Norwegian University of Life Sciences
Department of Animal and Aquacultural Sciences
(IHA)

Master Thesis 2014
60 credits

Mutation Screening of DNA Isolated from Formalin-Fixed Paraffin Embedded Tumor Blocks

Helene Røtterud

Acknowledgments

This study was performed at the Institute of Clinical Epidemiology and Molecular Biology (EpiGen) at Akershus University Hospital (Ahus), Lørenskog, Norway, as part of a Master's Degree program in biotechnology at the Norwegian University of Life Sciences (NMBU), Ås, Norway, The Department of Animal and Aquacultural Sciences (IHA).

Especially, I would like to thank my supervisor, scientist Torben Lüders at EpiGen, for his guidance and advice throughout this project, and for always being supportive and helpful. My deepest gratitude goes to you, for all your help.

Post. Doc. Jovana Klajic at EpiGen, I am ever grateful for all your help in double-checking my results and computing my Kaplan-Meier plots. Thanks a lot for your time.

I would also like to thank Pr. Vessela Kristensen for encouraging me to write an application for Internal Strategic Research Funding from Helse Sør-Øst. A daunting task I never thought I would manage, yet, the greatest reward was the 2012 grant (Strat 042). What a boost in confidence, eh!?

I would like to thank the entire staff at EpiGen for their support and thoughts in times of troubleshooting experimental procedures. Your advices have been very much appreciated, and indeed valuable.

Thanks to my internal supervisor, Pr. Dag Inge Våge, at the Department of Animal and Aquacultural Sciences, NMBU, for giving me support and exchange of thoughts when feeling blue.

At last, my family, who is ever so supporting. I owe it all to you.

Oslo, 10th of August 2014

Big Up
Helene Røtterud

Abstract

Colorectal carcinoma is one of the leading cancers in Norway. Molecular profiling of the different stages of colorectal cancers is still in demand as molecular markers to stratify patients for correct treatment response is lacking. This study was carried out to evaluate capillary electrophoresis based Sanger sequencing and fragment analysis on the ability to detect somatic *TP53* mutations in fresh frozen- and formalin fixed and paraffin waxed embedded (FFPE) tissue from Dukes' stage B and C colon adenocarcinomas. A small subset of samples was also subjected to ion semiconductor sequencing on the Ion Torrent PGM™ – ion semiconductor sequencing. These methods demonstrated that FFPE tissue performed well with the fragment analysis assay and on Ion Torrent PGM™, while the same samples required additional purification steps for Sanger sequencing. Sanger sequencing revealed the highest number of mutations, and deleted nucleotide composition that were difficult to interpret with Sanger sequencing were successfully verified on the Ion Torrent PGM™. Fragment analysis failed to detect all the mutations found by Sanger sequencing. The fragment analysis assay can detect mutations at 7 nucleotide positions (SNPs) located in exon 5, 7 and 8, of the *TP53* DNA binding domain. DNA was isolated from 37 surgically removed fresh frozen colon tumors and 40 surgically removed FFPE preserved tumors to perform *TP53* mutation analysis. Univariate, Kaplan-Meier analyses and the log-rank test for *TP53* and Dukes' stage parameters revealed differences in survival amongst Dukes' stage status, with no such difference for *TP53* status. Most mutations were evaluated as inactive mutants, and the mutation spectra corresponded well to data published in *TP53* databases. The quest for new clinically approved biomarkers continues as cancer genome studies reveals an ever growing repertoire of mutations previously assumed insignificant in carcinogenesis.

Sammendrag

Tykktarms- og endetarms kreft, kolorektal kreft, er en av de hyppigste kreftformene Norge. Molekylær profilering av ulike stadier kolorektal kreft er nødvendig da det mangler molekylære markører for bedre å stratifisere pasienter til korrekt behandling med gode prognoser. Denne studien ble utført for å evaluere to ulike kapillær elektroforese baserte metoder, Sanger sekvensering og fragment analyse, ved å undersøke somatiske *TP53* mutasjoner i ferskt frossent- og formalin fiksert og parafin støpt (FFPE) vev fra Dukes' stadium B og C tumorer. Et lite utvalg av prøver ble sekvensert på Ion-Torrent PGM™ - ion semikonduktor sekvensering. Disse metodene viste at FFPE vev fungerte godt med fragment analyse og Ion Torrent PGM™, mens FFPE prøver krevde ekstra rensetrinn før Sanger sekvensering. Sanger sekvensering detekterte flest antall mutasjoner, og DNA sekvenser som var vanskelig å lese med Sanger sekvensering ble bekreftet med Ion Torrent PGM™. Fragment analyse metoden inkluderer kun 7 nukleotid posisjoner i kodon 175, 245, 248, 273, og 306 i *TP53* DNA bindingsdomene og kan derfor detektere få mutasjoner i forhold til Sanger sekvensering. DNA ble isolert fra 37 ferske frosne- og 40 FFPE kolorektal svulster for *TP53* mutasjons analyse. Univariat, Kaplan-Meier analyse og log-rank test for parameterne *TP53* og Dukes' stadium avdekket forskjeller i overlevelse blant Dukes' stadium, med ingen forskjell ble funnet for *TP53* status. De fleste mutasjoner ble vurdert til å forårsake et inaktivt p53 protein. Distribusjon av mutasjoner stemmer overens med data publisert i *TP53* databaser. Søken etter nye klinisk godkjente biomarkører fortsetter da flere sekvenseringsstudier av kreftgenom avslører et stadig voksende repertoar av mutasjoner som tidligere er antatt å ha liten betydning for kreftutvikling.

TABLE OF CONTENTS

1 INTRODUCTION

1.1 Introduction to Cancer

Cancer is a heterogeneous and progressive genetic disease that evolves through multiple steps of alterations in the genome. There are hundreds of different types of cancers, most of which is known to differ in tumorigenesis, histopathology, biological and clinical behavior, as well as in treatment response. With such diversity there are some fundamental characteristics all cancer cells share in common and must acquire to reach full tumor development and malignancy. Through a succession of clonal expansions initiated by genetic alterations a cell must; (1) gain a growth advantage and become self-sufficient in growth-stimulating signals; (2) evade growth-suppressors to sustain proliferation; (3) gain the ability to invade surrounding tissue and metastasize to distant organs; (4) evade replicative cell senescence to gain unlimited replication; (5) induce angiogenesis to sustain blood supply/neovasculature; (6) evade the apoptotic machinery to resist programmed cell death; (7) alter energy metabolism to sustain active proliferation; and (8) evade destruction by the immune system. The acquisition of these hallmark characteristics evolves through two enabling hallmarks; (9) genome instability and mutations, and epigenetic changes that can trigger the initiation of clonal expansion; and (10) inflammations can contribute to the aforementioned characteristics and enhance tumorigenesis (Hanahan & Weinberg 2000; Hanahan & Weinberg 2011).

1.2 Colorectal Carcinoma

Cancers in the bowel region (bowel, colon, and rectum) are usually referred to as colorectal carcinoma (CRC). CRC evolves through well-defined morphological changes in the epithelia of the lumen. Colonic lesions or dysplasia can grow and develop into benign adenomatous polyps that further grow in size to adenoma. Unless surgically removed, the growing adenoma can evolve to malignant invasive carcinoma that may metastasize. Histology from each stage of CRC-tumor development alongside molecular characterization serves as a model system for multistep carcinogenesis (see section 1.7) (Fearon & Vogelstein 1990; Muto et al. 1975).

1.3 Epidemiology

Norway has the highest incidence of cancers in the colon and rectum among men and women in the western world. In 2012, 2717 (1294 men; 1423 women) new colon cancer cases was registered. The numbers for rectal-, rectosigmoid-, and anal cancers were collectively 1304 (540 women; 764 men) new cases in 2012. The most common cancer types for both sexes are prostate- (men), breast- (women), and lung cancer, including malignant melanoma. Although

CRC are most common among women over 70 years of age there is a slight increase in CRC among women in general. In men, colon cancer is one of the most common cancers, while rectal cancers have stabilized since the 1980s.

Mortality in 2012 from colon cancer was 1206 (595 women; 611 men), and 298 (181 women; 217 men) for rectal-, rectosigmoid-, and anal cancer. Prevalence at 5 years after diagnosis is 5906 for colon-, and 3458 for rectal-, rectosigmoid-, and anal cancer. The relative survival rate for colon cancer in 2012 was 60% for men and 63% for women, an increase of about 30% since the 1970s. In rectal cancers the relative survival rate in 2012 was 65% for women and 68% for men. The cumulative risk of developing colon cancer by the age of 75 (2008-2012) was 2.8% for women and 3% for men. For rectal-, rectosigmoid-, and anal cancer the cumulative risk was 1.3% for women and 2.1% for men (Norwegian Cancer Registry – Cancer in Norway 2012, www.kreftregisteret.no).

1.4 The Anatomy of the Large Intestine

The human large intestine (Figure 1A) is about 1.5 meters long and is structurally divided into the cecum, colon, rectum, and anal canal. The colonic segments are the ascending-, transverse-, descending-, and sigmoid parts, including two bends on each side of the transverse colon; the hepatic flexure on the right side, and the splenic flexure on the left side. The bowel wall surrounding the lumen of the large intestine is composed of four histological layers (Figure 1B); mucosa, submucosa, tunica muscularis, and serosa (peritoneum). Three sub-layers constitute the mucosa; a simple columnar epithelium, the lamina propria to support

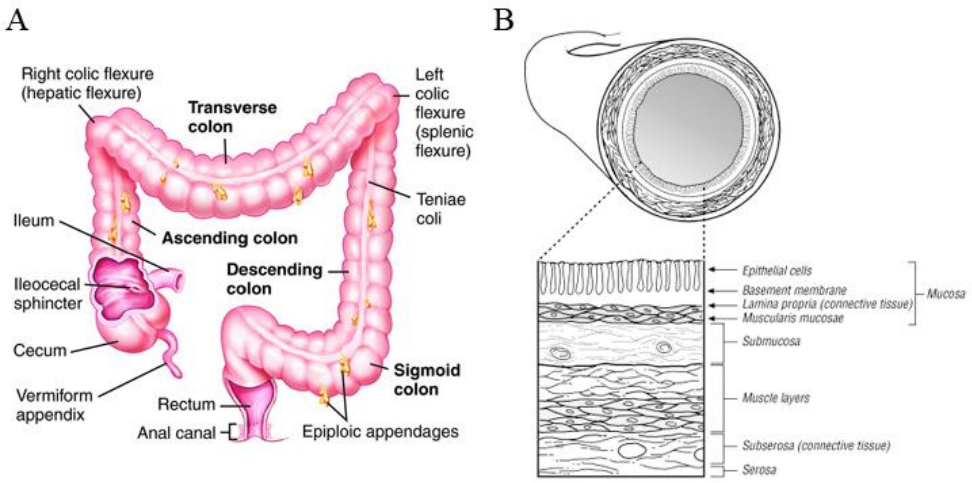


Figure 1: The large intestine. (**A**) The locations and sub-sites of the colon and rectum (large intestine) (Mosby's medical dictionary. (8th ed.) (2009). St Louis). (**B**) The tissue layers of the colon wall (www.oreilly.com).

the epithelium, and the muscularis mucosa; a thin smooth muscular layer external to the propria. The submucosa is a layer of highly vascular connective tissue, and the tunica muscularis consists of inner circular- and outer longitudinal muscle layers involved in peristalsis. The peritoneum of serosa is the outer-most layer and consists of loose connective tissue that confines the large intestine. Retro-peritoneal sections without serosa are the posterior- ascending and posterior descending colon which are covered in adventitia (Yeatman 2013).

1.5 Histopathology of the Colon – The Polyp-Cancer Sequence

The human large intestine is lined with a single layer of epithelial cells that covers the invaginated intestinal crypts of the digestive surface (as seen in the first photo of Figure 3). The epithelium is renewed every 4-5 days by dividing stem cells that lie protected at the bottom of the crypts which generates the four cell types found in the intestinal layer; (1) absorptive cells for uptake of nutrients; (2) mucus secreting goblet cells; (3) paneth cells for biological defense; and (4) regulatory enteroendocrine cells (Alberts et al. 2008). Most colorectal cancers start in the epithelium of the colonic mucosa as microscopic lesions called aberrant crypt foci (ACF) (Figure 2). ACF are the first visible signs of alteration in the colon and is generally characterized as larger than normal crypts with irregular shape and multilayered epithelium. ACF are histologically classified into non-dysplastic, dysplastic, hyperplastic, and mixed, each with its own distinct histopathological characteristics (Alrawi et al. 2006). ACF are presumed to be precursor (Alrawi et al. 2006) of benign polyps and adenomas that subsequently grow into malignant carcinomas according to the polyp-cancer sequence (Muto et al. 1975) (Figure 3). Three histological types of benign tumors exists; adenomatous polyp (tubular adenoma), villous adenoma (villous papilloma), and an intermediate type; tubule-villous adenoma, villo-glandular adenoma or papillary

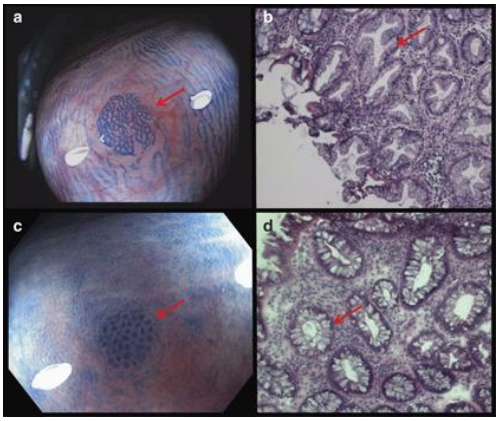


Figure 2: Macroscopic and histologic images of ACF (red arrows). (**a**) hyperplastic serrated ACF. (**b**) serrated hyperplastic colonic crypts characterized by stellate-shaped luminal crypts and a prominent component of columnar cells with microvesicular cytoplasm. (**c**) non-serrated hyperplastic ACF. (**d**) non-serrated hyperplastic crypts, which are characterized by a lack of crypt serration and prominent goblet cells. (a) and (c) are endoscopic images of ACF. (b) and (d) are hematoxylin and eosin stained images of ACF (Anderson et al. 2010).

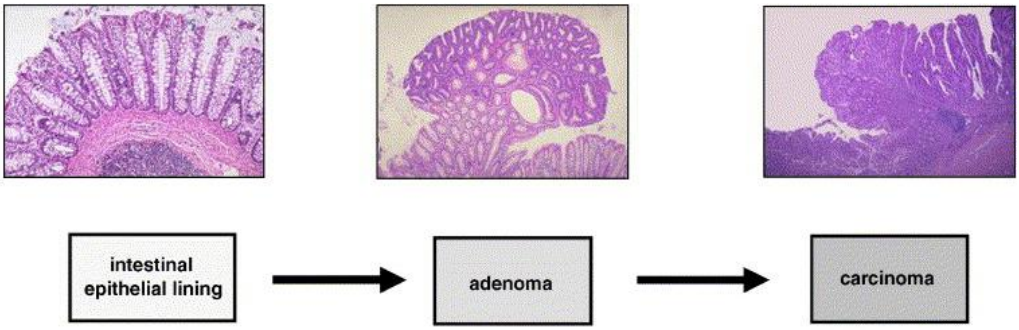


Figure 3: Polyp-cancer sequence. Histology slides of normal epithelia, adenoma and carcinoma according to the polyp-cancer sequence (Cardoso et al. 2007).

adenoma. Although adenomatous polyps are far more frequent than both villous-adenomas and the intermediate types, villous adenomas have a much higher malignant potential. The adenomatous polyps (tubular-adenoma) are structurally embedded in the lamina propria as branching tubules while villous-adenomas are pointed protrusions of lamina propria covered by a layer of epithelia that can be embedded in the muscularis mucosae. The intermediate tubulo-villous adenoma can consist of both the tubular and villous structures. The malignant potential of colonic polyps generally increases with size and only a fraction of all polyps grow large enough to become malignant during a human lifetime (Muto et al. 1975).

1.6 Oncogenes and Tumor Suppressor Genes

The malignant transformation of a normal cell is a complex multistep process caused by defective genes involved in cell growth, cell division, and cell death. Genetic defects can be induced by external chemical and physical agents, intracellular rearrangements, inherited germ line or somatic mutations, epigenetic changes and certain viral infections (e.g. Human Papillomavirus (HPV), Epstein Barr virus (EBV), and Human Herpesvirus-8 (HHV-8). Two distinct functionally and heterogeneous groups of defective genes associated with malignancy are the growth-promoting oncogenes and the growth-constraining tumor suppressor genes (TSG). Understanding the opposite mechanisms by which these two groups of genes operate has been fundamental for unraveling the multistep progress of carcinogenesis (Damania 2007; Land et al. 1983; Steele et al. 1998; Weinberg 1994).

1.6.1 Oncogenes

In normal cells oncogenes are called proto-oncogenes. Proto-oncogenes encode various regulatory proteins (e.g. growth factors, growth factor receptors, signal transducers, and nuclear transcription factors) involved in cell growth and cell division (Land et al. 1983;

Ramalakshmi & Muthuchelian 2011; Weinberg 1994). Oncogenes are activated through mutations in proto-oncogenes leading to amplification of growth-promoting genes whose encoded proteins can bypass cell cycle control mechanisms and activate downstream transcription (Martin 2004; Weinberg 1994). Mechanisms of activation can be somatic mutations (Reddy et al. 1982; Santos et al. 1983; Tabin et al. 1982; Taparowsky et al. 1983), gene amplification/gene copy number (Alitalo et al. 1983; Dalla-Favera et al. 1982b; Schwab et al. 1983), or chromosomal rearrangements/translocation (Dalla-Favera et al. 1982a; Leder et al. 1983; Mushinski et al. 1983). Viral infections can also induce tumorigenesis through insertion of oncogenic enhancer elements and promoter sequences to activate proto-oncogenes (Blair et al. 1981; Conrad & Botchan 1982; DeFeo et al. 1981; Gruss et al. 1981; Hayward et al. 1981). Oncogenes are said to be dominant as alteration in only one allele of the proto-oncogene is needed to acquire a gain-of-function (Figure 4A) phenotype (Weinberg 1991).

For example, somatic mutations in the *RAS* oncogene family (H-*RAS*, K-*RAS*, N-*RAS*) is found in about 30% of all human tumors (Fernandez-Medarde & Santos 2011), with K-*RAS* mutations found in 40-45% of colorectal tumors (Cho & Vogelstein 1992; Fearon & Vogelstein 1990; Fernandez-Medarde & Santos 2011). RAS are monomeric GTPases involved in signal transduction pathways affecting cell survival, growth, and proliferation. GTPases serves as molecular switches that alternates between an active (GTP-bound) and inactive (GDP-bound) conformational state. Inactive RAS is anchored to the cytoplasmic surface of the plasma membrane where it is activated by surface receptors to relay signals to downstream effector proteins. Upstream signals induce RAS guanine nucleotide exchange factors (RAS-GEFs) to activate RAS by dissociation of GDP, thereby affiliating the uptake of cytosolic GTP. Inactivation is induced by RAS GTPase-activating proteins (RAS-GAPs) that catalyzes RAS to hydrolyze its bound GTP to GDP (Malumbres & Barbacid 2003). Mutations in RAS is confined to residue 12 (Gly12) and 61 (Gln61) which obstruct its binding capacities to GAP and consequently affects GTP hydrolysis, leaving a constant flow of growth-promoting signals to activate downstream effectors that can promote tumorigenesis (Malumbres & Barbacid 2003; Reddy et al. 1982; Santos et al. 1983; Taparowsky et al. 1983).

1.6.2 Tumor Suppressor Genes

TSGs encode regulatory proteins involved in suppressing cell growth and proliferation, as to maintain cell- and tissue architecture through pathways like, cell cycle check-points, DNA repair mechanisms, senescence and apoptosis. Mutant inactive or lost (deleted) TSG protein

products lose their maintenance ability, and liberate the cell, by definition, to continue to grow and proliferate to neoplastic growth (Weinberg 1991). Mechanisms of TSG inactivation are inherited and somatic mutations (Knudson 1971), usually followed by chromosomal rearrangements and deletions (Baker et al. 1989; Dong 2001) leading to loss of heterozygosity (Kanazawa et al. 2002). Epigenetic changes can also silence TSGs through DNA methylation (Jones & Laird 1999). Viral infections induce tumorigenesis by inactivating TSGs through onco-proteins that target and form complexes with TSGs protein domains ("pockets") necessary for normal binding properties (Helt & Galloway 2003; Hu et al. 1990; Münger et al. 2004; Weinberg 1995). Unlike oncogenes, TSGs are recessive as alteration in both alleles of the gene is necessary for a loss-of-function (Figure 4B), known as the two-hit hypothesis (Knudson 1971; Weinberg 1991). The two-hit hypothesis was first attributed to the childhood cancer retinoblastoma where Knudson, 1971, reported that children with retinoblastoma had a dominantly inherited form with one germline mutation and that a second mutation had occured in somatic cells. Adults affected with retinoblastoma had a non-hereditary form where both mutations occurred in somatic cells. It is now clear that retinoblastoma is caused by mutations in both alleles of the retinoblastoma gene, *RB1*. pRB is involved in cell cycle regulation where it acts as a suppressor for transcription factors of the E2F family. Phosphorylation of pRB realeses E2F and leads to S-phase entry and replication in the cell cycle (Burke et al. 2012; Nevins 2001). However, TSGs can also show oncogenic behavior through a dominant-negative effect in which, for example, mutant p53 protein can dimerize with wild-type (wt) p53 dimers and inactivate otherwise functional p53 tetramers (Fearon & Vogelstein 1990).

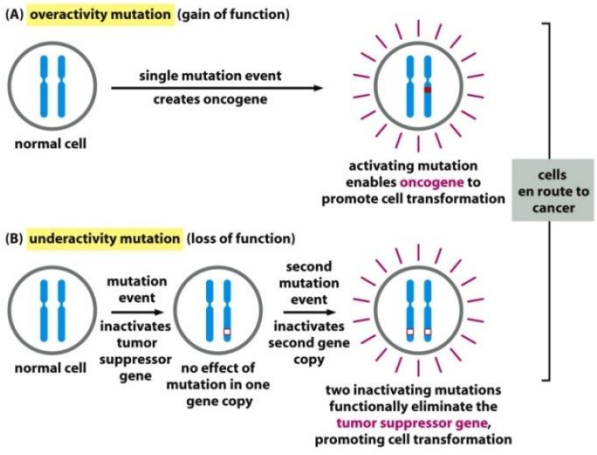


Figure 4: Dominant and recessive mutations in oncogenesis. Solid red boxes represents activating mutations, hollow red boxes represents inactivating mutations. (**A**) Dominant; a gain-of-function mutation in a single allele of oncogenes can drive tumor formation. (**B**) Recessive; a loss-of function in both alleles of tumor suppressor genes are necessary to drive tumor formation (Alberts et al. 2008)

1.7 Genetic Model of Colorectal Carcinoma - The Adenoma-Carcinoma Sequence

The adenoma-carcinoma sequence was first proposed by Fearon and Vogelstein in 1990 as a multistep genetic model for the development of colorectal carcinoma (Figure 5). The transition from normal epithelia to benign adenomas and malignant carcinomas involves the accumulation of genetic alterations in oncogenes (*KRAS*) and TSGs (*APC*, *DCC*, *TP53*).

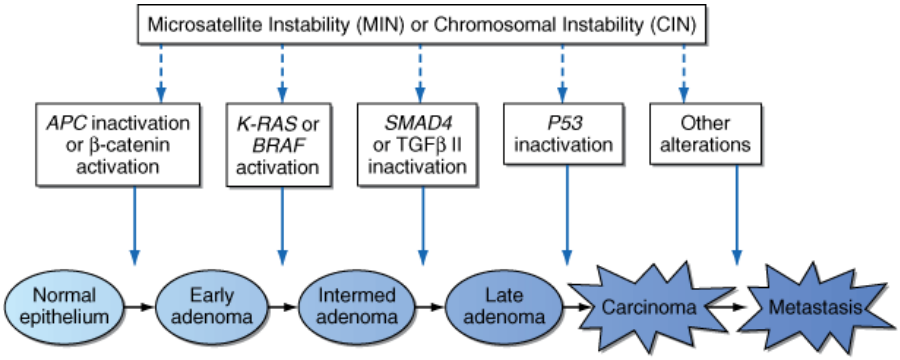


Figure 5: The adenoma-carcinoma sequence. Genetic model of tumor formation - from normal epithelium to carcinoma - involves the inactivation of TSGs and activation of oncogenes initiated through CIN or MIN (Fauci et al. 2008).

In 2012, The Cancer Genome Atlas (TCGA) project identified a total of 32 mutated somatic genes in colon and rectal cancers, and it is now recognized that CRC evolves through three main pathways; the chromosomal instability (CIN) pathway, the microsatellite instability (MSI) pathway, and through epigenetic mechanisms. CIN is characterized by loss or gain of whole chromosomes or chromosomal regions leading to aneuploidy, and is found in approximately 70-80% of CRC cases. MIN is insertions or deletions of short repetitive DNA segments (microsatellites) of 2-5 nucleotide bases and is associated with defective mismatch repair (MMR) genes observed in 15% of CRC cases. Epigenetic changes like promoter methylation of CpG (cytosine-phosphate-guanine) islands (CIMP) is a distinct phenotype in CRC associated with MIN (Cardoso et al. 2007; Kawasaki et al. 2008; TCGA 2012). Oncogenic pathways have distinct genetic characteristics, but in general, early events in carcinogenesis involves gain/loss of DNA methylation which may result in CIN followed by aneuploidy, and MIN and LOH in TSGs on chromosome 5q (*APC*), 17p (*TP53*), and 18q (*DCC*), and oncogenes on 12p (*KRAS*). Deletion of 5q and *RAS* mutations occurs in early stage adenomas while loss of 17p and 18q are events in late adenomas and carcinomas (Fearon & Vogelstein 1990). Less than 10% of CRCs have been shown to harbor mutations in all 3-4 genes mentioned here which are evident of multiple pathways to CRC, however, the inactivation of APC is considered the first event in CRC-development found in more than

80% CRC-cases. The transition from adenoma to carcinoma is associated with loss of 17p (*TP53*) in 75% of CRC and in several other cancers. Amplification and rearrangements are not common events in CRC (Bodmer 2006; Fearon & Vogelstein 1990).
APC mutations found in dysplastic ACF is associated with malignant carcinoma, while e.g. KRAS2 mutations found in heteroplastic ACF is associated with benign adenomas (Cardoso et al. 2007).

1.8 Cellular Pathways in Colorectal Cancer

Altered gene expression deregulates multiple cellular signaling pathways during tumor progression. Common pathways found altered in CRC are the Wnt, TGFβ, MAPK, PI3K and p53 pathways (Figure 6) (TCGA 2012). The Wnt (Wingless) pathway regulates β-catenin, a signal transducer leading to transcriptional activation of several critical cancer genes (*TCF-1*, *Myc* and, *CCND1*), but Wnt signaling also sustain the proliferation of stem cells in the intestinal crypts before they migrate to differentiate. The APC protein is a crucial component of the Wnt signaling pathway as a major element of a degradation complex consisting of APC, Axin and glycogen synthase kinase (GSK3β). In the absence of a Wnt signal the APC-complex ubiquinate β-catenin for degradation in the proteasome and no transcription is activated. In the presence of a Wnt signal, the APC-complex is disengaged leaving β-catenin

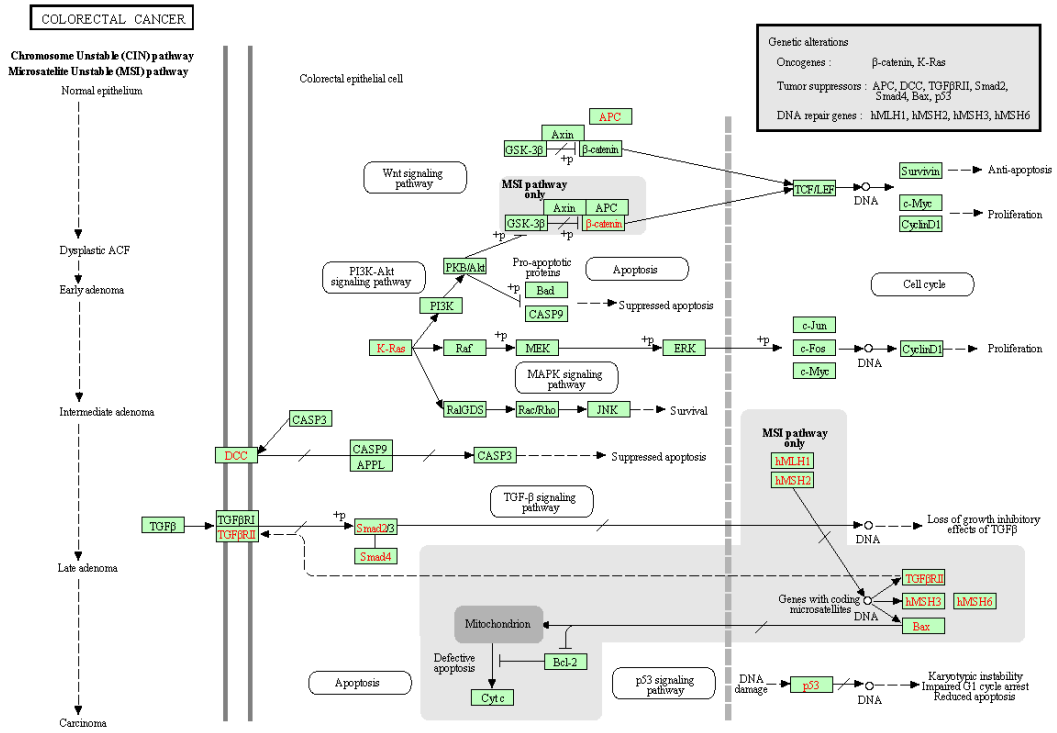


Figure 6: Cellular pathways in colorectal cancer. Genes in red are affected in several common pathways during colorectal cancer development (www.genome.jp (hsa05210)).

to accumulate and translocate to the nucleus where it stimulates transcription (Ougolkov et al. 2004). During tumorigenesis the Wnt pathway is thus activated by inactive APC rather than the normal cascade from extracellular ligands.

1.9 *TP53* – Tumor Suppressor Protein p53

p53 is a nuclear protein and a transcription factor found at basal levels in normal cells. In response to a wide range of stress factors (e.g. metabolic changes, nucleotide pool depletion, DNA damage, telomere erosion, hypoxia, UV and ionizing radiation, oncogene activation, nutrient and oxygen deprivation, and heat shock) p53 can respond to help the cell repair damages and survive, or, in more severe cases, it can induces either cell cycle arrest, apoptosis, or senescence (Figure 7) (Burns & El-Deiry 1999; Gottlieb & Vousden 2010).

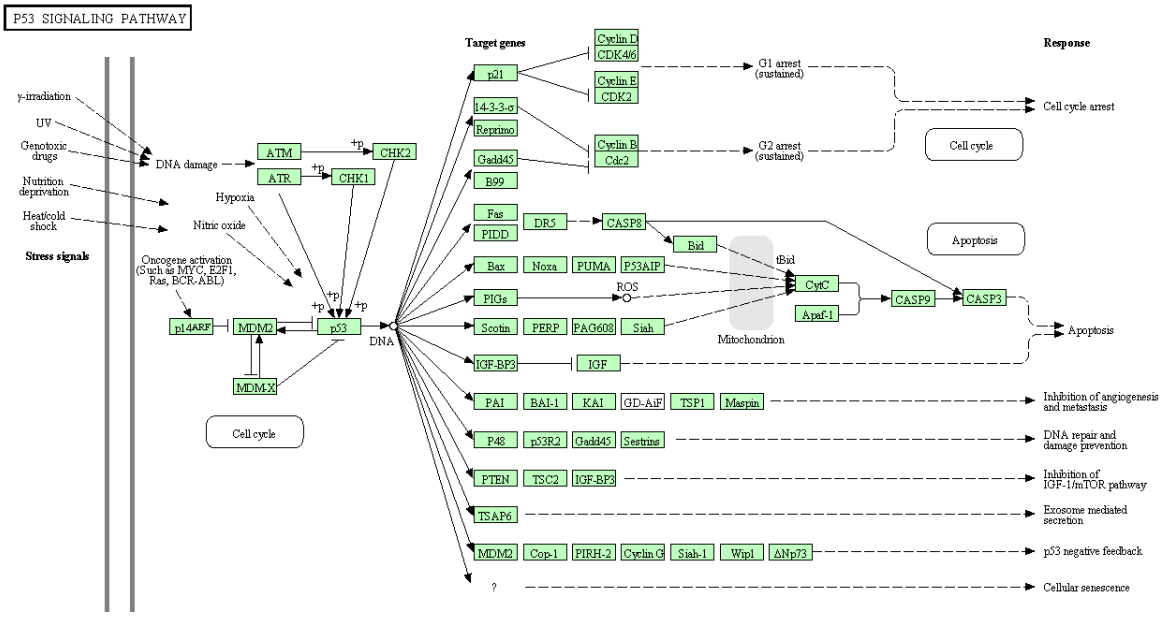


Figure 7: p53 signaling pathways. Stress signals induce p53 activation through several mechanisms leading to an array of cellular responses (www.genome.jp (hsa04115)).

The *TP53* gene is one of the most studied and frequently mutated gene in human cancers (Olivier et al. 2010). Located on the short arm of chromosome 17 (17p13) (Figure 8) (Benchimol et al. 1985; McBride et al. 1986) *TP53* is 20 kb long composed of 11 exons, of

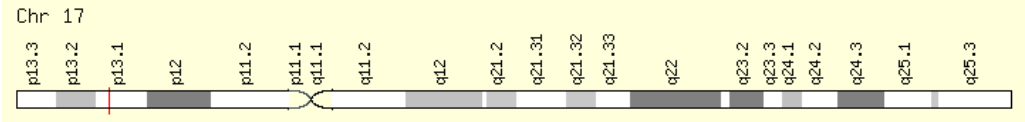


Figure 8: Chromosome 17. *TP53* is located on the short arm p13.1 indicated by the red line (www.genecards.org).

which exon 1 is non-coding followed by a 10 kb intron (Figure 9A) (Lamb & Crawford 1986) *TP53* encode a 393 amino acid long (Harlow et al. 1985), 53 kDa (Crawford et al. 1981; Lane & Crawford 1979) tumor suppressor protein, p53 (Finlay et al. 1989).

1.9.1 p53 Structure

Cellular p53 is a tetramer (dimer of dimer) of four subunits (Figure 9B) (Chen et al. 2010). Each subunit can be divided into 5 domains with distinct functions, and it contain 5 evolutionary highly conserved domains (HCD) (Hjortsberg et al. 2008); (1) The amino-terminus (N-terminus) (1-42) contains the acidic transactivation domain, the Mdm2 protein binding site, and the HCD I (Brady & Attardi 2010; Burns & El-Deiry 1999); (2) A proline rich domain (40-92) that contain a second transactivation domain involved in apoptotic activation (Venot et al. 1998); (3) The core DNA binding domain (101-306) for sequence-specific binding to response elements. It contains the HCD II to V, and it is the domain in which 90% of p53 mutations are found (Hjortsberg et al. 2008); (4) A tetramerization domain (307-355) that consist of a β-strand and an α-helix for quaternary stabilization, including a nuclear export signal (NES) (Joerger & Fersht 2010); (5) The carboxy-terminus (C-terminus) (356-393) regulatory domain (CAD) includes 3 nuclear localization signals (NLS) and a second NES, and is involved in non-sequence-specific binding to the promoter TATA-element to supress many target genes (Coutts et al. 2009; Seto et al. 1992). Both the amino- and carboxy-terminus are sites subjected to numerous post translational modifications (e.g. phosphorylation, acetylation, ubiquitination, sumoylation, and neddylation) to modulate p53 activity and localization during homeostasis and stress-induced responses (Kruse & Gu 2009).

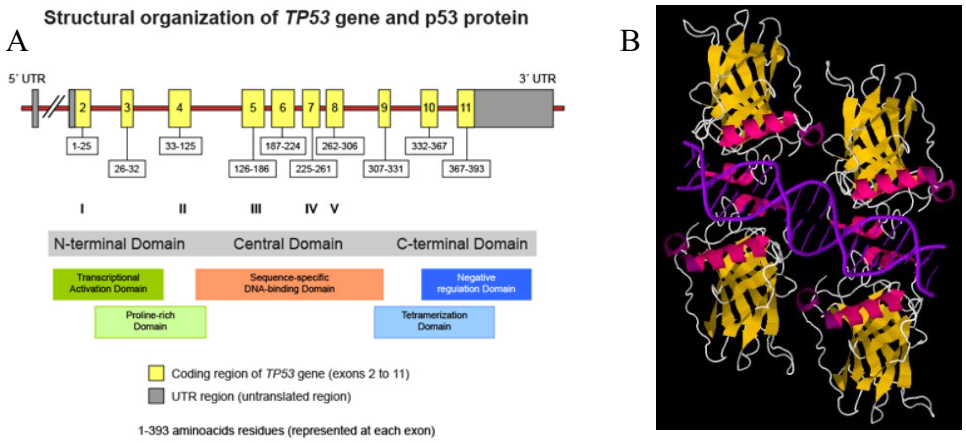


Figure 9: *TP53* domains and structure. (**A**) Exon boundaries with codon numbers and functional domains of the *TP53* gene (www.stjude.org). (**B**) Crystal structure (3KMD.pdb) of a self-assembled p53 tetramer with its core domain bound to a full consensus site in the DNA strand (www.pdb.org).

1.9.2 p53 Regulation

The transcriptional activity of p53 is kept under tight control in unstressed cells through Mdm2-mediated proteasomal degradation. Mdm2 binds to the p53 N-terminal region and undergoes a conformational change to fit a "ubiquitination signal" located in the core DNA binding domain in p53 to poly-ubiquinate and target p53 for 26s proteasomal degradation. p53 can either be degraded in the cytoplasm, or in the nucleus during down-regulation when rapid decrease in p53 transcriptional activity is necessary. Mdm2 also mediates p53 nuclear export by mono-ubiquitinate lysine residues in the C-terminal region and the core DNA binding domain in p53 which exposes the p53 C-terminal NES, leading to its nuclear export. Stable mono-ubiquitinated p53 proteins have been found translocated to control mitochondria-directed apoptosis, and are thought to have other, yet unknown, cytoplasmic functions.

The transcriptional activation of *Mdm2* is regulated by p53. In this way, p53 is regulated in a negative autoregulatory feedback-loop where poly-ubiquitination and degradation of p53 occurs during high Mdm2 levels, and activity directed mono-ubiquitination of p53 occurs during low Mdm2 levels. (Coutts et al. 2009; Kubbutat et al. 1997; Moll & Petrenko 2003).

1.9.3 p53 Activation

p53 is activated through three independent pathways; (1) in response to DNA damage p53 is phosphorylated by the protein kinases ATM, ATR, Chk1 and Chk2; (2) oncogene activation *of Ras or Myc triggers* aberrant growth signals to stimulate $p14^{ARF}$ protein to activate p53; and (3) chemical agents, radiation, and protein-kinase inhibitors triggers p53 phosphorylation by ATR and casein kinase II. Upon p53 activation, p53 degradation ceases, allowing p53 to carry out its transcriptional activation on target genes that ultimately leads to either cell-cycle arrest, apoptosis, or senescence (Vogelstein et al. 2000).

1.9.4 p53 in Cell-Cycle Control

During the cell cycle a cell grows and divides into two daughter cells. The cell cycle involves the interphase and mitosis. The interphase consists of Gap 1 (G_1), Synthesis (S), and Gap 2 (G_2), and the mitotic (M) phase consists of prophase, metaphase, anaphase, and telophase, and cytokinesis (Figure 10). During the G_1 phase the cell grows in size before entering the S phase for chromosome duplication, and then exceeds to the G_2 phase for further growth and preparation for mitosis. In the mitotic phase the sister chromatids are disassembled and distributed in two daughter cells (cytokinesis).

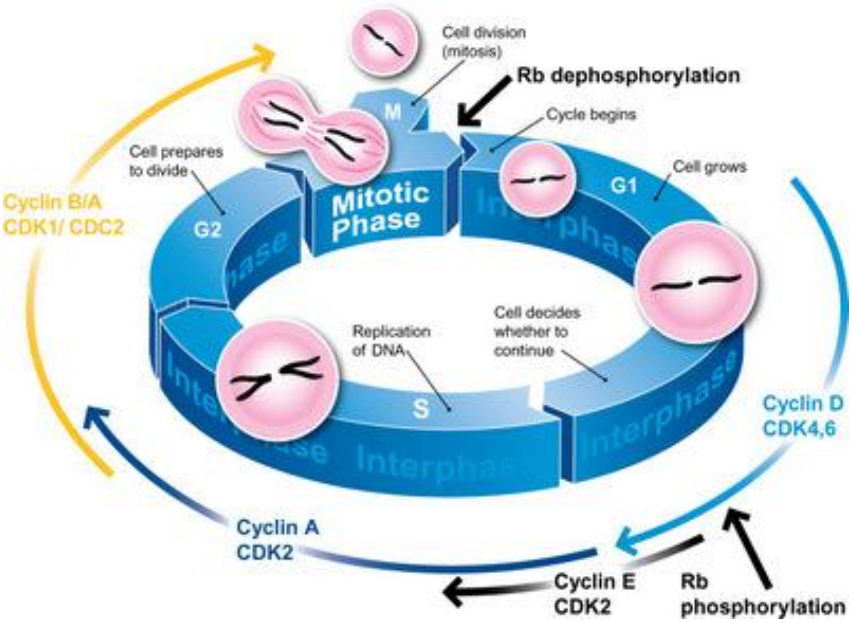


Figure 10: Cell cycle control system. Cell cycle starts in G_1. Each gap in the circle represents a major checkpoint controlled by the different cyclin dependent kinases and their cyclins (http://www.abcam.com).

The cell cycle is tightly regulated by the cell-cycle control system, mainly the cyclin-dependent kinases (Cdks) and their associated cyclins, at three major check-points positioned at the transition from G_1 to S, G_2 to M, and from metaphase to anaphase. Progression through each of these checkpoints is blocked in response to any complications that might cause damage in the duplicated cells, and cell cycle arrest occurs until the problem is fixed and the cell can proceed to the next phase. In response to DNA damage during the cell cycle, p53 is phosphorylated by kinases such as DNA-dependent kinase, ATM, Chk1 and Chk2, to mediate $p21^{WAF1/CIP1}$ expression. $p21^{WAF1/CIP1}$ is a negative regulator of Cdks responsible for G_1/S and G_2/M transition (Vogelstein et al. 2000). To sustain G_1 arrest, p53 mediated expression of p21 inhibits cyclin D/cdk4 and cyclin E/cdk2 kinase activity required for G_1/S transition. Normally, the G_1 cyclin/Cdks regulate the G_1/S transitional cyclins through the E2F transcription factor. E2F is repressed by pRb which dissociates E2F upon phosphorylation by G_1 cyclin/Cdks which promotes E2F transcription of cyclins required during S-phase (el-Deiry 1998; Helton & Chen 2007; Hiyama et al. 1998).

During S-phase there are two internal check-points to ensure correct replication of chromosomes. The intra-S check-point is activated in response to DNA damage, and the replication check-point prevents collapse of the replication fork on DNA polymerase during DNA damage repair. p53 is, however, not associated with S-phase arrest, instead p53 is redirected to promote DNA repair mechanisms, although there are indications that a p53

isoform, Δp53, might be involved in intra-S arrest by activating p21 and 14-3-3σ (Helton & Chen 2007). The G_2/M check-point prevents the segregation of damaged chromosomes during mitosis. In the event of G_2/M arrest, p53 mediates p21 expression to inhibit cyclin B/Cdk1 activity required for M-phase entry. p53 also mediates the expression of other proteins to sustain G_2 arrest, such as 14-3-3σ, GADD45, B99, and Reprimo. The expression of 14-3-3σ can direct cyclin B/Cdk1 out of the nucleus to sustain G_2 arrest (el-Deiry 1998; Helton & Chen 2007; Vogelstein et al. 2000).

1.9.5 p53 Mediated Apoptosis

Apoptosis is a way of eliminating damaged or oncogenic cells from tissues by programmed cell death. p53 induces apoptosis through a proteolytic cascade of caspases within the intrinsic mitochondrial and the extrinsic death receptor pathway. The intrinsic pathway is activated through ATM, ATR, Chk1, or Chk2 phosphorylation of p53 in response to damaging stimuli. Phosphorylated p53 induces transcription of pro-apoptotic BH3-only proteins within the Bcl-2 family, like Bax, Noxa, and Puma, and can repress transcription of anti-apoptotic genes like *Bcl-2*, including the inhibitor of apoptosis gene, *Survivin*. Upregulation of Bax, Noxa or Puma triggers the release of cytochrome c from the mitochondria, which binds to Apaf-1. Apaf-1 oligomerizes to form the apoptsome which recruit and activate initiator caspases. Activated caspases, in turn, activates downstream effector caspases of the proteolytic cascade that lead to cell death. A second approach is seen in p53-dependent PIG activation where PIG produces reactive oxygen species to induce cytochrome c release from the mitochondrion to trigger the proteolytic cascade.

The extrinsic death receptor pathway relies on tumor necrosis factor (TNF) receptors that contain a death receptor domain to bind incoming FasL and TRAIL (Tumor Necrosis Factor-related apoptosis-inducing ligand) ligands. FasL ligand binds to Fas/APO1 receptor or a decoy FAS receptor. Activation of Fas/APO1 recruits the adaptor protein FADD to initiate the formation of the death-inducing signaling complex (DISC) and recruit initiator caspases to be activated in the DISC. Activated caspases then activates effector caspases to induce cell death. The TRAIL receptors, KILLER/DR5 and DR4, and the decoy receptors TRID and TRUNDD, triggers a caspase cascade through adaptor proteins and initiator caspases that results in cell death. p53 can induce transcription of all four TRAIL receptors, although the most common p53-dependent apoptosis is through the intrinsic pathway (Amaral et al. 2010; Burns & El-Deiry 1999).

1.9.6 p53 in Senescence

Cellular senescence is also a way of eliminating damaged or oncogenic cells from tissue, except senescent cells does not undergo programmed cell death. Senescent cells appear to exhibit enlarged and flattened morphology and are characterized by a permanent, irreversible state of growth arrest without the ability to synthesize DNA or proliferate, while still maintaining active metabolism. Telomere shortening, DNA damage, oncogenic stimuli and tumor suppressive activity, or chromatin remodeling, can activate senescence in which p53 acts as a critical regulator in response to senescence biomarkers SA-β-Gal and SAHF (Senescence-Associated Heterochromatic Foci), with the ability to promote expression of p21, PML plasminogen activator inhibitor (PAI-1), and DEC1, which also accounts as senescence biomarkers. Many p53 isoforms and the p53 homologs p63 and p73 are implicated in cellular senescence (Itahana et al. 2001; Qian & Chen 2013).

1.10 *TP53* Mutations

While p53 plays an important role in regulating normal cell growth, the *TP53* gene is frequently mutated in human cancers, leading to a truncated or inactive p53 protein function. About 3.600 publications have reported more than 35.000 *TP53* mutations within numerous tumor types and cell lines (Figure 11). Currently there are two curated databases for *TP53* mutations, IARC TP53 Database (R16, November 2012) and Universal Mutation Database p53 (UMD-p53) (R1, June 2012). Both databases compile the occurrence of somatic mutations, germ line mutations, functional data on mutant proteins and *TP53* gene status of

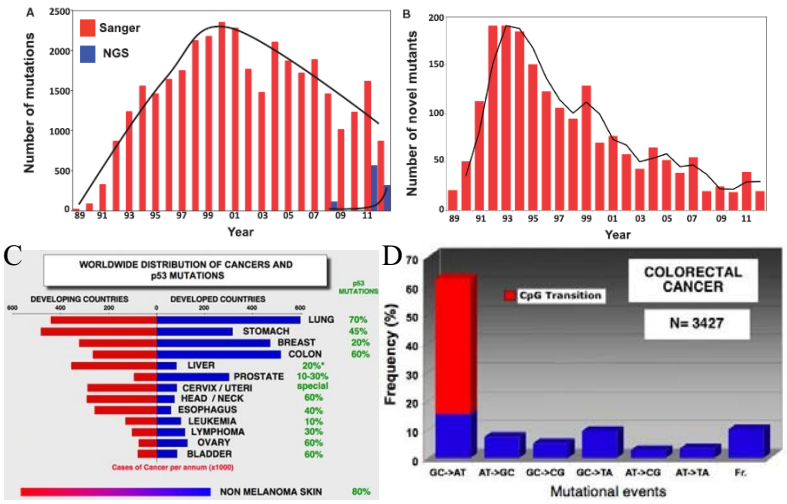


Figure 11: Trends in TP53 analysis. (**A**) Represents published mutations detected by Sanger methodology (red) or NGS (blue) pr. year. (**B**) Novel mutations registered pr. year. (**C**) Distribution of p53 mutations in different cancers worldwide. (**D**) Common mutational events found in CRC (Leroy et al. 2013).

cell-lines. The UMD-p53 database provides several datasets, user friendly analytical tools to manage p53 mutations and generate ready-to-use tables with information on each *TP53* mutation status, including software to analyze newly discovered *TP53* mutants – making this an interesting website to conduct research on *TP53* mutations. However, when interpreting *TP53* mutations, one must be cautious of the heterogeneity within the observed variety for specific loci (especially codon 175, 248, and 273), differences in frequency between tumor types, and the mutant effect on biological properties. Several hot-spot mutations (codon 175, 245, 248, 273, 282) found within the DNA binding domain is connected to key residues necessary for protein-protein interactions of p53 target genes.

More than 70% of *TP53* mutations are missense mutations found outside hot-spot regions, while roughly 30% accounts for rare variants listed only 1-3 times. Such rare variants were possibly generated from artifacts in PCR and sequencing errors, typing errors and incorrect codon assignment, and it is estimated that 2-5% of *TP53* mutations listed in the database are incorrect.

TP53 mutations exhibit distinct features in various cancers. The majority of mutations found in CRC patients are GC>AT transitions at CpG dinucleotides, while in lung cancer it is GC>TA transversions (Figure 11D). Tandem mutations in skin cancer are also linked to UV-exposure, something rarely seen in internal tumors.

TP53 contains 42 CpG dinucleotides frequently mutated. The mutational effect of CpG dinucleotides should be evaluated in detail as they can generate different mechanism depending on the residue affected in the coding sequence. These CpG sites can occur in 3 forms; CGN (type I), NCG (type II), or NNC GNN (type III). A transition at C or G in type I lead to amino acid substitution. In type II, only transition at the C residue will lead to an amino acid substitution. Due to the degeneration of the genetic code, transition at type II G residue has no effect, which is also the case for type III C residue. In type III, only transition at G will lead to an amino acid substitution (Hjortsberg et al., Leroy et al., 2013, Soussi and Beroud, 2003).

10.10.1 p53 Functional Categories

According to their *in vitro* transactivation ability towards eight transcription promoters in yeast, p53 mutations have been classified as active or inactive depending on the location of the affected residue in the protein, e.g. core domain, NH2 domain, COOH domain, α helix and β strand, etc. (Iacopetta et al. 2006; Kato et al. 2003; Soussi et al. 2005). By evaluating

the clinical significance of ~3000 *TP53* mutations registered in the UMD-p53 database, Iacopetta et al. (2006) found that only CRC patients with Dukes' stage D tumors were associated with significantly worse outcome when having inactive *TP53* mutations, while no such significance could be found for Dukes' stage A-C. As the mutation repertoire of tumor suppressor genes, such as *TP53* in the IARC database (http://p53.iarc.fr/) and UMD_p53 database (http://p53.fr/) is growing, the identification of new genetic markers is important for new drug development and to aid oncologists to the best suited treatment for their patients.

1.11 Clinical Characteristics of Colorectal Cancer

1.11.1 Staging

Staging of CRC is based on The TNM (Tumor extent, lymph Node status, Metastases status) Staging System, but are also traditionally classified as Dukes' stage A-D (Table 1). These systems rely on the morphological and histopathological characteristics of the tumor, and the extent of penetration of the tumor through the histological layers of the bowel wall (Figure 12A). Tumor grade describes how quickly the cancer is growing and is determined from the extent of cell differentiation defined as low, moderate (Figure 12B), or high. The malignant character (tumor stage) is defined by the ability of the neoplastic cells to invade the surrounding tissue layers (Figure 12C), and a full malignant cancerous phenotype is fulfilled through the ability to metastasize to distant organs.

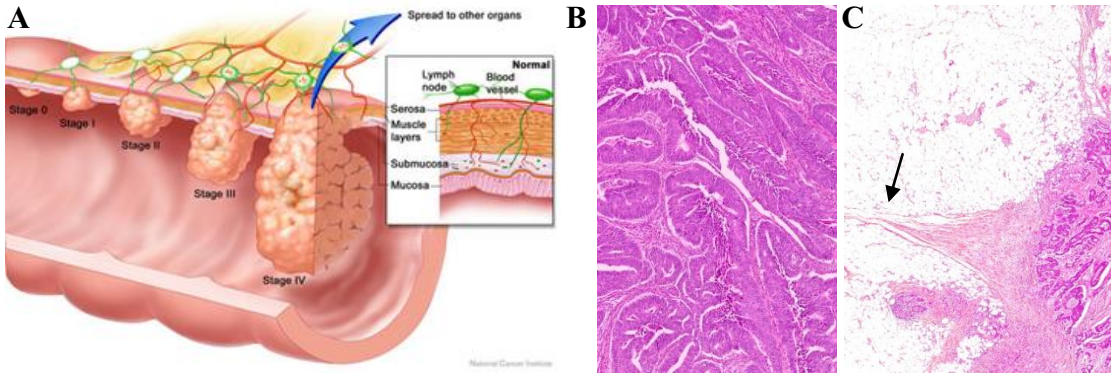


Figure 12: Tumor stage and grade. (**A**) Penetration of the tumor in to the bowel wall here represented by TNM staging (https://visualsonline.cancer.gov). (**B**) Moderately differentiated adenocarcinoma. (**C**) Tumor has grown through the wall and into the adipose layer (black arrow) (courtesy of pathologist Solveig Norheim Andersen, Dept. Pathology, Ahus).

About 85% of CRCs are moderately differentiated and the histology of Dukes stage B and C tumors is very similar. Metastases to distant organs can be either through lymph nodes or

through veins. Metastases through veins lead directly to liver metastases, and are usually more aggressive than metastases to lymph nodes (Compton & Greene 2004).

Table 1: Definition of Dukes' stage A-D, TNM stage, and TNM classification.

Dukes' stage	Dukes' definition	TNM stage	TNM classification
A	Tumor confined to the muscularis propria	I	T1,T2; N0; M0
B	Tumor penetrates the muscularis propria	IIA	T3; N0; M0
		IIB	T4; N0; M0
C	Metastases to lymph nodes	IIIA	T1, T2; N1; M0
		IIIB	T3, T4; N1; M0
		IIIC	Any T; N2; M0
D	Local or distant metastases	IV	Any T; Any N; M1

Table adapted from (Compton & Greene 2004; Dahl 2007). For detailed description on TNM classification see (Compton & Greene 2004).

1.11.2 Prognosis

For patients operated for CRC, the prognosis is dependent on cancer stage at the operation time and patients operated for Dukes' A have best prognosis, while the prognosis is worse for Dukes' C patients. The most challenging group of patients belongs to Dukes' group B patients where 5 years cancer specific survival varies between 75% and 90%. It is debated if some of the Dukes' B tumors may actually be Dukes' C (metastases to regional lymph nodes), and therefore should be offered adjuvant chemotherapy to improve survival (Dahl 2007). There are also genetic pin-points that left- and right sided colon cancer and rectum cancer differs in pathogeneses, and show different treatment response (Yamauchi et al. 2012).

1.11.3 Treatment Strategies

The specific treatment strategy is different for primary, advanced, or metastatic CRC, and depends on the type and stage of CRC based on available clinicopathological characteristics. Common chemotherapeutic drugs used in treating CRC is 5-fluorouracil (5-FU) alone or in combination with leucovorin to increase the effect of 5-FU, capecitabine (5-FU pill form), irinotecan, and oxaliplatin, in addition to monoclonal antibodies (mAb) for targeted therapy which include bevacizumab, cetuximab, and panitumumab (Winder & Lenz 2010). While many surviving patients suffer from recurring tumor growth, the mechanisms behind cytotoxic chemotherapy and anti-cancer drugs that cause cardiotoxicity also needs further elucidation (Raschi et al. 2010).

1.11.4 Prognostic and Predictive Markers

KRAS is the most common biomarker used by clinicians in metastatic CRC patients. Metastatic CRC patients with *KRAS* mutations do not respond to mAbs cetuximab when combined with oxaliplatin, and the addition of cetuximab can even have harmful side effects in some patients. However, the majority of metastatic CRC patients with wt-*KRAS* given cetuximab and FOLFIRI (leucovorin/5-FU/irinotecan combined) show improved response rate and progression free survival. Metastatic CRC patients with wt-*KRAS* are thus recommended to be treated with mAb panitumumab and cetuximab, and it is becoming a common practice to genotype metastatic CRC patients for *KRAS* mutations to determine treatment outcome.

Patients with wt-*BRAF* also show improved progression free- and overall survival over patients with *BRAF* mutations when treated with chemotherapy combined with cetuximab, although, *BRAF* status is not currently offered as a standard test to be used in treatment outcome (Kulendran et al. 2011).

Genetic alterations and several candidate genes show potential as molecular markers (e.g. 18qLOH, *TP53*, *SMAD4*, *VEGF*, *EGFR*, *TS*, *DPD*, *ERCC1*, *PIK3CA*; for a more detailed review see Winder and Lenz (2010) and Kulendran (2011)) but, there are conflicting results at such a detailed level where the difference in treatment and survival can be ruled down to a specific base substitution in a single allele (Kulendran et al. 2011; Winder & Lenz 2010). Patients with MSI tumors also show better prognosis than patients with CIN tumors, and when not subjected to adjuvant chemotherapy, MSI-H tumors show better prognosis than MSI-L tumors in resected stage II and III CRCs (Kulendran et al. 2011). Currently, only two molecular markers are available; MSI as an established prognostic marker, and *KRAS* mutations as an established predictor of anti-EGFR-targeted treatment with cetuximab or panitumumab (Winder & Lenz 2010). The existing histopathological characteristic is not enough to stratify patients with CRC, and there is lack of markers which can be used in selecting patients with Dukes' B tumors that could benefit from adjuvant chemotherapy (Dahl 2007; Kulendran et al. 2011).

1.12 Mutation Detection Methods

There are numerous techniques for detecting genetic alterations using sequence-, hybridization-, or cleavage-based methods depending on what kind of mutations are surveyed, whether at chromosome or nucleic acid level, or whether to determine chromosomal

rearrangement, chromosomal translocation, insertions and deletions, or point mutations. Only sequenced based methods have been employed in this study, and our theoretic focus will therefore be on sequence based technology.

1.12.1 DNA Sequencing Technology

There has been an enormous development in the field of DNA sequencing techniques only in the last 5 years. Since The Sanger method emerged in the 70's, DNA sequencing technologies has been classified as first, second, and now, third- or next generation sequencing (NGS). NGS is a common term used for a new line of technologies utilizing semiconductor-, microfluidics- and nanotechnology to achieve higher throughput, faster turnaround, length of reads, and consensus accuracy. Examples of NGS platforms include 454 GS Junior (Roche), MiSeq (Illumina) and Ion Torrent PGM (Life Technologies). DNA sequencing is now widely used in clinical applications, system- and comparative biology, epidemiology, for whole genome sequencing, including many other disciplines. These new emerging technologies are so advanced that the sequencing discipline is entering a new dimension with the possibilities to determine a wide range of biological mechanisms. Applications available for NGS makes it possible to sequencing a genome in just a few days, determine sites of DNA methylation on a genome-wide scale, monitor active regulatory chromatin depletion and histone-bound DNA activity, protein-DNA interactions, three-dimensional genome structure, and parallel analysis of RNA structure, just to mention a few as there are dozens of other available applications to choose from (Shendure & Aiden 2012).

1.12.2 First Generation Sequencing

Two prominent DNA sequencing techniques emerged in the 1970's; the enzymatic dideoxy chain-termination method developed by Sanger et al. (1977), and the chemical degradation method developed by Maxam and Gilbert (1977). As technology developed the Sanger method was better suited for automation due to the need of fewer toxic chemicals and less radioactive isotopes than the Maxam-Gilbert method. The introduction of polymerase chain reaction (PCR), cycle sequencing with fluorescent dyes, and capillary electrophoresis made the Sanger sequencing the method of choice, a position it still holds today, as the gold standard procedure for sequencing long read fragments (Schadt et al. 2010).

1.12.3 Sanger Dideoxy Sequencing

The Sanger Dideoxy Sequencing technique, also known as the dideoxy-mediated chain-termination method Sanger et al. (1977), used the principle of DNA replication in which DNA

polymerase can incorporate nucleoside analog to the deoxynucleoside triphosphate (dNTP); 2′, 3′- dideoxynucleoside triphosphate (ddNTP) that lack the hydroxyl group (-OH) on the 3′-carbon of the sugar, preventing phosphodiester bridge formation with the 5′-phosphate of the succeeding dNTP (Figure 13).

Essential components for Sanger sequencing requires a single- stranded DNA template (ssDNA), a sequencing primer, DNA polymerase (without exonuclease activity), deoxynucleotides (dNTPs; dA, dT, dC, dG), dideoxy-nucleotides (ddNTPs; ddA, ddT, ddC, ddG), and reaction buffer.Traditionally, either radioactive (^{32}P, ^{33}P, ^{35}S) internal labeled dATP or 5′^{32}P-labeled primers were used in four separate reactions, each containing all four dNTPs (dA, dT, dC, dG) and a small amount of one of the four ddNTP (ddA, ddT, ddC, ddG). DNA polymerase randomly incorporates either a dNTP or ddNTP to the 3′-end of the newly synthesized strand. Incorporation of a ddNTP that lack the -OH group prevents

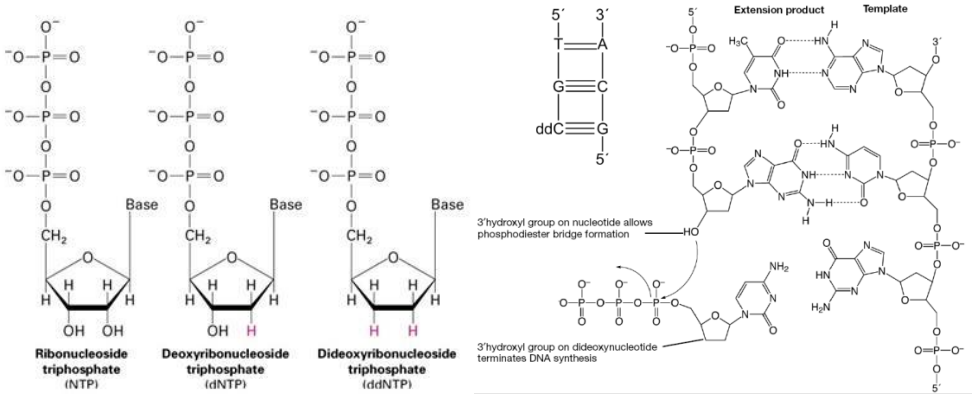


Figure 13: Nucleoside analogs. NTP, dNTP, and ddNTP, are phosphate-sugar nucleoside analogs. DNA synthesis continues with the incorporation of a dNTP. Incorporation of ddNTP terminates DNA synthesis (www3.appliedbiosystems.com).

phosphodiester bridge formation with the 5′-phosphate of the succeeding dNTP, which terminates synthesis in a base-specific manner at each nucleotide position of the synthesized strand. Each reaction of DNA fragments were then separated in adjacent lanes through denaturing polyacrylamide slab gel electrophoresis, and resolved on an autoradiograph to detect the radioactive bands. The DNA sequence were manually read in the 5′-3′ direction from the bottom of the gel and upwards (www3.appliedbiosystems.com).

1.12.4 Automated Cycle Sequencing

Modern automated cycle sequencing follows the principle of Sanger sequencing and requires the same essential components like ssDNA template, a sequencing primer, thermostable DNA polymerase (without exonuclease activity), dNTPs, ddNTPs, and reaction buffer. The

extension products are labeled with four different fluorescent dyes rather than radioactive dyes, either by dye-labeled ddNTPs (dye terminators) or 5`-dye-labeled primers (dye primers). Dye terminators and the essential sequencing components are combined in one single reaction. Dye primers are processed in four separate reactions. The extension products are amplified through cycles of annealing, extension, and denaturation on a thermal cycler, to produce DNA fragments that are terminated in a base-specific manner by one of the four dye-labeled ddNTPs. Each fragment is separated through automated capillary electrophoresis and the 3`-terminal ddNTP are resolved from the different wavelength each of the four fluorescent dyes emits when excited by light. The DNA sequence is then converted to digital data and resolved as an electopherogram (www3.appliedbiosystems.com).

1.12.5 Capillary Electrophoresis

Automated sequencing machines with capillary array electrophoresis (CE) made injection, separation, and optical detection of biological substances possible in a high-throughput fashion. In this study we used the ABI 3130xl Genetic Analyzer with a 50 cm 16-capillary array specialized for sequencing applications and fragment analysis. The capillaries are thin (50µm in diameter) glass arrays filled with a separation matrix, in this case with POP-7^{TM} Polymer. Separation matrices coat the capillary wall and are specifically optimized to control the electrokinetic forces (electrophoresis - the motion of charged particles when applied to an electric field, and electroosmosis – the volumetric liquid flow of the polymer in a capillary when applied to an electric field) inflicted from the external voltage applied during capillary electrophoresis.

The sequencing products are injected into the polymer-filled capillaries where the voltage applied makes the negatively charged DNA fragments migrate through the polymer towards the positive electrode. The mobility of the DNA fragments is proportional to the charged particle, which means that smaller fragments with a big charge move faster than larger fragments with a small charge. Each DNA fragments are thus injected into the capillary and separated by size prior to passing an Argon laser that emits light, which excites the fluorescent dyes of the terminating ddNTPs, while a sensor register the color sent back from the fluorescent dyes. The results are continuously transferred to a computer and resolved as an electopherogram. The velocity of charged particles through the polymer-filled capillary is affected by the applied voltage, the length of the capillary array, and the viscosity of the polymer, which is also affected by pH and temperature. Polymers with different properties can be selected depending on the analytes to be separated (DNA, RNA, proteins, etc.) and the

mobility of the analytes can be selectively controlled. Polymers are constantly optimized to improve efficiency, selectivity, migration time, and resolution (www3.appliedbiosystems.com).

1.12.6 Dye terminator sequencing kits

DYE Terminator kits are based on modification of the traditional dideoxy-mediated chain termination chemistry where each terminator is labeled with fluorescent dyes for automatic detection. Automated cycle sequencing with dye terminators have become the standard of choice for most DNA sequencing applications. In this study we used the BigDye® Terminator v1.1 Cycle Sequencing Kit, offered by ABI. Each of the four BigDye® terminators (ddG, ddA,ddT, ddC) are labeled with a fluorescein donor dye linked to one of the four dichlororhodamine (dRhodamine) acceptor dyes; dichloro[R6G] (ddA), dichloro[TAMRA] (ddC), dichloro[R110] (ddG), and dichloro[ROX] (ddT). Three of four terminators (ddG, ddT, and ddC) are attached to the dRhodamine dye through an ethylene oxide (EO) linker. The fluorescein donor absorbs the excitation energy from the argon ion laser which is transferred through the EO linker to the dRhodamine acceptor dye that emits fluorescence at different wavelengths captured by the sensor. This energy transfer is more efficient than the direct excitation of acceptor dyes from the laser which gives a sequencing method that is very sensitive and robust (www3.appliedbiosystems.com).

1.13 Fragment analysis by Capillary Electrophoresis

Fragment analysis is a powerful tool in molecular biology research used for genotyping, DNA profiling, and to detect mutations within a genome. Fragment analysis is the separation of DNA fragments by size using capillary electrophoresis, where PCR amplified DNA fragments are labelled with multiple fluorescent dyes of different color, including a color-labeled size standard of known length to determine the base-pair sizes of the sample product peaks (Figure 14). The analysis software genotypes the DNA fragments based on user-defined markers containing the analysis parameters of the nucleotide of interest (www3.appliedbiosystems.com).

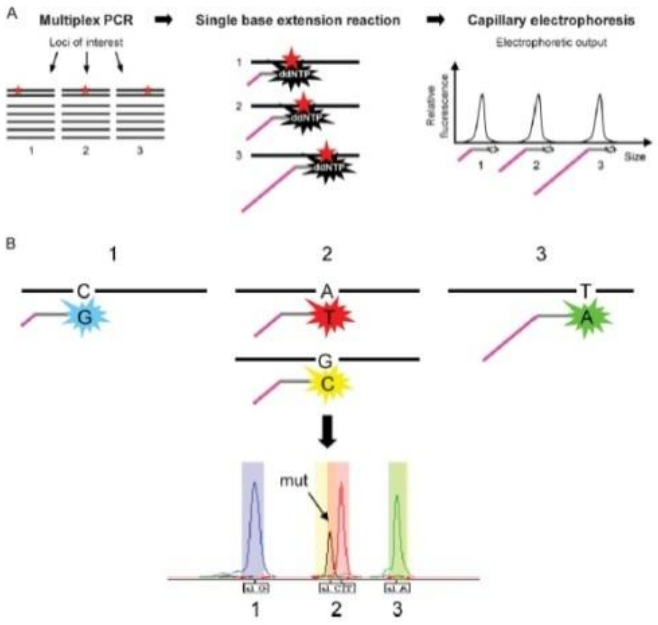


Figure 14: Overview of the SNaPshot® Multiplex method. The protocol consists of a multiplexed PCR step followed by a single-base extension reaction where allele specific probes are fluorescently labeled with ddNTPs of different color and sizes that are resolved by capillary electrophoresis (Dias-Santagata et al. 2010).

1.13.1 SNaPshot® Multiplex Kit

The ABI PRISM® SNaPshot® Multiplex Kit is designed to combine up to ten single nucleotide polymorphisms (SNPs) at known locations in ten DNA templates that can be run in a single tube. The chemistry is based on the dideoxy single-base extension method where unlabeled oligonucleotide primer binds to a complementary template and chain elongation is terminated by the incorporation of a ddNTP to the 3´ end of the growing strand. The terminators (ddG, ddA,ddT, ddC) are labeled with a fluorescein donor dye linked to one of the four dichlororhodamine (dRhodamine) acceptor dyes; dichloro[R6G] (ddA), dichloro[TAMRA] (ddC), dichloro[R110] (ddG), and dichloro[ROX] (ddT) – the same donor dyes used for the BigDye® terminators (www3.appliedbiosystems.com).

1.13.2 SNaPshot® Genotyping

Dias-Santagata et al. (2010) has developed a clinical genotyping assay based on The SNaPshot® Multiplex System from Applied Biosystems. The assay is designed to detect 58 mutations in 13 different cancer genes, and can detect a total of 120 multiple sequence variants (i.e. A>T; A>G; A>C) in many of the selected nucleotide positions, including low-level mutations in formalin fixed and paraffin embedded (FFPE) DNA samples. Only ~20 ng DNA pr. reaction is recommend by Dias-Santagata et al. (2010), and the low amount of isolated DNA from FFPE tissue required is a major improvement that makes it possible to

incorporate FFPE tumor blocks in genetic profiling of cancer patients. Although successful mutation profiling are currently conducted from FFPE material, several challenges exists, usually from limited amount of available tumor material, low quality of DNA extract, and non-reproducible sequence artifacts generated from PCR amplification, which greatly affects downstream results (Sah et al. 2013).

The modularity of the SNaPshot® system reflects the possibilities that lie ahead with additional assay modifications. The system is highly robust, sensitive and accurate, easily modified, and can be used in any laboratories with an automated DNA sequencer. In developing the SNaPshot® Multiplex assay, Dias-Santagata et al. (2010) focused on oncogenes over tumor suppressor genes. Oncogenes are often activated by a few mutated codons and identifying and inhibiting such activating pathways is a preferred target by clinical developers rather than inactivating tumor suppressor genes. Out of 30 CRC cases, Dias-Santagata et al. (2010) detected 14 mutations, including "Mutations or combination of mutations that are rare or not-previously described in the corresponding tumor type". Some of these, i.e. *TP53* mutations were found in combination with two other genes (KRAS G13D + TP53 R273H and NRAS Q61H + TP53 R175H), strongly suggesting the ability to discover novel genetic markers from FFPE tissue with the SNaPshot® Multiplex System.

1.14 Ion-Torrent PGM™ – Ion Semiconductor Sequencing

The new Ion-Torrent technology developed by Life Technologies takes advantage of a semiconductor chip device that directly translates chemical signals into digital data. Ion semiconductor sequencing is performed without light emitting signals that offer unprecedented speed, scalability, and low cost. The chemistry is based on detecting t change in pH from the release of protons when DNA polymerase incorporates a dNTP to a growing template (Figure 15), with no need for any light fixtures, scanning devices or recording cameras to monitor the sequencing by synthesis progression. The Ion Torrent semiconductor sequencing chip contains micro-wells filled with a million copies of a DNA molecule. The chip is flooded with one nucleotide after another which is incorporated to the DNA sequence when it complements the DNA sequence/template in a particular well. The ion sensor directly detects the change in pH of that solution and translates it to digital data.

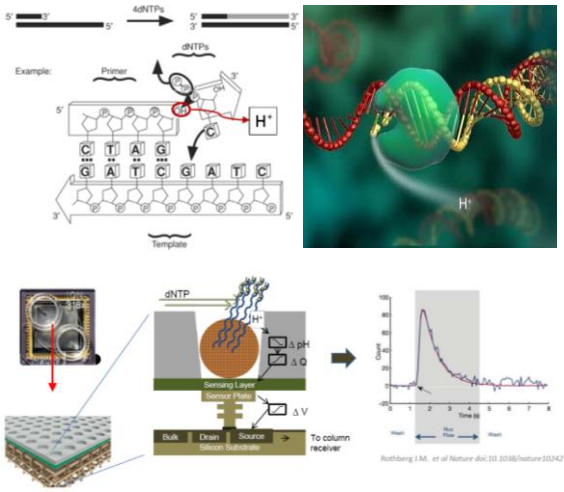


Figure 15: Ion Torrent chemistry and chip technology. When DNA polymerase incorporates a dNTP a hydrogen ion is released affecting the pH of the solution in the well (www3.appliedbiosystems.com).

First step in workflow is to prepare a library of DNA fragments flanked by the Ion Torrent adapters. The Ion Torrent adapters are either ligated to the PCR products or incorporated to the 5`-end of the PCR primers. Clonal amplification of the library fragments by emulsion PCR coats the Ion Sphere™ particles with the DNA template which is then applied to the Ion chip. The template coated Ion Sphere™ particles is deposited in the chip wells through a centrifugation step, and the chip is ready for sequencing (www3.appliedbiosystems.com).

2 AIM OF THE STUDY

Colorectal cancer is a leading cause of cancer related deaths in the western world. The challenges to stratify Dukes' stage B patients for correct treatment response are evident with the variations in cancer specific survival rates for Dukes' stage B patients. DNA from FFPE tissue is routinely used in many diagnostic laboratories to detect mutational biomarkers that can predict clinical response, although FFPE derived DNA is often fractionated and yields non-reproducible sequence artifacts not suitable for Sanger sequencing.

TP53 has been extensively studied and found mutated in about 50% of all cancers. The breast cancer group at EpiGen, Ahus, is participating in the Oslo Breast Cancer Consortium (OSBREACC) initiated by K.G. Jebsen Centre for Breast Cancer Research (KGJ CBCR). In this regard, *TP53* sequencing has been conducted on breast cancer material from Akershus University Hospital, but there has never been a screening of *TP53* mutations in colorectal cancer material from Akershus University Hospital. On this notion we chose *TP53* in anticipation to achieve new knowledge for colorectal cancer cases at Ahus.

In this study we are interested to evaluate the SNaPshot® fragment analysis method ideal for use on FFPE derived DNA against traditional Sanger sequencing by screening DNA isolated from fresh frozen tumors (FFT) and FFPE tumors.

Distribution of mutations will be analyzed in respect to known clinical and histopathological parameters in Dukes' stage B and C cases, and a general mutation analysis of the mutant effect on protein function will be theoretically evaluated using the Mut_Mat_II tool available from the UMD-p53 database.

In addition, during the course of this study The Ion Torrent PGM™ machine has been implemented at EpiGen and is currently under evaluation by scientist Torben Lüders. We thought this was a good opportunity to sequence a small subset of the tumors included in this study.

The aim of this study is therefor to:

- evaluate three different sequencing based methods on how they perform in detecting mutations from DNA isolated from FFT- and FFPE derived tissue
- analyze the spectrum of *TP53* mutations in Dukes' stage B and C cases

3 MATERIALS AND METHODS

3.1 Patient Materials

Available tumor samples from a consecutive series of 40 fresh frozen colon carcinomas (fresh frozen tumors, FFT samples) Dukes' stage B (n = 20) and C (n = 20), removed surgically between 2004 and 2007 at Akershus University Hospital (Ahus) (Lørenskog, Norway), were included in the present study and analyzed by direct Sanger sequencing for mutations within the *TP53* gene, exon 2 through 11, including flanking introns.

In addition, available tumor samples from a consecutive series of 40 paraffin-embedded colon carcinomas (FFPE samples) Dukes' stage B (n = 20) and C (n = 20), removed surgically between 2010 and 2011 at Ahus, were included in the present study and analyzed by SNaPshot® fragment analysis (FA) for mutations in 7 frequently mutated SNPs within codon 175, 245, 248, 273 and 306, distributed within exon 5, 7, and 8, located in the *TP53* DNA binding domain.

FFPE sections were provided and histologically confirmed by pathologist Solveig Norheim Andersen at the Department of Pathology, Ahus.

3.2 Tissue Preparation

FFPE-blocks were cut into 10 µm-thick sections (5-10 sections) at the Department of Pathology and received immediately after microdissection to avoid degradation from oxygen. Prior to DNA extraction, FFT tissue (≥25 mg) was finely cut and lysed overnight, while FFPE sections were deparaffinized with xylene (Sigma-Aldrich, St. Louis, USA). Residual xylene was removed through 2 ethanol (100%; Sigma-Aldrich) rinses.

3.3 DNA Extraction

Genomic DNA was extracted from each of the FFT- and FFPE tumors according to the DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany). This kit is based on a spin-column silica-membrane technology that binds DNA while cellular debris and other impurities are removed by centrifugation.

In principle, tissue lysis is loaded on to the spin column with optimal buffer conditions to increase the DNA binding affinity to the silica-membrane during centrifugation. Dissolved residual salts, cellular debris and contaminants are rinsed through the membrane by centrifugation during subsequent washing steps. Membrane-bound DNA is then eluted in

buffer solution or water, ready for use. Extracted DNA samples was kept at -20°C and diluted prior to amplification to obtain the required concentration specified in the protocols.

3.4 DNA Quantity and Purity Assessment

DNA concentration and sample purity was measured using a NanoDrop™ 8000 (Thermo Scientific, Wilmington, DE, USA) spectrophotometer. Maximum absorbance reading is 260 nm for nucleic acids and 280 nm for proteins. Sample purity assessment of nucleic acids with absorbance ratios of A260/280 nm around ~1.8, and A260/230 nm around ~2 is considered to be of good quality with little contaminants absorbed at the respective absorbance values. Any significant deviations from these values might indicate sample contamination.

3.5 Ethanol Precipitation

Isolated DNA from FFT samples with low (<1.0) A260/280 ratio were ethanol precipitated as follows; A 1/10 volume (adjusted individually according to each sample volume) of 3 M sodium acetate pH 5.2 (Sigma-Aldrich, St. Louis, MO, USA) were added to each sample and mixed well. Two volumes of cold 96% ethanol (Arcus, Nittedal, Norway) were added to each sample, mixed well, and placed on ice for >20 min, then centrifuged at full speed for 15 min. Supernatant was decanted and the pellet was washed in 1 mL cold 70% ethanol (Arcus), then mixed briefly before centrifuged at full speed for 2 min. The pellets were air dried and re-suspended in 50 µL AE buffer (Qiagen) before all tubes were warmed to 37°C for 30 min to ensure DNA is homogenized in solution and quantified as described above. Final salt concentration in sample should be 0.3 M. The sodium acetate volume is intended for DNA suspended in TE-buffer and should be adjusted if DNA is suspended in other solutions containing salt, while the ethanol volume should be adjusted according to the amounts of salt added. Although DNA was eluted in AE buffer (AE contains 0.5mM EDTA and TE contains 1 mM EDTA), no sodium acetate or ethanol volumes were adjusted.

3.6 Primers

Amplification and extension primers used for direct Sanger sequencing and FA (Supplementary Table A) were ordered from Life Technologies (Carlsbad, CA, USA). Lyophilized primers were re-suspended in nuclease-free water (Ambion, Austin, TX, USA) according to specifications in the primer product data sheet. The primers were diluted to obtain the required working solution specified in the respective protocols. Stock solutions and aliquots were stored at -20°C.

3.7 *TP53* Sanger Sequencing Analysis

Direct sequencing of the *TP53* gene, exon 2 through 11 including flanking introns, was performed on FFT tumors according to Vu et al. (2008) and IARC (2010) updated protocol (only exon 7) (http://p53.iarc.fr/ 2010).

Direct sequencing of the *TP53* gene, exon 5, 7, and 8, including flanking introns was performed on FFPE-tumors according Dias-Santagata et al. (2010).

Several modifications were made to all three protocols according to SPA-protocol (Staphylococcal Protein A protocol) regularly performed by the Dept. of Microbiology and Infection Control, Ahus.

3.7.1 PCR Amplification

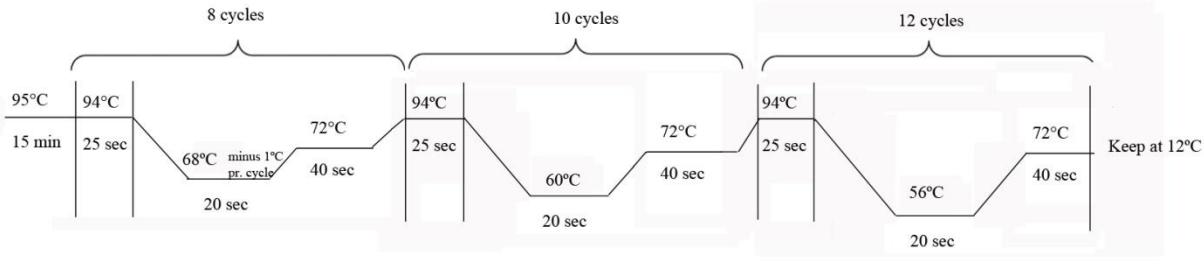
Different PCR conditions have been carried out in this study according to which protocol followed. Components used in all PCR reactions have been 10X PCR buffer with 15 mM $MgCl_2$ (Qiagen), dNTP (Qiagen), forward and reverse primer, HotStarTaq (Qiagen), Q-Solution (Qiagen), additional MgCl2 (Qiagen), total DNA template, and nuclease-free water (Ambion). The amount of each component and PCR cycle conditions are given in detail for each protocol (Table 2).

Table 2: PCR mix (a) and thermal cycle program (b).

1a. Vu et al. (2008)

Components	Volume/reaction
- 10X PCR buffer containing 15 mM $MgCl_2$	1X
- dNTP mix (2.5mM each)	2.5 nmol (2.5 mM each)
- Primer, forward and reverse	6 pmol (6µM each)
- HotStarTaq DNA polymerase (5U/µl)	0.5U
- Template DNA	25 ng
- Water, molecular biology grade	Qsp 10 µL

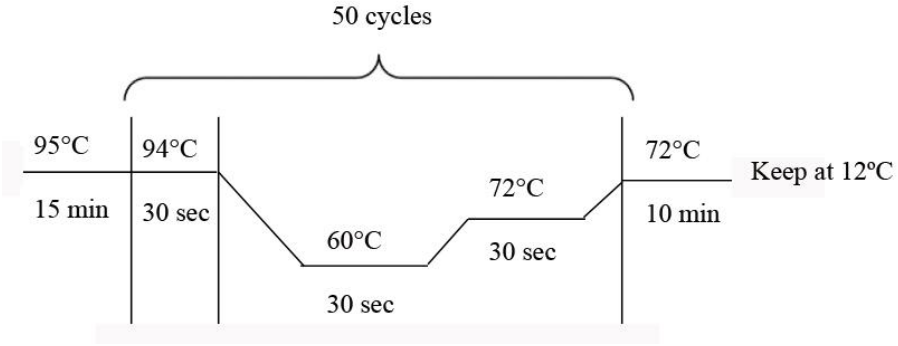
1b. Vu et al. (2008)



2a. IARC (2010)

Components	Volume/reaction
- 10X PCR buffer containing 15 mM $MgCl_2$	1X
- 5X Q-solution	1X
- dNTP mix	5 nmol (5 mM each)
- Primer, forward and reverse	10 pmol (10 µM each)
- HotStarTaq DNA polymerase (5U/µl)	0.5U
- Template DNA	50 ng
- Water, molecular biology grade	Qsp 20 µL

2b. IARC (2010)

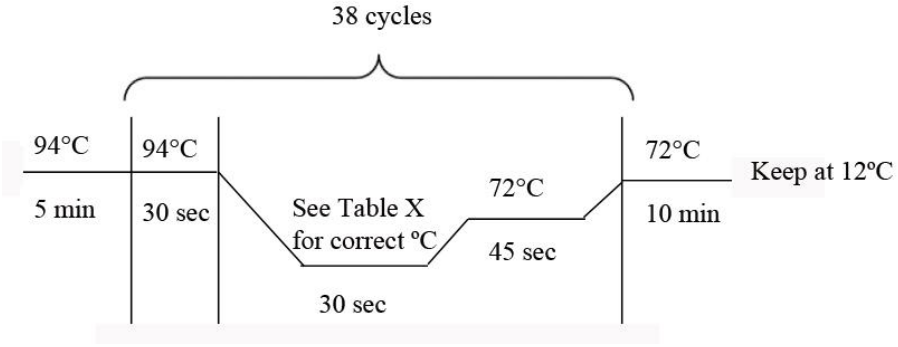


3a. Dias-Santagata et al. (2010)

Components	Amount
- 10X PCR buffer containing 15 mM $MgCl_2$	1X
- dNTP mix	4 nmol (4 mM each)
- Primer, forward and reverse	10 pmol (10 uM each)
- HotStarTaq DNA polymerase (5U/µl)	0.5U
- additional $MgCl_2$	Ref
- Template DNA	40 ng
- Water, molecular biology grade	Qsp 10 µL

Ref: Supplementary Table B

3b. Dias-Santagata et al. (2010)



3.7.2 Optimization of PCR- and Cycle Sequencing Parameters

Some FFPE samples' initial PCR amplification revealed no bands, or the final sequencing results failed due to poor data quality when processed according to the sequencing protocol by Dias-Santagata et al. (2010). PCR cycle conditions were in those cases increased from 38 to 48 cycles. Excess primers and unincorporated dNTPs were inactivated as described under section 3.7.4.

Newly amplified PCR products were then purified according to the silica-based QIAquick PCR Purification Kit (Qiagen), as recommended by the manufacturers' recommendations. Thermal cycling conditions were also increased from 38 to 48 cycles, and all samples were further analyzed as described.

3.7.3 Gel Electrophoresis

All PCR products were analyzed on 1.5% (w/v) Sea Kem LE agarose gel (Lonza, Rockland, ME, USA) in 1X Tris-Borate-EDTA (TBE) buffer (1L: 54 g Tris-base, 27.5 g boric acid, 20 mL 0.5 M EDTA (pH 8), nuclease-free water, Department of Microbiology and Infection Control, Ahus). The gel was stained with SYBR Safe® (Invitrogen, Carlsbad, CA, USA) DNA gel stain (2µL/50mL). 2 µL PCR products with 1.5 uL of 10X gel loading dye (Orange G) were loaded onto the gel. A Low Mass DNA ladder (Invitrogen) was included in all runs to quantify the amount of PCR products, along with positive (human DNA (Invitrogen)) and negative controls (blank).

The gel was run in 1X TBE at 200V for 25 min. The migrated DNA bands were visualized under Bio-Rad GelDoc XR (BIO-RAD, Hercules, CA, USA) imaging system and captured as *.jpg files.

3.7.4 Purifying PCR Products

PCR products from FFT samples were initially purified on a vacuum manifold with the ultrafiltration-based MinElute 96 UF PCR Purification Kit (Qiagen) according to manufacturers' recommendations.

For the modified Vu and IARC protocols, PCR products from FFT samples were diluted 1/10 with nuclease-free water prior to use in sequence reaction mix according to the SPA protocol. Excess primers and unincorporated dNTPs in FFPE samples were inactivated with 1U of shrimp alkaline phosphatase (SAP) (USB, Cleveland, OH, USA) and 5U of exonuclease I (Exo I) (USB) at 37°C for 20 min, followed by enzyme inactivation at 80°C for 15 min.

3.7.5 Sequence Reaction

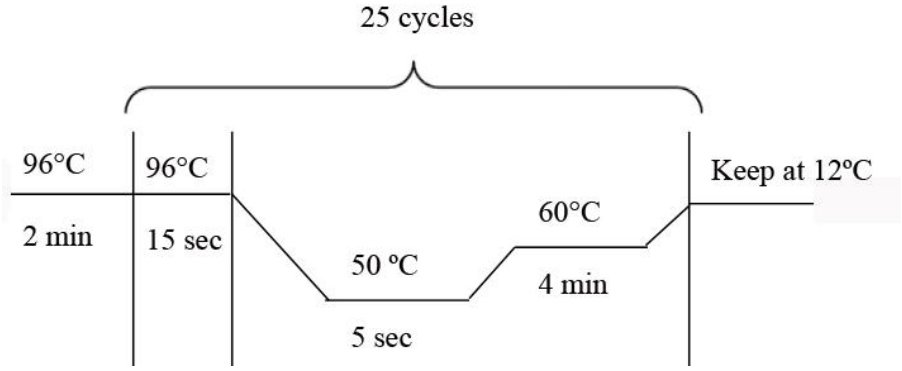
Components for each sequence reaction mix was carried out in a total volume of 10 µL containing 10X Sequencing buffer (Life Technologies, Carlsbad, CA, USA), Big Dye Terminator reaction mix v1.1 (Life Technologies), forward or reverse primer, purified PCR product template, and nuclease-free water (Ambion) up to final volume. The amount of each component and thermal cycle conditions for each protocol are given in detail (Table 3).

Table 3: Sequencing mix (a) and thermal cycle program (b)

1a. Vu et al. (2008)

Components	Volume/reaction
- 10X Sequencing buffer	1 µl
- Big Dye Terminator reaction mix v.1.1	2 µl
- Primer, fwd. or rev. 8 µM	1 µl
- PCR Product template (purified)	2 µl
- Water, molecular biology grade	Qsp 10 µl

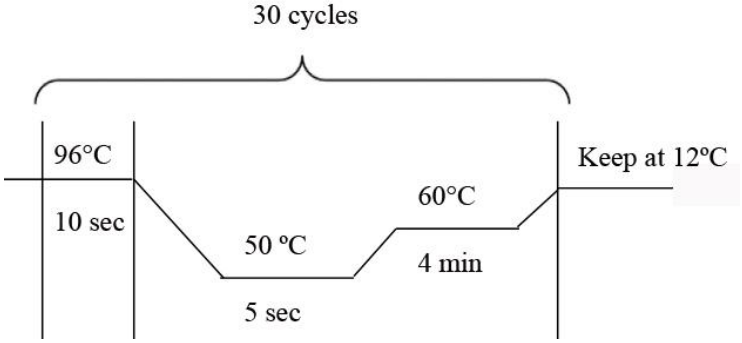
1b. Vu et al. (2008)



2a. Vu et al. (2008) and IARC (2010) according to SPA

Components	Volume/reaction
- 10X Sequencing buffer	1.5 µl
- Big Dye Terminator reaction mix v.1.1	1 µl
- Primer, fwd. or rev. 8/10 µM	1 µl
- PCR Product template (purified)	2 µl
- Water, molecular biology grade	Qsp 10 µl

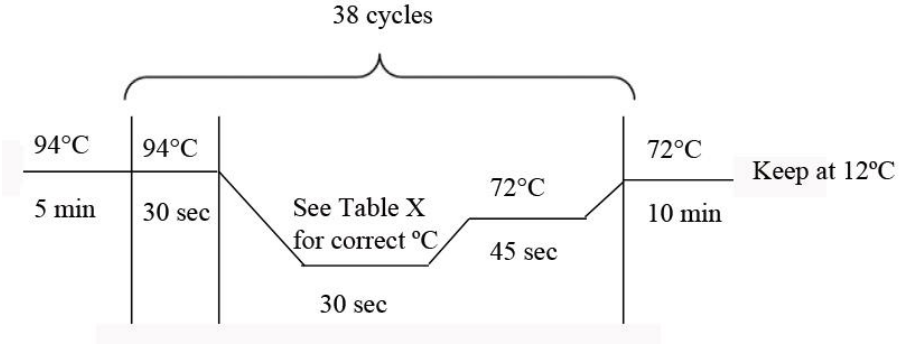
2b. IARC, 2010



3a. Dias-Santagata et al. (2010) according to SPA

Components	Volume/reaction
- 10X Sequencing buffer	1.5 µl
- Big Dye Terminator reaction mix v.1.1	1 µl
- Primer, fwd. or rev. 10 µM	1 µl
- PCR Product template (purified)	2 µl
- Water, molecular biology grade	Qsp 10 µl

3b. Dias-Santagata et al. (2010)



3.7.6 Purifying Sequencing Products

Sequencing products from FFT samples were initially purified on a vacuum manifold with the ultrafiltration-based Montage SEQ_{96} Sequencing Reaction Cleanup Kit (Millipore, Billerica, MA, USA) according to manufacturers' recommendations.

For the modified Vu, IARC, and Dias-Santagata protocols, sequencing products from FFT- and FFPE samples were purified by isopropanol precipitation according to the SPA protocol as follows; sequencing products were precipitated by adding 90 µL 69% isopropanol (Arcus) to each tube and incubated at the bench top for 15 min. The precipitate was centrifuged in an Eppendorf Centrifuge 5810R (Epperndorf, Hamburg, Germany) at 21°C and 3000 g for 30 min with the centrifuge acceleration set at 8 and the break set at 5. Carefully removed from

the centrifuge the tubes were turned upside-down for the isopropanol to quickly run off. The tubes were directly centrifuged upside-down on a tissue at 21°C and 700 g for 1 min, with the centrifuge acceleration set at 4 and the break set at 1. The slow acceleration and breaking is necessary to avoid the sequencing products to be thrown out of the tubes by excess force.

3.7.7 Sequencing – Capillary Electrophoresis

Vacuum purified sequencing products from FFT samples were directly loaded on to a sequencing plate which was covered with septa and briefly spun down. The samples were run on 3730 Genetic Analyzer (Applied Biosystems® (ABI), Carlsbad, CA, USA) with POP-7™ sieving polymer in 50cm capillaries of standard run module, and bases were called by KB™ Basecaller v1.4 with Dyeset E and mobility file KB_3730_POP7_BDTv1.mob (ABI).
For the modified Vu et al. (2008), IARC, and Dias-Santagata et al. (2010) protocols, sequencing products from FFT- and FFPE samples were processed according to the SPA-protocol as follows; 12µl HiDi formamide (Life Technologies) was added to the precipitated sequencing products and prepared appropriately as described above. The samples were run on 3130xl Genetic Analyzer (ABI) with POP-7™ sieving polymer in 50 cm capillaries with standard run module, and bases were called by KB™ Basecaller v1.4 with Dyeset E and mobility file KB_3130_POP7_BDTv1.mob (ABI).

3.7.8 Data Analysis - SeqScape® Software Version 2.6 (ABI)

Sequence data generated on ABI 3730 and 3130xl DNA ANALYZERs were analyzed for mutations with *SeqScape® Software Version 2.6* (ABI). A p53 project template with genomic reference sequence (NT_000017.9) of defined exon-intron boundaries and analysis settings for automatic analysis was obtained from Phoung Vu at the Department of Genetics, Norwegian Radium Hospital, Oslo, Norway.
Assembled sequencing results (*.abi trace files) were aligned into a consensus sequence and manually reviewed for basecalling accuracy against the reference sequence. Sequence coverage and clear range was checked for each samples' trace file, and filtered sequences with low quality regions were trimmed for clear range and assembled when possible.

3.8 *TP53* SNaPshot® Fragment Analysis

Fragment analysis of selected *TP53* nucleotide positions (Supplementary Table C) were performed on FFPE samples according to Dias-Santagata et al. (2010).

Also, a fraction of tissue was taken from 3 of the FFTs for formalin fixation and paraffin wax embedding. The three samples were then analyzed accordingly to verify mutations found by direct Sanger sequencing within these tumors. Paraffin blocks of FFT tissue were cast at the Department of Pathology, Ahus, and histologically confirmed by pathologist Solveig Norheim Andersen, Dept. of Pathology, Ahus.

3.8.1 PCR Amplification

PCR reactions were carried out in a total volume of 10 µL containing 10X PCR buffer with 15 mM $MgCl_2$ (Qiagen), 2.5 mM of each dNTP (Qiagen), 1U HotStarTaq (Qiagen), the correct amount of each forward and reverse primer (Supplementary Table D), ~25 ng genomic DNA template, and nuclease-free water (Ambion) up to final volume.
PCR cycle conditions were 95°C for 8 min of enzyme activation, followed by 45 cycles of denaturing at 95°C for 20 sec, annealing at 58°C for 30 sec, and extension at 72°C for 1 min, and one last cycle of extension at 72°C for 3 min, with indefinite storage ∞12°C.
DNA from FFPE tissue is often fragmented with low quality. This restricts the amplicon length parameters to a maximum of 200 nt. All final PCR product length varies between 75-187 nt.

3.8.2 Purifying PCR Products

Inactivation of excess primers and unincorporated dNTPs was done with 3.3U of SAP (USB) and 2.7U of Exo I (USB) at 37°C for 60 min, followed by enzyme inactivation at 75°C for 15 min.

The PCR products were analyzed by gel electrophoresis as described under section 3.7.3.

3.8.3 Labeled Primer Extension Reaction

Primer extension reaction mix was carried out in a total volume of 10 µL containing 1-2 µL of PCR product (determined from a dilution series test), 2.5 µL SNaPshot® Multiplex Ready Reaction mix (ABI), the correct amount of extension primers (Supplementary Table E), and nuclease-free water up to final volume.

Thermal cycling conditions were 96°C for 30 sec, followed by 25 cycles at 96°C for 10 sec, 50°C for 5 sec, and 60°C for 30 sec, with indefinite storage ∞12°C.

3.8.4 Purifying Extension Products

Inactivation of excess extension primers was done with 2U of SAP (USB) at 37°C for 60 min, followed by enzyme inactivation at 75°C for 15 min.

3.8.5 SNaPshot® Genotyping

0.5 µL inactivated labeled extension products were mixed with 0.2 µL GeneScan™ 120 LIZ® (ABI) and 9.3 µL HiDi™ Formamide (ABI) and denatured at 95°C for 5 min. The sequencing products were mixed thoroughly and transferred to a sequencing plate which was covered with septa and briefly spun down. The samples were run on ABI 3130xl Genetic Analyzer with POP-7™ sieving polymer in 50 cm capillaries, and ABI GeneScan E5 Run Module.

3.8.6 Data Analysis - GeneMapper® Software Version 4 (ABI)

SNaPshot® results were analyzed for mutations with The Applied Biosystems GeneMapper® Software Version 4, according to the default settings defined for SNaPshot® analysis. Panels and bin set parameters for automatic data analysis was provided as text files (Dias-Santagata et al. 2010).

All samples (*.fsa trace files) were manually reviewed for genotyping, and scoring criteria was calculated for all samples.

Each sample passed if the peak fluorescent height for the wild type allele was ≥ 50-100 times higher than the overall background noise, and if the peak fluorescent height for the wild type allele in the negative control (water sample) was < 10% of the height of the wild type allele in the clinical sample.

A positive mutation was called when the percentage of mutant allele was ≥ 10% (fluorescent peak height ratio of [mutant/(mutant+wild type)] alleles > 0.10), and if the peak fluorescence of the mutant allele was > 3 times above the background in the wild type control sample.

3.9 Ion-Torrent PGM™ Sequencing

Ion-Torrent sequencing was performed on a small selection of FFT- and FFPE samples with the Ion AmpliSeq™ Library Kit 2.9 and Cancer Hotspot Panel v2 according to manufacturers' procedures, carried out by scientist Torben Lüders at EpiGen, Ahus.

3.10 Statistical Analysis

Minitab® 16 was used to perform Fishers' exact test to establish association between the two groups of Dukes' stage B and C.

3.10.1 Clinicopathological Analysis

Univariate, Kaplan-Meier analyses and the log-rank test (α=0.05) for two parameters: *TP53* and Dukes' stage was performed to investigate how these parameters affect survival. KM-plot

was kindly computed by Post. Doc. Jovana Klajic (EgiGen, Ahus) using the SPSS Software. Overall survival was defined as date of operation to death from colon cancer, or to the last known follow-up date.

Microsoft Excel 2010 was used for Pearson´s correlation of coefficients between clinicopathological pairs of variables.

3.12 Stock Solutions

3.12.1 Orange G

10X Gel Loading Dye was prepared by dissolving 8 g sucrose (≥99.5% (GC) (Sigma-Aldrich) and 100 mg Orange G (Sigma-Aldrich) in 40 mL nuclease-free, then topped with nuclease-free water (Ambion) to a final volume of 50 mL.

3.12.2 Buffer PB

Qiagen Buffer PB stock solution was prepared by dissolving 47,765g Gu-HCl (Sigma-Aldrich) in 30 mL 100% isopropanol (Arcus) and topped with nuclease-free water (Ambion) to a final volume of 100 mL.

4 RESULTS

In this study we have used three different sequencing methods to investigate the ability to detect somatic mutation in CRC-tumor samples from FFT- and FFPE tissue (Figure 16). This chapter reflects the results obtained from these experiments.

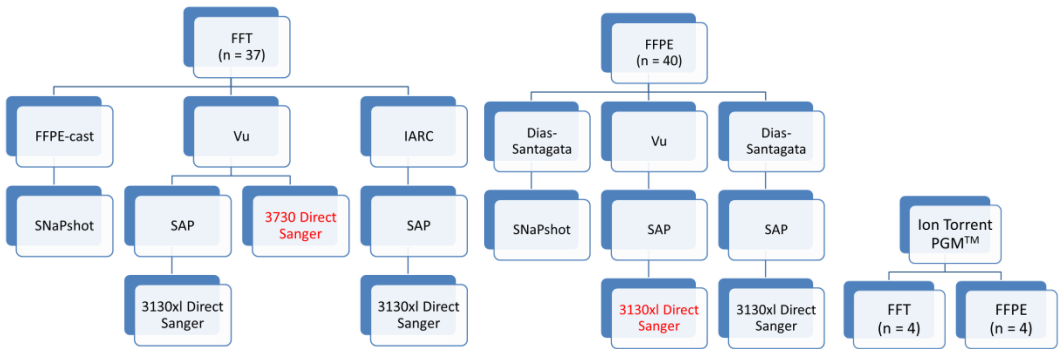


Figure 16: Overview of each experiment performed in this study on FFT- and FFPE-tissue. Text in red represents failed analyses.

From a total of 80 samples chosen for this study, 77 were successfully analyzed. Three FFT specimens were excluded; one specimen was not found, one specimen failed to yield any DNA extract, and one specimen was incorrectly labeled as Dukes' stage C when it in fact was Dukes' stage D.

4.1 DNA Sample Quality Assessment

To ensure reproducible results in downstream applications it is important to assess DNA sample purity to determine contaminants such as proteins, salts and solvents. DNA prepared from tissue usually has high protein content. A low (<1.8) A260/A280 ratio is indication on protein contamination, and a low (<1.5-1.8) A260/A230 ratio is indication on salt carryover from the DNA isolation process. Lower A260/A230 ratios indicate higher salt content. Absorbance readings of A260/A280 ratio measured in DNA purified from FFT tissue were in the range of 1.70-2.33, with the majority within 1.70-2.0. The A260/A230 ratio varied from 0.30-2.24. Twenty-one samples with A260/A230 ratio <1.0 were ethanol precipitated which increased the A260/A230 ratio from <1.0 to ~1.8-2.0 in the respective samples. Ideally, the A260/A230 ratio should be close to 1.8, but greater than 1.5. However, 8 FFT samples with A260/A230 ratio >1.0<1.5 were not ethanol precipitated.

DNA isolated from FFPE tissue had A260/A280 ratio of ~2.0. A low (<1.0) A260/A230 ratio were seen in most samples, however, these were not ethanol precipitated.

The complete list of absorbance reading values and extracted amount of DNA (ng/µl) from FFT- and FFPE samples, including ethanol precipitated samples, are given in Supplementary Table F.

4.2 *TP53* Direct Sanger Sequencing Analysis

We aimed to amplify exon 2-11 of the *TP53* gene in the FFT samples according to Vu et al. (2008) with slight modifications. Most samples produced target amplicons (verified by 1.5% agarose gel electrophoresis) of varying band strength from each primer pair covering exon 2-3, 4, 5-6, 7, 8-9, 10, and 11 (Figure 17), except one specific sample that failed in each run (see Figure 18 for representation). Fragment length varies between 396-733 bp (see Supplementary Table A for details).

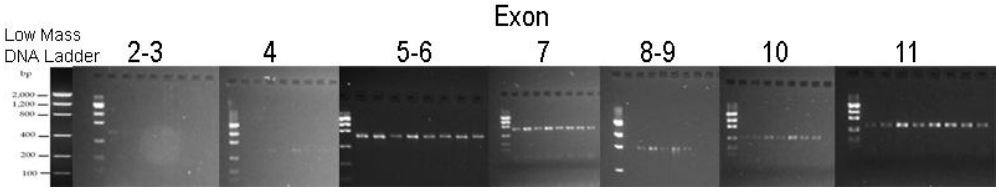


Figure 17: *TP53* target PCR amplicons obtained from FFT samples. Extract of 8 consecutive specimens identical in each gel image. Most samples failed amplification for exon 2-3 and 4.

Target amplicons for exon 7 from the new and improved IARC protocol (Figure 18) were obtained from all FFT samples, including the one sample that did not produce any amplicons for any of the primer pairs according to Vu et al. (2008).

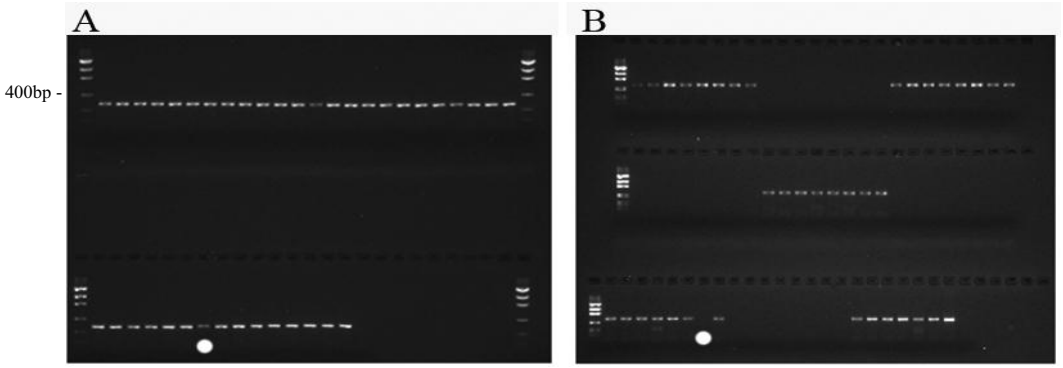


Figure 18: *TP53* target PCR amplicons obtained from FFT samples. (**A)** Target amplicons for exon 7 according to the improved IARC 2010 protocol. (**B)** Exon 11 according to Vu et al. (2008). Both gel images contains 38 samples in consecutive order, including positive (genomic human DNA), negative (blank) control, and Low mass DNA ladder as reference. Each sample in image B was run with an accompanying negative control, represented by the continuous gap of empty wells between bands of DNA. The white dot in image A and B represents one specific sample that failed (B) to obtain PCR target amplicons according to Vu et al. (2008) (results not shown for all exons) versus the amplicon obtained (A) according to the improved IARC protocol.

Initially we tried to amplify FFPE derived DNA for exon 5-6 and 8-9 according to Vu et al. (2008) and exon 7 according to IARC, since these primer pairs were already in our possession. Target amplicons were obtained for exon 7 according to IARC, but no amplicons were obtained for exon 5-6 and 8-9 according to Vu, et al. (2008) (results not shown). New primer pairs for exon 5, 7, and 8 (Figure 19), were ordered according to the sequencing analysis protocol from Dias- Santagata, et al. (2010). Since the Dias-Santagata protocol is

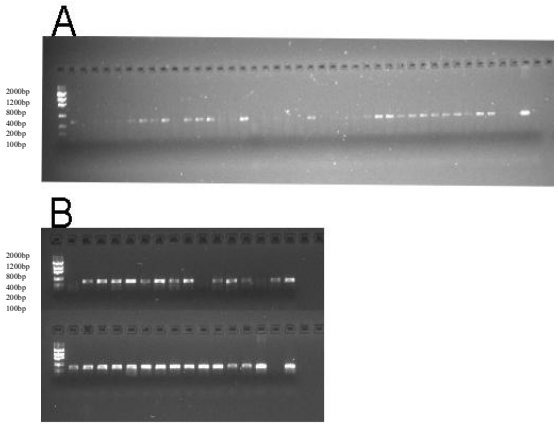


Figure 19: *TP53* PCR target amplicons obtained from FFPE samples. (A) Target amplicons for exon 8 failed for 22 of 40 samples at first attempt of PCR cycling. Gel image A contains 40 samples in consecutive order including positive (genomic human DNA), negative (blank) control, and Low mass DNA ladder as reference. **(B)** In total, 29 of 40 samples from image A were re-amplified with PCR cycles increased from 38 to 48 cycles, then spin-column purified.

optimized for FFPE derived DNA we decided to include primer pairs for exon 7, although this exon obtained successful amplicons with the IARC protocol. However, in a downstream test run sequencing results were undesirable, and to save costs from sequencing reactions we only decided to amplify and sequence exon 8 due to the fact that PCR amplification of FFPE samples for sequencing purpose was performed after the SNaPshot® fragment analysis was completed, and we found it crucial to verify the results obtained in codon 273 by sequencing exon 8 (Figure 26).

4.2.1 Sequence Data

ABI 3730

Most FFT samples resolved on the ABI 3730 DNA analyzer were unassembled and failed basecalling analysis due to poor data quality. The assembled sequences showed irregular and uneven trace peaks with dye blobs and background noise (**Figure 20**, **A1** and **A2**) throughout all chromatograms (not all results shown).

ABI 3130xl

FFT samples resolved on the ABI 3130xl analyzer were successfully assembled with regular and even trace peaks, and minimal background noise (**Figure 20**, **B1** and **B2**) throughout the chromatograms. However, forward sequences generated for exon 7 (Vu et al. 2008),

amplification primer 5′-GACCATC CTGGCTAACGG-3′ and universal M13 extension primer) resolved multiple sequences (**Figure 20**, **C1**) in all samples. New forward sequences generated for exon 7 (IARC) were successfully assembled with regular and even trace peaks, and minimal background noise (**Figure 20**, **C2**).

Sequencing results obtained from FFPE samples were generally of poor quality with the majority being unassembled. Assembled sequences showed irregular trace peaks with considerable background noise (Figure 21). From unassembled FFPE samples that were re-amplified with increase in PCR and cycle conditions and spin-column purified, the overall background noise was reduced (Figure 22). The reduced noise was not compared to sequencing results of corresponding samples prior to and after reamplification.

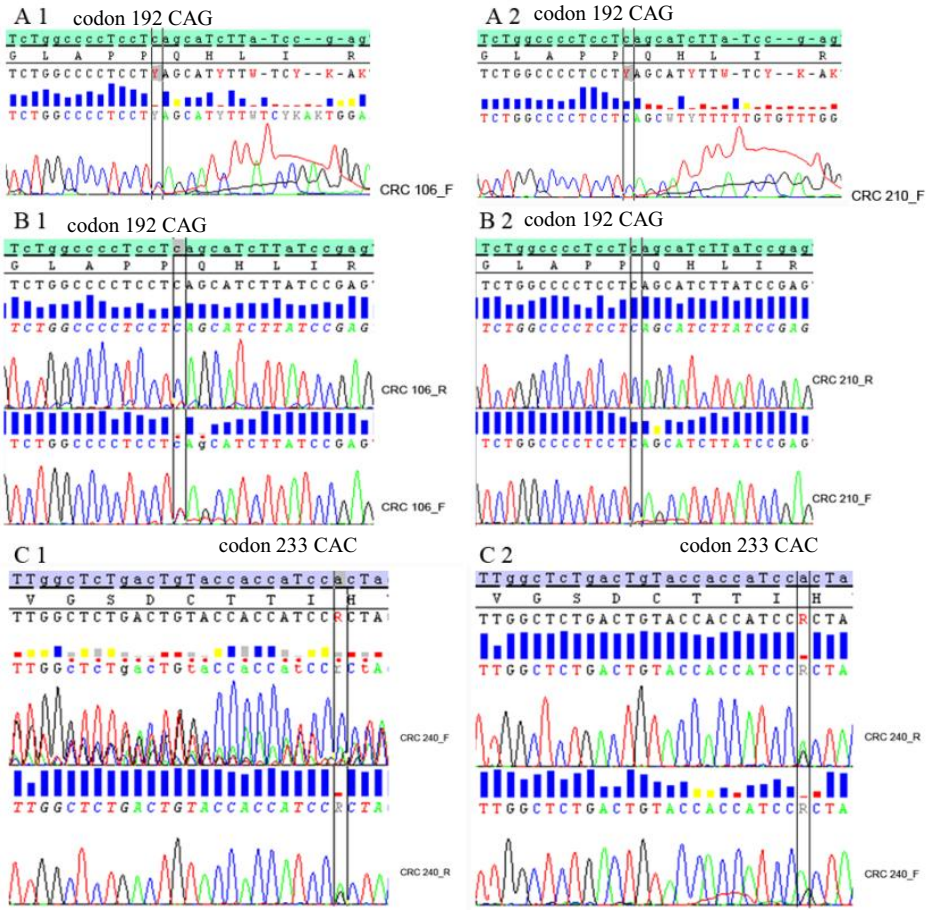


Figure 20: Sequencing results from FFT samples. Blue bars indicates QV scores (QV=20). **(A1** and **A2)** Both samples were resolved on ABI 3730 with dye blobs caused by unincorporated or incomplete removal of dye terminators (only forward sequences were assembled). **(B1** and **B2)** The same samples as shown in A1 and A2 after modifications according to SPA and resolved on ABI 3130xl. **(C1)** All forward sequences for exon 7 resolved multiple sequences according to Vu et al. (2008). Double peak in reverse strand can be a mutation or a PCR-artifact. **(C2)** Improved quality of exon 7, according to IARC/SPA, resolved a clean uniform sequence, and the double peak can be defined as a heterozygous mutation.

Five of 6 mutations detected by SNaPshot® genotyping were verified by Sanger sequencing and all false positive mutations detected by FA in codon 273 were verified as wild-type allele (Figure 21 and 22) (not all results shown).

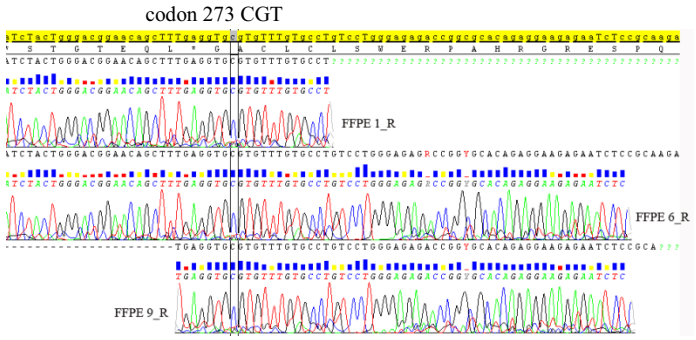


Figure 21: Sequencing results from FFPE samples. Vertical marker is at codon 273 (CGT), exon 8, and show wild-type in all three FFPE samples. Overall noisy trace files with many secondary peaks makes it difficult to interpret real mutations. FFPE 1 and 9 are fragmented with shortened clear range**.**

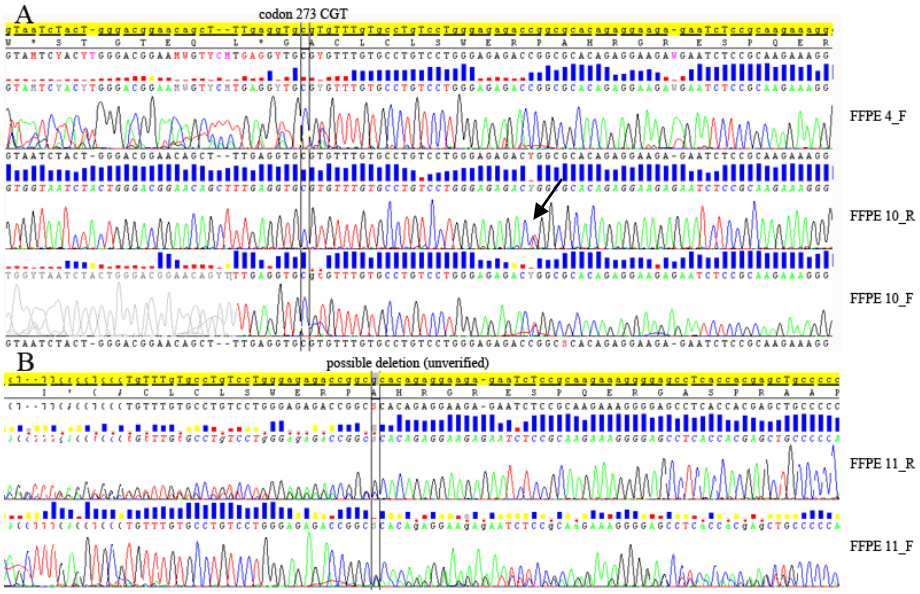


Figure 22: Sequencing results from purified FFPE samples. (**A**) Vertical marker is at codon 273 (CGT), exon 8 and show wild-type in all three samples. In sample FFPE 10 there is a heterozygous mutation in codon 282 (CGC) seen by a double red and blue peak (arrow). (**B**) Vertical marker defines a possible deletion in codon 283 (CGC). All 3 samples shown in this figure had been re-amplified and spin-column purified with increase in PCR and cycle conditions.

4.3 *TP53* SNaPshot® Fragment Analysis of FFPE

The SNaPshot® assay developed by Dias-Santagata et al. (2010) contains 13 mutated genes and can detect a total of 120 nucleotide variants. We exclusively focused on the *TP53* gene originally multiplexed into 5 panels. Two panels contain the same PCR amplification primer which was run as a uniplex PCR reaction (Figure 23). All FFPE samples produced successful target amplicons (verified by 1.5% agarose gel electrophoresis) (not all results shown).

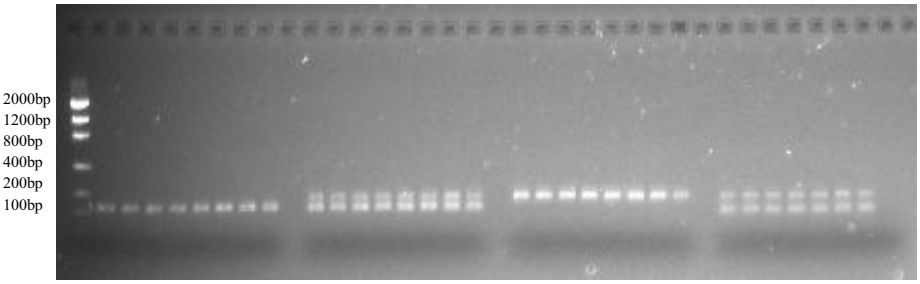


Figure 23: *TP53* target PCR amplicons obtained from FFPE samples. Gel image of 7 samples in consecutive order, including positive (genomic human DNA) in the first 3 rows of bands (8 wells) and negative (blank) control in the last row (7 wells), with Low mass DNA ladder as reference. Each row of bands represents SNaPshot panels containing the *TP53* nucleotide position of interest.

Sequencing results of FFT samples revealed a total of 3 mutations that can be detected by the SNaPshot® assay. A fraction of these tumors were formalin fixed and paraffin wax embedded, and isolated DNA was PCR amplified according to Dias-Santagata et al. (2010). Successful amplicons were obtained for 2 of 3 samples (results not shown).

4.3.1 SNaPshot® genotyping

SNaPshot® genotyping results from automatic data analysis were manually examined to check for each of the necessary parameters being passed by the GeneMapper software. Allele peak height was optimized through a series of extension product dilutions to determine the appropriate amount of extension products for genotyping (**Figure 24**). All FFPE samples were successfully genotyped (**Figure 25**) in 6 of 7 detectable mutations, except for codon 273 (c.817C>T), in which false positive results (**Figure 26**) were obtained in 39 of 40 FFPE samples. Paraffin embedded FFT samples did not pass automatic data analysis (**Figure 27**). A secondary peak is observed in sample CRC 204, although it was not automatically genotyped as a mutation. Panels and bin settings were evaluated for all results and scoring guidelines was calculated for each genotyped mutant allele (Supplementary Table G).

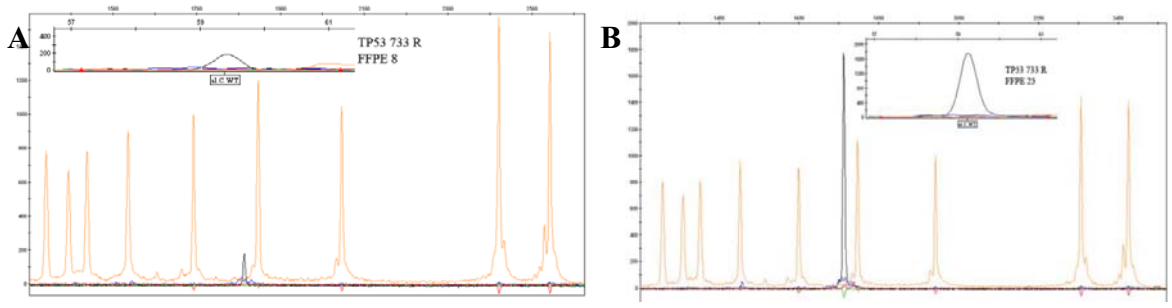


Figure 24: SNaPshot® genotyping results. Raw files with GeneScan™ 120 LIZ size standard and miniature chromatogram of wild-type sample. Both samples are from diluted PCR product (1:20). **(A)** A much diluted sample that failed the automatic data analysis settings. **(B)** Appropriately diluted sample that passed the automatic data analysis settings.

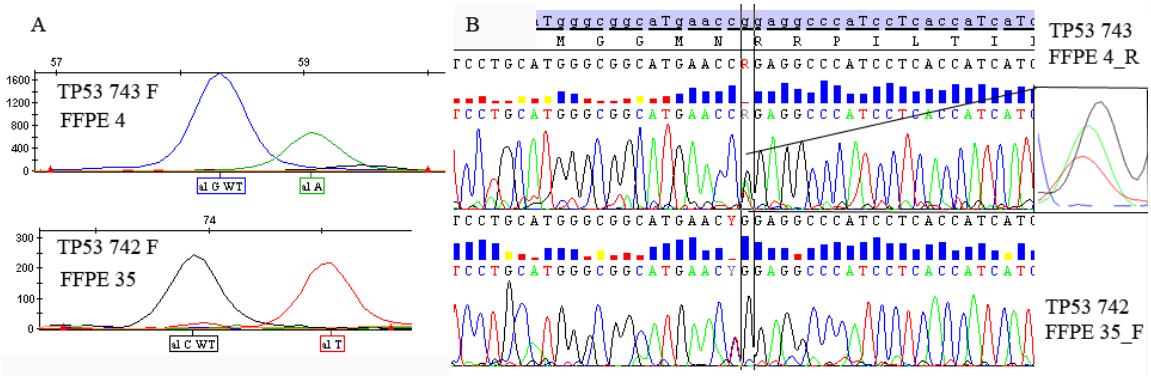


Figure 25: Successful SNaPshot® genotyping vs Sanger sequencing results. (A) Successful SNaPshot® genotyping of 2 FFPE samples. **(B)** Marker position is codon 248 (CGG). Mutations from SNaPshot® genotyping are verified by Sanger sequencing. Expanded top view show three peaks which makes the mutation difficult to interpret alone.

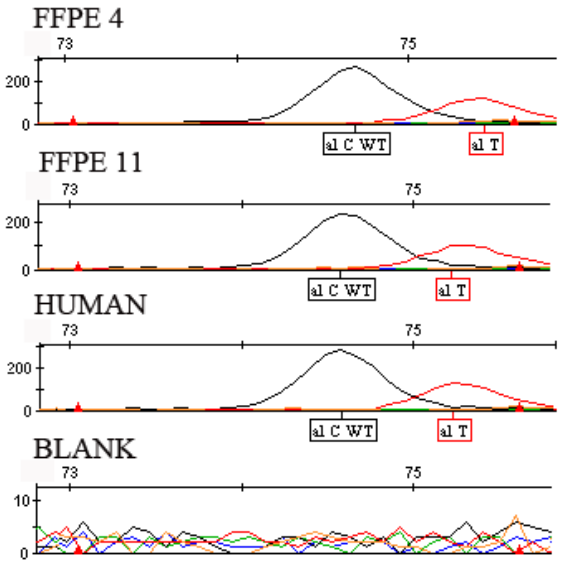


Figure 26: False positive genotyping. SNaPshot® genotyping of false positive mutations positioned at c.817C>T, codon 273. All samples passed data analysis.

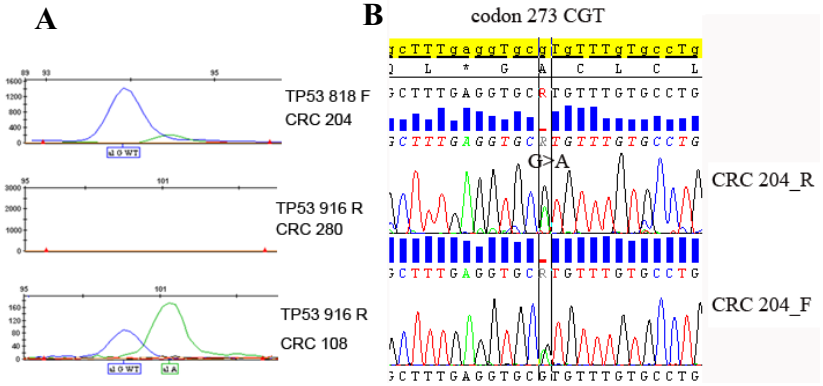


Figure 27: SNaPshot® Genotyping sensitivity vs Sanger sequencing. (**A**) SNaPshot® genotyping of paraffin embedded FFT samples. Each sample failed automatic data analysis criteria. (**B**) Heterozygous mutation detected by Sanger sequencing in CRC 204 that failed SNaPshot® genotyping (**A**).

4.4 Ion-Torrent PGM™

Ion Torrent sequencing detected each of the mutations found in the subset of 4 FFT- and 4 FFPE samples selected for semiconductor sequencing. The DNA base composition of suspected deletions found by Sanger sequencing were determined, and missense mutations were verified (Table 4). One FFPE sample was wild-type for all *TP53* exons (result not shown).

4.5 Cancer Analysis

4.5.1 *TP53* Mutations

Among the 77 tumors analyzed, 26 (33.8%) tumors revealed a total of 27 mutations within exon 5-9 of the *TP53* gene; an overall frequency of 33.8%. The specific DNA sequence alteration was identified for all 27 mutations (Table 5). The frequency of each individual mutation recorded in the IARC TP53 and UMD-p53 database (Figure 28) are mostly hot-spot and frequent mutations.

No statistical significance (2-sided Fisher's exact test) were seen between the distribution of mutations within Dukes' B (13/25) and C (13/26) cases ($p = 1$), or between the distribution of mutations by metastases ($p = 1$).

Table 4: Results obtained from Ion Torrent sequencing.

Smp. ID	Position [a]	Reverse sequence reference [b] (fwd)	Variant	Codon	WT	MUT	Type
CRC 140	7578269	GATGCT (AGCATC)	-	192-194	CAG CAT CTT	C-- --- -TT	DEL
CRC 164	7578475	G (C)	-	152	CCG	C-G	DEL
CRC 273	7579358	C (G)	A	110	CGT	CTT	SNP
CRC 285	7578275	GAG (CTC)	-	191-192	CCT CAG	C-- -AG	DEL
FFPE 4	7577538	C (G)	T	248	CGG	CAG	SNP
FFPE 11	7577090	C (G)	-	283	CGC	C-C	DEL
FFPE 26	7577094	G (C)	A	282	CGG	TGG	SNP

* g_description_hg19; fwd = forward; [a] Backwards reverse sequence

Table 5: Characterization of *TP53* mutations detected in CRC FFT- and FFPE samples.

Sequencing	ID	Dukes' stage	Nr. Mut	Exon	Codon	Base change	WT codon→MUT codon	WT AA→MUT AA	Effect	Comment	CpG site	Metastasis
Fresh tissue	88	C	1	5	157	G→T	GTC→TTC	Val→Phe	Missense		Yes	Yes
Fresh tissue	97	C	1	7	244	GC→TT	GGC→GTT	Gly→Val	Tandem		Yes	Yes
Fresh tissue	104	B	1	5	143	T→C	GTG→GCG	Val→Ala	Missense		No	No
Fresh tissue	108	C	1	8	306	C→T	CGA→TGA	Arg→Ter	Nonsense	SNaPshot®	Yes	Yes
Fresh tissue	127	C	1	8	286	G→A	GAA→AAA	Glu→Lys	Missense		No	Yes
Fresh tissue	140	B	1	6	192-194	DEL	CAG CAT CTT→C-- --- -TT		FS	Ion Torrent	No	No
Fresh tissue	164	B	1	5	152	DEL	CCG→C-G		FS	Ion Torrent	Yes	No
Fresh tissue	176	C	1	8	266	G→A	GGA→GAA	Gly→Glu	Missens		No	No
Fresh tissue	204	C	2	6	213	C→T	CGA→TGA	Arg→Ter	Nonsense		Yes	NA
				8	273	G→A	CGT→CAT	Arg→His	Missense	SNaPshot®	Yes	
Fresh tissue	207	B	1	8	279	G→A	GGG→GAG	Gly→Glu	Missens		No	No
Fresh tissue	210	C	1	5	176	G→A	TGC→TAC	Cys→Tyr	Missens		No	Yes
Fresh tissue	211	B	1	6	213	C→T	CGA→TGA	Arg→Ter	Nonsense	Homozygouz	Yes	No
Fresh tissue	240	B	2	7	233	A→G	CAC→CGC	His→Arg	Missense	Active*	No	No
Fresh tissue	280	B	1	5	175	G→A	CGC→CAC	Arg→His	Missense	SNaPshot®	Yes	No
Fresh tissue	285	B	1	6	191-192	DEL	CCT CAG→C-- -AG		FS	Ion Torrent	No	No
Fresh tissue	303	B	1	7	238	G→T	TGT→TTT	Cys→Phe	Missense		No	NA
Fresh tissue	345	C	1	7	237	G→A	ATG→ATA	Met→Ile	Missense		No	NA
SNaPshot® FA												
FFPE-tissue	4	C	1	7	248	G→A	CGG→CAG	Arg→Gln	Missense	Sanger/Ion Torrent	Yes	NA
FFPE-tissue	7	B	1	5	175	G→A	CGC→CAC	Arg→His	Missense	Sanger	Yes	NA
FFPE-tissue	18	B	1	5	175	G→A	CGC→CAC	Arg→His	Missense	Failed Sanger	Yes	No
FFPE-tissue	33	C	1	8	306	C→T	CGA→TGA	Arg→Ter	Nonsense	Sanger	Yes	No
FFPE-tissue	35	C	1	7	248	C→T	CGG→TGG	Arg→Trp	Missense	Sanger	Yes	No
FFPE-tissue	36	C	1	8	273	G→A	CGT→CAT	Arg→His	Missense	Sanger	Yes	No
Extra (Seq/Ion)												
FFPE-tissue	10	B	1	8	282	C→T	CGG→TGG	Arg→Trp	Missense	Sanger	Yes	No
FFPE-tissue	11	B	1	8	283	DEL	CGC→C-C		FS	Sanger/Ion Torrent	Yes	No
FFPE-tissue	26	C	1	8	282	C→T	CGG→TGG	Arg→Trp	Missense	Sanger/Ion Torrent	Yes	No
Fresh-tissue	273	B	1	4	110	G→T	CGT→CTT	Arg→Leu	Missense	Ion Torrent	Yes	No

* No significant loss of activity, see Supplementary Table H

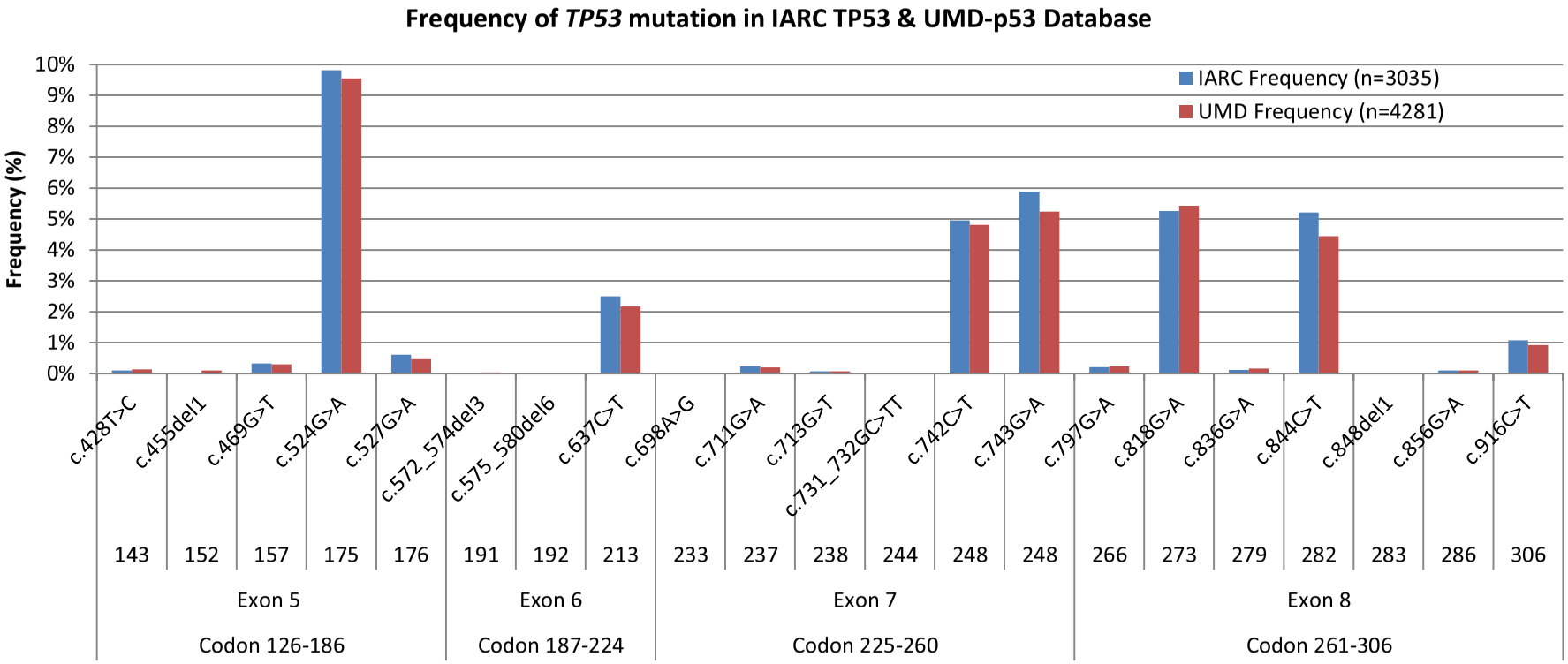


Figure 28: Distribution of somatic *TP53* mutations. Blue bars represents frequencies in IARC TP53 (R16, November 2012) and red bars represents frequencies in UMD-p53 Database (2012_R1). Coding region (cDNA) variant (IARC TP53 and UMD-p53 cDNA reference sequence NM_000546.5) with mutational event and distribution of mutated *TP53* codon with exon boundaries. CRC entries were exclusively selected from each database.

4.5.2 Effect of *TP53* Mutations

The effect of *TP53* mutations were in total 18 (66.7%) missense mutations, 4 (14.8%) nonsense mutations, 4 (14.8%) deletions, and 1 (3.7%) double missense (tandem) mutation (Figure 29A). The majority of the mutations were found within exon 8 (Figure 29B). Since only exons 5-9 were investigated in this study, these results are slightly biased as mutations are distributed throughout all coding exons (exons 2 through 11) of the *TP53* gene (Figure 29C).

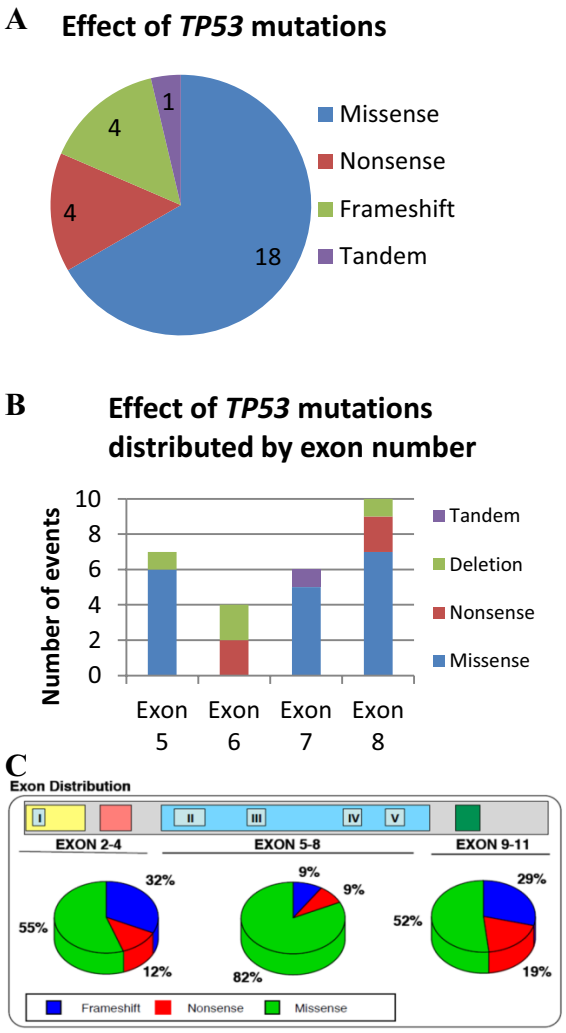


Figure 29: Effect of *TP53* mutations. (**A**) Pie-chart of mutational effect. Missense mutations accounts for the majority of mutations found in this study. (**B**) Effect of *TP53* mutations distributed by exon number. A majority of the mutations were found in exon 8. (**C**) Distribution of p53 mutations in the various functional domains of p53. Top bar: functional domains (Hjortsberg et al. 2008).

4.5.3 *TP53* Mutational Events and CpG dinucleotide

The most frequent type of mutations were purine to purine (G>A, 40.7%) transitions followed by pyrimidine to pyrimidine (C>T, 26%) transitions. Other transitions (A>G, 3.7%; T>C, 3.7%) and transversions (G>T, 7.4%), including deletions (14.8%) and double mutation (tandem) (3.7%) were less frequent (Figure 30A).

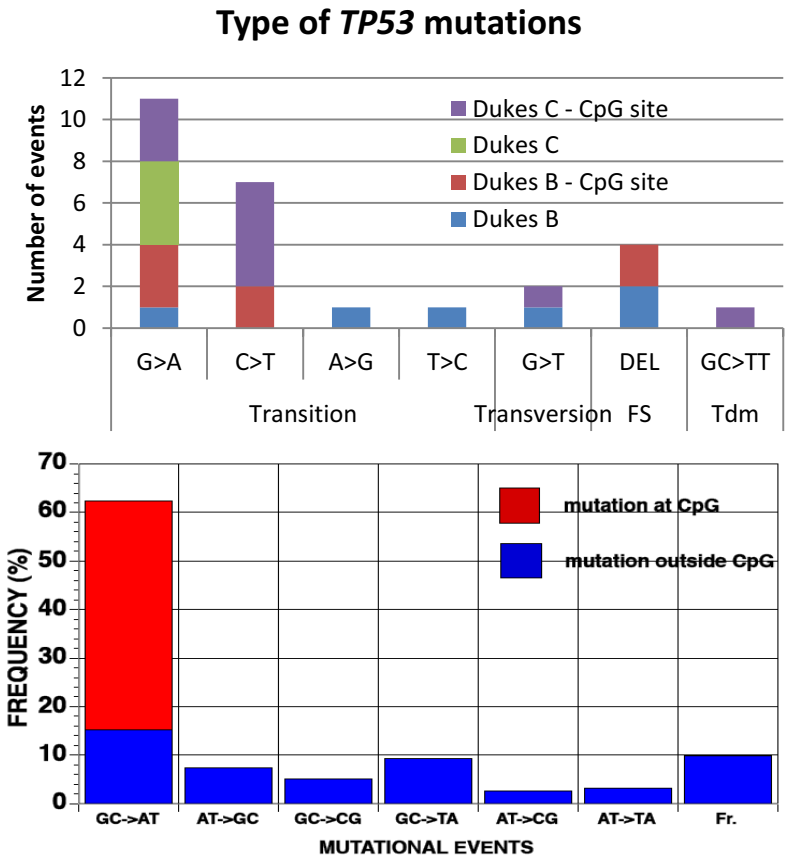


Figure 30: Distribution of mutational events at non-CpG and CpG dinucleotide. (**A**) Distribution of mutational events at non-CpG and CpG sites found in Dukes' stage B and C patients in current study. (**B**) Mutational events at non-CpG and CpG sites in colorectal carcinomas reported in the UMD-p53 database (Hjortsberg et al. 2008).

A total of 17 (63%) mutations were found at CpG dinucleotide, predominantly as G>A and C>T transitions (Figure 30A). However, mutations at CpG dinucleotide in colon carcinoma are dominated by GC->AT transitions with no reports of AT>GC transitions or any transversions in the UMD p53 Mutation database (Figure 30B).

4.5.4 p53 Mutant Activity

p53 mutant activity was analyzed through the UMD-p53 Mutation database spreadsheet, MUT_MAT_II (MUT-TP53 2.0, April 19th, 2014). Four deletions were excluded in the spreadsheet as only exonic missense mutations are allowed entered. Comments generated by the spreadsheet (Supplementary Table H) on mutant activity are the basis for the results reported in this section. Out of 23 mutations analyzed, 18 (78.3%) resulted in inactive proteins, 4 (17.4%) resulted in truncated proteins, and 1 (4.3%) did not display any significant

loss of protein activity. The excluded deletions were added to the pie-chart for a full representation of the mutation spectrum found in this study (Figure 31).

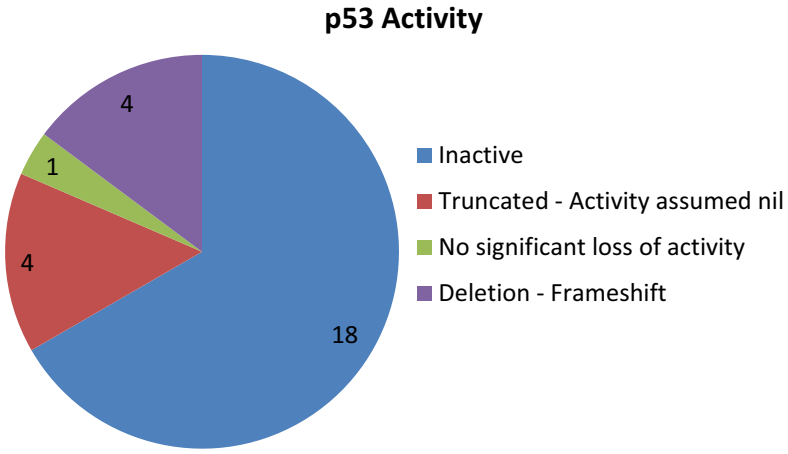


Figure 31: Activity of mutant p53 protein. Activity of mutant p53 protein was generated in UMD p53 spreadsheet MUT_MAT_II (MUT-TP53 2.0). Pie chart showing >50% of the mutations resulted in inactive protein. A truncated protein represents nonsense mutations.

4.5 Survival Analysis

All tumors were adenocarcinomas classified as Dukes' Stage B (n = 39) or C (n = 38). From a total of 77 patients, 9 patients (wt (n = 4) and mut (n = 5)) were excluded from further analysis because clinical information was missing for these patients (clinical data not shown). Survival time was defined from date of diagnosis to date of death from colorectal cancer, or to the last known follow-up date (censored observations), measured in months. Of the remaining 68 patients (Supplementary Table I) (Dukes' stage B (n = 34) and C (n = 34)), 21 were *TP53* mutant and 47 were *TP53* wild-type. At closing date (July 2013) 44 patients were alive of which 2 were wild-type with metastasis, and 27 wild-type and 15 mutants without metastasis. In total, 24 patients had died; 17 of 19 patients with metastasis (wt (n = 14) and mut (n =5)) had died from colorectal cancer, while 7 patients without metastasis had died from other causes. Median age for all patients was 70.5 years (range from 42 to 89 years). Median observed time of complete months lived (for all patients alive or dead) was 26.3 months (range from 0.1 to 97.6 months). Median observed time until last follow up date was 33 months (ranging from 0.7 to 97.6 months). Median observed time until metastasis was 3.3 months (range from 0 to 27.9 months). Median observed time until prime site recurrence (4 Dukes' stage B (wt (n = 2) and mut (n = 2)) was 37 months (range from 0.4 to 97.6 months). Median observed time from date of diagnosis until death occurred was 13.8 months (range from 0.1 to 54.8 months).

To investigate which parameters contribute to differences in survival we applied univariate analysis using Kaplan-Meier modeling and the log-rank test for two clinical parameters: *TP53* mutation status and Dukes' stage. Survival analysis identified Dukes' stage as significant predictor of overall survival. The Kaplan-Meier plot showed a non-significant difference in survival between *TP53* wild-type and mutant patients (Figure 32A). A significant difference in survival was seen between Dukes' stage B and C tumors where stage C patients have significantly worse prognosis (Figure 32B). Metastasis free survival stratified by *TP53* status (n = 7) could not be analyzed as all patients that had metastasis at time of diagnosis (n = 10) were excluded. Correlation between *TP53* status and clinicopathological characteristics are based on Pearson's correlation coefficient (Supplementary Table J).

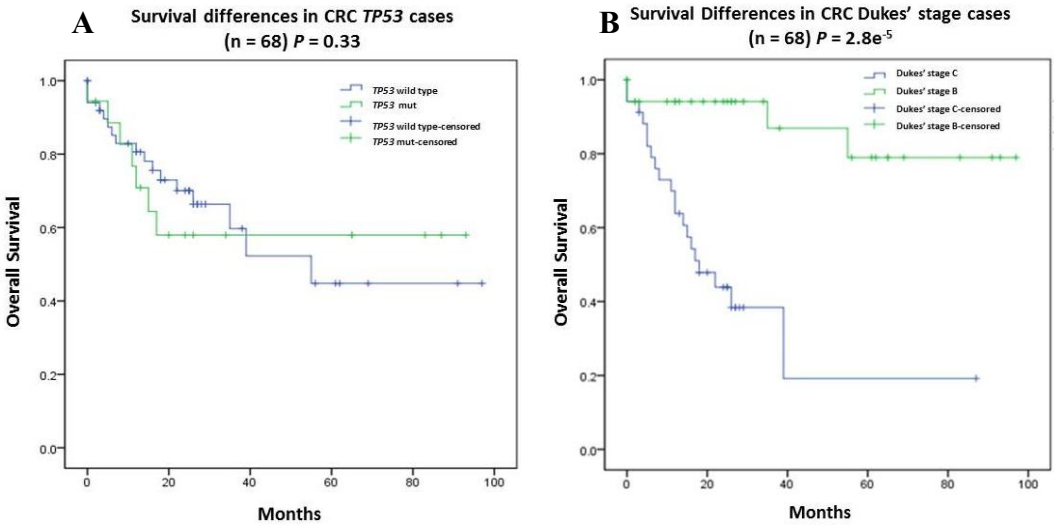


Figure 32: Overall patient survival according to the *TP53* mutation status and Dukes' stage.
(**A**) No difference in survival was identified between wild-type *TP53* and mutant *TP53* patients.
(**B**) Survival differences between Dukes' stage B and C identified Duke´s stage C tumors to have significantly worse prognosis.

5 DISCUSSION

5.1 Statistical Inconsistency of Methods

The number of mutations that can be detected by Sanger sequencing depends on the number of exons that is sequenced within the gene of interest. In contrast, fragment analysis is an allele-specific detection method that tests a sequence change at a specific site, which also makes it difficult to distinguish the mutational mechanism (e.g. point mutation vs. insertions or deletions).

Dias-Santagata et al. (2010) has designed a SNaPshot® Multiplex assay that can detect 58 mutations across 23 exons in 13 critical cancer genes. For this study we focused exclusively on the *TP53* gene that is 393 codons long, while the SNaPshot® assay is designed to detect 7 SNPs within 5 codons in the DNA binding domain.

Due to the different amount of somatic mutations that can be detected between Sanger sequencing (all exons and flanking introns) and fragment analysis (selected SNPs), any significance between results obtained cannot be statistically determined.

5.2 Differences in Mutation Detection

DNA sequencing technology has improved our understanding of biological processes, and the discovery of novel DNA biomarkers has helped to determine treatment outcome in many diseases. With the increased throughput of NGS, multiple genes and whole genomes can now be extensively assessed in contrast to sequencing single genes as traditionally performed with Sanger sequencing (Soussi 2011).

Tumor sequencing projects has discovered thousands of passenger mutations found coselected by driver mutations involved in sustaining carcinogenesis, and many of these passenger mutations are found in non-coding regions. Mutations are thus not confined to hot-spot or frequent mutations at known sites, but occur throughout the whole region of *TP53* (Soussi 2011).

The number of mutations that were detected by Sanger sequencing reflects many potentially undetected mutations when focusing on a few pre-selected SNPs (Supplementary Table E). In addition, our results are also biased and do not reflect the true *TP53* mutation spectrum potentially identified in the studied cohort, since we only focused on exon 5 through 9. The decision for this was twofold; after experiencing several problems in our initial sequencing runs (discussed below) we decided to focus on exon 5 through 9, since these regions contains

the SNPs that is targeted by the SNaPshot® assay, and we saw the need to reduce the workload for the purpose of this study.

5.3 Experimental Procedures and Optimization of Protocols

5.3.1 Sanger Sequencing

Our initial attempt to sequence *TP53* according to Vu et al. (2008) (with slight modifications) was unsuccessful in several ways. First, the vacuum manifold was defect and out of order (personal communication, Post. Doc. Tone Tannæs, EpiGen, Ahus). Secondly, the ABI 3730 DNA Analyzer was vacant at the onset of our experiments and the installed capillaries were most likely dried out and of old age (2 years). Although not elucidated in detail, both of these issues most likely had an effect on the poor results obtained on the ABI 3730 DNA analyzer. These initial problems led us to adapt parts of the SPA protocol already optimized and implemented at the Dept. of Microbiology and Infection Control, Ahus, which was performed on the ABI 3130xl DNA Analyzer, with very clean trace file results.

According to Vu et al. (2008), forward sequences for exon 7 produced trace phase shifting event (TPSE) with mixed peaks 3´ of the motif in the chromatogram (Figure 20 C1). These TPSEs were caused by PCR stutter when the polymerase slips during the amplification of mono- or di-nucleotide repeats of a certain length. Stutter in exon 7 forward sequences was a known problem with the current protocol and a new and improved protocol had been published in the IARC *TP53* database (personal communication, scientist Phuong Vu, Department of Genetics, Radium hospital, Oslo). The TPSEs were eliminated in all forward sequences according to the new IARC (2010) protocol (Figure 20 C2).

Since we only wanted to verify mutations detected by the SNaPshot® assay, Sanger sequencing of FFPE samples were confined to the respective exon where the mutation was detected, including exon 8 for all false-positive results (n = 39/40) (discussed below). Many FFPE samples did require additional PCR- and sequencing runs, including extra PCR-purification step, and these FFPE samples performed well with Sanger sequencing.

5.3.2 SNaPshot®

The SNaPshot® multiplex assay designed by Dias-Santagata et al. (2010) is ideal for use on FFPP derived DNA. Although designed for 13 genes, we solely focused on *TP53* and did not evaluate possible co-occurring mutations in other genes included in the assay.

Encountered issues involved optimizing the amount of extension products added to the genotype reaction. Many samples required several runs with different dilution factor of

extension product before passing automatic data analysis, and each sample seemed to require different amount of extension template for acceptable signal peak heights.

False-positive mutation was called for one SNP in codon 273 (CGT, c.817C>T) (Figure 26) in 39 of 40 FFPE samples. The high rate of false-positives is unknown. No troubleshooting was conducted as our intent was to verify genotyped mutant alleles by Sanger sequencing. If primer stock solution for this allele was contaminated, this would not have been detected as new primer working solutions were made during the experiment. However, no signal peak height was observed in the blank containing all the necessary components for a successful amplification reaction if DNA had been present (Figure 27), and contamination is thus unlikely. Primer concentration was kept constant as recommended in the protocol. All false-positives were verified as wild-type alleles by Sanger sequencing, and it will only be speculation as to the origin of these false-positives without having elucidated in detail.

5.3.3 Ion Torrent PGM™

Ion Torrent sequencing was performed by another scientist during an ongoing evaluation trial of the PGM™ machine implemented at EpiGen. No details will be discussed regarding these experimental procedures, as we will only focus on the mutations detected in the samples (n = 8) selected for Ion Torrent PGM™ sequencing.

5.4 Mutations Detected in Each Method

5.4.1 Sanger Sequencing

Since we only focused on exon 5 through 9, the number of mutations (n = 18) that were detected in FFT samples (n = 37) are restricted.

Although Sanger sequencing of FFPE samples was restricted to exon 8, an additional 3 mutations were detected in these samples apart from the SNPs selected in the SNaPshot® assay. Otherwise, 5 of 6 mutations in FFPE samples were verified by Sanger sequencing.

Deleted nucleotides were difficult to interpret with Sanger sequencing due to mixed peaks 3´ of the chromatogram. Tumors with suspected deletions were selectively chosen for Ion-Torrent PGM™ sequencing.

5.4.2 SNaPshot®

FFPE samples revealed a total of 6 mutations (6/40 tumors) within 5 of 7 SNPs included in the SNaPshot® assay.

To portray a better representation of tissue specificity, FFT tumors were FFPE cast prior to SNaPshot® analysis. However, 2 of 3 samples failed automatic data analysis. All calculated scoring criteria (Supplementary Table G) were within acceptable range, and since these mutations were already confirmed by Sanger sequencing they were not optimized with additional runs.

We did not attempt to validate assay sensitivity as this has been established by Dias-Santagata et al. (2010) as 3% input mutant DNA. Calculated amount of input mutant DNA material for all samples genotyped as a mutant (% MUT, Supplementary Table G) exceeds this limit.

5.4.3 Ion Torrent PGM™

Ion Torrent results demonstrated that deleted nucleotides can be accurately determined through ion semiconductor sequencing. In addition, 1 missense mutation was detected in exon 4 (Table 5), however, this mutation was not included for cancer analysis as our final analysis was restricted to exon 5 through 9. The respective sample was the one that failed PCR amplification for all primer pairs (Figure 18) according to Vu et al. (2008). In retrospect, after accurately determining the nucleotide composition of deletions, the inability to interpret the deleted nucleotides from Sanger sequencing might reflect the author's inexperience. Once determined these nucleotides also became more prominent in the Sanger sequencing chromatograms.

5.6 Cancer Analysis

Different cancers show different mutational pattern, and some type of mutations show tissue-specificity, like tandem mutations caused by UV-light and distinct mutations in lung cancer linked to cigarette smoke. Therefore, cancer analysis involves a detailed investigation of the specific mutant characteristics. From the cancer analysis conducted in this study, our results were in concordance with reported characteristics for *TP53* mutations (Figure 29C and 30B). Most mutations were missense mutations that resulted in inactive protein, with a few nonsense mutations that result in early termination of a truncated protein. Only 1 mutation (codon 233, CAC) was reported to have no significant loss of activity (sup table x), and should be carefully reviewed before publication. However, this mutant was detected in both strands (Figure 20, C2) and was concluded to be a true mutation that was included in cancer analysis. Type I and type II CpG sites were frequently mutated (17/26) in proteins reported to be inactive.

Evaluation of chromatograms revealed several PCR artifacts portraying heterozygous mutations with double peaks (results not shown). Generation of UMD-p53 MUT-MAT_II table reported these mutations as previously never described (results not shown). Since most of these artifacts were observed in single strands from samples where both forward and reverse sequences were assembled, these mutations were concluded to be true PCR artifacts and excluded from any analysis.

A detailed analysis of the mutational effect on protein function is important since incorrect mutant entries in the UMD-p53 database are estimated to account for 2-5% of all entries (Soussi & Beroud 2003). The curation of these mutations will continue to provide valuable information of the nature of *TP53* function in carcinogenesis, and simple errors in *TP53* mutation reports should be avoided by all means.

5.5 Survival Analysis

In this study we have examined the overall survival outcome by Dukes' stage and *TP53* mutation status in 68 primary colorectal adenocarcinomas. Our results demonstrate that Dukes' stage B is predictive of inferior survival with significantly worse prognosis in Dukes' stage C patients. No difference in survival is observed for *TP53* wild-type patients opposed to *TP53* mutant patients. However, 9 (wt (n = 4), mutant (n = 5) patients with missing clinicopathological data were excluded. Dukes' stage and *TP53* status were the only parameters available for these patients, without any information of date of operation or registered follow-up date at any given time. By having to exclude 5 *TP53* mutant patients (Dukes' stage B (n = 2) and C (n = 3)) due to missing clinicopathological data, these results does not reflect the true survival outcome in the studied cohort.

In addition, one patient with a mutation reported to have no loss in activity, were overlooked and included in the survival analysis table. This patient was alive at closing date and was treated as censored object for the Kaplan-Meier plots.

5.3 Implementation of New Protocols

This study was performed to evaluate the SNaPshot® fragment analysis assay, in contrast to Sanger sequencing, as a potential method to be introduced at Ahus for molecular profiling of FFPE material from CRC cases. Either of the performed protocols had previously been executed at EpiGen, and without an experienced conductor of these methods we faced many problems that could have been avoided with a detailed research regarding possible flaws in

protocol performance, and acquired more information about the instruments we used at EpiGen. Many aspects of this research have come about in a “learning-by-doing” manner. A valuable experience gained for prospective research work.

6 CONCLUSION

We have found the SNaPshot® assay designed by Dias-Santagata et al. (2010) to be highly robust, sensitive and accurate, and ideal for use on FFPE derived DNA. This system can be used in any laboratory with a capillary DNA sequencer.

Considering the nature of *TP53* mutation spectra it is important to include all coding regions with exon/intron boundaries when conducting a mutation analysis of *TP53*. Although we confined our research to selected exons within the *TP53* DNA binding domain, the SNaPshot® assay did not meet our expectations for our intended purpose of achieving new knowledge for colorectal cancer cases at Ahus.

A significant difference in overall survival was established for Dukes' stage B patients, but not for *TP53* wild-type patients. No mutations gave any indication of prognostic or predictive biomarkers

The identification of new genetic markers is important for new drug development and to aid oncologists to the best suited treatment for their patients. We consider the SNaPshot® assay to have more relevance for detecting established, clinically proven molecular markers to stratify patients for best treatment outcome and survival predicament.

REFERENCE LIST

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K. & Walter, P. (2008). *Molecular biology of the cell* 5th ed. New York, USA: Garland Scinece, Taylor & Francis Group

Alitalo, K., Schwab, M., Lin, C. C., Varmus, H. E. & Bishop, J. M. (1983). Homogeneously Staining Chromosomal Regions Contain Amplified Copies of an Abundantly Expressed Cellular Oncogene (c-myc) in Malignant Neuroendocrine Cells from a Human Colon Carcinoma. *Proceedings of the National Academy of Sciences of the United States of America*, 80 (6): 1707-1711.

Alrawi, S. J., Schiff, M., Carroll, R. E., Dayton, M., Gibbs, J. F., Kulavlat, M., Tan, D., Berman, K., Stoler, D. L. & Anderson, G. R. (2006). Aberrant Crypt Foci. *Anticancer Research*, 26 (1A): 107-119.

Amaral, J. D., Xavier, J. M., Steer, C. J. & Rodrigues, C. M. (2010). The role of p53 in apoptosis. *Discov Med*, 9 (45): 145-52.

Anderson, J. C., Pleau, D. C., Rajan, T. V., Protiva, P., Swede, H., Brenner, B., Heinen, C. D., Lambrecht, R. W. & Rosenberg, D. W. (2010). Increased Frequency of Serrated Aberrant Crypt Foci Among Smokers. *Am J Gastroenterol*, 105 (7): 1648-1654.

Baker, S. J., Fearon, E. R., Nigro, J. M., Hamilton, S. R., Preisinger, A. C., Jessup, J. M., vanTuinen, P., Ledbetter, D. H., Barker, D. F., Nakamura, Y., et al. (1989). Chromosome 17 deletions and p53 gene mutations in colorectal carcinomas. *Science*, 244 (4901): 217-21.

Benchimol, S., Lamb, P., Crawford, L. V., Sheer, D., Shows, T. B., Bruns, G. A. P. & Peacock, J. (1985). Transformation associated p53 protein is encoded by a gene on human chromosome 17. *Somatic Cell and Molecular Genetics*, 11 (5): 505-510.

Blair, D. G., Oskarsson, M., Wood, T. G., McClements, W. L., Fischinger, P. J. & Woude, G. G. V. (1981). Activation of the Transforming Potential of a Normal Cell Sequence: A Molecular Model for Oncogenesis. *Science*, 212 (4497): 941-943.

Bodmer, W. F. (2006). Cancer genetics: colorectal cancer as a model. *J Hum Genet*, 51 (5): 391-6.

Brady, C. A. & Attardi, L. D. (2010). p53 at a glance. *Journal of Cell Science*, 123 (15): 2527-2532.

Burke, J. R., Hura, G. L. & Rubin, S. M. (2012). Structures of inactive retinoblastoma protein reveal multiple mechanisms for cell cycle control. *Genes & Development*, 26 (11): 1156-1166.

Burns, T. F. & El-Deiry, W. S. (1999). The p53 pathway and apoptosis. *Journal of Cellular Physiology*, 181 (2): 231-239.

Cardoso, J., Boer, J., Morreau, H. & Fodde, R. (2007). Expression and genomic profiling of colorectal cancer. *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer*, 1775 (1): 103-137.

Chen, Y., Dey, R. & Chen, L. (2010). Crystal Structure of the p53 Core Domain Bound to a Full Consensus Site as a Self-Assembled Tetramer. *Structure*, 18 (2): 246-256.

Cho, K. R. & Vogelstein, B. (1992). Genetic alterations in the adenoma--carcinoma sequence. *Cancer*, 70 (6 Suppl): 1727-31.

Compton, C. C. & Greene, F. L. (2004). The Staging of Colorectal Cancer: 2004 and Beyond. *CA: A Cancer Journal for Clinicians*, 54 (6): 295-308.

Conrad, S. E. & Botchan, M. R. (1982). Isolation and characterization of human DNA fragments with nucleotide sequence homologies with the simian virus 40 regulatory region. *Molecular and Cellular Biology*, 2 (8): 949-965.

Coutts, A. S., Adams, C. J. & La Thangue, N. B. (2009). p53 ubiquitination by Mdm2: A never ending tail? *DNA Repair*, 8 (4): 483-490.

Crawford, L. V., Pim, D. C., Gurney, E. G., Goodfellow, P. & Taylor-Papadimitriou, J. (1981). Detection of a Common Feature in Several Human Tumor Cell Lines--A 53,000-Dalton Protein. *Proceedings of the National Academy of Sciences of the United States of America*, 78 (1): 41-45.

Dahl, O. (2007). Adjuvant kjemoterapi ved tykktarmskreft. *Tidskrift for den Norske Lægeforeningen*, 23 (127).

Dalla-Favera, R., Bregni, M., Erikson, J., Patterson, D., Gallo, R. C. & Croce, C. M. (1982a). Human c-myc onc gene is located on the region of chromosome 8 that is translocated in Burkitt lymphoma cells. *Proceedings of the National Academy of Sciences*, 79 (24): 7824-7827.

Dalla-Favera, R., Wong-Staal, F. & Gallo, R. C. (1982b). onc gene amplification in promyelocytic leukaemia cell line HL-60 and primary leukaemic cells of the same patient. *Nature*, 299 (5878): 61-63.

Damania, B. (2007). DNA tumor viruses and human cancer. *Trends in Microbiology*, 15 (1): 38-44.

DeFeo, D., Gonda, M. A., Young, H. A., Chang, E. H., Lowy, D. R., Scolnick, E. M. & Ellis, R. W. (1981). Analysis of two divergent rat genomic clones homologous to the transforming gene of Harvey murine sarcoma virus. *Proceedings of the National Academy of Sciences*, 78 (6): 3328-3332.

Dias-Santagata, D., Akhavanfard, S., David, S. S., Vernovsky, K., Kuhlmann, G., Boisvert, S. L., Stubbs, H., McDermott, U., Settleman, J., Kwak, E. L., et al. (2010). Rapid targeted mutational analysis of human tumours: a clinical platform to guide personalized cancer medicine. *EMBO Molecular Medicine*, 2 (5): 146-158.

Dong, J. T. (2001). Chromosomal deletions and tumor suppressor genes in prostate cancer. *Cancer Metastasis Rev*, 20 (3-4): 173-93.

el-Deiry, W. S. (1998). Regulation of p53 downstream genes. *Semin Cancer Biol*, 8 (5): 345-57.

Fauci, A., Kasper, D., Braunwald, E., Hauser, S., Longo, D., Jameson, J. & Loscalzo, J. (2008). *Harrison's Principal of Internal Medicine*, vol. 17th Edition: http://accessmedicine.com.

Fearon, E. R. & Vogelstein, B. (1990). A genetic model for colorectal tumorigenesis. *Cell*, 61 (5): 759-67.

Fernandez-Medarde, A. & Santos, E. (2011). Ras in cancer and developmental diseases. *Genes Cancer*, 2 (3): 344-58.

Finlay, C. A., Hinds, P. W. & Levine, A. J. (1989). The p53 proto-oncogene can act as a suppressor of transformation. *Cell*, 57 (7): 1083-93.

Gottlieb, E. & Vousden, K. H. (2010). p53 Regulation of Metabolic Pathways. *Cold Spring Harbor Perspectives in Biology*, 2 (4).

Gruss, P., Dhar, R. & Khoury, G. (1981). Simian virus 40 tandem repeated sequences as an element of the early promoter. *Proceedings of the National Academy of Sciences*, 78 (2): 943-947.

Hanahan, D. & Weinberg, R. A. (2000). The Hallmarks of Cancer. *Cell*, 100 (1): 57-70.

Hanahan, D. & Weinberg, Robert A. (2011). Hallmarks of Cancer: The Next Generation. *Cell*, 144 (5): 646-674.

Harlow, E., Williamson, N. M., Ralston, R., Helfman, D. M. & Adams, T. E. (1985). Molecular cloning and in vitro expression of a cDNA clone for human cellular tumor antigen p53. *Molecular and Cellular Biology*, 5 (7): 1601-1610.

Hayward, W. S., Neel, B. G. & Astrin, S. M. (1981). Activation of a cellular onc gene by promoter insertion in ALV-induced lymphoid leukosis. *Nature*, 290 (5806): 475-480.

Helt, A.-M. & Galloway, D. A. (2003). Mechanisms by which DNA tumor virus oncoproteins target the Rb family of pocket proteins. *Carcinogenesis*, 24 (2): 159-169.

Helton, E. S. & Chen, X. (2007). p53 modulation of the DNA damage response. *Journal of Cellular Biochemistry*, 100 (4): 883-896.

Hiyama, H., Iavarone, A. & Reeves, S. A. (1998). Regulation of the cdk inhibitor p21 gene during cell cycle progression is under the control of the transcription factor E2F. *Oncogene*, 16 (12): 1513-23.

Hjortsberg, L., Rubio-Nevado, J. M., Hamroun, D., Claustres, M., Béroud, C. & Soussi, T. (2008). *The p53 Mutation HandBook v 2. (Oct 2008).* *http://p53.free.fr/Database/p53_mutation_HB.html*.

http://p53.iarc.fr/. (2010). *Detection of TP53 mutations by direct sequencing*. Available at: http://p53.iarc.fr/Download/TP53_DirectSequencing_IARC.pdf (accessed: 11. June).

Hu, Q. J., Dyson, N. & Harlow, E. (1990). The regions of the retinoblastoma protein needed for binding to adenovirus E1A or SV40 large T antigen are common sites for mutations. *EMBO J*, 9 (4): 1147-55.

Iacopetta, B., Russo, A., Bazan, V., Dardanoni, G., Gebbia, N., Soussi, T., Kerr, D., Elsaleh, H., Soong, R., Kandioler, D., et al. (2006). Functional categories of TP53 mutation in colorectal cancer: results of an International Collaborative Study. *Annals of Oncology*, 17 (5): 842-847.

Itahana, K., Dimri, G. & Campisi, J. (2001). Regulation of cellular senescence by p53. *European Journal of Biochemistry*, 268 (10): 2784-2791.

Joerger, A. C. & Fersht, A. R. (2010). The tumor suppressor p53: from structures to drug discovery. *Cold Spring Harb Perspect Biol*, 2 (6): a000919.

Jones, P. A. & Laird, P. W. (1999). Cancer epigenetics comes of age. *Nat Genet*, 21 (2): 163-7.

Kanazawa, T., Watanabe, T., Kazama, S., Tada, T., Koketsu, S. & Nagawa, H. (2002). Poorly differentiated adenocarcinoma and mucinous carcinoma of the colon and rectum show higher rates of loss of heterozygosity and loss of E-cadherin expression due to methylation of promoter region. *International Journal of Cancer*, 102 (3): 225-229.

Kato, S., Han, S. Y., Liu, W., Otsuka, K., Shibata, H., Kanamaru, R. & Ishioka, C. (2003). Understanding the function-structure and function-mutation relationships of p53 tumor suppressor protein by high-resolution missense mutation analysis. *Proc Natl Acad Sci U S A*, 100 (14): 8424-9.

Kawasaki, T., Ohnishi, M., Nosho, K., Suemoto, Y., Kirkner, G. J., Meyerhardt, J. A., Fuchs, C. S. & Ogino, S. (2008). CpG island methylator phenotype-low (CIMP-low) colorectal cancer shows not only few methylated CIMP-high-specific CpG islands, but also low-level methylation at individual loci. *Mod Pathol*, 21 (3): 245-255.

Knudson, A. G., Jr. (1971). Mutation and cancer: statistical study of retinoblastoma. *Proc Natl Acad Sci U S A*, 68 (4): 820-3.

Kruse, J. P. & Gu, W. (2009). Modes of p53 regulation. *Cell*, 137 (4): 609-22.

Kubbutat, M. H., Jones, S. N. & Vousden, K. H. (1997). Regulation of p53 stability by Mdm2. *Nature*, 387 (6630): 299-303.

Kulendran, M., Stebbing, J. F., Marks, C. G. & Rockall, T. A. (2011). Predictive and Prognostic Factors in Colorectal Cancer: A Personalized Approach. *Cancers*, 3 (2): 1622-1638.

Lamb, P. & Crawford, L. (1986). Characterization of the human p53 gene. *Molecular and Cellular Biology*, 6 (5): 1379-1385.

Land, H., Parada, L. F. & Weinberg, R. A. (1983). Cellular Oncogenes and Multistep Carcinogenesis. *Science*, 222 (1625): 771-778.

Lane, D. P. & Crawford, L. V. (1979). T antigen is bound to a host protein in SY40-transformed cells. *Nature*, 278 (5701): 261-263.

Leder, P., Battey, J., Lenoir, G., Moulding, C., Murphy, W., Potter, H., Stewart, T. & Taub, R. (1983). Translocations among antibody genes in human cancer. *Science*, 222 (4625): 765-771.

Leroy, B., Fournier, J. L., Ishioka, C., Monti, P., Inga, A., Fronza, G. & Soussi, T. (2013). The TP53 website: an integrative resource centre for the TP53 mutation database and TP53 mutant analysis. *Nucleic Acids Res*, 41 (Database issue): D962-9.

Malumbres, M. & Barbacid, M. (2003). RAS oncogenes: the first 30 years. *Nat Rev Cancer*, 3 (6): 459-65.

Martin, G. S. (2004). The road to Src. *Oncogene*, 23 (48): 7910-7917.

McBride, O., Merry, D. & Givol, D. (1986). The gene for human p53 cellular tumor antigen is located on chromosome 17 short arm (17p13). *Proceedings of the National Academy of Sciences*, 83 (1): 130-134.

Moll, U. M. & Petrenko, O. (2003). The MDM2-p53 interaction. *Mol Cancer Res*, 1 (14): 1001-8.

Mosby's medical dictionary. (8th ed.) (2009). St Louis, M. M. E.

Mushinski, J. F., Potter, M., Bauer, S. R. & Reddy, E. P. (1983). DNA Rearrangement and Altered RNA Expression of the c-myb Oncogene in Mouse Plasmacytoid Lymphosarcomas. *Science*, 220 (4599): 795-798.

Muto, T., Bussey, H. J. & Morson, B. C. (1975). The evolution of cancer of the colon and rectum. *Cancer*, 36 (6): 2251-70.

Münger, K., Baldwin, A., Edwards, K. M., Hayakawa, H., Nguyen, C. L., Owens, M., Grace, M. & Huh, K. (2004). Mechanisms of Human Papillomavirus-Induced Oncogenesis. *Journal of Virology*, 78 (21): 11451-11460.

Nevins, J. R. (2001). The Rb/E2F pathway and cancer. *Human Molecular Genetics*, 10 (7): 699-703.

Olivier, M., Hollstein, M. & Hainaut, P. (2010). TP53 Mutations in Human Cancers: Origins, Consequences, and Clinical Use. *Cold Spring Harbor Perspectives in Biology*, 2 (1).

Ougolkov, A., Zhang, B., Yamashita, K., Bilim, V., Mai, M., Fuchs, S. Y. & Minamoto, T. (2004). Associations Among β-TrCP, an E3 Ubiquitin Ligase Receptor, β-Catenin, and NF-κB in Colorectal Cancer. *Journal of the National Cancer Institute*, 96 (15): 1161-1170.

Qian, Y. & Chen, X. (2013). Senescence regulation by the p53 protein family. *Methods Mol Biol*, 965: 37-61.

Ramalakshmi, S. & Muthuchelian, K. (2011). Cancer and Oncogenes - An Overview. *Academic Journal of Cancer Research*, 4 (1): 10-17.

Reddy, E. P., Reynolds, R. K., Santos, E. & Barbacid, M. (1982). A point mutation is responsible for the acquisition of transforming properties by the T24 human bladder carcinoma oncogene. *Nature*, 300 (5888): 149-152.

Sah, S., Chen, L., Houghton, J., Kemppainen, J., Marko, A., Zeigler, R. & Latham, G. (2013). Functional DNA quantification guides accurate next-generation sequencing mutation detection in formalin-fixed, paraffin-embedded tumor biopsies. *Genome Medicine*, 5 (8): 77.

Santos, E., Reddy, E. P., Pulciani, S., Feldmann, R. J. & Barbacid, M. (1983). Spontaneous Activation of a Human Proto-Oncogene. *Proceedings of the National Academy of Sciences of the United States of America*, 80 (15): 4679-4683.

Schadt, E. E., Turner, S. & Kasarskis, A. (2010). A window into third-generation sequencing. *Human Molecular Genetics*, 19 (R2): R227-R240.

Schwab, M., Alitalo, K., Varmus, H. E., Bishop, J. M. & George, D. (1983). A cellular oncogene (c-Ki-ras) is amplified, overexpressed, and located within karyotypic abnormalities in mouse adrenocortical tumour cells. *Nature*, 303 (5917): 497-501.

Seto, E., Usheva, A., Zambetti, G. P., Momand, J., Horikoshi, N., Weinmann, R., Levine, A. J. & Shenk, T. (1992). Wild-type p53 binds to the TATA-binding protein and represses transcription. *Proc Natl Acad Sci U S A*, 89 (24): 12028-32.

Shendure, J. & Aiden, E. L. (2012). The expanding scope of DNA sequencing. *Nat Biotech*, 30 (11): 1084-1094.

Soussi, T. & Beroud, C. (2003). Significance of TP53 mutations in human cancer: a critical analysis of mutations at CpG dinucleotides. *Hum Mutat*, 21 (3): 192-200.

Soussi, T., Kato, S., Levy, P. P. & Ishioka, C. (2005). Reassessment of the TP53 mutation database in human disease by data mining with a library of TP53 missense mutations. *Human Mutation*, 25 (1): 6-17.

Soussi, T. (2011). Advances in carcinogenesis: A historical perspective from observational studies to tumor genome sequencing and TP53 mutation spectrum analysis. *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer*, 1816 (2): 199-208.

Steele, R. J. C., Thompson, A. M., Hall, P. A. & Lane, D. P. (1998). The p53 tumour suppressor gene. *British Journal of Surgery*, 85 (11): 1460-1467.

Tabin, C. J., Bradley, S. M., Bargmann, C. I., Weinberg, R. A., Papageorge, A. G., Scolnick, E. M., Dhar, R., Lowy, D. R. & Chang, E. H. (1982). Mechanism of activation of a human oncogene. *Nature*, 300 (5888): 143-149.

Taparowsky, E., Shimizu, K., Goldfarb, M. & Wigler, M. (1983). Structure and activation of the human N-ras gene. *Cell*, 34 (2): 581-586.

TCGA, T. C. G. A. (2012). Comprehensive molecular characterization of human colon and rectal cancer. *Nature*, 487 (7407): 330-337.

Venot, C., Maratrat, M., Dureuil, C., Conseiller, E., Bracco, L. & Debussche, L. (1998). *The requirement for the p53 proline-rich functional domain for mediation of apoptosis is correlated with specific PIG3 gene transactivation and with transcriptional repression*, vol. 17. 4668-4679 pp.

Vogelstein, B., Lane, D. & Levine, A. J. (2000). Surfing the p53 network. *Nature*, 408 (6810): 307-310.

Weinberg, R. A. (1991). Tumor suppressor genes. *Science (New York, N.Y.)*, 254 (5035): 1138-1146.

Weinberg, R. A. (1994). Oncogenes and tumor suppressor genes. *CA: A Cancer Journal for Clinicians*, 44 (3): 160-170.

Weinberg, R. A. (1995). The Molecular Basis of Oncogenes and Tumor Suppressor Genes. *Annals of the New York Academy of Sciences*, 758 (1): 331-338.

Winder, T. & Lenz, H.-J. (2010). Molecular predictive and prognostic markers in colon cancer. *Cancer Treatment Reviews*, 36 (7): 550-556.

Yamauchi, M., Lochhead, P., Morikawa, T., Huttenhower, C., Chan, A. T., Giovannucci, E., Fuchs, C. & Ogino, S. (2012). Colorectal cancer: a tale of two sides or a continuum? *Gut*.

Yeatman, T. J. (2013). Colon Cancer. In: eLS. John Wiley & Sons Ltd, Chichester. http://www.els.net [doi: 10.1038/npg.els.0001891].

APPENDIX A: Supplementary Tables

Supplementary Table A: Properties for *TP53* amplification primer sequences according to Vu et al. (2008) and IARC (2010) protocols.

Vu et al. (2008)		
Exon	**Amplification primer sequences**	**PCR fragment length (Bp)**
2, 3	F - 5′-GGAGTGCTTGGGTTGTGGT-3′ R - 5′-CGGCAAGGGGGACTGTA-3′	586
4	F - 5′-GACTTCCTGAAAACAACG-3′ R - 5′-CACACATTAAGTGGGTAAAC-3′	593
5, 6	F - 5′-TTTCTTTGCTGCCGTCTTC-3′ R - 5′-TTGCACATCTCATGGGGTTA-3′	588
7	F - 5′-GACCATCCTGGCTAACGG-3′ R - 5′-CACAGGTTAAGAGGTCCCAAA-3′	595
8, 9	F - 5′-TTTGGGACCTCTTAACCTGT-3′ R - 5′-CAGGCAAAGTCATAGAACCAT-3′	733
10	F - 5′-CATGTTGCTTTTGTACCGTC-3′ R - 5′-GGCAAGAATGTGGTTATAGGA-3′	396
11	F - 5′-AAGGGAAGATTACGAGACT- 3′ R - 5′-TAAGCTGGTATGTCCTACTC-3′	500
Extension primer - Universal M13 sequences		
-21 M13	F - 5′-TGTAAAACGACGGCCAGT- 3′	NA
-M13 REV	R - 5′-CAGGAAACAGCTATGACC- 3′	NA
2010 updated IARC protocol		
Amplification and extension primer sequences		
7	P-333 F - 5′-CTTGCCACAGGTCTCCCCAA-3′ P313 R - 5′-AGGGGTCAGAGGCAAGCAGA-3′	237

F = Forward, R = Reverse; Amplification primers are linked to universal M13 sequences in the 5` position

Supplementary Table B: Properties for *TP53* Sanger sequencing amplification and extension primer sequences according to Dias-Santagata et al., (2010).

Primers	**Amplification and extension primer sequences**	**Annealing Temp**	**$MgCl_2$**
TP53_ex5_ Seq_a1 TP53_ex5_ Seq_a2	5′-CTTGTGCCCTGACTTTCAAC-3′ 5′-ACCAGCCCTGTCGTCTCTC-3′	64ºC	40 nmol
TP53_ex7_ Seq_a1 TP53_ex7_ Seq_a2	5′-TCATCTTGGGCCTGTGTTATC-3′ 5′-GAAATCGGTAAGAGGTGGGC -3′	58ºC	50 nmol
TP53_ex8_ Seq_a1 TP53_ex8_ Seq_a2	5′-TTTCCTTACTGCCTCTTGCTTC-3′ 5′-GGAAAGGTGATAAAAGTGAATCTG -3′	58ºC	50 nmol

Supplementary Table C: SNPs targeted by the SNaPshot assay.

Codon	Nucleotide position	CpG site
175	c.524G	Yes
245	c.733G	Yes
248	c.742C	Yes
248	c.743G	Yes
273	c.817C	Yes
273	c.818G	Yes
306	c.916C	Yes

Supplementary Table D: Forward and reverse primer amount for SNaPshot® analysis PCR reaction mix according to Dias-Santagata et al. (2010).

Multiplex PCR primers	Primer amount[1]
Panel II - 742	
TP53_exon7_(a1+a2)	0.6 pmol
Panel IV - 733	
TP53_exon7_(a1+a2)	0.6 pmol
Panel V - 743/817	
TP53_exon7_(a1+a2)	0.6 pmol
TP53_exon8_(a1+a2)	0.6 pmol
Panel VI - 818	
TP53_exon8_(a1+a2)	1.2 pmol
Panel VIII – 524/916	
TP53_exon5_(a1+a2)	1.2 pmol
TP53_exon8_(a1+a2)	0.6 pmol

[1]"The value indicated represents the amount of each forward and reverse primer added per 10 µl of multiplex PCR (X pmol of a1 and X pmol of a2)."

Supplementary Table E: Properties for *TP53* SNaPshot® amplification- and extension primer sequences according to Dias-Santagata et al., (2010).

Amplification primers	Amplification primer sequences *
TP53_exon 5_a1	5′-*ACGTTGGATG*CAAGCAGTCACAGCACATGA-3′
TP53_exon 5_a2	5′-*ACGTTGGATG*CTGCTCACCATCGCTATCTG-3′
TP53_exon 7_a1	5′-*ACGTTGGATG*TGGCTCTGACTGTACCACCA-3′
TP53_exon 7_a2	5′-*ACGTTGGATG*CCAGTGTGATGATGGTGAGG-3′
TP53_exon 8_a1	5′-*ACGTTGGATG*CTACTGGGACGGAACAGCTT-3′
TP53_exon 8_a2	5′-*ACGTTGGATG*GCTTCTTGTCCTGCTTGCTT-3′
Extension primers	**Extension primer sequences**
TP53_524_extF	5′-GACTGACTGACTGACTGACTGACTGACTGACTGACTGA CTGACTGACTGACTGACTGACTGACTGACCGGAGGTTGT G AGGC-3′
TP53_733_extR	5′-GACTGACTGACTGACTGACTGACTGACTGACTGACTC C TCCGGTTCATGC-3′
TP53_742_extF	5′-ACTGACTGACTGACTGACTGACTGACTGACTGACTGAC TGACTGACTGACTGGGGCGGCATGAAC-3′
TP53_743_extF	5′-CTGACTGACTGACTGACTGACTGACTGACTGACTGACG GCGGCATGAACC-3′
TP53_817_extF	5′-CTGACTGACTGACTGACTGACTGACTGACTGACTGACT GACTGACTGAGGAACAGCTTTGAGGTG-3′
TP53_818_extF	5′-ACTGACTGACTGACTGACTGACTGACTGACTGACTGAC TGACTGACTGACTGACTGACTGACTGAGAACAGCTTTGAG GTGC-3′
TP53_916_extR	5′-TGACTGACTGACTGACTGACTGACTGACTGACTGACTG ACTGACTGACTGACTGACTGACTGACTGACTGAGTCCTGCTGC TTACCTC-3′

*10 nt long 5′- anchor tail (5′-ACGTTGGATG-3′) attached to all sequences; a1 = Forward, a2 = Reverse.

Supplementary Table F: Absorbance reading values of extracted DNA.

Sample ID	DNA ng/ul	A260	A280	260/280	260/230
CRC 88	58,70	1,174	0,594	1,98	0,84
CRC 89	133,43	2,669	1,365	1,95	0,53
CRC 90	110,60	2,212	1,297	1,71	0,31
CRC 97	564,15	11,283	6,484	1,74	1,15
CRC 104	87,94	1,759	0,912	1,93	0,46
CRC 106	89,47	1,789	0,941	1,90	0,30
CRC 107	132,52	2,650	1,525	1,74	0,52
CRC 108	207,05	4,141	2,304	1,80	0,50
CRC 112	72,99	1,460	0,743	1,96	1,71
CRC 120	92,17	1,843	0,981	1,88	0,48
CRC 124	156,36	3,127	1,754	1,78	0,45
CRC 127	192,40	3,848	2,199	1,75	0,32
CRC 134	43,08	0,862	0,441	1,96	1,92
CRC 140	107,31	2,146	1,152	1,86	0,40
CRC 143	79,14	1,583	0,790	2,00	1,22
CRC 147	51,91	1,038	0,555	1,87	1,62
CRC 164	102,97	2,059	1,044	1,97	0,81
CRC 176	33,17	0,663	0,324	2,05	1,21
CRC 200	104,33	2,087	1,059	1,97	0,75
CRC 204	211,25	4,225	2,594	1,63	0,38
CRC 207	106,06	2,121	1,054	2,01	1,28
CRC 210	39,39	0,788	0,415	1,90	0,58
CRC 211	129,15	2,583	1,296	1,99	1,74
CRC 222	156,42	3,128	1,792	1,75	0,39
CRC 232	94,94	1,899	0,952	2,00	1,39
CRC 240	106,70	2,134	1,042	2,05	1,82
CRC 253	88,20	1,764	0,886	1,99	2,01
CRC 255	89,27	1,785	0,892	2,00	2,06
CRC 269	32,10	0,642	0,325	1,98	1,21
CRC 273	286,50	5,730	3,359	1,71	0,34
CRC 280	85,95	1,719	0,945	1,82	0,39
CRC 285	154,32	3,086	1,727	1,79	0,52
CRC 290	46,29	0,926	0,472	1,96	1,26
CRC 303	138,62	2,772	1,604	1,73	0,47
CRC 304	38,08	0,762	0,406	1,88	1,51
CRC 315	39,73	0,795	0,451	1,76	0,74
CRC 345	38,75	0,775	0,396	1,96	1,46
CRC-FFPE 108	470,37	9,407	4,610	2,04	2,05
CRC-FFPE 204	199,54	3,991	2,049	1,95	1,66
CRC-FFPE 280	367,31	7,346	3,619	2,03	2,10
FFPE 1	291,60	5,832	2,849	2,05	0,75
FFPE 2	49,23	0,985	0,445	2,21	0,26
FFPE 3	101,40	2,028	1,027	1,97	0,47
FFPE 4	94,84	1,897	0,901	2,11	0,38
FFPE 5	146,10	2,923	1,430	2,04	0,81
FFPE 6	267,70	5,355	2,599	2,06	0,95
FFPE 7	354,80	7,096	3,463	2,05	1,24
FFPE 8	227,80	4,557	2,209	2,06	0,94
FFPE 9	285,20	5,704	2,790	2,04	1,13
FFPE 10	89,08	1,782	0,839	2,12	0,59
FFPE 11	371,80	7,435	3,594	2,07	0,93
FFPE 12	258,70	5,174	2,472	2,09	1,15
FFPE 13	141,90	2,837	1,365	2,08	0,74
FFPE 14	191,30	3,827	1,847	2,07	0,53
FFPE 15	166,30	3,325	1,617	2,06	0,53
FFPE 16	266,40	5,328	2,531	2,11	0,71
FFPE 17	67,67	1,353	0,646	2,09	0,32
FFPE 18	131,20	2,624	1,330	1,97	0,65
FFPE 19	94,82	1,896	0,979	1,94	0,61
FFPE 20	103,90	2,079	1,003	2,07	0,55
FFPE 21	40,50	0,810	0,395	2,05	0,37
FFPE 22	263,60	5,273	2,503	2,11	0,78
FFPE 23	108,50	2,171	1,003	2,16	0,31
FFPE 24	152,90	3,058	1,488	2,06	0,77
FFPE 25	187,10	3,743	1,839	2,03	0,52
FFPE 26	189,10	3,783	1,860	2,03	0,80
FFPE 27	293,10	5,861	2,873	2,04	1,04
FFPE 28	246,90	4,938	2,387	2,07	0,94
FFPE 29	224,20	4,484	2,135	2,10	0,82
FFPE 30	252,20	5,045	2,378	2,12	0,63
FFPE 31	221,70	4,434	2,053	2,16	1,05
FFPE 32	197,40	3,948	1,874	2,11	0,99
FFPE 33	253,30	5,065	2,409	2,10	0,74
FFPE 34	281,42	5,628	2,747	2,05	1,06
FFPE 35	244,77	4,895	2,449	2,00	1,08
FFPE 36	126,10	2,522	1,201	2,10	0,84
FFPE 37	181,18	3,624	1,750	2,07	0,92
FFPE 38	258,03	5,161	2,527	2,04	0,82
FFPE 39	139,69	2,794	1,417	1,97	0,34
FFPE 40	190,34	3,807	1,911	1,99	0,82
EtOH percipitated					
CRC 88	181,65	3,633	1,841	1,97	1,77
CRC 89	280,96	5,619	2,796	2,01	2,00
CRC 90	188,36	3,767	1,982	1,90	1,32
CRC 97	373,70	7,473	3,833	1,95	1,89
CRC 104	225,55	4,511	2,240	2,01	2,12
CRC 106	180,38	3,608	1,806	2,00	2,06
CRC 107	192,88	3,858	1,882	2,05	2,33
CRC 108	70,10	1,402	0,703	1,99	1,87
CRC 120	87,64	1,753	0,892	1,97	1,81
CRC 124	84,37	1,687	0,864	1,95	1,83
CRC 127	60,20	1,204	0,624	1,93	1,19
CRC 140	87,56	1,751	0,880	1,99	2,15
CRC 164	167,30	3,346	1,646	2,03	1,99
CRC 200	151,16	3,023	1,507	2,01	2,05
CRC 204	132,14	2,643	1,304	2,03	1,96
CRC 210	54,76	1,095	0,555	1,97	1,91
CRC 222	158,73	3,175	1,672	1,90	1,68
CRC 232	90,26	1,805	0,876	2,06	1,70
CRC 253	531,47	10,629	5,461	1,95	2,04
CRC 273	237,70	4,754	2,536	1,87	1,10
CRC 280	166,64	3,333	1,683	1,98	1,59
CRC 285	158,49	3,170	1,656	1,91	1,62
CRC 303	71,69	1,434	0,745	1,92	1,86
CRC 315	34,42	0,688	0,375	1,83	1,82

Supplementary Table G: Calculated scoring values for SNaPshot® mutations analyzed in GeneMapper.

Sample ID	MUT aph	WT aph	Noise ph	% MUT [a]	MUT aph>3x control [b]	cDNA	Codon	Sanger Sequencing	Low level mut [c]
4	682	1733	10	28,24	42,63	743_F	248	Verified	
7	427	450	10	48,69	17,08	524_F	175	Verified	
18	266	309	10	46,26	10,64	524_F	175	No PCR	
33	104	175	10	37,28	2,81	916_R	306	Verified	
35	217	245	10	46,97	31,00	742_F	248	Verified	
36	402	605	10	39,92	26,80	818_F	273	Verified	
108*	175	90	10	66,04	4,73	916_R	305	Verified	17,5
204*	202	1428	10	12,39	13,47	818_F	273	Verified	20,2

aph = allele peak height (ph); [a] % Mutant allele > 10 % (fluorescent peak height ratio of [mutant /(mutant+wt)] alleles >0.10); [b] peak fluorescent of mutant allele > 3x noise in human control; [c] % of mutant allele was > = 5% and peak fluorescent of the mutant allele was > 5 times above background: * Failed automatic analysis settings in GeneMapper

Supplementary Table H: Activity of mutant p53 proteins. Comments generated from UMD-p53 Mutation database, Mut_Mat_II (MUT-TP53 2.0.).

ID	Codon	WT codon	MUT codon	WT AA	WT AA	MUT AA	MUT AA	p.Mutant (3)	p.Mutant (1)	c.Mutant	Frequency	Activity	Comment 1	Comment 2	Comment 3
88	157	GTC	TTC	Val	V	Phe	F	p.Val157Phe	p.V157F	c.469G>T	205	9,06	**This is a hot spot mutant**	This mutant is inactive	No Problem
97	244	GGC	GTT	Gly	G	Val	V	p.Gly244Val	p.G244V	c.[731G>T ;732C>T]	25	0,00	**This mutant is frequent**	This mutant is inactive	Unusual mutational event except in skin cancer: check the analysis
104	143	GTG	GCG	Val	V	Ala	A	p.Val143Ala	p.V143A	c.428T>C	23	11,35	**This mutant is frequent**	This mutant is inactive	No Problem
108	306	CGA	TGA	Arg	R	X	X	p.Arg306X	p.R306X	c.916C>T	179	ND	**This mutant is frequent**	The activity of truncated p53 is assumed to be nil	No Problem - This codon is close to the end of an exon : splicing can be altered
127	286	GAA	AAA	Glu	E	Lys	K	p.Glu286Lys	p.E286K	c.856G>A	92	11,07	**This mutant is frequent**	This mutant is inactive	No Problem
176	266	GGA	GAA	Gly	G	Glu	E	p.Gly266Glu	p.G266E	c.797G>A	82	0,00	**This mutant is frequent**	This mutant is inactive	No Problem
204	213	CGA	TGA	Arg	R	X	X	p.Arg213X	p.R213X	c.637C>T	363	ND	**This is a hot spot mutant**	The activity of truncated p53 is assumed to be nil	No Problem
204	273	CGT	CAT	Arg	R	His	H	p.Arg273His	p.R273H	c.818G>A	888	1,01	**This is a hot spot mutant**	This mutant is inactive	No Problem
207	279	GGG	GAG	Gly	G	Glu	E	p.Gly279Glu	p.G279E	c.836G>A	50	0,27	**This mutant is frequent**	This mutant is inactive	No Problem
210	176	TGC	TAC	Cys	C	Tyr	Y	p.Cys176Tyr	p.C176Y	c.527G>A	106	14,82	**This mutant is frequent**	This mutant is inactive	No Problem
211	213	CGA	TGA	Arg	R	X	X	p.Arg213X	p.R213X	c.637C>T	363	ND	**This is a hot spot mutant**	The activity of truncated p53 is assumed to be nil	No Problem
240	233	CAC	CGC	His	H	Arg	R	p.His233Arg	p.H233R	c.698A>G	2	53,24	This mutant is very rare	This mutant does not display a significant loss of activity	CHECK CAREFULLY BEFORE PUBLICATION
280	175	CGC	CAC	Arg	R	His	H	p.Arg175His	p.R175H	c.524G>A	1314	12,41	**This is a hot spot mutant**	This mutant is inactive	No Problem
303	238	TGT	TTT	Cys	C	Phe	F	p.Cys238Phe	p.C238F	c.713G>T	46	0,82	**This mutant is frequent**	This mutant is inactive	No Problem
345	237	ATG	ATA	Met	M	Ile	I	p.Met237Ile	p.M237I	c.711G>A	204	0,43	**This is a hot spot mutant**	This mutant is inactive	No Problem
4	248	CGG	CAG	Arg	R	Gln	Q	p.Arg248Gln	p.R248Q	c.743G>A	1016	0,00	**This is a hot spot mutant**	This mutant is inactive	No Problem
7	175	CGC	CAC	Arg	R	His	H	p.Arg175His	p.R175H	c.524G>A	1314	12,41	**This is a hot spot mutant**	This mutant is inactive	No Problem
18	175	CGC	CAC	Arg	R	His	H	p.Arg175His	p.R175H	c.524G>A	1314	12,41	**This is a hot spot mutant**	This mutant is inactive	No Problem
33	306	CGA	TGA	Arg	R	X	X	p.Arg306X	p.R306X	c.916C>T	179	ND	**This mutant is frequent**	The activity of truncated p53 is assumed to be nil	No Problem - This codon is close to the end of an exon : splicing can be altered
35	248	CGG	TGG	Arg	R	Trp	W	p.Arg248Trp	p.R248W	c.742C>T	803	0,00	**This is a hot spot mutant**	This mutant is inactive	No Problem
36	273	CGT	CAT	Arg	R	His	H	p.Arg273His	p.R273H	c.818G>A	888	1,01	**This is a hot spot mutant**	This mutant is inactive	No Problem
10	282	CGG	TGG	Arg	R	Trp	W	p.Arg282Trp	p.R282W	c.844C>T	683	0,55	**This is a hot spot mutant**	This mutant is inactive	No Problem
26	282	CGG	TGG	Arg	R	Trp	W	p.Arg282Trp	p.R282W	c.844C>T	683	0,55	**This is a hot spot mutant**	This mutant is inactive	No Problem

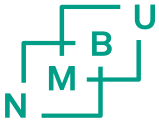
Supplementary Table I: Distribution of vital status for Dukes' stage B and C patients (n = 68) with metastasis (w/ Met) and without metastasis (wo/ Met).

Dukes' stage	NA		Alive				Dead							
							Colon cancer				Other			
			w/ Met		wo/Met		w/ Met		wo/Met		w/ Met		wo/ Met	
	wt	mut	wt	mut	wt	mut	wt	mut	wt	mut	wt	mut	wt	mut
B	3	2	1	0	19	10	0	0	0	0	0	0	3	1
C	1	3	1	0	8	5	12	5	0	0	0	0	3	0

NA; Not Applicable

Supplementary Table J: Correlation between *TP53* mutation status and clinicopathological data.

	N	%	WT		MUT		*r*-value	*p*-value
TOT	68		**N**	**%**	**N**	**%**		
Age								
<50	5	7.4	4	5.9	1	1.5	0,0664	0,5234
>50	63	92.6	43	63.2	20	29.4		
Gender								
Female	42	61.8	30	44.1	12	17.6	0,0636	0,5225
Male	26	38.2	17	25	9	13.2		
Tumor stage								
B	34	50	23	33.8	11	16.2	0,0318	0,5113
C	34	50	24	35.3	10	14.7		
Metastasis								
Yes	19	28	14	20.6	5	7.4	-0,0615	0,4783
No	49	72.1	33	48.5	16	23.5		



Norwegian University
of Life Sciences

Postboks 5003
NO-1432 Ås, Norway
+47 67 23 00 00
www.nmbu.no