Pharmacogenomic Profiling and Clarification of the Role of the Mismatch Repair Genes in Response to the Chemotherapeutic Agent 5-Fluorouracil in a South African Colorectal Cancer Cohort

Jacqueline Meyer (B.Sc [Med] Hons)
MYRJAC003

Thesis presented for the degree of
Master of Science in Medicine (M.Sc. Med)
In the Faculty of Health Sciences,
and the Division Of Human Genetics
The University of Cape Town
July 2009

PLAGIARISM DECLARATION

1. I know that plagiarism is wrong. Plagiarism is using another's work and to pretend that it is ones own.

2. I have used the Journal Of Human Genetics as the convention for citation and referencing. Each significant contribution to, and quotation in, this thesis **'Pharmacogenomic Profiling and Clarification of the Role of the Mismatch Repair Genes in Response to the Chemotherapeutic Agent 5-Fluorouracil in a South African Colorectal Cancer Cohort'**, from the work, or works of other people has been attributed to them and has been cited and referenced.

3. This thesis **'Pharmacogenomic Profiling and Clarification of the Role of the Mismatch Repair Genes in Response to the Chemotherapeutic Agent 5-Fluorouracil in a South African Colorectal Cancer Cohort'** is my own work.

4. I have not allowed, and will not allow, anyone to copy my work with the intention of passing it off as his or her own work.

5. I acknowledge that copying someone else's assignment or essay, or part of it, is wrong, and declare that this is my own work.

Jacqueline Meyer
MYRJAC003

SIGNATURE: Signed by candidate

DATE: 12/11/2009

TABLE OF CONTENTS

ABSTRACT

Hereditary non-polyposis colorectal cancer (HNPCC) describes a unique inherited clinico-pathological entity, caused by a mutation in any one of a set of genes that are known to monitor the fidelity and facilitate repair of DNA, the mismatch repair (MMR) genes.

To date, surgery is the mainstay treatment for HNPCC. Adjuvant chemotherapy and radiotherapy are often used to reduce systemic and locoregional recurrence, respectively, after curative surgical resection. The main chemotherapeutic agent is 5-Fluorouracil (5-FU). Studies have attempted to elucidate whether the MMR status of a colorectal cancer (CRC) cohort will define a response to 5-FU therapy. However, no difference in long term response or survival after 5-FU treatment between patients with MMR-proficient and MMR-deficient tumours were detected.

Individual patients can vary in their response to 5-FU. A genetic susceptibility to adverse drug effects is purported to exist. 5-FU is mainly catabolised in the liver by the enzyme dihydropyrimidine dehydrogenase (DPYD) and 5-FU metabolites interacts with the enzyme thymidylate synthase (TYMS). Thus polymorphisms in the genes (*DPYD, TYMS*) encoding these enzymes might influence the efficacy of 5-FU.

The current pilot study set out to utilise a range of informative sources and molecular techniques, in order to identify the opportunities and challenges in setting up any comprehensive cohort study and to develop a set of pharmacogenetic assays pertinent to the use of 5-Fluorouracil (5-FU) in the treatment of colorectal cancers. Additionally this study evaluated in a pilot cohort, whether CRC patients with a defective MMR system respond differently to the chemotherapeutic agent, 5-FU, than their MMR efficient counterparts.

A research cohort of 231 individuals affected with CRC was chosen and their

treatment, treatment outcomes and 5-year survival evaluated. In addition, five polymorphisms in *TYMS* and *DPYD*, were genotyped, in an attempt to determine the frequencies of the variants in patients and controls of Caucasian, Black and Mixed Ancestry in South Africa.

Kaplan-Meier analysis determined that local or distant recurrence of cancer was not a function of chemotherapeutic intervention with 5-FU in either MMR proficient or deficient individuals. The current study also contributes knowledge about the frequency of five variants within two important pharmacogenes, *DPYD* and *TYMS* in the larger CRC affected SA population. The ultimate analysis of the pharmacogenes should be regarded as a pilot study, setting a template for future larger scale studies.

This study is also a critical survey of the referral system at Grootte Schuur Hospital and The University of Cape Town, in order to identify means of streamlining data and sample capture for the future. Furthermore, this research contributes substantially to form part of a future network of population-based registries until such time that the government passes legislation to make cancer a notifiable disease.

AKNOWLEDGEMENTS

Firstly, I would like to thank my supervisor, Professor Raj Ramesar, for giving me the opportunity to make use of the facilities at The Division of Human Genetics, and for his valuable and insightful assistance on this thesis.

Dr Barbera Robertson at the Groote Schuur Hospital Oncology ward, for her help in arranging access to medical records, and her assistance in the writing of this dissertation.

My dutiful Honours student, Maureen Akinyi, for her help in the practical side of the experiments.

To Sarah Owens, my grateful thanks for being so willing to read my thesis on such short notice! You lead by example.

To Alvera Vorster, without whom this thesis would not be possible! Thank you for teaching me with such patience, and for always making time for my issues, big or small!

To my colleagues at HumGen, thanks for the lunches, the advice, and for keeping me sane amidst chaos! You ladies always know how to make me laugh when I am feeling down!

To the friends, and family that endured my workaholic silences and mood swings, thanks for hanging in there, and for being willing to stay my friends!

To the love of my life, Darren, words cannot explain how much I appreciate and love you for the Saturdays you sacrificed to spend it in the lab with me! Thanks for calming me down when things went awry, for spoiling me unexpectedly, and being my ever-ready shoulder to cry on! I cherish you!

Aan my ouers, Lizel and Christoffel, baie dankie dat julle dit vir my moontlik gemaak het om hierdie skripsie te voltooi, al moes dit langer vat as wat almal beplan het! Dankie dat julle elke tree saam met my gestap het! Ek is baie lief vir julle!

Lastly, I want to give thanks to and praise my Lord Jesus Christ, for His unwavering, unconditional, merciful grace and guidance, and for holding on when I felt like letting go.

LIST OF TABLES AND FIGURES

TABLES

FIGURES

ABBREVIATIONS

5-FU	5-Fluorouracil
B-CLL	B-cell Chronic Lymphocytic Leukemias
BKA	Black Ancestry
BKAF	Black African
BV	Bevacizumab
CAU	Caucasian
CIS	Cisplatin
CRC	Colorectal Cancer
CTX	Cetuximab
dATP	Deoxy Adenine Triphosphate
ddNTP	Dideoxynucleotides
DHFR	Dihydrofolate Reductase
DNA	Deoxyribonucleic Acid
dNTP	Deoxynucleotide
DPD	Dihydropyrimidine Dehydrogenase
DPYD	Dihydropyrimidine Dehydrogenase
dTMP	Deoxythymidine Monophosphate
dTTP	Deoxythymidine Triphosphate
dUDP	Deoxyuridine Diphosphate
dUMP	Deoxyuridine Monophosphate
dUTP	Deoxyuridine Triphosphate
E.g.	For Example
EGFR	Epidermal Growth Factor Receptor
EtBr	Ethidium Bromide
F-ddNTP	Fluoro-deoxynucleotide
FAP	Familial Adenomatous Polyposis
FDA	US Food and Drug Administration
FdUMP	Fluorodeoxyuridine Monophosphate
FdUTP	Fluorodeoxyuridine Triphosphate
FOLFIRI	5-FU/Leuocovorin, Irinotecan

FOLFOX	5-FU/LV, Oxaliplatin
FUDR	Fluorodeoxyuridine
FUTP	Fluorouridine Triphosphate
GIT	Gastrointestinal Tract
GSH	Groote Schuur Hospital
HCT116	*MLH1*-deficient human colorectal carcinoma cell line
HER	Herero
HIV	Human Immunodeficiency Virus
HNPCC	Hereditary Non-polyposis Colorectal Cancer
HPLC	High Performance Liquid Chromotography
ICG-HNPCC	International Collaborative Group on HNPCC
IFL	5-FU/LV, Calcium Folate and Irinotecan
IHC	Immuno-histochemical
IR	Irinotecan
KHS	Khoisan
LOH	Loss of Heterozygosity
lrSNP	Local Representative SNPs
LV	Leucovorin/ 5'-formyltetrahydrofolate
MA	Mixed Ancestry
mCRC	Metastatic Colorectal Cancer
Min	Minute/s
miRNA	MicroRNA
MRC	Medical Research Council
MMR	Mismatch Repair
MrTHF	5-methyltetrahydrofolate
MSI	Microsatellite Instability
MSI-H	MSI-high
MSI-L	MSI-low
MSS	MS-stable
MTHFR	Methylenetetrahydrofolate Reductase
MTX	Methotrexate
NCR	National Cancer Registry

NDOH	National Department of Health
NICE	UK National Institute for Health and Clinical Excellence
NTC	No Template Control
OX	Oxaliplatin
PCR	Polymerase Chain Reaction
RFU	Relative Fluorescent Units
RT	Radiotherapy
S	Second/s
SA	South Africa
SAP	Shrimp Alkaline Phosphatase
siRNA	Small Interfering RNA
SNP	Single Nucleotide Polymorphism
STS	Sotho/Tswana
Ta	Annealing Temperature
TEAA	Triethylammonium Acetate Ions
tSNP	Tagging SNPs
TYMS	Thymidylate Synthase
UDG	Uracil-DNA Glycosylase
UTR	Untranslated Region
VNTR	Variable Number of Tandem Repeats
XHS	Xhosa
ZUL	Zulu

PLAN OF THESIS

This thesis represents research carried out towards an M.Sc. (Med) degree, with the specific aim of i) understanding the heterogeneity of colorectal cancer; ii) identifying the patterns of referral within the academic and service centers; iii) understanding the treatment of colorectal cancer iv) assessing treatment outcomes and v) providing an overall critique of the process.

The dissertation is divided into several chapters and the layout is as follows:

Chapter one provides some background information about colorectal cancer (CRC), more specifically about the inherited form, hereditary non-polyposis colorectal cancer or HNPCC. Additionally chemotherapeutic interventions for CRC is discussed as well as the utility and implication of pharmacogenomic studies of the agent, 5-Fluorouracil.

Chapter two is the experimental aspect of the study. It comprises the various methods utilised in ascertaining the study cohort, and the methods whereby specific variants in drug metabolising genes for 5-FU were analysed in a pilot cohort. Furthermore, this chapter contains the statistical analyses performed after the various genotypes for the variants were ascertained for the study cohort.

Chapter three provides the results from the analyses described in chapter two.

The thesis is concluded in **Chapter four**, which provides comments and critique on each of the abovementioned aspects that the study aimed to accomplish, followed by concluding remarks in **Chapter five**.

CHAPTER 1
INTRODUCTION

1.1 Epidemiology of Cancer

Cancer is a life threatening disease which affects nearly 6.7 million people world-wide each year. Almost 10 million people are diagnosed every year, and it is estimated that, globally, there are 24.6 million people living with cancer at present (Parkin *et al.* 2005). Lung cancer has the highest incidence (1.35 million) and is the leading cause of cancer-related deaths (1.18 million) (Albrecht 2006). Colorectal, prostate, breast and bladder cancer are predominant in developed countries, whereas developing countries such as South Africa (SA) present with more cervical, stomach, liver, and esophageal cancers (Albrecht 2006).

In 1998-9, the five most prevalent cancers in males in SA were reported to be those of the prostate, lung, oesophagus, colorectum and bladder. In females the most prevalent cancers were those of the breast, cervix, colorectum, oesophagus and lung (Mqoqi *et al.* 2004).

Parkin *et al.* (2005) have suggested that cancer cases are severely under-reported in SA. Nearly 60,000 new cancer cases were reported to the National Cancer Registry (NCR) in 1999 (Mqoqi *et al.* 2004), however, mortality data from the Burden of Disease Research Unit of the Medical Research Council (MRC), states that nearly 66,000 deaths were caused by cancer in 2000 (Bradshaw *et al.* 2004). Thus the published mortality rates are higher than the reported incidence (Mqoqi *et al.* 2004; Bradshaw *et al.* 2004). It is estimated that the annual incidence could realistically be closer to 114 000 cases per annum and that cancer surveillance by the NCR only covers approximately 52% of the cancer population (Albrecht 2006). Moreover, the National Department of Health (NDOH), which governs the Disease Notification System under the auspices of the Directorate: Epidemiology and Surveillance, does not consider cancer a notifiable disease

in South Africa (http://www.doh.gov.za/docs/dns-f.html), because of the overwhelming burden of communicable diseases. Hence cancer is under-emphasised, with 33 medical conditions, such as cholera, malaria and tuberculosis preferentially demanding priority with the NDOH. Patient data is intermittently reported to the NCR on a goodwill basis only, and the quality of the data is variable (Parkin *et al.* 2008). However, cooperation has been sporadic due to patient confidentiality issues. The registry has not been updated since 1999 and contains only histological diagnoses as well as demographic data, and no geographic information (Albrecht 2006).

It is evident that, currently, cancer can only be monitored by mortality data failing accurate incidence reporting as indicated by the disharmony in the mortality and incidence rates. Unfortunately, cause of death certification and registration in SA is of poor quality, and has been cited by the World Health Organisation (WHO) to be incomplete, ill-defined and under reported (Mathers *et al.* 2007). Despite recent improvements, and overcoming a backlog of cause of death statistics production, the accuracy of death registration is still poor. A high proportion (92%) of death notification forms from the greater Cape Town area contain errors (Burger *et al.* 2007). Burger *et al.* (2007) have recommended that clinicians should primarily be better trained in death certification in order to alleviate the certification error rate.

1.2 Colorectal Cancer (CRC)

Colorectal cancer (CRC) accounts for approximately 9.4% of the total number of cancer cases diagnosed globally (Parkin *et al.* 2005). The occurrence of CRC in developing countries as in Africa, is low in comparison with other developed countries such as the United States of America (USA), Australia, Europe and Japan. Comparative mortality data in the USA and SA revealed that the incidence of CRC is 200-300% lower in SA than in the USA (Albrecht 2006). This geographical difference is probably largely due to different environmental exposures such as diet and lifestyle, although the effect of under reporting of cancer in SA should not be excluded. Of the approximately 60,000 cancer cases that are reported each year in SA, nearly

1200 are CRC cases (Mqoqi *et al.* 2004) which translates to about 11.3 males per 100,000 and 8.3 females per 100,000 being diagnosed with this form of cancer (Parkin *et al.* 2005). CRC has, despite a high incidence worldwide, relatively low mortality rates in both males and females in SA (Albrecht 2006).

1.2.1 Inherited CRC

A large body of evidence shows that some individuals have an increased predisposition to developing CRC. This is most clearly seen in the familial forms of the disease, which collectively account for approximately 10-15% of all CRC; namely familial adenomatous polyposis (FAP) and hereditary non-polyposis colorectal cancer (HNPCC), or Lynch syndrome (Vasen 2005). A short introduction to FAP will follow, but the focus of this review and the subsequent study, will be on HNPCC.

A) Familial adenomatous polyposis (FAP)

Clinically, patients with FAP present with thousands of adenomatous polyps in the colon, and although mostly benign, they have malignant potential. Extracolonic manifestations of FAP include brain tumours and retinal lesions (Arnold *et al.* 2005). The molecular pathogenesis is attributed to causative mutations in the *APC* gene, which is located on the long arm of chromosome 5 (Herrera *et al.* 1986; Kinzler *et al.* 1991).

B) Hereditary non-polyposis colorectal cancer (HNPCC)

HNPCC describes a unique inherited clinico-pathological entity, caused by a mutation in any one of a set of genes that are known to facilitate the repair of DNA (Lynch *et al.* 1993). Lynch and colleagues described this hereditary disease which is also known as Lynch syndrome, although it was first reported by Aldred Warthin in 1913 (Lynch *et al.* 1993). HNPCC is an autosomal dominant disease accounting for nearly 4% of all CRC (Barnetson *et al.* 2006). Clinically, Lynch syndrome is classified into two major groups; Lynch syndrome types I and II, based on the absence or presence of extra-colonic malignancies, respectively. Lynch syndrome I is site-specific,

characterised by tumours mostly in the proximal colon, whereas Lynch II manifests with extra-colonic cancers such as those of the ovaries, endometrium, small bowel, pancreas and brain. More recently, the existence of a third group, Lynch Syndrome III, was proposed by Felton *et al.* (2007). This group comprises patients presenting with a different spectrum of early onset cancers, as a result of homozygous or compound heterozygous mutations in one of the mismatch repair genes. These patients present with brain or haematological malignancies early in life. Biallelic, germline mutations in these genes also result in greatly impaired MMR function, necessitating a distinct classification based on molecular pathology and site of cancer (Felton *et al.* 2007).

i) Clinical diagnosis of HNPCC

HNPCC is a heterogeneous disease, and as such, no single defining feature exists on which diagnosis can be based. Instead, the clinical diagnosis of HNPCC is facilitated by considering family history, histopathology and age at onset, as set out in the Amsterdam and Bethesda criteria.

In 1991, the International Collaborative Group on HNPCC (ICG-HNPCC) developed the Amsterdam Criteria in order to accurately define HNPCC for comparative studies. At the time, the consensus was to diagnose HNPCC based on the following minimum criteria: i) at least three relatives should have a histologically verified CRC, one of whom is a first degree relative of the other two; (ii) at least two consecutive generations should be affected; and (iii) in one of the relatives, CRC should be diagnosed before the age of 50 years (Vasen *et al.* 1991). Therefore, HNPCC is broadly defined as "colorectal cancer in multiple family members from multiple generations, with cancers arising, on average, in the fifth decade" (Lynch and Smyrk 1998). A major weakness of the Amsterdam Criteria was the exclusion of extracolonic malignancies in the diagnosis, since the incidence of cancers of the endometrium, small bowel, stomach, kidney and ovary is increased in HNPCC (Lynch and Smyrk 1998). For this reason the Amsterdam II Criteria included all malignancies associated with HNPCC. Some patients with CRC

do not meet the Amsterdam criteria for HNPCC and hence could not be distinguished from sporadic CRC. This necessitated another set of criteria, the Bethesda guidelines, to be developed which describe all clinical conditions, diagnosed before age 45, for which screening of tumour tissue DNA for the presence of microsatellite instability (MSI) or immuno-histochemical staining (IHC) is indicated, in order to identify potential HNPCC patients (Rodriguez-Bigas *et al.* 1997). The revised Bethesda criteria was later instituted to include all patients who were less than 60 years of age and diagnosed with cancer (Umar *et al.* 2004).

ii) Molecular pathogenesis of HNPCC

In each cell, complex enzymatic systems exist which monitor DNA damage and replication errors. The mismatch repair (MMR) system is comprised of several proteins that work together to monitor the fidelity of DNA and act as tumour suppressor genes (Hemminki *et al.* 1994). Carcinogenesis may result from compromising any component of this critical enzymatic system, as is reflected in the molecular pathogenesis of HNPCC. HNPCC is caused by the mutational inactivation of both alleles of a MMR gene. The one allele is usually inactivated by an inherited/germ-line mutation, whilst the other may be rendered non-functional by an acquired somatic mutation. The acquired mutation is also referred to as the "second hit" since the result is a complete loss of function for the MMR gene in question (Nystrom-Lähti *et al.* 1994; Peltomaki 1997).

The postreplicative correction of mismatches and insertion/deletion loops has been attributed to five pivotal MMR genes viz. *MLH1, PMS2, MSH2, MSH3* and *MSH6* (Hemminki *et al.* 1994; Jiricny and Nystrom-Lähti 2000). Functional studies in *Escherichia coli* (*E. coli*) and *Saccharomyces cerevisiae* have contributed to our understanding of the intricate workings of the human MMR system. This has been fully reviewed in Jacob and Praz (2002) and will only briefly be discussed. The *E. coli* MMR system is mediated by the complex interaction of three proteins: MutS, MutL and MutH, encoded by genes *mutS*, *mutL* and *mutH*. MutS efficiently recognises mismatches and insertion/deletion loops. Once MutS binds to a mismatch, it interacts with

MutL, facilitated by an ATP-directed conformational change and enables repair. The five human homologs of MutS, with partly redundant functions, are MSH2, MSH3, MSH4, MSH5 and MSH6. Only MSH2, MSH3, and MSH6 collaborate to repair nuclear DNA. MSH4 and MSH5 are responsible for other important cellular functions (Bocker *et al.* 1999; Edelmann *et al.* 1999). During repair, MSH2 forms heteroduplexes with MSH3 and MSH6 to form MutSβ and MutSα, respectively (Acharya *et al.* 1996; Alani 1996). In human cells, MutSα is present in much higher levels than its twin, MutSβ, and as such, the majority of mismatch recognition is performed by MutSα. MutL has four human homologs, namely MLH1, MLH3, PMS1 and PMS2. MutSα, or alternatively, MutSβ can interact with MutLα (heterodimer of MLH1 and PMS1) or MutLβ (heterodimer of MLH1 and MLH3), depending on the type of mutation (Räschle *et al.* 1999). A ternary complex is thus formed, consisting of the mismatch, MutSα/β and MutLα/β which allows for strand discrimination and removal of the erroneous base (Prolla *et al.* 1994).

Currently, information on mutations and polymorphisms in the MMR genes are maintained in several databases including The Memorial University of Newfoundland Mismatch Repair Genes Variant Database (*http://www.med.mun.ca/MMRvariants*), and INSiGHT (*www.nfdht.nl*), curated and maintained by the ICG-HNPCC. More than 448 mutations in the MMR genes have been implicated in HNPCC and, of these, nearly 50% occur in *MLH1*, 39% in *MSH2* and 7% in *MSH6* (Peltomaki 1997; Jiricny and Nystrom-Lähti 2000; Marcos *et al.* 2006; *http://www.nfdht.nl*).

iii) Molecular diagnosis of HNPCC

Patients who meet the Amsterdam or Bethesda criteria, and are suspected of having HNPCC, are subjected to several molecular analyses in order to verify the diagnosis. Generally, screening involves testing the DNA from tumour tissue for MSI. Additionally, IHC staining for the MMR proteins of colorectal cancer specimens from suspected HNPCC patients can be performed as an alternative method of screening. Loss of MMR protein expression may reflect a pathogenic mutation in the MMR genes (Arnold *et al.* 2004). If there are noticeably lower levels of any of the MMR proteins in tumour tissue

compared to normal tissue, targeted mutational analysis of that specific gene can be performed to identify the pathogenic mutation.

Tumorigenesis in a colonocyte can be driven by a particular pattern of DNA damage, resulting in normal cells being transformed to tumour cells. Most tumours from patients with HNPCC show genomic instability or MSI, and this has been attributed to inactivation of the MMR system that is responsible for maintaining fidelity during DNA replication (Boland *et al.* 1998). If the caretaker function of the MMR system malfunctions, errors that occur during DNA replication are not repaired, and mutations are allowed to accumulate. Simple repetitive sequences, called microsatellites, that are omnipresent in the genome, are used to gauge the extent of mismatch repair deficiency, giving rise to the phenotype termed MSI (Bhattacharyya *et al.* 1994; Boland *et al.* 1998). MSI studies in human colorectal adenocarcinoma cell lines have provided evidence of a 100-fold increase in mutation rates (Bhattacharyya *et al.* 1994; Branch *et al.* 1995). Mutations within key genes, resulting in loss of protein function, are thought to contribute to tumorigenesis. It should be noted that a proportion (15%) of non-hereditary, or sporadic, colorectal tumours also show MSI (Boland *et al.* 1998). The absence of MMR proteins in such samples after IHC may be due to hypermethylation of the MMR promoter (Valle et al. 2007). When an allele is hypermethylated, it may either repel transcription factors or it may attract proteins that inhibit transcription, resulting in down-regulated expression (Bartolomei and Tilghman 1997). Once the cytosine residue of a CG dinucleotide couple — contained within CpG islands in promoter regions — is methylated, the gene is transcriptionally silenced (Niv 2007).

MSI at a particular marker is defined as a change of any length, due to either the expansion/contraction of repeating units in a microsatellite within tumour DNA, when compared to normal tissue (Boland *et al.* 1998). The recommended panel of markers is composed of two mononucleotide repeats *(BAT26* and *BAT25)* and three dinucleotide repeats *(D5S346, D2S123* and *D17S250)* that should be investigated in order to report on the mutator phenotype (Boland *et al.* 1998). Tumours showing instability in the majority

of the markers (two or more of the five microsatellite sequences) are designated MSI-high (MSI-H), whereas those showing only one of the 5 markers to be mutated are designated MSI-low (MSI-L). A third specific tumour phenotype was described as MS-stable (MSS), where there was no evidence of MSI (Boland *et al.* 1998).

1.3 Cancer Management

Patient care in South Africa comprises an interaction between private and state-funded facilities. Patients who have healthcare insurance, usually through employment, are treated in the private sector, whilst the majority of individuals (85.7%) are provided for in the public sector (General Household Survey 2007). Both sectors are expected to provide high-quality care, however the latter are often under-resourced due to lack of appropriate state funding. Patients diagnosed with any cancer, including CRC are first assessed for treatment, whether this is curative or aimed at improving quality of life (Abratt and Vorobriof 2003). In South Africa, the stage of cancer, according to the Dukes' staging system, determines whether chemotherapy is appropriate. For instance, patients newly diagnosed with Dukes stage C CRC (an invasive tumour with some nodal metastases) at Groote Schuur Hospital (GSH), Cape Town, usually undergo surgery to remove the bulk of the tumour growth. Thereafter, patients are generally treated with adjuvant (defined as "after surgery") chemotherapy to prevent systemic recurrence. With Dukes stage D CRC, distant metastatic disease is present, and depending on their fitness, patients receive chemotherapy and are closely monitored to gauge whether the disease will remain stable or progress. Unlike private sector patients, public sector hospitals are generally unable to switch to chemotherapeutic agents that are effective in the refractory setting in the event of recurrence or disease progression, and most patients must resort to symptom-directed care (Goldberg *et al.* 2007; Dr B Robertson, GSH Oncologist, private communication).

1.3.1 Chemotherapy

For the last forty years there was only a single chemotherapeutic agent, 5-Fluorouracil (5-FU), in the armamentarium for treating colorectal cancer. Today, however, the strategies and drugs are relatively plentiful. In order to enhance the cytotoxicity of 5-FU, biochemical modulation by other agents are applied, resulting in a continuum of management possibilities. Some regimens either alter the metabolism of 5-FU, or result in the retention of active metabolites in the cell. Drugs in this category include Leucovorin and Methotrexate. Another category focuses on agents which will produce synergistic effects, resulting in greater cytotoxicity. Oxaliplatin falls into this class. Lastly, strategies that alter the pharmacokinetics of 5-FU (e.g Capecitabine) are also utilised (Grem 2000). However, not all strategies have proven to be successful. It is still unclear which patients should receive which drug to attain the best possible outcome. Some of the most popular 5-FU drug additives, mediating cytotoxicity, shall be discussed.

A) CRC and chemotherapy

Unlike for cervical cancer, which is frequently caused by the human papilloma virus, no vaccine exists against CRC (Markowitz *et al.* 2007). The US Food and Drug Administration (FDA) (*http:/www.fda.gov/*) has approved the use of Capecitabine, 5-FU, Irinotecan, Levamisole, Oxaliplatin and Leucovorin to treat CRC in the USA. Currently, despite a plethora of chemotherapeutic agents being available, the mainstay regimen for CRC and by default, HNPCC, in state-funded hospitals in Cape Town, involves 5-FU. In SA, the popularity of 5-FU as a chemotherapeutic agent for CRC, is mainly because it is relatively inexpensive and thus freely available in the public sector (Meyers *et al.* 2001). Nevertheless, apart from the response rate to 5-FU as a monotherapy being only 10-20%, the quality of life in patients on 5-FU is reported to be severely compromised, with many debilitating side effects (Meyers *et al.* 2001).

B) Modulators of 5-FU to enhance clinical benefit

i) Leucovorin

The non-responsiveness of tumours to 5-FU as a monotherapy has forced oncologists and researchers to assess the use of combination and sequential combination therapy (Venook 2005).

Leucovorin (LV, 5'-formyltetrahydrofolate) is a folinic acid-based biomodulator of 5-FU. LV contributes the co-factors needed for the interaction of the 5-FU metabolite, fluorodeoxyuridine monophosphate (FdUMP) with the enzyme thymidylate synthase (this is discussed in depth in section 1.4.1). A meta-analysis of several clinical trials (The Advanced Colorectal Cancer Meta-Analysis Project 1992) showed a substantial increase in response rates of patients on combination therapy. Patients benefitted significantly ($p<0.001$) from combination therapy (23%) in terms of tumour response rate, compared to single-agent 5-FU treatment (11%). Since then, adjuvant therapy with 5-FU has seldom been administered without LV modulation. A subsequent meta-analysis in 2004, which incorporated an extended follow-up period, only served to affirm the positive interaction between the two drugs.

ii) Methotrexate

Methotrexate (MTX) and 5-FU act synergistically to inhibit thymidylate synthase (section 1.4.1) although MTX does this in an indirect fashion. MTX inhibits one of the key enzymes in folate metabolism; dihydrofolate reductase (DHFR) (Goulian *et al.* 1980a). Several clinical trials were included in a meta-analysis which has provided evidence that 5-FU modulation by MTX nearly doubles the response rate of patients with metastatic CRC (mCRC) (The Advanced Colorectal Cancer Meta-Analysis Project 1994).

iii) Irinotecan

Irinotecan (IR) *(Camptosar®; Pfizer Pharmaceuticals Inc., New York, NY, http://www.pfizer.com*) is a topoisomerase I inhibitor. Topoisomerase I is a nuclear enzyme which is crucial for the unwinding of DNA during replication and is therefore needed for cell division (Saltz *et al.* 2000). Douillard *et al.* (2000) reported a 49% increase in response rates of mCRC patients treated

with infusional 5-FU/LV, calcium folate and IR (FOLFIRI). The overall survival for patients on FOLFIRI was 17.4 months compared to 14.1 months for patients without IR, after the start of treatment. A median overall survival of more than 20 months for patients on FOLFIRI has since been reported (Kohne *et al.* 2004; Ji *et al.* 2005).

Another study assessed the clinical benefits of IR with bolus 5-FU/LV (IFL) for previously untreated patients with mCRC. The IFL regimen proved to be more effective than the 5-FU/LV dual combination, based on its ability to reduce tumour size and delay disease progression. The response rates for patients on IFL were 50% compared to 28% for the 5-FU/LV combination and the overall survival was 14.8 months compared to 12.6 months (Saltz *et al.* 2000).

iv) Oxaliplatin

Oxaliplatin (OX) *(Eloxatin®; Sanofi-Synthélabo Inc., New York, NY, http://www.sanofisynthelabo.us)* is a platinum derivative, currently part of the mainstay treatment regimen for mCRC in the USA. Patients who receive OX in combination with 5-FU and LV (FOLFOX) show great synergistic drug activity, having longer progression-free survival (20 months versus 14 months) than control populations who receive only 5-FU/LV (de Gramont *et al.* 2000) or IR/OX alone (Ashley *et al.* 2007; Sanoff *et al.* 2008). Additionally, more than 9% of patients have a 5-year survival on this regimen (Sanoff *et al.* 2008). Drug-related side effects are decreased (albeit dose dependant) when OX is included in the 5-FU/LV regimen with the exception of reversible neuropathy (Hospers *et al.* 2006). A combination of OX, Folinic acid, 5-FU and IR, however, has failed to show any significant survival benefit compared to Folinic acid, 5-FU and IR alone (Souglakos *et al.* 2006).

v) Capecitabine

Capecitabine *(Xeloda®; Hoffmann-La Roche Inc., Nutley, NJ, http://www.roche.us)* is an oral derivative/pre-prodrug of 5-FU with a more favourable toxicity profile than 5-FU. It escapes degradation and hence is more active in tumour tissue (Schellens 2007). It is reportedly more tolerable,

convenient and at least as effective as 5-FU/LV (Cassidy *et al.* 2002). It has been shown that Capecitabine has a more favourable outcome, and could potentially replace 5-FU as combination therapy with OX (Cassidy *et al.* 2004).

C) Targeted agents as supplement to chemotherapy for CRC

The success of initial chemotherapeutic agents has been hampered by a lack of specificity and dose-dependent adverse effects. Advances in discovering the molecular pathogenesis of cancer, has paved the way for developing target-specific agents. Unlike chemotherapy, which assaults all dividing cells, both healthy and cancerous, some targeted therapies act at key points in the development of tumours, and hence focus only on cancerous tissue making them less harmful to normal, healthy cells. This strategy has the potential to also reduce the side effects associated with therapy. The success of targeted therapies, however, is hindered by the costs involved in treatment.

In 2006, the UK National Institute for Health and Clinical Excellence (NICE), evaluated bevacizumab (*Avastin®; Genentech Inc., South San Francisco, CA, http://www.gene.com*), and cetuximab (*Erbitux®; ImClone Systems Inc., New York, NY, http://www.imclone.com*) as targeted therapeutic agents for CRC (Barnett *et al.* 2006).

Bevacizumab (BV) is a humanised immunoglobulin G1 monoclonal antibody that inhibits vascular endothelial growth factor, which is a mediator of tumour angiogenesis (Presta *et al.* 1997). Angiogenesis is defined as the formation of tumor blood vessels, allowing tumor cells to access the systemic circulation and to metastasise. Tumours are unable to progress without adequate blood supply which provides nutrients and oxygen. Thus, the rationale behind using BV is that if angiogenesis is prevented, the tumour will not progress further (Venook 2005).

Several clinical trials were performed to assess the efficacy of BV used in conjunction with chemotherapy. In a trial by Kabbinavar *et al.* (2003), BV was added to 5-FU/LV therapy, in previously untreated CRC patients. By

comparing it to standard chemotherapy (5-FU/LV) alone, the authors showed that addition of a targeted agent is successful in increasing response rates and prolonging survival (21.5 versus 13.8 months). Several subsequent studies confirmed the latter, whereby the addition of BV prolonged overall median survival (16.6 versus 12.9 months), progression-free survival, and increased the response rate by 11% (Emmanouilides *et al.* 2004; Kabbinavar *et al.* 2005). In another trial, Hurwitz *et al.* (2004) showed that the BV/IFL regimen (section 1.3.1A*iii*) significantly prolonged overall survival, progression-free survival, and response rates in therapy-naive patients. Patients whose regimens were supplemented with BV showed an average of 20.3 months overall survival compared to 15.6 months for those individuals on a placebo.

It has been shown that expression of the epidermal growth factor receptor (EGFR) in colonic carcinomas is accompanied by a more aggressive disease and a poor prognosis (Prewett *et al.* 2002). Cetuximab (CTX) is a chimeric IgG1 monoclonal antibody that has a high affinity for EGFR, and by binding to it, prevents its natural ligand from binding and inducing phosphorylation of EGFR. Cunningham *et al.* (2004) investigated the efficacy of CTX in combination with IR in patients with CRC who were unresponsive to monotherapy with IR. A significantly higher response rate (22.9%) in the combination therapy group was observed when compared to the monotherapy group (10.8%) as well as longer time to progression.

D) A defective Mismatch Repair system influences response to chemotherapy

It has been shown that HNPCC patients who present with Dukes stage B or C CRC, show better survival rates in general, compared to their sporadic CRC counterparts [87.5% versus 44.8%] (Elsakov and Kurtinaitis 2006). Thus a survival benefit is indicated for HNPCC patients in the absence of any chemotherapy (Ribic *et al.* 2003); in this regard one might assume that treatment with 5-FU is not necessarily beneficial to everyone taking it (Longley *et al.* 2003; Allen and Johnston 2005).

Several studies attempted to elucidate whether the MMR or MSI status of a CRC cohort will define a response to 5-FU therapy (Ribic *et al.* 2003; Jover *et al.* 2006). Most studies distinguished a tumour as MMR-deficient when a loss of e.g. MLH1 protein expression or MSI occurred. However, no difference in long term response or survival after 5-FU treatment between patients with MMR-proficient and MMR-deficient tumours were detected (Ribic *et al.* 2003; Jover *et al.* 2006), and patients presenting with MSI-H tumours, did not benefit from 5-FU treatment when compared to individuals with MSI-L tumours (Warusavitarne and Schnitzler 2006). Conversely, in patients receiving no adjuvant treatment, those with MSI-H tumours had longer overall survival than stage-matched sporadic cancer patients with MSI-L tumours or MSS tumours (Warusavitarne and Schnitzler 2006).

Some studies have suggested that the MMR proteins mediate a response to other DNA-damaging agents. The trend amongst researchers has been to study hypersensitivity or alternatively, resistance, to cytotoxic agents in MMR-deficient and proficient cell lines. Work with the *MLH1*-deficient human colorectal carcinoma cell line (HCT116) has been well characterised. MMR function restoration is accomplished in these cells by transfection of one copy of chromosome 3 containing the *MLH1* gene into the cells. Thereafter cells usually manifest sensitivity to certain DNA damaging agents (Aebi *et al.* 1996; Aebi *et al.* 1997; Takahashi *et al.* 2005).

Differential sensitivity of MMR-deficient cells to OX and cisplatin (CIS) have been reported (Aebi *et al.* 1996; Sergent *et al.* 2002). Carboplatin, OX and CIS are platinum compounds that form adducts with DNA, thereby signaling apoptosis (Claij and Riele 1999). Mutations in *MLH1* and *MSH6* result in resistance only to CIS, but not to OX as evidenced by the preferential recognition of CIS adducts over OX adducts by the bacterial MMR protein, MutS (Vaisman *et al.* 1998; Zdraveski *et al.* 2002). Furthermore, tumours obtained from ovarian cancer patients, which displayed an MSI phenotype and *MLH1* deficiency, were reported to be resistant to CIS-based therapy (Watanabe *et al.* 2001). Thus it is expected that *in vivo*, MMR-deficient tumours will not respond to treatment with CIS, but rather with OX. However,

Sergent *et al.* (2002) published a contradictory report, citing spontaneous resistance to OX and CIS in colon cancer cell lines with a defective MMR system. It has also been shown by Anthoney *et al.* (1996) that MSI is not a causal factor in resistance to CIS, but rather a byproduct of acquired resistance after CIS treatment. Resistance to topoisomerase I inhibitors, Topotecan and IR have also been demonstrated in *MLH1*- and *MSH2*-deficient cell lines (Takahashi *et al.* 2005), but contradictory evidence shows that MSI-H cell lines and tumours are hypersensitive to IR (Bras-Gonçalves *et al.* 2000; Fallik *et al.* 2003). Thus, a variable response to several chemotherapeutic agents has been attributed to MMR status, and these inconsistent results provide scope for further research.

It is evident that hypermethylation of the MMR genes is also associated with a resistance to chemotherapy. Some studies have shown that loss of *MLH1* expression, and corresponding hypermethylation of promoter CpG islands accompanied the acquired resistance of ovarian cancer cell lines to CIS (Strathdee *et al.* 1999; Wei *et al.* 2003). A landmark paper by Arnold *et al.* (2003) promotes the use of de-methylating agents *in vitro* to treat MMR-deficient, *MLH1* hypermethylated cell lines. In so doing, the authors were able to restore *MLH1* expression and sensitise a 5-FU resistant cell line to treatment. This provides a similar effect as had previously been accomplished by chromosome 3-based transfections into *MLH1*-deficient cell lines. Gifford *et al.* (2004) showed that CpG methylation of DNA could be detected in blood plasma samples, and corresponded to the status in tumour cells. The proportion of samples in which the promoter of *MLH1* was methylated, increased substantially after treatment with chemotherapeutic agents such as Paclitaxel, Carboplatin and Docetaxel. Some samples that had relapsed after treatment also showed increased methylation patterns compared to pre-chemotherapy samples, and hence the authors speculated that methylation is in fact acquired after chemotherapy, and contributes to resistance (Gifford *et al.* 2004).

Several authors studied mechanisms of resistance to 5-FU *in vitro*. Zhu *et al.* (2005), disclosed that cells which were resistant to 5-FU over-expressed a

common anti-apoptotic protein, Bcl-L, although the parental lines did not. Expression of other members of the Bcl-2 family of proteins were also upregulated in 5-FU resistant cells compared to 5-FU sensitive cells. However, these proteins are pro-apoptotic, and the authors speculated that the ratio of the two types of proteins might affect apoptotic behaviour after 5-FU treatment.

E) MicroRNA and Cancer

MicroRNA (miRNA) are essential for temporal control of a range of post-embryonic developmental events and form part of an endogenous RNA interference (RNAi) pathway (Lee *et al.* 1993).

Studies indicate that modification of miRNA gene expression may play an intricate part in the development of cancer. Therefore it is also possible that chemotherapeutic agents could alter miRNA gene expression patterns, given that most anti-neoplastic agents interfere with nucleic acid metabolism and gene expression. Calin *et al.* (2002) were the first to associate miRNA with cancer, and thereafter miRNA expression profiles in lung (Yanaihara *et al.* 2006), breast (Ma *et al.* 2007) and colon cancer have been investigated. Calin *et al.* (2002) reported that several miRNA genes, referred to as miR-15a and miR-16a, are located in a chromosomal region that is often deleted in B-cell chronic lymphocytic leukemias (B-CLLs). The ensuing deletion analysis showed that miR-15a and miR-16a are located on chromosome 13q14, within 30-kb of a region that is lost in CLL. An accompanying expression analysis demonstrated that the majority of CLL cases show a down-regulation of miR-15a and miR-16a (68%). Further studies by Calin *et al.* (2004) included a systematic search to identify whether the genomic position of a large number of miRNAs correlate with the location of cancer associated genomic regions. The authors speculated that miRNA genes might be located in fragile chromosomal regions which are often susceptible to amplification, deletion, or translocation during the course of tumour development. Thus it is evident that normalisation of the miRNA expression pattern might result in restoration of the deregulated post-transcriptional control and therefore could have a therapeutic effect. Their search led them to tabulate a significant number of

miRNAs, including several located in regions implicated in colon cancer. Micheal *et al.* (2003) identified two miRNAs that exhibit reduced levels in precancerous and neoplastic colorectal tissue. The authors' approach included a Northen Blot analysis on enriched small fragments of RNA from both a colonic adenocarcinoma and normal mucosa. Most sequences show similar steady-state levels of mature miRNAs, as well as precursor miRNA in both normal and cancerous tissues. However, the authors confirmed that several sequences (miR-143 and miR-145) consistently exhibit significantly reduced steady-state levels of *mature* miRNA in neoplastic epithelium as well as adenomatous polyps, relative to normal epithelium. The unprocessed precursors of both miRNA were in equal abundance in both tissues which implies that the reduction is due to post-transcriptional processes. The authors found it imperative to also establish the putative targets of miR-143 and miR-145. Represented transcripts include components of signal transduction pathways, proteins involved in metabolic processes, as well as processing of RNAs. How the observed reduction of mature miR-143 and miR-145 levels is associated with the translation of these putative targets is still unknown. Later, Akao *et al.* (2006) showed that *let-7* is down-regulated in human colon cancer cell lines and tumours. Transfection with the precursor miRNA of *let-7, let-7a-1,* caused significant suppression of growth in the cancer cells and suggests that let-7 miRNA is involved in growth of colon cancer cells. In recent years, Lanza *et al.* (2007) did a genome-wide expression analysis of microRNA in MSS and MSI-H colon cancer samples. They were able to produce a molecular signature of differentially expressed miRNA genes that could distinguish between MSS and MSI-H tumours.

Rossi *et al.* (2007) hypothesised that treatment with 5-FU somehow alters miRNA gene expression and showed evidence to that effect. Not only were 19 miRNA genes upregulated in colon cancer cells after 5-FU treatment, but three were also shown to be down-regulated. Surprisingly, miRNA genes that are already induced to over-express in transformed neoplastic colorectal tissue, showed additional drug-induced up-regulation. On the other hand, some miRNA genes, commonly up-regulated in tumour tissue, were down-regulated after treatment.

Other authors discuss the potential of using siRNA to enhance the chemosensitivity of cells to a particular drug. According to Lei *et al.* (2007), it is postulated that over expression of several anti-apoptotic proteins in cancer cells lead to resistance to apoptosis. The authors established that by silencing these proteins using siRNA, cells became more sensitive to a chemotherapeutic agent, and apoptosis of cancer cells was increased. Lei *et al.* (2007) used siRNA mediated silencing, *in vitro*, to show that cells in which knockdown of anti-apoptotic genes, *Bcl-2* and *Bcl-xl* occur, displayed increased sensitivity to 5-FU, and hence underwent increased cell-death. Thus knockdown of these genes may act as a chemosensitising event.

1.4 Pharmacogenomics of 5-FU

As mentioned before, individual patients can vary in their response to 5-FU. It is thought that a genetic susceptibility to adverse drug effects may exist (Giacomini *et al.* 2007). However, before the pharmacogenomic possibilities of 5-FU can be discussed, it is imperative that the intercellular dynamics and kinetics of 5-FU is made explicit.

1.4.1 Mechanism of action of 5-FU

For 5-FU to exert its cytotoxic effects it must first be activated. After transport into the cell, 5-FU is sequentially anabolised to give several active compounds; fluorouridine triphosphate (FUTP), fluorodeoxyuridine monophosphate (FdUMP) and fluorodeoxyuridine triphosphate (FdUTP) [See Figure 1.4.1A] (Longley *et al.* 2003).

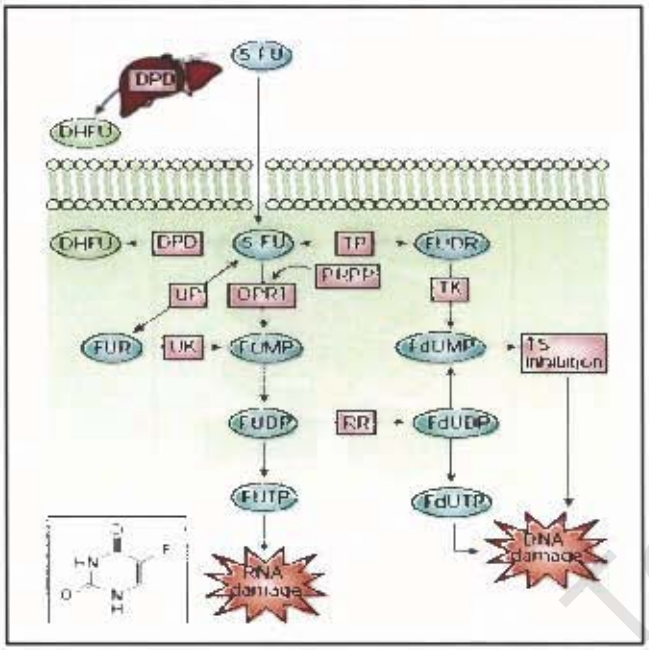


Figure 1.4.1A: Mechanism of action of 5-FU (Longley *et al.* 2003)

The 5-FU active compounds exert their cytotoxic function by RNA and DNA directed actions. Primarily, misincorporation of the antimetabolites into the RNA strand results in mRNA splicing alterations (Wilkinson and Pitot 1973; Wilkinson and Crumley 1977 and 1978; Kufe *et al.* 1981; Doong and Dolnicks 1988). The metabolite FdUMP also behaves as a quasi-substrate of the enzyme thymidylate synthase (TS) (Santi *et al.* 1974; Lönn and Lönn 1984; Sawyer *et al.* 1984; Schuetz *et al.* 1986). As previously mentioned, this interaction is modulated by LV, which contributes the co-factors needed for the interaction of FdUMP with TS. Thus, these two agents are co-inhibitors of the TS pathway in DNA synthesis. In the cell, LV is converted step-wise to 5' 10'-methylenetetrahydrofolate (CH_2-THF). The enzyme methylene-tetrahydrofolate reductase (MTHFR) then reduces CH_2-THF to 5-methyltetrahydrofolate (MrTHF). Under normal circumstances, this reaction would favour DNA synthesis by accelerating the TS pathway, since CH_2-THF provides methyl groups for the nonreversible, reductive methylation of dUMP. However, the presence of 5-FU and its interaction with and subsequent inhibition of TS, is strengthened by the formation of a ternary complex

involving CH_2-THF and MrTHF (Figure 1.4.1B) (Wisotzkey *et al.* 1999, Sohn *et al.* 2004).

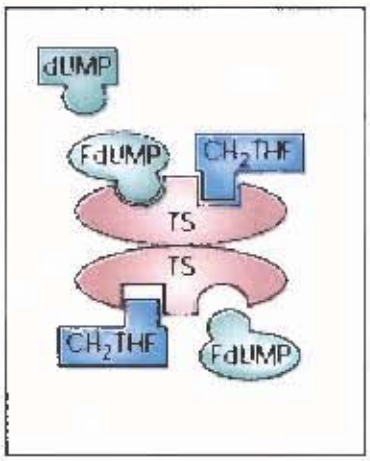


Figure 1.4.1B: Stable ternary complex between FdUMP, TS and CH2-THF (Longley *et al.* 2003)

TS is translated from the *TYMS* gene on chromosome 18p11.32 (NC_000018.8), producing a 313 amino acid protein which functions as a dimer. In unperturbed circumstances, TS serves to convert the DNA replication precursor, deoxyuridine monophosphate (dUMP) to deoxythymidine monophosphate (dTMP) by reductive methylation (Santi *et al.* 1974). Thymine nucleotides, which are crucial for DNA replication, are thereby provided. However, during chemotherapy, the conversion of 5-FU to fluorodeoxyuridine (FUDR) is catalysed by thymidine phosphorylase, and the protein is subsequently phosphorylated by thymidine kinase to FdUMP (Wilkinson and Crumley 1977; Schwartz *et al.* 1995). FdUMP chemically resembles the uracil moiety in dUMP, and as such is a competitive inhibitor of this natural substrate of TS (Figure 1.4.1C). FdUMP will preferentially bind to the active site of TS with a covalent bond, significantly decreasing dTMP synthesis, causing the cells to become depleted of their only source of *de novo* thymidine, and resulting in a deoxynucleotide (dNTP) pool imbalance (Santi *et al.* 1974). This thymine-deprived state has been associated with DNA fragmentation (Curtin *et al.* 1991). If the cells are unable to salvage thymidine, the subsequent dNTP pool imbalance (particularly the adenine to thymidine (dATP:dTTP) ratio (Houghton *et al.* 1995) will signal an endonuclease which would produce direct DNA double strand cuts, disrupting

DNA synthesis and ultimately resulting in cell death. It has also been hypothesised that the stability of duplex DNA and resulting DNA double-strand breaks are affected in a sequence specific manner (Yoshioka *et al.* 1987; Sahasrabudhe *et al.* 1995).

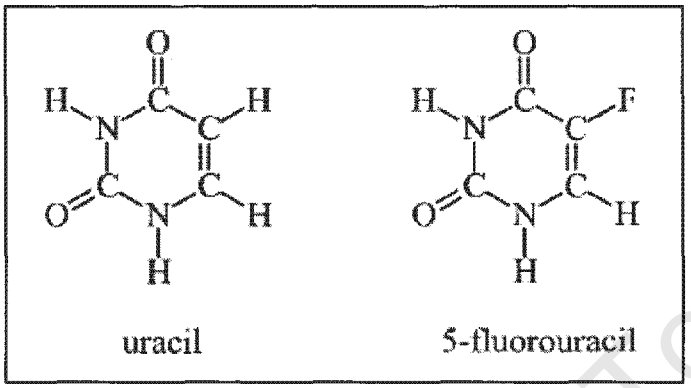


Figure 1.4.1C: The structure of a uracil nucleotide compared to its uracil analogue, 5-FU (Grem, 2000)

In a thymine-deprived state, because of substantially more dUMP in the cell, the levels of deoxyuridinediphosphate and deoxyuridinetriphosphate increase, and these ribonucleotides, as well as other unconventional bases (including FdUTP), are misincorporated into the growing DNA strand. Subsequent recognition and removal of the erroneous bases from the DNA by the base-excision repair enzyme, uracil-DNA glycosylase (UDG), would leave an apyrimidinic site. Once this site is targeted for repair, another futile round of dUTP reincorporation would take place in the absence of dTTP which would eventually also contribute to the abovementioned DNA strand breakage (Ingraham *et al.* 1982; Sawyer *et al.* 1984; Curtin *et al.* 1991; Houghton *et al.* 1995; Fisher *et al.* 2007). DNA fragmentation occurs after the addition of FdUDP, suggesting that the removal of the bases by UDG is followed by endonucleolytic cleavage, and subsequent DNA breaks (Cheng and Nakayama 1983; Shuetz *et al.* 1986). The presence of UDG supports the theory that the cytotoxicity of 5-FU could be as a result of the removal of mispaired bases, instead of the actual incorporation into DNA (Shuetz *et al.* 1986).

The drug Methotrexate (section 1.3.1Aii) also substantially contributes to the reduction of the levels of dTTP and presents an effective way to modulate 5-

FU therapy. It essentially mimics the 5-FU mechanism of action, as Methotrexate also promotes the erroneous incorporation of uracil into DNA (Goulian *et al.* 1980b).

1.4.2 Personalised 5-FU treatment

i) Clinical toxicity of 5-FU

The toxic effects of 5-FU treatment are extensive and varied. Understandably, and because replicating DNA and cell division are retarded, rapidly proliferating tissues, such as the bone marrow and gastrointestinal tract (GIT) are affected quite severely. Ulceration of the epithelial lining of the GIT causes symptomatic manifestations such as mucositis, gastritis, colitis, diarrhea, and nausea or vomiting. Blood-related side effects such as granulocytopenia, minor thrombocytopenia and anaemia are frequently described. Photosensitivity, alopecia (hair loss), and discoloration of veins are some dermatological toxicities that have been reported. Ocular toxicities include conjunctivitis, whilst hepatic toxicity is infrequently described. Neurological-related toxicities are sporadic, but characterised by cerebellar ataxia (*www.PharmGKB.org*; Grem 2000).

ii) Thymidylate synthase (TYMS)

As previously mentioned, the 5-FU metabolite, FdUMP, interacts with the enzyme TS (Santi *et al.* 1974) which is responsible for the conversion of dUMP to dTMP. Thus it is safe to assume that polymorphisms in the *TYMS* gene might influence the efficacy of this interaction. The best described polymorphism in *TYMS* was first discovered by Horie and colleagues (1995). A variable number of tandem repeats (VNTR) in the 5' untranslated region (UTR) is thought to predict the effectiveness of 5-FU therapy (Figure 1.4.2A). Two or three tandem copies of the repeat (designated 2R/2R or 3R/3R respectively), or the heterozygous combination genotype (2R/3R), were most commonly found in populations, until Marsh *et al.* (1999 and 2000) published studies on ethnic variation, proposing up to nine repeats in certain African and Asian populations. These repeats, situated in the *TYMS* promoter, act as an enhancer and it has been shown that an increased number of the 28bp repeat leads to an increase in *TYMS* expression and activity, with the 3R showing up

to a 2.6 fold increase in expression compared to the 2R (Kawakami *et al.* 1995; Kawakami *et al.* 1999). Later a G>C SNP was identified within the second 28bp repeat, located in the 12th nucleotide (Figure 1.4.2A). This common SNP abolishes the ability of the repeat to act as an enhancer. Thus the 3R allele will only show increased expression relative to the 2R allele of the VNTR if this particular SNP is present (Kawakami and Watanabe 2003; Mandola *et al.* 2003). It is evident from many studies that this variant could explain inter-individual differences in response to 5-FU. Pullerkat *et al.* (2001) showed that individuals who were homozygous for the double tandem repeat (2R/2R), had a better response to the chemotherapy than their 3R/3R counterparts.

A 6-bp deletion in *TYMS* mRNA has been observed at bp 1494 in the 3'UTR (NM_0001071, Figure 1.4.2), and shown to have a prevalence of 0.29 in the Caucasian population (Ulrich *et al.* 2000). Dotor *et al.* (2006) showed a correlation between this variant and well-differentiated tumours, and proposed that the deletion is protective insofar as no deaths were observed in a homozygous deletion (6-/6-) group. Although the function of this deletion remains unknown, it has been speculated that the 3'UTR is rich in microRNA genes, which are known to play a role in mRNA stability and translation.

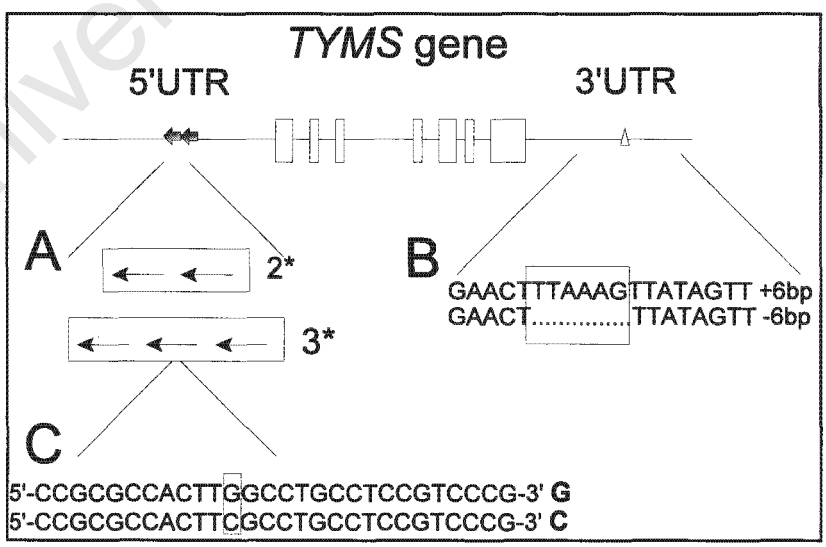


Figure 1.4.2: Most common variants in *TYMS* associated with enzyme activity. A: a 28-bp 5'UTR VNTR; B: G>C SNP in 2nd repeat of VNTR; C: 3'UTR 6bp deletion (Figure adapted from Marsh 2005).

iii) Dihydropyrimidine dehydrogenase (DPYD)

It is reported that 5-FU is mainly catabolised in the liver by the enzyme dihydropyrimidine dehydrogenase (DPD) (Naguib and el Kouni 1985). This enzyme determines the efficacy of the drug since the metabolism of 5-FU is limited by the availability of DPD (Tuchman *et al.* 1985). DPD deficiency shows a familial clustering, with autosomal recessive inheritance. A family history of pyrimidinemia and pyrimidinuria were often observed in patients with decreased haematological DPD activity, who go on to develop severe toxicities after 5-FU treatment (Diasio *et al.* 1988). The first causative mutation was discovered by Meinsma *et al.* (1995).

Since then, the precise mechanism of action of DPD has been elucidated. DPD initiates the catabolism of the pyrimidine bases, uracil and thymine. Because 5-FU is a pyrimidine analogue (Figure 1.4.1C), it is also catabolised via this pathway. The primary initial metabolite is dihydrofluorouracil/ fluoro-5,6-dihydrouracil, which is rapidly eliminated from the blood plasma (Heggie *et al.* 1987). It is the rate limiting step in the catabolism of the fluoropyrimidines. Of the administered dose, 60-90% of the catabolites are excreted via the urine, and 2-3% via the biliary system (Heggie *et al.* 1987).

Disease-causing mutations in the *DPYD* gene have contributed to a deficiency in DPD (Yokota *et al.* 1994). A patient presenting with a complete lack of DPD activity showed defective mRNA in fibroblasts (Meinsma *et al.* 1995). Later, it was observed that as a result of a G>A point mutation in intron 14 (IVS14 +1G>A), a 5'-splice donor recognition sequence is abolished, and the resulting *DPYD* mRNA subsequently has a 165 base pair deletion leading to a truncated, non-functional protein (Wei *et al.* 1996). This mutation is perceived to be the most prevalent and occurs at a frequency of 0.94% in the Caucasian population. The identification of this mutation has helped many researchers explain 5-FU toxicity, since nearly 25% of all patients presenting with 5-FU-related toxicities, have this particular mutation (Raida *et al.* 2001). There are reports of patients, homozygous for the IVS14 +1G>A mutation, dying after 5-FU administration due to a complete deficiency of DPD activity and subsequent lethal toxicity (Van Kuilenburg *et al.* 2001). Likewise

patients with low DPD activity as a result of a heterozygous splice mutation, have reduced capacity to clear 5-FU (Maring *et al.* 2002).

Despite its prevalence, the IVS14 +1G>A mutation is not singularly responsible for DPD deficiency. Functional expression studies in *E. coli* revealed novel missense mutations such as R29C, a C>T missense mutation in exon 2 and 1627A>G (I543V) in exon 13, which have been shown to lead to enzymatically compromised DPD (Vreken *et al.* 1998). In some instances, no sequence variants could be identified to which a deficiency in DPD could be attributed, which led researchers to believe that epigenetic promoter hypermethylation could be the cause of downregulated *DPYD* expression. For instance, Ezzeldin *et al.* (2005) studied the promoter of *DPYD* using a novel adaptation of denaturing high performance liquid chromatography and found evidence of an association between altered methylation and DPD enzyme deficiency in 80% of patients studied.

1.4.3 Prognostic pharmacogenetic tests

With the advent of pharmacogenomics, there is a potential to revolutionise health care, by providing personalised medicine. Many pharmaceutical companies are focussed on harnessing an individual's genetic profile, and tailoring a drug treatment regimen according to specific genetic variants to predict the outcome of treatment. This is accomplished by studies of how the body deals with an administered drug ("pharmacokinetics"), involving its absorption, distribution, metabolism and excretion from the body, and how a drug exerts its influence upon the body ("pharmacodynamics").

A genetic test can have several straightforward purposes; predictive (predicts the response to treatment), diagnostic and, more recently, prognostic purposes (predicts the outcome of treatment). Prognostic tests provide information which will aid in selecting suitable therapeutics and dosages.

A literature search in the NCBI database, Pubmed, *http://www.ncbi.nlm.nih.gov/pubmed/*, using the phrase "cancer genetic test" reveals a plethora of citations describing various genes and/or variants

associated with cancer. However, it is evident that these associated genes or variants will not be established as validated commercial clinical tests in the near future. Some of the major commercial diagnostic laboratories in the United States (US), such as The Laboratory Corporation of America® (LabCorp®, Burlington, NC) and Quest Diagnostics (Teterboro, NJ) have detailed listings of cancer genetic tests and these tests are available in commercial laboratories in the USA and academic institutions. One such test assays for the CD52 marker, an antigen expressed on the surface of malignant cells, which determines the eligibility of chronic lymphocytic leukemia patients to receive Campath (alemtuzumab), an anti-CD52 drug. Another predictive test identifies breast cancer patients who are eligible to receive Herceptin. Herceptin is a drug which actively targets the HER-2/neu oncogene which is over expressed in HER2-positive breast cancers (HER-2 is human epidermal growth factor receptor family 2) (Chin *et al.* 2006). Additionally, the Food and Drug Administration (FDA, USA) has cleared the MammaPrint® assay, the first commercialised prognostic test for estrogen receptor-positive or -negative breast cancer patients, which predicts whether cancer may metastasise (Ross *et al.* 2008).

Currently, a few prognostic genetic tests for CRC, predicting treatment outcome, are available for use in a clinical setting, and mainly in developed countries like the USA. Tests include a chromosome 18q assay which is based on the premise that patients with deletions of chromosome 18 are more likely to have recurrence of their cancer and hence a shorter disease-free survival time (Kirley *et al.* 2005). A *CRC Pharmacogenomic Panel* was developed by Quest Diagnostics for the management of CRC (*www.questdiagnostics.com*). Unapproved by the FDA, this in-house laboratory test aims to assay variants in seven genes, which serves to evaluate the toxicity response of patients to treatment with 5-FU, Oxaliplatin (OX) and Irinotecan (IR). The genes included in the assay are *UGT1A1* (for IR), *TYMS* and *DPYD* (for 5-FU), *XRCC1, ERCC1, GSTP1* and *XPD* (for OX). The *DPYD* IVS14 +1G>A mutation, *TYMS* 5'UTR and 3'UTR variants, as well as a polymorphic TA promoter repeat in *UGT1A1* is examined. Furthermore, one pharmacorelevant SNP in each of the OX-metabolising genes are

analysed (Park *et al.* 2001; Stoehlmacher *et al.* 2002; Viguier *et al.* 2005). Additionally, this company developed simple mutational analyses of *DPYD* or *UGT1A1* only, which can predict toxicity from the pyrimidine-based agents, 5-FU and Capecitabine and IR respectively. Their repertoire extends also to a vascular endothelial growth factor (VEGF) test to assay response to Bevicazumab (section 1.3.1B) and a test which determines the eligibility of CRC patients to receive Cetixumab, an epidermal growth factor receptor (EGFR) inhibitor (section 1.3.1B).

Validation of prognostic genetic tests should be required as an antecedent before pharmacogenetic testing can be performed. Laboratories can commercialise genetic tests via two main regulatory bodies in the USA; The FDA and the Clinical Laboratory Improvement Amendments of 1988 (CLIA). However, many genetic tests, albeit diagnostic or prognostic, are being sold directly to the consumer (DTC), without first being substantiated by a health-care provider to ensure that the test results are not misinterpreted (Katsanis *et al.* 2008). It is not required that laboratories demonstrate clinical validity before new genetic tests are deployed, only that the laboratory and test be certified under the CLIA. Also, the FDA can only approve manufactured test kits from reputable companies and is unable to manage in-house laboratory based testing (Katsanis *et al.* 2008). Recently, the American Medical Association has adopted stricter policies on the regulation of DTC genetic tests (2008, News Release).

It is not clear whether a government body exists in SA which monitors genetic testing activities, but it is imperative that stricter quality control is exercised, globally and nationally, to determine the applicability and limitations to these tests. This would ensure that clinical validation occurs, and that the information pertaining to a test is accurately conveyed to the public and health-care providers. For example, the US laboratory, Quest Diagnostics (*www.questdiagnostics.com*), maintains a strict policy that tests can only be ordered through a reputable physician, and test results may only be released to the attending physician, and not to the patients themselves. On the other hand, it is possible to order genetic tests via the internet in SA, with the

results following suit. There should be a mechanism to guarantee that genetic tests are properly endorsed by the appropriate authorities before being made publically available.

The question remains whether it is unethical to withhold prognostic testing when available, or to provide such tests but without the appropriate control.

1.5 Aims and Objectives

The aim of this pilot study was to utilise a range of informative sources and molecular techniques, in order to identify the opportunities and challenges in setting up any comprehensive cohort study and to develop a set of pharmacogenetic assays pertinent to the use of 5-Fluorouracil (5-FU) in the treatment of colorectal cancers.

In order to accomplish the aim of the study, the following objectives were set:

a) To identify mechanisms for accessing patient records and biomaterial.

b) To interrogate and further improve the current patient database in the Division of Human Genetics, University of Cape Town (UCT). This would serve to establish the first of a network of population-based registries, until such time that cancer becomes a notifiable disease. This will facilitate the flow of information to the National Cancer Registry in South Africa, alleviating, at least to some extent, the disharmony which currently exists in cancer reporting.

c) To perform a critical survey of the referral system at Groote Schuur Hospital and UCT, in order to identify means of streamlining data and sample capture for the future.

d) To identify a subject cohort, affected with colorectal cancer from the patient registry in the Division of Human Genetics. The cohort will be differentiated into individuals who were shown to carry a mismatch repair mutation (Mut+) and those that were shown not to (Mut-), and to assess specific disease-

related features within and between the groups.

e) To understand the allele frequencies, in patient and control cohorts, of potential pharmacovariants in two genes, *TYMS* and *DPYD*, known to be responsible for interacting with and metabolising 5-FU, which is used in the treatment of colorectal cancer in South Africa.

CHAPTER 2

MATERIALS AND METHODS

2.1 COHORT SELECTION

2.1.1 Selection of pilot cohort

The Division of Human Genetics at UCT has performed genetic studies on individuals affected with or predisposed to Lynch syndrome type cancers, including colorectal cancer, since the late 1980s. Information about these individuals is contained in a patient registry located in the Division. Individuals affected with cancer are intermittently enlisted primarily from the Oncology Department at Groote Schuur Hospital (GSH), a neighbouring medical facility, where they receive follow-up. Predisposed individuals comprise unaffected family members who are subsequently recruited and found to have disease-predisposing mutations. Blood samples are given a unique identity number for easy access and to protect the patient's identity. The unique identity number is allocated chronologically according to the order in which blood was taken and the families identified. The identity number starts with a three letter code describing the specific disease (e.g. **NPC** for **N**on-**p**olyposis **C**olorectal Cancer), followed by a numerical assignment for each member of a family (e.g. **30.1**—family 30, member 1, usually the proband), and finally the first three letters of the individual's first name (e.g. **ANN** for Anne).

For the current study, all screening was performed on archived DNA samples previously obtained for abovementioned genetic studies. The patient registry in the Division contains information on 450 affected individuals, which includes probands and their relatives. Of these 450 individuals, 240 have been screened comprehensively for mutations in the mismatch repair (MMR) genes, *MLH1*, *MSH2* and *MSH6*. The remaining 210 individuals are affected with colorectal cancer, but have yet to be screened for mutations in the MMR genes. These 210 unscreened individuals were not included in the current study. The study cohort initially comprised the 240 screened patients. The

database contains the relevant clinico-pathological information and a thorough family history of only 144 patients, as well as some follow-up information. Additionally, the Division of Human Genetics maintains a concurrent database of background control individuals from different ethnic backgrounds. Hence 100 control subjects, 50 from Mixed Ancestry and 50 from Caucasian Ancestry, were admitted to the study. Informed written consent was obtained from patients pertaining to the current study (see Addendum B for a copy of the form).

By liaising with the Oncology Department at GSH, the Radiotherapy (RT) folders or follow-up records of 80 individuals were obtained with consent from patients and the referring physician. A concerted effort was made to populate or append the existing registry with phenotypic information *inter alia* pathology, staging of cancers, age at diagnosis, treatment outcomes, survival, and prognosis for these 80 individuals, in order to initiate a local population-based cancer registry. A positive response to therapy was defined as a 5-year progression-free survival after curative surgery and, if applicable, the successful completion of treatment. Recurrence before five years was delineated as resistance to therapy (if any) and subsequent relapse. Side-effects from therapy were noted. Patients or family members were periodically contacted in order to obtain updated information relevant to the study.

2.1.2 DNA isolation from peripheral blood lymphocytes and formalin fixed paraffin-embedded tumour tissue sections

Genomic DNA was previously isolated from peripheral blood lymphocytes using the PUREGENE™ DNA Isolation kit (Gentra Systems). DNA can become degraded over time, thus the concentration and purity of each DNA sample was investigated using the NanoDrop® ND-1000 (NanoDrop Technologies). Quality control was performed by electrophoresis on a 1.5% agarose gel (Addendum A) for 25 minutes at 160V. General-purpose sizing was carried out using a GeneRuler™ 100bp DNA Ladder Plus molecular weight marker (0.05µg/µl) (Fermentas Life Sciences, Addendum A). A loading buffer containing 40% sucrose and 0.25µl bromophenol blue (Merck)

was used. The DNA was visualised by the nucleic acid stain, ethidium bromide (EtBr; 5µg/ml, SIGMA®). The DNA was viewed using a transilluminator (UVITEC).

The High Pure PCR Template Preparation Kit (Roche, Mannheim, Germany) was used to isolate DNA from formalin-fixed paraffin-embedded tissue sections. Three tissue sections (5-10µm thick) were used to attempt to amass enough DNA for microsatellite instability testing. The manufacturer's protocol was adhered to, except that volumes were downscaled two-fold in order to compensate for the small amounts of tissue present. DNA quality control was performed as described above.

2.2 IDENTIFICATION OF MMR-DEFICIENT AND MMR-PROFICIENT PATIENTS

2.2.1 Determination of MSI status of affected patients using the Bethesda panel of microsatellite markers

In order to make a clear and detailed comparison between patients with and without an intact MMR system, microsatellite instability (MSI) testing was attempted on 35 DNA samples of patients who had received chemotherapy. The Bethesda panel of recommended microsatellite markers was used to assay for MSI (section 1.2.1iii). The panel comprises two mononucleotide markers; BAT25 and BAT26, and three dinucleotide markers; D2S123, D17S250 and D5S346. Primer sequences were obtained from Loukola *et al.* (2001). Information pertaining to the markers and their associated primers, including the repeat motif of the marker, the primer sequence, fluorescent label on the primer, and the size of the amplicon is described in Table 2.2.1a (Addendum B). To analyse MSI, a polymerase chain reaction (PCR) experiment was performed on DNA isolated from peripheral blood lymphocytes and from formalin fixed paraffin embedded tumour tissue, with marker-specific fluorophore-labelled primers, and the products genotyped. The PCR reaction was performed in a final volume of 10µl, and contained 1.5-2.0mM magnesium chloride ($MgCl_2$; Merck), 200µM deoxynucleotides (dNTPs; Bioline), 0.4µM of the forward and reverse primer, 0.1units/µl Taq

DNA polymerase (Promega) with 1x of the appropriate Go Taq buffer and 0.1ng/µl DNA, made up to 10µl with sterile distilled water (sdH_2O). Optimisation of primer specificity necessitated an increase in $MgCl_2$ concentration (1.5mM to 2mM) for marker D17S250. Amplification was performed on the Thermo Electron Corporation PX2 Thermal Cycler. The reaction mix is identical for all five primer pairs and is depicted in Table 2.2.1b (Addendum B), and the correct primer pair is replaced for the various microsatellite markers. The cycling conditions are depicted in Table 2.2.1c (Addendum B), and consisted of an initial denaturation step for 5 minutes (min) at 95°C, followed by 35 cycles of amplification consisting of: 94°C for 30 seconds (s), followed by 30s at 50°C, and 72°C for 40s. This is followed by a final elongation step of 7min at 72°C. The cycling conditions were identical for each primer pair, with the exception that the melting temperature was adjusted to 53°C for marker D2S123 to increase the yield. PCR products were electrophoresed according to protocol (section 2.1.2) on a 1.5% agarose gel for 30min at 160V to determine if amplification was successful.

In order to determine whether the DNA isolated from tumour tissue displays microsatellite instability compared to blood, PCR products from DNA from both sources were separated on the basis of size on the ABI Prism® 3100 DNA Genetic analyser (Applied Biosystems, Foster City, CA) and the raw data collected by the ABI Prism® 3100 Data Collection Software. The PCR products were diluted by a factor of ten (1/10) and 1µl of the diluted DNA was allotted to its respective well in a Thermo-Fast® 96 Detection Plate. A cocktail of Hi-Di™ Formamide (8.5µl/sample; Applied Biosystems) and the GeneScan size standard 500 Rox™ (0.4µl/sample; Applied Biosystems) was added to the designated wells on the plate. The PCR products were denatured at 95°C for 5min before being placed on ice and loaded onto the autosampler of the ABI 3100. The results were analysed with GeneMapper™ V3.0 software (Applied Biosystems).

2.2.2 Whole genome amplification of purified genomic DNA

The majority (80%) of the DNA samples extracted from formalin-fixed paraffin-embedded tissue sections were degraded and the entire panel of markers

failed to amplify despite extensive optimisation. In order to maximise the starting template, whole genome amplification was attempted on two samples using the REPLI-g® Mini Kit (Quigen, Hilden, Germany) according to the manufacturer's protocol, and PCR amplification repeated on the amplified products as described in section 2.2.1.

2.3 DEVELOPMENT OF PHARMACOGENOMIC ASSAYS AND SUBJECT PROFILING

2.3.1 Determination of pharmacorelevant genes and variants

A list (n=7) of genes involved in 5-FU metabolism was compiled from an extensive review of the literature (PubMed; *http://www.ncbi.nlm.nih.gov/pubmed*) and from the PharmGKB database (*www.pharmgkb.org*, Klein *et al.* 2001) using the keywords ***5-Fluorouracil, 5-FU toxicity, 5-FU metabolism, 5-FU genes***. Two genes were selected which had been shown to influence 5-FU toxicity and the 5-FU responder phenotype; dihydropyrimidine dehydrogenase *(DPYD)* and thymidylate synthase *(TYMS)* respectively. Information on variants in both *TYMS* and *DPYD* that could possibly influence the manifestation of toxic side effects and the response to chemotherapy were extracted. Table 2.3.1 contains the relevant information for each of the chosen variants; a VNTR in the 5' untranslated region (UTR) (rs45445694) and an insertion/deletion (indel) in the 3'UTR of *TYMS* (rs16430), and three single nucleotide polymorphisms (SNP) in *DPYD* (rs3918290, rs1801265, rs1801159). A variant had to comply with one or more preset parameters, including a high population heterozygosity as published on NCBI (*http://www.ncbi.nlm.nih.gov*) and/or functional studies relating to loss of protein function.

2.3.2 Gene annotation

The gene information and sequence of the human *TYMS* (GeneID: 7298; reference sequence: NC_000018.8) and *DPYD* (GeneID: 1806; reference sequence: NC_000001.9) genes were obtained from the NCBI website. Likewise, the sequences for *TYMS* and *DPYD* were procured from the

Table 2.3.1: Information regarding the inclusion criteria pertaining to studied variants in *DPYD* and *TYMS*

Variant	Sequence/Nucleotide change	Amino acid change	Location	NCBI Heterozygosity	Reference
DPYD					
rs3918290 (SNP)	NM_000110.3: c.1905+1 G>A (IVS14+ G>A)	Not detected (ND)	Intron 14	ND	Meinsma *et al.* (1995); Wei *et al.* (1996)
rs1801159 (SNP)	NM_000110.3: c.1627A>G	I543V	Exon 13	0.326	Vreken *et al.* (1998)
rs1801265 (SNP)	NM_000110.3: c.85T>C	R29C	Exon 2	0.316	Vreken *et al.* (1998)
TYMS					
rs45445694 (VNTR)	NM_001071.2: c.43ins28	ND	5'UTR	0.503	Horie *et al.* (1995)
rs16430 (deletion)	NM_001071.2: c*447-*452del6	ND	3'UTR	0.49	Ulrich *et al.* (2000); Dotor *et al.* (2006)

Ensembl website (*http://www.ensembl.org*). Approximately 5kb of sequence upstream and downstream of each gene was included. The gene annotations were based on the version of the genome available on 1 January 2008. These sequence files were utilised as input files in a Command prompt script to annotate the gene with gene information, marker features and variation features. PerlV5 was used as the programming language with the scripts, Annotv9 and Annotv9Ev2 courtesy of Dr. G. Rebello (personal communication). The output files showed the position of the start and stop codon, 5' and 3'UTR, SNP variations (as it appears in dbSNP, with accompanying *rs* numbers) and branch sites. Exonic sequences were in upper case and intronic sequences were in lower case. These annotations facilitated primer design (see Addenum B).

2.3.3 Design of primers to assay relevant variants

A primer pair, forward and reverse, was designed for each of the variants in *DPYD* and *TYMS*. Most primers were designed using a selection of bioinformatic tools. Firstly, primers were selected using Primer3 (Rozen *et al.* 2000), which is freely available on-line at *http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi*, and manually modified to optimise specificity. Several parameters including a Human Mispriming Library; primer size of 18-27 bases; primer melting temperature of 50-60°C; primer GC content of 40-60% and a maximum of three consecutive repetitive base pairs were specified. The resulting Primer3 primers were subjected to a hairpin analysis, self-dimer analysis and a hetero-dimer analysis, using Oligo Analyser courtesy of Integrated DNA Technologies *(www.idtdna.com/analyzer/Applications/OligoAnalyzer/Default.aspx)*. Primer sequences susceptible to these secondary structures were avoided due to the effect they may have on the PCR efficacy. The primers were analysed using BLAST *(www.ncbi.nlm.nih.gov/BLAST/)* to ascertain whether non-specific amplification for the primer pair could be likely elsewhere in the genome (Altschul *et al.* 1990). If Primer3 failed to give satisfactory primers, sequences were chosen manually and analysed with the same strict parameters. After satisfactory primers were designed, the sequences were ordered from the UCT core facility, where they were synthesised using the

Oligo 1000M DNA Synthesiser (Beckman Instruments Inc). Table 2.3.3 (Addendum B) contains the sequence, length, melting temperature and GC content of the final primers. The primer sequences in relation to the variants are depicted in Addendum B.

2.3.4 Design and optimisation of a multiplex PCR assay

The amplification of the variants in *TYMS*, the 5'UTR VNTR (rs45445694) and the 3'UTR indel (rs16430), will be discussed separately in sections 2.3.6 and 2.3.7 respectively. The three SNPs in *DPYD* (rs3918290, rs1801265 and rs1801159) were analysed in a cohort of 179 individuals, derived from the larger cohort of 240 individuals, after several lymphocytic DNA samples failed the quality checks.

In order to facilitate the analysis of the three SNPs in *DPYD* (rs3918290, rs1801265 and rs1801159), a multiplex experiment was designed. Before amplification of the cohort DNA could occur, the efficiency of the interaction between the six primers (three primer pairs, one for each variant) was optimised using various methods. Several parameters were altered in order to achieve optimal PCR amplification. Primarily, a temperature gradient PCR, with primer annealing temperatures increasing in 12 increments was performed for each of the separate primer sets individually, using control DNA, on a Multigene Thermocycler (Labnet International Inc., WhiteSci). The starting temperature was discretionary for each set. Thereafter, the primer pairs were pooled in various combinations. To evaluate the success of the amplification and establish the optimum annealing temperature of the primers, the 12 amplified control DNA samples were electrophoresed according to protocol (section 2.1.2) on a 2% agarose gel at 120V for 60 min. Failure to obtain distinct and separate bands on the agarose gel signified a subsequent optimisation step, which entailed modifying the buffer composition and pH, using a PCR Optimisation Kit (Roche Diagnostics). This method comprised 16 buffers (100mM Tris HCl, 500mM KCl) with varying concentrations of $MgCl_2$ and at variable pH. The final optimised PCR setup and cycling conditions are depicted in Table 2.3.4a and 2.3.4b (Addendum B). Briefly, the PCR was performed in a final 25µl reaction mixture, containing 100ng

template DNA, 0.4µM of each primer, 0.5U/µl of Taq polymerase (Promega), 1x buffer, 1.5mM $MgCl_2$ (Merck) and 200µM dNTPs (Bioline). Each PCR was performed on a Multigene Thermocycler (Labnet Inc.) and comprised an initial 5min denaturation step at 95°C, followed by 30 cycles of 94°C (1min), 56°C (1min) and 72°C (80s), with a final 7min extension at 72°C.

2.3.5 Multiplex genotyping of SNPs (rs3918290, rs1801265, rs1801159) in *DPYD* using the SNaPShot™ reaction

The method that was selected to perform SNP genotyping was the ABI Prism® SNaPShot™ Multiplex Kit (Applied Biosystems). It is a high throughput minisequencing-based approach which makes use of an unlabelled internal oligonucleotide primer that anneals one base pair upstream (5') of the relevant SNP, and is extended by a single fluorochrome-labelled dideoxynucleotide ((F)-ddNTP). The internal primers are designed specifically so that the incorporated (F)-ddNTP reflects the corresponding SNP (Lindblad-Toh *et al.* 2000). The alleles are designated by separating the extended products by capillary electrophoresis, and detecting the fluorescence. Specific fluorescent dyes are allotted to each individual (F)-ddNTP, facilitating analysis with GeneMapper™ v3.0 Analysis Software. The assignments are shown below in Table 2.3.5a.

Table 2.3.5a: Dye Assignments of ddNTPs for SNaPShot™ reaction.

Variant	Strand Primer was designed from	Alleles expected	Fluorescent Label of ddNTP	Colour for analysis
rs3918290	positive	A	dR6G	**Green**
		G	dR110	**Blue**
rs1801159	negative	C	dTAMRA™	**Black**
		T	dROX™	**Red**
rs1801265	positive	C	dTAMRA™	**Black**
		T	dROX™	**Red**

The software does not have predefined kit, panel and bin information for SNaPShot™ multiplex samples since different primer sets are used by different users. Therefore, before samples could be analysed, panels and

bins needed to be created from reference data. A bin is a base pair range and dye colour that defines each allele of a SNP. It is specified by boundaries that are user defined. A panel is a set of markers that will be analysed. Thus an initial optimisation step (data not shown) allowed the primers to be assayed separately and bin sets and panels allocated before being pooled and analysed simultaneously. This is an essential step to collect the necessary parameters needed to analyse each individual SNP. The height of a specific data point was given in terms of relative fluorescent units (RFU). A peak height of 150RFU was viewed by the analyst as the cut-off point, below which a peak was disregarded. Analysis was consistent with regards to no-template control samples and cohort samples. When a result could not be determined conclusively, the genotyping was repeated to yield an unambiguous result.

The primers for the SNaPShot™ reaction were chosen manually. No flexibility is permitted regarding the location of the 3' end of the SNaPShot™ primers, therefore, primers could be synthesised from either the positive (+) strand or the negative (-) strand. These primers were analysed for possible secondary structures (see section 2.3.3). The amplicons obtained from primary PCR amplification of *DPYD* using the external primers (section 2.3.4, Table 2.3.3, Addendum B) are used in the SNaPShot™ reaction. BioEdit Sequence Alignment Editor *(www.mbio.ncsu.edu/BioEdit/bioedit.html,* courtesy of Tom Hall; Ibis Therapeutics*)* was used to prepare a local database consisting only of the primary *DPYD* amplicon, and a local BLAST search was subsequently performed to prevent the SNaPShot™ primers from inadvertently amplifying another region of the main *DPYD* amplicon.

The SNaPShot™ primers were designed with a non-homologous 5' tail (5'-CAATCAA-3' tandem repeat) to produce variably sized extension products which differed by 6bp. This enabled sufficient separation by capillary electrophoresis. The detail of the SNaPShot™ primers are shown in Table 2.3.3 (Addendum B). The strand from which each SNaPShot™ primer was designed is important to note, since four possible dNTPs may be incorporated depending on the orientation of the primer .

2.3.5a Post-PCR purification of *DPYD* products

Primary PCR products require a purification step prior to the SNaPShot™ reaction in order to remove unused dNTPs and primers that may interfere with the subsequent thermal cycling primer-extension reaction. To perform the purification, five units of *Shrimp Alkaline Phosphatase* (SAP) (Promega) and one unit of *Exo1* (New England BioLabs) were added to 7.45µl of PCR product in a final volume of 10µl. SAP removes the 5' phosphoryl groups from nucleotides, and *Exo1* degrades the remaining primer sequences. The reaction mixture was incubated at 37°C for 60min, followed by a 15min period of incubation at 75°C to inactivate the enzymes.

2.3.5b SNaPShot™ thermal cycling reaction (minisequencing)

The SNaPShot™ Multiplex Ready Reaction Mix contains AmpliTaq® DNA Polymerase, reaction buffer and (F)-ddNTPs. The SNaPShot™ primers were tested in a singleplex reaction before multiplex reactions were attempted. The amount of water in the reaction was adjusted to accommodate a change in primer or template volumes. A master mix was made if several samples were analysed simultaneously. The SNaPShot™ primers were diluted to 20µM, and pooled in a 2:2:1 (rs1801265: rs1801159:rs3918290) ratio after significant optimisation, in order to achieve a high signal intensity. The minisequencing reaction was carried out in a final volume of 10µl, containing 4µl of cleaned-up primary PCR products, 2.5µl of SNaPShot™ Multiplex Ready Reaction Mix and 3.5µl of pooled primers. Amplification was performed for 25 cycles of denaturation at 96°C for 10s, annealing at 50°C for 5s, and single base extension at 60°C for 30s.

2.3.5c Post-extension purification treatment

Post-extension treatment entailed the addition of one unit of SAP to the entire SNaPShot™ reaction product and subsequent incubation at 37°C for 60min, followed by a 15min deactivating step at 75°C. The SAP removes the 5' phosphoryl groups from the unincorporated (F)-ddNTPs. If the SNaPShot™ reaction product is left untreated, these (F)-ddNTPs may interfere with the genotyping reaction by co-migrating with the fragment of interest.

2.3.5d Capillary electrophoresis on ABI Prism® 3100

The samples were prepared for electrophoresis by mixing 2µl of the purified SNaPShot™ products with 9µl of Hi-Di™ formamide solution (Applied Biosystems) and 0.3µl of the internal size standard, GeneScan™-120 LIZ™ (Applied Biosystems) and loaded into a Thermo-Fast® 96 Detection Plate (Applied Biosystems). Before electrophoresis on the ABI Prism® 3100 Genetic Analyser, the double-stranded DNA was denatured for five min at 95°C on a Touchdown (HYBAID) heat block and thereafter kept on ice for a minimum of two min. The raw data was collected using the ABI Prism® 3100 Data Collection software.

2.3.5e Verification by cycle sequencing

Verification of the SNaPShot™ results were performed by cycle sequencing. The *DPYD* PCR products were electrophoresed on a 1% agarose gel according to protocol (section 2.1.2) for 45min at 120V and the resulting bands excised from the gel using sterile blades. The DNA was purified using the QIAquick® PCR Purification Kit (Quigen) according to the manufacturers protocol, and eluted in a final volume of 30µl. The subsequent purified products were electrophoresed as per protocol (section 2.1.2) to determine integrity and the concentration approximated by comparison with the molecular weight marker. The purified PCR products were subjected to bidirectional cycle sequencing (both the forward and reverse primer were utilised, see Table 2.3.3) using a BigDye™ Terminator Cycle Sequencing kit, version 3.1 (Applied Biosystems, Foster City, CA). The sequencing reaction mix and cycling conditions are depicted in Table 2.3.5b and Table 2.3.5c (Addendum B). The sequencing reaction was performed in a final volume of 10µl and comprised 1µl Big Dye Terminator Cycle Sequencing Mix (Applied Biosystems), 1x Big Dye Sequencing Buffer (Applied Biosystems), 2µM of the appropriate primer, and approximately 50ng/µl purified DNA. The cycling conditions consisted of an initial denaturation step of 96°C for 5min, followed by 25 cycles of 96°C (30s), 50°C (15s) and 72°C (4min). Post cycling, the sequencing reaction was purified of primers, dNTPs and salts. This was achieved by ethanol precipitation by the addition of 2.5µl of sodium acetate (3M at pH 5.0, Merck) and 2.5 volumes of 100% ethanol (Merck). The

samples were incubated at -20°C for approximately 60 min to precipitate the DNA. Thereafter the products were centrifuged for 10 min at 9,300xg. The supernatant was discarded and 30µl of 70% ethanol added to the pelleted DNA. The sample was centrifuged for another 10 min at 9,300xg. After the supernatant was discarded, the sample was allowed to air-dry before rehydration in 10µl Hi-Di™ formamide and electrophoresed on an ABI Prism® 3100 Genetic Analyser and examined using BioEdit Sequence Alignment Editor (*www.mbio.ncsu.edu/BioEdit/bioedit.html,* courtesy of Tom Hall; Ibis Therapeutics*)*.

2.3.6 Genotyping of the 5'UTR VNTR (rs45445694) in *TYMS*

2.3.6.1 PCR amplification

A group of 179 individuals were genotyped for this variant. This group was derived from the larger initial cohort of 240, but a large number of individuals were excluded after their DNA samples failed the quality control. In order to amplify this variant in patient DNA samples, extensive optimisation was required. The variant is located in the 5' promoter region of the *TYMS* gene, which is a GC rich area. Primers, fluorescently labelled with the FAM fluorophore, were designed as previously described (section 2.3.3). The primer sequences are depicted in Table 2.3.3. Parameters altered to obtain optimum PCR amplification included increasing the amount of template DNA to 200ng, increasing the final amount of $MgCl_2$ to 3.5mM and extended cycling (denaturing, annealing and extension) parameters. Furthermore, a Touchdown PCR was performed in order to attempt to increase the product yield with conditions as follows: 95°C for 5 min, followed by 7 cycles with decreasing annealing temperatures (T_a) of 95°C for 45 seconds, 63°C→56°C for 45 seconds, 72°C for 60 seconds, 30 cycles of 95°C for 45 seconds, 54°C for 45 seconds, 72°C for 60 seconds and finally 72°C for 10 min. An attempt was made to amplify the VNTR using The FailSafe™ PCR system buffer (Epicentre® Biotechnologies, Madison, WI), that is optimised for amplifying GC rich areas, but was unsuccessful. Finally, by addition of 5% glycerol to the PCR reaction, amplification was obtained. The final optimised PCR setup is as described in section 2.3.4, omitting the $MgCl_2$ and replacing it with 5% glycerol. The reaction mixture is depicted in Table 2.3.6a (Addendum B).

The PCR cycling conditions are similar to those described in section 2.2.1, with the exception of a T_a of 52°C. The conditions are portrayed in Table 2.3.6b (Addendum B).

In order to determine the size of the alleles in a DNA sample and thereby derive the number of repeats, genotyping was performed on the ABI Prism® 3100 Genetic Analyser as described in section 2.2.1.

2.3.6.2 Verification by cycle sequencing

Verification of the genotyping results was performed using cycle sequencing as described in section 2.3.5e. The sequencing reaction mix and cycling conditions were adapted to facilitate the analysis of a GC rich amplicon. Primarily, 5% glycerol was added to the reaction mix (Table 2.3.6c, Addendum B) and the cycling conditions adjusted as follows: an initial 98°C for 10min, followed by 40 cycles of 98°C (1min), 55°C (1min) and 72°C (2min) (Table 2.3.6d, Addendum B).

2.3.7 Analysis of *TYMS* 3'UTR insertion/deletion (rs16430)

2.3.7.1 PCR amplification

The reaction mix and cycling conditions for the amplification of the *TYMS* 3'UTR insertion/deletion (rs16430) in 192 patient DNA samples are depicted in Table 2.3.7a and 2.3.7b (Addendum B). A group of 192 individuals were genotyped for this variant. This group was derived from the larger initial cohort of 240, but a large number of individuals were excluded after their DNA samples failed the quality control. The PCR was performed in a final reaction volume 25µl, containing 100ng template DNA, 0.4µM of each primer, 0.5U/µl of Taq polymerase, 1x buffer, and 200µM dNTPs. Each PCR comprised an initial 5min denaturation step at 95°C, followed by 30 cycles of 94°C (1min), 54°C (1min) and 72°C (80s), with a final 7min extension at 72°C.

2.3.7.2 Sizing of fragments by non-denaturing high performance liquid chromatography

In order to effectively size the 6-bp insertion/deletion, non-denaturing high performance liquid chromotography (ndHPLC) was performed using the Transgenomic WAVE® DNA Fragment Analysis System (Transgenomic Inc, USA). Sizing of fragments are possible under non-denaturing conditions and the elution behaviour of a fragment is independent of the sequence of the DNA. The concept of this method relies on DNA that bind to a positively charged stationary column within the system, containing triethylammonium acetate ions (TEAA), via the negatively charged phosphate ion backbone. When an increasing concentration of a mobile phase containing a counter ion, such as acetonitrile (thus an increase in the fraction of organic solvent), flows through the column, the DNA, and TEAA bridging molecule, will be displaced off the column in a linear gradient separation manner. The DNA is then detected through UV absorption. Fragments will elute in a size dependent manner, and since smaller fragments would contain fewer phosphates to bind to the column, they will elute first (WAVEMAKER™ Software Manual 2002). The 3'UTR PCR products were sized at 50°C. The samples were separated within a linear acetonitrile gradient at a flow rate of 1.5ml/min, within a 2%/min increase in buffer B concentration over 4.5 min. The starting/ending buffer B concentration for the amplicon was 56-70%. The product peaks eluted at 61%B, between 4 and 5 min.

Three types of samples are expected after WAVE® analysis; heterozygous samples (eluted at approximately 4.4 and 4.8 min), homozygous insertion (eluted at approximately 4.8 min) and homozygous deletion samples (expected to elute at approximately 4.4 min). The samples could be distinguished based on the retention time and peak profile. In some cases the distinction was not clear, and subsequent partially-denaturing HPLC was performed to verify the results.

2.3.7.3 Partially denaturing HPLC

The principle of partially denaturing HPLC is based on the detection of heteroduplices which is formed between (in this case) the deletion and

insertion amplicons in an individual that is heterozygous for the deletion variant. During heteroduplex formation, two species of DNA (a known homozygous sample and a query sample, either heterozygous for the variant in question or homozygous for the variant) is mixed under denaturing conditions, and allowed to reanneal by cooling. Thereafter two or more species of DNA exist in the sample with different thermodynamic properties; perfectly matched homoduplex molecules, and mismatched heteroduplex molecules. Essentially, the heteroduplex molecules are less stable than homoduplex molecules, due to the mismatched DNA (Rudolph *et al.* 2002). The reaction sample containing the mixed DNA is loaded onto an autosampler tray and injected into a dHPLC column. The DNA is absorbed onto the column (stationary phase) and eluted off the column by means of an increase in the concentration of the acetonitrile mobile phase at a constant partially denaturing temperature (52-75°C). Under partially denaturing conditions, because of their unstable thermodynamic properties and hence lower affinity for the column, the heteroduplexes are expected to elute from the column earlier than homoduplexes (Xiao and Oefner 2001).

2.3.7.3a Sample preparation and heteroduplex formation

For 104 samples, equimolar amounts of the unknown (suspected heterozygous) and a known homozygous insertion PCR sample (as established during non-denaturing conditions, section 2.3.7.2) is mixed and denatured at 95°C for 5min. The samples are allowed to reanneal by slowly decreasing the temperature to 25°C at a rate of 0.1°C/4s.

2.3.7.3b Temperature and method optimisation

Prior to deletion detection, the text version of the amplicon was analysed using WAVE® software. The software predicts the temperature at which heteroduplex detection will occur, based on the melting characteristics of the fragment. Certain sequence motifs may affect the melting temperatures of the amplicon. SNPs (or deletions) resolve optimally if the helical fraction ranges from 70%-85% (thus 70% of the fragment is single stranded at a specific temperature). The 3'UTR amplicon had a uniform melting profile (Figure 2.3.7A), allowing for deletion detection at a single temperature (Figure

2.3.7B). Additionally, the gradient parameters and flow rates were determined empirically based on the melting profile of the amplicon.

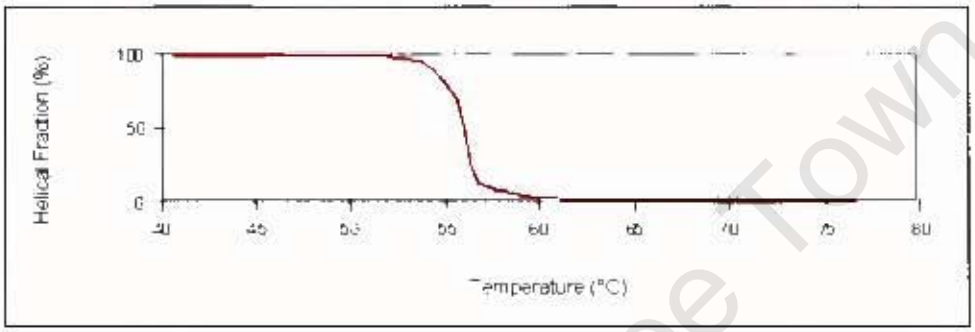


Figure 2.3.7A: Melting profile of *TYMS* 3'UTR amplicon showing a uniform melting profile. This uniformity allowed deletion detection to occur at only one temperature. Y-axis shows the predicted percentage helicity of the fragment, and the X-axis shows the melting temperature.

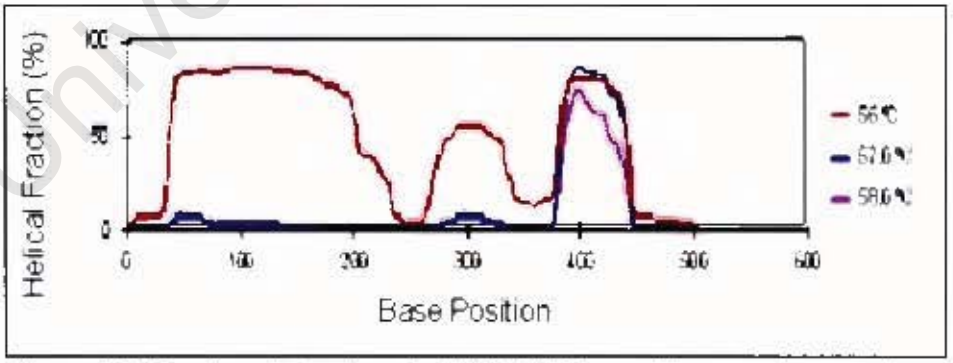


Figure 2.3.7B: Helical fraction of *TYMS* 3'UTR amplicon at three predicted temperatures. The optimal detection temperature for this amplicon is where 70% of the fragment is single stranded; in this case it is represented by the red line, at ~56°C. The Y-axis shows the predicted percentage helicity of the amplicon, and the X-axis represents the base pair positions within the PCR amplicon.

2.3.7.3c WAVE® analysis

Post heteroduplex formation, the samples were separated within a linear acetonitrile gradient at a flow rate of 1.5ml/min, adjusted by means of a -0.2 time shift, with a 5%/min increase in buffer B concentration over 2.6min. The product peaks eluted at 60%B, between 1.5 and 2.5min. The starting/ending buffer B was 57-73%. Heteroduplex detection occurred at 55°C.

If, during the mixing experiment, the unknown sample was a homozygous insertion, only homoduplexes would occur after heteroduplexing, and be evident on the elution profile as uniform peaks eluting at approximately 2.1 min. If the unknown sample was heterozygous, hetero- and homoduplexes would be present after heteroduplex formation, and the elution profile would contain a double peak, eluting at approximately 1.9 and 2.1min. Similarly, a homozygous deletion sample would contain three species after the heteroduplex reaction, eluting at approximately 1.7, 1.9 and 2.1min, resulting in a triple peak elution profile.

2.3.7.4 Verification by cycle sequencing

Six samples displayed an aberrant elution profile compared to the majority profiles, and were subjected to bidirectional cycle sequence analysis as described in section 2.3.5e, using both forward and reverse primers. Sequencing results were analysed using BioEdit Sequence Alignment Editor *(www.mbio.ncsu.edu/BioEdit/bioedit.html,* courtesy of Tom Hall; Ibis Therapeutics) .

2.4 STATISTICAL ANALYSES

2.4.1 Analysis of age at diagnosis and recurrence-free survival

Survival curves (based on age at diagnosis of Lynch spectrum type cancer) were estimated by the Kaplan-Meier product limit method (Stata; StataCorp LP) for 145 and 201 individuals to evaluate the probability that subjects with a particular stage (145) or site of cancer (201), respectively, would have an earlier age at onset. The individuals included in this analysis are derived from the larger cohort of 240, but several individuals were excluded from the

analyses failing accurate clinico-pathological and follow-up data. The failure event was defined as diagnosis of colorectal cancer. The observation period started at birth and ended at the failure event. Some patients were recruited and entered into a comprehensive screening programme by the Division, following the identification of a disease- causing mutation in a first or second degree relative with colorectal cancer. These individuals were thus cancer free when entered into the screening programme and some only developed cancer after the fact. These patients were censored on the screening date and not the date of the failure event since they undergo regular colonoscopies and are monitored. This would institute an assessment bias and hence tumour progression is altered. The comparison between the curves was made by Log-Rank test with a p-value of <0.05 considered as being significant.

Additionally, the Kaplan-Meier product limit method (Stata; StataCorp LP) was used to evaluate the survival probability for disease recurrence with or without chemotherapeutic intervention for subjects with a particular MMR genotype. Patients included into the analysis had received curative surgery and hence were disease free at the start of the observation period. Curative surgery was defined as a complete removal of the disease. Patients with metastatic disease (and receiving palliative chemotherapy) were subsequently excluded from the analysis. The failure event was defined as local (reappearance of cancer at previously treated site) or systemic/distant (metastases to other organs such as liver, lymphatic system and lungs) recurrence. Total analysis time for participants on chemotherapy started from the date of their first treatment to either a) the date of first local or distant recurrence; b) last known recurrence-free date; or c) date of death if no recurrence was observed. Data were censored on either a, b, or c. Total analysis time for participants without chemotherapeutic treatment started from the date of curative surgery for primary cancer, to either a, b, or c. Kaplan-Meier survival curves were plotted and the difference between the curves assessed using the Log-Rank test. p-values less than 0.05 were considered to be statistically significant.

2.4.2 *DPYD* and *TYMS* haplotype analysis

In order to obtain phase information for the cohort from the genotyping data, haplotype reconstruction was performed using PHASE version 2.1 (v2.1) (Stephens *et al.* 2001; Stephens and Scheet 2005) for both affected and control study groups. PHASE v2.1 is a mathematical algorithm that uses Bayesian statistics to reconstruct haplotypes from population data and infer missing data. Haplotypes of the three bi-allelic *DPYD* SNPs were constructed for each individual from SNaPShot™ genotyping data. Similarly, haplotypes of the *TYMS* 5'UTR VNTR and 3'UTR insertion/deletion were deduced. Some genotypes could not be determined experimentally due to poor DNA integrity. These missing haplotypes were inferred from known genotype or haplotype frequencies using PHASE v2.1 and used in subsequent data analyses. The study cohort comprised three population groups; Caucasian (CAU), Mixed Ancestry (MA) and Black Ancestry (BKA), and these subjects were analysed separately in order to infer population-specific haplotypes. In order to obtain reliable results, posterior estimates for the missing phase haplotypes were based on 10,000 iterations, with a thinning interval of 1 and a burn-in of 10,000. Family-based data enabled the use of "known"/specified haplotypes which improves performance of the algorithm (Stephens *et al.* 2001). Additionally, the algorithm was run multiple times using random seed numbers to determine whether haplotype frequencies would differ significantly from one estimation to the next.

The PHASE program produced several files used for the purpose of this study. The primary output document (*outputfilename.out*) was formatted to contain a list of the best haplotype guess for each individual. If a phase was difficult to infer, the haplotype was noted in parentheses, and if it was uncertain, it was noted by square brackets (Addendum C). Additionally, a second file (*outputfilename.freq*) contained the posterior estimates for the sample haplotype frequencies which could also be considered as estimates for the population haplotype frequencies (Addendum C).

Additionally a case-control permutation test was performed to test for

significant differences in haplotype frequencies in the case and control groups. A control population consisting of 50 CAU and 50 MA individuals had previously been genotyped (Honours Project, 2008, Division of Human Genetics, unpublished work). The PHASE v2.1 algorithm was applied for each population group for both *DPYD* and *TYMS*. One hundred permutations were performed in order to test the null hypothesis that the case and control haplotypes are more similar than different.

2.4.3 Allele frequencies of pharmacovariants in study populations and Hardy-Weinberg equilibrium

The frequencies of the five variants in *DPYD* and *TYMS* were determined for each control (CAU and MA) population group, as well as 142 unrelated, affected subjects from CAU (28), MA (94) and BKA (20), and compared using Pearson's chi-square test (one degree of freedom). In circumstances where the expected number of observations (or allele frequencies) were less than five, Fisher's Exact test was performed instead. Both Fisher's Exact and Pearson's chi-square require that individuals are unrelated, however some of the subjects in the CRC study populations were part of large families, and thus only unrelated individuals were included in the analysis. Pearson's goodness of fit chi-square test (one degree of freedom) was also performed in order to determine whether the observed genotype frequencies differ from what is expected for each of the studied variants under Hardy-Weinberg Equilibrium. A p-value of 0.05 was regarded as significant, under which the null hypothesis, stating no significant difference, was rejected. Bonferroni correction was applied for multiple comparisons.

2.4.4 *In silico* comparative analysis of allele frequencies of pharmacoSNPs in *TYMS* and *DPYD* in indigenous African populations and Caucasian populations.

2.4.4.1 Construction of variation database

In a recent collaborative study, the Division of Human Genetics and the

University of The Witwatersrand assayed genome-wide variation for five indigenous African populations (BKAF) (Unpublished Work, 2007). These were Sotho/Tswana (STS), Xhosa (XHS), Zulu (ZUL), Khoisan (KHS) and Herero (HER). The allele frequencies of 933,684 SNPs were determined for 126 unrelated individuals (STS:25, XHS:34, ZUL:20, KHS:22, HER:25) using the Affymetrix Genome Wide SNP 6.0 Array (Homo Sapiens, Genome Assembly: NCBI Build 36, UCSC hg18). This data was compared to the four population groups defined in The HapMap project (The International HapMap Consortium, 2005). The Affymetrix Array data was stored in a local Biomart database with a Martview user interface, which allows a user to compare allele frequencies of any SNP assayed by the Affy Array within the five BKAF and four HapMap populations. The interface allows SNPs to be queried by preset fields such as a dbSNP *rs* identifier, the associated gene and the chromosomal position of the variant.

2.4.4.2 Frequency estimation of 5-FU pharmacoSNPs in the indigenous African populations

The Affymetrix array assays 933,684 SNPs across the genome. Thus not all SNPs in all genes are represented. Therefore, the Affymetrix Array data was queried to determine whether any of the pharmacorelevant SNPs featured in this study (section 2.3.1), in the 5-FU metabolising gene, *DPYD*, were also genotyped in the 5 BKAF Populations. The respective dbSNP *RS* identifier for the SNPs (rs3918290, rs1801265, rs1801159) were used in the query.

None of the three SNPs were featured on the array, so no direct allele frequency comparisons could be performed. Therefore all SNPs in *TYMS* and *DPYD* which were assayed using the chip were corroborated against tagging SNPs (tSNPs) or local representative SNPs (lrSNPs) associated with the three *DPYD* SNPs on the Ensembl database, as well as with the *TYMS* VNTR (rs45445694) and deletion (rs16430). However, none of the SNPs featured on the array tagged any of the pharmacovariants of interest.

2.4.4.3 Comparison of African genotyping data with data from The HapMap project

A comparison of the IrSNPs in *TYMS* and *DPYD* was made between the five BKAF and The Hapmap European Caucasian populations (CEPH) in order to identify SNPs with significantly different allele frequencies that may affect treatment outcomes in a population-specific way.

Once the allele frequencies were obtained for all selected SNPs, a pair wise comparison was performed for all the BKAF populations versus the CEPH population. The Fisher's Exact Test was used to determine the extent of the difference in allele frequencies between these populations. Using a p-value cut-off of 0.05, it was determined which SNPs show significant different allele frequencies between all the BKAF populations and the CEPH population. Thereafter a p-value cut-off of 0.001 was used in order to detect the most significant SNPs .

CHAPTER 3
RESULTS

3.1 COHORT SELECTION

3.1.1 Selection of pilot cohort

An attempt was made to obtain follow-up records (RT-folders) of 240 affected individuals, currently in the patient registry of the Division of Human Genetics, from the Oncology Department at Groote Schuur Hospital (GSH). These patients have been comprehensively screened for mutations in the mismatch repair system (MMR). The patient registry had been previously populated with the phenotypic and clinico-pathological information of these patients, *inter alia* age at diagnosis, pathology, and staging of cancers. The RT-folders provided information *viz.* treatment, treatment outcomes, 5-year survival, and long term prognosis. Information was gathered on these aspects in order to make a clear and detailed comparison between patients with or without an intact MMR system, and to simultaneously initiate a local, population based cancer registry. The primary aim was to contribute to the understanding of whether the molecular pathological differences between these two subgroups influenced response to chemotherapy.

The registry contained the screening history, clinical, pathological and follow-up information on 92 individuals with mutations in either one of six MMR genes (Mut+), and 148 individuals without a mutation in the MMR genes (Mut-), therefore, the patient registry could serve as a valuable resource for the study of pharmacogenetics in colorectal cancer, and an entry point from whence a local cancer registry could be pioneered which could provide incentive to the already existing, yet under reported, National Cancer Registry. As shown in Table 3.1.1a, only 87 Mut+ individuals gave consent for their RT-folders to be accessed. RT-folders were available and collected for 31% (n = 27/87) of the Mut+ individuals of whom only 13 individuals received chemotherapy. Three of these subjects received treatment at

facilities outside Cape Town, with no information available pertaining to treatment and treatment outcomes. RT-folders could not be located for 69% (n = 60/87) of the Mut+ individuals. The possible reasons for this are elaborated on in Chapter 4. Of the 148 Mut- patients, four patients did not consent for their folders to be accessed, however, of the remaining 144 patients, 53 folders had been obtained and 94 (64%) could not be located. Of the 53 patients, only 25 received chemotherapy.

Table 3.1.1a: Comprehensive summary of the final number of study participants

# Entire Affected Cohort in Division patient registry	450		
	Screened for mutations in MMR genes 240		Unscreened 210
Final recruited Cohort	Mutation Positive	Mutation Negative	
# Individuals with Consent (n=231)	87	144	N/A
# RT folders collected (n=80)	27 (31%)	53 (36%)	N/A
# Individuals treated with 5-FU (n=38)	13 (15%)	25 (17%)	N/A

— "number of"
N/A — not applicable

Comprehensive clinico-pathological information was available from the registry for only 144 of the 231 individuals who gave consent, with 80 RT-folders supporting the information. The distribution of the clinico-pathological characteristics of the patient cohort of 144, as obtained from the registry, is tabulated in Table 3.1.1b. For the remaining patients, information was discontinuous. Cancers of the distal colon appeared in 66% of Mut- patients, whereas 76% of Mut+ patients presented with cancers of the proximal colon. It was determined that cancers of the distal colon appeared more frequently in Mut- patients, whereas Mut+ patients presented with cancers of the proximal colon more often ($p<0.001$, Fisher's Exact). The Dukes' stage of a tumour at diagnosis was not contingent on mutation status ($p=0.57$; Fisher's Exact). Patients presenting with extra-colonic tumours (endometrial, cervical

and duodenal), did not receive any chemotherapeutic treatment. In this cohort, chemotherapy was given only for tumours classified as Dukes' stage B and C (n=29). Few patients who had received chemotherapy experienced local or distant recurrence (n=4).

Table 3.1.1b: Clinico-pathological characteristics of 144 mismatch repair mutation-positive and negative patients

Variable		Mut -		Mut +	
		Chemo	No Chemo	Chemo	No Chemo
Age at diagnosis	<45 years	19	52	5	34
	>45 years	3	16	5	10
Gender	Male	12	26	4	29
	Female	10	42	6	15
Site of Cancer	Proximal◇ Colon	5	25	7	34
	Distal❖ Colon	17	42	3	7
	Extra-Colonic❖	0	1	0	3
Dukes# Stage of Cancer	A	0	6	0	4
	B	7	26	3	14
	C	14	31	7	22
	D	0	5	0	3
Recurrence	YES (local or distant metastasis ^)	3	7	1	5
	NO	16	12	9	11
	Unknown Status	0	31	0	22
	Primary disease is metastatic	3	18	0	6

Mut - : no mutation in mismatch repair genes

Mut +: mutation in mismatch repair genes

> more than

< less than

◇ Comprising cancers of the ceacum, ascending colon, hepatic and splenic flexure, appendix and transverse colon

❖ Comprising cancers of the rectum, sigmoid and descending colon

❖ Includes cancers of the duodenum, endometrium and cervix

Dukes staging per definition: **A**—tumour penetrates through the mucosal layer of the bowel; **B**—tumour penetrates into/through the muscular layer of the bowel; **C**— Tumour penetrates into/through the muscle layer of the bowel, and affects several local lymph nodes; **D** Cancer has metastasised to other organs such as liver

^ Metastasis defined as: migration of cancer cells through lymphatic system or blood from original tumour site to establish a secondary area of disease elsewhere in the body, such as in the liver, lungs and brain

The extent of the toxic side effects of 5-FU treatment in the cohort varied. Patients mainly presented with diarrhoea (n=9), nausea (n=12) and/or vomiting (n=6) with minor occurrences of hair loss (n=1), cramps (n=1), and constipation (n=2). Only one fatality after 5-FU treatment was recorded, although the patient was diagnosed with an advanced stage of disease, and it is not clear whether the cause of death was due to disease or from a severe reaction to the chemotherapy.

Attendance of the treatment schedules was erratic and non-compliance was common. Three patients were active participants in existing clinical trials and no information was available about their chemotherapeutic regimen and outcomes. Most of the patient cohort received adjuvant ("after surgery") chemotherapy, although three patients received palliative treatment for metastatic disease.

The medical records of the majority of the patient cohort could not be traced successfully. In an attempt to improve the ascertainment of treatment outcome records, substantial remedial actions were taken. Patients, or their relatives, for whom treatment records were missing, were contacted telephonically in order to determine who their treating physician was and what their treatment status or survival status was. The information obtained from this exercise was negligible.

Figure 3.1 describes the organisation of the study cohort as described in section 3.1 to 3.4. This figure will be referred to throughout Chapter 3, in order to facilitate interpretation of the partitioning of the cohort through the course of the study.

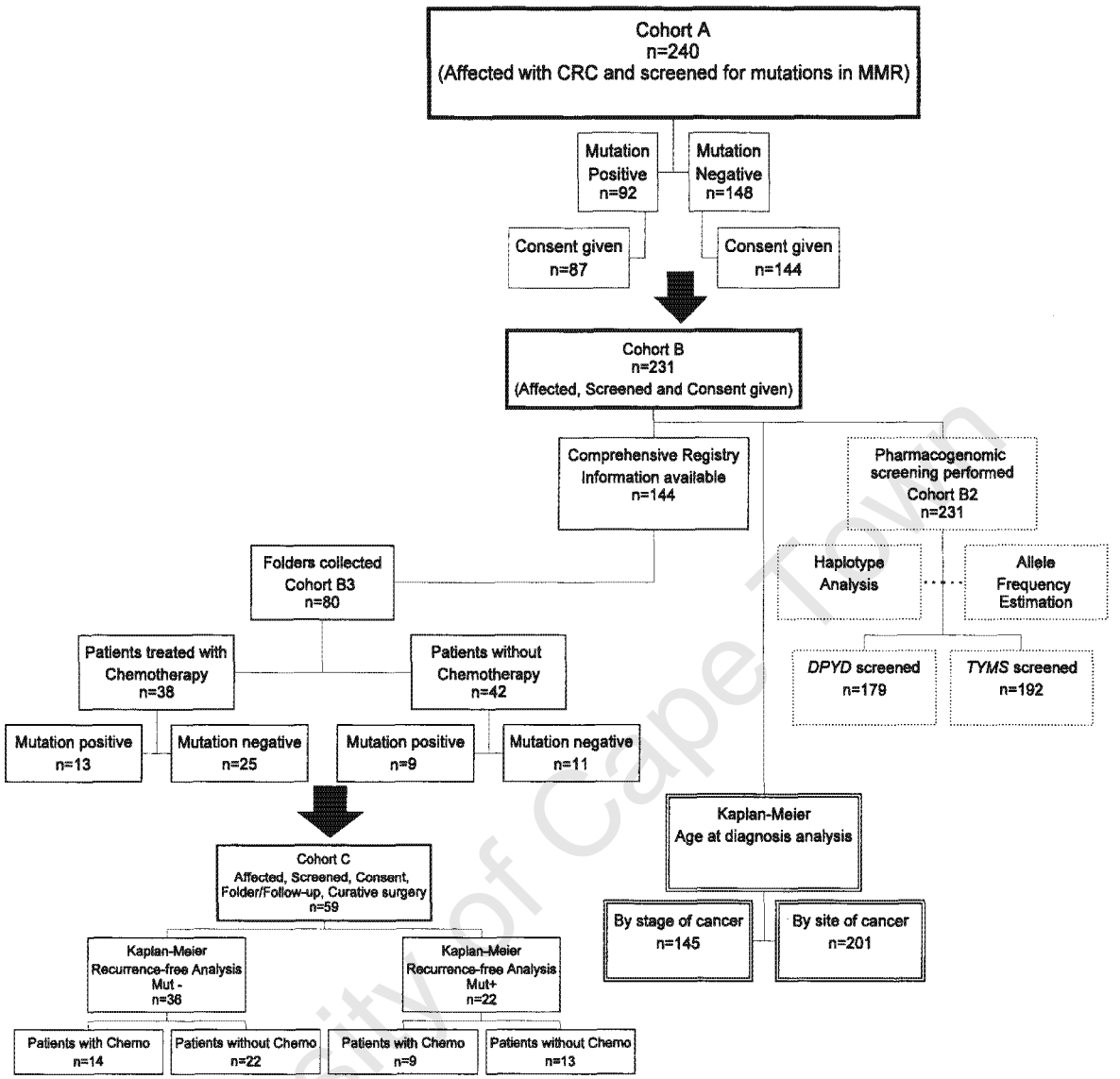


Figure 3.1: Organisation of study cohort facilitating in text referencing

3.2 IDENTIFICATION OF MMR-DEFICIENT AND MMR-PROFICIENT PATIENTS

3.2.1 Evaluation of MSI status of affected patients using the Bethesda panel of microsatellite markers

The original intention was to classify the cohort according to their mismatch repair (MMR) status using the Bethesda panel of microsatellite markers

(section 1.2.1iii). In order to make a comparison between patients with or without an intact MMR system, microsatellite instability (MSI) was assayed. An attempt was made to compare the microsatellite profile of each of the five markers in the panel in DNA extracted from peripheral blood lymphocytes and tumour tissue. The panel of markers were successfully amplified in DNA extracted from peripheral blood for 37 samples. However, the quality of the DNA extracted from formalin-fixed paraffin-embedded tissue sections proved to be severely compromised. Despite several optimisation methods, the Bethesda panel of markers could not be amplified successfully in DNA from all tissue sections. Whole genome amplification was attempted, however this also proved unsuccessful. Therefore, the mutation status of each individual was not based on MSI, but rather on previous comprehensive, reliable MMR mutation screening studies. A brief summary of the MMR partitioning appears in Table 3.1.1a.

3.3 DEVELOPMENT OF PHARMACOGENOMIC ASSAYS AND SUBJECT PROFILING

In order to establish the frequencies of six pharmacorelevant variants in a South African (SA) colorectal cancer (CRC) cohort (*Cohort B2, Figure 3.1*) consisting of Black Ancestry (BKA; n=23), Mixed Ancestry (MA; n=169) and Caucasian (CAU; n=39) subjects, various methods were used.

Interrogation of *DPYD* SNPs

Three SNPs in *DPYD* (rs3918290, rs1801265 and rs1801159) were analysed. In order to facilitate their analysis, rs3918290, rs1801265, and rs1801159 were, after appropriate optimisation, simultaneously amplified and interrogated using the SNaPShot™ multiplex genotyping kit. Of the attempted 231 samples, only 179 could be accurately ascertained, despite repeated efforts to optimise amplification. The subdivision of the observed genotypes is presented in Table 3.3a. For rs1801159 (I543V), the 'TT' genotype occurred in the majority (76% CAU; 63% MA; 63% BKA) of the patients in all three population groups. Similarly, the rs3918290 (IVS14

+1G>A) 'GG' genotype occurred in almost all of the patients (99-100%). The 'GA' genotype was only detected in a single individual in both the CAU and MA population groups. No homozygous 'AA' individuals were observed. For the purpose of this study no BKA control individuals were genotyped for any of the *DPYD* variants.

Table 3.3a: Partitioning of SNP genotypes in *DPYD* in all cohorts

		Population				
		CAU	*CAU control*	*MA*	*MA control*	*BKA*
Number of individuals		29	50	134	50	16
Variant	**Genotype**					
85C>T (R29C)	*TT*	13	32	52	19	2
	CT	14	13	57	23	8
	CC	1	5	25	8	6
	Unknown for this variant	1				
IVS14+1G>A	*GG*	28	50	133	50	16
	GA	1	0	1	0	0
	AA	0	0	0	0	0
1627A>G (I543V)	*TT*	22	29	84	31	10
	CT	7	16	42	18	6
	CC	0	4	8	4	0

CAU: Caucasian
MA: Mixed Ancestry
BKA: Black Ancestry

Figure 3.3A shows a typical SNaPShot™ electropherogram for a compound heterozygous individual. Bidirectional sequencing confirmed the specificity and sensitivity of the technique (Figure 3.3A2).

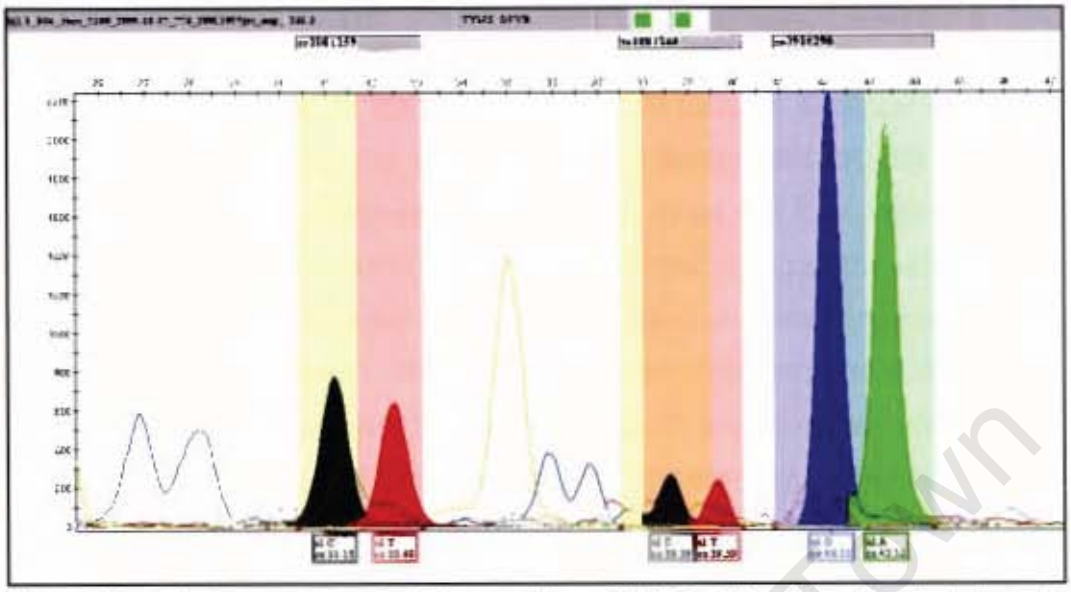


Figure 3.3A: Electropherogram depicting a multiplex SNaPShot™ profile. The shaded peaks represent the genotype of individual NPC 241.1MAR, who is heterozygous for all three bi-allelic SNPs. The shorter fragment, SNP rs1801159 eluted first, followed by SNP rs1801265 and then SNP rs3918290. The alleles are indicated at the bottom of each peak. A red peak represents a "T" allele, a black peak corresponds to a "C" allele, a blue and green peak exemplifies a "G" and an "A" allele respectively. The shaded vertical bands indicate the bins that were manually set for each SNP. A bin is a base pair range and dye colour that defines each allele of a SNP (section 2.3.5). The grey panel at the top of each pair of alleles denotes the marker that is being visualised. The unshaded yellow peak represents the 120 LIZ™ size standard. The unshaded peaks are background noise. The relative fluorescent units are shown on the Y-axis and the size of the fragment is shown on the X-axis.

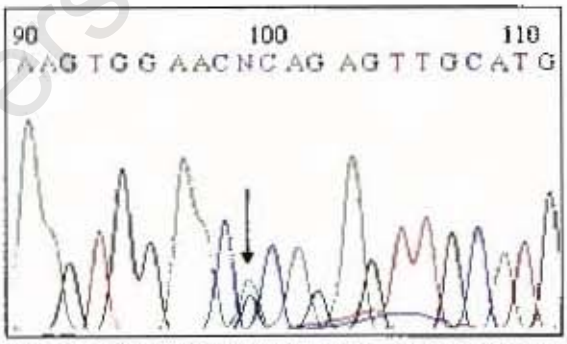


Figure 3.3A2: Electropherogram depicting cycle sequencing of *DPYD* SNP, rs1801265. Patient NPC 240.1AUB was heterozygous for the variant, having both an 'A' and a 'G' allele, which is indicated on the sequence by an arrow and designated 'N' (the sequencing was performed using the reverse primer, and therefore reverse complimentary to the template DNA).

Examination of *TYMS* variants

In *TYMS*, the 28bp 5'UTR VNTR (rs45445694) was interrogated by fluorescent genotyping on the ABI Prism® 3100 Genetic Analyser. Several alleles were possible for the VNTR, and were designated 2R, 3R and 4R, based on the number of repeats present (two, three and four, respectively). It was evident that a consistent discrepancy in the sizing of the alleles was present. The expected target amplicon size based on primer locations, were 369 (two tandem repeats), 397bp (three tandem repeats) and 425bp (four tandem repeats). However, the 2R, 3R and 4R alleles were reflected as fragments of roughly 358, 385 and 413bp respectively. Empirically, the reason for this discrepancy is suggested to be a function of the GC rich sequence properties of the *TYMS* promoter region, which affects the secondary structure of the DNA. An example of an electropherogram is shown in Figure 3.3B. Of the attempted 231 colorectal cancer samples, only 192 could be amplified successfully, similarly, 5 control samples were unable to be amplified. The observed genotype partitions in each of the study cohorts appear in Table 3.3b, and it is evident that the most common genotypes in all populations were 2R/3R and 3R/3R. The 4R allele was absent in all populations except MA, and only observed heterozygous with the 2R and 3R allele. Verification of the sensitivity of the technique was attempted by means of cycle sequencing, but was unsuccessful due to the GC rich nature of the sequence which made analysis difficult.

Table 3.3b: Partitioning of *TYMS* 5' VNTR genotypes in study cohorts

Population	Number of individuals in group	Genotypes				
		2R2R	2R3R	3R3R	2R4R	3R4R
CAU	29	3	19	7	0	0
CAU control	47	9	21	17	0	0
MA	143	22	57	62	2	0
MA control	48	5	19	21	1	2
BKA	20	3	8	9	0	0

CAU— Caucasian; MA— Mixed Ancestry; BKA Black Ancestry

2R— Two tandem repeats; 3R— Three tandem repeats; 4R— Four tandem repeats

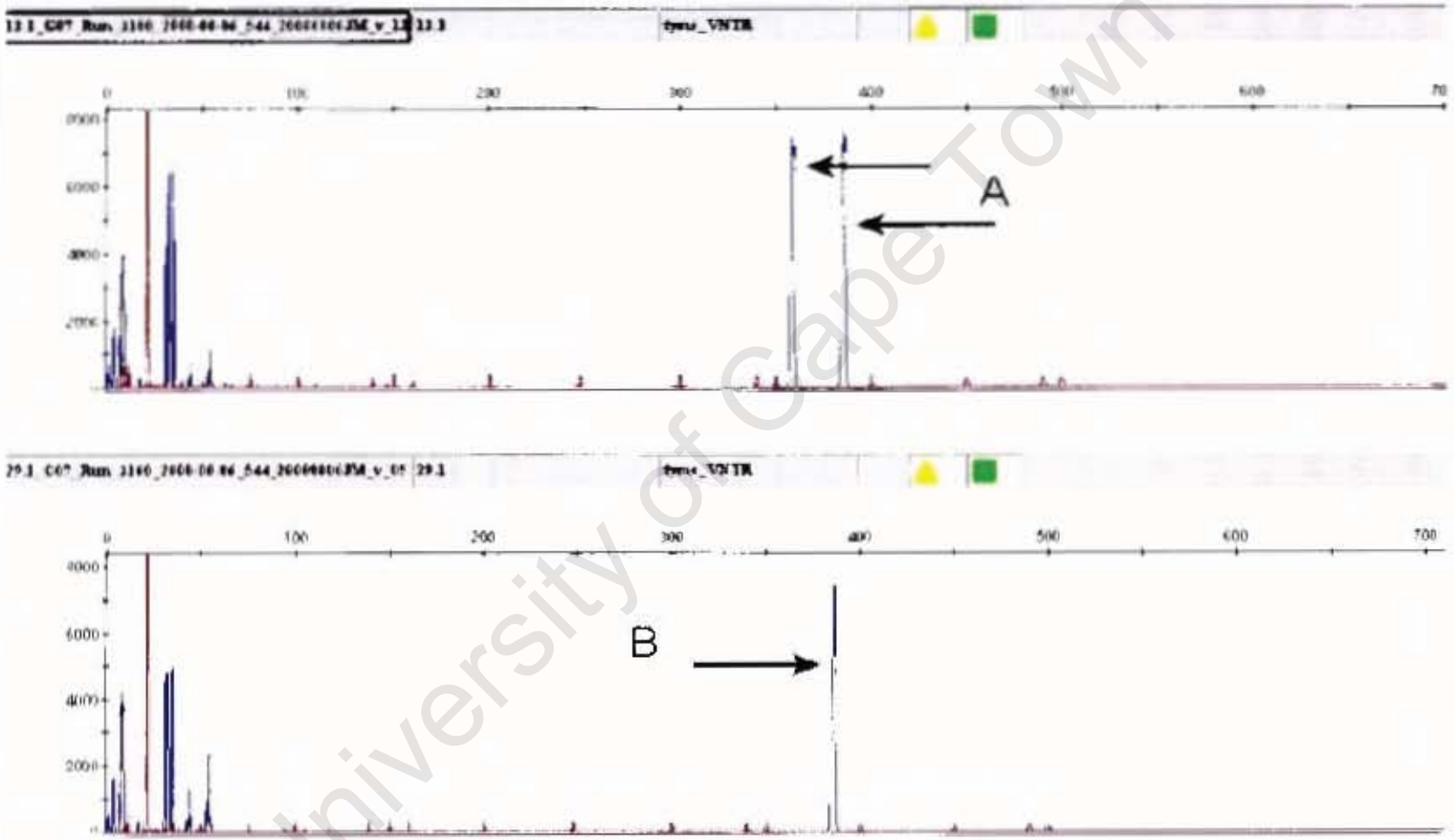


Figure 3.3B: Electropherogram depicting the genotyping results for 5'UTR *TYMS* VNTR. The relative fluorescent units are shown on the Y-axis, whilst the size of the fragment is shown on the X-axis. The arrows indicate A) the individual NPC13.1C YN who is heterozygous for two and three repeats and B) the individual NPC 29.1ABD, who is homozygous for three repeats. The red peaks denote the 500 ROX™ size standard

The 6-bp 3'UTR insertion/deletion polymorphism (indel) (rs16430) was successfully interrogated by non-denaturing and partially denaturing high performance liquid chromatography (dHPLC) or WAVE® analysis in 82 control individuals of CAU and MA, as well as 192 CRC patients (BKA-20, CAU-29 and MA-143). This method successfully identified 130 homozygous samples containing two insertion alleles (6+/6+), as well as 60 heterozygous individuals with both an insertion and deletion allele (6+/6-). The samples could be distinguished based on the retention time and peak profile. A homozygous individual would display a single peak on an elution profile, as exemplified in Figure 3.3C.

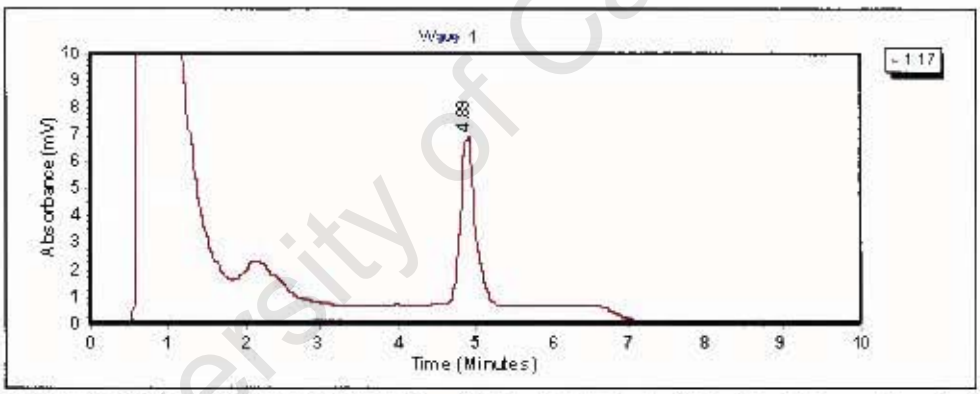


Figure 3.3C: dHPLC chromatogram. The homozygous 6-bp insertion amplicon of sample NPC 1.17JES, elutes off the WAVE™ column at 4.99 minutes. In an elution profile, the UV absorption at which detection occurs appears on the Y-axis, whilst the elution time appears on the X-axis. The additional peaks at ~1 and 2 minutes, represent the acetonitrile injection and the detection of the primers in the reaction, respectively. Flow rate: 1.5ml/min; column temperature: 50°C; acetonitrile gradient: 56-70 (%B).

Similarly a heterozygous individual would display a double peak profile (Figure 3.3D). No homozygous deletion individuals (6-/6-) could be distinguished from 6+/6+ individuals with this technique.

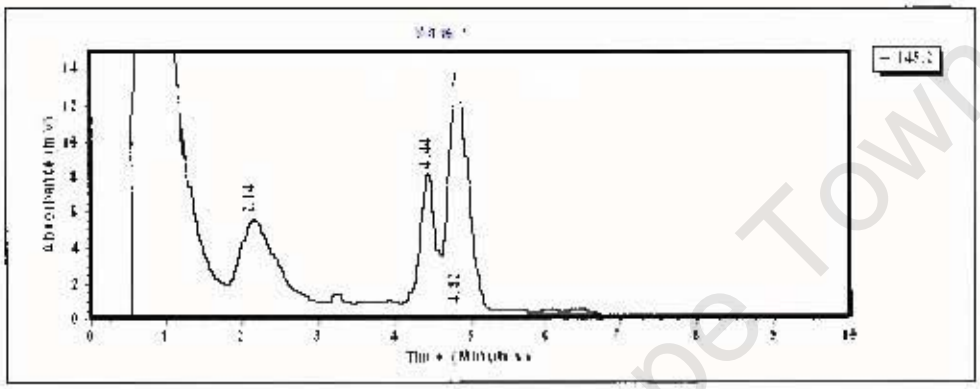


Figure 3.3D: dHPLC chromatogram. The 6-bp insertion and deletion amplicons of the heterozygous sample NPC 145.2MAR, elutes off the WAVE™ column at 4.44 and 4.82 minutes, respectively. In an elution profile, the UV absorption at which DNA detection occurs appears on the Y-axis, whilst the elution time appears on the X-axis. The additional peaks at ~1 and 2.14 minutes represent the acetonitrile injection and the detection of the primers in the reaction, respectively. Flow rate: 1.5ml/min; column temperature: 50°C, acetonitrile gradient: 56-70 (%B).

Subsequent partially-denaturing HPLC was performed for 100 samples to verify the results. A control sample with a homozygous 6+/6+ genotype was heteroduplexed with each sample to be identified. If the unknown sample was homozygous 6+/6+, only homoduplexes would be present in the sample, and a uniform peak was expected on the elution profile. If the unidentified sample was heterozygous, hetero- and homoduplexes would be present, and the elution profile would contain a double peak. Similarly, if the unknown sample was homozygous 6-/6-, three species would be present in the reaction; homoduplexes of the insertion allele, heteroduplexes, and homoduplexes of the deletion allele, resulting in a triple peak profile. Some profiles were anomalous, and subjected to bidirectional cycle sequencing to resolve the results.

Cycle sequencing in both forward and reverse directions confirmed the presence of either one or two 6-bp deletion alleles in all query samples. Due to the nature of the sequence in this region, four hypothetical 6-bp deletions were possible (Table 3.3c). However, only two, rs16430 and rs11280056 are published and curated. These deletions (TTAAAG/- and AAGTTA/-, respectively) are separated by a 3-bp difference on the *TYMS* genomic sequence, and were not distinguished by dHPLC analysis. Figures 3.3E and 3.3F illustrate the aberrant WAVE® elution profiles and subsequent sequencing characterisation. The partitions of the 3'indel genotypes are represented in Table 3.3d. Only two patients were homozygous for the 3'UTR deletion, and both were from the MA population group, as highlighted in Table 3.3d.

Table 3.3c: Putative 6-bp deletion sequences as determined by the nature of the sequence in the 3'UTR region of *TYMS*

OBSERVED DELETION SEQUENCE:

5' GTGGTTATGAACTTTATAGTTGTTTTA 3'

HYPOTHETICAL 6-bp DELETIONS:

1) dbSNP: rs16430 (ttaaag/-)

5' GTGGTTATGAACTTTAAAGTTATAGTTGTTTTA 3'

2) dbSNP: rs11280056 (aagtta/-)

5' GTGGTTATGAACTTTAAAGTTATAGTTGTTTTA 3'

3) Possibility A (taaagt/-)

5' GTGGTTATGAACTTTAAAGTTATAGTTGTTTTA 3'

4) Possibility B (aaagtt/-)

5' GTGGTTATGAACTTTAAAGTTATAGTTGTTTTA 3'

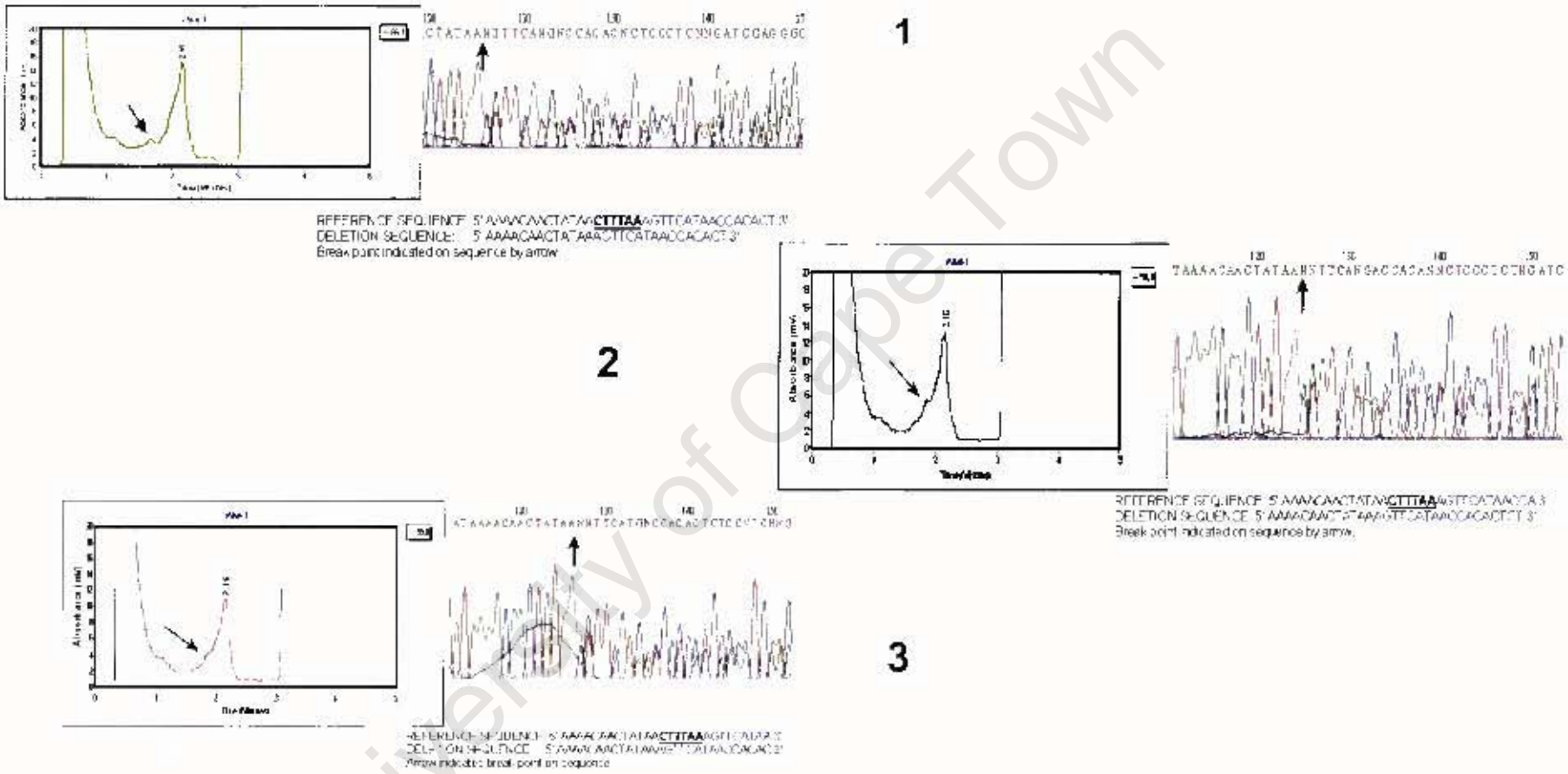


Figure 3.3E: Aberrant dHPLC elution profiles and subsequent sequence analyses depicting 6-bp deletion. Profiles **1-3** indicate individuals NPC 66.1REG, NPC 94.1MAR and NPC 95.1CHA respectively. All three patients are heterozygous for the 6-bp deletion. dHPLC performed in an acetonitrile gradient of 57-73 (%B), flow rate of 1.5ml/min, and column temperature at 55°C.

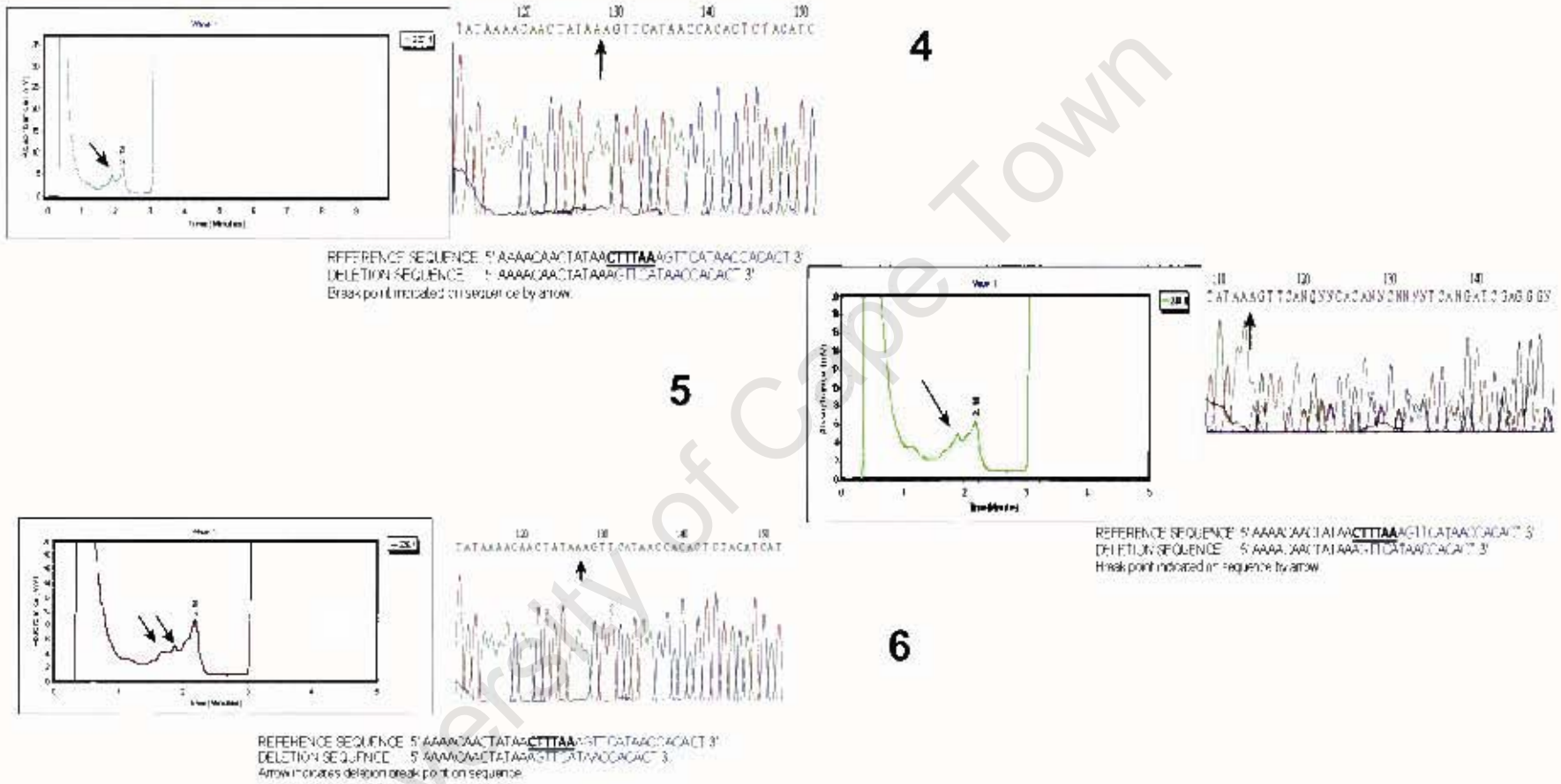


Figure 3.3F: Aberrant dHPLC elution profiles and subsequent sequence analysis depicting the 6-bp deletion. Profiles 4-6 indicate individuals NPC 202 1PET, NPC240.1AUB and NPC 226.1MAR respectively Both NPC 226.1MAR and NPC 202 1PET are homozygous for the 6-bp deletion and NPC 240.1AUB is heterozygous. dHPLC performed in an acetonitrile gradient of 57-73 (%B), flow rate of 1.5ml/min, and column temperature at 55°C.

Table 3.3d: Partitioning of *TYMS* 3'UTR insertion/deletion genotypes in study cohorts

Population group	Number of individuals in population group	Genotype		
		6-/6-	6+/6-	6+/6+
CAU	29	0	9	20
CAU Control	42	0	21	21
MA	143	2 ❖	44	97
MA Control	40	0	23	17
BKA	20	0	7	13

CAU: Caucasian

MA: Mixed Ancestry

BKA: Black Ancestry

❖ Only two individuals from MA were identified to carry both deletion alleles.

3.4 STATISTICAL ANALYSES

3.4.1 Analysis of age at diagnosis of disease and recurrence-free survival

Of the initial cohort of 231 patients, clinico-pathological data on Dukes' staging of cancer was available in the registry for a total of 145 patients. These patients had presented with either Dukes' stage A, B, C or D cancer (*cohort B, Figure 3.1*). Additionally data on the site of the primary tumour was available for 201 patients, who had been diagnosed with either proximal or distal tumours (*cohort B, Figure 3.1*). Hence the age at diagnosis of disease was compared between patients with a particular stage or site of cancer. A cohort of 145 patients with either Dukes' stage A (10), B (51), C (75) or D (9) cancer were firstly analysed. Kaplan-Meier survival analysis showed that there is no significant difference in the age at diagnosis/cancer-free survival time from birth, regardless of stage of cancer. In other words, patients presenting with Dukes' stage A CRC do not display an earlier or later age at diagnosis than, e.g, stage C CRC patients (Figure 3.4.1A).

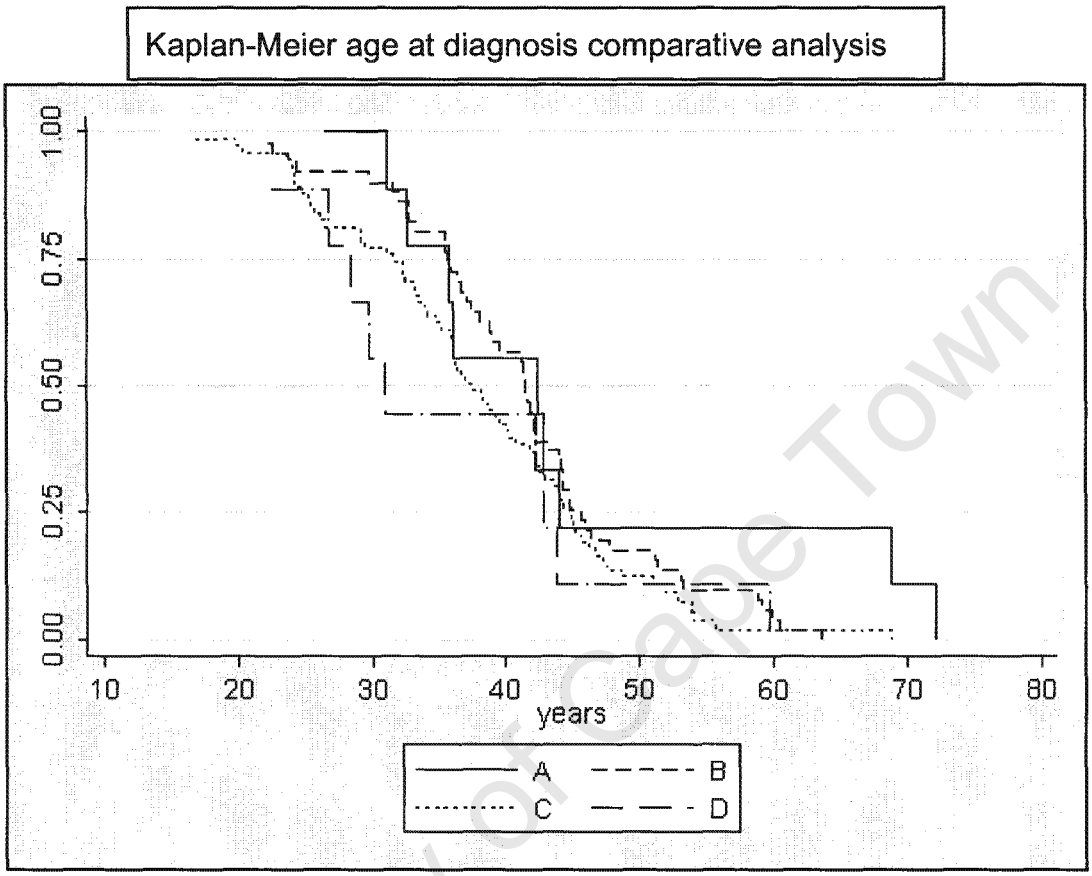


Figure 3.4.1A: Kaplan-Meier age at diagnosis comparison of CRC patients according to Dukes staging of cancer (Log-Rank test; p=0.212). X-axis depicts age at diagnosis in years. Y-axis depicts proportion of individuals cancer-free at one time. The legends A, B, C and D indicate the stage of cancer according to the Dukes stage. A-tumour invaded mucosa; B-tumour invaded into but not through muscle layer; C-tumour metastasised to local lymph nodes; D-tumour metastasised to distant lymph nodes and other organs.

No significant difference was detected in the age at diagnosis between patients with tumours of the proximal colon (includes appendix, caecum, ascending colon, hepatic flexure, transverse colon and splenic flexure) compared to patients with tumours of the distal colon (includes descending colon, sigmoid and rectum). Patients diagnosed with distal tumours, are not diagnosed with disease later than patients with proximal tumours (Log-Rank test, p=0.057, Figure 3.4.1B). It should be noted that the p-value that was

obtained in this analysis is approaching significance, and warrants further research in a larger cohort. The difference in age of onset between patients with Lynch spectrum extracolonic and colonic tumours could not be determined due to small sample sizes which would confer limited statistical power (data not shown).

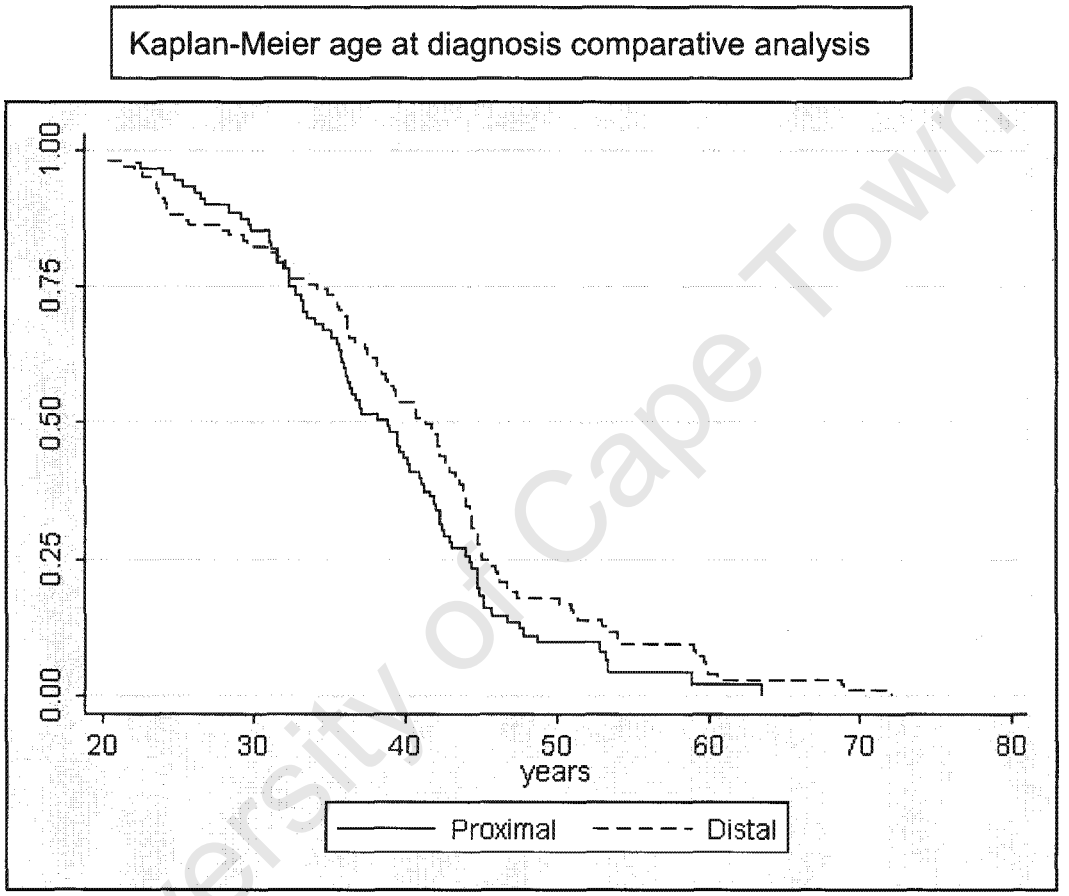


Figure 3.4.1B: Kaplan-Meier age at diagnosis comparison of CRC patients according to the site of the tumour (Log-Rank test; p=0.057). Age at diagnosis depicted on x-axis in years; proportion of patients surviving at one time depicted on y-axis. The legend "proximal" indicated tumours of the ascending colon, hepatic and splenic flexure and transverse colon; "distal" indicates tumours of the rectum, sigmoid and descending colon.

As previously mentioned, the medical records of the majority of the patient cohort could not be traced successfully. Although some clinico-pathological data and intermittent follow-up was recorded in the patient registry, no comprehensive follow-up data was available to indicate whether they received chemotherapy, or whether cancer recurrence had occurred. These

patients were classified as of "unknown status" and were excluded from the analysis. Despite having accessed the medical records of 80 patients (*cohort B3, Figure 3.1*), only 58 individuals could be included in the survival analysis (*cohort C, Figure 3.1*). The 21 patients who were excluded had not received curative surgery and hence failed to meet the criteria for inclusion. Instead, these patients had either received examination under anaesthesia with no other action taken, or a biopsy was taken of the tumour/s for diagnostic purposes only.

Kaplan-Meier recurrence-free survival analysis was performed in 59 individuals based on their mutation status in order to ascertain if a patient's mutation status determines the time until recurrence of disease. No significant difference was observed between Mut+ and Mut- individuals (p=0.267, Figure 3.4.1C).

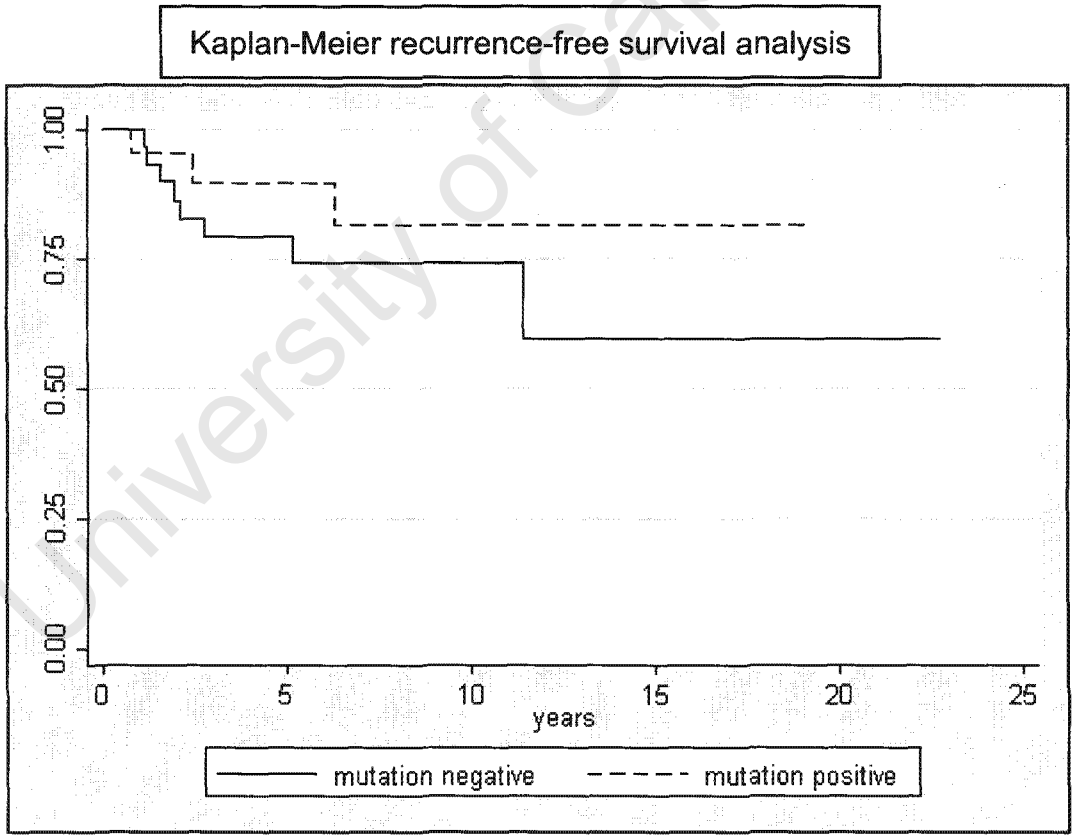


Figure 3.4.1C: Kaplan-Meier recurrence-free survival curve comparing CRC patients with known MMR status. The dotted line indicates patients with mutations in any of the MMR genes, whilst the solid line represents patients with no known mutation in any of the MMR genes. Recurrence-free time on x-axis in years, proportion of patients surviving at one particular time on y-axis.

Kaplan-Meier recurrence-free survival analysis was performed in 36 MMR mutation-negative (Mut-) CRC patients in order to test the null hypothesis that there is no significant difference in recurrence-free survival between subjects with (n=14), and without (n=22) chemotherapeutic intervention (*cohort C, Figure 3.1*). Analysis with the Log-Rank test showed that there was no appreciable difference in recurrence-free survival times (p=0.717, Figure 3.4.1D).

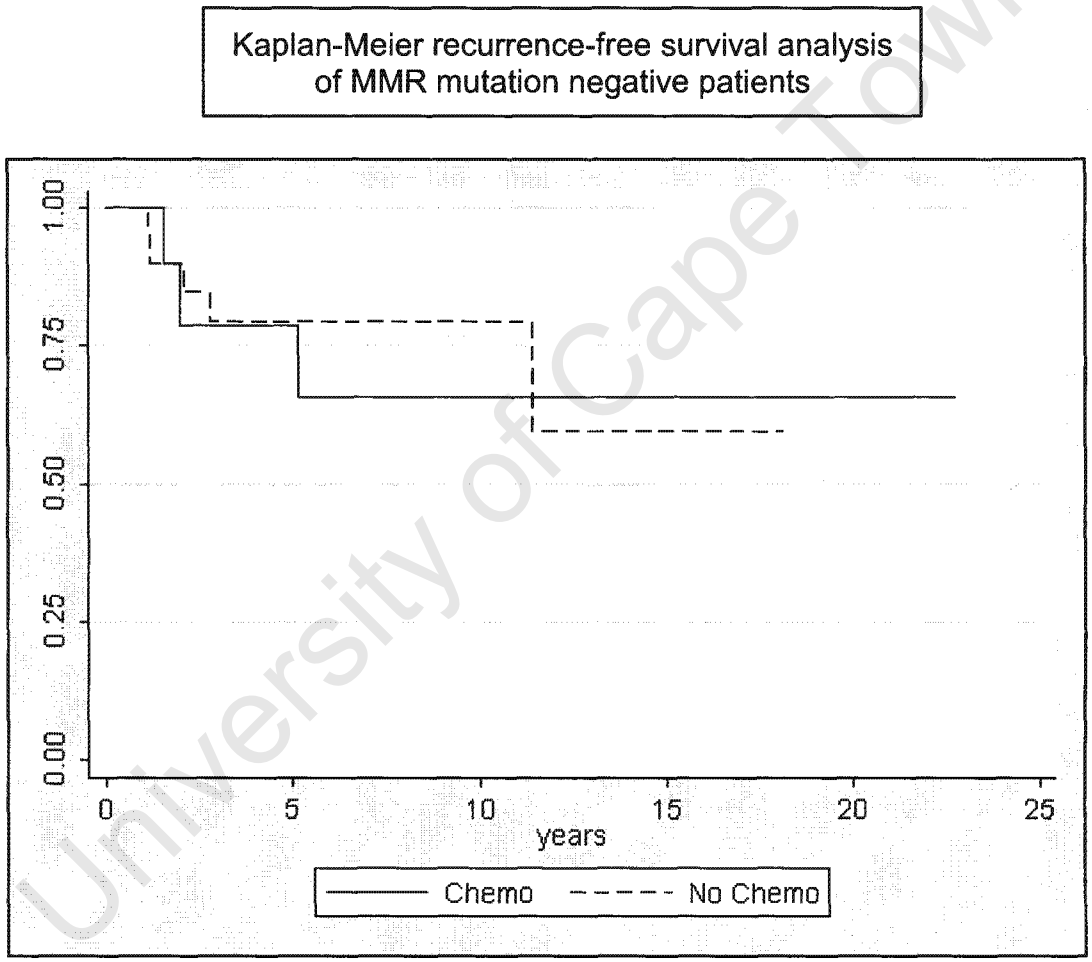


Figure 3.4.1D: Kaplan-Meier recurrence-free survival curve of mismatch repair mutation negative patients with and without chemotherapeutic intervention (Log-Rank test; p=0.717). Recurrence-free survival time in years depicted on x-axis. Proportion of patients surviving at one time depicted on y-axis. Patients who had received chemotherapy indicated by solid line, patients who did not receive any chemotherapy represented by dotted line.

Additionally, Kaplan-Meier recurrence-free analysis was performed in 22 MMR mutation-positive (Mut+) CRC patients in order to test the null hypothesis that there is no significant difference in recurrence-free survival between subjects with (n=9), and without (n=13) chemotherapeutic intervention (*cohort C, Figure 3.1).* The results of a Log-Rank test showed that there was no difference (p=0.881) in recurrence-free survival times (Figure 3.4.1E).

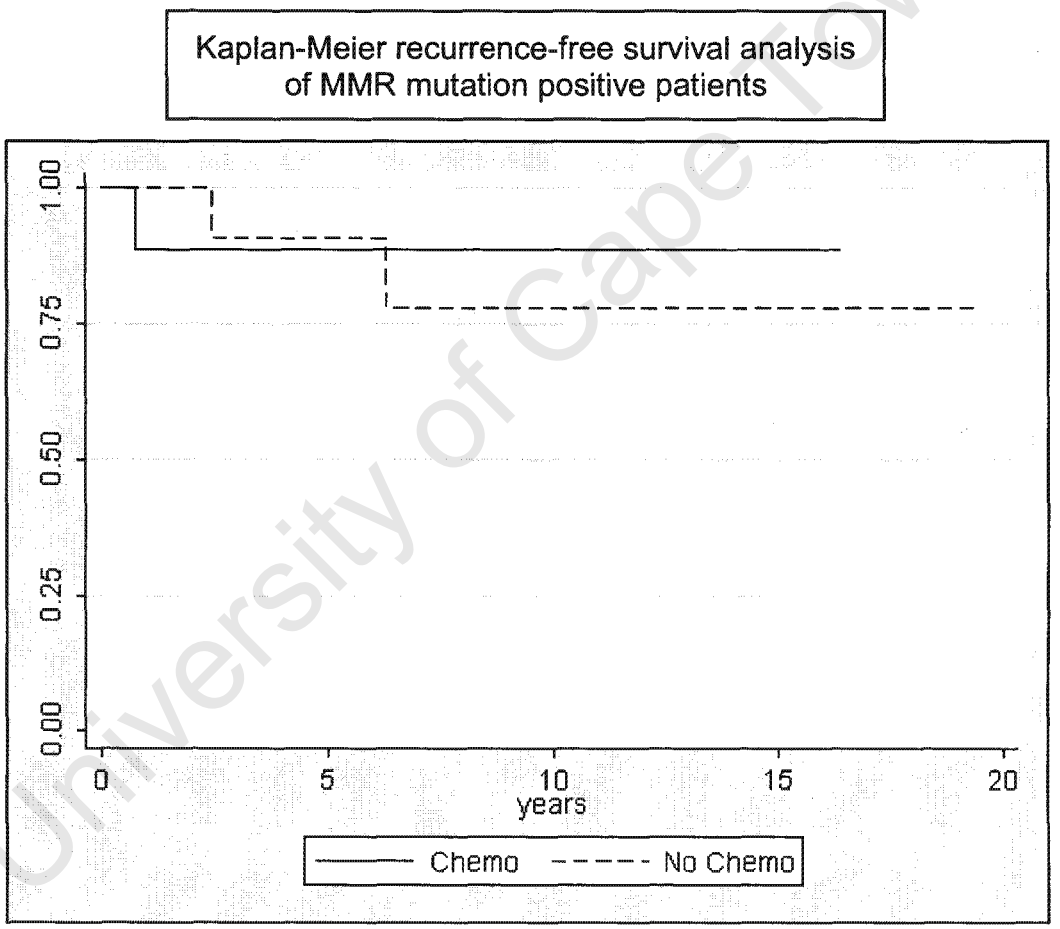


Figure 3.4.1E: Kaplan-Meier recurrence-free survival curve of mismatch repair mutation positive patients with and without chemotherapeutic intervention (Log-Rank test; p=0.881). Recurrence-free survival time in years depicted on x-axis, proportion of individuals surviving at one time depicted on y-axis. Patients who had received chemotherapy indicated by solid line; patients who had not received chemotherapy indicated by dotted line.

3.4.2 *DPYD* and *TYMS* haplotype analysis

DPYD and *TYMS* haplotypes were inferred in the MA (n=133 [*DPYD*] and n= 143 [*TYMS*]), CAU (n=27) and the BKA (n=16 [*DPYD]*, n=20 [*TYMS*]) study populations using PHASE, version 2.1 (Stephans *et al.* 2001; Stephens and Scheet 2005). PHASE uses linkage disequilibrium patterns to infer a missing locus, based on an observed locus. For *DPYD,* missing loci could only be inferred for patients with genotype data at one or two out of the three loci. Hence 55 individuals from the initial 231 (*cohort B2*) were excluded for *DPYD* haplotype analysis. Similarly, 41 individuals had to be excluded for *TYMS* haplotype analysis. A number of output files were produced during the course of PHASE analysis, and these are presented as supplementary files (Addendum C). A summary output file contained a list of the best haplotype guess for each individual, which were used in subsequent analyses. Some phases and alleles were difficult to infer, and indicated appropriately by the software. The software, moreover, estimated the various haplotype frequencies for the sample populations, which act as indicators of the greater population haplotype frequencies.

Eight haplotypes were deduced for *DPYD* in the MA and CAU study populations, but only four haplotypes could be inferred in the Black Ancestry BKA population. In the BKA study population, all four haplotypes occurred at a frequency greater than 5% and no distinguishing haplotype could account for the variation in *DPYD* within the population. Of the eight possible haplotypes in MA and CAU, three and four haplotypes, respectively, account for the majority of the variation observed and occur at a frequency greater than 5%. The haplotype which reportedly confers the greatest risk for toxicity (ATC or IVS14+1A—R29C—I543V) appears at very low frequencies in both CAU (0.008) and MA (0.0009), and was absent in the BKA group.

Similarly, four to six haplotypes were inferred for *TYMS* in the BKA, MA and CAU study populations, respectively. The *TYMS* haplotype which purportedly bestows the highest risk for decreased tumoral TS levels, and hence non-responsiveness to 5-FU therapy (2R6- or two tandem repeats and

a deletion allele) occur at low frequencies in each population group: BKA (0.0590), MA (0.0569) and CAU (0.0549). The estimated haplotype frequencies for each of the genes in each of the study populations are presented in a series of figures (Figures 3.4.2A-F).

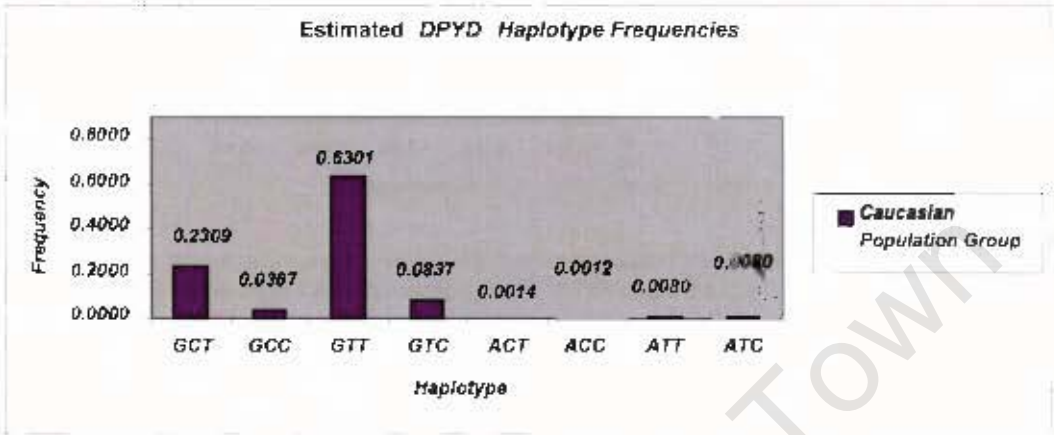


Figure 3.4.2A: A histogram depicting the estimated haplotype frequencies for *DPYD* in the Caucasian population. The SNPs are depicted in the haplotype order IVS14+1G>A — R29C — I543V.

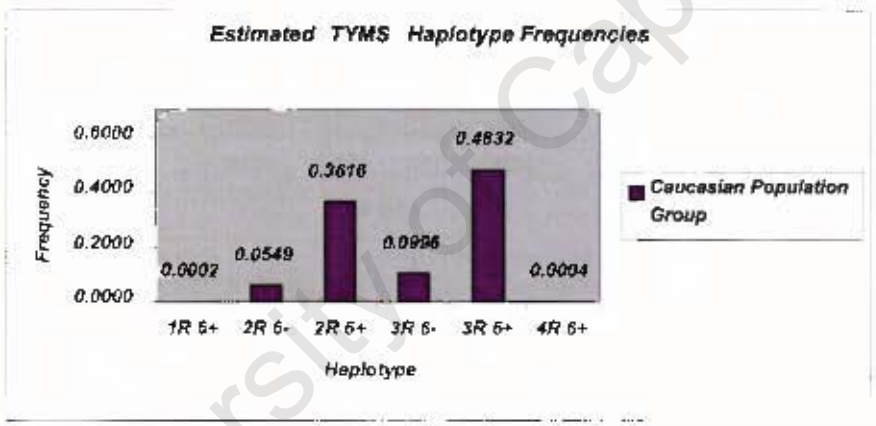


Figure 3.4.2B: A bar graph illustrating the estimated haplotype frequencies for *TYMS* in the Caucasian population. *1R6+* depicts the 1 repeat allele of the 5' VNTR and the insertion allele of the 3' indel.

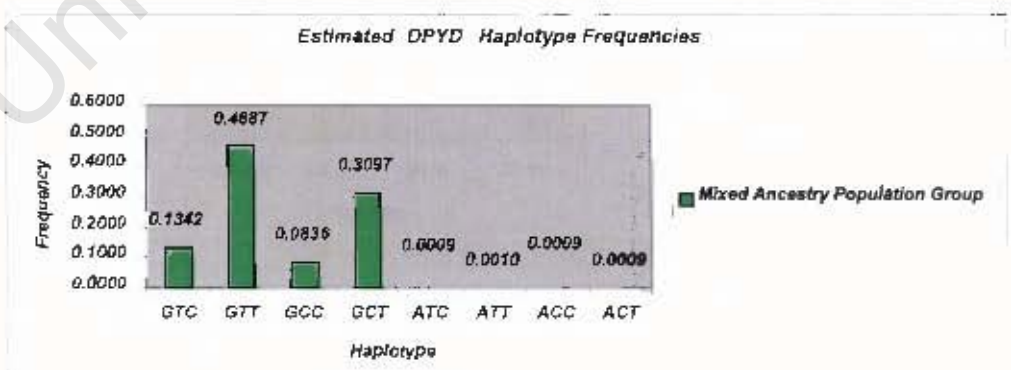


Figure 3.4.2C: A histogram depicting the estimated haplotype frequencies for *DPYD* in the Mixed Ancestry population. The SNPs are depicted in the haplotype order IVS14+1G>A — R29C — I543V.

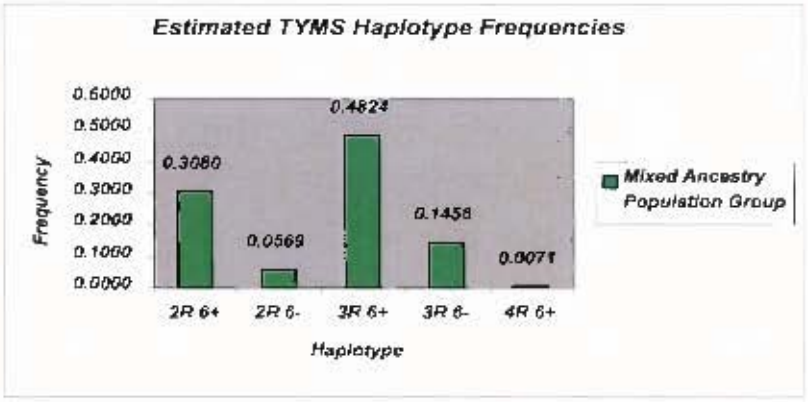


Figure 3.4.2D: A bar graph illustrating the estimated *TYMS* haplotype frequencies in the Mixed Ancestry population. *2R6+* describe the 2 repeat allele of the 5'VNTR and the insertion allele of the 3'indel.

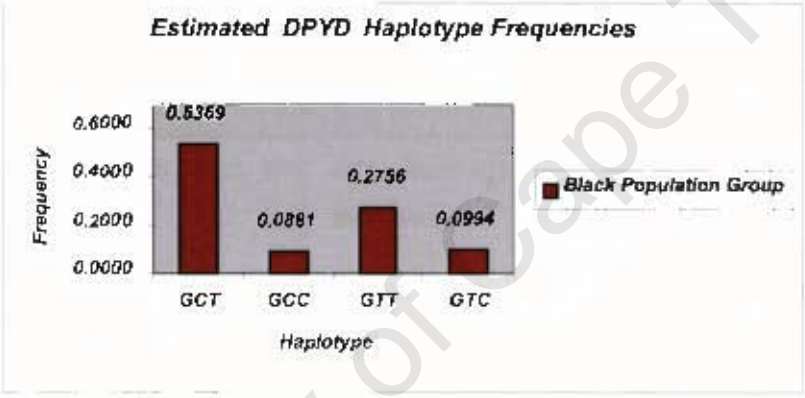


Figure 3.4.2E: A histogram depicting the estimated *DPYD* haplotype frequencies in the Black Ancestry population group. The SNPs are depicted in the haplotype order IVS14+1G>A — R29C — I543V.

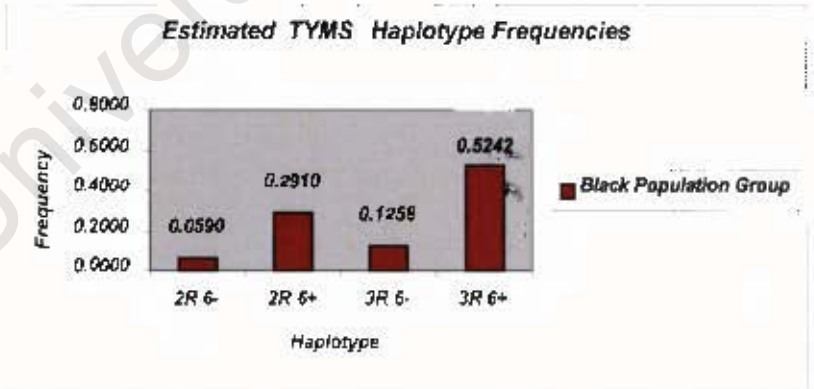


Figure 3.4.2F: A bar graph illustrating the estimated *TYMS* haplotype frequencies in the Black Ancestry population group. *2R6-* describe the 2 repeat allele of the 5'VNTR and deletion allele of the 3'indel.

A comparison of the derived haplotype frequencies in *TYMS* and *DPYD* for each of the study populations is presented in Figures 3.4.2G and 3.4.2H.

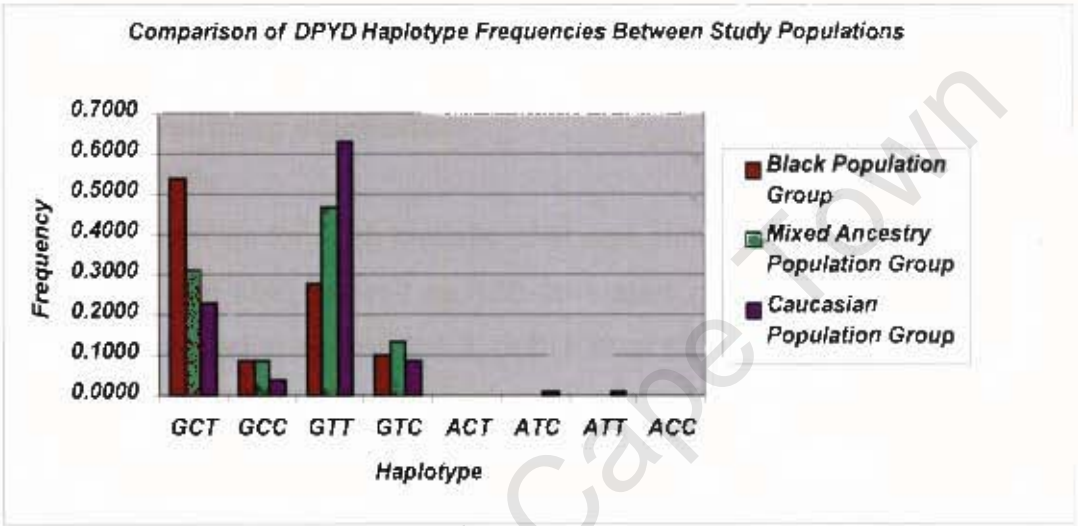


Figure 3.4.2G: A comparative histogram illustrating the differences in haplotype frequencies of *DPYD* between the relevant study populations.

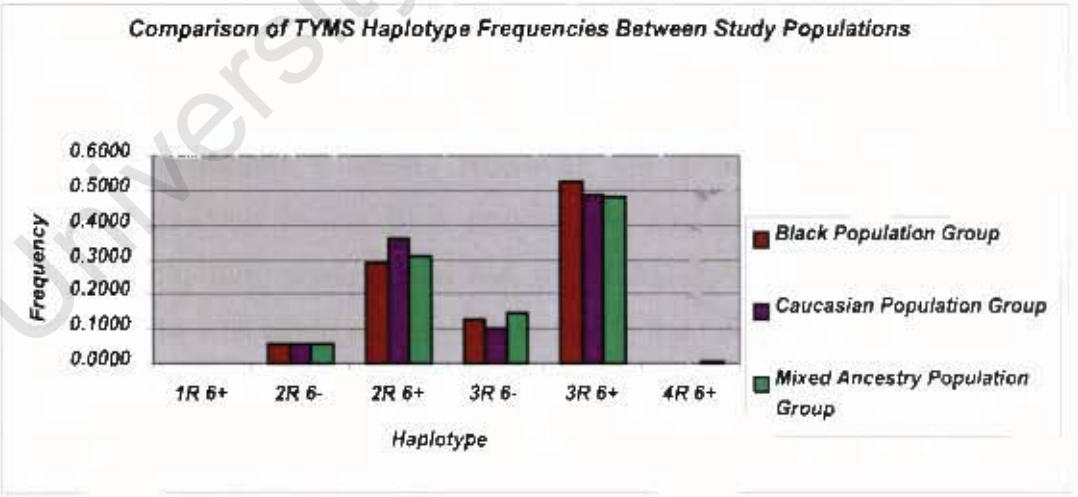


Figure 3.4.2H: A comparative illustration of the differences in *TYMS* haplotype frequencies between the relevant study populations.

Additionally a case-control permutation test was performed in order to test the null hypothesis that the case and control haplotypes in the study populations are more similar than different. No significant differences were observed between the derived *DPYD* haplotype frequencies of the cases and controls in either MA (p=0.95), or CAU (p=0.18). Similarly, the estimated *TYMS* haplotype frequencies for MA and CAU showed no significant deviations (p=0.11, p=0.4 respectively).

3.4.3 Allele frequencies of pharmacovariants in study populations and Hardy-Weinberg equilibrium

DPYD SNPs

The allele frequencies for each biallelic SNP was estimated in the 100 control subjects (CAU and MA) as well as 130 unrelated colorectal cancer (CRC) cases. As mentioned in section 2.4.3, both Fisher's Exact and Pearson's chi-square require that individuals are unrelated, however some of the subjects in the CRC study populations were part of large families, and thus only unrelated individuals were included in the analysis. Genotypes were incomplete for 12 of the individuals as a result of degraded DNA which hindered amplification of the variants. Table 3.4.3a outlines the respective allelic frequencies. For rs1801159 (I543V), the 'T' allele occurred in the majority of the patients. The missense 'C' allele appeared at low frequencies in CAU, BKA and MA population groups (0.12, 0.19 and 0.19 respectively). In contrast, the major allele in both CAU and MA populations for rs1801265 (R29C) was the mutant 'T' allele. However, the 'T' allele of this SNP was infrequently observed in the BKA population, and the frequency differed significantly from that observed in MA and CAU ($p<0.01$, Pearson's Chi^2 test). The IVS14 +1A allele was only detected in the CAU and MA population groups at a very low frequency (0.02 and 0.01 respectively).

There was no appreciable divergence in allele frequencies between the CRC cohort and control individuals, for both MA and CAU population groups, for the R29C variation (p=0.77 and 0.73 respectively, Pearson's Chi^2). The same applies to the I543V sequence variation (p=0.87 and 0.07 respectively,

Pearson's Chi^2). No difference in allele frequencies were observed for the IVS14+1G>A mutation between each of the populations (p>0.05, Fisher's Exact) as well as the control and CRC patient cohort in the MA (p=0.63, Fisher's Exact) and CAU population groups (0.35, Fisher's Exact). The allele frequencies for BKA control individuals were not established.

Table 3.4.3a: Observed allele frequencies of *DPYD* SNPs in Caucasian, Caucasian control, Black Ancestry, Mixed Ancestry and Mixed Ancestry control populations

	Mutation	Allele	Amino acid change	Allele Frequency	
				CRC patient cohort	Control cohort
Caucasian (n=28)	rs1801159	**C**	I543V	0.12	0.24
		T	I543I	**0.88**	0.76
	rs1801265	C	R29R	0.28	0.23
		T	R29C	**0.72**	0.77
	rs3918290	G	-	**0.98**	1
		A	Abolished splice site	0.02	ND
Mixed Ancestry (n=86)	rs1801159	C	I543V	0.19	0.2
		T	I543I	0.81	0.8
	rs1801265	C	R29R	0.37	0.39
		T	R29C	0.63	0.61
	rs3918290	G	-	0.99	1
		A	Abolished splice site	0.01	ND
Black Ancestry (n=16)	rs1801159	C	I543V	0.19	N/A
		T	I543I	0.81	N/A
	rs1801265	C	R29R	0.62	N/A
		T	R29C	0.38	N/A
	rs3918290	G	-	1	N/A
		A	Abolished splice site	ND	N/A

ND — "not detected"

N/A — "not applicable"

CRC — colorectal cancer

TYMS 5'UTR VNTR

Several alleles are possible for this variant. The frequency distribution in all populations is presented in Table 3.4.3b. The majority of the patients in all ethnic groups carried the 2R and 3R allele. The 4R allele was not identified in any population other than MA and heterozygous with the 2R allele. The observed frequency of the 4R allele in the MA population is 1.0%. Allele frequencies were established for 140 out of 142 individuals. The genotypes of two individuals were undefined due to poor DNA quality, and hence excluded from the analysis.

Table 3.4.3b: Observed *TYMS* 5'UTR VNTR allele frequency in Caucasian, Caucasian control, Black Ancestry, Mixed Ancestry and Mixed Ancestry control populations

Population	Number of individuals in population group	Allele Frequency		
		2R	*3R*	*4R*
Caucasian	27	46,3%	53,7%	ND
Caucasian control	47	41,5%	58,5%	ND
Mixed Ancestry	93	33,9%	65,1%	1,0%
Mixed Ancestry control	48	31,25%	66,6%	3,125%
Black Ancestry	20	35%	65%	ND

ND — "not detected"

2R — Two tandem repeats; 3R — Three tandem repeats; 4R — Four tandem repeats

TYMS 3'Indel

Table 3.4.3c shows the frequency distribution of the alleles for the 3'indel in each of the study and control populations. The frequency of the insertion allele (6+) was 0.85, 0.82 and 0.82 in the CAU, MA and BKA patients respectively. The deletion allele was in the minority and only two patients of MA were homozygous. There was no significant difference in allele frequencies between populations ($p>0.05$, Pearson's Chi^2), nor between control and CRC patients from the CAU population ($p=0.15$, Pearson's Chi^2). However, within the MA population group, a trend towards a significant difference was noted between the 6+ frequencies of the control population

and the CRC population (p=0.056, Pearson's Chi^2). The frequency of the 3'UTR indel could not be established in BKA control populations.

Table 3.4.3C: Observed *TYMS* 3'UTR deletion in Caucasian, Caucasian control, Black Ancestry, Mixed Ancestry and Mixed Ancestry control populations

	Allele	Allele Frequency	
		CRC	Control
Caucasian (n=27)	6+	0.85	0.75
	6-	0.15	0.25
Mixed Ancestry (n=93)	6+	0.82	0.71
	6-	0.18	0.29
Black Ancestry (n=20)	6+	0.82	N/A
	6-	0.18	N/A

6+ — insertion allele; 6- — deletion allele

N/A — "not applicable"

CRC — colorectal cancer patient cohort

No significant difference in genotype frequencies were discovered between observed and expected values under Hardy-Weinberg equilibrium for two of the SNPs analysed in *DPYD*, SNP rs1801159 (I543V) and rs1801265 (R29C) for CAU, MA and BKA individuals (p>0.05; Pearson's Chi^2; Table 3.4.3d). SNP rs3918290 (IVS14 +1G>A) was in Hardy-Weinberg equilibrium for only CAU and MA individuals, but deviated for BKA patients. Even after Bonferroni correction (p<0.125), the deviation was still significant.

The 3'UTR *TYMS* indel (rs16430) did not depart from Hardy-Weinberg equilibrium (p<0.05, Pearson's Chi^2) in the MA population, but deviated in the CAU group (p=0.03, Pearson's Chi^2). In 100 control individuals of CAU and MA, the IVS14 +1G>A SNP showed significant divergence (p<0.05, Pearson's Chi^2) in both population groups, even after Bonferroni adjustments were made. The 3' indel showed a similar departure from Hardy-Weinberg equilibrium, but this was restricted to the MA population group.

Table 3.4.3d: p-values calculated for each variant in *DPYD* and *TYMS* to test for deviation from Hardy-Weinberg equilibrium

		Population Group	P-value from Pearson's Chi Square	Significant after Bonferroni adjustment (n=4; p<0.0125)
DPYD	rs3918290 IVS14 +1G>A	CAU	0.92	NO
		CAU control	**<0.001**	YES
		BKA	**<0.001**	YES
		MA	0.95	NO
		MA control	**<0.001**	YES
	rs1801159 I543V	CAU	0.449	NO
		CAU control	0.412	NO
		BKA	0.355	NO
		MA	0.15	NO
		MA control	0.81	NO
	rs1801265 R29C	CAU	0.29	NO
		CAU control	0.06	NO
		BKA	0.789	NO
		MA	0.9	NO
		MA control	0.37	NO
TYMS	rs16430 3' Indel	CAU	**0.003**	YES
		CAU control	0.366	NO
		BKA	0.34	NO
		MA	0.44	NO
		MA control	**0.01**	YES

CAU: Caucasian

MA: Mixed Ancestry

BKA: Black Ancestry

3.4.4 *In silico* comparative analysis of allele frequencies of pharmacoSNPs in the drug metabolising genes *TYMS* and *DPYD* in indigenous African populations and Caucasian populations.

Using Fisher's Exact Test, the allele frequencies of 298 SNPs in the pharmacogenes *TYMS* and *DPYD* were compared between indigenous Black African populations (BKAF), and The HapMap European Caucasian population group (CEPH), to calculate the significance of the difference between all possible population pairings (data available as supplementary data, Addendum C). When a particular SNP is compared between two populations, the p-value cut-off of 0.05 gives an indication of whether the difference in allele frequencies between the given populations are significant. Table 3.4.4 summarises the number of SNPs that were consistently significantly different between every one of the five BKAF and the CEPH population. As evidenced by the data, there is more variation harboured in *DPYD* than in *TYMS* in the BKAF populations. The allele frequencies of a single SNP in *TYMS* was significantly different between the BKAF and CEPH populations. SNP rs2612095 is an intronic variation, located between the pharmacovariants of interest, rs16430 and rs45445694 studied in section 2.3.1 (Figure 3.4.4A).

The variation contained in *DPYD* was more extensive (Figure 3.4.4B). A total of 89 SNPs showed a significant difference in allele frequencies ($p<0.05$) and 40 SNPs registered a more substantial difference ($p<0.001$), as indicated in Table 3.4.4. According to Ensembl, all of the SNPs are located in intronic regions, and their functions are unknown. Moreover, these SNPs do not tag any of the variants that were studied in the South African CAU, BKA and MA populations *viz.* rs 3918290, rs1801265 and rs1801159 (section 2.3.1).

Table 3.4.4: Extent of variation harboured in *DPYD* and *TYMS* in indigenous African (BKAF) populations

Gene	Chromosome	Gene Size	Total number of SNPs in gene	Number of SNPs differing significantly between BKAF and HapMap CEU (% of total)		Number of SNPs compared (%)
				<0.05	<0.001	
DPYD	1		2613	89 (3.4%)	40 (1.53%)	280 (10.7%)
TYMS	18		235	1 (0.42%)	1 (0.42%)	18 (7.66%)

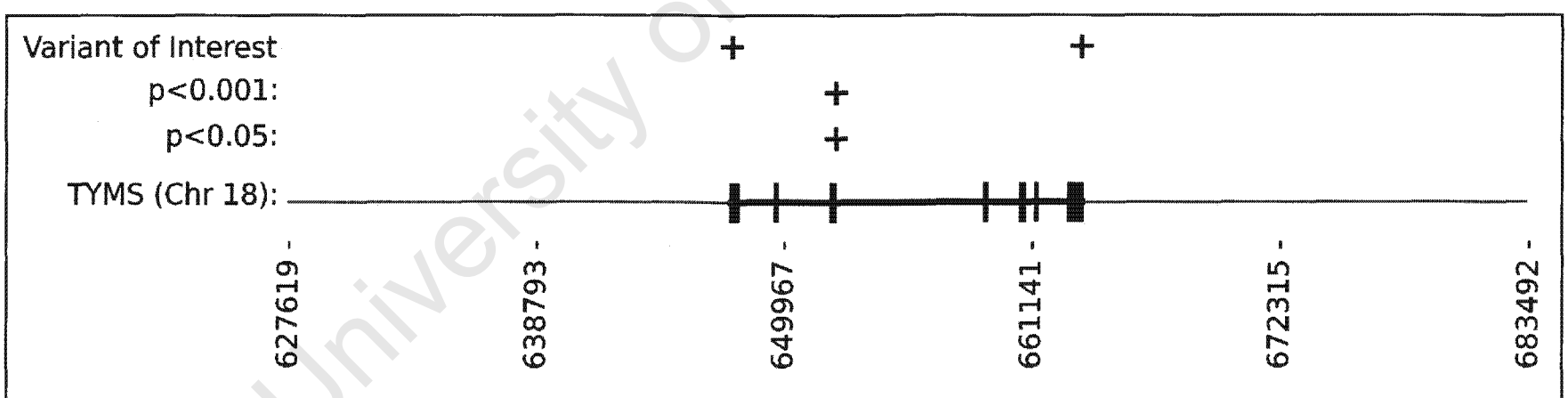


Figure 3.4.4A: Graphic illustration of the position of the variants of interest in *TYMS* (rs16430, rs45445694) and the SNPs which have significantly different allele frequencies (p<0.05, p<0.001) between the indigenous African populations and The HapMap Caucasian population group. The positions of the variants are indicated by +, and their base-pair position on chromosome 18 is shown on the X-axis. The black, shaded horizontal line represents the *TYMS* gene, and the seven vertical bars indicates the seven exons in *TYMS*.

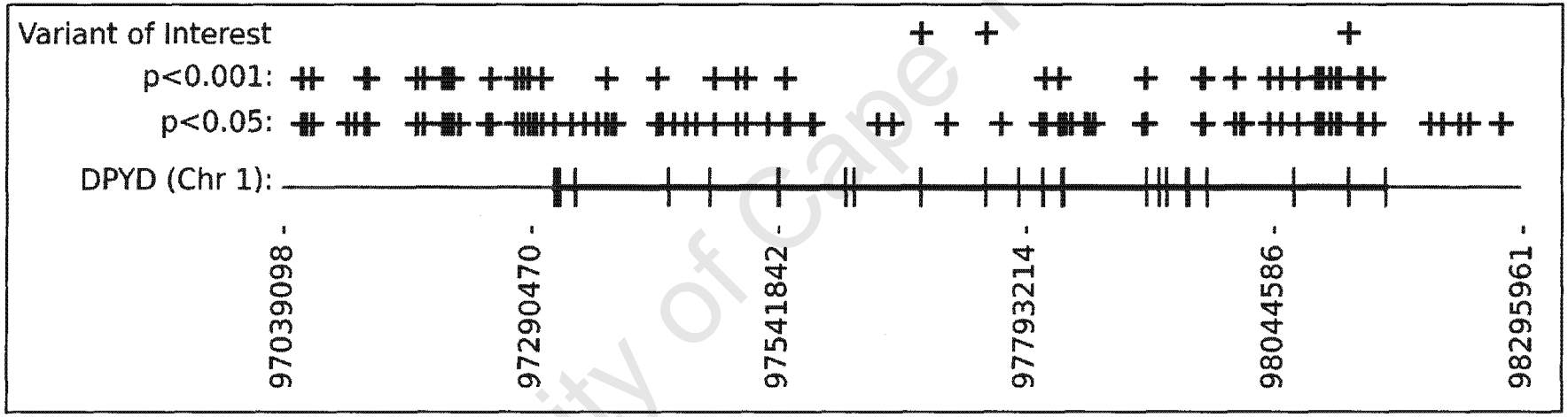


Figure 3.4.4B: Graphic illustration of the position of each of the variants of interest (rs3918290, rs1801265, rs1801159) in *DPYD* and the SNPs which show a significant difference ($p<0.05$; $p<0.001$) in allele frequencies between the indigenous African populations and The HapMap Caucasian population group. The positions of the SNPs are indicated by +, and their base-pair positions on Chromosome 1 are shown on the X-axis. The black, shaded horizontal line indicates the position of the *DPYD* gene on chromosome 1, and the vertical bars represents the exons.

CHAPTER 4
DISCUSSION

The Division of Human Genetics at The University of Cape Town (UCT) has performed genetic studies on individuals affected with colorectal cancer, since the late 1980s. Currently, demographic, clinical and follow-up data, and biological material is archived for 450 probands, and their relatives, who are either affected, or at risk for disease (comprising a total of approximately 1800 subjects). The information about these individuals is contained in a computerised patient registry located in the Division.

In order to correlate disease and clinical outcomes with genetic data (i.e. to do with the primary predisposing mutation, and/or variants influencing chemotherapeutic outcomes) methods/protocols for accessing patient records and biomaterial, and the establishment of molecular genetic assays was required. Several academic and clinical units are responsible for managing each of the aspects of patient care, with their respective peculiarities, including challenges and shortcomings. This required a critical survey of the referral system at Groote Schuur Hospital and UCT, in order to identify flaws, suggest solutions and, where possible, to rectify the problems.

The present pilot study, which compared a range of features between cancers due to defects in Mismatch Repair (MMR) genes (Mut+), and those that were not due to MMR defects (Mut-), identified the opportunities and challenges for setting up such studies.

It was determined that cancers of the distal colon appeared more frequently in Mut- patients, whereas Mut+ patients presented with cancers of the proximal colon more often ($p<0.001$, Fisher's Exact). This observation is in line with international literature regarding a predominance of proximal tumours in Lynch syndrome when compared to sporadic CRC (Lynch *et al.* 1993). Since the majority of the cohort entering the research registry are generally recruited under the criterion of 'cancer under the age of 50 years',

there was no attempt made at deducing whether a greater proportion of younger individuals with CRC are likely to have mutations in the MMR genes. The Dukes' stage of a tumour at diagnosis was not contingent on mutation status (p=0.57; Fisher's Exact).

However, this work has also shown that there is a disparity in terms of the amount and quality of information pertinent to patients, at each point of care, ultimately meaning that patients may not benefit from the potential of 'genetic studies/testing', in the current setting, if and when such genetic information is shown to be of consequence and value.

DNA and diagnostic information (i.e. diagnosis with CRC) was available on each of the 240 subjects in the registry, and for whom mutation analysis was carried out. The substantial attrition of subjects between diagnosis and surgery for CRC, to treatment with e.g. 5-FU, is partially explained by the fact that only a proportion of individuals diagnosed with disease, actually end up being further treated with chemotherapy. Even for individuals, who might have been treated with chemotherapy, in order to assemble the information pertaining to treatment and outcomes, each individual patient had to be actively consented, before their 'treatment-folder' could be physically interrogated. The ultimate smaller cohort of Mut+ and Mut- individuals who were compared with regard to 5-FU treatment and outcomes reflected individuals for whom there was reliable information available from admission, through surgery and pathology, and ultimately for their oncological chemotherapeutic treatment and follow-up.

In order to capture the data for future studies, an empirical form was drawn up onto which the relevant data, regarding treatment modalities and outcomes, could be entered. The data on the form was to be captured onto the patient database, and would negate the need for arduous paperwork for anybody interested in this data in future. A copy of the form is shown in Addendum B.

This study successfully improved the current patient database by adding very relevant fields pertinent to treatment and outcomes (where patient information was accessible). Hence, despite an attrition in information, between original patient admission to eventual discharge and follow-up, this project ultimately assisted in improving the process of data capturing in the Division. So, currently, when a newly-recruited patient enters the research/screening programme of the Division, not only is a research blood sample collected, but an extensive clinico-pathological history is required, and information from the patient's radiotherapy (RT)-folder, if the patient has received additional chemotherapeutic treatment, is captured on the database. A research assistant has been appointed to actively retrieve new and/or outstanding RT-folders from Groote Schuur Hospital and the relevant pathology, treatment schedules and treatment outcomes are being entered into the database and monitored on a regular basis. This will facilitate more comprehensive studies on this valuable cohort in future

The completion of the sequencing of the Human Genome has paved the way for individualising cancer treatment. Several enzyme systems are responsible for metabolising chemotherapeutic agents, and it is assumed that variations in the genes encoding these enzymes might influence response to treatment. Toxicity from, and resistance to, chemotherapy could be explained by differential expression of metabolising genes, and genetic profiles of patients can potentially aid in discovering effective treatment options (Stoelmacher *et al.* 2002).

Oncologists in South Africa (SA) do not use mutation status to guide therapy. However, there are several prognostic genetic tests currently available in the USA, which show promise in changing the way drugs are prescribed. These tests are limited in as much as they are developed to analyse polymorphisms identified in populations outside of SA. In order for hospitals in SA to adequately utilise these tests, the frequency of the variant alleles in populations in SA would have to be determined, and thus, whether these tests would be appropriate to determine prognosis in SA populations. Hence,

another aspect of this pilot study was formulated to develop pharmacogenetic assays which could be applied in future large-scale studies, to determine whether variants in drug-metabolising genes determine the efficacy of 5-FU in different populations in SA. In this regard, this project established (i) a range of pharmacogenetic tests pertinent to treatment with 5-FU, and (ii) the frequencies of polymorphisms in the genes known to influence 5-FU toxicity and metabolism in the subject population.

Although a substantial subject cohort was interrogated for the purpose of establishing population frequencies of pharmacogenetic alleles in the respective genes, a relatively small proportion of the patients entering the system for genetic studies were ultimately triaged for such studies, as explained previously. This cohort reflected individuals for whom there was reliable information available from admission, through surgery and pathology, and ultimately for their oncological chemotherapeutic treatment and follow-up. The population and patient pilot cohort were studied at five variants within the 5-FU pharmaco-relevant genes, *thymidylate synthase (TYMS)* and *dihydropyrimidine dehydrogenase (DPYD).* None of the abovementioned polymorphisms have previously been screened in a diverse population such as in SA.

i) *Dihydropyrimidine dehydrogenase* gene *(DPYD)*

The correlation between genotype and phenotype are still lacking for the majority of known *DPYD* mutations, and these may only be common polymorphisms which do not influence enzymatic activity. However, some mutations are functional and result in an inactive protein. The current data reports the frequency of three SNPs in 179 SA Caucasian (CAU), Black (BKA) and Mixed Ancestry (MA) CRC patients (of the original cohort of 240 patients), as well as in 100 control CAU and MA subjects from the background population. The inability to reliably genotype the remaining 61 patients for all three polymorphisms in *DPYD* may be due to factors such as low quality lymphocytic DNA samples.

A G>A point mutation in intron 14 (IVS14 +1G>A), abolishes a 5'-splice donor recognition sequence, resulting in a non-functional protein (Wei *et al.* 1996). The IVS14 +1G>A mutation is purportedly the most common cause of protein deficiency and normally occurs at a frequency of ~1% in the general German CAU population. A large number of German cancer patients (25%) suffering from severe 5-FU-related toxicities, have been reported with this mutation (Raida *et al.* 2001). Similar to other studies, the mutation appeared in the present study in the heterozygous state in only 2% of the SA CAU patients, and in only 1% of MA patients. No patients of BKA carried this mutation. None of the patients who had received chemotherapy carried this mutation, and no excessive side effects were reported. A significant ($p<0.05$) divergence from Hardy-Weinberg equilibrium (HWE) was detected in the genotype frequencies for this mutation in the background CAU and MA population groups. One would expect these control, outbred populations to adhere to HWE, and deviations may indicate errors in the data sets (Salanti *et al.* 2005). It is hypothesised that the power to detect the deviations might have been too low in the current study, due to the relatively small sample sizes.

There are conflicting results about the involvement of the c.85C>T (R29C) and c.1627A>G (I543V) variants with respect to toxic side effects from 5-FU treatment. It has been reported that a patient with a homozygous R29C mutation suffered from severe toxic side effects after 5-FU therapy (Vreken *et al.* 1997), and this was substantiated by other studies, further suggesting that the presence of the mutant allele leads to an enzymatically inactive protein (Vreken *et al.* 1998). On the contrary, these mutations have been suggested to be common innocent polymorphisms as evidenced by studies describing patients homozygous for the R29C mutation, but free from any side effects (Collie-Duguid *et al.* 2000; Van Kuilenberg *et al.* 2000). The amino acid positions, 29 and 543, are also reported to be evolutionarily unconserved (Mattison *et al.* 2002). The relatively high frequency of the R29C in patient cohorts is understandable in light of the high incidence of the R29C and I543V mutations in the German CAU background population

(0.194 and 0.137 respectively) (Gross *et al.* 2003). In the current study, there was no appreciable difference in allele frequencies between the CRC cohort and control individuals for both MA and CAU population groups for either the R29C, I543V and IVS14 +1G>A variants. This is to be expected, accepting that these polymorphisms are defined as common in the background population, and chosen specifically for their purported role in 5-FU metabolism. Although no individual difference was noted, the three variants, R29C, I543V and IVS14+1G>A may have a cumulative effect. Some ethnic variability was observed in the allele frequencies of the R29C variant between the study (affected) BKA, MA and CAU populations. The reference 'C' allele occurred much more frequently in the BKA population than in the CAU or MA groups ($p<0.05$, Pearson's Chi^2).

ii) *Thymidylate synthase* gene (*TYMS*)

In the present study, 192 CAU, MA and Black patients and 100 controls individuals were assayed for the 5'UTR VNTR polymorphism in the gene for thymidylate synthase (TS) which converts dUMP to dTMP. Of the initial cohort of 240, only 192 DNA samples were of sufficient quality for subsequent molecular assays. No novel VNTR alleles were identified in this study. Two or three tandem copies of the repeat (designated 2R/2R or 3R/3R respectively), or the heterozygous combination genotype (2R/3R) was found most often in the current study populations. However, the four-tandem repeat (4R) allele was identified in two patients from the MA population and found to be heterozygous with the 2R allele at a frequency of 1.0%. Although the 4R allele was not found in any other population, it is possible that this allele exists in these populations, but at a frequency too low to observe in a small scale study. Marsh *et al.* (2000) identified the 4R allele at a frequency of 2-7% in control Black African populations (originating mainly from Kenya and Ghana). Thus it was anticipated that this allele would occur more frequently in the current BKA study population. However, the BKA study sample size was too small to adequately expose the allele, if it does occur in the BKA population in SA. It has been suggested that the 4R allele occurs at approximately a 1% frequency (~1%) in British CAU populations (Marsh *et al.*

2000). Nevertheless, it was not seen in the SA CAU study cohort. If it does occur, it is very likely to be present at a substantially lower frequency (<0.5%), since it was not evident in the study cohort. It is certainly worthwhile pursuing a more accurate estimation of these allele frequencies in the various populations in South Africa particularly because of the purported role of copy number on drug metabolism and treatment outcome.

Marsh *et al.* (1999 and 2000) proposed up to nine tandem repeats of the VNTR in certain African and Asian populations. However, the role of ethnicity in response to TS targeted chemotherapy has not yet been evaluated. These repeats are enhancer elements and a 2.6 fold increase in expression has been reported in individuals harbouring three repeats, compared to two repeats (Kawakami *et al.* 1995; Kawakami *et al.* 1999). Over-expression of the TS protein in tumours has been associated with a poor prognosis. Lenz *et al.* (1998) showed that patients with increased tumoral TS levels had disease progression on 5-FU chemotherapy. In support of this, Pullerkat *et al.* (2001) demonstrated that individuals, who were homozygous for the double tandem repeat (2R/2R), had a better response to the chemotherapy than those harbouring two 3R alleles.

The 6-bp insertion/deletion (indel) polymorphism in the 3'UTR of *TYMS* was also examined in the 192 subjects mentioned above. Two 6-bp indel variations, separated by 3-bp on the *TYMS* genomic sequence, are currently published and curated *viz.* rs16430 and rs11280056 (TTAAAG/- and AAGTTA/-, respectively). Validation by cycle sequencing affirmed that the deletion observed in the population could theoretically be either of the abovementioned, due to the nature of the sequence in that area, and that they could not be distinguished using dHPLC analysis (refer to Table 3.3c). Further restriction enzyme analysis will have to be performed to verify which 3'indel was observed in the current cohort. For the purpose of result analysis, the variation is referred to as the 3'UTR indel.

The results show that the deletion allele (6-) occurs at a frequency of 0.18, 0.15 and 0.18 in the BKA (n=7/20), CAU (n=8/27) and MA (n=34/93) populations, respectively. This finding is in contrast with other studies which have shown the 6- allele to be present at a frequency of 0.41 in American CAU populations (n=63) (Mandola *et al.* 2004). It is evident that some ethnic variation of this polymorphism exists and that it is in strong linkage disequilibrium with the *TYMS* VNTR polymorphism (Mandola *et al.* 2004; Dotor *et al.* 2006). However, no ethnic variability was noted between the BKA, CAU and MA study population groups ($p>0.05$, Pearson's Chi^2).

Studies have shown that patients harbouring two deletion alleles have a four-fold decrease in *TYMS* mRNA expression, whereas the heterozygous genotype demonstrated intermediate levels of expression (Lecomte *et al.* 2004; Mandola *et al.* 2004). Lenz *et al.* (1998) proposed that patients with decreased tumoral TS levels would respond better to 5-FU, and this was corroborated by Dotor *et al.* (2006) who suggested that the deletion allele is protective since no deaths after treatment were observed in their homozygous deletion (6-/6-) study group (n=10/129). Although the function of this deletion remains unknown, it has been speculated that the 3'UTR is rich in microRNA coding sequences, which are known to play a role in mRNA stability and translation. In the present study only two patients out of a potential cohort of 231, were homozygous for the 6- allele. These patients had not undergone any chemotherapeutic treatment.

An additional dimension to better understand the potential variation in pharmacogenes in South African populations, presented itself from results of a recent study within the Division of Human Genetics (unpublished work, 2007). The Affymetrix GenomeWideSNP 6.0 Array (http://www.affymetrix.com/products_services/arrays/specific/genome_wide_snp6/genome_wide_snp_6.affx) was used to establish the genetic variation harboured in the pharmacogenes of five indigenous Black African population groups and to compare the data to that of the four populations in The HapMap Project (The HapMap Consortium, 2005). The SNP 6.0 chip

provides information on 906 600 SNPs and 946 000 Copy Number Variants (CNVs), per individual sample. The data was used to construct a local Biomart database with a Martview user interface, which allows a user to compare allele frequencies of a particular SNP within the nine populations.

By comparatively analysing the allele frequencies of several SNPs in *DPYD* and *TYMS*, it could be established which SNPs are significantly different and could affect treatment outcomes. In retrospect, the data from the Affymetrix GenomeWide SNP 6.0 Array should have been included at the outset of the study, in order to prioritise the SNPs to be studied in our SA populations. However, the data became available for use only at a later stage in the study, and was deemed to be valuable in integrating current research with related state of the art technology and analysis.

It was evident that more variation is present in *DPYD* in the indigenous Black African populations, since 40 SNPs showed a significant difference in allele frequencies ($p<0.001$) compared to only a single SNP in *TYMS* which was significantly different between the population groups. *DPYD* contains approximately 2613 published SNPs, of which 280 could be compared using the Affymetrix array (~11%). *TYMS* harbours as little as 235 SNPs, and 18 are captured and comparable on the array (~8%). If one takes this into consideration, *DPYD* still has a large number of SNPs (3.4% vs 0.42% in *TYMS*) which have significantly different allele frequencies. This suggests that *DPYD* may be an attractive candidate gene for future studies into the pharmacogenetic variation in Africa.

A significant difference in allele frequencies of a single SNP (rs2612095) in *TYMS* was detected between the indigenous Black African and European CAU populations ($p<0.001$). The SNP is an intronic variation with no discernible functional properties. This SNP may, however, still have an effect on the structure and conformation of the RNA transcript, or affect RNA splicing.

Unfortunately, not all SNPs in *DPYD* and *TYMS* were featured on the array, and as such the comparison was limited to only a subset of SNPs. Custom-made arrays provide an opportunity to design experiments which contain pharmacorelevant SNPs which are likely to be more applicable to SA populations, through biotechnology companies such as Illumina® and Affymetrix®. This would offer one the freedom of choosing appropriate SNPs for a specified study, and is highly recommended for future studies into pharmacogenomics in SA.

This present study was successful in its intent to evaluate if colorectal cancer (CRC) patients with a defective mismatch repair system (MMR) respond differently to the chemotherapeutic agent, 5-FU when compared to their MMR-efficient counterparts. For statistical analysis, the current study cohort was divided into two groups based on their mismatch repair status; mutation-positive (Mut+) and -negative (Mut-). Kaplan-Meier analysis of 36 Mut- CRC individuals determined that local or distant recurrence of cancer was not a function of chemotherapeutic intervention with 5-FU. In 14 patients who had treatment with 5-FU, there was no significant difference in recurrence-free survival compared to 22 patients without intervention ($p=0.643$). This is contrary to what was expected, given that 5-FU is the main agent used for treating CRC, and should increase the survival term of cancer patients (Grem 2000). However, this research is presented as a pilot study with relatively small cohorts, and hence these results should be considered with caution. It is important to note that Kaplan-Meier analysis requires large sample sizes for adequate statistical power. It is therefore essential that this trend be explored further in larger study groups.

Several studies attempted to elucidate whether the MMR status of a colorectal cancer cohort will define a response to 5-FU therapy (Ribic *et al.* 2003; Jover *et al.* 2006). These studies distinguished a tumour as MMR-deficient when a loss of eg. MLH1 protein expression or microsatellite instability (MSI) occurred. However, no difference in long term response or survival after 5-FU treatment between patients with MMR-proficient (n=330)

and MMR-deficient (n=23) tumours were detected (Ribic *et al.* 2003). Patients presenting with high MSI (MSI-H) (n=53) tumours, did not benefit from 5-FU treatment when compared to individuals with low MSI (MSI-L) (n=230) tumours. Conversely, in patients receiving no adjuvant treatment, those with MSI-H tumours (n=42) displayed longer overall survival than stage-matched sporadic cancer patients with MSI-L tumours or MSI stable tumours (n=245) (Jover *et al.* 2006). In the current study, Kaplan-Meier analysis of 22 Mut+ study participants (i.e MMR-deficient), nine of whom had undergone chemotherapeutic treatment, showed that there is no significant difference in recurrence-free survival times between patients with, and without 5-FU therapy (p=0.208). Thus, in this cohort, neither MMR-deficient patients, nor their MMR-proficient counterparts, seemed to benefit from 5-FU therapy. However, as previously discussed, these results should be assessed with caution, due to the small size of the study population.

The age at diagnosis (AAD) between patients with a particular stage or site of cancer was compared. Kaplan-Meier survival curve analysis exhibited no significant difference in the AAD (cancer-free survival time from birth) of disease, regardless of stage of cancer. In other words, patients were not diagnosed with Dukes' stage A CRC at a younger age than, e.g, stage C CRC patients. Moreover, patients did not present with tumours of the proximal colon (includes appendix, ceacum, ascending colon, hepatic flexure, transverse colon and splenic flexure) at an earlier age compared to patients with tumours of the distal colon (includes descending colon, sigmoid and rectum) (p=0.057). It should be noted that the p-value for this analysis approached significance, and therefore warrants further investigation in a larger study cohort.

Now that a formal system is in place, offering Lynch syndrome families genetic testing, counselling and clinical surveillance, it is likely that the surveillance process would identify Mut+ individuals with disease at an earlier stage than individuals affected with sporadic or non-Lynch CRC.

Limitations of this study

As mentioned, this study is considered and presented as a pilot, comparing clinoco-pathological features and treatment outcomes between Mut+ and Mut- patients and provided an opportunity to identify potential problems and suggest solutions for future studies. Due to the retrospective nature of the study, a significant proportion of data was missing, despite remedial action, and hence the results should be treated with caution.

The aim was to recruit subjects affected with CRC, who have been treated with 5-FU and to relate their subsequent response to their MMR status. Of the 240 eligible participants, screened for MMR mutations and being managed by the Division of Human Genetics, 38 (13 mutation-positive and 25 mutation-negative) individuals met the pre-set criteria. These criteria included providing consent for access to their treatment data, physical hospital folders which were accessible, and treatment with 5-FU.

The patient database in the Division of Human Genetics is kept current with relevant patient information by nurses, divisional staff members and students working on the project. Although genotyping data is reliable, patient and clinical data may be incomplete, and in some cases, erroneous. In order to obtain follow-up (RT) records for the patients, the relevant RT number or Groote Schuur Hospital (GSH) folder number were obtained. Despite a fully functioning colorectal cancer clinic at GSH, follow-up records and current status of the patients were inadequate or missing from the GSH Oncology storage facility for a significant proportion of the cohort. In some cases the absence of a particular folder could be explained by patients choosing to receive follow-up treatment at other hospitals or clinics, nationally, and thus inaccessible in the circumscribed time that the study was being conducted. Alternatively, some patients were receiving chemotherapy actively at a particular time, or receiving follow-up care, and thus their folders were kept on hand by the treating physician for the duration of the treatment. Having identified that only hard copies of patient-related data is the norm at GSH, a solution to this problem is to computerise all patient RT records. An interim

solution is to ensure that a liaison person attempts to track the physical files relevant to every patient entering the chemotherapy environment, and which has, as a result of this study, already been implemented.

Not all CRC patients at GSH receive chemotherapeutic treatment. Many patients undergo a curative resection of the tumour and recuperate without the need for 5-FU treatment. Of the 80 patient RT folders that were obtained, only 38 individuals (47.5%) had undergone treatment with 5-FU, of which three received palliative treatment for metastatic disease, and had to be excluded from the analysis. Often, radiotherapy is received instead of 5-FU. However, the number of individuals receiving radiotherapy and the number receiving 5-FU is not mutually exclusive.

It is clear that the treatment and outcome information that was accessed is an under-representation of the whole study cohort. It should be noted that remedial actions were attempted throughout the course of the study in order to maximise the primary cohort, and include as many CRC-affected patients as possible. However, the majority of the patients recruited into this current study are obliged to make use of the state-funded medical facilities at GSH. They do not have the financial capacity to receive treatment at private medical facilities, hence treatment folders are mainly accessed through GSH. Telephonic contact was attempted for patients and/or family members for whom no records were accessible. Nevertheless, a sizable proportion was either unapproachable, non-responsive or did not know their treatment details, and the subsequent information obtained was negligible.

The National Cancer Institute (NCI) supervises the endeavours against cancer in the USA. Results of approximately 21 criteria are regularly available as free domain on the internet. Only two of these criteria, namely, cancer incidence and mortality rates, are captured in SA by the Medical Research Council (MRC) and the National Health Laboratory Service (SA does not have an USA-NCI equivalent). However, it is recognised that the incidence of cancer is severely under-reported and hence unreliable.

According to the National Cancer Registry (NCR) and the MRC respectively, the incidence data and mortality data do not harmonise. The reported number of people dying from cancer is purportedly higher than the incidence (Albrecht 2006). It is furthermore speculated that the under-reporting of cancer cases is more extensive in Black South African population groups than in White South African groups. This is due to the disparity in clinical services between the different populations groups, as a result of past government policies. Therefore it is imperative that the NCR is informed of South Africans with unreported cancer diagnoses. Likewise, mortality rates in SA have been deemed erroneous by the World Health Organisation (Mathers *et al.* 2007) stating inaccurate death certification. The NCR is in no way responsible for these discrepancies, for they are only able to use the data that is being provided to them from medical institutions.

An important aspect of the study of Lynch Syndrome is the technical capacity to carry out microsatellite instability analysis. Although the designed markers were resolved in lymphocytic DNA, their amplification was problematic in paraffinised tumour tissue. It is likely that preservation artefacts, or age of stored samples could influence the outcome or attempts to analyse MSI status in preserved tumour tissue. In mitigation of the current project, however, the study cohort was chosen based on the basis of comprehensive germline MMR gene analysis, resulting in subcohorts of Mut+ and Mut- subjects, respectively. This is considered to be a much more reliable indication of Lynch predisposition and cause of cancer (i.e. Mut+). There is nonetheless, a caveat that at least a small proportion of the tumours in Mut+ patients could be a sporadic phenocopy.

Future prospects

Several prognostic tests are currently being marketed in the USA for a variety of chemotherapeutic agents. Irinotecan (IR) is a topoisomerase I inhibitor and a drug frequently used to treat colorectal cancer (CRC) under the trade name, Camptosar® *(Pfizer Pharmaceuticals Inc., http://www.pfizer.com).* The hepatic enzyme, uridine diphosphate glucoronosyltransferase 1A1

(UGT1A1) participates in the metabolism of SN-38, the activated form of IR (Iyer *et al.* 1998). A well-studied polymorphism in *UGT1A1*, an additional TA repeat in the 5' promoter region, has been implicated in the toxic effects of IR and is responsible for causing decreased enzymatic activity (Miners *et al.* 2002; Rouits *et al.* 2004). In 2005, the FDA added a warning to the Camptosar® product insert. It recommends that individuals who are homozygous for the TA repeat polymorphism (UGT1A1*28 allele) are at risk of developing neutropenia, and should receive lower initial doses of the drug. Treating-physicians in the USA can make use of an FDA-approved prognostic genetic test offered by companies such as Quest Diagnostics (www.questdiagnostics.com) to assay this polymorphism in patients. Since this variant is a short tandem repeat, it was not featured on the Affymetrix GenomeWideSNP 6.0 Array and hence the allele frequencies could not be established in the five indigenous Black SA populations. Therefore future studies can be conducted to establish the incidence of the variant in SA to determine whether the UGT1A1*28 prognostic test is suitable and widely applicable. Beutler *et al.* (1998) reported a high prevalence of the additional TA repeat in people of African origin, making it a putative marker to study in SA.

A *CRC Pharmacogenomic Panel* has also recently been developed by Quest Diagnostics for the management of CRC. The test is aimed at evaluating the toxicity response of patients treated with 5-FU, Oxaliplatin (OX; *Eloxatin*®) and IR. Variants in a panel of seven genes are assayed including *UGT1A1, TYMS, XRCC1, ERCC1, GSTP1, XPD* and *DPYD*. As previously mentioned, variance in 5-FU and IR metabolism can be attributed to polymorphisms in *TYMS, DPYD* and *UGT1A1* respectively (Kawakami *et al.* 1995; Wei *et al.* 1996; Ulrich *et al.* 2000; Raida *et al.* 2001; Rouits *et al.* 2004). This test also assays the *GSTP1*—I105V, *XPD*—K751Q, *XRCC1*—R399Q and *ERCC1*—T118C SNPs, as indicators of prognosis in order to guide chemotherapy selection. These polymorphisms have previously been shown through clinical studies to affect how patients respond to OX (Park *et al.* 2001; Stoehlmacher *et al.* 2002; Viguier *et al.* 2005). However, several

issues confound the utility of this test in SA. Not only does the haplotype of *DPYD* polymorphisms, which reportedly confers the greatest risk for toxicity (ATC or IVS14+1A—R29C—I543V), appear at very low frequencies in both CAU (0.008) and MA (0.0009) population groups, but OX is currently not used as a chemotherapeutic agent in the public sector, and specifically at Groote Schuur Hospital (GSH). Instead, Cisplatin (CIS), another platinum derivative is the agent of choice for, *inter alia,* bladder, cervical and oesophageal cancer (Personal communication, Dr Barbara Robertson, Department of Oncology, Groote Schuur Hospital). Moreover, the SNPs assayed by the *CRC Pharmacogenomic Panel* were not featured on the Affymetrix GenomeWideSNP 6.0 Array and, therefore, the respective allele frequencies could not be immediately identified in the five indigenous Black SA populations. Therefore, it is doubtful whether this test would be useful in the public sector in SA at this point in time.

It remains, however, that potential studies could determine the difference in allele frequencies between SNPs in CIS- and OX-metabolising genes in a SA CRC cohort. In so doing, it could possibly be established which of the two drugs would be most beneficial and, hence, recommended for use in SA hospitals. Of note, are several studies reporting that MMR- deficient cells are resistant to CIS but not OX (Aebi *et al.* 1996; Vaisman *et al.* 1998; Zdraveski *et al.* 2002). Aebi *et al.* (1996) carried out experiments on isogenic (genetically identical) pairs of cell lines that were either deficient in *MLH1* (HCT116) and *MSH2* (HEC59) or proficient in MMR. The authors reported that both deficient cell lines were two-fold resistant to the anti-neoplastic effects of CIS compared to their counterparts which were expressing *MLH1* and *MSH2* (Aebi *et al.* 1996). Aebi *et al.* (1997) later independently confirmed these findings. This differential sensitivity of MMR-deficient cells should be taken into consideration when prescribing platinum compounds to HNPCC patients.

The same applies to another simple mutational analysis of the *DPYD* gene, produced by the same company, Quest Diagnostics, which can predict

toxicity from the pyrimidine-based agents, 5-FU and Capecitabine. However, the assay has yet to be approved by the FDA, and as such the company cannot provide this test to be used as a proven prognostic/diagnostic screen. At the moment, the 5-FU package insert contains no information pertaining to pharmacogenetic screening for toxic responses. The reason for this is because no acceptable, validated test is commercially available yet, despite convincing proof that *DPYD* is involved.

It would also have to be established whether such a test would be cost effective if made available to individuals within the public and private health sector of SA. SA is a developing country, with limited resources. However, this current study provided valuable insight into several pharmacogenetic variants, which could contribute to the development of future SA-specific prognostic tests, which would be more affordable and applicable.

CHAPTER 5

CONCLUDING REMARKS

This pilot study was formulated to identify mechanisms for accessing patient records and biomaterial, and developing assays and analyses, towards larger scale studies on an established CRC cohort. Moreover, a critical survey was performed of the referral system at Groote Schuur Hospital and UCT, in order to identify means of improving data and sample capture, for more comprehensive and meaningful analyses in future.

This research has laid the foundation for future studies which will aim, amongst other things, to determine whether South African (SA) colorectal cancer (CRC) patients with a defective mismatch repair system (MMR) respond differently to the chemotherapeutic agents when compared to their MMR proficient counterparts. In order to perform statistical analysis with sufficient power, future patient cohorts will have to be maximised. This would very likely entail recruiting individuals not only from Groote Schuur Hospital (GSH), but also from other private and state institutions in SA, and obtaining complete treatment records. Additionally, implementing a reliable informatics system, at every point of patient care and follow-up, to capture patient data at GSH, and elsewhere, would facilitate the collection of valuable information for research aimed at understanding the heterogeneity of disease, its manifestation, targeted treatment, and improved prognosis.

Cancer surveillance and control in SA is dealt with in a fragmented fashion, lacking an optimal infrastructure, and without a definitive policy on cancer control. It is therefore advisable that the National Department of Health (NDOH) of SA establishes a Cancer Control Programme, which will include key role players such as clinical and research institutions, whose main focus will be to alleviate these discrepancies (Albrecht 2006). As mentioned previously, cancer is a non-notifiable disease in SA. Therefore it is

imperative that the surveillance infrastructure is remedied by changing the legislation such that all cancer diagnoses are notifiable to the NCR. In the interim, a network of population-based registries needs to be established, until such time that cancer becomes a notifiable disease. The research conducted in this study contributes substantially to this endeavour. This improved patient registry, and the reinforcement of relationships between role-players at each point in patient care and management, at GSH and UCT, will facilitate the flow of information to the NCR, where it can be evaluated (Albrecht 2006).

In addition, the current study was intended to evaluate whether variants in 5-FU-metabolising genes can be studied in future to determine the efficacy of the drug in different populations in SA. Five variants within the 5-FU pharmacorelevant genes, thymidylate synthase *(TYMS)* and dihydropyrimidine dehydrogenase (*DPYD*) were studied in order to gain valuable insight into the frequencies of these five variants in SA, which could contribute to the development of SA-specific prognostic tests, which would be more affordable and applicable.

This could serve as valuable groundwork in order to expand pharmacogenetic studies of other cancers and chemotherapeutic modalities, towards individualised cancer treatment, and improved prognosis. Toxic side effects and resistance to chemotherapy could be explained by differential expression of metabolising genes, and genetic profiles of patients can potentially aid in discovering effective treatment options. Oncologists in SA do not use mutation status to guide therapy. There are several prognostic genetic tests currently available in the USA, which shows promise in changing the way drugs are prescribed. These tests are, however, limited in as much they are developed to analyse polymorphisms identified in populations outside of SA, emphasising the need to develop larger scale pharmacogenetic studies, in sizable cohorts of subjects encompassing the various population groups present in this country. More importantly, and as repeatedly highlighted in the present study, there is a strong need to ensure

integrity of patent-related data from point of admission and diagnosis, surgery, treatment and subsequent follow-up.

REFERENCES

1. Abratt RP and Vorobiof DA (2003) Cancer in Africa. Lancet Oncol 4:394-396

2. Aebi S, Kurdihaidar B, Gordon R, Cenni B, Zheng H, Fink D, Christen RD, Boland CR, Koi M, Fishel R, Howell SB (1996) Loss of DNA mismatch repair in acquired resistance to cisplatin. Cancer Res 56(13):3087-3090

3. Aebi S, Fink D, Gordon R, Kim HK, Zheng H, Fink JL, Howell SB (1997) Resistance to Cytotoxic Drugs in DNA Mismatch Repair-deficient Cells. Clin Cancer Res 3:1763-1767

4. Akao Y, Nakagawa Y, Naoe T (2006) let-7 MicroRNA Functions as a Potential Growth Suppressor in Human Colon Cancer Cells. Biol Pharm Bull 29(5):903-906

5. Albrecht C, Independent Medical Research Consultant (2006) Overview of the South African Cancer Research Environment as a Basis for Discussions Concerning the Activation of CARISA (Cancer Research Initiative of South Africa)

6. Allen WL and Johnston PG (2005) Role of Genomic Markers in Colorectal Cancer Treatment. J Clin Oncol 23:4545-4552

7. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215(3):403-10

8. Anthoney DA, McIlwrath AJ, Gallagher WM, Edlin ARM, Brown R (1996) Microsatellite instability, apoptosis, and loss of p53 function in drug-resistant tumor cells. Cancer Res 56(6):1374-1381

9. Arnold CN, Goel A, Boland CR (2003) Role of hMLH1 promoter hypermethylation in drug resistance to 5-fluorouracil in colorectal cancer cell lines. Int J Cancer 106(1):66-73

10. Arnold CN, Goel A, Compton C, Marcus V, Niedzwiecki D, Dowell JM, Wasserman L, Inoue T, Mayer RJ, Bertagnolli MM, Boland CR (2004) Evaluation of Microsatellite Instability, hMLH Expression and hMLH1 Promoter Hypermethylation in Defining the MSI Phenotype of Colorectal Cancer. Cancer Biol Ther 3(1):73-78

11. Arnold CN, Goel A, Niedzwiecki D, Dowell JM, Wasserman L, Compton C, Mayer RJ, Bertagnolli MM, Boland CR (2005) APC Promoter Hypermethylation Contributes to the loss of APC Expression in Colorectal Cancers with Allelic Loss on 5q. Cancer Biol Ther 3(10) :960-964

12. Barnett D, Stevens A, Longson C (2006) Appraisal of bevacizumab and cetuximab for treatment of metastatic colorectal cancer in the UK. Lancet 7:807-808

13. Barnetson RA, Tenesa A, Farrington SM, Nicholl ID, Cetnarskyj R, Porteous ME, Campbell H, Dunlop MG. (2006) Identification and survival of carriers of mutations in DNA mismatch repair genes in colon cancer. N Engl J Med 354:2751-2763

14. Bartolomei MS and Tilghman SM (1997) Genomic imprinting in mammals. Annu Rev Genet 31:493-525

15. Bradshaw D, Nannan N, Laubscher R, Groenewald P, Joubert J, Nojilana B, Normn R, Pieterse D, Schneider M (2004) South African National Burden of Disease Study 2000: Estimates of Provincial Mortality. Cape Town, South African Medical Research Council

16. Beutler E, Gelbart T, Demina A (1998) Racial variability in the UDP-glucuronosyltransferase (UGT1A1) promoter: A balanced polymorphism for regulation of bilirubin metabolism? Proc Natl Acad Sci USA 95:8170-8174

17. Bhattacharyya NP, Skandalis A, Ganesh A, Groden J, Meuth M (1994) Mutator phenotypes in human colorectal carcinoma cell lines. Proc Nat Acad Sci USA 91:6319-6323

18. Bocker T, Barusevicius A, Snowden T, Rasio D, Guerrette S, Robbins D, Schmidt C, Burczak J, Croce CM, Copeland T, Kovatich AJ, Fishel R (1999) hMSH5: A Human MutS Homologue That Forms a Novel Heterodimer with hMSH4 and Is Expressed during Spermatogenesis. Cancer Res 59:816-822

19. Boland CR, Thibodeau SN, Hamilton SR, Sidransky D, Eshleman JR, Burt RW, Meltzer SJ, Rodriguez-Bigas MA, Fodde R, Ranzani GN, Srivastava S (1998) A National Cancer Institute Workshop on Microsatellite Instability for cancer detection and familial predisposition: development of international criteria for the determination of microsatellite instability in colorectal cancer. Cancer Res 58:5248-57

20. Branch P, Hampson R, Karran P (1995) DNA Mismatch Binding Defects, DNA Damage Tolerance, and Mutator Phenotypes in Human Colorectal Carcinoma Cell Lines. Cancer Res 55:2304-2309

21. Bras-Gonçalves RA, Rosty C, Laurent-Puig P, Soulié P, Dutrillaux B, Poupon M-F (2000) Sensitivity to CPT-11 of xenografted human colorectal cancers as a function of microsatellite instability and p53 status. Br J Cancer 82(4):913-923

22. Burger EH, van der Merwe L, Volmink J (2007) Errors in the completion of the death notification form. S Afr Med J 97(11):1077-1081

23. Calin GA, Dumitru CD, Shimizu M, Bichi R, Zupo S, Noch E, Aldler H, Rattan S, Keating M, Rai K, Rassenti L, Kipps T, Negrini M, Bullrich F, Croce CM (2002) Frequent deletions and down-regulation of micro-RNA genes *miR15* and *miR16* at 13q14 in chronic lymphocytic leukemia. Proc Natl Acad Sci USA 99:15524-15529

24. Calin GA, Sevignani C, Dumitru CD, Hyslop T, Noch E, Yendamuri S, Shimizu M, Rattan S, Bullrich F, Negrini M, Croce CM (2004) Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers. Proc Natl Acad Sci USA 101:2999-3004

25. Camptosar [package insert] (2006) New York, NY: Pfizer Inc

26. Cassidy J, Twelves C, Van Cutsem E, Hoff P, Bajetta E, Boyer M, Bugat R, Burger U, Garin A, Graeven U, Mckendrick J, Maroun J, Marshall J, Osterwalder, Perez-manga G, Rosso R, Rougier P, Schilsky RL (2002) First-line oral capecitabine therapy in metastatic colorectal cancer: a favourable safety profile compared with intravenous 5-fluorouracil/leucovorin. Ann Oncol 13:566-575

27. Cassidy J, Tabernero J, Twelves C, Brunet R, Butts C, Conroy T, Debraud F, Figer A, Grossmann J, Sawada N, Schöffski P, Sobrero A, Van Cutsem E, Díaz-Rubio E (2004) XELOX (Capecitabine Plus Oxaliplatin): Active First-Line Therapy for Patients With Metastatic Colorectal Cancer. J Clin Oncol 22:2084-2091

28. Cheng Y and Nakayama K (1983) Effects of 5-fluoro-2'-deoxyuridine on DNA metabolism in HeLa cells. Mol Pharmacol 23:171-174

29. Chin KM, Wessler B, Chew P, Lau J (2006) Genetic Tests for Cancer. Tufts-NEMC EPC

30. Claij N and Riele H (1999) Microsatellite instability in human cancer: a prognostic marker for chemotherapy? Exp Cell Res 246(1):1-10

31. Clinical Laboratory Improvement Amendments of 1988, U.S. Code, Publ. Law No. 100-578.

32. Collie-Duguid ES, Etienne MC, Milano G, McLeod HL (2000) Known variant *DPYD* alleles do not explain DPD deficiency in cancer patients. Pharmacogenetics 10(3):217–223

33. Cunningham D, Humblet Y, Siena S, Khayat D, Bleiberg H, Santoro A, Bets D, Mueser M, Harstrick A, Verslype C, Chau I, Van Cutsem E (2004) Cetuximab monotherapy and cetuximab plus irinotecan in irinotecan-refractory metastatic colorectal cancer. N Engl J Med 351:337-45

34. Curtin NJ, Harris AL, Aherne GW (1991) Mechanism of Cell Death following Thymidylate Synthase Inhibition: 2'-Deoxyuridine-5'-triphosphate Accumulation, DNA Damage, and Growth Inhibition following Exposure to CB3717 and Dipyridamole. Cancer Res 51:2346-2352

35. De Gramont A, Figer A, Seymour M, Homerin M, Hmissi A, Cassidy J, Boni C, Cortes-Funes H, Cervantes A, Freyer G, Papamichael D, Le Bail N, Louvet C, Hendler D, de Braud F, Wilson C, Morvan F, Bonetti A (2000) Leucovorin and Fluorouracil With or Without Oxaliplatin as First-Line Treatment in Advanced Colorectal Cancer. J Clin Oncol 18:2938-2947

36. Diasio RB, Beavers TL, Carpenter JT (1988) Familial Deficiency of Dihydropyrimidine Dehydrogenase: Biochemical Basis for Familial Pyrimidinemia and Severe 5-Fluorouracil-induced Toxicity. J Clin Invest 81:47-51

37. Doong S and Dolnick BJ (1988) 5-Fluorouracil Substitution Alters Pre-mRNA Splicing *in vitro*. J Biol Chem 263(9):4467-4473

38. Dotor E, Cuatrecases M, Martínez-Iniesta M, Navarro M, Vilardell F, Guino´ E, Pareja L, Figueras A, Molleví DG, Serrano T, de Oca J, Peinado MA, Moreno V, Ramo´n Germa` J, Capella´ G, Villanueva A (2006) Tumor Thymidylate Synthase 1494del6 Genotype As a Prognostic Factor in Colorectal Cancer Patients Receiving Fluorouracil-Based Adjuvant Treatment. J Clin Oncol 24(10):1603-1611

39. Douillard JY, Cunningham D, Roth AD, Navarro M, James RD, Karasek P, Jandik P, Iveson T, Carmichael J, Alakl M, Gruia G, Awad L, Rougier P (2000) Irinotecan combined with fluorouracil compared with fluorouracil alone as first-line treatment for metastatic colorectal cancer: a multicentre randomised trial. Lancet 355:1041-7

40. Elsakov PA and Kurtinaitis J (2006) Survival from colorectal carcinoma in HNPCC families as compared to the general population in Lithuania - initial results. Fam Cancer 5:369-371

41. Emmanouilides C, Pegram M, Robinson R, Hecht R, Kabbinavar F, Isacoff W (2004) Anti-VEGF antibody bevacizumab (Avastin) with 5FU/LV as third line treatment for colorectal cancer. Tech Coloproctol 8:S50-S52

42. Ezzeldin HH, Lee AM, Mattison LK, Diasio RB (2005) Methylation of the DPYD Promoter: An Alternative Mechanism for Dihydropyrimidine Dehydrogenase Deficiency in Cancer Patients. Clin Cancer Res 11(24):8699-8705

43. Fallik D, Borrini F, Boige V, Viguier J, Jacob S, Miquel C, Sabourin J, Ducreux M, Praz F (2003) Microsatellite Instability Is a Predictive Factor of the Tumor Response to Irinotecan in Patients with Advanced Colorectal Cancer. Cancer Res 63:5738-5744

44. Felton KEA, Gilchrist DM, Andrew SE (2007) Constitutive deficiency in DNA mismatch repair: is it time for Lynch III? Clin Genet 71:499-500

45. Fisher F, Barenfaller K, Jiricny J (2007) 5-Fluorouracil is efficiently removed from DNA by the base excision and mismatch repair system. Gastroenterology 133(6):1858-1865

46. Giacomini KM, Krauss RM, Roden DM, Eichelbaum M, Hayden MR, Nakamura Y (2007) When good drugs go bad. Nature 446:975-977

47. Gifford G, Paul J, Vasey PA, Kaye SB, Brown R (2004) The Acquisition of *hMLH1* Methylation in Plasma DNA after Chemotherapy Predicts Poor Survival for Ovarian Cancer Patients. Clin Cancer Res 10:4420-4426

48. Goldberg RM, Rothenberg ML, Van Cutsem E, Benson AB, Blanke CD, Diasio RB, Grothey A, Lenz HJ, Meropol NJ, Ramanathan RK, Becerra CHR, Whickham R, Amstrong D, Viele C (2007) The continuum of Care: A Paradigm for the Management of Metastatic Colorectal Cancer. Oncologist 12:38-50

49. Goulian M, Bleile B, Tseng BY (1980) The Effect of Methotrexate on Levels of dUTP in Animal Cells. J Biol Chem 255(22):10630-10637 **(a)**

50. Goulian M, Bleile B, Tseng BY (1980) Methotrexate-induced misincorporation of uracil into DNA. Proc Natl Acad Sci USA 77(4):1956-1960 **(b)**

51. Grem JL (2000) 5-Fluorouracil: forty-plus and still ticking. A review of its preclinical and clinical development. Invest New Drugs 18:299-313

52. Gross E, Ullrich T, Seck K, Mueller V, de Wit M, von Schilling C, Meindl A, Schmitt M, Kiechle M (2003) Detailed Analysis of Five Mutations in Dihydropyrimidine Dehydrogenase Detected in Cancer Patients With 5-Fluorouracil-related Side Effects. Hum Mutat 22:498

53. Hamilton SR, Liu B, Parsons RE, Papadopoulos N, Jen J, Powell SM, Krush AJ, Berk T, Cohen Z, Tetu B, Burger PC, Wood PA, Taqi F, Booker SV, Peterson GM, Offerhaus GJA, Tersmette AC, Giardiello FM, Vogelstein B, Kinzler KW (1995) The molecular basis of Turcot's syndrome. N Engl J Med 332: 839-47

54. Heggie GD, Sommadossi J-P, Cross DS, Huster WJ, Diasio RB (1987) Clinical pharmacokinetics of 5-fluorouracil and its metabolism in plasma, urine, and bile. Cancer Res 47:2203-2206

55. Hemminki A, Peltomaki P, Mecklin J, Jarvinen H, Salovaara R, Nystrom-Lahti, de la Chapelle A, Aaltonen LA (1994) Loss of the wild type MLH1 gene is a feature of hereditary nonpolyposis colorectal cancer. Nat Genet 8:405-410

56. Herrera L, Kakati S, Gibas L, Pietrzak E, Sandberg AA (1986) Gardner syndrome in a man with an interstitial deletion of 5q. Am J Med Genet 25:473-476 [ABSTRACT]

57. Honchel R, Halling KC, Schaid DJ, Pittelkow M, Thibodeau SN (1994) Microsatellite instability in Muir-Torre syndrome. Cancer Res 54:1159 -1163.

58. Horie N, Aiba H, Oguro K, Hojo H, Takeishi K (1995) Functional analysis and DNA polymorphism of the tandemly repeated sequences in the 5'-terminal regulatory region of the human gene for thymidylate synthase. Cell Struct Funct 20:191-197

59. Hospers GAP, Schaapveld M, Nortier JWR, Wils J, van Bochove A, de Jong RS, Creemers GJ, Erjavec Z, de Gooyer DJ, Slee PHThJ, Gerrits CJH, Smit JM, Mulder NH (2006) Randomised Phase III study of biweekly 24-h infusion of high-dose 5FU with folinic acid and oxaliplatin versus monthly plus 5-FU/folinic acid in first-line treatment of advanced colorectal cancer. Ann Oncol 17 :443-449

60. Houghton JA, Tillman DM, Harwood FG (1995) Ratio of 2'-deoxyadenosine-5'-triphosphate/Thymidine-5'-triphosphate Influences the Commitment of Human Colon Carcinoma Cells to Thymineless Death. Clin Cancer Res 1:723-730

61. Hurwitz H, Fehrenbacher L, Novotny W, Cartwright T, Hainsworth J, Heim W, Berlin J, Baron A, Griffing S, Holmgren E, Ferrara N, Fyfe G, Rogers B, Ross R, Kabbinavar F (2004) Bevacizumab plus irinotecan, fluorouracil, and leucovorin for metastatic colorectal cancer. New Engl J Med 350:2335-2342

62. Ingraham HA, Dickey L, Goulian M (1986) DNA fragmentation and cytotoxicity from increased cellular deoxyuridylate. Biochemistry 25(11):3225-30

63. Iyer L, King CD, Whitington PF, Green MD, Roy SK, Tephly TR, Coffman BL, Ratain MJ (1998) Genetic Predisposition to the Metabolism of Irinotecan (CPT-11) Role of Uridine Diphosphate Glucuronosyltransferase Isoform 1A1 in the Glucuronidation of its Active Metabolite (SN-38) in Human Liver Microsomes. J Clin Invest 101(4):847-854

64. The International HapMap Consortium (2003) The International HapMap Project. Nature 426:789-796

65. Jacob S and Praz F (2002) DNA mismatch repair defects: role in colorectal carcinogenesis. Biochimie 84(1):27-47

66. Ji SH, Park YS, Lee J, Lim DH, Park B, Park KW, Kang JH, Lee S, Park JO, Kim K, Kim WS, Jung CW, Im Y, Kang WK, Park K (2005) Phase II Study of Irinotecan, 5-Fluorouracil and Leucovorin as First-line Therapy for Advanced Colorectal Cancer. Jpn J Clin Oncol 35(4):214-217

67. Jiricny J and Nyström-Lahti M (2000) Mismatch repair defects in cancer. Curr Opin Genet Dev 10:157-161

68. Jover R, Zapater P, Castells A, Llor X, Andreu M, Cubiella J, Piñol V, Xicola RM, Bujanda L, Reñé JM, Clofent J, Bessa X, Morillas JD, Nicolás-Pérez D, Payá A, Alenda C (2006) Mismatch repair status in the prediction of benefit from adjuvant fluorouracil chemotherapy in colorectal cancer. Gut 55:848-855

69. Kabbinavar F, Hurwitz HI, Fehrenbacher L, Meropol NJ, Novotny WF, Lieberman G, Griffing S, Bergsland E (2003) Phase II, randomised control trial comparing bevacizumab plus fluorouracil (FU)/leucovorin (LV) with FU/LV alone in patients with metastatic colorectal cancer. J Clin Oncol 21:60-65

70. Kabbinavar FF, Schulz J, McCleod M, Patel T, Hamm JT, Hecht JR, Mass R, Perrou B, Nelson B, Novotny WF (2005) Addition of bevacizumab to bolus fluorouracil and leucovorin in first-line metastatic colorectal cancer: results of a randomized phase II trial. J Clin Oncol 23:3697-705

71. Katsanis SH, Javitt G, Hudson K (2008) A Case Study of Personalized Medicine. Science 320:53-54

72. Kawakami K, Salonga D, Omura K, Park J, Danenberg K, Watanabe Y, Danenberg P (1995) Effects of polymorphic tandem repeat sequence on the *in vitro* translation of messenger RNA. Proc Am Assoc Cancer Res 40 :436-437 [ABSTRACT]

73. Kawakami K, Omura K, Kanehira E, Watanabe Y (1999) Polymorphic tandem repeats in the thymidylate synthase gene is associated with its protein expression in human gastrointestinal cancers. Anticancer Res 19:3249-3252

74. Kawakami K and Watanabe G (2003) Identification and Functional Analysis of Single Nucleotide Polymorphism in the Tandem Repeat Sequence of *Thymidylate Synthase* Gene. Cancer Res 63:6004-6007

75. Kinzler KW, Nilbert MC, Vogelstein B, Bryan TM, Levy DB, Smith KJ, Preisinger AC, Hamilton SR, Hedge P, Markham A, Carlson M, Joslyn G, Groden J, White R, Miki Y, Miyoshi Y, Nishisho I, Nakamura Y (1991) Identification of a gene located at chromosome 5q21 that is mutated in colorectal cancers. Science 251:1366-1370

76. Kirley SD, D'Apuzzo M, Lauwers GY, Graeme-Cook F, Chung DC, Zukerberg LR (2005) The Cables Gene on Chromosome 18Q Regulates Colon Cancer Progression In Vivo. Cancer Biology Therapy 4(8):861-863

77. Klein TE, Chang JT, Cho MK, Easton KL, Ferguson R, Hewett M, Lin Z, Liu Y, Oliver DE, Rubin DL, Shafa F, Stuart JM, Altman RB (2001) Integrating genotype and phenotype information: an overview of the PharmGKB project. Pharmacogenetic Research Network and Knowledge Base. Pharmacogenomics J 1:167-170

78. Kohne CH, Van Cutsem E, Wils J, Bokemeyer C, El-Serafi M, Lutz MP, Lorenz M, Reichardt P, Rückle-Lanz H, Frickhofen N, Fuchs R, Mergenthaler HG, Langenbuch T, Vanhoefer U, Rougier P, Voigtmann R, Müller L, Genicot B, Anak O, Nordlinger B (2005) Phase III study of weekly high-dose infusional fluorouracil plus folinic acid with or without irinotecan in patients with metastatic colorectal cancer: European Organisation for Research and Treatment of Cancer Gastrointestinal Group Study 40986. J Clin Oncol. 23:4856-4865

79. Kufe DW, Majors PP, Egan EM, Loh E (1981) 5-Fluoro-2'-deoxyuridine Incorporation in L1210 DNA. J Biol Chem 256(17):8885-8888

80. Lanza G, Ferracin M, Gafà R, Veronese A, Spizzo R, Pichiorri F, Liu C, Calin GA, Croce CM, Negrini M (2007) mRNA/microRNA gene expression profile in microsatellite unstable colorectal cancer. Mol Cancer 6:54

81. Lei X, Zhong M, Feng L, Zhu B, Tang S, Liao D (2007) siRNA-mediated Bcl-2 and Bcl-xl Gene Silencing Sensitizes Human Hepatoblastoma Cells to Chemotherapeutic Drugs. Clin Exp Pharmacol Physiol 34:450-456

82. Lenz HJ, Hayashi K, Salonga D, Danenberg KD, Danenberg PV, Metzger R, Banerjee D, Bertino JR, Groshen S, Leichman LP, Leichman CG (1998) p53 point mutations and thymidylate synthase messenger RNA levels in disseminated colorectal cancer: An analysis of response and survival. Clin Cancer Res 4:1243-1250

83. Lindblad-Toh K, Winchester E, Daly MJ Wang DG, Hirschhorn JN, Laviolette JP, Ardlie K, Reich DE, Robinson E, Sklar P, Shah N, Thomas D, Fan JB, Gingeras T, Warrington J, Patil N, Hudson TJ, Lander ES (2000) Large-scale discovery and genotyping of single-nucleotide polymorphisms in the mouse. Nat Genet 24:381-386

84. Longley DB, Harkin DP, Johnston PG (2003) 5-Fluorouracil: Mechanisms of action and clinical strategies. Nature 3:330-338

85. Lönn U and Lönn S (1984) Interaction between 5-Fluorouracil and DNA of Human Colon Adenocarcinoma. Cancer Res 44, 3414-3418

86. Loukola A, Eklin K, Laiho P, Salovaara R, Kristo P, Järvinen H, Mecklin J, Launonen V, Aaltonen LA (2001) Microsatellite Marker Analysis in Screening for Hereditary Nonpolyposis Colorectal Cancer (HNPCC). Cancer Res 61:4545-4549

87. Lynch HT, Smyrk TC, Watson P, Lanspa SJ, Lynch JF, Lynch PM, Cavalieri RJ, Boland CR (1993). Genetics, Natural History, Tumour Spectrum, and Pathology of Hereditary Nonpolyposis Colorectal Cancer: An Updated Review. Gastroenterology 104:1535-1549

88. Lynch HT and Smyrk TC (1998) Identifying hereditary nonpolyposis colorectal cancer. N Engl J Med. 338:1537-8

89. Ma L, Teruya-Feldstein J, Weinberg RA (2007) Tumour invasion and metastasis initiated by microRNA-10b in breast cancer. Nature 1-8

90. Mandola MV, Stoehlmacher J, Muller-Weeks S, Cesarone G, Yu MC, Lenz HJ, Ladner RD (2003) A novel single nucleotide polymorphism within the 5' tandem repeat polymorphism of the thymidylate synthase gene abolishes USF-1 binding and alters transcriptional activity. Cancer Res 63:2898-2904

91. Mandola MV, Stoelmacher J, Zhang W, Groshen S, Yu MC, Iqbal S, Lenz HJ, Ladner RD (2004) A 6-bppolymorphism in the thymidylate synthase gene causes message instability and is associated with decreased intratumoural TS mRNA levels. Pharmacogenomics 14:319-327

92. Marcos I, Borrego S, Urioste M, García-vallés C, Antiñolo G (2006) Mutations in the DNA Mismatch Repair Gene MLH1 Associated With Early-onset Colon Cancer. J Pediatr 148:837-9

93. Maring JG, van Kuilenburg ABP, Haasjes J, Piersma H, Groen HJM, Uges DRA, Van Gennip AH, De Vries EGE (2002) Reduced 5-FU clearance in a patient with low DPD activity due to heterozygosity for a mutant allele of the *DPYD* gene. Br J Cancer 86:1028-33

94. Markowitz LE, Dunne EF, Saraiya M, Lawson HW, Chesson H, Unger ER; Centers for Disease Control and Prevention (CDC) Advisory Committee on Immunization Practices (ACIP) (2007) Quadrivalent Human Papillomavirus Vaccine: recommendations of the Advisory Committee on Immunization Practices. MMWR Recomm Rep 56(RR-2):1-24

95. Marsh S, Ameyaw MM, Githanga' J, Indalo A, Ofori-Adjei D, McLeod HL (2000) Novel thymidylate synthase enhancer region alleles in African populations. Hum Mutat 16:528-524

96. Marsh S, Collie-Duguid ES, Li T, Liu X, McLeod HL (1999) Ethnic variation in the thymidylate synthase enhancer region polymorphism among Caucasian and Asian populations. Genomics 58:310-312

97. Marsh S (2005) Thymidylate synthase pharmacogenetics. Invest New Drugs 23:533-537

98. Mathers CD, Ma Fat D, Inoue M, Rao C, Lopez A (2005) Counting the dead and what they died from: an assessment of the global status of cause of death data. Bull World Health Organ 83:171-177.

99. Mattison LK, Johnson MR, Diasio RB (2002) A comparative analysis of translated dihydropyrimidine dehydrogenase cDNA; conservation of functional domains and relevance to genetic polymorphisms. Pharmacogenetics 12(2):133-144

100. Meinsma R, Fernandez-Salguero P, van Kuilenburg AB, Van Gennip AH, Gonzalez FJ (1995) Human polymorphism in drug metabolism: mutation in the dihydropyrimidine dehydrogenase gene results in exon skipping and thymine uraciluria. DNA Cell Biol 14:1-6

101. Meyers M, Wagner MW, Hwang HS, Kinsella TJ, Boothman DA (2001) Role of the *hMLH1* DNA Mismatch Repair Protein in Fluoropyrimidine-mediated Cell Death and Cell Cycle Responses. Cancer Res 61:5193-5201

102. Michael MZ, O' Connor SM, van Holst Pellekaan NG, Young GP, James RJ (2003) Reduced Accumulation of Specific MicroRNAs in Colorectal Neoplasia. Mol Cancer Res 1:882-891

103. Miners JO, McKinnon RA, Mackenzie PI (2002) Genetic polymorphisms of UDP-glucuronosyltransferases and their functional significance. Toxicology 181/182:453-456

104. Mqoqi N, Kellett P, Sitas F, Jula M (2004) Incidence of Histologically Diagnosed Cancer in South Africa 1998 - 1999. National Cancer Registry

105. Naguib FNM and el Kouni MH (1985) Enzymes of uracil catabolism in normal and neoplastic human tissues. Cancer Res 45:5405-12

106. Niv Y (2007) Microsatellite instability and MLH1 promoter hypermethylation in colorectal cancer. World J Gastroenterol 13(12):1767-1769

107. Park DJ, Stoehlmacher J, Zhang W, Tsao-Wei DD, Groshen S, Lenz H (2001) A Xeroderma Pigmentosum Group D Gene Polymorphism Predicts Clinical Outcome to Platinum-based Chemotherapy in Patients with Advanced Colorectal Cancer. Cancer Res 61:8654-8658

108. Parkin DM, Pisani P, Ferlay J (1999) Global Cancer Statistics. CA Cancer J Clin 49:33-64

109. Parkin DM, Bray F, Ferlay J, Pisani P (2005) Global Cancer Statistics, 2002. CA Cancer J Clin 55:74-108

110. Parkin DM, Sitas F, Chirenje M, Stein L, Abratt R, Wabinga H (2008) Part 1: Cancer in Indigenous Africans-burden, distribution, and trends. Lancet Oncol 9:683-692

111. Presta LG, Chen H, O'Connor SJ, Chisholm V, Meng YG, Krummen L, Winkler M, Ferrara N (1997) Humanization of an anti-vascular endothelial growth factor monoclonal antibody for the therapy of solid tumors and other disorders. Cancer Res 57:4593-4599

112. Prewett MC, Hooper AT, Bassi R, Ellis LM, Waksal HW, Hicklin DJ (2002) Enhanced antitumor activity of anti-epidermal growth factor receptor monoclonal antibody IMC-C225 in combination with irinotecan (CPT-11) against human colorectal tumor xenografts. Clin Cancer Res 8:994-1003

113. Pullerkat ST, Stoelmacher J, Ghaderi V, Xiong Y-P, Ingles SA, Sherrod A, Warren R, Tsao-Wei D, Groshen S, Lenz H-J (2001) Thymidylate synthase gene polymorphism determines response and toxicity of 5-FU chemotherapy. Pharmacogenomics J 1:65-70

114. Raida M, Scvhawabe W, Hausler P, Van Kuilenburg ABP, Van Gennip AH, Behnke D, Hoffken K (2001) Prevalence of a Common Point Mutation in the Dihydropyrimidine Dehydrogenease (DPD) Gene within the 5'-Splice Donor site of Intron 14 in Patients with Severe 5-fluorouracil (5-FU)-related Toxicity Compared to Controls. Clin Cancer Res 7:2832-2839

115. Ribic CM, Sargent DJ, Moore MJ, Thibodeau SN, French AJ, Goldberg RM, Hamilton SR, Laurent-Puig P, Gryfe R, Shepherd LE, Tu D, Redston M, Gallinger S (2003) Tumor Microsatellite-Instability Status as a Predictor of Benefit from Fluorouracil-Based Adjuvant Chemotherapy for Colon Cancer. N Engl J Med 349:247-57

116. Rodriguez-Bigas MA, Boland CR, Hamilton SR, Henson DE, Jass JR, Khan PM, Lynch H, Perucho M, Smyrk T, Sobin L, Srivastava S. (1997) A National Cancer Institute Workshop on Hereditary Non-Polyposis Colorectal Cancer Syndrome: Meeting Highlights and Bethesda Guidelines. J Nat Cancer Inst 89(23):1758-1762

117. Rozen S, Skaletsky H (2000) Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol 132:365-86

118. Ross JS, Hatzis C, Symmans WF, L Pusztai, Hortobágyi GN (2008) Commercialized Multigene Predictors of Clinical Outcome for Breast Cancer. Oncologist 13:477-493

119. Rossi L, Bonmassar E, Faraoni I (2007) Modification of miR gene expression pattern in human colon cancer cells following exposure to 5-fluorouracil *in vitro*. Pharmacol Res 56:248-253

120. Rouits E, Boisdron-Celle M, Dumont A, Guérin O, Morel A, Gamelin E (2004) Relevance of Different UGT1A1 Polymorphisms in Irinotecan-Induced Toxicity: A Molecular and Clinical Study of 75 Patients. Clin Cancer Res 10:5151-5159

121. Rudolph JG, White S, Sokolsky C, Bozak D, Mazzanti C, Lipsky RH, Goldman D (2002) Determination of Melting Temperature for Variant Detection Using dHPLC: A Comparison Between an Empirical Approach and DNA Melting Prediction Software. Genet Test 6(3):169-176

122. Sahasrabudhe PV, Ponland RT, Gmeiner WH (1995) Effects of site-specific substitution of 5-fluorouridine on the stabilities of duplex DNA and RNA. Nucleic Acids Res 23(19):3916-3921

123. Saltz LB, Cox JV, Blanke C, Rosen LS, Fehrenbacher L, Moore M J, Maroun JA, Ackland SP, Locker PK, Pirotta N, Elfring GL, Miller LL (2000) Irinotecan plus fluorouracil and leucovorin for metastatic colorectal cancer. Irinotecan study group. N Engl J Med 343:905-914

124. Santi DV, McHenry CS, Sommer H (1974) Mechanism of Interaction of Thymidylate Synthetase with 5-Fluorodeoxyuridylate. Biochemistry 13(3):471-481

125. Sawyer RC, Stolfi RL, Martin DS, Spiegelman S (1984) Incorporation of 5-fluorouracil into murine bone marrow DNA in vivo. Cancer Res 44:1847-1851

126. Schellens JHM (2007) Capecitabine. Oncologist 12:152-155

127. Schuetz JD, Collins JM, Wallace HJ, Diasio RB (1986) Alteration of the secondary structure of newly synthesized DNA from murine bone marrow cells by 5-fluorouracil. Cancer Res 46:119-123

128. Schwartz EL, Baptiste N, Wadler S, Makower D (1995) Thymidine Phosphorylase Mediates the Sensitivity of Human Colon Carcinoma Cells to 5-Fluorouracil. J Biol Chem 270(32)19073-19077

129. Sergent C, Franco N, Chapusot C, Lizard-Nacol S, Isambert N, Correia M, Chauffert B (2002) Human colon cancer cells surviving high doses of cisplatin or oxaliplatin *in vitro* are not defective in DNA mismatch repair proteins. Cancer Chemother Pharmacol 49:445-452

130. Sohn K, Croxford R, Yates Z, Lucock M, Kim Y (2004) Effect of Methylenetetrahydrofolate Reductase C677T Polymorphism on chemosensitivity of Colon and Breast Cancer Cells to 5-Fluorouracil and Methotrexate. J Natl Cancer Inst 96:134-144

131. Souglakos J, Androulakis N, Syrigos K, Polyzos A, Ziras N, Athanasiadis A, Kakolyris S, Tsousis S, Kouroussis C, Vamvakas L, Kalykaki A, Samonis G, Mavroudis D, Georgoulias V (2006) FOLFOXIRI (folinic acid, 5-fluorouracil, oxaliplatin and irinotecan) vs FOLFIRI (folinic acid, 5-fluorouracil and

irinotecan) as first-line treatment in metastatic colorectal cancer (MCC): a multicentre randomised phase III trial from the Hellenic Oncology Research Group (HORG). Br J Cancer (2006) 94, 798 - 805

132. Statistics South Africa (2007) General Household Survey. Release P3018:1-53

133. Stephens M and Scheet P (2005) Accounting for decay of linkage disequilibrium in haplotype inference and missing data imputation. Am J Hum Genet 76:449-462.

134. Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. Am J Hum Genet 68:978-989

135. Stoehlmacher J, Park DJ, Zhang W, Groshen S, Tsao-Wei DD, Yu MC, Lenz HJ (2002) Association Between Glutathione S-Transferase P1, T1, and M1 Genetic Polymorphism and Survival of Patients With Metastatic Colorectal Cancer. J Natl Cancer Inst 94(12):936-942

136. Strathdee G, MacKean M, Illand M, Brown R (1999) A role for methylation of the hMLH1 promoter in loss of hMLH1 expression and drug resistance in ovarian cancer. Oncogene 18:2335-41

137. Takahashi T, Min Z, Uchida I, Arita M, Watanabe Y, Koi M, Hemmi H (2005) Hypersensitivity in DNA mismatch repair-deficient colon carcinoma cells to DNA polymerase reaction inhibitors. Cancer Lett 220:85-93

138. The Advanced Colorectal Cancer Meta-Analysis Project (1992). Modulation of fluorouracil by leucovorin in patients with advanced colorectal cancer: evidence in terms of response rate. J Clin Oncol 10 :896-903

139. The Advanced Colorectal Cancer Meta-analysis Project (1994) Meta-analysis of randomized trials testing the biochemical modulation of fluorouracil by methotrexate in metastatic colorectal cancer. J Clin Oncol 12:960-969 [ABSTRACT]

140. Tuchman M, Stoeckeler JS, Kiang DT, O'Dea RF, Ramnaraine ML (1985) Familial pyrimidine and pyrimidinemia associated with severe fluorouracil toxicity. N Engl J Med 313:245-249

141. Ulrich CM, Bigler J, Velicer CM, Greene EA, Farin FM, Potter JD (2000) Searching Expressed Sequence Tag Databases: Discovery and Confirmation of a Common Polymorphism in the Thymidylate Synthase Gene. Cancer Epidemiol Biomarkers Prev 9:1381-1385

142. Umar A, Boland CR, Terdiman JP, Syngal S, de la Chapelle A, Rüschoff J, Fishel R, Lindor NM, Burgart LJ, Hamelin R, Hamilton SR, Hiatt RA, Jass J, Lindblom A, Lynch HT, Peltomaki P, Ramsey SD, Rodriguez-Bigas MA, Vasen HFA, Hawk ET, Barrett JC, Freedman AN, Srivastava S. (2004) Revised Bethesda Guidelines for Hereditary Non-polyposis Colorectal Cancer (Lynch Syndrome) and Microsatellite Instability. J Nat Cancer Inst 96(40):261-268

143. Vaisman A,Varchenko M, Umar A, Kunkel TA, Risinger JI, Barrett JC, Hamilton TC, Chaney SG (1998) The role ofhMLH1, hMSH3, and hMSH6 defects in cisplatin and oxaliplatin resistance: correlation with replicative bypass of platinum-DNA adducts. Cancer Res 58:3579-85

144. Valle L, Carbonell P, Fernandez V, Dotor AM, Sanz M, Benitez J, Urioste M (2007) MLH1 germline epimutations in selected patients with early-onset non-polyposis colorectal cancer. Clin Genet 71:232-237

145. Van Kuilenburg ABP, Muller EW, Haasjes J, Meinsma R, Zoetekouw L, Waterham HR, Baas F, Richel DJ, van Gennip AH (2001) Lethal Outcome of a Patient with a Complete Dihydropyrimidine Dehydrogenase (DPD) Deficiency after Administration of 5-Fluorouracil: Frequency of the Common IVS14+1G>A Mutation Causing DPD Deficiency. Clin Cancer Res 7:1149-1153

146. Vasen HFA (2005) Clinical description of the Lynch syndrome [hereditary nonpolyposis colorectal cancer (HNPCC)]. Fam Cancer 4:219-225

147. Venook A (2005) Critical evaluation of current treatments in metastatic colorectal cancer. Oncologist 10:250-261

148. Viguier J, Boige V, Miquel C, Pocard M, Giraudeau B, Sabourin J, Ducreux M, Sarasin A, Praz F (2005) ERCC1 Codon 118 Polymorphism Is a Predictive Factor for the Tumor Response to Oxaliplatin/5-Fluorouracil Combination Chemotherapy in Patients with Advanced Colorectal Cancer. Clin Cancer Res 11(17):6212-6217

149. Vreken P, Van Kuilenburg ABP, Meinsma R, Beemer FA, Duran M, Van Gennip AH (1998) Dihydropyrimidine dehydrogenase deficiency: A novel mutation and expression of missense mutations in E. coli. J Inher Metab Dis 21:276-279

150. Vreken P, Van Kuilenburg ABP, Meinsma R, van Gennip AH (1997) Dihydropyrimidine dehydrogenase (DPD) deficiency: identification and expression of missense mutations C29R, R886H and R235W. Hum Genet 101:333–338.

151. Warusavitarne J and Schnitzler M (2006) The role of chemotherapy in microsatellite unstable (MSI-H) colorectal cancer. Int J Colorectal Dis 22(7):739-748

152. Watanabe Y, Koi M, Hemmi H, Hoshai H, Noda K (2001) A change in microsatellite instability caused by cisplatin-based chemotherapy of ovarian cancer. Br J Cancer 85:1064-1069

153. WAVEMAKER™ Software Manual (2002)

154. Wei SH, Brown R, Huang TH (2003) Aberrant DNA methylation in ovarian cancer: is there an epigenetic predisposition to drug response? Ann N Y Acad Sci 983:243-50

155. Wei X, McLeod HL, McMurrough J, Gonzalez FJ, Fernandez-Salguero P (1996) Molecular Basis of the Human Dihydropyrimidine Dehydrogenase Deficiency and 5-Fluorouracil Toxicity. J Clin Invest 98(3):610-615

156. Wilkinson DS and Pitot HC (1973) Inhibition of Ribosomal Ribonucleic Acid Maturation in Novikoff Hepatoma Cells by 5-Fluorouracil and 5-Fluorouridine. J Biol Chem 248(1):63-68

157. Wilkinson DS and Crumley J (1976) The Mechanism of 5-Fluorouridine Toxicity in Novikoff Hepatoma Cells. Cancer Res 36:4032-4038

158. Wilkinson DS and Crumley J (1977) Metabolism of 5-Fluorouracil in Sensitive and Resistant Novikoff Hepatoma Cells. J Biol Chem 252(3):1051-1056

159. Xiao W and Oefner PJ (2001) Denaturing high-performance liquid chromotography: A review. Hum Mutat 17(6):439-474

160. Yanaihara N, Caplen N, Bowman E, Seike M, Kumamoto K, Yi M, Stephens RM, Okamoto A, Yokota J, Tanaka T, Calin GA, Liu C, Croce CM, Curtis C. Harris CC (2006) Unique microRNA molecular profiles in lung cancer diagnosis and prognosis. Cancer Cell 9 :189-198

161. Yokota H, Fernandez-Salguero P, Furuya H, Lin K, McBride OW, Podschun B, Schnackerz KD, Gonzalez FJ (1994) cDNA cloning and chromosome mapping of human dihydropyrimidine dehydrogenase, an enzyme associated with 5-fluorouracil toxicity and congenital thymine uraciluria. J Biol Chem 269:23192-23196

162. Yoshioka A, Tanaka S, Hiraoka O, Koyama Y, Hirota Y, Ayusawa D, Seno T, Garrett C, Wataya Y (1987) Deoxyribonucleoside triphosphate imbalance. 5-Fluorodeoxyuridine-induced DNA double strand breaks in mouse FM3A cells and the mechanism of cell death. J Biol Chem 262:8235-8241

163. Zdraveski ZZ, Mello JA, Farinelli CK, Essigmann JM, Marinus MG (2002) MutS preferentially recognizes cisplatin- over oxaliplatin-modified DNA. J Biol Chem 277:1255-60.

164. Zhu H, Guo W, Zhang L, Davis JJ, Teraishi F, Wu S, Cao X, Daniel J, Smythe WR, Fang B (2005) Bcl-XL small interfering RNA suppresses the proliferation of 5-fluorouracil-resistant human colon cancer cells. Mol Cancer Ther 4(3):451-456

Websites utilised

1) http://ihg2.helmholtz-muenchen.de/cgi-bin/hw/hwa1.pl
For determination of Hardy-Weinberg equilibrium
2) www.PharmGKB.org
3) www.ncbi.nlm.nih.gov
4) www.questdiagnostics.com
5) www.ensemble.org
6) www.nfdht.nl
7) http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi
8) www.idtdna.com/analyzer/Applications/OligoAnalyzer/ Default.aspx
9) www.ncbi.nlm.nih.gov/BLAST/
10) www.mbio.ncsu.edu/BioEdit/bioedit.html
11) www.fda.gov/
12) www.med.mun.ca/MMRvariants
13) www.pfizer.com
14) www.sanofisynthelabo.us
15) www.roche.us
16) www.gene.com
17) www.imclone.com
18) www.doh.gov.za/docs/dns-f.html

LIST OF ADDENDUMS

ADDENDUM A

General recipes and protocols

Dilution of primers to a working stock of 20µM

$$\frac{OD \times 35}{-nmer \times 330} = X\mu M$$

$$\frac{20\mu M}{X\mu M} \times 200\mu l = volume\,Y$$

The Optical Density (OD) is a measure of the concentration of primer in the stock solution and -nmer specifies the length of the primer. Volume Y is the volume of concentrated primer solution that is made up to 200µl with sterile distilled water and vortexed briefly.

Tris borate EDTA Buffer (TBE)

10×TBE buffer: 216g Tris [hydroxymethyl] amino methane (TRIS) (USB); 110g boric Acid *(Promega)*, 14.8g EDTA made up to 2L with distilled water. Dilute to 1× as required.

Agarose gels (1%, 2%, 3%)

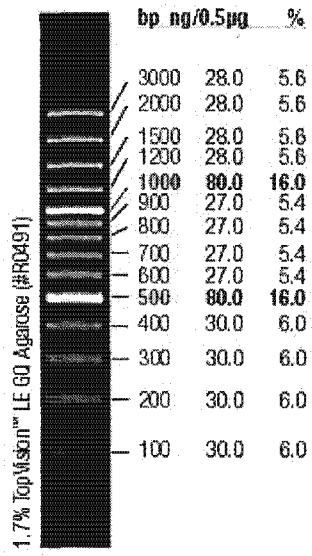
1%: 1g of Agarose powder dissolved in a final volume of 100ml 1×TBE buffer; with 5µl ethidium bromide (5ng/µl).

2%: 2g of Agarose powder dissolved in a final volume of 100ml 1×TBE buffer; with 5µl ethidium bromide (5ng/µl).

3%: 3g of Agarose powder dissolved in a final volume of 100ml 1×TBE buffer; with 5µl ethidium bromide (5ng/µl).

DNA Size Standard:

GeneRuler™ 100bp DNA Ladder Plus (Fermentas International Inc, Canada)



Agarose DNA Loading Dye

0.25% bromophenol blue (0.125g); 40% sucrose (20g) made up to a final volume of 50ml with distilled water.

ADDENDUM B

Supplementary information to experiments

Index:

REQUEST FOR MOLECULAR STUDIES (DNA)

Rommar laboratory
Division of Human Genetics (Rm N3.14)
Faculty of Health Sciences
University of Cape Town
Anzio Road
Observatory 7925

Tel: (021) 406 6373 Fax: (021) 4066826

Please fill in all the information requested:

Blood should be drawn in 2 plastic EDTA Tubes (Purple top) +/- 10ml each using a yellow barrel. Each tube should be inverted to mix and should be clearly labelled with the patient's name and DOB. Keep blood in fridge at 4°C until able to send to laboratory.

Please DO NOT send specimens on ice or frozen.

Surname: VAN NIEKERK First Name(s) _____

New Family: Yes ☐ No ☐ Please fill in Family name / Lab Nr: _____

Medical aid: _____ Medical aid No: _____

Sex: M ☐ F ☐ Date of Birth: Year: _____ Month: _____ Day: _____

Ethnic Origin: Black ☐ Indian ☐ Mixed ancestry ☐ Caucasian ☐ Other ☐ _____

Contact Address: _____ Fax: _____

_____ Tel: _____

_____ Cell phone: _____

E-mail: _____

Referring Doctor/Sister: _____

Hospital or Address: _____ Fax: Tel: _____

Bloods taken for: Research ☐
Disease Status: Affected ☐ At Risk ☐ Spouse ☐

Bloods taken for: Diagnostic ☐
Disease Status: Affected ☐

Date of Next Visit: / /

Bloods taken for: Predictive ☐
Disease Status: At Risk ☐ Spouse ☐

Bloods taken for: Post Test ☐

Forward my predictive test results to: _____ Address: _____

_____

For Laboratory use only:

DNA number: _____ *Vol.Blood:* _____ *(ml)* *Other:* _____

Date Received: Year: _____ *Month:* _____ *Day:* _____ *Computer Index No:* _____

***** NOTE - PLEASE INSERT A FAMILY PEDIGREE DRAWING ON THE REVERSE OF THIS FORM**

CONSENT FOR DNA ANALYSIS AND STORAGE

1. I, _____, request that an attempt be made using genetic material to assess the probability that I / my child / my unborn child **(DELETE WHERE NOT APPLICABLE)** might have inherited a disease-causing mutation in the gene for:

   _____

2. I understand that the genetic material for analysis is to be obtained from: blood cells / skin sample / other (specify) **(DELETE WHERE NOT APPLICABLE)**

3. I request that no portion of the sample be stored for later use. ☐ **(MARK IF APPLICABLE)**
   OR
   I request that a portion of the sample be stored indefinitely for **(DELETE WHERE NOT APPLICABLE)**
   (a) possible re-analysis
   (b) analysis for the benefit of members of my immediate family
   (c) research purposes, subject to the approval of the University of Cape Town Research Ethics Committee provided that any information from such research will remain confidential.

4. The results of the analysis carried out on this sample of stored biological material will be made known to me, via my doctor, in accordance with the relevant protocol, if and when available. In addition, I authorise that they may be made known to: **(DELETE WHERE NOT APPLICABLE)**

   (a) other doctors involved in my care _____

   (b) the following family members: _____

   other: _____

5. I authorise / do not authorise my doctor(s) to provide relevant clinical details to the Division of Human Genetics, UCT. **(DELETE WHERE NOT APPLICABLE)**

6. I have been informed that:
   (a) There are risks and benefits associated with genetic analysis and storage of biological material and these have been explained to me.
   (b) The analysis procedure is specific to the genetic condition mentioned above and cannot determine the complete genetic makeup of an individual.
   (c) The genetics laboratory is under an obligation to respect medical confidentiality.
   (d) Genetic analysis may not be informative for some families or family members.
   (e) Even under the best conditions, current technology of this type is not perfect and could lead to incorrect results.
   (f) Where biological material is used for research purposes, there may be no direct benefit to me.

7. I understand that I may withdraw my consent for any aspect of the above at any time without this affecting my future medical care.

8. **ALL OF THE ABOVE HAS BEEN EXPLAINED TO ME IN A LANGUAGE THAT I UNDERSTAND AND MY QUESTIONS ANSWERED BY:**

Signature _____ **Date:** ____/__/__

****Patient signature** _____ **Witnessed consent:** _____

RADIATION AND CHEMOTHERAPY FOLLOW-UP RECORD
NPC NUMBER:.....RT-NUMBER:.....

1. Follow-up Record

The most important fact to establish from the follow-up record is to determine whether a patient had a recurrence of any cancer since they were admitted to the hospital for their primary colon cancer.

1. Photocopy the follow-up summary document in the front of the RT-folder.
2. Briefly go through the detailed notes to establish whether a patient had presented with another cancer since their initial admission for the primary cancer
   .....
3. Photocopy all pathology forms that are found in the RT-folder.

2. Chemotherapy record (please record the following information)

A. The Primary Site of the cancer .....
B. The Start date/ First date for chemotherapy .....
C. The End date for chemotherapy .....
D. The patient's: i. Weight in kg .....
   ii. Height in cm .....
   iii. Surface area in m^2 .....
E. The drugs prescribed for the patient: Drug1: .....
   Drug2: .....
F. The dose ordered in mg: Drug1: .....
   Drug2: .....
G. The number of chemotherapy cycles that had been completed for a patient:
   .....
H. If the prescribed cycles were stopped due to complications, please indicate the reason for doing so (if listed)
   .....
   .....

3. Radiotherapy treatment sheet (please record the following information)

A. The site that was treated.....
B. The start date of radiotherapy treatment.....
C. The end date of the radiotherapy treatment.....
D. The instrument/machine that was used for treatment .....
E. SSD .....
F. Field size .....
G. Norm TD/CD .....
H. Dose/Frac .....
I. Number (#) of fractions .....
J. Total Dose .....
K. Complications

.....
.....
.....
.....

Table 2.2.1a: Bethesda panel of microsatellite markers and associated primers (adapted from Dietmaier *et al.* 1997)

Marker	Repeat Motif & Primer name	Fluorescent Label	OD/ml	Concentration (µM)	Size Range
D2S123	Dinucleotide **$(CA)_{13}TA(CA)_{15}(T/G\ A)_7$**				
	D2S123 F: 5'AAACAGGATGCCTGCCTTTA 3'	FAM	336.1	1782.3	197-227bp
	D2S123 R: 5'GGACTTTCCACCTATGGGAC 3'		499.2	2647.2	
D17S250	Dinucleotide **$(TA)_7$..... $(CA)_{24}$**				
	D17S250 F: 5'GGAAGAATCAAATAGACAAT 3'	FAM	388.9	2062.3	140-170bp
	D17S250 R: 5'GCTGGCCATATATATATTTAAACC 3'		506.5	2238.3	
D5S346	Dinucleotide **$(CA)_{26}$**				
	D5S346 F: 5'ACTCACTCTAGTGATAAATCGGG 3'	TET	157.4	758	96-122bp
	D5S346 R: 5'AGCAGATAAGACAAGTATTACTAG 3'		446.9	1887	
BAT25	Mononucleotide **TTTT.T.TTTT. $(T)_7A(T)_{25}$**				
	BAT25 F: 5'TCGCCTCCAAGAATGTAAGT 3'	HEX	264.8	1337	90-125bp
	BAT25 R: 5'TCTGGATTTTAACTATGGCTC 3'		344.8	1741	
BAT26	mononucleotide **T_5....A_{26}**				
	BAT26 F: 5'TGACTACTTTTGACTTCAGCC 3'	FAM	63.2	306	80-120bp
	BAT26 R: 5'AACCATTCAACATTTTTAACC 3'		377.7	1796	

Table 2.2.1b: Reaction set-up for PCR amplification to determine MSI

Reagent	**Stock Concentration**	**1× (µl)**	**Final Concentration**
sdH_2O (SABAX)	N/A	5.67	-
Go Taq Buffer (Promega)	5×	2	1×
$MgCl_2$	50mM	0.4	1.5mM/2.0mM
dNTPs (Bioline)	5mM	0.5	200µM
Forward Primer	20µM	0.25	0.4µM
Reverse Primer	20µM	0.25	0.4µM
Go Taq Polymerase (Promega)	5 Units/µl	0.02	0.1 Units/µl
DNA	10ng/µl	1	0.1ng/µl
Final Volume		10	

Table 2.2.1c: General cycling conditions for PCR assay of microsatellite instability

Denaturation	95°C	for	5 min	1 cycle
Amplification	94°C	for	30 s	35 cycles
	50°C	for	30 s	
	72°C	for	40 s	
Final Extension	72°C	for	7 min	1 cycle

Annotation of *DPYD* with positions of primers relative to variant sequences

IVS14+1 G>A

5'caagaagtaatttagctaaagctaacaaatacttatttacctttttatttgcaagtagtttatgttcaattctaatttaatgt
atattaaaaattcctctgcaaaaatgtgagaagggacctcataaaatattgtcatatggaaatgagcagataataaag
atlalagcttttcttgtcaaaaggagactcaatatcttact>>>>>**ctttcatcagGACATTGTGAC**AAATG
TTTCCCCCAGAATCATCCGGGGAACCACCTCTGGCCCCATGTATGGCCCTGGA
CAAAGCTCCTTTCTGAATATTGAGCTCATCAGTGAGAAAACGGCTGCATATTGG
TGTCAAAGTGTCACTGAACTAAAGGCTGACTT^TCCAGACAA^C^g (476002) (c/t)
(dbSNP:rs3918290)taagtgtgatttaacatctaaaacaagagaattggcataagttgggtgaatg**ttt**attaaa
catccaattcataggcttataaatattaatgtgtatattttattaaagaatctgccagttgctttgctgatgcatagaaagat
aaaaaagaaagaaaagctcaagaactcatlaaaaacccacacaatgtgaagctttgttataaatgggtgccatgta
agatggaagaagtatctacataagcagaaggaagagaaatgaaatactcattttattgagttggttttcactgtatgtg
gctggtatttatgaaggtgatgaccttaggaagaaattgtaaactatlaaatcatttlaaatataaacctgaggcagaag
cagcatatcttcctatgaagtct^atatttttttcagtgggaaataatttattaaataattttaacactcctatactaaattagtc
aactatctttataaaatctttttttttctcttaaagaaccatgcctaaatattgcagtaaaatcaggcttgaatataatagata
atattttctcttcaatagtctgattagggtgatccttaatg^tcattcatttattgtcoctatcoctcccacctcoctctcccactt
cctatttttttcctactcctttttctctctctcatccttcctctctttcccctcccctcctcaatttctgtctc^tttcaaa
tatacacaacttatttctctttctttcattctttctttcaatgaattgttagt**gaatggaggttgaggtttgag**<<<<<<<gc
aatgtgtg^ggtttatgtca

R29C

tgagccaggtggatcgcttgagaccaggagttcaagaccagcctgggcaacatggcgaaacctcatctctattaaa
aatataaaaattagccaggtgtggtagcgtacgcctgtagtccgagctacttgggagactaaggtgggaggatcgcc
tgagcctggaaggcagaggttgcagtgaactgagattgtaccactgcactccagcctgggtgacaaagtgagaga
gaccgtgtctcaaaataaataaaaataaaataaaaataagtttaaacaaa^tgccaacatatttccatatatttacaca
ttgttt>>>>>>**atgctgtctttagAGTATCCTGG**CTTTAAATCCTCGAACACAAACTCATG
CAACTCTG^C (42731) (g/a) (dbSNP:rs1801265) GTTCCACTTCGGCCAAGAAAT
TAGACAAGAAACATTGGAAAAGAAATCCTGATAAGAACTGCTTTgtaagtaccactgat
acactatttcatgctgaaaattacct**cactccacacattgtaaggcaa**<<<<<<gattttacatacacttagaaata
aggtattaaagtacagccgtgattttaaaaatattttaaatatttaatattt

I543V

actgcctttgaaattaaaaggccatattatatgaattcagttcactgctcactgacttttˆtgccagttttttttcattgtaaaca
gactgttgtaatctatattaattc^ggatgctgtgttgaagtgatttttttggtaaaatattatatggacaatttaga^tgtaat
atgaaaccaagtattggtttgtattttgcagTCACAATATGGAGC>>>>>>**TTCCGTTTCTGCCAA**
GCCTGAACTACCCCTCTTTTACACTCCTATTGATCTGGTGGACATTA^GTGTAGA
AATGGCCGGATTGAAGTTT^**A(410221) (t/c) (dbSNP:rs1801159)** TAAATCCTTT
TGGTCTTGCTAGCGCAACTCCAGC**CACCAGCACATCAATGATTC**<<<<<<GAAG
AGCTTTTGAAGCTGGATGGGGTTTTGCCCTCACCAAAACTTTCTCTCTTGATAA
Ggtaagaaaatattattgaagtcatatagaaatgtctat^c^atatattttaattaaggtataaacatt

Annotation of *TYMS* with positions of primers relative to variant sequence

5'UTR VNTR

cttagagaaggcgcggtcgaccagacggttcccaaagggcgcagtcctcc<ca>gccaccgcacctg<ca>tccaggttcc

>>>>>>cgggtttcctaagactctcagctgtggccctgggctccgttctgtgccacacccgtggctc

ctgcg*tttccccctggcgcacgctctctagagcg*gggggccg<cc>gcgaccccgccgagcaggaagaggcggagcgcggga

cggccgcgggAAAAGGCGCGCGGAAGGGGTCCTGCCA**CCGCGCCACTTGGCCTGCCT**

CCGTCCCG/CCGCGCCACTTGGCCTGCCTCCGTCCCGCCGCGCCACTTCGCCTG

CCTCCGTCCCCCGCCCGCCGCGCCATGCCTGTGGCCGGCTCGGAGCTGCCGCGCCG

GCCCTTGCCCCCCGCCGCACAGGAGCGGGACGCCGAGCCGCGTCCGCCGCACGGGG

AGCTGCAGTACCTGGG**GCAGATCCAACACATCCTC<<<<<<**CGC

TGCGGCGTCAGGAAGGACGACCGCACGGGCACCGGCACCCTGTCGTATTCGGCATG

CAGGCGCGCTACAGCCTGAGAGgtgacgccgcgggcccctgcgggacgggtggcgggaaggagggaggcg

cggctggggagagcgctcgggagctgccgggcgctgcggaccccgttagtcctaacctcaatcctg<cg>agggaggggac

gcatcgtcctcctcgccttacagacgccgaaacgg

3'UTR indel

caacaggtcgtacaattatggcaaa^ataatggccttattttgttttagCTTCAGC^GAGAACCCAGACCTTTCCC

AAAGCTCAGGATTCTTCGAAAAGTTGAGAAAATTGATGACTTCAAAGCTGAAGACTTTCA

GATTGAAGGGTACAATCCGCATCCAACTATTAAAATGGAAATGGCTGTT^TAGGGTGCTT

TCAAAGGAGCT^CGAAGGATATTGTCAGTCTTTAGGGGTTGGGC^TGGATGCCGAGGTA

AAAGTTCTTTTTGCTCTAAAAGAA^AAAGGAACTAGGTCAAA**>>>>>>AATCTG**

TCCGTGACCTATCAGTTATTAATTTTTAAGGATG^TTGCCACTGGCAAATG

TAACTGTGCCAGTTCTTTCCATAATAAAAGGCTTTGAGTTAACTCACTGAGG^GTATCTG

ACAATGCTGAGGTTATGAACAAAGTGAGGAGAATGAAATGTATGTGCTCTTAGCAAAAA^

C^ATGTATGTGCATTT^CAATCCCACGTACTTATAAAGAAGGTTGGTGAATTTCACAAGCT

ATTTTTGGAATATTTTTAGAATATTTTAAGAATTTCACAAGCTATTCCCTCAAATCTGA^GG

GAGCTGAGTAACACCATCGATCATGATGTAGAGTGTGGTTATGAACT**^TTAAAG**

(20794)(ttaaag/-)(dbSNP:rs16430) TTA^TAGTTGTTTTATATGTTGCTATAATAAAGA

AGTGTTCTGCattcgtccacgctttgttcattctgtactgccacttatctgctcagttccttcctaaaatagattaaagaactctcctt

aagtaaacatgtg**ctgtattctggtttggatgcta<<<<<<**cttaaa^agagtatattttaga

aataatagtgaatatattttgccctatttttctcatttaactgcatcttatcctcaaaatataatgaccatttaggatagagttttttttttttt^t

ttaaactt^ttataaccttaa^agggttatttaaaataatctatggactaccattttgccctc^attagcttcagcatggtgtgacttctcta

ataatatgcttagattaagcaaggaaaagatgcaaaaccacttcggggttaatcagtgaaatatttt^tcccctcgtgcataccagat

accccccggtgttgcacgactatttttatttctgctaatttatgacaagtgttaaacagaacaaggaattattccaacaagttatgcaaca

^tgttgcttattttcaa

Table 2.3.3: Primer information for analysis of variants within the pharmacogenes *TYMS* and *DPYD*

Primer Name	Length	Tm (°C)	GC%	Primer Sequence 5' to 3'	Designer
DPYD_IVS14_F	36	60	43	TGTAAAACGACGGCCAGTCTTTCATCAGCACATTCTGAC	J Meyer
DPYD_IVS14_R	21	60	43	CACCAACTTATGCCAATTCTC	J Meyer
DPYD_543_F	36	60	56	TGTAAAACGACGGCCAGTTTCCGTTTCTCCCAACCC	J Meyer
DPYD_543_R	20	60	45	GAATCATTGATGTGCTGGTG	J Meyer
DPYD_29_F	22	53	43	ATGCTCTCTTTAGACTATCCTCG	M Akinya
DPYD_29_R	22	53	45	TTGCCTTACAATGTGTGGAGTG	M Akinya
TYMS_VNTR_F	21	54	52	FAM *CGCGTTTCCTAAGACTCTCAG	J Meyer
TYMS_VNTR_R	20	54	55	GAGGATGTGTTGGATCTGCC	J Meyer
TYMS_del_F	21	52	50	AATCTGTCCGTGACCTATCAG	M Akinya
TYMS_del_R	22	51	41	TAGCATCCAAACCACAATACAG	M Akinya
SNaPShot™ Internal Primers					
rs3918290 (DPYD IVS14 +1G>A SNP)	42	N/A	N/A	✷**CAATCAACAATCAACAA**ICAACAGGCTGACTTTCCACACAAC	M Akinya
rs1801159 (DPYD I543V SNP)	30	N/A	N/A	**CAATCAA**CCTAGCAAGACCAAAAGGATTTA	M Akinya
rs1801265 (DPYD_R29C_SNP)	38	N/A	N/A	**CAATCAACAATCAACAATCA**CAAACTCATCCAACT	M Akinya

✷ Non-homologous 5' tail tandem repeat

*FAM: Fluorecent label

Table 2.3.4a: Reaction set-up for multiplex PCR assay of pharmacogenomic variants

Reagent	Stock Concentration	1× (µl)	Final Concentration
sdH_2O (SABAX)	N/A	VARIABLE	N/A
Go Taq Buffer (Promega)	5×	5	1×
$MgCl_2$ (Merck)	50mM	1	1.5mM
dNTPs (Bioline)	5mM	1	200µM
Forward Primer of rs3918290, rs1801265 and rs1801159	20µM	0.5/variant	0.4µM
Reverse Primer of rs3918290, rs1801265 and rs1801159	20µM	0.5/variant	0.4µM
Go Taq Polymerase (Promega)	5 Units/µl	0.1	0.5 Units/µl
DNA	100ng/µl	1	4ng/µl
Final Volume		25µl	

Table 2.3.4b: General cycling conditions for multiplex PCR assay of pharmacorelevant variants

Denaturation	95°C	for	5 min	1 cycle
Amplification	94°C	for	60 s	30 cycles
	56°C	for	60 s	
	72°C	for	80 s	
Final Extension	72°C	for	7 min	1 cycle

Table 2.3.5b: Reaction mix for Cycle Sequencing of *DPYD* variants, rs3918290, rs1801265 and rs1801159

Reagent	**Stock Concentration**	**1x (µl)**
Termination mix (Applied Biosystems)	N/A	1
Big Dye Sequencing Buffer (Applied Biosystems)	5x	2
Reverse/Forward primer of rs3918290, rs1801265 and rs1801159	20µM	1
Purified PCR products	~50ng/µl	3
sdH_2O	N/A	3
Final Volume		**10**

Table 2.3.5c: Cycling conditions for sequencing of *DPYD* variants

96°C	for	5 min	1 cycle
96°C	for	30 s	25 cycles
50°C	for	15 s	
72°C	for	4 min	

Table 2.3.6a: Reaction set-up for PCR assay of *TYMS* VNTR (rs45445694)

Reagent	Stock Concentration	1× (µl)	Final Concentration
sdH_2O	N/A	14.4	N/A
Go Taq Buffer	5×	5	1×
Glycerol (Merck)	50%	2.5	5%
dNTPs	5mM	1	200µM
Forward Primer	20µM	0.5	0.4µM
Reverse Primer	20µM	0.5	0.4µM
Go Taq Polymerase	5 Units/µl	0.1	0.5 Units/µl
DNA	100ng/µl	1	4ng/µl
Final Volume		25µl	

Table 2.3.6b: General cycling conditions for PCR assay *TYMS* VNTR (rs45445694)

Denaturation	95°C	for	5 min	1 cycle
Amplification	94°C	for	30 s	35 cycles
	52°C	for	30 s	
	72°C	for	40 s	
Final Extension	72°C	for	7 min	1 cycle

Table 2.3.6c: Reaction mix for cycle sequencing of *TYMS* 5' UTR VNTR (rs45445694)

	Stock Concentration	1x (µl)
Termination mix	N/A	1
Big Dye Sequencing Buffer	5x	2
Reverse primer of rs45445694	20µM	1
Purified PCR products	~50ng	3
Glycerol	50%	1
sdH_20	N/A	2
Final Volume		**10**

Table 2.3.6d Cycling conditions for cycle sequencing of *TYMS* 5'UTR VNTR (rs45445694)

98°C	for	10 min	1 cycle
98°C	for	1 min	40 cycles
55°C	for	1 min	
72°C	for	2 min	

Table 2.3.7a: Reaction set-up for amplification of *TYMS* 3'UTR insertion/deletion (rs16430)

Reagent	Stock Concentration	1× (µl)	Final Concentration
sdH_2O	N/A	16.9	N/A
Go Taq Buffer	5×	5	1×
dNTPs	5mM	1	200µM
Forward Primer of rs16430	20µM	0.5	0.4µM
Reverse Primer of rs16430	20µM	0.5	0.4µM
Go Taq Polymerase	5 Units/µl	0.1	0.5 Units/µl
DNA	100ng/µl	1	4ng/µl
Final Volume		25µl	

Table 2.3.7b: Cycling conditions for amplification of *TYMS* 3'UTR insertion/deletion (rs16430)

Denaturation	95°C	for	5 min	1 cycle
Amplification	94°C	for	30 s	30 cycles
	54°C	for	30 s	
	72°C	for	40 s	
Final Extension	72°C	for	7 min	1 cycle

ADDENDUM C

Index of supplementary files on CD

A. EXCEL document with Frequencies of *DPYD* SNPs on AFFY chip

B. EXCEL document with Frequencies of *TYMS* SNPs on AFFY chip

C. Folder containing Fisher's Exact Test files:

   Folder containing files with allele frequencies of *DPYD* and *TYMS* variants for all Black African population groups

   Folder containing files with the results from Fisher's Exact test for all population comparisons

   Data for Figure 3.4.4A

   Data for Figure 3.4.4B

   Ensembl information about significant SNPs of interest

D. Folder containing Output files from PHASE v2.1 analysis of *TYMS* in Mixed Ancestry cohort

E. Folder containing Output files from PHASE v2.1 analysis of *DPYD* in Mixed Ancestry cohort

F. Folder containing Output files from PHASE v2.1 analysis of *TYMS* in Black Ancestry cohort

G. Folder containing Output files from PHASE v2.1 analysis of *DPYD* in Black Ancestry cohort

H. Folder containing Output files from PHASE v2.1 analysis of *TYMS* in Caucasian cohort

I. Folder containing Output files from PHASE v2.1 analysis of *DPYD* in Caucasian cohort