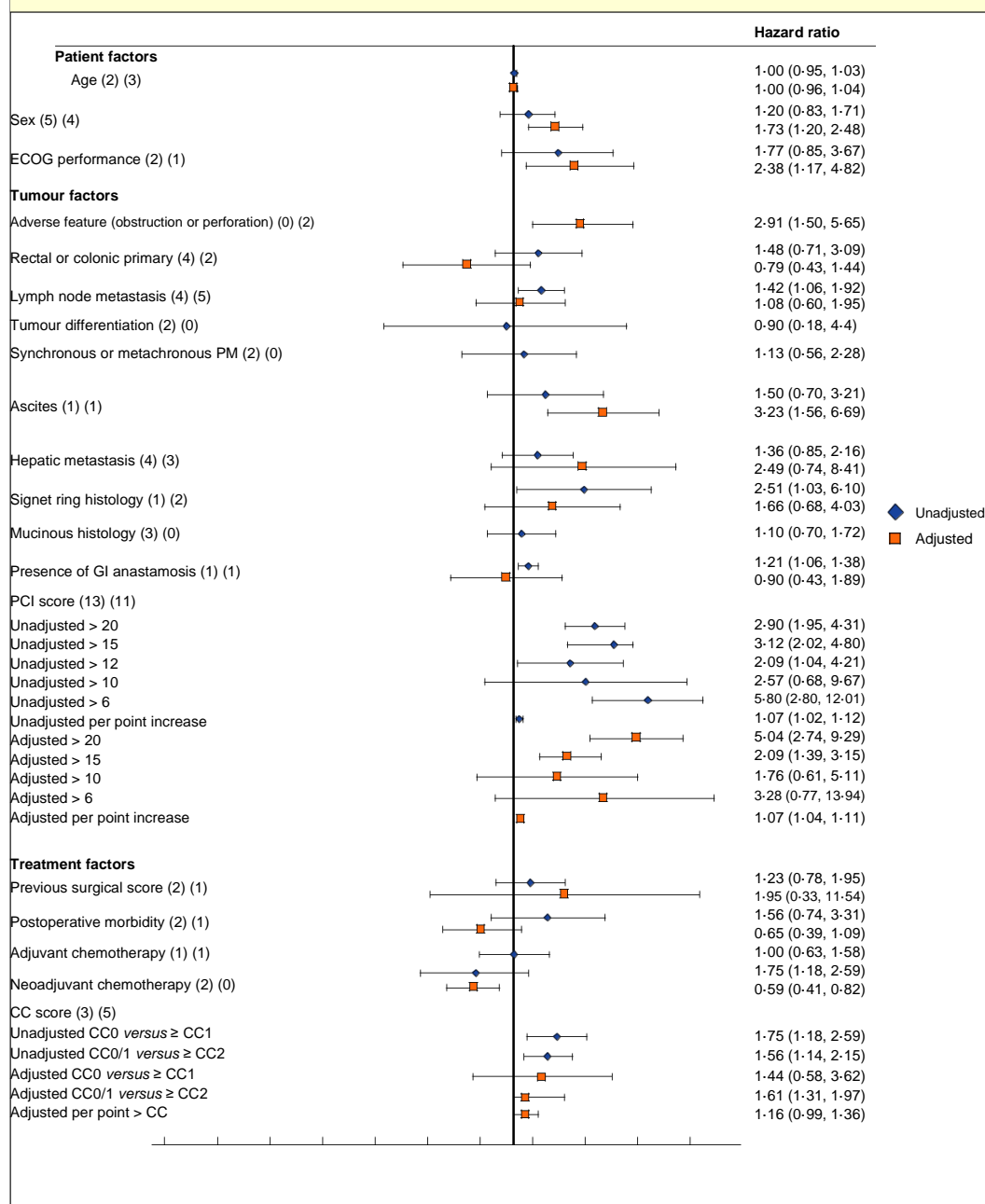
Completeness of cytoreduction

Eight studies^{11,32,35,40,42,43,51,55} reported on the effect of the completeness of cytoreduction on OS; seven presented adequate data to be included in the meta-analysis. A CC score above zero was predictive of a reduction in OS (un-adjusted HR 1.75, 95 per cent c.i. 1.18 to 2.59) ($P = 79.5$ per cent), as was a CC score of 2 or more (HR 1.61, 1.31 to 1.97) ($P = 0$ per cent) (Fig. 2; Fig. S1, supporting information).

Discussion

This systematic review and meta-analysis examined the effect of prognostic factors on OS following CRS + HIPEC for CPM. Emergency presentation with obstruction or perforation of the primary tumour as well as the extent of CPM and the completeness of resection, as described by the PCI and the CC score respectively, were the only significant prognostic factor

Fig. 2 Effect of prognostic factors on overall survival



Hazard ratios are shown with 95 per cent confidence intervals. Values in parentheses after each factor indicate the numbers of studies providing unadjusted and adjusted hazard ratios respectively. ECOG, Eastern Cooperative Oncology Group; PM, peritoneal metastasis; GI, gastrointestinal; PCI, Peritoneal Carcinomatosis Index; CC, completeness of cytoreduction.

An emergency presentation of the primary tumour with obstruction or perforation was predictive of reduced OS (for both synchronous and metachronous CPM). A number of factors may contribute to this. In the primary setting, colorectal cancers presenting with obstruction or perforation are associated with decreased cancer-specific survival and increased postoperative mortality⁵⁶. Obstructed or perforated colorectal cancer is, by definition, advanced in stage and increases the risk of metastasis; additionally, emergency presentation limits the possibility of neoadjuvant treatment and may delay adjuvant treatment owing to postoperative morbidity. The extent of peritoneal metastasis as described by the PCI was predictive of reduced OS as a continuous variable and when the PCI score was 12 or above. A complete cytoreduction (CC0) was predictive of improved OS. Improving patient selection is therefore reliant on the ability to predict accurately the extent of peritoneal metastasis and the ability to resect it completely. Specialist radiologists have demonstrated good concordance between radiological and surgical PCI estimations, particularly when combining modalities; however, these tend to be most accurate in patients with high PCI scores^{57,58}. In some centres this is used in combination with diagnostic laparoscopy before CRS + HIPEC. This is feasible in the majority of patients and may help to reduce the laparotomy rate in patients for whom CRS + HIPEC may not be possible⁵⁹.

Included patients were relatively young at 54 (range 45–69) years compared with the incident age of colorectal cancer (80–90 years)⁶⁰. In addition, performance status was not reported by the majority of studies in the review, and limited to an ECOG score of less than 2 by a further nine. Within these limits, no other patient or tumour factor was predictive of OS.

Details of neoadjuvant and adjuvant treatments were reported poorly in the included studies. Within these limits, the use of neoadjuvant or adjuvant chemotherapy was not predictive of OS after CRS + HIPEC. One meta-analysis, by Kwakman and colleagues⁶¹ from 2016, examined the effect of clinicopathological variables only on OS following CRS + HIPEC. Significant prognostic factors identified in the present review (adverse features of the primary tumour, PCI, CC score) were not comparable with those from Kwakman *et al.*⁶¹ as they were not examined. In contrast to the present study, Kwakman and co-workers⁶¹ found performance status, the presence of lymph node or hepatic metastasis, tumour differentiation, signet ring histology, a rectal primary and the use of neoadjuvant chemotherapy to be predictive prognostic factors. A number of differences may explain this variation in findings: the

exclusion of 73 papers described as unavailable in full text may introduce a potential selection bias⁶¹. In addition, a number of studies included by Kwakman and colleagues were excluded in the present study for the following reasons: presentation of inadequate data to estimate the HR accurately^{13,52,54,62}, the inclusion of mixed primary tumours^{48,63,64}, and the inclusion of patients having repeat CRS + HIPEC procedures. Finally, the meta-analysis⁶¹ combined all studies regardless of the presentation of unadjusted or adjusted HR.

The strengths of the present study include the comprehensive and systematic literature search including 3128 patients with CPM undergoing CRS + HIPEC, all potential prognostic factors were included, and a consistent association was found between predictive prognostic factors across different studies. The low heterogeneity associated with these factors adds to the strength and generalizability of the findings. The application of strict inclusion criteria limits the potential impact of factors such as mixed primary tumour origin. The present analysis takes into account the adjustment factors used in primary studies to ensure meta-analysis of time to event data was performed only when data were comparable.

A number of limitations must, however, be acknowledged. This review is limited by the quality of primary studies and the heterogeneity of the population. Prognostic factor systematic reviews, by their nature, represent one of the most difficult categories due to their retrospective and observational nature. As CRS + HIPEC is performed at tertiary centres, data concerning the primary tumour and its treatment may not have been captured or reported fully. Additionally, the statistical analysis and presentation of data necessary for accurate meta-analysis varied widely across primary studies, which may introduce a degree of inclusion bias.

There was a lack of molecular and genetic data concerning patients with CPM undergoing CRS + HIPEC in comparison with studies of primary colorectal cancer, and this is an area for future research. Discordance between primary genetic mutations and those in peritoneal metastases may identify novel therapeutic targets and prognostic markers in this metastatic group. Recent research by Schneider and colleagues⁶⁵ found that *RAS/RAF* mutations impair survival after CRS/HIPEC, although this is the only study that has considered this prognostic factor and validation is required. Further large-scale research is needed to account for these factors; this would require collaboration between CRS + HIPEC centres, standardization of the analysis and presentation of prognostic factor time to event data.

Disclosure

The authors declare no conflict of interest.

References

1. Jayne DG, Fook S, Loi C, Seow-Choen F. Peritoneal carcinomatosis from colorectal cancer. *Br J Surg* 2002; **89**: 1545–1550.
2. Segelman J, Granath F, Holm T, Machado M, Mahteme H, Martling A. Incidence, prevalence and risk factors for peritoneal carcinomatosis from colorectal cancer. *Br J Surg* 2012; **99**: 699–705.
3. Quere P, Facy O, Manfredi S, Jooste V, Faivre J, Lepage C *et al.* Epidemiology, management, and survival of peritoneal carcinomatosis from colorectal cancer: a population-based study. *Dis Colon Rectum* 2015; **58**: 743–752.
4. Chu DZ, Lang NP, Thompson C, Osteen PK, Westbrook KC. Peritoneal carcinomatosis in nongynecologic malignancy. A prospective study of prognostic factors. *Cancer* 1989; **63**: 364–367.
5. Klaver YL, Simkens LH, Lemmens VE, Koopman M, Teerenstra S, Bleichrodt RP *et al.* Outcomes of colorectal cancer patients with peritoneal carcinomatosis treated with chemotherapy with and without targeted therapy. *Eur J Surg Oncol* 2012; **38**: 617–623.
6. van Oudheusden TR, Razenberg LG, van Gestel YR, Creemers GJ, Lemmens VE, de Hingh IH. Systemic treatment of patients with metachronous peritoneal carcinomatosis of colorectal origin. *Sci Rep* 2015; **5**: 18632.
7. Franko J, Shi Q, Goldman CD, Pockaj BA, Nelson GD, Goldberg RM *et al.* Treatment of colorectal peritoneal carcinomatosis with systemic chemotherapy: a pooled analysis of north central cancer treatment group phase III trials N9741 and N9841. *J Clin Oncol* 2012; **30**: 263–267.
8. Verwaal VJ, Bruin S, Boot H, van Slooten G, van Tinteren H. 8-year follow-up of randomized trial: cytoreduction and hyperthermic intraperitoneal chemotherapy *versus* systemic chemotherapy in patients with peritoneal carcinomatosis of colorectal cancer. *Ann Surg Oncol* 2008; **15**: 2426–2432.
9. Glehen O, Kwiatkowski F, Sugarbaker PH, Elias D, Levine EA, De Simone M *et al.* Cytoreductive surgery combined with perioperative intraperitoneal chemotherapy for the management of peritoneal carcinomatosis from colorectal cancer: a multi-institutional study. *J Clin Oncol* 2004; **22**: 3284–3292.
10. Glehen O, Gilly FN, Boutitie F, Bereder JM, Quenet F, Sideris L *et al.*; French Surgical Association. Toward curative treatment of peritoneal carcinomatosis from nonovarian origin by cytoreductive surgery combined with perioperative intraperitoneal chemotherapy: a multi-institutional study of 1290 patients. *Cancer* 2010; **116**: 5608–5618.
11. Cashin PH, Graf W, Nygren P, Mahteme H. Cytoreductive surgery and intraperitoneal chemotherapy for colorectal peritoneal carcinomatosis: prognosis and treatment of recurrences in a cohort study. *Eur J Surg Oncol* 2012; **38**: –515.
12. Elias D, Lefevre JH, Chevalier J, Brouquet A, Marchal F, Classe JM *et al.* Complete cytoreductive surgery plus intraperitoneal chemohyperthermia with oxaliplatin for peritoneal carcinomatosis of colorectal origin. *J Clin Oncol* 2009; **27**: 681–685.
13. Franko J, Ibrahim Z, Gusani NJ, Holtzman MP, Bartlett DL, Zeh HJ III. Cytoreductive surgery and hyperthermic intraperitoneal chemoperfusion *versus* systemic chemotherapy alone for colorectal peritoneal carcinomatosis. *Cancer* 2010; **116**: 3756–3762.
14. Hompes D, D'Hoore A, Van Cutsem E, Fieuws S, Ceelen W, Peeters M *et al.* The treatment of peritoneal carcinomatosis of colorectal cancer with complete cytoreductive surgery and hyperthermic intraperitoneal peroperative chemotherapy (HIPEC) with oxaliplatin: a Belgian multicentre prospective phase II clinical study. *Ann Surg Oncol* 2012; **19**: 2186–2194.
15. Moran B, Cecil T, Chandrakumaran K, Arnold S, Mohamed F, Venkatasubramanian A. The results of cytoreductive surgery and hyperthermic intraperitoneal chemotherapy in 1200 patients with peritoneal malignancy. *Colorectal Dis* 2015; **17**: 772–778.
16. Cavaliere F, Valle M, De Simone M, Deraco M, Rossi CR, Di Filippo F *et al.* 120 peritoneal carcinomatoses from colorectal cancer treated with peritonectomy and intra-abdominal chemohyperthermia: a S.I.T.I.L.O. multicentric study. *In Vivo* 2006; **20**: 747–750.
17. Glehen O, Cotte E, Schreiber V, Sayag-Beaujard AC, Vignal J, Gilly FN. Intraperitoneal chemohyperthermia and attempted cytoreductive surgery in patients with peritoneal carcinomatosis of colorectal origin. *Br J Surg* 2004; **91**: 747–754.
18. Shen P, Hawksworth J, Lovato J, Loggie BW, Geisinger KR, Fleming RA *et al.* Cytoreductive surgery and intraperitoneal hyperthermic chemotherapy with mitomycin C for peritoneal carcinomatosis from nonappendiceal colorectal carcinoma. *Ann Surg Oncol* 2004; **11**: 178–186.
19. Vaira M, Cioppa T, D'Amico S, de Marco G, D'Alessandro M, Fiorentini G *et al.* Treatment of peritoneal carcinomatosis from colonic cancer by cytoreduction, peritonectomy and hyperthermic intraperitoneal chemotherapy (HIPEC). Experience of ten years. *In Vivo* 2010; **24**: 79–84.
20. Frøysnes IS, Larsen SG, Spasojevic M, Dueland S, Flatmark K. Complete cytoreductive surgery and hyperthermic intraperitoneal chemotherapy for colorectal peritoneal metastasis in Norway: prognostic factors and oncologic outcome in a national patient cohort. *J Surg Oncol* 2016; **114**: 222–227.
21. Desantis M, Bernard JL, Casanova V, Cegarra-Escolano M, Benizri E, Rahili AM *et al.* Morbidity, mortality, and oncological outcomes of 401 consecutive cytoreductive procedures with hyperthermic

- intraperitoneal chemotherapy (HIPEC). *Langenbecks Arch Surg* 2015; **400**: 37–48.
- 22 Simkens GA, van Oudheusden TR, Luyer MD, Nienhuijs SW, Nieuwenhuijzen GA, Rutten HJ *et al*. Predictors of severe morbidity after cytoreductive surgery and hyperthermic intraperitoneal chemotherapy for patients with colorectal peritoneal carcinomatosis. *Ann Surg Oncol* 2016; **23**: 833–841.
- 23 Simkens GA, van Oudheusden TR, Braam HJ, Luyer MD, Wiezer MJ, van Ramshorst B *et al*. Treatment-related mortality after cytoreductive surgery and HIPEC in patients with colorectal peritoneal carcinomatosis is underestimated by conventional parameters. *Ann Surg Oncol* 2016; **23**: 99–105.
- 24 Moher D, Liberati A, Tetzlaff J, Altman DG; PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *J Clin Epidemiol* 2009; **62**: 1006–1012.
- 25 Tierney JF, Stewart LA, Ghersi D, Burdett S, Sydes MR. Practical methods for incorporating summary time-to-event data into meta-analysis. *Trials* 2007; **8**: 16.
- 26 Guyot P, Ades AE, Ouwens MJ, Welton NJ. Enhanced secondary analysis of survival data: reconstructing the data from published Kaplan–Meier survival curves. *BMC Med Res Methodol* 2012; **12**: 9.
- 27 Hayden JA, van der Windt DA, Cartwright JL, Côté P, Bombardier C. Assessing bias in studies of prognostic factors. *Ann Intern Med* 2013; **158**: 280–286.
- 28 DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986; **7**: 177–188.
- 29 Int'Hout J, Ioannidis JP, Rovers MM, Goeman JJ. Plea for routinely presenting prediction intervals in meta-analysis. *BMJ Open* 2016; **6**: e010247.
- 30 Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ* 2003; **327**: 557–560.
- 31 Baratti D, Kusamura S, Iusco D, Bonomi S, Grassi A, Virzi S *et al*. Postoperative complications after cytoreductive surgery and hyperthermic intraperitoneal chemotherapy affect long-term outcome of patients with peritoneal metastases from colorectal cancer: a two-center study of 101 patients. *Dis Colon Rectum* 2014; **57**: 858–868.
- 32 Benizri EI, Bernard JL, Rahili A, Benchimol D, Bereder JM. Small bowel involvement is a prognostic factor in colorectal carcinomatosis treated with complete cytoreductive surgery plus hyperthermic intraperitoneal chemotherapy. *World J Surg Oncol* 2012; **10**: 56.
- 33 da Silva RG, Sugarbaker PH. Analysis of prognostic factors in seventy patients having a complete cytoreduction plus perioperative intraperitoneal chemotherapy for carcinomatosis from colorectal cancer. *J Am Coll Surg* 2006; **203**: 878–886.
- 34 Duraj FF, Cashin PH. Cytoreductive surgery and intraperitoneal chemotherapy for colorectal peritoneal and hepatic metastases: a case–control study. *J Gastrointest Oncol* 2013; **4**: 388–396.
- 35 Elias D, Gilly F, Boutitie F, Quenet F, Bereder JM, Mansvelt B *et al*. Peritoneal colorectal carcinomatosis treated with surgery and perioperative intraperitoneal chemotherapy: retrospective analysis of 523 patients from a multicentric French study. *J Clin Oncol* 2010; **28**: 63–68.
- 36 Elias D, Borget I, Farron M, Dromain C, Ducreux M, Goéré D *et al*. Prognostic significance of visible cardiophrenic angle lymph nodes in the presence of peritoneal metastases from colorectal cancers. *Eur J Surg Oncol* 2013; **39**: 1214–1218.
- 37 Elias D, Mariani A, Cloutier AS, Blot F, Goéré D, Dumont F *et al*. Modified selection criteria for complete cytoreductive surgery plus HIPEC based on peritoneal cancer index and small bowel involvement for peritoneal carcinomatosis of colorectal origin. *Eur J Surg Oncol* 2014; **40**: 1467–1473.
- 38 Faron M, Macovei R, Goéré D, Honoré C, Benhaim L, Elias D. Linear relationship of peritoneal cancer index and survival in patients with peritoneal metastases from colorectal cancer. *Ann Surg Oncol* 2016; **23**: 114–119.
- 39 Franko J, Gusani NJ, Holtzman MP, Ahrendt SA, Jones HL, Zeh HJ III *et al*. Multivisceral resection does not affect morbidity and survival after cytoreductive surgery and chemoperfusion for carcinomatosis from colorectal cancer. *Ann Surg Oncol* 2008; **15**: 3065–3072.
- 40 Huang CQ, Yang XJ, Yu Y, Wu HT, Liu Y, Yonemura Y *et al*. Cytoreductive surgery plus hyperthermic intraperitoneal chemotherapy improves survival for patients with peritoneal carcinomatosis from colorectal cancer: a phase II study from a Chinese center. *PLoS One* 2014; **9**: e108509.
- 41 Huang Y, Alzahrani NA, Chua TC, Liauw W, Morris DL. Impacts of peritoneal cancer index on the survival outcomes of patients with colorectal peritoneal carcinomatosis. *Int J Surg* 2016; **32**: 65–70.
- 42 Ihemelandu C, Sugarbaker PH. Management for peritoneal metastasis of colonic origin: role of cytoreductive surgery and perioperative intraperitoneal chemotherapy: a single institution's experience during two decades. *Ann Surg Oncol* 2017; **24**: 898–905.
- 43 Lorimier G, Linot B, Paillocher N, Dupoirion D, Verrière V, Wernert R *et al*. Curative cytoreductive surgery followed by hyperthermic intraperitoneal chemotherapy in patients with peritoneal carcinomatosis and synchronous resectable liver metastases arising from colorectal cancer. *Eur J Surg Oncol* 2017; **43**: 150–158.
- 44 Maggiori L, Goéré D, Viana B, Tzanis D, Dumont F, Honoré C *et al*. Should patients with peritoneal carcinomatosis of colorectal origin with synchronous liver metastases be treated with a curative intent? A case–control study. *Ann Surg* 2013; **258**: 116–121.
- 45 Ng JL, Ong WS, Chia CS, Tan GH, Soo KC, Teo MC. Prognostic relevance of the peritoneal surface disease severity score compared to the peritoneal

- cancer index for colorectal peritoneal carcinomatosis. *Int J Surg Oncol* 2016; **2016**: 2495131.
- 46 Simkens GA, Razenberg LG, Lemmens VE, Rutten HJ, Creemers GJ, de Hingh IH. Histological subtype and systemic metastases strongly influence treatment and survival in patients with synchronous colorectal peritoneal metastases. *Eur J Surg Oncol* 2016; **42**: 794–800.
- 47 Simkens GA, van Oudheusden TR, Braam HJ, Wiezer MJ, Nienhuijs SW, Rutten HJ *et al*. Cytoreductive surgery and HIPEC offers similar outcomes in patients with rectal peritoneal metastases compared to colon cancer patients: a matched case control study. *J Surg Oncol* 2016; **113**:548–553.
- 48 Sluiter NR, de Cuba EM, Kwakman R, Meijerink WJ, Delis-van Diemen PM, Coupé VM *et al*. Versican and vascular endothelial growth factor expression levels in peritoneal metastases from colorectal cancer are associated with survival after cytoreductive surgery and hyperthermic intraperitoneal chemotherapy. *Clin Exp Metastasis* 2016; **33**: 297–307.
- 49 Teo MC, Ching Tan GH, Lim C, Chia CS, Tham CK, Soo KC. Colorectal peritoneal carcinomatosis treated with cytoreductive surgery and hyperthermic intraperitoneal chemotherapy: the experience of a tertiary Asian center. *Asian J Surg* 2015; **38**: 65–73.
- 50 Ung L, Chua TC, David LM. Peritoneal metastases of lower gastrointestinal tract origin: a comparative study of patient outcomes following cytoreduction and intraperitoneal chemotherapy. *J Cancer Res Clin Oncol* 2013; **139**: 1899–1908.
- 51 Winer J, Zenati M, Ramalingam L, Jones H, Zureikat A, Holtzman M *et al*. Impact of aggressive histology and location of primary tumor on the efficacy of surgical therapy for peritoneal carcinomatosis of colorectal origin. *Ann Surg Oncol* 2014; **21**: 1456–1462.
- 52 Varban O, Levine EA, Stewart JH, McCoy TP, Shen P. Outcomes associated with cytoreductive surgery and intraperitoneal hyperthermic chemotherapy in colorectal cancer patients with peritoneal surface disease and hepatic metastases. *Cancer* 2009; **115**: 3427–3436.
- 53 Votanopoulos KI, Swett K, Blackham AU, Ihmelandu C, Shen P, Stewart JH *et al*. Cytoreductive surgery with hyperthermic intraperitoneal chemotherapy in peritoneal carcinomatosis from rectal cancer. *Ann Surg Oncol* 2013; **20**: 1088–1092.
- 54 Rivard JD, McConnell YJ, Temple WJ, Mack LA. Cytoreduction and heated intraperitoneal chemotherapy for colorectal cancer: are we excluding patients who may benefit? *J Surg Oncol* 2014; **109**: 104–109.
- 55 Füzün M, Sökmen S, Terzi C, Canda AE. Cytoreductive approach to peritoneal carcinomatosis originated from colorectal cancer: Turkish experience. *Acta Chir Iugosl* 2006; **53**: 17–21.
- 56 Chen HS, Sheen-Chen SM. Obstruction and perforation in colorectal adenocarcinoma: an analysis of prognosis and current trends. *Surgery* 2000; **127**: 370–376.
- 57 Dohan A, Hoefel C, Soyer P, Jannot AS, Valette PJ, Thivolet A *et al*. Evaluation of the peritoneal carcinomatosis index with CT and MRI. *Br J Surg* 2017; **104**: 1244–1249.
- 58 Flicek K, Ashfaq A, Johnson CD, Menias C, Bagaria S, Wasif N. Correlation of radiologic with surgical peritoneal cancer index scores in patients with pseudomyxoma peritonei and peritoneal carcinomatosis: how well can we predict resectability? *J Gastrointest Surg* 2016; **20**: 307–312.
- 59 Iversen LH, Rasmussen PC, Laurberg S. Value of laparoscopy before cytoreductive surgery and hyperthermic intraperitoneal chemotherapy for peritoneal carcinomatosis. *Br J Surg* 2013; **100**: 285–292.
- 60 Office for National Statistics. *Cancer Registration Statistics, England: 2014*. <https://cy.ons.gov.uk/releases/cancerregistrationstatisticsengland2014> [accessed 3 May 2019].
- 61 Kwakman R, Schrama AM, van Olmen JP, Otten RH, de Lange-de Klerk ES, de Cuba EM *et al*. Clinicopathological parameters in patient selection for cytoreductive surgery and hyperthermic intraperitoneal chemotherapy for colorectal cancer metastases: a meta-analysis. *Ann Surg* 2016; **263**:1102–1111.
- 62 Chua TC, Yan TD, Ng KM, Zhao J, Morris DL. Significance of lymph node metastasis in patients with colorectal cancer peritoneal carcinomatosis. *World J Surg* 2009; **33**: 1488–1494.
- 63 Elias D, Glehen O, Pocard M, Quenet F, Goéré D, Arvieux C *et al*. Association Française de Chirurgie. A comparative study of complete cytoreductive surgery plus intraperitoneal chemotherapy to treat peritoneal dissemination from colon, rectum, small bowel, and nonpseudomyxoma appendix. *Ann Surg* 2010; **251**: 896–901.
- 64 Baumgartner JM, Tobin L, Heavey SF, Kelly KJ, Roeland EJ, Lowy AM. Predictors of progression in high-grade appendiceal or colorectal peritoneal carcinomatosis after cytoreductive surgery and hyperthermic intraperitoneal chemotherapy. *Ann Surg Oncol* 2015; **22**: 1716–1721.
- 65 Schneider MA, Eden J, Pache B, Laminger F, Lopez-Lopez V, Steffen T *et al*. Mutations of RAS/RAF proto-oncogenes impair survival after cytoreductive surgery and PEC for peritoneal metastasis of colorectal origin. *Ann Surg* 2018; **268**: 845–853.

Supporting information

Additional supporting information can be found online in the Supporting Information section at the end of the article



OPEN The transition from primary colorectal cancer to isolated peritoneal malignancy is associated with an increased tumour mutational burden

Sally Hallam, Joanne Stockton, Claire Bryer, Celina Whalley, Valerie Pestinger, Haney Youssef & Andrew D. Beggs 

Colorectal Peritoneal metastases (CPM) develop in 15% of colorectal cancers. Cytoreductive surgery and heated intraperitoneal chemotherapy (CRS & HIPEC) is the current standard of care in selected patients with limited resectable CPM. Despite selection using known prognostic factors survival is varied and morbidity and mortality are relatively high. There is a need to improve patient selection and a paucity of research concerning the biology of isolated CPM. We aimed to determine the biology associated with transition from primary CRC to CPM and of patients with CPM not responding to treatment with CRS & HIPEC, to identify those suitable for treatment with CRS & HIPEC and to identify targets for existing repurposed or novel treatment strategies. A cohort of patients with CPM treated with CRS & HIPEC was recruited and divided according to prognosis. Molecular profiling of the transcriptome (n = 25), epigenome (n = 24) and genome (n = 21) of CPM and matched primary CRC was performed. CPM were characterised by frequent Wnt/ β catenin negative regulator mutations, TET2 mutations, mismatch repair mutations and high tumour mutational burden. Here we show the molecular features associated with CPM development and associated with not responding to CRS & HIPEC. Potential applications include improving patient selection for treatment with CRS & HIPEC and in future research into novel and personalised treatments targeting the molecular features identified here.

Abbreviations

CRC	Colorectal cancer
CPM	Colorectal peritoneal metastasis
CRS & HIPEC	Cytoreductive surgery and heated intraperitoneal chemotherapy
DFS	Disease free survival
DMR	Differentially methylated regions
OS	Overall survival
FFPE	Formalin fixed paraffin embedded

Background

Little is known about the biology of isolated colorectal peritoneal metastasis (CPM), which although a relatively rare phenomenon is one with a high mortality rate¹. Understanding tumour biology may identify which patients with primary colorectal cancer (CRC) are at risk of developing CPM, and which are suitable for treatment with cytoreductive surgery and heated intra-peritoneal chemotherapy (CRS & HIPEC). CRS & HIPEC (usually using an agent such as mitomycin C or more recently, oxaliplatin) aims to achieve macroscopic tumour resection with multiple visceral and peritoneal resections and ablation of microscopic disease. Five-year survival however varies

widely, and morbidity and mortality are relatively high². There is a need therefore to improve patient selection, allowing alternative existing or novel treatment strategies to be used for patients unlikely to respond.

Primary CRC research has identified markers of response to specific treatments, for example KRAS mutation in selection for anti-EGFR mAb therapy³. Gene expression signatures have been developed and are in clinical use for prognostication and therapeutic stratification in breast cancer^{4–7}. Gene expression profiling in primary CRC has identified signatures associated with the development of metastasis⁶. One small study combining a small number of CPM with a larger cohort of appendix adenocarcinoma identified a signature predictive of reduced overall survival (OS) following CRS & HIPEC; these are however two biologically distinct tumours, appendix having significantly improved prognosis⁷.

The dysregulation of methylation is a key step in tumorigenesis CpG island promoter methylation (CIMP) appears to be stable between matched primary CRC and hepatic metastasis suggesting an epigenetic methylation programme is established prior to the development of metastasis^{8–10}. Hypermethylation of KRAS, Wnt modulators, tumour suppressor genes, CIMP and hypomethylation of oncogenes are associated with an unfavourable response to chemotherapy and anti-EGFR antibodies as well as tumour recurrence and reduced OS in primary and metastatic CRC^{11–16}. Chromosomal instability is ubiquitous in cancer, increased copy number alteration, indicative of chromosomal instability is found in metastatic CRC^{17,18}. Lopez-Garcia et al.¹⁹ demonstrated that the evolution of chromosomal instability is depending on cellular tolerance, either via dysregulation of TP53 or via alternate escape mechanisms such as dysfunction of BCL9L regulated caspase signalling.

CRC metastatic drivers are less clearly defined, apart from TP53 which is well characterised as being present in metastatic cancer²⁰. Some studies have found mutations exclusive to metastatic sites^{21,22}, whereas others found similar patterns of mutation between primary and metastasis²³. Studies have examined the somatic mutations in CPM and their prognostic implications. These studies are limited to individual or small panels of mutations routinely tested for in clinical practice with limited evidence to suggest which genes should be included in panel sequencing in CPM. Schneider et al. examined the KRAS and BRAF mutation status of patients with CPM who underwent CRS & HIPEC²⁴. They found mutations of RAS/RAF were associated with reduced OS independent of the use of targeted anti-EGFR treatment²⁴. Sasaki et al. examined the KRAS, BRAF and PIK3CA mutation status of patients with metastatic CRC, with or without CPM²⁵. They found the incidence of BRAF mutation was significantly associated with the presence of CPM but not with prognosis²⁵.

The landscape of metastatic colorectal cancer was studied by the MSK-IMPACT²⁰ group which undertook panel based sequencing of 1134 metastatic colorectal cancers. Of these 39 patients were defined as “peritoneal” malignancy, it is unclear whether these were isolated peritoneal metastasis. Only 14 of these patients had metastectomy. 7 of these had peritonectomy suggesting isolated disease suitable for resection. These tumours were also not studied with matched primary tumour of origin.

There is a need to improve the outcomes for patients with CPM and significant variation in survival despite patient selection for treatment using known prognostic factors. There is a paucity of knowledge concerning CPM tumour biology. Understanding tumour biology will identify patients with primary CRC at risk of developing CPM, those suitable for treatment with CRS & HIPEC or alternative existing and novel treatment strategies. This study aims to determine the landscape of gene expression, methylation, and somatic mutation profile associated with the transition from primary CRC to isolated CPM and determine the association between these and prognosis following CRS & HIPEC in order to identify therapeutic targets.

Methods

Patient cohorts. This study obtained ethical approval from the North West Haydock Research Ethics Committee, (15/NW/0079), project ID (17/283). Participants gave informed consent. All experiments were performed in accordance with relevant guidelines and regulations. Consecutive retrospective patients were recruited from an internally held database of all patients undergoing CRS & HIPEC at Good Hope hospital from 2011 to 2017. Patients with CPM (adenocarcinoma), no extra-abdominal metastasis, a complete resection (CC0) and a peritoneal carcinomatosis index (PCI) of < 12 were eligible for inclusion. The completeness of cytoreduction score describes the degree of macroscopic tumour remaining after CRS and the likelihood of benefit from intraperitoneal chemotherapy²⁶. Patients with no residual tumour score CC0, residual tumour < 0.25 cm, CC1, residual tumour 0.25–2.5 cm CC2. The extent of peritoneal metastasis is described by the PCI score. A PCI of ≥ 12 is poor prognostic factor for patients undergoing CRS & HIPEC²⁷. Patients were divided into two groups. CRS & HIPEC is a long operation associated with a protracted inpatient and high dependency (HDU) or intensive care (ITU) stay an associated mortality of 1–12% and morbidity of 7–63% and a prolonged post-operative recovery^{28–37}. With palliative chemotherapy DFS is 11–13 months and therefore patients post-treatment (CRS & HIPEC) with disease free survival (DFS) < 12 months were

defined as “non-responders”³⁸. Patients undergoing therapy with DFS > 12 months were defined as “responders”. Patients were imaged with CT which was reported by an experienced CPM radiologist, diagnostic laparoscopy was not used, not all patients with recurrence are suitable for iterative CRS & HIPEC and so this is not a standard procedure in their follow up. Adhesions following primary excision and CRS & HIPEC may also preclude accurate assessment of peritoneal recurrence in all areas with laparoscopy. Disease recurrence was determined when confirmed by CT and MDT review.

Demographic, tumour and treatment details were compared between the prognostic cohorts. For continuous variables, the students T-test was applied to normally distributed data and Mann Whitney-U to non-normally distributed data. Categorical variables were compared with the Chi-squared test or Fishers exact test. A *p* value of < 0.05 was considered statistically significant. DFS survival between the responders and non-responders was compared using the Kaplan Meier method. Statistical analysis was performed in IBM SPSS Statistics for Windows, Version 24.0³⁹.

Nucleic acid extraction. DNA and RNA were extracted from histologically confirmed Formalin fixed, paraffin embedded (FFPE) scrolls using the Covaris E220 evolution focused-ultrasonicator and the truTRAC FFPE total NA Kit. All peritoneal metastases samples were taken at the commencement of surgery. Nucleic acid concentration was quantified using the Qubit 3.0 Fluorometer and Qubit RNA / DNA HS (high sensitivity) assay kit. Nucleic acid quality was measured by electrophoresis using the Agilent 2200 TapeStation Nucleic Acid System, Agilent 2200 TapeStation Software A.01.05 and the Agilent High Sensitivity RNA / DNA ScreenTape and reagents.

RNA library preparation, sequencing and bioinformatics. RNA library preparation was performed using the Lexogen Quant Seq 3' mRNA-Seq Library Prep kit. RNA libraries were denatured, diluted, loaded onto a 75-cycle High output flow cell and sequenced using the NextSeq500 at 2.5–5 million r reads⁴⁰.

Quality control, trimming and alignment to the reference genome, (NCBI build 37, hg19) was performed with the Partek Flow genomics suite software package (Partek, St Louis, MI, USA). The gene expression profiles of primary and CPM and responders and non-responders were compared using gene Specific Analysis (GSA) Modelling using Partek flow with a false discovery rate (FDR) of < 0.1. Gene specific enrichment analysis (GSEA) and gene expression pathway analysis was performed using Partek flow, a *p* value of ≤ 0.05 was considered statistically significant.

CMS and CRIS classifications were performed using ‘CMScaller’ (v0.99.1) in the R package, version 2.10.2^{38,41,42}. Fishers exact test was used to compare contingency between primary and CPM and responders and non-responders in IBM SPSS Statistics for Windows, Version 24.0³⁹. A *p* value of < 0.05 was considered significant.

Methylation array and bioinformatics. DNA was treated with sodium bisulphite using the Zymo EZDNA methylation kit, according to manufacturer’s instructions. Degraded FFPE DNA was restored prior to methylation array with the Infinium HD FFPE restore kit, according to manufacturer’s instructions. Methylation array was performed according to the Infinium MethylationEPIC BeadChip Kit manufacturer’s instructions. BeadChips were imaged using the Illumina iScan system. Initial data quality was checked using GenomeStudio Methylation Module Software.

Raw data was loaded into the RStudio version 3.5.0 software using the minifi package. Bioinformatics analysis was performed using the Chip Analysis Methylation Pipeline (ChAMP) R package, version 2.10.2^{43,44}. Probes with signals from less than three functional beads, low confidence with a detection *p* value > 0.01, covering SNPs, non-CpG and those located on the X and Y chromosome were filtered. Beta-mixture quantile normalization (BMIQ) was applied and a singular value decomposition (SVD) performed to identify batch effects. The association between methylation and prognosis was determined using the Bioconductor R package limma and bumphunter functions. Copy number alteration calling was performed using the CHAMP CNA function with a significance threshold of, *p* value < $p < \times 10^{-10}$.

Exome capture, high-throughput sequencing and bioinformatics. DNA was sheared using the Covaris E220 evolution focused-ultrasonicator to produce a uniform 150 bp fragment size. Libraries were prepared using the TruSeq Exome Kit then denatured, diluted, loaded onto a 150-cycle High output flow cell and sequenced using the NextSeq500.

Sequencing reads were assessed using FastQC. Sequences with a Phred score of < 30 were removed giving a base call accuracy of 99.9%. Sequence reads were aligned to the human reference genome, (hg19) using the Burrows–Wheeler Aligner (BWA) package⁴⁵. SAMTools was used to generate chromosomal coordinate-sorted BAM files and Picard was used to remove PCR duplicates⁴⁶. Somatic variants were called from matched tumour/normal samples using Strelka2 in tumour/normal mode⁴⁷. Somatic variants were viewed, filtered and annotated in genomics workbench⁴⁸. Mutations with a MAF of > 1% in known variant databases, (dbSNP and 100,000 genomes) were filtered. Mutations were annotated with information from known variant databases, (dbSNP and 100,000 genomes), PhastCons score and functional consequences. The prognostic groups were compared using Fischer exact test to identify potential candidate driver mutations for non-responders. Somatic mutations were entered into the IntOGen platform for further analysis⁴⁹. The IntOGen-mutation platform incorporates a number of pipelines to identify cancer driver mutations and activated pathways⁴⁹. The OncodriveFM pipeline identifies mutations with a high functional impact using three scoring methods (Sorting Intolerant From Tolerant, (SIFT)⁵⁰, PolyPhen2⁵¹, and Mutation

Assessor scores)^{49,52}, and assesses the likelihood that such mutations are cancer drivers. The OncodriveCLUST pipeline assesses the clustering of mutations to identify relevant activated pathways⁴⁹. MSI assessment was carried out using MSI_classifier_v3 (https://rpubs.com/sigve n/msi_classification_v3).

Ethics approval and consent to participate. North West Haydock Research Ethics Committee, (15/ NW/0079), project ID (17/283).

Results

Patient cohort. From 2011 to 2017 a total of $n = 161$ patients underwent CRS & HIPEC at University Hospitals Birmingham, $n = 88$ patients for metachronous CPM.

Patients were excluded for the following reasons: other primary tumour (appendix, pseudomyxoma peritonei, ovarian) $n = 49$, synchronous colorectal cancer $n = 26$, no primary tumour available $n = 53$ CC2 resection $n = 8$ ²⁶, PCI of ≥ 12 $n = 20$, follow up period of ≤ 12 months $n = 27$, leaving $n = 28$ patients. Complete information regarding the primary CRC pathology and treatment was available for $n = 26$ patients who form the basis of this study.

Each patient had matched normal, primary CRC and CPM samples.

		Responders	Non-responders	<i>p</i> Value
Age, mean \pm SD		58 \pm 13	58 \pm 9	0.97
Gender, male		$n = 7$ (54)	$n = 7$ (54)	0.68
Tumour location	Right	$n = 9$ (69)	$n = 6$ (46)	
	Transverse	$n = 1$ (8)	$n = 0$ (0)	
	Left	$n = 3$ (23)	$n = 7$ (54)	0.33
T stage primary	3	$n = 3$ (23)	$n = 3$ (23)	
	4a	$n = 5$ (38.5)	$n = 7$ (54)	
	4b	$n = 5$ (38.5)	$n = 3$ (23)	0.66
N stage primary	0	$n = 4$ (31)	$n = 1$ (8)	
	1	$n = 7$ (54)	$n = 5$ (38)	
	2	$n = 2$ (15)	$n = 7$ (54)	0.86
DFI months		25 \pm 9	24 \pm 12	0.83
PCI score, median (range)		5 (3–12)	8 (2–12)	0.019
CC score	CC0	$n = 13$ (100)	$n = 13$ (100)	1
	CC1	$n = 0$ (0)	$n = 0$ (0)	
	CC2	$n = 0$ (0)	$n = 0$ (0)	
Follow up, months, median (range)		29 (19–72)	16 (5–55)	0.11
Adjuvant treatment	Yes	$n = 11$ (85)	$n = 12$ (92)	0.38
	No	$n = 2$ (15)	$n = 1$ (8)	
DFS, median (range)		24 (15–72)	6 (2–11)	< 0.0001
OS, median (range)		29 (19–72)	16 (5–55)	0.12

Table 1. Comparison of responders and non-responders to CRS & HIPEC. *N* number value in parenthesis, percentage, *DFI* disease free interval, time from primary CRC to metachronous CPM, *PCI* peritoneal carcinomatosis index, *CC score* completeness of cytoreduction, *DFS* disease free survival, *OS* overall survival. Log rank $p < 0.0001$.

Thirteen patients had a DFS of 24 months (15–72 range) following CRS & HIPEC and formed the ‘responders cohort, thirteen patients had a DFS of 6 months (2–11 range) and formed the ‘non-responders’. There were no significant differences between cohorts in demographics, primary CRC or CPM tumour, treatment or follow up (Table 1). No patients had neoadjuvant therapy for their primary tumour. Three patients (all in the responders group) had poorly differentiated, mucinous adenocarcinoma, one had signet ring adenocarcinoma (in the nonresponders group) and all the others had moderately differentiated adenocarcinoma.

Following nucleic acid extraction all patients had adequate CPM RNA for RNAseq ($n = 13$ responders, $n = 13$ non-responders), $n = 25$ had matched primary CRC samples. For methylation array $n = 24$ patients ($n = 12$ responders, $n = 12$ non-responders) had adequate DNA. As the Infinium methylation array comprises a 32-prep kit, $n = 4$ responders and $n = 4$ non-responders primary tumours were matched to these. For exome sequencing $n = 24$ patients ($n = 12$ responders, $n = 12$ non-responders) had

adequate DNA from both the primary and CPM samples, extraction of DNA from normal tissue resulted in n = 21 samples (n = 9 responders, n = 12 non-responders).

Exome sequencing. Across all six sequencing runs, we obtained a median of 60X coverage (42–166) with a median uniformity of 88% (71–89).

Somatic mutations identified in the primary and matched CPM cohort. In the matched CPM cohort, a total of n = 244,531 somatic SNV's were identified (CPM-primary subtraction) significantly more than found in the matched primary cohort (n = 112,420).

Nine CPM samples, 9/24 (56%) had a high tumour mutational burden $TMB \geq 10 \text{ mut/Mb}$ ⁵³ compared with 7/24 (30%) samples in the matched primary cohort. Mutations were identified in n = 69 of n = 95 known CRC driver genes, n = 51 were shared between the primary and CPM, n = 13 were novel (supplementary table S1)⁵⁴. Of the somatic variants identified in CPM, n = 58,958 (29%) were present in the primary CRC, n = 205,552 variants occurred exclusively in the CPM suggesting a significant accumulation of mutations in the transition to CPM (Fig. 1). OncodriveFM identified n = 265 potential driver genes with high levels of functional mutation (Q-value < 0.05) in the CPM cohort: *FLNB*, *SPTB*, *PPL*, *TP53*, *PDE4DIP*, *RIOK2*, *CDC16*, *NUP98*, *CDC16* and *SVEP1* (supplementary table S2), however these results must be treated with caution due to the bias of the hypermutator phenotype. KEGG pathway analysis of mutations demonstrated enrichment in pathways concerning the immune system, signalling, metabolism and cancer (supplementary table S1). In the CPM group KRAS or BRAF status was not significantly associated with prognosis (chi2 $p = 1.00$).

Clonality analysis with SuperFreq showed significant (Wilcoxon rank $p = 0.007$) differences between the responders and non-responders groups, with a median of 2 clones in the responders group of primary tumours (range 1–4) and 3 clones in the non-responders group (range 2–7). In the peritoneal metastases there were a

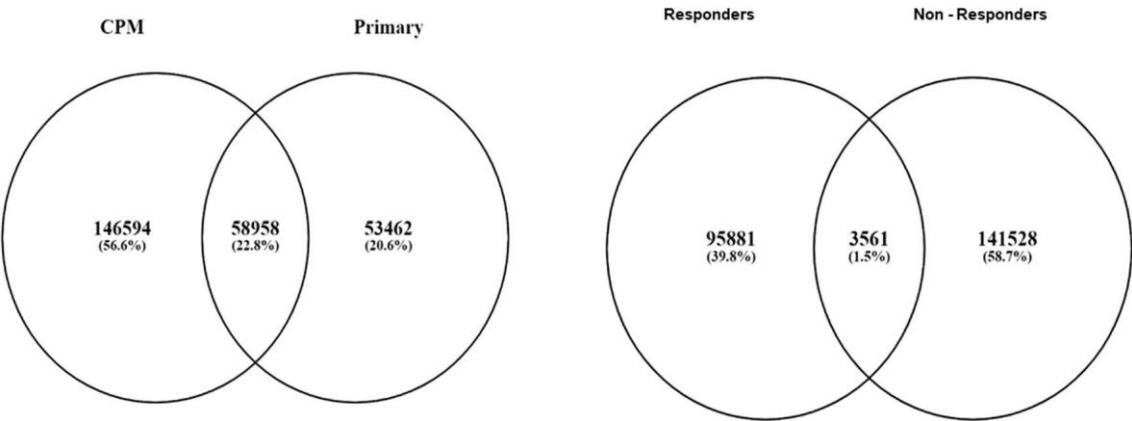


Figure 1. Venn diagrams depicting the frequency of mutations exclusive to and shared between primary CRC and matched CPM and responders and non-responders.

Chr	Position	Reference	Allele	<i>p</i> Value	FDR	Sample (case)	frequency	Sample (control)	frequency	Gene ID
4	93,084,410	C	G	0.007	0.53	62.5		0		<i>FAM13A</i>
18	11,552,313	G	C	0.023	0.53	50		0		<i>PIEZO2</i>

Table 2. Potential candidate variants, non-responders to CRS & HIPEC. CPM identified through Fisher exact test, genomics workbench (Chr, chromosome, FDR, false discovery rate).

median of 3 clones in both the responders (range 1–4) and non-responders (range 2–5) groups. Of note, in the non-responders group during clonal expansion, the dominant clone in the peritoneal metastasis group arose de-novo rather than being a prior clone that existed in the primary tumour (Supplementary Fig. 1, S1e primary tumours, 9/21 were MSI (47.4%) and 10/21 were MSS (52.6%) whereas in the isolated peritoneal metastasis group, 4/21 (19.0%) were MSS and 17/21 MSI (81.0%) Demonstrating that there was a significantly higher rate of MSI in the isolated peritoneal metastasis group ($p < 0.05$, Chi2).

Non-responders had a higher frequency of somatic mutations: 60% of all mutations in CPM cohort vs. 40%. Non-responders more commonly had a high tumour mutational burden, TMB ≥ 10 mut/Mb⁵³, 56% vs. 44%. Of the somatic mutations identified in non-responders, n = 35,461 (30%) were present in responders, n = 145,089

variants occurred exclusively in non-responders, suggesting a high tumour mutational burden was associated with non-response to CRS & HIPEC (Fig. 1). Mutational signature analysis of the MSI tumours demonstrated a predominance of signature 5 (associated with mutational “clock” effects), signature 26 (associated with defective mismatch repair) and signature 20 (associated with defective mismatch repair).

Comparison of somatic mutations in responders and non-responders identified two potential candidate genes to identify non-responders, *FAM13A* and *PIEZO2* (Fishers exact $p < 0.05$, FDR = 0.53) (Table 2).

Differential gene expression. *Differential gene expression between primary CRC and matched CPM.* Primary CRC and matched CPM showed differential expression of n = 65 genes with an FDR < 0.1. (Fig. 2) Sixteen genes showed significantly decreased expression in CPM compared with primary CRC (Table 3). Forty-nine genes showed significantly increased expression in CPM compared with primary CRC (Table 3). A KEGG pathway analysis was performed to identify the enriched biological functions among the differentially expressed genes (Supplementary Table 1). The expression of *FABP6*, an intercellular bile acid transporter, was decreased 34.30-fold in CPM. *OLFM4* is a target of the Wnt/ β -catenin pathway, its expression was reduced 3.77-fold in CPM. *DCN* and *PTEN* are able to initiate a number of signalling pathways including *ERK* and *EGFR* leading to growth suppression, their expression was increased 3.3-fold and 3.25 fold in CPM, this was unexpected and in contrast to the literature⁵⁵. *NF- κ BIA* expression was increased 3.24-fold in CPM, its upregulation may reflect increased *NF- κ B* activity in the development of CPM⁵⁶.

Gene specific enrichment analysis (GSEA) results are presented in supplementary table 5 We identified 848 upregulated gene ontology categories in CPM and 14 upregulated gene pathways. which may contribute to the pathogenesis of CPM: the mTOR pathway as well as immune pathways including the intestinal immune network for IgA production, Leukocyte transendothelial migration and the actin cytoskeleton pathway.

Differential gene expression between non-responders and responders to CRS & HIPEC. One hundred and forty-nine genes showed increased expression in non-responders (Fig. 3). Five genes showed decreased expression in non-responders, however none had a fold change ≥ 1.5 suggesting minimal difference in expression between the responders and non-responders (Supplementary Table 2). KEGG pathway analysis demonstrated enrichment in endocytosis, metabolism, phagocytosis, cell movement and architecture, bacterial and viral cell infection, transcription and the expression of genes controlling apoptosis, cell cycle, oxidative stress resistance and longevity (Table 3). The expression of *CEACAM1*, a member of the carcinoembryonic antigen (*CEA*) immunoglobulin family, was increased 8.27-fold in non-responders⁵⁷.

AXIN1 encodes a cytoplasmic protein which forms the β -Catenin destruction complex, a negative regulator of the WNT signalling pathway⁵⁸. *AXIN1* expression was increased 5.42-fold in non-responders⁵⁹.

Gene specific enrichment analysis (GSEA) results are presented in supplementary table 6. We identified 591 upregulated gene ontology categories in CPM and 15 upregulated gene pathways. which may contribute to the pathogenesis of CPM: Endocytosis, the adherens junction pathway and immune pathways such as those regulating the bacterial invasion of epithelial cells.

Amongst the n = 51 primary CRC and CPM samples n = 29 were representative of each CMS subtype, the remaining n = 22 samples did not have a consistent pattern (Fig. 4). Comparison of the CMS subtypes in primary and CPM and prognostic groups revealed an apparent transition from primary CRC to CPM. No primary CRC samples were classified as CMS4 (mesenchymal subtype characterized by prominent transforming growth factor activation, stromal invasion and angiogenesis) compared to 31% of CPM ($p = 0.085$). Secondly, non-responders were more commonly CMS4, 46% vs. 15% ($p = 0.005$, Table 4).

Methylation. *Differential methylation between primary CRC and matched CPM.* Thirty-two samples in total were hybridised successfully to the Illumina HumanMethylation EPIC microarrays. DMPs were called between the primary CRC and CPM. The top ranked differentially methylated probe was cg04146982, BF 34.5, adjusted p value 5.67×10^{-16} (chr8:144,943,810–144,943,810, hg19 coordinates), which tags a CpG dinucleotide 3651 bp upstream of the transcription start site of gene Epiplakin 1, (*EPPK1*)⁶⁰. *EPPK1* is part of the Plakin family an important component of the cell cytoskeleton⁶¹. The other DMP was cg12209861, BF 7.1, adjusted p value 0.059 (chr4:37,459,078–37,459,078, hg19 coordinates), 3526 bp upstream of the transcription start site of gene Chromosome 4 Open Reading Frame 19, (*C4orf19*). DMRs were called between primary CRC and CPM via the dmrLasso function of the CHAMP pipeline (Supplementary Table 3). The top 10 most DMRs were in the region of *IGF2*, *ZNF461*, *RASGFR1*, *CELF4*, *ZSCAN18*, *EDNRB*, *ZBED9*, *VTRNA2-1*, *ZNF256* and *EGFLAM*. KEGG pathway analysis did not reveal any significantly enriched pathways.

Comparison of CNA between primary and CPM via methylation arrays did not identify any significant differences in CNA between primary and CPM at a stringent p value of $< \times 10^{-10}$ however a number of CNA were identified at a lower significance

threshold, $p = 2.78 \times 10^{-07}$ (Supplementary Table 4). Genes showing CNA gains of known significance in patients with CPM included; *TRIM3*, 5, 6, 21 and 22, *MT1A*, 2A, 3, 4 encode proteins of the metallothionein family.

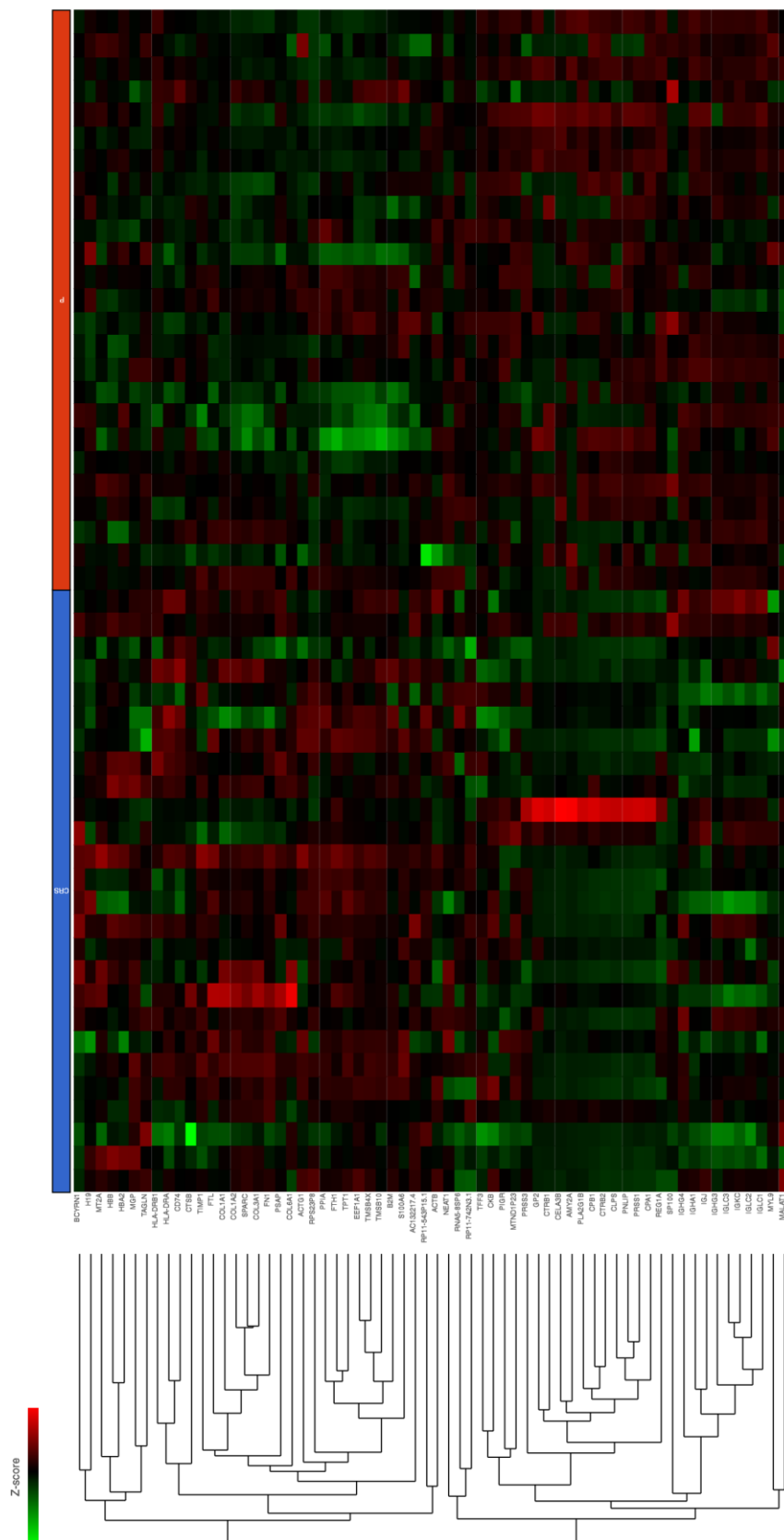


Figure 2. Heatmap of differential gene expression in 100 highest genes ranked by variance between primary CRC (P, red) and colorectal peritoneal metastasis (CRS, blue). Sample type is indicated at the X axis of the heatmap with individual IDs of each patient are below the indicators of primary or CRS sample. Gene expression as indicated by the Z-score is displayed as colour ranging from green to black to red as shown in the legend. Created in Partek Flow.

Rank	Gene name	Function	Fold change	FDR <i>p</i> value
Reduced expression CPM samples vs. primary CRC				
1	FABP6	Intracellular bile acid transporter	− 34.30	1.74×10^{-06}
2	DEFA6	Cytotoxic peptide involved in host intestine defence	− 8.15	8.55×10^{-06}
3	DMBT1	Tumour suppressor	− 6.06	2.43×10^{-04}
4	TTC38	Protein coding gene	− 4.56	5.80×10^{-05}
5	OLFM4	Wnt/ β -catenin pathway target	− 3.77	1.01×10^{-04}
6	IGHA1	Immune receptor	− 3.66	4.23×10^{-05}
7	CES2	Intestinal enzyme controlling drug clearance	− 3.20	6.84×10^{-05}
8	NDUFS6	Enzyme in electron transport chain of mitochondria	− 2.70	7.74×10^{-05}
9	P2RY11	G-protein coupled receptor	− 2.53	6.37×10^{-04}
10	MUC2	Encodes a mucinous intestinal coating	− 2.34	7.22×10^{-04}
Increased expression CPM samples vs. primary CRC				
1	CD53	Tetraspanin	7.29	5.87×10^{-05}
2	CYR61	Extracellular signalling protein	4.24	3.12×10^{-04}
3	CXCL12	G-protein coupled receptor	3.64	9.25×10^{-04}
4	NR2F1	Nuclear hormone receptor and transcriptional regulator	3.53	7.09×10^{-04}
5	CTGF	Connective tissue growth factor	3.49	1.55×10^{-04}
6	CSTB	Cystatin	3.41	6.13×10^{-04}
7	TSC22D3	Anti-inflammatory protein glucocorticoid (GC)-induced leucine zipper	3.36	3.94×10^{-04}
8	DCN	Tumour suppressor gene	3.30	6.19×10^{-05}
9	PTEN	Tumour suppressor gene	3.25	9.28×10^{-04}
10	NF- κ BIA	Inhibits the <i>NF-κB</i> transcription factor	3.24	1.06×10^{-04}

Table 3. The top 10 genes with significantly altered expression (FDR < 0.1) in CPM samples compared with primary CRC samples.

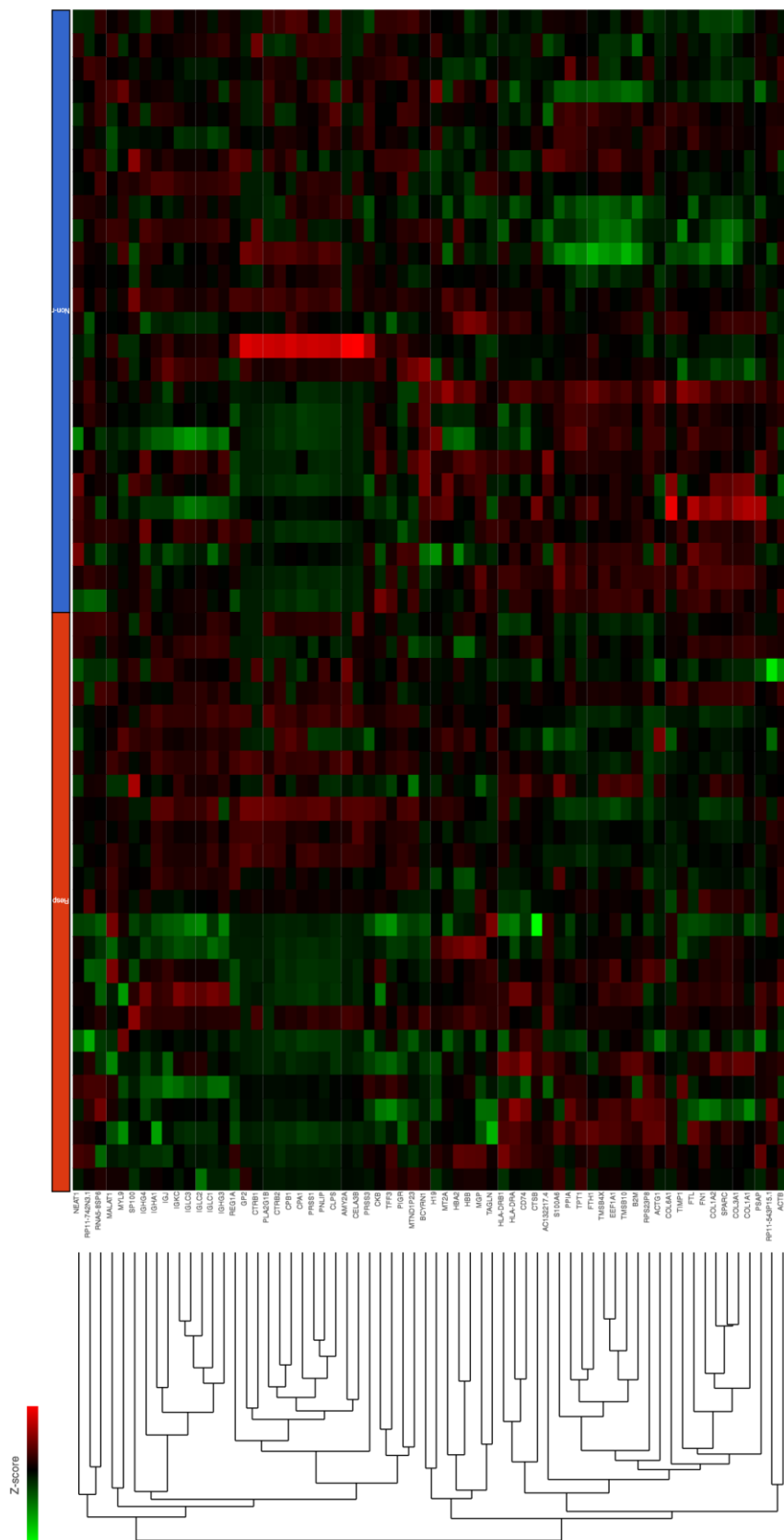
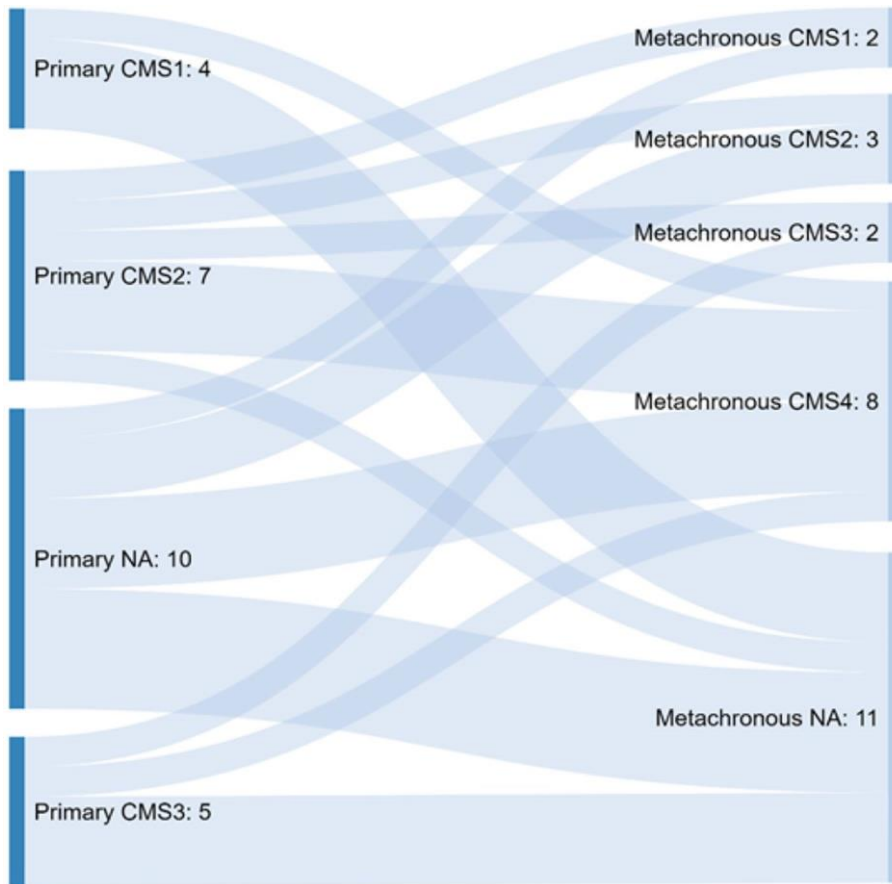


Figure 3. Heatmap differential gene expression of top 100 genes as ranked by variance between responders (blue) and non-responders (red) Sample type is indicated at the transverse border of the heatmap with individual genes on the longitudinal border. Gene expression is displayed as colour ranging from green to red as shown in the legend. Created in Partek Flow.



	Non-consensus	CMS1	CMS2	CMS3	CMS4	Total
Primary	10 (40)	4 (16)	6 (24)	5 (20)	0 (0)	25
CPM	12 (46)	1 (4)	3 (12)	2 (7)	8 (31)	26

Figure 4. Sankey diagram depicting the transition in consensus molecular subtypes (CMS) from primary to CPM. CMS classifications were performed using ‘CMScaller’ (v0.99.1) in the R /Bioconductor statistics package. Classifications include CMS1 to CMS4, non-consensus samples do not have a consistent pattern of subtype label association. Primary CRC samples, classification and number are shown to the left of the diagram with CPM samples, classification and number to the right of the diagram. Fishers exact p value 0.085, values in parenthesis percentages.

	Non-consensus	CMS1	CMS2	CMS3	CMS4		Total
Responders	10 (77)	0 (0)	0 (0)	1 (8)	2 (15)		13
Non-responders	2 (15)	1 (8)	3 (23)	1 (8)	6 (46)		13

Table 4. CMS classification responders vs. non-responders to CRS & HIPEC. CMS Fishers exact p value 0.005, CRIS Fischer’s exact p value 0.148, values in parenthesis percentages.

Differential methylation between non-responders and responders to CRS & HIPEC. The top ranked differentially methylated probe was cg07951355, BF = 6, (chr1:40,123,717) which tags an intergenic region 1076 bp before gene *NT5C1A*. Cg25909064,

BF 4 adjusted *p* value 0.47 (chr11:120,081,487–120,082,345) which tags an intron of gene *OAF* and cg12977942, BF 4 adjusted *p* value 0.47 (chr5:92,839,309–92,839,309) which tags an intron of gene *NR2F1-AS1*⁶⁰. Six significant DMRs (Supplementary Table 3) were identified in the regions of *NKX6-2*, *CHFR*, *GATA3*, *IRX5*, *HCK* and *BC019904*. KEGG pathway analysis did not reveal any significantly enriched pathways.

Comparison of CNA between the CPM prognostic groups identified recurrent gene losses at chromosomes 3, 4, 14, 15, 17 and 19 (Supplementary Table 4). CNA losses clustered in the RAS-MAPK-ERK signalling pathway suggesting dysregulation in non-responders.

Comparison of CNA between the CPM prognostic groups identified *n* = 19 gene gains at chromosomes 9, 10 and 11. Genes showing CNA gains in non-responders included: *SIT1*, *RNF38*, *MELK*, *PAX5*, *SHB*, *ZEB1*, *DEAF1*, *ANTXR*, *EPS8L2* and *PIDD1*.

Discussion

This study determined the gene expression, CNA, methylation and somatic mutation profile of primary CRC and matched isolated CPM to determine whether there were changes associated with the development of CPM or predicting prognosis for patients with CPM. To our knowledge, this is the first such analysis in a cohort of patients with isolated CPM suitable for treatment with CRS & HIPEC. The MSKCC cohort of metastatic cancer²⁰ had a diverse range of metastatic cancer, none of whom overlapped with the type we have studied, which is isolated colorectal peritoneal metastasis, with matched primary samples, suitable for cytoreduction.

Within this study responders and non-responders to CRS & HIPEC were well matched by demographics, tumour stage, treatment and follow up. PCI varied between groups with responders having a median PCI of 5 (3–12) and non-responders a median PCI of 8 (2–12). A PCI of greater than 12 is associated with reduced survival following CRS & HIPEC, no significant difference is consistently found at PCI levels below this²⁷.

Comparison of patients with primary CRC and metachronous CPM identified biological changes associated with the transition from primary CRC to CPM. Hypermethylation, CNA and hypermutation resulted in the inactivation of tumour suppressors and oncogene activation in CPM, (TP53, VTRA2-1, TRIM proteins). These changes suggest a rapid rate of tumour growth unchecked by tumour suppressor or apoptotic mechanisms.

Increased MAPK and Wnt/ β -catenin pathway activation was noted in CPM. Gene expression of negative regulators of the Wnt pathway was reduced, (OLFM4, DEAF6), negative Wnt regulators contained somatic mutations, (APC, RNF43, FAM123B and TSC1), and the MAPK marker, RASFGFR1 was hypermethylated suggesting persistent activation of MAPK and Wnt pathways. Multiple mutations of negative Wnt signalling regulators make this an attractive therapeutic target. Porcupine inhibitors mediate the palmitoylation of Wnt ligands, blocking Wnt signalling. The porcupine inhibitor LGK974 inhibits the upstream negative Wnt regulator mutant RNF43 and is a potential therapeutic target in CPM⁶².

CPM contained a high proportion of MSH6 somatic mutations suggesting deficiency in the mismatch repair pathway and MSI. MSH6 mutations are commonly found in isolated peritoneal metastasis⁵⁹. As expected for tumours with mismatch repair deficiency both the primary CRC and CPM cohort had a high tumour mutational burden, crucially this suggests they may have a good response to treatment with immune checkpoint inhibitors such as pembrolizumab⁶³, a new therapeutic avenue for these difficult to treat patients. The frequency of hypermutation seen in our study (48%) was considerably higher than that observed for both the MSKCC metastatic disease cohort (5%) and the TCGA Colorectal⁶⁴ cohort (10%). The expression of genes regulating innate immunity however was downregulated, (DEFA6, DMBT1, MUC2) or altered via somatic mutations, (HLA-A antigen) suggesting immune evasion in the transition to CPM which may reduce the likelihood of successful PD-1 therapy.

The expression of genes suppressing invasion, migration and EMT was downregulated or hypermethylated, (MUC2, MMP26, ILK, FLNB, SPTB, PPL, and SVEP1) and those triggering these processes upregulated, (CYR61, CXCL12, CTGF, and CSTB). These changes suggest a mechanism by which CPM cells metastasise from the primary CRC. In keeping with changes in EMT regulators there appeared to be a transition in CMS subtypes towards CMS4 from primary CRC to CPM. The CMS4 subtype is an interesting therapeutic target, *TGF β* signalling inhibitors and targeted immunotherapies have been trialled with success in pre-clinical models to block cross talk between the tumour microenvironment and halt disease progression of stromal rich CMS4 CRC^{65,66}.

Methylation appeared to be dysregulated in CPM with a bias towards a hypermethylator phenotype caused by somatic mutation of the TET2 tumour suppressor and CDH7 chromatin regulator. Active DNA demethylation by TET enzymes is an important tumour suppressor mechanism in a variety of cancers^{67–69}. Downregulation of CES2, a gene known to activate the prodrug irinotecan, a chemotherapy used as part of the FOLFIRI regimen in the UK in the adjuvant treatment of primary CRC and CPM was seen in this cohort. Resistance to the treatment of primary CRC may in part explain the development of CPM.

CEACAM1 expression correlates with metastasis and reduced survival in CRC and was upregulated in this cohort of patients⁷⁰. Novel therapies in the form of CEA TCB IgG-based T-cell bispecific antibodies (Cibisatamab) may therefore be of benefit⁷¹. Additionally there was a downregulation of gene expression of negative regulators of the Wnt pathway, (*AXIN1*) and somatic mutations of key Wnt regulators, (*FAM13A*) and hypermethylation of MAPK and TGF- β pathway markers, (*RAB8A*, *RAB34*, *FGF5* and *BMP3*) suggesting persistent activation of MAPK, TGF- β and Wnt in non-responders to CRS & HIPEC.

A recent randomised controlled trial has called into question the use of HIPEC in CPM, PRODIGE-7 treated patients with CPM with CRS & HIPEC or CRS alone in addition to systemic chemotherapy. PRODIGE-7 suggests no added benefit from HIPEC however this study was not powered to stratify the impact of HIPEC according to PCI score, on subgroup analysis patients with a PCI of 11–15 had significantly improved median survival with the addition of HIPEC 41.6 months vs. 32.7 months *p* value 0.0209⁷².

A relative weakness of this study is the small cohort of patients, the biological changes identified here form a starting point in identifying the tumour biology associated with the development of CPM and predicting nonresponders to CRS & HIPEC. However, we have identified multiple potential targets for therapy, along with the important finding that CPM appears to be a hypermutated, hypermethylated, immune evasive cancer which allows it to be potentially targeted by emerging novel therapeutics. Our study findings have implications for the recent addition of oxaliplatin to HIPEC, as the FOXTROT study of neoadjuvant therapy in colorectal cancer showed that oxaliplatin has no effect in dMMR tumours.

Conclusions

Patients with colorectal peritoneal metastasis (CPM) secondary to colorectal cancer have limited survival with the best available treatments. Despite selection for treatment using known prognostic factors survival varies widely and can be difficult to predict. There is a paucity of knowledge concerning the biology of CPM, it is likely that there are additional biological markers of response to currently available as well as novel or re-purposed alternative treatments. Here we have comprehensively profiled a cohort of patients with isolated CPM and identified a number of therapeutically targetable alterations including mutations in Wnt/ β catenin regulators (via Porcupine inhibitors), the mismatch repair pathway (via PD-1/CTLA-4 immunotherapy) and methylation regulators. We suggest that these are urgently investigated in a larger cohort with the development of pre-clinical models as, in particular, the finding that these patients may be sensitive to immunotherapy may radically change the therapy options available for this difficult to treat group of patients.

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Received: 20 May 2020; Accepted: 13 October 2020

Published online: 03 November 2020

References

- Mimezami, R. *et al.* Cytoreductive surgery in combination with hyperthermic intraperitoneal chemotherapy improves survival in patients with colorectal peritoneal metastases compared with systemic chemotherapy alone. *Br. J. Cancer* **111**, 1500–1508. <https://doi.org/10.1038/bjc.2014.419> (2014).
- Koumpa, F. S. *et al.* Colorectal peritoneal metastases: a systematic review of current and emerging trends in clinical and translational research. *Gastroenterol. Res. Pract.* **2019**, 5180895. <https://doi.org/10.1155/2019/5180895> (2019).
- Lièvre, A. *et al.* KRAS mutation status is predictive of response to cetuximab therapy in colorectal cancer. *Cancer Res.* **66**, 3992–3995. <https://doi.org/10.1158/0008-5472.CAN-06-0191> (2006).
- Cardoso, F. *et al.* 70-gene signature as an aid to treatment decisions in early-stage breast cancer. *N. Engl. J. Med.* **375**, 717–729. <https://doi.org/10.1056/NEJMoA1602253> (2016).
- van de Vijver, M. J. *et al.* A gene-expression signature as a predictor of survival in breast cancer. *N. Engl. J. Med.* **347**, 1999–2009. <https://doi.org/10.1056/NEJMoA021967> (2002).
- Friederichs, J. *et al.* Gene expression profiles of different clinical stages of colorectal carcinoma: toward a molecular genetic understanding of tumor progression. *Int. J. Colorectal Dis.* **20**, 391–402. <https://doi.org/10.1007/s00384-004-0722-1> (2005).
- Levine, E. A. *et al.* Gene expression profiling of peritoneal metastases from appendiceal and colon cancer demonstrates unique biologic signatures and predicts patient outcomes. *J. Am. Coll. Surg.* **214**, 599–606; discussion 606–597. <https://doi.org/10.1016/j.jamcollsurg.2011.12.028> (2012).
- Ju, H. X. *et al.* Distinct profiles of epigenetic evolution between colorectal cancers with and without metastasis. *Am. J. Pathol.* **178**, 1835–1846. <https://doi.org/10.1016/j.ajpat.2010.12.045> (2011).
- Konishi, K. *et al.* DNA methylation profiles of primary colorectal carcinoma and matched liver metastasis. *PLoS ONE* **6**, e27889. <https://doi.org/10.1371/journal.pone.0027889> (2011).

- Murata, A. *et al.* Methylation levels of LINE-1 in primary lesion and matched metastatic lesions of colorectal cancer. *Br. J. Cancer* **109**, 408–415. <https://doi.org/10.1038/bjc.2013.289> (2013).
- Miranda, E. *et al.* Genetic and epigenetic changes in primary metastatic and nonmetastatic colorectal cancer. *Br. J. Cancer* **95**, 1101–1107. <https://doi.org/10.1038/sj.bjc.6603337> (2006).
- Sharrard, R. M., Royds, J. A., Rogers, S. & Shorthouse, A. J. Patterns of methylation of the c-myc gene in human colorectal cancer progression. *Br. J. Cancer* **65**, 667–672 (1992).
- Siemens, H. *et al.* Detection of miR-34a promoter methylation in combination with elevated expression of c-Met and β -catenin predicts distant metastasis of colon cancer. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* **19**, 710–720. <https://doi.org/10.1158/10780432.CCR-12-1703> (2013).
- Tang, D. *et al.* Diagnostic and prognostic value of the methylation status of secreted frizzled-related protein 2 in colorectal cancer. *Clin. Invest. Med.* **34**, E88–95 (2011).
- Zhang, X. *et al.* CpG island methylator phenotype is associated with the efficacy of sequential oxaliplatin- and irinotecan-based chemotherapy and EGFR-related gene mutation in Japanese patients with metastatic colorectal cancer. *Int. J. Clin. Oncol.* **21**, 1091–1101. <https://doi.org/10.1007/s10147-016-1017-6> (2016).
- Kim, S. H. *et al.* CpG island methylator phenotype and methylation of Wnt pathway genes together predict survival in patients with colorectal cancer. *Yonsei Med. J.* **59**, 588–594. <https://doi.org/10.3349/ymj.2018.59.5.588> (2018).
- Lee, S. Y. *et al.* Comparative genomic analysis of primary and synchronous metastatic colorectal cancers. *PLoS ONE* **9**, e90459. <https://doi.org/10.1371/journal.pone.0090459> (2014).
- Turajlic, S. *et al.* Deterministic evolutionary trajectories influence primary tumor growth: TRACERx renal. *Cell* **173**, 595–610 e511. <https://doi.org/10.1016/j.cell.2018.03.043> (2018).
- Lopez-Garcia, C. *et al.* BCL9L dysfunction impairs caspase-2 expression permitting aneuploidy tolerance in colorectal cancer. *Cancer Cell* **31**, 79–93. <https://doi.org/10.1016/j.ccell.2016.11.001> (2017).
- Yaeger, R. *et al.* Clinical sequencing defines the genomic landscape of metastatic colorectal cancer. *Cancer Cell* **33**, 125–136.e123. <https://doi.org/10.1016/j.ccell.2017.12.004> (2018).
- Goryca, K. *et al.* Exome scale map of genetic alterations promoting metastasis in colorectal cancer. *BMC Genet.* **19**, 85. <https://doi.org/10.1186/s12863-018-0673-0> (2018).
- Huang, D. *et al.* Mutations of key driver genes in colorectal cancer progression and metastasis. *Cancer Metastasis Rev.* **37**, 173–187. <https://doi.org/10.1007/s10555-017-9726-5> (2018).
- Mlecnik, B. *et al.* The tumor microenvironment and immunoscore are critical determinants of dissemination to distant metastasis. *Sci. Transl. Med.* **8**, 327ra326. <https://doi.org/10.1126/scitranslmed.aad6352> (2016).
- Schneider, M. A. *et al.* Mutations of RAS/RAF proto-oncogenes impair survival after cytoreductive surgery and HIPEC for peritoneal metastasis of colorectal origin. *Ann. Surg.* **268**, 845–853. <https://doi.org/10.1097/SLA.0000000000002899> (2018).
- Sasaki, Y. *et al.* Value of KRAS, BRAF, and PIK3CA mutations and survival benefit from systemic chemotherapy in colorectal peritoneal carcinomatosis. *Asian Pac. J. Cancer Prev.* **17**, 539–543 (2016).
- Sugarbaker, P. H. Successful management of microscopic residual disease in large bowel cancer. *Cancer Chemother. Pharmacol.* **43**(Suppl), S15–25 (1999).
- Hallam, S., Tyler, R., Price, M., Beggs, A. & Youssef, H. Meta-analysis of prognostic factors for patients with colorectal peritoneal metastasis undergoing cytoreductive surgery and heated intraperitoneal chemotherapy. *BJS Open* **3**, 585–594. <https://doi.org/10.1002/bjs5.50179> (2019).
- Cashin, P. H., Graf, W., Nygren, P. & Mahteme, H. Cytoreductive surgery and intraperitoneal chemotherapy for colorectal peritoneal carcinomatosis: prognosis and treatment of recurrences in a cohort study. *Eur. J. Surg. Oncol.* **38**, 509–515. <https://doi.org/10.1016/j.ejso.2012.03.001> (2012).
- Glehen, O. *et al.* Intraperitoneal chemohyperthermia and attempted cytoreductive surgery in patients with peritoneal carcinomatosis of colorectal origin. *Br. J. Surg.* **91**, 747–754. <https://doi.org/10.1002/bjs.4473> (2004).
- Shen, P. *et al.* Cytoreductive surgery and intraperitoneal hyperthermic chemotherapy with mitomycin C for peritoneal carcinomatosis from nonappendiceal colorectal carcinoma. *Ann. Surg. Oncol.* **11**, 178–186 (2004).
- Cavaliere, F. *et al.* 120 peritoneal carcinomatoses from colorectal cancer treated with peritonectomy and intra-abdominal chemohyperthermia: a SITILO multicentric study. *Vivo* **20**, 747–750 (2006).
- Vaira, M. *et al.* Treatment of peritoneal carcinomatosis from colonic cancer by cytoreduction, peritonectomy and hyperthermic intraperitoneal chemotherapy (HIPEC). Experience of ten years. *In Vivo* **24**, 79–84 (2010).
- Hompes, D. *et al.* The treatment of peritoneal carcinomatosis of colorectal cancer with complete cytoreductive surgery and hyperthermic intraperitoneal peroperative chemotherapy (HIPEC) with oxaliplatin: a Belgian multicentre prospective phase II clinical study. *Ann. Surg. Oncol.* **19**, 2186–2194. <https://doi.org/10.1245/s10434-012-2264-z> (2012).

Froynes, I. S., Larsen, S. G., Spasojevic, M., Dueland, S. & Flatmark, K. Complete cytoreductive surgery and hyperthermic intraperitoneal chemotherapy for colorectal peritoneal metastasis in Norway: prognostic factors and oncologic outcome in a national patient cohort. *J. Surg. Oncol.* **114**, 222–227. <https://doi.org/10.1002/jso.24290> (2016).

Desantis, M. *et al.* Morbidity, mortality, and oncological outcomes of 401 consecutive cytoreductive procedures with hyperthermic intraperitoneal chemotherapy (HIPEC). *Langenbecks Arch. Surg.* **400**, 37–48. <https://doi.org/10.1007/s00423-014-1253-z> (2015).

Simkens, G. A. *et al.* Predictors of severe morbidity after cytoreductive surgery and hyperthermic intraperitoneal chemotherapy for patients with colorectal peritoneal carcinomatosis. *Ann. Surg. Oncol.* **23**, 833–841. <https://doi.org/10.1245/s10434-015-4892-6> (2016).

Simkens, G. A. *et al.* Treatment-related mortality after cytoreductive surgery and HIPEC in patients with colorectal peritoneal carcinomatosis is underestimated by conventional parameters. *Ann. Surg. Oncol.* **23**, 99–105. <https://doi.org/10.1245/s10434-0154699-5> (2016).

Team, R. C. R: A Language and Environment for Statistical Computing. *R Foundation for Statistical Computing* (2020).

Corp, I. *IBM SPSS Statistics for Windows, Version 24.0*, Released 2016.

Liu, Y., Zhou, J. & White, K. P. RNA-seq differential expression studies: more sequence or more replication?. *Bioinformatics* **30**, 301–304. <https://doi.org/10.1093/bioinformatics/btt688> (2014).

Rinaudo, P., Boudah, S., Junot, C. & Thévenot, E. A. biosigner: a new method for the discovery of significant molecular signatures from omics data. *Front. Mol. Biosci.* **3**, 26. <https://doi.org/10.3389/fmolb.2016.00026> (2016).

Eide, P. W., Bruun, J., Lothe, R. A. & Sveen, A. CMScaller: an R package for consensus molecular subtyping of colorectal cancer pre-clinical models. *Sci. Rep.* **7**, 16618. <https://doi.org/10.1038/s41598-017-16747-x> (2017).

Morris, T. J. *et al.* ChAMP: 450k chip analysis methylation pipeline. *Bioinformatics* **30**, 428–430. <https://doi.org/10.1093/bioinformatics/btt684> (2014).

Aryee, M. J. *et al.* Minfi: a flexible and comprehensive bioconductor package for the analysis of infinium DNA methylation microarrays. *Bioinformatics* **30**, 1363–1369. <https://doi.org/10.1093/bioinformatics/btu049> (2014).

Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* **26**, 589–595. <https://doi.org/10.1093/bioinformatics/btp698> (2010).

Li, H. *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352> (2009).

Saunders, C. T. *et al.* Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinformatics* **28**, 1811–1817. <https://doi.org/10.1093/bioinformatics/bts271> (2012).

Qiagen. *CLC Genomics Workbench 11.0*, <https://www.qiagenbioinformatics.com>.

Gonzalez-Perez, A. *et al.* IntOGen-mutations identifies cancer drivers across tumor types. *Nat. Methods* **10**, 1081–1082. <https://doi.org/10.1038/nmeth.2642> (2013).

Ng, P. C. & Henikoff, S. SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Res.* **31**, 3812–3814 (2003).

Adzhubei, I., Jordan, D. M. & Sunyaev, S. R. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr. Protoc. Hum. Genet.* <https://doi.org/10.1002/0471142905.hg0720s76> (2013).

Reva, B., Antipin, Y. & Sander, C. Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic Acids Res.* **39**, e118. <https://doi.org/10.1093/nar/gkr407> (2011).

Büttner, R. *et al.* Implementing TMB measurement in clinical practice: considerations on assay requirements. *ESMO Open* **4**, e000442. <https://doi.org/10.1136/esmoopen-2018-000442> (2019).

IntOGen. *IntOGen Colorectal cancer driver genes*, <https://www.intogen.org/search>.

Copija, A. *et al.* PTEN—clinical significance in colorectal cancer. *OncoReview* **6**, A86-90 (2016).

Perkins, N. The diverse and complex roles of NF- κ B subunits in cancer. *Nat. Rev. Cancer* **12**, 121–132 (2012).

Safran, M. *et al.* GeneCards Version 3: the human gene integrator. *Database (Oxford)* **2010**, baq020. <https://doi.org/10.1093/database/baq020> (2010).

Morin, P. *et al.* Activation of β -catenin-Tcf signaling in colon cancer by mutations in β -catenin or APC. *Science* **275**, 1787–1790 (1997).

Kim, C. G. *et al.* Effects of microsatellite instability on recurrence patterns and outcomes in colorectal cancers. *Br. J. Cancer* **115**, 25–33. <https://doi.org/10.1038/bjc.2016.161> (2016).

Kent, W. J. *et al.* The human genome browser at UCSC. *Genome Res.* **12**, 996–1006. <https://doi.org/10.1101/gr.229102> (2002).

Hu, L., Huang, Z., Wu, Z., Ali, A. & Qian, A. Mammalian Plakins, giant cytolinkers: versatile biological functions and roles in cancer. *Int. J. Mol. Sci.* <https://doi.org/10.3390/ijms19040974> (2018).

Bahrami, A. *et al.* Therapeutic potential of targeting Wnt/ β -catenin pathway in treatment of colorectal cancer: rational and progress. *J. Cell Biochem.* **118**, 1979–1983. <https://doi.org/10.1002/jcb.25903> (2017).

Schrock, A. B. *et al.* Tumor mutational burden is predictive of response to immune checkpoint inhibitors in MSI-high metastatic colorectal cancer. *Ann. Oncol. Off. J. Eur. Soc. Med. Oncol.* <https://doi.org/10.1093/annonc/mdz134> (2019).

Cancer Genome Atlas, N. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **487**, 330–337. <https://doi.org/10.1038/nature11252> (2012).

Becht, E. *et al.* Immune and stromal classification of colorectal cancer is associated with molecular subtypes and relevant for precision immunotherapy. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* **22**, 4057–4066. <https://doi.org/10.1158/1078-0432.CCR-15-2879> (2016).

Calon, A. *et al.* Stromal gene expression defines poor-prognosis subtypes in colorectal cancer. *Nat. Genet.* **47**, 320–329. <https://doi.org/10.1038/ng.3225> (2015).

Tahara, T. *et al.* Colorectal carcinomas with CpG island methylator phenotype 1 frequently contain mutations in chromatin regulators. *Gastroenterology* **146**, 530–538.e535. <https://doi.org/10.1053/j.gastro.2013.10.060> (2014).

Jones, P. A. & Baylin, S. B. The fundamental role of epigenetic events in cancer. *Nat. Rev. Genet.* **3**, 415–428. <https://doi.org/10.1038/nrg816> (2002).

Carter, C. A. *et al.* No patient left behind: the promise of immune priming with epigenetic agents. *Oncoimmunology* **6**, e1315486. <https://doi.org/10.1080/2162402X.2017.1315486> (2017).

Beauchemin, N. & Arabzadeh, A. Carcinoembryonic antigen-related cell adhesion molecules (CEACAMs) in cancer progression and metastasis. *Cancer Metastasis Rev.* **32**, 643–671. <https://doi.org/10.1007/s10555-013-9444-6> (2013).

Bacac, M. *et al.* A novel carcinoembryonic antigen T-cell bispecific antibody (CEA TCB) for the treatment of solid tumors. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* **22**, 3286–3297. <https://doi.org/10.1158/1078-0432.CCR-15-1696> (2016).

Quenet, *et al.* A UNICANCER phase III trial of hyperthermic intra-peritoneal chemotherapy (HIPEC) for colorectal peritoneal carcinomatosis (PC): PRODIGE 7. *J. Clin. Oncol.* **36**, LBA3503 (2018).

Author contributions

Study design: A.D.B., S.H., H.Y. Patient recruitment: S.H., H.Y. Molecular analysis: S.H., J.S., C.W., C.B., V.P.

Funding

The study was funded by a grant from the Good Hope Hospital Charity. ADB is funded by a Cancer Research UK Advanced Clinician Scientist Award (C31641/A23923).

Competing interests

The authors declare no competing interests.


Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-020-75844-6>.

Correspondence and requests for materials should be addressed to A.D.B.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access  This article is licensed under a Creative Commons Attribution 4.0 International

License, which permits use, sharing, adaptation, distribution and reproduction in any medium or

format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020

