Investigate Tumour Heterogeneity and Genetic Pathways in Colorectal Cancer

HERSH ABDUL HAM KARIM

MSc

Thesis submitted to the University of Nottingham
for the degree of Doctor of Philosophy

2017

Abstract

Background: Colorectal cancer is a major global public health problem that has predominantly been considered a genetic disease following a precise series of molecular events. These are characterized by sequential accumulation of genetic and epigenetic alteration in several oncogenes and tumour suppressor genes. Understanding of the genetic mechanisms that explain the initiation and evolution of colorectal cancer are key to improving risk prediction, prognostication and treatment.

Aims: The aim of this study was to understand the basic principles of the molecular biology of colorectal cancer based on genomic, transcriptomic, and proteomic profiles analyses.

Methods: DNA extracted from 147 formalin fixed paraffin embedded (FFPE) samples from 83 patients with colorectal cancer (first cohort) including (83 primary colorectal cancer (CRC), 22 matched liver metastases, 25 matched biopsies and 17 normal colon tissue) were screened for mutation in 26 genes (Trusight tumour kit) using Targeted Next Generation Sequencing. Additionally, exonuclease domain region of *POLE* and *POLD1* were also screened using High Resolution Melting and Sanger Sequencing methods. These data used to:

- Investigate mutation profiles of CRCs among 83 primary samples.
- Investigate the difference between chromosomal instable-CRC (46 primary sample) and chromosomal and microsatellite stable-CRC (35 primary sample).
- Compare mutations in 26 genes of 25 paired biopsy samples and corresponding resection specimens.
- Investigate genetic discrepancies between 22 primary colorectal cancers and their respective metastases. Additionally, expression of a panel of six miRNAs (miR-20a, miR-21, miR-29a, miR-31, miR-92a and miR-224) was measured using RT-qPCR and protein expression of 20 genes was measured using Reverse Phase Protein Array (RPPA).

In a second cohort including 81 primary CRC and their matched normal samples, expression of the six miRNAs and mRNA of six genes (*SMAD4*, *PTEN*, *BCL2*, *TGFBRII*, *KLF4* and *RASA1*) targeted by the six miRNAs were measured using RT-qPCR. Additionally, expression of proteins of the targeted six genes was also measured using immunohistochemistry (IHC).

Cell-free DNA (cfDNA) extracted from 16 blood samples (third cohort), which were taken from 5 CRC patients at different time points (pre- and post-surgery) were screened for mutations in *KRAS*, *TP53*, *PTEN*, *SMAD4*, *BRAF* and *PIK3CA* genes. Additionally, expression of the six miRNAs was measured using RT-qPCR.

Results: **In the first part**; investigating mutation profile of the first cohort 83 CRC showed high frequency of mutation in TP53 (75%), *APC* (57%) and *KRAS* (53%). Approximately 93% CRCs have mutation in at least one of *APC*/*TP53*/*KRAS*/*BRAF*/*SMAD4*/*PIK3CA*/*PTEN*/*FBXW7* genes. Moreover, mutations were found in the exonuclease domain regions of *POLE* in 9.6% and *POLD1* in 2.4%. Regarding biopsy vs resection, the mutant allele frequency was 1.03-fold higher in resection specimens than biopsies and there was no mutation in the biopsy specimens that were not seen in the resection specimens.

In the second part; Comparison of CIN-CRC vs MACS-CRC, which were included in the first cohort CRCs showed similar mutation frequencies of mutation in all 28 genes except *KRAS (41%CIN vs 68%MACS), POLE (15%CIN vs 2%MACS), GNAS (0%CIN vs 11%MACS).* Statistically there was a significant difference (each p=0.01) which was lost following multiple testing correction.

In the third part; comparison of primary CRC vs matched metastasis showed that a total of 61 non-synonymous somatic variations in 12 genes were found in primary 22 specimens whereas 60 were found in metastasis cases. The mutant allele frequency was 1.01-fold higher in primary than metastasis CRCs. Evaluated expression levels of six miRNAs and protein expression of other 20 genes, did not show any significant differences between primary CRC and matched metastasis.

In the fourth part; Expression of the six miRNAs and mRNA and protein of the six targeted genes were tested in the second cohort 81 samples. Statistical analysis revealed significant increase in the expression level of miR-20a (p=0.04), miR-21 (p=0.01) miR-29a (p=0.03) and miR-31 (p=0.01) and decrease in the mRNA expression level of TGFBRII and RASA1 in tumour samples compared to normal tissues. IHC staining showed low expression level of SMAD4 in 51 (63%), PTEN 67 (83%), TGFBRII 65 (80%), RASA1 61 (75%) BCL2 47 (58%) and high expression of KLF4 36 (44%). High miR-21 and miR-224 expression were associated with low expression of TGFBRII. In addition, over expression of both miR-29a and miR-31 inversely correlated with RASA1.

In the fifth part; Mutation in the cfDNA was detected in 5 cases. two of these showed a loss of the mutant signal post-operatively. Whereas the mutant signal was persistent in the rest 3 of the cases for all the samples taken post-operatively. Although miRNAs expression was fluctuated between these time points, paired t-test showed a non-significant difference when comparing pre- and post-surgical miRNAs level. However, level of the cfmiRNAs is changed by more than 2 folds (upregulated) in the day of surgery compared to normal plasma as follow, miR-20a in 1/5 (20%), miR-21 in 4/5 (80%), miR-29a in 3/5 (60%), miR-92a in 4/5 (80%) and miR-224 in 2/5 (40%).

Conclusion: Investigation profiles of CRCs from both cohorts indicated that, different mutated genes and upregulated miRNAs, which are involved in different signalling pathways, may have roles in CRC carcinogenesis. Significant difference was neither noticed between MACS and CIN group and nor between primary and metastasis tumour. miRNAs from tissues and cfmiRNAs from plasma, can differentiate CRC from healthy group and have potential clinical value in early CRC detection. In addition to the resection specimens, the study found that it is acceptable to use biopsy material for predictive testing and cfNAs can be used for a variety of clinical and investigational applications.

Publications arising from this thesis

Articles:

Ebili Henry O., **Ham-Karim Hersh Abdul** and Ilyas Mohammad**.** A Simple Two-Stage PCR Method for Quality Amplification of Degenerate DNA. Int J Cancer Clin Res 2015, 2:2.

Henry O Ebili,[1,2] James Hassall,[1] Abutaleb Asiri,[1] **Hersh Ham-Karim**,[1] Wakkas Fadhil,[1,3] Ayodeji Johnson Agboola[2], Mohammad Ilyas[1,3]. QMC-PCRx: a novel method for rapid mutation detection. J Clin Pathol 2017;0:1-10.

Hersh Ham-Karim[1], Henry Okuchukwu Ebili[1,2], Wakkas Fadhil[1]Abutaleb Asiri[1], James Hassall[1], Mohammad Ilyas[1]. **cold-hrm:** A combination of methods to infer the nature of somatic mutations. Adv Cytol Pathol 2017; 2(2): 1-9.

Hersh A. Ham-Karim[1], Kirsty Manger[2], Wakkas Fadhil[1], Susan D. Richman[3], Narmeen S. Ahmad[4], Gareth Cross[2], Enric Domingez[5], Ian P Tomlinson[5] and Mohammad Ilyas[1]**.** Targeted next generation sequencing (NGS) demonstrates a common early genetic pathway between colorectal cancers with chromosomal instability (CIN) and those with microsatellite and chromosome stability (MACS) (Manuscript in preparation).

Hersh A. Ham Karim, Ola H Negm, Kirsty Manger, James C. Hassall, Wakkas Fadhil, Narmeen S. Ahmad, Agrawal Utkarsh, Mohammad Ilyas. Comparative analysis of primary tumour and matched metastases in colorectal cancer patients; An investigation of the genomic, proteomic and post-transcriptional regulation profiles (Manuscript in preparation).

Hersh A. Ham-Karim[1], Wakkas Fadhil[1] and Mohammad Ilyas[1]**.** Identification of Somatic Mutations in Exonuclease Domain Regions that Alters Proofreading Activity of both POLE and POLD1 Genes (Manuscript in preparation).

Hersh A. Ham-Karim[1], Kirsty Manger[2], Wakkas Fadhil[1], Narmeen S. Ahmad[3], Gareth Cross[2] and Mohammad Ilyas[1]. Targeted Next Generation Sequencing validates the utility of diagnostic biopsies for predictive testing in colorectal cancer using (Manuscript in preparation).

Abutaleb Asiri[1], James Hassall[1], **Hersh Ham-Karim**[1], Henry Okuchukwu Ebili[1], Wakkas Fadhil[1], Oliver Ng[2,3], Peter Fardouly[3], Austin G Acheson[2,3] and Mohammad Ilyas[1]. Circulating free DNA in colorectal cancer as a marker of surgical clearance (Manuscript in preparation).

Posters and Oral presentations

Oral Presentation "Comparative Analysis of Primary Tumour and Matched Metastasis in Colorectal Cancer Patients; Investigate Genomic, Transcription and Translational Profiles" 9th Joint Meeting of the British Division of the International Academy of Pathology and the Pathological Society of Great Britain & Ireland 28 June – 1 July 2016.

Poster Presentation "A Simple Strategy for Delineating Allelic Imbalance in Tumours Without Requiring Normal Tissue" 9th Joint Meeting of the British Division of the International Academy of Pathology and the Pathological Society of Great Britain & Ireland 28 June – 1 July 2016.

Poster Presentation "Aberrant High Resolution Melting Pattern: Germline Variant or Somatic Mutation?" 9th Joint Meeting of the British Division of the International Academy of Pathology and the Pathological Society of Great Britain & Ireland 28 June – 1 July 2016.

Oral Presentation “Targeted next generation sequencing validates the utility of diagnostic biopsies for predictive testing in colorectal cancer using” 26th European Congress of Pathology Pathology – Understanding disease 30 August – 3 September 2014, ExCeL, London, United Kingdom.

Poster Presentation “Comparison of Chromosomal Instability (CIN) and Microsatellite and Chromosome Stability (MACS) using targeted next generation sequencing” 26th European Congress of Pathology Pathology – Understanding disease 30 August – 3 September 2014, ExCeL, London, United Kingdom.

Acknowledgements

Thanks to GOD for helping me to fulfil this study

I would like to thank Professor Dr. Mohammad Ilyas for his guidance, encouragement, support, and patience with me throughout this work. Without his method of monitor ship, I would not have developed this research as much professionally as I have.

I wish to express my deepest appreciation and grateful thanks to Kirsty Manger for her scientific support, valuable aid and guidance throughout the work.

I am deeply grateful to Dr. Wakkas Fadhil for his supervision, support, sincere cooperation and much needed moral support throughout my PhD study.

I wish to express my deep thanks to Dr. Ola Negm for her support, and sincere cooperation.

I am also grateful for all my colleagues and the staff of the pathology research group at the University of Nottingham for their assistances and patience during the work.

My heartfelt appreciation goes to my father, and my mother for their support and encouragement during my study.

My deepest gratitude goes to my wife (Dr. Narmeen Ahmad) and my little prince (Aleen) for their patience and support during my studies.

Dedication

Words fail me to express my appreciation to my wife (Dr. Narmeen S. Ahmad) whose dedication, love and persistent confidence in me, has taken the load off my shoulder.

List of Abbreviations

APC	Adenomatous Polyposis Coli
BAX	BCL2-Associated X Protein
BCL2	B-cell lymphoma 2
CDH1	Epithelial Cadherin 1
CEA	Carcinoembryonic antigen
cfDNA	Cell-free DNA
cfNA	Cell-free nucleic acid
CIMP	CpG island methylator phenotype
CIN	Chromosomal instability
Cold-PCR	CO-amplification at Lower Denaturation temperature PCR
CpG	5′-CG-3′ dinucleotide
CRC	Colorectal cancer
CTNNB1	Catenin Beta 1
DNA	DeoxyriboNucleic Acid
dNTP	DeoxyriboNucleotide TriPhosphate
DPBS	Dulbecco‟s Phosphate Buffered Saline
DTT	Dithiothreitol
ED-	Exonuclease domain
EDM	Exonuclease domain mutation

EDTA	Ethylene Diamine Tetraacetic Acid
EGFR	Epidermal growth factor receptor
EMT	Epithelial to mesenchymal transformation
FAP	Familial Adenomatous Polyposis
FBXW7	F-box and WD repeat domain-containing 7
FFPE	Formalin Fixed Embedded Paraffin
FGFR2	Fibroblast growth factor receptor 2
FOBT	Faecal occult blood tests
FOXL2	Forkhead Box L2
GAPDH	GlycerAldehyde-3-Phosphate DeHydrogenase
gDNA	Genomic DNA
GDP	Guanosine di-phosphate
GNAS	Guanine Nucleotide-Binding Protein
GTP	Guanosine tri-phosphate
HNPCC	Hereditary Non-Polyposis Colorectal Cancer
HPRT	Hypoxanthine PhosphoRibosyl Transferase
HRM	High Resolution Melting
KRAS	Kirsten-Retrovirus Associated DNA Sequences
LNA	Locked nucleic acid
LOH	Loss of heterozygosity
MACS	Microsatellite and chromosomal stable

MCR	Mutation cluster region
MiR-	MicroRNA
MLH1	Mutl homologue 1
MMR	Mismatch Repair
mRNA	Messenger RiboNucleic Acid
MSH2	Muts homologue 2
MSH6	Muts homologue 6
MSI	Microsatellite instability
MSI-H	Microsatellite instability High
MSI-L	Microsatellite instability Low
MSS	Microsatellite stable
mtDNA	Mitochondrial DNA
NCBI	National centre for biotechnology information
NGS	Next Generation Sequencing
NRAS	Neuroblastoma RAS
PCR	Polymerase Chain Reaction
PDM	Prediagnostic multiplex
PI3K	Phosphoinositide 3-kinase
PI3KCA	Phosphoinositide-3-kinase, catalytic, alpha polypeptide
PMS	Protein homologue 2
POLD1	Polymerase (DNA) delta 1, Catalytic Subunit

POLE	Polymerase (DNA) Epsilon, Catalytic Subunit
PTEN	Phosphatase and tensin homolog
QCP	Quality control primer
QCT	Quality control template
QMC-PCR	Quick-multiplex consensus PCR
qPCR	Quantitative Polymerase chain reaction
RNA	RiboNucleic Acid
RT-	Reverse Transriptase
SDS	Sodium dodecyl sulfate
SMAD4	Mothers against decapentaplegic (MAD[12]) and the Caenorhabditis elegans protein SMA
SNP	Single nucleotide polymorphism
SNV	Single nucleotide variant
SSD	Single specific diagnostic
TBS	Tris Buffered Saline
Tc	Critical temperature
TGFBRII	Transforming growth factor-βeta receptor 2
TMA	Tissue microarray
TP53	Tumour protein p53
UTR	3-untranslated region
WTB-PCR	Wild-type blocking PCR

List of contents

List of figures

Chapter One

Chapter Two

Chapter Three

Chapter Five

Chapter Six

Chapter Seven

List of Tables

Chapter One

Chapter Two

Chapter Three

Chapter Four

Chapter Five

Chapter Six

Chapter Seven

Chapter one: Literature Review

1.1 Colorectal Cancer (CRC)

Colorectal cancer is a complex and heterogeneous disease [1] which takes decades to develop [2]. It is the third and second most commonly diagnosed cancer in males and females respectively. The World Health Organisation (WHO) reported that more than one million (~1.2 million) colorectal cancer patients have been diagnosed yearly with approximately 600,000 deaths from the disease annually [3]. Genetic factors, environmental factors, and advanced age are predisposing factors which have an effect on the risk of CRCs [4]. The risk of developing CRC increases with age, and over 90% of sporadic CRCs occur in individuals over the age of 50 [5]. Other factors include family history of CRC, alcohol, cigarette smoking, obesity, sedentary occupation, a diet low in fibre and folate and high in fat and red meat, diabetes. There is variation in the geographical area as well [6]. For example, CRCs rate is relatively low in Asia, Africa, and parts of Latin America while it is high in the ‘‘Western’’ countries, including Northern Europe, Australia, New Zealand, and United States of America [7]. The incidence of CRC is lower in developing countries than in developed countries possibly because of the Western sedentary life style along with a diet which is poor in fruits and vegetables while rich in fat [8]. Generally, 30% of risk of CRC is hereditary although only 2-5% of CRCs cases are attributable to familial cancer syndromes with a highly penetrant inherited mutation such as Familial Adenomatous Polyposis (FAP) and Hereditary Non-Polyposis Colorectal Cancer (Lynch syndrome). However, about 70% of CRC cases seem to arise sporadically. [9, 10].

1.2 The concept of Genetic pathways in CRC

As in other types of cancer, CRCs develop through the accumulation of genetic and epigenetic alterations. These occur in a sequential manner that leads to the transformation of normal colorectal epithelium to colorectal adenocarcinoma [11]. Wood et al, [12] found that although in CRC there are 80 common exonic mutations (termed mountain genes), 15 of these are essential for the initiation and promotion of cancer and are called “driver” mutations. The rest represent “passenger mutations”, which are not important for carcinogenesis. The non-

random accumulation of mutations during carcinogenesis can be considered as the “genetic pathway” [12]. The first pathway of genetic alterations was described by Fearon and Vogelstein (figure 1.1) [4]. This proposed that mutational inactivation of tumour suppressor genes such as *APC* and *TP53* coupled with mutational activation of oncogenes such as *KRAS* lead to colorectal neoplasia. The pathway also conjectured that CRCs develop as a result of mutations in more than 5 genes and finally the accumulation of genetic alterations rather than their order is responsible for determining the biologic behaviour of the tumour. As Zlobec et al, [13] highlighted, this model is responsible for FAP and approximately 85% cases of sporadic CRC. Despite its strengths, the Fearon and Vogelstein model is not able to illuminate several phenomena such as the presence of other genetic pathways, the low level of concomitant mutations in the key genes and the difference in the frequency of adenoma and carcinoma cases, as in CRCs there is the vast difference in the rate of malignant conversion of adenomas in different clinical conditions [14, 15].

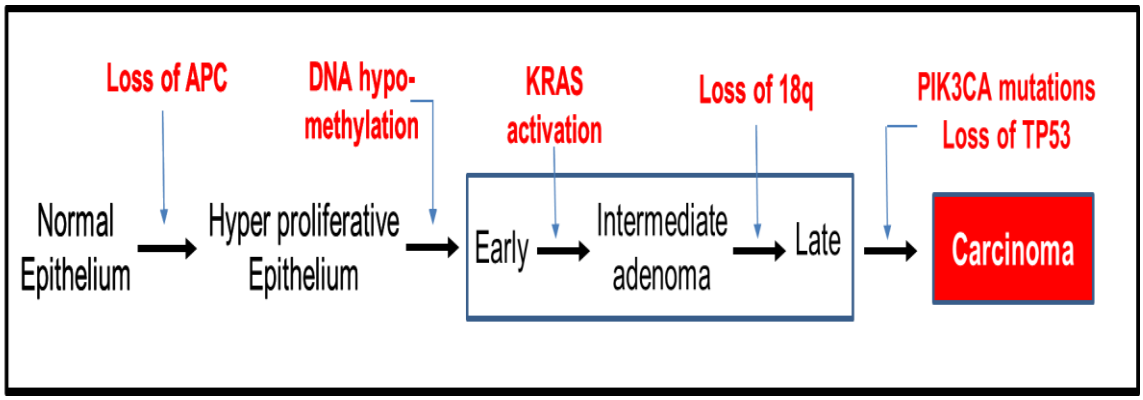

Figure 1.1: The Fearon and Vogelstein model. It is characterised by classic tubular adenoma histology, early acquisition of APC mutations, frequent activating mutations of the KRAS oncogene at the early adenoma stage, loss of heterozygosity at chromosome 18q (18qLOH) in late adenomas and TP53 mutations that facilitate the transition to malignancy. (updated from Mudassar et al, [16])

1.3 Genetic pathways defined by genomic instability

In CRC, two distinct patterns of genomic instability are described and can be used to divide tumours into distinct groups. The most common is termed the

chromosomal instability pathway, in which aneuploidy, amplification and translocation are common [17]. The second pattern of genomic instability is the microsatellite instability (MSI) pathway which is less common and arises due to loss of proficiency of the DNA mismatch repair (MMR) system. Recently a third pathway of CRC tumours have been described without either MSI or CIN, and referred to as microsatellite and chromosome stable (MACS) [18].

1.3.1 Chromosomal Instability (CIN)

In a healthy body, genomic stability firmly controls cell homeostasis. Disruption of this will lead to enhanced mutational rate, clonal diversity of the mutated cells and eventual progress to cancer [19]. Chromosomal instability represents the most common type of genomic instability in colorectal cancers approximately ranging between 60% - 80% [20] and is defined either as numerical CIN or structural CIN [21]. Alteration in the number of chromosomes due to gain or loss of whole chromosomes is referred to as numerical CIN while amplifications, deletions, or translocations at the chromosomal level referred to as structural CIN [20]. Alteration of the mitotic spindle checkpoints, which normally and tightly regulate cell cycle progression by controlling normal duplication and segregation of chromosomes, seems likely to contribute to genomic instability and enhance chromosome aneuploidy and polyploidy [22]. In addition, McGranhan et al, [21] found that irregular chromosome attachment to the mitotic spindle, aberrant sister chromatid cohesion and centrosome amplification are other mechanisms which might be contribute to numerical CIN. As a result, tumour cells with CIN have karyotypic variability [23], chromosomal rearrangements, loss of heterozygosity (LOH), and an imbalance in chromosomal number either aneuploidy or polyploidy [17]. The molecular mechanism of genomic instability of CIN is not well established although many genes have been associated with the chromosomal instability [24]. For example, dysfunction of the adenomatous polyposis coli (APC) gene, involved in the *Wnt* signal transduction pathway and associated with the early stage of tumourigenesis [25]. The tumour suppressor gene *TP53*,

participates in G1 arrest and apoptosis, is another gene that may promote the development of aneuploidy in CRCs.

Genes that are frequently targeted by point mutations in CIN-pathway tumours are the oncogene *KRAS*, activation of which leads to growth promotion, and *SMAD4*, a component of the transforming growth factor beta pathway [26]. Moreover, abnormal number or function of *Aurora* kinase A, STK15/BTAK, which is a centrosome-associated serine threonine kinase, and mutations in *BUB1* and *BUB1B* (BUBR1), genes encoding mitotic checkpoint proteins, may be causes of CIN [11]. Significantly, several clinical studies have highlighted that CIN is an indicator of bad prognosis in CRC [24] and it seems to be occurring in the left/distal side of colon and attributed by well or moderately differentiated, rarely mucinous without any lymphocytic infiltration. In addition, more lymph node involvement with high risk of metastasis are feature of CRCs with CIN [2].

1.3.1.1 Adenomatous Polyposis Coli

The *APC* gene, which makes the APC protein and plays a critical role in several cellular processes, is a tumour suppressor gene. It was first mapped to chromosome 5q21 in 1987 and it was cloned in 1991 [27, 28]. *APC* is a relatively large gene composed of 15 exons and which encodes an 8.5-kb mRNA and 312-kD protein with multiple functional domains. The APC protein negatively regulates *Wnt* signalling via targeting β-catenin for ubiquitin-mediated proteasomal degradation through making a destruction complex with AXIN, Casein kinase 1 (CK1), glycogen synthase kinase 3 (GSK-3β), and Protein Phosphatase 2A (PP2A) [29, 30]. In addition, *APC*, not only degrades β-catenin but also inhibits its nuclear localization and therefore controlling cell cycle progression. *APC* is involved in stabilizing microtubules thus promoting chromosomal stability [31]. Wnt signalling occurs when the destruction complex is lost and the resultant stabilization of intracellular β-catenin facilitates its translocation to the nucleus, where it binds to nuclear partners TCF/LEF family (T cell factors/lymphoid enhancer factor) to promote the transcription of different genes involved in cellular

activation [29, 32]. Therefore, the interaction with β-catenin is essential for the tumour suppressor activity of *APC* [33].

Mutation of *APC* occur somatically in sporadic CRC and in the germline, is the cause of the familial adenomatous polyposis (FAP) [34]. *APC* mutation is regarded as the initiating event in CRC and is found in classic tubular adenomas of the CIN pathway [35-37]. In both patients with FAP and patients with sporadic CRC, more than 90% of *APC* mutations occur in the mutation cluster region (MCR), which has been identified in exon 15 located between codons 1286 and 1513 and which comprises less than 10% of the total *APC* coding sequence [38, 39]. Germ-line *APC* mutations cause FAP an inherited cancer predisposition syndrome, which is characterized by development of more than 100 adenomatous polyps; in people with germline mutations in the *APC* gene, the risk of CRC by the age of 40 years is almost 100% [29, 32, 40]. Moreover, somatic mutations in APC have been reported in 60% of sporadic CRC and occur not only in the advanced adenoma, small adenoma and carcinoma [41, 42], but also in the microscopic adenoma with few dysplastic glands [38].

1.3.1.2 K-RAS

The Ras family of Small-G proteins consist of three isoforms; Kirsten RAS (*KRAS*), Harvey RAS (*HRAS*) and Neuroblastoma RAS (*NRAS*). These act as the molecular switches downstream of growth factor receptors [43, 44]. The RAS gene codes for a protein normally present in an inactive GDP bound form. In the presence of extracellular signals, the GDP is exchanges for GTP and the GTP-bound form of *KRAS* participates in the RAS/RAF/MEK/ERK signalling pathway and, transfers signals that regulate different biological processes involved in cell proliferation, cell survival, and migration [45]. KRAS protein, which encodes a 189 amino acid protein is located on the short arm of chromosome 12, is composed of four coding exons and a 5' non-coding exon [46] and is a prominent component in the development of CRCs [47]. In human CRC, more than quarter of traditional adenomatous polyps have *KRAS* mutation [48] where as 40% of all CRCs have *KRAS* mutation [48-50]. [50]. Generally, the frequency of *KRAS* mutation

increases with the size of adenoma. Around 10% of cases who have small adenoma less than 1cm, show *KRAS* mutation while this frequency is raised to 40%-50% in those cases that have adenoma larger than 1cm [25]. Approximately 90%-95% of *KRAS* mutations are found in codon 12 (80%-85%) and codon 13 (10%-15%) while other mutations less frequently occur in codons 59, 61, 117, 146, and 154 [51, 52]. The role of *KRAS* mutations in prognosis for patients with CRCs remains controversial. For example, the data collected by RASCAL group from 2721 patients from 13 different countries indicated that presence of *KRAS* mutations is associated with recurrence and bad prognosis in CRCs [53, 54]. Additionally, Quasar 2 trial, which was done in 170 hospitals in 7 countries, collected data from 7475 patients aged 18 years or older with CRC that was histologically proven to be stage III or high-risk stage II. In their multivariate analyses, the presence of *KRAS* mutation was significantly associated with poorer prognosis [55]. In contrast, other studies Roth et al., [56] collected data from 1564 cases and Ogino et al., [57] collected data from 1264 cases demonstrated that *KRAS* mutations do not have any prognostic value in patients with stage II and stage III CRCs. Presence of variation in the results of different studies might be related to the difference in sample size or samples that have been taken from patients at different stages [46]. The studies conducted by Astrid et al., [58] and Khambata-Ford et al., [59] have demonstrated that KRAS has predictive value i.e. cancer with *KRAS* mutation do not have any response to cetuximab treatment. Moreover, European Medicines Evaluation Agency, American Society of Clinical Oncology (ASCO), and National Comprehensive Cancer Network (NCCN) announced that testing for *KRAS* gene should be performed in advance for CRC patient's prior to receive treatment with panitumumab or cetuximab [46].

1.3.1.3 PIK3CA

The PI3K family is composed of three classes (I, II, and III). The class I, which represents the best characterised class among the PI3K classes, is divided into class IA and Class IB. The Class IA PI3Ks, which are more commonly associated

with human cancer, are activated by small G-protein RAS, receptor tyrosine kinase (RTKs) and G-protein– coupled receptors (GPCRs) [60] (figure 1.2). The Class IA PI3Ks, which are heterodimers consisting of a catalytic protein kinase domain (P110α) and a regulatory (P85) subunit, phosphorylate phosphatidylinositol 4,5 bisphosphate (PIP2) to form phosphatidylinositol 3,4,5-trisphosphate (PIP3) which in turn recruits 3-phosphoinositide-dependent kinase 1(PDK1) to activate the AKT pathway and consequently regulate different cellular processes such as cell growth and survival [61]. Mutations in the *PIK3CA* gene, encoding the P110α subunit, are commonly associated in human cancers such as prostate, colon, endometrium and breast cancer [62]. Occurrence of somatic mutations in *PIK3CA* has been reported in between 14% to 25% of CRCs cases [63], and more than 80% of these mutations occur in a small hot spot area in exon 9 and exon 20 [64]. Hence *PIK3CA* mutations, which are oncogenic in CRC, commonly occur late in tumourigenesis [65]. In mutant *PIK3CA* the activity of p110α catalytic domain drastically increases by two main molecular mechanisms [62]. Mutations in exon-9 (helical-domain) raise activity of P110α by interacting with RAS-GTP binding independent of binding to the p85 regulatory subunit while mutations in exon-20 (kinase-domain) activate the P110α catalytic subunit after binding with p85 independent on the interaction with RAS-GTP [66]. The Velho et al., [63] study accentuated that no differences were reported in the frequency of *PIK3CA* mutations between MSI and MSS subsets and no association has been observed between *PIK3CA* mutations and CIMP or *MLH1* methylation in CRCs. However, in CRCs, mutant *PIK3CA* is likely to occur concomitantly with mutations that activate *KRAS* and *BRAF* [67]. Moreover, a European consortium highlighted that the activity of cetuximab is reduced in the presence of exon-20 *PIK3CA* mutations whereas presence of exon-9 *PIK3CA* mutations do not show any impact on cetuximab efficacy [68]. Additionally, the study conducted Liao et al, [69] by suggested that regular use of aspirin after the diagnosis of colorectal cancer is significantly associated with increased survival among patients with mutated-*PIK3CA* tumours but not among patients with wild-type *PIK3CA* tumours.

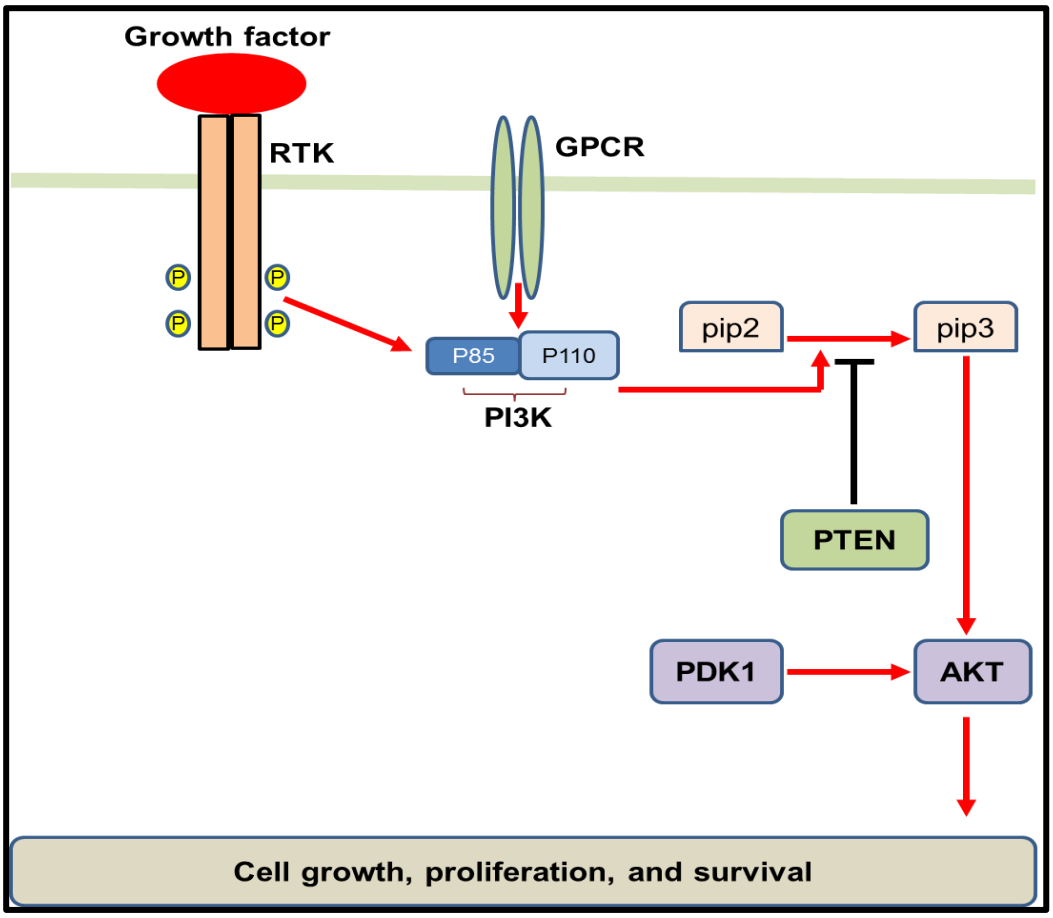

Figure 1.2: Simplified PI3K signalling, updated from Keniry and Parsons 2008 [70].

1.3.1.4 PTEN

The phosphatase and tensin homolog (*PTEN*) gene, located on chromosome 10 is a major antagonist of the phosphatidylinositol 3- kinase (PI3K) survival pathway and it mediates dephosphorylation of PIP3 to PIP2 [71]. Some studies reported that the frequency of inactivating *PTEN* somatic mutations is around 10% of CRCs [12] and others shown a higher frequency loss of *PTEN* protein expression of 15–20% of CRCs [72]. Loss of *PTEN* expression is observed in both *KRAS* mutant and *KRAS* wild type tumours. Similar to *PIK3CA* oncogenic mutations, deregulation of *PTEN* tumour suppressor gene function most likely acts to enhance effects downstream of KRAS protein via PIP3-mediated activation of AKT [71].

1.3.1.5 Chromosome 20q

One of the shortest chromosomes in human genome is chromosome 20 which composed of 63 mega base pairs and bears more than 566 genes on the q-arm (NCBI). Recurrent amplification and gain of the q-arm of chromosome 20 (20q13) is one of the most frequent chromosomal alterations among several chromosomal anomalies that have been reported in CRC [73]. In early colorectal adenoma, chromosome 20q gain was only reported from 0%– 20% [74, 75] while this range increased dramatically to 60% in advanced adenoma and carcinoma [76]. This indicates that occurrence of chromosome 20q gain promotes the malignant transformation of colorectal adenoma to carcinoma [74]. It is associated with bad prognosis [77], aggressive tumour phenotype and metastasis formation and consequently reduces the survival rates in CRCs [73]. In the past, few decades, the effects of several genes, which are located on chromosome 20q, have been reported in the progression of CRC. For instance, *Aurora* kinase A (AURKA), which participate in cell cycle kinase regulation, mitotic entry, centrosome maturation and spindle formation [78], and TPX2, which regulates binding of Aurora-A to the spindle microtubules [79], are two main genes enhanced chromosome 20q gain and accelerate colorectal adenoma-to-carcinoma progression [80]. Chromosome segregation 1-like (*CSE1L*) gene, which is located on chromosome 20q13, is another gene highly expressed in CRC and play a role in apoptosis, adhesion and invasion and tumours with strong *CSE1L* expression have poor prognosis [81]. Another gene located on chromosome 20q and implicated in CRC is *BCL2L1*. It is a member of BCL2 protein families which participate in mitochondrial pathway of apoptosis through releasing of pro- (Bcl-xS) and anti-apoptotic (Bcl-xL) factors [82]. As Krajewska et al., (114) highlighted, both of pro- (Bcl-xS) and anti-apoptotic (Bcl-xL) splice variant have low expression in normal tissue compared to adenomas whereas BCL-XL mRNA and protein strongly expressed in CRCs. In addition, several other genes harboured on the chromosome 20q are also involved and highly expressed in CRCs such as *DIDO1*, *RBM39*, *HM13*, *TCFL5*, *SLC17A9*, *SRC*, *ZNF217* and *PRPF6* [77, 80, 82].

1.3.1.6 FBXW7

FBXW7 (F-box and WD repeat domain-containing 7), also known as *FBW7* and *CDC4*, which is one of the F-box proteins component of the SCF (SKP1/CUL1/F-box protein) E3 ubiquitin ligase, has been found to be involved in numerous cellular processes including cell proliferation, apoptosis, cell cycle and differentiation [83-85]. In humans *FBXW7* is located on chromosome 4 and encodes three transcripts (isoforms α, β and γ) derived from the same gene locus by alternative splicing [86]. Importantly, *FBXW7* is considered as a tumour suppressor protein which induces the degradation of positive cell-cycle regulators (oncoproteins) such as c-Myc, cyclin E, c-Jun, mTOR (mammalian target of rapamycin), Notch, peroxisome proliferator-activated receptor-γ coactivator-1α (PGC-1α), Presenilin SREBP Aurora-A Krüppel-like factor 5 (KLF5) DEK proto-oncogene and c-Myb [86-88]. Loss of its function leads to genetic instability by mechanisms that might involve the activation of Aurora-A and tumor growth [86, 89]. *FBXW7* is a tumor suppressor that is deactivated by deletions and point mutations in a variety of human neoplasms [90] such as in T-cell acute lymphocytic leukaemia [91], prostate cancer [92], CRC [93], glioma [94], and gastric cancer [94]. In CRC, deactivated *FBXW7 due to somatic mutations* has been reported in less than 8 to 10% of cases [95-97] and results in a CIN phenotype due to a defect in execution of metaphase and poor prognosis [93]. Recently, several studies showed that in human cancer cell drug resistance is mediated by suppressing *FBXW7* and increased levels of pro-survival factor MCL1 and mTOR [98-100]. In addition, Wang et al. demonstrated that in CRC cells, loss of *FBXW7* leads to rapamycin drug-resistant by inducing Epithelial-Mesenchymal Transition (EMT) [101]. The vast majority of the mutations detected in *FBXW7* were single nucleotide changes, predominantly missense substitutions rather than truncating mutations and LOH is rare [102, 103].

1.3.1.9 18q

Allelic imbalance or loss of heterozygosity (LOH) is a common mechanism of loss of tumour-suppressor gene (TSG) functions in cancer [104-107]. It results from

several mechanisms including chromosomal deletion, mitotic non-disjunction, somatic recombination leading to uniparental disomy, gene conversion, etc [105, 107-109]. While the first and second mechanism lead to copy number loss LOH, the others cause a copy number neutral LOH [104, 109]. LOH represents the second hit of the TSG loss-of-function in the Knudson two-hit hypothesis of tumourigenesis. The first hit being either an inherited or somatic mutation of one copy of the gene [104-106]. LOH is a key feature that identifies whether tumours are arising from the CIN pathway or the MSI pathway. Allelic imbalance at chromosome 18q is a one of the potentially promising markers of CRC prognosis and associated with more aggressive clinical behaviour. In early stage tumours (eg; small adenomas) allelic losses of 18q are infrequent, but they are common in primary CRC particularly at advanced stages and nearly 100% of hepatic metastases coming from colorectal primaries, indicating that chromosome 18q LOH may contribute more to progression rather than initiation of colorectal cancer. [110, 111]. Chromosome 18q LOH has been associated with bad prognosis in CRC patients in stage II and stage III as well as decrease responsiveness to fluorouracil-based adjuvant chemotherapy in stage III CRC patients. [111]. Candidate tumour suppressor genes including *Deleted* in Colorectal Carcinoma (*DCC*), *SMAD2, DPC4*/*SMAD4* (*D*eleted in *P*ancreatic *C*ancer *4*), and *Cables* are located on this region [112]. The Smad group including *SMAD2* and *SMAD4* genes code for intracellular proteins involved as a downstream regulator in transmitting the signals from the transforming growth factor-β (TGF-β) superfamily onto the nucleus [113]. The *SMAD4* protein forms complexes with the receptor-phosphorylated *SMAD2* and *SMAD3* and then translocate from the cytoplasm to the nucleus where they regulate gene transcription include inhibitory *SMAD7* [114, 115] and cyclin-dependent kinase inhibitors such as p15 (ink4B) [116] in conjunction with a wide range of co-regulator proteins. Loss of *SMAD4* function may abrogate this signalling pathway and loss of transcription of genes critical to cell-cycle control [117] and consequently promotes tumourigenesis. In sporadic CRCs, somatic mutations

that abrogate *SMAD4* and *SMAD2* proteins are found in approximately 10% and 7% of patients, respectively [116, 118-120].

1.3.1.8 TP53

The p53 transcription factor acts as a tumour suppressor protein that inhibits or enhances the expression of some genes involved in several pathways such as cell cycle checkpoints (G1/S) and (G2/M), apoptosis and angiogenesis [25]. P53 protein is composed of 393 amino acids [23] encoded by the *TP53* tumour suppressor gene known as “the guardian of the genome’’ located on the chromosome 17p [121]. In addition to the functions performed in the nucleus, the p53 protein participates in regulating centrosome duplication, in inhibiting autophagy, in regulating apoptosis at the mitochondria and other functions in the cytoplasm [25]. To date, it has been found that inactivation of *TP53* transcription factor not only occurs as a result of mutation in *TP53* gene, but also occurs due to allelic loss on chromosome 17p and mutations in other factors that activate or degrade p53 protein such as ATM, ChK1, and the E3 ubiquitin ligase [122]. Loss of function of p53 has been reported in 4%–26% of adenomas, 50% of adenomas with invasive foci, and in 50%–75% of CRCs, defining its role in the transition from an adenoma to carcinoma [123]. Approximately 75-80% of mutations occurring in *TP53* gene are missense mutations particularly in 5 hotspot codons 175, 245, 248, 273, and 282 [23, 25]. Generally, there is an inverse relation between p53 mutation and MSI within a tumour. It is found that *TP53* mutation is less frequently observed in proximal and mucinous tumour, whereas it is more common in the distal colon and MSS tumour [124]. In addition to mutations and allelic loss, rearrangement, deletion, nonsense and frameshift mutations are also observed in CRCs with lower frequency [25, 125]. According to the Goh and Smith [126] study, the effect of radiation and chemotherapeutic agents are higher on the tumour with functional *TP53* gene than on tumour with mutant *TP53* gene and these results are potentially of great clinical importance. However, several available retrospective studies have not confirmed prognostic and predictive roles of mutant *TP53* in CRC [127].

1.3.1.7 Gain of 8q allele

In the past, several studies reported that a high degree chromosomal instability associates with CRC metastatic to the lymph nodes. In particular, gain of the long chromosomal arms of chromosome 8 most exclusively occur in CRCs and it is rarely observed in lymph node-negative carcinomas [128]. The researchers assumed that Chromosome 8q contains several oncogenes which participate in progression and metastasis of CRCs [129]. For example, PRL-3 gene that map to chromosome 8q24.3 encodes a small, 22-kD tyrosine phosphatase that is located at the cytoplasmic membrane. It is highly overexpressed in metastatic colorectal cancer [130]. In addition, c-MYC gene is located on chromosome band 8q24.21 and consists of three exons. It is a transcription factor that regulates activation of several genes. Activation of the c-myc gene, which occurs in several ways, leads to the overexpression and enhances the development of colorectal cancer [131]. Furthermore, single nucleotide polymorphisms at the 8q24 locus are correlated with an increased risk for the development of colon cancer [132].

1.3.2 Microsatellite insatiability (MSI)

The human genome contains several repetitive sequence known as microsatellites which consist of mono-, di- or tetra nucleotides such as (A)n or (CA)n [133]. The mismatch repair system is responsible for correcting all errors (insertion and deletion mutations) which might occur in these regions [134]. The only important factor to affect mutation rate that has so far been discovered is microsatellite length — mutation rate increases with an increasing number of repeat units [135]. These CRCs are known as deficient Mismatch repair (MMR) while the majority of CRCs have proficient MMR [34]. Mismatch repair DNA is a pathway that is highly conserved and plays an important role in the maintenance of genomic stability, (figure 1.3 describes how MMR system works). Therefore, germline mutation in mismatch repair genes (MMR) such as (*MLH1*, *MSH2*, *MSH6, PMS2 and EPCAM*) and somatic hypermethylation of promoter region of the *MLH1* gene lead to microsatellite instability which is present in 10-20% of colorectal cancers [136]. As a result, mutation rates increase and can be seen in

more than 30 driver genes that are implicated in cellular growth and function. For example, transforming growth factor β1 receptor type II (*TGF*β*1RII*) regulates cell growth, proliferation and differentiation, insulin-like growth factor II receptor (*IGFIIR*) receptor for *IGFII*, *BAX* gene promote apoptosis [137, 138]. Germline mutations in one of the MMR genes mentioned above increase the possibility of a hereditary type of CRCs known as hereditary non-polyposis colorectal cancer (Lynch syndrome). The majority of known germline mutations occur in *MLH1* and *MSH2* [139]. In contrast, transcriptional silencing of *MLH1* gene caused by promoter hypermethylation, is responsible for more than 75% of sporadic CRCs with MSI, which arise commonly via serrated neoplasia pathway (figure 1.4) [140]. For the identification of MSI the National Cancer Institute (NCI) recommended a Bethesda panel of five microsatellite loci which consist of two mononucleotide repeats (BAT25 and BAT26) and three dinucleotide repeats (D5S346, D2S123, and D17S250) [141]. However, additional data over the time also recommended some alternative markers in the reference panel of microsatellite markers. Using the Bethesda panel, CRC sample with 2 or more unstable loci are designated as a high-frequency microsatellite instability (MSI-H) whereas presence of instability at one locus in a sample is considered as a low-frequency microsatellite instability as MSI-L. Moreover, absence of unstable loci in a sample is classified as microsatellite stable MSS [142]. Studies by Halford et al., [143] highlighted that all or almost all tumours exhibit at least MSI-L if enough microsatellites markers are tested and analyzed. According to Goel et al, [144], most sporadic CRCs with MSI-H are a consequence of promoter hypermethylation of *MLH1* gene and *MGMT* methylation is strongly associated with sporadic MSI-L cancers, though is not restricted to this subset [145]. Cancer with MSI-L is limited to dinucleotide markers and MSI should probably be grouped as MSI-L when 2NCI panel dinucleotide markers are involved [146]. Generally, CRCs with MSI have some distinct features such as occurring in the right/proximal side of colon, poorly differentiated, none or few lymph node involvement, less often distant metastasis, a mucinous phenotype with prominent numbers of tumour-infiltrating lymphocytes. CRC in patients with MSI have a favourable prognosis when

compared with chromosomal instability [2, 147]. Additionally, the study conducted by Kerr, Rachel et al, which collected data from 7475 cases, found that for patients with microsatellite-stable tumours, 5-year disease-free and overall survival were shorter with the addition of bevacizumab, but no significant difference in outcome (disease-free or overall survival) with the addition of bevacizumab to capecitabine in patients with microsatellite-unstable tumours [55]. Through the MSI pathway, CRC progression is enhanced by a rapid mutation accumulation especially in coding repetitive sequences of target genes with key cellular roles, such as genes related with DNA repair, apoptosis, signal transduction, and cell cycle. Target genes of MSI are involved in various cellular functions including (*BRAF*, *TGFBRII*, *CTNNB1*, *RNF43*, *ZNRF3*, *BAX, POLE* and *POLD1*) are described below.

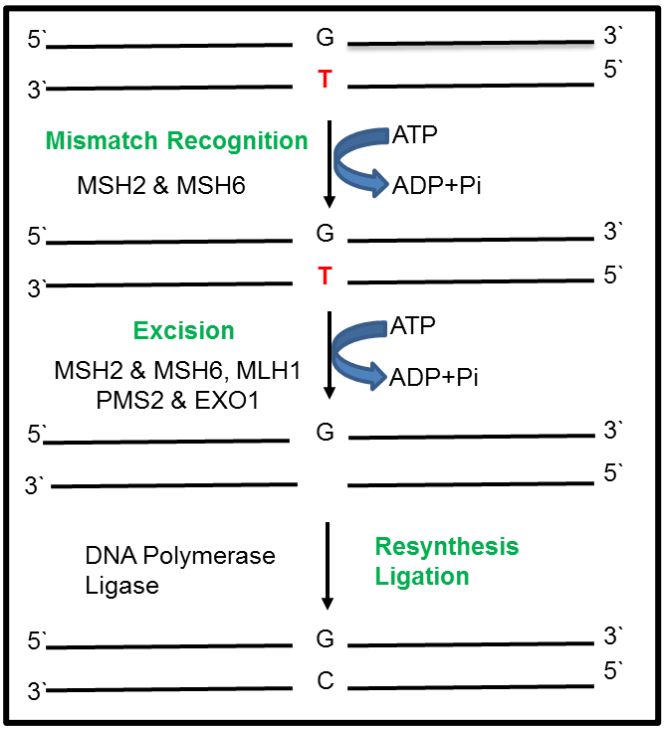

Figure 1.3: Mechanism of mismatch repair (MMR) in eukaryotic, updated from Christmann et al, [148] and Martin et al, [149]

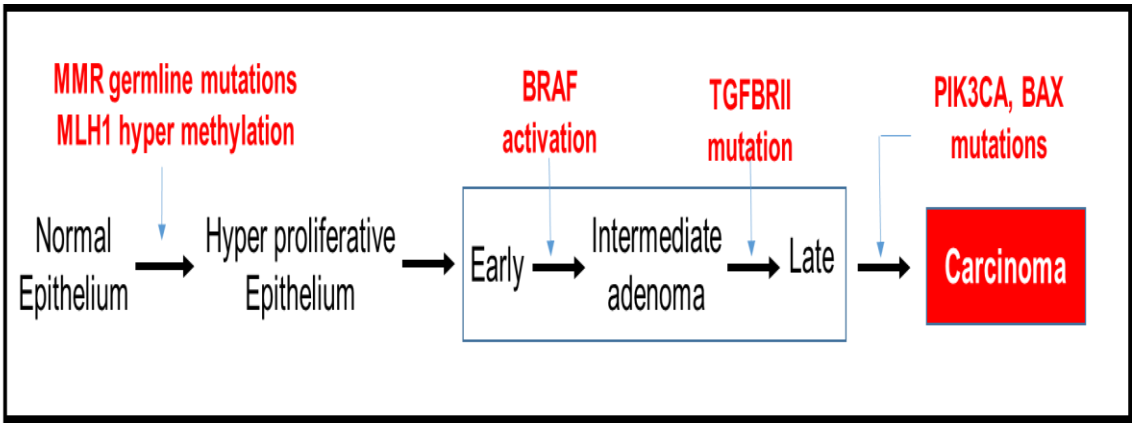

Figure 1.4: The microsatellite instability pathway MSI. It is characterized by early acquisition of MMR gene mutations, frequent activating mutations of BRAF oncogenes at the early adenoma stage and TGFBRII mutations in late adenomas and are not associated with 18qLOH or TP53 mutations. Sporadic MSI cancers arise commonly via the serrated neoplasia pathway. (updated from Mudassar et al, [16])

1.3.2.1 BRAF

RAF is a MAP kinase kinase kinase (MAP3K) protein, which is phosphorylated and stimulated by activated RAS, and appears as three isoforms *ARAF*, *BRAF*, and *CRAF*. *BRAF* is a stronger inducer of MEK phosphorylation when compared to *ARAF*, and only requires the stimulatory signals that activate RAS-GTP for its activation. However, *ARAF* and *CRAF* proteins require the presence of RAS-GTP on the cell membrane for their activation [150]. The *BRAF* gene encodes a protein (serine–threonine protein kinase) which is part of the RAS/RAF/ERK/MAPK pathway that regulates different cellular processes [151]. *BRAF* encodes a 766 amino acid protein and is located on the chromosome 7. It is composed of 18 coding exons (NCBI). Several studies have reported that *ARAF* and *CRAF* mutations occur rarely in cancers while *BRAF* mutations appear in approximately 7% of all cancers [45]. The frequency of *BRAF* mutations has been reported as 5-15% of CRC [152, 153]. In CRCs, *BRAF* mutations commonly appear to be in exon 11 and 15 [45]. The single base change at the codon 600 (V600E), which leads to change from glutamic acid to valine, represents the most common type of *BRAF* mutation in CRC [24].

BRAF with this mutation has a greater kinase activity than wild type *BRAF* or *BRAF* with other mutations. Moreover, this change has a role in enhancing

promoter methylation of *MLH1* (76). According to Pritchard and Grady [24] and Deng et al., [153] *BRAF* mutations more frequently occurred in sporadic CRCs with microsatellite instability MSI than in CRCs with microsatellite stable MSS. Goel and Balaguer[48] and Li et al., [152] demonstrated that *BRAF* mutations are completely absent in MSI-H cancers arising in Lynch syndrome CRCs but are very frequent in sporadic CRCs with MSI-H. Furthermore, other studies Pritchard and Grady [24] and Markowitz and Bertagnolli [154], highlighted that *BRAF* mutations are also present in a high frequency in serrated adenomas, hyperplastic polyps and proximal colon cancers, mainly in those tightly linked to the CpG island methylator phenotype (CIMP) phenotype, while it is rare in conventional adenomatous polyps [48, 154]. As *BRAF* may represent another downstream effector of the EGFR pathway, *BRAF* mutations might have a role in reducing the effect of anti-EGFR monoclonal antibodies treatment in CRCs such as cetuximab or panitumumab [24, 155]. However, anti-BRAF drugs such as vemurafenib (a selective $BRAF^{V600E}$ inhibitor) , which slow the growth of tumours in some people whose metastatic melanoma has a *BRAF* mutant [156], did not show meaningful clinical activity in patients with BRAF gene change metastatic CRC [157].

1.3.2.2 CTNNB1

The CTNNB1 gene, which is located on the short arm of chromosome 3, encompasses 23.2kb of DNA and encodes the 781 amino acid β-catenin protein. β-catenin is an E-cadherin binding protein that regulates cell-cell adhesion [158]. and is the major mediator of the *Wnt*-signalling pathway, which is important in embryogenesis and cancer [159]. Activated *Wnt* signalling or a disruption of the regulatory complex leads to release of cytosolic β-catenin with subsequent nuclear translocation, and TCF/LEF1 family-mediated transcriptional activation, resulting in induces cell cycle progression by induction of target genes such as *cyclin D1 [160]* and *c-myc* [161] and also genes necessary for invasive growth, such as *uPAR* [162], *Fibronectine* [163], *Matrilysin* [164] and *CD44* [165].

APC is a critical component in the destruction of β-catenin and, inactivating mutations in the *APC* gene lead to a cytoplasmic and subsequent nuclear accumulation of β-catenin [166]. Alternatively, in some cases without *APC* mutations, cellular accumulation of β-catenin is acquired because of stabilizing mutations in the β-catenin gene [167] that involve missense mutation or, occasionally, deletion of exon 3 serine/threonine residues [168] Mutations in *β-catenin* gene are more common in adenoma (12.5%) while, uncommon in CRC (~1.4%) [169]. *APC* and exon 3 *β-catenin* mutations are mutually exclusive events which inhibit serine/threonine phosphorylation of amino acid residues, thus impairing *β-catenin* degradation [167]. Interestingly, several studies reported *β-catenin* mutations have been associated with microsatellite instability (MSI) in colorectal carcinomas [170] also found in hereditary non-polyposis colorectal cancer (Lynch syndrome) [171]. However, other studies reported that MSI status and beta-catenin overexpression represent two different pathways to colorectal tumorigenesis [172, 173].

1.3.2.3 RNF43 and ZNF43

RNF43 and *ZNRF3* are transmembrane E3 ubiquitin-protein ligase proteins which negatively regulate the *Wnt* signalling pathway by mediating the ubiquitination, endocytosis and degradation of *Wnt* receptor complex components Frizzled. This suggests that these gene products would act as tumour suppressors [174]. RNF43/ZNRF3-induced ubiquitination is neutralized by binding R-spondin (ligand) to LGR (receptor) which interacts with RNF43/ZNRF3, and consequently membrane disappearance of the complete LGR/RNF43-ZNRF3 complex. In the absence of E3 ligases, Frizzled receptors would accumulate at the membrane and enhances *Wnt* signalling [175]. Thus, the *R-spondin-ZNRF3/RNF43* signalling is an important system to fine-tune *Wnt/β-catenin* signalling and control self-renewal and differentiation of intestinal stem cells. The inhibitory function of *ZNRF3* and *RNF43* in *Wnt* signalling is supported by mouse genetic studies [176]. Koo et al. [177] showed that double knockout of *ZNRF3* and *RNF43* in mouse intestine induce a marked expansion of the

intestinal stem cell zone, resulting in intestinal adenomas with strong expression of β-catenin. Recently a series of experiments revealed that *RNF43* and *ZNRF3* were frequently mutated in CRC with MSI but not with MSS. The frequency of *RNF43* mutations has been reported as 33-87% of CRC with MSI, whereas *ZNRF3* was mutated less frequently than *RNF43*, occurred in around 30% CRC with MSI. However, mutations of *RNF43* and *ZNRF3* were reported in CRC with MSS at frequencies ranging 1-4.8%. [174, 178-182]. Two hotspot mutations G659fs and R117fs were detected in RNF43, which are close to mononucleotide repeat sequences (i.e. microsatellites) and this may explain why cancers such as CRC, gastric and endometrial cancers with deficient MMR, have frequent *RNF43* mutations [182]

Additionally, mutations of *RNF43* were subsequently identified in a variety of cancers including pancreatic cancer 4%, gastric cancer 16%, ovarian cancer 9% and endometrial cancer 18% and in those tumour types *RNF43* and *ZNRF3* mutations were mutually exclusive with *APC* and *β-catenin* mutations. This may be explained that *RNF43* and *ZNRF3* mutations have an active role in tumorigenesis in the colon [182].

1.3.2.4 TGFBRII

Transforming growth factor beta (TGFβ) superfamily is a group of multifunctional proteins that are involved in the control of several biological processes including cell proliferation, differentiation, migration and apoptosis [183] through binding to TGFβ receptors. Three types of TGFβ receptors (TGFBR I, TGFBR II, TGFBR III) are identified in most cells [184]. TGFβ exerts its effects upon binding of the ligand TGF-ß1 to the transmembrane TGFBRII receptor, which causes the recruitment of the TGFBR1 with subsequent activation of the receptor complex. Activated TGFBR1 then phosphorylates two downstream transcription factors *SMAD2* and *SMAD3* [185-187]. These phosphorylated SMAD proteins form a heterotrimeric complex with the common mediator Smad, *SMAD4*. This trimeric complex translocate into the nucleus and regulates the expression of a variety of genes like *SMAD7* [188] or *SERPINE* [189]. Additionally, TGFβ can activate SMAD-

independent pathways such as phosphatidylinositol 3-kinase/ Akt pathway, Ras/MAPK pathway and JNK pathway [190] .

Members of the TGFβ signalling pathway are common targets for mutation and altered cellular signalling pathways in colon cancers [191]. Mutational inactivation of *TGFβRII*, which is the most common mechanism for deregulating TGFβ signalling has been identified in 90% of MSI colon cancers [192]. In colon cancer, however, early data implicated *TGFβRII* as a tumour suppressor gene although recent in vitro data have suggested that *TGFβRII* has contextual effect in carcinogenesis and may be both tumour-promoting and tumour-suppressing [193-196]. In early stages of carcinogenesis, *TGFβRII* act as a tumour suppressor gene, but in advanced stages it accelerates tumour progression by suppressing tumour cell death, immune repression and also induces tumour progression, invasion, and metastasis through enhancing epithelial to mesenchymal transitions (EMT) [197]. Regarding the clinical utility of *TGFβRII* in the management of CRC, Watanabe demonstrated that in stage III colon cancer mutation in *TGFβRII* is slightly related with improved 5-year overall survival. The rate of disease-free survival at 5 years, among stage III colon cancer patients who had both *TGFβRII* mutation and MSI was 79% as compared to 40% among those whose tumours had high levels of MSI with no mutation of the *TGFβRII* gene [198].

1.3.2.5 BAX

BAX (BCL2-Associated X Protein) is a Protein which plays an important role in regulating the mitochondrial apoptotic pathways and contributes in performing p53-mediated apoptosis. The expression and transcription of *BAX* is regulated by the tumour suppressor *TP53* although Bax is also able to promote apoptosis in a *TP53*-deficient background [199]. The protein encoded by this gene belongs to the *BCL2* protein family, which is categorized into two functional subtypes: anti-apoptotic and pro-apoptotic. In response to the induction of wild-type *TP53*, *BAX* which is localized in the cytoplasm, is translocated to mitochondria, where it causes cell death by inducing the formation of ion-permeable pores that disrupt

the mitochondrial membrane barrier [200, 201]. As a result cytochrome *c* and *Smac/DIABLO* are released from mitochondria, which helps caspase activation followed by nuclear fragmentation [202]. *BAX* also participates in the apoptotic response by forming a heterodimer with anti-apoptotic proteins of the BCL2 family via its BH3 domain and inhibiting their functions [203, 204]. In addition, the *BAX* gene serves as an essential factor in protecting cells from DNA damage, thus any damage to the *BAX* gene can contribute to tumour progression by interrupting the apoptotic cascade, [199]. Mutations in the coding mononucleotide repeats of *BAX* and other genes such as (*TGFBRII* and *IGF2R*) have been found in MSI-high CRC [205-207]. A stretch of eight consecutive G residues (poly G8) in the third exon of the *Bax* gene has been described as a frequent site for frameshift mutations in cancers with a MSI phenotype, including 51% (21/41) of hereditary nonpolyposis colorectal cancers, 41% (9/ 22) of other CRC, and 33% (5/15) of gastric cancers [199, 208]. Although several studies have been conducted to assess the role of *BAX* in CRC, the prognostic role of *BAX* mononucleotide mutation still remains uncertain in MSI-high tumours [209]. However, the mutations that occur in *BAX* at a late stage of tumourigenesis are independent prognostic indicators of poor survival [210]. In addition, Zhang et al. [211] demonstrated that the absence of *BAX* entirely abolished the apoptotic response to the chemo preventive agent sulindac and other nonsteroidal anti-inflammatory drugs (NSAIDs) as well as contributing to the development of resistance to anticancer treatments in which cytotoxicity involves apoptosis, and thus might be regarded as a prognostic factor.

1.3.2.6 Epsilon and Delta Polymerases

Nowadays, in the mammalian cell more than 14 polymerases have been detected (Table 1:1) [212], in particular, four members of family β-DNA polymerases including pol α, pol δ, pol ε and pol ζ are considered to play a role in DNA replication and repair [213]. The primers necessary for synthesis of both leading and lagging strands during DNA replication are synthesized by Pol α [214], and pol ζ is engaged in translesion DNA synthesis [213], whereas both pol δ, and pol

ε have a role in the synthesis of lagging and leading strand with a high accuracy in nucleotide integration into the DNA strand with error rates of 10^{4}-10^{-5}. As well, both pol δ, and pol ε raise replication fidelity ~100-fold by displaying DNA repair function [215]. The polymerase domains of *POLE* and *POLD1* are able to recognize correct and incorrect deoxynucleotides (dNTPs). Errors are corrected primarily by the intrinsic proofreading exonuclease (EXO) domain present in each polymerase. Errors that escape proofreading are fixed by the MMR system [216].

Table 1. 1: Summary of eukaryotic DNA polymerase functions [217]

Polymerase	Cellular function
α (alpha)	Chromosome replication (initiation, Okazaki fragment priming) Double-strand break repair
δ (delta)	Chromosome replication (elongation) Nucleotide-excision repair, Base-excison repair, Mismatch repair, Double-strand break repair
ε (epsilon)	Chromosome replication (elongation) Nucleotide-excision repair, Break-excision repair, Mismatch repair, Double-strand break repair
ζ (zeta)	Translesion synthesis
γ (gamma)	Mitochondrial DNA replication
Θ (theta)	DNA repair
β (beta)	Base-excision repair
ʎ (lambda)	Base-excision repair
μ (mu)	Nonhomologous end joining
σ (sigma)	Sister chromatin cohesion
Telomerase	Telomere maintenance
η (eta)	Translesion synthesis
ι (iota)	Translesion synthesis
κ (kappa)	Translesion synthesis

Epsilon polymerase is encoded by the *POLE* gene located at chromosome 12 and composed of 2286 AA (protein length) while delta polymerase encoded by *POLD1* gene located at chromosome 19 and composed of 1107AA (protein

length) (http://www.uniprot.org/uniprot/P28340). The pol δ, and pol ε have exonuclease proofreading domain in their N-terminal regions, these domains are responsible for DNA repair function of both genes [218]. Three highly conserved amino acid motifs (Exos I, II, and III) represents exonuclease active site which provide proof reading activity in both pol δ, and pol ε [219]. Consistent with their leading- and lagging-strand functions, pol δ, and pol ε proofread opposite DNA strands [214]. In the past, several studies have shown that in eukaryotic cells both polymerases are engaged in different repair pathways. For example, the repair synthesis step in nucleotide excision repair NER performed by DNA polymerase delta and epsilon depends on the DNA replication factors including the single stranded DNA binding protein replication protein A (RPA), replication factor C (RFC), *proliferating cell nuclear antigen (*PCNA) [220, 221]. In addition, in long-patch base excision repair (BER), the gap created by removing nucleotides during repair processes is filled by DNA polymerase delta and epsilon together with PCNA and RF-C [148, 222]. Moreover, Hsieh, and Yamane [223] and Prindle, and Loeb, [224] underscored the role of polδ as a repair polymerase in the gap filling step of MMR with the aid of PCNA. On the other hand, Pospiech, and Syväoja [225] highlighted that Pol ε and the 5´-specific FEN-1, and possibly PCNA are also involved in nonhomologous end-joining (NHEJ) activity. In yeast the incidence of mutation is raised dramatically by inactivation of 3'→5' proofreading exonuclease activity in either Pol ε or Pol δ [214, 218]. Furthermore, Goldsby et al., [226] and Lange et al., [221] demonstrated that the deficient exonuclease activity of Pol ε and Pol δ in mice increases the possibility of mutagenesis and drives carcinogenesis resulting in epithelial, mesenchymal and teratomatous types of cancer. Two recent studies conducted by Briggs and Yoshida et al., found occurrence of somatic and germline mutations in the proofreading exonuclease domains (ED-) in both *POLE* and *POLD1* genes. In the *POLE* gene, germline mutation encoding the L424V variant, in some individuals, predisposes to a dominant syndrome of multiple colorectal adenoma similar to that produces as a result of defect in base excision repair (*MUTYH* associated polyposis). Conversely, in other individuals, the L424V variant leads to large adenomas which

are similar to Lynch syndrome. One study reported that ‘’Although the action of genetic modification leads to the phenotypic variation caused by L424V, it might have other causes, for instance, different environments, random mutation, reduction in risk due to therapeutic surgery or other tumour prophylaxis, under-reporting of adenomas, especially in older family members, or premature death’’ [227].The study also demonstrated that colorectal phenotype produced as a result of the *POLD1* S478N variant cannot be distinguished from that produced due to *POLE* L424V variant. They also have shown that *ED-POLE* and *ED-POLD1* mutations characterize a group of hypermutated, microsatellite stable (MSS) colorectal tumours [227, 228]. Additionally, the Cancer Genome Atlas (TCGA) exome sequencing project, which published the result of full genomic profiling of 224 CRC samples, confirmed many mutations in *POLE* exonuclease domain in a small subset of MSS CRC that are nevertheless hypermutated. These findings suggest that CRC associated with *POLE* mutations may represent a distinct CRC subtype [229].

1.3.3 Microsatellite and chromosomal stable pathway

In the past, researchers have assumed that CRC occurs as a result of genetic instability in the form of either microsatellite Instability or chromosomal instability. However, recent data have described CRCs without either MSI or CIN and it may represent a potentially distinctive form of the disease known as Microsatellite and Chromosome Stable (MACS) CRC. Regarding the proportion of MACS in sporadic CRCs, studies showed different proportion ranging from 17% to 45% [230-233]. In comparison to other genetic pathways in CRC, MACS-CRCs are poorly described and their clinicopathological features are also not well described. According to the data published before by different studies, mixed information has been reported about the MACS-CRCs group. For example, the study conducted by Andrew, et al., [234] demonstrated that MACS are more probably to be left than right side while other studies found a significant association of diploid tumors with the right side of the colon. In contrast, Hawkins et al. [230] showed that MACS-CRCs not associated with tumour sidedness. With regard to the prognosis

of MACS-CRCs, Sanjay et al., [232] revealed that MACS-CRCs have better survival prospects compared to CIN-CRCs, other study has presented contrary results and showing that MACS-CRCs have worse prognosis and represent an aggressive subset of CRC because distant metastasis is more common in this group of CRCs compared to MSI and aneuploid tumors [230]. In the contrary to other groups, MACS-CRCs are equally likely to occur in males as females and not associated with lymphocytic infiltration within the tumour. In an epigenetic aspect, MACS tumours have a high rate CpG island methylator phenotype similar to that in MSI tumours and higher than that in CIN tumours. Moreover, MACS tumours have a lower rate of loss of *APC* mutation than CIN tumours and less loss of *MLH1* or *BAX* protein than MSI tumours [233]. Tang, et al., [231] highlighted that MACS tumours have a reduced rate of *TP53* mutations (10%).

1.4 Genetic pathways defined by epigenetic change

Epigenetic changes, which are usually caused by DNA methylation or histone modifications at the 5′-CG-3′ (CpG) dinucleotide, cause alterations in gene expression and function without changing the DNA sequence of that particular gene [4]. The finding of aberrant methylation of promoter region CpG islands in sporadic MSI colon cancers showed the role of epigenetic changes as potential pathogenetic alterations in cancer [235-237]. CpG islands are regions rich in CpG dinucleotide located at the promoter area of about 50% of human genes. Normally gene expression can be regulated by methylation of promoter CpG islands. Any defect or alteration in the rate of DNA methylation seems to promote carcinogenesis, because it is associated with the transcriptional silencing of genes which control some important cellular functions such as tumour suppression, cell cycle control, DNA repair or invasion [147]. Regardless of the transcriptional status of the gene, nearly all CpG island regions are not methylated in normal colon mucosa. However, In CRC promoter CpG island hypermethylation is frequent and affects gene regulating the WNT pathway, the NOTCH pathway, *PI3K* signaling, *TP53*, and IGF signaling. In addition, genes involved in other pathways which regulate cell cycle regulation, DNA repair,

apoptosis, angiogenesis, invasion and metastasis [238] may be silenced by promoter methylation. The CpG island methylator phenotype (CIMP) is a discrete phenotype in CRC which has specific clinicopathological and molecular features [37]. Some authors classified CIMP into three different subtypes (No-CIMP, CIMP-Low, and CIMP-High) depending on the presence of methylated marker loci [239, 240]. On the other hand, another group classified CIMP into three separate subtypes CIMP1, CIMP2 and CIMP negative as a consequence of unsupervised cluster analysis of a large panel of methylation markers [241]. Regardless of MSI status, CRCs with CIMP-high have a distinct molecular profile and clinicopathological features including proximal tumour location in colon, female sex, poor differentiation, high *BRAF* and low *TP53* mutation rates [242, 243]. However, CRCs with CIMP-low are associated with frequent *KRAS* mutation (92%), male sex, smaller number of genes are affected by a denser pattern of methylation and positively associated with MSI-low [173, 244, 245]. In addition, loss of heterozygosity 18q can be seen as remarkable point which is positively associated with CIMP-negative while it is negative in CIMP-low [246]. Generally CIMP phenotype tumours are assumed to develop along an alternate serrated adenoma pathway in which hypermethylation rather than mutation is used to inactivate tumour suppressor genes [247]. Collectively in colorectal cancer, the above data might be used to construct molecular differences between CIMP-low, CIMP-high and CIMP-negative.

Regarding the classification of CRCs, it is not reliable to classify CRC based on CIN because methods which are used to analyse CIN are not uniform. Therefore, Jeremy Jass proposed a new method for CRCs classification based on MSI and CIMP [248]. In addition, CRCs can be molecularly classified on the basis of some other features including *Kras mutation status, Braf mutation status,* and methylation status of O6-methylguanine DNA methyltransferase *(MGMT)* [147]. Jeremy Jass classified CRCs into five molecular subtypes: CIMP high/MSI high/ *BRAF*+ and MLH1 methylation (12%), CIMP high/MSI low or MSS/ *BRAF*+/ methylation of multiple genes (8%), both of them originating from serrated adenoma. CIMP low/MSI low or MSS/ CIN+, *KRAS*+/*MGMT* methylation+ (20%)

originates in tubular, tubulovillus, or serrated adenomas. CIMP- /MSS/ CIN+ (57%) originates in traditional adenoma. Finally, CRCs arising in lynch syndrome (3% of CRC cases) show CIMP-/MSI high/ negative for *BRAF* mutation [173]. Although, Zlobec et al., [13] study confirmed the Jass model with a different percentage of each subtype, it also shows other subtypes of CRCs: (1) CIMP-H/*BRAF* wild type, (2) CIMP-/ MSI-H/ *BRAF*+, (3) CIMP-low.

1.5 Novel CRC risk loci

The paper published by Genome Wide Association and Sequencing studies on CRC [9] show results reported by three meta-analysis studies that found some additional risk loci. The first meta-analysis conducted by Richard et al., [249], detected four novel CRC risk loci at 14q22 (*BMP4*), 16q22 (*CDH1*), 19q13 (*RHPN2*) and 20p12. Bone morphogenetic protein (BMP) is a ligand of the TGFβ that plays an important role in CRC by suppressing *Wnt*- β-catenin signaling and blocks intestinal stem cell renewal. *CDH1* encodes the protein cadherin-1 (also known as E-cadherin), defect of *CDH1 by inactivating* mutation or promoter methylation, leading to increased activity of the β-catenin–TCF transcription. *RHPN2* encodes a Rho GTPase, which has a role in the regulation of the actin cytoskeleton and cell motility.

The second meta-analysis conducted by Richard et al. [250], identified further four genetic loci at 1q41, 3q26 (*MYNN*), 12q13 (*ATF1/DIP2B*) and 20q13 (*LAMA5*). *ATF1* is the 3' partner in the recurrent translocations with *EWSR1* that contributes to the development of soft-tissue sarcomas. *LAMA5* is required for the production of noggin, which plays a role in inactivating signalling proteins in the TGFβ pathway.

Finally, the third meta-analysis was performed by conglomerating data from five GWAS. Malcolm et al. [251] identified three risk loci at 6p21 (*CDKN1A*), 11q13 (*POLD3*) and Xp22 (*SHROOM2*). *CDKN1A* gene encodes p21, which has a role in regulating multiple tumour suppressor pathways (e.g. p53, TGFβ and BMP signalling) and also interferes with DNA polymerase δ activity to inhibit DNA replication and alter DNA repair. *POLD3* gene encodes a component in the DNA

polymerase- complex, which participates in base excision repair, DNA mismatch, and suppression of homologous recombination. *SHROOM2* gene has a role in morphogenesis during epithelial and endothelial tissue development.

1.6 Consensus Molecular Subtypes

Recently and after analysing 18 different CRC gene expression datasets, including data from TCGA in conjunction with molecular data on mutations and SCNAs for a subset of the samples, an international consortium [252] described four consensus molecular subtypes (CMS1-4). CMS1 (microsatellite instability immune, 14%) CRC, which is equated with MSI CRC subgroup, are hypermutated, MSI and display a widespread hypermethylation status with frequent BRAF mutations, has low prevalence of somatic copy number alterations (SCNAs) and strong immune activation associated with a diffuse immune infiltrate, mainly composed of NK cell, TH1 cell and cytotoxic T cells. Moreover, CMS1 tumours were frequently diagnosed in females, with right-sided lesions and presented with higher histopathological grade and has a very poor survival rate after relapse. In contrast, CMS2-4 display high chromosomal instability and all characterised by high levels of SCNAs. CMS2 (Canonical, 37%) CRC display epithelial differentiation, marked WNT and MYC signalling activation and frequent copy number losses in tumour suppressor genes and gains in oncogenes than in the other subtypes. CMS2 tumours were mainly left-sided and has superior survival rate after relapse compared with the other subtypes. CMS3 (metabolic, 13 %) CRC subtype display higher prevalence of CIMP-low, fewer SCNAs, more hypermutated, which overlapped with MSI status, than CMS2 and CMS4, along with frequent KRAS activating mutations and a slightly higher prevalence of CIMP-low. It also displays epithelial differentiation and evident metabolic dysregulation. The CMS4 subtype (mesenchymal, 23 %) CRC subtype showed clear increase expression of genes implicated in epithelial mesenchymal transition, evidence of prominent transforming growth factor-β activation, with gene expression profile implicated in angiogenesis, stromal invasion, complement inflammatory system, matrix remodelling and mesenchymal

activation. CMS4 tumours, which is tend to be diagnosed at more advanced stages (III and IV), has a worse overall survival than patients of the other groups. Finally, there are some samples with mixed features (13%) possibly represent a transition phenotype or intratumoral heterogeneity.

1.7 Gene interactions in cancer

Obviously, cancerous cells arise from normal cells as a result of the interaction of altered gene activities not due to the independent action of these genes [253]. Gene interactions are the phenomenon of the modification of the effects of one gene by the activity of one or more other genes. It is also can be described as epistasis and then further sub-classified as either enhancement or suppressive effects [254]. The term epistasis (Bateson,1909) was firstly used to describe the ability of a gene at one locus to “mask” the mutational influence of a gene at another locus [255]. As an example of suppressive interaction; combination of mutations in two genes, which have effect on a specific phenotype such as cell fitness, produces a less profound effect than would be expected from the combination of individual gene effects. In contrast, genetic enhancement interaction produces greater effects when both genes are mutated than would be predicted by a simple combination of individual gene effects [253]. Moreover, cell death or a significant reduction in the fitness of cells might occur when the gene interactions are extreme.

Examples of enhancement interaction are the interaction between the transcription factor gene *MYC* and the GTPase gene RAS, which contributing to the processes of transformation and immortalization [254], as well as increasing β-catenin activity by *KRAS* through inhibition of β-catenin degradation by *GSK3β* [256]. Whereas a relationship between *TP53* gene (gatekeeper) and *BRCA1* or *BRCA2* tumour suppressor gene (caretaker) mutations, represent an example of suppressor interactions, in which *BRCA1* and *BRCA2* protect the genome against tumorigenic effects of DNA damage, while *TP53* induce cell death or cell-cycle arrest of cells that have lost caretaker function [253]. Another study found that simultaneous alteration of the both *RAS* and *TP53* genes caused a synergistic

alteration in the expression of a set of 95 genes known as Cooperation Response Genes (CRGs), which are involved in the metabolism, apoptosis, cell signalling, regulation of transcription, transport and adhesion [257]. Among 24 CRGs that were tested, 14 have a significant contribution to the malignant pathogenesis, whereas non-CRGs showed minor contribution to the process. This clearly shows that cancer is a cooperative process dependent on the alteration of different genes [257]. For example, combine effort of STAT3 and NF-κB enhance cancer hallmarks by controlling the expression of different genes including *cyclin D1, c-Myc*, *BCL2* and *BCL-XL* and others [258]. Moreover, the activity of cell cycle inhibitor *p27* is repressed by the cooperation of two main adenoma-carcinoma sequence proteins: *APC* and *SMAD4* cooperate [259].

1.8 Molecular features of metastasis in CRC

Haematogenous dissemination of tumour cells from the primary site to distant areas and development of secondary tumours represents the final stage of CRC. Although the exact molecular mechanisms which promote metastatic spread largely remain unknown, definitive genetic and epigenetic change is required for progressing each step in metastasis [260]. In CRCs tumour, cells metastasize to several vital organs and develop metastasis. For example, liver metastasis occurs in 30-70%, lung metastasis occurs in 10-20% [261], and brain metastasis occurs in 3-9% [262]. Though the main reason for targeting liver as one of the most significant site for metastasis in CRCs not elucidated yet, partially the portal drainage of the gastrointestinal tract is responsible. However, different molecular mechanisms are certain to underlie whether CRC is likely to metastasize into the liver [263]. Colorectal liver metastasis is formalized into a series of discrete stages, which includes escape of cells from the primary tumour and intravasation into vascular or lymphatic system. The cells then travel through the portal vein system and escape from host defence mechanism, which finally extravasates into the liver [264]. Metastasis from CRC, like primary CRCs show molecular and genetics diversity and the site and route of metastasis might be indicated by the type of genetic alterations [265]. There are different studies that have compared

the genetic differences between primary colon cancer and their metastasis. Studies conducted by Daniele et al. [266], Salvatore et al. [267] and Santini et al. [268] have shown a high concordance of *KRAS* and *BRAF* status between primary and metastatic CRCs lesions. This high grade of concordance confirms that *KRAS* and BRAF mutation represents a very early mutational step in CRC pathogenesis and plays a central role in tumour progression. Whereas Yu-Fen et al. [269] and Kastrinakis et al. [270] demonstrated that there is a big difference in the frequency of mutations targeting *PIK3CA* and *TP53* between primary and metastatic CRCs and these results confirms that *PIK3CA* and *TP53* mutations represent late mutational steps in CRC pathogenesis and support the notion believed that PI3K and *TP53* play a great role in metastasis. Regarding chromosomal abnormalities, several studies have shown that chromosome 20q gain is more frequent in metastatic CRCs compared to primary CRCs [271-273], the high frequency of chromosome 20q in metastatic CRCs also indicated that 20q gain plays a major role in tumour progression and metastasis in CRCs. In addition, several other proteins participate in progression and metastasis in CRCs. For example, E-cadherin is a glycoprotein that regulates cell-cell adhesion by binding to other extracellular cadherin's and sticks to the cytoskeleton via β-catenins [274, 275]. Down regulation of E-cadherin and release of β-catenins into cytoplasm and entering into the nucleus involved in a process known as EMT, which is important to promote migration and invasion of stroma. In CRC, Kang and Massagué [276] and Raftopoulos et al., [277] showed that a low expression of E-cadherin and high expression of β-catenin may be used as an indicator of metastatic potential and unfavourable prognosis.

Matrix metalloproteinases (MMP), are a group of proteolytic enzymes characterized by their ability to degrade basement membrane and extracellular matrix as well causing detachment of cell from the primary site and progression of EMT process [278, 279]. Several studies have reported that in CRCs over expression of MMPs play crucial roles in cell motility, gain the capacity of cancer cell to invade and promote metastasis to distant area particularly in liver metastasis [280]. Furthermore, Epidermal growth factor receptor (EGFR), a cell-

surface receptor when bound to the ligand such as epidermal growth factor, activates downstream RAS-RAF-MAPK and PI3K signalling cascades, SRC/FAK, Phospholipase-C and STAT (signal transducer and activators of transcription protein). Alteration of the EGFR signaling pathway promotes angiogenesis by activating VEGF, promotes invasiveness by activating serine protease and activating RAS-RAF-MAPK pathway and the PI3K-Akt pathway which have prominent roles in carcinogenesis [260]. Overexpression of EGFR is another phenomenon which has clearly been seen in 72-82% of metastatic CRCs and some studies have reported that expression of EGFR in CRC is correlated with aggressiveness and metastasis ability [280]. Additionally, VEGF a proangiogenic factors play a crucial role in angiogenesis, which is important to transfer oxygen and nutrient to neoplasm [260]. In CRCs, it has been reported that predominant expression of VEGF factor is an indicator of progression and formation of liver metastasis [280].

1.9 MicroRNA biomarkers in CRC

1.9.1 microRNA

MicroRNAs (miRNAs) are evolutionary conserved small (18-25 nucleotides long) non-coding RNA sequences encoded by eukaryotic genomic DNA. They post-transcriptionally regulate gene expression via base-pairing with complementary sequences in a target messenger RNA (mRNA) leading to translational repression or degradation of target mRNAs [281]. Originally, in 1993, Lee and co-workers described this type of regulation in genes that control timing of larval development in *Caenorhabditis Elegans.* It has been estimated that miRNAs regulate 30 to 60% of protein-coding genes [282] and are involved in the regulation of genes related to many biological processes, such as cell proliferation, differentiation, apoptosis and metabolism, However, to establish and maintain the differentiated status of many cell types are the main function of miRNAs [283]. Today more than 2042 miRNAs have been listed in the miRNA registry called miRBase, which constituting 1 to 3% of the genes in the human genome [282, 284].

1.9.1.1 MiRNA biogenesis and function

MiRNA genes are located within exons and intergenic regions but are mostly detected within introns of coding or non-coding genes [283]. microRNA coding sequences either have their own promoters or use the same promoters as protein coding genes and they are mostly transcribed as longer precursors by RNA polymerase II into primary transcripts (pri-miRNAs) of variable length (1 kb- 3 kb) [285]. A ribonucleoprotein complex, Drosha/DGCR8, then cleaves pri-miRNAs resulting in a hairpin intermediate of about 60–100 nucleotides, called (pre-miRNAs) [286], which consists of two miRNA strands complementary to one-another, where one is called 3-p and the other one 5-p. After recognizing and exporting into the cytoplasm by the nuclear export protein, Exportin 5, and its co-factor Ran-GTP, the pre-miRNA structure is further processed into miRNA: miRNA* duplex of about 17-25nt by the RNase III endoribonuclease enzyme Dicer, and its partner, trans-activator RNA-binding protein (TRBP) [287]. This duplex then unwinds into mature miRNA and passenger miRNA (miRNA*), and are discriminated either by 3-p and 5-p, or with one dominantly processed and the recessive one star-labeled (*), which function differently against different target genes [288]. The most abundant strand between 3-p and 5-p, which is transferred into an Ago protein, is called the mature strand whereas the other strand, which is usually degraded, is determined as the star (*)-strand [289]. However, trace miRNA* strands are found to be conserved and have downstream effects on transcription and translation of RNA and DNA and play a great role in cell homeostasis [285]. After incorporating mature miRNA into the RNA-induced silencing complex (RISC), then it binds to the 3-untranslated region (UTR) of the target gene. Depending on the degree of complementarity (imperfect or perfect) with the target DNA sequence, the binding of miRNAs complex with mRNA results in inhibition of protein synthesis (imperfect homology) or mRNA degradation (perfect homology) [290]. The core sequence of 6-7 nucleotides of the miRNAs indicates the specificity of miRNAs. As a consequence, a single miRNA can significantly affect hundreds of genes or a single gene can be targeted by several

different miRNAs [291]. Although mechanistic details of miRNA mediated repression and particularly the effects on mRNA translation are still lacking, miRNAs use different mechanisms to degrade mRNA. Removal of the Poly-A tail from mRNA by a deadenylase complex such as CCR4-NOT is one of them and another is by the miRNA/RISC complex which cleaves and degrades the mRNA transcript [292-294]. Shortening of the poly-A tail leads to the removal of the mRNA 5' cap resulting in 5' to 3' exoribonucleases such as Xrn1 can easily remove mRNAs from the cell (see figure 1-5) [295].

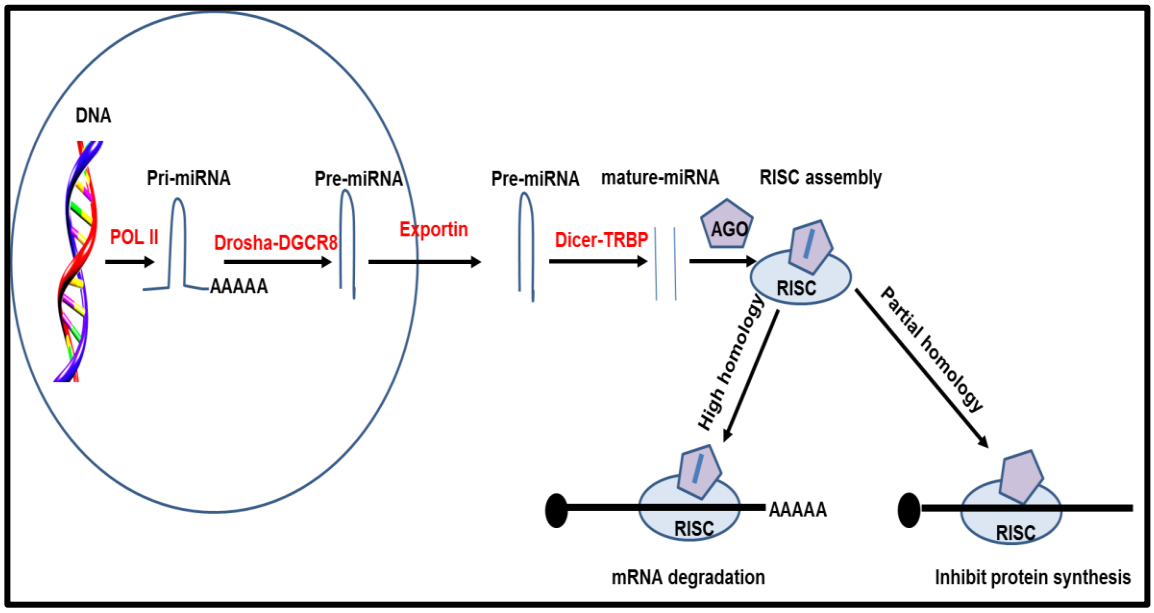

Figure 1. 5: miRNA biogenesis, updated from Shee K., [283]

1.9.1.2 MiRNA in CRC

Over 50% of the miRNAs genes are frequently located in cancer-associated genomic regions and at fragile sites and viral integration sites. In addition, they are also present in common breakpoint regions causing deletions and mutations, Minimal regions of loss of heterozygosity and minimal regions of amplification can causes dysregulation of miRNAs [296, 297]. Deregulation of miRNA expression can influence carcinogenesis in various types of solid tumours including colorectal, lung, breast cancers and leukaemia [286]. MiRNAs can act as either oncogenes or tumour suppressor genes. Upregulation of mature miRNA may

occur due to amplification or transcriptional activation of the miRNA encoding gene, whereas silencing or reduced expression may result from deletion of a particular chromosomal region, epigenetic silencing, or defects in their biogenesis [298].

Many miRNAs are aberrantly expressed in tumour tissue, leading to a potentially faulty regulation of their target mRNAs. After detecting the relationship between miRNA and human cancers for the first time in 2000, others confirmed that miRNAs are consistently and reproducibly altered in CRC compared to normal tissue [299-302]. Several studies have shown that aberrant transcript miRNAs are involved in the pathogenesis of CRC. The first miRNAs detected in CRC by Michael and colleagues were miR-143 and miR-145 which showed that microRNA expression patterns were altered in CRC and other studies demonstrated that these miRNAs (miR-143 and miR-145) have tumour suppressive functions and reduced their expression pattern in CRC [302-304]. Lately more miRNAs, which are relatively over-expressed or down regulated, have been identified that affect many proteins involved in key signalling pathways of CRC.

Different studies have shown contradictory results about the fact that spectrums of dysregulated miRNAs are associated with CRC genesis, progression, and therapeutic response [162-164]. This may be because of tumour location or genetic background including different MSI and CIMP tumours [305]. Additionally, miRNA expression patterns are different between microsatellite instable (MSI) and microsatellite stable (MSS) tumours [306], and this may help classify different phenotypic subgroups of CRC. For example, DNA mismatch repair proteins MLH1, MSH2 and MSH6, which contribute to the MSI phenotype in CRC, are targeted by miR-155. MSH2, which is thought to increase tumour resistance to 5 fluorouracil based therapies, is targeted by miR-21 as well [307].

1.9.1.3 Biomarkers and screening tests of CRC

Generally, a biomarker is defined as any substance, structure or process found in blood, other body fluids, or tissues that could be used as a measurable phenotypic parameter for measuring the normal physiological state, abnormal process, or of a condition or disease [308] In cancer, several molecules or substances (most typically proteins, glycolipids) that might be either secreted or released by cancerous cells or by the body as a response to the presence of cancer, are used as tumour markers for early diagnosis of cancer, identifying the possibility of recurrence of cancer and in monitoring the progress of treatment, efficacy of neoadjuvant therapy, surgery, chemotherapy and adjuvant radiation therapy and follow-up for possible recurrence. Based on biomolecules, cancer biomarkers may be DNA, RNA, microRNA (miRNA), epigenetic changes or antibodies [309]. In CRC, biomarkers are most often used to evaluate the patient's response to cancer treatment or to monitor for a recurrence. For example, carcinoembryonic antigen CEA used during therapy to assess response to treatment or after completion of therapy to monitor for recurrence [310]. Detects haemoglobin in feces which indicates bleeding from GIT, is another markers is an easy, non-invasive, and simple diagnostic test called FOBT, for CRC [311]. Additionally, With the recent progress in understanding the molecular mechanisms of cancer development, the following markers discovered throughout the recent years continue to be closely examined: MSI, 18qLOH, KRAS, BRAF, PTEN, PIK3CA mutations, epigenetic markers etc [312]. In UK, to detect cases of bowel cancer sooner, the NHS offers two types of bowel cancer screening; (1) all men and women aged 60-74 are invited to carry out a faecal occult blood (FOB) test. However, people younger than 60 aren't eligible for the FOB screening test, but can have bowel scope screening. (2) bowel scope screening is a relatively new test to help prevent bowel cancer. It is gradually being rolled out to all men and women in England aged 55. Bowel scope screening uses a thin flexible tube with a tiny camera on the end to look at the large bowel. It can find and remove small growths called polyps from the bowel. Polyps do not usually cause symptoms but some might turn into cancer if they

are not removed. The technical term for bowel scope screening is flexible sigmoidoscopy screening.

1.9.1.4 MiRNAs as diagnostic biomarkers for CRC

Undoubtedly detection of CRC at an early stage at the time of diagnosis provides the best chance for successful treatment and the chance of survival reduces dramatically at later stages with metastases to lymph nodes and other organs. Current strategies such as foecal occult blood tests FOBT and colonoscopies have improved survival rates for CRC and potentially reduce disease mortality, but several factors have impeded compliance and use. For example, FOBT has low sensitivity, specificity and requires dietary restriction [313]. Although colonoscopy is a reliable screening tool, many people are hesitant to have a colonoscopy due to the embarrassment, invasive, expensive, and fear of pain and discomfort that are associated with the procedure.[314]. Another biomarker is carcinoembryonic antigen, but low sensitivity 29.4% and specificity 30-70% limits its clinical use, particularly for early detection of CRC giving many false positive and false negative cases of CRC. Nowadays, new and better screening methods are needed because current biomarkers have low sensitivity and specificity [315]. Recently several studies have shown miRNAs in plasma as possible diagnostic biomarkers. For instance, the study conducted by Ng and colleagues shown that elevation of circulatory level of miRNA92a is a candidate biomarker to discriminate CRC patients from healthy controls with a quite promising sensitivity of 89% and specificity of 70%. Notably, the expression of miR-92 dropped significantly following surgical removal of the tumour, showing the marker was likely coming from the colorectal lesion [316]. Similar studies found that circulating levels of more miRNA candidates were elevated in CRC. For example, miR-141 has been shown to be elevated in metastatic CRC and its expression is associated with poor prognosis [317, 318]. Cheng et al, [318] tested diagnostic value of miRNA29a and miRNA92 in early lesion in the development of CRC and found a significant elevated expression when compared to those in normal control. Moreover, Kanaan et al, [319] highlighted that *miR-21* level was statistically significantly elevated in CRC serum compared with healthy control

subjec*ts, whereas no statistically significant differences were noted in serum* miR-31.

1.9.1.5 MicoRNAs as prognostic and predictive biomarkers for CRC

Due to tumour recurrence after resection and variable clinical responses to adjuvant therapy, there have been recent studies to find more reliable predictive and prognostic biomarkers [320]. The study conducted by Roldo et al, [321] has shown that over-expressed miR-21 correlates with the presence of liver metastasis. Later [322] confirmed that in CRC patients, upregulation of miR-A21 is associated with lymph node positivity and development of distant metastasis. Moreover, Schetter et al, [323] identified that highly expressed miR-21 was associated with poor survival in the CRC patients. This association between upregulation of miR-21 expression and worse survival outcomes in CRC has been validated by other studies conducted by Kulda et al, and Nielsen et al, [324, 325]. Dysregulation of miR-31 is another factor for tumour metastasis in CRC tissues and Bandrés et al, [302] identified that the miR-31 was significantly upregulated in stage IV compared with stage II CRC. After that the association between miR-31 level and pathological stage, including local invasion was confirmed by Wang et al, [326].

Although significant improvements in cancer therapy have been achieved, drug resistance is still a major obstacle in its treatment. Predictive biomarkers provide information to monitor a particular treatment type. Svoboda et al, [327] found that a short-course of capecitabine-based chemoradiotherapy was significantly associated with increased miR-125b and miR-137 expression and higher induction of miR-125b and miR-137 were associated with worse response to the treatment. Later the study done by Ragusa et al, [328] found that over expression of miR-17-3p and downregulation of let-7b and let-7e were potential predictive markers of cetuximab resistance.

1.10 Genetic implication in colorectal cancer treatments

Understanding the genetic mechanisms of CRCs is a key to improve risk prediction, prognostication and treatment. In the past few years, an improvement

has been seen in the CRC treatments and developing more effective targeted therapies. Monoclonal antibodies (MoAbs) targeting the EGFR, such as cetuximab and panitumumab, are one of the therapeutic options for metastatic colorectal cancer, but they are effective only in a small subset of 10% metastatic CRC patients [329]. To detect the main reason, several investigations found that genetic alteration in the downstream components of EGFR have been associated with resistant to anti-EGFR therapies [330]. Nowadays, researchers have found that mutations in several oncogenes seem to inhibit the efficacy of anti-EGFR therapies such as *KRAS*, *BRAF*, *PIK3CA*, *PTEN*, and *NRAS*. Moreover, activation of ERBB2 signaling, amplification of *HER2* [331] and *MET* gene [330] produce resistance to anti-EGFR therapies. Although researchers discovered most causes of resistant to anti-EGFR therapies, reasons for resistance of anti-EGFR therapies in more than one out of four CRCs patients are unknown [329]. Therefore, Identification of the additional resistance mechanisms could help further refine the rest of CRC patients likely to benefit from cetuximab or cetuximab-based combination therapies.

1.11 Hypothesis and Aims of the study

1.11.1 Hypothesis

1- CRC is a complex disease and several genes, which are involved in different pathways, have role in initiation and progression of the disease and raise the strong heterogeneity both between and inside of tumours.

2- Heterogeneity of the CRC increases the genetic discrepancy between CIN and MACS pathways and between the primary CRC and matched metastasis as well.

3- The driver mutations, which are responsible for early clonal sweeps through the adenoma–carcinoma sequence, predominantly present in most of the tumour cells and consequently would be present in the biopsy samples.

4- Small nuclear RNA (e.g. miRNAs) has role in defecting gene expression and ultimately increase risk of CRC through controlling different genes either by suppressing their mRNA or protein.

1.11.2 Aims of the study

The aims of this study were to investigate molecular basis of colorectal cancers via screening mutations of different genes, measuring the expression of the miRNAs, mRNA and proteins expressed in colorectal cancer in order to;

1- Investigate molecular profile of CRC to define clinically or biologically relevant signatures of the disease via screening of mutations in 28 genes using NGS and HRM in the first cohort 83 CRC samples.

2- Within the disease investigate genetic discrepancy by comparing the mutation profiles of CIN and MACS-types tumours in the first cohort samples.

3- Investigate genetic differences between primary CRC and their respective metastasis in the first cohort samples through; screening of mutations using NGS and HRM, measuring expression levels of the six miRNAs and 20 proteins using RT-qPCR and RPPA respectively.

4- Investigate expression of the six miRNA molecules in paired tumour/normal samples from the second cohort 81 CRC patients using RT-qPCR method to see whether miRNAs can serve as diagnostic markers in CRC patients. Additionally, investigate whether miRNA molecules differentially expressed have roles in carcinogenesis through targeting genes involved in several signalling pathways.
5- In addition to the molecular basis of CRC, the study aimed to investigate whether the biopsies are representative of the tumour by comparing mutations in 26 genes of paired biopsy samples and corresponding resection specimens.
6- As sub study, we aimed to investigate cfNAs in CRC to see whether; (a) is it possible to use liquid biopsy to investigate molecular alteration (b) genomic alteration in cfDNA is sufficiently sensitive to be used as a biomarker in clinical medicine.

Chapter Two: Materials and Methods

2.1 Materials

This study was approved by Nottingham Research Ethics Committee Reference number (REC reference C02.310).

2.1.1 Patient samples

2.1.1.1 Cohort one: genetic analysis

One hundred and two Formalin Fixed Paraffin Embedded (FFPE) samples of primary CRC had been previously recovered from the archives of the pathology department of the Nottingham University Hospitals NHS trust. From this group, 48 of the cases had matched samples from secondary deposits (mostly liver), 30 cases had matched normal mucosa were well away from the tumour (in separate different blocks) and, in 30 cases, the matched pre-surgical diagnostic biopsy was available. This spectrum meant that it was possible to perform sub-group analysis on the whole series of cases.

All patients had undergone surgery between 2003 and 2010 at Queen`s Medical Centre (QMC) Nottingham, UK. Cases were selected on the basis of the availability of clinicopathological data and the presence of at least 50% of tumour tissue in the biopsy specimens and tumour block as well, (table 2.1). The haematoxylin and eosin stained slides were reviewed by Dr. Wakkas Fadhil (PhD Student, Department of Pathology, Pathology Research group).

Of these cases, 83 (primary CRC) were found to suitable for genetic analysis (see below). Comparison of the original 102 CRCs and the final data set of 83 CRCs showed a similar profile (table 2.2) thereby confirming that there had been no bias introduced by the removal of the poor-quality cases.

The ploidy and microsatellite status of the first cohort 83 samples, were previously tested by Wakkas Fadhil et al [233]. in brief, ploidy status was examined using flow cytometry, whereas a panel of six markers (BAT25, BAT26, NR21, NR22, NR24 and B-CAT25) was used to evaluate instability at mononucleotide repeat sequences and analysis was performed by PCR following HRM.

Table2. 1: Clinicopathological features of patient's cohort n=102

Variable	Classification	Frequency n (%)
Sex	Male	57 (56%)
	Female	37 (36%)
	unknown	8 (8%)
Age	Median	70 (43-88)
Dukes' stage	A	6 (6%)
	B	29 (28%)
	C	41 (40%)
	D	2 (2%)
	unknown	24 (23%)
Vascular invasion	V0	49 (48%)
	V1	31 (30%)
	V2	1 (1%)
	unknown	21 (21%)
Tumour Stage	pT1	3 (3%)
	pT2	6 (6%)
	pT3	47 (46%)
	pT4	24 (23%)
	unknown	22 (22%)

Table2. 2: Clinicopathological features of patient's cohort n=83

Variable	Classification	Frequency n (%)
Sex	Male	49 (59%)
	Female	34 (41%)
Age	Median	71 (43-80)
Dukes' stage	A	4 (5%)
	B	29 (35%)
	C	39 (47%)
	unknown	11 (13%)
Vascular invasion	V0	42 (51%)
	V1	29 (35%)
	V2	1 (1%)
	unknown	11 (13%)
Tumour Stage	pT1	3 (4%)
	pT2	4 (5%)
	pT3	43 (52%)
	pT4	21 (25%)
	unknown	12 (14%)

2.1.1.2 Cohort two: miRNA analysis and Immunohistochemistry (IHC)

In a separate cohort, paired Formalin Fixed Embedded Paraffin (FFPE) cases (primary CRC and matched normal mucosa samples) were obtained from 81 patients. Normal tissue samples were a mixture of margin blocks and normal immediately adjacent to the tumour and all the normal samples were only extracted from the mucosal tissue. All patients had undergone surgery between 2012 and 2014 at Queen`s Medical Centre (QMC) Nottingham, UK. Cases were selected on the basis of the availability of clinicopathological data and the presence of at least 50% tumour cells in the tumour block, (table 2.3). The haematoxylin and eosin stained slides were reviewed by Dr. Wakkas Fadhil (PhD Student, Department of Pathology, Pathology Research group). These samples were used for evaluation of miRNA profiles and testing of protein expression using IHC (see below). Although it would have been preferable to perform the tests on

cohort 1, this was not possible as cohort 1 was originally put together for technical studies and samples were thus completely anonymised.

Table2. 3: Clinicopathological features of patient's cohort two, n=81

Variable	Classification	Frequency n (%)
Dukes' stage	A	12 (15%)
	B	37 (46%)
	C	32 (39%)
Vascular invasion	V0	40 (49%)
	V1	40 (49%)
	V2	1 (2%)
Tumour grade	1	2 (3%)
	2	74 (91%)
	3	5 (6%)
Tumour Stage	pT1	3 (4%)
	pT2	12 (15%)
	pT3	46 (57%)
	pT4	20 (24%)

2.1.1.3 Blood

Studies were performed on DNA and miRNA extracted from blood. In total, up to 10 mL of whole blood was collected from a peripheral vein and placed in EDTA-containing tubes from five colon cancer patients diagnosed at the department of oncology at the Queen`s Medical Centre (QMC)/Nottingham/UK and from 3 healthy volunteers as well. Patients undergoing surgery for colorectal cancer were consented by Dr Oliver Ng and blood taken on the day of operation and consecutive days after operation till the patients been discharged from the hospital. Access to tissue was approved by the Nottingham Sciences Biobank (Biobank Reference ID ACP000007). After collection, blood samples were sent to the Biobank/QMC for anonymisation before plasma was separated from blood cells by double centrifugation. The first centrifugation was done at 1900 g for 10min at 4^{o}C and the supernatant was transferred into 1.5 ml tubes. This was

followed by second centrifugation at 16,000g for 10 min at 4°C to completely remove cellular components. Plasma aliquots were immediately stored at -80°C until use. The time from venesection to first centrifugation was under 1 hr. the clinicopathological features of all 5 patients are shown in table 2.4.

Table2. 4: Clinicopathological features of patient's cohort three, n=5

Variable	Classification	Frequency n
Dukes' stage	B	2
	C	3
Vascular invasion	V0	4
	V1	1
Tumour grade	2	5
Tumour Stage	pT2	2
	pT3	2
	pT4	1

2.2 Methods

2.2.1 Mutation Screening

2.2.1.1 FFPE

DNA samples from cell lines (VACO5, HCT116, HT29, DLD1, RKO) and the first cohort of FFPE samples were screened for mutations in exonuclease domain region of both *POLE* and *POLD1* genes using Polymerase Chain Reaction (PCR) followed by High Resolution Melting Analysis (HRM). Additionally, *APC, KRAS, TP53, BRAF, NRAS, PIK3CA, PTEN, FBXW7, CTNNB1, CDH1, MET, AKT1, MSH6, ALK, FOXL2, KIT, PDGFRA, EGFR, GNAQ, GNAS, FGFR2, MAP2K1, KRBB2, SMAD4, STK11 and SRC* genes were screened for mutations using Targeted Next generation sequencing (Illumina Miseq platform).

2.2.1.2 Plasma

cfDNA from plasma of CRC patients (pre – and post operation) were screened for mutations in *KRAS, BRAF, PTEN, PIK3CA, SMAD4* and *TP53* using PCR followed by HRM. Blood samples were taken from five CRC patients the day of surgery (before) and within one week after surgery and blood was also taken from three health volunteers.

2.2.1.3 DNA extraction

2.2.1.3.1 Cell line and FFPE samples

All DNA from Cell lines and 150 FFPE (cohort one) used in this project, were prepared by our colleague Salih Ibrahim and Wakkas Fadhil (Molecular Pathology department/QMC). The identity of the cell lines was verified by their morphology and examining *APC, TP53, KRAS* and *BRAF* mutation.

2.2.1.3.2 cfDNA from plasma

QIAamp DNA Blood Mini Kit (Qiagen, Cat. No. 51104) was used for cfDNA extraction from plasma taken from colon cancer patients and healthy volunteers. In brief, up to 200 µl of plasma was added into the 1.5 ml microcentrifuge tube, which contained 20 µl proteinase K (QIAGEN Protease) and then 200 µl of AL buffer was added into the mixture before being vortexed for 15 sec and incubated at 56^{o}C for 10 min. After that, the tube was centrifuged briefly to remove drops from the inside of the lid and 200 µl of absolute ethanol (96-100%) was added into the mixture and vortexed for 15 sec before transferring the entire mixture into the QIAamp Mini column and centrifuging at 8x10^{3} rpm for 15 sec. Thereafter the spin column was placed in a new 2 ml tube and washed with 500µl of buffer AW1 and spinning at 8x10^{3} rpm for 15 sec before adding 500µl of buffer AW2 and again centrifuging at 8x10^{3} rpm for 2 min. Finally, for removing any remaining ethanol the column was then centrifuged for 3 min at 13x10^{3}rpm and placed in a 1.7ml

tube and DNA was eluted by addition of 30µl of the AE buffer, waiting for 1 min, and centrifugation at 13x10^{3}rpm for 1 min.

2.2.1.3.3 DNA concentration measurement and purity assessment

The NanoDrop ND-1000 UV-Vis Spectrophotometer (LabTech International Ltd, Ringmer, UK) was used to measure the concentration and purity of the extracted DNA. The instrument measures absorbance (A) at 280 nanometres (nm) for protein and at 260nm for DNA. A260/A280 ratio of ~1.8 indicates relatively pure samples (Fleige and Pfaffl, [332]). The concentration and purity of DNA extracted from FFPE were assessed by Wakkas Fadhil and Salih Ibrahim. Whereas I and Abutalib Assiri assessed cfDNA extracted from the plasma samples.

2.2.1.4 High Resolution Melting Analysis protocol

2.2.1.4.1 Primer Design

In order to amplify all the DNA sequences, two sets of primers were designed, one for initial prediagnostic multiplex (PDM) reaction and the second one for a single specific diagnostic (SSD) reaction as previously described [333]. The publicly available genomic DNA sequence of the target regions were taken from NCBI (http://www.ncbi.nlm.nih.gov/pubmed/). Primer 3 software (http://biotools.umassmed.edu/bioapps/primer3_www.cgi) was used to design primers for all exons involved in the regions of interest. After that all primers were tested by using in silico PCR (http://genome.ucsc.edu/cgi-bin/hgPcr?command=start) to ensure that all primers are specific for the length of the genes of interest as expected. As well, MFEprimer-2.0 software (http://biocompute.bmi.ac.cn/CZlab/MFEprimer-2.0/index.cgi/check_dimer) was used to check the possible primer dimers and hairpins. Primers covered exons involved in the exonuclease domains in both *POLE* (exons 9, 10, 11, 12, 13, and 14) and *POLD1* genes (exon 8, 9, 10, 11, 12, and 13). Each exon in both genes was amplified by the primer pairs leading to a short amplicon whereas for exon 8 and exon 13 in *POLD1* were amplified each by two set of nested primer.

Additionally, primers for other genes were used in the present study to validate NGS data using PCR followed by HRM. Primers for both *TP53* and *APC* gene were previously designed by Seth et al, [334] and Blaker et al, [335], respectively, while, primers for the rest of genes were designed by (PhD Students, Department of Pathology, Pathology Research group) (Appendix A- Table 1). The genes and their NCBI reference sequence number are shown in (Appendix A- Table 2).
The primers were designed to get small products to (1) avoid occurrence of Single Nucleotide Polymorphism (SNP), (2) minimise the adverse effect of the fragmentation of the DNA from FFPE tissues on PCR amplification, and (3) improve the results of HRM analysis as the clearest discrimination is usually seen with the shortest amplicons [336].

2.2.1.4.2 PCR optimization

To detect the optimal annealing temperature for all outer and nested primers used in this study, gradient PCR was performed using 12 different annealing temperatures. For initial optimization of PCR protocols, high-quality DNA was obtained from cell lines. The reaction was performed in 25 µl reaction which contained 1 µl (1u) DNA *Taq* polymerase, 2.5mM $MgCl_2$, PCR buffer 2.5µl, dNTPs 400µM (Thermo scientific, Frementas, UK), forward and reverse primers pair at 250nM, 20 ng template DNA and water to complete the total volume. A gradient thermal Cycler (PeQLab Biotechnology) was used with the following protocol: initial denaturation at (95^{o}C/ 5 min) 1 cycle; [denaturation (95^{o}C/1min), annealing (12 different temperatures /1min), extension (72^{o}C/1min)) 40 cycle]; followed by a final elongation (72^{o}C/10 min) 1 cycle. Analyzing PCR products were undertaken on gel electrophoresis on a 1.5% agarose gel and 1xTBE buffer (0.089M Tris, 0.089M boric acid, 0.002M EDTA, pH 8.2 – 8.3). Validation of PCR protocols were also undertaken on DNA extracted from FFPE tissue, as this is representative of material usually available in the diagnostic setting.

2.2.1.4.3 Polymerase Chain Reaction Protocols

2.2.1.4.3.1 Quick-multiplex-consensus (QMC)-PCR protocols

The study used the QMC-PCR protocol, which is a nested procedure in which an initial prediagnostic multiplex (PDM) reaction is followed by a single specific diagnostic (SSD) reaction (23), The PDM reaction was undertaken in a single tube which contain all outer primer sets. The PDM reaction was performed in a final volume of 25µl. Although *Taq* DNA polymerase (Thermo scientific, Frementas, UK) gave a nice PCR product, it was not optimal for HRM analysis and it was decided to try different master mixes (table 2.5). Only the 1xHotShot master mix (Cadama Medical Ltd), got good PCR product and HRM results. Each reaction contained 1xHotShot master mix (Cadama Medical Ltd), 250nM of each primer and 20ng template DNA. PCR was performed using a two-step cycling protocol: initial denaturation (95^{o}C/5 min) 1 cycle; (denaturation (95^{o}C/ 1 sec) and annealing (Ta ^{o}C/1 sec)) 25 cycles whereas for POLD1 primers, initial denaturation (95^{o}C/5 min) 1 cycle; (denaturation (95^{o}C/ 10s), annealing (- ^{o}C/30 sec) and extension (72 ^{o}C/10 sec)) 40 cycles. The SSD reaction was in a final volume of 10µl, which contained 1xHotShot master mix, 1 primer pair with each primer at 250 nM final concentration, 1x LC Green PLUS (Idaho Technology) and water to complete the total volume. The template consisted of 1µl of a 1:100 dilution of the PCR product from the PDM reaction, and PCR was performed using a twostep protocol: initial denaturation (95^{o}C/5 min) 1 cycle; (denaturation (95^{o}C/1 sec) and annealing (55^{o}C/1 sec) cycle 30.

Table2. 5: List of master mix used for optimization HRM analysis

Master Mix	Company
Hotshot	Clent Life science
SensiFAST™ HRM Kit	Bioline
Gotaq HRM master mix	Promega
Taq DNA polymerase	Thermo Scientific
Quanta	Biocience

2.2.1.4.3.2 Standard PCR

In the standard PCR, each reaction contained 1xHotShot master mix (Cadama Medical Ltd), 250nM of each primer and 20ng template DNA. PCR was performed using a three-step cycling protocol: initial denaturation (95^{o}C/5 min) 1 cycle; [(denaturation (95^{o}C/ 10 sec), annealing (- ^{o}C/20-45 sec) (extension 72^{o}C/1 min) 40 cycle] and final extension (72^{o}C/5) 1 cycle.

2.2.1.4.3.3 Wild-type blocking PCR (WTB-PCR)

In order to improve the limit of detection of the HRM, we tried to enrich for mutant alleles using wild-type blocking probes. It is a modified form of the standard PCR protocol by adding Locked Nucleic Acid (LNA) probes to bind to wild-type sequences during the PCR and thereby prevent amplification. This would then favour amplification of mutant DNA during PCR. LNAs are nucleic acid analogs that contain a 2`O to 4`C methylene bridge that locks the ribose group into a C3`-endo conformation (figure 2.1). The introduction of LNA monomers into oligonucleotides is complementary to a region of the wild-type sequence and increases the melting temperature of DNA between 1°C and 8°C per modification [337, 338]. The probe would be mismatched if bound to mutant sequence (even if the difference between wild-type and mutant is only one base) and therefore its binding would be less efficient thereby enabling amplification of the mutant DNA [339]. A phosphorylation group was modified at 3`-terminus of the LNA to avoid the extension of WTB by DNA polymerase in PCR procedures. And more, the study used Klenow Fragment, LC (2 U/µl) (Thermo Scientific Ca. No. EP0054), which lacks 5`→3`, to avoid hydrolysis of WTB by 5`→3` exonuclease activities of traditional DNA polymerase. WTB-PCR was performed in a volume of 20µl mixture containing 1 units of Klenow fragment polymerase (Thermo Scientific), 2µl Klenow fragment DNA Polymerase Buffer, dNTPs 400µM (Thermo scientific, Frementas, UK), forward and reverse primers pair at 250nM, and with series concentrations of WTB (CTCTTGCCTACGCCACC and A+CG+C+C+A+C+C+AG+CT underlined bases were LNA) (Exiqon). The amplifications were performed with the following cycling conditions: denaturation

at 95ºC/5 min; 1 cycle; 95ºC/20 sec, -ºC/45 sec and 72ºC/1 min; 45 cycles; and a final extension at 72ºC/5 min; 1 cycle. DNA derived from HCT116, which contains a heterozygous KRAS GGT13GAT mutation, and VACO5, which is wild type at KRAS codon 12 and 13.

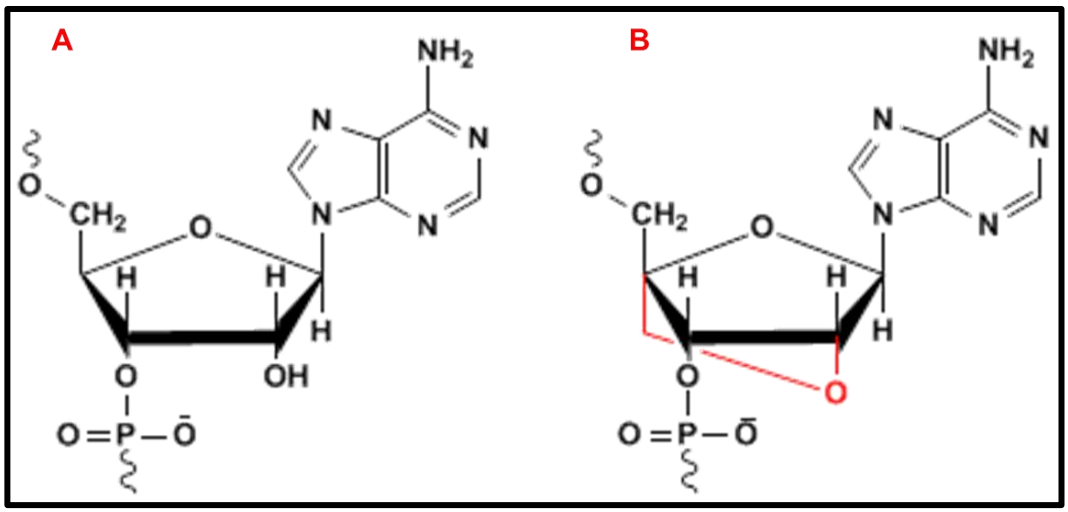

Figure 2. 1: Structures of (A) Normal nucleic acid and (B) Locked nucleic acid

2.2.1.4.3.4 Cold PCR

The limit of detection of HRM analysis is limited to mutant alleles present at 5% among wild type alleles [340-347], Co-amplification at lower denaturation temperature-PCR (Cold-PCR) has been introduced, which increases the sensitivity of mutation-scanning assays through enrichment of mutant alleles without adding to the costs and complexity of the experiment [348].

Cold-PCR is a new form of modified standard PCR protocol developed to enrich 'minority alleles' from mixtures of wild-type and mutation-containing sequences. As a result Cold-PCR amplification of genomic DNA, PCR products containing high percentages of variant alleles are obtained thus permitting their detection [137]. This method is based on the observation that each DNA sequence has a critical denaturation temperature (Tc), which is lower than its melting temperature (Tm). DNA amplicons differing by a single nucleotide have different amplification efficiencies when PCR denaturation temperature is set to Tc [349]. Generally, two

form of cold PCR described, fast and full cold PCR. Full Cold-PCR can theoretically enhance detection of any type of mutation via selective denaturation of heteroduplexes produced after annealing of wild and mutant pairs. Fast Cold-PCR enriches G>C to A>T mutations that slightly lower the melting temperature of the PCR amplicon, by using a specific Tc that favours PCR amplification of the mutant allele [138].

This study preferred to use full cold PCR instead of fast cold PCR because it was looking for unknown mutations. Since occurrence of single nucleotide mismatch generates a predictable change in melting temperature, the study performed gradient PCR in order to find the lowest denaturing temperature with the following cycling conditions [initial denaturation (95°C/5min) 1 cycle; [denaturation (95°C/1 sec) / annealing (55.5°C/1 sec)] 10 cycle; [denaturation (Tm±4°C/10 sec) / annealing (55°C/15 sec) / extension (72°C/10 sec)] 25 cycle (23). After analysing products using HR-1, the lowest denaturation temperature that produced amplification was chosen as the critical temperature Tc for each exon, which is enough to melt heteroduplexes preferentially and allow better amplification. After introducing two extra steps to standard PCR protocol, the following cycling conditions were used: [initial denaturation (95^{o}C/5 min) 1cycle; [denaturation (95^{o}C/30 sec)/ hybridization (70^{o}C/30 sec)/(denaturation at Tc/30 sec)/(annealing (55.5^{o}C/1 min)/ extension(72^{o}C/1 min)) 45 cycle; (72^{o}C/10 min) 1 cycle (figure 2.2).

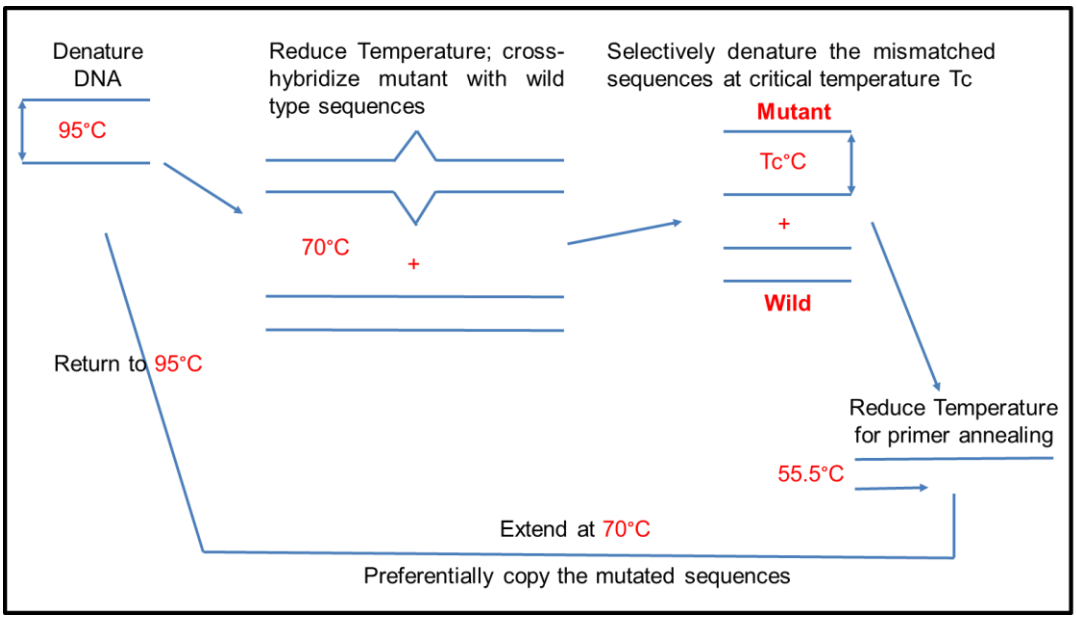

Figure 2. 2:Description of full cold PCR protocol, updated from Li et al, [350]

2.2.1.4.4 High Resolution Melting Analysis (HRM)

HRM is a new method for DNA analysis developed in 2002 by collaboration between academics (University of Utah, UT, USA) and industry (Idaho technology, UT, USA). [351]. HRM is a post-PCR analysis and highly sensitive screening method used for detecting genetic variation as small as one base difference between samples in nucleic acid sequences. Although HRM was developed for detection of SNPs, it is applicable for mutation analysis. A SNP is just a single base change in a DNA sequence and SNPs normally occur throughout a person's DNA (approximately every 1000 bases) and represent the most common type of genetic variation among people. Most SNPs have no risk on health but some of these genetic differences may influence disease risk, could help predict an individual's response to certain drugs, susceptibility to environmental factors and so on. Sensitivity and specificity of this method is around 100% for PCR product less than 400bp [135]. However, studies that have been conducted on the HR-1 and LightScanners instruments, revealing reduced sensitivity as the length increases above 400 bps [352, 353]. The technique of high resolution melting curve (HRM) analysis is a method applied for detection of mutations based on specific sequence-related melting profiles in the PCR

product. HRM analysis of PCR products depends on the differences in DNA melting and it is enhanced by heteroduplexes between the mutant and wild type alleles which are less stable and melt at a lower temperature in comparison to pure wild type or mutant sequence. Initially within the DNA sequence the region of interest is amplified using the PCR, and to follow the transition of double-stranded DNA (dsDNA) to single stranded (ssDNA) special saturation dyes "known as intercalating dye", is added in to the reaction. The dye emits more strongly when bound to dsDNA than ssDNA, with increasing temperature from around 50^{o}C to 95^{o}C, the fluorescence intensity decreases as two strands of the dsDNA melt apart (figure 2.3). The main advantages of the method include cheap, fast, high sensitivity and specificity, a closed-tube format that greatly reduces contamination risk and is a non-destructive method in which after HRM further analysis of the sample can still be performed by other methods, such as DNA sequencing or gel-electrophoresis [354]. However, some of limitations of HRM are defined such as it can identify the presence of sequence change but cannot identify the specific nucleotide changes and it cannot distinguish between germline variants SNPs and cancer-specific somatic mutations. HRMA data therefore require confirmation with other techniques such as Sequencing.

Although HRM is simple, rapid and inexpensive technique, several factors have to be taken into consideration to get accurate HRM results. For example, quality of DNA template, PCR specificity, length of the amplicon (below 400bp) [355], type of mutations (small insertions and deletions may be more difficult to detect than substitutions) [356], GC content [353, 357], dye (for example, LC Green Plus which is better than SYTOs 9, which is better than EvaGreens, which is better than SYBR Green I) [358], melting analysis software and resolution of the instrument because the ability of the instruments to resolve shape and absolute temperature differences is not similar [355, 358, 359]. For these reasons and in order to get a highly accurate HRM results, the present study used pure DNA samples (the purity and quality of the DNA samples were assessed as mentioned above), primers were designed to amplify a short amplicon (most less than 200bp), LC Green plus dye, and the specificity of the PCR product were checked

initially by blasting the primers against NCBI human genomic database, amplicon size was checked with gel electrophoresis.

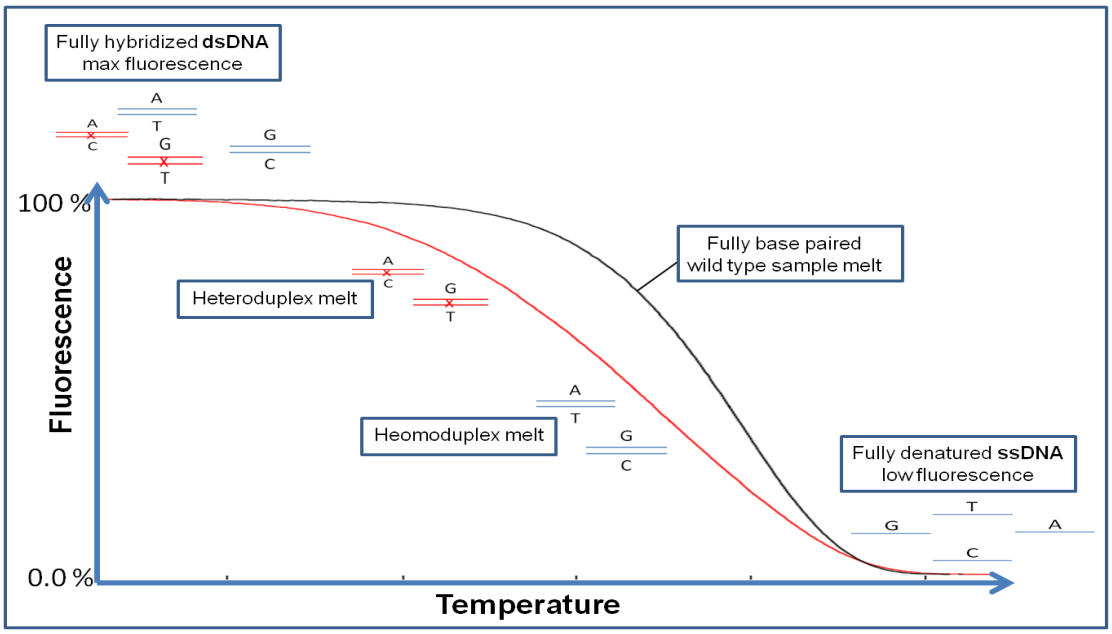

Figure 2. 3:Illustrated principle of work of HRM analysis, updated from (www.BIO-RAD.com). After formation of heteroduplexes between mutant and wild type alleles and fluorescence data showing a linear decrease of fluorescence at low temperature.

HRM was performed on two different machines i.e. the HR-1 and the Lightscanner. With the HR-1 platform, the PCR products were transferred to Light Cycler capillaries (20 µl) (Roche). The products were loaded in the HR-1 HRM instrument (Idaho Technology) and melted by raising temperature at a rate of 0.3^{o}C/s, 200 data points were collected /^{o}C, with a starting temperature of 70^{o}C, a final temperature of 95^{o}C and fluorescence being measured between 80-95^{o}C. The HR-1 analysis tool custom software was used to analyse all data, and both derivative plots and difference plots were generated after normalizing, which allows visual clustering of different genotypes to have the same beginning and ending fluorescence. Temperature shifting is a temperature adjustment forcing each curve through the same temperature range to overcome the effect of absolute temperature variation and this increases the accuracy of the results. In order to separate out the mutants from the wild-type samples, the difference plot

was visually inspected by using a threshold of 4% difference in fluorescence as previously described [134].

The LightScanner-96 platform is different to the HR-1 in the following way; (1) The HR-1 uses capillaries and can analyse only one sample at a time while Lightscanner is a plate system facilitating the analysis of 96 samples simultaneously (2) data point/^{o}C in HR-1 is 200 whereas in the Lightscanner it is 10 (3) Lightscanner software has auto grouping to do analysis of the samples while in HR-1 the data only analysed manually.

To analyse DNA melting on the light scanner, the following melting parameters were used. The temperature range was set at 65^{0}C – 95^{0}C with a ramp rate of 0.1^{0}C/sec and to 'Auto' exposure. Sample equilibration was performed at 62^{0}C. The high-resolution CCD camera captured the fluorescence data for analyses. Thereafter, analyses were performed with the Lightscanner Call-IT software version 2.0.0.1.331 using the Expert scanning module. The negative control samples as well as the poorly amplified samples were excluded from further analyses using the negative filter. Normalization of the melting data was done as per manufacturer's instructions. The normalized melting curves were shifted at a 0.05 level, while sensitivity was set to normal at zero level. Both the 'Auto group' and 'Common vs Variant' functions were used to group the normalized and shifted melting curves according to their melting patterns. The results were viewed in the 'Shifted melt curve' and 'Difference curves' outputs.

2.2.1.5 Sanger direct sequencing

2.2.1.5.1 Performance

The PCR products of tumour samples showing abnormal HRM patterns and suspected to contain mutations underwent direct bidirectional Sanger sequencing. The QIAquick PCR Purification Kit (Qiagen, UK) was applied to purify PCR product according to the manufacturer's instructions. In brief, one volume of PCR sample was mixed with the five volumes of buffer PB and then the mixture was added to the QIAquick column and centrifuged at $8x10^{3}$ rpm for 1 min. Next, 750 µl of the washing buffer, PE, was added to the column and centrifuged at

$8x10^3$ rpm for 1 min. The elution was carried out in a clean microcentrifuge tube using 30 µl of PCR grade water and, following this, the concentration was measured by the NanoDrop ND-1000 UV-Vis Spectrophotometer. The appropriate products and their corresponding primers were directly sequenced at the Biosynthesis and Biopolymer Unit, School of Biomedical sciences, University of Nottingham, UK, and sequenced using Applied Biosystems BigDye® Terminator v3.1 Cycle Sequencing Kit and 3130xl Genetic Analyzer.

2.2.1.5.2 Sequence Analysis

The Finch TV software V1.5.0 (http://www.technelysium.com.au/chromas.html) was used to view and interpret the generated chromatograms. Finally, the BLAST 2 (http://www.ncbi.nlm.nih.gov/blast/bl2seq/wblast2.cgi) sequences program was used to analyze the sequencing data by comparison against wild-type sequence of the target genes.

2.2.1.6 Targeted Next Generation Sequencing

Next generation sequencing (NGS), massively parallel or deep DNA sequencing technology perform sequencing of millions of fragments of DNA in parallel. In Targeted NGS, regions of the genome are selected and sequenced. The present study used Illumina MiSeq sequencing by synthesis chemistry as this gives the highest accuracy, lowest error rate and the highest percentage of base calls above Q30. Generally, 4 basic steps are included in Illumina NGS; (1) Library preparation: the DNA is randomly fragmented, adapter ligated to the 5` and 3` ends, then adapter ligated DNA is amplified by PCR and purified before putting into the machine for cluster generation. (2) Cluster generation: the library is loaded into the cartridge and the fragments are captured on a lawn of surface bound oligos complementary to the library adapter, then each fragment is amplified into the clonal cluster. (3) Sequencing and (4) data analysis.

2.2.1.6.1 Trusight^{TM}tumour kit

The present study used Trusight^{TM}tumour kit (Illumina, USA) which provides a comprehensive view of somatic variation in solid tumours including lung, colon, melanoma, gastric and ovarian. The genes involved with solid tumours were selected for this panel to include 26 driver genes from CAP and NCCN guidelines, relevant publications, and late phase pharmaceutical clinical trials (Illumina, USA). The Trusight tumour content set provides coverage of entire exons in regions where variation has been catalogued in the COSMIC database in oncogenes and coverage of all exons in tumour suppressor genes (Illumina, USA). The targeted regions included 82 exons of 26 genes involved in solid tumours and total length captured was 21.6Kb (Appendix A- Table 3). The amplicons were 165-195bp long, and the read depth of each amplicon in the target region was 1000x as a minimum coverage, with 7000x mean (Illumina, USA).

2.2.1.6.2 Target enrichment of genomic DNA and sequencing

The workflow of the NGS mutation analysis is represented in (Figure 2.4) including a description of DNA library preparation, NGS processing and data analysis.

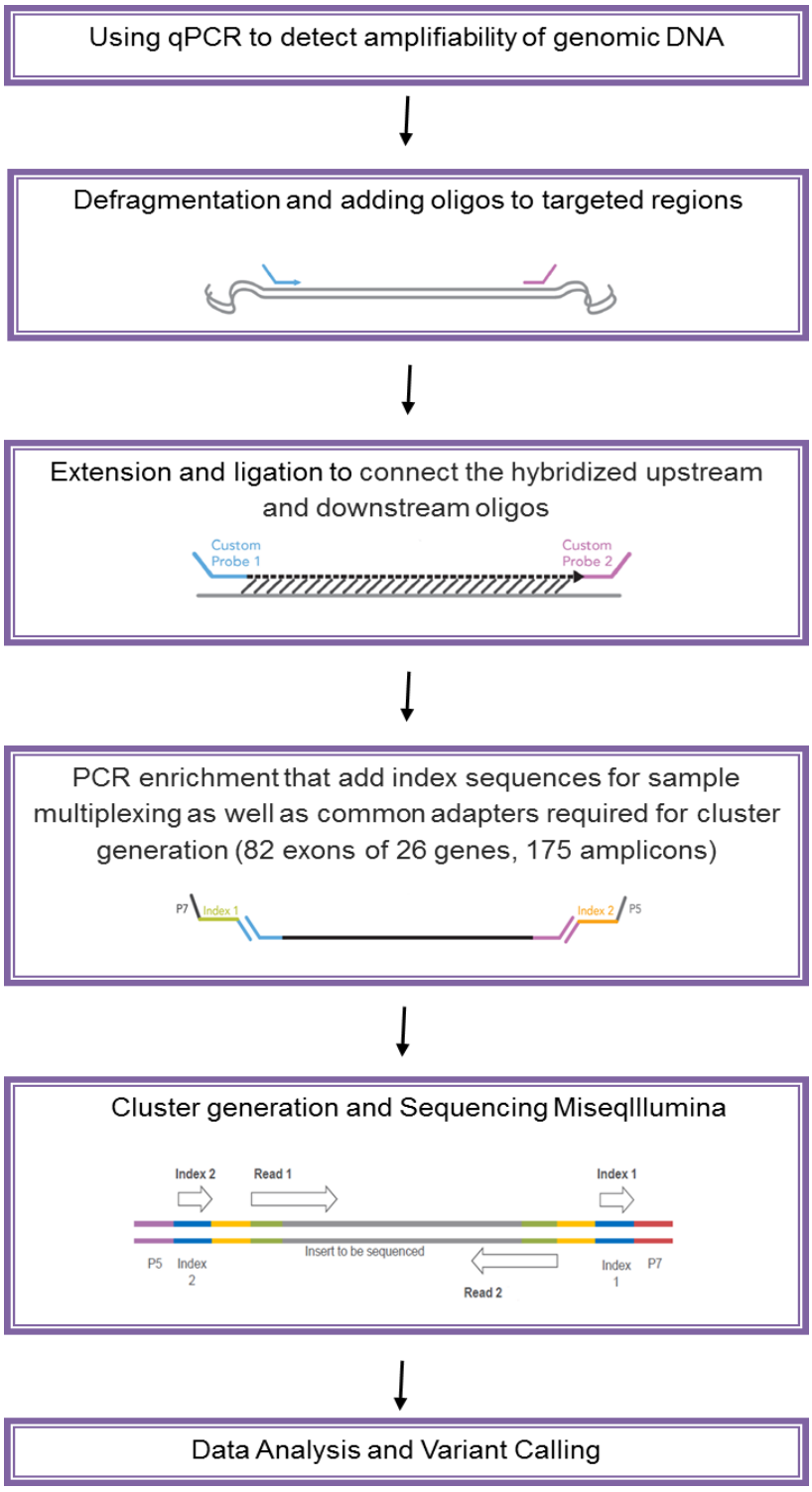

Figure 2.4: Trusight tumour kit workflow

2.2.1.6.2.1 Quality control of DNA extracted from FFPE samples and cell lines

Assessment of the quality of genomic DNA extracted from FFPE samples to ensure it is suitable for NGS is a first step in preparing DNA libraries. DNA libraries are DNA sequences of interested regions after adding common adapters required for cluster generation. As described in the manufacturer's protocol (Illumina, kit), qPCR was used to determine the amplifiability of genomic DNA samples. As an initial step the QCP primer (Quality control primer) and QCT (Quality control template), which is supplied by the Illumina, were diluted as follows: 1 µl of primer was diluted in 9µl of Nuclease-free water to make a 10-fold dilution (sequence of primers and targeted regions were not mentioned in the kit), and 5 µl of control template was diluted in 495µlof Nuclease-free water to make a 100-fold dilution (quantity of neither QCP nor QCT were mentioned in the kit). The reaction mixture was performed in 10 µl which contained 2 µl of diluted DNA (1 µl of Qiagen-extracted genomic DNA was added to 99µl of Nuclease-free water in microfuge tubes to make a 100-fold dilution), 1 µl of diluted QCP primer (quality control primer) 5 µl of Sybermastermix (Life technologies) and 1 µl of water. Thermal cycler (Applied Biosystems) was used with the following condition: 50ºC/2min (1cycle)/ 95ºC/5 min (1cycle)/ [(95ºC/30 sec; 60ºC/30 sec; 72ºC/30 sec) (40cycle)]. By comparing the amplifiability of FFPE DNA relative to that of the QCT non-FFPE reference gDNA, a ΔCt value was calculated for each sample and used to predict its performance in the TruSight Tumor Sample Preparation assay (Illumina does not support the use of gDNA samples giving a delta Ct value of greater than 4), the (table 2.6) determine the fold dilution required for each calculated delta Ct.

Table2. 6: DNA dilution rates for NGS

Delta Ct	-2.5 to -1.5	-1.5 to -0.5	-0.5 to 0.5	0.5 to 1.5	1.5 to 4.0
Dilution	16x	8x	4x	2x	No dilution

2.2.1.6.2.2 Fragmentation and hybridization of oligo pool

In the fragmentation and size selection steps, which serve to break the DNA template into smaller sequence-able fragments, 30-300ng genomic DNA was enzymatically fragmented to fragments between 300-330bp (the kit does not mention how much and which enzyme used). Upstream and downstream oligos specific to targeted regions of interest were hybridized to the genomic DNA samples. For each amplicon two pairs of oligos are designed. One pair is complementary to one strand and another pair to the opposite strand (figure 2.4). 10 µl of each sample was added into two wells of the plate, once the first well contain 5 µl upstream and downstream oligos for the forward strand and the second well contain 5µl upstream and downstream oligos for the reverse strand, then 35 µL of OHS3 (Oligo hybridization for sequencing 3) was added, which facilitates the process of hybridization between oligos and template. After that the plate was sealed with aluminium foil and heated up to 95^{o}C for 1 min (this step is required to denature all double stranded DNA into single strand) before incubated at 40^{o}C for 14-15 hrs (Illumina, Trusight Tumour Kit).

2.2.1.6.2.3 Extension and ligation

This process connects the hybridized upstream and downstream oligos after removing unbound oligos from genomic DNA using a filter capable of size selection as described in manufacturer's protocol (Illumina, kit). A DNA polymerase extends the upstream oligo through the targeted region, and is ligated to the 5'end of the downstream oligo using a DNA ligase after incubating reaction mixture in 37^{o}C for 45min. This results in the formation of products containing the targeted regions of interest flanked by sequences required for amplification (Trusight tumour kit, Illumina).

2.2.1.6.2.4 PCR Amplification

In this step, the extension-ligation products are amplified using primers that added index sequences for sample multiplexing, by which software can now uniquely identify the sample, as well as common adapters required for cluster generation. After detaching DNA from the filter plate using 25 µl of 0.05N NaOH, the PCR reaction mixture was performed in 60 µl which contained 20 µl of DNA, 22µl of PMM2/TDP1 PCR master mix (PMM2 is a PCR master mix and TDP1 is a TruSeq DNA polymerase) and 9 µl of forward and reverse primers. The PCR cycling condition was 95^{o}C for 3 min and 27cycles of 95^{o}C for 30 sec, 62^{o}C for 30 sec, and 72^{o}C for 30 sec and the final extension 72^{o}C for 5 min (Trusight tumour kit, Illumina). As a result, bar-coded fragment sequencing libraries were made using a Paired-end DNA sample preparation kit. Paired-end means reading both the forward and reverse template strands of each cluster during one paired-end read. Both reads contain long range positional information, allowing for highly precise alignment of reads.

2.2.1.6.2.5 PCR clean-up and Library quantification

After ligation with the Illumina adaptor, the libraries were cleaned-up using AMPure beads (Beckman Coulter, Inc.) as described in the manufacturer's protocol (Illumina, kit) and in order to achieve the highest quality of data and to create optimum cluster densities on the Illumina MiSeq sequencing platform, the quantity of DNA libraries were measured using Qubit® 2.0 Fluorometer (Life Technologies). Moreover, analysing PCR products was undertaken on gel electrophoresis on a 2% agarose gel and 1xTBE buffer (0.089M Tris, 0.089M boric acid, 0.002M EDTA, pH 8.2 – 8.3), 1kb ladder to confirm the presence of the 300–330 bp library products.

2.2.1.6.2.6 Library Normalization

This process normalizes the quantity of each library to ensure more equal library representation in your pooled sample. In this step, samples were diluted using EBT (Elution buffer with Tris) to a final concentration of 4nM.

2.2.1.6.2.7 Prepare samples for sequencing

In preparation for cluster generation and sequencing, equal volumes of normalized library are denatured, then are combined, diluted in hybridization buffer, and heat denatured prior to sequencing on the MiSeq. In this step 5 µL of each 4nM normalized library DNA (12 in total) was transferred into its own tube containing 15 µl of the NaOH/EBT. Then 10 µl of each Library/NaOH/EBT was transferred into a single tube and mixed before transferring 8µl of pooled libraries into the tube contained 792 µl of hybridization buffer plus 8µl of 20 pMPhiX library, which is an internal control for sequencing, then incubated at 96^{o}C for 2 min and cooled on ice for 5minutes before transferred into the template position in the Miseq reagent cartridge and sequenced on the MiSeq system using Miseq reagent kits v2 as recommended by Illumina.

2.2.1.6.2.8 Cluster generation and sequencing

On the Illumina Miseq analyzer captured libraries were amplified and sequenced as paired-end reads of 2x121bp plus 2x8bp of two index sequences reads. Clustering is a process wherein each fragment molecule is amplified on the flowcell, which is a glass slide coated with two types of oligos complimentary to those adapters attached to the DNA library. Everything except sample preparations are completed on the flowcell including; template annealing, template amplification, sequencing primer hybridization, sequencing by-synthesis reaction and generation of fluorescent signal. After formation of millions of clusters, sequencing begins with the extension of the first sequencing primer to produce the first read with each cycle, then the index one read primer is hybridized to the template and read the first index. Afterwards the second index is read before extend the second flowcell oligo and forming complementary strand (figure

2.5). After that, the original forward strand is cleaved off and washed away then the second read begins with the introduction of the read two sequencing primer. Finally, both reads will overlap in the middle to ensure complete coverage of the amplicon (Illumina, USA). Library preparation, target enrichment and sequencing were carried out in Molecular Genetics at the Nottingham University Hospital (Nottingham, UK).

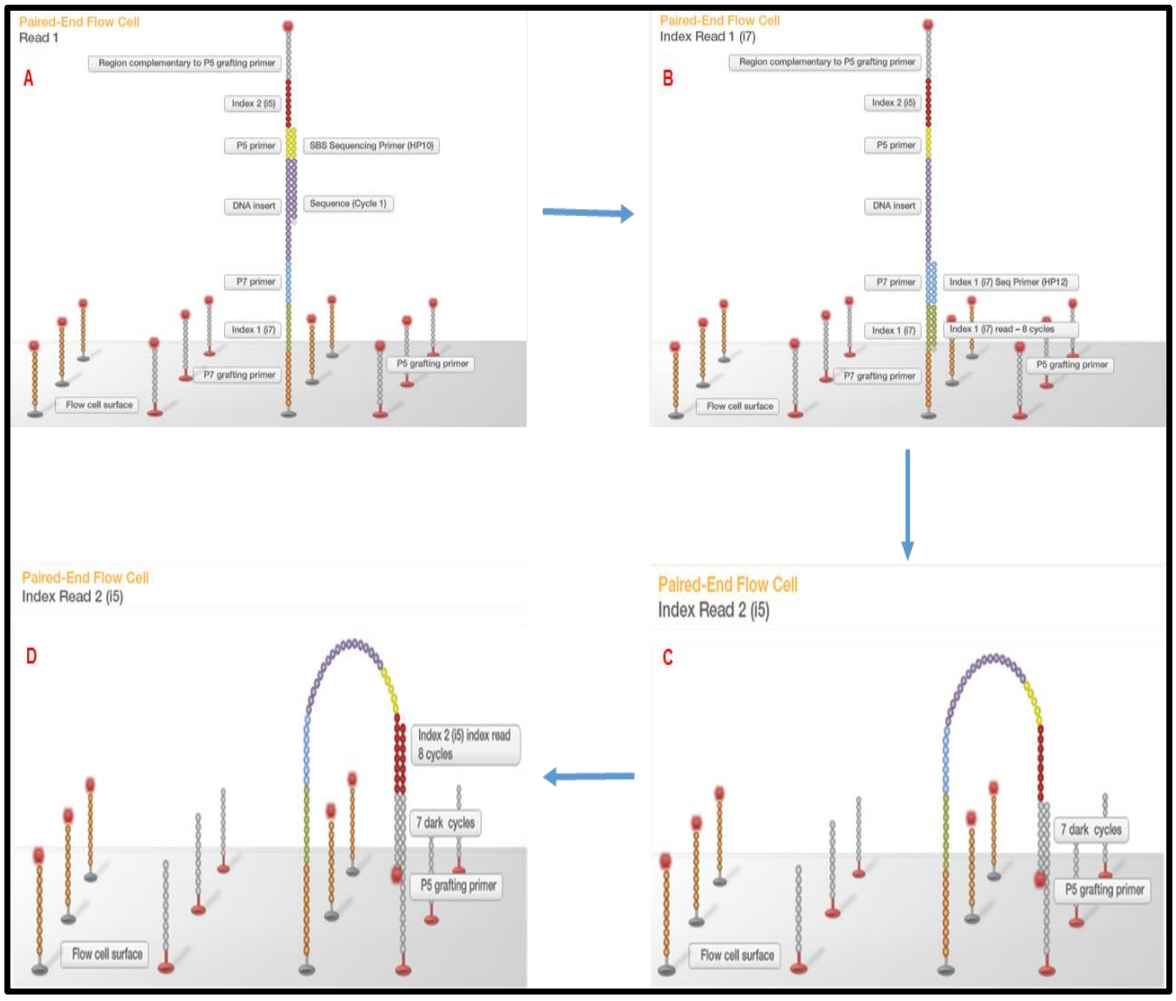

Figure 2. 5:Cluster generation and sequencing; (A) Sequencing begins with the extensions of the first sequencing primer to produce the first read with each cycle, (B) After the completion of the first read, the read product is washed away. In this step the index one read primer is introduced and hybridized to the template, (C) After completion of the index read, the read product is washed away. The 3' end of the template is d-protected. The

template now falls over and binds the second oligos on the flowcell, (D) Index 2 is read in the same manner as index one. Index 2 is washed off at the completion of this step.

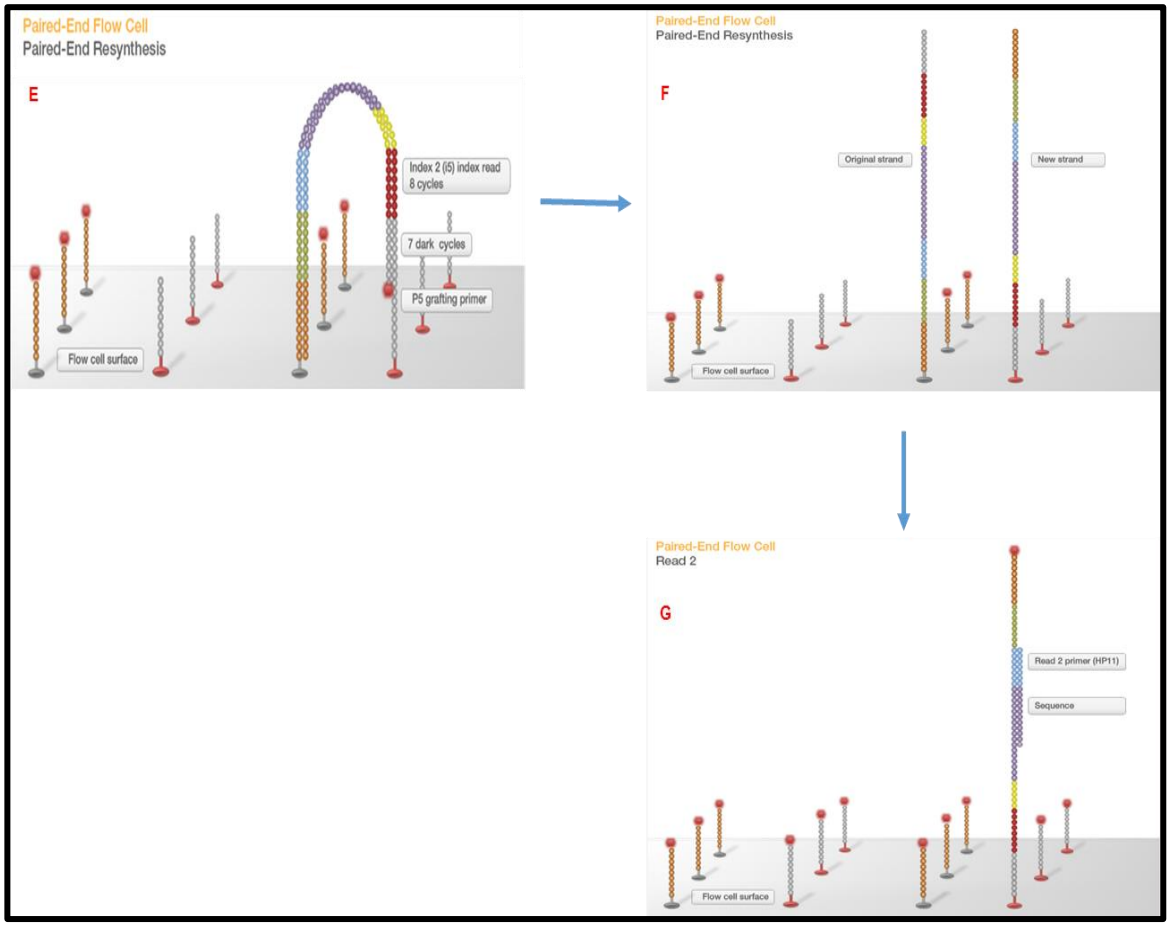

Figure 2. 6: (E) Polymerase extend the second flowcell oligo forming a double stranded bridge, (F) The double stranded DNA is then linearized and the 3' end blocked. The original forward strand is cleaved off and washed away, (G) Read two begins with the introduction of the read two sequencing primer. The reads will overlap in the middle to ensure complete coverage of the amplicon

2.2.1.6.2.9 NGS Data analysis

After sequencing, the raw signal data were analysed using MiSeq Reporter v2.1. The pipeline includes signal processing, base calling, quality score assignment, trimming of adapter sequences, filtering for high quality reads, PCR duplicate removal, read alignment to human genome (hg19) sequence, coverage analysis, and variant calling. After primary data analysis, detected sequence variants

(SNVs) and (indels) were assembled in a variant call file format generated by the MiSeq Reporter Program. Variant filtering and annotation was performed with Variantstudio™ v2.1analyser. Single nucleotide variants (SNVs) or insertion/deletions (indels) were defined based on the following criteria: (1) were present in both pools (i.e. were present in both forward and reverse strand sequencing), (2) read depth of 1000x or an average depth of 500x per pool, (3) have ≥3% or variant frequency in the merged variant call (VCF) files, (4) have a minimum Q-score of 20 (phredQ score: Q-score is a prediction of the possibility of an incorrect base call. A higher Q-score means that a base call is less likely to be incorrect and more reliable), (5) must not be polymorphisms according to dbSNP (Illumina, USA). For the evaluation of the mutation artefact, criterion (1) and (3) were excluded. For the analysis of the limit of detection, criterion (3) was excluded.

2.2.1.6.3 NGS assay performance

Template derived from FFPE tissue is known to have limitations i.e. the size of the PCR amplicon is limited, the presence of contaminants may inhibit the PCR, there are frequent C→T mutation artefacts due to deamination and there is a higher frequency of spontaneous PCR errors due to degraded template.

In order to assess the utility of the TruSight tumour kit for use in FFPE tissue, we evaluated the frequency of spontaneous mutations and C→T mutation artefacts comparing the number of changes occurring in one direction only with the total number of reads in that direction. We also tested the short-term precision (intra-assay variability), the long-term precision (inter-assay variability) and the limit of detection of the assay. For short term precision, one of the samples was tested in 8 replicates in the same run. For long term precision, the same sample was tested on 3 different runs. In each case, both the depth of sequencing and the mutant allele frequency were evaluated. For limit of detection, a series of dilution experiments were carried out using DNA from two diploid CRC cell lines. DNA from Vaco5 harboring a BRAF V600E (c.1799 T>A) heterozygous mutation was spiked into DNA extracted from HCT116, which is wild type at codon 600. DNAs

were mixed to produce samples containing mutant alleles at the following percentages; 50%, 25%, 12.5%, 6%, 3%, 1.5% and 0.75%.

2.2.2 Quantification of miRNAs and mRNAs expression levels

Expression level of a panel of six miRNAs (miR-20a, miR-21, miR-29a, miR-31, miR-92a and miR-224) and mRNAs of six genes (*SMAD4*, *PTEN*, *TGFBRII*, *BCL2*, *KLF4* and *RASA1*) were measured using RT-qPCR on cell lines, the second cohort of FFPE samples and a small set of plasma samples

2.2.2.1 MiRNA in formalin-fixed and paraffin-embedded (FFPE) samples

FFPE tissues are widely used archive materials representing a major resource for the study of clinical samples, biomarker discovery and validation with possible long-term follow-up data [360]. However, these types of samples represent a challenge for mRNA profiling due to damage and degradation during fixation [361], and is thus of limited value for gene expression analysis [362]. In contrast, miRNAs are better preserved, because of their stability and small size; therefore, their RNA are extracted more readily from FFPE samples [363]. This phenomenon thus enhances our ability to consider miRNAs as cancer biomarkers.

2.2.2.1.1 Macrodissection

Due to the fact that stromal cells can confound the interpretation of the tumour gene expression profiles, tumour specimens were macrodissected after haematoxylin–eosin slide evaluation by a pathologist to ensure a minimum of 50% tumour tissue content as recommended by Chretien et al, [364]. Two 20µm-thick serial sections were cut from each paraffin block and placed on glass slides (unstained section) by Andrew Murphy (Histopathology Department at Nottingham Queen`s Medical centre). The area containing the region of tumour was identified by Dr. Wakkas Fadhil by examining haematoxylin–eosin (H&E)

slides that were used as templates to mark the region on unstained sections. The unstained sections were all prepared in the same orientation as the original H&E slides and the region of tumours and normal stroma were correctly identified. Before starting macrodissection the area of tumour was marked on the underside by using an indelible marked pen. Then the area within the mark settled scraped off with the disposable scalpel and collected in Eppendorf® vials. The same procedure has been done for the normal stroma as well. RNA isolation was performed using the miRNAeasy FFPE kit (Cat. No. 217504, Qiagen, GmbH) protocol.

2.2.2.1.2 Total RNA Extraction

Purification of total RNA, including miRNA was started by adding 160 µl of deperafinization solution (Cat. No. 19093) into the Eppendorf (2 m) tube containing FFPE sections, mixed vigorously and heated up to 56^{o}C for 3min to remove paraffin. After cooling down at room temperature, 180 µl of PKD buffer (a Proteinase K Digestion Buffer) was added, then vortexed vigorously and centrifuged at 10x10^{3}rpm for 1min, which makes two phases (the lower clear phase and upper turbid phases). After that to release RNA from section 10 µl of proteinase K was added into the clear phase and incubated at 56^{o}C for 15min before incubation at 80^{o}C for 15min to reverse formaldehyde modification of nucleic acids. After that the lower phase was transferred into a new 2mL tube, and incubated on ice for 3min before being centrifuged for 15 min at 13.5x10^{3} rpm and then the supernatant was transferred into the new 2 ml tube. Next, 16µl of DNAse booster buffer and 10 µl of DNase I (to digest DNA) were added into the tube and mixed gently by inverting it and incubating at room temperature for 15 minutes. Following this, the lysate was mixed with 320 µl buffer RBC and to provide appropriate binding conditions for RNA 1120 µl absolute ethanol (100%) was added to the sample and mixed thoroughly by vortexing. The lysate (700 µl) was then applied to an RNeasy MinElute spin column and centrifuged for 15sec at 10x10^{3} rpm, and repeated till the entire lysate has been passed through

RNeasy MinElute spin column. The membrane then washed twice with 500 µl of buffer RPE, the first wash was followed by spinning at $10x10^3$ rpm for 15 sec and the second wash at $10x10^3$ rpm for 2 min (use of new collection tube after each wash). The column was then centrifuged for 3min at $13x10^3$ rpm for removing any remaining ethanol. Thereafter to evaporate the residual ethanol, opened spin columns lid was centrifuged for 5 min at $13.5x10^3$ rpm, then RNA was eluted by addition of 25 µl of the RNase-free water, waiting for 5min and centrifugation at $13.5x10^3$ rpm for 1 min.

2.2.2.1.3 RNA concentration measurement and purity assessment

The extracted RNA concentration and purity were measured using Thermo Scientific NanoDrop™ 1000 Spectrophotometer. The instrument measures absorbance (A) at 260 nanometer (nm) for DNA and at 280nm for protein (figure 2-6). An OD 260/280 ratio of ~2.0 is usually acceptable indicator of pure RNA quality. After quantification stock samples were stored at -80°C.

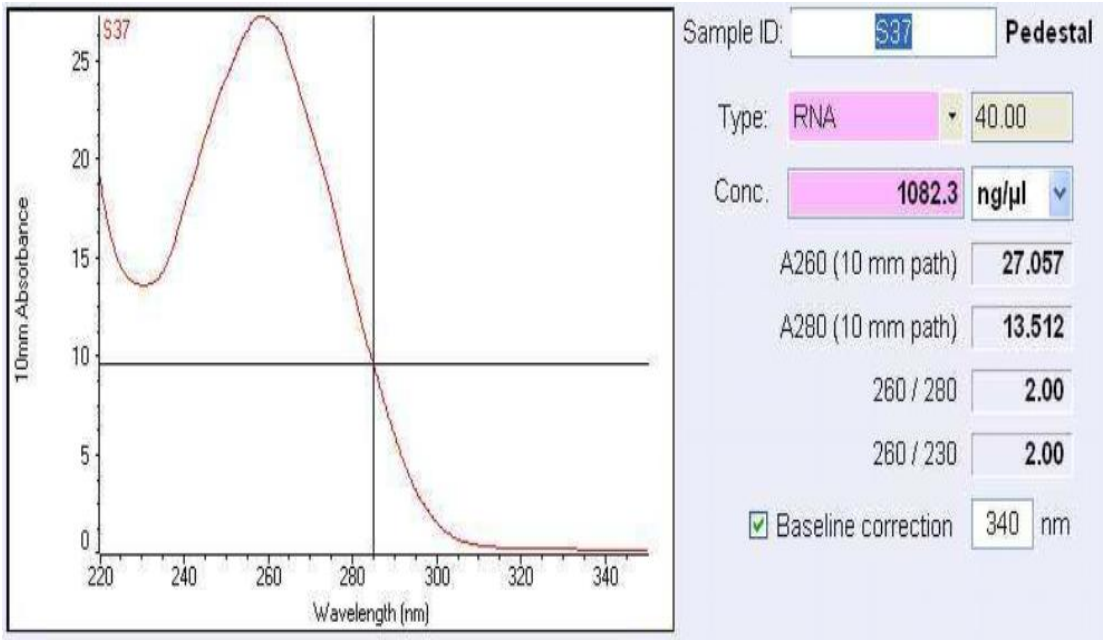

Figure 2. 7:Measurement of the RNA concentration obtained from FFPE by Nanodrop. ***The A260/A280 absorption ratio was 2.0 indicative of high RNA amount relative to protein.***

2.2.2.2 MiRNA in Plasma

RNA isolation from plasma was performed using the miRNAeasy serum/plasma kit (Cat. No. 217184, Qiagen, GmbH) protocol. The frozen lysates were incubated at 37ºC in a water bath until samples were completely thawed and 200µl of plasma was mixed with 5 volume (1 ml) of Qiazol lysis reagent and then incubated at room temperature for 5 min. Thereafter 5 µl miRNeasy Serum/Plasma Spike-In Control (1.6 x 10^8 copies/µl working solution) was added and mixed thoroughly before adding 200 µl of chloroform and vortexing for 15 sec. After incubation of the tubes containing lysates at room temperature for 3 min and centrifugation for 15 min at 12,000 x *g* at 4°C, the upper aqueous phase was transferred into a new collection tube and mixed thoroughly with 1.5 volume of absolute ethanol 96-100%. The lysate (700 µl) was then applied to an RNeasy MinElute spin column and centrifuged for 15sec at 10x10^3rpm. Centrifugation was repeated till the entire lysate has been passed through RNeasy MinElute spin column. The membrane then was washed with 500µl of buffer RPE followed by spinning at 10x10^3rpm for 2 min (use of new collection tube after each wash). Thereafter to remove the

residual ethanol, opened spin columns lid was centrifuged for 5 min at 13.5x10^{3}rpm. RNA was eluted by addition of 14 µl of the RNase-free water, waiting for 5 min and centrifugation at 13.5x10^{3}rpm for 1 min.

2.2.2.2.1 Preparation of miRNeasy Serum/Plasma Spike-In Control

In order to evaluate expression of miRNAs in plasma, the study used the miRNeasy Serum/Plasma Spike-In Control, which is a *C. elegans* miR-39 miRNA mimic (Cat.No. 219610, GmbH), as an internal endogenous control. In total, 10 pmol lyophilized miR-39 was reconstituted in 300µl RNsae-free water, resulting in a 2 x 10^{10} copies/µl stock. After that to get the recommended concentration of spike-in control for RNA purification, 4µl of 2 x 10^{10} copies/µl miRNeasy Serum/Plasma Spike-In Control stock was added to 16µl RNase-free water, resulting in a 4 x 10^9 copies/µl dilution. Thereafter, 2µl of the 4 x 10^9 copies/µl was then added to 48 µl of RNase-free water to provide a 1.6 x 10^8 copies/µl working solution.

2.2.2.3 Reverse Transcriptase-Quantitative Polymerase Chain Reaction

2.2.2.3.1 qPCR primers

The primers used in the present study, are of different sources. Primer pairs for *SMAD4, KLF4, RASA1, TGFBRII, PTEN* and *BCL2* were designed by the primer3 software (http://biotools.umassmed.edu/bioapps/primer3_www.cgi) and their annealing temperature optimised (Appendix A- Table 4). In order to avoid any genomic DNA amplification; the primers were designed to span exon-exon junction, in which the primers bind to two consecutive exons and cannot bind if there are intronic sequences present. *HPRT* primers were previously published [365]. The publicly available mRNA sequences of the genes were taken from NCBI (http://www.ncbi.nlm.nih.gov/pubmed/) and were blasted against the Refseq mRNA database using the free online available service provided by NCBI at, (http://www.ncbi.nlm.nih.gov/tools/primerblast/index). After that all primers were tested by using in silico PCR (http://genome.ucsc.edu/cgi-

bin/hgPcr?command=start) to ensure that all primers are specific for the length of the genes of interest as expected. As well, use MFEprimer-2.0 software (http://biocompute.bmi.ac.cn/CZlab/MFEprimer-2.0/index.cgi/check_dimer) to check the possible primer dimers and hairpins. Finally, PCR products generated from all primers were checked for size by gel electrophoresis as detailed earlier. In (Appendix A- Table 4), the genes mRNA NCBI reference sequence number, the sequences of the qPCR primers are shown.

2.2.2.3.2 Generation of Complementary Deoxyribonucleic acid (cDNA)

2.2.2.3.2.1 mRNA

In order to quantify the amount of messenger RNA (mRNA) product of a specific gene, mRNA was converted into cDNA following the provider's instruction. In summary, RT+ (reaction mixture with Reverse Transcriptase) and RT- (reaction mixture without Reverse Transcriptase) reactions for each sample were performed by mixing 0.5 µg random hexamers (Thermo Scientific 50142 200 ng/µl) (2.5 µl) with 1 µg RNA and the mix was made up to a total volume of 15 µl of RNase free water and incubated at 70°C for 5 minutes to remove any secondary structure. Next, the master mix was prepared which is composed of M-MLV RT 5X buffer (Promega), 1µl of Moloney Murine Leukemia Virus Reverse Transcriptase enzyme (M-MLVRT) (Promega, 200Unit/µl), and 0.5 mM of .Deoxyribonucleotides triphosphate (dNTP) (Thermo Scientific). The master mix was added into the samples and was incubated at 37°C for 60 min. For inactivation of the enzyme the sample was heated for 10 min at 90°C. The generated cDNA samples stored at -80°C for subsequent use.

2.2.2.3.2.2 miRNA

Reverse transcription to create cDNA was performed using the Miscript II RT kit (Cat. No. 218161, Qiagen, GmbH) protocol. In brief, 1 µg of FFPE RNA containing miRNA was polyadenylated by poly (A) polymerase and reverse transcribed to cDNA using miScript Reverse Transcription kit (Qiagen). The reverse transcription master mix was prepared on ice as follows, 1 µg of extracted RNA

including miRNA was mixed with 4 µl of 5x miScript HiSpec Buffer (to prepare cDNA for subsequent mature miRNA), 2 µl of 10x nucleic acid, 2 µl of miscript reverse transcriptase mix and made up to a total volume of 20µl of H_2O. Following this, Samples were incubated at 37°C for 1 hr (strand synthesis) followed by 95°C for 5 min (enzyme denaturation). Samples were diluted 1:10 and stored at -20°C.

2.2.2.3.3 Real-time PCR amplification efficiencies and linearity

The efficiency (E) of PCR is defined as the fraction of target molecules that are copied in one PCR cycle [366]. Amplification efficiency, E, is calculated from the slope of the standard curve using the following formula: E = 10–1/slope. Determination of the efficiency of the primers is the very first thing that needs to be done before getting set up for quantitative PCR (qPCR) because knowing the assay efficiency is critical to accurate data interpretation. The standard curve remains the most powerful way to estimate PCR assay efficiency that is broadly accepted by the community [367]. A serial dilution (1:10, 1:100 and 1:1000) has been made from a sample of unknown quantity (cDNA) and the standard curve is created by plotting the log of the dilution factor (for unknown quantities) against the CT value obtained during amplification of each dilution. The hallmarks of an optimized qPCR assay are: Linear standard curve (R2 > 0.980), high amplification efficiency (90–105%) and consistency across replicate reactions. Perfect PCR efficiency will demonstrate a change of 3.3 cycles between 10 fold dilutions of template [368].

2.2.2.3.4 Quantification by real-time quantitative RT-PCR

2.2.2.3.4.1 miRNA

In FFPE CRC specimens, a panel of 6 miRNAs, which were selected based on previous reports [288, 369-372], were quantified using SYBR green qRT-PCR assay. Realtime qPCR was performed using miScript SYBR Green PCR kit (Qiagen) with the manufacturer provided miScript Universal primer and the

miRNA-specific forward primers. Q-PCR was undertaken in a 7500 Fast Real-Time PCR System (Applied Biosystems, USA).

MiRNA forward primers were reconstituted by adding 550 µL Tris-EDTA buffer solution P.H 8.0 (Cat. No. T9285, Sigma-Aldrich, UK). The miRNA specific forward primer sequences are shown in (Appendix A- Table 5). Each reaction was performed in a final volume of 10 µl containing 2 µl of the cDNA, 1 µl 10x of each primer and 1x SYBR Green PCR Master mix (Qiagen). The amplification profile was: denaturation at 95ºC for 15 min, followed by 40 cycles of 94ºC for 15 sec, 55ºC for 30 sec and 70ºC for 30 sec, in which fluorescence was acquired. At the end of the PCR cycles, melting curve analyses were performed to validate the specific generation of the expected PCR product. Each sample was run in triplicates for analysis. The expression levels of miRNAs were normalized to RNU6B. MiRNA expression was quantified as ΔCt values, where Ct = threshold cycle, ΔCt = (Ct target miRNA - average Ct of RNU6B), and ΔΔCt values = (ΔCt target miRNA tumor tissue - ΔCt target miRNA matched normal tissue) were used to quantify miRNA expression of tumor compared with matched mucosa. Fold change of gene was calculated by the equation $2^{-\Delta\Delta Ct}$.

2.2.2.3.4.2 mRNA

In FFPE CRC specimens, a panel of 6 genes, which were selected based on previous reports, were quantified using SYBR green qRT-PCR assay. Realtime qPCR was performed using miScript SYBR Green PCR kit (Qiagen) and 7500 Fast Real-Time PCR System (Applied Biosystems). After 10-fold dilution of cDNA with water, PCR was performed in a final volume of 10 µl containing 2 µl of the cDNA, 0.250 µM of each primer and 1x SYBR Green PCR Master mix (Qiagen) and the reaction brought up to the total volume with water. The amplification profile was: denaturation at 95ºC for 15 min, followed by 40 cycles of 94ºC for 15 sec, annealing 55ºC for 30 sec and 70ºC for 30 sec. At the end of the PCR cycles, melting curve analyses were performed to validate the specific generation of the expected PCR product. Each sample was run in triplicate for analysis. The expression levels of mRNAs were normalized to HPRT. mRNA

expression was quantified as ΔCt values, where Ct = threshold cycle, ΔCt = (Ct target gene - average Ct of HPRT), and ΔΔCt values = (ΔCt target gene mRNA tumor tissue - ΔCt target mRNA matched normal tissue) were used to quantify mRNA expression of tumor compared with matched mucosa. Fold change of the gene was calculated by the equation $2^{-\Delta\Delta Ct}$.

2.2.3 Evaluation of Protein Expression

Protein expression of six genes (*SMAD4, PTEN, TGFBRII, BCL2, KLF4 and RASA1*) were measured using immunohistochemistry staining (IHC) on the second cohort of FFPE samples. Additionally, protein expression of 20 target genes (*CD31, CD34, D2-40, WT1, SMAD4, BCL2, TGFBRII, KLF4, RASA1, RAS, E-cadherin, p85, p110, PTEN, phospho-PTEN, phoshpo-AKT Serin, phosphor-AKT Threonine, mTOR, pCRAF and pGSK*) were also measured using Reverse Phase Protein Array (RPPA) on a small set of FFPE samples (20 primary CRCs and matched metastasis).

2.2.3.1 Reverse Phase Protein Assay

2.2.3.1.1 Protein extraction for RPPA

For protein extraction, 20 µm thick sections were cut from FFPE tumour blocks. Tissue was deparaffinised in xylene and the xylene was removed using graded ethanol (100%, 96%, and 70%). Excess alcohol was removed by centrifugation at 16,000g for two minutes. Next, 40 µl of lysis buffers (20% SDS, 0.5M DTT and 0.5M Tris-HCl solution at pH 8) were added to the tissue pellet and incubated at 100°C for 60 min using a thermomixer. Finally, the tubes were centrifuged at 14,000 rpm for 20 min and the supernatants were collected and stored at 80°C until used.

The next steps including protein quantification (Fast green stain) and then run samples with RPPA were performed by Dr. Ola Negm (Research Fellow Immunology/ School of Life Science/ Nottingham University Hospital).

2.2.3.2 Western Blot

2.2.3.2.1 Electrophoresis and blotting

To check the primary antibody specificities using Western Blotting, the proteins, which were taken from cultured cells by our group, were denatured at 95ºC for 5min after mixing 3 volumes of lyaste with one volume of 4X loading buffer (100mM Tris-HCl [pH 6.8], 200 mM DTT, 4% SDS, 0.2% glycerol and 0.2% bromophenol blue) freshly supplemented with β-mercaptoethanol (5% V/V final concentration) (Sigma Aldrich, UK) and then chilled on ice for another 5min prior to loading. The denatured protein was then separated on NuPAGE 4-12% Bis-Tris Mini Gels (Novex Life Technology, USA). The gel was placed into the electrophoresis tank (Invitrogen life Technology) containing 1X NuPAGE MOPS SDS running buffer (Novex Life Technology, USA).

The samples (15 µl), 3 µl of magic maker (Invitrogen life Technology) and 2 µl Amersham rainbow marker (GE Healthcare) were loaded on the gels and run at 125 V for 1.5 hr. After this time, the gel filter sandwich was prepared to transfer the separated proteins onto an Amersham nitrocellulose membrane (GE Healthcare) in NuPAGE transfer buffer (Novex Life Technology, USA) at 25 V for 30 min. The gel filter sandwich includes the gel, nitrocellulose membrane and filter papers. The gel was placed onto the filter paper, which was soaked in transfer buffer, then the membrane, which was wetted with methanol and washed with transfer buffer, was placed onto the gel and finally placed the remaining filter papers on the top of the membrane. The membrane was then put in the blocking buffer (consisting of 5% non-fat milk powder, 1X PBS and 0.1% Tween-20) to block the nitrocellulose membrane at room temperature for 1hr on the shaker. After that the blocked membrane was incubated with the diluted primary antibodies for overnight at 4ºC with agitation. Following this three 5min with PBS plus 0.1% Tween-20, were performed after the primary antibody incubation. After the last wash, the membranes were incubated for 1hr at room temperature with horseradish peroxidase (HRP)-conjugated secondary antibody solution (1:1000 diluted in blocking buffer; DAKO), depending on species of the primary antibody (i.e. polyclonal rabbit anti-mouse and polyclonal goat anti-rabbit; Dako). Next,

three washes of 5min each were done with PBS plus 0.1% Tween-20. Nitrocellulose membrane was visualized by enhanced chemiluminescence (ECL) reagents (Supersignal West Pico Chemiluminescent Substrate, Thermo scientific, UK) and exposed to an Amersham hyperfilm ECL (Kodak, UK) for different periods to visualise any bands.

2.2.3.3 Immunohistochemistry (IHC)

2.2.3.3.1 Optimisation of primary antibodies for IHC

Before doing IHC staining, Western blotting analysis was applied to test the specificity of the primary antibodies. Those antibodies which did not give non-specific bands and specifically bound to the target proteins were chosen for IHC staining. Later on, colorectal tumour composite tissue sections were used to find optimal concentrations of primary antibodies using Novolink Polymer Detection Systems (Leica Microsystems). After testing different concentrations (higher and lower than the recommended one) of primary antibodies, the antibody concentration that demonstrated a heterogeneous staining pattern without background staining was picked as the best one for IHC staining.

2.2.3.3.2 Tissue Microarray

After testing and identifying specificity and optimal concentration of all primary antibodies, TMA sections of colorectal cancer tissue were stained to measure the expression of proteins. A TMAs is a recent innovation and high throughput technology in the field of pathology for evaluation of biomarker expression in a large number of tissue samples. Tissue microarrays are paraffin blocks produced from minute tissue samples (i.e. single core/ tumour) from different paraffin donor blocks and collected in an array manner (microarray) block to allow multiplex histological analysis. Details of antibodies used in the present study are shown in (Appendix A- Table 6). In the present study, TMA sections were prepared at Nottingham Health Science Biobank/QMC/Nottingham/UK, 4-µm paraffin-embedded CRC TMA sections were stained with anti-SMAD4, anri-KLF4, anti-

RASA1, anti-PTEN, anti-TGFBRII and anti-BCL2antibodies using Novolink Polymer Detection Systems (Leica Microsystems). A positive and negative control to confirm successful of experiments were included in each run.

2.2.3.3.3 Immunohistochemical staining

TMA slides were preheated at 60^{o}C for 10 min to melt the paraffin and then left on the bench for 10 further minutes to cool down to room temperature before being loaded onto an Autostainer XL Staining System ST5010 (Leica Microsystems) for dewaxing and rehydrating. The slides were dewaxed in two xylene baths for 5 min each then rehydrated in a sequence of IMS (methylated spirit industrial 0.89 S.G. 74 O.P., Fisher Scientific) concentrations each for 2 minutes followed by washing with water for 5 min. Once this has finished the antigen retrieval was performed by placing the slides in either 10 mM sodium citrate buffer pH 6.0 or EDTA (pH 9.0) and boiling the slides in a microwave at 750 W for 20 min. Sections were placed under running tap water to gently bring their temperature down to room temperature. Following antigen retrieval, sections were loaded into Shandon Sequenza coverplates (Thermo Fisher Scientific). To exclude any leak or possibility of presence of air bubbles, the coverplate was filled with pH 7.6 TBS (Tris Buffered Saline). Next, slides were loaded with 100 µl Peroxidase block for 5 min and washed twice with TBS each for 5 min. After washing slides were incubated with 100 µl Protein block for 5min and washed twice with TBS each for 5 min before incubated them for 1 hr with the primary antibody, which diluted to optimal staining conditions using Leica Antibody Diluent, and washed twice with TBS each for 5 min. Slides were loaded with Post primary block for 30 min and then washed twice with TBS each for 5 min, followed by Novolink Polymer for 30 min and twice with TBS each for 5 min.

The slides were loaded with diaminobenzidine (DAB) for 5 min (DAB chormogen and DAB substrate buffer in a ratio of 1:20) and washed twice with TBS each for 5 min before loading with counter staining haematoxylin for 6 minute and then washed twice with TBS each for 5 min. Finally, slides were dehydrated in a series

IMS (methylated spirit industrial 0.89 S.G. 74 O.P., Fisher Scientific) solutions and followed by two baths of xylene solutions for 5 min each and mounted with DPX.

2.2.3.3.4 Assessment of protein expression

Initially the stained TMA slides were checked with the light microscope to confirm the validity and staining, followed by scanning slides with a Nanozoomer Digital Pathology scanner (Hamamatsu Photonics) at 20x magnification, a semi-quantitative method (H-score) was used to assess protein expression in tumour cells. In the H-scoring method, presence and intensity of immunoreactivity were assessed. Staining intensity of each core was assessed as (0 was negative, 1 was weak, 2 was moderate and 3 was strong staining), then H-scores were calculated by multiplying the percentage of positive tumour cells (minimum 0 and maximum 100) by the staining intensity. [373]. After that to ensure reproducibility, slides were all assessed by a second scorer (Dr. Karzan Marf Murad) and intraclass correlation coefficient was applied to assess concordance between both scorers.

2.3 Statistical Analysis

Statistical analysis was performed using the SPSS software package (version 22). Categorical data was tested for assessing the statistical significance using a chi-square test and continuous data was tested for assessing the statistical significance of the differences between data sets using Wilcoxon test. Fisher's exact test correlation was used for association between unpaired tumour groups. Spearman's correlation was used for detecting correlation between targets. Multiple corrections testing (Bonferroni step-down (Holm) correction) and for both statistical analyses, P-values of <0.05 were statistically significant.

Chapter Three: Investigate mutation profile of CRCs

3.1 Abstract

Background: In the field of genomics, detecting and understanding genetic variation between populations and individuals is important. Mutation screening of genes is a key step in the search for potentially pathogenic genetic variants in disease susceptibility genes. This study is an attempt to investigate heterogeneity and mutation profiles of colorectal cancer.

Materials and Methods: DNA extracted from first cohort 83 CRC, matched normal mucosa of 17 samples, and 25 biopsy specimens were tested using the Trusighttumour kit allowing interrogation of 26 cancer driver genes. Samples were run on an IlluminaMiSeq genome sequencer. Additionally, the exonuclease domain regions of both POLE and POLD1 genes were screened for mutations using High Resolution Melting analysis and Sanger sequencing.

Results: The NGS assay had a limit of detection of 3% and intra- and inter-assay Coefficient of Variation of 12.3% and 10.6% respectively. Mutations were detected in 17/28 (60%) and common mutations (occurring in >5% of tumours) seen in 10/28 (35%) genes at expected frequencies. Approximately 8.4% (7/83) CRCs displayed mutations on only TP53 and only 5/83 (6%) tumours showed no evidence of mutation in any of the genes included in the panel. Quantification of mutant allele frequency showed clonal heterogeneity in 25%. Quantification of SNPs in *APC* and *TP53* showed that, allelic imbalance of APC SNP was seen in 16/34 (47%) of CRCs and allelic imbalance of TP53 SNP was seen in 10/17 (59%) of CRCs. As a sub-study a total of 81 mutations were found in the resection specimens and all but 3 (3.7%) were also present in the biopsy specimens. The mutant allele frequency was 1.03-fold higher in resection specimens than biopsies reflecting the presence of sub clones in the resection specimens.

Conclusion: Our study concluded that, the presence of mutations in several genes indicating different signalling pathways involved in CRC carcinogenesis. Mutation screening of ED-POLE and -POLD1 genes have shown that CIN and hypermutation may occur together. Additionally, for CRC at least, NGS demonstrates that it is acceptable to use biopsy material for predictive testing.

3.2 Introduction

In the past, depending on the hereditability of the malignancy CRC was only divided into hereditary and sporadic and all patients with CRCs were treated the same. However, advancement of different technologies enhanced developed of molecular biology and found that the division of CRCs into these two groups was insufficient. A rich history of investigations has revealed various types of genomic alterations ranging from small changes such as point mutations or small indels to large chromosomal copy number changes or rearrangements. These were detected in several genes and pathways, and have great roles in CRC development including Wnt/ß-catenin pathway, which is the most affected one, found in 93% of tumours regardless of subtype, epidermal growth factor receptor (EGFR, HER1), transforming growth factor beta (TGF-ß), phosphoinositide 3-kinase (PI3K) signalling activation and downstream mitogen-activated protein kinase (MAPK) [374]**.** Despite this background, a fully integrated view of the genetic and genomic alterations and thir significance for colorectal tumorigenesis is not available. Further investigation on these alterations may enable better understanding of the pathophysiology of CRC. The TCGA network project, which conducted a genome-scale analysis of 276 CRC samples, revealed that CRC could be split in two major groups by mutation rate, hypermutated and non-hypermutated cancers [229] match well with the previously discussed MSI and CIN pathways, respectively. The hypermutated category was further subdivided in two subgroups; hypermutated with microsatellite instability due to defective mismatch repair (dMMR) with a high mutation rate of 12–40 mutations/Mb and CIMP and hypermutated without microsatellite instability with somatic mutations in related DNA repair genes, such as *POLE* and rarely *POLD1* high mutation rate of >40 mutations/Mb [375]. In the recent past, NGS has begun to supplant other technologies which has helped research efforts to jump from investigation of mutations of individual genes to genome-wide identification of genetic alterations in cancer [9, 376]. NGS-based studies conducted to detect genomic profiles of CRC and combining results highlighted that among the non-hypermutated tumours, the eight most frequently mutated genes were *TP53*, *APC*, *KRAS*,

PIK3CA, *PTEN*, *SMAD4*, *FBXW7* and *NRAS*. Additional frequent mutations were identified *FAM123B, DOCK, ARID1A, FAT4,* TCF7L2, *SOX9, CDH10*, *SMAD2* and EGFR. Whereas in the hypermutated tumours, *BRAF*, *TGFBRII*, *MSH2*, *MSH6*, *ACVR2A*, *TCF7L2*, and *SLC9A9* [229, 377-381]**.** Additionally, *RNF43*, *ZNRF3*, *CTNNB1*, *POLE* and *POLD1* were also mutated [170, 176, 180, 375]. However, two genes that were frequently mutated in the non-hypermutated cancers were significantly less frequently mutated in hypermutated tumours: *APC* and *TP53 [229]*. These results demonstrate that non-hypermutated and hypermutated tumours progress through different sequences of genetic events. This observation is relatively consistent the Catalogue Of Somatic Mutations In Cancer (COSMIC) (http://cancer.sanger.ac.uk). However, higher rates of *TP53*, *APC*, and *KRAS* mutations are seen in the TCGA database, compared to the Catalogue of Somatic Mutations in Cancer (COSMIC) database.

The high degree of heterogeneity of the genetic and epigenetic changes among the individual tumours reflect not only many different clones of cancer cells, but also cancer cells show important dissimilarities among each other. The presence of different sub-clonal populations within the CRC makes it a complex disease with a variable clinical course and divergences in the response to therapy, even in tumours with similar histopathological characteristics [382]. Hence, increasing information of inter-tumour and intra-tumour heterogeneity is important for molecular classifications which enable healthcare professionals to provide the **'**right treatment, for the right person, at the right time**.** [383]. Use of biological agents against the epidermal growth factor receptor (EGFR) is the best example of the 'stratified medicine' approach, because the presence of mutations in the *KRAS* gene can predict lack of benefit from EGFR-targeted antibodies [384]. The gain-of-function mutation in *KRAS* induces constitutive Ras/MAPK signalling, which activates the epidermal growth factor signalling pathway at a point downstream of EGFR. In order to prevent adverse effects of the EGFR antibodies as well as reduce excessive treatment costs *KRAS* mutation analysis prior to anti-EGFR therapy is mandatory [385].

3.3 Aims

The aims of this chapter were to investigate the mutation profiles of a series of CRC and, in a sub-study, to ascertain whether diagnostic biopsy specimens are suitable for predictive testing using NGS. As it is a relatively new assay, we also tested the robustness with poor quality template. Mutations detected by NGS were validated using Quick-Multiplex-Consensus QMC-PCR in conjunction with high resolution melting. Additionally, we hypothesised that Cold-PCR could be used to distinguish germline SNP from somatic mutation.

3.4 Results

3.4.1 NGS results

3.4.1.1 QC of samples

The amplifiability of DNA from [102 primary CRC (resection), 30 biopsy, 32 metastasis and 17 normal colon] FFPE tissues was tested by comparing to that of the QCT non-FFPE reference gDNA, a ΔCt value was calculated for each sample and results showed that the quality of DNA from [83 primary, 25 biopsy, 22 metastasis and 17 normal colon] samples were suitable and the DNA from the rest 33 FFPE samples have bad quality and were not suitable for NGS analysis (Illumina does not support the use of gDNA samples giving a delta Ct value of greater than 4). The study was able to detect pathogenic mutations in tumour samples and cell lines, whereas all matched normal samples were negative for mutations.

3.4.1.2 NGS sequencing

A total of 12 cases were pooled into each flow cell. Each case was individually bar-coded. The pooled libraries in each flow cell averaged approximately 255 000 mapped sequence reads with the mean read length of 121 bp. This constitutes an average of approximately ~30 Mb of sequence per sample. In addition, all runs showed a very good read quality and nearly all bases (above 95%) had scores > Q30, which indicate that almost all of the reads were perfect and having zero errors and ambiguities. The depth sequence per mutational hotspot averaged 3581 reads (range 41-22962 reads) and the depth of sequence over all targets averaged 14354 reads (range 1112-51761).

3.4.1.3 Assay performance of NGS

Only samples which had passed the QC step underwent testing by NGS. As would be expected from a PCR based assay, the sequencing depth was variable between amplicons with the lowest depth seen for TP53 Exon 2 (mean depth 7088, range 1112-26264) and the greatest depth seen for PIK3CA Exon 20 (mean depth 25241, range 5417-51761). Since both strands are sequenced, changes seen in one strand but not in the other can be picked up as sequence artefacts.

Analysis of the raw data showed a mean frequency of C→T mutation artefacts of 24.4% and a spontaneous PCR error frequency of 8.9%.

Precision is a measure of the reproducibility of an assay and is best indicated by the coefficient of variation (CV) in a series of replicate tests. The short-term precision assay showed a mean CV of 12.3% (range 8.6% – 15.3%) for sequencing depth and 2.5% (range 1.6%-4.4%) for mutant allele frequency. The long-term precision assay showed a mean CV of 10.6% (range 3.2% – 15.1%) for sequencing depth and 2.2 % (range 0.01%-6.1%) for mutant allele frequency. Values of 10% and 15% respectively are regarded as acceptable for intra- and inter-assay CV and our data show that this kit and this platform perform reproducibly in FFPE tissue-derived template which has passed the QC.

A series of spiking experiments was carried out, to determine the limit of detection of mutant allele frequency. High quality DNA template derived from cell lines was used to obtain optimal data. The limit of detection was 3% of mutant alleles.

3.4.1.4 Sequencing of normal tissue

DNA from 17 normal tissue were analysed using NGS, the depth of sequences in normal tissue samples averaged 14422 reads (range 5028-42182), which is similar to the depth of sequence in matched tumour samples which averaged 13908 (range 3066-35510). Interestingly all the somatic variants were detected in tumour samples are totally absent in normal tissues. Additionally, a number of SNP were detected in tumour samples which were also found in matched normal samples. The list of SNPs found in matched tumour and normal samples are shown in table 3.1.

3.4.1.5 Validation of mutations

A total of 255 somatic non-synonymous mutations were observed in 15 of the 26 genes in the TruSight tumour panel (see below, figure 3.1). QMC-PCR with HRM was used to validate the mutations identified and initially 246/255 (96.4%) of the mutations were successfully validated (figure 3.2). The remaining 9 mutations (3.6%) were not detected. These were samples with low mutant allele frequencies

on NGS and, since the limit of detection of this NGS platform is lower than HRM, we optimized the Cold-PCR protocol (see below), to enrich for the mutations. This confirmed the presence of the mutations detected by NGS (figure 3.3) and showed that there were no false positives.

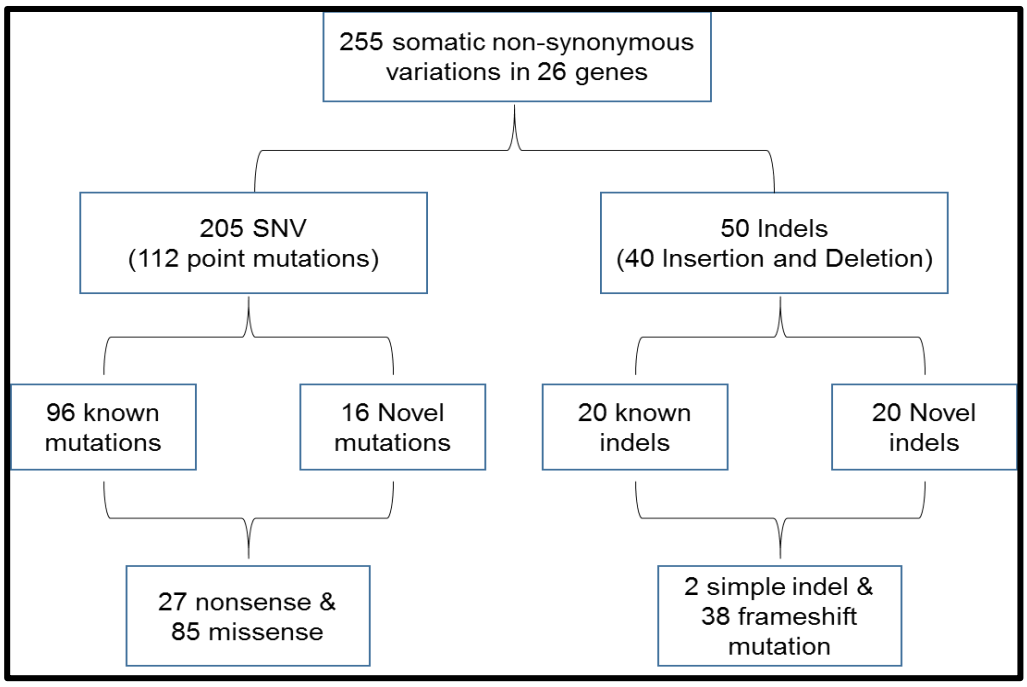

Figure 3. 1: summary of all mutations detected in all 83 CRCs. SNV, single nucleotide variant; indel, insertion and deletion using NGS.

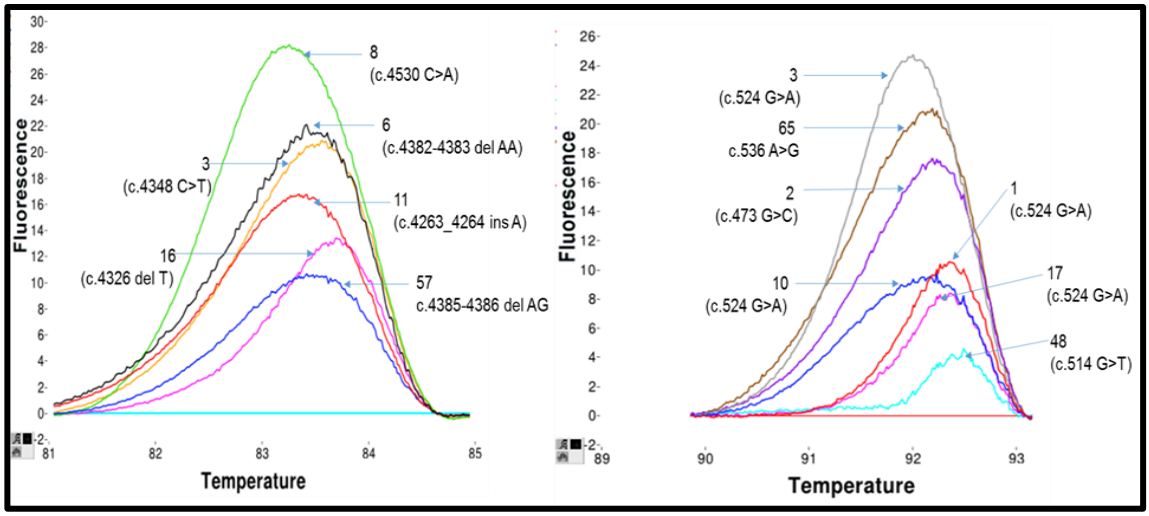

Figure 3. 2: Mutations identified by NGS were confirmed using HRM. (A) This shows the difference plots obtained for APC mutations in some of the samples. (B) This shows the difference plots obtained for TP53 mutations in some of the samples.

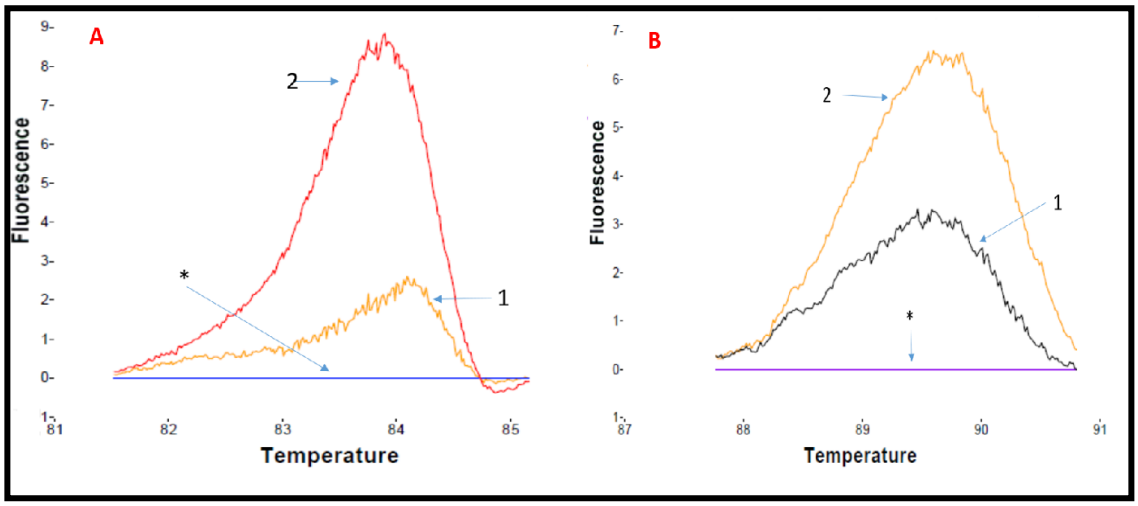

Figure 3. 3: COLD-PCR was used to enrich for low frequency mutations. This shows mutations which were detected by NGS. Initially undetectable by QMC-PCR/HRM (trace 1), it was confirmed by COLD-PCR/HRM (trace 2). (A) TP53 c.886C>T mutation. (B) PIK3CA c.1633G>A mutation. (*): wild.

Table 3. 1:List of SNPs identified in 17 paired samples (Tumour and matched normal tissue)

Samples	SNPs	Allele Frequency	
		Normal	Tumour
1	APC c.4479G>A	99.5	99.4
	PDGFRA c.1701A>G	99.7	99.6
	TP53 c.215C>G	99.7	99.6
	EGFR c.2361G>A	50.3	50.2
2	APC c.4479G>A	50.4	46.8
	TP53 c.215C>G	99.7	99.7
	EGFR c.2361G>A	50.6	50.4
3	APC c.4479G>A	49.7	48.3
	KIT c.2586G>C	49.7	49.3
	MET c.4071G>A	99.5	99.6
	TP53 c.215C>G	99.6	99.6
4	APC c.4479G>A	50.8	48.5
	PDGFRA c.1701A>G	99.7	99.7
	TP53 c.215C>G	50.6	50.7
5	APC c.4479G>A	99.4	99.5
	PDGFRA c.1701A>G	99.6	99.8
	MET c.4071G>A	48.3	48.78
	TP53 c.215C>G	99.7	99.6
6	APC c.4479G>A	99.4	99.6
	PDGFRA c.1701A>G	99.5	99.7
	TP53 c.215C>G	51.6	49.8
7	APC c.4479G>A	51.1	62.1
	TP53 c.215C>G	50.6	51.4
	STK11 c.816C>T	50.4	49.5
8	APC c.4479G>A	50.8	64.7
	TP53 c.215C>G	50.2	53.9
9	APC c.4479G>A	99.3	99.3
	TP53 c.215C>G	99.5	99.6
10	APC c.4479G>A	50.4	50.1
	KIT c.2586G>C	50.3	49.5
	TP53 c.215C>G	99.6	99.7
11	APC c.4479G>A	99.5	99.4
	PDGFRA c.2472C>T	48.8	48.6
	MET c.4071G>A	99.6	99.5
	TP53 c.215C>G	99.5	99.6
12	APC c.4479G>A	48.2	38.7
	PDGFRA c.2472C>T	48.9	48.9
	KIT c.2586G>C	48.9	49.6
	TP53 c.215C>G	51.1	50.3
13	APC c.4479G>A	50.8	54.1
	PDGFRA c.1701A>G	99.7	99.5
	KIT c.2586G>C	50.7	49.4
14	APC c.4479G>A	99.3	99.8
	TP53 c.215C>G	99.4	99.6
15	EGFR c.2361G>A	50.6	50.7
	TP53 c.215C>G	51.2	53.1
16	KIT c.2586G>C	50.1	49.5
	TP53 c.215C>G	99.6	99.7
17	APC c.4479G>A	49.1	36.0
	TP53 c.215C>G	50.4	24.6

3.4.1.6 Optimization of Cold-PCR

The purpose of this study was to optimise protocols to enrich and detect low mutant allele frequencies. Cold-PCR has recently been described as a variation on the usual protocol that can enrich for mutant alleles. As described in section (2.2.1.4.3.4) the study empirically derived the Tc for *KRAS* exon 2 on our thermal cycler and tested the full cold protocols on the spiked DNA mixtures. Tc is the lowest denaturation temperature that produced amplification for each exon, which is enough to melt heteroduplexes preferentially and allow better amplification. The limit of mutant allele detection was performed to identify the lowest level of mutant alleles were present in the samples. Variant quantities of DNA from the cell lines (HCT116 contains a heterozygous *KRAS* G13D mutation whilst Vaco5 is wild type for *KRAS*) were mixed to produce mixtures containing *KRAS* G13D mutant alleles at a frequency of 25%, 12.5%, 6.25%, 3.75%, 1.5% and 0.75% and the resulting mixtures were screened for *KRAS* G13D mutation. The results have shown that Cold-PCR/HRM increased the limits of detection down to 1.5% (figure 3.4). Additionally, the Difference curve of the HRM analysis shows the change in fluorescence of a test sample compared with a control which consists only of homoduplexes. The data show that the melting patterns between the samples were different and that the greater the proportion of mutant DNA (and therefore a higher proportion of heteroduplexes), the greater the difference in fluorescence between test sample and control. The differences were not linearly related to the mutant allele proportion and cannot therefore be used to precisely quantify mutant alleles. These results do however show that HRM can quantify the proportion of variant DNA in a mixture with wild type template.

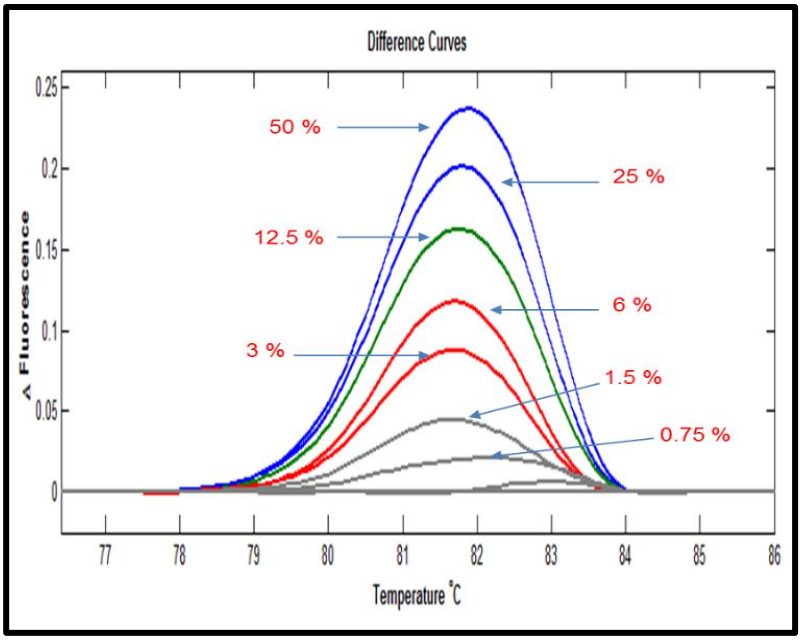

Figure 3. 4: Limit of detection of the PCR followed HRM was tested by spiking DNA from HCT116 into Vaco5 and testing for KRAS mutation. The figures show a difference plot confirming the differing quantities of spiked DNA. A difference of 4% is usually considered as a positive result and shows a limit of detection of 1.5 % mutant alleles for this mutation by Cold-PCR/HRM.

3.4.2 ED-POLE- and ED-POLD1 mutation analysis

In addition to the 26 genes that were analysed using NGS, exonuclease domain regions of *POLE* and *POLD1* were also screened for mutations using PCR followed HRM analysis.

3.4.2.1 Confirmation of primer specificity

3.4.2.1.1 Primer design and PCR optimisation

Primers were designed for a nested PCR screen of the entire exonuclease region of both *POLE* and *POLD1*. The exonuclease domain regions and the areas of the primers are shown in (figure 3.5). In order to get the optimal temperature for all outer and inner primers, they were tested on a gradient PCR machine. This showed that 60^{o}C is an optimal annealing temperature for all outer (8 pairs) *POLD1* primers while inner (8 pairs) *POLD1* and all outer (6 pairs) and inner (6 pairs) *POLE* primers relatively obtained clear and single band on gel electrophoreses at 55.5^{o}C annealing temperature (figure 3.6 and 3.7).

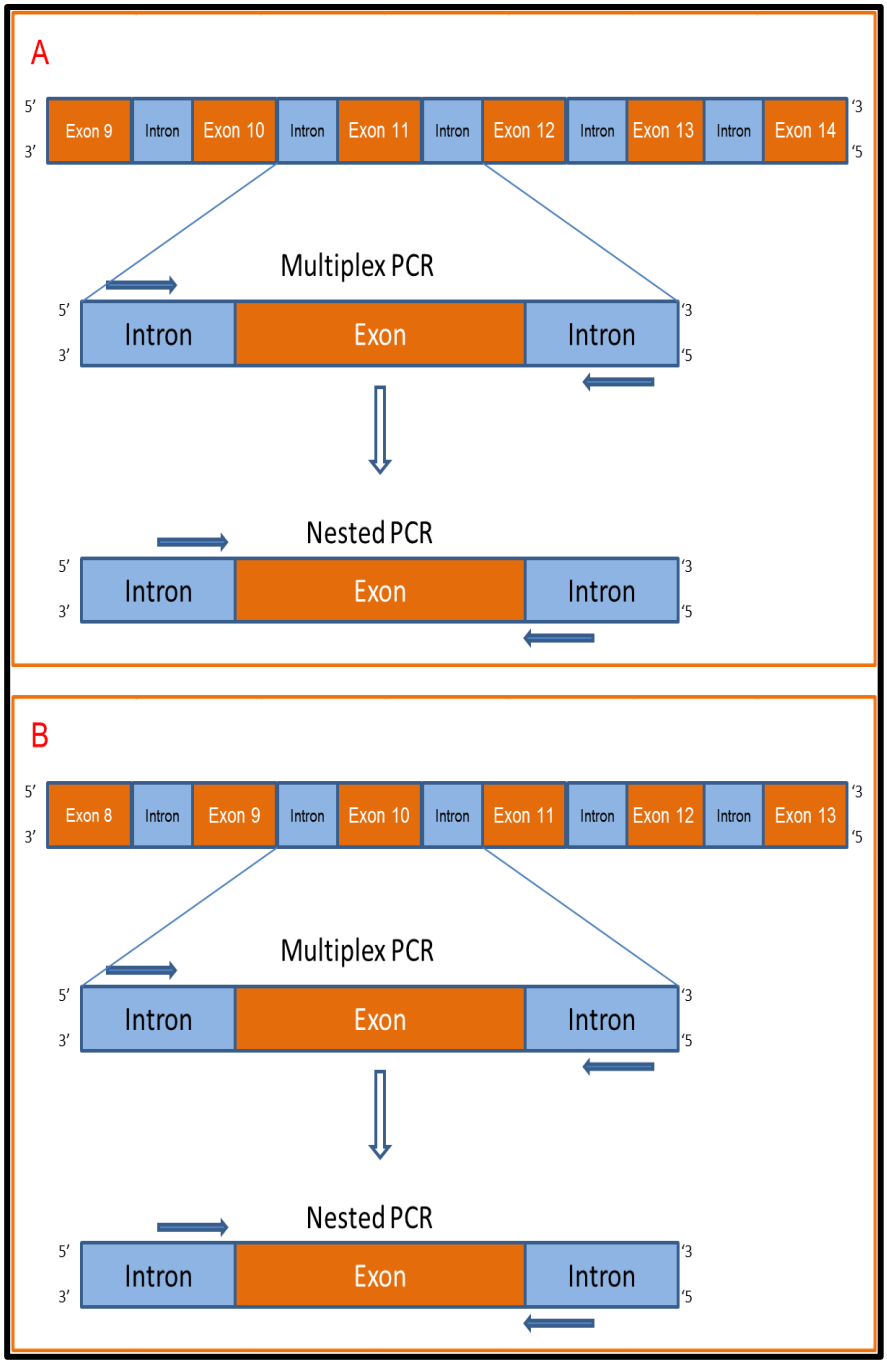

Figure 3. 5: Represent Exonuclease Domain Regions and the areas of the primers of (A) POLE and (B) POLD1 genes

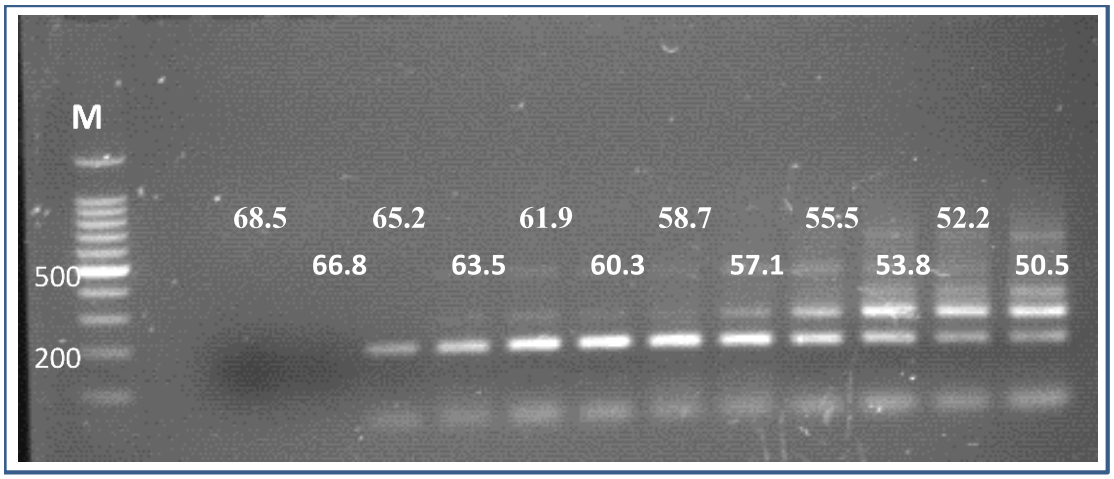

Figure 3. 6: Denaturing gradient gel electrophoresis of the PCR amplified exonuclease domain regions of POLD1 gene from colon cancer cell lines. Numbers at the top of the gel indicate temperature, and M = DNA marker.

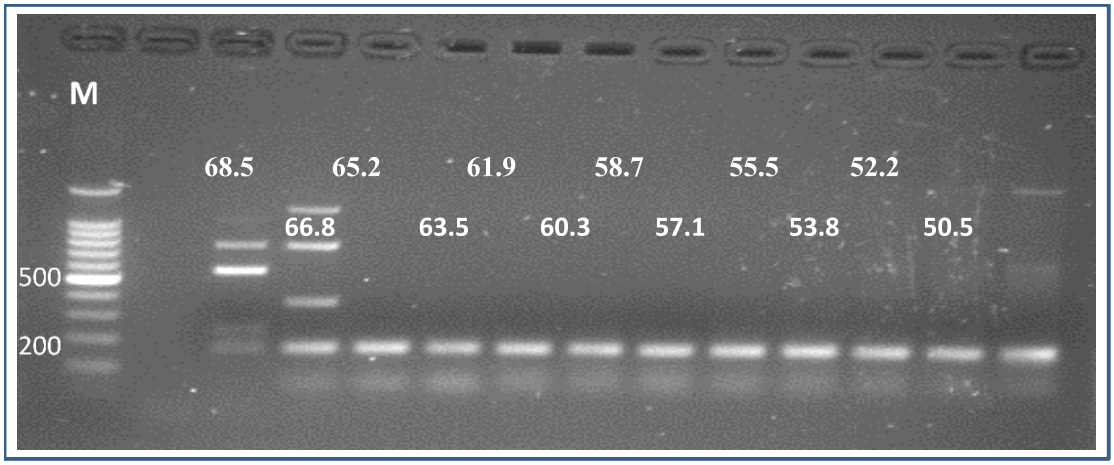

Figure 3. 7: Denaturing gradient gel electrophoresis of the PCR amplified exonuclease domain regions of POLE gene from colon cancer cell lines. Numbers at the top of the gel indicate temperature, and M = DNA marker.

3.4.2.2 QMC-PCR

After detecting the best annealing temperature, the study used the QMC-PCR protocol to amplify all exons that cover the genomic regions encoding the exonuclease domain regions of human *POLE* and *POLD1* genes in DNA extracted from cell lines, 83 primary CRCs, 22 matched metastases and 17 normal corresponding tissues. As shown in (figure 3.8 and 3.9) the products of

the nested PCR primers, which were visualized by electrophoresis on 1.5% agarose gel, obtained nice and clear single band with accurate size and specificity and indicated the high efficiency of QMC-PCR protocol used in this study.

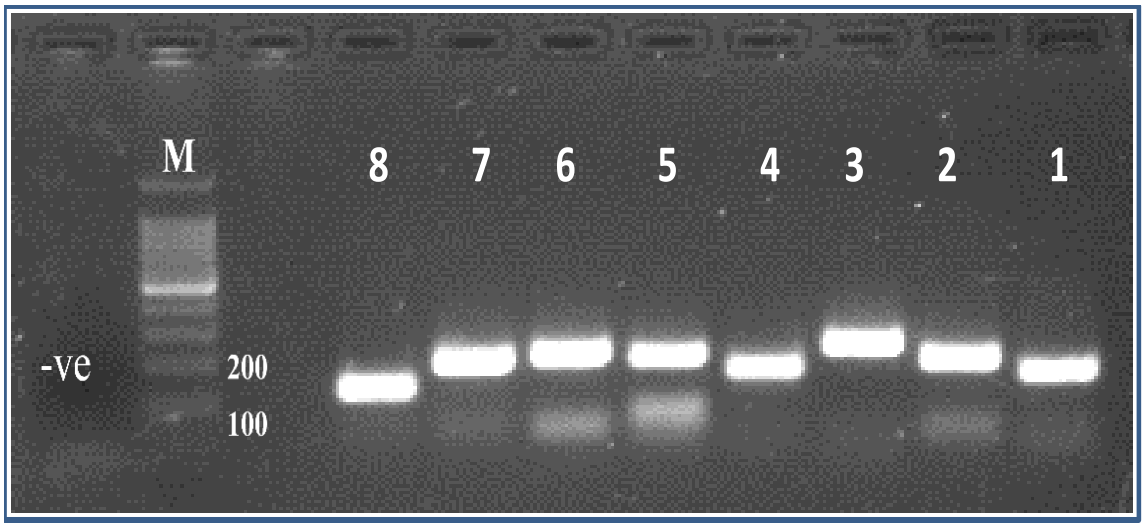

Figure 3. 8: PCR products represent all exons amplification (separately) involved in exonuclease domain region in POLD1 gene. Lane 1&2: Exon 8. Lane 3: Exon 9. Lane 4: Exon 10. Lane 5: Exon 11. Lane 6: Exon 12 Lane 7&8: Exon 13. M: DNA ladder. -ve: Negative control.

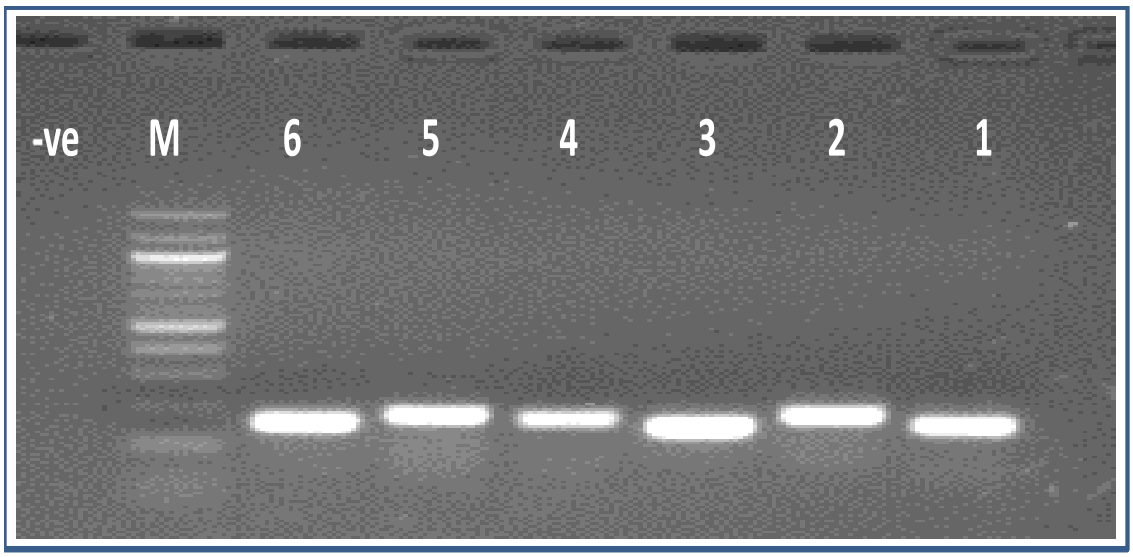

Figure 3. 9: PCR products represent all exons amplification (separately) involved in exonuclease domain region in POLE gene. Lane 1: Exon 9. Lane 2: Exon 10. Lane 3: Exon 11. Lane 4: Exon 12. Lane 5: Exon 13 Lane 6: Exon 14. M: DNA ladder. -ve: Negative control.

3.4.2.3 Validate HRMA

To validate the HRM analysis methodology, the study analyzed two colon cancer cell lines with known *POLD1* mutations; two missense mutation in Lovo cell line at exon 8 [CTG Val 312 Met ATG] and in DLD-1 cell line at exon 13 [CGC Arg 506 His CAC] (20). As shown in figure 3.10 A&B the difference plot generated after normalizing and temperature shifting, depict significant deviations from mutant samples in relative to the spread of other cell lines (SW48, HT29, HCT116 and HCA) which have wild type DNA sequence in both exons 8 and 13 and also the aberrant melting curve was confirmed using Sanger sequencing (figure 3.10 C&D).

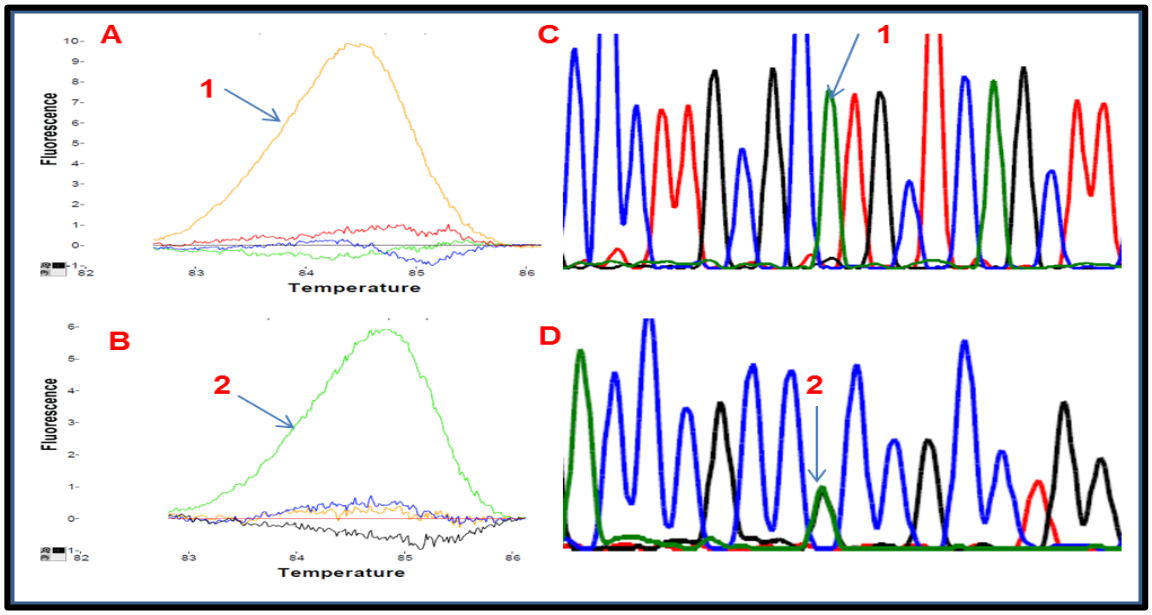

***Figure 3. 10: HRM analysis and sequencing results obtained from Lovo and DLD-1 cell lines with known mutations in exonuclease domain region in POLD1 gene. (A) Difference plot of* POLD1 *exon 8 conducted on five cell lines showing how one of the cell lines containing a mutation (Lovo) melts differently from the four wild-type alleles. (B) Difference plot of* POLD1 *exon 13 conducted on five cell lines showing how one of the cell lines containing a mutation (DLD-1) melts differently from the four wild-type alleles. (C and D) direct DNA sequencing confirms results obtained from QMC-PCR and HRM analysis techniques.* Black=G, Blue=C, Green=A, Red=T.**

3.4.2.4 Mutations found in the ED-POLE and ED-POLD1

DNA from 83 primary CRCs and 17 normal corresponding tissue samples were tested for mutations in *ED-POLE* and *ED-POLD1* with the optimised QMC-PCR followed HRM. In total, 45 samples looked aberrant for *ED-POLE* as follows (22 cases in exon 10 and 23 cases in other (9,11,12,13 and 14) exons, whereas only 25 (7 cases in exon 13, 18 cases in other exons) samples were looked aberrant for *ED-POLD1*. After the aberrant samples, had been sent for sequencing in total seven different mismatches were identified; 2 synonymous (in 2 different CRCs) and 5 non-synonymous causing amino-acid substitutions were identified in the *ED-POLE* of 5 CRCs as follow; in exon 9: c.834 G>T (T278T), in exon 13: c.1240 G>A (D414N) and c.1295 T>C (L432P). In exon 14; c.1365 G>A (L455L), c.1364 T>A (L455Q), c.1369 A> G (T457A) and c. 1380 C>A (S461Stop codon). In addition, the study found a previously reported SNP (rs4077170). While only two (non-synonymous) different missense mutations were detected in *ED-POLD1* of 2 matched primary and metastasis CRCs as follow; in exon 10: c. 1231 C>T (P404L) and c. 1229 C>T (A410V) and a synonymous mutations c.1230 T>C (A410A) in exon 10. Moreover, a previously reported SNP (rs2230245) was found c.1485 C>T in exon 13 (T495T) (Appendix B- figures 2.7). However, in 16 samples, a mutation was not detected by sequencing despite the presence of unequivocal aberrant melting on HRM. It was decided to perform cold PCR as an attempt to enrich for variant alleles and find the cause of aberrant melting pattern in these samples.

3.4.2.4.2 Cold PCR results

After detecting Tc as the lowest melting temperature for each exon the study performed cold PCR for those 16 samples. After comparing the HRM results obtained from both QMC-PCR and Cold-PCR, 9/16 samples showed different difference plot with Cold-PCR when compared with that obtained from QMC-PCR. Direct DNA sequencing picked up 4 mismatches in the *ED-POLE* gene among 4/9 samples as follow; in exon 9: c.833 C>T (T278M), in exon 12: c.1121 C>T

(A374V) and in exon 13: c.1298 G>A (G433D) and c.1314 G>A (E438E) (see figure 3.11).

In total and after screening 83 primary CRCs and 22 metastasis samples, 13 mutations were detected as follow, 11 mutations (3 synonymous and 8 non-synonymous) in *ED-POLE* and 3 mutations (1 synonymous and 2 non-synonymous) in *ED-POLD1*. No combination has been seen between *POLE* and *POLD1* genes (mutually exclusive) and only 4 samples (4/10) have a hypermutated phenotype, were they more likely to have full house of APC/KRAS/SMAD4/TP53 mutations, the rest 6/10 were more likely to have sub-clones (Details of mutations in *ED-POLE* and *ED-POLD1* are shown in Table 3-2). Additionally, two SNPs (rs4077170 and rs2230245) were also detected in *ED-POLE* and *ED-POLD1*, respectively.

Table 3. 2:Details of mutations in ED-POLE and ED-POLD1 and other genes

Subject	Gene/ Exon	Nucleotide position	Codon	a.a	MS status	Ploidy status	APC	KRAS	TP53	PIK3CA	SMAD4	PTEN	BRAF	MET	FBXW7
19	POLE/ 9	c.833	ACG	T > M	MSS	A	M	-	M	-	M	-	M	-	-
41	POLE/	c.1121	GCC	A > V	MSS	A	M	-	M	-	-	-	-	-	-
29	POLE/	c.1240	GAC	D > N	MSS	A	M	M	M	-	M	-	-	-	-
22	POLE/	C1295	CTA	L > P	MSS	A	M	M	M	-	M	-	-	-	-
37	POLE/13	c.1298	GGC	G > D	MSS	A	-	-	M	-	-	-	M	-	-
6	POLE/	c.1364	CTG	L > Q	MSS	A	M	M	M	-	M	-	-	-	M
66	POLE/	c.1369	ACG	T > A	MSS	D	M	M	-	-	-	-	-	-	-
83	POLE/	c.1382	TCA		MSI	D	-	-	M	-	-	M	-	-	-
45	POLD1/	c.1211	CCG		MSS	A	-	-	M	-	-	-	-	-	-
40	POLD1/	c.1229	GCC		MSS	A	-	-	M	-	-	-	-	M	-

M= mutant, A=Aneuploid, D=Diplod

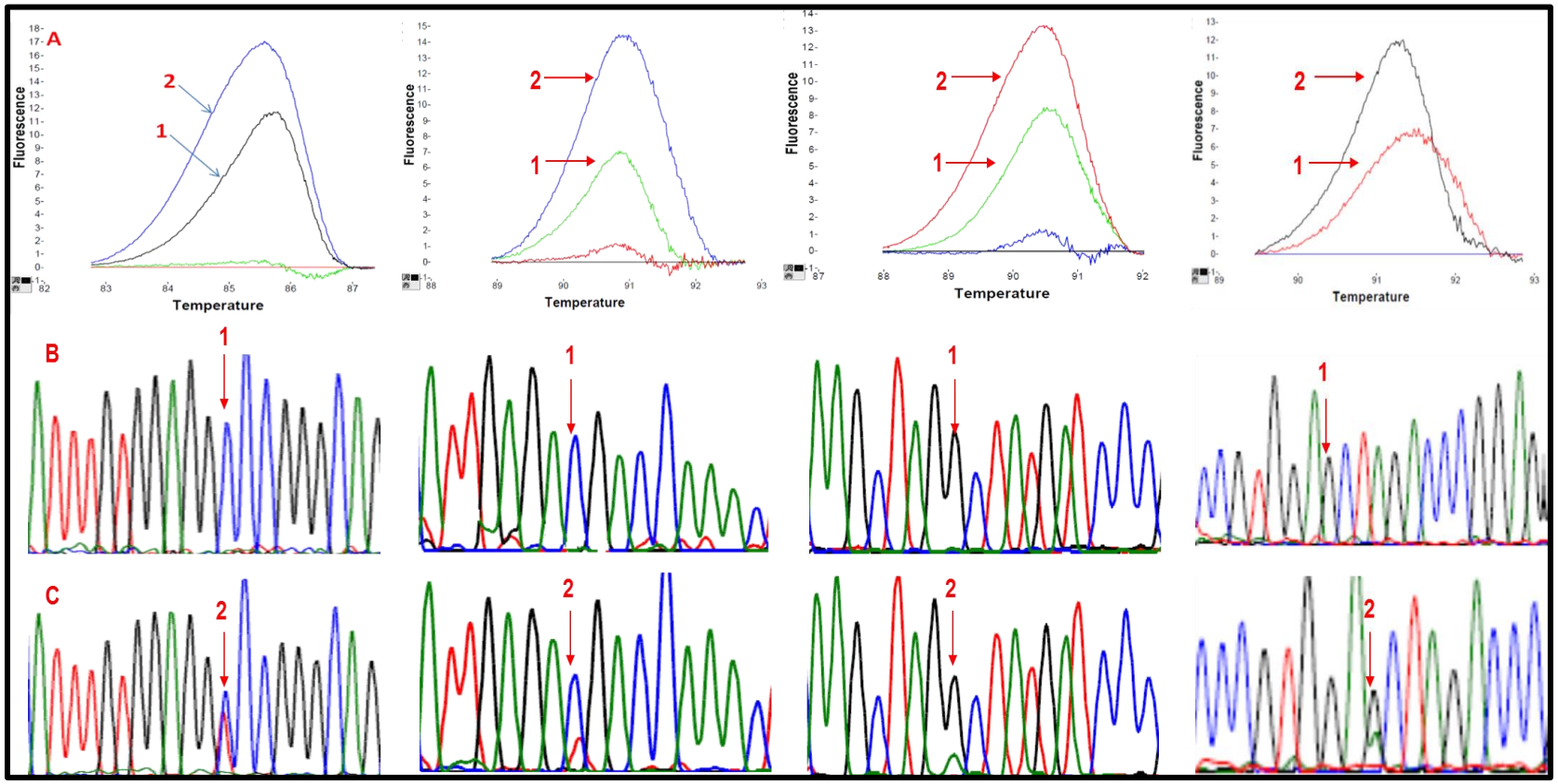

Figure 3. 11: comparison of difference plot and DNA sequencing obtained from QMC-PCR and Cold PCR protocol. Line 1 is a tumour samples obtained from QMC-PCR, line 2 is the same tumor sample obtained from cold PCR. In panel (A) the difference plot clearly show marked change between the results obtained for the same aberrant sample with two different protocols QMC-PCR and Cold PCR. In panel (C) clear enrichment of mutant allele as results of Cold-PCR can be seen if compared to that sequence obtained from QMC-PCR (B). Black=G, Blue=C, Green=A, Red=T.

3.4.2.6 Discriminate germline from somatic alterations

During screening *POLE* and *POLD1* genes for somatic mutations, we have understood that, conventional PCR followed HRM is not able to discriminate somatic from germline variants without analysing matched normal samples and applying Sanger sequencing, In this project, comparison of data obtained from tumour tissues with somatic mutations showed that, there were differences in the melting pattern between QMC-PCR- and COLD-PCR-amplified tumour samples (see figure 3-11).

Our hypothesis was based on the prediction that a minor allele (whether due to somatic mutation or LOH at a SNP locus) will undergo enrichment with COLD-PCR thus altering the HRM pattern whilst a heterozygous allele (at a frequency of 50%) will remain unchanged. The lack of change was seen in the experiments with the SNP in *POLE* SNP rs4077170 (G>C) in a series of 10 cases of matched tumour and normal. In all 20 samples, the melting pattern was identical whether QMC-PCR or COLD-PCR was performed (figure 3.12A). Sequencing of 10 of the cases (5 tumour and 5 normal samples) similarly showed that there was no change in the ratio of alleles (figure 3.12 B-E: only 1 normal and 1 tumour samples shown).

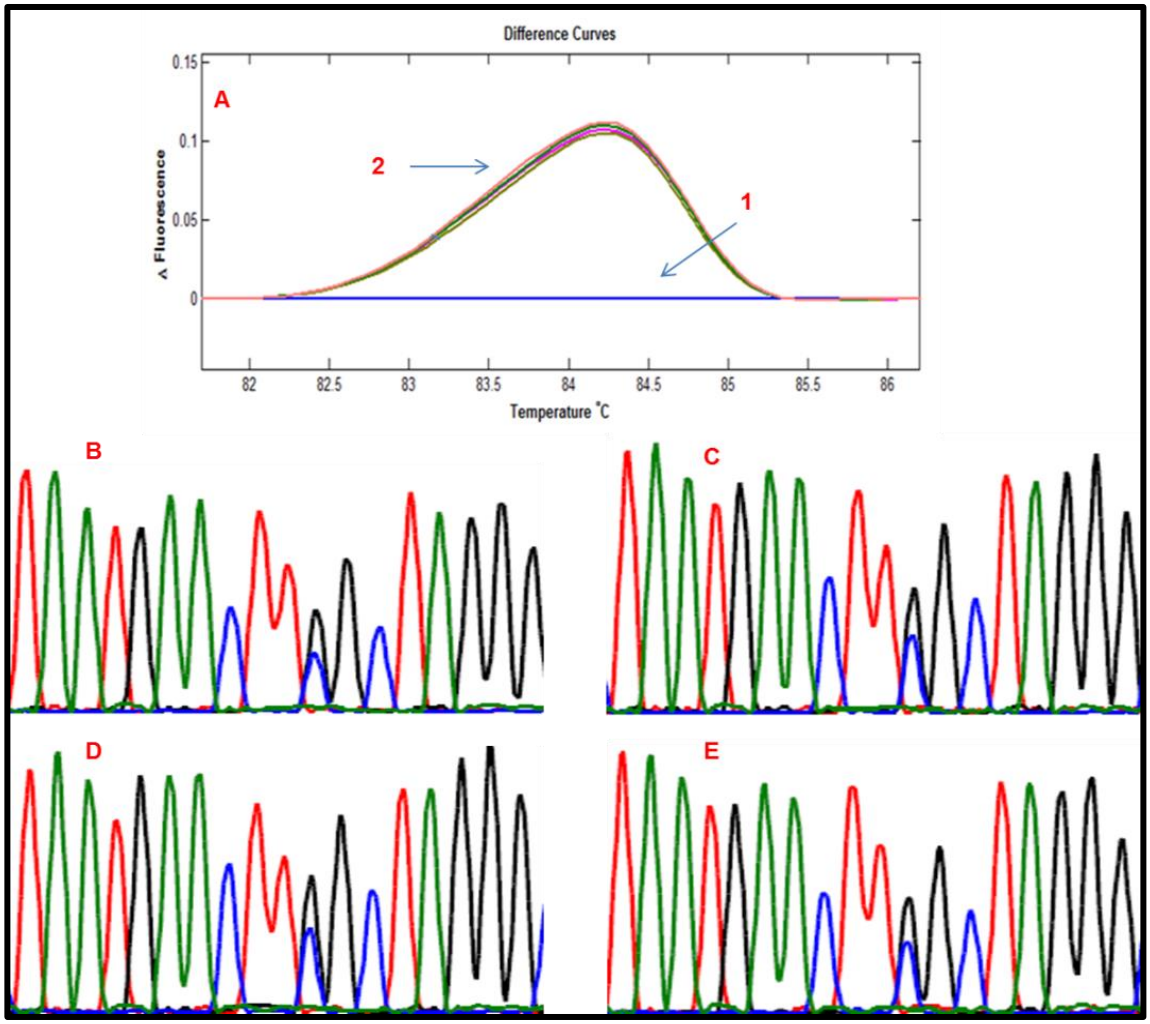

Figure 3. 12: HRM of tumour and matched normal sample having the rs4077170 SNP at POLE showing no significant change in melting pattern between standard PCR- and COLD PCR-amplified products. 1: Normal without SNP, 2: matched normal and tumour with the rs4077170 SNP.Electropherogram of 2 samples with POLE rs4077170 SNP. No change is seen in the peak of any of the alleles between standard PCR- and COLD PCR-amplified products. B: Tumour (Standard PCR), C: Normal (Standard PCR), D: Tumour (Cold-PCR), E: Normal Cold-PCR). *Black=G, Blue=C, Green=A, Red=T.*

3.4.2.7 COLD PCR-HRMA reveals allelic imbalance

Loss of Heterozygosity (LOH) is another type of somatic alteration in cancer [1-3]. Detection of these changes is important as it delineates the profile of a cancer and it may also provide prognostic and predictive information. Tumours frequently show allelic imbalance (AI) at gene loci. If a locus shows heterozygosity for SNPs, then AI will result in a change in the ratio of the SNPs, a move towards

homozygosity and creation of a minor allele. The latter feature allows AI to be tested by comparing tumour and matched normal tissue using HRMA. DNA samples from 20 cases of CRC and matched normal tissue were analysed for germline SNPs at the rs12455792 locus in SMAD4 and the rs9831477 locus in TGFBRII using standard and COLD-PCR amplification followed HRM. These samples were tested previously using standard PCR followed HRM and it was found that 11 out of 20 have SNP at the rs12455792 locus in SMAD4 and 7 out of 20 have SNP at the rs9831477 locus in TGFBRII. HRM of the products of both PCR methods were analysed for each sample. The results showed a differential melting pattern between the COLD PCR-amplified and the standard PCR-amplified products from the tumour samples, characterised by an increase in the height of the COLD PCR amplification products over those of the standard PCR (figure 3.13 and 3.15). The normal DNA samples did not show this change (figure 3.16). This is an indication of the enrichment of the minor allele in the tumour samples by the COLD PCR process. In 11/20 cases of CRC, had a SNP in SMAD4, and four of them (4/11 cases) showed a difference in the pattern between the COLD-PCR-amplified and the standard PCR-amplified products (figure 3.14). One sample which was homozygous at this locus was used as control. The amplified products were also subjected to Sanger sequencing. The results showed an increase in the height of the wild type C allele peak in three tumour samples, and an increase in the height of the variant T allele peak in the 4th tumour sample (figure 3.13). 7/20 cases of CRC showed SNP in TGFBRII and 3 of them (3/7 cases) showed a difference in the pattern between the COLD-PCR-amplified and the standard PCR-amplified products (figure 3.15). One sample which was homozygous at this locus was used as controls. (Sanger sequencing data not available for this set)

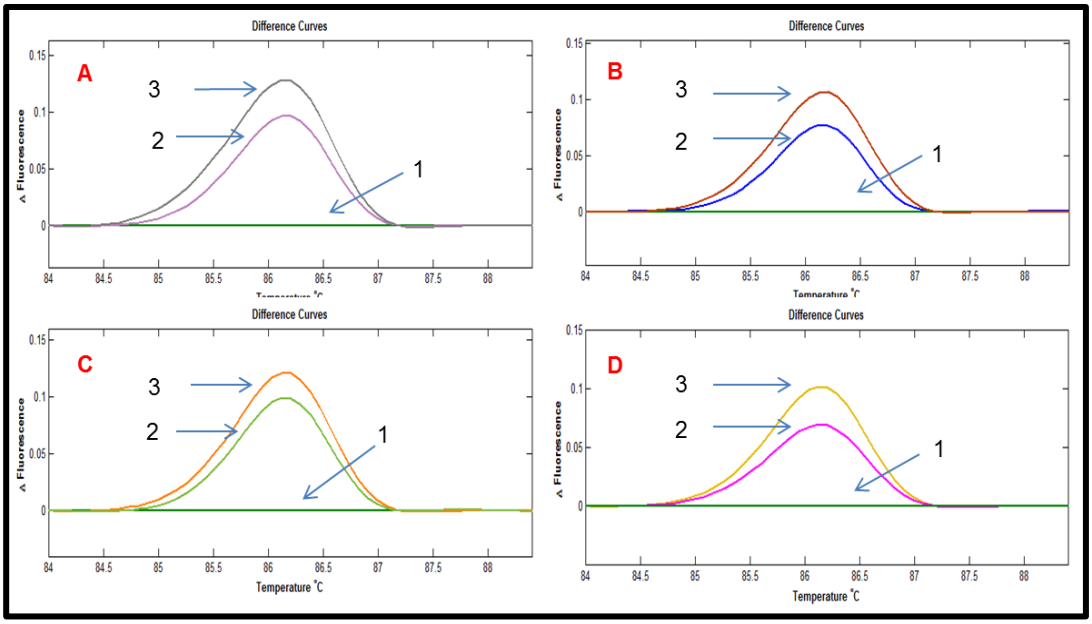

Figure 3. 13: HRM of the standard and full COLD PCR products of SMAD4 rs12455792 in homozygous reference, tumour and matched heterozygous normal DNA shows enrichment of the minor allele by COLD PCR in the tumour DNA, a confirmation of LOH in the tumour DNA at rs12455792 . 1: control without SNP, 2: normal with SNP (standard and Cold-PCR), 3: Tumour with SNP (standard-PCR), 4: Tumour with SNP (Cold-PCR).

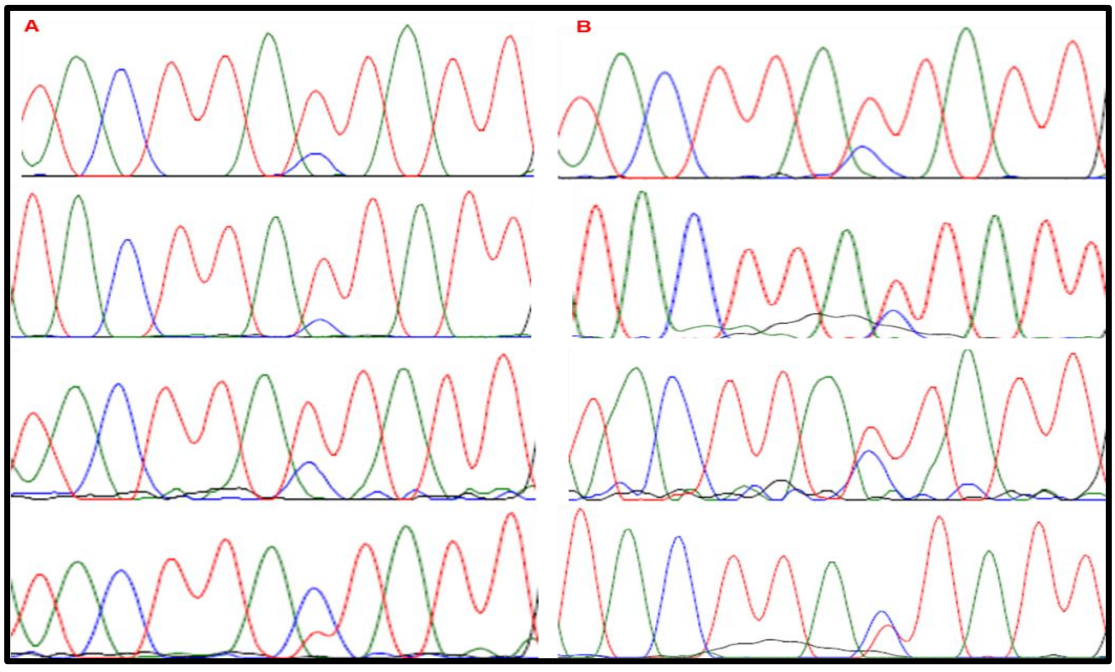

Figure 3. 14: Sanger sequencing confirms an enrichment of the minor allele in the tumour DNA sample (arrow). (A) heterozygous tumour samples amplified by standard PCR; (B) heterozygous tumour samples amplified by Cold-PCR. *Black=G, Blue=C, Green=A, Red=T.*

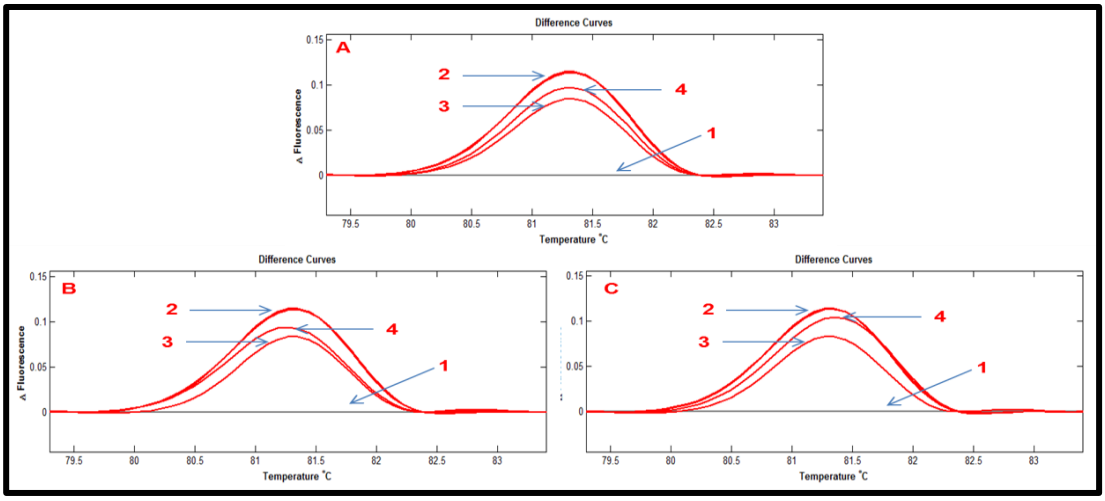

Figure 3. 15: HRM of the standard and full COLD PCR products of TGFBRII rs9831477 in homozygous reference, tumour and matched heterozygous normal DNA shows enrichment of the minor allele by COLD PCR in the tumour DNA, a confirmation of LOH in the tumour DNA at rs9831477. 1: control without SNP, 2: normal with SNP (standard and Cold-PCR), 3: Tumour with SNP (standard-PCR), 4: Tumour with SNP (Cold-PCR).

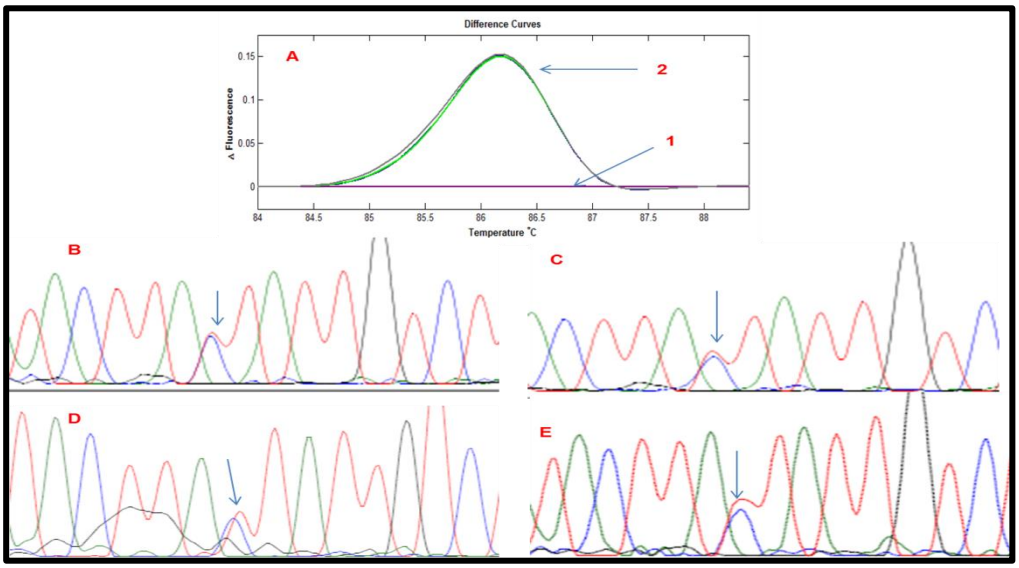

Figure 3. 16: HRM of the standard and full cold PCR products of SMAD4 rs12455792 locus in homozygous reference, tumour and matched heterozygous normal DNA showing tumour and a normal which are heterozygous SNPs which do not change in HRM pattern and which show identical peaks in the sequencing. (A)1: control without SNP; 2: normal and matched tumour with SNP (standard and Cold-PCR). Electropherogram of normal and matched tumour samples with SMAD4 SNP. No change is seen in the peak of any of the alleles between standard PCR- and COLD PCR-amplified products. B: Normal (Standard PCR), C: Tumour (Standard PCR), D: Normal(Cold-PCR), E: Tumour(Cold-PCR). Black=G, Blue=C, Green=A, Red=T.

3.4.3 Overall mutation profile

Eighty-three primary CRC samples were analysed for the presence of somatic mutations in the 26 genes of the Illumina TruSight panel, and mutations in *ED-POLE* and *ED-POLD1* using PCR followed HRM. Only 5/83 (6%) tumours showed no evidence of mutation in any of the genes included in the panel. The frequencies of mutations detected in the sample set are listed in detail in (table 3.3). It is of interest that only 28/83 (33.7%) of tumours contained the full house of *APC*/ (*KRAS* or *BRAF* or *NRAS*)/*TP53* mutations of the Fearon and Vogelstein pathway and approximately 93% CRCs have mutation in at least one of *APC/TP53/KRAS/BRAF/SMAD4/PIK3CA//FBXW7* gene. Furthermore, the frequency of *APC* mutations (47/83, 57%) was lower than that of *TP53* mutations (62/83, 75%). Overall approximately 8.4% (7/83) had a profile of *TP53* only mutation. Analysis of the association of mutations showed a significant positive association between mutations of *APC* and *KRAS* (p=0.01) and 33/47 (70%) tumours with an *APC* mutation had a co-incident *KRAS* mutation. Associations were also seen between *APC* and *PTEN* (p=0.01), *KRAS* and *PIK3CA* (p=0.01), *KRAS* and *SMAD4* (p=0.02) and *PIK3CA* and *GNAS* (p=0.009). There was no association between APC and TP53 mutation or KRAS and TP53 mutations. As expected, mutations in *KRAS* and *BRAF* were mutually exclusive. Details of all mutations are listed (see Appendix B- Table 1).

Table 3. 3:Mutation profiles of CRCs

Gene	Mutation status	Overall
KRAS	Wild type	39 (47%)
	Mutant	44 (53%)
GNAS	Wild type	79 (95%)
	Mutant	4 (5%)
PIK3CA	Wild type	64 (77%)
	Mutant	19 (23%)
PTEN	Wild type	75 (90%)
	Mutant	8 (10%)
CDH1	Wild type	82 (99%)
	Mutant	1 (1%)
FGFR2	Wild type	81 (97%)
	Mutant	2 (3%)
BRAF	Wild type	76 (91%)
	Mutant	7 (9%)
APC	Wild type	36 (43%)
	Mutant	47 (57%)
FBXW7	Wild type	73 (88%)
	Mutant	10 (12%)
SMAD4	Wild type	71 (85%)
	Mutant	12 (15%)
TP53	Wild type	21 (25%)
	Mutant	62 (75%)
KIT	Wild type	81 (97%)
	Mutant	2 (4%)
NRAS	Wild type	80 (96%)
	Mutant	3 (3%)
CTNNB1	Wild type	82 (99%)
	Mutant	1 (1%)
ET	Wild type	82 (99%)
	Mutant	1 (1%)
ED-POLE	Wild type	75 (90.4%)
	Mutant	8 (9.6%)
ED-POLD1	Wild type	81 (97.6%)
	Mutant	2 (2.4%)

3.4.4 Clonal heterogeneity

Tumour heterogeneity may hinder personalized molecular-target treatment that depends on the somatic mutation profiles, very recently several studies have been performed to detect heterogeneity by analysing mutations in multiple spatially separated samples obtained from multi-regional primary CRC [386, 387]. Founder mutations (i.e. those occurring early in the development of a tumour) would be expected to be present in all tumour cells whilst mutations occurring later may only be present in a small sub-clone. Working on the assumption that *APC* mutation, when present, is likely to represent the founder mutation, a ratio of the mutant allele frequency of *APC* to that of the other tumour suppressor genes (*TP53*, *SMAD4*, *PTEN* and *FBXW7*) was calculated. In the complete data set, the mean ratio was 1.06 with a standard deviation of 0.73 and thus ratios of >1.79 or <0.33 were considered to represent minor sub-clones. In total, 11/44 (25%) tumours showed the presence of a sub-clone indicating heterogeneity. In 8 of these cases the ratio was >1.79 thus supporting the assumption that *APC* mutation is usually a founder mutation (table 3.4).

3.4.5 Allelic Imbalance

Quantification of heterozygous SNPs in *APC* and *TP53* was used to test for allelic imbalance. From Knudson's two hit hypothesis [388], it can be inferred that allelic imbalance at these loci represents loss of heterozygosity. Based on the precision testing, SNP frequencies between 48–52% were considered to show heterozygosity and values outside this range were considered as showing allelic imbalance. The NM_000038.5:c.4479G>A SNP in *APC* gene was present in 34/83 (41%) CRCs and allelic imbalance of this SNP was seen in 16/34 (47%) of CRCs. The NM_000546.5:c.215C>G SNP in *TP53* was present in 19/83 (23%) CRCs and allelic imbalance of this SNP was seen in 10/17 (59%) of CRCs. In the tumours overall, allelic imbalance was associated with mutation of the tumour suppressor (Table 3.5, $p < 0.001$).

Table 3. 4:Mutant allele frequency in tumour suppressor genes among CRC cases

	Frequency of Mutant Alleles (%)					Ratio of mutant allele frequency			
No.	**APC**	**TP53**	**SMAD4**	**PTEN**	**FBWX7**	**APC:TP53**	**APC:SMAD4**	**APC:PTEN**	**APC:FBWX7**
1	10	11				0.91			
2	21	32			7	0.66			3.00
3	32	47				0.68			
4	12	27				0.44			
6	30	51	53		32	0.59	0.57		0.94
7	30	42			44	0.71			0.68
8	47	24			48	1.96			0.98
11	17	52	11			0.33	1.55		
19	16	24	23			0.67	0.70		
21	28	25		37		1.12		0.76	
22	35	39	9			0.90	3.89		
25	37	41				0.90			
26*		38			47				0.81
27	58	55				1.05			
28	26	14		30		1.86		0.87	
29	27	42	36			0.64	0.75		
33	30	57				0.53			
34	25	38			19	0.66			
41	15	30				0.50			
44	23	34				0.68			
46	26	31				0.84			
48	5	6		11		0.83		0.45	
50			46		48				0.96
52	25	14				1.79			
53	6	6				1.00			
55		13			12				1.08

	Frequency of Mutant Alleles (%)					Ratio of mutant allele frequency			
No.	**APC**	**TP53**	**SMAD4**	**PTEN**	**FBWX7**	**APC:TP53**	**APC:SMAD4**	**APC:PTEN**	**APC:FBWX7**
57	7	6	5	21		1.17	1.40	0.33	
59	3	6	4			0.50	0.75		
61	24	35				0.69			
63	70	76		69		0.92		1.01	
65	19	22	28			0.86	0.64		
67	24		11				2.18		
68	23	19				1.21			
69	36	24				1.50			
70	6	35				0.17			
71	10	15		7		0.67		1.43	
72	19	27				0.70			
73	38	41				0.93			
75	25	33				0.76			
78	37	21				1.76			
80	31	8		43		3.88		0.72	
81	36	37				0.97			
82	27	31			10	0.87			2.7
83* **		15		14					

***, **, ***= samples without APC mutations**

Table 3. 5:Allelic imbalance at the APC and TP53 loci

APC c.4479G>A				TP53 c.215C>G			
Sample ID	**Mutation**	**AI**	**SNP%**	**Sample ID**	**Mutation**	**AI**	**SNP%**
1	N	N	49.26	5	N	N	48.46
5	N	N	48.31	10	Y	Y	63.53
6	N	N	48.24	14	Y	Y	75.14
9	Y	Y	31.16	19	N	N	50.79
10	Y	Y	62.1	20	Y	Y	67.12
11	Y	Y	64.73	22	N	N	51.66
13	N	N	50.37	28	N	N	49.04
16	N	N	48.25	34	Y	Y	71.6
17	N	N	52.04	39	N	N	50.55
19	N	N	47.75	42	Y	Y	69.83
21	N	Y	62.75	45	Y	Y	76.41
26	Y	Y	62.94	52	N	N	51.12
28	N	N	49.29	62	N	N	51.79
31	N	N	49.14	75	Y	Y	57.46
32	Y	Y	78.07	76	Y	Y	31.86
34	Y	Y	41.17	77	Y	Y	70.95
36	Y	Y	65.71	78	N	N	51.85
37	Y	Y	23.91	79	N	N	52.07
39	N	N	49.05	83	Y	Y	53.7
43	N	Y	76.56				
49	N	N	49.97				
52	Y	Y	35.64				
53	N	N	49.12				
55	N	N	52.99				
56	N	N	49.53				
58	N	N	50.31				
63	Y	Y	21.28				
64	N	N	49.43				
68	Y	Y	88.54				
72	Y	Y	41.5				
77	Y	Y	31.2				
78	N	N	51.24				
82	Y	Y	36.03				
83	N	N	48.3				

N= Wild (no mutation), Y= Mutant

3.4.6 Validates the utility of diagnostic biopsies for predictive testing in colorectal cancer

To investigate whether diagnostic biopsy specimens are appropriate for predictive testing, 25 pre-surgical diagnostic biopsies that are matched to 25 of the resection specimens were evaluated for the presence of mutations.

3.4.6.1 Comparison of biopsy and resection specimens

A total of 81 non-synonymous somatic variations in 10 genes were found in resection specimens and all but 3 (3.7%) were also present in the biopsy specimens (details of all mutations are shown in Appendix B- table 2). The mutant allele frequency was 1.03-fold higher in resection specimens than biopsies and all the missed mutations, were present in the resection specimen at frequencies of <4%. There was no mutation in the biopsy specimens that were not seen in the resection specimens. Although the paired biopsy and resection specimens among 25 cases were not identical 100%, chi-square test, shown that there is not a significant difference between them (p=0.11).

In the 3 discrepant cases, there were other mutations present which were seen in both biopsy and resection. As shown in (table 3.6), in case number 1, the biopsy was shown to be wild type for GNAS, whereas exon 8 mutation in GNAS was detected in the resection specimen. This case also had APC, KRAS, PIK3CA and TP53 (in exon5 and exon6) mutations and these were identical in the biopsy and resection specimens. In case number 2, the biopsy was shown to be wild type for GNAS, whereas an exon 8 mutation in GNAS was detected in the resection specimens. This case also had KRAS, TP53 and PIK3CA mutations and these were identical in the biopsy and resection specimens. Finally, in case number 3, the biopsy was shown to be wild type for exon 2 in TP53, whereas an exon 2 mutation in TP53 was detected in the resection specimens. This case also had TP53 mutation (in exon 5) and these were identical in the biopsy and resection specimens. Additionally, Allelic imbalance of (c.4479G>A) SNP in APC and (c.215C>G) SNP in TP53 genes were seen in 5 out of 12 (42%) and 4 out of 7 (57%) of CRC cases, respectively (in both biopsy and resection specimens).

Comparison of biopsy and resection specimens showed no overt difference in the frequency of allelic imbalance of the tumours. (table 3.7).

Table 3. 6:Summary of mutations detected in samples have shown discrepancy between biopsy and resection

Case No	Gene	Mutation	codon	Exon	Frequency of	
					Resection	Biopsy
1	APC	G>G/GA	1554	15	25.1	14.91
	KRAS	GGT>GTT	12	2	29.5	15.94
	TP53	CGC>CAC	175	5	13.53	20.14
	TP53	CGA>TGA	196	6	3.9	4.1
	PIK3CA	AAAG>AAAG/A	111	1	10.23	17.37
	GNAS	CGT>CAT	201	8	3.38	Nil
2	KRAS	GGT>GTT	12	2	10.15	4.82
	TP53	CTGGGAGCT>	79-81	4	7.38	4.6
	PIK3CA	TTT>TTT/AAA	83	1	6.59	3.77
	GNAS	TCT>TTT	205	8	3.05	Nil
3	TP53	CCT>CTT	8	2	3.22	Nil
	TP53	CGC>CAC	175	5	18.71	22.23

Table 3. 7:Loss of heterozygosity according to APC and TP53 genes

APC				TP53			
genotype	SNP c.4479G>A	Resection cases	Biopsy cases	genotype	SNP c.4479G>A	Resection cases	Biopsy cases
M	No AI	48.3	50.5	M	AI	74.1	75.1
W	No AI	48.2	50.7	M	AI	40.2	33.7
W	No AI	47.7	47.4	W	No AI	51.1	52.8
M	AI	38.3	35.6	W	No AI	51.0	51.2
W	No AI	49.1	51.4	M	AI	24.62	34.2
W	No AI	51.9	50.7	M	AI	63.53	64.22
M	AI	37.2	31.2	M	No AI	52.95	52.39
M	AI	62.9	62.0				
M	AI	64.7	63.5				
M	AI	36.0	24.6				
W	No AI	48.3	49.9				
W	No AI	50.4	47.2				

M= mutant, W= Wild, AI= Allelic imbalance

3.5 Discussion

Nowadays the availability of genetic information plays a great role in making decisions, such as in treatment of solid tumours including EGFR mutation in non-small cell lung cancer, KRAS mutation in colorectal cancer and ERBB2 (HER2) gene amplification in breast and gastric cancer [389-392]. There are many different methods which can be used to obtain this information, each with its own strengths and weaknesses. The most efficient method for obtaining maximum genetic information is NGS and this is likely to emerge as the main technology used for global profiling of tumours.

NGS

Since only low quality DNA template derived from FFPE tissue will be available for many cases, it is important that any proposed technology can work with this type of template. As part of our experimental work, we performed a technical evaluation of the Trusight targeted sequencing assay. We found this to be a robust assay producing accurate and reproducible data with FFPE-derived DNA template. The mean sequencing depth overall was 15509 (ranging from means of 1008 – 25401 for lowest and highest performing amplicons) and short term and long term precision assays showed an acceptable coefficient of variation. The data obtained from the spiking experiments with the cell lines has shown that trusight tumour kit has a limit of detection that allowed confident detection of mutation when ~3% mutant alleles despite very deep sequencing. This is much better than the current 'gold standard' of Sanger sequencing (which has a limit of detection of 20% [333]) and it probably does mean that, in routine practice, microdissection of tissue sections will not be necessary. This particular methodology sequences both strands which allowed us to quantify of deamination artefacts at 24.4% and technical sequencing errors at 8.9%. The former is rather high and both reiterate the caution needed when interpreting sequencing data from FFPE tissue. All 255 somatic mutations detected by NGS were validated using QMC-PCR and HRM or Cold-PCR when the mutant allele frequency was less than 5% (the limit of detection of PCR-HRM) confirming that there was no

false positive arising from the technology despite the massive scale of sequencing. Additionally, we have demonstrated that both this NGS kit and platform make a reliable and sensitive assay for testing of tumours with low quality DNA template.

Mutation profile

The main purpose of the work was to evaluate the mutation profile in a series of CRCs. Overall, the frequency of gene mutation was within the range of that which has been published for tumours [25, 49, 393-395]. When viewed globally, a number of interesting features emerge from our data. Firstly, only one third of the tumours showed the classical profile of tumours developing in accordance with the Fearon and Vogelstein model i.e. combined disruption of the Wnt/MAPK/P53 pathways. In addition, the most common mutations were in *TP53* with a mutation frequency of 75% whilst *APC* mutation was found in 57% of tumours. This is consistent with data published by Smith *et al.* [7] and Gay *et al*. [396] although other studies have found higher frequencies of *APC* mutation. Interestingly, 11% of tumours had mutation of *TP53* only (without mutation in any other genes). These data reinforce the importance of *TP53* in CRCs and show that, although a significant number of tumours develop in accordance with the Fearon and Vogelstein model, there are other pathways. It is uncertain whether these involve deregulation of the Wnt or MAPK signalling pathways through disruption of other components (such as R-spondin, RNF43a, ZRNF3) or whether some CRCs can develop without involvement of these pathways.

In the present study, the frequency of somatic ED-*POLE* mutations (9.6%) was relatively in the range reported previously in the literature (3-12.3%) [229, 397] whereas a lower frequency of somatic *ED-POLD1* mutations (2.4%) was detected. Additionally, the study observed that only one tumour sample with T457A in *ED-POLE* was MSI, whereas the rest of tumour with mutations in *ED-POLE*- and *ED-POLD1* were MSS, with these data being in agreement with those reported by the TCGA project [229] and Palles et al, [227]. It is therefore likely that CRCs with *ED-POLE and ED-POLD1* mutations define a hypermutated

phenotype and MSS. However, most of those 10 cases (9/10) were detected as chromosomal unstable. There was not any association between *ED-POLE* and *ED-POLD1* mutations (mutually exclusive) and only four sample 4/10 with *ED-POLE* and ED-POLD1 mutations, have the full house of APC/KRAS/SMAD4/TP53 mutations. So, it is not clear whether this group of tumours follows one of the previous established subtypes of CRC pathways or it represents a new pathogenetic pathway.

Quantification of mutant alleles and SNPs allowed us to examine the tumours for the presence of clonal heterogeneity and allelic imbalance. Overall the tumours showed a moderate amount of heterogeneity with 11/44 (25%) tumours showed the presence of a sub-clone. In most cases the clone containing the *APC* mutation was the dominant clone confirming *APC* mutation as an early event. In three cases, however the reverse was true demonstrating that *APC* mutation can occur as a later event. Additionally, allelic imbalance was usually seen at the *APC* and *TP53* loci and was tightly associated with gene mutation ($p < 0.001$), which indicates that tumours follow Knudson's two hit hypotheses for inactivation of tumour suppressor genes.

Cold-PCR/HRM

Cold-PCR is able to enrich 'minority alleles' from mixtures of wild-type and mutation-containing sequences. This study combined and optimized COLD-PCR and HRM (called "Cold-HRM") to firstly validate samples with low mutant allele frequencies. After that the limit of detection of HRM was increased down to 1% following Cold-PCR amplification. Cold-HRM validated the 9 mutations with low mutant allele frequencies on NGS and enriched minor mutant allele in POLE gene for 4 CRC cases which was validated by sequencing. Comparison of the data showed that in these cases, there were differences in the melting pattern between standard PCR- and COLD-PCR-amplified tumour samples in every case. Based on this result and based on the assumption that heterozygous SNPs would remain in a 50/50 ratio following COLD-HRM (since there was no minor allele), this study sought to use COLD-HRM to discriminate somatic sequence mutations from

germline SNPs. Both standard and COLD PCR were performed, and PCR results were validated by sequencing analysis in some of these cases. The *POLE* SNP was tested in a series of 10 cases of matched tumour and normal. In all 20 samples, the melting pattern was identical whether standard PCR or COLD-PCR was performed which was validated by sequencing. This data confirms that Cold-HRM can discriminate somatic from germline SNPs without the need to compare with matched normal DNA or using other methods such as Sanger sequencing. Additionally, since LOH like somatic mutation induces a state of major and minor alleles of the target gene [1-3], the present study applied Cold-HRM for the LOH analysis. For this purpose, we tested 12 CRCs, 11 of which were known to have a SNP in the *SMAD4* gene. In 4 cases (36.3%) there was a difference in melting pattern between standard PCR and COLD-PCR. Minor allele enrichment was confirmed by sequencing and thereby validating this method as a robust means of detecting LOH. This method is quicker than FISH and, since it does not require normal matched DNA samples, it is more flexible than testing for LOH using microsatellites. It is also quicker than massively parallel technologies (such as microarrays or next generation sequencing) and its role would be more in low throughput targeted research or diagnostics. In addition, it would have a role in laboratories where the budget is limited. Since LOH is a form of allelic imbalance, it is conceivable that the same principle could be used to test for gene amplification if a SNP is identified within the amplified region. In this case, the non-amplified allele may be the minor allele.

The method we have described is simple and robust but it does have some caveats. Firstly, the PCR must be done twice i.e. a standard protocol and the COLD-PCR protocol and this requires the identification of the Tc for each target. This is however outweighed by the advantage of not requiring normal DNA for each tumour. It is possible to calculate the Tc using software which can predict the melting temperature of heteroduplexes therefore removing the need to individually test each target [30]. Another caveat is that this method depends on the presence of a minor allele. In most cases, a mutant allele will be the minor allele due to the presence of stroma within a tumour. It is possible that, if there is

a high proportion of tumour cells and there are events such as copy-neutral LOH, the frequency of the mutant allele approaches 50%. This may give a false negative result although it is anticipated that this will not occur often.

Biopsy vs Resection

Using of neoadjuvant therapy in colon cancer is limited, whereas it is commonly used in other types of cancer. Now in the UK a trial of neoadjuvant therapy [Fluoropyrimidine Oxaliplatin and Targeted-Receptor pre-Operative Therapy (FOXTROT)] has been started in locally advanced colonic cancer [398]. If the Foxtrot trial is successful, it will require predictive testing on diagnostic biopsy specimens. Thus, a new question arises whether biopsy specimens, which represent a tiny fragment of the tumour, are adequately representative of the tumour to use in patient stratification.

Targeted NGS analysis, was performed on 25 pairs of pre-surgical biopsy and the subsequent resection specimens. The data were concordant for 78/81 mutations. However, 3 (3.7%) mutations were detected in the resection specimens but not found in the corresponding biopsy specimens. In these three cases, there were other mutations found in both the biopsy and resection indicating that the discrepancy was not due to insufficient tumour within the biopsy. The discrepancies are most probably due to sub-clonal heterogeneity and indicate that some sub-clonal mutations could be missed through insufficient sampling. Additionally, the frequency of the missed mutation is lower than the other mutations, it indicates that it must be a sub-clone. Non-significant differences (p>0.05) seen in median numbers of mutation per tumour were [in resection specimens: 3.12 (range 0–7), in biopsy specimens: 2.92 (range, 0-6)] and nor in mutated gene per tumour [in resection specimens: 2.64 (range 0–5), in biopsy specimens: 2.52 (range 0-5)]. Thus, it is possible to see a high concordance between resection and biopsy specimens. Results in the present study show a high concordance between biopsy and resection specimens and indicate that biopsies are generally adequately representative of the tumour and suitable for predictive testing.

Although, our data suggested that, biopsy specimens represent a feasible material increase the probability of sampling of the dominant clone, some factors should be considered when interpreting data from tumour biopsy specimens. For example, where the tissue taken from. Baldus et al. [399] demonstrated discrepancy in the frequency of mutations in KRAS, BRAF and PIK3CA by 8%, 1% and 5% respectively between the centre and the invasive edge of colorectal tumours [399], and one explanation of this discrepancy is invasive edges are probably more prone to stromal contamination than the central portions of the tumour. Another factor is related to the fact that some of the early driver mutations are occasionally selected late, and may be present in a sub-clone that has not yet gained dominance [400]. In addition, the limit of detection may affect the detection of mutations. Our study used the targeted NGS methodology, which we have shown is able to detect as few as 3% mutant alleles.

Quantification of SNPs allowed us to examine the tumours for the presence of allelic imbalance as well. Overall the tumours showed allelic imbalance was usually seen at the loci of tumour suppressor genes and was usually associated with mutation of the tumour suppressor.

Based on the fact that somatic alterations were not detected in matched normal mucosa of the tested cases, we could say that variants are true mutations and NGS can be reliably implemented in diagnostic tests.

In summary, the present study has shown that this particular targeted NGS platform is robust even with low quality DNA derived from FFPE tissue. Our data show that there is a large degree of overlap in the mutation profiles, the presence of heterogeneity and the frequency of allelic imbalance indicating a common genetic pathway for these tumour types in the early stages. Our data support the theory that aneuploidy is a late occurring feature in cancer progression although the cause of the aneuploidy has not been identified. We were also able to show that, diagnostic biopsy specimens are suitable for predictive testing. Additionally, combining COLD PCR with HRM (COLD-HRM) allows inference of LOH and somatic sequence mutation without the requirement for matched normal DNA.

The role of genomic instability in the development of cancer is a contentious issue, and in the past controversial studies showed a new pathway called microsatellite and chromosomal stable MACS- in colorectal cancer. The next chapter (chapter 4) investigates the mutation profile of MACS-CRC compared to CIN-MACS.

Chapter Four: Investigate genetic discrepancies between CIN- and MACS-CRCs

4.1 Abstract

Background: Approximately 85-90% of sporadic colorectal cancers (CRCs) are microsatellite stable. Of these, 60-70% are aneuploid and are considered to have chromosomal instability (CIN). The remainder are diploid and are termed 'Microsatellite and Chromosomal Stable' (MACS) tumours. Here we have compared the mutation profiles of CIN and MACS-type tumours using targeted next generation sequencing (NGS).

Materials and Methods: DNA (from formalin-fixed paraffin embedded tissue) was extracted from 46 CIN-CRCs and 35 MACS-CRCs (first cohort). Mutation profiling across 26 known cancer-associated genes was undertaken using the Illumina TruSight tumour kit. Mutations identified by NGS were validated using Quick-Multiplex-Consensus (QMC)-PCR and high resolution melting (HRM) analysis. The tumour groups were compared for mutation frequency, clonal heterogeneity and allelic imbalance. Additionally, exonuclease domain regions of both POLE and POLD1 genes were screened using QMC-PCR followed by HRM.

Results: Comparison of CIN vs MACS showed similar mutation frequencies of mutation in all genes except *KRAS (41%CIN vs 68%MACS), GNAS (0%CIN vs 11%MACS) and POLE (15% CIN vs 2%MACS).* Statistically there was a significant difference (each p=0.01) which was lost following multiple testing correction. Approximately 20% (9/46) CIN-CRCs displayed mutations on only TP53, a profile not seen in the MACS-CRCs (p=0.008). Quantification of mutant allele frequency showed clonal heterogeneity in 28% MACS-CRCs and 23% CIN-CRCs. Quantification of Single Nucleotide Polymorphisms (SNPs) in *APC* and *TP53* showed that, despite differences in nuclear DNA content, there was no difference in allelic imbalance at these loci.

Conclusion: The mutation profiles of CIN and MACS tumours are similar, suggesting that they evolve along similar genetic pathways in their early stages.

The events allowing aneuploidy (or forcing retention of diploidy) remain unknown and may occur after the mutations which cause malignant change.

4.2 Introduction

In recent years, a number of different molecular classifications for sporadic CRCs have been produced [2, 4, 401]. Although there is some variation, almost all classifications subscribe to the dogma of two major genetic pathways which can be discriminated by the nature of the genetic instability present in a tumour. Approximately 10-15% of sporadic CRCs are deemed to have Microsatellite Instability (MSI) whilst the remainder are considered to have Chromosomal Instability (CIN) [173, 402]. MSI is seen overwhelmingly in right sided (i.e. proximal to the splenic flexure) colonic tumours [403] and arises due to loss of mismatch repair (MMR) function. In most cases, the MSI phenotype arises as a result of epigenetic inactivation of the *MLH1* gene. Mutation of the *MLH1* gene itself (or other MMR genes such as *MSH2, MSH6* or *PMS2)* may be the cause in a small subset of MSI tumours. [136, 404-407]. These tumours have a characteristic genetic profile (displaying mutations in genes such as *TGF*β*1RII, IGFIIR and BAX* [137, 138]), they are associated with a CpG island methylator (CIMP) phenotype, and have a diploid or near-diploid genotype [408].

The cause of CIN is unknown, but CRCs with CIN (CIN-CRC) are characterised by large-scale chromosomal changes such as alteration in whole chromosome number, chromosomal deletion and translocation [20]. CIN-CRCs are characterised by an aneuploid genotype [233] and mutation in genes such as *APC, TP53,* and *FBXW7* [19, 22, 23, 408].

Until recently, tumours which did not have MSI were assumed to have CIN. However, a third subset of CRCs which have neither MSI nor CIN has been identified. These have become known as Microsatellite and Chromosome Stable (MACS) although they have also been reported in the literature as ‘X-type’ [409, 410]. Data on the proportion of sporadic CRCs displaying the MACS phenotype (MACS-CRC) are varied, with figures ranging between 17% and 45% [19, 410, 411]. Furthermore, data on the clinicopathological features associated with MACS-CRCs are conflicting and unclear. Silver *et al* [234] demonstrated that MACS-CRCs are more likely to be left sided whilst Cai *et al* [412] showed an association with the right colon and Hawkins *et al*. [230] failed to find any

association with location. In terms of the prognosis of MACS tumours, Kakar *et al* [411] demonstrated that MACS tumours had improved survival (compared to CIN tumours) whilst Hawkins *et al* [230] reported that MACS-CRCs had a poorer prognosis than both MSI and true CIN-CRCs .

As MACS is a recently described sub-group, genetic data are predictably sparse. MACS tumours are defined by an absence of MSI and a diploid or near-diploid genotype. They are reported to display a high rate of CpG island methylator phenotype, similar to that of MSI tumours [412]. They may display a lower rate of loss of APC protein compared to CIN tumours and a lower rate of loss of expression of MLH1 and BAX proteins compared to MSI tumours [412].

4.3 Aims

For this study, we sought to compare CIN-CRCs and MACS-CRCs using a targeted sequencing approach and tested a series of 26 cancer-related genes using the Illumina TruSight tumour kit. Additionally, screening ED-POLE and ED-POLD1 genes using HRMA.

4.4 Results

4.4.1 Comparison of CIN-CRCs and MACS-CRCs

4.4.1.1 Mutation profile

NGS analysis revealed similar mutation frequencies within the CIN and MACS tumour populations. Additionally, no significant clinico-pathological differences seen between them. There was no significant difference in the frequency of tumours which were wild type for all genes in the panel (3/35 (8%) MACS-CRCs and 3/46 (6%) CIN-CRCs). Approximately 20% CIN-CRCs had a profile of TP53 only mutation whilst this profile was not seen in the MACS-CRCs (9/46 vs 0/35, p=0.008). However, following multiple correction testing using the Bonferroni correction, it failed to retain significance p=0.1. There were also significant differences in the frequency of *KRAS*, *GNAS* and *POLE* mutation between the two tumour groups (p=0.02, p=0.03 and p=0.02 respectively). However, following multiple correction testing using the Bonferroni correction, they both failed to retain significance (p=0.3, p=0.4 and p=0.3 respectively, table 4.1).

4.4.1.2 Clonal heterogeneity

Clonal heterogeneity was tested in both groups as described in the previous chapter. In the complete data set (from 46 CIN- and 35 MACS- CRCs), the mean ratio was 1.04 with a standard deviation of 0.71 and thus ratios of >1.76 or <0.33 were considered to represent minor sub-clones. In total, 11/42 (26%) tumours showed the presence of a sub-clone and in 8 of these cases the ratio was >1.76 thus supporting the assumption that *APC* mutation is usually a founder mutation. Comparison between the two groups (tables 4.2 & 4.3) showed no difference with 6/21 (28%) of MACS-CRCs and 5/21 (23%) CIN-CRCs showing clonal heterogeneity.

Table 4. 1:clinic-pathological and molecular features of sporadic CRCs classified as CIN or MACS

		MACS-CRCs	CIN-CRCs	*p* value	Corrected *p* value
Sex	M	19 (54%)	25 (54%)	0.5	-
	F	16 (46%)	21 (54%)		
Age	Median	71	65	0.3	-
	Mean	72	66		
Dukes' stage	A/B	12 (43%)	16 (45%)	0.5	-
	C/D	16 (57%)	20 (55%)		
EMVI	V0	9 (26%)	15 (33%)	0.4	-
	V1	18 (51%)	23 (50%)		
	unknow	8 (23%)	8 (17%)		
Location	Rt side	10 (28%)	12 (26%)	0.4	-
	Lt side	25 (72%)	34 (74%)		
KRAS	Wild	11 (32%)	27 (59%)	0.02	0.3
	Mutant	24 (68%)	19 (41%)		
GNAS	Wild	31 (88%)	46 (100%)	0.03	0.4
	Mutant	4 (11%)	0 (0%)		
PIK3CA	Wild	24 (68%)	39 (85%)	0.08	0.7
	Mutant	11 (32%)	7 (15%)		
PTEN	Wild	30 (86%)	44 (96%)	0.1	-
	Mutant	5 (14%)	2 (4%)		
FGFR2	Wild	35 (100%)	44 (96%)	0.2	-
	Mutant	0 (0%)	2 (4%)		
CDH1	Wild	34 (98%)	46 (100%)	0.2	-
	Mutant	1 (2%)	0 (0%)		
CTNNB1	Wild	35 (100%)	45 (98%)	0.3	-
	Mutant	0 (0%)	1 (2%)		
MET	Wild	35 (100%)	45 (98%)	0.3	-
	Mutant	0 (0%)	1 (2%)		
BRAF	Wild	33 (95%)	41 (89%)	0.4	-
	Mutant	2 (5%)	5 (11%)		

		MACS-CRCs	CIN-CRCs	*p* value	Corrected *p* value
FBXW7	Wild	32 (92%)	40 (87%)	0.5	-
	Mutant	3 (8%)	6 (13%)		
APC	Wild	14 (40%)	21 (46%)	0.6	-
	Mutant	21 (60%)	25 (54%)		
TP53	Wild	10 (29%)	11 (24%)	0.6	-
	Mutant	25 (71%)	35 (76%)		
SMAD4	Wild	29 (83%)	40 (87%)	0.6	-
	Mutant	6 (17%)	6 (13%)		
KIT	Wild	34 (98%)	45 (98%)	0.8	-
	Mutant	1 (2%)	1 (2%)		
NRAS	Wild	33 (94%)	45 (98%)	0.8	-
	Mutant	2 (6%)	1 (2%)		
POLE	Wild	34 (98%)	40 (85%)	0.02	0.3
	Mutant	1 (2%)	6 (15%)		
POLD1	Wild	35 (100%)	44 (96%)	0.2	
	Mutant	0 (0.0 %)	2 (4%)		

Table 4. 2:Mutant allele frequency in tumour suppressor genes among MACS-CRC cases

	Frequency of Mutant Alleles (%)					Ratio of mutant allele frequency			
No.	**APC**	**TP53**	**SMAD4**	**PTEN**	**FBWX7**	**APC:TP53**	**APC:SMAD4**	**APC:PTEN**	**APC:FBWX7**
2	5	6		11		0.83		0.45	
4			46		48				0.96
6	25	14				1.79			
7	6	6				1.00			
9*		13			12				1.08
11	7	6	5	21		1.17	1.40	0.33	
13	3	6	4			0.50	0.75		
15	24	35				0.69			
17	70	76		69		0.92		1.01	
19	19	22	28			0.86	0.64		
21	24		11				2.18		
22	23	19				1.21			
23	36	24				1.50			
24	6	35				0.17			
25	10	15		7		0.67		1.43	
26	19	27				0.70			
27	38	41				0.93			
29	25	33				0.76			
32	37	21				1.76			
34	31	8		43		3.88		0.72	
35	36	37				0.97			

Table 4. 3:Mutant allele frequency in tumour suppressor genes among CIN-CRC cases

	Frequency of Mutant Alleles (%)					Ratio of mutant allele frequency			
No.	APC	TP53	SMAD4	PTEN	FBWX7	APC:TP53	APC:SMAD4	APC:PTEN	APC:FBWX7
1	10	11				0.91			
2	21	32			7	0.66			3.00
3	32	47				0.68			
4	12	27				0.44			
6	30	51	53		32	0.59	0.57		0.94
7	30	42			44	0.71			0.68
8	47	24			48	1.96			0.98
11	17	52	11			0.33	1.55		
19	16	24	23			0.67	0.70		
21	28	25		37		1.12		0.76	
22	35	39	9			0.90	3.89		
25	37	41				0.90			
26		38			47				0.81
27	58	55				1.05			
28	26	14		30		1.86		0.87	
29	27	42	36			0.64	0.75		
33	30	57				0.53			
34	25	38			19	0.66			
41	15	30				0.50			
44	23	34				0.68			
46	26	31				0.84			

4.4.1.3 Allelic Imbalance

Quantification of heterozygous SNPs in *APC* and *TP53* was used to test for allelic imbalance. From Knudson's two hit hypothesis [388], it can be inferred that allelic imbalance at these loci represents loss of heterozygosity. Based on the precision testing, SNP frequencies between 48–52% were considered to show heterozygosity and values outside this range were considered as showing allelic imbalance. In the tumours overall, allelic imbalance was associated with mutation of the tumour suppressor (Table 5-4, $p < 0.001$).

The NM_000038.5:c.4479G>A SNP in *APC* gene was present in 20/46 (43%) CIN-CRCs and 12/35 (34%) MACS-CRCs. Allelic imbalance of this SNP was seen in 10/20 (50%) of CIN and 5/12 (42%) of MACS groups.

The NM_000546.5:c.215C>G SNP in *TP53* was present in 11/46 (24%) CIN-CRCs and 7/35 (20%) MACS-CRCs. Allelic imbalance of this SNP was seen in 6/11 (54%) of CIN-CRCs and 3/6 (50%) of MACS-CRCs.

CIN-CRCs are characterised by large changes in chromatin but despite this, comparison of CIN-CRCs and MACS-CRCs showed no difference in the frequency of allelic imbalance of the tumours at *APC* and *TP53* (table 4.4)

Table 4. 4:Allelic imbalance at the APC and TP53 loci

	CIN-CRC				MACS-CRC		
APC c.4479G>A							
Sample ID	Mutation	AI	SNP%	Sample ID	Mutation	AI	SNP%
1	N	N	49.26	3	N	N	49.97
5	N	N	48.31	6	Y	Y	35.64
6	N	N	48.24	7	N	N	49.12
9	Y	Y	31.16	9	N	N	52.99
10	Y	Y	62.1	10	N	N	49.53
11	Y	Y	64.73	12	N	N	50.31
13	N	N	50.37	17	Y	Y	21.28
16	N	N	48.25	18	N	N	49.43
17	N	N	52.04	22	Y	Y	88.54
19	N	N	47.75	26	Y	Y	41.5
21	N	Y	62.75	31	Y	Y	31.2
26	Y	Y	62.94	32	N	N	51.24
28	N	N	49.29				
31	N	N	49.14				
32	Y	Y	78.07				
34	Y	Y	41.17				
36	Y	Y	65.71				
37	Y	Y	23.91				
39	N	N	49.05				
43	N	Y	76.56				
TP53 c.215C>G							
Sample ID	Mutation	AI	SNP%	Sample ID	Mutation	AI	SNP%
5	N	N	48.46	6	N	N	51.12
10	Y	Y	63.53	16	N	N	51.79
14	Y	Y	75.14	29	Y	Y	57.46
19	N	N	50.79	30	Y	Y	31.86
20	Y	Y	67.12	31	Y	Y	70.95
22	N	N	51.66	32	N	N	51.85
28	N	N	49.04	33	N	N	52.07
34	Y	Y	71.6				
39	N	N	50.55				
42	Y	Y	69.83				
45	Y	Y	76.41				

N= Wild (no mutant), Y=Mutant

4.5 Discussion

In this study, we compared the mutation profiles of CIN-CRCs and MACS-CRCs using targeted NGS and QMC-PCR followed by HRM. These are tumours which do not have microsatellite instability and which can be discriminated on the basis of ploidy status. We used the commercially available TruSight tumour kit (Illumina, USA) which allows simultaneous interrogation of 26 cancer-associated genes. Additionally, ED-POLE and ED-POLD1 were screened.

The microsatellite and ploidy status for all DNA used in this study were evaluated before by Dr. Wakkas Fadhil [233]. The tumours which had an aneuploid genotype were considered to show chromosomal instability (CIN) and were designated as CIN-CRCs whilst the diploid group was designated as MACS-CRCs because of their stable karyotype and the presence of both microsatellite and chromosomal stability (MACS). A series of 81 colorectal cancer (first cohort) were analysed using NGS, from that number 35 cases were diploid (MACS-CRCs) and 46 cases were aneuploid (CIN-CRCs).

The present study sought to evaluate molecular differences between MACS-CRCs and CIN-CRCs. Comparison of clinico-pathological features, has revealed no significant differences between these two groups of samples (CIN- and MACS-CRCs). However, other studies published mixed data and had shown a significant association of MACS-CRC group with tumour sidedness and prognosis of the disease. For example, Silver et al, [234] demonstrated that diploid non-MSI tumour is associated with the left side of the colon whilst other studies found a significant association of these tumors with the right side of the colon [413].

Comparison of the two groups showed that, for most of the genes in the panel, there was no difference in frequency of mutation between MACS-CRCs compared to CIN-CRCs. There was a difference in the proportion of tumours showing a profile of *TP53* mutation only with approximately 9/46 (20%) CIN-CRCs having this profile in contrast with 0/35 MACS-CRCs (p=0.008), However following multiple correction testing using the Bonferroni correction, it failed to retain significance p=0.1). The significance of this is uncertain and, despite an enrichment of tumours with this profile, there was no difference in the overall

frequency of *TP53* mutation between the two groups. Similarly, there was no difference in the frequency of *FBXW7* mutation between the two groups. Both of these genes has previously been associated with the development of aneuploidy [96, 414] but this role is not however supported by our study.

There was a slightly greater frequency of *KRAS, PIK3CA and GNAS* mutations in MACS-CRCs which was statistically significant for KRAS and GNAS (Fisher's exact test, p=0.02 and p=0.03 respectively) suggesting a preference of MACS tumours to development along a KRAS-dependent and GNAS-dependent pathway although this was lost after multiple testing correction. However, CIN-CRCs showed a greater frequency of ED-POLE mutation (p=0.02) which was lost after multiple testing correction. The large overlap in mutation profiles also suggests that factors contributing to aneuploidy most likely occur late during in the carcinogenic process, after occurrence of these mutations. This is consistent with the theory that aneuploid clones frequently occur in late stage adenomas [415-417].

As mentioned in chapter 3, mutant alleles and SNPs were quantified to detect the presence of heterogeneity and allelic imbalance in both groups and results demonstrated that in terms of heterogeneity comparison between the two groups showed no difference with 6/21 (28%) of MACS-CRCs and 5/21 (23%) CIN-CRCs showing clonal heterogeneity. Additionally, the lack of a difference in allelic imbalance was noticed. Presence of high association between allelic imbalance with gene mutations indicate that both tumour types follow Knudson's two hit hypotheses for inactivation of tumour suppressor genes. One can infer that the allelic loss in MACS-CRCs occurs due to small genetic changes which do not result in a large change in the nuclear chromatin content. Interestingly, MACS-CRC group is very similar with consensus molecular subtype 3 (CMS3), as it has frequent KRAS activating mutations and few somatic copy number alterations (SCNAs) [252].

In the present study, the frequency of somatic *ED-POLE* mutation was 15% in CIN- and 2% in MACS-CRC groups. Whereas the frequency of somatic *POLD1*-EDMs was 4% in CIN- and 0.0% in MACS-CRC groups. Although all 10 samples

with *ED-POLE* and *ED-POLD1* mutations were MSS, which is consistent with other data reported before by the TCGA project [229] and Palles et al, [227]. Most those 10 cases (9/10) were detected as chromosomal instable, which have shown that CIN and hypermutation occur together. So, it is not clear whether this group of tumours follows one of the previous established subtypes of CRC pathways or it represents a new pathogenetic pathway.

In summary, our data show that there is a large degree of overlap in the mutation profiles, the presence of heterogeneity and the frequency of allelic imbalance indicating a common genetic pathway for these tumour types in the early stages. Our data support the theory that aneuploidy is a late occurring feature in cancer progression although the cause of the aneuploidy has not been identified. In the future, we would expand our study by investigating larger number of genes for screening of mutations, methylation, LOH and miRNAs expression in larger cohort of CRCs, to get more accurate results about the difference between the sub-groups. In the next chapter, we sought to compare genetic alteration between primary CRC and matched metastasis not only to see is there any difference between them but also to find whether those genes. which are selected, have roles in metastasis.

As with primary CRC, metastasis is a multistep process and the heterogeneity between primary tumours and metastases seemed to be an additional reason for the failure of targeted therapies. The next chapter (chapter 5) evaluates molecular discrepancies between primary CRC samples and matched metastasis.

Chapter Five: Comparative analysis of primary CRC and matched metastasis

5.1 Abstract

Background: Approximately 50% of patients with primary CRC develop liver metastases. This study proposes to investigate genetic discrepancies between primary colorectal cancer (CRC) and their respective metastasis.

Methods: A total of 22 pairs of primary CRC and metastasis (first cohort) were tested. Mutation profiling across 26 cancer-associated genes was undertaken. Expression of a panel of six miRNAs was tested and protein expression of 20 genes was measured using Reverse Phase Protein Array (RPPA).

Results: Among the primary tumours, the mutation frequencies were: TP53 (86%), APC (64%), KRAS (41%), PIK3CA (9%), SMAD4 (9%), NRAS (9%), BRAF (4%), GNAS (4%), FBXW7 (4%), and CDH1 (4%). In primary vs metastasis, four mutations were detected in primary which were not detected in metastasis and two mutations were detected in metastasis which were not present in the primary. However, no significant differences were seen in mutation profiles between primary CRC and metastasis. Of the six miRNAs, four had significantly higher expression levels in the metastases than the corresponding primary tumour (miR-20a ($p<0.02$), miR-21 ($p<0.02$), miR-31 ($p<0.03$) and miR-92a ($p<0.02$)). Regarding RPPA data, only the protein expression of CD34 was significantly increased in the metastases compared to the primary cases ($p<0.04$).

Conclusion: Whilst there was no statistically significant difference between the mutation profile in primary and metastatic tumours, 5/22 had a difference for one mutation and 2/22 had a difference in an actionable mutation i.e. KRAS. Differences of 23% and 9% are concerning. The lack of statistical significance is likely due to underpowering and tumour cells not being representative of widespread metastasis. There are some differences in miRNA and protein expression although the mechanism of change and functional significance of these remains uncertain.

5.2 Introduction

Although substantial advances were achieved in adjuvant chemotherapy in the past few decades; the 5 year survival rates are still poor and more than half of all CRC patients are expected to die from metastasis within this period [418]. Thus; the understanding of the biological steps controlling CRC cells to develop metastases is of substantial importance to enable the establishment of new diagnostic and therapeutic strategies that would be able to tackle the disease in its early stages and consequently hinder its progression to the metastatic state [419]**.** Though metastases are the main cause of colorectal cancer deaths, the knowledge about the underlying molecular changes in primary carcinomas is much higher than that in more advanced disease stages because of limited studies performed on metastatic CRC [420]. Overwhelming clinical evidence indicates that colorectal tumours which show identical histological features have dramatically different prognosis and response to treatment, interestingly, this phenomenon supports the notion that CRC is a heterogeneous disease with rather unravelled molecular ambiguities that could explain the diverse clinical outcomes [421, 422]. Moreover, the initiation and progression of tumours in an individual occurs in a unique manner that is not identical amongst all tumours [423]. As a consequence, researchers now focus on the molecular basis of this malignancy which include the susceptibility, growth, progression, response to treatment and metastatic spread [154]. In general, the mutations which lead to primary cancers are also present in metastases, but additional mutations can occur after transformation [424, 425].

Generally multiple and consecutive genetic alterations are needed for cancer development and some patients may have coexistent alterations in two or more different signalling pathways [25, 426, 427]. CRCs accrue survival benefits via the process of Darwinian evolution followed by clonal expansion and this can create genetic heterogeneity. As cancer cells metastasise, different survival benefits are required to adapt to a new environment, which can lead to further genetic changes. Moreover, it may be an older clone which spreads to a distant site whilst the primary tumour progresses. The heterogeneity between primary and

metastatic cancers is a leading cause of increased resistance to therapy, which is the major cause of cancer-related death [424, 428]. A full understanding of the possible genetic discrepancies, changes in the expression of specific proteins and miRNAs between the primary and their corresponding metastatic tumours, is required for developing efficient targeted mechanisms to stop tumours from progressions to metastatic disease and accordingly potentially improve patients' survival. Recently, Neerincx et al [429] using NGS screened 220 fresh-frozen samples, including paired primary and metastatic tumor tissue and non-tumorous tissue from 38 patients for quantification of 2245 known unique mature miRs and 515 novel candidate miRs. The authors interestingly found that only eight miRNAs were differentially expressed between primary CRC and the corresponding metastasis, consisting of five known miRs (miR-320b, miR-320d, miR-3117, miR-1246 and miR-663b) and three novel candidate miRs (chr 1-2552-5p, chr 8-20656-5p and chr 10-25333-3p).

Additionally, the precise characterization of oncogenic mutations present at the time of metastatic disease that could represent potential therapeutic targets is of utmost importance. However, it is considered challenging to obtain tissue biopsies from metastatic tumours, and patients and clinicians still prefer using the least invasive method possible for genetic testing. Recently, there is a growing evidence suggesting inter- and intra-tumour genetic heterogeneity in a range of solid tumours including CRC which raises concerns that molecular profiling of primary tumours may not be representative of metastatic disease [430-432]. In CRC, comparative sequencing studies found a high degree of concordance between primary CRC tumours and their matched metastases in a study with a small number of cases and also in larger studies [433, 434]. Additionally, the study conducted by Kim et al [435], which collected data from 468 CRC samples by assessing 1321 individual genes using NGS, found a high level of concordance for potential somatic alterations, suggesting that genomic profile of primary site is very similar to that of the metastatic site for the majority of interrogated cancer genes. In direct a contrast, a study used NGS for analysing 1264 genes in 21 paired primary CRC and their corresponding metastasis reported a high degree

of mutational discordance between primary and metastatic samples of 21 patients [436].

5.3 Aims

This study aimed to investigate the molecular differences between primary CRC samples and their matched metastases in order to find events which may cause metastasis. NGS was used to compare the genetic composition of the 26 genes associated with different pathways and processes. Additionally, a panel of six miRNAs were tested using RT-QPCR and protein expression of further genes was evaluated using Reverse Phase Protein Array (RPPA) to explore the possible differences in the main signalling pathways in CRC between the primaries and their matched metastases.

5.4 Results

5.4.1 Screening of mutations

5.4.1.2 Mutations in matched pairs of primary and metastases CRCs

NGS analysis was performed in a series of 22 cases of primary tumour and matched metastasis as described in ch3 (details of paired samples are shown in table 5.1). A total of 61 non-synonymous somatic variations in 12 genes were found in primary 22 specimens whereas 60 were found in metastases (table 5.2). In addition, 2 more mutations were detected (1 mutation in *ED-POLE* in one metastasis sample and 1 mutation in *ED-POLD1* in one pair of primary and metastasis) using PCR-HRM (figure 5.1). The mutant allele frequency was 1.03-fold higher in primary than metastasis CRCs. Although the paired primary and metastases CRCs among 22 cases were not identical, chi-square test, has shown that there is not a significant difference between them (p>0.05). Discrepancy between the primary and metastasis in CRCs was seen in 5 cases. As shown in (table 5-2). In case number 1, the primary was shown to be mutant type for *KRAS* (G12V) and (H179R) in *TP53* whereas these were wild type in the metastasis. This case also had *APC*, *SMAD4*, *KRAS* (Q61H) and *TP53* (V175A) mutations and these were identical in the primary and metastasis. In case number 2, the metastasis was shown to be wild type for *GNAS*, whereas an exon 8 mutation in *GNAS* was detected in the primary specimens. This case also had *APC*, *KRAS* and *SMAD4* mutations and these were identical in the primary and metastasis. In case number 3, the metastasis was shown to be mutant for *NRAS* (Q61K) and *TP53* (I254N) whereas the primary was wild type at these two hotspots. This case also had *APC*, and *TP53* (G266E) mutations and these were identical in the primary and metastasis. Moreover, in case number 4, the metastasis was shown to be wild type for exon 8 in *CDH1*, whereas an exon 8 mutation in *CDH1* (T342S) was detected in the primary specimens. This case also had KRAS and TP53 mutations and these were identical in the primary and metastasis. In case number 5, the primary was shown to be wild type for exon 13 in *ED-POLE*, whereas an exon 13 mutation in *ED-POLE* (G433D) was detected in the metastasis specimens. This case also hd *BRAF* and *TP53* mutations and these were identical

in the primary and metastasis. According to data reported in the COSMIC database (http://cancer.sanger.ac.uk/cosmic) apart of mutations (c.1024A>T and c.1298 G>A) in *CDH1* and *POLE* genes respectively, which are novel mutations and not reported before, others five mutations are pathogenic mutations. Generally, the overall mutations detected in 12 genes in metastasis are almost similar with that found in overall 83 primary CRC samples (see table 5.3).

Table5. 1:Details of 22 paired (primary CRC and matched metastasis) samples

Sample ID/ Primary	Year of resection	Sample ID/ Metastasis	Year	Sex	Age at primary	Site of Metastasis
6	2003	43	2004	F	46	Hepatic
7	2002	41	2003	M	61	Hepatic
10	2004	53	2005	M	67	Hepatic
13	2001	54	2003	M	77	Hepatic
18	2004	57	2004	F	80	Hepatic
22	2005	33	2006	M	63	Hepatic
44	2005	45	2006	M	43	Hepatic
48	2005	67	2006	M	62	Hepatic
62	2004	39	2006	M	84	Hepatic
65	2002	47	2005	M	50	Hepatic
72	2008	73	2008	M	66	Hepatic
83	2003	84	2005	F	79	Hepatic
85	2005	86	2007	M	71	Hepatic
87	2005	88	2007	M	56	Hepatic
89	2005	90	2008	M	65	Hepatic
93	2003	94	2006	F	58	Hepatic
95	2006	96	2007	M	76	Hepatic
97	2004	98	2005	M	68	Hepatic
101	2003	103	2005	F	69	Hepatic
108	2005	109	2006	M	77	Hepatic
110	2006	111	2007	F	49	Hepatic
112	2005	113	2006	M	79	Hepatic

Table5. 2:Summary of mutations detected in 22 paired CRC samples.

Case No.	Gene	Mutation	Codon	Exon	Primary			Metastasis		
					Read depth	Alt read	Alt variant	Read depth	Alt read	Alt variant
1	APC	c.3871C>T	1291	15	28980	5605	19.34	18950	5239	27.65
		c.4468_4474delCATTTTG	1490-92	15	23882	4060	17.0	14723	3218	21.8
	KRAS	c.35G>T	61	3	22400	4509	20.13	13210	4005	30.32
		c.183A>T	12	2	10804	690	6.39			wild
	TP53	c.518T>C	179	5	3222	779	24.18			Wild
		c.536A>G	173	5	2969	644	21.69	1414	530	37.48
	SMAD4	c.1496G>A	499	9	14227	3918	27.54	9972	3698	37.08
2	APC	c.3707_3708delCA	1236	15	20975	1232	5.87	5178	725	14.
		c.4033G>T	1345	15	29576	1232	5.38	11444	1716	14.99
		c.4348C>T	1450	15	29576	1592	24.33	2944	1206	29.33
	KRAS	c.38G>A	13	2	10800	4144	38.37	2929	2913	40.96
	SMAD4	c.1478A>G	493	9	28808	3081	10.69	6588	2274	34.52
	GNAS	c.2531G>A	201	8	12005	806	6.71			wild
3	APC	c.4012C>T	1338	15	13490	5146	38.15	8516	313	3.68
	NRAS	c.181C>A	61	3			wild	10968	3777	34.44
	TP53	c.797G>A	266	7	6597	2674	40.53	2317	175	7.55
		c.761T>A	254	7			wild	6824	2265	33.19
4	KRAS	c.139G>C	47	2	17060	1399	8.2	30655	10397	33.92
	TP53	c.517G>A	173	5	9823	964	9.81	10892	6900	63.34
	CDH1	c.1024A>T	342	8	12277	578	4.71			wild

Case No.	Gene	Mutation	Codon	Exon	Primary			Metastasis		
					Read	Alt	Alt	Read	Alt	Alt
5	BRAF	c.1799T>A	600	15	19542	4665	23.87	5140	1403	27.29
	TP53	c.723delC	241	7	13348	4498	33.7	8870	4207	45.43
	POLE	c.1298 G>A					wild			mutant
6	APC	c.3944C>A	1315	15	39561	22962	58.04	14255	1473	10.33
	TP53	c.54_55insAT	19	2	3665	2009	54.82	4452	560	12.58
7	-	-					-			-
8	-	-					-			-
9	APC	c.3949G>C	1317	15	12564	6159	49.02	10235	4807	46.97
		c.4360delA	1454	15	6759	2004	29.65	4500	1218	27.07
	KRAS	c.35G>C	12	2	6467	2715	41.98	4449	1996	44.86
	TP53	c.743G>A	248	7	4813	2752	57.18	5015	2561	51.07
10	APC	c.4059_4060insT	1376	15	8231	2071	25.16	13207	5705	43.2
	TP53	c.845_846insCGGT	282	8	2034	763	37.51	5928	3090	52.3
	PIK3CA	c.1633G>A	545	9	18638	2958	15.87	21081	6472	30.7
	FBXW7	c.1168dupT	390	7	18748	3588	19.14	16308	5389	33.05
11	KRAS	c.38G>A	13	2	7915	312	3.94	3377	161	4.77
	TP53	c.821T>A	274	8	7872	250	3.18	6268	201	3.21
12	APC	c.4468_4474delCATTTTG	1490-92	15	21364	14882	69.66	15924	6049	37.99
	KRAS	c.35G>T	12	2	10604	3399	32.05	7012	1557	22.2
	TP53	c.536A>G	179	5	4482	3418	76.26	2525	1095	43.37

Case No.	Gene	Mutation	Codon	Exon	Primary			Metastasis		
					Read	Alt	Alt	Read	Alt	Alt
13	APC	c.3916G>T	1306	15	24089	5497	22.82	15561	7589	48.77
	TP53	c.524G>A	175	5	4632	870	18.78	3616	1341	37.09
14	APC	c.4660_4661insA	1554	15	26838	5150	19.19	12919	2317	17.93
	KRAS	c.35G>A	12	2	4509	1230	27.28	1921	506	26.34
	TP53	c.745A>G	249	7	16319	4368	26.77	10866	2881	26.51
	PIK3CA	c.1624G>A	542	9	14632	1619	11.06	2303	299	12.98
15	APC	c.3957delT	1319	15	18197	4190	23.03	17442	4444	25.48
	KRAS	c.35G>A	12	2	7266	1791	24.65	7832	2152	27.48
	TP53	c.422G>A	141	5	5426	1862	34.32	5812	2186	37.61
16	NRAS	c.181C>A	61	3	21442	8883	41.43	9857	3957	40.14
	TP53	c.761T>A	254	7	7658	2834	37.01	3016	1083	35.91
17	APC	c.4455delT	1485	15	15133	5401	35.69	2138	712	33.3
	KRAS	c.35G>C	12	2	3148	690	21.92	1721	198	11.51
	TP53	c.223_224dupCC	75	4	4299	1580	36.75	2954	387	11.74
18	APC	c.4033G>T	1345	15	15282	916	5.99	6215	101	16.25
	TP53	c.524G>A	175	5	6706	2366	35.28	2312	978	42.3
19	APC	c.4479G>A	1306	15	35103	8448	24.07	12068	4990	41.35
	TP53	c.406delC	136	5	10324	3635	35.21	1346	584	43.39

Case No.	Gene	Mutation	Codon	Exon	Primary			Metastasis		
					Read	Alt	Alt	Read	Alt	Alt
20	TP53	c.638G>A	213	6	8932	1718	19.23	1332	240	18.02
		c.844C>T	282	8	5264	1969	37.41	1546	610	40.11
	POLD1	c.1231 C>T					mutant			mutant
21	APC	c.3980C>G	1327	15	10459	1558	14.9	4634	913	19.7
	TP53	c.437G>A	146	5	3801	1157	30.44	1413	803	56.84
22	TP53	c.454_466delCCGCCCGG	152-156	5	5400	1046	19.7	1311	69	5.26

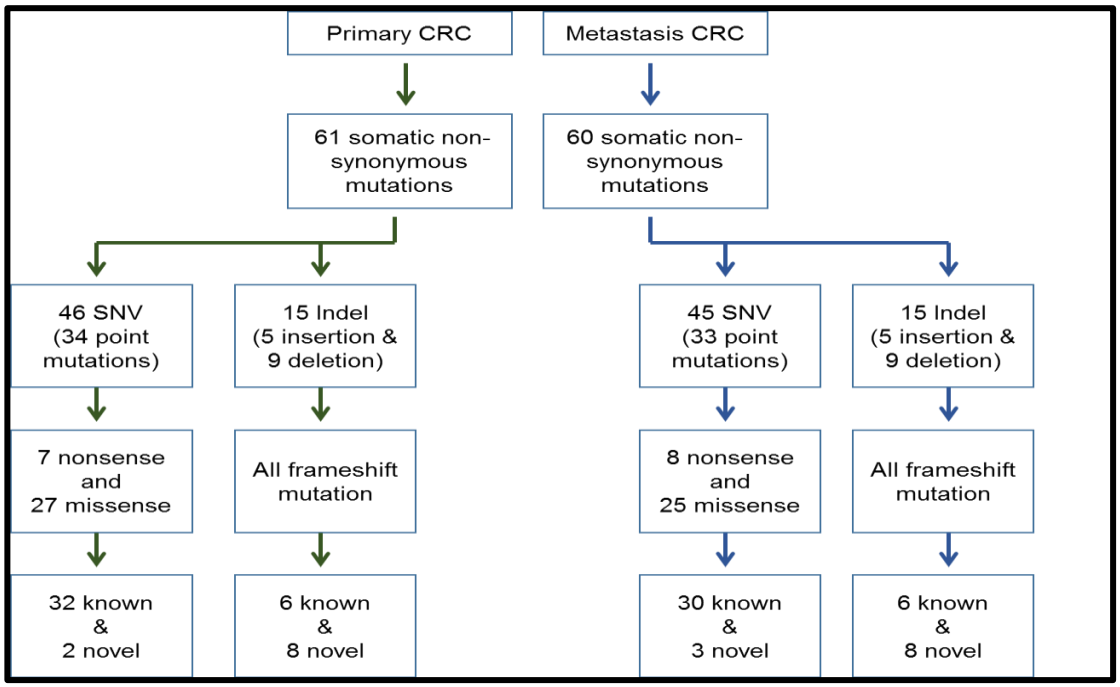

Figure 5. 1: summary of all mutations detected in all 22 paired samples. SNV, single nucleotide variant; indel, insertion and deletion.

Table 5.3: comparison of overall mutations detected in 22 metastasis samples to those found in overall 83 CRC samples.

Gene	Mutation status	Overall 83 primary	Mutation status	Overall 22 Metastasis
KRAS	Wild type Mutant	39 (47%) 44 (53%)	Wild type Mutant	12 (55%) 10 (45%)
GNAS	Wild type Mutant	79 (95%) 4 (5%)	Wild type Mutant	21 (95%) 1 (5%)
PIK3CA	Wild type Mutant	64 (77%) 19 (23%)	Wild type Mutant	20 (91%) 2 (9%)
CDH1	Wild type Mutant	82 (99%) 1 (1%)	Wild type Mutant	21 (95%) 1 (5%)
BRAF	Wild type Mutant	76 (91%) 7 (9%)	Wild type Mutant	21 (95%) 1 (5%)
APC	Wild type Mutant	36 (43%) 47 (57%)	Wild type Mutant	8 (37%) 14 (63%)
FBXW7	Wild type Mutant	73 (88%) 10 (12%)	Wild type Mutant	21 (95%) 1 (5%)
SMAD4	Wild type Mutant	71 (85%) 12 (15%)	Wild type Mutant	20 (91%) 2 (9%)
TP53	Wild type Mutant	21 (25%) 62 (75%)	Wild type Mutant	4 (19%) 18 (81%)
NRAS	Wild type Mutant	80 (96%) 3 (3%)	Wild type Mutant	20 (91%) 2 (9%)
ED-POLE	Wild type Mutant	75 (90.4%) 8 (9.6%)	Wild type Mutant	21 (95%) 1 (5%)
ED-POLD1	Wild type Mutant	81 (97.6%) 2 (2.4%)	Wild type Mutant	21 (95%) 1 (5%)

5.4.2 MiRNA quantification

For this project a panel of 6 miRNAs were selected based on previous studies where they reported that all six miRNAs are oncogenic i.e. their expression was elevated at different stages in CRC initiation and progression [298, 304, 437-446].

5.4.2.1 Measurement of the quality and quantity of miRNA

Total RNA was extracted from 22 paired of primary CRCs and matched metastasis tissue samples. The median RNA yield from the purification of FFPE was 496.5 ng and all samples had a 260/280-ratio ranged between 1.86 - 2. All samples had at least 50% tumour cell content. After that, a qPCR assay was developed to measure the expression of miRNAs in FFPE CRC and normal healthy specimens. In this qPCR, relative quantification was applied and thus RNU6B, which is a small nuclear RNA, was used as an appropriate internal normalisation control to normalise sample-to-sample variations.

5.4.2.2 Confirmation of primer specificity

Specificity of qPCR products was checked with gel electrophoresis and resulted in a single band with the anticipated length (RNU6B, miR-20a, miR-21, miR-29a, miR-31, miR-92a and miR-224) have the same length 110 bp (figure 5.2). This list was on the basis of their roles in CRC development and metastasis, (see table 6.1 in chapter 6). In addition, a qPCR dissociation curve analysis was performed which resulted in single sharp peak indicating no primer-dimers were generated. The specific melting temperatures as follows: RNU6B, 74.64°C; miR-20a 75.17°C; miR-21, 76.29°C; miR-29a, 75.47°C; miR-31, 76.78°C; miR-92a, 77.42°C; and miR-224, 76.34°C; (Appendix C- figure 1).

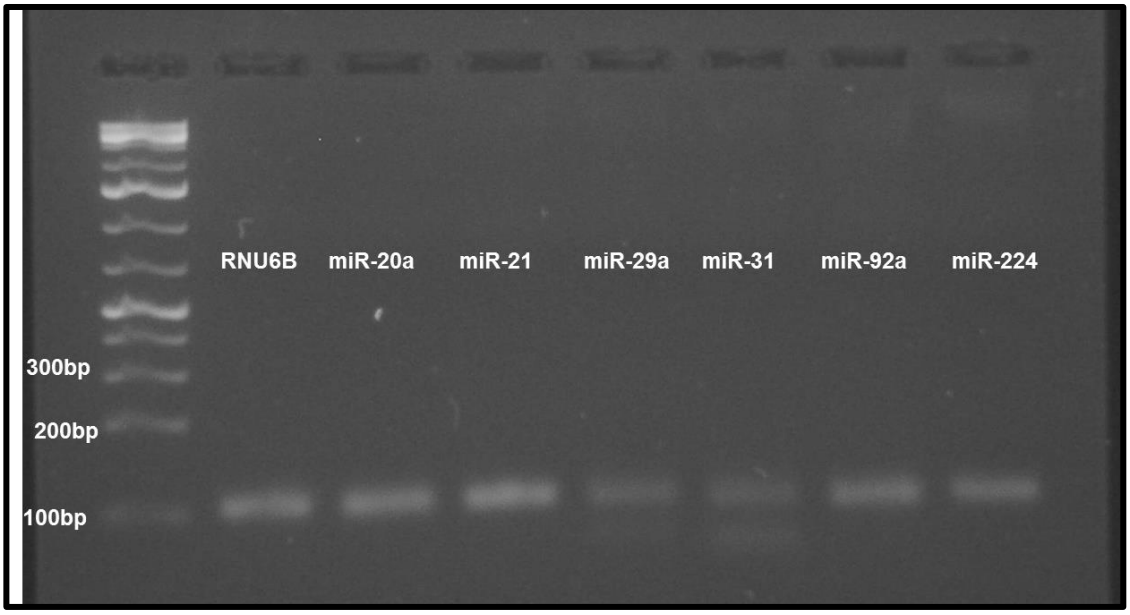

Figure 5. 2: PCR products represent all miRNAs (miR-RNU6B, miR-20a, miR-21, miR-29a, miR-31, miR-92a and miR-224).

5.4.2.3 Real-time PCR amplification efficiencies

The reproducibility of an assay is best indicated by the coefficient of variation (CV) in a series of replicate tests. The short-term precision (intra-assay variability), which was determined in three repeats, showed a mean CV of the primers as follow: 0.84% (range 0.25%-1.74%) for miRNAs primers. The long-term precision (inter-assay variability), which investigated in three different experimental runs performed on 3 days, showed a mean CV of the primers as follow: 3.68% (range 1.23% - 6.52%) for miRNAs primers. Tested primers showed high real-time PCR efficiency rates; RNU6B 102%, RNU61 97%, miR-39 103%, miR-20a 93%, miR-21 96%, miR-29a 105%, miR-31 96%, miR-92a 96, and miR-224 90% (Appendix C- figure 2 & 3)

5.4.2.4 Stability of endogenous controls

Initially before working on real samples, two small nuclear RNA (RNU6B) and (RNU61) to be used as an endogenous control. We evaluated the expression and stability of the two-chosen snRNA (RNU6B) and (RNU61) to see which one is more suitable to be used as an internal normalisation control. Our data demonstrated that both RNU6B and RNU61 were readily detectable in the 20-paired normal and tumour tissues of patients with CRC. No significant difference was observed in terms of Ct values of RNU6B ($p=0.49$, Wilcoxon test) and RNU61 ($p=0.11$, Wilcoxon test) between normal and tumour tissue samples [447]. Although both controls showed similar abundance in the FFPE tissue, expression of RNU6B was more stable and less variable between normal and tumour tissues than RNU61. After that the expression level of RNU6B was evaluated in the metastasis samples by comparing with it is expression level in their matched primary CRC samples. Our results showed that RNU6B was detectable in the metastasis samples and no significant difference was seen in term of Ct values ($p=0.13$, Wilcoxon test).

5.4.2.5 Cut-off point detection of miRNAs

Before running tumour samples to detect expression levels of miRNAs, we sought to find the cut-off point to show high or low expression. To achieve this goal, we extracted RNA from 20 pure normal colon tissue that were pooled with an equal amount from each sample. The expression level of all miRNAs was measured in all normal colon tissue samples after comparing with the pooled sample. On average the minimum fold of expression of all miRNAs in normal colon tissue was 0.5 and the highest was 1.5 (using <0.5 fold as showing downregulation and >1.5 fold as showing up-regulation).

5.4.2.6 MiRNA expression in primary and metastasis CRC specimens

To identify aberrantly expressed miRNAs, the study screened miRNA levels in 44 CRC. All assays were done in triplicate and the Ct value of all targets in all samples were less than 25 ranging between (15.5-24.3) with SD less than 0.5 between replicates Ct value. The miRNAs with significantly different expression in the primary CRC samples compared with matched metastasis were identified by statistical analysis. Of the six candidates, four micro-RNAs including miR-20a ($p<0.02$), miR-21 ($p<0.02$), miR-31 ($p<0.03$) and miR-92a ($p<0.02$), had significantly higher expression levels in the metastasis CRC group than in the primary CRC group. However, following multiple correction testing using the Bonferroni correction, they all failed to retain significance (table 5.4). Other miRNAs including miRNA29a ($p<0.12$) and miRNA224 ($p<0.48$) did not show a significant difference in their expression levels between primary and metastasis CRCs.

Table 5.4: Compare expression profiles of candidate microRNAs between primary CRC and metastasis.

Target genes	Expression in pri vs met	p value	Post Bonferroni correction
miRNA20a	Pri < Met	**0.02**	0.11
miRNA21	Pri < Met	**0.02**	0.11
miRNA29a	Pri < Met	**0.03**	0.16
miRNA31	Pri < Met	**0.02**	0.11
miRNA92a	Pri = Met	0.12	0.53
miRNA224	Pri = Met	0.48	0.98

Pri = Primary, Met= Metastasis

5.4.3 Evaluate protein expression

5.4.3.1 Protein expression profiles

In addition to the mutation screening and miRNA expression, the present study tested protein expression of a further 20 genes, some of which are targeted by the six miRNAs. This choice of proteins was due to mutation not being unique cause leading to the alteration of protein expression. For example, Zhu et al [269] found that PI3K mutation rate is not in accordance with the expression rate of PI3K protein, indicating that PI3K mutations are not the only cause increasing the expression of PI3K protein in CRC samples.

Whole protein lysates from 21 informative primary CRCs and matched metastatic tumours were generated to detect the expression of 20 different proteins by reverse phase protein array. The samples were run in triplicate, after taken the mean of these replicates for each target in each sample and for comparative analysis, the geometric mean (calculated by the square root of the product of the values for each biomarker) of intensity signals from epithelial markers BerEP4 and AE1/3 were used to normalize signals generated from other markers. Then after normalization the median was used to stratify the specimens into low and high expression groups. From 20 protein targets, the Wilcoxon test revealed that CD34 was found to be significantly higher in metastasis when compared with the primary tumour ($p<0.04$). However as shown in (table 5.5), following multiple correction testing using the Bonferroni correction, it also failed to retain significance ($p=0.55$). Additionally, (figure 5.3) illustrates an unsupervised one-way hierarchical cluster analysis of the expression of 20 biomarkers using RPPA. For testing the protein expression levels in this part of the project, I have extracted protein from FFPE samples, and Dr. Ola Negm (Research Fellow Immunology/ School of Life Science/ Nottingham University Hospital) done protein quantification and run samples with RPPA. I have also done the data analysis with the help of the statistician Agrawal Utkarsh at the school of computer science/Nottingham University.

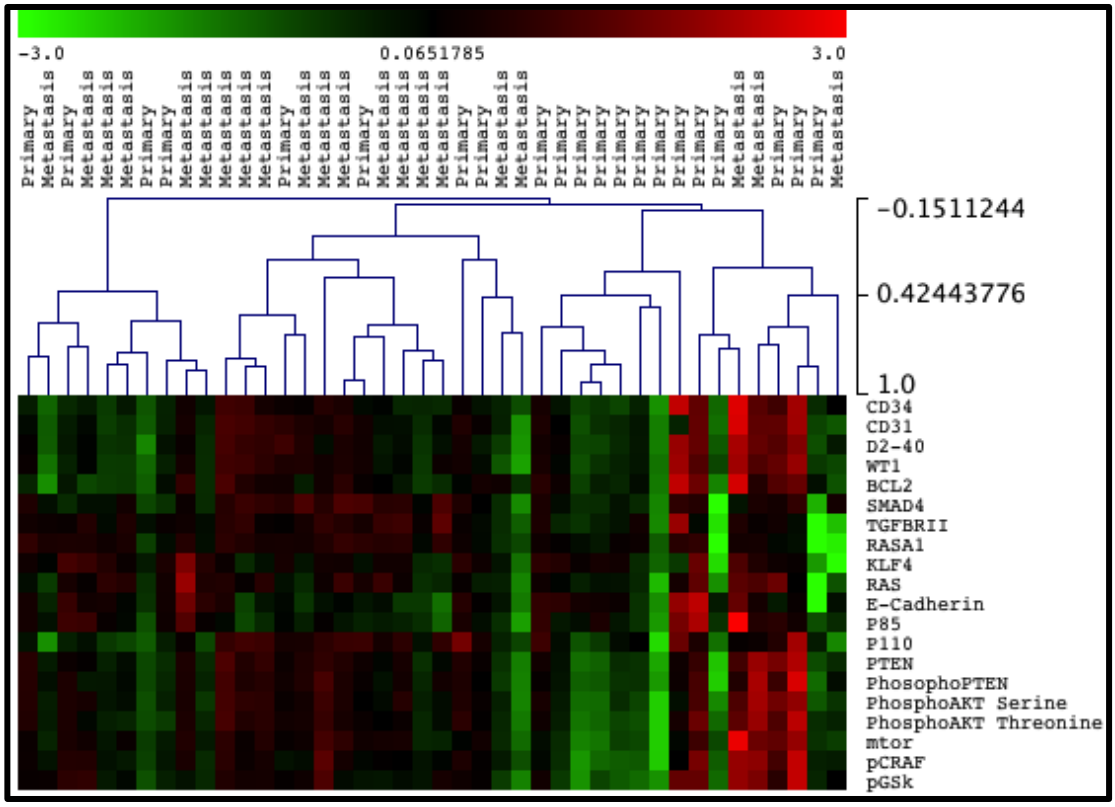

Figure 5.3: Heat-map showing different signalling pathway intermediates studied in primary CRC (21) and metastasis (21) using RPPA. Rows represent the different signalling molecules studied. Green and red denote markers that are present at lower and higher levels, respectively. Data represent the mean of four technical replicates per sample.

Table 5.5: compare protein expression profile of candidate genes between primary CRC and metastasis.

Target genes	P value	Post Bonferroni correction
CD34	0.04	0.55
CD31	0.3	0.99
D2-40	0.3	0.99
WT1	0.2	0.98
BCL2	0.05	0.64
SMAD4	1.0	1.0
TGFBRII	0.06	0.70
RASA1	0.3	0.99
KLF4	0.6	1.0
RAS	1.0	1.0
E-Cadherin	0.5	1.0
P85	0.3	0.99
P110	0.2	0.98
PTEN	1.0	1.0
Phospho-PTEN	0.3	0.99
Phospho-AKT Serin	0.5	1.0
Phospho-AKT	0.5	1.0
mTOR	0.3	0.99
pCRAF	0.3	0.99
pGSK	0.3	0.99

5.5 Discussion

Now with the targeting molecular alterations that mediate cancer progression as a strategy to develop cancer therapies, genotyping of patients with advanced CRC is performed as a part of clinical practice. In particular, testing of *KRAS* and *NRAS* mutation is used as a component for the treatment of patients with metastatic CRC. However, it remains unknown whether the primary lesion is sufficient to perform such testing. Although in the past some studies founded a significant discordance in mutation profile between primary and their matched metastasis [448-451], the National Comprehensive Cancer Network recommends testing either the primary tumour or metastatic lesion based on several studies that highlighted high concordance (>95%) of *KRAS* mutation between primary CRCs and metastases [41, 266, 452, 453]. The present study sought to identify the incidence of mutations and their differences in paired primary CRC and metastatic samples. Mutations could be classified into three types: mutations that are (1) present in both the primary tumour and the metastasis), (2) limited to the primary tumour or (3) in the metastasis only. Overall 22 paired (primary and their corresponding metastasis) tumours were genotyped and found that in both subtypes (primary CRCs and matched metastasis) TP53, APC and KRAS were more frequently mutated where as PIK3CA, SMAD4, NRAS, BRAF, FBXW7 CDH1, POLE, POLD1 and GNAS mutated in only two or one cases and no mutations were found in the rest of genes included in the panel. Additionally, the mutation profile and frequency of mutated genes in this subgroup is almost like that detected in the overall (83) population (see table 5.3). Consistent with the model those alterations in APC, K*RAS* and *BRAF* occur early in CRC pathogenesis [454] high similarity seen for *APC*, K*RAS, NRAS* and *BRAF* in primary and metastasis, suggesting that they share the same origin. The results demonstrated that the frequency of mutations in the above genes did not differ significantly between primary tumours and metastases. The data generated are in line with recent reports which found no significant difference in the frequency of mutations between primary CRC and their metastasis [267, 454-456]. However, other studies demonstrated that there are big differences in the

frequency of mutations targeting *PIK3CA* and *TP53* between primary and metastatic CRCs and showed that mutations in these genes may play a great role in the establishment of metastases [269, 457].

GNAS and CDH1 were mutated only in primary CRC and not found in their corresponding metastasis, it may be an older clone which spreads to a distant site whilst the primary tumour progresses. In addition, other discrepancies were seen in 2 more patients, mutations targeting KRAS (G12V) and TP53 (H179R) in primary case number1, are absent in its metastasis. In contrast, mutations targeting NRAS (Q61K) and TP53 (I254N) occurred in in the metastasis in case number 3 but were not seen in its primary. It is most likely that these discrepancies are generated by late stage mutations when the primary tumour has already metastasised and both tumours undergo separate evolution and clonal expansion under different conditions which generate different mutations. The study did not only demonstrate the concordance of mutations targeting most of the mentioned genes in primary and paired metastases, but also found that the type of mutations was identical, which indicates stability of these mutations during metastatic progression.

Beside comparing mutation profiles, the study sought to compare the expression levels of miRNAs between primary CRC and metastasis, as it is unknown to what extent tumour derived miRNAs are differentially expressed between primary CRCs and metastatic lesions. Before starting to identify the expression of the miRNAs, the study evaluated the stability of the RNU6B, which is used as an endogenous control. The study did not find a significant difference in the expression of RNU6B between tumour and normal tissues ($p=0.49$, Wilcoxon test). This result is consistent with another study which reported that that RNU6B is a good candidate to be used as a control [447, 458-461]. However, Huang et al, [317] and Carlsson et al, [462] found that RNU6B appeared to be less stable and an unsuitable reference for normalization.

The present study demonstrated that the miRNA expression profile of primary CRCs closely resembles that of their corresponding metastases. Comparison of the two groups showed that, of the six miRNAs', four had significantly higher

expression levels in the metastases than the corresponding primary tumour. However, following multiple correction testing using the Bonferroni correction, they all failed to retain significance. This result is in line with other studies which showed there is not a significant difference in miRNAs profile expression between primary CRC and their corresponding metastasis [463-465]. However, Ke et al,[466] Ling et al, [467] and Tang et al, [468] demonstrated that miR92a, miR224 and miR29a, are increased and promote CRC metastasis. Based on our results, if miRNA expression is to be used as a biomarker for prognosis and response to treatment, they can be tested in tissue obtained from either primary or secondary tumour. This is of clinical importance, because tissue samples from the primary tumour are often readily available, while these are not routinely collected from metastases.

Protein expressions of 20 targets that are involved in different signalling pathways were also analysed using reverse-phase protein assay (RPPA) analysis. The study used BerEP4 as a good epithelial housekeeper, which is not expressed in liver epithelium and accurately reflects epithelial contents of a tissue sample thus allowing biomarker expression to be normalised to epithelial contents [469]. All 20 targets that were analysed by RPPA demonstrated relatively similar expression between primary CRCs and matched tumour samples. Interestingly, our results are relatively similar to Chloe et al, who found concordant PTEN expression in 98% of CRC primary and liver metastasis [470]. However, Losi et al, [471] showed significantly lower SMAD4 expression in metastasis than in primary CRC. In addition, Sawai et al, [472]suggested PTEN was more strongly expressed in primary CRC specimens than in patients with liver metastasis. Moreover, Yu-Fen et al, [269] demonstrated that high PI3K expression is associated with CRC metastasis. The inconsistent results may due to the different quantification methods or small sample sizes as well as intra-tumour heterogeneity.

Our study has several limitations common to most published findings in this field. First, the sample is relatively small and underpowering. Second, oligometastatic disease not necessarily being representative of widespread metastases. Third,

despite NGS data being validated by PCR/HRM method, miRNA and protein evaluations were not confirmed by other methods. Despite these limitations, this study gives some clinically meaningful suggestions. Our study demonstrated the existence of a relative concordance between primary tumours and their corresponding metastases in patients with CRC. These results emphasize that genetic testing could be performed in primary tissues as a surrogate genetic profile that covers both the primary and their matched metastatic tissues.

In summary, the study suggests that in most clinical scenarios, analysis of the primary colorectal tumour is sufficient for determining mutational status in those genes included in the panel that was selected, additionally, it demonstrated that none of those genes have a differential expression between the primary and liver metastasis, therefore we could not point to a specific gene that could have a particular significant role in the development of colorectal metastasis. Although, our results showed lack of significant difference between primary CRC and matched metastasis, bigger study needs to be done to decide whether miR expression profiles can be used as predictive biomarkers for prognosis and response to treatment regardless of metastatic tumour profile.

In addition to the genetic alteration, a novel group of biomarkers, has recently been discovered and are believed to have an impact on laboratory medicine as new diagnostic and prognostic markers. The next chapter (chapter 6) investigate the feasibility of using miRNAs as molecular markers for colorectal cancer detection.

Chapter Six: Investigate miRNAs as potential marker for colorectal cancer diagnosis

6.1 Abstract

Background: Recently studies reported that miRNAs play vital roles in pathogenesis of many human diseases including cancer, which may serve as diagnostic, prognostic markers and may provide means for cancer treatment. The study aimed to investigate the feasibility of using miRNAs as molecular markers for colorectal cancer detection.

Methods: Paired tumour/normal tissues of 81 CRC patients (second cohort) were investigated to measure the expression level of six miRNAs (miR-20a, 21, 29a, 31, 92a and 224), evaluate mRNA and protein expressions in a further six genes (SMAD4, PTEN, TGFBRII, BCL2, KLF4 and RASA1) genes using RT-qPCR and immunohistochemistry analysis respectively.

Results: In relative to the normal tissue mucosa, statistical analysis revealed a significant increase in the tumour expression level of (miR-21 ($p=0.0017$), miR-29a ($p=0.006$) and miR-31 ($p=0.0018$) with a decrease in the mRNA expression level of TGFBRII. Spearman's rank order correlation demonstrated that high miR-20a expression was inversely correlated with PTEN-mRNA level ($r= -0.238$, $p=0.033$) and PTEN-protein level ($r= -0.253$, $p= 0.023$). High miR-21 and miR-224 expression were associated with low expression of TGFBRII-mRNA [($r= -0.358$, $p=0.001$), ($r= -0.276$, $p=0.013$)] and TGFBRII-protein [($r= -0.328$, $p= 0.003$), ($r= -0.319$, $p=0.004$)] respectively. In addition, over expression of both miR-29a and miR-31 inversely correlated with RASA1-mRNA level [($r= -0.217$, $p=0.014$), ($r= -0.276$, $p=0.013$)] and RASA1-protein level [($r= -0.222$, $p=0.046$), ($r= -0.209$, $p=0.010$)] respectively.

Conclusions: The presence of a high degree of correlation between upregulated miRNAs and downregulation of some of target genes involved in different signalling pathways, indicated that miRNAs may have roles in CRC carcinogenesis. Additionally, upregulation of (miR-20a, 21, 29a and 31) may be suitable to differentiate CRC with a high degree of accuracy from a normal mucosa of CRC patients and can play a critical role on screening CRC in general population.

6.2 Introduction

There is convincing evidence that screening and early detection of CRC has been a pivotal strategy for reducing the incidence and mortality rates of the disease [473]. For example, the 5-year survival rate is as high as 93.2% for TNM stage I as compared to only 8.1% for stage IV [474]. However, existing screening tools such as (1) colonoscopy screening, which is currently the most reliable screening tool, has been hampered because of its invasive nature and high cost, (2) the faecal occult blood test (FOBT), which has low sensitivity and requires dietary restriction, impedes compliance and use. Additionally, studies have investigated several molecular biomarkers for CRC detection, such as carcinoembryonic antigen (CEA), and shown that high CEA levels are associated with CRC progression. However, its utility in the disease screening is limited due to the serum level of CEA not being elevated after the tumour has entered the serosa membrane [475]. In view of these caveats, there is an urgent need for new specific molecular markers to improve the diagnosis of CRC. In the recent past, researchers have focussed on miRNAs due to the roles they play in a variety of cellular processes including development, cell cycle progression, cell differentiation, proliferation and apoptosis [476, 477]. Others found aberrant miRNA expression has been associated with several types of cancers [478, 479] as these act as either tumour suppressors or oncogenes [307]. Furthermore, miRNAs have been shown to successfully discriminate various types of cancers and predict outcomes in both haematological and solid malignancies [480]. In CRC difference in profiles of miRNA expression between tumour and paired adjacent colorectal normal tissue [323], highlights their potential for early diagnostic and prognostic applications [317, 447]. However, inconsistencies about the diagnostic accuracy of differentially expressed miRNAs still exists.

In the present study, we have chosen a panel of 6 miRNAs including (miR-20a, 21, 29a, 31, 92a and 224) which previous studies have shown to be upregulated in CRC and which could be used as diagnostic and prognostic markers [302, 316, 317, 322, 323, 326, 481]. (see table 6.1)

Table 6.1: Summarizes selected miRNAs by location, target genes and expression in tumours

Name	location	targets	Expression in tumours
miR-20a	chromosome 13 (13q31.3) [437]	E-cadherin [438] PTEN [482] *SMAD4* [483].	Downregulated: in pancreatic carcinoma and breast cancer [300, 438] Upregulated in gliomas, gastric, and colon [323, 484, 485]
miR-21	chromosome 17 (17q23.1)	TGFBR2, TPM1, PTEN, RASA1, RhoB and programmed cell death PDCD [439, 486].	Upregulated in: colon cancer, breast cancer, lung cancer, prostate cancer, pancreatic cancer, stomach cancer, hepatocellular carcinomas, ovarian cancer, and oesophagus [300, 487-489]
miR-29a	chromosome 7 (7q32.3) [490]	KLF4 and E-cadherin [468]	Downregulated in: Acute myeloid Leukaemia AML [491]. Upregulated in: colorectal, pancreatic, lung, [468, 492]
miR-31	chromosome 9 (9p21.3) [304]	RASA1 [493]	Downregulated in: prostate carcinoma, gastric cancer, breast cancer, and urothelial carcinomas [494-496]. Upregulated in: hepatocellular carcinoma, squamous cell carcinoma and head, neck and colorectal cancer [493, 497, 498]
miR-92a	chromosome 13q31-q32 [499]	BCL2, SMAD4 [283],	Upregulated in Liver, Bladder, colorectal cancer and ovarian [500-503]
miR-224	X-chromosome (Xq28)	Cdc42, SMAD4, TGF, AKT and RKIP [504-506].	Downregulated in oral cancer, lung cancer, prostate, breast cancer, and ovarian cancers [507-511] Upregulated in: clear cell renal cell carcinoma, pancreatic ductal adenocarcinomas , hepatocellular carcinoma, and Colon cancer [371, 443, 512-514]

6.3 Aims

The study aimed to investigate whether this panel of miRNAs: (1) have roles in the activation of common singling pathways involved in CRC carcinogenesis, by measuring mRNA and protein expression of some of genes that are targeted by the panel of miRNAs. (2) are suitable to use as screening biomarkers for CRC. miRNA and mRNA were measured using RT-qPCR and immunohistochemistry was used to measure protein expression in a series of 81 CRC samples and their matched metastasis.

6.4 Results

6.4.1 MiRNA quantification

6.4.1.1 MicroRNA quantification by real-time quantitative RT-PCR

To measure the expression of miRNAs including (miR- 20a, 21, 29a, 31, 92a and 224), the study screened miRNA levels in 81 CRC samples and matched normal mucosa through RT-qPCR assay, normalised to RNU6B. All assays were done in triplicate and the Ct value of all targets in all samples were less than 27 (range 16.1-26.8) with SD less than 0.5 between replicates Ct value. The miRNAs with significantly different expression in the CRC samples compared with the normal mucosa were identified by statistical analysis (Wilcoxon test (because data are not normally distributed) with an expression fold greater than (1.5). Of the six miRNAs, four had significantly higher expression levels in CRC samples compared with the corresponding normal mucosa. MiR-31 was averaged 10.83 fold higher in comparison to adjacent normal colon tissue (0.52-161.69 fold, p=0.003), miR-29a was averaged 8.11 fold higher in comparison to adjacent normal colon tissue (0.52-108.36 fold, p=0.001), miR-21 was 6.42 fold higher in CRC than in normal tissue (0.5-63.84 fold, p=0.003), miR-20a was 3.27 fold higher in CRC than in normal tissue (0.53-109.16 fold, p=0.007), miR-92a was 2.2 fold higher in CRC than in normal tissue (0.37-34.8, fold p=0.2) and miR-224 was 2.68 fold higher in CRC than in normal tissue (0.51-19.35-fold p=0.042). However, following multiple correction testing using the Bonferroni correction, miR-224 failed to retain significance (p=0.22) the rest remain in significance (miR-20a (p=0.04), miR-21 (p=0.01), miR-29a (p=0.003) and miR-31 (p=0.001) (see table 6.2).

Table 6. 2: Expression profiles of candidate microRNAs

Genes	Fold of change	*p* value	Bonferroni correction test
miRNA20a	3.27	0.007	0.04
miRNA21	6.42	0.003	0.01
miRNA29a	8.11	0.001	0.03
miRNA31	10.83	0.003	0.01
miRNA92a	2.2	0.29	0.87
miRNA224	2.68	0.042	0.22

6.4.1.2 Association of the expression of biomarkers and clinicopathological variables

In the previous chapter (chapter 5), we have defined normal and high level of the selected miRNAs level (using <0.5 fold as showing downregulation and >1.5 fold as showing up-regulation). Pearson Chi-square test was applied to identify association between miRNAs and clinicopathological features, and the results showed that normal miRNA92a expression was associated with grade two (X^2=7.037, d.f.=2, p=0.03). High miRNA21 expression was associated with Duke`s B stage (X^2=6.115, d.f.=2, 0.04). However, following multiple correction testing using the Bonferroni correction, they both failed to retain significance (p=0.28 and p=0.36, respectively) (see table 6.3).

Table 6. 3: Association between miRNAs expression and clinic-pathological variables

miR-20a					
variables		**Normal**	**high**	**P value**	**correction**
Tumour grade	well	1 (2.6%)	1 (2.3%)	0.8	0.2
	good	34 (89.5%)	40 (93.0%)		
	poor	3 (7.9%)	2 (4.7%)		
Nodal state	pN 0	21 (55.3%)	29 (67.4%)	0.3	0.7
	pN I	12 (31.6%)	12 (27.9%)		
	pN II	5 (13.2%)	2 (4.7%)		
Duke`s stage	A	5 (13.2%)	7 (16.3%)	0.5	0.9
	B	16 (42.1%)	22 (51.2%)		
	C	17 (44.7%)	14 (32.6%)		
EMVI	0	15 (39.5%)	26 (60.5%)	0.1	0.3
	1	22 (57.9%)	16 (37.2%)		
	2	1 (2.6%)	1 (2.3%)		
miR-21					
Tumour grade	well	2 (6.3%)	0 (0.0%)	0.1	0.3
	good	29 (90.6%)	45 (91.8%)		
	poor	1 (3.1%)	4 (8.2%)		
Nodal state	pN 0	15 (46.9%)	35 (71.4%)	0.08	0.2
	pN I	13 (40.6%)	11 (22.4%)		
	pN II	4 (12.5%)	3 (6.1%)		
Duke`s stage	A	2 (6.3%)	10 (20.4%)	0.04	0.1
	B	13 (40.6%)	25 (51.0%)		
	C	17 (53.1%)	14 (28.6%)		
EMVI	0	12 (37.5%)	29 (59.2%)	0.05	0.1
	1	20 (62.5%)	18 (36.7%)		
	2	0 (0.0%)	2 (4.1%)		
miR-29a					
Tumour grade	well	1 (2.8%)	1 (2.2%)	0.9	0.9
	good	33 (91.7%)	41 (91.1%)		
	poor	2 (5.6%)	3 (6.7%)		
Nodal state	pN 0	22 (61.1%)	28 (62.2%)	0.6	0.9
	pN I	12 (33.3%)	12 (26.7%)		
	pN II	2 (5.6%)	5 (11.1%)		

variables		Normal	high	P value	correction
Duke`s stage	A	4 (11.1%)	8 (17.8%)	0.6	0.9
	B	18 (50.0%)	20 (44.4%)		
	C	14 (38.9%)	17 (37.8%)		
EMVI	0	19 (52.8%)	22 (48.9%)	0.4	0.8
	1	17 (47.2%)	21 (46.7%)		
	2	0 (0.0%)	2 (4.4%)		
miR-31					
Tumour grade	well	1 (2.9%)	1 (2.1%)	0.1	0.3
	good	29 (85.3%)	45 (95.7%)		
	poor	4 (11.8%)	1 (2.1%)		
Nodal state	pN 0	22 (64.7%)	28 (59.6%)	0.1	0.3
	pN I	7 (20.6%)	17 (36.2%)		
	pN II	5 (14.7%	2 (4.3%)		
Duke`s stage	A	2 (5.9%)	10 (21.3%)	0.1	0.3
	B	19 (55.9%)	19 (40.4%)		
	C	13 (38.2%)	18 (38.3%)		
EMVI	0	14 (41.2%)	27 (57.4%)	0.3	0.7
	1	19 (55.9%)	19 (40.4%)		
	2	1 (2.9%)	1 (2.1%)		
miR-92a					
Tumour grade	well	1 (1.8%)	1 (4.2%)	0.03	0.2
	good	55 (96.5%)	19 (79.2%)		
	poor	1 (1.8%)	4 (16.7%)		
Nodal state	pN 0	38 (66.7%)	12 (50.0%)	0.2	0.6
	pN I	14 (24.6%)	10 (41.7%)		
	pN II	5 (8.8%)	2 (8.3%)		
Duke`s stage	A	9 (15.8%)	3 (12.5%)	0.3	0.7
	B	29 (50.9%)	9 (37.5%)		
	C	19 (33.3%)	12 (50.0%)		
EMVI	0	31 (54.4%)	10 (41.7%)	0.3	0.7
	1	24 (42.1%)	14 (58.3%)		
	2	2 (3.5%)	0 (0.0%)		

miR-224					
variables		**Normal**	**high**	**P value**	**correction**
Tumour grade	well	1 (2.3%)	1 (2.7%)	0.4	0.8
	good	39 (88.6%)	35 (94.6%)		
	poor	4 (9.1%)	1 (2.7%)		
Nodal state	pN 0	29 (65.9%)	21 (56.8%)	0.6	0.9
	pN I	11 (25.0%)	13 (35.1%)		
	pN II	4 (9.1%)	3 (8.1%)		
Duke`s stage	A	6 (13.6%)	6 (16.2%)	0.5	0.9
	B	23 (52.3%)	15 (40.5%)		
	C	15 (34.1%)	16 (43.2%)		
EMVI	0	23 (52.3%)	18 (48.6%)	0.9	0.9
	1	20 (45.5%)	18 (48.6%)		
	2	1 (2.3%)	1 (2.7%)		

6.4.2 Measurement of mRNA level of targeted genes

6.4.2.1 Confirmation of primer specificity

Specificity of qPCR products was checked with gel electrophoresis and resulted in a single band with the anticipated length (SMAD4 163bp, PTEN 177bp, BCL2 115bp, TGFBRII 147bp, KLF4 189bp and RASA1107 bp) (figure 6.1). In addition, a qPCR melting curve analysis was performed, which resulted in single sharp peak generated specific melting temperatures as follows: SMAD4, 78.32°C; PTEN, 76.86°C; TGFBRII, 80.05°C; BCL2, 83.72°C; KLF4, 85.43°C; RASA1, 85.08°C (Appendix D- figure 1).

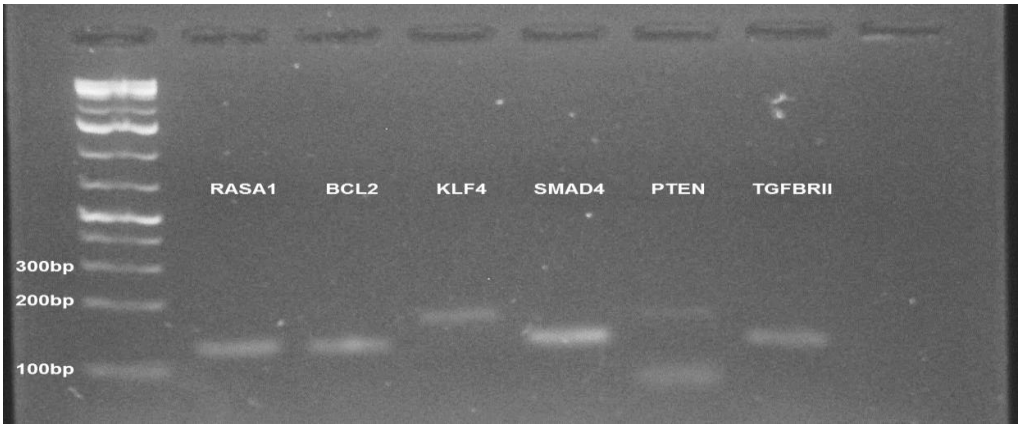

Figure 6. 1: PCR products represent exon-exon mRNA primers of SMAD4, PTEN, BCL2, TGFBRII, KLF4 and RASA1

6.4.2.2 Real-time PCR amplification efficiencies

Accuracy and reproducibility of an assay are best indicated by the coefficient of variation (CV) in a series of replicate tests. The short-term precision (intra-assay variability), which was determined in three repeats, showed a mean CV of the primers 1.6% (range 0.8% – 2.0%) for mRNA primers. The long-term precision (inter-assay variability), which investigated in three different experimental runs performed on 3 days, showed a mean CV of the miRNAs primers and 3.2% (range 1.8% – 3.9%). Investigated primers showed high real-time PCR efficiency rates; SMAD4 96%, PTEN 90%, TGFBRII 92%, BCL2 93%, KLF4 96% and RASA1 92%. (Appendix D- figure 2 & 3)

6.4.2.3 Expression of endogenous controls

In order to evaluate the expression of the HPRT to see if it is suitable to be used as an internal normalisation control, which is required to normalise sample-to-sample variations. The expression of HPRT was measured in matched 20 cases (20 colorectal samples with their matched normal mucosa) and the data demonstrated that HPRT was readily detectable in the normal and tumour tissues of patients with CRC. No significant difference was observed in terms of Ct values of HPRT (p=0.58, Wilcoxon test) between normal and tumour tissue samples [447].

6.4.2.4 Cut-off point to detect mRNAs

Before running samples to detect expression level of mRNAs, we sought to find cut-off pint to show what the high or low expression is. To achieve this goal extracted RNA from 20 pure normal colon tissue were pooled with an equal volume. The expression level of all mRNAs was measured in all normal colon tissue samples after comparing with pooled sample. In average the minimum fold of expression of all mRNAs in normal colon tissue was 0.6 and the highest was 1.8. (using <0. fold as showing downregulation and >1.8 fold as showing up-regulation).

6.4.2.5 mRNA quantification by real-time quantitative RT-PCR

The mRNAs with significantly different expression in the CRC samples compared with the normal mucosa that were identified by statistical analysis (paired t-test) using <0.6 fold as showing downregulation and >1.8 fold as showing up-regulation. All assays were done in triplicate and replicates with a Ct SD greater than 0.5 were omitted from further analysis. Of the six targets, two mRNAs of two genes including RASA1 (p=0.002 with fold change 0.66 ± SD 0.74, 95% CI 0.5-0.82) and TGFBRII (p=0.0001 with fold change 0.46 ± SD 0.78, 95%CI 0.29-0.63) had significantly lower expression levels in the tumour samples than normal tissues samples. whereas as mRNA of BCL2 (p=0.05 with fold change 3.72 ± SD 8.2, 95%CI 1.93-5.53) had a significantly higher expression level. mRNA of the rest genes including SMAD4 (p= 0.12 with fold change 0.98 ± SD 0.92, 95% CI 0,78-1.18), PTEN (p=0.22 with fold change 0.97 ± SD 0.91, 95% 0.77-1.17), and KLF4 (p=0.8 with fold change 1.16 ± SD 1.6, 95%CI 0.81-1.51) had relatively similar expression in tumour and normal tissues. However, following multiple correction testing using the Bonferroni correction, RASA1 and TGFBRII retain significance (p=0.01) and (p=0.006) BCL2 failed to retain significance (p=0.2) (table 6.4).

Table 6. 4:Expression profiles of candidate target genes mRNA

Genes	Fold of change from tumour to normal	*p* value	Bonferroni multiple correction
SMAD4	0.98	0.12	
PTEN	0.97	0.22	
BCL2	3.72	0.05	0.2
TGFBRII	0.46	0.001	0.006
KLF4	1.18	0.8	
RASA1	0.66	0.002	0.01

6.4.3 Protein evaluation

6.4.3.1 Optimisation of primary antibodies for IHC

Before staining the proteins of interest, in CRC tissues, using IHC, the specificity of anti-SMAD4, anti-TGFBR2, anti-RASA1 and anti-KLF4 antibodies were tested using Western blotting. The primary antibodies are against SMAD4, TGFBR2, RASA1 and KLF4 proteins with molecular weights of 65 KDa, 75 KDa, 140 KDa, and 55 KDa, respectively. As shown in (figure 6.2A), anti-SMAD4 antibody detected the expected band (65 KDa) in SW480 cell lysate, but not in other human colorectal cell lines HCT116, HT29, RKO and Lovo. The Western blotting of RASA1 (figure 6.2B) showed the expected band (140KDa) in HT29 and Lovo while KLF4 (figure 6.2C) antibody showed predicted bands and 55 KDa, respectively) in the HT29 cell lysate, but not in other human colorectal cell lines. Finally, as shown in (figure 6.2D), TGFBR2 antibody detected the predicted band at around 75 KDa in HT29 and SW480. Such Western blotting results demonstrated the specificity of these antibodies, which were to be used to stain the target proteins in CRC TMAs using IHC. The remaining antibodies used in our project were taken from our groups and the histopathological department at Nottingham Queens Medical Centre, and the specificity had already been tested. The concentration of each antibody applied in IHC was optimized further. Different concentrations of each antibody were stained to get an optimal concentration. As shown in (figure 6.3) a heterogeneous staining pattern, with no background staining, was obtained for anti-SMAD4 at 1:100, while, of the other concentrations

(1:50) was too high and 1:200 was too low to detect expression. TGFBR2 gave a heterogeneous staining at 1:400, but too high at 1:200, and 1:300. Anti-RASA1 at 1:40 gave a heterogeneous staining pattern with no background staining. The other concentrations (1:50, and 1:100) gave weak staining of RASA1. The optimal concentration of anti-KLF4 was 1:100. Background staining of anti-KLF4 was observed at 1:50, while it gave weak staining at 1:200. Regarding BCL2 and PTEN, Histopathology department at Queens Medical Centre and James Hassall (PhD student) performed staining for BCL2 and PTEN respectively.

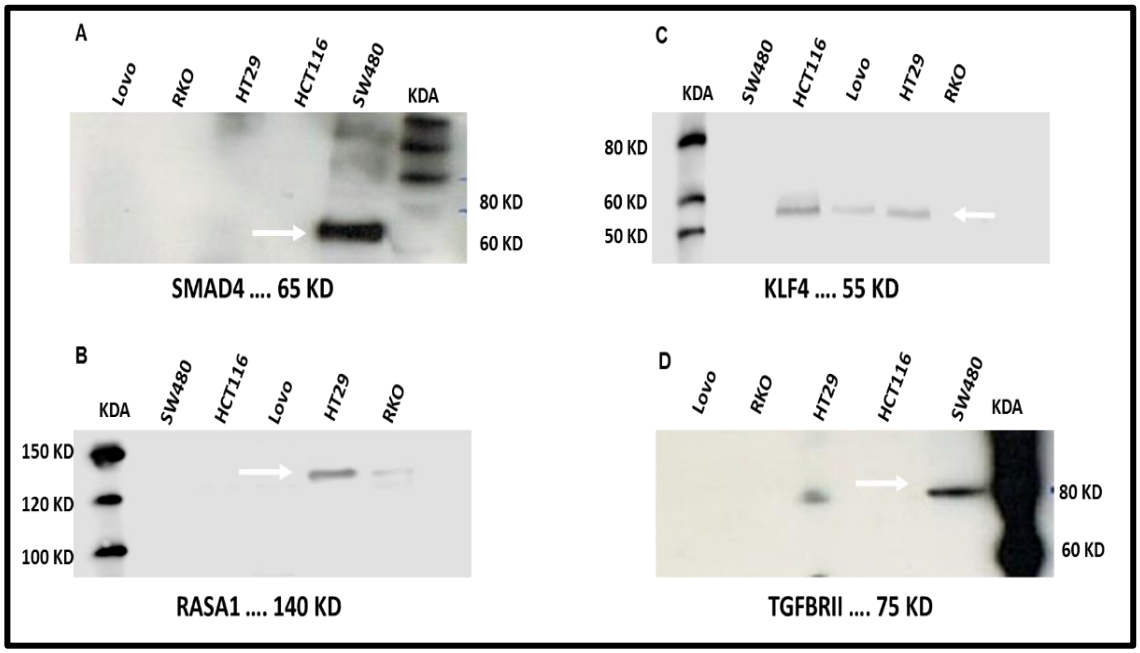

Figure 6. 2: Expression of SMAD4, RASA1, KLF4 and TGFBRII in different cell lines. Lysates from human colon cancer were subjected to Western blotting to assess the expression of (A) SMAD4, (B) RASA1, (C) KLF4, and (D) TGFBRII.

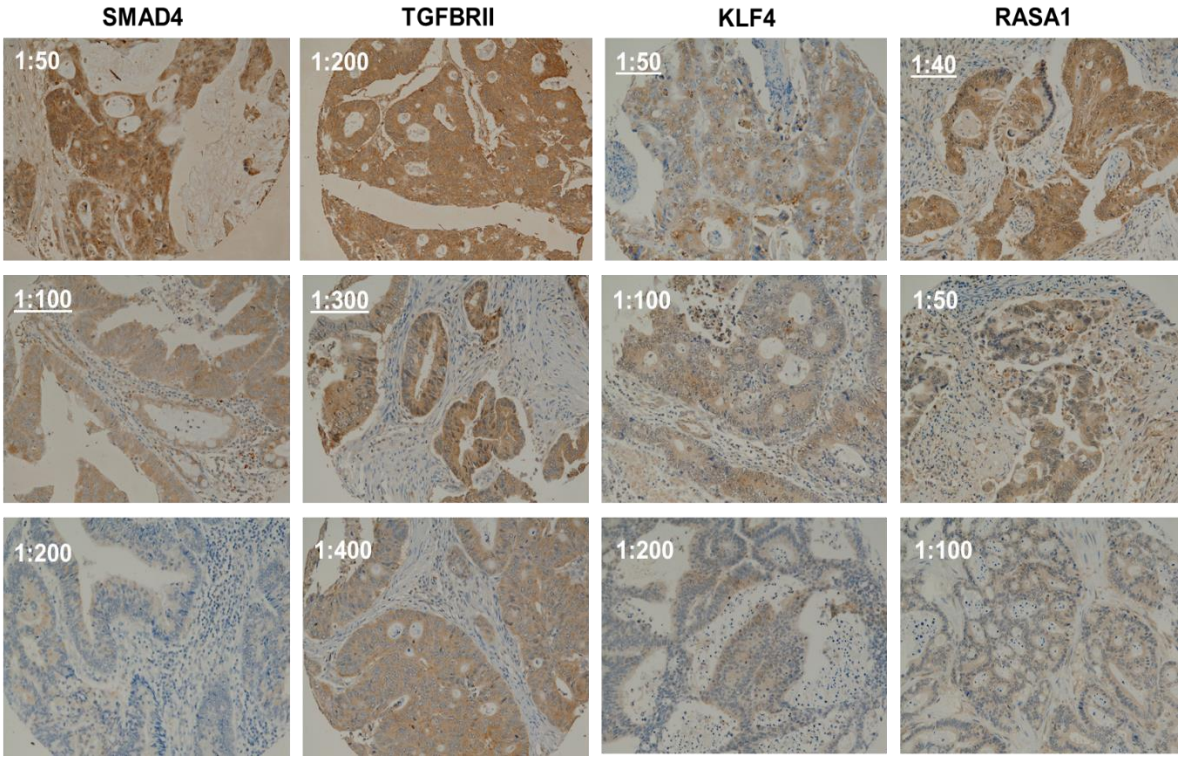

Figure 6.3: Optimisation of anti-SMAD4, anti-TGFBRII, anti-KLF4 and anti-RASA1 antibodies. Different concentrations of each antibody were applied on colorectal tumour composite sections, and optimal concentration was chosen depending on heterogeneous staining pattern with no background staining (200X magnification)

6.3.2 Protein expression of target genes

In order to further explore expression of miRNAs target genes, expression of SMAD4, PTEN, TGFBRII, BCL2, KLF4 and RASA1 was assessed in a cohort of primary operable invasive CRC patients. The staining pattern for all markers was heterogeneous between, as well as within, certain tumour cores varying from weak to intense (figure 6.4). We stained three different cores per case and after scoring average of 3 cores were taken for further analysis. The H-score was determined for the six markers; SMAD4 had a median H-score of 85, range between 0 and 300, PTEN had a median H-score of 95, range between 0 and 300, TGFBRII had a median H-score of 70, range between 0 and 225, BCL2 had a median H-score of 125, range between 0 and 300, KLF4 had a median H-score of 90, range between 0 and 300, RASA1 had a median H-score of 65, range between 0 and 200. As mentioned above, single measure intra class correlation

coefficients (ICC) between scorers were 0.78, 0.81, 0.73, 0.77, 0.71 and 0.82 for SMAD4, PTEN, TGFBRII, BCL2, KLF4 and RASA1, respectively, showing excellent concordance between both scorers. The mean of scores was used to stratify the specimens into low and high expression groups. For SMAD4 51 (63%) CRC cases having low expression and 30 (37%) cases having high expression than normal mucosa, for PTEN 67 (83%) CRC cases having low expression and 14 (17%) cases characterized as high expression than normal mucosa, for TGFBRII 65 (80%) CRC cases being characterised as low expression and 16 (20%) CRC cases having high expression than normal mucosa; for BCL2 47 (58%) CRC specimens having low expression and 34 (42%) CRC specimens were characterised as high expression than normal mucosa; for KLF4 36 (44%) CRC cases having a low expression and 45 (56%) CRC cases characterised as high expression than normal mucosa; and for RASA1 61 (75%) CRC cases having a low H-score, and 20 (25%) CRC cases showing high expression than normal mucosa. Chi-square test was conducted to assess the correlation between mRNA and protein level of the markers used in this study. Results demonstrated a significant correlation between mRNA and protein levels in SMAD4 ($r=0.466$, $p<0.0001$), TGFBRII I ($r=0.708$, $p<0.0001$), BCL2 ($r=0.623$, $p<0.0001$), and RASA1 ($r=0.728$, $p<0.0001$). No correlation was observed between mRNA and protein levels in PTEN ($r= -0.085$, $p= 0.450$), and KLF4 ($r=0.114$, $p=0.313$).

6.4.4 Association the expression of biomarkers (target genes) and miRNAs

Spearman's rank order correlation was conducted to assess the correlation between the markers used in this study. High miR-21 and miR-224 expression were associated with low expression of TGFBRII-mRNA [($r= -0.358$, $p=0.001$), ($r= -0.276$, $p=0.01$)] and TGFBRII-protein [($r= -0.328$, $p= 0.003$), ($r= -0.319$, $p=0.004$)] respectively. In addition, over expression of both miR-29a and miR-31 inversely correlated with RASA1-mRNA level [($r= -0.217$, $p=0.01$), ($r= -0.276$, $p=0.01$)] and RASA1-protein level [($r= -0.222$, $p=0.004$), ($r= -0.209$, $p=0.01$)] respectively, (see table 6.5). Significant correlation was observed between miR-20a and miR-29 ($r=$

0.380, p=0.0001) and miR-31 (r= 0.403, p=0.0001). Significant correlation was seen as well between miR-21 and miR-29a (r= 0.526, p=0.0001) and miR-31 (r= 0.285, p=0.01). Moreover, high correlation was observed between miR-29a and miR-31 (r= 0.275, p=0.01) and miR-92a (r= 0.324, p=0.003). Significant correlation was seen as well between miR-31 and miR-224 (r= 0.328, p=0.003) and between miR-92a and miR-224 (r= 0.382, p=0.0001) (see table 6.6).

Table 6. 5: Association between miRNAs and target genes (mRNA and protein)

Variables		TGFBRII-mRNA	TGFBRII-protein	RASA1-mRNA	RASA1-protein
miR-21	CC	-0.358	-0.328		
	p-value	0.001	0.003		
miR-29a	CC			-0.217	-0.222
	p-value			0.01	0.004
miR-31	CC			-0.276	-0.209
	p-value			0.01	0.01
miR-224	CC	-0.276	-0.319		
	p-value	0.01	0.004		

CC= correlation coefficient

Table 6. 6: Association between miRNAs

Variables		miR-29a	miR-31	miR-92a	miR-224
miR-20a	CC	0.380	0.403		
	p-value	0.0001	0.0001		
iR-21	CC	0.526	0.285		
	p-value	0.0001	0.01		
miR-29a	CC		0.275	0.324	
	p-value		0.01	0.003	
miR-31	CC				0.328
	p-value				0.003
miR-92a	CC				0.382
	p-value				0.0001

CC= correlation coefficient

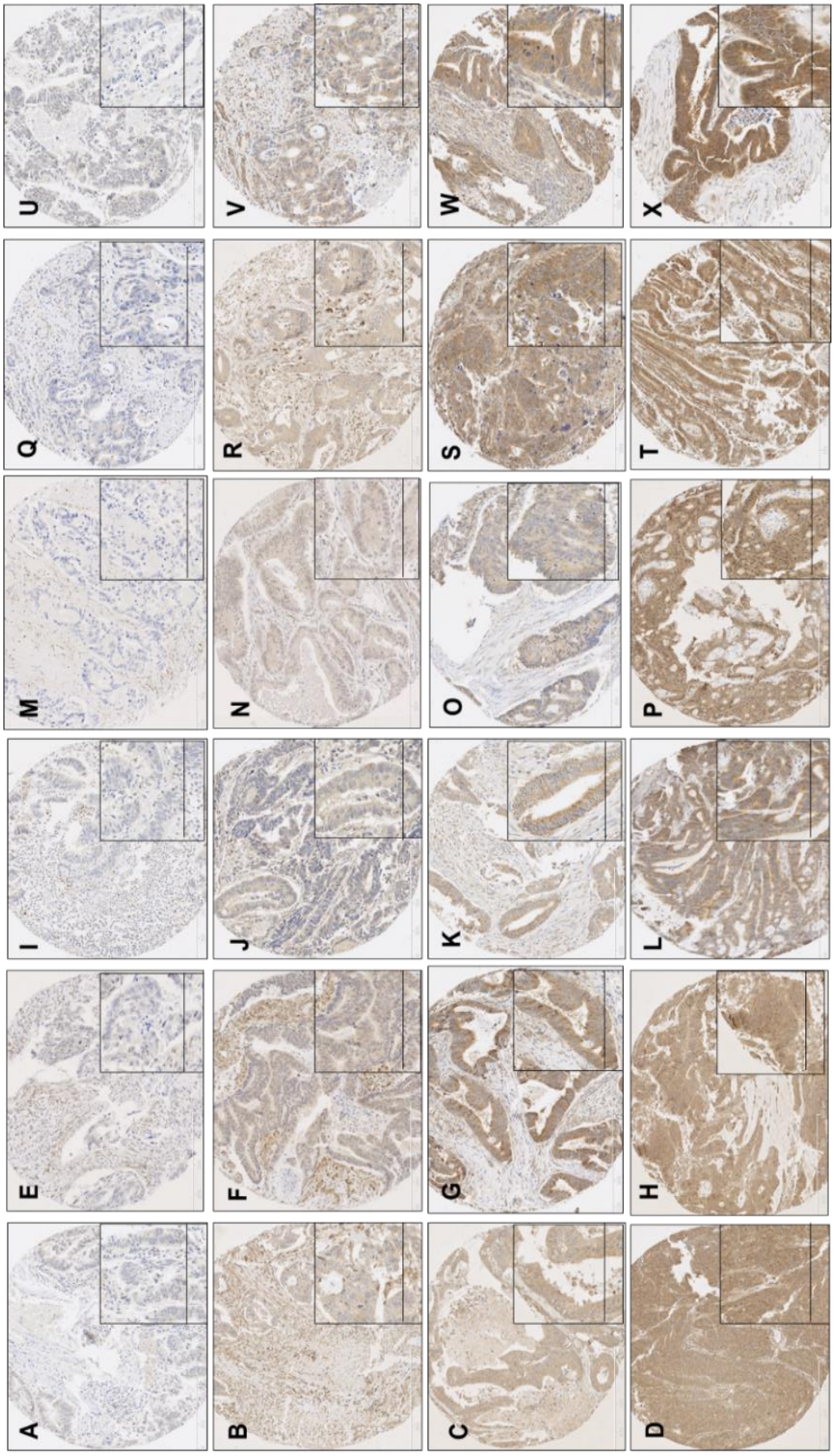

Figure 6. 4: Representative photomicrographs of staining patterns of markers in colorectal tissue. (A) None, (B) weak, (C) moderate, and (D) strong staining of SMAD4, (E) negative, (F) weak, (G) moderate, and (H) strong staining of PTEN, (I) negative, (J) weak, (K) moderate, and (L) strong staining of BCL2, (M) negative, (N) weak, (O) moderate, (P) strong staining of TGFBRII, (Q) negative, (R) weak, (S) moderate, and (T) strong staining of KLF4, and (U) negative, (V) weak, (W) moderate, (X) strong staining of RASA1 (Photomicrographs; 100X magnification, inset box; 200X magnification; scale bar,100 µm

6.5 Discussion

It is undeniable to say that a supreme screening method must have both high sensitivity and high specificity. Biomarkers with high false positive or negative rates will be considered ineffective and cannot be used [515, 516]. Then, an improved, reliable, accurate and non-invasive biomarker is still a need to improve the detection of CRC, mainly at early disease stages before the cancer metastasizes and becomes incurable [285]. Now researchers and clinicians are focused on miRNAs as biomarkers for cancer screening because recently different studies reported that miRNAs play a vital role in the development and progression of CRC [517, 518]. In this part of our study we investigated the expression of a panel of six oncogenic miRNAs to see whether these miRNAs could be suitable to be used as a marker for CRC detection.

The accurate quantification of mRNA of the targeted genes is reliant upon the selection of a good endogenous control for normalizing quantitative qPCR data. The HPRT was tested and results showed that expression is almost similar between tumour and normal corresponding tissue samples. This result is in line with the study identified HPRT as the best reference gene that could be used as an accurate endogenous control for the measurement of multiple housekeeping genes [519]. Next, the efficiency of the primers was evaluated and results demonstrated that all primers used to estimate expression of targeted mRNAs are more reliable as R2 linear regression ranged between 90-99% [520].

The present study evaluated the expression level of six miRNAs which are potential diagnostic biomarkers and prognostic factors in cancers. For example, For example, Chai et al, [413] reported that miR-20a has an oncogenic effect in CRC tissue samples and overexpression contributed to the resistance of colorectal adenocarcinoma to chemotherapeutics. Other studies highlighted that overexpression of miR-21 is associated with poor survival and response to chemotherapy in CRC [322,323,521,522]. Moreover, others found that miR-21 expression correlates with clinical stage

and could be a potential diagnostic marker, a prognostic marker and it could predict the pathological tumour response to chemotherapy [440, 523-525]. Furthermore, Tang et al, [468] found that high expression of miR-29a has a great role in cancer metastasis in CRC tumorigenesis through upregulation of MMP2 and downregulation of E-cad via targeting.

The study profiled 81 paired tumours and matched normal mucosa samples and statistical analysis identified significantly elevated levels of MiR- miR-20a (p=0.04), miR-21 (p=0.01), miR-29a (p=0.003) and miR-31 (p=0.001), this result is in line with other data reported before and detecting that the above miRNAs expression increases dramatically and are diagnostic and prognostic markers in CRC tissue [326, 369, 370, 468, 526-528]. However, Ahmed et al, [313] and Tao-Wei et al, [529] showed down regulation of miR29a and miR224 in CRC. On the other hand, the study also found a non-significant increase in the expression of, miR-92a and miR-224 in CRCs when compared to normal matched mucosa. This result is not consistent with other studies conducted before and demonstrated an increased level of expression of miR92a and miR224 in CRCs [304,317,447,528,530]. Differential expression of selected miRNAs makes it possible to use them as diagnostic biomarkers for CRC.

Although the sample size in this study is limited, clinical associations were also analysed and results highlighted no significant correlation between miRNAs and clinic-pathological features. Whereas other studies reported contrary results, for example, Expression level of miR-21 was clearly discernible, with significantly higher levels in stage IV patients compared with stage I or II patients [319]. Schepeler et al, [304] found that miR-31 is significantly increased in stage IV tumours compared to stage II CRC tumours and Zhou et al, [531] showed that high expression of miR-92a correlated with advanced clinical stage, lymph node metastases, and distant metastases.

As many studies, have demonstrated that miRNAs, which have been chosen in this study, are overexpressed and promote colon cancer

formation by either up-regulating oncogenic or down-regulating tumour suppressor genes. For example, the miR-20a regulated tumour suppressor gene PTEN, which has a vital role in the inhibition of progression of survival pathways including PI3K/Akt and mitogen activated protein kinase pathways [482]. Thus, the study analysed mRNA and protein levels of the targeted genes to see is there any correlation between target genes and selected miRNAs, if yes then find whether miRNAs suppress or activate targeted genes at mRNA or protein levels. Despite finding high correlation between mRNA and protein expression in *SMAD4*, *TGFBRII*, *BCL2* and *RASA1*, only down regulation of RASA1 and TGFBRII were significantly associated with the miRNAs upregulation as follow; High miR-21 and miR-224 expression were associated with low expression of TGFBRII, and over expression of both miR-29a and miR-31 inversely correlated with RASA1. This result is in line with previous data reported before [369, 493, 532]. However, into our knowledge no one reported any correlation between miR 224 and miR 29a with TGFBRII and RASA1 expression respectively. Moreover, downregulation of RASA1 and TGFBRII in CRC samples is in line with other data reported before and detecting that the RASA1 and TGFBRII decreases dramatically in CRC tissue [369, 533]. Additionally, finding no correlation between mRNA and protein levels of PTEN in one hand and high correlation between upregulated miR-20a and low protein expression of PTEN not mRNA in the other hand indicated that miR-20a suppress PTEN at protein level. Despite the present study did not do any functional invitro study, the high correlation between miRNAs and target genes indicated that selected miRNAs may have roles in CRC carcinogenesis as they control some genes which have role in different signalling pathways involved in the development of CRC. For example, up-regulation of miR-31 has the potential to drive tumour progression by down regulating RASA1, which acts as a suppressor of RAS function [534]. Suppressing RASA1 leads to the activation of the Ras protein, thereby leading to aberrant intracellular signalling through the RAS-Raf-MAPK and RAS-PI3K-AKT pathways, causing an increase in cell

proliferation, anti-apoptosis pro-survival signals and induce cell malignant transformation [535].

The design of the current study has some strengths. For example, tissues of CRC tumours and matched normal mucosa were used and by including the miR expression profiles of adjacent normal tissue, the influence of non-tumorous miRs on the tumours miRs expression was further minimized. However, this study has several limitations. First, the number of samples and the amount of measured miRs were relatively small. Second, the data was not validated by other methods. Third we did not do screening of mutations to find what the correlation is between them.

In summary, upregulation of four miRNAs (has-mir-20a, 21, 29a and 31) in CRC samples may have roles in CRC progression by targeting genes involved in several signalling pathways. Additionally, these miRNAs may be used as diagnostic biomarkers as they differentially expressed in CRCs.

In the future, we would continue to find different pathways through targeting different genes commonly involved in CRC and finding the correlation if present between miRNAs with gene mutations and consequently roles of miRNAs in CRC pathways.

In addition to the nucleic acids (DNA, RNA), which are taken from cells and used as biomarkers, the next chapter (chapter 7) describe the feasibility of using cell free nucleic acids (cfNAs) as a diagnostic, prognostic and predictive markers in patients with CRC.

Chapter Seven: Circulating free DNA in colorectal cancer as a marker of surgical clearance

7.1 Abstract

Background: cfDNA and miRNA are released from tumour tissue and circulate in the blood of cancer patients. Genomic alteration and change in the level of circulating nucleic acids have been associated with tumour burden and malignant progression. Circulating nucleic acids may serve as a novel class of minimally invasive biomarkers for surveillance, cancer screening, prognosis, detection of recurrence, monitoring minimal residual disease and monitoring therapeutic response. The study sought to characterize the use of monitoring genomic alterations and the level of circulating miRNA as a means of measuring the level of surgical clearance.

Method: cfDNA and total RNA were extracted from blood collected from 5 CRC patients and from three healthy volunteers. For the patients, blood was collected prior to surgery and where possible, daily post-surgery until discharge. Genotyping of cfDNA was performed using Cold-PCR/HRM and miRNA quantification was done using RT-qPCR methods.

Results: Three mutations were detected in patients no.1 and no.4, which disappeared after the 2nd day post-operation. Whereas other mutations which were detected in patients no.2, no.3 and no. 5 were present even at the 5th day post-operation. Although miRNA expression fluctuated between these time points, paired t-test showed a non-significant difference when comparing pre- and post-surgical miRNA levels.

Conclusion: Patients can be divided into a group which either loses or retains mutant cfDNA and all have shown non-significant change in the expression levels of miRNA following operation. However, it is possible but still unclear to conclude whether the use of cfNAs are beneficial biomarkers for surgical clearance due to the small number of patients.

7.2 Introduction

Although surgical treatment is the key to prolonged survival of patients with CRC, post-surgical follow-up is important for the early detection of relapsing or of disease progression. Currently no effective diagnostic means are available to identify which patients are disease-free after surgery (i.e. cured) and which have residual disease which may give rise to recurrence. As a result, patients with high-risk clinical and pathologic criteria are indiscriminately treated with adjuvant chemotherapy despite the fact that this therapy is frequently not needed because a large number are cured and do not need such potentially toxic therapy [536]. Currently clinical and pathological parameters (mainly the tumour-node-metastasis staging system) are the most important criterion used to predict which patients are likely to have residual disease and therefore need adjuvant therapy.

The use of DNA as a biomarker in clinical medicine has been a significant advancement, particularly for early diagnosis, prognosis, and monitoring of therapy[537]. In advanced cancer patients, when repeated tumour biopsies are not always feasible or when tissue specimens are difficult to obtain, use of circulating DNA assay is an attractive alternative option because circulating tumour DNA fragments contain all cancer related molecular alterations including somatic point mutations, DNA methylation changes, translocations, polymorphisms, loss of heterozygosity (LOH) and gene copy number changes [538]. Recently, Diehl et al, [539] has found that cfDNA is sufficiently sensitive to detect minimal residual disease after surgical resection. Moreover, in patients with early stage breast cancer, Beaver et al, [540] used tumour cfDNA to detect *PIK3CA* mutations and to monitor residual disease after surgery. On the other hand, assessment of cfDNA concentration was used to monitor the presence of residual tumour cells after surgery. For example Catarino et al, [541] found that following surgery in cancer patients the levels of cfDNA may decrease to levels that are observed in healthy individuals. However, remaining high level of cfDNA is an indicator of the presence of residual tumour cells [542]. Additionally,

Danese et al, reported increasing levels of plasma cfDNA in early stage CRC indicating that cfDNA may be suitable to be used as a marker for identifying high-risk individuals [543]. Others demonstrated that high levels of circulating cfDNA, which have been observed in several tumours such as prostate cancer, lung cancer, pancreatic cancer, leukaemia and lymphoma, was correlated with tumour metastasis, response to therapy, prognosis and recurrence [544]. The study conducted by Luo et al, demonstrated that cfDNA might be a promising screening test in non-small cell lung cancer through detection EGFR mutations in cfDNA [545]. However, other studies found that the amount of cfDNA without knowledge of tumour mutations, is not a useful diagnostic tool and that the utility of cfDNA analysis is limited [546].

7.2.1 Circulatory cell-free nucleic acid

For the first time in 1948, cell-free nucleic acid (cfNA) in human blood was reported by Mandel and Métais and they were able to detect circulating cfNA in the plasma of healthy and sick individuals. In the 1960s and 1970s further interest in this field was developed when researchers discovered high levels of circulating cfNA in patients with systemic lupus erythematosus, and cancer. Furthermore, it was demonstrated that the amount of circulating cfNA was higher in individuals with metastatic disease and, in a number of cases, the levels of circulating cfNAs decreased with successful chemotherapy [547]**.** After that a study conducted by Stroun *et al.* proposed that the circulating DNA that is found in cancer patients reveals certain features of tumour DNA [548]. In 1994 this important suggestion attracted the attention of the scientific community when the importance of cfNA was recognized as a consequence of the discovery of mutated RAS gene fragments in the plasma of cancer patients [549]. In 1996 Nawroz et al, [550] had shown microsatellite alterations on cell-free DNA (cfDNA) in cancer patients. These latter reports support that the tumour cells are able to release their cfNA into the circulation. Although the plasma of healthy

individual contains little cfNA, the increased levels of cfNA in blood of patients indicated that it may be used as a liquid biopsy for the non-invasive diagnosis of a variety of clinical conditions and would avoid the need for tumour tissue biopsies. Generally, cfNA yields are higher in cancerous and non-cancerous patients such as diabetes, trauma, inflammatory diseases, stroke, rheumatoid arthritis etc. [551]. In addition studies of circulating nucleic acids have shown that foetal-derived nucleic acids in maternal plasma in pregnant women can be used in prenatal diagnosis to detect and monitor foetal diseases and pregnancy associated complications [544]. Normally liver and kidney remove nucleic acids from the blood, which have a variable half-life in the circulation ranging from 15 minutes to several hours [552]. In cancer patients, the amount of cfNA in blood is not stable and continually changes by several factors. For example, state and size of the tumour and some physiological factors such as clearance, degradation and other filtering events of the blood and lymphatic circulation are other factors have the influence on the proportion of cfNA [553]**.**

7.2.2 CFDNA

Recently several studies demonstrated that a small quantity of extracellular nucleic acids are also circulating in the bloodstream [554]. In healthy individuals, concentration of this free circulating DNA cfDNA, is not more than a few nangrams per millilitre (ng/ml), [536]. This is composed of both genomic DNA (gDNA) and mitochondrial DNA (mtDNA). In plasma in contrast to circulating mtDNA, which exists in both free and particle-associated forms, gDNA generally circulates in a cell-free form [555]. Although cfDNA levels in healthy plasma samples are lower than cancer patients, since most non-living cells are removed efficiently from the circulation by phagocytes, the concentrations of cfDNA vary considerably in both groups [542, 556, 557]. In healthy control donors, a range of 0-100ng/ml, with an average of 30ng/ml cfDNA has been measured. By comparison cancer patients have concentration between 0-1000ng/ml cfDNA of blood, with an average of 180 ng per ml cfDNA [552]. Although

apoptosis is a main source of DNA in circulation, several minor sources include break down of blood cells, spontaneous release of newly synthesized nucleic acids, cell lysis by the necrotic pathway, and break down of pathogens such as viruses and bacteria [537, 558, 559]. The source of circulating DNA might be detected by molecular weight. For instance, due to the DNA release by apoptosis being mainly 100-1000bp in size, fragments form a ladder when is separated by gel electrophoresis. In contrast necrosis leads to DNA fragments of about 10,000bp due to incompletely and non-specifically digested DNA and thus smears are seen on electrophoretic separation [548]. In oncology, however there are 3 main challenges to measurement of circulating tumour DNA (ctDNA): discrimination of cfDNA, which released by tumour cell and known as circulating tumour DNA (ctDNA) from normal cfDNA, presence of low levels of ctDNA and the accurate quantification of the number of mutant fragments in a sample. The presence of somatic mutations, which are commonly single base-pair, is the main way to discriminate ctDNA from normal cfDNA because these mutations are present only in the genomes of cancer cells or precancerous cells and are not present in the DNA of normal cells of the same individual.

7.2.3 CFMiRNA

Circulating cell-free miRNA is another type of cfNA, and recent studies revealed that it is stably expressed in plasma or serum, which enables it to be a potential biomarker for non-invasive diagnosis of cancer and other diseases [560]. Although expression of tumour miRNAs affect the level of circulating miRNAs, the mechanism of releasing into the circulation is largely unclear. Circulating miRNAs are most likely protected within protein or lipid vesicles [561] and also can be passively leaked from apoptotic or necrotic cells. [562]. After secretion of miRNAs by the cell, they can bind to specific proteins and associate with multivesicular bodies (MVBs) and exosomes. Since most cells secret exosomes, all species of miRNAs might be detected in circulation [563]. Recently various studies have confirmed circulating

miRNAs are differentially expressed in CRC patients in comparison to the healthy individual and are suitable to use for diagnosis. For example, miR-31 [369], miR-20a [564], miR221 [565], were differentially expressed in the preoperative plasma samples of CRC patients. This indicated that these miRNAs may be a non-invasive molecular marker for CRC diagnosis. Other studies have shown that the high level of circulatory cell free miRNAs return to a normal level after tumour resection and this makes it possible to be used as a biomarker for monitoring and surgical clearness. For example, up-regulated miR-21, miR-29a, miR-92a, in CRC, miR-106b in gastric cancer, miR-195 in breast cancer were significantly decreased after resection of tumours [566, 567].

7.3 Aim of the study

The study sought to optimize the methodology for extracting, quantifying, and assessing cfNAs. We then aimed to assess surgical clearance through screening mutations of some genes (KRAS, TP53, PIK3CA, BRAF, PTEN and SMAD4) in cfDNA using HRM and quantifying the level of circulating miRNA (miR20a, 21, 29a, 31, 92a and 224) using RT-qPCR method in a time series pre-surgery (day of surgery) and post-surgery (the day after surgery and daily after that until patients discharged from hospital).

7.4 Results

7.4.1 Screening of mutations in cfDNA

7.4.1.1 QIAamp mini Kit

Based on data that were reported previously, this study used the Qiamp mini kit which was one of the best kits to purify cfDNA from plasma samples. Additionally, we used two measures to assess DNA recovery from plasma samples. DNA was extracted from 3 plasma samples collected from three healthy donors and quantified by nanodrop results, which as shown in table 7.1, which revealed that the mean 260/280 ratio of cfDNA was 1.84 and overall the average cfDNA yield from 200µl of sixteen blood samples was 11.1µg/µl (range 8.7 to 13.4µg/µ). For the second measure, DNA was assessed by PCR following spiking of known mutations into bovine calf serum as follow; DNA was extracted from HCT116 (containing a heterozygous G13D mutation in *KRAS*) using Qiamp mini kit and spiked into bovine calf serum to get different concentration of targeted DNA/µL as follow; (50 ng, 25 ng, 12.5 ng, 6.25 ng, 3.125 ng, 1.56 ng, 0.78 ng, 0.39 ng, 0.194 ng, 0.097 ng, 0.049 ng, 0.025 ng and 0.012 ng. The DNA from the resulting diluted mixtures was re-purified and amplified by PCR and the resulting sequences were analysed using HRM analysis. Figure 7.1 shows amplification for *KRAS* from DNA extracted from a mixture containing as little as 25 pg of DNA in 200 µl liquid. We were also getting DNA recovery even without spiking in any HCT116 which represents the bovine DNA and which did not show amplification with primers for human *KRAS*.

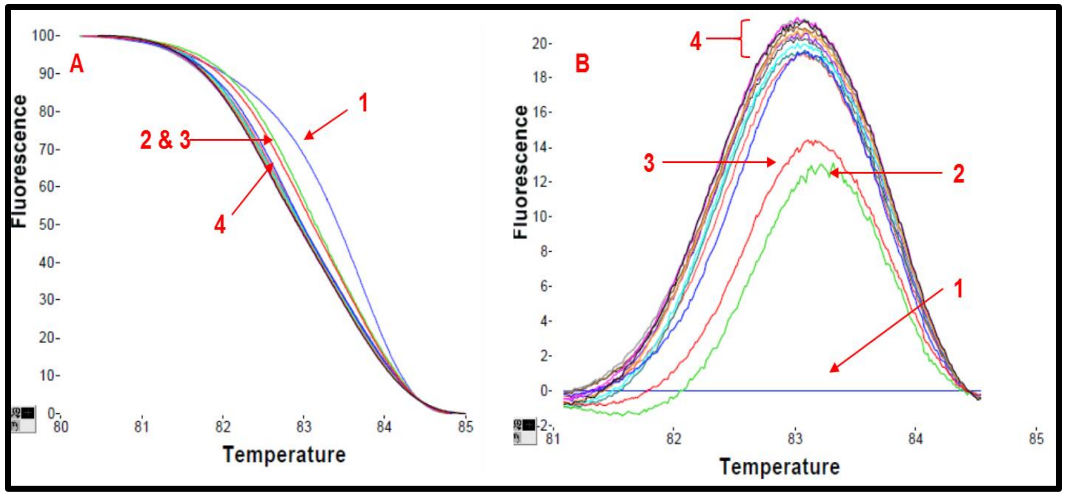

Figure 7. 1: HRM analysis shows that the Qiamp mini kit could purify as few as 25pg of DNA in liquid sample. (1) Vaco5 (wild for KRAS), (2) diluted HCT116 (0.025ng), (3) diluted HCT116 (0.05ng), (4) HCT116 (from undiluted down to 0.1ng). There is no PCR amplification for DNA from diluted HCT116 (0.012ng) and bovine calf serum.

7.4.1.2 Extraction of cfDNA

In the present study for the analysis of genetic alterations in cfDNA, it was necessary to obtain sufficient amount of pure DNA free from any RNA or protein contaminations. For approaching this goal, the Qiamp mini kit was used for extracting cfDNA from overall samples (plasma) which were collected from five CRC patients. Overall the average cfDNA yield from 200µl of sixteen blood samples was 7.45µg/µL (range 2.5 to 15.8µg/µL). The DNA purity of all samples was estimated, and the results showed that the mean 260/280 ratio of cfDNA was 1.76 (SD 0.085) (table 8-1).

Table 7.1: shows the dates, amount and purity of cfDNA collected from plasma. (Day1=day of surgery means collect blood just before operation, Day2 is the first day post-surgery, Day3 is the second day post surgery etc., N is a normal healthy case)

Patients No.	Days of collection	Total DNA (ng/µl	DNA purity (260/280) ratio	Total RNA (ng/µl	RNA purity (260/280) ratio
1	Day1	7.1	1.78	41.3	1.98
	Day2	6.3	1.79	38.5	1.87
	Day5	2.8	1.75	39.8	1.91
	Day6	2.5	1.72	43.1	1.94
2	Day1	8.7	1.87	58.4	1.88
	Day2	8.0	1.73	64.2	2.1
	Day5	10.2	1.71	71.3	2.05
3	Day1	9.3	1.88	82.7	1.84
	Day4	15.8	1.74	72.6	1.87
4	Day1	4.3	1.72	32.1	1.94
	Day2	6.7	1.73	38.9	1.82
	Day3	11.0	1.75	49.5	1.99
	Day4	7.1	1.73	41.7	1.81
5	Day1	4.3	1.75	92.4	2.07
	Day4	6.4	1.75	76.8	201
	Day5	8.7	1.72	94.1	1.98
1N	-	8.7	1.86	56.1	1.86
2N	-	13.4	1.91	79.7	1.84
3N	-	11.2	1.75	48.4	1.93

7.4.1.3 Wild type blocking PCR

Due to the amount of cfDNA being very low and screening mutations in cfDNA needing a very sensitive method, the study sought to improve the limit of detection of the HRM method to less than 1%. To achieve this goal, the study used Wild-type blocking PCR, which is a modified form of standard PCR by adding LNA probes to favour mutant DNA amplification during PCR. Two LNA/DNA chimera oligonucleotide (mentioned in section 2.2.1.4.3.3) were used to inhibit PCR amplifications of human KRAS gene/exon2 wild-type codons 12 and 13. Despite using these two LNA sequences and using

previously optimized protocols, the study couldn't achieve this more sensitive PCR with HRM. The designed probes were not only not able to completely block wild sequence but they also suppressed amplification of the mutant DNA sequence as well (figure 7.2).

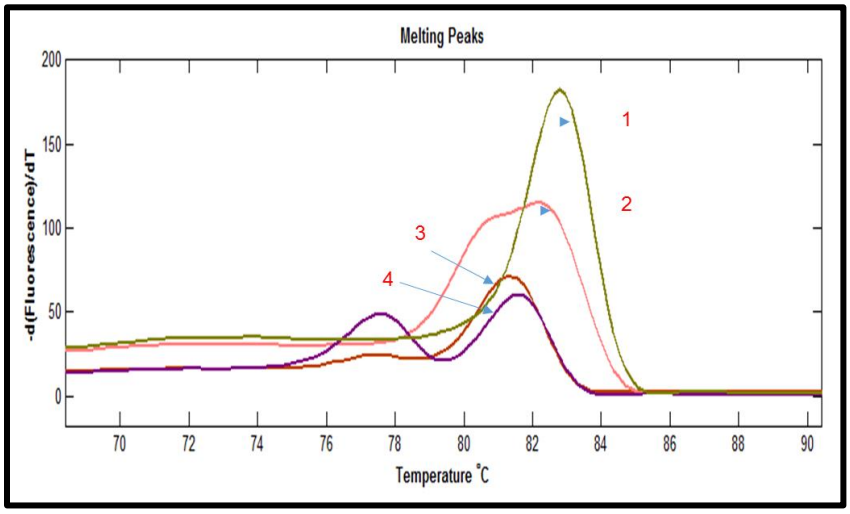

Figure 7.2: HRM of a standard and WTB-PCR-amplified sample (HCT116 cell line) with known germline mutation at KRAS codon 13 and VACO5 wild type showing a decrease in the curve peak on the derivative plots after adding. 1: VACO5 (wild), 2: HCT116 (mutant) (Standard PCR), 3: VACO5 (wild) (Cold-PCR), 4: HCT116 (mutant) (WTB-PCR)

7.4.1.4 Mutation screening

DNA extracted from sixteen plasma samples from five CRC patients were analysed for the presence of somatic mutations in 5 genes including (*KRAS*, *BRAF*, *TP53*, *PIK3CA* and *SMAD4*) using standard PCR followed by HRM. Taking Ct value less than 30 for all tested targets and an amplification curve with exponential phase and amplitude, identified good quality of PCR (figure 7.3). In patient no.1 *KRAS*/exon2 and TP53/exon6 mutations were detected, the *KRAS*/exon2 mutation, which appeared in the (1st day), disappeared in the plasma samples that were collected the days after operation (figure 7.4A), whereas *TP53*/exon6 mutation was observed in the (1st and 2nd days) and disappeared after that (figure 7.4B). *PIK3CA*/exon9

mutation was detected in plasma samples collected in the 1st day and days after operation from patient no.2 and no.3 (figure 7.5A&B). In patient no.4 *BRAF*/exon15 and *TP53*/exon5 mutations were detected in the (1st day) plasma sample before operation whereas they both disappeared from the samples collected the days after operation (figure 7.6A&B). Patient no. 5 showed *TP53*/Exon8 mutation in the plasma samples collected before and after operation (figure 7.7). Mutated genes are listed in (table 7.2).

Table 7. 2: List of mutation detected in plasma samples, collected as pre- and post-surgery

Case No.	Gene mutated	Exon	Mutation detected					
			Day 1	Day 2	Day 3	Day 4	Day 5	Day 6
1	KRAS	2	Y	N	N/A	N/A	N	N/A
	TP53	6	Y	Y	N/A	N/A	N	N/A
2	PIK3CA	9	Y	Y	N/A	N/A	Y	N/A
3	PIK3CA	9	Y	N/A	N/A	Y	N/A	N/A
4	BRAF	15	Y	N	N	N	N/A	N/A
	TP53	5	Y	N	N	N	N/A	N/A
5	TP53	8	Y	N/A	N/A	Y	Y	N/A

Y=mutation, Y=wild, N/A=non-available

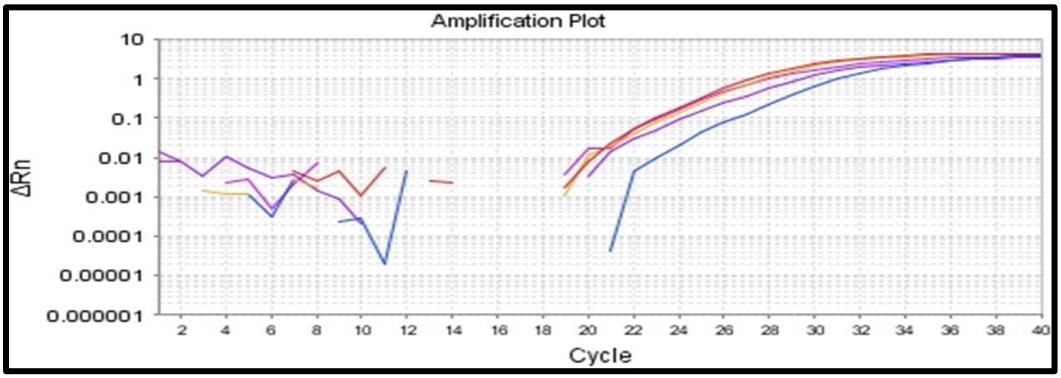

Figure 7.3: Amplification plot (Log graph of ΔRN vs Cycle), of KRAS exon 2, of KRAS exon 2. KRAS exon 2 was amplified from 5ng of original cfDNA templates of Plasma samples collected from CRC patient (pre- and post-surgery). The results show that adequate amplifications were achieved for the samples at ct values of 21.92-25.45.

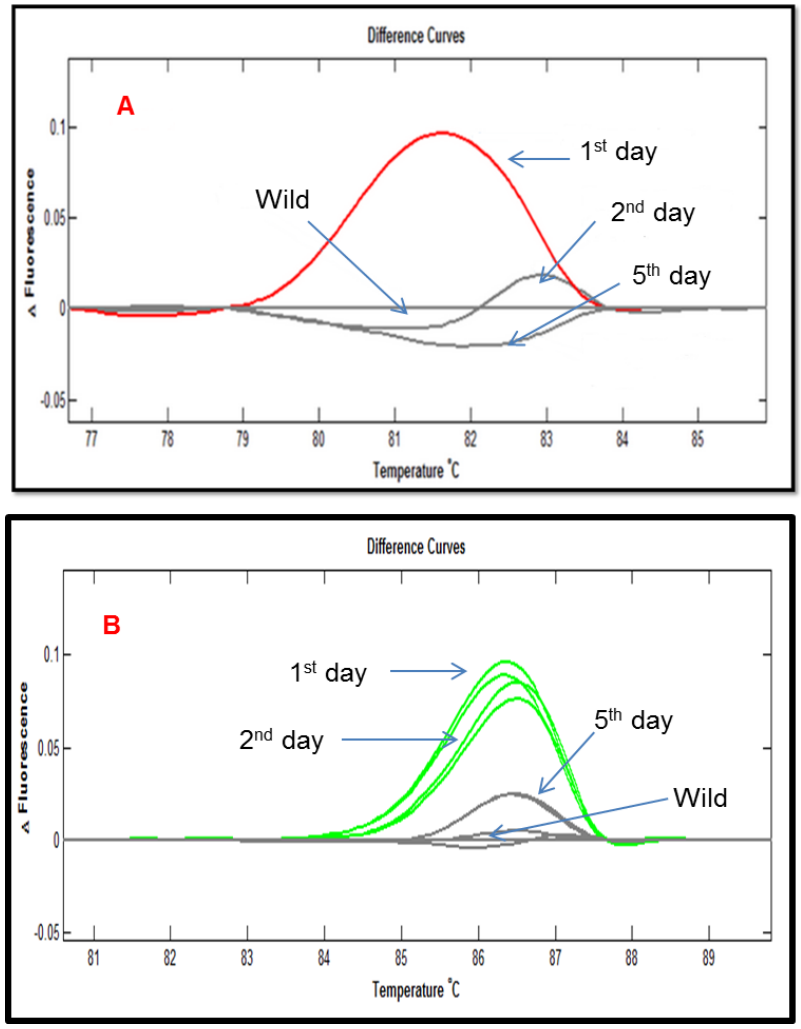

Figure 7.4: HRM analysis obtained from plasma samples. (A) In patient no.1 KRAS/exon2 mutation was detected, which appeared in the (1st day), and disappeared in the the days after operation, whereas (B) shows TP53/exon6 mutation, which was observed in the (1st and 2nd days) and was disappeared after that.

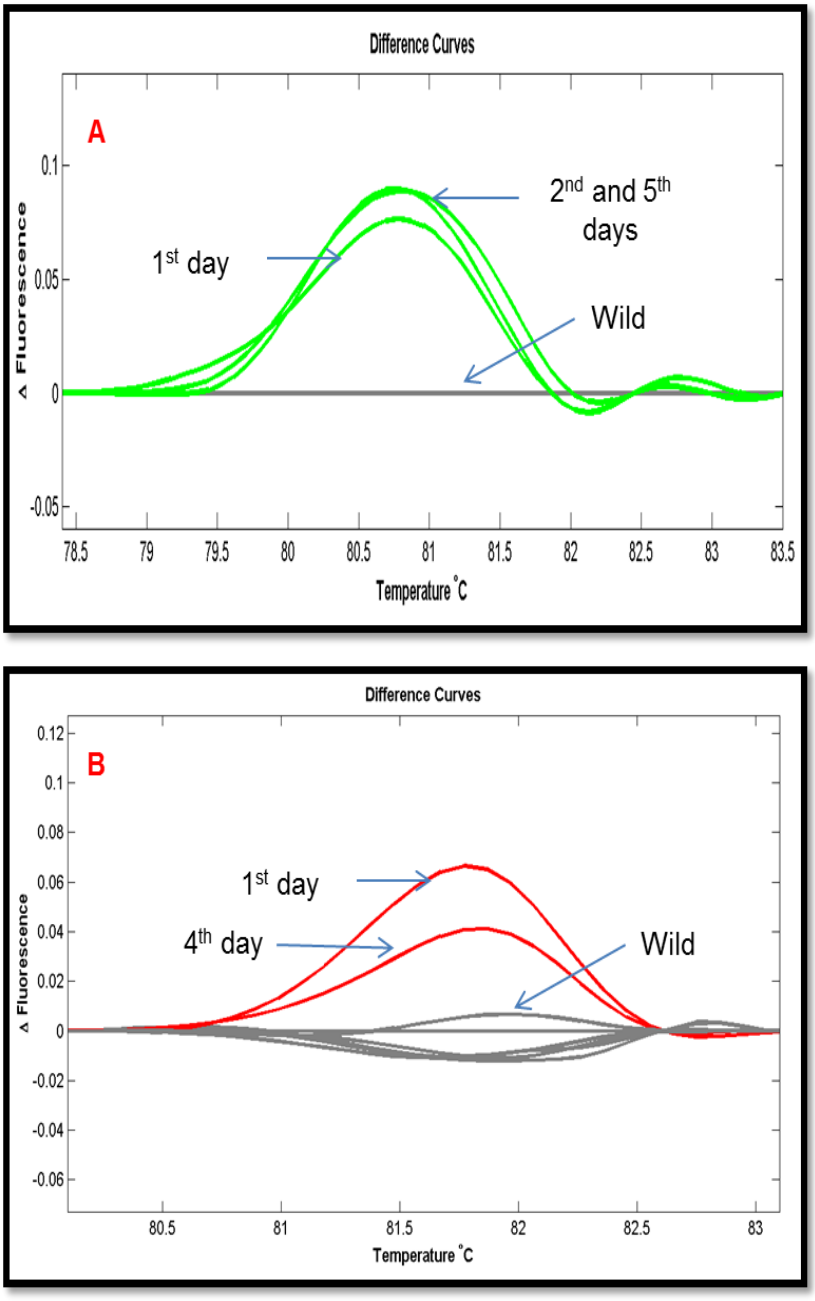

Figure 7.5: HRM analysis obtained from plasma samples PIK3CA/exon9 mutation was detected in plasma samples collected in the 1st day and days after operation from (A) patient no.2 and (B) patient no.3.

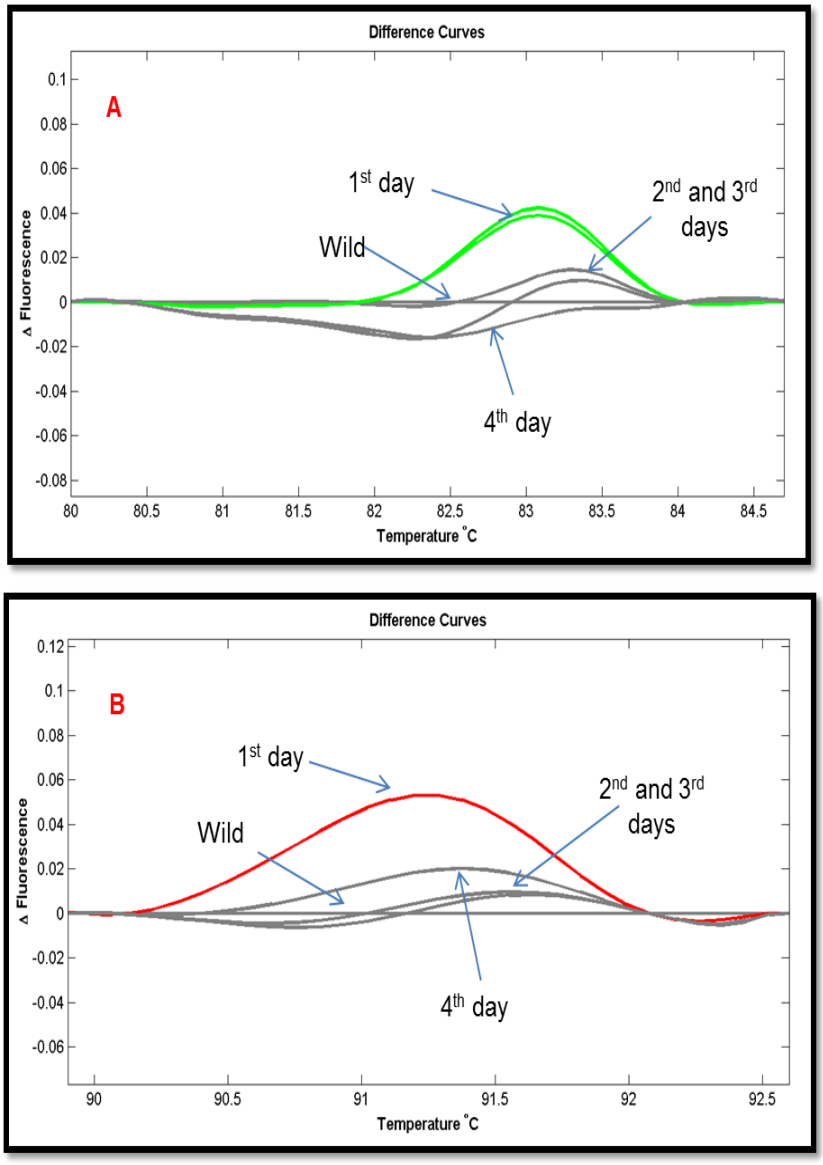

Figure 7.6: HRM analysis obtained from plasma samples. In patient no.4 (A) BRAF/exon15 and (B) TP53/exon5 mutations were detected in the (1st day) plasma sample and disappeared from the samples collected the days after operation.

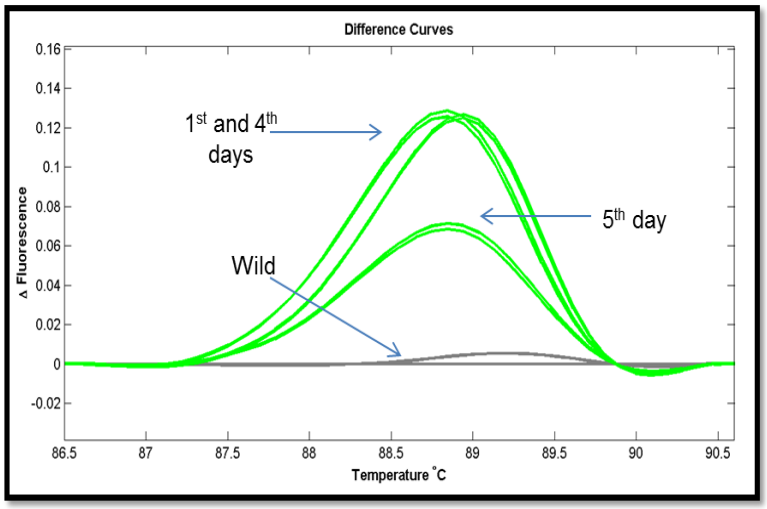

Figure 7.7: HRM analysis obtained from plasma samples Patient no. 5 showed TP53/Exon8 mutation in the plasma samples collected before and after operation.

7.4.2 MiRNA quantification

7.4.2.1 Measurement the quantity and quality of miRNA

Total RNA was extracted from 19 plasma samples (3 from healthy donors and 16 from patients with CRC). The median RNA yield from the purification of plasma was 59.1ng and all samples had a 260/280-ratio ranged 1.81 – 2.1. QPCR assay was developed to quantify miRNAs in CRC patients and normal healthy plasma samples. In this qPCR, relative quantification was applied and thus spiked in miR-39 was used as an appropriate internal normalisation control to normalise sample-to-sample variations.

7.4.2.2 Stability of endogenous controls

To evaluate miRNAs expression in plasma samples, the study has chosen synthetic RNA (miR-39) as an exogenous gene for normalization based on data reported previously and highlighted that adding the same amount of spiked-in RNAs with an equal volume of serum provides a stable reference control. As mentioned in (section 2.2.2.2.1), we spiked in miR-39 in an equal volume into the plasma of all samples and later on total RNA was extracted.

In order to evaluate the stability of the miR-39 to see is it suitable to be used as an internal normalisation control, which required to normalise sample-to-sample variations. Our data demonstrated that miR-39 was readily detectable in the plasma samples taken from three normal healthy donors and 5 CRC patients. No significant difference was observed in terms of Ct values of miR-39 (p=0.68) between plasma samples taken from healthy donors and patients with CRC.

7.4.2.3 Compare miRNAs expression between pre- and post-surgical plasma samples

In order to investigate the level of miRNAs expression among plasma samples collected before and after surgical operation from CRC patients and three healthy normal control, which were pooled an equal volume to generate a reference healthy plasma sample, normalised to spiked in miR-39, a panel of six miRNAs (miR- 20a, 21, 29a, 31, 92a and 224) was analysed using RT-qPCR method. All assays were done in triplicate. Samples having CT values greater than 35 or CT replicate standard deviations greater than 0.5 were excluded from analysis. Although some variations were noticed between pre- and postsurgical plasma miRNA levels, paired t-test analysis revealed no significant differences between these time points ($p > 0.5$). However, the level of miRNAs is changed by more than 2-fold (upregulated) on the day of surgery as follow, miR-20a in 1/5 (20%), miR-21 in 4/5 (80%), miR-29a in 3/5 (60%), miR-92a in 4/5 (80%) and miR-224 in 2/5 (40%) see (table 7.3).

Table 7.3: The fold of change of miRNAs expression levels in list of upregulated miRNAs (change .2-fold as a cut-off level) in plasma of CRC patients compared to healthy controls.

Patient No.	Sample	miR-20a	miR-21	miR-29a	miR-31	miR-92a	miR-224
1	Day 1	1.05	2.88	0.86	0.14	2.22	0.55
	Day 2	0.28	0.95	0.33	0.04	0.17	0.22
	Day 5	0.96	1.61	0.51	0.10	2.04	0.31
	Day 6	1.24	3.86	0.81	0.21	2.19	0.50
2	Day 1	2.66	4.88	3.20	7.33	7.62	6.62
	Day 2	0.50	2.49	5.33	5.34	8.80	6.90
	Day 5	0.28	3.48	4.37	4.88	10.51	7.89
3	Day 1	0.91	3.45	5.10	13.43	13.96	16.63
	Day 4	4.10	2.55	4.20	15.65	15.90	14.01
4	Day 1	0.81	0.74	2.05	1.85	2.01	1.63
	Day 2	0.66	0.83	0.64	1.38	2.63	1.23
	Day 3	0.98	2.61	0.01	0.99	1.91	0.65
	Day 4	0.48	0.97	0.01	1.55	1.69	1.24
5	Day 1	1.64	2.06	0.01	1.71	0.73	0.82
	Day 4	0.72	0.62	0.43	0.88	0.71	0.90
	Day 5	0.01	0.00	0.14	1.67	0.82	0.79

7.5 Discussion

Due to the simplicity and non-invasiveness of sample collecting, the use of blood samples for detection and monitoring of various disease conditions such as cancer diagnosis or surveillance is very attractive. Different studies in the past have highlighted the importance of using the DNA extracted from blood plasma or serum in investigating mutations and methylation in multiple genes. The detection of such mutations in plasma or serum will not only provide solid evidence that nucleic acids are released into the circulation by tumours, but also could be used as promising non-invasive biomarkers for diagnosis, prognosis, selection and monitoring of treatment effect in cancer, which may lead to increasing clinical impact in the near future.

The present study investigated variation in the levels of 6 miRNAs and screening of mutations among 5 genes within a specified time period (pre- and post-surgery time points) as a potential result of surgical removal of a tumour. To our knowledge, it is the first investigation for genotyping cfDNA in patient with CRC at more than two time points after surgery and very few studies have investigated circulating miRNA in patient with CRC at more than two time points after surgery [568]. In this study, 16 blood samples were collected from 5 CRC patients. Pre-surgery blood was drawn at the day of surgical operation and post-surgery blood draws were implemented within a week after removal of the tumour (range: 2–6 days).

Based on previous reports [569, 570] Qiamp mini kit and miRNeasy Serum/Plasma kits were chosen for cfDNA and miRNA purification from plasma samples. Qiamp mini kit was assessed and the results demonstrated that it is more efficient and was able to purify as few as 25pg/µl of DNA from plasma samples. Additionally, miRNeasy Serum/Plasma kits was able to purify enough good amount of RNA up to 59ng/µl with the acceptable 260/280-ratio ranged 1.81 – 2.1.

Generally using molecular biomarkers and particularly DNA somatic mutations has a great advantage, because these mutations are most probably coming from tumour cells[571]. In the present study, although the cfDNA was screened for several mutations without analysis of the tumour tissue, testing many genes would validate the results. Disappearance of mutations were not the same among the genes and patients. For example, in patients no.1 mutations disappeared at slightly different times (KRAS mutation gone the day after surgery where as TP53 mutation disappeared at the 4th post-surgery day) so there were some differences in clearance. In case no.4 all mutations in all genes disappeared at the 1st day post-surgery thus this suggested that tumour DNA had really gone and no more tumour cells are present to release their DNA content into the circulation. In patients no.2, no.3 and no.5 mutations persisted in all the genes, may also be because according to the clinic-pathological features (table 2.4, chapter 2) in those 3 cases there is another tumour as they were at pT3 and pT4 stages of the diseases'. However, based on the finding that HRM can be quantitative (section 3.4.1.6), the difference in the melting pattern of mutations between samples collected pre- and post-surgery, indicate a slight decrease in the amount of the mutant allele which could be reflecting the amount of tumour cells.

miRNA expression was also evaluated at different time points (pre- and post-surgery resection). Although the results showed their expression fluctuated, paired t-test analysis revealed no significant differences between these time points ($p > 0.5$). This result is in contrast to studies, which collected data from 35 CRC patients, that recently reported significant decrease of circulating miRNAs including miR-20a and miR-92a after comparing pre- and post-surgical miRNAs level [568]. It is undeniable to say that inflammation is a common post-operative complication and due to inflammation there may be increase expression of oncogenic miRNAs [572, 573], that may be one of the reason behind upregulation of the miRNAs level even at postoperative period. However, the sample size of this

methodologic study was limited which may also be a possible explanation for the lack of statistically significant difference between these time points.

In conclusion, the study was able to extract an acceptable amount of cfNAs (cfDNA and cfmiRNA), which are good enough for screening and quantification of cfNAs. Additionally, screening of mutations in cfDNA may be suitable for monitoring surgical clearness.

In the future, further analysis is needed on a larger number of samples not only to investigate surgical clearness, but also for monitoring treatment response / surveillance for metastases and recurrence.

Chapter Eight: General discussion

8.1 Discussion

Colorectal cancers encompass a heterogeneous group of diseases, and each CRC patient has a unique disease that has been caused by distinctive genetic/epigenetic background. Investigating the molecular biology of colon cancer is important to provide an insight into the role of alterations of genetic, epigenetic and miRNAs in the progression of a polypoid lesion into a carcinoma at the molecular level and consequently finding similar features among individual tumours, to foresee the pathogenesis and biological behaviour of a tumour. In the present study, we investigated the molecular biology of colorectal cancer, through screening genes for mutations, measuring aberrantly expressed mRNA, miRNAs, and evaluation of protein expression. Thus, providing an insight into the role of these alterations in the progression of a polypoid lesion into a carcinoma at the molecular level and consequently finding similar features among individual tumours, to foresee the pathogenesis and biological behaviour of a tumour. We did work to identify the difference in mutation profile between CIN-CRC and MACS-CRC and investigated genetic discrepancies between primary colorectal cancer and their respective metastases. As sub study, we have also worked on the biopsy and plasma samples to see whether are they suitable to be used as an alternative for molecular investigation in colorectal cancer. Furthermore, the study analysed robustness of NGS technologies and developed the Cold-HRM method to increase limit of detection and applied to discriminate germline from somatic alterations without need of matched normal samples.

According to the data reported by COSMIC and TCGA database, several genes including APC, TP53, KRAS, BRAF, PIK3CA, PTEN, SMAD4 and FBXW7, which are involved in different pathways (Wnt, PI3K, RAS-MAPK, TGF-β, p53 and DNA repair pathway), were frequently mutated in CRC. These pathways that are involved and cooperate to maintain CRC malignant phenotype interact, and regulate expression of each other. The deeper understanding of the mechanisms of CRC genetic alteration and related

consequences are important of drug development for the predication of efficacy of treatment. In the present study Investigating mutation profile of CRC cases is consistent with data published. For example, we found that TP53, APC, KARS, PIK3CA, SMAD4, FBXW7 and BRAF are mutated frequently and the frequencies of mutations detected in the sample set are fall within the ranges published in the literature [25, 49, 393-395]. Additionally, presence of concomitant mutations in these genes indicated alteration of pathways which are commonly altered in CRC. For example, (33.7%) of tumours contained the full house of APC/ (KRAS or BRAF or NRAS)/TP53 mutations of the Fearon and Vogelstein pathway, suggesting that other genes may be involved in the tumorigenic process. Furthermore, the KRAS, Wnt, PI3K and p53 signalling pathways are commonly altered in CRC and we speculated that some sort of crosstalk interactions would exist between them. Approximately 93% CRCs have mutation in at least one of the above-mentioned genes. Furthermore, the frequency of *APC* mutations (47/83, 57%) was lower than that of *TP53* mutations (62/83, 75%) and this is consistent with data published by Smith *et al.* [7] and Gay *et al*. [396]. In addition to the most commonly frequently mutated genes, we found mutations in ED-POLE and ED-POLD1 genes, which are involved in DNA repair pathway. Recently several studies identified the occurrence of somatic and germline mutations in exonuclease domains region in the *POLE* and *POLD1* genes [574], which enhanced mutation rate during DNA replication and increase risk of CRCs through a defect in the normal proofreading activity in both genes. Mutations result in amino acid substitutions in the POLE/POLD1 exonuclease domain dramatically reduced 3'-5' exonuclease activity comparative to the wild-type POLE [575, 576]. The result in the present study detected that most of the *POLE*-EDM and *POLD1*-EDM (9/10) mutations were found in CRC with MSS. This result is consistent with that hypothesis believed that most EDM-POLE cancers are MSS [215, 227] and this result suggesting that defective proofreading provides an alternative mechanism to achieve genomic instability and

tumourigenesis. Although, data in both database highlighted that POLE (and POLD1) proofreading mutations associated with ultra-mutated cancer, which is different from hypermutated MSI and non-hypermutated CIN, and rarely seen in the non-hypermutated, 7/10 mutations found in the present study were in the CIN and non-hypermutated cases. This may show a distinct CRC group has a new pathogenic pathway. However, more investigation needs to validate this finding.

After investigated mutation profiles of the 83 (first cohort) CRC cases, we believed that the deeper understanding of the mechanisms of CRC genetic alterations and related consequences are important, because of their value in detecting different pathways of carcinogenesis, and their direct relevance to prognosis. The present study tried to identify differences between CIN- and MACS-CRC groups as mixed information was reported regarding MACS-CIN as some defined as a distinct group [230, 577, 578]. Others described MACS as an early stage of CIN and suggests that factors contributing to aneuploidy most likely occur late during in the carcinogenic process. In the present study screening of mutations in the 28 genes was used to compare (46 CRC-CIN) and (35 CRC-MACS) groups included in the first cohort in term of mutation profile, clonal heterogeneity, and allelic imbalance. The results highlighted that there is a large degree of overlap in the mutation profiles, the presence of heterogeneity and the frequency of allelic imbalance indicating a common genetic pathway for these tumour types in the early stages. Furthermore, comparison of clinic-pathological features, has revealed no significant differences between these two groups of samples (CIN- and MACS-CRCs), which is in line with previous study conducted by Fadhil et al, [233]. Our data was consistent with this hypothesis, as we did not find any differences between these two groups. Other point which support this finding is, no differences was seen in term of clinicopathological features between these two groups as well.

In clinical practice, due to the difficulties to obtain tissues of metastatic tumours most results of gene tests were carried out mainly on primary tumours. However, the concordance of gene mutation status is controversial between the primary tumours and their metastases because some results raise concern that genetic profiling of primary tumours not representative of metastatic disease [579, 580]. After detecting mutations in several genes involved in different pathways in CRC, the study sought to determine whether analysis of the primary tumour is sufficient or whether a metastatic lesion should be studied in patients with metastases. In this study, in addition to screening mutations of previously mentioned 28 genes which are frequently and infrequently mutated in 22 paired (first cohort) primary CRC and their corresponding metastasis, we compared primary CRC and metastasis through measuring expression levels of miRNAs and evaluate protein expression of other genes. The reason behind measuring of protein is, mutation is not the only cause of change in the expression of proteins. The most important finding was that a high degree of concordance existed between the primary tumour and its metastasis as mutation, miRNA and protein expression profiles of the metastasis closely resembles that of their corresponding primary CRCs. This result is in line with other studies and reported a high degree of concordance between primary tumors and metastases [581, 582]. Additionally, Brannon et al [434] recently performed targeted next generation sequencing analysis of 230 cancer-associated genes on primary, metastatic, and normal tissue from 69 colorectal cancer patients. KRAS, NRAS, and BRAF mutations were identical in both the primary and metastatic tumours. This mutation profile is concordance with the mutation frequencies for non-hypermutated samples reported by The Cancer Genome Atlas (TCGA) and in line with findings of the current study.[583]. However, in the same study, 18 private mutations, defined as mutations called only in the primary or the metastatic tumour, were found in APC, PIK3CA, SMAD4, and TP53. Following this the authors sequenced 97 additional samples from spatially separate regions of the primary tumors (n

= 62) and metastases (n = 37) from 22 patients, encompassing 46 discordant mutations. Sequencing of multiple regions and samples resolved 17/46 discordant mutations, including 12/22 (55%) mutations that were originally detected only in the metastasis but were subsequently found to be subclonal in the primary tumor and accordingly, the authors concluded that the small proportion of discordant mutations we observed may be an overestimate of the actual discordance when intra-tumour heterogeneity is considered as an effective confounding factor. Interestingly, in agreement with the current study, they found that PIK3CA had private events specific to the primary tumour suggesting that despite the excellent genomic concordance between primary and metastatic CRC, heterogeneity in potentially targetable genes, such as members of the PI3K pathway, is existing in at least a minor subset of patients. This could have a therapeutic implication particularly with ongoing clinical trials on PI3K inhibitors [584]. A limitation for the current study is we did not test different areas of the tumours in both the primary and metastases, which could potentially be the culprit behind the few cases that showed discordance between primary and metastatic tumours. Our data identified that despite molecular profiles of primary tumours and metastases are not similar, but statistically were not significant and we can conclude that primary tumour for molecular profiling may be suitable for clinical decision making in metastatic CRC.

After investigating genetic alteration among genes involved in many signalling pathways, which have roles in the pathogenic mechanisms of CRC as mentioned above, the study also investigated differential expression of miRNAs and some of their target genes that regulate these pathways and having a key role in cancer development, progression, and resistance to chemotherapy [585-587]. Despite we did not do any functional invitro study, based on our finding we could say that upregulation of miRNAs (miR-20a, miR-21, miR-29a and miR-31) may have a role in CRC carcinogenesis by affecting different target genes which they have role in different signalling pathways. For example, upregulation of miR-20a is significantly deregulate

PTEN (has a role in PI3K pathway) protein expression level. Upregulation of miR-21 with decreased expression of both mRNA and protein expressions of TGFBRII (has a role in TGFB pathway). This is same as the study conducted by Lizarbe et al, [588] who showed that overexpression of miR-21 is able to downregulate the expression of TGFBR2 gene. High expression of miR-29a and miR-31 inversely correlated with low RASA1-mRNA level (has a role in RAS-MAPK pathway). This results also is consistent with the study conducted by Xu et al, [589] Chen et al, [590]. Additionally, the expression level of miR-92a and miR-224, which are detected in this study as not significantly changed in CRC samples, are different from that recently reported [283, 371, 444, 483, 502, 564, 591, 592]. This discrepancy may be due to limit number of samples, methods of analysis, low quality DNA template derived from FFPE tissue. This part of the study was performed on the second cohort of FFPE samples and in the future, we can investigate genetic alteration in this cohort and find a correlation between miRNAs and mutations as well.

Due to Recent studies have reported the correlation between cfDNA mutational status and changes in tumour burden before and after surgery [593-595]. This study aimed to investigate the use of cfNAs as biomarker to monitor the change in mutation frequency in pre-and post-surgical samples from CRC patients. Utilizing cfNAs as a noninvasive methodology has more advantages over the standard methods of genetic analysis via serial biopsies of tumour and surgical resections in cancer patients. Our study was successful in optimizing a method for extracting enough cfDNA and miRNA to be used for further analysis. After investigating cfDNA and cfmiRNA, the results also highlighted that cfNAs are suitable to be used for mutation screening in more than one genes and measuring miRNAs as well. Serial sampling of cfDNA pre-and post-operatively in CRC patient has been reported recently to monitor the tumour changes and observe the correlation with tumour progression and clearance after surgery [593, 596]. This is same as what the present study obtained from 2/5 patients, as mutations

which were detected in plasma samples from CRC patients before surgery were completely absent were totally disappear after surgical operation. However, the same results did not get from the plasma samples obtained from other three CRC patients. Additionally, we did not see any significant difference in the amount of expression of miRNAs in pre- and post-surgery. Although the study couldn't find any differences between pre- and post-surgical genomic alteration and miRNAs level. No conclusion can be drawn from our data due to the low sample size.

Beside investigating genetic alterations in the CRC samples, the study sought to investigate whether biopsy and circulating-cell free nucleic acid (in plasma) are suitable to be used as an alternative of resection samples for investigating genetic alteration in CRCs. Nowadays, tumour material from the resection specimen might be used for predictive testing of adjuvant therapy whereas the use of neoadjuvant therapy in patients with CRC is likely to increase. Currently, patients with rectal tumours have taken neoadjuvant therapy whilst a clinical trial of neoadjuvant chemotherapy for locally advanced colonic cancer is recently started in the UK. If the use of neoadjuvant therapy shows desirable outcomes, then the diagnostic biopsy specimens may become the only material available for predictive testing to confirm the presence of invasive malignancy [597]. Due to CRC develops as a consequence of waves of clonal expansion, which resulting from mutations called 'driver mutations' that give a selective advantage [598]. Driver mutation is responsible for a new clone that expands and becomes predominant in the tumour and ultimately be present throughout the tumour. A hypothesis believed that the driver mutations, which are responsible for early clonal sweeps through the adenoma–carcinoma sequence, predominantly present in most of the tumour cells and consequently would be present in the biopsy samples. In the current study, sensitivity, and feasibility of using diagnostic biopsy specimens was investigated and the results detected that for certain mutations biopsy specimens are suitable for predictive testing. This is same as the study conducted by Fadhil et al, [597]

showed that diagnostic biopsy specimens, even though they are a tiny sample of the tumour, are sufficiently representative for use in predictive testing for early driver mutations in colorectal cancer.

Although, in the past high resolution melting analysis was applied to detect LOH and to discriminate SNPs from somatic mutation, this method often required non-tumour DNA from the same patient [104, 599-602]. As sub studies, we developed a method not only applied for mutation screening for several genes, but also could discriminate germline from somatic alteration by combining of standard and COLD PCR followed by HRMA without the requirement for matched normal DNA. This method can be applied for testing DNA from FFPE tissue with high degree of sensitivity and specificity with reasonable cost. Additionally, it reduces the number of the samples need to be validated by other expensive methods such as Sanger sequencing which is characterised by its relative low sensitivity. Another advantage point in this method is, it improves the detection sensitivity of HRM screening, allowing for the detection of mutant mixtures at frequencies of as little as 1.0%. Recently published reports of COLD-PCR with HRM have shown relatively similar levels of mutation enrichment and improved levels of mutation detection sensitivity [348, 603, 604]. False positive is the main concern about this method because of damaging DNA by limitation of this method is formalin fixation and storage for long time (Quach et al., 2004). Additionally, despite the study able to delineate germline from somatic, but it cannot tell the type of mutations.

Although HRM is simple, cheap, rapid and inexpensive technique, several factors have to be taken into consideration to get high accurate HRM results. For example, quality of DNA template, PCR specificity, length of the amplicon (below 400bp) [355], type of mutations (small insertions and deletions may be difficult to detect than substitutions) [356], GC content [353, 357], dye (for example, LCG reens Plus which is better than SYTOs 9, which is better than EvaGreens, which is better than SYBR Green I) [358], melting analysis software and resolution of the instrument because the

ability of the instruments is not similar to resolve shape and absolute temperature differences [355, 358, 359]. For these reasons and in order to get a high accurate HRM results, the present study used; pure DNA samples (the purity and quality of the DNA samples were assessed as mentioned in chapter 2), primers were designed to amplify a short amplicon (almost less than 200bp), LC Green plus dye, and the specificity of the PCR product were checked initially by blasting the primers against NCBI human genomic database, amplicon size was checked with gel electrophoresis

8.3 Conclusion

This study showed that (1) in addition to find that mutation profile of CRC in this study is almost similar with that reported before in several studies and in COSMIC and TCGA databases, it was also found a number of miRNAs could play vital role in CRC progression by directly targeting components of several signalling pathways in CRC development. Thus, they may serve as a diagnostic marker and therapeutic target for CRC because their expressions were upregulated in tumour when compared with normal tissues. (2) there is a large degree of overlap in the mutation profiles between CIN-CRC and MACS-CRC, the presence of heterogeneity and the frequency of allelic imbalance indicating a common genetic pathway for these tumour types in the early stages. Our data support the theory that aneuploidy is a late occurring feature in cancer progression although the cause of the aneuploidy has not been identified. (3) the study demonstrated the existence of a relative similarity between primary tumours and their corresponding metastases in patients with CRC. These reminded that gene test can only be conducted in primary tissues, and the metastatic specimen not needs to be re-examined. (4) the study also found that it is acceptable to use biopsy material for predictive testing. Additionally, circulating cell-free nucleic acid may be suitable for molecular diagnosis and for monitoring surgical clearness. (6) Finally, we have proposed a refined interpretation of

COLD-HRM which allows inference of LOH and somatic sequence mutation based on presence or absence of melting pattern following COLD-PCR.

Limitation of the study

1. No power calculations have been undertaken to ensure adequate powering.
2. Low statistical power (because of low sample size of studies, small effects, or both) negatively affects the likelihood that a nominally statistically significant finding reflects a true effect.
3. The use of relatively old FFPE samples might have adverse effects on the outcome of all assays. The results might be of compromised quality when compared to newer tissues.
4. Tumour samples were not collected from multiple sites of metastasis which may cause the level of heterogeneity became less significant.
5. Could not find any correlation between mutations and gene expressions (miRNAs, mRNAs and protein) as the investigation did on different cohorts of CRC samples.
6. For IHC, no staining was done on whole sections to confirm homogeneity of staining.

8.4 Suggestions and future direction

1- Screening larger number of genes of bigger numbers of samples to not only identify mutation profiles, but also to investigate different genetic pathways interrogated in CRC, which may therefore help elucidate the nature of genetic defects that confer susceptibility to development of these subclasses of colorectal cancer.

2- Functional studies will be required to understand how POLE and POLD1 mutations contribute to carcinogenesis and their clinical implications, and in the advanced CRC to investigate how the

hypermutated phenotype related with *these* mutations may be used therapeutically.

3- Further studies will be required to investigate the heterogeneity of genomic profiles between the primary tumour and metastases, by interrogating a bigger number of genes, larger sets of samples and samples from multiple sites of metastasis.

4- The miRNAs biomarkers were analysed in this study need to be validated in a large number of samples using alternative approaches and their roles in the mechanism of CRC carcinogenesis should be explored by means of functional experiments.

5- Investigate cell free nucleic acids (cfDNA and miRNA) in a larger cohort of blood samples from CRC patients (pre- and postsurgical (2–7 days after surgery) and 6 months' follow-up), as a means of measuring level of surgical clearance. As well as using genomic alterations as methods of screening/ monitoring treatment response / surveillance for metastases and recurrence. to investigate tumour heterogeneity.

Chapter Nine: Appendix

Appendix A: Supplementary information of chapter two

Table 1: List of genes, primers and amplicon size. (outer is primers were used for initial prediagnostic multiplex, Inner is primers were used for single specific diagnostic)

Gene	Primer Sequence	Designed by:	Amplicon size (bp)
KRAS Exon2	Forward 5`-AGGCCTGCTGAAAATGACTG-`3	Hersh	174
	Reverse 5`-ATCAAAGAATGGTCCTGCAC-`3		
KRAS Exon2	Forward 5`-GCCTGCTGAAAATGACTGAA-`3	Hersh	162
	Reverse 5`-GGTCCTGCACCAGTAATATGC-`3		
SMAD4 Exon9	Forward 5'-GCTCCTGAGTATTGGTGTTCC-'3	Hersh	183
	Reverse 5'- TGCTCTCTCAATGGCTTCTG-'3		
SMAD4 Exon9	Forward 5'- TCCTTCAAGCTGCCCTATTG-'3	Hersh	108
	Reverse 5'- TCAATGGCTTCTGTCCTGTG-'3		
SMAD4 Exon11	Forward 5'- TGCTGGAATTGGTGTTGATG-'3	Hersh	191
	Reverse 5'- AAGGTTGTGGGTCTGCAATC'-3		
SMAD4 Exon11	Forward 5'- GACCGGATTACCCAAGACAG-'3	Hersh	109
	Reverse 5'- GCAATCGGCATGGTATGAAG-'3		
NRAS Exon2	Forward 5'- GGTTTCCAACAGGTTCTTGC-'3	Hersh	191
	Reverse 5'- TCCGACAAGTGAGAGACAGG-'3		
NRAS Exon2	Forward 5'- TACAAACTGGTGGTGGTTGG-'3	Hersh	115
	Reverse 5'- CACTGGGCCTCACCTCTATG-'3		
TP53 Exon2	Forward 5'- ATCCCCACTTTTCCTCTTGC-'3	Hersh	198
	Reverse 5'- TCCCACAGGTCTCTGCTAGG-'3		
TP53 Exon2	Forward 5'- AGACTGCCTTCCGGGTCAC-'3	Hersh	115
	Reverse 5'- TCCAATGGATCCACTCACAG-'3		
TP53 Exon4	Forward 5'- AGATGAAGCTCCCAGAATGC-'3	Hersh	202
	Reverse 5'- ACTGACCGTGCAAGTCACAG-'3		
TP53 Exon4	Forward 5'- CTGCACCAGCAGCTCCTAC-'3	Hersh	125
	Reverse 5'- AGAATGCAAGAAGCCCAGAC-'3		
TP53 Exon10	Forward 5`- TGTTGCTTTTGATCCGTCAT-`3	Hersh	257
	Reverse 5`- TAGGAAGGCAGGGGAGTAGG-`3		
TP53 Exon10	Forward 5`- TTCTCCCCCTCCTCTGTTG -`3	Hersh	169
	Reverse 5`- AGGGGAGTAGGGCCAGTAAG -`3		
GNAS Exon8	Forward 5`-TTGGCTTTGGTGAGATCCAT-`3	Hersh	177
	Reverse 5`- GGACTGGGGTGAATGTCAAG-`3		
GNAS Exon8	Forward 5`- TGTTTCAGGACCTGCTTCG-`3	Hersh	101
	Reverse 5`- AAGGTAACAGTTGGCTTACTGGA`-3		
PTEN Exon3	Forward 5`- TCATTTTTGTTAATGGTGGCTTT-`3	Hersh	182
	Reverse 5`- AACTCTACCTCACTCTAACAAGCAGA-`3		

Gene	Primer Sequence	Designed by:	Amplicon
PTEN Exon3	Forward 5`- GGCTTTTTGTTTGTTTGTTTTG-`3	Hersh	158
	Reverse 5`- CCTCACTCTAACAAGCAGATAACTTTC-`3		
APC Cdc 0	Forward 5'- TTCCAACCACATTTTGGACA-'3	Hersh	200
	Reverse 5'- TCTTCATGCTGTTCTTCTTCAGAG -'3		
APC Cdc 0	Forward 5'-CAGCAGGAATGTGTTTCTCC-'3	Hersh	112
	Reverse 5'-CTTGACACAAAGACTGGCTTACA-'3		
POLE Exon9	Forward 5'ATGGGGAGTTTAGAGCTTGG'3	Hersh	192
	Reverse 5'CCATCCCAGGAGCTTACTTC'3		
POLE Exon9	Forward 5'TGCTTATTTTGTCCCCACAG'3	Hersh	153
	Reverse 5'TACTTCCCAGAAGCCACCTG'3		
POLE Exon10	Forward 5'GCAGCCTCTGACTTGTGC'3	Hersh	200
	Reverse 5'CCGTTCTTCCCACAATACC'3		
POLE Exon10	Forward 5'CCTCTGACTTGTGCTGATTGC'3	Hersh	179
	Reverse 5'CCGGGTAGTTTCCCAAGTG'3		
POLE Exon11	Forward 5'GCAGACCTCTGACTGCTGTG'3	Hersh	167
	Reverse 5'CCCTCCCTCTCAAATGCTG'3		
POLE Exon11	Forward 5'CTGCTGTGACTTGGGTTCAG'3	Hersh	133
	Reverse 5'CAAATGCTGCCCAGTTACTC'3		
POLE Exon12	Forward 5'GGCATTAGAGCCTGACCTG'3	Hersh	214
	Reverse 5'CACGGACAGCAGTGAGGAG'3		
POLE Exon12	Forward 5'TGTCTGTCCTCTTTCCAACC'3	Hersh	153
	Reverse 5'AGCCATGCTGCTCTGTGG'3		
POLE Exon13	Forward 5'CGGGCTGCATGTTAGAATC'3	Hersh	205
	Reverse 5'AGCCGGGATGTGGCTTAC'3		
POLE Exon13	Forward 5'TGGCTTCTGTTCTCATTCTCC'3	Hersh	179
	Reverse 5'CGGGATGTGGCTTACGTG'3		
POLE Exon14	Forward 5'TTTGATGGCCCTGCTCTC'3	Hersh	236
	Reverse 5'TCCATTCAGCTCCAGTGC'3		
POLE Exon14	Forward 5'CTCTGGCGTTCTCTCCTCAG'3	Hersh	158
	Reverse 5'GCCGACAGGACAGATAATGC'3		
POLD1 Exon8A	Forward 5`-CTGGGAAATACGCCCTGAG'3	Hersh	220
	Reverse 5' CGATATCGAAGCTGAGCAC'3		
POLD1 Exon8A	Forward 5'TGAGGCTGAAGGAGAAGGTG'3	Hersh	157
	Reverse 5'CTTCCGGTGGGTGACTGAC'3		
POLD1 Exon8B	Forward 5' ACGTGCTGTGGTCTGACG'3	Hersh	201
	Reverse 5' CCTTCATGCTTGCCATTCC'3		
POLD1 Exon8B	Forward 5'GCTGTGGTCTGACGTGGTC'3	Hersh	190
	Reverse 5'GCTTGCCATTCCCTCCAC'3		
POLD1 Exon9	Forward 5' AGCCTCCCTGCTGTGTTG'3	Hersh	304
	Reverse 5' GGAGCTGATGGCTCAGGAC'3		
POLD1 Exon9	Forward 5'GTGAGGGGCAGGAGTCAG'3	Hersh	228
	Reverse 5'GCGTGGAGCGAGAGCTAC'3		

Gene	Primer Sequence	Designed by:	Amplicon
POLD1 Exon10	Forward 5' CACCAGGGTGACCCAATG'3	Hersh	255
	Reverse 5' GAGGCACACAGAGCAAGATG'3		
POLD1 Exon10	Forward 5'AGGGTGACCCAATGTGCTC'3	Hersh	165
	Reverse 5'GAGAGAAGCCTCCCACCTG'3		
POLD1 Exon11	Forward 5' TGTGGTTGGTCTCAATCTCC'3	Hersh	263
	Reverse 5' CCTTCCAAGGACAGGGACAC'3		
POLD1 Exon11	Forward 5'CACACCCTGCCTCTCCTC'3	Hersh	200
	Reverse 5'CCAAGGACAGGGACACACC'3		
POLD1 Exon12	Forward 5' GGGTGTGTCCCTGTCCTTG'3	Hersh	293
	Reverse 5' CAGCAGGTCAAAGAGTGAGG'3		
POLD1 Exon12	Forward 5'GTCCTTGGAAGGCCACTG'3	Hersh	
	Reverse 5' GTCAGAGGTTGGGGTGAGAG'3		
POLD1 Exon13A	Forward 5' AGGCTACCTCACCCTGACC'3	Hersh	247
	Reverse 5' GCCACGACTGAGCAGGTAG'3		
POLD1 Exon13A	Forward 5'CCTTCTCCTGCTCCACCTC'3	Hersh	189
	Reverse 5'CAGTGACCCTCGCCATCTC'3		
POLD1 Exon13B	Forward 5'ATGCCTACCTGCCACTGC'3	Hersh	215
	Reverse 5'GGATGCCATTGCCTGGAG'3		
POLD1 Exon13B	Forward 5'TCATGGTGCTGGTGAACG'3	Hersh	130
	Reverse 5'GACAAGTCTCGGCTACTGACC'3		
APC Ex15			
Cdc1	Forward 5' TGCCACTTGCAAAGTTTCTTC'3	Salih	153
	Reverse 5' TCTGCTTCCTGTGTCGTCTG'3		
Cdc 2	Forward 5'CATCAGCTGAAGATGAAATAGGA'3	Salih	165
	Reverse 5'TGCTGGATTTGGTTCTAGGG'3		
Cdc 3	Forward 5'GTGAGCGAAGTTCCAGCAGT'3	Salih	248
	Reverse 5'GCAATCGAACGACTCTCAA'3		
Cdc 4	Forward 5'ATGTTCAGGAGACCCCACTC'3	Salih	229
	Reverse 5' ACTTCTCGCTTGGTTTGAGC'3		
Cdc 5	Forward 5' CATGCCACCAAGCAGAAGTA'3	Salih	233
	Reverse 5' CACTCAGGCTGGATGAACAA'3		
Cdc 6	Forward 5' GGGTCCAGGTTCTTCCAGAT'3	Salih	226
	Reverse 5' GCCTCTTTCTCTTGGTTTTCA'3		
Cdc 7	Forward 5' GCCTCCAGTTCAGGAAAATG'3	Salih	245
	Reverse 5' TTGCCACAGGTGGAGGTAAT'3		
APC Ex15 Inner			
Cdc1	Forward 5'- CACTTGCAAAGTTTCTTCTAT-'3	Salih	
	Reverse 5'- GCTTCCTGTGTCGTCTGATT-'3		
Cdc 2	Forward 5'- TGGATTTGGTTCTAGGGTGC-'3	Salih	
	Reverse 5'- CAGCTGAAGATGAAATAGGATGT-'3		

Gene	Primer Sequence	Designed by:	Amplicon
Cdc 3	Forward 5'- AGCGAAGTTCCAGCAGTGTC-'3	Salih	
	Reverse 5'- ATCGAACGACTCTCAAAACT-'3		
Cdc 4	Forward 5'- TTCAGGAGACCCCACTCATG-'3	Salih	
	Reverse 5'- TCTCGCTTGGTTTGAGCT-'3		
Cdc 5	Forward 5'- GCCACCAAGCAGAAGTAAAAC-'3	Salih	
	Reverse 5'- TCAGGCTGGATGAACAAGAA-'3		
Cdc 6	Forward 5'- TCCAGGTTCTTCCAGATGCT-'3	Salih	
	Reverse 5'- TCTTTCTCTTGGTTTTCATTTG-'3		
Cdc 7	Forward 5'- TCCAGTTCAGGAAAATGACA-'3	Salih	
	Reverse 5'- CCACAGGTGGAGGTAATTTTG-'3		
PIK3CA Exon1	Forward 5'-CACGACCATCATCAGGTGAA-'3	Wakkas	168
	Reverse 5'-GGAGGGGGTATTTTCTTGCT-'3		
PIK3CA Exon1	Forward 5'-ATGCCCCCAAGAATCCTAGT-'3	Wakkas	129
	Reverse 5'-GAGGGGGTATTTTCTTGCTTC-'3		
PIK3CA Exon2 A	Forward 5'-TCTACAGAGTTCCCTGTTTGC-'3	Abutalib	213
	Reverse 5'-TGAGGTGAATTGAGGTCCCTA-'3		
PIK3CA Exon2 A	Forward 5'- TCTACAGAGTTCCCTGTTTGC-'3	Abutalib	202
	Reverse 5'- GAGGTCCCTAAGATCCACAGC-'3		
PIK3CA Exon2 B	Forward 5'- GGCATGCCAGTGTGTGAAT-'3	Abutalib	258
	Reverse 5'- AGACACAGGTAGAAGACTGCACTA-'3		
PIK3CA Exon2 B	Forward 5'- AAGTACAGGACTTCCGAAGAAA-'3	Abutalib	218
	Reverse 5'- AGACACAGGTAGAAGACTGCACTA-'3		
PIK3CA Exon9	Forward 5'-CTGTGAATCCAGAGGGGAAA-3'	Wakkas	197
	Reverse 5'-GCACTTACCTGTGACTCCATAGAA-3'		
PIK3CA Exon9	Forward 5'-AAGGGAAAATGACAAAGAACAG-3'	Wakkas	103
	Reverse 5'-CACTTACCTGTGACTCCATAGAAA-3'		
PIK3CA Exon20	Forward 5'-TGAGCAAGAGGCTTTGGAGT-3'	Wakkas	201
	Reverse 5'-CCTATGCAATCGGTCTTTGC-3'		
PIK3CA Exon20	Forward 5'-GCAAGAGGCTTTGGAGTATTTC-3'	Wakkas	115
	Reverse 5'-TTTTCAGTTCAATGCATGCTG-3'		
KRAS Exon2	Forward 5'-GCCTGCTGAAAATGACTGAA-`3	Wakkas	114
	Reverse 5'-TTGGATCATATTCGTCCACAA-3'		
KRAS Exon2	Forward 5'-CCTGCTGAAAATGACTGAATATAA-3'	Wakkas	112
	Reverse 5'-TGGATCATATTCGTCCACAAAA-3'		
KRAS Exon3	Forward 5'-CCAGACTGTGTTTCTCCCTTC-3'	Wakkas	152
	Reverse 5'-AAAGAAAGCCCTCCCCAGT-3'		
KRAS Exon3	Forward 5'-TGTGTTTCTCCCTTCTGAGGA-3'	Wakkas	145
	Reverse 5'-AAGAAAGCCCTCCCCAGT-3'		
KRAS Exon4	Forward 5'-AGACACAAAACAGGCTCAGGA-3'	Wakkas	160
	Reverse 5'-TTGAGAGAAAAACTGATATATTAAATGAC-		
KRAS Exon4	Forward 5'-GACACAAAACAGGCTCAGGACT-3'	Wakkas	105
	Reverse 5'-CAGATCTGTATTTATTTCAGTGTTA-3'		

Gene	Primer Sequence	Designed by:	Amplicon
BRAF Exon11	Forward 5'-TGTTTGGCTTGACTTGACTTT-3'	Wakkas	189
	Reverse 5'-CTTGTCACAATGTCACCACATTACATA-3'		
BRAF exon11	Forward 5'-GACGGGACTCGAGTGATGAT-3'	Wakkas	135
	Reverse 5'-TGTCACAATGTCACCACATTACA-3'		
BRAF Exon15	Forward 5'-ATCTACTGTTTTCCTTTACTTACTACAC-3'	Wakkas	205
	Reverse 5'-CAGCATCTCAGGGCCAA-3'		
BRAF Exon15	Forward 5'-TGTTTTCCTTTACTTACTACACCTCA-3'	Wakkas	143
	Reverse 5'-CCACAAAATGGATCCAGACA-3'		
TP53 Exon5	Forward 5'-AACTCTGTCTCCTTCCTCTTC-3'	Wakkas	240
	Reverse 5'-CTGTCGTCTCTCCAGCC-`3		
TP53 Exon5	Forward 5'-AGTACTCCCCTGCCCTCAAC-3'	Wakkas	194
	Reverse 5'-CTGCTCACCATCGCTATCTG-3'		
TP53 Exon6	Forward 5'- AGGCCTCTGATTCCTCACTGAT-3'	Wakkas	187
	Reverse 5'- ACCCTTAACCCCTCCTCCCA-`3		
TP53 Exon6	Forward 5'-CCTCTGATTCCTCACTGATTGC-3'	Wakkas	181
	Reverse 5'-CTTAACCCCTCCTCCCAGAG-3'		
TP53 Exon7	Forward 5'-ATCTTGGGCCTGTGTTATCT-3'	Wakkas	156
	Reverse 5'-GGGTGGCAAGTGGCTCCT-3'		
TP53 Exon7	Forward 5'-TTGGGCCTGTGTTATCTCCT-3'	Wakkas	150
	Reverse 5'-TGGCAAGTGGCTCCTGAC-3'		
TP53 Exon8	Forward 5'-CTCTTGCTTCTCTTTTCCTATCC-3'	Wakkas	192
	Reverse 5'-ACCGCTTCTTGTCCTGCTTG-3'		
TP53 Exon8	Forward 5'-TTGCTTCTCTTTTCCTATCCTGA-3'	Wakkas	186
	Reverse 5`- GCTTCTTGTCCTGCTTGCTT-3'		
PTEN Exon5A	Forward 5`-GGTTATCTTTTTACCACAGTTGCAC-`3	Wakkas	118
	Reverse 5`-GATTGTCATCTTCACTTAGCCATT-`3		
PTEN Exon5B	Forward 5`-TTCACTGTAAAGCTGGAAAGG-`3	Wakkas	117
	Reverse 5`-TGGTCCTTACTTCCCCATAGAA-`3		
PTEN Exon7	Forward 5`-GTTCCCTCAGCCGTTACCT-`3	Wakkas	191
	Reverse 5`-CACCTGCAGATCTAATAGAAAACAA-`3		
PTEN Exon8	Forward 5`-CGTGCAGATAATGACAAGGAA`-3	Wakkas	190
	Reverse 5`-TCTTCATCAGCTGTACTCCTAGAAT-`3		
FBXW7 Exon9	Forward 5'- CGGACACTCAAAGTGTGGAA-3'	Wakkas	130
	Reverse 5'- GAAGTCCCAACCATGACAAGA-3'		
FBXW7 Exon11	Forward 5'- GGAATTGCATTCACACGTTA-3'	Wakkas	122
	Reverse 5'- TCCTGTTTTGATATCCCAGATTTT-3'		
CTNNB1 Exon 3A	Forward 5'-CAATGGGTCATATCACAGATTCTT-'3	Abutalib	232
	Reverse 5'-CCTCAGGATTGCCTTTACCA-'3		
CTNNB1 Exon 3A	Forward 5'-CAATGGGTCATATCACAGATTCTT-'3	Abutalib	229
	Reverse 5'-CAGGATTGCCTTTACCACTCA-'3		
CTNNB1 Exon 3B	Forward 5'-GAATCCATTCTGGTGCCACT-'3	Abutalib	181
	Reverse 5'-TTCTGACTTTCAGTAAGGCAATG-'3		
CTNNB1 Exon 3B	Forward 5'-TTCTGGTGCCACTACCACAG-'3	Abutalib	174
	Reverse 5'-TTCTGACTTTCAGTAAGGCAATG-'3		

Gene	Primer Sequence	Designed by:	Amplicon
MET Exon 4	Forward 5'-GCTCTTTCCACCCCTTCTCT-'3	Abutalib	223
	Reverse 5'-TGCCAGCTGTTAGAGATTCCT-'3		
MET Exon 4	Forward 5'-TCCACCCCTTCTCTTCACAG-'3	Abutalib	216
	Reverse 5'-GCCAGCTGTTAGAGATTCCTAC-'3		
FGFR2 Exon 6	Forward 5'-CTCCTTTCTTCCCTCTCTCCA-'3	Abutalib	253
	Reverse 5'-CCAGTTGTGGGTACCTTTAGATTC-'3		
FGFR2 Exon 6	Forward 5'-TTTCTTCCCTCTCTCCACCA-'3	Abutalib	236
	Reverse 5'-CCTTTAGATTCAGAAAGTCCTCAC-'3		
KIT Exon 17	Forward 5'-TGGTTTTCTTTTCTCCTCCAAC-'3	Abutalib	184
	Reverse 5'-GCAGGACTGTCAAGCAGAGA-'3		
KIT Exon 17	Forward 5'-GGTTTTCTTTTCTCCTCCAACC-'3	Abutalib	176
	Reverse 5'-TGTCAAGCAGAGAATGGGTACT-'3		
CDH1 Exon 8	Forward 5'-CCTGGTCCTGACTTGGTTGT-'3	Abutalib	205
	Reverse 5'-GACCTTTCTTTGGAAACCCTCT-'3		
CDH1 Exon 8	Forward 5'-GTCCTGACTTGGTTGTGTCG-'3	Abutalib	196
	Reverse 5'-TTCTTTGGAAACCCTCTAAGGAG-'3		

Table 2: Genes tested for mutations using PCR/HRM and their NCBI reference sequence (accession) number

Genes	Exons Tested	NCBI Reference Sequence	Source of the Primers
POLE	9	NG_033840.1	Designed
	10		
	11		
	12		
	13		
	14		
POLD1	8	NG_033800.1	Designed
	9		
	10		
	11		
	12		
	13		
APC	15	NG_008481.4	(Published Blaker et al., 2004)
KRAS	2	NG_007524.1	Designed
	3		
	4		
	5		

Genes	Exons Tested	NCBI Reference Sequence	Source of the Primers
TP53	2	NG_017013.1	Published Seth et al., 2009b
	4		
	5		
	6		
	7		
	8		
	10		
PIK3CA	1	NG_012113.2	Designed
	2		
	9		
	20		
PTEN	3	NG_007466.2	Designed
	5		
	7		
	8		
BRAF	11	NG_007873.1	Designed
	15		
FBXW7	9	NG_029466.1	Designed
	11		
SMAD4	9	NG_013013.2	Designed
	11		
GNAS	8	NG_016194.1	Designed
CTNNB1	3	NG_013302.1	Designed
MET	4	NG_008996.1	Designed
FGFR	6	NG_007726.3	Designed
KIT	17	NG_007456.1	Designed
CDH1	8	NG_008021.1	Designed
NRAS	2	NG_007572.1	Designed

Table 3: List of genes were tested using NGS

Gene	Exon	Chromosome	Start	End	Detail
AKT1	2	14	105246425	105246558	NM_005163
ALK	23	2	29443642	29443706	NM_004304
APC	15 (H1)	5	112173743	112173988	NM_000038
APC	15 (H2)	5	112174526	112174778	NM_000038
APC	15 (H3)	5	112174975	112175370	NM_000038
APC	15 (H4)	5	112175469	112176130	NM_000038
BRAF	11	7	140481371	140481498	NM_004333
BRAF	15	7	140453072	140453195	NM_004333
CDH1	8	16	68846033	68846171	NM_004360
CDH1	9	16	68847213	68847403	NM_004360
CDH1	12	16	68855899	68856128	NM_004360
CTNNB1	2	3	41266012	41266161	NM_001904
EGFR	18	7	55241609	55241741	NM_005228
EGFR	19	7	55242412	55242518	NM_005228
EGFR	20	7	55248981	55249176	NM_005228
EGFR	21	7	55259407	55259572	NM_005228
KRBB2	20	17	37880974	37881169	NM_004448
FBXW7	7	4	153250819	153250942	NM_033632
FBXW7	8	4	153249355	153249546	NM_033632
FBXW7	9	4	153247153	153247388	NM_033632
FBXW7	10	4	153245331	153245551	NM_033632
FBXW7	11	4	153244028	153244303	NM_033632
FGFR2	6	10	123279488	123279688	NM_000141
FOXL2	1	3	138665109	138665221	NM_023067
GNAQ	4	9	80412431	80412569	NM_002072
GNAQ	5	9	80409374	80409513	NM_002072
GNAQ	6	9	80343425	80343588	NM_002072
GNAS	6	20	57480433	57480540	NM_000516
GNAS	8	20	57484403	57484483	NM_000516
KIT	9	4	55592018	55592221	NM_000222
KIT	11	4	55593577	55593713	NM_000222
KIT	13	4	55594175	55594292	NM_000222
KIT	17	4	55599231	55599363	NM_000222
KIT	18	4	55602662	55602780	NM_000222
KRAS	1	12	25398203	25398323	NM_004985
KRAS	2	12	25380163	25380351	NM_004985
KRAS	3	12	25378543	25378712	NM_004985
KRAS	4	12	25362724	25362850	NM_004985
MAP2K1	2	15	66727437	66727561	NM_002755
MET	1	7	116339543	116340298	NM_000245
MET	4	7	116380901	116381084	NM_000245
MET	13	7	116411898	116412048	NM_000245

Gene	Exon	Chromosome	Start	End	Detail
MET	15	7	116417441	116417528	NM_000245
MET	16	7	116418825	116419016	NM_000245
MET	17	7	116422037	116422156	NM_000245
MET	18	7	116423353	116423526	NM_000245
MET	20	7	116435936	116436183	NM_000245
MSH6	5	2	48030554	48030829	NM_000179
NRAS	1	1	115258677	115258786	NM_002524
NRAS	2	1	115256416	115256604	NM_002524
NRAS	3	1	115252185	115252354	NM_002524
NRAS	4	1	115251151	115251280	NM_002524
PDGFRA	11	4	55141003	55141145	NM_006206
PDGFRA	13	4	55144058	55144178	NM_006206
PDGFRA	17	4	55152003	55152135	NM_006206
PIK3CA	1	3	178916609	178916970	NM_006218
PIK3CA	2	3	178917473	178917692	NM_006218
PIK3CA	7	3	178927969	178928131	NM_006218
PIK3CA	9	3	178936039	178936127	NM_006218
PIK3CA	20	3	178951877	178952157	NM_006218
PTEN	1	10	89624222	89624310	NM_000314
PTEN	2	10	89653777	89653871	NM_000314
PTEN	3	10	89685266	89685319	NM_000314
PTEN	4	10	89690798	89690851	NM_000314
PTEN	5	10	89692765	89693013	NM_000314
PTEN	6	10	89711870	89712021	NM_000314
PTEN	7	10	89717605	89717781	NM_000314
PTEN	9	10	89725041	89725152	NM_000314
SMAD4	8	18	48591788	48591981	NM_005359
SMAD4	11	18	48604621	48604842	NM_005359
SRC	14	20	36031584	36031787	NM_005417
STK11	1	19	1206908	1207149	NM_000455
STK11	4	19	1220367	1220400	NM_000455
STK11	6	19	1221221	1221344	NM_000455
STK11	8	19	1222979	1223176	NM_000455
TP53	2	17	7579834	7579917	NM_000546
TP53	3	17	7579695	7579726	NM_000546
TP53	4	17	7579307	7579586	NM_000546
TP53	5	17	7578366	7578559	NM_000546
TP53	6	17	7578172	7578294	NM_000546
TP53	7	17	7577498	7577613	NM_000546
TP53	8	17	7577014	7577160	NM_000546
TP53	9	17	7576848	7576931	NM_000546
TP53	10	17	7573922	7574036	NM_000546
TP53	11	17	7572922	7573013	NM_000546

Table 4: the genes mRNA NCBI reference sequence number, the sequences of the qPCR primers.

Gene	NCBI Reference	Sequence of the Primers	Source of the
SMAD4	NM_005359.5	Forward 5`GTGCATATATAAAGGTCTTTGAT`3	Designed
PTEN	NM_001304717	Forward 5`GGACCAGAGACAAAAAGGGAGT`3	Designed
TGFBRII	NM_003242.5	Forward 5`CAGGTGGGAACTGCAAGATA`3	Designed
BCL2	NM_000657.2	Forward 5`GCTGGGATGCCTTTGTGGAA`3	Designed
KLF4	NM_001314052.1	Forward 5`CCGCTCCATTACCAAGAGC`3	Designed
RASA1	NM_002890.2	Forward 5`GGCCGGTATTATAACAGCATT`3	Designed

Table 5: List of miRNA genes, sequences of forward primers and Qiagen Cat. No.

MiRNAs	Sequence of Forward Primers	Qiagen Cat. No.
RNU6B	5`ACGCAAATTCGTGAAGCGTT`3	MS00033740
RNU61		MS00033705
miR-39	5`UCACCGGGUGUAAAUCAGCUUG`3	MS00019789
miR-20a-5p	5`UAAAGUGCUUAUAGUGCAGGUAG`3	MS00003199
miR-21-5p	5`UAGCUUAUCAGACUGAUGUUGA`3	MS00009079
miR-29a-3p	5`UAGCACCAUCUGAAAUCGGUUA`3	MS00003262
miR-31-5p	5`AGGCAAGAUGCUGGCAUAGCU`3	MS00003290
miR-92a-3p	5`UAUUGCACUUGUCCCGGCCUGU`3	MS00006594
miR-224-5p	5`UAUUGCACUUGUCCCGGCCUGU`3	MS00003878

Table 6: List of Antibodies

Antibody	KD	Cat No.	Type	Company
CD34	110	NCL-L-END	Mouse	Leica
CD31	130	M0823	Mouse	Dako
D2-40	38	N1607	Mouse	Dako
WT1	129	M3561	Mouse	Dako
BCL2	25	Mo887	Mouse	Dako
AE1-3	50	M3515	Mouse	Dako
BerEP4	50	M0804	Mouse	Dako
SMAD4	65	Ab40759	Rabbit	Abcam
TGFBRII	75	Ab61213	Rabbit	Abcam
RASA1	140	Ab40677	Rabbit	Abcam
KLF4	50	12173S	Rabbit	Cell Signaling
RAS	21	3965S	Rabbit	Cell Signaling
E-Cadherin	120	3195S	Rabbit	Cell Signalling
P85	85	4257	Rabbit	Cell Signalling
P110	110	3011	Rabbit	Cell Signalling
PTEN	54	9554	Rabbit	Cell Signalling
Phosphor-	54	9549	Rabbit	Cell Signalling
Phospho- AKT	60	9275	Rabbit	Cell Signalling
Phospho- AKT	60	9271	Rabbit	Cell Signalling
mTOR	289	2972	Rabbit	Cell Signalling
pCRAF	74	9421	Rabbit	Cell Signalling
pGSK	46	9336	Rabbit	Cell Signalling

Appendix B: Supplementary information of chapter Three

Table 1: List of overall mutations detected in all 83 CRC samples

Sample	Ploidy MS status	Gene	Mutation	Codon	Exon	Read depth	Alt Red depth	All variant frequency
1	A/MSS	APC	c.4348 C>T	1450	15	15678	1556	9.92
		KRAS	c.35 G>T	12	2	8239	384	4.66
			c.34 G>A	12	2	8269	2004	24.24
		TP53	c.379 T>C	127	4	11906	1278	10.73
			c.524 G>A	175	5	8602	953	11.08
			c.844 C>T	282	8	3074	480	15.61
		PIK3CA	c.353 G>A	118	2	16328	528	3.23
2	A/MSS	APC	c.4660-4661 ins A	1554	15	16965	3630	21.4
		BRAF	c.1799 T>A	600	15	31431	6759	21.5
		TP53	c.473 G>C	158	5	6918	2226	32.18
		FBXW7	c.1393 C>T	465	8	40817	2697	6.61
3	A/MSS	APC	c.4348 C>T	1450	15	20030	6350	31.7
		KRAS	c.35 G>T	12	2	12019	4058	33.76
		TP53	c.524 G>A	175	5	4260	2022	47.46
		PIK3CA	c.353 G>A	118	2	28935	2497	8.63
			c.1634 A>A	545	9	51761	5768	11.14
4	A/MSS	APC	c.3997 del A	1333	15	12904	1156	8.96
		APC	c.4216 C>T	1406	15	25233	3014	11.94
		NRAS	c.35 G>A	12	2	13572	2791	20.56
		TP53	c.271 T>T	91	4	4784	1372	28.68
5	A/MSS	APC	c.3379 C>T	1127	15	11971	2632	21.99
		KRAS	c.34 G>T	12	2	10275	3442	33.5
6	A/MSS	APC	c.4382-4383 del AA	1461	15	6409	1929	30.1
		KRAS	c.35 G>A	12	2	17219	10529	61.15
		TP53	c.428 T>A	143	5	6243	3171	50.79
		FBXW7	c.1396 G>A	465	8	27409	8791	32.07
		SMAD4	c.1609 G>T	537	11	21170	11192	52.87
		POLE	c.1364 T>A	455	14	-	-	-
7	A/MSS	APC	c.4011-4012 dle GC	1437	15	11729	3489	29.75
		KRAS	c.38 G>A	13	2	9000	1470	16.33
		TP53	c.427 C>T	273	8	7236	3036	41.96
		FBXW7	c.2065 C>T	689	11	8557	3773	44.09

Sample	Ploidy MS status	Gene	Mutation	Codon	Exon	Read depth	Alt Red depth	All variant frequency
8	A/MSS	APC	c.4529 del G	1510	15	14781	6772	45.82
			c.4530 C>A	1510	15	14647	6873	46.92
		KRAS	c.436 G>A	146	4	44577	22746	51.03
		TP53	c.638 G>A	213	6	6594	1599	24.25
		PIK3CA	c.317 G>G	106	1	22486	3781	16.81
		FBXW7	c.1513 C>T	505	9	19798	9603	48.5
9	A/MSS	BRAF	c.1799 T>A	600	15	35332	6958	19.69
		TP53	c.404 G>T	135	4	6373	1836	28.81
10	A/MSS	TP53	c.23 C>T	8	2	5086	164	3.22
		TP53	c.524 G>A	175	5	15744	2946	18.71
11	A/MSS	APC	T>T/TA	1421	15	31777	5301	16.68
		APC	c.4264-4271 del	1422-	15	31777	5267	16.57
		KRAS	c.35 G>T	12	2	10231	1736	16.97
		TP53	c.874 A>G	292	8	5623	2905	51.66
		PIK3CA	c.1633 G>A	545	9	35510	6285	17.7
		SMAD4	c.1091 T>G	364	9	15094	1650	10.93
			c.1094 G>Ac	365	9	14850	1477	9.95
12	A/MSS	TP53	c.524 G>A	175	5	8447	1818	21.52
13	A/MSS	TP53	c.659 A>G	220	6	3379	750	22.2
14	A/MSS	APC	c.4375-4376 Ins C	1459	15	17179	2218	12.91
		KRAS	c.35 G>T	12	2	7925	2812	35.48
15	A/MSS	-	-	-	-	-	-	-
16	A/MSS	APC	c.4326 del T	1442	15	26131	3500	13.39
		KRAS	c.35 G>T	12	2	11570	2834	24.49
		PIK3CA	c.1633 G>A	545	9	41341	5185	12.54
17	A/MSS	TP53	c.524 G>A	175	5	5376	536	9.97
18	A/MSS	TP53	c.743 G>A	248	7	15723	3435	21.85
19	A/MSS	APC	c.2540 del A	847	15	30496	2845	9.33
			c.4222 G>T	1408	15	22698	3713	16.36
		BRAF	c.1781 A>G	594	15	23949	4371	18.25
		TP53	c.1024 C>T	342	10	8666	2098	24.21

Sample	Ploidy MS status	Gene	Mutation	Codon	Exon	Read depth	Alt Red depth	All variant frequency
		SMAD4	c.1083 G>A	361	9	13533	3113	23
		POLE	c.833 C>T	278	9	-	-	-
20	A/MSS	APC	c.4732 del T	1578	15	9729	930	9.56
		KRAS	c.436 G>A	146	4	31484	2974	9.45
21	A/MSS	APC	C3842 C>G	1281	15	10619	2954	27.82
		TP53	c.761 T>A	254	7	11125	2814	25.29
		PTEN	c.795 del A	265	7	2378	871	36.63
22	A/MSS	APC	c.4285 C>T	1429	15	16207	5691	35.11
		KRAS	c.35 G>A	12	2	8516	3532	38.94
		TP53	c.524 G>A	175	5	6958	2437	38.94
		SMAD4	c.1609 G>C	537	11	8986	794	8.84
		POLE	C1295 T>C	432	13	-	-	-
23	A/MSS	TP53	c.565 del G	189	6	3904	1357	34.76
24	A/MSS	-	-	-	-	-	-	-
25	A/MSS	APC	c.3921-3925 del	1442	15	33628	12448	37.02
		TP53	c.527 C>T	176	5	6190	2555	41.28
26	A/MSS	TP53	c.742 C>T	248	7	18483	6998	37.86
		FBXW7	c.1787 C>G	596	10	43759	20408	46.64
27	A/MSS	APC	c.3944 C>A	1315	15	39561	22962	58.04
		TP53	c.55-56 Ins AT	19	2	3665	2009	54.82
28	A/MSS	APC	c.4239 del G	1413	15	13655	3535	25.89
			c.2627 G>A	876	15	16289	41.63	25.56
		KRAS	c.35 G>T	12	2	8515	2080	24.43
		TP53	c.562 C>T	282	8	3883	557	14.34
		PTEN	c.333 G>T	111	5	20244	6065	29.96
		KIT	c.2411 G>A	804	17	13493	3537	26.21
29	A/MSS	APC	C3871 C>T	1291	15	28098	7490	26.66
		KRAS	c.183 A>T	61	3	21990	6110	27.79
		TP53	c.518 T>C	173	5	3196	1331	41.65
		SMAD4	c.1496 G>A	499	11	16156	5857	36.25
		POLE	c.1240 G>A	414	13	-	-	-

Sample	Ploidy MS status	Gene	Mutation	Codon	Exon	Read depth	Alt Red depth	All variant frequency
30	A/MSS	-	-	-	-	-	-	-
31	A/MSS	KRAS	c.35 G>T	12	2	5424	1073	19.78
		SMAD4	c.1477 G>A	493	11	11722	2559	21.83
		CTNNB1	c.121 A>G	41	2	5048	1524	30.19
32	A/MSS	KRAS	c.149 C>T	50	3	11134	531	4.77
		TP53	c.725 G>T	242	7	10433	1780	17.06
33	A/MSS	APC	c.4360 del A	1454	15	6759	2004	29.65
		KRAS	c.35 G>C	12	2	6467	2715	41.98
		TP53	c.743 G>A	248	7	4813	2752	57.18
34	A/MSS	APC	c.4059-4060 Ins T	1376	15	8231	2071	25.16
		TP53	c.845-846 Ins CGGT	282	8	2034	763	37.51
		PIK3CA	c.1633 G>A	545	9	18638	2958	15.87
		FBXW7	c.1168 dup T	390	7	18748	3588	19.14
35	A/MSS	BRAF	c.1799 T>A	600	15	22600	8580	37.96
		TP53	c.701 A>G	234	7	2094	1027	49.04
		FGFR2	c.907 C>T	303	6	1362	41	3.01
36	A/MSS	TP53	c.747 G>T	249	7	14793	2276	15.39
37	A/MSS	BRAF	c.1799 T>A	600	15	19542	4665	23.87
		TP53	c.732 del C	241	7	13348	4498	33.7
		POLE	c.1298 G>A	433	13	-	-	-
38	A/MSS	TP53	c.454-466 del	152-..	5	5400	1046	19.37
		FGFR2	c.907 C>T	303	6	3628	116	3.2
39	A/MSS	TP53	c.742 C>T	248	7	14267	4122	28.89
40	A/MSS	TP53	c.1090 G>A	364	10	9575	296	3.09
			c.831 del T	277	8	16654	8418	50.55
		MET	c.1645 G>A	549	4	28652	868	3.03
		POLD1	c.1229 C>T	410	10	-	-	-
41	A/MSS	APC	c.3980 C>G	1327	15	10459	1558	14.9
		TP53	c.437 G>A	146	5	3801	1157	30.44
		POLE	c.1121 C>A	374	12	-	-	-

Sample	Ploidy MS status	Gene	Mutation	Codon	Exon	Read depth	Alt Red depth	All variant frequency
42	A/MSS	APC	c.3340 C>T	1114	15	2694	362	13.44
		KRAS	c.35 G>A	12	2	3662	799	21.82
		PIK3CA	c.1633 G>A	545	9	8857	776	8.76
43	A/MSS	KRAS	c.38 G>A	13	2	7915	312	3.94
		TP53	c.821 T>A	274	8	7872	150	3.18
44	A/MSS	APC	c.3957 del T	1319	15	18197	4190	23.03
		KRAS	c.35 G>A	12	2	7266	1791	24.65
		TP53	c.422 G>A	141	5	5426	1862	34.32
45	A/MSS	TP53	c.638 G>A	213	7	8932	1718	19.23
			c.844 C>T	282	8	5264	1969	37.41
		POLD1	c.1231C>T	404	10	-	-	-
46	A/MSS	APC	c.4033 G>T	1345	15	24722	6477	26.2
		TP53	c.701 A>G	234	7	14627	4579	31.31
47	D/MSS	BRAF	c.599 T>A	600	15	30018	4985	16.61
		PIK3CA	c.1495 C>T	499	9	14227	3339	23.47
		GNAS	c.602 G>A	201	8	10729	3683	34.33
48	D/MSS	APC	c.3921-3925 del	1307-	15	45726	2340	5.12
		KRAS	c.183 A>C	61	3	31431	1671	5.32
		TP53	c.514 G>T	172	5	5214	326	6.25
		PIK3CA	c.263 G>A	88	1	18523	960	5.18
		PTEN	c.202 T>C	68	3	42647	4614	10.82
49	D/MSS	KRAS	c.35 G>A	12	2	11472	5125	44.67
		SAMD4	c.1057 T>A	353	9	11402	4141	36.32
50	D/MSS	KRAS	c.35 G>A	12	2	12302	3976	32.32
		PIK3CA	c.344 G>T	115	1	12894	1370	10.63
			c.1624 G>A	542	9	28035	2823	10.07
		SMAD4	c.1573 T>G	525	11	12233	5911	48.32
		FBXW7	c.1394 G>A	465	8	32431	5932	18.29
51	D/MSS	-	-	-	-	-	-	-
52	D/MSS	APC	c.4660-4661 ins A	1554	15	35765	8978	25.1
		KRAS	c.35 G>T	12	2	9924	2928	29.5
		TP53	c.524 G>A	175	5	3046	412	13.52

Sample	Ploidy MS status	Gene	Mutation	Codon	Exon	Read depth	Alt Red depth	All variant frequency
			c.586 C>T	196	6	902	36	3.9
		PIK3CA	c.331-333 del AAG	111	1	31127	3184	10.23
		GNAS	c.602 G>A	201	8	16759	566	3.38
53	D/MSS	APC	c.3925 G>T	1309	15	29226	1616	5.53
			c.3940-3941 del AG	1314	15	29086	1938	6.66
			c.3947 C>T	1316	15	29202	1995	6.83
		KRAS	c.38 G>A	13	2	7884	548	6.95
		TP53	c.623 A>G	208	6	6652	386	5.8
54	D/MSS	APC	c.4216 C>T	1406	15	31948	6371	19.94
		KRAS	c.34 G>T	12	2	10915	2044	18.73
		PIK3CA	c.290 C>A	97	1	18426	1223	6.64
55	D/MSS	KRAS	c.35 G>A	12	2	9296	2044	21.99
		TP53	c.523 C>T	175	5	15627	2008	12.85
		FBXW7	c.1177 C>T	393	7	14850	2346	12.49
			c.1513 C>T	505	9	18781	2062	13.89
56	D/MSS	KRAS	c.35 G>T	12	2	8756	889	10.15
		TP53	c.186-193 del	79-81	6	26264	1937	7.38
		PIK3CA	c.247-249 Inv TTT	83	1	8608	567	6.59
		GNAS	c.614 C>T	205	8	17158	524	3.05
57	D/MSS	APC	c.4385-4386 del AG	1462	15	8603	575	6.68
		PIK3CA	c.1637 A>T	546	9	29501	1611	5.46
		FBXW7	c.1136 A>T	379	7	11711	708	6.05
		PTEN	c.209-	267	7	7763	1909	20.73
		SMAD4	c.1082 G>A	361	9	13633	746	5.47
58	D/MSS	-	-	-	-	-	-	-
59	D/MSS	APC	c.4348 C>T	1450	15	7905	265	3.35
		KRAS	c.38 G>A	13	2	4037	341	8.45
		NRAS	c.35 G>T	12	2	7341	363	4.94
		TP53	c.641 A>G	214	6	6018	379	6.3
		SAMD4	c.1051 G>A	351	9	8542	355	4.16
60	D/MSS	BRAF	c.1080 G>A	594	15	24432	7497	30.69
		TP53	c.917 C>T	273	8	6144	3083	50.18
61	D/MSS	APC	c.3916 G>T	1306	15	35103	8448	24.07

Sample	Ploidy MS status	Gene	Mutation	Codon	Exon	Read depth	Alt Red depth	All variant frequency
		TP53	c.406 del C	136	5	10324	3635	35.21
62	D/MSS	KRAS	c.35 G>A	12	2	14568	5250	36.04
		TP53	c.892 G>T	298	8	5367	2436	45.39
		PIK3CA	c.1258 T>C	420	7	28507	5670	19.89
63	D/MSS	APC	c.4468-4474 del	1490-	15	21364	14882	69.66
		KRAS	c.35 G>T	12	2	10604	3399	32.05
		TP53	c.536 A>G	179	5	4482	3418	76.26
		PTEN	c.209+1-209+4 del	68-70	3	27585	18973	68.78
64	D/MSS	BRAF	c.1799 T>A	600	15	15910	7720	48.52
		TP53	c.527 G>T	176	5	5398	2929	54.26
65	D/MSS	APC	c.3871 C>T	1291	15	28980	5605	19.34
			c.4468-4474 del	1490-	15	23882	4060	17
		KRAS	c.35 G>T	12	2	10804	690	6.39
			c.183 A>T	61	3	22400	4509	20.13
		TP53	c.518 T>C	173	5	2969	644	21.69
			c.536 A>G	179	5	3222	779	24.18
		SMAD4	c.1496 G>A	499	11	14227	3918	27.54
66	D/MSS	APC	c.4305 A>C	1435	15	11096	4779	43.07
			c.4308 del T	1436	15	11243	4815	42.83
		KRAS	c.35 G>A	12	2	12208	5651	46.29
		POLE	c.1369 A>G	457	14	-	-	-
67	D/MSS	APC	c.3707-3708 del CA	1236	15	20975	1232	5.87
			c.4033 G>T	1345	15	29576	1592	5.38
			c.4348 C>T	1450	15	19718	4798	24.33
		KRAS	c.38 G>A	13	2	10800	4144	38.37
		SMAD4	c.1478 A>G	493	11	28808	3081	10.69
		GNAS	c.602 G>A	201	8	12005	806	6.71
68	D/MSS	APC	c.3916 G>T	1306	15	24089	5497	22.82
		TP53	c.524 G>A	175	5	4632	870	18.78
69	D/MSS	APC	c.4385-4388 del	1464	15	7513	2710	36.07
			c.4660-4661 ins A	1554	15	18967	5895	31.08
		KRAS	c.35 G>T	12	2	10025	4638	46.26
		TP53	c.743 G>A	248	7	4406	1037	23.54
		PIK3CA	c.3140 A>G	1047	20	24875	2055	8.26

Sample	Ploidy MS status	Gene	Mutation	Codon	Exon	Read depth	Alt Red depth	All variant frequency
70	D/MSS	APC	c.4033 G>T	1345	15	15282	916	5.99
		TP53	c.524 G>A	175	5	6706	2366	35.28
71	D/MSS	APC	c.3944 C>T	1315	15	6564	641	9.77
		KRAS	c.436 G>A	146	4	8307	696	8.38
		TP53	c.1024 C>T	342	10	1756	258	14.69
			c.54-55 ins AT		2	2180	274	12.57
		PTEN	c.209+1-209+4 del	Intronic	3	6237	458	7.34
72	D/MSS	APC	c.4660-4661 ins A	1554	15	26838	5150	19.19
		KRAS	c.35 G>A	12	2	4509	1230	27.28
		TP53	c.745 A>G	249	7	16319	4368	26.77
		PIK3CA	c.1624 G>A	542	9	14632	1619	11.06
73	D/MSS	APC	c.4012 C>T	1338	15	13490	5146	38.15
		TP53	c.797 G>A	266	8	6597	2674	40.53
74	D/MSS	KRAS	c.139 G>C	47	3	17060	1399	8.2
		TP53	c.517 G>A	173	5	9823	964	9.81
		CDH1	c.1024 A>T	342	8	12277	578	4.71
75	D/MSS	APC	c.3368 del A	1123	15	8135	2070	25.42
			c.3872 del A	1291	15	25856	2048	7.92
		KRAS	c.35 G>A	12	2	5671	1197	21.11
		TP53	c.844 C>T	282	15	4953	1610	32.51
76	D/MSS	NRAS	c.181 G>A	61	3	21442	8883	41.43
		TP53	c.761 T>A	254	7	7658	2834	37.01
77	D/MSS	-	-	-	-	-	-	-
78	D/MSS	APC	c.4385-4388 del	1464	15	5922	2100	35.46
			c.4660-4661 ins A	1554	15	14478	5358	37.01
		KRAS	c.35 G>T	12	2	8095	3697	45.67
		TP53	c.743 G>A	248	7	5884	1238	21.04
		PIK3CA	c.3140 A>G	1047	20	16812	1414	8.41
79	D/MSS	KRAS	c.38 G>A	13	2	2452	112	4.57
		TP53	c.821 T>A	274	8	5548	252	4.54
80	D/MSS	APC	c.2626 C>T	876	15	22487	7061	31.4
			c.4239 del G	1413	15	23263	6869	29.53

Sample	Ploidy MS status	Gene	Mutation	Codon	Exon	Read depth	Alt Red depth	All variant frequency
		KRAS	c.35 G>T	12	2	5580	1764	31.61
		TP53	c.844 C>T	282	8	4064	340	8.37
		PTEN	c.333 G>T	111	5	23402	10058	42.98
		KIT	c.2411 G>A	804	17	12532	4016	32.05
81	D/MSS	APC	c.4455 del T	1485	15	15133	5401	35.69
		KRAS	c.35 G>C	12	2	3148	690	21.92
		TP53	c.223-224 dup CC	75	2	4299	1580	36.72
82	D/MSI	APC	c.2663 C>T	888	15	20303	2768	13.63
			c.4222 G>T	1408	15	20542	5688	27.69
		KRAS	c.35 G>A	12	2	7544	1915	25.38
		TP53	c.229 C>T	77	4	1444	737	51.04
		PIK3CA	c.325_327 del GAA	109	1	25505	3506	13.75
		FBXW7	c.1393 C>T	465	8	28121	2956	10.51
83	D/MSI	TP53	c.844 C>T	282	8	3584	548	15.29
		PTEN	c.795 del A	267	7	3924	569	14.5
		POLE	c.1382 C>A	461	14	-	-	-

Table 2: Details and frequency of all variants detected in both biopsy and resection specimens.

Case No.	Gene	Mutation	Codon	Exon	Resection			Biopsy		
					Read depth	Alt read depth	Alt variant Freq %	Read depth	Alt read depth	Alt variant Freq %
1	APC	c.3997delA	1333	15	12904	1156	8.96	32117	1267	3.94
	APC	c.4216C>T	1406	15	25233	3014	11.94	25829	1050	4.07
	NRAS	c.35G>A	12	2	13572	2791	20.56	17257	1401	8.12
	TP53	c.273G>A	91	4	4784	1372	28.68	11000	527	4.79
2	APC	c.4216C>T	1127	15	11971	2632	21.99	14951	1938	12.96
	KRAS	c.34G>T	12	2	10275	3442	33.5	12183	2735	22.45
3	APC	c.4382_4383delA	1461	15	6409	1929	30.1	10671	3164	29.65
	KRAS	c.35G>A	12	2	17219	10529	61.15	13894	8510	61.25
	TP53	c.428T>A	143	5	6243	3171	50.79	7578	3787	49.97
	FBXW7	c.1394G>A	465	8	27409	8791	32.07	29735	9191	30.91
	SMAD4	c.1609G>T	537	11	21170	11192	52.87	23362	12762	54.63
4	APC	c.4375_4376insC	1459	15	17179	2218	12.91	13800	1669	12.09
	KRAS	c.35G>T	12	2	7925	2812	35.48	6180	2076	33.59
	TP53	402_403delTT	332	10	9646	4360	45.2	12112	5533	45.68
5	TP53	c.994-2A>C	332	10	1366	425	31.11	3856	1378	35.74

6	APC	c.4326delT	1442	15	26131	3500	13.39	11468	690	6.02
	KRAS	c.35G>T	12	2	11570	2834	24.49	6137	727	11.85
	PIK3CA	c.1633G>A	545	9	41341	5185	12.54	14994	841	4.79
7	APC	c.4732delT	1578	15	11433	1333	11.66	9729	930	9.56
	KRAS	c.436G>A	146	4	38504	4554	11.83	31484	2974	9.45
8	-	-	-	-	-	-	-	-	-	-
9	APC	c.4660_4661insA	1554	15	35765	8978	25.1	18723	2791	14.91
	KRAS	c.35G>T	12	2	9924	2928	29.5	6392	1019	15.94
	TP53	c.524G>A	175	5	3046	412	13.53	7680	1547	20.14
		c.886C>T	196	6	1902	75	3.9	2104	81	3.8
	PIK3CA	c.331_333delAA	111	1	31127	3184	10.23	15502	2693	17.37
	GNAS	c.2543C>T	201	8	16759	566	3.38			
10	APC	c.3925G>T	1309	15	29226	1616	5.53	28474	4277	15.2
		c.3940_3941delA	1314	15	29086	1938	6.66	28154	4700	16.69
		c.3946C>T	1316	16	29202	1995	6.83	28403	4818	16.96
	KRAS	c.38G>A	13	2	7884	548	6.95	6162	1030	16.72
	TP53	c.623A>G	208	6	6652	386	5.8	5336	961	18.01
11	APC	c.4216C>T	1406	15	31948	6371	19.94	16480	4529	27.48
	KRAS	c.34G>T	12	2	10915	2044	18.73	5866	1442	24.58
	PIK3CA	c.290C>A	97	1	18426	1223	6.64	14432	1193	8.27
12	KRAS	c.35G>A	12	2	9296	2044	21.99	5667	2287	40.36
	TP53	c.523C>T	175	5	15627	2008	12.85	5659	1742	30.78
	FBXW7	c.1177C>T	393	9	14850	2346	12.49	7879	2195	27.86

		c.1513C>T	505	11	18781	2062	13.89	11290	3140	27.81
13	KRAS	c.35G>T	12	2	8756	889	10.15	7205	347	4.82
	TP53	c.186_193delAG	79-81	4	26264	1937	7.38	24144	1123	4.6
	PIK3CA	c.247_249invTTT	83	1	8608	567	6.59	11986	452	3.77
	GNAS	c.2531G>A	205	8	17158	524	3.05			
14	APC	c.4385_4386delA	1462	15	8603	575	6.68	5907	758	12.83
	PIK3CA	c.1637A>T	546	9	29501	1611	5.46	18036	2035	11.28
	FBXW7	c.1136A>T	379	7	11711	708	6.05	11048	1297	11.74
	PTEN	c.801+1G>A	*	7	12950	1308	10.1	7763	1609	20.73
	SMAD4	c.1082G>A	361	9	13633	746	5.47	8609	293	3.4
15	-	-	-	-	-	-	-	-	-	-
16	BRAF	c.1780G>A	594	15	24432	7497	30.69	21500	7588	35.29
	TP53	c.817C>T	273	8	6144	3083	50.18	3904	2354	60.3
17	APC	c.4011_4012del	1437-	15	20739	5044	24.32	11729	3489	29.75
	KRAS	c.38G>A	13	2	6077	813	13.38	9000	1470	16.33
	TP53	c.817C>T	273	8	6772	2345	34.63	7236	3036	41.96
		c.874A>G	292	8	3481	267	13.03	5621	365	15.4
	PIK3CA	c.1633G>A	545	9	14717	184	3.44	21438	843	3.93
	FBXW7	c.2065C>T	689	11	8557	3773	44.09	6326	2409	3808
18	APC	c.4529delG	1510	15	14781	6772	45.82	20798	3095	14.88
		c.4530C>A	1510	15	14647	6873	46.92	20477	3144	15.35
	KRAS	c.436G>A	146	4	44577	22746	51.03	31842	8980	28.2
	PIK3CA	c.316G>C	106	1	22486	3781	16.81	19929	4402	22.09

	FBXW7	c.1513C>T	505	9	19798	9603	48.5	10881	1955	17.97
19	BRAF	c.1799T>A	600	15	35332	6958	19.69	24135	6644	27.53
	TP53	c.404G>T	135	5	6373	1836	28.81	8358	4088	48.91
20	TP53	c.23C>T	8	2	5086	164	3.22			
	TP53	c.524G>A	175	5	15744	2946	18.71	10716	2382	22.23
21	APC	c.4263_4264insA	1421	15	31777	5301	16.68	16064	2303	14.34
	APC	c.4264_4271del	1422-	15	31777	5267	16.57	16064	2278	14.18
	KRAS	c.35G>T	12	2	10231	1736	16.97	4723	862	18.25
	TP53	c.874A>G	292	8	5623	2905	51.66	4856	2536	52.22
	PIK3CA	c.1633G>A	545	9	35510	6285	17.7	14794	2421	16.36
	SMAD4	c.1091T>G	364	9	15094	1650	10.93	6080	751	12.35
		c.1094G>A	365	9	14850	1477	9.95	5963	416	6.98
22	TP53	c.524G>A	175	5	8447	1818	21.52	15164	4553	30.03
23	APC	c.2663C>T	888	15	20303	2768	13.63	15845	3810	24.05
		c.4222G>T	1408	15	20542	5688	27.69	13631	6669	48.93
	KRAS	c.35G>A	12	2	7544	1915	25.38	7822	3261	41.69
	TP53	c.229C>T	77	4	1444	737	51.04	2030	970	47.78
	PIK3CA	c.325_327delGA	109	1	25505	3506	13.75	16829	4469	26.56
	FBXW7	c.1393C>T	465	8	28121	2956	10.51	17997	3383	18.8
24	TP53	c.844C>T	282	8	3584	548	15.29	4430	215	4.85
	PTEN	c.795delA	267	7	3924	569	14.5	4541	359	7.91
25	TP53	c.659A>G	220	7	3379	750	22.2	5888	2132	36.21

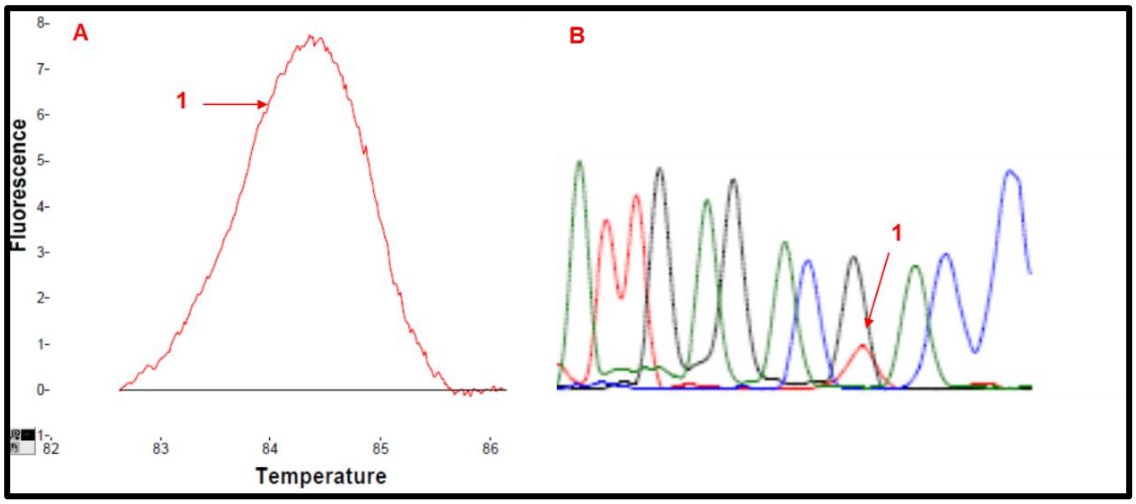

Figure 2: HRM analysis and sequencing results obtained from FFPE CRCs samples looked aberrant in exonuclease domain region in *POLE* gene. (A) Difference plot of *POLE* exon 9 conducted on FFPE CRCs samples showing how three of the FFPE samples containing a mutation and melts differently from the rest samples. (B) Direct DNA sequencing showing heterozygous mutations: (1) G>T. Black=G, Blue=C, Green=A, Red=T.

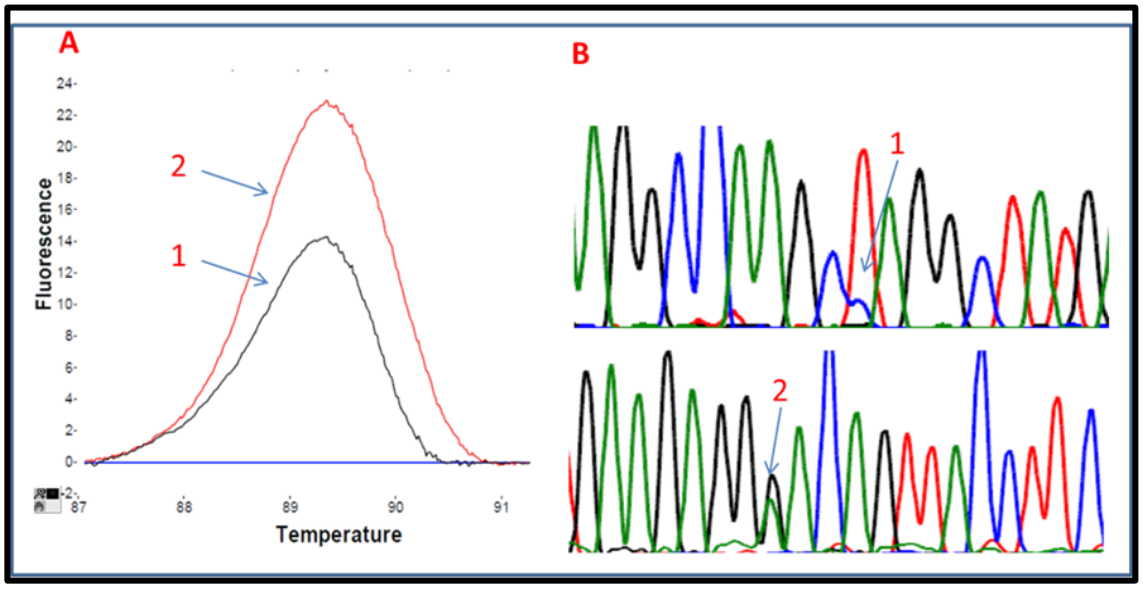

Figure 3: HRM analysis and sequencing results obtained from FFPE CRCs samples looked aberrant in exonuclease domain region in *POLE* gene. (A) Difference plot of *POLE* exon 13 conducted on FFPE CRCs samples showing how three of the FFPE samples containing a mutation and melts differently from the rest samples. (B) Direct DNA sequencing showing heterozygous mutations: (1) T > C, (2) G > A. Black=G, Blue=C, Green=A, Red=T.

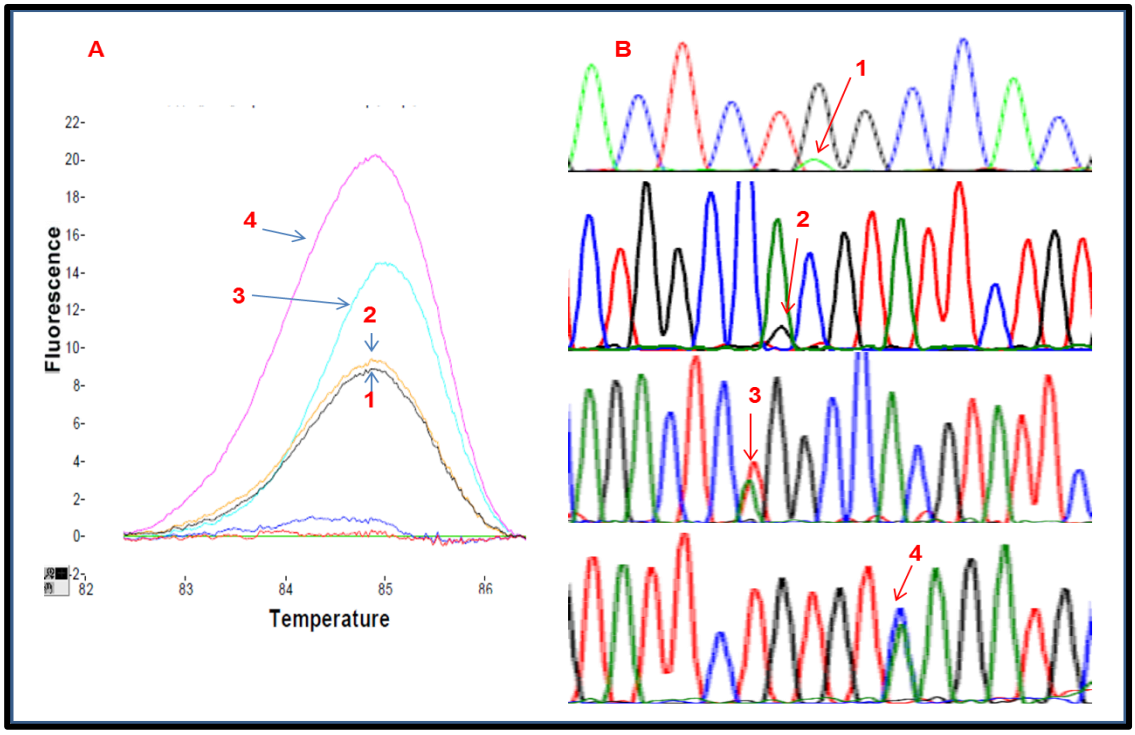

Figure 4: HRM analysis and sequencing results obtained from FFPE CRCs samples looked aberrant in exonuclease domain region in POLE gene. (A) Difference plot of *POLE* exon 14 conducted on FFPE CRCs samples showing how four of the FFPE samples containing a mutation and melts differently from the rest samples. (B) Direct DNA sequencing showing a heterozygous mutation: (1) G>A, (2) A > G, (3) T > A, (4) C > A. Black=G, Blue=C, Green=A, Red=T.

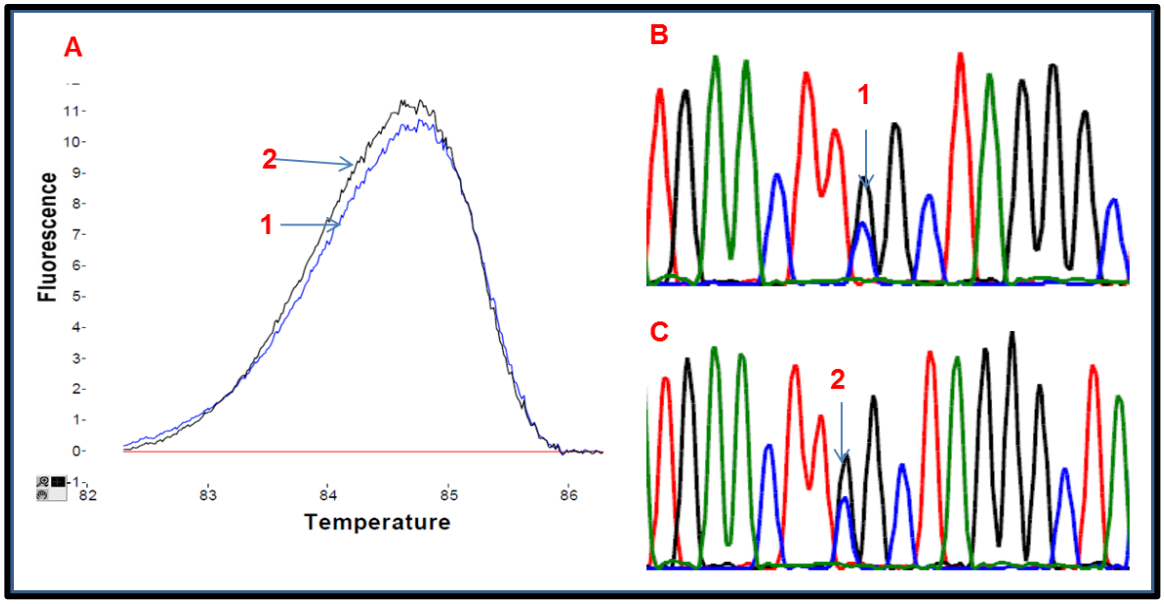

Figure 5: HRM analysis and sequencing results obtained from FFPE CRC sample and their normal corresponding tissue looked aberrant in exonuclease domain region in POLE gene. (A) Difference plot of *POLE* conducted on FFPE CRCs samples showing how (CRC sample and their corresponding normal tissue) containing a mutation and melts differently from the wild sample. (B) and (C) direct DNA sequencing showing a heterozygous mismatch at the intornic region near the 5' end of exon 10: (1) G > C, in CRC sample and their corresponding normal tissue respectively. Black=G, Blue=C, Green=A, Red=T.

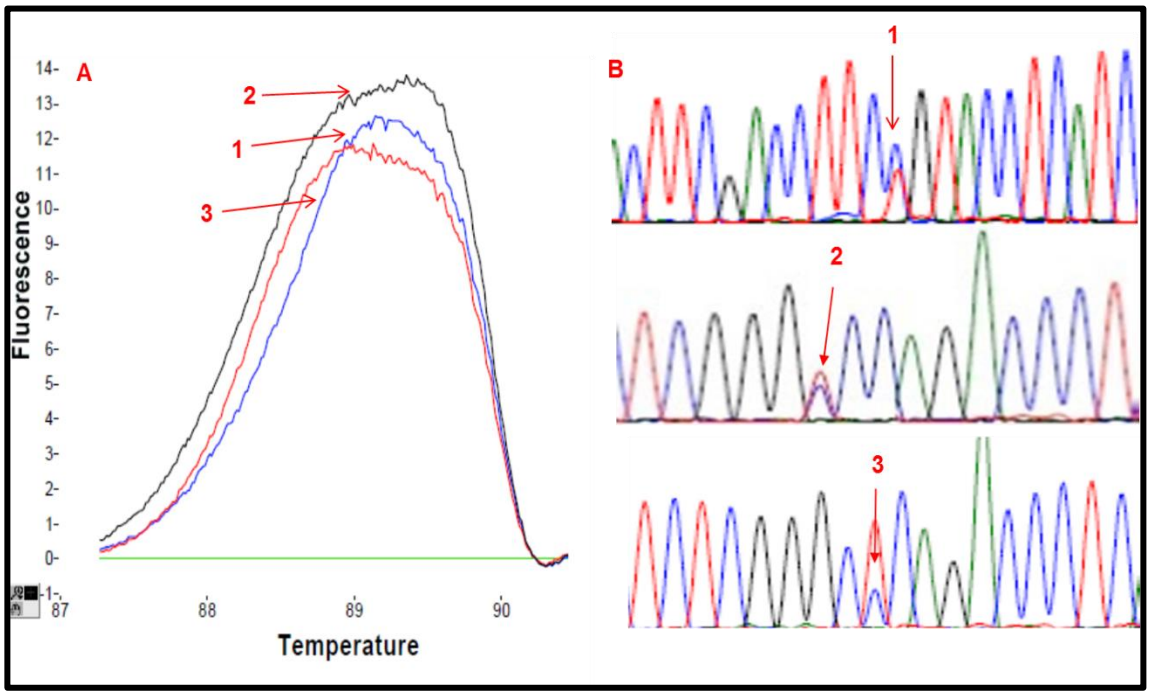

Figure 6: HRM analysis and sequencing results obtained from FFPE CRCs samples looked aberrant in exonuclease domain region in POLD1 gene. (A) Difference plot of *POLD1* exon 12 conducted on FFPE CRCs samples showing how three of the FFPE samples containing a mutation and melts differently from the rest samples. (B) Direct DNA sequencing showing a heterozygous mutation: (1) C>T, (2) C>T, (3) T>C. Black=G, Blue=C, Green=A, Red=T.

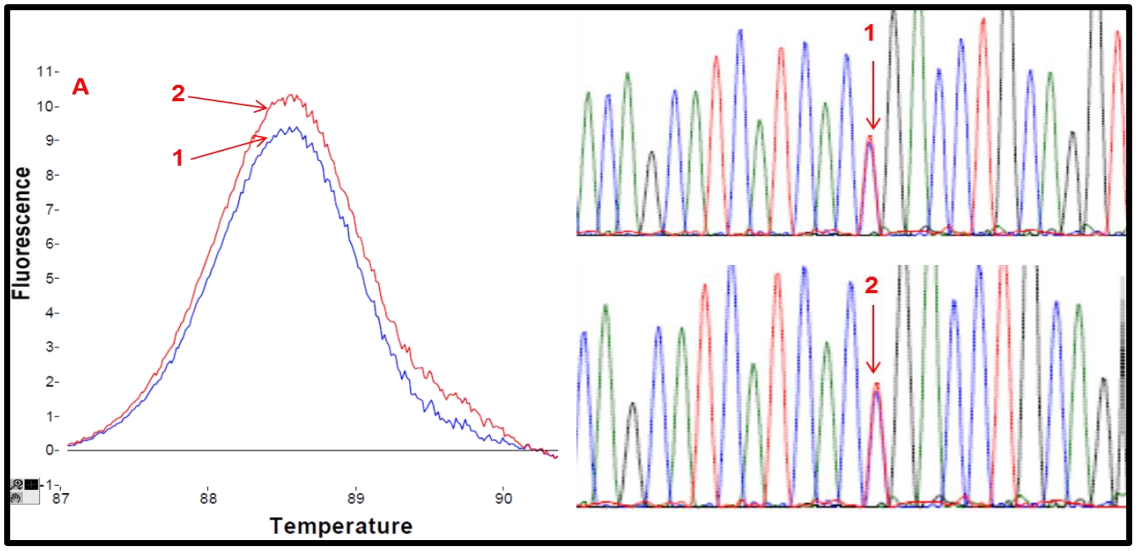

Figure 7: HRM analysis and sequencing results obtained from FFPE CRC sample and their normal corresponding tissue looked aberrant in exonuclease domain region in POLD1 gene. (A) Difference plot of *POLD1* conducted on FFPE CRCs samples showing how (CRC sample and their corresponding normal tissue) containing a mutation and melts differently from the wild sample. (B) direct DNA sequencing showing a heterozygous mismatch in exon 13 near conserved region V: (1) T > C, in CRC sample and their corresponding normal tissue respectively. Black=G, Blue=C, Green=A, Red=T.

Appendix C: Supplementary information of chapter Six

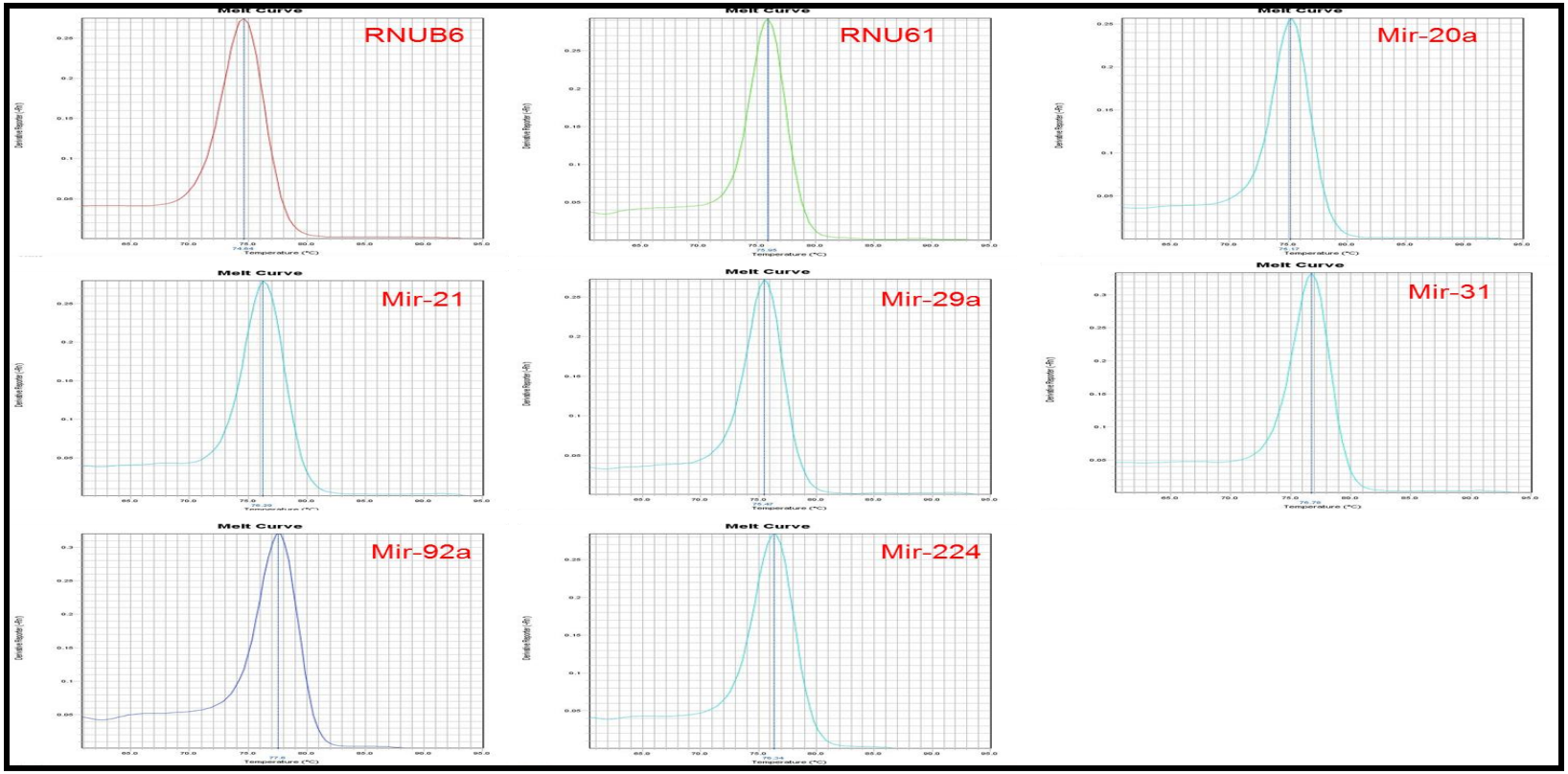

Figure 1: a qPCR melting curve analysis was performed which resulted in single sharp peak indicated no primer-dimers were generated specific melting temperatures.

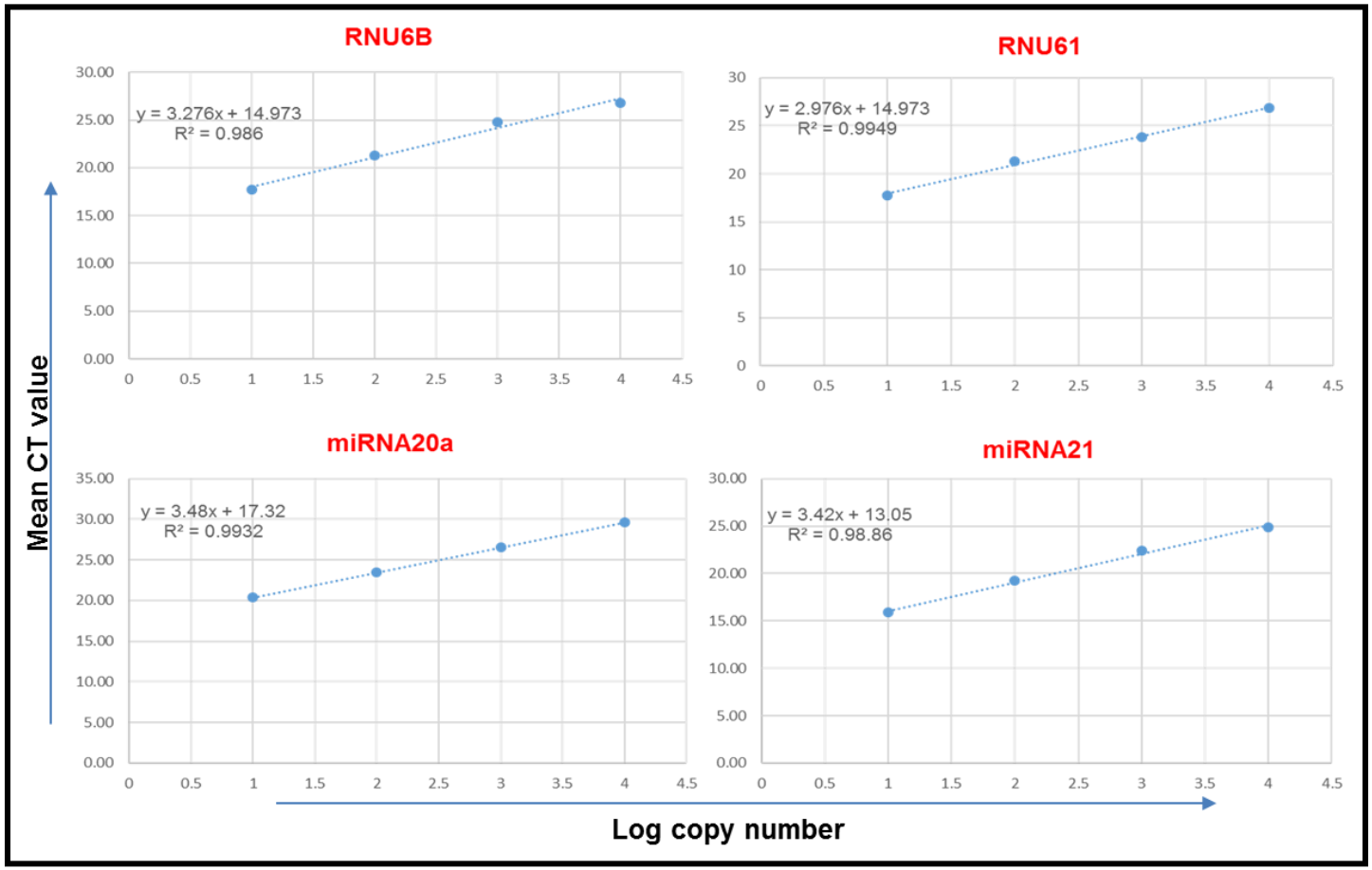

Figure 2: Real-Time PCR Standard Curve representing high PCR Efficiency. A 100% efficient reaction will yield a 10-fold increase in PCR amplicon every 3.32 cycles during the exponential phase of amplification ($\log_2 10 = 3.3219$).

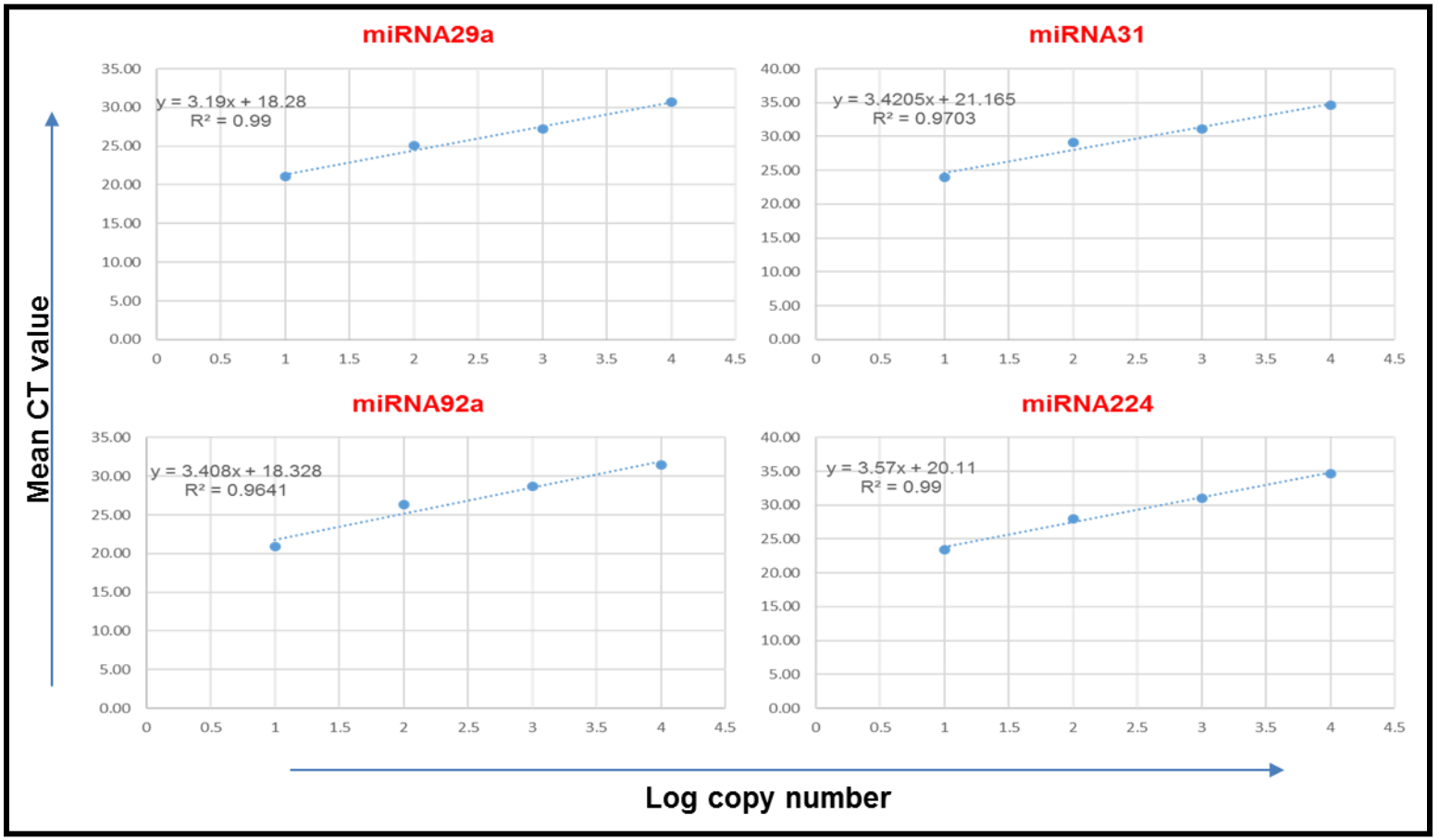

Figure 3: Real-Time PCR Standard Curve representing high PCR Efficiency. A 100% efficient reaction will yield a 10-fold increase in PCR amplicon every 3.32 cycles during the exponential phase of amplification ($\log_2 10 = 3.3219$).

Appendix D: Supplementary information of chapter Eight

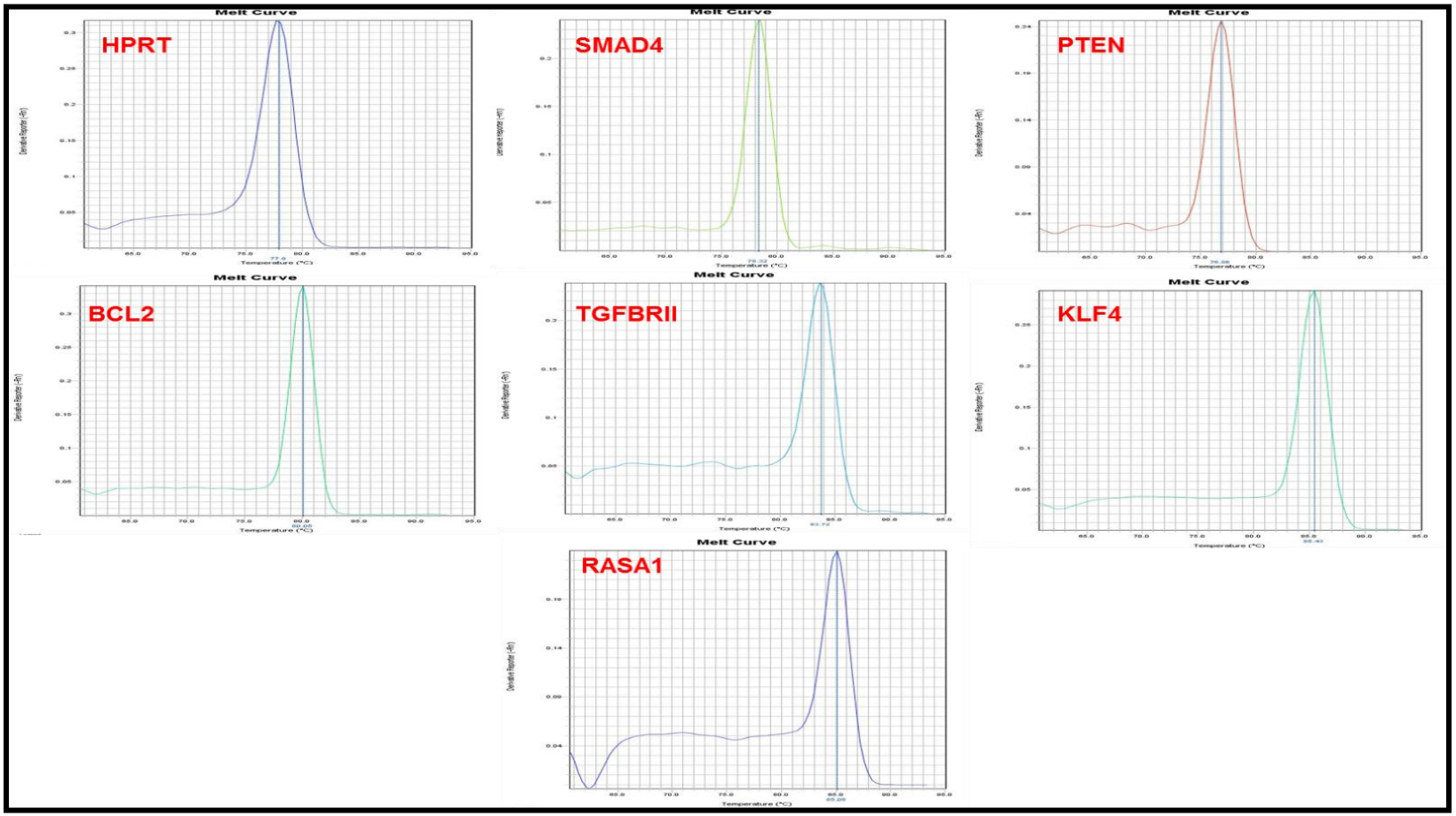

Figure 1: a qPCR melting curve analysis was performed which resulted in single sharp peak indicated no primer-dimers were generated specific melting temperatures

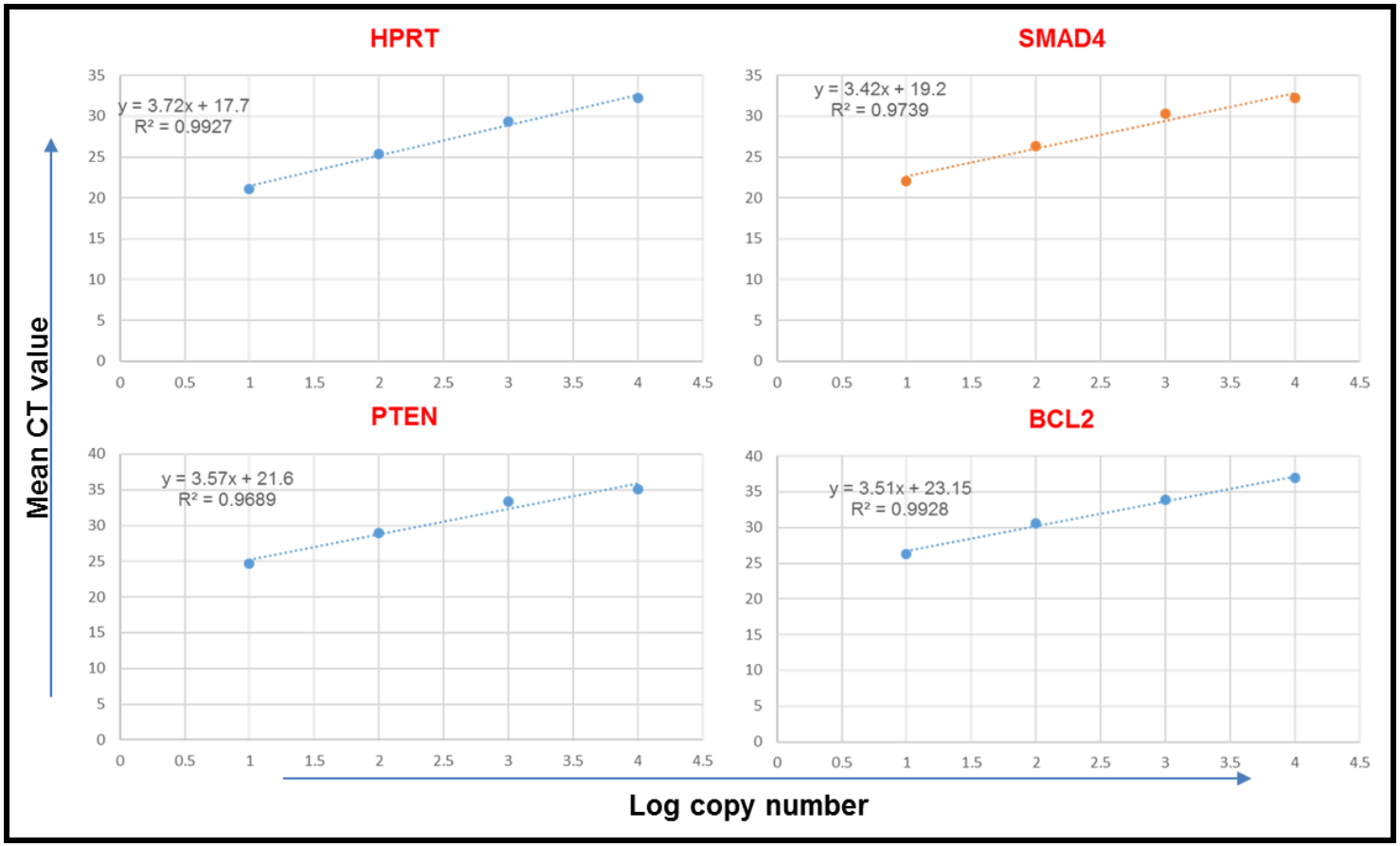

Figure 2: Real-Time PCR Standard Curve representing high PCR Efficiency. A 100% efficient reaction will yield a 10-fold increase in PCR amplicon every 3.32 cycles during the exponential phase of amplification ($\log_2 10 = 3.3219$).

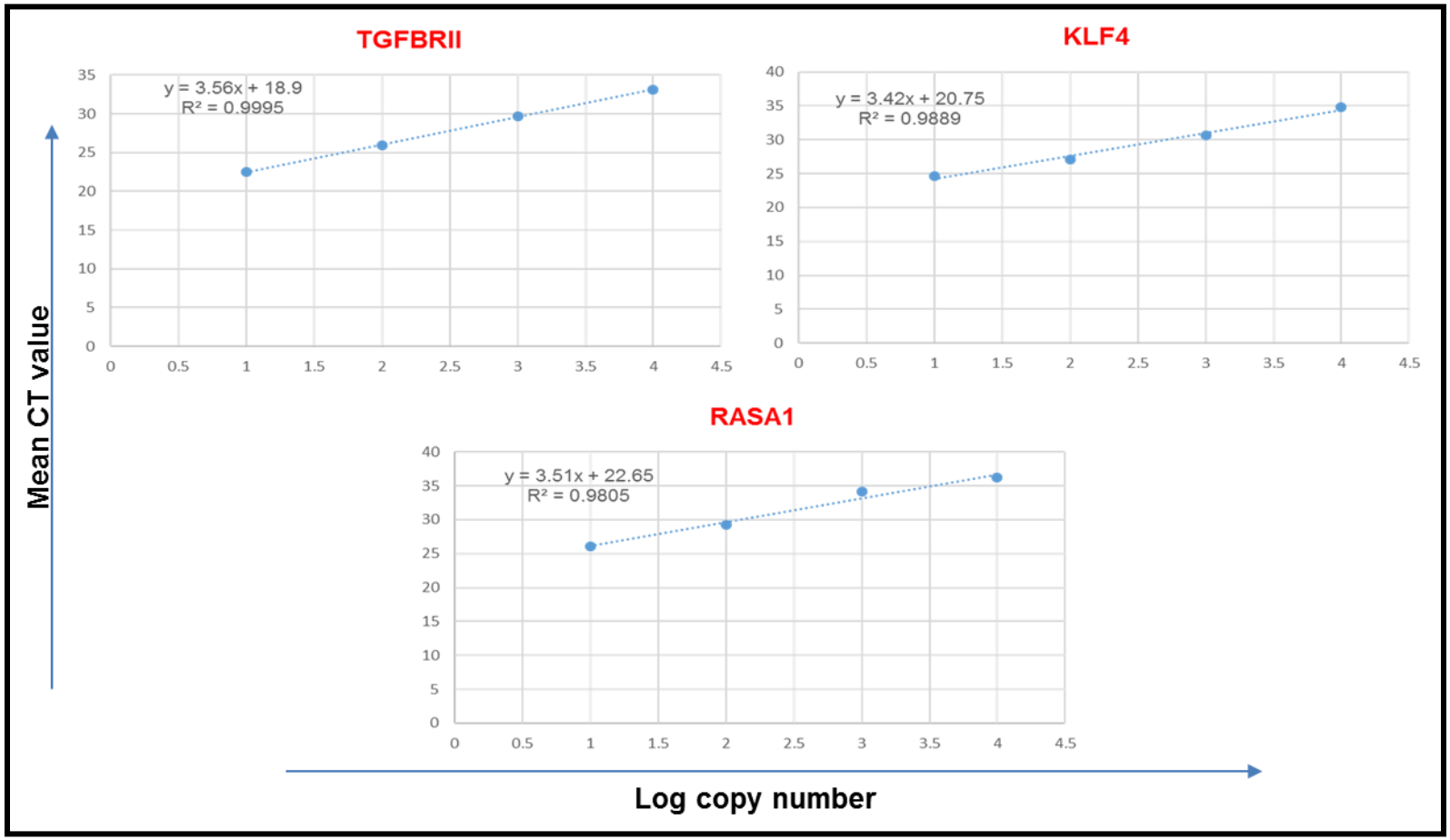

Figure 3: Real-Time PCR Standard Curve representing high PCR Efficiency. A 100% efficient reaction will yield a 10-fold increase in PCR amplicon every 3.32 cycles during the exponential phase of amplification ($\log_2 10 = 3.3219$).

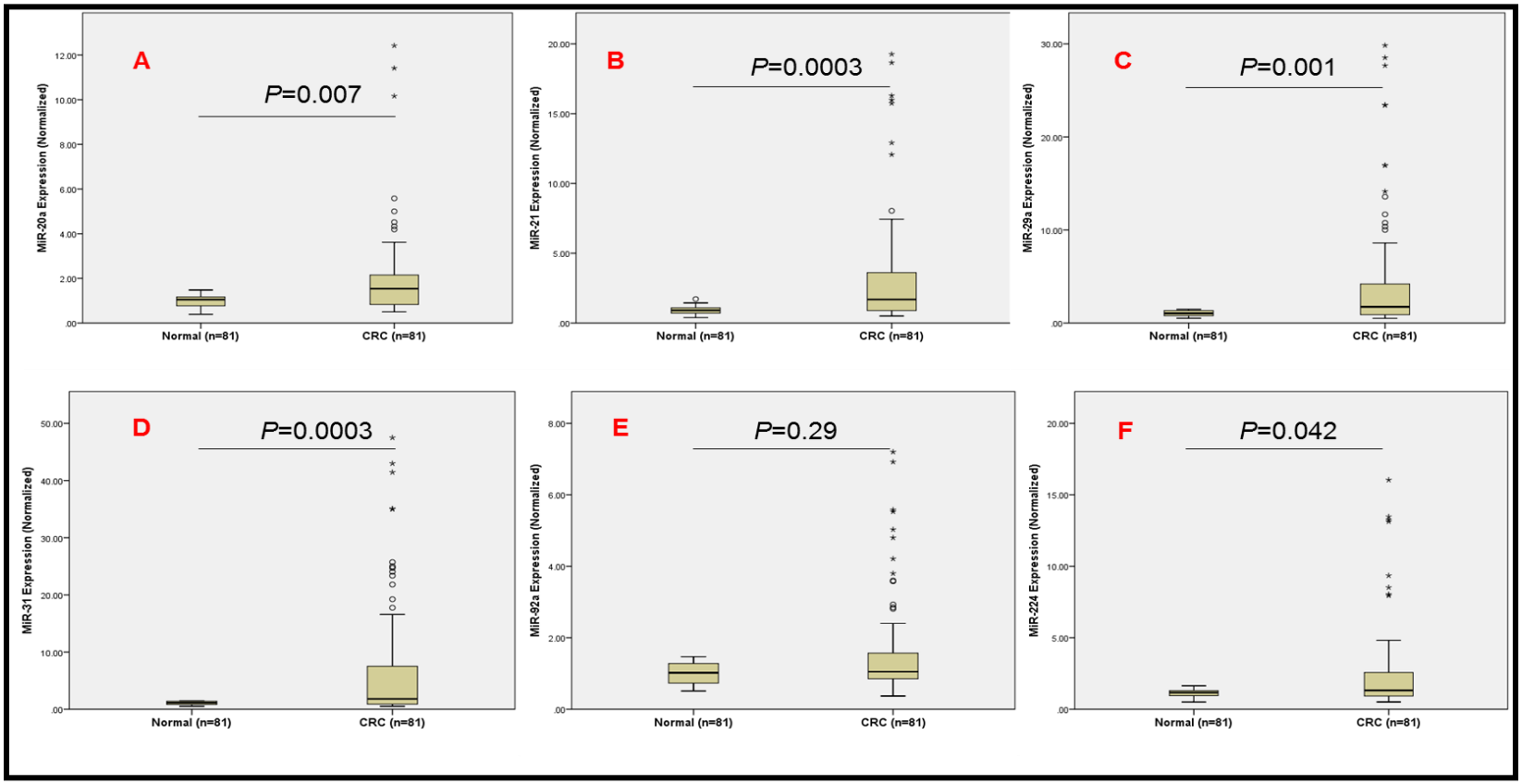

Figure 4: MicroRNA selection and validation (Phase II) by quantitative reverse transcription polymerase chain reaction (RT-PCR) analysis. Box plots of plasma levels of (A) miR-20a, (B) miR-21 (C) miR29a, (D) miR-31, (E) miR-92a and (F) miR-224 in healthy normal (N) subjects (n=81) and patients with colorectal cancer (CRC) (n=81). Expression levels of the miRNAs (log10 scale at y-axis) are normalised to RNU6B. The lines inside the boxes denote the medians. The boxes mark the interval between the 25th and 75th percentiles. The whiskers denote the interval between the 10th and 90th percentiles. Filled circles indicate data points outside the 10th and 90th percentiles. Statistically significant differences were determined using Wilcoxon tests.

References

1. Bedeir, A. and A.M. Krasinskas, *Molecular diagnostics of colorectal cancer.* Archives of pathology & laboratory medicine, 2011. **135**(5): p. 578-587.
2. Söreide, K., et al., *Microsatellite instability in colorectal cancer.* British journal of surgery, 2006. **93**(4): p. 395-406.
3. Jemal, A., et al., *Global cancer statistics.* CA: a cancer journal for clinicians, 2011. **61**(2): p. 69-90.
4. Al-Sohaily, S., et al., *Molecular pathways in colorectal cancer.* Journal of gastroenterology and hepatology, 2012. **27**(9): p. 1423-1431.
5. Cappell, M.S., *Pathophysiology, clinical presentation, and management of colon cancer.* Gastroenterology Clinics of North America, 2008. **37**(1): p. 1-24.
6. Wei, E.K., et al., *Cumulative risk of colon cancer up to age 70 years by risk factor status using data from the Nurses' Health Study.* American journal of epidemiology, 2009. **170**(7): p. 863-872.
7. Smith, G., et al., *Mutations in APC, Kirsten-ras, and p53—alternative genetic pathways to colorectal cancer.* Proceedings of the National Academy of Sciences, 2002. **99**(14): p. 9433-9438.
8. Benson, A., *Epidemiology, disease progression, and economic burden of colorectal cancer.* Journal of managed care pharmacy, 2007. **13**(6 Supp C): p. 5-18.
9. Wong, S.H., et al. *Genome-wide association and sequencing studies on colorectal cancer*. in *Seminars in cancer biology*. 2013. Elsevier.
10. Calvert, P.M. and H. Frucht, *The genetics of colorectal cancer.* Annals of internal medicine, 2002. **137**(7): p. 603-612.
11. Ogino, S. and A. Goel, *Molecular classification and correlates in colorectal cancer.* The Journal of Molecular Diagnostics, 2008. **10**(1): p. 13-27.
12. Wood, L.D., et al., *The genomic landscapes of human breast and colorectal cancers.* Science, 2007. **318**(5853): p. 1108-1113.
13. Zlobec, I., et al., *Stratification and prognostic relevance of Jass's molecular classification of colorectal cancer.* Frontiers in oncology, 2012. **2**: p. 7.
14. Rowan, A., et al., *Refining molecular analysis in the pathways of colorectal carcinogenesis.* Clinical Gastroenterology and Hepatology, 2005. **3**(11): p. 1115-1123.
15. Jass, J.R., *Limitations of the adenoma–carcinoma sequence in colorectum.* Clinical Cancer Research, 2004. **10**(17): p. 5969-5970.
16. Mudassar, S., M.S. Khan, and N.P. Khan, *Possible Role of Proto-Oncogenes in Colorectal Cancer—A Population Based Study.* 2014.
17. Jass, J.R., J. Young, and B.A. Leggett, *Evolution of colorectal cancer: change of pace and change of direction.* Journal of gastroenterology and hepatology, 2002. **17**(1): p. 17-26.
18. Banerjea, A., et al., *Microsatellite and chromosomal stable colorectal cancers demonstrate poor immunogenicity and early disease recurrence.* Colorectal Disease, 2009. **11**(6): p. 601-608.
19. Pancione, M., A. Remo, and V. Colantuoni, *Genetic and epigenetic events generate multiple pathways in colorectal cancer progression.* Pathology research international, 2012. **2012**.
20. Muleris, M., et al., *Chromosomal instability in near-diploid colorectal cancer: a link between numbers and structure.* PLoS One, 2008. **3**(2): p. e1632.
21. McGranahan, N., et al., *Cancer chromosomal instability: therapeutic and diagnostic challenges.* EMBO reports, 2012. **13**(6): p. 528-538.

22. Mihaylov, I.S., et al., *Control of DNA replication and chromosome ploidy by geminin and cyclin A.* Molecular and cellular biology, 2002. **22**(6): p. 1868-1880.
23. Lengauer, C., K.W. Kinzler, and B. Vogelstein, *Genetic instabilities in human cancers.* Nature, 1998. **396**(6712): p. 643-649.
24. Pritchard, C.C. and W.M. Grady, *Colorectal cancer molecular biology moves into clinical practice.* Gut, 2010: p. gut. 2009.206250.
25. Fearon, E.R., *Molecular genetics of colorectal cancer.* Annual Review of Pathology: Mechanisms of Disease, 2011. **6**: p. 479-507.
26. van Wezel, T., et al., *A review of the genetic background and tumour profiling in familial colorectal cancer.* Mutagenesis, 2012. **27**(2): p. 239-245.
27. Bodmer, W., et al., *Localization of the gene for familial adenomatous polyposis on chromosome 5.* 1987.
28. Preisinger, A.C., et al., *Identification of FAP locus genes from chromosome 5q21.* Science, 1991. **253**: p. 661-5.
29. Goss, K.H. and J. Groden, *Biology of the adenomatous polyposis coli tumor suppressor.* Journal of Clinical Oncology, 2000. **18**(9): p. 1967-1979.
30. Hart, M.J., et al., *Downregulation of β-catenin by human Axin and its association with the APC tumor suppressor, β-catenin and GSK3β.* Current Biology, 1998. **8**(10): p. 573-581.
31. Galiatsatos, P. and W.D. Foulkes, *Familial adenomatous polyposis.* The American journal of gastroenterology, 2006. **101**(2): p. 385-398.
32. Vogelstein, B. and K.W. Kinzler, *The genetic basis of human cancer*. 2002: McGraw-Hill.
33. Jemal, A., et al., *Cancer occurrence.* Cancer Epidemiology, 2009: p. 3-29.
34. Sinicrope, F.A., et al., *DNA mismatch repair status and colon cancer recurrence and survival in clinical trials of 5-fluorouracil-based adjuvant therapy.* Journal of the National Cancer Institute, 2011. **103**(11): p. 863-875.
35. Jass, J., et al., *Morphology of sporadic colorectal cancer with DNA replication errors.* Gut, 1998. **42**(5): p. 673-679.
36. Hampel, H., et al., *Feasibility of screening for Lynch syndrome among patients with colorectal cancer.* Journal of Clinical Oncology, 2008. **26**(35): p. 5783-5788.
37. Barault, L., et al., *Hypermethylator phenotype in sporadic colon cancer: study on a population-based series of 582 cases.* Cancer Research, 2008. **68**(20): p. 8541-8546.
38. Kinzler, K.W. and B. Vogelstein, *Lessons from hereditary colorectal cancer.* Cell, 1996. **87**(2): p. 159-170.
39. Nakamura, Y., *The role of the adenomatous polyposis coli (APC) gene in human cancers.* Advances in cancer research, 1993. **62**: p. 65-87.
40. Lynch, H.T., et al., *Hereditary colorectal cancer syndromes: molecular genetics, genetic counseling, diagnosis and management.* Familial cancer, 2008. **7**(1): p. 27-39.
41. Fearon, E.R. and B. Vogelstein, *A genetic model for colorectal tumorigenesis.* Cell, 1990. **61**(5): p. 759-767.
42. Polakis, P., *The many ways of Wnt in cancer.* Current opinion in genetics & development, 2007. **17**(1): p. 45-51.
43. Malumbres, M. and M. Barbacid, *RAS oncogenes: the first 30 years.* Nature Reviews Cancer, 2003. **3**(6): p. 459-465.
44. Hisamuddin, I.M. and V.W. Yang, *Molecular genetics of colorectal cancer: an overview.* Current colorectal cancer reports, 2006. **2**(2): p. 53-59.

45. Seth, R., et al., *Concomitant mutations and splice variants in KRAS and BRAF demonstrate complex perturbation of the Ras/Raf signalling pathway in advanced colorectal cancer.* Gut, 2009. **58**(9): p. 1234-1241.
46. Tan, C. and X. Du, *Du X. KRAS mutation testing in metastatic colorectal cancer.* World J Gastroenterol, 2012. **18**(37): p. 5171-5180.
47. Jiang, Y., et al., *Assessment of K-ras mutation.* Cancer, 2009. **115**(16): p. 3609-3617.
48. Goel, A. and F. Balaguer, *Serrated Pathway to Colorectal Carcinogenesis: A Molecular Perspective.* Current Colorectal Cancer Reports, 2011. **7**(1): p. 50-57.
49. Macedo, M.P., et al., *Multiple mutations in the Kras gene in colorectal cancer: review of the literature with two case reports.* International journal of colorectal disease, 2011. **26**(10): p. 1241-1248.
50. Mäkinen, M., *Colorectal serrated adenocarcinoma.* Histopathology, 2007. **50**(1): p. 131-150.
51. Guan, R.J., et al., *Association of K-ras mutations with p16 methylation in human colon cancer.* Gastroenterology, 1999. **116**(5): p. 1063-1071.
52. Richman, S.D., et al., *How close are we to standardised extended RAS gene mutation testing? The UK NEQAS evaluation.* Journal of clinical pathology, 2016: p. jclinpath-2016-203822.
53. Andreyev, H.J.N., et al., *Kirsten ras mutations in patients with colorectal cancer: the multicenter "RASCAL" study.* Journal of the National Cancer Institute, 1998. **90**(9): p. 675-684.
54. Andreyev, H., et al., *Kirsten ras mutations in patients with colorectal cancer: the 'RASCAL II'study.* British journal of cancer, 2001. **85**(5): p. 692.
55. Kerr, R.S., et al., *Adjuvant capecitabine plus bevacizumab versus capecitabine alone in patients with colorectal cancer (QUASAR 2): an open-label, randomised phase 3 trial.* The Lancet Oncology, 2016. **17**(11): p. 1543-1557.
56. Roth, A.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, 2010. **28**(3): p. 466-474.
57. Ogino, S., et al., *KRAS mutation in stage III colon cancer and clinical outcome following intergroup trial CALGB 89803.* Clinical cancer research, 2009. **15**(23): p. 7322-7329.
58. Lievre, A., et al., *KRAS mutation status is predictive of response to cetuximab therapy in colorectal cancer.* Cancer research, 2006. **66**(8): p. 3992-3995.
59. Ford, S.K., et al., *Expression of epiregulin and amphiregulin and K-RAS mutation status predict disease control in metastatic colorectal cancer patients treated with cetuximab (Erbitux®).* Cancer Research, 2007. **67**(9 Supplement): p. 5670-5670.
60. Zhang, J., T.M. Roberts, and R.A. Shivdasani, *Targeting PI3K signaling as a therapeutic approach for colorectal cancer.* Gastroenterology, 2011. **141**(1): p. 50-61.
61. Johnson, S.M., et al., *Novel expression patterns of PI3K/Akt/mTOR signaling pathway components in colorectal cancer.* Journal of the American College of Surgeons, 2010. **210**(5): p. 767-776.
62. Zhao, L. and P.K. Vogt, *Class I PI3K in oncogenic cellular transformation.* Oncogene, 2008. **27**(41): p. 5486-5496.
63. Velho, S., et al., *BRAF, KRAS and PIK3CA mutations in colorectal serrated polyps and cancer: primary or secondary genetic events in colorectal carcinogenesis?* BMC cancer, 2008. **8**(1): p. 1.
64. De Roock, W., et al., *KRAS, BRAF, PIK3CA, and PTEN mutations: implications for targeted therapies in metastatic colorectal cancer.* The lancet oncology, 2011. **12**(6): p. 594-603.

65. Samuels, Y., et al., *High frequency of mutations of the PIK3CA gene in human cancers.* Science, 2004. **304**(5670): p. 554-554.
66. Zhao, L. and P.K. Vogt, *Helical domain and kinase domain mutations in p110α of phosphatidylinositol 3-kinase induce gain of function by different mechanisms.* Proceedings of the National Academy of Sciences, 2008. **105**(7): p. 2652-2657.
67. Samuels, Y. and V.E. Velculescu, *Oncogenic mutations of PIK3CA in human cancers.* Cell cycle, 2004. **3**(10): p. 1221-1224.
68. De Roock, W., et al., *Effects of KRAS, BRAF, NRAS, and PIK3CA mutations on the efficacy of cetuximab plus chemotherapy in chemotherapy-refractory metastatic colorectal cancer: a retrospective consortium analysis.* The lancet oncology, 2010. **11**(8): p. 753-762.
69. Liao, X., et al., *Aspirin use, tumor PIK3CA mutation, and colorectal-cancer survival.* New England Journal of Medicine, 2012. **367**(17): p. 1596-1606.
70. Keniry, M. and R. Parsons, *The role of PTEN signaling perturbations in cancer and in targeted therapy.* Oncogene, 2008. **27**(41): p. 5477-5485.
71. Chalhoub, N. and S.J. Baker, *PTEN and the PI3-kinase pathway in cancer.* Annual review of pathology, 2009. **4**: p. 127.
72. Laurent-Puig, P., et al., *Analysis of PTEN, BRAF, and EGFR status in determining benefit from cetuximab therapy in wild-type KRAS metastatic colon cancer.* Journal of Clinical Oncology, 2009. **27**(35): p. 5924-5930.
73. Tabach, Y., et al., *Amplification of the 20q chromosomal arm occurs early in tumorigenic transformation and may initiate cancer.* PLoS One, 2011. **6**(1): p. e14632.
74. Carvalho, B., et al., *Multiple putative oncogenes at the chromosome 20q amplicon contribute to colorectal adenoma to carcinoma progression.* Gut, 2009. **58**(1): p. 79-89.
75. Meijer, G.A., et al., *Progression from colorectal adenoma to carcinoma is associated with non-random chromosomal gains as detected by comparative genomic hybridisation.* Journal of clinical pathology, 1998. **51**(12): p. 901-909.
76. De Angelis, P., et al., *Chromosomal gains and losses in primary colorectal carcinomas detected by CGH and their associations with tumour DNA ploidy, genotypes and phenotypes.* British journal of cancer, 1999. **80**(3-4): p. 526.
77. Sillars-Hardebol, A.H., et al., *CSE1L, DIDO1 and RBM39 in colorectal adenoma to carcinoma progression.* Cellular Oncology, 2012. **35**(4): p. 293-300.
78. Kollareddy, M., et al., *Aurora kinases: structure, functions and their association with cancer.* Biomed Pap Med Fac Univ Palacky Olomouc Czech Repub, 2008. **152**(1): p. 27-33.
79. Wu, J.-C., et al., *Identification of V23RalA-Ser194 as a critical mediator for Aurora-A-induced cellular motility and transformation by small pool expression screening.* Journal of Biological Chemistry, 2005. **280**(10): p. 9013-9022.
80. Sillars-Hardebol, A.H., et al., *TPX2 and AURKA promote 20q amplicon-driven colorectal adenoma to carcinoma progression.* Gut, 2012. **61**(11): p. 1568-1575.
81. Alnabulsi, A., et al., *Cellular apoptosis susceptibility (chromosome segregation 1-like, CSE1L) gene is a key regulator of apoptosis, migration and invasion in colorectal cancer.* The Journal of pathology, 2012. **228**(4): p. 471-481.
82. Sillars-Hardebol, A.H., et al., *BCL2L1 has a functional role in colorectal cancer and its protein expression is associated with chromosome 20q gain.* The Journal of pathology, 2012. **226**(3): p. 442-450.

83. Welcker, M. and B.E. Clurman, *FBW7 ubiquitin ligase: a tumour suppressor at the crossroads of cell division, growth and differentiation.* Nature Reviews Cancer, 2008. **8**(2): p. 83-93.
84. Minella, A.C. and B.E. Clurman, *Mechanisms of tumor suppression by the SCFFbw7.* Cell cycle, 2005. **4**(10): p. 1356-1359.
85. Lau, A.W., H. Fukushima, and W. Wei, *The Fbw7 and beta-TRCP E3 ubiquitin ligases and their roles in tumorigenesis.* Frontiers in bioscience: a journal and virtual library, 2012. **17**: p. 2197.
86. Wang, Z., et al., *Tumor suppressor functions of FBW7 in cancer development and progression.* FEBS letters, 2012. **586**(10): p. 1409-1418.
87. Naganawa, Y., et al., *Decreased expression of FBXW7 is correlated with poor prognosis in patients with esophageal squamous cell carcinoma.* Experimental and therapeutic medicine, 2010. **1**(5): p. 841-846.
88. Li, N., et al., *FBXW7-mutated colorectal cancer cells exhibit aberrant expression of phosphorylated-p53 at serine-15.* Oncotarget, 2015. **6**(11): p. 9240.
89. Mao, J.-H., et al., *Fbxw7/Cdc4 is a p53-dependent, haploinsufficient tumour suppressor gene.* Nature, 2004. **432**(7018): p. 775-779.
90. Grim, J.E., et al., *Fbw7 and p53 cooperatively suppress advanced and chromosomally unstable intestinal cancer.* Molecular and cellular biology, 2012. **32**(11): p. 2160-2167.
91. Mansour, M.R., et al., *Prognostic implications of NOTCH1 and FBXW7 mutations in adults with T-cell acute lymphoblastic leukemia treated on the MRC UKALLXII/ECOG E2993 protocol.* Journal of Clinical Oncology, 2009. **27**(26): p. 4352-4356.
92. Koh, M.S., et al., *CDC4 gene expression as potential biomarker for targeted therapy in prostate cancer.* Cancer biology & therapy, 2006. **5**(1): p. 78-83.
93. Iwatsuki, M., et al., *Loss of FBXW7, a cell cycle regulating gene, in colorectal cancer: clinical significance.* International Journal of Cancer, 2010. **126**(8): p. 1828-1837.
94. Hagedorn, M., et al., *FBXW7/hCDC4 controls glioma cell proliferation in vitro and is a prognostic marker for survival in glioblastoma patients.* Cell division, 2007. **2**(1): p. 1.
95. Kemp, Z., et al., *CDC4 mutations occur in a subset of colorectal cancers but are not predicted to cause loss of function and are not associated with chromosomal instability.* Cancer research, 2005. **65**(24): p. 11361-11366.
96. Rajagopalan, H., et al., *Inactivation of hCDC4 can cause chromosomal instability.* Nature, 2004. **428**(6978): p. 77-81.
97. Akhoondi, S., et al., *FBXW7/hCDC4 is a general tumor suppressor in human cancer.* Cancer research, 2007. **67**(19): p. 9006-9012.
98. Mao, J.-H., et al., *FBXW7 targets mTOR for degradation and cooperates with PTEN in tumor suppression.* Science, 2008. **321**(5895): p. 1499-1502.
99. Wertz, I.E., et al., *Sensitivity to antitubulin chemotherapeutics is regulated by MCL1 and FBW7.* Nature, 2011. **471**(7336): p. 110-114.
100. Inuzuka, H., et al., *SCFFBW7 regulates cellular apoptosis by targeting MCL1 for ubiquitylation and destruction.* Nature, 2011. **471**(7336): p. 104-109.
101. Wang, Y., et al., *Rapamycin inhibits FBXW7 loss-induced epithelial–mesenchymal transition and cancer stem cell-like characteristics in colorectal cancer cells.* Biochemical and biophysical research communications, 2013. **434**(2): p. 352-356.
102. Cai, Z., et al., *APC, FBXW7, KRAS, PIK3CA, and TP53 gene mutations in human colorectal cancer tumors frequently detected by next-generation DNA sequencing.* J Mol Genet Med, 2014. **8**(145): p. 1747-0862.1000145.

103. Jardim, D.L., et al., *FBXW7 mutations in patients with advanced cancers: clinical and molecular characteristics and outcomes with mTOR inhibitors.* PloS one, 2014. **9**(2): p. e89388.
104. Zheng, H.-T., et al., *Loss of heterozygosity analyzed by single nucleotide polymorphism array in cancer.* World journal of gastroenterology, 2005. **11**(43): p. 6740-6744.
105. Lindblad-Toh, K., et al., *Loss-of-heterozygosity analysis of small-cell lung carcinomas using single-nucleotide polymorphism arrays.* Nature biotechnology, 2000. **18**(9): p. 1001-1005.
106. Milbury, C.A., et al., *COLD-PCR: improving the sensitivity of molecular diagnostics assays.* Expert review of molecular diagnostics, 2011. **11**(2): p. 159-169.
107. Alexandrov, L.B., et al., *Signatures of mutational processes in human cancer.* Nature, 2013. **500**(7463): p. 415-421.
108. Jia, P., W. Pao, and Z. Zhao, *Patterns and processes of somatic mutations in nine major cancers.* BMC medical genomics, 2014. **7**(1): p. 1.
109. Piccolo, S.R. and L.J. Frey. *Somatic mutation signatures of cancer*. in *AMIA*. 2008.
110. Roper, J. and K.E. Hung, *Molecular mechanisms of colorectal carcinogenesis*, in *Molecular Pathogenesis of Colorectal Cancer*. 2013, Springer. p. 25-65.
111. Mehlen, P. and E.R. Fearon, *Role of the dependence receptor DCC in colorectal cancer pathogenesis.* Journal of clinical oncology, 2004. **22**(16): p. 3420-3428.
112. Popat, S., et al., *Relationship between chromosome 18q status and colorectal cancer prognosis: a prospective, blinded analysis of 280 patients.* Anticancer research, 2007. **27**(1B): p. 627-633.
113. Sameer, A.S., et al., *Colorectal cancer, TGF-signaling and SMADs.* International Journal of Genetics and Molecular Biology, 2010. **2**(6): p. 101-111.
114. Brodin, G., et al., *Efficient TGF-β induction of the Smad7 gene requires cooperation between AP-1, Sp1, and Smad proteins on the mouse Smad7 promoter.* Journal of Biological Chemistry, 2000. **275**(37): p. 29023-29030.
115. von Gersdorff, G., et al., *Smad3 and Smad4 mediate transcriptional activation of the human Smad7 promoter by transforming growth factor β.* Journal of Biological Chemistry, 2000. **275**(15): p. 11320-11326.
116. Feng, X.H., X. Lin, and R. Derynck, *Smad2, Smad3 and Smad4 cooperate with Sp1 to induce p15Ink4B transcription in response to TGF-β.* The EMBO journal, 2000. **19**(19): p. 5178-5193.
117. de Caestecker, M.P., E. Piek, and A.B. Roberts, *Role of transforming growth factor-β signaling in cancer.* Journal of the National Cancer Institute, 2000. **92**(17): p. 1388-1402.
118. Maitra, A., et al., *Loss of Dpc4 expression in colonic adenocarcinomas correlates with the presence of metastatic disease.* The American journal of pathology, 2000. **157**(4): p. 1105-1111.
119. Howe, J.R., et al., *Mutations in the SMAD4/DPC4 gene in juvenile polyposis.* Science, 1998. **280**(5366): p. 1086-1088.
120. Houlston, R., et al., *Mutations in DPC4 (SMAD4) cause juvenile polyposis syndrome, but only account for a minority of cases.* Human molecular genetics, 1998. **7**(12): p. 1907-1912.
121. Calistri, D., et al., *KRAS, p53 and BRAF gene mutations and aneuploidy in sporadic colorectal cancer progression.* Analytical Cellular Pathology, 2006. **28**(4): p. 161-166.
122. Tasdemir, E., et al., *A dual role of p53 in the control of autophagy.* Autophagy, 2008. **4**(6): p. 810-814.

123. Du, C.T.a.X., *KRAS mutation testing in metastatic colorectal cancer.* World J Gastroenterol, 2012. **18**(37): p. 10.
124. Iacopetta, B., *TP53 mutation in colorectal cancer.* Human mutation, 2003. **21**(3): p. 271-276.
125. Muneer, A., et al., *Molecular prognostic factors in penile cancer.* World journal of urology, 2009. **27**(2): p. 161-167.
126. Goh, H.-S., J. Yao, and D.R. Smith, *p53 point mutation and survival in colorectal cancer patients.* Cancer research, 1995. **55**(22): p. 5217-5221.
127. Hoff, P.M., *Is there a role for routine p53 testing in colorectal cancer?* Journal of clinical oncology, 2005. **23**(30): p. 7395-7396.
128. Ghadimi, B.M., et al., *Gain of chromosome 8q23–24 is a predictive marker for lymph node positivity in colorectal cancer.* Clinical cancer research, 2003. **9**(5): p. 1808-1814.
129. Camps, J., et al., *Chromosomal breakpoints in primary colon cancer cluster at sites of structural variants in the genome.* Cancer research, 2008. **68**(5): p. 1284-1295.
130. Saha, S., et al., *A phosphatase associated with metastasis of colorectal cancer.* Science, 2001. **294**(5545): p. 1343-1346.
131. Gardner, L., L. Lee, and C. Dang, *The c-Myc oncogenic transcription factor.* Encyclopedia of Cancer, 2002. **2**: p. 2002.
132. Zanke, B.W., et al., *Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24.* Nature genetics, 2007. **39**(8): p. 989-994.
133. Choi, S.-W., et al., *Genetic classification of colorectal cancer based on chromosomal loss and microsatellite instability predicts survival.* Clinical cancer research, 2002. **8**(7): p. 2311-2322.
134. Vilar, E. and J. Tabernero, *Molecular dissection of microsatellite instable colorectal cancer.* Cancer discovery, 2013. **3**(5): p. 502-511.
135. Ellegren, H., *Microsatellites: simple sequences with complex evolution.* Nature reviews genetics, 2004. **5**(6): p. 435-445.
136. Poynter, J.N., et al., *Molecular characterization of MSI-H colorectal cancer by MLHI promoter methylation, immunohistochemistry, and mismatch repair germline mutation screening.* Cancer Epidemiology Biomarkers & Prevention, 2008. **17**(11): p. 3208-3215.
137. Yang, V.W., *The molecular genetics of colorectal cancer.* Current gastroenterology reports, 1999. **1**(5): p. 449-454.
138. Duval, A. and R. Hamelin, *Mutations at Coding Repeat Sequences in Mismatch Repair-deficient Human Cancers Toward a New Concept of Target Genes for Instability.* Cancer research, 2002. **62**(9): p. 2447-2454.
139. Ward, R., et al., *Microsatellite instability and the clinicopathological features of sporadic colorectal cancer.* Gut, 2001. **48**(6): p. 821-829.
140. Domingo, E., et al., *Use of multivariate analysis to suggest a new molecular classification of colorectal cancer.* The Journal of pathology, 2013. **229**(3): p. 441-448.
141. Murphy, K.M., et al., *Comparison of the microsatellite instability analysis system and the Bethesda panel for the determination of microsatellite instability in colorectal cancers.* The Journal of Molecular Diagnostics, 2006. **8**(3): p. 305-311.
142. Lin, C.-H., et al., *Molecular profile and copy number analysis of sporadic colorectal cancer in Taiwan.* Journal of biomedical science, 2011. **18**(1): p. 1.
143. Halford, S., et al., *Low-level microsatellite instability occurs in most colorectal cancers and is a nonrandomly distributed quantitative trait.* Cancer research, 2002. **62**(1): p. 53-57.

144. Goel, A., et al., *Characterization of sporadic colon cancer by patterns of genomic instability.* Cancer research, 2003. **63**(7): p. 1608-1614.
145. Whitehall, V.L., et al., *Methylation of O–6-methylguanine DNA methyltransferase characterizes a subset of colorectal cancer with low-level DNA microsatellite instability.* Cancer research, 2001. **61**(3): p. 827-830.
146. Jass, J.R., et al., *Emerging concepts in colorectal neoplasia.* Gastroenterology, 2002. **123**(3): p. 862-876.
147. Hagland, H.R., et al., *Molecular pathways and cellular metabolism in colorectal cancer.* Digestive surgery, 2013. **30**(1): p. 12-25.
148. Christmann, M., et al., *Mechanisms of human DNA repair: an update.* Toxicology, 2003. **193**(1): p. 3-34.
149. Martin, A. and M.D. Scharff, *AID and mismatch repair in antibody diversification.* Nature Reviews Immunology, 2002. **2**(8): p. 605-614.
150. Nandan, M.O. and V.W. Yang, *An update on the biology of RAS/RAF mutations in colorectal cancer.* Current colorectal cancer reports, 2011. **7**(2): p. 113-120.
151. Tejpar, S., et al., *Prognostic and predictive biomarkers in resected colon cancer: current status and future perspectives for integrating genomics into biomarker discovery.* The oncologist, 2010. **15**(4): p. 390-404.
152. Li, W.Q., et al., *BRAF mutations are associated with distinctive clinical, pathological and molecular features of colorectal cancer independently of microsatellite instability status.* Molecular cancer, 2006. **5**(1): p. 1.
153. Deng, G., et al., *BRAF mutation is frequently present in sporadic colorectal cancer with methylated hMLH1, but not in hereditary nonpolyposis colorectal cancer.* Clinical Cancer Research, 2004. **10**(1): p. 191-195.
154. Markowitz, S.D. and M.M. Bertagnolli, *Molecular basis of colorectal cancer.* New England Journal of Medicine, 2009. **361**(25): p. 2449-2460.
155. Di Nicolantonio, F., et al., *Wild-type BRAF is required for response to panitumumab or cetuximab in metastatic colorectal cancer.* Journal of Clinical Oncology, 2008. **26**(35): p. 5705-5712.
156. Chapman, P.B., et al., *Improved survival with vemurafenib in melanoma with BRAF V600E mutation.* n Engl J Med, 2011. **2011**(364): p. 2507-2516.
157. Kopetz, S., et al., *Phase II pilot study of vemurafenib in patients with metastatic BRAF-mutated colorectal cancer.* Journal of Clinical Oncology, 2015. **33**(34): p. 4032-4038.
158. Ozawa, M., M. Ringwald, and R. Kemler, *Uvomorulin-catenin complex formation is regulated by a specific domain in the cytoplasmic region of the cell adhesion molecule.* Proceedings of the National Academy of Sciences, 1990. **87**(11): p. 4246-4250.
159. Funayama, N., et al., *Embryonic axis induction by the armadillo repeat domain of beta-catenin: evidence for intracellular signaling.* The Journal of Cell Biology, 1995. **128**(5): p. 959-968.
160. Kaler, P., L. Augenlicht, and L. Klampfer, *Macrophage-derived IL-1β stimulates Wnt signaling and growth of colon cancer cells: a crosstalk interrupted by vitamin D3.* Oncogene, 2009. **28**(44): p. 3892-3902.
161. Sekine, S., et al., *Target disruption of the mutant beta-catenin gene in colon cancer cell line HCT116: preservation of its malignant phenotype.* Oncogene, 2002. **21**(38): p. 5906-5911.
162. Bachelder, R.E., et al., *Glycogen synthase kinase-3 is an endogenous inhibitor of Snail transcription implications for the epithelial–mesenchymal transition.* The Journal of cell biology, 2005. **168**(1): p. 29-33.

163. Kaler, P., et al., *Tumor associated macrophages protect colon cancer cells from TRAIL-induced apoptosis through IL-1β-dependent stabilization of Snail in tumor cells.* PLoS One, 2010. **5**(7): p. e11700.
164. Kaler, P., et al., *The NF-κB/AKT-dependent induction of Wnt signaling in colon cancer cells by macrophages and IL-1β.* Cancer Microenvironment, 2009. **2**(1): p. 69-80.
165. Polyak, K. and R.A. Weinberg, *Transitions between epithelial and mesenchymal states: acquisition of malignant and stem cell traits.* Nature Reviews Cancer, 2009. **9**(4): p. 265-273.
166. Korinek, V., et al., *Constitutive transcriptional activation by a β-catenin-Tcf complex in APC−/− colon carcinoma.* Science, 1997. **275**(5307): p. 1784-1787.
167. Morin, P.J., et al., *Activation of β-catenin-Tcf signaling in colon cancer by mutations in β-catenin or APC.* Science, 1997. **275**(5307): p. 1787-1790.
168. Ilyas, M., et al., *β-Catenin mutations in cell lines established from human colorectal cancers.* Proceedings of the National Academy of Sciences, 1997. **94**(19): p. 10330-10334.
169. Samowitz, W.S., et al., *β-catenin mutations are more frequent in small colorectal adenomas than in larger adenomas and invasive carcinomas.* Cancer research, 1999. **59**(7): p. 1442-1444.
170. Shitoh, K., et al., *Frequent activation of the β-catenin-Tcf signaling pathway in nonfamilial colorectal carcinomas with microsatellite instability.* Genes, Chromosomes and Cancer, 2001. **30**(1): p. 32-37.
171. Miyaki, M., et al., *Frequent mutation of β-catenin and APC genes in primary colorectal tumors from patients with hereditary nonpolyposis colorectal cancer.* Cancer research, 1999. **59**(18): p. 4506-4509.
172. Kawasaki, T., et al., *Correlation of β-catenin localization with cyclooxygenase-2 expression and CpG island methylator phenotype (CIMP) in colorectal cancer.* Neoplasia, 2007. **9**(7): p. 569-577.
173. Jass, J., *Classification of colorectal cancer based on correlation of clinical, morphological and molecular features.* Histopathology, 2007. **50**(1): p. 113-130.
174. Jo, Y.S., et al., *Frequent frameshift mutations in 2 mononucleotide repeats of RNF43 gene and its regional heterogeneity in gastric and colorectal cancers.* Human pathology, 2015. **46**(11): p. 1640-1646.
175. Zebisch, M., et al., *Structural and molecular basis of ZNRF3/RNF43 transmembrane ubiquitin ligase inhibition by the Wnt agonist R-spondin.* Nature communications, 2013. **4**.
176. Bruens, L., *The role of receptor-associated E3 ligases in WNT signaling and disease.* 2014.
177. Koo, B.-K., et al., *Tumour suppressor RNF43 is a stem-cell E3 ligase that induces endocytosis of Wnt receptors.* Nature, 2012. **488**(7413): p. 665-669.
178. Giannakis, M., et al., *RNF43 is frequently mutated in colorectal and endometrial cancers.* Nature genetics, 2014. **46**(12): p. 1264-1266.
179. Wang, K., et al., *Whole-genome sequencing and comprehensive molecular profiling identify new driver mutations in gastric cancer.* Nature genetics, 2014. **46**(6): p. 573-582.
180. Bond, C.E., et al., *RNF43 and ZNRF3 are commonly altered in serrated pathway colorectal tumorigenesis.* Oncotarget, 2016.
181. Sekine, S., et al., *Frequent PTPRK–RSPO3 fusions and RNF43 mutations in colorectal traditional serrated adenoma.* The Journal of pathology, 2016. **239**(2): p. 133-138.

182. Hao, H.-X., X. Jiang, and F. Cong, *Control of Wnt Receptor Turnover by R-spondin-ZNRF3/RNF43 Signaling Module and Its Dysregulation in Cancer.* Cancers, 2016. **8**(6): p. 54.
183. Massagué, J., S.W. Blain, and R.S. Lo, *TGFβ signaling in growth control, cancer, and heritable disorders.* Cell, 2000. **103**(2): p. 295-309.
184. Lièvre, A., et al., *KRAS mutations as an independent prognostic factor in patients with advanced colorectal cancer treated with cetuximab.* Journal of Clinical Oncology, 2008. **26**(3): p. 374-379.
185. Markowitz, S.D. and A.B. Roberts, *Tumor suppressor activity of the TGF-β pathway in human cancers.* Cytokine & growth factor reviews, 1996. **7**(1): p. 93-102.
186. Fynan, T. and M. Reiss, *Resistance to inhibition of cell growth by transforming growth factor-beta and its role in oncogenesis.* Critical reviews in oncogenesis, 1992. **4**(5): p. 493-540.
187. Massagué, J., *TGFβ signaling: receptors, transducers, and Mad proteins.* Cell, 1996. **85**(7): p. 947-950.
188. Nakao, A., et al., *Identification of Smad7, a TGFβ-inducible antagonist of TGF-β signalling.* Nature, 1997. **389**(6651): p. 631-635.
189. Hua, X., et al., *Specificity in transforming growth factor β-induced transcription of the plasminogen activator inhibitor-1 gene: Interactions of promoter DNA, transcription factor μE3, and Smad proteins.* Proceedings of the National Academy of Sciences, 1999. **96**(23): p. 13130-13135.
190. Elliott, R.L. and G.C. Blobe, *Role of transforming growth factor Beta in human cancer.* Journal of Clinical Oncology, 2005. **23**(9): p. 2078-2093.
191. Akhurst, R.J., *TGFbeta signaling in health and disease.* Nature genetics, 2004. **36**(8): p. 790-792.
192. Biswas, S., et al., *Transforming growth factor β receptor type II inactivation promotes the establishment and progression of colon cancer.* Cancer research, 2004. **64**(14): p. 4687-4692.
193. Wakefield, L.M. and A.B. Roberts, *TGF-β signaling: positive and negative effects on tumorigenesis.* Current opinion in genetics & development, 2002. **12**(1): p. 22-29.
194. Yang, Y.-a., et al., *Lifetime exposure to a soluble TGF-β antagonist protects mice against metastasis without adverse side effects.* The Journal of clinical investigation, 2002. **109**(12): p. 1607-1615.
195. Hahm, K.B., et al., *Conditional loss of TGF-β signalling leads to increased susceptibility to gastrointestinal carcinogenesis in mice.* Alimentary pharmacology & therapeutics, 2002. **16**(s2): p. 115-127.
196. Tang, B., et al., *Transforming growth factor- 1 is a new form of tumor suppressor with true haploid insufficiency.* Nature medicine, 1998. **4**(7): p. 802-807.
197. Thiery, J.P., *Epithelial–mesenchymal transitions in tumour progression.* Nature Reviews Cancer, 2002. **2**(6): p. 442-454.
198. Watanabe, T., et al., *Molecular predictors of survival after adjuvant chemotherapy for colon cancer.* New England Journal of Medicine, 2001. **344**(16): p. 1196-1206.
199. Cao, S.-n., et al., *Frameshift mutations in the Bax gene are not involved in development of ovarian endometrioid carcinoma.* Modern pathology, 2003. **16**(10): p. 1048-1052.
200. Goping, S., et al., *Regulated targeting of BAX to mitochondria.* The Journal of cell biology, 1998. **143**(1): p. 207-215.
201. Marzo, I., et al., *Bax and adenine nucleotide translocator cooperate in the mitochondrial control of apoptosis.* Science, 1998. **281**(5385): p. 2027-2031.

202. Miquel, C., et al., *Role of bax mutations in apoptosis in colorectal cancers with microsatellite instability.* American journal of clinical pathology, 2005. **123**(4): p. 562-570.
203. Zha, H., et al., *Proapoptotic protein Bax heterodimerizes with Bcl-2 and homodimerizes with Bax via a novel domain (BH3) distinct from BH1 and BH2.* Journal of Biological Chemistry, 1996. **271**(13): p. 7440-7444.
204. Hanada, M., et al., *Structure-function analysis of Bcl-2 protein identification of conserved domains important for homodimerization with Bcl-2 and heterodimerization with Bax.* Journal of Biological Chemistry, 1995. **270**(20): p. 11962-11969.
205. Iacopetta, B., F. Grieu, and B. Amanuel, *Microsatellite instability in colorectal cancer.* Asia-Pacific Journal of Clinical Oncology, 2010. **6**(4): p. 260-269.
206. Rampino, N., et al., *Somatic frameshift mutations in the BAX gene in colon cancers of the microsatellite mutator phenotype.* Science, 1997. **275**(5302): p. 967-969.
207. Shin, K.-H., Y.J. Park, and J.-G. Park, *PTEN gene mutations in colorectal cancers displaying microsatellite instability.* Cancer letters, 2001. **174**(2): p. 189-194.
208. Pryczynicz, A., et al., *Bax protein may influence the invasion of colorectal cancer.* World Journal of Gastroenterology: WJG, 2014. **20**(5): p. 1305.
209. Shima, K., et al., *TGFBR2 and BAX mononucleotide tract mutations, microsatellite instability, and prognosis in 1072 colorectal cancers.* PLoS One, 2011. **6**(9): p. e25062.
210. Ionov, Y., et al., *Mutational inactivation of the proapoptotic gene BAX confers selective advantage during tumor clonal evolution.* Proceedings of the National Academy of Sciences, 2000. **97**(20): p. 10872-10877.
211. Zhang, L., et al., *Role of BAX in the apoptotic response to anticancer agents.* Science, 2000. **290**(5493): p. 989-992.
212. Zhou, Q., et al., *Mutations/Polymorphisms in the 55 kDa Subunit of DNA Polymerase ϵ in Human Colorectal Cancer.* Cancer Genomics-Proteomics, 2009. **6**(6): p. 297-304.
213. Tahirov, T.H., et al., *Evolution of DNA polymerases: an inactivated polymerase-exonuclease module in Pol ε and a chimeric origin of eukaryotic polymerases from two classes of archaeal ancestors.* Biology direct, 2009. **4**(1): p. 1.
214. Albertson, T.M., et al., *DNA polymerase ε and δ proofreading suppress discrete mutator and cancer phenotypes in mice.* Proceedings of the National Academy of Sciences, 2009. **106**(40): p. 17101-17104.
215. Church, D.N., et al., *DNA polymerase ε and δ exonuclease domain mutations in endometrial cancer.* Human molecular genetics, 2013. **22**(14): p. 2820-2828.
216. Preston, B.D., T.M. Albertson, and A.J. Herr. *DNA replication fidelity and cancer*. in *Seminars in cancer biology*. 2010. Elsevier.
217. Hübscher, U., G. Maga, and S. Spadari, *Eukaryotic DNA polymerases.* Annual review of biochemistry, 2002. **71**(1): p. 133-163.
218. Tran, H.T., D.A. Gordenin, and M.A. Resnick, *The 3'→ 5' exonucleases of DNA polymerases δ and ε and the 5'→ 3' exonuclease Exo1 have major roles in postreplication mutation avoidance in Saccharomyces cerevisiae.* Molecular and cellular biology, 1999. **19**(3): p. 2000-2007.
219. Kirchner, J., H. Tran, and M. Resnick, *A DNA polymerase ε mutant that specifically causes+ 1 frameshift mutations within homonucleotide runs in yeast.* Genetics, 2000. **155**(4): p. 1623-1632.
220. de Boer, J. and J.H. Hoeijmakers, *Nucleotide excision repair and human syndromes.* Carcinogenesis, 2000. **21**(3): p. 453-460.

221. Lange, S.S., K.-i. Takata, and R.D. Wood, *DNA polymerases and cancer.* Nature Reviews Cancer, 2011. **11**(2): p. 96-110.
222. Wood, R.D. and M. Shivji, *Which DNA polymerases are used for DNA-repair in eukaryotes?* Carcinogenesis, 1997. **18**(4): p. 605-610.
223. Hsieh, P. and K. Yamane, *DNA mismatch repair: molecular mechanism, cancer, and ageing.* Mechanisms of ageing and development, 2008. **129**(7): p. 391-407.
224. Prindle, M.J. and L.A. Loeb, *DNA polymerase delta in DNA replication and genome maintenance.* Environmental and molecular mutagenesis, 2012. **53**(9): p. 666-682.
225. Pospiech, H. and J.E. Syväoja, *DNA polymerase e-more than a polymerase.* The Scientific World Journal, 2003. **3**: p. 87-104.
226. Goldsby, R.E., et al., *High incidence of epithelial cancers in mice deficient for DNA polymerase δ proofreading.* Proceedings of the national academy of sciences, 2002. **99**(24): p. 15560-15565.
227. Palles, C., et al., *Germline mutations affecting the proofreading domains of POLE and POLD1 predispose to colorectal adenomas and carcinomas.* Nature genetics, 2013. **45**(2): p. 136-144.
228. Seshagiri, S., et al., *Recurrent R-spondin fusions in colon cancer.* Nature, 2012. **488**(7413): p. 660-664.
229. Network, C.G.A., *Comprehensive molecular characterization of human colon and rectal cancer.* Nature, 2012. **487**(7407): p. 330-337.
230. Hawkins, N., et al., *Microsatellite-stable diploid carcinoma: a biologically distinct and aggressive subset of sporadic colorectal cancer.* British journal of cancer, 2001. **84**(2): p. 232.
231. Tang, R., et al., *Colorectal cancer without high microsatellite instability and chromosomal instability—an alternative genetic pathway to human colorectal cancer.* Carcinogenesis, 2004. **25**(5): p. 841-846.
232. Kakar, S., et al., *Clinicopathologic characteristics, CpG island methylator phenotype, and BRAF mutations in microsatellite-stable colorectal cancers without chromosomal instability.* Archives of pathology & laboratory medicine, 2008. **132**(6): p. 958-964.
233. Fadhil, W., et al., *DNA content analysis of colorectal cancer defines a distinct 'microsatellite and chromosome stable'group but does not predict response to radiotherapy.* International journal of experimental pathology, 2014. **95**(1): p. 16-23.
234. Silver, A., et al., *A distinct DNA methylation profile associated with microsatellite and chromosomal stable sporadic colorectal cancers.* International Journal of Cancer, 2012. **130**(5): p. 1082-1092.
235. Breivik, J. and G. Gaudernack. *Genomic instability, DNA methylation, and natural selection in colorectal carcinogenesis*. in *Seminars in cancer biology*. 1999. Elsevier.
236. Herman, J.G., et al., *Incidence and functional consequences of hMLH1 promoter hypermethylation in colorectal carcinoma.* Proceedings of the National Academy of Sciences, 1998. **95**(12): p. 6870-6875.
237. Veigl, M.L., et al., *Biallelic inactivation of hMLH1 by epigenetic gene silencing, a novel mechanism causing human MSI cancers.* Proceedings of the National Academy of Sciences, 1998. **95**(15): p. 8698-8702.
238. van Engeland, M., et al., *Colorectal cancer epigenetics: complex simplicity.* Journal of Clinical Oncology, 2011. **29**(10): p. 1382-1391.
239. Privette, L.M., et al., *Loss of CHFR in human mammary epithelial cells causes genomic instability by disrupting the mitotic spindle assembly checkpoint.* Neoplasia, 2008. **10**(7): p. 643-652.

240. Hinoue, T., et al., *Genome-scale analysis of aberrant DNA methylation in colorectal cancer.* Genome research, 2012. **22**(2): p. 271-282.
241. Shen, L., et al., *Integrated genetic and epigenetic analysis identifies three different subclasses of colon cancer.* Proceedings of the National Academy of Sciences, 2007. **104**(47): p. 18654-18659.
242. Hawkins, N., et al., *CpG island methylation in sporadic colorectal cancers and its relationship to microsatellite instability.* Gastroenterology, 2002. **122**(5): p. 1376-1387.
243. Ogino, S., et al., *Evaluation of markers for CpG island methylator phenotype (CIMP) in colorectal cancer by a large population-based sample.* The Journal of Molecular Diagnostics, 2007. **9**(3): p. 305-314.
244. Ogino, S., et al., *Molecular correlates with MGMT promoter methylation and silencing support CpG island methylator phenotype-low (CIMP-low) in colorectal cancer.* Gut, 2007. **56**(11): p. 1564-1571.
245. Ogino, S., et al., *CpG island methylator phenotype-low (CIMP-low) in colorectal cancer: possible associations with male sex and KRAS mutations.* The Journal of molecular diagnostics, 2006. **8**(5): p. 582-588.
246. Ogino, S., et al., *18q loss of heterozygosity in microsatellite stable colorectal cancer is correlated with CpG island methylator phenotype-negative (CIMP-0) and inversely with CIMP-low and CIMP-high.* BMC cancer, 2007. **7**(1): p. 1.
247. Ang, P.W., et al., *Comprehensive profiling of DNA methylation in colorectal cancer reveals subgroups with distinct clinicopathological and molecular features.* BMC cancer, 2010. **10**(1): p. 1.
248. Kang, G.H., *Four molecular subtypes of colorectal cancer and their precursor lesions.* Archives of pathology & laboratory medicine, 2011. **135**(6): p. 698-703.
249. Houlston, R.S., et al., *Meta-analysis of genome-wide association data identifies four new susceptibility loci for colorectal cancer.* Nature genetics, 2008. **40**(12): p. 1426-1435.
250. Houlston, R.S., et al., *Meta-analysis of three genome-wide association studies identifies susceptibility loci for colorectal cancer at 1q41, 3q26. 2, 12q13. 13 and 20q13. 33.* Nature genetics, 2010. **42**(11): p. 973-977.
251. Edler, D., et al., *Common variation near CDKN1A, POLD3 and SHROOM2 influences colorectal cancer risk.* 2013.
252. Guinney, J., et al., *The consensus molecular subtypes of colorectal cancer.* Nature medicine, 2015.
253. Ashworth, A., C.J. Lord, and J.S. Reis-Filho, *Genetic interactions in cancer progression and treatment.* Cell, 2011. **145**(1): p. 30-38.
254. Gao, H., J.M. Granka, and M.W. Feldman, *On the classification of epistatic interactions.* Genetics, 2010. **184**(3): p. 827-837.
255. Cordell, H.J., *Epistasis: what it means, what it doesn't mean, and statistical methods to detect it in humans.* Human molecular genetics, 2002. **11**(20): p. 2463-2468.
256. Li, J., et al., *Oncogenic K-ras stimulates Wnt signaling in colon cancer through inhibition of GSK-3β.* Gastroenterology, 2005. **128**(7): p. 1907-1918.
257. McMurray, H.R., et al., *Synergistic response to oncogenic mutations defines gene class critical to cancer phenotype.* Nature, 2008. **453**(7198): p. 1112-1116.
258. Grivennikov, S.I. and M. Karin, *Dangerous liaisons: STAT3 and NF-κB collaboration and crosstalk in cancer.* Cytokine & growth factor reviews, 2010. **21**(1): p. 11-19.
259. Arends, J.W., *Molecular interactions in the Vogelstein model of colorectal carcinoma.* The Journal of pathology, 2000. **190**(4): p. 412-416.

260. Kanthan, R., J.-L. Senger, and S.C. Kanthan, *Molecular events in primary and metastatic colorectal carcinoma: a review.* Pathology research international, 2012. **2012**.
261. Shibayama, M., et al., *Prediction of metastasis and recurrence in colorectal cancer based on gene expression analysis: ready for the clinic?* Cancers, 2011. **3**(3): p. 2858-2869.
262. Zang, Y.-W., et al., *Brain metastases from colorectal cancer: microenvironment and molecular mechanisms.* International journal of molecular sciences, 2012. **13**(12): p. 15784-15800.
263. Jin, K., et al., *Mechanisms regulating colorectal cancer cell metastasis into liver (Review).* Oncology letters, 2012. **3**(1): p. 11-15.
264. Kawada, K., et al., *Molecular mechanisms of liver metastasis.* International journal of clinical oncology, 2011. **16**(5): p. 464-472.
265. Messick, C.A., et al., *Genetic and molecular diversity of colon cancer hepatic metastases.* Surgery, 2009. **146**(2): p. 227-231.
266. Santini, D., et al., *High concordance of KRAS status between primary colorectal tumors and related metastatic sites: implications for clinical practice.* The oncologist, 2008. **13**(12): p. 1270-1275.
267. Artale, S., et al., *Mutations of KRAS and BRAF in primary and matched metastatic sites of colorectal cancer.* Journal of Clinical Oncology, 2008. **26**(25): p. 4217-4219.
268. Santini, D., et al., *High concordance of BRAF status between primary colorectal tumours and related metastatic sites: implications for clinical practice.* Annals of oncology, 2010. **21**(7): p. 1565-1565.
269. Zhu, Y.-F., et al., *PI3K expression and PIK3CA mutations are related to colorectal cancer metastases.* World J Gastroenterol, 2012. **18**(28): p. 3745-3751.
270. Kastrinakis, W., et al., *Increased incidence of p53 mutations is associated with hepatic metastasis in colorectal neoplastic progression.* Oncogene, 1995. **11**(4): p. 647-652.
271. Korn, W.M., et al., *Chromosome arm 20q gains and other genomic alterations in colorectal cancer metastatic to liver, as analyzed by comparative genomic hybridization and fluorescence in situ hybridization.* Genes, Chromosomes and Cancer, 1999. **25**(2): p. 82-90.
272. Nanashima, A., et al., *Gain of chromosome 20 is a frequent aberration in liver metastasis of colorectal cancers.* Digestive diseases and sciences, 1997. **42**(7): p. 1388-1393.
273. Parada, L.A., et al., *Cytogenetic analyses of secondary liver tumors reveal significant differences in genomic imbalances between primary and metastatic colon carcinomas.* Clinical & experimental metastasis, 1999. **17**(6): p. 471-479.
274. Peinado, H., F. Portillo, and A. Cano, *Transcriptional regulation of cadherins during development and carcinogenesis.* 2004.
275. Katoh, Y. and M. Katoh, *Hedgehog signaling, epithelial-to-mesenchymal transition and miRNA (review).* International journal of molecular medicine, 2008. **22**(3): p. 271.
276. Kang, Y. and J. Massagué, *Epithelial-mesenchymal transitions: twist in development and metastasis.* Cell, 2004. **118**(3): p. 277-279.
277. Raftopoulos, I., et al., *Level of α-catenin expression in colorectal cancer correlates with invasiveness, metastatic potential, and survival.* Journal of surgical oncology, 1998. **68**(2): p. 92-99.
278. Nagase, H., R. Visse, and G. Murphy, *Structure and function of matrix metalloproteinases and TIMPs.* Cardiovascular research, 2006. **69**(3): p. 562-573.
279. Orlichenko, L.S. and D.C. Radisky, *Matrix metalloproteinases stimulate epithelial-mesenchymal transition during tumor development.* Clinical & experimental metastasis, 2008. **25**(6): p. 593-600.

280. Zhang, Y.-Y., B. Chen, and Y.-Q. Ding, *Metastasis-associated factors facilitating the progression of colorectal cancer.* Asian Pacific Journal of Cancer Prevention, 2012. **13**(6): p. 2437-2444.
281. Svoboda, M., et al., *MicroRNA expression profile associated with response to neoadjuvant chemoradiotherapy in locally advanced rectal cancer patients.* Radiation oncology, 2012. **7**(1): p. 1.
282. Schoof, C., et al., *MicroRNAs in cancer treatment and prognosis.* Am J Cancer Res, 2012. **2**(4): p. 414-433.
283. Schee, K., *Prognostic biomarkers in colorectal cancer with emphasis on microRNA.* 2012.
284. Mazeh, H., et al., *The diagnostic and prognostic role of microRNA in colorectal cancer-a comprehensive review.* J Cancer, 2013. **4**(3): p. 281-95.
285. Aslam, M.I., et al., *MicroRNAs are Novel Biomarkers for Detection of Colorectal Cancer*. 2012: INTECH Open Access Publisher.
286. Mosakhani, N., *MicroRNAs as Predictive and Prognostic Biomarkers inHuman Neoplasia: With Specific Focus on Colorectal Cancer, Giant Cell Tumor of Bone, and Leukemias.* 2013.
287. O'HARA, S.P., et al., *MicroRNAs: key modulators of posttranscriptional gene expression.* Gastroenterology, 2009. **136**(1): p. 17.
288. Ye, J.-J. and J. Cao, *MicroRNAs in colorectal cancer as markers and targets: Recent advances.* World J Gastroenterol, 2014. **20**(15): p. 4288-4299.
289. Bartel, D.P., *MicroRNAs: target recognition and regulatory functions.* cell, 2009. **136**(2): p. 215-233.
290. van Kouwenhove, M., M. Kedde, and R. Agami, *MicroRNA regulation by RNA-binding proteins and its implications for cancer.* Nature Reviews Cancer, 2011. **11**(9): p. 644-656.
291. Santarpia, L., M. Nicoloso, and G.A. Calin, *MicroRNAs: a complex regulatory network drives the acquisition of malignant cell phenotype.* Endocrine-related cancer, 2010. **17**(1): p. F51-F75.
292. Eulalio, A., et al., *Deadenylation is a widespread effect of miRNA regulation.* Rna, 2009. **15**(1): p. 21-32.
293. Eulalio, A., et al., *Target-specific requirements for enhancers of decapping in miRNA-mediated gene silencing.* Genes & development, 2007. **21**(20): p. 2558-2570.
294. Behm-Ansmant, I., et al., *mRNA degradation by miRNAs and GW182 requires both CCR4: NOT deadenylase and DCP1: DCP2 decapping complexes.* Genes & development, 2006. **20**(14): p. 1885-1898.
295. Huntzinger, E. and E. Izaurralde, *Gene silencing by microRNAs: contributions of translational repression and mRNA decay.* Nature Reviews Genetics, 2011. **12**(2): p. 99-110.
296. Calin, G.A., et al., *Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers.* Proceedings of the National academy of Sciences of the United States of America, 2004. **101**(9): p. 2999-3004.
297. Calin, G.A. and C.M. Croce, *MicroRNA signatures in human cancers.* Nature Reviews Cancer, 2006. **6**(11): p. 857-866.
298. Slaby, O., et al., *MicroRNAs in colorectal cancer: translation of molecular biology into clinical application.* Molecular cancer, 2009. **8**(1): p. 1.
299. Lu, J., et al., *MicroRNA expression profiles classify human cancers.* nature, 2005. **435**(7043): p. 834-838.
300. Volinia, S., et al., *A microRNA expression signature of human solid tumors defines cancer gene targets.* Proceedings of the National academy of Sciences of the United States of America, 2006. **103**(7): p. 2257-2261.

301. Cummins, J.M., et al., *The colorectal microRNAome.* Proceedings of the National Academy of Sciences of the United States of America, 2006. **103**(10): p. 3687-3692.
302. Bandrés, E., et al., *Identification by Real-time PCR of 13 mature microRNAs differentially expressed in colorectal cancer and non-tumoral tissues.* Molecular cancer, 2006. **5**(1): p. 1.
303. Akao, Y., et al., *Role of anti-oncomirs miR-143 and-145 in human colorectal tumors.* Cancer gene therapy, 2010. **17**(6): p. 398-408.
304. Schepeler, T., et al., *Diagnostic and prognostic microRNAs in stage II colon cancer.* Cancer research, 2008. **68**(15): p. 6416-6424.
305. Slattery, M.L., et al., *MicroRNAs and colon and rectal cancer: differential expression by tumor location and subtype.* Genes, Chromosomes and Cancer, 2011. **50**(3): p. 196-206.
306. Lanza, G., et al., *mRNA/microRNA gene expression profile in microsatellite unstable colorectal cancer.* Molecular cancer, 2007. **6**(1): p. 1.
307. Schetter, A.J., H. Okayama, and C.C. Harris, *The role of microRNAs in colorectal cancer.* Cancer journal (Sudbury, Mass.), 2012. **18**(3): p. 244.
308. Langan, R.C., et al., *Colorectal cancer biomarkers and the potential role of cancer stem cells.* J Cancer, 2013. **4**(3): p. 241-50.
309. Lech, G., et al., *Colorectal cancer tumour markers and biomarkers: Recent therapeutic advances.* World journal of gastroenterology, 2016. **22**(5): p. 1745.
310. Nguyen, M.T. and D.S. Weinberg, *Biomarkers in Colorectal Cancer Screening.* Journal of the National Comprehensive Cancer Network, 2016. **14**(8): p. 1033-1040.
311. Lieberman, D., *Progress and challenges in colorectal cancer screening and surveillance.* Gastroenterology, 2010. **138**(6): p. 2115-2126.
312. Lech, G., R. Slotwinski, and I. Krasnodebski, *The role of tumor markers and biomarkers in colorectal cancer.* Neoplasma, 2013. **61**(1): p. 1-8.
313. Ahmed, F.E., et al., *Diagnostic microRNA markers to screen for sporadic human colon cancer in blood.* Cancer Genomics-Proteomics, 2012. **9**(4): p. 179-192.
314. Wang, S., et al., *A plasma microRNA panel for early detection of colorectal cancer.* International Journal of Cancer, 2015. **136**(1): p. 152-161.
315. Nielsen, N., et al., *Circulating microRNAs as a diagnostic biomarker for colorectal cancer*. 2014.
316. Ng, E.K., et al., *Differential expression of microRNAs in plasma of colorectal cancer patients: a potential marker for colorectal cancer screening.* Gut, 2009.
317. Huang, Z., et al., *Plasma microRNAs are promising novel biomarkers for early detection of colorectal cancer.* International journal of cancer, 2010. **127**(1): p. 118-126.
318. Cheng, H., et al., *Circulating plasma MiR-141 is a novel biomarker for metastatic colon cancer and predicts poor prognosis.* PloS one, 2011. **6**(3): p. e17745.
319. Toiyama, Y., et al., *Serum miR-21 as a diagnostic and prognostic biomarker in colorectal cancer.* Journal of the National Cancer Institute, 2013. **105**(12): p. 849-859.
320. Wu, C.-W. and J.J.-Y. Sung, *Colorectal cancer screening: are stool and blood based tests good enough?* Chinese clinical oncology, 2012. **2**(1).
321. Roldo, C., et al., *MicroRNA expression abnormalities in pancreatic endocrine and acinar tumors are associated with distinctive pathologic features and clinical behavior.* Journal of Clinical Oncology, 2006. **24**(29): p. 4677-4684.
322. Slaby, O., et al., *Altered expression of miR-21, miR-31, miR-143 and miR-145 is related to clinicopathologic features of colorectal cancer.* Oncology, 2008. **72**(5-6): p. 397-402.
323. Schetter, A.J., et al., *MicroRNA expression profiles associated with prognosis and therapeutic outcome in colon adenocarcinoma.* Jama, 2008. **299**(4): p. 425-436.

324. Kulda, V., et al., *Relevance of miR-21 and miR-143 expression in tissue samples of colorectal carcinoma and its liver metastases.* Cancer genetics and cytogenetics, 2010. **200**(2): p. 154-160.
325. Nielsen, B.S., et al., *High levels of microRNA-21 in the stroma of colorectal cancers predict short disease-free survival in stage II colon cancer patients.* Clinical & experimental metastasis, 2011. **28**(1): p. 27-38.
326. Wang, C.-J., et al., *Clinicopathological significance of microRNA-31,-143 and-145 expression in colorectal cancer.* Disease markers, 2009. **26**(1): p. 27-34.
327. Svoboda, M., et al., *Micro-RNAs miR125b and miR137 are frequently upregulated in response to capecitabine chemoradiotherapy of rectal cancer.* International journal of oncology, 2008. **33**(3): p. 541.
328. Ragusa, M., et al., *Specific alterations of microRNA transcriptome and global network structure in colorectal carcinoma after cetuximab treatment.* Molecular cancer therapeutics, 2010. **9**(12): p. 3396-3409.
329. Valtorta, E., et al., *KRAS gene amplification in colorectal cancer and impact on response to EGFR-targeted therapy.* International Journal of Cancer, 2013. **133**(5): p. 1259-1265.
330. Bardelli, A., et al., *Amplification of the MET receptor drives resistance to anti-EGFR therapies in colorectal cancer.* Cancer discovery, 2013. **3**(6): p. 658-673.
331. Yonesaka, K., et al., *Activation of ERBB2 signaling causes resistance to the EGFR-directed therapeutics antibody cetuximab.* Cancer Research, 2012. **72**(8 Supplement): p. 4833-4833.
332. Fleige, S. and M.W. Pfaffl, *RNA integrity and the effect on the real-time qRT-PCR performance.* Molecular aspects of medicine, 2006. **27**(2): p. 126-139.
333. Fadhil, W., et al., *Quick-multiplex-consensus (QMC)-PCR followed by high-resolution melting: a simple and robust method for mutation detection in formalin-fixed paraffin-embedded tissue.* Journal of clinical pathology, 2010. **63**(2): p. 134-140.
334. Seth, R., et al., *The putative tumour modifier gene ATP5A1 is not mutated in human colorectal cancer cell lines but expression levels correlate with TP53 mutations and chromosomal instability.* Journal of clinical pathology, 2009. **62**(7): p. 598-603.
335. Bläker, H., et al., *Analysis of somatic APC mutations in rare extracolonic tumors of patients with familial adenomatous polyposis coli.* Genes, Chromosomes and Cancer, 2004. **41**(2): p. 93-98.
336. Liew, M., et al., *Genotyping of single-nucleotide polymorphisms by high-resolution melting of small amplicons.* Clinical chemistry, 2004. **50**(7): p. 1156-1164.
337. Pfundheller, H.M., et al., *Locked nucleic acid synthesis.* Oligonucleotide Synthesis, 2005: p. 127-145.
338. Vester, B. and J. Wengel, *LNA (locked nucleic acid): high-affinity targeting of complementary RNA and DNA.* Biochemistry, 2004. **43**(42): p. 13233-13241.
339. Mouritzen, P., et al., *Single nucleotide polymorphism genotyping using locked nucleic acid (LNA™).* Expert review of molecular diagnostics, 2003. **3**(1): p. 27-38.
340. Do, H., et al., *High resolution melting analysis for rapid and sensitive EGFR and KRAS mutation detection in formalin fixed paraffin embedded biopsies.* BMC cancer, 2008. **8**(1): p. 1.
341. Fukui, T., et al., *Prospective study of the accuracy of EGFR mutational analysis by high-resolution melting analysis in small samples obtained from patients with non–small cell lung cancer.* Clinical Cancer Research, 2008. **14**(15): p. 4751-4757.
342. Krypuy, M., et al., *High resolution melting for mutation scanning of TP53 exons 5–8.* BMC cancer, 2007. **7**(1): p. 1.

343. Krypuy, M., et al., *High resolution melting analysis for the rapid and sensitive detection of mutations in clinical samples: KRAS codon 12 and 13 mutations in non-small cell lung cancer.* BMC cancer, 2006. **6**(1): p. 1.
344. Nomoto, K., et al., *Detection of EGFR mutations in archived cytologic specimens of non–small cell lung cancer using high-resolution melting analysis.* American journal of clinical pathology, 2006. **126**(4): p. 608-615.
345. Poláková, K.M., et al., *High-resolution melt curve analysis: initial screening for mutations in BCR-ABL kinase domain.* Leukemia research, 2008. **32**(8): p. 1236-1243.
346. Takano, T., et al., *Epidermal growth factor receptor mutation detection using high-resolution melting analysis predicts outcomes in patients with advanced non–small cell lung cancer treated with gefitinib.* Clinical Cancer Research, 2007. **13**(18): p. 5385-5390.
347. Vandersteen, J.G., et al., *Identifying common genetic variants by high-resolution melting.* Clinical chemistry, 2007. **53**(7): p. 1191-1198.
348. Kristensen, L.S., et al., *Increased sensitivity of KRAS mutation detection by high-resolution melting analysis of COLD-PCR products.* Human mutation, 2010. **31**(12): p. 1366-1373.
349. Carotenuto, P., et al., *Detection of KRAS mutations in colorectal cancer with Fast COLD-PCR.* International journal of oncology, 2012. **40**(2): p. 378.
350. Li, J., et al., *Replacing PCR with COLD-PCR enriches variant DNA sequences and redefines the sensitivity of genetic testing.* Nature medicine, 2008. **14**(5): p. 579-584.
351. Reed, G.H., J.O. Kent, and C.T. Wittwer, *High-resolution DNA melting analysis for simple and efficient molecular diagnostics.* 2007.
352. McKinney, J.T., et al., *Mutation Scanning and Genotyping in Plants by High-Resolution DNA Melting.* The Handbook of Plant Mutation Screening: Mining of Natural and Induced Alleles, 2010: p. 149-165.
353. Reed, G.H. and C.T. Wittwer, *Sensitivity and specificity of single-nucleotide polymorphism scanning by high-resolution melting analysis.* Clinical chemistry, 2004. **50**(10): p. 1748-1754.
354. Liu, Y.-P., et al., *Diagnostic accuracy of high resolution melting analysis for detection of KRAS mutations: a systematic review and meta-analysis.* Scientific reports, 2014. **4**.
355. Montgomery, J.L., L.N. Sanford, and C.T. Wittwer, *High-resolution DNA melting analysis in clinical research and diagnostics.* Expert review of molecular diagnostics, 2010. **10**(2): p. 219-240.
356. van der Stoep, N., et al., *Diagnostic guidelines for high-resolution melting curve (HRM) analysis: An interlaboratory validation of BRCA1 mutation scanning using the 96-well LightScanner™.* Human mutation, 2009. **30**(6): p. 899-909.
357. Taylor, C.F., *Mutation scanning using high-resolution melting.* Biochemical Society Transactions, 2009. **37**(2): p. 433-437.
358. Herrmann, M.G., et al., *Amplicon DNA melting analysis for mutation scanning and genotyping: cross-platform comparison of instruments and dyes.* Clinical chemistry, 2006. **52**(3): p. 494-503.
359. Akiyoshi, K., et al., *KRAS mutations in patients with colorectal cancer as detected by high-resolution melting analysis and direct sequencing.* Anticancer research, 2013. **33**(5): p. 2129-2134.
360. Khoshnaw, S.M., et al., *Detection and quantification of microRNAs in laser-microdissected formalin-fixed paraffin-embedded breast cancer tissues.* Laser Capture Microdissection: Methods and Protocols, 2011: p. 119-142.

361. Macabeo-Ong, M., et al., *Effect of duration of fixation on quantitative reverse transcription polymerase chain reaction analyses.* Modern pathology, 2002. **15**(9): p. 979-987.
362. Srinivasan, M., D. Sedmak, and S. Jewell, *Effect of fixatives and tissue processing on the content and integrity of nucleic acids.* The American journal of pathology, 2002. **161**(6): p. 1961-1971.
363. Dijkstra, J., et al., *MicroRNA expression in formalin-fixed paraffin embedded tissue using real time quantitative PCR: the strengths and pitfalls.* Journal of cellular and molecular medicine, 2012. **16**(4): p. 683-690.
364. Chretien, A.S., et al., *Optimization of routine KRAS mutation PCR-based testing procedure for rational individualized first-line-targeted therapy selection in metastatic colorectal cancer.* Cancer medicine, 2013. **2**(1): p. 11-20.
365. Ahmed, M.A., et al., *CD24 is upregulated in inflammatory bowel disease and stimulates cell motility and colony formation.* Inflammatory bowel diseases, 2010. **16**(5): p. 795-803.
366. Alvarez, M.J., et al., *Model based analysis of real-time PCR data from DNA binding dye protocols.* BMC bioinformatics, 2007. **8**(1): p. 1.
367. Bustin, S.A., et al., *The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments.* Clinical chemistry, 2009. **55**(4): p. 611-622.
368. Higuchi, R., et al., *Simultaneous amplification and detection of specific DNA sequences.* Bio/technology, 1992. **10**(4): p. 413-417.
369. Sun, D., et al., *MicroRNA-31 activates the RAS pathway and functions as an oncogenic MicroRNA in human colorectal cancer by repressing RAS p21 GTPase activating protein 1 (RASA1).* Journal of Biological Chemistry, 2013. **288**(13): p. 9508-9518.
370. Xia, X., et al., *Prognostic role of microRNA-21 in colorectal cancer: a meta-analysis.* PloS one, 2013. **8**(11): p. e80426.
371. Zhang, X., et al., *miR‐224 promotes colorectal cancer cells proliferation via downregulation of P21WAF1/CIP1.* Molecular medicine reports, 2014. **9**(3): p. 941-946.
372. Ahmed, F.E., et al., *Diagnostic microRNA markers to screen for sporadic human colon cancer in stool: I. Proof of principle.* Cancer Genomics-Proteomics, 2013. **10**(3): p. 93-113.
373. Abd El-Rehim, D.M., et al., *High-throughput protein expression analysis using tissue microarray technology of a large well-characterised series identifies biologically distinct classes of breast cancer confirming recent cDNA expression analyses.* Int J Cancer, 2005. **116**(3): p. 340-50.
374. Tabernero, J. *The evolution of our molecular understanding of colorectal cancer: what we are doing now, what the future holds, and how tumor profiling is just the beginning*. 2014. American Society of Clinical Oncology.
375. Müller, M.F., A.E. Ibrahim, and M.J. Arends, *Molecular pathological classification of colorectal cancer.* Virchows Archiv, 2016. **469**(2): p. 125-134.
376. Woodford-Richens, K., et al., *SMAD4 mutations in colorectal cancer probably occur before chromosomal instability, but after divergence of the microsatellite instability pathway.* Proceedings of the National Academy of Sciences, 2001. **98**(17): p. 9719-9723.
377. Dallol, A., et al., *Clinical significance of frequent somatic mutations detected by high-throughput targeted sequencing in archived colorectal cancer samples.* Journal of translational medicine, 2016. **14**(1): p. 1.

378. FRIEDMAN, K., H. SAFRAN, and M.B. RESNICK, *Mutation Profiling of Clinically Advanced Cancers Using Next-Generation Sequencing for Targeted Therapy: A Lifespan Experience.* Rhode Island medical journal (2013), 2015. **98**(10): p. 16.
379. Ciombor, K.K., S. Haraldsdottir, and R.M. Goldberg, *How can next-generation sequencing (genomics) help us in treating colorectal cancer?* Current colorectal cancer reports, 2014. **10**(4): p. 372-379.
380. Kim, T.-M., S.-H. Lee, and Y.-J. Chung, *Clinical applications of next-generation sequencing in colorectal cancers.* World J Gastroenterol, 2013. **19**(40): p. 6784-93.
381. Youssef, O., et al., *Mutations by Next Generation Sequencing in Stool DNA from Colorectal Carcinoma Patients–A Literature Review and our Experience with this Methodology.* Journal of Analytical Oncology, 2016. **5**(1): p. 24-32.
382. Blanco-Calvo, M., et al., *Colorectal cancer classification and cell heterogeneity: a systems oncology approach.* International journal of molecular sciences, 2015. **16**(6): p. 13610-13632.
383. Turtoi, A., et al., *Organized proteomic heterogeneity in colorectal cancer liver metastases and implications for therapies.* Hepatology, 2014. **59**(3): p. 924-934.
384. Zenonos, K. and K. Kyprianou, *RAS signaling pathways, mutations and their role in colorectal cancer.* World journal of gastrointestinal oncology, 2013. **5**(5): p. 97.
385. Krol, L., et al., *Concordance in KRAS and BRAF mutations in endoscopic biopsy samples and resection specimens of colorectal adenocarcinoma.* European Journal of Cancer, 2012. **48**(7): p. 1108-1115.
386. Lu, Y.-W., et al., *Colorectal Cancer Genetic Heterogeneity Delineated by Multi-Region Sequencing.* PloS one, 2016. **11**(3): p. e0152673.
387. Kogita, A., et al., *Inter-and intra-tumor profiling of multi-regional colon cancer and metastasis.* Biochemical and biophysical research communications, 2015. **458**(1): p. 52-56.
388. Knudson, A.G., *Two genetic hits (more or less) to cancer.* Nature Reviews Cancer, 2001. **1**(2): p. 157-162.
389. Slamon, D.J., et al., *Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2.* New England Journal of Medicine, 2001. **344**(11): p. 783-792.
390. Bang, Y.-J., et al., *Trastuzumab in combination with chemotherapy versus chemotherapy alone for treatment of HER2-positive advanced gastric or gastro-oesophageal junction cancer (ToGA): a phase 3, open-label, randomised controlled trial.* The Lancet, 2010. **376**(9742): p. 687-697.
391. Mok, T.S., et al., *Gefitinib or carboplatin–paclitaxel in pulmonary adenocarcinoma.* New England Journal of Medicine, 2009. **361**(10): p. 947-957.
392. Van Cutsem, E., et al., *Cetuximab plus irinotecan, fluorouracil, and leucovorin as first-line treatment for metastatic colorectal cancer: updated analysis of overall survival according to tumor KRAS and BRAF mutation status.* Journal of Clinical Oncology, 2011: p. JCO. 2010.33. 5091.
393. Lüchtenborg, M., et al., *APC mutations in sporadic colorectal carcinomas from The Netherlands Cohort Study.* Carcinogenesis, 2004. **25**(7): p. 1219-1226.
394. Li, W.Q., et al., *BRAF mutations are associated with distinctive clinical, pathological and molecular features of colorectal cancer independently of microsatellite instability status.* Molecular cancer, 2006. **5**(1): p. 2.
395. Fleming, N.I., et al., *SMAD2, SMAD3 and SMAD4 mutations in colorectal cancer.* Cancer research, 2013. **73**(2): p. 725-735.

396. Gay, L.J., et al., *Dietary, lifestyle and clinicopathological factors associated with APC mutations and promoter methylation in colorectal cancers from the EPIC-Norfolk study.* The Journal of pathology, 2012. **228**(3): p. 405-415.
397. Stenzinger, A., et al., *Mutations in POLE and survival of colorectal cancer patients–link to disease stage and treatment.* Cancer medicine, 2014. **3**(6): p. 1527-1538.
398. Dighe, S., et al., *Accuracy of radiological staging in identifying high-risk colon cancer patients suitable for neoadjuvant chemotherapy: a multicentre experience.* Colorectal Disease, 2012. **14**(4): p. 438-444.
399. Baldus, S.E., et al., *Prevalence and heterogeneity of KRAS, BRAF, and PIK3CA mutations in primary colorectal adenocarcinomas and their corresponding metastases.* Clinical Cancer Research, 2010. **16**(3): p. 790-799.
400. Yun, J., et al., *Glucose deprivation contributes to the development of KRAS pathway mutations in tumor cells.* Science, 2009. **325**(5947): p. 1555-1559.
401. Sanchez, J., et al., *Genetic and epigenetic classifications define clinical phenotypes and determine patient outcomes in colorectal cancer.* British Journal of Surgery, 2009. **96**(10): p. 1196-1204.
402. Ilyas, M., et al., *Genetic pathways in colorectal and other cancers.* European Journal of Cancer, 1999. **35**(14): p. 1986-2002.
403. Nilbert, M., et al., *Microsatellite instability is rare in rectal carcinomas and signifies hereditary cancer.* European Journal of Cancer, 1999. **35**(6): p. 942-945.
404. Cunningham, J.M., et al., *Hypermethylation of the hMLH1 promoter in colon cancer with microsatellite instability.* Cancer research, 1998. **58**(15): p. 3455-3460.
405. Cunningham, J.M., et al., *The frequency of hereditary defective mismatch repair in a prospective series of unselected colorectal carcinomas.* The American Journal of Human Genetics, 2001. **69**(4): p. 780-790.
406. Kane, M.F., et al., *Methylation of the hMLH1 promoter correlates with lack of expression of hMLH1 in sporadic colon tumors and mismatch repair-defective human tumor cell lines.* Cancer research, 1997. **57**(5): p. 808-811.
407. Thibodeau, S.N., et al., *Microsatellite instability in colorectal cancer: different mutator phenotypes and the principal involvement of hMLH1.* Cancer Research, 1998. **58**(8): p. 1713-1718.
408. Kouri, M., et al., *Diploid predominance in hereditary nonpolyposis colorectal carcinoma evaluated by flow cytometry.* Cancer, 1990. **65**(8): p. 1825-1829.
409. Linnebacher, M., et al., *Single nucleotide polymorphism array analysis of microsatellite-stable, diploid/near-diploid colorectal carcinomas without the CpG island methylator phenotype.* Oncology letters, 2013. **5**(1): p. 173-178.
410. Ostwald, C., et al., *Chromosomally and microsatellite stable colorectal carcinomas without the CpG island methylator phenotype in a molecular classification.* International journal of oncology, 2009. **35**(2): p. 321-327.
411. Kakar, S., et al., *Clinicopathologic characteristics, CpG island methylator phenotype, and BRAF mutations in microsatellite-stable colorectal cancers without chromosomal instability.* Archives of pathology & laboratory medicine, 2008. **132**(6): p. 958.
412. Cai, G., et al., *Clinicopathologic and molecular features of sporadic microsatellite-and chromosomal-stable colorectal cancers.* International journal of colorectal disease, 2008. **23**(4): p. 365-373.
413. Chai, H., et al., *miR-20a targets BNIP2 and contributes chemotherapeutic resistance in colorectal adenocarcinoma SW480 and SW620 cell lines.* Acta biochimica et biophysica Sinica, 2011: p. gmq125.

414. Williams, A.C., et al., *The effect of different TP53 mutations on the chromosomal stability of a human colonic adenoma derived cell line with endogenous wild type TP53 activity, before and after DNA damage.* Genes, Chromosomes and Cancer, 1997. **20**(1): p. 44-52.
415. Quirke, P., et al., *DNA aneuploidy in colorectal adenomas.* British journal of cancer, 1986. **53**(4): p. 477.
416. van den Ingh, H.F., G. Griffioen, and C.J. Cornelisse, *Flow cytometric detection of aneuploidy in colorectal adenomas.* Cancer research, 1985. **45**(7): p. 3392-3397.
417. Goh, H. and J. Jass, *DNA content and the adenoma-carcinoma sequence in the colorectum.* Journal of clinical pathology, 1986. **39**(4): p. 387-392.
418. Sadahiro, S., et al., *Recurrence patterns after curative resection of colorectal cancer in patients followed for a minimum of ten years.* Hepato-gastroenterology, 2002. **50**(53): p. 1362-1366.
419. Eccles, S.A. and D.R. Welch, *Metastasis: recent discoveries and novel treatment strategies.* The Lancet, 2007. **369**(9574): p. 1742-1757.
420. Kleivi, K., et al., *Gene expression profiles of primary colorectal carcinomas, liver metastases, and carcinomatoses.* Molecular cancer, 2007. **6**(1): p. 1.
421. Deschoolmeester, V., et al., *A review of the most promising biomarkers in colorectal cancer: one step closer to targeted therapy.* The oncologist, 2010. **15**(7): p. 699-731.
422. Kahng, L.S., *Genetic aspects of non-polypoid colorectal neoplasms.* Gastrointestinal endoscopy clinics of North America, 2010. **20**(3): p. 573-578.
423. Ogino, S., et al., *Molecular pathological epidemiology of colorectal neoplasia: an emerging transdisciplinary and interdisciplinary field.* Gut, 2011. **60**(3): p. 397-411.
424. Talmadge, J.E. and I.J. Fidler, *AACR centennial series: the biology of cancer metastasis: historical perspective.* Cancer research, 2010. **70**(14): p. 5649-5669.
425. Siegmund, K.D., et al., *Many colorectal cancers are "flat" clonal expansions.* Cell Cycle, 2009. **8**(14): p. 2187-2193.
426. Worthley, D.L. and B.A. Leggett, *Colorectal cancer: molecular features and clinical opportunities.* Clin Biochem Rev, 2010. **31**(2): p. 31-8.
427. Migliore, L., et al., *Genetics, cytogenetics, and epigenetics of colorectal cancer.* BioMed Research International, 2011. **2011**.
428. Palena, C., et al., *Strategies to target molecules that control the acquisition of a mesenchymal-like phenotype by carcinoma cells.* Experimental Biology and Medicine, 2011. **236**(5): p. 537-545.
429. Neerincx, M., et al., *MiR expression profiles of paired primary colorectal cancer and metastases by next-generation sequencing.* Oncogenesis, 2015. **4**(10): p. e170.
430. Bedard, P.L., et al., *Tumour heterogeneity in the clinic.* Nature, 2013. **501**(7467): p. 355-64.
431. Gerlinger, M., et al., *Intratumor heterogeneity and branched evolution revealed by multiregion sequencing.* N Engl J Med, 2012. **366**(10): p. 883-92.
432. Yachida, S., et al., *Distant metastasis occurs late during the genetic evolution of pancreatic cancer.* Nature, 2010. **467**(7319): p. 1114-7.
433. Jones, S., et al., *Comparative lesion sequencing provides insights into tumor evolution.* Proc Natl Acad Sci U S A, 2008. **105**(11): p. 4283-8.
434. Brannon, A.R., et al., *Comparative sequencing analysis reveals high genomic concordance between matched primary and metastatic colorectal cancer lesions.* Genome Biol, 2014. **15**(8): p. 454.

435. Kim, R., et al., *Co-evolution of somatic variation in primary and metastatic colorectal cancer may expand biopsy indications in the molecular era.* PloS one, 2015. **10**(5): p. e0126670.
436. Vermaat, J.S., et al., *Primary colorectal cancers and their subsequent hepatic metastases are genetically different: implications for selection of patients for targeted treatment.* Clin Cancer Res, 2012. **18**(3): p. 688-99.
437. Fan, M.-Q., et al., *Decrease expression of microRNA-20a promotes cancer cell proliferation and predicts poor survival of hepatocellular carcinoma.* Journal of Experimental & Clinical Cancer Research, 2013. **32**(1): p. 1.
438. Yan, H., et al., *MicroRNA-20a overexpression inhibited proliferation and metastasis of pancreatic carcinoma cells.* Human gene therapy, 2010. **21**(12): p. 1723-1734.
439. Yang, C.H., et al., *MicroRNA miR-21 regulates the metastatic behavior of B16 melanoma cells.* Journal of Biological Chemistry, 2011. **286**(45): p. 39172-39178.
440. Xiong, B., et al., *MiR-21 regulates biological behavior through the PTEN/PI-3 K/Akt signaling pathway in human colorectal cancer cells.* International journal of oncology, 2013. **42**(1): p. 219-228.
441. Giráldez, M.D., et al., *Circulating microRNAs as biomarkers of colorectal cancer: results from a genome-wide profiling and validation study.* Clinical Gastroenterology and Hepatology, 2013. **11**(6): p. 681-688. e3.
442. Luo, X., et al., *Identification and evaluation of plasma microRNAs for early detection of colorectal cancer.* PloS one, 2013. **8**(5): p. e62880.
443. Mosakhani, N., et al., *MicroRNA profiling differentiates colorectal cancer according to KRAS status.* Genes, Chromosomes and Cancer, 2012. **51**(1): p. 1-9.
444. Sengupta, D. and S. Bandyopadhyay, *Topological patterns in microRNA–gene regulatory network: studies in colorectal and breast cancer.* Molecular bioSystems, 2013. **9**(6): p. 1360-1371.
445. Wang, Y., et al., *Profiling microRNA expression in hepatocellular carcinoma reveals microRNA-224 up-regulation and apoptosis inhibitor-5 as a microRNA-224-specific target.* Journal of Biological Chemistry, 2008. **283**(19): p. 13205-13215.
446. Zhang, Y., et al., *Involvement of microRNA-224 in cell proliferation, migration, invasion, and anti-apoptosis in hepatocellular carcinoma.* Journal of Gastroenterology and Hepatology, 2013. **28**(3): p. 565-575.
447. Ayaz, L., et al., *Differential expression of microRNAs in plasma of patients with laryngeal squamous cell carcinoma: potential early-detection markers for laryngeal squamous cell carcinoma.* Journal of cancer research and clinical oncology, 2013. **139**(9): p. 1499-1506.
448. Oliveira, C., et al., *KRAS and BRAF oncogenic mutations in MSS colorectal carcinoma progression.* Oncogene, 2007. **26**(1): p. 158-163.
449. Albanese, I., et al., *Heterogeneity within and between primary colorectal carcinomas and matched metastases as revealed by analysis of Ki-ras and p53 mutations.* Biochemical and biophysical research communications, 2004. **325**(3): p. 784-791.
450. Bossard, C., et al. *Does Testing of KRAS in Patients with Metastatic Colorectal Cancer Offer Valuable Information in Deciding Treatment Options?* in *LABORATORY INVESTIGATION*. 2010. NATURE PUBLISHING GROUP 75 VARICK ST, 9TH FLR, NEW YORK, NY 10013-1917 USA.
451. Houlle, S., et al. *Metastatic CRCs KRAS and BRAF genotyping in routine diagnosis: Results and pitfalls*. in *LABORATORY INVESTIGATION*. 2010. NATURE PUBLISHING GROUP 75 VARICK ST, 9TH FLR, NEW YORK, NY 10013-1917 USA.

452. Etienne-Grimaldi, M.-C., et al., *K-Ras mutations and treatment outcome in colorectal cancer patients receiving exclusive fluoropyrimidine therapy.* Clinical Cancer Research, 2008. **14**(15): p. 4830-4835.
453. Italiano, A., et al., *KRAS and BRAF mutational status in primary colorectal tumors and related metastatic sites: biological and clinical implications.* Annals of surgical oncology, 2010. **17**(5): p. 1429-1434.
454. Vakiani, E., et al., *Comparative genomic analysis of primary versus metastatic colorectal carcinomas.* Journal of Clinical Oncology, 2012. **30**(24): p. 2956-2962.
455. Voutsina, A., et al., *Combined analysis of KRAS and PIK3CA mutations, MET and PTEN expression in primary tumors and corresponding metastases in colorectal cancer.* Modern Pathology, 2013. **26**(2): p. 302-313.
456. Kopetz, S., et al. *Mutation and copy number discordance in primary versus metastatic colorectal cancer (mCRC)*. in *ASCO Annual Meeting Proceedings*. 2014.
457. Forslund, A., et al., *Mutations and allelic loss of p53 in primary tumor DNA from potentially cured patients with colorectal carcinoma.* Journal of clinical oncology, 2001. **19**(11): p. 2829-2836.
458. Parasramka, M.A., et al., *A role for low-abundance miRNAs in colon cancer: the miR-206/Krüppel-like factor 4 (KLF4) axis.* Clinical epigenetics, 2012. **4**(1): p. 1.
459. Takamizawa, J., et al., *Reduced expression of the let-7 microRNAs in human lung cancers in association with shortened postoperative survival.* Cancer research, 2004. **64**(11): p. 3753-3756.
460. Choong, M.L., H.H. Yang, and I. McNiece, *MicroRNA expression profiling during human cord blood-derived CD34 cell erythropoiesis.* Experimental hematology, 2007. **35**(4): p. 551-564.
461. Corney, D.C., et al., *MicroRNA-34b and MicroRNA-34c are targets of p53 and cooperate in control of cell proliferation and adhesion-independent growth.* Cancer research, 2007. **67**(18): p. 8433-8438.
462. Carlsson, J., et al., *Validation of suitable endogenous control genes for expression studies of miRNA in prostate cancer tissues.* Cancer genetics and cytogenetics, 2010. **202**(2): p. 71-75.
463. Feiersinger, F., et al., *MiRNA-21 Expression Decreases from Primary Tumors to Liver Metastases in Colorectal Carcinoma.* PloS one, 2016. **11**(2): p. e0148580.
464. Wang, L.-g. and J. Gu, *Serum microRNA-29a is a promising novel marker for early detection of colorectal liver metastasis.* Cancer epidemiology, 2012. **36**(1): p. e61-e67.
465. Yuan, K., et al., *Decreased levels of miR-224 and the passenger strand of miR-221 increase MBD2, suppressing maspin and promoting colorectal tumor growth and metastasis in mice.* Gastroenterology, 2013. **145**(4): p. 853-864. e9.
466. Ke, T.-W., et al., *MiR-92a promotes cell metastasis of colorectal cancer through PTEN-mediated PI3K/AKT pathway.* Annals of surgical oncology, 2015. **22**(8): p. 2649-2655.
467. Ling, H., et al., *The clinical and biological significance of MIR-224 expression in colorectal cancer metastasis.* Gut, 2015: p. gutjnl-2015-309372.
468. Tang, W., et al., *MicroRNA-29a promotes colorectal cancer metastasis by regulating matrix metalloproteinase 2 and E-cadherin via KLF4.* British journal of cancer, 2014. **110**(2): p. 450-458.
469. Negm O H, H.J.C., Fadhil W, Pitiot A, Tighe P J, Marchessoux C and Ilyas M, *BerEP4 and AE1/3 are Reliable Markers of Epithelial Content for Biomarker Discovery Using Reverse Phase Protein Arrays (RPPA).* Int J Pathol Clin Res, 2016. **2**(2): p. 1-7.

470. Atreya, C.E., et al., *PTEN expression is consistent in colorectal cancer primaries and metastases and associates with patient survival.* Cancer medicine, 2013. **2**(4): p. 496-506.
471. Losi, L., H. Bouzourene, and J. Benhattar, *Loss of Smad4 expression predicts liver metastasis in human colorectal cancer.* Oncology reports, 2007. **17**(5): p. 1095-1100.
472. Sawai, H., et al., *Loss of PTEN expression is associated with colorectal cancer liver metastasis and poor patient survival.* BMC gastroenterology, 2008. **8**(1): p. 1.
473. Anwar, R., *Screening for colorectal cancer in the UK.* Digestive and liver disease, 2006. **38**(4): p. 279-282.
474. O'Connell, J.B., M.A. Maggard, and C.Y. Ko, *Colon cancer survival rates with the new American Joint Committee on Cancer sixth edition staging.* Journal of the National Cancer Institute, 2004. **96**(19): p. 1420-1425.
475. Huerta, S., *Recent advances in the molecular diagnosis and prognosis of colorectal cancer.* Expert review of molecular diagnostics, 2008. **8**(3): p. 277-288.
476. Herranz, H. and S.M. Cohen, *MicroRNAs and gene regulatory networks: managing the impact of noise in biological systems.* Genes & development, 2010. **24**(13): p. 1339-1344.
477. Zhou, X., et al., *Identifying miRNA/mRNA negative regulation pairs in colorectal cancer.* Scientific reports, 2015. **5**.
478. Hannafon, B.N., et al., *Expression of microRNA and their gene targets are dysregulated in preinvasive breast cancer.* Breast cancer research, 2011. **13**(2): p. 1.
479. Miko, E., et al., *Differentially expressed microRNAs in small cell lung cancer.* Experimental lung research, 2009. **35**(8): p. 646-664.
480. Duttagupta, R. and K.W. Jones, *The curious case of miRNAs in circulation: potential diagnostic biomarkers?* Wiley Interdisciplinary Reviews: RNA, 2013. **4**(1): p. 129-138.
481. Diaz, R., et al., *Deregulated expression of miR-106a predicts survival in human colon cancer patients.* Genes, Chromosomes and Cancer, 2008. **47**(9): p. 794-802.
482. Sansal, I. and W.R. Sellers, *The biology and clinical relevance of the PTEN tumor suppressor pathway.* Journal of Clinical Oncology, 2004. **22**(14): p. 2954-2963.
483. Zhang, G.J., et al., *miR‐20a is an independent prognostic factor in colorectal cancer and is involved in cell metastasis.* Molecular medicine reports, 2014. **10**(1): p. 283-291.
484. Malzkorn, B., et al., *Identification and functional characterization of microRNAs involved in the malignant progression of gliomas.* Brain pathology, 2010. **20**(3): p. 539-550.
485. Liu, R., et al., *A five-microRNA signature identified from genome-wide serum microRNA expression profiling serves as a fingerprint for gastric cancer diagnosis.* European journal of cancer, 2011. **47**(5): p. 784-791.
486. Gong, B., et al., *MiR-21/RASA1 axis affects malignancy of colon cancer cells via RAS pathways.* World Journal of Gastroenterology: WJG, 2015. **21**(5): p. 1488.
487. Krichevsky, A.M. and G. Gabriely, *miR-21: a small multi-faceted RNA.* Journal of cellular and molecular medicine, 2009. **13**(1): p. 39-53.
488. Shen, J., et al., *Plasma microRNAs as potential biomarkers for non-small-cell lung cancer.* Laboratory investigation, 2011. **91**(4): p. 579-587.
489. Komatsu, S., et al., *Circulating microRNAs in plasma of patients with oesophageal squamous cell carcinoma.* British journal of cancer, 2011. **105**(1): p. 104-111.
490. Lagos-Quintana, M., et al., *Identification of novel genes coding for small expressed RNAs.* Science, 2001. **294**(5543): p. 853-858.
491. Wang, X.-S., et al., *MicroRNA-29a and microRNA-142-3p are regulators of myeloid differentiation and acute myeloid leukemia.* Blood, 2012. **119**(21): p. 4992-5004.

492. Muniyappa, M., et al., *MiRNA-29a regulates the expression of numerous proteins and reduces the invasiveness and proliferation of human carcinoma cell lines.* European Journal of Cancer, 2009. **45**(17): p. 3104-3118.
493. Cekaite, L., et al., *MiR-9,-31, and-182 deregulation promote proliferation and tumor cell survival in colon cancer.* Neoplasia, 2012. **14**(9): p. 868-IN21.
494. Schaefer, A., et al., *Diagnostic and prognostic implications of microRNA profiling in prostate carcinoma.* International journal of cancer, 2010. **126**(5): p. 1166-1176.
495. Guo, J., et al., *Differential expression of microRNA species in human gastric cancer versus non-tumorous tissues.* Journal of gastroenterology and hepatology, 2009. **24**(4): p. 652-657.
496. Veerla, S., et al., *MiRNA expression in urothelial carcinomas: important roles of miR-10a, miR-222, miR-125b, miR-7 and miR-452 for tumor stage and metastasis, and frequent homozygous losses of miR-31.* International Journal of Cancer, 2009. **124**(9): p. 2236-2242.
497. Liu, X., et al., *MicroRNA profiling and head and neck cancer.* Comparative and functional genomics, 2009. **2009**.
498. Wong, Q.W., et al., *MiR-222 overexpression confers cell migratory advantages in hepatocellular carcinoma through enhancing AKT signaling.* Clinical Cancer Research, 2010. **16**(3): p. 867-875.
499. Ota, A., et al., *Identification and characterization of a novel gene, C13orf25, as a target for 13q31-q32 amplification in malignant lymphoma.* Cancer research, 2004. **64**(9): p. 3087-3095.
500. Shigoka, M., et al., *Deregulation of miR-92a expression is implicated in hepatocellular carcinoma development.* Pathology international, 2010. **60**(5): p. 351-357.
501. Adam, L., et al. *Plasma microRNA profiles for bladder cancer detection*. in *Urologic Oncology: Seminars and Original Investigations*. 2013. Elsevier.
502. Wu, C.W., et al., *Detection of miR-92a and miR-21 in stool samples as potential screening biomarkers for colorectal cancer and polyps.* Gut, 2011: p. gut. 2011.239236.
503. Resnick, K.E., et al., *The detection of differentially expressed microRNAs from the serum of ovarian cancer patients using a novel real-time PCR platform.* Gynecologic oncology, 2009. **112**(1): p. 55-59.
504. Huang, L., et al., *MicroRNA-224 targets RKIP to control cell invasion and expression of metastasis genes in human breast cancer cells.* Biochemical and biophysical research communications, 2012. **425**(2): p. 127-133.
505. Yao, G., et al., *MicroRNA-224 is involved in transforming growth factor-β-mediated mouse granulosa cell proliferation and granulosa cell function by targeting Smad4.* Molecular endocrinology, 2010. **24**(3): p. 540-551.
506. Yi, R., et al., *MicroRNAs as diagnostic and prognostic biomarkers in colorectal cancer.* World journal of gastrointestinal oncology, 2016. **8**(4): p. 330.
507. Scapoli, L., et al., *MicroRNA expression profiling of oral carcinoma identifies new markers of tumor progression.* International journal of immunopathology and pharmacology, 2010. **23**(4): p. 1229-1234.
508. Yanaihara, N., et al., *Unique microRNA molecular profiles in lung cancer diagnosis and prognosis.* Cancer cell, 2006. **9**(3): p. 189-198.
509. Mavridis, K., K. Stravodimos, and A. Scorilas, *Downregulation and prognostic performance of microRNA 224 expression in prostate cancer.* Clinical chemistry, 2013. **59**(1): p. 261-269.

510. Giricz, O., et al., *Hsa-miR-375 is differentially expressed during breast lobular neoplasia and promotes loss of mammary acinar polarity.* The Journal of pathology, 2012. **226**(1): p. 108-119.
511. Iorio, M.V., et al., *MicroRNA signatures in human ovarian cancer.* Cancer research, 2007. **67**(18): p. 8699-8707.
512. Mencia, N., et al., *Underexpression of miR-224 in methotrexate resistant human colon cancer cells.* Biochemical pharmacology, 2011. **82**(11): p. 1572-1582.
513. Rossi, L., E. Bonmassar, and I. Faraoni, *Modification of miR gene expression pattern in human colon cancer cells following exposure to 5-fluorouracil in vitro.* Pharmacological research, 2007. **56**(3): p. 248-253.
514. Wang, Y.X., et al., *Initial study of microRNA expression profiles of colonic cancer without lymph node metastasis.* Journal of digestive diseases, 2010. **11**(1): p. 50-54.
515. Søreide, K., *Receiver-operating characteristic curve analysis in diagnostic, prognostic and predictive biomarker research.* Journal of clinical pathology, 2009. **62**(1): p. 1-5.
516. Park, S.H., J.M. Goo, and C.-H. Jo, *Receiver operating characteristic (ROC) curve: practical review for radiologists.* Korean Journal of Radiology, 2004. **5**(1): p. 11-18.
517. Corté, H., et al., *MicroRNA and colorectal cancer.* Digestive and Liver Disease, 2012. **44**(3): p. 195-200.
518. Ma, X., et al., *Interaction of the oncogenic miR-21 microRNA and the p53 tumor suppressor pathway.* Carcinogenesis, 2013: p. bgt044.
519. de Kok, J.B., et al., *Normalization of gene expression measurements in tumor tissues: comparison of 13 endogenous control genes.* Laboratory investigation, 2005. **85**(1): p. 154-159.
520. Svec, D., et al., *How good is a PCR efficiency estimate: Recommendations for precise and robust qPCR efficiency assessments.* Biomolecular detection and quantification, 2015. **3**: p. 9-16.
521. Shibuya, H., et al., *Clinicopathological and prognostic value of microRNA-21 and microRNA-155 in colorectal cancer.* Oncology, 2011. **79**(3-4): p. 313-320.
522. Chang, K., et al., *MicroRNA-21 and PDCD4 expression in colorectal cancer.* European Journal of Surgical Oncology (EJSO), 2011. **37**(7): p. 597-603.
523. Bullock, M., et al., *Pleiotropic actions of miR-21 highlight the critical role of deregulated stromal microRNAs during colorectal cancer progression.* Cell death & disease, 2013. **4**(6): p. e684.
524. Zhang, J., et al., *miR-21, miR-17 and miR-19a induced by phosphatase of regenerating liver-3 promote the proliferation and metastasis of colon cancer.* British journal of cancer, 2012. **107**(2): p. 352-359.
525. Faltejskova, P., et al., *Identification and functional screening of microRNAs highly deregulated in colorectal cancer.* Journal of cellular and molecular medicine, 2012. **16**(11): p. 2655-2666.
526. Reid, J.F., et al., *miRNA profiling in colorectal cancer highlights miR-1 involvement in MET-dependent proliferation.* Molecular Cancer Research, 2012. **10**(4): p. 504-515.
527. Schee, K., et al., *Deep sequencing the microRNA transcriptome in colorectal cancer.* PloS one, 2013. **8**(6): p. e66165.
528. Earle, J.S., et al., *Association of microRNA expression with microsatellite instability status in colorectal adenocarcinoma.* The Journal of Molecular Diagnostics, 2010. **12**(4): p. 433-440.
529. Ke, T.-W., et al., *MicroRNA-224 suppresses colorectal cancer cell migration by targeting Cdc42.* Disease markers, 2014. **2014**.

530. Diosdado, B., et al., *MiR-17-92 cluster is associated with 13q gain and c-myc expression during colorectal adenoma to adenocarcinoma progression.* British journal of cancer, 2009. **101**(4): p. 707-714.
531. Zhou, T., et al., *Overexpression of miR-92a correlates with tumor metastasis and poor prognosis in patients with colorectal cancer.* International journal of colorectal disease, 2013. **28**(1): p. 19-24.
532. Chang, K.H., et al., *MicroRNA signature analysis in colorectal cancer: identification of expression profiles in stage II tumors associated with aggressive disease.* International journal of colorectal disease, 2011. **26**(11): p. 1415-1422.
533. Mishra, S., et al., *Androgen receptor and microRNA-21 axis downregulates transforming growth factor beta receptor II (TGFBR2) expression in prostate cancer.* Oncogene, 2014. **33**(31): p. 4097-4106.
534. Donovan, S., K.M. Shannon, and G. Bollag, *GTPase activating proteins: critical regulators of intracellular signaling.* Biochimica et Biophysica Acta (BBA)-Reviews on Cancer, 2002. **1602**(1): p. 23-45.
535. Gong, B., et al., *MiR-21/RASA1 axis affects malignancy of colon cancer cells via RAS pathways.* World journal of gastroenterology, 2015. **21**(5): p. 1488-1497.
536. Lecomte, T., et al., *Circulating free tumor DNA and colorectal cancer.* Gastroentérologie clinique et biologique, 2010. **34**(12): p. 662-681.
537. Elshimali, Y.I., et al., *The clinical utilization of circulating cell free DNA (CCFDNA) in blood of cancer patients.* International journal of molecular sciences, 2013. **14**(9): p. 18925-18958.
538. Siravegna, G. and A. Bardelli, *Genotyping cell-free tumor DNA in the blood to detect residual disease and drug resistance.* Genome biology, 2014. **15**(8): p. 1.
539. Diehl, F., et al., *Circulating mutant DNA to assess tumor dynamics.* Nature medicine, 2008. **14**(9): p. 985-990.
540. Beaver, J.A., et al., *Detection of cancer DNA in plasma of patients with early-stage breast cancer.* Clinical Cancer Research, 2014. **20**(10): p. 2643-2650.
541. Catarino, R., et al., *Quantification of free circulating tumor DNA as a diagnostic marker for breast cancer.* DNA and cell biology, 2008. **27**(8): p. 415-421.
542. Wimberger, P., et al., *Impact of platinum-based chemotherapy on circulating nucleic acid levels, protease activities in blood and disseminated tumor cells in bone marrow of ovarian cancer patients.* International Journal of Cancer, 2011. **128**(11): p. 2572-2580.
543. Kadam, S.K., M. Farmen, and J.T. Brandt, *Quantitative measurement of cell-free plasma DNA and applications for detecting tumor genetic variation and promoter methylation in a clinical setting.* The Journal of Molecular Diagnostics, 2012. **14**(4): p. 346-356.
544. Ferrari, M., L. Cremonesi, and S. Galbiati, *Circulating nucleic acids as diagnostic tool.* biochimica clinica, 2008. **32**(1).
545. Luo, J., L. Shen, and D. Zheng, *Diagnostic value of circulating free DNA for the detection of EGFR mutation status in NSCLC: a systematic review and meta-analysis.* Scientific reports, 2014. **4**.
546. Sozzi, G., et al., *Analysis of circulating tumor DNA in plasma at diagnosis and during follow-up of lung cancer patients.* Cancer Research, 2001. **61**(12): p. 4675-4678.
547. Lo, Y.D., *Circulating nucleic acids in plasma and serum: an overview.* Annals of the New York Academy of Sciences, 2001. **945**(1): p. 1-7.
548. Stroun, M., et al., *Neoplastic characteristics of the DNA found in the plasma of cancer patients.* Oncology, 1989. **46**(5): p. 318-322.

549. Sorenson, G.D., et al., *Soluble normal and mutated DNA sequences from single-copy genes in human blood.* Cancer Epidemiology Biomarkers & Prevention, 1994. **3**(1): p. 67-71.
550. Nawroz, H., et al., *Microsatellite alterations in serum DNA of head and neck cancer patients.* Nature medicine, 1996. **2**(9): p. 1035-1037.
551. Swarup, V. and M. Rajeswari, *Circulating (cell-free) nucleic acids–a promising, non-invasive tool for early detection of several human diseases.* FEBS letters, 2007. **581**(5): p. 795-799.
552. Fleischhacker, M. and B. Schmidt, *Circulating nucleic acids (CNAs) and cancer—a survey.* Biochimica et Biophysica Acta (BBA)-Reviews on Cancer, 2007. **1775**(1): p. 181-232.
553. Schwarzenbach, H., D.S. Hoon, and K. Pantel, *Cell-free nucleic acids as biomarkers in cancer patients.* Nature Reviews Cancer, 2011. **11**(6): p. 426-437.
554. McLarty, J. and C. Yeh, *Circulating cell-free DNA: The blood biopsy in cancer management.* MOJ Cell. Sci. Rep, 2015. **2**: p. 0021.
555. Chiu, R.W., et al., *Quantitative analysis of circulating mitochondrial DNA in plasma.* Clinical chemistry, 2003. **49**(5): p. 719-726.
556. Boddy, J.L., et al., *Prospective study of quantitation of plasma DNA levels in the diagnosis of malignant versus benign prostate disease.* Clinical cancer research, 2005. **11**(4): p. 1394-1399.
557. Kamat, A.A., et al., *Plasma cell-free DNA in ovarian cancer.* Cancer, 2010. **116**(8): p. 1918-1925.
558. Agostini, M., et al., *Circulating cell-free DNA: a promising marker of regional lymphonode metastasis in breast cancer patients.* Cancer Biomarkers, 2012. **11**(2, 3): p. 89-98.
559. Gahan, P.B. and R. Swaminathan, *Circulating nucleic acids in plasma and serum.* Annals of the New York Academy of Sciences, 2008. **1137**(1): p. 1-6.
560. Agostini, M., et al., *Predictive response biomarkers in rectal cancer neoadjuvant treatment.* Front Biosci (Schol Ed), 2014. **6**: p. 110-119.
561. Cortez, M.A., et al., *MicroRNAs in body fluids—the mix of hormones and biomarkers.* Nature reviews Clinical oncology, 2011. **8**(8): p. 467-477.
562. Ji, X., et al., *Plasma miR-208 as a biomarker of myocardial injury.* Clinical chemistry, 2009. **55**(11): p. 1944-1949.
563. Théry, C., L. Zitvogel, and S. Amigorena, *Exosomes: composition, biogenesis and function.* Nature Reviews Immunology, 2002. **2**(8): p. 569-579.
564. Luo, X., et al., *MicroRNA signatures: novel biomarker for colorectal cancer?* Cancer Epidemiology Biomarkers & Prevention, 2011.
565. Pu, X.x., et al., *Circulating miR-221 directly amplified from plasma is a potential diagnostic and prognostic marker of colorectal cancer and is correlated with p53 expression.* Journal of gastroenterology and hepatology, 2010. **25**(10): p. 1674-1680.
566. Fesler, A., et al., *Circulating microRNA testing for the early diagnosis and follow-up of colorectal cancer patients.* Molecular diagnosis & therapy, 2014. **18**(3): p. 303-308.
567. Wang, J., et al., *Tumor-associated circulating microRNAs as biomarkers of cancer.* Molecules, 2014. **19**(2): p. 1912-1938.
568. Ristau, J., et al., *Suitability of circulating miRNAs as potential prognostic markers in colorectal cancer.* Cancer Epidemiology Biomarkers & Prevention, 2014. **23**(12): p. 2632-2637.
569. McAlexander, M.A., M.J. Phillips, and K.W. Witwer, *Comparison of methods for miRNA extraction from plasma and quantitative recovery of RNA from cerebrospinal fluid.* The

origin, function and diagnostic potential of extracellular microRNA in human body fluids, 2014: p. 45.
570. Page, K., et al., *Influence of plasma processing on recovery and analysis of circulating nucleic acids.* PLoS One, 2013. **8**(10): p. e77963.
571. Francis, G. and S. Stein, *Circulating cell-free tumour DNA in the management of cancer.* International journal of molecular sciences, 2015. **16**(6): p. 14122-14142.
572. Ortega, F.J., et al., *Inflammation triggers specific microRNA profiles in human adipocytes and macrophages and in their supernatants.* Clinical epigenetics, 2015. **7**(1): p. 1.
573. Raisch, J., A. Darfeuille-Michaud, and H. Nguyen, *Role of microRNAs in the immune system, inflammation and cancer.* World J Gastroenterol, 2013. **19**(20): p. 2985-96.
574. Valle, L., et al., *New insights into POLE and POLD1 germline mutations in familial colorectal cancer and polyposis.* Human molecular genetics, 2014. **23**(13): p. 3506-3512.
575. Miyabe, I., T.A. Kunkel, and A.M. Carr, *The major roles of DNA polymerases epsilon and delta at the eukaryotic replication fork are evolutionarily conserved.* PLoS Genet, 2011. **7**(12): p. e1002407.
576. Shinbrot, E., et al., *Exonuclease mutations in DNA polymerase epsilon reveal replication strand specific mutation patterns and human origins of replication.* Genome research, 2014. **24**(11): p. 1740-1750.
577. Chan, T.L., et al., *Early-onset colorectal cancer with stable microsatellite DNA and near-diploid chromosomes.* Oncogene, 2001. **20**(35).
578. Georgiades, I., et al., *Heterogeneity studies identify a subset of sporadic colorectal cancers without evidence for chromosomal or microsatellite instability.* Oncogene, 1999. **18**(56).
579. Bedard, P.L., et al., *Tumour heterogeneity in the clinic.* Nature, 2013. **501**(7467): p. 355-364.
580. Gerlinger, M., et al., *Intratumor heterogeneity and branched evolution revealed by multiregion sequencing.* New England journal of medicine, 2012. **366**(10): p. 883-892.
581. Jones, S., et al., *Comparative lesion sequencing provides insights into tumor evolution.* Proceedings of the National Academy of Sciences, 2008. **105**(11): p. 4283-4288.
582. Brannon, A.R., et al., *Comparative sequencing analysis reveals high genomic concordance between matched primary and metastatic colorectal cancer lesions.* Genome biology, 2014. **15**(8): p. 1.
583. *Comprehensive molecular characterization of human colon and rectal cancer.* Nature, 2012. **487**(7407): p. 330-7.
584. Domingo, E., et al., *Evaluation of PIK3CA mutation as a predictor of benefit from nonsteroidal anti-inflammatory drug therapy in colorectal cancer.* J Clin Oncol, 2013. **31**(34): p. 4297-305.
585. Esquela-Kerscher, A. and F.J. Slack, *Oncomirs—microRNAs with a role in cancer.* Nature Reviews Cancer, 2006. **6**(4): p. 259-269.
586. Garofalo, M. and C.M. Croce, *microRNAs: Master regulators as potential therapeutics in cancer.* Annual review of pharmacology and toxicology, 2011. **51**: p. 25-43.
587. Alajez, N.M., *Cancer stem cells. From characterization to therapeutic implications.* Saudi medical journal, 2011. **32**(12): p. 1229-1234.
588. Lizarbe, M.A., et al., *Colorectal cancer: from the genetic model to post-transcriptional regulation by non-coding RNAs.*
589. Xu, R.-S., et al., *The tumor suppressor gene RhoBTB1 is a novel target of miR-31 in human colon cancer.* International journal of oncology, 2013. **42**(2): p. 676-682.

590. Chen, T., et al., *MicroRNA-31 contributes to colorectal cancer development by targeting factor inhibiting HIF-1α (FIH-1).* Cancer biology & therapy, 2014. **15**(5): p. 516-523.
591. Yu, G., et al., *Prognostic values of the miR-17-92 cluster and its paralogs in colon cancer.* Journal of surgical oncology, 2012. **106**(3): p. 232-237.
592. Wang, Y., et al., *MicroRNA-224 targets SMAD family member 4 to promote cell proliferation and negatively influence patient survival.* PloS one, 2013. **8**(7): p. e68744.
593. Guo, N., et al., *Circulating tumor DNA detection in lung cancer patients before and after surgery.* Scientific Reports, 2016. **6**.
594. Sato, K.A., et al., *Tumor-unique mutation detection in cell-free DNA to monitor colorectal tumor burden using a cancer-associated gene sequencing panel*. 2015, AACR.
595. Reinert, T., et al., *Analysis of circulating tumour DNA to monitor disease burden following colorectal cancer surgery.* Gut, 2015: p. gutjnl-2014-308859.
596. Garcia-Murillas, I., et al., *Mutation tracking in circulating tumor DNA predicts relapse in early breast cancer.* Science translational medicine, 2015. **7**(302): p. 302ra133-302ra133.
597. Fadhil, W., et al., *The utility of diagnostic biopsy specimens for predictive molecular testing in colorectal cancer.* Histopathology, 2012. **61**(6): p. 1117-1124.
598. Greaves, M. and C.C. Maley, *Clonal evolution in cancer.* Nature, 2012. **481**(7381): p. 306-313.
599. Argos, M., et al., *Genomewide scan for loss of heterozygosity and chromosomal amplification in breast carcinoma using single-nucleotide polymorphism arrays.* Cancer genetics and cytogenetics, 2008. **182**(2): p. 69-74.
600. Bertheau, P., et al., *Allelic loss detection in inflammatory breast cancer: improvement with laser microdissection.* Laboratory investigation, 2001. **81**(10): p. 1397-1402.
601. Ramos, T.D. and L.M. Amorim, *Molecular biology techniques for loss of heterozygosity detection: the glioma example.* Jornal Brasileiro de Patologia e Medicina Laboratorial, 2015. **51**(3): p. 189-196.
602. Ryland, G.L., et al., *Loss of heterozygosity: what is it good for?* BMC medical genomics, 2015. **8**(1): p. 1.
603. Boisselier, B., et al., *COLD PCR HRM: a highly sensitive detection method for IDH1 mutations.* Human mutation, 2010. **31**(12): p. 1360-1365.
604. Song, C., et al., *Rapid and sensitive detection of KRAS mutation after fast-COLD-PCR enrichment and high-resolution melting analysis.* Diagnostic Molecular Pathology, 2011. **20**(2): p. 81-89.