ANALYSIS OF VARIATIONS IN HYPOXIA-PATHWAY GENES AND MITOCHONDRIAL DNA AS PROGNOSTIC MARKERS IN COLORECTAL CANCER PATIENTS

ASAN MEERA SAHIB HAJA MOHIDEEN

Analysis of variations in hypoxia-pathway genes and mitochondrial DNA as prognostic markers in colorectal cancer patients

by

©Asan Meera Sahib Haja Mohideen

A thesis submitted to the School of Graduate Studies

In partial fulfilment of the requirements for the degree of Master of Science – Medicine

Discipline of Genetics,

Faculty of Medicine,

Memorial University of Newfoundland

October 2013

St. John's Newfoundland and Labrador

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

LIST OF ABBREVIATIONS

5-FU:	5–Fluorouracil
AD:	Allelic Discrimination
AFAP:	Attenuated Familial Adenomatous Polyposis
AHR:	Aryl Hydrocarbon Receptor
AJC:	American Joint Committee
ANT:	Adenine Nucleotide Translocase
APC:	Adenomatous Polyposis Coli
ATP:	Adenosine Triphosphate
AQ:	Allelic Quantification
ARNT:	Aryl Hydrocarbon Receptor Nuclear Translocator
ATM:	Ataxia Telangiectasia Mutated
BER:	Base excision repair
bHLH:	Basic-Helix-Loop-Helix
CHK:	Cell cycle checkpoint kinase
CIMP:	CpG Island Methylator Phenotype

CIN:	Chromosomal Instability
CNV:	Copy Number Variation
CONAN:	Copy Number Analysis
COSMIC:	Catalogue of Somatic Mutations in Cancer
COX2:	Cyclooxygenase-2
DNAPK:	DNA-dependent Protein Kinase
EGFR:	Epidermal Growth Factor Receptor
EPAS 1:	Endothelial PAS Domain Protein 1
FAP:	Familial Adenomatous Polyposis Coli
GADD:	Growth Arrest and DNA Damage
GRP:	Glucose Regulatory Protein
HIC:	Human Investigation Committee
HIF:	Hypoxia Inducible Factor
HNPCC:	Hereditary Non-Polyposis Colorectal Cancer
HRE:	Hypoxia Responsive Element
HBS:	HIF Binding Sites
JPEG:	Joint Photographic Experts Group

KRAS:	Kirsten Rat Sarcoma-2
LD:	Linkage Disequilibrium
LOH:	Loss of Heterozygosity
LOX:	Lysyl Oxidase gene
MGB:	Minor Groove Binder
MIF:	Macrophage Migration Inhibitory Factor
MMR:	Mismatch Repair System
MSDS:	Material Safety Data Sheet
MSI:	Microsatellite Instability
MSI-H:	Microsatellite Instability-High
MSI-L:	Microsatellite Instability-Low
MSS:	Microsatellite stable
mtDNA:	Mitochondrial DNA
nDNA:	Nuclear DNA
NFCCR:	Newfoundland Colorectal Cancer Registry
NL:	Newfoundland and Labrador
ORP:	Oxygen Regulatory Protein

OXPHOS:	Oxidative Phosphorylation
PAS:	Period/ARNT/Single-minded
PCR:	Polymerase Chain Reaction
PPE:	Personal Protective Equipment
PFS:	Progression Free Survival
qPCR:	Quantitative PCR
R_n:	Fluorescence Intensities
ΔR_n:	Normalized Fluorescence Intensities
ROS:	Reactive Oxygen Species
RPM:	Revolution Per Minute
RT:	Room Temperature
SDS:	Sequence Detection Systems
SDS-RQ:	SDS Relative Quantitation
SNP:	Single Nucleotide Polymorphism
TagSNP:	Tagging SNP
TAS:	Terminated Associated Sequences
TBE:	Tris-Borate-EDTA

TE:	Tris-EDTA
TLR:	Toll-like Receptor
TNM:	Tumour-Node-Metastasis
Tris-HCl:	Tris-Hydrochloride
TSG:	Tumour Suppressor Gene
VEGF:	Vascular Endothelial Growth Factor
WHIMS:	Workplace Hazardous Materials Information System
WHO:	World Health Organization
XRE:	Xenobiotic Response Elements

ABSTRACT

Colorectal cancer is a common malignancy, characterized by high incidence and mortality rates. Hypoxia induces angiogenesis, metastasis and aggressive tumour phenotype and hence promotes cancer progression. Similarly, dysfunctions in mitochondria have been shown to contribute to cancer progression. Therefore, in this thesis, genetic markers in both the hypoxia pathway genes and the mitochondrial DNA are hypothesized to be candidate genetic prognostic markers in colorectal cancer.

The aim of this study is to test genetic markers that can predict outcome in colorectal cancer patients. This study consists of two projects: 1) the mitochondrial DNA variations project and 2) the hypoxia pathway Single Nucleotide Polymorphisms (SNPs) project. In the mitochondrial DNA (mtDNA) variations project, six mtDNA polymorphisms and the mtDNA copy number change were tested for their prognostic associations. However, none of the mtDNA markers selected was found to be associated with outcome. The hypoxia pathway SNPs project was done in two phases. In the first phase, 49 tagging SNPs from six hypoxia genes (*HIF1A*, *HIF1B*, *HIF2A*, *LOX*, *MIF* and *CXCL12*) were genotyped in a cohort of 272 colorectal cancer patients (cohort I). In the second phase, 77 tagging SNPs from seven hypoxia genes (*HIF1A*, *HIF1B*, *HIF2A*, *HIF2B*, *HIF3A*, *LOX* and *CXCL12*) were genotyped in a separate cohort of 536 colorectal cancer patients (cohort II).

In phase I, the TT and AT genotypes of the *HIF2A* rs11125070 polymorphism was associated with increased disease free survival (DFS) in the multivariable model

(p=0.004; HR=0.619; 95% CI: 0.446-0.859). In phase II, patients with TC and CC genotypes for the *HIF2A* rs4953352 polymorphism and patients with the GG genotype for the *HIF2B* rs12593988 polymorphism were associated with both reduced overall survival (OS) and DFS in multivariate models. However, the association of the *HIF2A* rs11125070 polymorphism with DFS detected in the phase I cohort was not replicated in the phase II patient cohort, suggesting this association was possibly a false-positive association. Similarly, association of the *HIF2A* rs4953352 polymorphism with OS and DFS detected in the phase II cohort failed to be detected in the phase I patient cohort, which may be due to the small sample size of cohort I. Association of the *HIF2B* rs12593988 polymorphism remains to be tested in another colorectal cancer cohort.

In conclusion, this study shows no evidence of associations between majority of genetic markers in hypoxia pathway genes and mtDNA with prognosis in colorectal cancer. However, whether or not the *HIF2B* rs12593988 polymorphism is associated with prognosis needs to be confirmed by investigating other patient cohorts.

LIST OF RESEARCH OUTPUT AND AWARDS

Manuscripts in preparation:

1) **Asan M. S. Haja Mohideen**, Elizabeth Dicks, Patrick Parfrey, Roger Green, Sevtap Savas. "Relation between genetic variations in mitochondrial DNA (mtDNA) and outcome in colorectal cancer".
2) **Asan M. S. Haja Mohideen**, Angela Hyde, Jessica Squires, Jing Wang, Patrick Parfrey, Ban Younghusband, Roger Green, Sevtap Savas. "Two polymorphisms in the hypoxia inducible factor genes (*HIF2A* and *HIF2B*) are associated with outcome in colorectal cancer".

Oral presentations:

1. **Asan M. S. Haja Mohideen,** Elizabeth Dicks, Ban Younghusband, Patrick Parfrey, Roger Green, Sevtap Savas. "Two Polymorphisms in the Hypoxia Inducible Factor Genes *HIF2A* and *HIF2B* are Associated with Outcome in Colorectal Cancer". The BHCRI Annual Cancer Research Conference, Halifax, NS, Canada, November 5-6, 2012.

2. **Asan M. S. Haja Mohideen**, Elizabeth Dicks, Ban Younghusband, Patrick Parfrey, Roger Green, Sevtap Savas. Genetic variations in mitochondrial DNA and hypoxia pathway genes and clinical outcomes in colorectal cancer. The TargetMeeting 2nd World Cancer Online Conference, Jan 8-11, 2013.

Poster presentation:

1. **Asan M. S. Haja Mohideen**, Angela Hyde, Jessica Squires, Patrick Parfrey, Ban Younghusband, Roger Green, Sevtap Savas. "Relation between polymorphisms in hypoxia pathway and outcome in colorectal cancer", Annual Human Statistical Genetics Meeting, Niagara, ON, Canada, April 29-May 3, 2012.

2. **Asan M. S. Haja Mohideen**, Angela Hyde, Jessica Squires, Elizabeth Dicks, Patrick Parfrey, Ban Younghusband, Roger Green, Sevtap Savas. "Polymorphisms in the hypoxia pathway genes and survival in colorectal cancer patients", Second Annual Human Statistical Genetics Meeting, Quebec, April 21-24, 2013.

Travel Award:

1. $500 travel award by the Beatrice Hunter Cancer Research Institute, Dalhousie University, to attend the Annual BHCRI Research Conference held in Halifax, NS, Canada, November 5-6, 2012.

2. $500 travel award by the Beatrice Hunter Cancer Research Institute, Dalhousie University, to attend the BHCRI's cancer research training program (CRTP) cancer workshop at Moncton, Newbrunswick, Canada.

ACKNOWLEDGMENTS

First of all, I would like to thank my thesis supervisor Dr. Sevtap Savas without whom this thesis would not be possible. She has been supportive of all my endeavours and she rendered support to me throughout my progress as a graduate student at Memorial University. I would also like to thank my thesis committee members, Dr. Roger Green and Dr. Patrick Parfrey for their critical and productive comments.

I would also like to express my gratitude to all the members of the Board of Regents, Memorial University of Newfoundland. Especially, I would like to thank Dr. Gary Kacahnoski, President and Vice-Chancellor, Memorial University of Newfoundland and Dr.Susan Dyer Knight, Chancellor, Memorial University of Newfoundland for their non-academic support.

I also wanted to thank Mr. Amit Negandhi, who is my fellow lab mate, well-wisher and my best friend in St. John's. He has been supportive of me both academically and non-academically. He helped me in going through all the procedures that are required by various organisations within and outside the Memorial University. He also helped me learn laboratory techniques such as gel electrophoresis.

I sincerely thank Ms. Jessica Squires, who performed the genotyping reactions for some of the polymorphisms in phase I of the hypoxia pathway project; for being kind and affable to me and for teaching me how to perform SNP genotyping reactions using the TaqMan® SNP genotyping technique.

I also appreciate the NFCCR for providing the DNA samples, genetic, clinicopathological and prognostic data for my Master's thesis projects. I am also thankful to all personnel who are directly or indirectly involved with the NFCCR and all the patients and their family members who were a part of the NFCCR study. I also express my sincere gratitude to the funding agencies that provided funds to (the thesis supervisor and the NFCCR) and made these studies possible (mainly Memorial University, Faculty of Medicine, Medical Research Foundation-Cox Award-2010, Canadian Institutes of Health Research (CIHR), Research and Development Corporation of Newfoundland (RDC) and Atlantic Canada Opportunities Agency-Atlantic Innovation Fund).

I would like to thank Danielle Collins, Administrative and Systems co-ordinator, YMCA of Northeast Avalon, NL and Lana Rostenberg, YMCA of Northeast Avalon, NL for proof reading my thesis. I would also like to thank the administrative assistant of my academic unit, Mrs. Deborah Quinlan and others in Research and Graduate Studies for timely processing of my queries.

I am also thankful to my fellow graduate students in Memorial University, namely Erica Clarke, Amanda Dohey, Sridhar Srinivas, Daniel Evans, Cindy K. Penney, Ashar Seyed Pizarda and Krista Mahoney for being helpful. I would also like to thank the staff in the Division of Bio-Medical Sciences namely Mrs. Madonna, Mrs. Deborah Parrott and Mrs. Janes for their moral support.

Finally, it would be ungrateful if I failed to thank my family members, all my friends in India who had always been with me and shared some good memories and many Newfoundlanders who have helped me on several occasions non-academically.

CHAPTER 1: INTRODUCTION

1.1) Cancer

Cancer is characterized by uncontrolled proliferation of abnormal cells. It is a multifactorial disease caused by an individual's genetic factors in combination with several environmental factors. According to the World Health Organization (WHO), in 2008, cancer accounted for 7.06 million deaths and was the leading cause of death worldwide (2). WHO estimates that the number of cancer related deaths will continue to rise and 13.1 million people will die because of cancer in 2030 (2). According to the Canadian Cancer Statistics 2012, it is estimated that twenty one Canadians will be diagnosed with some type of cancer and nine Canadians will die due to cancer every hour (3). Despite several significant advances in science and technology, survival rates for the majority of cancer types have not improved substantially. Therefore, cancer remains as a major public health concern in Canada and worldwide.

The transformation of a normal cell to malignant cancer is a multi-step process that involves accumulation of mutations to form pre-cancerous cells which are then subjected to additional mutations to transform into malignant cancer cells (4). Genes that control and suppress cancer formation and proliferation are termed tumour suppressors, while other genes that contribute to cancer formation are referred to as proto-oncogenes (4). In brief, in cancer cells, a group of tumour suppressor genes are deactivated and a group of oncogenes are activated (4).

Cancer cells have distinct characteristics when compared to normal cells. For example, cancer cells are characterized by abnormal metabolism and energy production to meet the energy needs of rapidly dividing cancer cells. In order to be metabolically active, normal cells derive energy produced through the Oxidative Phosphorylation Pathway (OXPHOS) (5). However, cancer cells undergo a metabolic shift to produce energy through aerobic glycolysis instead of OXPHOS (5,6). In addition, certain pathways that control cell proliferation such as programmed cell death (apoptosis) are deactivated, while signalling pathways that facilitate cell proliferation are activated. Hence, cancer cells have certain distinct characteristics when compared to normal cells.

1.2) Colorectal cancer

Colorectal cancer is a form of cancer occurring in either the colon or the rectum. The incidence and mortality rates of colorectal cancer are increasing worldwide due to lifestyle changes and dietary factors (7). According to the International Agency for Research on Cancer, colorectal cancer is the third most frequent cancer and the fourth leading cause of death worldwide (8). In particular, the western world is at an increased risk of colorectal cancer which is attributed to diet and sedentary lifestyle (7). According to the Canadian Cancer Statistics 2012, colorectal cancer is the third most common cancer type and is the second leading cause of cancer related deaths in Canada (3). It has been projected that one in every 13 Canadians will develop colorectal cancer and one in every 28 Canadians will die from it (9).

According to the American Joint Committee on Cancer (AJCC), the survival rates of colorectal cancer range from 10-90%, depending on the stage at which patients are diagnosed (10). For patients diagnosed with colorectal cancer at stage I, the 5 year survival rate is 93.2%, whereas it is only 8.1% for the stage IV patients. In Canada, overall, the five year relative survival rate for colorectal cancer is 63% (9). Thus, colorectal cancer has a significant impact on the mortality rates of Canadians.

In Newfoundland and Labrador (NL) the incidence and mortality rates of colorectal cancer are higher when compared to western provinces in Canada. Colorectal cancer rates in NL are two times higher when compared to those in British Columbia (9). Therefore, effective management of this disease requires serious attention especially in NL.

Classification of colorectal cancer

Colorectal cancer can be classified into inherited and sporadic forms based on familial clustering of this disease.

Previous findings suggest that genetic factors contribute to 35% of the risk of colorectal cancer formation and only 6% of the patients characterised with colorectal cancer can be attributed to inherited mutations in specific genes (11). This shows that only a small portion of patients diagnosed with colorectal cancer have inherited components identified so far (12). Thus, for the majority of the colorectal cancer cases, the genetic basis of this disease is unknown.

Inherited colorectal cancer

In inherited colorectal cancer cases, the disease is inherited through mutations in genes and can be sub-classified as 1) Lynch syndrome, 2) Familial colorectal cancer type X (FCCTX), 3) Familial adenomatous polyposis (FAP), 4) MUTYH associated polyposis (MAP) and 5) Attenuated familial adenomatous polyposis (AFAP). Lynch syndrome is an autosomal dominant disease caused by mutations in DNA repair genes *MLH1*, *MSH2*, *MSH6*, and *PMS2*. In FCCTX, there is no recognizable pattern of Mendelian inheritance – although it is believed that FCCTX predominantly follows an autosomal dominant mode of inheritance. Moreover, the disease pathogenesis seems to be distinct from other types of cancer (13,14). Nevertheless, it is believed that there are specific genes and mechanisms which are responsible for this disease which are yet to be identified (15). Hereditary non-polyposis colorectal cancer (HNPCC) comprises both Lynch syndrome and FCCTX (16).

FAP is an autosomal dominant inherited disease (17) characterised by inherited mutations in the adenomatous polyposis coli (*APC*) gene (18). This gene is involved in normal cellular functions such as cellular adhesion, communication and microtubule assembly (17). Mutant *APC* is linked to high expression of cyclooxygenase (*COX*) gene which in turn increases the adhesion of extracellular matrix and inhibits apoptosis (19,20). MUTYH associated polyposis (MAP) is caused by mutations in the *MUTYH* gene and is inherited as an autosomal recessive condition (21). AFAP is also caused by mutations in the *APC* gene and is similar to FAP; however patients with AFAP have very few polyps when compared to patients with FAP (22).

Sporadic colorectal cancer

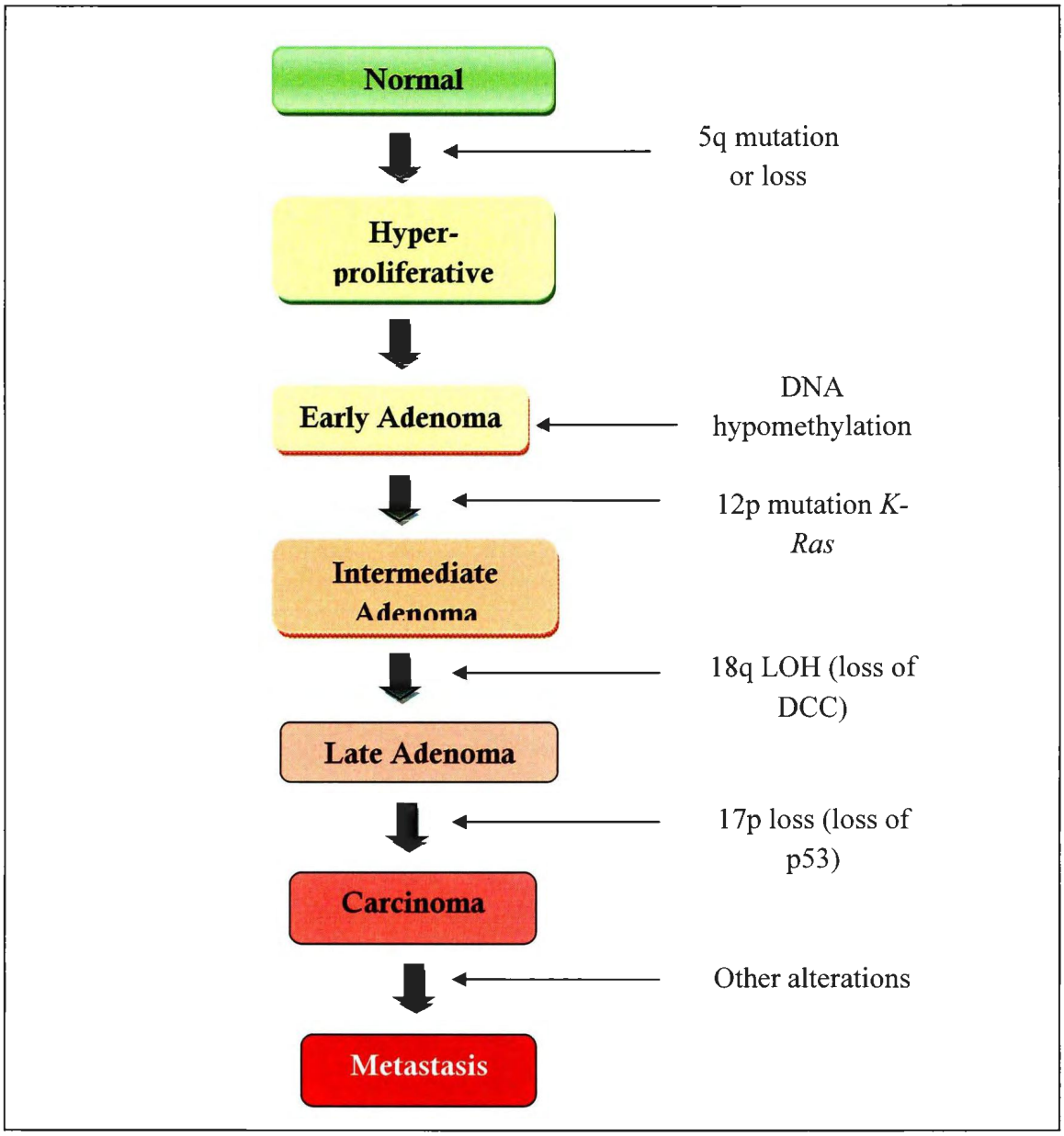
Sporadic cancers account for the majority of the colorectal cancer cases (12). Sporadic colorectal cancers do not show specific inheritance patterns, but they may be associated with a genetic component (17). It is assumed that sporadic colorectal cancer is caused by low susceptibility alleles with small effect sizes whose effect is modified by interaction with environmental factors or other genetic and epigenetic factors (11,23). So far, several low susceptibility alleles have been identified in colorectal cancer through genome wide association studies (24-26). Sporadic colorectal cancer patients are also characterised by one or a combination of somatic mutations in *APC*, *DCC*, *K-RAS*, *N-RAS* and *TP53* genes as well as the deletion of chromosome 18q (17).

1.2.1) Colorectal cancer tumourigenesis and progression pathways

Vogelstein initially proposed that colorectal cancer tumourigenesis is a multistep process that involves sequential mutations for progression from adenoma to malignant carcinoma (27). The steps involved in this process are depicted in **Figure 1**.

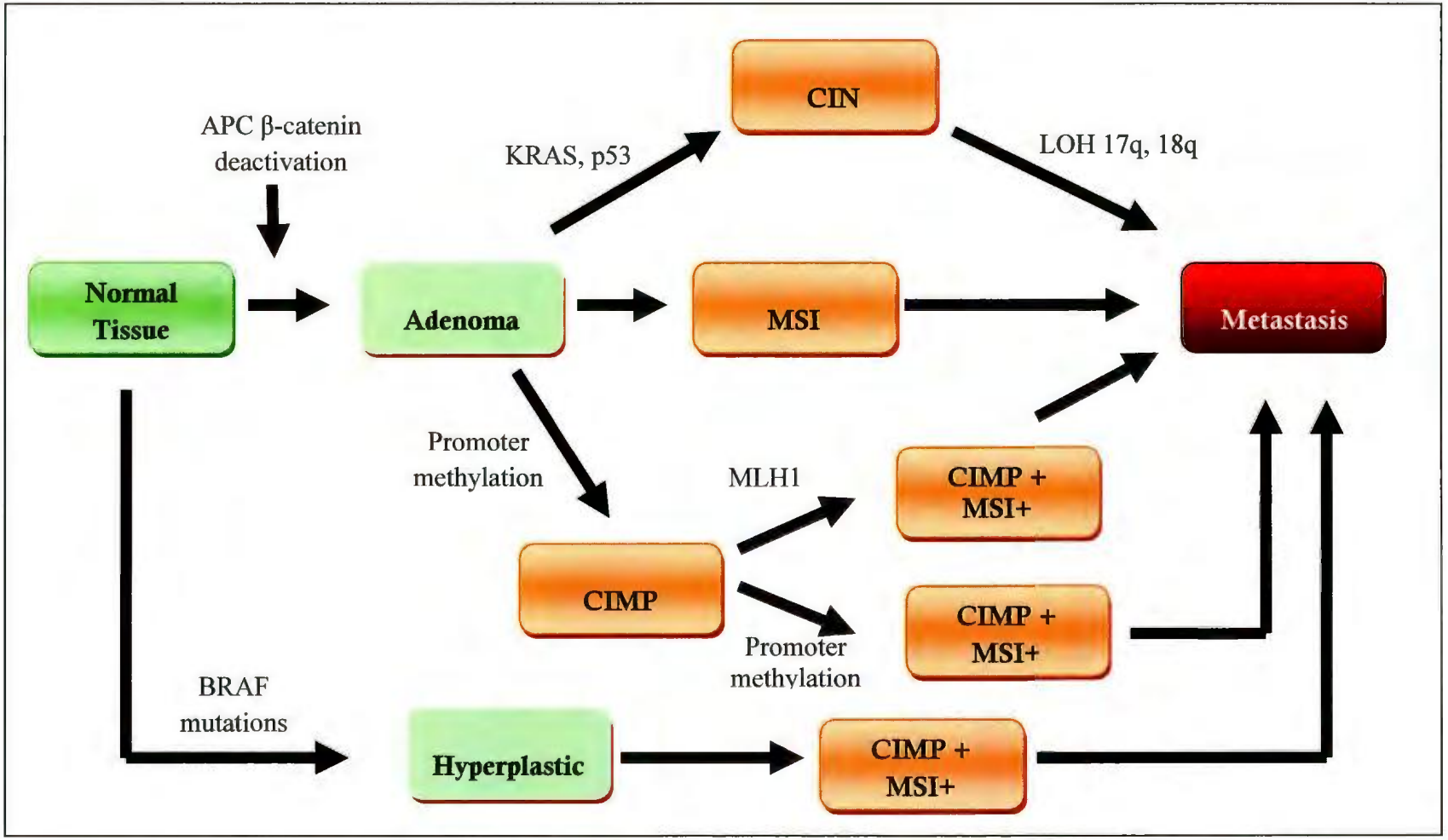
There are biological pathways that have been elucidated through molecular research, which have added more evidence to the adenoma-carcinoma model by Vogelstein as illustrated in **Figure 2**.

Figure 1: Vogelstein's genetic model for tumourigenesis



This figure shows the steps involved in the process of carcinogenesis starting from the early stages to the later stages as proposed by Vogelstein. (Adapted from the genetic model proposed by Vogelstein) (27). *5q mutation or allelic loss refers to mutations or allelic loss in APC gene. 12p mutations refers to the mutations that occur in the RAS gene (predominantly K-RAS gene), 18q LOH (loss of DCC) refers to the loss of heterozygosity on chromosome 18q which is predominantly characterized by allelic loss of the DCC locus. Order of these alterations might vary; however, what is more important is that the accumulation of these mutations is required for cancer initiation. Combined with other alterations they also facilitate cancer progression.*

Figure 2: Colorectal cancer tumourigenesis



A recent model of carcinogenesis with additional pathways included into the Vogelstein's model (adapted from (29)*). APC: Adenomatous polyposis coli; CIN: Chromosomal instability; MSI: Microsatellite instability (MSI+); CIMP+: CpG island methylator phenotype; LOH: Loss of heterozygosity. These pathways are explained in detail in* ***Section 1.2.1****.*

Chromosomal Instability (CIN) pathway

The gain or loss of chromosomal fragments or chromosomal rearrangements is referred to as chromosomal instability (CIN). Nearly 75-85% colorectal tumours are characterised by CIN (28). Usually, the events that lead to CIN are aneuploidy,

chromosomal translocations, gene amplification, allelic loss and mutations in specific genes such as *K-RAS, TP53, APC* and *SMAD4* (28). In addition, CIN tumours are characterized by the loss of either the paternal or the maternal allele which is referred to as Loss of Heterozygosity (LOH) (30). Furthermore, mutations of genes involved in mitotic spindle checkpoint, such as *MAD2, BUB1* and *BUBR1,* and the genes involved in the DNA damage control check point, such as *ATM, ATR, BRCA1, BRCA2,* and *TP53,* can also lead to CIN (30). Telomere length and telomerase enzyme also can play a crucial role in preventing chromosomal instability (30).

Microsatellite Instability (MSI) pathway

Microsatellite instability (MSI) is caused by length changes in repetitive microsatellite sequences. Up to 15-20% of the colorectal cancer tumours are characterised by MSI-High (MSI-H) phenotype (31). It occurs by inactivation of the mismatch repair system (MMR). *MLH1, MSH2, MSH3, MSH6, PMS1, PMS2* and *MLH3* genes encode proteins that are essential for an effective MMR system (30). Inactivation of MMR occurs by either mutations or by epigenetic alterations in either one, or more than one, MMR gene (31).

The MSI phenotype is categorized into three groups based on the instability in markers recommended by the National Institute of Health (NIH) such as BAT-25, BAT-26, NR-21, NR-24 and MONO-27 (32). Tumours with instability in one or two markers are categorized as microsatellite instability-low (MSI-L), and tumours with no instability in these markers are categorized as microsatellite stable (MSS) (32). If tumours are characterised with instability in more than two markers, then they are categorized as MSI-H. Lynch syndrome patients are characterised by MSI-H tumours and this occurs by inherited mutations in the mismatch repair genes (mainly *MSH2* and *MLH1*) (33). In the case of sporadic colorectal cancer, up to 15% of the tumours are categorized as MSI-H and this occurs mainly by epigenetic modification of the *MLH1* gene and may also be characterized by somatic mutation in *MLH1* gene (33).

CpG island methylator phenotype pathway

Defects in methylation patterns are also observed in colorectal cancer. For example, defects in the DNA methylation lead to genomic instability in colorectal cancer cell lines (34-36).

The 5' region of genes contains the promoter and the transcription initiation site (37). Regions rich in CpG sequences are referred to as CpG island and nearly 50% of human genes contain CpG island sequences in their 5' region (37). In colorectal cancer, aberrant methylation of CpG dinucleotides takes place in tumour suppressor genes which results in the CpG island methylator phenotype (CIMP). CIMP serves as an alternative pathway for tumour progression (37). Defects in the CIMP are the predominant epigenetic alterations found in colorectal cancer tumours (38). CIMP is classified into

CIMP-1 (CIMP high), CIMP-2 (CIMP low) and CIMP (-) (CIMP negative) based on the degree of methylation (37). CIMP is associated with MSI-H and *BRAF* mutations in sporadic cases (39). For example, nearly 80% of the colorectal cancer tumours with CIMP-1 have MSI-H phenotype and nearly 50% of tumours carry mutations in the *BRAF* gene (37). The same study also reported that while fewer patients with CIMP-1 tumours have mutations in *KRAS* (16%) and *TP53* (11%) genes, the majority of the colorectal cancer patients with CIMP-2 tumours are characterised by *KRAS* (93%) and *TP53* mutations (31%). But, CIMP-2 colorectal tumours are characterised by very low frequency of *BRAF* mutations (4%) and MSI-H (0%) phenotype. This study also showed that in CIMP(-) colorectal cancer cases, the frequency of mutations is high in *TP53* (71%) and the tumours are characterised by low frequency of MSI-H (12%) and very low frequency of *BRAF* mutations (2%).

1.2.2) Colorectal cancer treatment options and strategies

In colorectal cancer, treatment usually involves chemotherapy or radiotherapy (40). Radiotherapy normally involves exposure to X-rays, gamma rays and may be used in combination with chemotherapy (40). Chemotherapy drugs aim to stop the uncontrolled proliferation of cancer cells. Some of them are DNA damaging agents such as alkylating agents and intercalating platinum complexes (41-44). There are also a group of drugs that inhibit mitosis and tumour progression (such as Paclitaxel) (45). In addition, some drugs inhibit DNA synthesis during DNA replication, thereby preventing tumour progression (such as Methotrexate (administered along with Leucovorin), 5-Fluorouracil

(5-FU), Irinotecan and Mercaptopurine) (46-51). Biological agents such as monoclonal antibodies which act on signalling proteins to stop tumour progression are also used in treatment of colorectal cancer (52).

1.3) Prognosis

Prognostic estimates predict the risk of developing a particular outcome based on patients' clinical and non-clinical characteristics (53). In cancer, these outcomes may include: death, cancer recurrence, cancer progression, metastasis, changes in quality of life or other clinically important outcomes (54-56).

Prognostic markers

Clinical variables such as stage, histology, location and tumour grade are used as prognostic markers (38). Prognostic markers can also be biological markers such as protein, ribonucleic acid or DNA markers (57). For example, genetic markers such as single nucleotide polymorphisms (SNPs) (58), copy number variations (CNVs), inversions or variability in the trinucleotide CAG repeats are candidate markers. In addition, the degree of DNA methylation can also be used to predict prognosis (i.e. epigenetic markers) (59,60), although more research is required to use these markers in the clinic. These and other markers - also referred to as predictors – can be used to predict the outcome in patients diagnosed with a particular disease condition.

Prognosis of patients may be influenced by a variety of factors such as age, sex, familial history, quality of life, symptoms and other clinically relevant variables (61). For

example, in colorectal cancer, several reports suggest that older patients have a worse prognosis when compared to young patients; female patients have a better prognosis when compared to male patients; patients with tumours characterized by mucinous tumour phenotype have a worse prognosis when compared to patients with tumours characterized by non-mucinous tumour phenotype (62); patients diagnosed at stage IV have a worse prognosis when compared to patients diagnosed at earlier stages; patients with tumours that show vascular invasion have a worse prognosis when compared to patients with tumours without vascular invasion; and patients with MSI-H tumours have a better prognosis when compared to patients with microsatellite stable tumours (63). These and other prognostic variables in colorectal cancer are discussed in detail in **Section 1.4**.

Prognostic studies and models

According to Hemingway et al (64), the two main aims for prognostic studies are: "*1) to identify individual biomarkers that are associated with the outcome independent of other biomarkers,* and *2) development of multivariable prognostic models that can predict an individual outcome using prognostic index or risk score*". Prognostic information is commonly used to stratify patients into two or more groups depending upon the prognostic criteria.

Since the characteristics of patients change from population to population and within a population (65), it is almost impossible to predict the patient outcome with a single predictor (61). In a clinical setting, multiple indicators are used to predict outcome. Determining the probability of a dependent event (such as death), using a combination of predictors (independent variables) such as disease stage, sex and genotypes, requires

specialised multivariable prediction tools (61). These prediction tools referred to as prognostic or prediction models (61) are based on statistical methods that simultaneously analyze multiple independent predictors (66).

Before designing a prognostic study, the endpoint or clinical outcome must be defined explicitly. It should be measurable, reproducible and there should not be any bias involved in the ascertainment of the endpoint (54). Death and recurrence are the most frequently investigated endpoints in colorectal cancer.

In prognostic research, the association of study variables with patient outcome may be tested individually in a univariate analysis (61,65,67). However, univariate analysis does not consider the effect of confounding by other variables. Hence, multivariate analysis is used to eliminate the confounding effects of other variables. The number of predictor variables should be kept as small as possible since complex models are difficult to integrate into a clinical setting (68). The number of predictor variables can also be minimized by using selection techniques such as the forward selection, backward selection and the best subset method (67).

Sample size is related to study power, i.e detecting an association. In prognostic research, more than the sample size, it is the number of events per variable that influences power and variance (69). In an ideal prognostic study, there should be at least ten patients who have experienced the event of interest per predictor variable (61). The variables that are included in a prognostic study may have different effect sizes, i.e., each variable may contribute to the outcome with different magnitudes of effect (70).

Patients in a prognostic study are followed from a common start point, such as the date of diagnosis, until they experience the outcome event of interest (endpoint) (71). The phase between the common start point and the endpoint is referred to as the "follow-up phase". During the follow-up, events of interest experienced by the patients are recorded. The observed time between the start and endpoint is referred to as the survival time (71). Some patients may not experience the outcome of interest at the end of the follow-up period and they are referred to as "censored individuals" during statistical analyses (67). The association between the predictor and the outcome is then analysed by using a statistical method suited to the study objective.

Depending on the ease of collecting data, a prognostic study will be carried out either prospectively or retrospectively. Both of these methods have advantages and disadvantages. Retrospective studies use pre-recorded clinical data, such as hospital records, and archived DNA or tissue specimens (72). One of the advantages of a retrospective study is that since existing records or archived biospecimens are used, retrospective studies take less time and cost less. In prospective studies, patients are recruited based on pre-determined eligibility requirements. Their baseline characteristics are recorded, and they are followed until the end of the study (73). The main advantage is that patients can be followed for any predetermined time interval. Unlike retrospective studies, in prospective studies, the follow-up time can be varied to increase the number of patients with outcomes. The main disadvantage of prospective studies is the cost and time required to recruit study subjects.

Statistical data analysis and validation of prognostic models

After obtaining the required data, different statistical techniques are applied for data analysis depending upon the study objectives. The Kaplan Meier method (univariate) compares the survival characteristics with respect to time between two or more groups of patients who are categorized based on specific criteria (71,74). The Cox Proportional Hazard model computes the hazard ratio of experiencing the outcome of interest in groups of patients categorized based on specific variable characteristics, together with the 95% confidence intervals (CIs) (75). Similar to Kaplan Meier method, in univariate Cox regression analysis, the relation between each variable and the outcome is assessed individually. However, in Cox multivariate regression method, several variables are assessed for their relation to the outcome simultaneously. Since multivariate models can detect and adjust for confounding by other variables, the majority, if not all, of the prognostic models are used in the clinic are based on the multivariate analysis.

A multivariate prognostic model may not have a clinical significance unless its credibility and reproducibility have been successfully validated by additional analyses. To do so, prognostic models may be first internally validated in an existing data set by split-group, jackknife (cross validation) or bootstrap methods (66). Additionally, a prognostic model should be validated in a patient cohort from a different geographical area or preferably in a cohort which has characteristics entirely different from the patient cohort that was used to develop the model (external validation) (68,76). By doing so, how well the model fits in a different setting may be determined, so that expert panels may

investigate these findings and consider recommending it for implementation in clinical settings.

1.4) Prognostic indicators in colorectal cancer

1.4.1) Pathological prognostic markers in colorectal cancer categorized by the American Pathologists Consensus Statement

There are many markers that have been reported to be of prognostic significance in colorectal cancer. However, conflicting results obtained in different studies call into question the reliability of a particular marker in predicting outcome in cancer patients. For assessing the usefulness of prognostic markers in colorectal cancer, the College of American Pathologists (CAP) reviewed medical literature and grouped markers according to their prognostic significance (77).

Category I markers include those that have been validated using multiple robust studies; category IIA includes markers that have been studied extensively, are of sufficient importance, but have not yet been validated; category IIB includes other promising markers. Category III includes markers that have not been studied extensively to elucidate their prognostic significance; and category IV includes variables that have been studied extensively but lack prognostic significance. Select markers that fall under these four categories in colorectal cancer are discussed below.

Tumour stage

The tumour stage remains the most significant prognostic indicator for prediction of prognosis in colorectal cancer patients (38). Although colorectal cancer patients with same cancer stage may have variable survival times, the tumour stage is an important indicator when predicting whether patients will need therapy (38). The tumour stages were classified according to the following criteria by American Joint Committee on Cancer Sixth Edition Staging as follow: *"T1 = tumor invades submucosa; T2 = tumor invades muscularis propria; T3 = tumor invades through the muscularis propria into the subserosa or into nonperitonealized pericolic tissues; T4 = tumor directly invades other organs or structures and/or perforates visceral peritoneum; N0 = no regional lymph node metastasis; N1 = metastasis to one to three regional lymph nodes; N2 = metastasis to four or more regional lymph nodes; M0 = no distant metastasis; M1 = distant metastasis"*(78). According to recent nomenclature (10), colorectal cancer stages are classified as stage I, stage II (subgroups A and B), and stage III (subgroups A, B, C, D, and E) and stage IV (subgroup A and B) as shown in **Table 1**. In this table, the 5 year survival rates of different stages and their subgroups are also shown (10).

It is evident from **Table 1** that stage I has a favorable prognosis and stage IV has the poorest prognosis in colorectal cancer. In short, disease stage remains the most useful indicator in predicting the outcome (79) and is a category I marker (77).

Table 1: Stage classification of colorectal cancer patients and their corresponding 5 year survival rates.

Stage	Subgroup	5 year survival rates (%)
Stage I	-	93.2
Stage II	A	84.7
	B	72.2
Stage III	A	83.4
	B	62.4
	C	52.3
	D	43
	E	26.8
Stage IV	-	8.1

Adapted from the American Joint Committee on Cancer Sixth Edition Staging (10)

Tumour differentiation

The degree of cell differentiation in tumours is referred to as tumour grade. Although the American Pathologists' Consensus Statement recommended that grade be classified as either high or low grade it is generally classified as grade I, grade II and grade III (38). In grade I, cells are well differentiated and uniform without any nuclear stratification or polarity (38). Grade II cells are moderately differentiated, with disorganized or lost nuclear polarity and with irregular glandular formations (38). In

grade III, cells are poorly differentiated, accompanied with loss of tubular differentiation and polarity (38). There are several studies examining the prognostic role of tumour grade in colorectal cancer. For example, it has been demonstrated that tumour differentiation is an independent prognostic factor in colorectal cancer in a multivariate analysis (38). In this study, grade III (poorly differentiated) tumours were associated with increased invasion, nodal metastasis and recurrence. Hence, grade III colorectal cancer patients were reported to have poor prognosis (38). In short, tumour grade falls under the category IIA of the American Pathologists' Consensus Statement (77), and hence is a prognostic marker that is of sufficient importance. However, its clinical value remains to be validated before it can be used in the clinic for predicting the prognosis of colorectal cancer patients.

Tumour histology

Depending on their histological characteristics, colorectal tumours can be mainly classified as mucinous and non-mucinous tumours. Mucinous tumours are characterised by abundant extracellular mucin and as tumours with more than 50% extracellular mucin (80). Mucinous tumour phenotype is stage dependent and is associated with recurrence, metastasis and advanced stage in colorectal cancer (38). Mucinous histology is found in colorectal tumours with a high degree of DNA methylation and is associated with MSI-H in both sporadic colorectal and Lynch syndrome cases (38). Tumour histology falls under category III of the American Pathologists' Consensus Statement (77), which indicates that it has to be studied extensively to elucidate its prognostic significance.

Lymphatic and venous invasion

Lymphatic and venous invasions are usually considered important stage independent prognostic markers (81-83). Both lymphatic and venous invasions contribute to metastasis (79,84,85). Specifically, minimal lymphatic penetration is enough to implant distant metastases (86) and invasion of thick walled blood vessels by tumours is prognostically more significant than invasion of thin walled blood vessels (81,82). Lymphatic and venous invasions come under category III of the American Pathologists' Consensus Statement, which indicates that their predictive potential has not been studied extensively (77).

Angiogenesis

Angiogenesis is the process of new blood vessels development and is a naturally occurring process for tissue repair and growth (87). However, in cancer, this process facilitates disease progression. Growth of cancer cells depends on their ability to have blood vessels around them (87). Pro-angiogenic factors favour angiogenesis and this increases the capability of cancer cells to metastasize to distant organs (87). Therefore, angiogenesis can influence patient outcome and thus it is classified as a category III marker according to the American Pathologists' Consensus Statement, which indicates that angiogenesis-markers have to be studied extensively (77).

1.4.2) Genetic prognostic markers in colorectal cancer categorized by the American Pathologists Consensus Statement

Microsatellite instability high (MSI-H)

MSI-H is caused by inactivation of mismatch repair genes in the case of Lynch syndrome (88) or by hypermethylation of *MLH1* gene promoter in the case of some sporadic colorectal cancer (79). Around 15-20% of colorectal cancer patients have MSI-H tumours which are characterised by lymphatic invasion, poor differentiation, mucinous histological type and proximal location (79).

There are conflicting reports over the prognostic significance of the MSI phenotype. Previously, it has been reported that patients with MSI-H tumours have a better prognosis when compared to those with MSS tumours (31,33). It has also been reported that colorectal cancer patients with MSI-H tumours show resistance to 5-FU, the most commonly used chemotherapeutic agent in the treatment of colorectal cancer. Therefore, MSI-H was suggested as a potential marker to differentiate patients who may not benefit from 5-FU treatment (89). However, there are also conflicting reports showing that the MSI-H phenotype can neither predict outcome nor response to 5-FU (90-92). Nevertheless, MSI-H is categorized as a category IIA marker by the American Pathologists' Consensus Statement (77), which indicates that it is a promising molecular marker in colorectal cancer.

Loss of heterozygosity of 18q in tumour cells

Seventy percent of colorectal tumours have allelic imbalance at chromosome 18q and the biological effect of this chromosomal loss is directly implicated in colorectal carcinogenesis (89). LOH of 18q causes haploinsuffciency, which results in relatively low protein expression. *DCC* is an example of a gene affected by LOH at 18q. 18q LOH is associated with aggressive tumours in stage II colorectal cancer patients (39) and is associated with worse prognosis (89), including poor response to 5–FU based adjuvant chemotherapy (38) in colorectal cancer patients. A meta-analysis of 17 studies involving 2189 colorectal cancer patients concluded that LOH of 18q can be used to stratify patients for adjuvant chemotherapy and can also predict outcome (93). The American Pathologists' Consensus Statement (77) categorizes LOH 18q as a category IIB marker, which indicates that it is a candidate marker, but at the time being lacks sufficient data to confirm its prognostic significance.

KRAS mutations

KRAS (Kirsten rat sarcoma-2) belongs to the RAS gene family. It encodes p21 protein, a 21 kDa membrane protein, with GTPase activity, which controls cell proliferation and differentiation by acting as a molecular switch in response to extracellular mitogenic signals (89). Over-expression of *KRAS* results in increased production of proteases which increase degradation of the extracellular matrix and thus increase the ability of cells to metastasize (89). *KRAS* mutations are found in 15-20% of all human cancers and in 30-40% of colorectal cancer tumours (94). *KRAS* plays an important role in the EGFR pathway and mutant KRAS protein confers resistance to anti-

EGFR agents, such as Cetuximab or Panitumumab (94). Clinical prognostic markers such as tumour budding and podia formation are associated with *KRAS* mutations (95) and colorectal cancer patients with *KRAS* mutated tumours have poor prognosis (89). The presence of *KRAS* mutations is also a strong indicator for adjuvant chemotherapy (89). According to the American Pathologists Consensus statement (77), *KRAS* mutations are category III markers which need to be studied further to elucidate their potential prognostic significance.

Epithelial growth factor receptor (EGFR)

The *EGFR* pathway plays an important role in cell proliferation, migration and differentiation (89). *EGFR* is differentially expressed in normal, pre-cancerous and advanced cancer cells (94). More importantly, the expression of *EGFR* seems to be critical in metastatic colorectal cancer (89,96). *EGFR* expression is related to advanced disease stage, poor histological grade and lymphatic invasion (97-99). However, there are conflicting reports showing that EGFR status is not associated with histology, grade, stage or patient survival (100-102). In another study, inhibition of *EGFR* expression showed an anti-tumour activity and hence *EGFR* inhibition may be a treatment option for colorectal cancer patients (79). The American Pathologists Consensus statement categorizes *EGFR* under category III, which indicates there is insufficient data to conclude its prognostic significance.

1.4.3) Other potential prognostic markers in colorectal cancer

BRAF mutations

BRAF belongs to the RAF gene family which codes for kinases that mediate the cellular response to growth factor signals (89). *BRAF* mutations are found in 5-15% of colorectal cancer tumours (89). Colorectal cancer patients having tumours with MSS and *BRAF* mutations have worse prognosis (38). Patients with tumours carrying the *BRAF* activating Val600Glu somatic mutation show resistance to anti-cancer drugs and have `poor prognosis (103). *BRAF* mutations are also used to predict resistance to anti-*EGFR* targeted therapies in colorectal cancer patients (39). In addition, it has been reported that *BRAF* mutations are present at very low frequency in the tumours of stage II and stage III patients, but those patients who have tumours with *BRAF* mutation have a poor overall survival rate when compared to patients with the wild type *BRAF* (104). Clearly additional studies are needed to identify whether *BRAF* somatic mutations have prognostic importance in colorectal cancer.

Single nucleotide polymorphisms (SNPs) as genetic prognostic markers

Recently, there has been an interest to test SNPs as prognostic markers in cancer in addition to point mutations, deletions, copy number variations and gene expression profiles. Many studies have investigated SNPs from various pathways such as angiogenesis (58), hypoxia (105,106), DNA repair (107) and micro RNA genes (108). However, these kinds of studies are still in the research phase and the majority of these SNPs have to be validated for their prognostic significance before they can be used as

prognostic predictors in the clinic. The research described in this Master's thesis focuses on SNPs and their potential prognostic significance in colorectal cancer.

1.5) General rationale and objectives

The main objective of this project is to identify genetic markers that can predict outcome in colorectal cancer patients. This thesis focuses on genetic markers from mitochondrial DNA (Chapter 2) and hypoxia pathway genes (Chapter 3). Genetic markers in mitochondrial DNA have been implicated in cancer progression. However, the prognostic role of these and other variations have not been well characterized. Similarly, hypoxic conditions have been shown to favor aggressive tumor phenotype, and to facilitate resistance to anti-cancer therapies. Currently, there are no comprehensive studies which look at the association of the vast majority of the SNPs in the hypoxia pathway genes with the patient prognosis (Chapter 3). Therefore, in this thesis, we hypothesize that, genetic variations in the mitochondrial DNA and hypoxia pathway genes may be prognostic markers in colorectal cancer.

The specific objectives of this thesis are to:

i. Test the association of select genetic markers in the mitochondrial DNA with prognosis in a colorectal cancer patient cohort (Chapter 2).

ii. Test the association of select genetic markers in the hypoxia pathway genes with prognosis in two colorectal cancer patient cohorts (Chapter 3).

CHAPTER 2: MITOCHONDRIAL DNA (mtDNA) VARIATIONS AND THEIR RELATION TO PROGNOSIS IN COLORECTAL CANCER

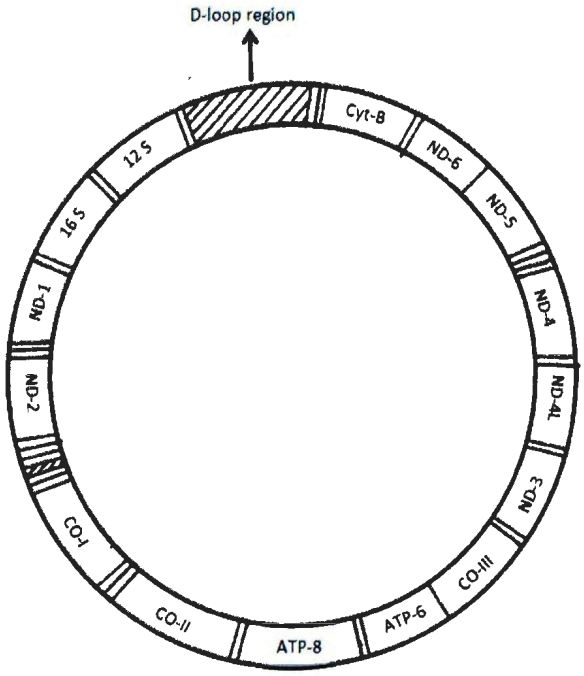
2.1) Introduction to mtDNA and cancer

Organisation of the mtDNA genome

Mammals have two genomes - a nuclear genome and a mitochondrial genome (109). The mitochondrial DNA (mtDNA) is a double stranded, circular molecule of 16,568 bp (109) (**Figure 3**). It has 37 genes, of which 13 code for polypeptides which are involved in respiration and oxidative phosphorylation (OXPHOS), two code for rRNAs, and 22 genes code for tRNAs (109). The mitochondrial genome undergoes replication and transcription independent of the nuclear DNA (nDNA) (110). MtDNA lacks histones and introns, and has short reading frames when compared to nDNA (34). The D-loop region lacks coding sequences and it is where the replication and transcription of mtDNA start (4). A portion of the D-loop is triple stranded (7S DNA) because of repetitive synthesis (4). MtDNA is repaired as short patches through base pair excision repair (BER) and other than BER, the mtDNA does not have sophisticated DNA repair mechanisms (34). Hence, the mutation rate of mtDNA is ten times higher than that of nDNA (34).

MtDNA has a heavy (H) strand and a light (L) strand. The light and heavy strands have high cytosine and guanine contents, respectively (4). The replication of

Figure 3: Mitochondrial DNA organisation



Adapted from Park and Larsson (111). *ATP-6: mitochondrial encoded ATP synthase 6; ATP-8 mitochondrial encoded ATP synthase 8; CO-I: cytochrome c oxidase subunit 1; CO-II: cytochrome c oxidase subunit II; CO-III: cytochrome c oxidase subunit III; ND-1: NADH dehydrogenase 1; ND-2: NADH dehydrogenase 2; ND-3: NADH dehydrogenase 3; ND-4: NADH dehydrogenase 4; ND-4L: NADH dehydrogenase 4L; ND-5: NADH dehydrogenase 5; Cyt-B: cytochrome B gene; 12S and 16S: rRNAs; and 22 tRNAs shown as vertical rectangular blocks. Drawn not to scale.*

each strand takes place in opposite directions. Initially the H-strand is synthesized using the L-strand as template, and when two-thirds of H-strand is synthesized, the synthesis of the L-strand takes place in the opposite direction using the H-strand as a template (4).

Function of mitochondria

Also known as the power house of the cell, the mitochondria regulate programmed cell death, energy metabolism and cell proliferation (6,112). Mitochondria are involved in energy production through OXPHOS, a metabolic process which produces reactive oxygen species (ROS) (113). ROS are mutagenic agents and can induce mutations in DNA (6). Generation of ROS and absence of histones as well as sophisticated DNA repair mechanism make mtDNA susceptible to high mutation rates (5,113). A mammalian cell has approximately 1000 mitochondria with multiple copies of mtDNA in each. The total number of mitochondria and thus mtDNA molecules differs from one cell type to another depending on the energy requirements of the cell (5). Since apoptosis is energy dependent, an increase or decrease in mtDNA copy number may affect apoptosis and thus may have an effect on cancer progression. Furthermore, the amounts of transcriptional products of mitochondria are directly proportional to its mtDNA copy number. Hence, the mtDNA copy number might act as a surrogate marker of mitochondrial function as well as its dysfunction.

Role of mtDNA in carcinogenesis and cancer progression

There is mounting evidence that mitochondrial dysfunctions are associated with tumourigenesis and cancer progression.

There are also reports regarding the metabolic and molecular differences between normal and malignant cells (4). In 1930, Warburg (114) proposed that alterations in mitochondrial function might favour glycolysis. Production of adenosine tri-phosphate (ATP) molecules occurs in normal cells by oxidative phosphorylation pathway, whereas

in cancer cells, energy is produced mainly through glycolysis. Hence mitochondrial alterations that favor glycolysis may be implicated in cancer progression. In addition, alterations of activities of mitochondrial proteins such as ATPase and cytochrome C oxidase (5) are observed in cancer cells, indicating that alteration of mitochondrial function may be involved in tumourigenesis and cancer progression.

Apoptosis, the process of programmed cell death, is energy dependent. Therefore alterations in 13 mtDNA genes coding for proteins involved in ATP production may also affect apoptosis. In addition, mutations impaired mitochondrial BER, could result in impaired DNA repair system, that could lead to higher mutation rate and may facilitate tumour progression (34). Therefore, it is evident that there is a link between cancer and mtDNA abnormalities.

Alterations in mtDNA - either qualitative (e.g. point mutations, microsatellite instability in the hyper variable D-loop region) or quantitative (e.g. copy number variations) - are frequent in colorectal cancer tumours (112). For example, mtDNA somatic variations were detected in 70% of the colorectal cancer cell lines (seven of 10) and in 45% (20 of 45) of the colorectal tumours (115). There are also other homoplasmic (i.e. observed in all mtDNA samples of an individual or a cell) and heteroplasmic (i.e. observed in only a portion of the mtDNA molecules of an individual or a cell) mtDNA alterations that may contribute to progression in colorectal cancer (109). A comparative proteomic study also identified differences in the mitochondrial proteins produced in the colorectal tumour tissues when compared to adjacent normal tissues (116). These findings illustrate the role of mtDNA in cancer progression and suggest that it may also play a role

in influencing the prognosis. Therefore, mtDNA markers may also be prognostic markers in colorectal cancer (109).

Role of mtDNA in colorectal cancer prognosis

MtDNA point mutations and copy number changes are associated with tumour progression and drug resistance. Therefore, they may be candidate prognostic markers in predicting the survival of colorectal cancer patients (113). For instance, the presence of somatic mutations in the D-loop of mtDNA was previously associated with poor prognosis in colorectal cancer patients (112,117). In another study, stage II colorectal cancer patients with the mtDNA D-loop mutations had resistance to 5-FU based chemotherapy (5). Similarly, it has been reported in colorectal cancer patients that mitochondrial dysfunction results in lower level of β-F1-adenosine triphosphatase (ATP) and intracellular adenosine triphosphate, which triggers anti-apoptotic activity, resulting in resistance to 5-FU (118). In addition to this study, Lievre and colleagues (119) reported that the three year survival rate of colorectal cancer patients with D-loop mutations was lower when compared to those without D-loop mutations.

Recently, a study (120) examined 140 tagging mtDNA variants for their potential relationship with the colorectal cancer mortality in a Scottish cohort. This study found that two SNPs (G752A and G1440A) in the 12S ribosomal RNA gene and one SNP (G4770A) in the nicotinamide adenine dinucleotide dehydrogenase subunit 2 (*ND-2*) gene were significantly associated with colorectal cancer mortality (120). To our knowledge, this is the only study which comprehensively investigated the association of mtDNA polymorphisms and prognosis in colorectal cancer.

Copy number changes in mtDNA are also suggested to have an effect on the prognosis of patients. Since mutations in the D-loop region affect the copy number of mtDNA and mtDNA copy number is related to mitochondrial function, copy number variation of mtDNA can be a prognostic marker in cancer prognosis (6,113,121). Both an increase and decrease in the mtDNA copy number have been reported in colorectal tumours (122). In one study, it has been shown that an increase in copy number of mtDNA harboring the 4977 bp common deletion (in tumour tissues when compared to normal tissues) were associated with advanced stages and metastasis in colorectal cancer patients (112). Conflicting with this finding, it has also been reported that there was an association between the decreased mtDNA copy number (in colorectal cancer cells versus normal cells) with higher TNM stages and poor differentiation (118). This study also reported that an increase in the mtDNA copy number was associated with longer OS in colorectal cancer patients (118). Due to the conflicting nature of these results, it can be suggested that a prognostic role for changes in mtDNA copy number in colorectal cancer has not been established yet.

2.2) Rationale, hypothesis and objectives described in Chapter 2

The main function of mitochondria is energy production. Cancer cells are usually characterised by molecular aberrations which include abnormal energy production and metabolism. However, despite being implicated in cancer progression, the potential prognostic roles of mtDNA variations in colorectal cancer are understudied. In this study,

we hypothesize that alterations in mtDNA might facilitate cancer progression and may be potential prognostic markers in colorectal cancer.

The objective of the study described in this chapter is to:

i. test the association of six mtDNA polymorphisms and its copy number variation with outcome in colorectal cancer.

2.3) Contributions and credits

Asan M. S. Haja Mohideen: Obtained the genotypes of 2 mtDNA SNPs (mt16189 (T/C) and mt10398 (A/G)) using the TaqMan® SNP genotyping experiments, performed the qPCR analysis to determine the relative mtDNA/nDNA copy number ratio in tumour and non-tumour tissues, organised and coded the genotypes prior to statistical analysis, performed the statistical analyses and interpreted the results.

Dr. Sevtap Savas: Processed and coded the prognostic data for the NFCCR cohort and performed the Mann-Whitney U and Chi square statistical tests to compare the differences between the entire NFCCR cohort (n=736) and sub-group of this cohort used in the qPCR analysis (mtDNA copy number analysis cohort; n=279) as well as the genotype analysis (NFCCR SNP genotyping cohort; n=537), provided the baseline table for the entire NFCCR cohort.

Dr. Roger Green and Dr. Patrick Parfey: Provided the DNA samples, clinicopathological and prognostic information for the patients, as well as the genotype data for the four mtDNA SNPs obtained by a genomewide SNP genotyping method investigated in this study.

2.4) Materials and methods

2.4.1) Ethics approval

Ethics approval was obtained from the Human Investigation Committee (HIC) of Memorial University of Newfoundland (HIC #: 11.102).

2.4.2) Solutions

In this project, 5X and 1X Tris-Borate-EDTA (TBE; adjusted to pH 8.3) and 1X Tris-EDTA (TE) buffers were used. Further information on these buffers can be found in **Appendix 1**.

2.4.3) Patient cohorts

NFCCR cohort

This study investigates patients registered with the Newfoundland Colorectal Cancer Registry (NFCCR). Patients in NL diagnosed with colorectal carcinomas, who were under 75 years of age at the time of diagnosis, were recruited to the NFCCR over a 5 year period between January 1, 1999 and December 31, 2003 (123). The NFCCR obtained written consent from the patients to withdraw their blood samples, as well as to permit access to their tumour tissue samples and medical records (123). When patients were unable to provide their consent, the NFFCR obtained consent from their proxies (123).

The pathological and molecular data for patient tumours were also ascertained by the NFCCR as follows: the familial risk stratification was done according to Amsterdam and revised Bethesda criteria (123,124); MSI status was ascertained using immunohistochemistry analyses and by analysing five different microsatellite markers recommended by the National Cancer Institute (123); and *BRAF* somatic mutation status (Val600Glu) was ascertained by an allele specific PCR method (123). Stage and other pathological features were ascertained based on the tumour size, nodal status and metastasis data. The NFCCR also collected other patient clinicopathological and prognostic information such as date of death, recurrence of colorectal cancer, age, sex and other characteristics studied in this project. Patients were followed until 2010.

NFCCR patient sub-cohorts investigated in this study

DNA samples used in this study were obtained from the NFCCR. While the NFCCR cohort has 736 stage I-IV patients, for this project, we have used two sub-cohorts of NFCCR patients (**Table 2**). Out of 736 patients, 537 patients with available DNA (extracted from blood cells) were included into SNP analyses in both Chapter 2 and Chapter 3. This subcohort of 537 NFCCR patients is named the "*NFCCR SNP genotyping cohort*" throughout this thesis. Similarly, for the mtDNA copy number analysis, DNA samples extracted from tumour and the adjacent non-tumour tissue were available only for 276 patients out of the 537 NFCCR patients when this study was conducted. This subcohort is named as the "*mtDNA copy number analysis cohort*" throughout this thesis. The baseline characteristics of the entire NFCCR cohort, the NFCCR SNP genotyping cohort, and the mtDNA copy number analysis cohort are summarised in **Table 3.**

Table 2: Two NFCCR sub-cohorts used in this project

Sub-cohort	Source of DNA	Number of patients
NFCCR SNP genotyping cohort	Blood	537
mtDNA copy number analysis cohort	Tumour and adjacent non-tumour tissue	276

For SNP genotyping and mtDNA copy number analysis, each DNA plate containing a stock DNA concentration of 10 ng/µl were provided by NFCCR. The stock DNA solutions were diluted to 4 ng/µl for TaqMan® SNP genotyping reactions and 5 ng/µl for quantitative polymerase chain reaction (qPCR) with sterile dH_20.

Table 3: Baseline and prognostic characteristics of the entire NFCCR, mtDNA copy number analysis and the NFCCR SNP genotyping cohorts

Variables	Entire NFCCR cohort Total=736 n (%)	mtDNA copy number analysis cohort Total=276 n (%)	*p-values	NFCCR SNP genotyping cohort Total=537 n (%)	**p-values
Sex			0.387		0.964
female	286 (38.9)	116 (42.03)		208 (38.73)	
male	450 (61.1)	160 (57.97)		329 (61.27)	
Location			0.496		0.355
colon	506 (68.5)	183 (66.3)		356 (66.29)	
rectum	230 (31.5)	93 (33.7)		181 (33.71)	
Histology			0.915		0.537
non-mucinous	644 (87.5)	243 (88.04)		476 (88.64)	
mucinous	92 (12.5)	33 (11.96)		61 (11.36)	
Stage			0.167		**<0.001**
I	112 (15.2)	31 (11.23)		98 (18.25)	
II	244 (33.2)	94 (34.06)		207 (38.55)	
III	227 (30.8)	101 (36.6)		179 (33.33)	
IV	153 (20.8)	50 (18.11)		53 (9.87)	
Grade			0.489		0.089
well/moderately differentiated	651 (88.4)	243 (88.04)		494 (92)	
poorly/undifferentiated	73 (10)	32 (11.6)		39 (7.26)	
unknown	12 (1.6)	1(0.36)		4 (0.74)	

Vascular invasion			0.454		**0.013**
-	398 (54.1)	139 (50.36)		327 (60.9)	
+	282 (38.3)	111 (40.22)		171 (31.84)	
unknown	56 (7.6)	26 (9.42)		39 (7.26)	
Lymphatic invasion			0.455		**0.034**
-	389 (52.9)	137 (49.64)		316 (58.85)	
+	285 (38.7)	113 (40.94)		179 (33.33)	
unknown	62 (8.4)	26 (9.42)		42 (7.82)	
Familial risk			0.571		0.52
Low	354 (48.1)	142 (51.45)		256 (47.67)	
High/intermediate	361 (49)	133 (48.19)		281 (52.32)	
unknown	21 (2.8)	1 (0.36)			
MSI-H status			0.813		0.601
Yes	634 (86.1)	248 (89.86)		457 (85.10)	
No	73 (10)	26 (9.42)		58 (10.80)	
unknown	29 (3.9)	2 (0.72)		17 (3.17)	
***BRAF1* Val600Glu mutation status**			0.503		0.318
-	589 (80)	219 (79.35)		437 (81.38)	
+	80 (10.9)	35 (12.68)		49 (9.12)	
unknown	67 (9.1)	22 (7.97)		51 (9.5)	
Prognostic and follow up information					
OS status					
alive	380 (51.6)	141 (51.09)		354 (65.92)	
dead	355 (48.2)	134 (48.55)		182 (33.89)	
unknown	1 (0.2)	1 (0.36)		1 (0.19)	
Median OS follow-up:	5.6 years (range: 0.04-11.1)	6.2 years (range:0.036-10.88)		6.3 years (range: 0.38-10.88)	

DFS status					
recurrence/metastasis/death (+)	348 (47.2)	149 (53.99)		213 (39.66)	
recurrence/metastasis/death (-)	387 (52.6)	126 (45.65)		323 (60.15)	
Unknown	1 (0.2)	1 (0.36)		1 (0.19)	
Median DFS follow-up:	5 years (range: 0.04-11.1)	5.1 years (range:0.036-10.88)		5.97 years (range: 0.22-10.88)	
Age			0.95		0.219
Median Age:	62.3 years (range: 20.7-75)	62.4 years (31.75-74.94)		61.22 years (range: 20.7-74.98)	

*(+): presence, (-): absence, DFS: Disease Free Survival, MSI-H: Microsatellite Instability-High, OS: Overall Survival. * p-values for differences between the entire NFCCR and the mtDNA copy number analysis cohort, ** p-values for differences between the entire NFCCR and the NFCCR SNP genotyping cohort. A p-value (obtained by Chi-square or Mann-Whitney U test) less than 0.05 for a variable denotes significant differences between the compared cohorts. Familial risk refers to risk stratification based on Amsterdam and revised Bethesda criteria. Please note that Disease specific survival data for these cohorts were not available.*

2.4.4) Selection of genetic markers in mtDNA

mtDNA polymorphisms genotyped in this study

The first mtDNA polymorphism investigated in this study, the mt16189 (T/C) substitution, is located in the D-loop region (125). This 16189 T to C substitution leads to the formation of a homopolymeric cytosine (homopolymeric C) tract between the positions 16189-16194 resulting in a heteroplasmic length variation of the number of cytosine residues (126). It has been reported that there are differences in the mean mtDNA copy numbers based on the differences in the number of poly-cytosine (poly-C) variants (125). In this study, samples with an uninterrupted poly-C tract (cases with mt16189 C allele) had the lowest mtDNA copy number and samples with the interrupted poly-C tract (cases with mt16189 T allele) had the highest mtDNA copy number (125). There is also a similar study where heteroplasmic length variation has been suggested to affect the mtDNA replication (127). This may be explained by the fact that the SNP 16189 (T/C) is located in the Terminated Associated Sequences that play a major role in the termination of mtDNA synthesis (128). In addition, this polymorphism has been reported to be associated with increased oxidative damage and altered anti-oxidative status (129).

The second mtDNA polymorphism investigated in this study, mt10398 (A/G), is located in the *ND-3* gene, which encodes one of the respiratory subunits of the oxidative phosphorylation pathway. In a study performed on cell lines derived from breast cancer patients, cell lines containing G allele have been reported to have elevated levels of ROS and de-polarized mitochondria when compared to cell lines containing A allele for

mt10398 polymorphism (125,130). In addition, cell lines with the 10398 G allele showed increased colony forming ability. The same study also found that the 10398 G allele promotes metastases in mice (125,130). Therefore, this polymorphism may facilitate cancer progression and worsen outcome in patients.

Since both polymorphisms may disrupt vital biological mitochondrial functions, in this study we hypothesized that they may be associated with outcome in colorectal cancer patients.

Selection of mtDNA SNPs from the genome-wide SNP genotyping data

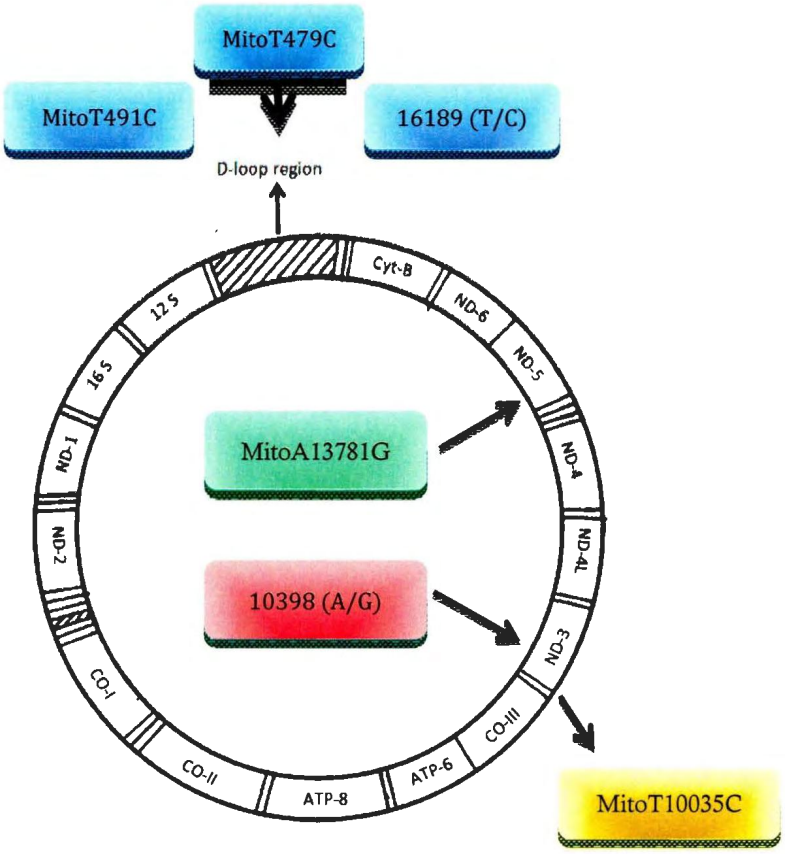
Using the Illumina® HumanOmni-1 Quad BeadChip genome wide SNP genotyping platform the NFCCR obtained genotype data for approximately one million SNPs. The genotyping reaction was performed by the Centrillion® Genomic Services (CA, USA). These genotype data also contained the genotypes for 25 mtDNA SNPs for 537 NFCCR patients with available prognostic data (i.e. NFCCR SNP genotyping cohort) (**Appendix 2**). Upon further analysis, we found that seven of these SNPs were mono-allelic, some SNP-flanking sequences had a significant sequence homology with nDNA sequences (thus it was not clear whether these SNPs were mtDNA or nDNA SNPs, n=7), seven SNPs had very low MAFs (<4%) (**Appendix 2**) and as a result, 4 out of 25 SNPs from the genomewide SNP genotype data were included in the statistical analysis for this project (**Table 4**). The locations of these four mtDNA SNPs as well as the two additional mtDNA SNPs genotyped by the TaqMan® SNP genotyping method in this study are shown in **Figure 4**.

Table 4: The mtDNA polymorphisms obtained from the genomewide SNP genotyping data that are included in this study

SNP ID	Number of missing genotypes	Number of samples with major allele	Number of samples with minor allele	Number of samples that are heteroplasmic	MAF (%)
MitoT479C	6	506	25	0	4.07
MitoT491C	32	478	27	0	5.35
MitoT10035C	1	501	35	0	6.53
MitoA13781G	13	490	34	0	6.49

Heteroplasmic: patients with both alleles for a particular polymorphism, MAF: minor allele frequency

Figure 4: Locations of the six mtDNA polymorphisms investigated in this study



MitoT491C, MitoT479C and 16189(T/C) are located in the D-loop region. MitoA13781G is located in the ND-5 gene, 10398 (A/G) is located in the ND-3 gene and Mito T10035C is located in a tRNA gene that is specific for glycine codon. Drawn not to scale.

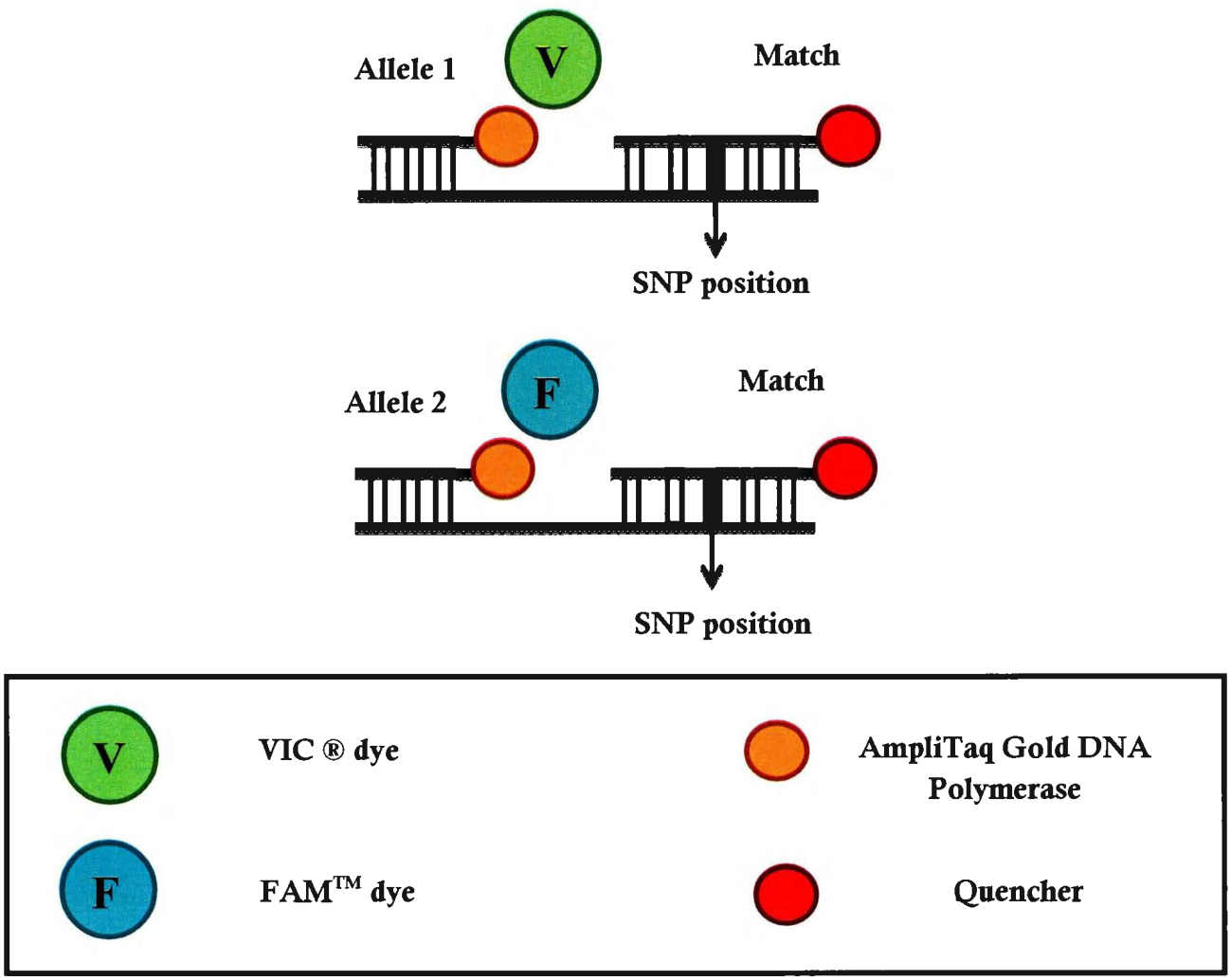
2.4.5) Experimental procedures

2.4.5.1) TaqMan® SNP genotyping (5' nuclease) assays

Principle

The TaqMan® SNP genotyping assays use fluorogenic probes and utilize the 5' nuclease activity of the Taq DNA polymerase. A typical TaqMan® probe contains a reporter fluorescent dye at the 5' end and a quencher dye on the 3' end. When the quencher dye is in close proximity to the reporter dye (which happens before the reporter dye cleavage), the quencher dye suppresses the fluorescence emitted by the reporter dye. During the PCR reaction, if the target sequence is present, the probe binds to the specific target sequence and the forward and reverse primers also bind to their complementary sequences (Figure 5). During the polymerization step, the probe is cleaved by the 5' nuclease activity of the Taq DNA polymerase. By this process the reporter dye is removed from the probe. Since the reporter dye is no longer in close proximity to the quencher, the fluorescent signal of the reporter dye increases. During each cycle, additional reporter dye molecules are cleaved from the probe and thus the reporter fluorescence signal increases every cycle and is proportional to the number of amplicon produced. Depending upon the genotype, the fluorescence intensity of each probe differs for each amplified sample and this information is used to determine the genotypes of the DNA samples.

Figure 5: Principle of TaqMan® genotyping assay



The probe binds to the specific target sequence and the forward and reverse primers also bind to their complementary sequences. The probe is cleaved by the 5' nuclease activity of the Taq DNA polymerase

Primer and probe information for the mtDNA 16189 (T/C) TaqMan® SNP genotyping assay

The TaqMan® SNP genotyping assay containing the primers and probes (one probe for each allele) for the mtDNA 16189 (T/C) SNP was custom designed using the Applied Biosystems' assay design tool (https://www5.appliedbiosystems.com/tools/cadt/) (**Table 5**) as follows: the updated complete mtDNA reference genome sequence (the revised Cambridge sequence; (NC_012920.1)) was obtained from the NCBI's (National Center for Biotechnology Information) nucleotide database. The location of the mtDNA 16189 (T/C) was located on the reference mtDNA genome sequence by manual inspection. The SNP flanking sequences (~80 bases) in both directions were then

Table 5: Primer and probe information for the mtDNA 16189 (T/C) TaqMan® SNP genotyping assay

Forward primer	**Reverse primer**
CACCTGTAGTACATAAAAACCCAATCCA	GGGTTGATTGCTGTACTTGCTTGTA
VIC probe (T allele)	**FAM probe (C allele)**
CCCCCTCCCCATGCT	CCCCCCCCCATGCT

The underlined nucleotides in the sequences above correspond to location of the polymorphic site, the mtDNA 16189 (T/C), which is recognized by the VIC and FAM probes.

imported. The location of other SNPs adjacent to the mtDNA 16189 (T/C) SNP within the SNP flanking sequences were obtained from the mtdbSNP database (131) and were

marked as "N" (i.e. masked nucleotides) before entering into the assay design tool, as recommended by Applied Biosystems.

Primer and probe information for the mt10398 (A/G) TaqMan® SNP genotyping assay

The probe and primer sequences for the mt10398 (A/G) SNP were obtained from a published article (132) (**Table 6**), and manufactured by Applied Biosystems (USA) and Integrated DNA Technologies (IDT) (USA), respectively.

Table 6: Primer and probe information for the mtDNA 10398 (A/G) TaqMan® SNP genotyping assay

Forward primer	**Reverse primer**
CTAGCCCTAAGTCTGGCCTATGA	AATGAGTCGAAATCATTCGTTT TGTTTAAACT
VIC probe (A allele)	**FAM probe (G allele)**
ATTAGACTGAACCGAATTG	ACTGAGCCGAATTG

The underlined nucleotides in the sequence above correspond to location of the polymorphic site, the mt10398 (A/G), which is recognized by the VIC and FAM probes.

Procedure for preparing and storing the TaqMan® SNP genotyping assays

The 40X TaqMan® assay mix (Applied Biosystems, CA, USA) for mtDNA 16189 (T/C) SNP was directly purchased by the Applied Biosystems (CA, USA). The 40X TaqMan® assay mix was aliquoted into separate tubes and were preserved at -20°C. The stock solution was preserved until after the aliquot tubes were consumed up. The assay mix (40X) solutions were diluted to 20X using 1X TE buffer and stored at -20°C.

For the mtDNA 10398 (A/G) SNP, probes and primers were obtained in separate tubes from different manufacturers. Therefore, the 40X assay mixture containing the primers and probes were prepared in our laboratory. The primers were received in pelletized form from the manufacturer (Integrated DNA Technologies, USA) and were suspended in appropriate amounts of de-ionized water to have primer solutions with a concentration of 100μM. The stock primer tubes were then tapped and vortexed gently to ensure that the contents were dissolved completely. 100 μl of the stock primer solution was then aliquoted into 2-3 aliquot tubes and stored at -20^0C.

The 20X TaqMan® SNP genotyping assay mix for the 10398 (A/G) was prepared by adding 56 μl of 1X TE buffer, 18 μl of forward and reverse primer (100 μM each), 4 μl of stock probe 1 and 2 (100 μM each) and by mixing in a 1.5 ml microtube. The 20X assay mix was then spinned down in a mini-centrifuge, aliquoted, and stored at -20^0C.

TaqMan® SNP genotyping methodology

These reactions were performed in 96-well amplification plates. For each reaction plate, the concentration and volumes of reagents in the reaction mix are provided in **Table 7** below.

Table 7: Volume of reagents needed for setting up a TaqMan® SNP genotyping reaction mix (for a 96 well plate reaction)

Contents	Volume
PCR Master Mix (2X)	5µl x 110= **550µl**
TaqMan® Assay Mix (20X)	0.25µl x 110= **27.5µl**
Sterile water	3.75µl x 110= **412.5µl**
Sub-total	**990 µl**

Before starting each experiment, all reaction mixes as well as the DNA plates were spinned down briefly. 80 µl of the master reaction mixes were transferred into the well of a 12-well strip tube. These aliquot mixes were then pipetted (9 µl each) into each of the 8 columns of a MicroAmp® Fast Optical 96-well reaction plate (Part number. 4346906, Applied Biosystems, CA, USA) using a twelve channel pipette. These reaction plates are compatible with the 7900HT Fast Real Time PCR System (Part number. 4330966, Applied Biosystems, CA, USA). 1 µl of the DNA solution from the DNA plates with a concentration of 4ng/µl were added to the corresponding wells in the 96-well reaction plate. The 96-well reaction plate was covered with a compatible optical adhesive film (Part number. 4360954, Applied Biosystems, CA, USA). The optical film was sealed tightly at the edges as well as on the surface to make sure that the reaction solution does not evaporate during the PCR amplification process. The base of the plate was also covered with a Kimwipe to avoid any possible source of contamination until it was placed into the equipment. After that step, the reaction plate was centrifuged at 1000 revolution per minute (rpm) in a bench top centrifuge (catalogue number: 75004367, Sorvall Legend

T+ Centrifuge, ThermoFisher Scientific, MA, USA) for approximately 10 seconds. The reaction plate was then placed in the Applied Biosystems 7900HT Fast Real Time PCR system (Applied Biosystems -USA) or the Veriti thermocycler (Applied Biosystems-USA) for PCR amplification.

There are two file formats that are associated with Applied Biosystems 7900HT Sequence Detection Systems (SDS) software: allelic discrimination file (AD) and the allelic quantification (AQ) file. Both the AD and AQ files were created prior to running the PCR reactions using the SDS software package as described in the *'Applied Biosystems 7900HT Fast Real-time PCR System Allelic Discrimination Getting Started Guide'* manual. In short, the AD files were used to discriminate the different alleles of a particular SNP and it contains the marker (detector) information and reporter dye information (VIC and FAM). On the other hand, the AQ file is required to run the amplification reaction on the Applied Biosystems 7900HT Fast Real Time PCR system equipment. A typical AD file was used for two processes, the pre-read and post-read processes; both were performed in the 7900HT Fast Real time PCR system equipment (Applied Biosystems, CA, USA).

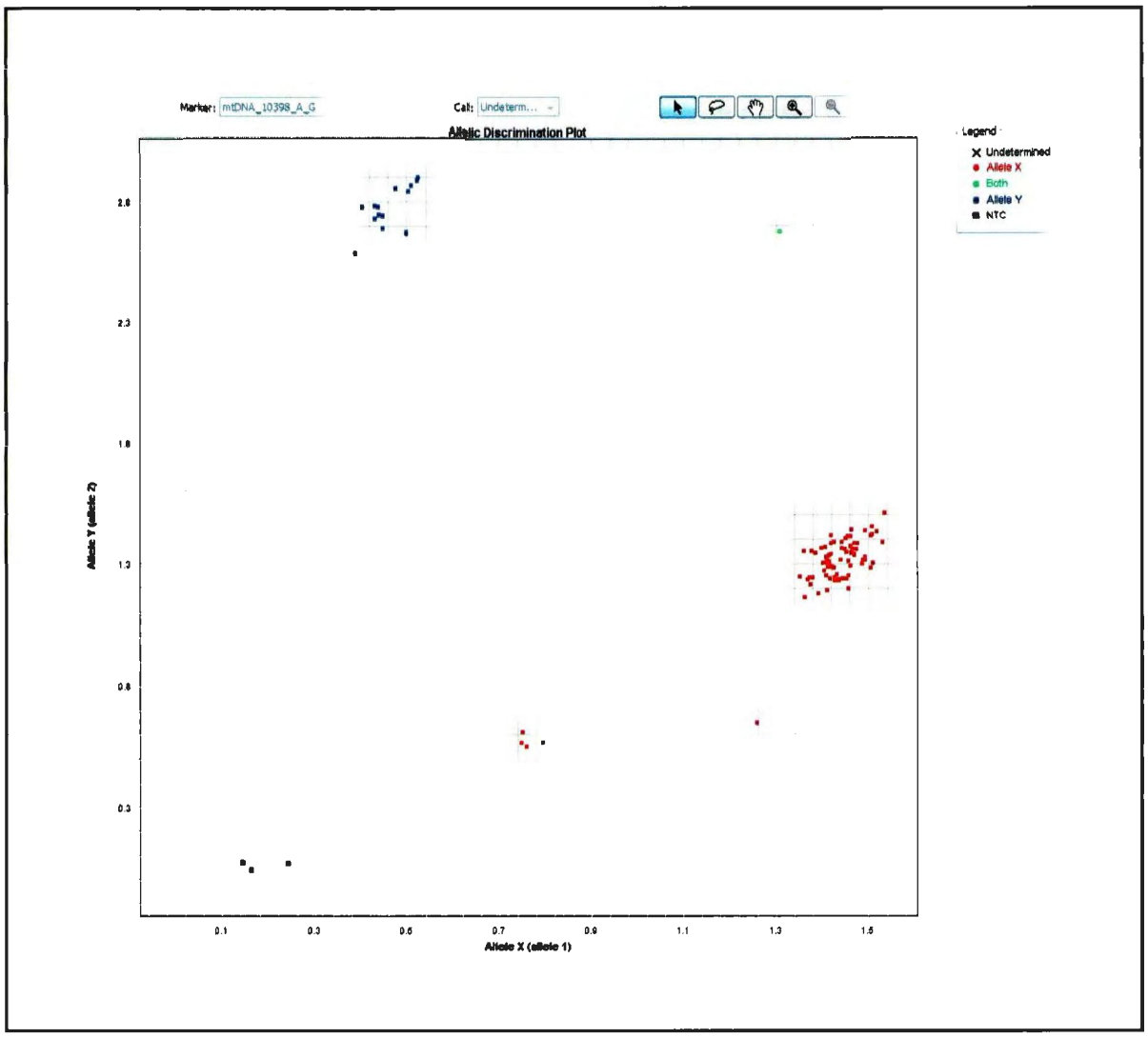
The pre-read step was performed before the PCR amplification process. During the pre-read step, the fluorescence intensity (background signal) of the passive reference dye (ROX) is measured. After that step, the PCR amplifications were performed using either the Veriti thermocycler (Part number: 4375305, Applied Biosystems, CA, USA) or the Applied Biosystems 7900HT Fast Real Time PCR equipment (Part number: 4330966,

Applied Biosystems, CA, USA). The PCR thermocycling conditions are provided in **Appendix 3**.

After the PCR amplification was completed, the post-read step was performed in the Applied Biosystems 7900HT Fast Real-time PCR equipment using the AD file. During the post-read, the final fluorescence intensities at the end of the reaction are recorded by the Applied Biosystems 7900HT Fast Real Time PCR System. The pre-read and post-read data stored in the AD file were then processed by the SDS software during which the fluorescence intensity at the end of the PCR reaction was subtracted from the passive reference signal and the fluorescence intensities were plotted in a scatter plot by the SDS software (**Figure 6**).

In a typical scatter plot, there are two fluorescence intensity scales mapped to X (X_n) and Y (Y_n) axis; one for each probe (allele) respectively. Thus, in a typical plot (**Figure 6**), one can see three genotypes; homozygous for X allele (XX – inclined to X axis), homozygous for Y allele (YY- inclined towards Y axis); and heterozygous (XY- midway between the two axes). Both mtDNA SNPs that are genotyped in this project were almost always homoplasmic. Therefore, only two genotypes in the scatter plots were observed (except one heteroplasmic sample for each mtDNA SNP). The NTCs typically should occupy the region closer to the center of origin of both the axes, unless there is a contamination. The genotypes called automatically were also inspected manually by an independent scientist (Dr. Savas) to confirm the genotype calls.

Figure 6: An example of an allelic discrimination (AD) plot for the mtDNA 16189 (T/C) polymorphism



The black squares near the center of origin show no amplification and are NTCs. The red dots represent those samples which are homoplasmic for the T allele. The blue dots are those samples which are homoplasmic for the C allele. The green dot, which is midway between the two homoplasmic clusters, is a heteroplasmic sample. There a few outlier red dots that represents DNA samples with low amplification which might be due to the low quality of DNA samples or due to the presence of PCR inhibitors.

2.4.5.2) qPCR analysis

Selection of target gene sequences to be amplified in qPCR reaction

In brief, to quantify mtDNA copy number with respect to nDNA in tumour versus non-tumour cells using the qPCR method, we first selected target genes to be amplified based on previously published studies. Because the aim of this study is to identify the differences between the mtDNA copy number between the tumour and adjacent non-tumour tissue, we performed additional analyses to make sure the selected gene sequences are not altered (i.e. deleted or somatically mutated) in colorectal tumours. The presence of polymorphisms at primer and probe binding sequences were also avoided. Since this experiment was quantitative, we also checked to make sure that the control gene in the nDNA did not have a copy number variation (CNV) among individuals. These procedures are described in detail below.

For selecting gene regions for qPCR amplification, initially about 100 nuclear genes were randomly investigated. For this purpose, first, genes were analyzed for deletions (LOH) in gastrointestinal tumour tissue samples using the information published in the CONAN (copy number analysis) database of the Wellcome Trust Sanger Institute, as of October 2011 (133). The *FASLG* showed the least structural variation (one LOH (3% in analyzed cell lines (n=39)) among several genes examined this way (133). A previous study had primer and TaqMan® probe sequences for the target sequence in this gene (134). The same study (134) also had the primer and TaqMan® probe sequences for an mtDNA gene fragment (*ND-2*). These primers and probes were selected to be used in the qPCR method. Additional steps were taken to ensure that known sequence alterations

would not affect binding of the primers and the probe. These steps are explained in **Appendix 4**. The primer and probe sequences for *FASLG* and *ND-2* are provided in **Table 8**.

Table 8: Primers and probe for the *FASLG* and *ND-2* gene fragments used in the qPCR method

FASLG	
GenDIR (primer)	GGC TCT GTG AGG GAT ATA AAG ACA
GenREV (primer)	CAA ACC ACC CGA GCA ACT AAT CT
Probe	VIC-CTGTTCCGTTTCCTGCCGGTGC-TAMRA
ND-2	
mtDNA DIR (primer)	CAC AGA AGC TGC CAT CAA GTA
mtDNA REV (primer)	CCG GAG AGT ATA TTG TTG AAG AG
Probe	FAM-CCT CAC GCA AGC AAC CGC ATC C - TAMRA

Primer and probe information was obtained from a previous literature report (134). *These primers and probes were separately ordered from the Integrated DNA Technologies (USA) and Applied Biosystems (USA), respectively.*

The qPCR methodology

A set of reactions were performed to identify the optimum primer and probe concentration for this reaction. To prepare a 50 µl of 20X qPCR assay mix, primers and probes were mixed together as shown in **Table 9**.

Table 9: Volumes of reagents needed to prepare 50 µl of qPCR assay mix (20X)

	Reagent	**Volume (µl)**
mtDNA *(ND-2)*	Forward primer (100 µM)	2.25
	Reverse primer (100 µM)	2.25
	Probe (100 µM)	0.5
nDNA *(FASLG)*	Forward primer (100 µM)	18
	Reverse primer (100 µM)	18
	Probe (100 µM)	4
	1X TE buffer	5
	Total	50

All qPCR reactions were performed in the 7900 HT Fast Real-Time PCR system. For each patient, two separate qPCR reactions were performed in triplicates: one for the tumour extracted DNA and the other for the non-tumour extracted DNA. The DNA samples extracted from tumour and non-tumour samples from each patient were amplified in the same reaction plate to minimize inter-assay variability. In qPCR reaction, each well contained the following reagents as shown in **Table 10**.

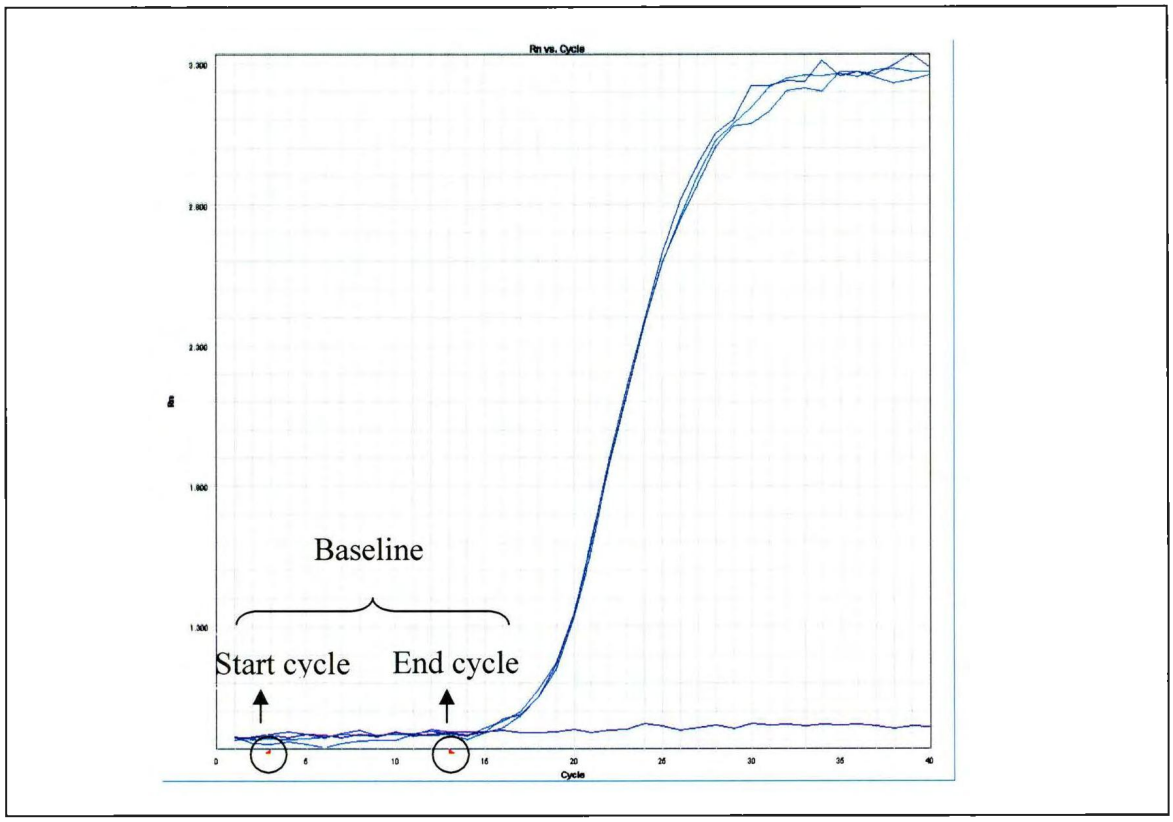
Table 10: The qPCR reaction mix

Reagent	Volume
PCR Master Mix (2X)	5µl
TaqMan® Assay Mix (20X)	0.25µl
Sterile water	3.25µl
DNA solution (5 ng/µl)	1.5 µl
Total volume	10.0 µl

According to the Applied Biosystems' guidelines, the SDS Relative Quantitation (SDS-RQ) file was set up prior to the reaction. The reaction condition for qPCR was as follow: 1) Activation of AmpErase® UNG at 50 °C for 2 mins, 2) AmpliTaq Gold Polymerase activation at 95 °C for 10 mins, and 34-35 cycles of 3) denaturation of DNA at 95 °C for 15 sec and 4) primer annealing and extension at 60 °C for 1 min.

After the qPCR reaction was over, the empty wells were excluded from the analysis to eliminate background noise. The SDS file was saved and the SDS-RQ file was analyzed using the RQ manager software. The RQ manager software generates two different plots, namely fluorescence intensities (R_n) versus cycle number and baseline normalized fluorescence intensities (ΔR_n) versus cycle number. As recommended by Applied Biosystems, in order to reduce the background noise, a baseline correction was performed. This was performed by selecting the start and end cycle values from the R_n versus cycle number plot (**Figure 7**). The start cycle range value was set to cycle number three and the end cycle value was selected by choosing the amplification curve that had the lowest C_T

Figure 7: R_n versus cycle number plot in qPCR

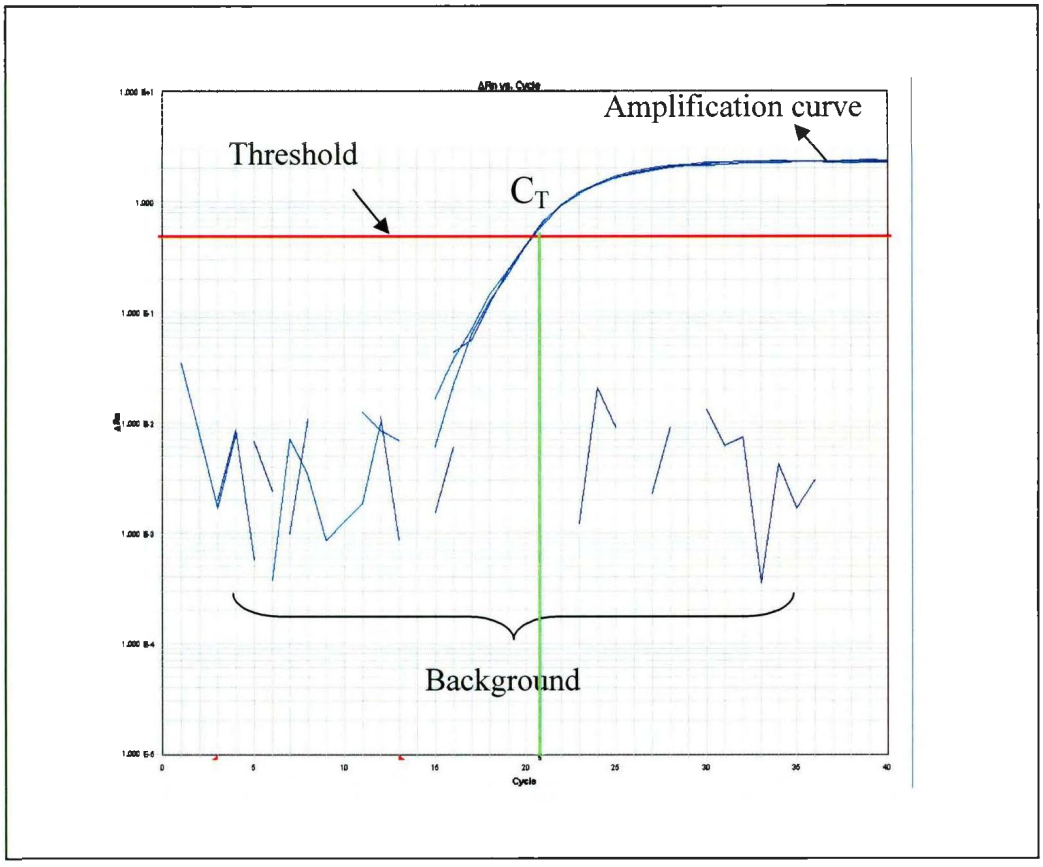


Endcycle is fixed three cycles below the cycle number at which a detectable amplification starts

(the cycle value at which amplification begins) value in the R_n versus cycle number plot and subtracting three cycles from it. The corrected baseline cycle values were then set for each reaction file. In addition, the RQ manager software generates a threshold line for threshold cycle (C_T) (the cycle number at which the threshold line crosses the amplification curve) calculation, separately for mtDNA and nDNA in the delta R_n plots (**Figure 8**). This threshold was set either automatically by the software or manually in the exponential phase as recommended by Applied Biosystems. The cycle value which intersects the threshold line is called the threshold cycle (C_T).

The exported C_T values were then organized and analyzed for standard deviation (SD) among the three replicate amplifications (triplicates) for each sample using a Microsoft Excel® program. Since qPCR for each DNA sample was performed in triplicates, initially there were three C_T values for every DNA sample. For reliable qPCR results, according to the Applied Biosystems guidelines, the C_T values should be normalized based on at least two repeated (replicated) reactions (i.e. within triplicates of each DNA sample) and with a SD value of C_T <0.30.

Figure 8: ΔR_n versus cycle number plot in qPCR



In other words, the variability among the repeated amplifications of a DNA sample should be minimal. Samples that did not satisfy the criterion (i.e. SD <0.3) between triplicates were reanalyzed by excluding an outlier C_T value. SD was then re-calculated based on the remaining two C_T values. If still the SD was >0.3 in these two C_T values, the qPCR reactions for these patients (for both the tumour and non-tumour tissue extracted DNA) were repeated. Once C_T values that satisfied the above criteria were obtained, average C_T values were calculated for each sample group (i.e. tumour and non-tumour DNA of each patient). From the average C_T value, the mtDNA copy number change between the tumour and non-tumour tissue for each patient were determined using the delta-delta C_T ($\Delta\Delta C_T$) method (135), as shown in the equation below:

$$\text{mtDNA copy number change} = 2^{-\Delta\Delta C_T}$$

$$= 2^{-(\Delta C_{T\,(\text{tumour})} - \Delta C_{T\,(\text{non-tumour})})}$$

Where,

$$\Delta C_{T\,(\text{tumour})} = C_{T\,(\text{tumour mtDNA})} - C_{T\,(\text{tumour nDNA})}$$

$$\Delta C_{T\,(\text{non-tumour})} = C_{T\,(\text{non-tumour mtDNA})} - C_{T\,(\text{non-tumour nDNA})}$$

In the above equation, for each sample, the quantity of the mtDNA amplification (i.e. mtDNA copy number) is normalized to the quantity of nDNA amplification to calculate the relative number of mtDNA with respect to nDNA. This information is then used to calculate the relative mtDNA copy number of the tumour tissue when compared

with the non-tumour tissue sample of each patient. When the mtDNA copy number ratio is less than one it means there is a decrease in the number of mtDNA copies in the tumour tissue when compared to normal tissue. When the mtDNA copy number ratio is one or greater than one then there is an increase in the mtDNA copy number in tumour when compared to non-tumour tissue. For statistical analyses, mtDNA ratios were rounded to three decimals.

2.4.6) Quality control measures and statistical analyses

Calculation of genotype replication rate

When performing genotyping reactions using the TaqMan® SNP genotyping method, genotyping reactions for a number of samples were repeated to check if results obtained were consistent. The duplication rates for the samples were calculated by the following equation:

$$Duplication\ rate = \frac{number\ of\ samples\ genotyped\ succesfully\ at\ least\ twice}{total\ number\ of\ samples\ genotyped\ successfully}$$

Calculation of genotype concordance rate

The concordance rate was calculated using the genotyping results for samples genotyped twice by the TaqMan® SNP genotyping method using the following formula:

$$Concordance\ rate = \frac{Number\ of\ duplicated\ samples\ with\ concordant\ genotypes}{Number\ of\ samples\ duplicated}$$

Determination of minor allele frequencies

The genotypes for each of the 6 mtDNA SNPs (obtained by both the TaqMan® SNP genotyping method and the genomewide SNP genotyping method) were organized and examined in Microsoft® Excel sheets to identify the major homozygotes, minor homozygotes and heterozygotes for each SNP. For each SNP, the minor allele frequency (MAF) was also computed using the numbers of major homozygotes, heterozygotes and minor homozygotes using the following equation:

$$MAF = \frac{((2 * aa) + Aa)}{2 * (AA + Aa + aa)}$$

where, "A" is the major allele and "a" is the minor allele.

Coding of categorical variables for statistical analysis

For statistical analysis of mtDNA SNPs in relation to prognosis, the organized genotype data were coded for statistical analysis based on the genotypes: the major genotypes were coded as zero and the minor genotypes were coded as one. This way, the minor genotype is compared to the major genotype during the statistical analysis. In the case of mtDNA 16189 T/C and mtDNA 10398 A/G polymorphisms, there was one (separate) patient who was heteroplasmic for these SNPs. These patients were excluded from the statistical analysis. Other clinicopathological variables were coded as follows:

male=1 and female=0; mucinous histology=1 and non-mucinous histology =0; rectal cancer=1 and colon cancer=0; poorly or undifferentiated tumours=1 and well or moderately differentiated tumours=0; presence of vascular invasion in tumours=1 and those with no vascular invasion=0; presence of *BRAF* (Val600Glu) mutation =1 and absence of *BRAF* (Val600Glu) mutation=0; high or moderate familial risk=1 and low familial risk=0; MSI-H=1 and MSS/MSI-L absence of MSI-H=0; and stage I=1, stage II=2, stage III=3 and stage IV=4. Since age is a continuous variable it was not dichotomised; rather it was analyzed as a continuous variable. For statistical analysis of the per results, the increase in mtDNA in tumour tissue with respect to non-tumour tissue was coded as one and samples with decreased mtDNA copy number in tumour tissues with respect to non-tumour tissues were coded as zero.

Measures of outcome

Survival analyses were performed to test whether the mtDNA genetic variations investigated were associated with outcome. For this analysis, we used two measures of outcome: overall survival (OS) was the primary outcome and disease free survival (DFS) was the secondary outcome. **OS** is the time interval from diagnosis to the last follow up or to the date of death from any cause. **DFS** is the time interval from diagnosis to the last follow up or to the date of death from any cause, recurrence or metastasis, whichever has occurred earlier.

Kaplan Meier curves

The Kaplan Meier method is used to compare the survival characteristics over specified time interval for two or more group of patients stratified based on specific

clinicopathological characteristics (136). Kaplan Meir method was used to generate the survival curves in this project.

Univariate Cox regression analysis

In this method, patients are grouped in two or more groups based on their variable characteristics. The probability of occurrence of an event in one group of patient when compared to another group of patients is called the hazard ratio (HR) (71). In the univariate analysis, each of the variables is analyzed individually to find out whether they are correlated with the outcome separately. In our study, for univariate analysis, age was analyzed as a continuous variable. The remaining variables were categorized into two groups, except for stage which had four categories (categories coded as "0" were used as reference group) except in stage where stage II, stage III, and stage IV patients were separately compared to stage I patients (the reference group).

Chi-square test

The Chi-square test statistic compares the difference between the observed and expected frequencies for groups of patients for a categorical variable. This test was used to compare the differences between patient cohorts, as well as the associations of genotypes with clinicopathological features.

Mann-Whitney U test

The Mann-Whitney U test was used to compare the differences in the age distribution between the patient cohorts, (**Table 3**), as age was not normally distributed in these patient cohorts (*data not shown*).

All statistical analyses were performed using the Statistical Package for the Social Sciences (SPSS) (version 19), unless otherwise stated. The statistical significance level was set at $p<0.05$ and all statistical tests were double-sided. Due to the hypothesis generating nature of this study, no correction for multiple testing was performed (137).

2.5) Results

2.5.1) mtDNA polymorphisms

Characteristics of the NFCCR SNP genotyping cohort

The baseline characteristics of the cohort investigated for the mtDNA SNPs (i.e. the NFCCR SNP genotyping cohort) are shown in **Table 3**. In this cohort (n=537), the median age at diagnosis was 61.2 years, the median OS follow up time was 6.3 years, and the median DFS follow up time was ~6 years. When compared to the entire NFCCR cohort, this cohort had fewer stage IV patients (20.8% versus 9.9%; $p<0.001$), fewer patients with tumours characterized by vascular invasion (38.3% versus 31.8%, $p=0.013$), and fewer patients with tumours characterized by lymphatic invasion (38.7% versus 33.3%; $p=0.034$).

Statistical analysis results of the mtDNA polymorphisms

The genotype frequencies of the six mtDNA SNPs included in this study are shown in **Table 11**. The genotype frequencies were generally similar to the frequencies

reported in the dbSNP database (138) or previous findings published in the literature (127,139-142) for Caucasian populations.

For the two mtDNA SNPs (10398 (A/G) and 16189 (T/C)) (**Table 11**), 13.4% and 6.5% of the genotypes were duplicated, respectively. The replicate genotypes were 100% concordant.

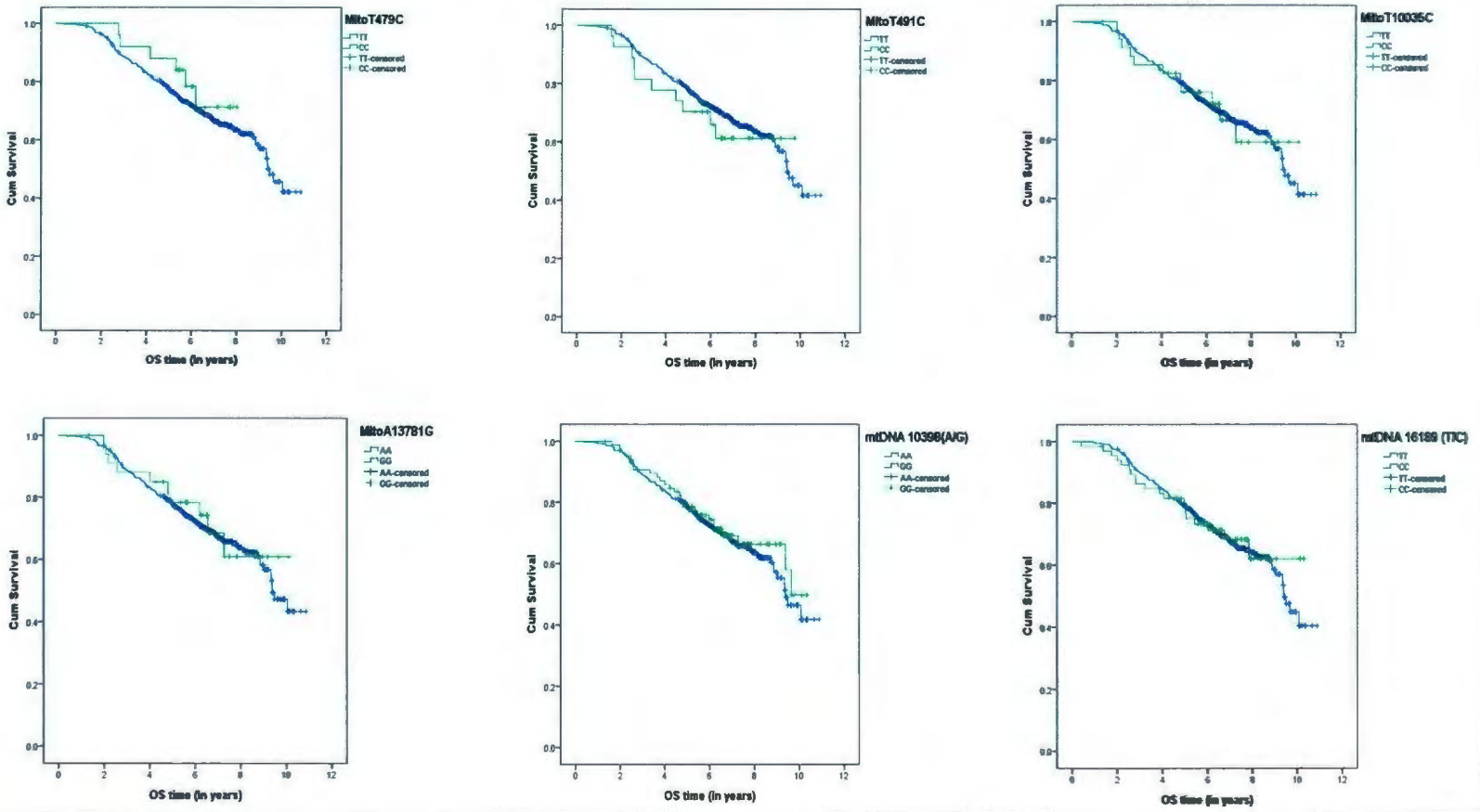
Table 11: Genotype frequencies of the six mtDNA polymorphisms investigated in this study

SNP	Major genotype (%)	Minor genotype (%)	Heteroplasmy (%)	No.of patients with missing genotype
10398 (A/G)	A (83.6)	G (16.2)	0.19	--
16189 (T/C)	T (87.15)	C (12.66)	0.19	--
MitoT479C	T (95.3)	C (4.7)	--	6
MitoT491C	T (94.65)	C (5.35)	--	32
MitoT10035C	T (93.47)	C (6.53)	--	1
MitoA13781G	A (93.51)	G (6.49)	--	13

Using univariate analysis to test the association of these six mtDNA SNPs with survival, the survival differences of the patients grouped based on their genotypes were not found to be significantly different. The univariate Kaplan-Meier survival curves for the six mtDNA SNPs for OS and DFS are presented in **Figures 9** and **10**, respectively. The univariate Cox regression analysis results for OS and DFS are also summarized in **Tables 12** and **13**, respectively.

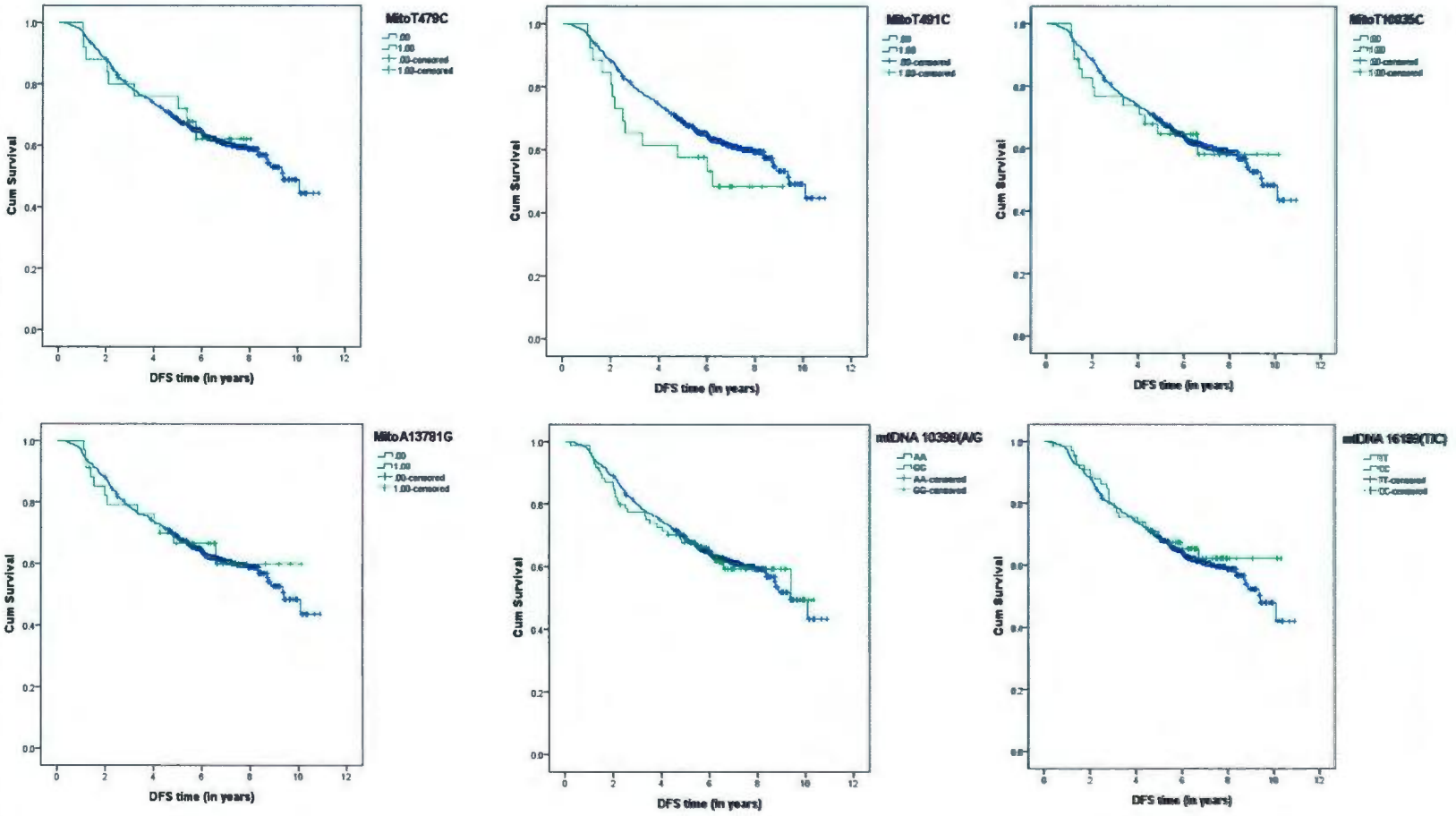
In OS univariate analysis, sex, stage, vascular invasion, lymphatic invasion and MSI status were significantly associated with OS (**Table 12**). Specifically, male patients had increased risk of death when compared to female patients. Stage III and IV patients, when compared to the stage I patients had shorter OS times. Patients with tumours characterized by vascular or lymphatic invasions had increased risk of death when compared to patients without lymphatic or vascular invasion. Patients with MSI-H tumours had longer OS compared to patients with the non-MSI tumours. For these variables, similar results were obtained in the DFS analysis as well (**Table 13**). In addition, in the DFS analysis, rectal cancer patients had increased risk of event (i.e. recurrence, metastasis, or death; **Table 13**) when compared to colon cancer patients.

Figure 9: Kaplan Meier plots for six mtDNA polymorphisms investigated in this study (overall survival)



OS time is defined as the time interval from the time of diagnosis to the date of last follow up or the date of death. Cumulative survival refers to the proportion of patients who are alive at a particular point of time. Survival plots were generated by the Kaplan Meier method.

Figure 10: Kaplan Meier plots for six mtDNA polymorphisms investigated in this study (disease free survival)



DFS time is defined as the time interval from the time of diagnosis to the date of last follow up or the date of recurrence or metastasis or death. Cumulative survival refers to the proportion of patients who are alive at a particular point of time. Survival plots were generated by Kaplan Meier method.

Table 12: Univariate analysis results for the six mtDNA polymorphisms and clinico-pathological variables (overall survival)

Variables	n	Number of patients in each category	p-value	HR	95% CI	
					Lower	Upper
MitoT479C (C vs T)	530	(25 vs 505)	0.455	0.733	0.324	1.656
MitoT491C (C vs T)	504	(27 vs 477)	0.625	1.172	0.619	2.221
MitoT10035C (C vs T)	535	(35 vs 500)	0.923	0.97	0.527	1.785
MitoA13781G (G vs A)	523	(34 vs 489)	0.74	0.897	0.474	1.699
mtDNA 10398 (G vs A)	529	(86 vs 443)	0.605	0.897	0.595	1.353
mtDNA 16189 (C vs T)	529	(66 vs 463)	0.803	0.943	0.592	1.502
Sex (male vs female)	536	(329 vs 207)	**0.012**	**1.493**	**1.092**	**2.042**
Histology (mucinous vs non-mucinous)	536	(61 vs 475)	0.898	0.97	0.61	1.544
Location (rectum vs colon)	536	(180 vs 356)	0.193	1.22	0.905	1.645
Stage			**<.001**			
Stage (II vs I)	536	(206 vs 98)	0.176	1.455	0.845	2.507
Stage (III vs I)		(179 vs 98)	**0.004**	**2.199**	**1.289**	**3.751**
Stage (IV vs I)		(53 vs 98)	**<.001**	**10.14**	**5.769**	**17.822**
Grade (poorly differentiated/undifferentiated vs well/moderately differentiated)	532	(38 vs 494)	0.673	0.877	0.476	1.614
Vascular invasion (+ vs -)	497	(171 vs 326)	**<.001**	**1.724**	**1.274**	**2.333**
Lymphatic invasion (+ vs -)	494	(179 vs 315)	**0.003**	**1.581**	**1.171**	**2.136**
Familial risk (high/moderate vs low)	536	(281 vs 255)	0.65	1.07	0.799	1.433
MSI-H status (Yes vs No)	514	(58 vs 456)	**<.001**	**0.233**	**0.103**	**0.527**
BRAF mutations status (+ vs -)	485	(49 vs 436)	0.404	0.797	0.468	1.357
Age	536	*	0.438	1.006	0.99	1.023

(+): presence, (-): absence, CI: Confidence Interval, HR: Hazard Ratio, MSI-H: Microsatellite Instability-High, vs: versus. Values that are bolded are statistically significant (p value <0.05).
**Age is analyzed as a continuous variable.*

Table 13: Univariate analysis results for six mtDNA polymorphisms and clinico-pathological variables (disease free survival)

Variables	n	Number of patients in each category	p-value	HR	95% CI	
					Lower	Upper
MitoT479C (C vs T)	529	(25 vs 504)	0.923	0.968	0.496	1.889
MitoT491C (C vs T)	503	(26 vs 477)	0.179	1.47	0.838	2.582
MitoT10035C (C vs T)	534	(35 vs 499)	0.963	1.014	0.578	1.776
MitoA13781G (G vs A)	522	(34 vs 488)	0.841	0.942	0.526	1.688
mtDNA 10398 (G vs A)	528	(85 vs 443)	0.934	1.016	0.7	1.474
mtDNA 16189 (C vs T)	528	(66 vs 462)	0.544	0.874	0.567	1.349
Sex (male vs female)	535	(328 vs 207)	**0.009**	**1.473**	**1.103**	**1.967**
Histology (mucinous vs non-mucinous)	535	(61 vs 474)	0.785	0.941	0.611	1.452
Location (rectum vs colon)	535	(180 vs 355)	**0.029**	**1.359**	**1.031**	**1.79**
Stage			**<.001**			
Stage (II vs I)	535	(206 vs 97)	0.244	1.327	0.825	2.133
Stage (III vs I)		(179 vs 97)	**0.002**	**2.121**	**1.332**	**3.378**
Stage (IV vs I)		(53 vs 97)	**<.001**	**5.737**	**3.458**	**9.518**
Grade (poorly differentiated/ undifferentiated vs well/moderately differentiated)	531	(38 vs 493)	0.487	0.813	0.454	1.456
Vascular invasion (+ vs -)	496	(171 vs 325)	**0.001**	**1.644**	**1.241**	**2.177**
Lymphatic invasion (+ vs -)	493	(179 vs 314)	**0.003**	**1.534**	**1.161**	**2.028**
Familial risk (high/moderate vs low)	535	(280 vs 255)	0.291	1.157	0.883	1.517
MSI-H status (Yes vs No)	513	(57 vs 456)	**0.001**	**0.347**	**0.183**	**0.655**
BRAF mutations status (+ vs -)	485	(49 vs 436)	0.66	0.899	0.56	1.444
Age	535	*	0.9	1.001	0.986	1.016

(+): presence, (-): absence, CI: Confidence Interval, HR: Hazard Ratio, MSI-H: Microsatellite Instability-High, vs: versus. Values that are bolded are statistically significant (p value <0.05)

**age is analyzed as a continuous variable.*

Other clinicopathological variables such as age, histology, grade, familial risk status, and *BRAF*-Val600Glu mutation status were not significantly associated with either OS or DFS, while certain other clinicopathological variables remained significant (**Tables 12** and **13**).

2.5.2) mtDNA copy number

Characteristics of the mtDNA copy number change cohort

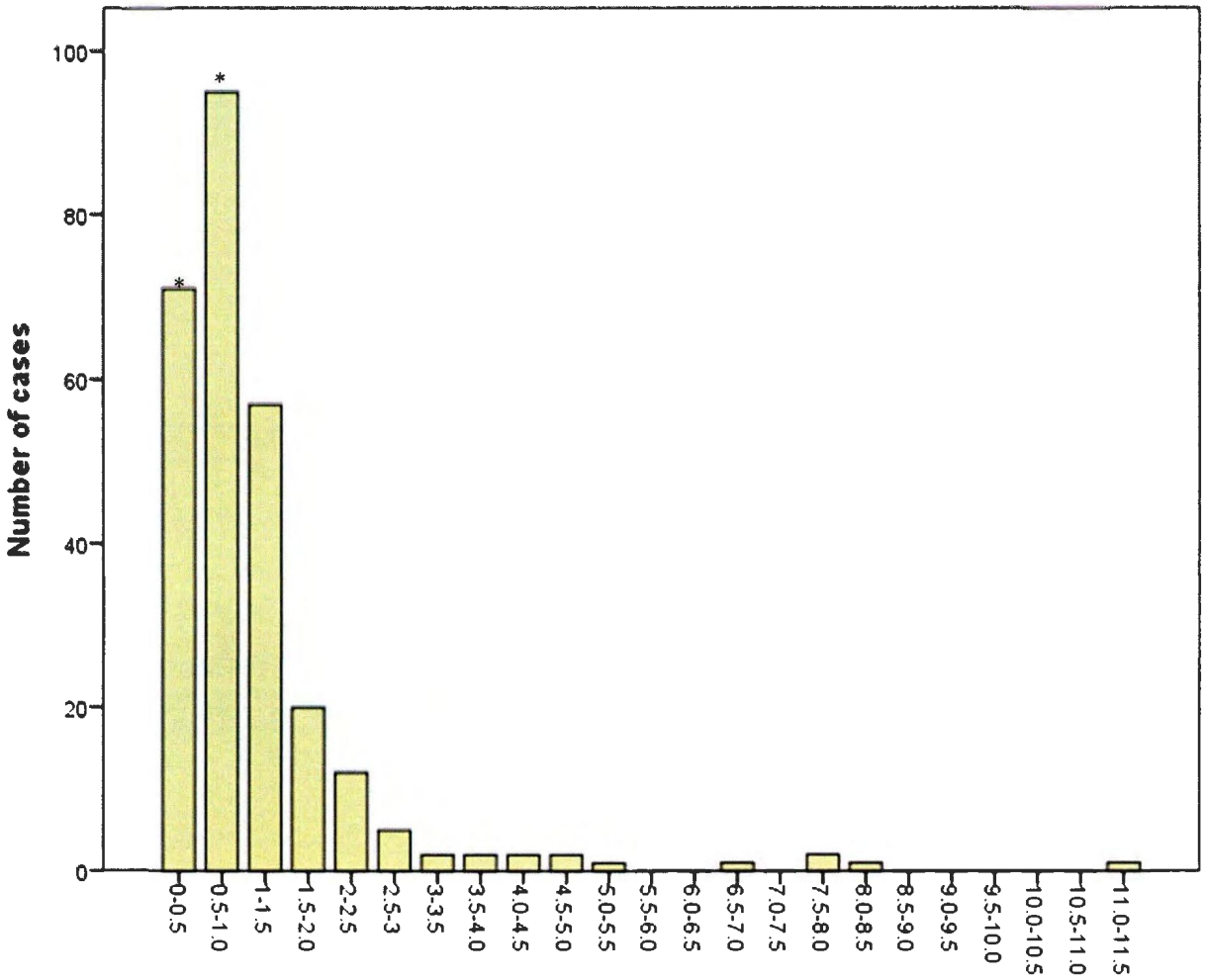
The baseline characteristics of 276 patients investigated in this analysis (the mtDNA copy number analysis cohort) are summarised in **Table 3**. The median age was 62.3 years and the median OS and DFS follow up times were 6.2 and 5.1 years, respectively. The clinico-pathological features of these 276 patients did not differ from those of the entire NFCCR cohort (**Table 3**).

Distribution of the mtDNA copy number change

The mtDNA copy number ratio was estimated for 274 patients. The distribution of the mtDNA copy number ratio for this patient cohort is illustrated in **Figure 11**.

As shown in **Figure 11**, a wide range of mtDNA copy number ratio was detected in this patient cohort (i.e. 0.003-11.02). The majority of patients had a ratio between 0.5-2.0 (n=172, 62.8%). A decrease of mtDNA copy number in tumour tissue was observed in 166 (60.6%) of the patients and increase in the mtDNA copy number was observed in 108 (39.4%) of the patients.

Figure 11: Distribution of the relative mtDNA/nDNA ratio in tumour tissue when compared to non-tumour tissue



The histogram shows the distribution of the mtDNA copy number ratio in 274 colorectal cancer patients. Bars with mtDNA ratio less than 1 (i.e. the bars with asterisks) corresponds to patients with decreased mtDNA copy number in tumour tissue with respect to non-tumour tissues, whereas patients with ratio equal to or greater than one had increase in mtDNA copy number in tumour tissue with respect to non-tumour tissues. mtDNA ratios were rounded to three decimals for this analysis.

According to the univariate analysis the mtDNA copy number change was not associated with either OS (**Table 14**) or DFS (**Table 15**). Yet, there were other clinic-pathological variables that were significant in both OS and DFS analysis.

To see if there were differences in the baseline characteristics of patients with increased or decreased mtDNA copy number, a Chi-square analysis was also performed. As a result, no significant association between any of these variables and the mtDNA copy number change was detected (**Table 16**).

Table 14: Overall survival univariate analysis for the mtDNA copy number change and clinicopathological variables

Variables compared	n	No of patients in each category	p-value	HR	95.0% CI for HR	
					Lower	Upper
Sex (male vs female)	273	(158 vs 115)	0.055	1.412	0.993	2.009
Histology (mucinous vs non-mucinous)	273	(33 vs 240)	0.73	0.91	0.531	1.557
Location (rectum vs colon)	273	(92 vs 181)	0.376	0.85	0.594	1.217
Stage	**273**	**31**	**<0.001**			
Stage (II vs I)		(91 vs 31)	0.297	1.597	0.663	3.847
Stage (III vs I)		**(101 vs 31)**	**<0.001**	**3.065**	**1.314**	**7.15**
Stage (IV vs 1)		**(50 vs 31)**	**<0.001**	**17.45**	**7.385**	**41.25**
Grade (poorly differentiated/undifferentiated vs well/moderately differentiated)	**272**	**(31 vs 241)**	**0.019**	**1.796**	**1.103**	**2.926**
Vascular invasion (+ vs -)	**247**	**(111 vs 136)**	**<0.001**	**2.187**	**1.527**	**3.131**
Lymphatic invasion (+ vs -)	**247**	**(113 vs 134)**	**<0.001**	**1.937**	**1.355**	**2.77**
Familial risk (high/moderate vs low)	272	(132 vs 140)	0.973	1.006	0.715	1.415
MSI status (MSI-H vs MSS/MSI-L)	**271**	**(26 vs 245)**	**0.004**	**0.23**	**0.085**	**0.623**
BRAF status 600ValGlu (+ vs -)	251	(35 vs 216)	0.195	1.374	0.85	2.22
Age	**273**	*	**0.025**	**1.023**	**1.003**	**1.043**
mtDNA copy number (increase vs decrease)	273	(108 vs 165)	0.389	1.163	0.825	1.64

(+): presence, (-): absence, CI: Confidence Interval, HR: Hazard Ratio, MSI-H: Microsatellite Instability-High, vs: versus. Values that are bolded are statistically significant (p value <0.05). age was analyzed as a continuous variable.*

Table 15: Disease free survival univariate analysis for the mtDNA copy number change and clinicopathological variables

Variables compared	n	No of patients in each category	p-value	HR	95.0% CI for HR	
					Lower	Upper
Sex (male vs female)	272	(157 vs 115)	**0.046**	**1.41**	**1.007**	**1.964**
Histology (mucinous vs non-mucinous)	272	(33 vs 239)	0.757	0.923	0.557	1.531
Location (rectum vs colon)	272	(92 vs 180)	0.969	0.993	0.709	1.393
Stage	272	30	**<0.001**			
Stage (II vs I)		(91 vs 30)	0.324	1.474	0.682	3.185
Stage (III vs I)		(101 vs 30)	**0.006**	**2.85**	**1.36**	**5.985**
Stage (IV vs 1)		(50 vs 30)	**<0.001**	**11**	**5.155**	**23.558**
Grade (poorly differentiated/undifferentiated vs well/moderately differentiated)	271	(31 vs 240)	**0.029**	**1.69**	**1.055**	**2.716**
Vascular invasion (+ vs -)	246	(111 vs 135)	**<0.001**	**2.05**	**1.457**	**2.877**
Lymphatic invasion (+ vs -)	246	(113 vs 133)	**<0.001**	**1.88**	**1.34**	**2.644**
Familial risk (high/moderate vs low)	271	(131 vs 140)	0.567	1.099	0.795	1.52
MSI status (MSI-H vs MSS/MSI-L)	270	(25 vs 245)	**0.014**	**0.39**	**0.181**	**0.826**
***BRAF* status Val600Glu (+ vs -)**	251	(35 vs 216)	**0.049**	**1.56**	**1.003**	**2.421**
Age	272	*	0.103	1.015	0.997	1.034
mtDNA copy number (increase vs decrease)	272	(108 vs 164)	0.389	1.154	0.833	1.601

*(+): presence, (-): absence, CI: Confidence Interval, HR: Hazard Ratio, MSI-H: Microsatellite Instability-High, vs: versus. Values that are bolded are statistically significant (p value <0.05). * age was analyzed as a continuous variable.*

Table 16: Differences in the clinicopathological features of patients with increased or decreased mtDNA copy number (in tumour tissue when compared to non-tumour tissue)

Variables compared	p-value
Sex	0.352
Histology	0.998
Location	0.34
Stage	0.075
Grade	0.306
Vascular invasion	0.935
Lymphatic invasion	0.977
Familial risk	0.754
MSI status	0.074
BRAF mutation status	0.148

BRAF mutation status: presence or absence of Val600Glu mutation status MSI: Microsatellite Instability. P-values are calculated by Chi square test.

2.6) Discussion

Mitochondrial dysfunctions are implicated in cancer progression (113). Therefore, it can be hypothesized that genetic alterations in mtDNA (such as polymorphisms) can influence cancer progression and affect outcome. Despite this biological relevance, genetic alterations in mtDNA have not been investigated for their prognostic significance in colorectal cancer extensively. In this study, the genotypes obtained for the six mtDNA polymorphisms in 537 colorectal cancer patients were analyzed for their associations with patient prognosis. In addition, change in the mtDNA copy number in tumour tissue with respect to non-tumour tissue was investigated in relation to prognosis in 276 colorectal cancer patients.

mtDNA SNPs and prognosis in colorectal cancer

The results obtained in this study showed that none of the six mtDNA SNPs genotyped using the DNA samples extracted from blood cells was correlated with prognosis in colorectal cancer.

Out of the six SNPs investigated in this study, the mtDNA 10398 (A/G) SNP is a non-synonymous substitution (Thr114Ala) located in the *ND3* gene (**Figure 3**), which encodes for one of the respiratory subunits of the oxidative phosphorylation pathway. This SNP has been reported in functional studies to promote metastasis, increase the production of reactive oxygen species and promote resistance to apoptosis (130).

The second SNP, mtDNA 16189 (T/C), is located in the D-loop region (**Figure 3**), which is the non-coding regulatory region of mtDNA (125). It has been shown that the mtDNA 16189 (T/C) may result in a homopolymeric C-tract and this might result in replication slippage, which in turn results in heteroplasmic length variation in the number of cytosine residues (126). This heteroplasmic length variation has also been hypothesized to affect the mtDNA replication and hence the mtDNA copy number (127).

In the case of the remaining four mtDNA SNPs included in this study, two SNPs MitoT479C and MitoT491C are located in the D-loop, MitoA13781G is a non-synonymous substitution (Ile482Val) located in the *ND5* gene, and the MitoT10035C is located in a tRNA gene specific for glycine codon (**Figure 3**). These four SNPs have been previously investigated for their associations with all cause and cancer specific mortalities in a Scottish cohort of colorectal cancer patients (142). Similar to the findings obtained in this study, these authors did not find these SNPs to be associated with the patient outcomes (142). Although these SNPs have been investigated in relation to prognosis in colorectal cancer before and were not associated with outcome, these SNPs were investigated in this study as there may be cohort specific differences such as event rate, length of follow up. This was the main rationale in testing these four SNPs in the current study

One limitation of the present study was the significant difference between the entire NFCCR cohort and the NFCCR SNP genotyping cohort in terms of baseline clinicopathological features, such as stage ($p<0.001$) and vascular ($p=0.013$) and lymphatic invasions ($p=0.034$) as only patients with available DNA samples extracted

from blood were included in this study (**Table 3**). Specifically, there were a lower proportion of stage IV patients (9.9% versus 20.8%) and a higher proportion of patients with tumours characterized by the presence of vascular (31.8% versus 38.3%) and lymphatic invasion (33.3% versus 38.7%) in the NFCCR SNP genotyping cohort compared to the entire NFCCR cohort. Because of these clinical differences, the NFCCR SNP genotyping cohort was not representative of the NFCCR cohort (a population-based patient cohort). This study also did not include all known polymorphisms on the mtDNA. On the other hand, this study is the first study that investigated the association of the mtDNA 10398 (A/G) and the mtDNA 16189 (T/C) SNPs with prognosis in colorectal cancer patients. The sample size was quite large (n=537), though it is possible that the sample size was not large enough to detect the possible prognostic effects of the mtDNA polymorphisms analyzed in this study.

mtDNA copy number change and prognosis in colorectal cancer

Increase or decrease in the mtDNA copy number in tumour tissue with respect to the non-tumour tissue has been detected in several cancer types before. For example, increase in mtDNA copies in tumour tissues have been observed in renal oncocytomas (143,144), salivary gland oncocytomas (144), head and neck cancers (145), papillary thyroid carcinomas (146), endometrial cancer (147), ovarian cancer (148) and prostate cancer (149). On the other hand, decrease in the mtDNA copy numbers has also been observed in renal carcinomas (143,144,150,151,151), hepatocellular carcinomas (151-153), gastric cancer (154) and breast cancer (146,155,156). In colorectal cancer, increase in mtDNA copy number was observed in 40% of the patients in one study (122) and 39%

of the patients in another study (118). It should also be noted that there is variability between studies in detecting and defining the mtDNA copy number change and hence the categorization of colorectal cancer patients based on the mtDNA copy number status also differs among these studies. Therefore, our results cannot be compared directly to previously published results. In our study, increase in the mtDNA copy number in tumour tissues was observed in 39.4% colorectal cancer patients with respect to non-tumour DNA.

Previously, change in mtDNA copy number and its association with patient prognosis has also been investigated in different cancer types. For example, decrease in the mtDNA copy number was associated with poor patient prognosis in breast cancer (157) and increase in the mtDNA copy number was associated with higher grade tumours in ovarian cancer (148).

There are a limited number of reports, often with conflicting results regarding the prognostic significance of the mtDNA copy number change in colorectal cancer. For example, Chen and his co-authors investigated the mtDNA copy number change (in tumour versus non tumour cells) in 104 colorectal cancer patients from China and found that the increased mtDNA copy number in patients (also harboring a 4977 bp mtDNA deletion) was associated with advanced stage and high risk of metastasis in colorectal cancer patients (112). On the other hand, Lin and his co-authors investigated 153 colorectal cancer patients from Taiwan and found that decreased mtDNA copy number (in tumour versus non tumour cells) was associated with advanced stage and poor differentiation (118). In the same study, it has also been reported that the increase in the

mtDNA copy number was associated with longer OS. Similarly, Chang and his co-authors investigated the mtDNA copy number change in 194 colorectal cancer patients from Taiwan and found that the decrease in the mtDNA copy number change in tumour tissues with respect to normal tissues was associated with worse outcome (158).

In this thesis project, no association between the mtDNA copy number change and prognosis in colorectal cancer (in either the OS or the DFS analysis) was detected (**Tables 14** and **15**). Likewise, no association between mtDNA copy number change and patient clinicopathological features, including the disease stage, was detected (**Table 16**). Therefore, the results obtained in this project are different from the results published previously (112,118,158), which may be attributed to reported associations being false-positive associations, different definitions and categorizations of mtDNA copy number change in different studies, or the ethnic or other differences between these cohorts and the cohort investigated in this study. In addition, positive results are more preferred to be published over negative results and hence due to this selective reporting, there is a possibility that other studies similar to the finding reported in this study might not have been published.

To our knowledge, this thesis study is the first to be conducted on Caucasian colorectal cancer patients looking at the association of mtDNA copy number change with prognosis. In addition, the cohort investigated in this study is larger than other studies described above (112,118,158), and so likely to produce more reliable results. Furthermore, there were no significant differences between the cohort investigated for the mtDNA copy number change in this study and the entire NFCCR cohort (**Table 3**). Thus,

patients analyzed in the mtDNA copy number analysis in this thesis represent the entire NFCCR cohort, which is a population based cohort.

In conclusion, although it is possible that associations of these genetic variations with prognosis may be detected in larger patient cohorts, the results presented in this thesis do not support a role for mtDNA SNPs (mtDNA 10398 (A/G), mtDNA 16189 (T/C), MitoT479C, MitoT491C, MitoA13781G and MitoT10035C) and the mtDNA copy number change in prognosis of colorectal cancer patients.

CHAPTER 3: POLYMORPHISMS IN THE HYPOXIA PATHWAY GENES AND THEIR RELATION TO PROGNOSIS IN COLORECTAL CANCER PATIENTS

3.1) Introduction to hypoxia and its role in cancer

Hypoxia

The normal oxygen (O_2) concentration fluctuates in tissues between 2-9%, which is called normoxic condition. Hypoxia (<2% oxygen concentration) frequently occurs in solid tumours due to restricted blood flow. Subsequent cellular response to hypoxia is facilitated by Hypoxia Inducible Factors (HIFs).

Hypoxia Inducible Factors (HIFs)

HIFs are heterodimeric transcription factors consisting of α and β subunits (159). In mammals, α and β subunits dimerize together to form a transcriptionally active HIF. *HIF1A, HIF1B, HIF2A, HIF2B,* and *HIF3A* genes codes for these subunits. Transcriptional regulation by HIFs is accomplished by binding of HIFs to the hypoxia responsive elements (HREs) located near the promoter region of target genes. These target genes include genes that control functions such as angiogenesis, metastasis, oxygen and glucose regulation, cell motility, glycolysis and cell adhesion. Hypoxia-regulated genes are thus implicated in cancer progression and include genes such as *MIF* (160,161), *CXCL12* (162,163) and *LOX* (164).

Hypoxia and its role in cancer

Hypoxic conditions promote cancer progression and invasiveness in solid tumours. When human embryo cells grown *in vitro* under both extreme hypoxia and normoxic conditions were exposed to radiation, cells grown under hypoxic conditions show a 2.5 fold increase in colony forming ability compared to cells grown under normoxic conditions (165). It has also been shown that hypoxia confers stem cell like properties to cancer cells and also increases the colony forming ability (1). Since embryonic cells share a similarity with cancer cells in terms of rapid proliferation, cancer cells exposed to hypoxic conditions may have an aggressive tumour phenotype and have pronounced invasiveness (165). In addition, HIFs are over-expressed in various types of cancers and are also implicated in metastasis and tumour invasion (166). Hypoxic conditions have also been reported to increase the metastatic potential of tumour cells (167) and to stimulate the production of stress proteins, which in turn have been reported to be responsible for abnormal cell growth and metabolism (168) and promote tumour progression through angiogenesis (169). These findings suggest an inherent relation between hypoxia and cancer progression.

It has also been shown that solid tumours characterised by hypoxic conditions show resistance to anti-cancer therapies. In one study, the damage caused by exposure to radiation under well-oxygenated conditions was found to be three times greater than the damage caused by radiation under hypoxic conditions (170,171). Under normoxic conditions when cells are irradiated, free radicals are produced and these free radicals are responsible for causing cellular or DNA damage. But under hypoxic conditions, due to

low oxygen concentrations, an insufficient level of free radicals is produced and this may confer the hypoxic cells resistance to radiation (170,171). In addition, some chemotherapeutic agents require sufficient oxygen concentrations for their cytotoxic activity and therefore cells that are characterized by hypoxic conditions are found to be resistant to certain commonly used chemotherapeutic drugs (172). In summary, hypoxia is one of the factors that drives the selection of metastatic and treatment-resistant tumour phenotypes and maybe a marker for poor survival in cancer patients (173).

3.2) Role of hypoxia pathway genes in cancer progression

HIF1A

There are various reports investigating the potential role of *HIF1A* in facilitating tumour progression under hypoxic conditions (174-177), albeit with conflicting results. In some studies it has been reported that increased *HIF1A* expression is correlated with increased metastatic capacity, tumour growth, invasion and survival, and angiogenesis (174-177). On the other hand, other studies reported that *HIF1A* expression or its increased expression was correlated with tumour inhibition by either inducing programmed cell death or cell cycle arrest (178-181). Therefore, the functional significance of *HIF1A* expression in relation to clinical outcome in cancer is currently not well understood.

The role of *HIF1A* expression in clinical outcome in colorectal cancer is also not well established. For example, increased expression of *HIF1A* has been shown to be

correlated with poor prognosis, vascular invasion and TNM stage in rectal cancer patients (182). In another study, *HIF1A* over expression was found to be significantly associated with worse prognosis after adjusting for other variables such as MSI, Line 1 hypomethylation, *BRAF* and *PIK3CA* somatic mutations, and CpG island methylator phenotype (183). Some other studies also have reported *HIF1A* gene expression as an independent prognostic marker in colorectal cancer (184,185). However, these findings were not replicated in other studies (186,187).

So far there have been only a few studies that tested the association of genetic polymorphisms in *HIF1A* and their relation to outcome. Szkandera and his co-authors investigated the *HIF1A* C1772T (rs11549465) and G1790A (rs11549467) polymorphisms in a Caucasian colorectal cancer cohort (n=336) and found no association between the polymorphisms and patient prognosis (105). Previously, individuals carrying the variant allele for the *HIF1A* C1772T and *HIF1A* G1790A polymorphisms were shown to have higher transcriptional activity of *HIF1A* gene when compared to persons carrying the wild type alleles (188). Lee and co-authors investigated the same polymorphisms in relation to outcome in 445 Korean colorectal cancer patients and obtained similar results; these two SNPs were not associated with outcome (189). However, in another study, the C1772T polymorphism was found to be significantly associated with increased progression-free survival (PFS) in a univariate analysis (190).

ARNT (HIF1B)

Aryl Hydrocarbon Receptor Nuclear Translocator (ARNT), otherwise known as HIF1B, is a regulatory subunit of HIF1 (107). It is a basic-Helix-Loop-Helix (bHLH)

protein that controls expression of several genes by forming heterodimeric transcriptional factors with other bHLH proteins (191). Unlike HIF1A, which is degraded by ubiquitin mediated protein degradation under hypoxic conditions, HIF1B is expressed and is active under both hypoxic and normoxic conditions (192,193). Dimerization of this subunit with the α subunit of HIF1 (HIF1A) or the α subunit of HIF2 (HIF2A) is necessary for the regulation of genes involved in the hypoxia response pathway (159,194,195).

There are several reports about the role of *HIF1B* in tumourigenesis. In a study, HIF1B down-regulation (mediated by estrogen receptor beta or IFN-δ) resulted in HIF1 repression through disrupting the formation of HIF1A-HIF1B complex (196,197). It has also been reported that HIF1B recruits the co-activators necessary for HIF transcriptional complex formation and HIF transactivation (198,199). In addition, epidermal growth factor (EGF) activates the HIF1B signalling pathway, which in turn up-regulates cyclo-oxygenase gene expression. That, in turn, has been implicated in squamous cell carcinoma formation (200). In another study, blocking HIF1B under normoxic conditions has been reported to inhibit P21 (WAF1/CIP1) (201), which is a tumour suppressor protein (202). It has also been reported that HIF1B plays a significant role in tumour growth and progression (203). These findings illustrate the functional significance of *HIF1B* in cancer.

So far, there is only one study that has investigated a polymorphism (rs2228099 G>C) in the *HIF1B* gene in relation to outcome in colorectal cancer. This polymorphism was investigated in relation to OS and PFS in a small colorectal cancer cohort with mixed ethnicities, but no association was observed (190). Thus, the role of *HIF1B*

polymorphisms and their relation to prognosis in colorectal cancer remains unclear due to insufficient data.

EPAS1 (HIF2A)

EPAS1 (HIF2A) is a PAS (Period/ARNT/Single-minded) domain protein and one of the heterodimeric subunits of the HIF2 subclass of the HIF protein family. There are conflicting reports about *HIF2A* and its relation to cancer progression and clinical outcome. A number of knock-out and expression studies showed that *HIF2A* expression favours cancer progression (204-206), but other reports showed that *HIF2A* expression favours tumour inhibition (207-209). There are also conflicting reports on the role of HIF2A on prognosis. For example, in rectal cancer patients, *HIF2A* was reported to be expressed constitutively, but its expression was not correlated with patient survival (182). Similarly, in a study performed on 731 colorectal cancer patients, *HIF2A* expression was not found to be significantly associated with the patient outcome (183). On the other hand, in another study investigating colorectal cancer patients, *HIF2A* expression was found to be statistically significantly associated with poor prognosis (210).

To our knowledge, previously SNPs in this gene have not been investigated in relation to prognosis in colorectal cancer patients.

ARNT2 (HIF2B)

Similar to HIF1B, HIF2B belongs to the basic helix–loop–helix (bHLH) PAS domain protein family. It dimerizes with HIF2A and results in the formation of transcriptionally active HIF2. Similar to HIF1B, HIF2B has also been implicated in xenobiotic metabolism (211). There is one functional study investigating the relation between HIF2B and cancer. In this study, it was reported that HIF2B plays a role in tumour angiogenesis and hypoxia response in neurons (212). However, to date, there are no functional or prognostic studies investigating the relationship between *HIF2B* and colorectal cancer.

HIF3A

HIF3A encodes for the alpha heterodimeric subunit of HIF3. It has been shown in functional studies performed in embryonic stem cells that *HIF3A* up-regulates the *HIF2A* expression and blocks the expression of *HIF1A.* Therefore, it is evident that *HIF3A* act as a control unit in regulating the hypoxia-response in cells. To our knowledge, as of October 2012, there were no published studies investigating the prognostic significance of *HIF3A* SNPs in colorectal cancer patients.

MIF

Macrophage migration inhibitory factor *(MIF)* gene encodes for a T-cell derived polypeptide, which modifies the mobility of the macrophages in response to biological and physiological factors (213,214). It is secreted from the anterior pituitary and its main function is to mitigate the immunosuppressive effects induced by glucocorticoids (215,216). In the mid-60s, MIF was reported to have macrophage activation functions

such as adherence, phagocytosis, tumouricidial activity, and phage activation (217-220). But recently, its possible role in cell proliferation, survival, angiogenesis and response to hypoxia has started to emerge. For example, in a study performed on colorectal cancer cell lines that looked at the effect of hypoxia on MIF function, it was found that MIF is influential in triggering apoptosis of cells that are exposed to hypoxic conditions (221). The same study also indicated that MIF had a direct effect on the survival of the colorectal cancer cell lines under hypoxic conditions (221). Therefore, the authors suggested that *MIF* expression may be used as a marker for anti-angiogenic therapies or therapies that rely on hypoxia-induced apoptosis (221).

MIF is one of the hypoxia regulated genes. A HRE is located in the 5' end of *MIF* gene and *MIF* gene is over-expressed under hypoxic conditions (160,222). HIF1A activation of *MIF* results in premature cell senescence (223) and MIF is needed for stabilization of HIF1A under hypoxic conditions (224). It also regulates the *VEGF* expression under both hypoxic (224) and normoxic conditions, and mediates inhibition of *TP53* tumour suppressor gene at chronic inflammatory sites (225,226). It has also been shown that MIF is an important factor required for endothelial cell proliferation initiated by tumour growth and this has a profound effect on the tumour size. For example, chronic inflammation favours tumour growth (227). Therefore, it has been hypothesized that there is a link between the *MIF* expression, chronic inflammation condition and cancer development (228). Expression profile studies have also revealed that MIF is over-expressed in metastatic tumours when compared to normal tissue (229) and it is associated with aggressive tumours and tumours that have a higher potential to

metastasize (230-232). Additionally, other studies showed that MIF is a pro-angiogenic factor and suppressing *MIF* expression using a MIF inhibitory antibody leads to suppressed tumour growth and suppressed tumour associated angiogenesis (233). *MIF* expression also activates other pathways and induces the production of proteins that promote tumour growth, thereby proving its role as a tumour promoter (161,234). In summary, literature findings indicate that the *MIF* expression is regulated by hypoxia and it may promote tumour progression.

There have been conflicting reports about the *MIF* gene and its relation to patient prognosis in cancer. Both increased (230,235,236) and decreased (231,237) expression of *MIF* gene have been reported to be associated with poor outcome in various different types of cancers. In colorectal cancer, one study found that *MIF* was expressed in higher amounts in colorectal cancer when compared to malignant melanoma (238).

Currently, there is no study published investigating the relationship between SNPs in the *MIF* gene and prognosis in colorectal cancer. Therefore, additional studies are needed to elucidate the genetic variations in this gene and their potential role in prognosis.

CXCL12

CXCL12 is a 8 kilo Dalton CXC chemokine family member (239). CXCL12 is primarily involved in immunological processes, but its various pleiotropic effects have also been reported, such as chemotaxis, hematopoiesis and neo-vascularization (240-242).

CXCL12 has two HIF1 binding sites and its expression is upregulated under hypoxic conditions (162).

Although many pleiotropic effects that have been documented, *CXCL12*'s main function is to trigger ankiosis, which is a form of programmed cell death caused by the loss of anchorage of cells to the extracellular matrix. Most of the metastatic cancers are characterized by presence of ankiosis. There have also been reports that CXCL12 reduces the capacity of cells to metastasize by activating ankiosis. In a study performed on mammary carcinoma cells, epigenetic silencing of *CXCL12* resulted in increased metastatic capacity (243). Also, in a study performed on colorectal carcinoma cells, *CXCL12* expression has been found to induce apoptosis (244). In addition, high expression of *CXCL12* was found to be associated with poor prognosis when compared to its low expression; and patients with high *CXCL12* expression and high grade tumours were reported to have the worst outcome in colorectal cancer (245). It has also been shown that *CXCL12* expression is correlated with tumour cell proliferation and migration in colorectal cancer (246). These finding suggest a role of CXCL12 in tumour progression.

Contrary to these reports, down-regulation of *CXCL12* expression by hypermethylation has also been shown to promote metastasis of colorectal cancer carcinomas (247). A similar finding was reported in another study where the authors have showed that *CXCL12* prevents cells from metastasizing and thus, maintaining *CXCL12* expression may reduce tumour progression in colorectal cancer (244).

Six SNPs in *CXCL12* have been investigated so far in relation to outcome in colorectal cancer patients. In one study performed in a Caucasian cohort (n=328), a G/A polymorphism in the 3'-UTR of *CXCL12* (rs1801157) was not found to be associated with DFS and other prognostic variables such as stage, sex, age of onset, histological grade or tumour size (248). In another study performed with a Asian cohort (n=424), a *CXCL12* SNP (C14478T; rs1065297) was not found to be associated with the lymph node metastasis (106). In the same study, the authors also reported that the A allele of another SNP (G801A, rs1801157) was correlated with higher *CXCL12* mRNA levels and the GG genotype was found to be associated with better DFS in patients without the lymph node metastasis, but not in patient with the lymph node metastasis (106). In the same study, four other *CXCL12* SNPs, namely G8906A (rs2297630), G6201A (rs266085) G5887A (rs2839693) and G5753A (rs754618) were not found to be associated with DFS (106).

LOX

The Lysyl oxidase gene (*LOX*) is located on 5q23.1 and encodes the lysyl oxidase enzyme which is a copper dependent amine-oxidase (249). It is responsible for the deamination of peptidyl lysine residues to form α-aminoadipic δ-semialdehydes. These α-aminoadipic δ-semialdehydes undergo a condensation reaction with other aldehydes or other amino acids resulting in a cross-linking between collagen and elastin. This cross-linking plays a crucial role in maintaining the structural integrity of the connective tissues (164,250). *LOX* has a HRE in its promoter region and interaction of this HRE with HIF1 results in increased expression of the *LOX* gene (251). In a microarray differential expression analysis, increased expression of *LOX* has been observed in tumour cells

under hypoxic conditions, (251); suggesting that LOX expression is regulated by oxygen concentration.

The *LOX* gene has roles in several processes such as tumourigenesis, chemotaxis and cell aging. In one study blocking the expression of the *LOX* showed less cell proliferation in agar medium, probably because of the inability of cells to communicate with each other resulting in inhibition of tumour formation (252,253). Loss of LOX function during tumour development as a result of somatic mutations has been shown to play a role in colon cancer pathogenesis. Thus, *LOX* gene may have a tumour suppressor role (254).

A few studies have examined the over expression of the *LOX* in relation to prognosis in colorectal cancer. For example, increased expression of *LOX* has been observed in colon cancer cells (255). Increased expression of the *LOX* has also been reported in other tumours (256-258). To our knowledge, SNPs in *LOX* have not been investigated in relation to prognosis in colorectal cancer.

3.3) Rationale, hypothesis and objectives specific to Chapter 3

Hypoxic conditions promote angiogenesis, resistance to conventional anti-cancer therapy, aggressive tumour phenotype and increased metastatic potential in solid tumours. Therefore, it was hypothesized that polymorphisms in hypoxia pathway genes may be associated with outcome of colorectal cancer patients.

i. The main objective of this project is to test the associations of genetic variants in the hypoxia pathway genes with outcome in colorectal cancer patients.

3.4) Contributions and credits

Asan M. S. Haja Mohideen: organized and coded the genotype data prior to statistical analysis for the polymorphisms in phase I of this project; performed the statistical analyses for phase I and phase II polymorphisms; and interpreted the results.

Dr. Sevtap Savas: coded the pilot study and the NFCCR SNP genotyping cohort clinicopathological and prognostic data; genotyped the *HIF1B*-rs10847 using the TaqMan® SNP genotyping method; provided the **Table 19**; provided the baseline table for the pilot study samples (**Table 17**); constructed the LD maps; selected the tag SNPs in phases I and II and coded the SNP genotypes in phase II for statistical analysis.

Jessica Squires: Genotyped 13 polymorphisms in phase I using the TaqMan® SNP genotyping method and compiled their genotype data.

Dr. Angela Hyde: Provided the clinicopathological and prognostic data for the pilot study samples (phase I).

Dr. Roger Green and Dr. Patrick Parfey: Provided the clinicopathological and prognostic data for the NFCCR cohort; provided the DNA samples of pilot study samples; provided the genotype data of the patient cohort in phase II.

3.5) Materials and methods

Ethics approval for this project was obtained from HIC (HIC #: 11.102).

3.5.1) Project phases and genes and polymorphisms studied

In this project (hypoxia pathway SNPs project) a number of genes and their genetic variations were investigated in relation to prognosis in two separate colorectal cancer patient cohorts in two project phases. A summary of these phases are shown in **Figure 12.**

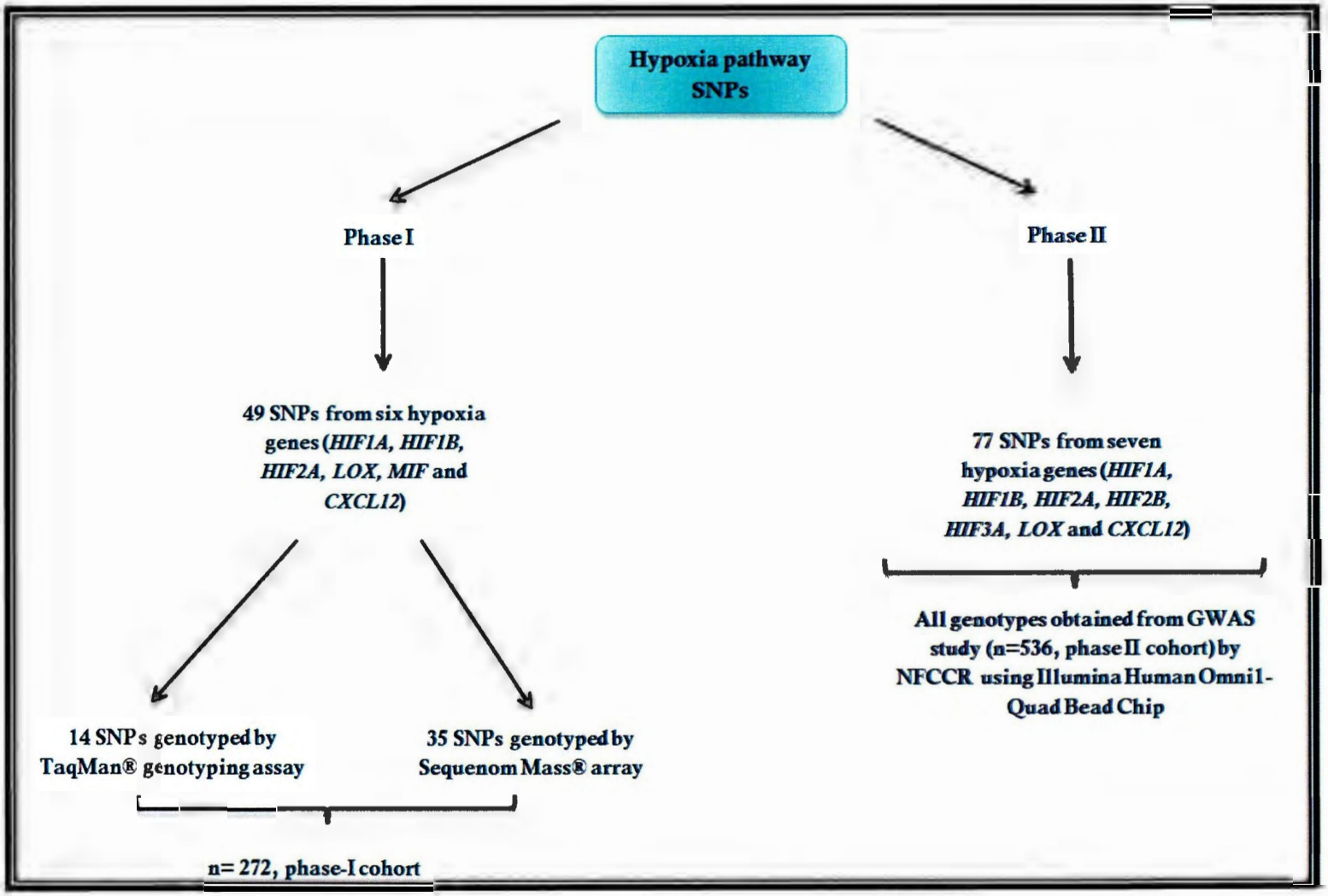
Among the several genes involved in the hypoxia pathway, a number of candidate genes were prioritized in this study based on literature evidence showing their biological roles under hypoxic conditions or in regulating the response to hypoxic conditions (**Section 3.2**). Specifically, this project investigated five HIF-coding genes (*HIF1A*, *HIF1B*, *HIF2A*, *HIF2B*, and *HIF3A*) and three biologically well-characterized genes regulated by HIFs (*CXCL12*, *MIF*, and *LOX*).

In order to avoid investigation of highly correlated SNPs in the same analysis, the genotype data for polymorphisms in selected genes was extracted from the HapMap

database (HapMap Data Rel24/Phase II Nov08, on NCBI B36 assembly, dbSNP B126) for Caucasian samples (259). SNPs with a MAF >10% were used to calculate the correlation coefficients (r^2; minimum 0.8 for highly correlated SNPs) between SNPs and to select tagging SNPs (tagSNPs) using the pair wise tagger function in the Haploview program (ver 4.2) (260). r^2 values were used to remove the highly correlated SNPs from analysis to reduce data redundancy.

Overall, in phase I of this study, 49 SNPs in six prioritized genes (*HIF1A, HIF1B, HIF2A, LOX, MIF* and *CXCL12*) were investigated in 272 colorectal cancer patients (a patient cohort called the pilot project study cohort (**Section 3.5.2**)). In phase II of this project, 77 SNPs from five genes (*HIF1A, HIF1B, HIF2A, LOX* and *CXCL12)* included in phase I and two additional genes (*HIF2B* and *HIF3A*) were investigated in 536 colorectal cancer patients (the NFCCR SNP genotyping cohort; **Section 2.4.3**).

Figure 12: Flowchart of the hypoxia pathway project: phase I and phase II



3.5.2) Patient cohorts

This project involves investigation of two separate colorectal cancer patient cohorts from NL.

Patient cohort studied in phase I (the pilot project study samples)

In the first phase of this project, we have analyzed a colorectal cancer cohort, which is termed "the pilot project study cohort", by our team. These patients (n=280) were all from the Avalon Peninsula in Newfoundland, and were diagnosed with primary colorectal cancer during 1997-1998. These patients were recruited based on the following criteria (Dr. Angela Hyde, *personal communication*):

1) Patients diagnosed with carcinoma in the polyp were included provided that there was invasion of carcinoma into the stalk.
2) Patients with primary carcinoma of the colon or rectum were included.
3) Patients were included irrespective of their age at diagnosis.
4) Patients with carcinoma in situ, mucosal carcinoma, polyps without invasive carcinoma, carcinoid tumours and tumours arising due to FAP were excluded.

The baseline characteristics of patients included in the pilot project study are shown in **Table 17**. DNA samples from 272 patients were available for the genotype analysis.

Clinicopathological and prognostic data for this cohort were collected using the medical records by Dr. Angela Hyde. The last date of patient follow up was July 2009.

Table 17: Baseline and prognostic characteristics of the pilot project study samples (phase I).

Variables	n	%
Sex		
Male	150	53.6
Female	130	46.4
Median age at diagnosis	68.42 years range (25.29-91.61)	
Grade		
poorly differentiated/undifferentiated	42	15
well/moderately differentiated	234	83.6
Unknown	4	1.4
Histology		
Mucinous	43	15.4
non-mucinous	237	84.6
Location		
Rectum	57	20.4
Colon	223	79.6
Lymphatic invasion of tumor		
Lymphatic invasion (+)	110	39.3
Lymphatic invasion (-)	70	25
Unknown	100	35.7
Stage		
I	54	19.3
II	94	33.6
III	76	27.1
IV	47	16.8
Unknown	9	3.2
MSI-H status		
Yes	34	12.1

No	246	87.9
Prognostic and follow up information		
OS status at the time of last follow up		
Dead	172	61.4
Alive	108	38.6
Unknown	0	0
Median OS and DSS (follow up) time	5.31 years range (0-12.52 years)	
DFS status at the time of last follow up		
recurrence/metastasis/death (+)	184	65.7
recurrence/metastasis/death (-)	96	34.3
Unknown	0	0
DFS (follow up) time	3.37 years range (0-12.52 years)	
DSS status at the time of last follow up		
death from colorectal cancer	113	40.4
death from other causes/alive	167	59.6
Unknown	0	0

(+): presence, (-): absence, DFS: Disease Free Survival, DSS: Disease Specific Survival, MSI-H: Microsatellite Instability-High, MSI-L: Microsatellite instability-Low, MSS: Microsatellite Stable, n: number, OS: Overall Survival.

Patient cohort studied in phase II (the NFCCR SNP genotyping cohort)

The patient cohort investigated in phase II of the hypoxia project is almost identical to the NFCCR SNP genotyping cohort described in **Section 2.4.3** (**Table 3**), except for the fact that there was one fewer patient in the cohort investigated in phase II (total n=536).

3.5.3) Genotyping of DNA samples extracted from blood

Phase I

In phase I, 35 SNPs were genotyped using the Sequenom MassArray® technique at an outsourced facility (Analytical Genetics Technology Centre, Toronto, Canada) and 14 other SNPs were genotyped using the TaqMan® SNP genotyping technique. The TaqMan® SNP genotyping methodology is described in **Section 2.6.1.** (the reaction conditions were the same except for the cycle number which was 40 in the case of these SNPs) and the assay IDs are shown in **Appendix 5**. Genotyping reactions were repeated two additional times for DNA samples that failed the TaqMan® SNP genotyping reactions in the first attempt. Overall, 49 of 53 selected SNPs were successfully genotyped and were included in phase I of this study (**Table 18**). Further information on the 53 polymorphisms is shown in **Appendix 6**.

Table 18: List of SNPs included in phase I.

SNP	Genotyping technique	*Concordance rate (%)	*Duplication rate (%)	Missing genotype (n)	Missing genotype (%)	Missing data >15%	MAF <10%	**Deviation from HWE
LOX rs2956540	MassArray	100	5.2	9	3.21			
LOX rs2288393	MassArray	100	5.2	9	3.21			
LOX rs10040971	MassArray	100	5.2	8	2.86			yes
LOX rs10519694	MassArray	100	5.2	8	2.86			
HIF1A rs2301106	MassArray	100	5.2	9	3.21			
HIF1A rs2301111	MassArray	100	5.2	8	2.86			
HIF1A rs2301113	MassArray	100	5.2	8	2.86			
HIF1A rs11158358	MassArray	100	5.2	8	2.86			
HIF1B rs2228099	MassArray	100	5.2	8	2.86			
HIF1B rs3738483	MassArray	100	5.2	8	2.86		yes	
HIF1B rs11204737	MassArray	100	5.2	9	3.21			
HIF1B rs10847	TaqMan®	100	10.6	19	6.79			yes
CXCL12 rs2236534	MassArray	100	5.2	9	3.21			yes
CXCL12 rs2839688	MassArray	100	5.2	16	5.71			
CXCL12 rs2236533	TaqMan®	100	7.6	8	2.86			yes
CXCL12 rs11592974	TaqMan®	100	37.83	8	2.86			yes
MIF rs2096525	MassArray	100	5.2	17	6.07			
HIF2A rs2121266	TaqMan®	100	8.76	29	10.36			
HIF2A rs2346175	TaqMan®	100	12.97	65	23.21	yes		
HIF2A rs3768730	TaqMan®	100	9.31	33	11.79			
HIF2A rs4952818	TaqMan®	100	8.8	30	10.71			
HIF2A rs4953340	TaqMan®	100	12.9	32	11.43			
HIF2A rs6753127	TaqMan®	100	8.33	28	10		yes	
HIF2A rs7583558	TaqMan®	100	11.98	38	13.57			
HIF2A rs9679290	TaqMan®	100	6.35	28	10			

HIF2A rs9973653	TaqMan®	100	13.22	38	13.57		yes
HIF2A rs10178633	TaqMan®	100	12.4	40	14.29		
HIF2A rs11687512	TaqMan®	100	11.6	30	10.71	yes	
HIF2A rs10199201	MassArray	100	5.2	30	10.71		
HIF2A rs11125070	MassArray	100	5.2	33	11.79		
HIF2A rs12614710	MassArray	100	5.2	32	11.43		
HIF2A rs13019414	MassArray	100	5.2	29	10.36		
HIF2A rs13412887	MassArray	100	5.2	31	11.07		
HIF2A rs1374748	MassArray	100	5.2	30	10.71		
HIF2A rs1562453	MassArray	100	5.2	32	11.43		
HIF2A rs1868084	MassArray	100	5.2	31	11.07		
HIF2A rs1868087	MassArray	100	5.2	30	10.71		
HIF2A rs1992846	MassArray	100	5.2	31	11.07		
HIF2A rs2034327	MassArray	100	5.2	37	13.21		
HIF2A rs2044456	MassArray	100	5.2	41	14.64		
HIF2A rs2346176	MassArray	100	5.2	29	10.36		
HIF2A rs3768728	MassArray	100	5.2	29	10.36		
HIF2A rs4145836	MassArray	100	5.2	29	10.36		yes
HIF2A rs4953344	MassArray	100	5.2	32	11.43		
HIF2A rs4953349	MassArray	100	5.2	35	12.5		
HIF2A rs4953353	MassArray	100	5.2	29	10.36		
HIF2A rs6706003	MassArray	100	5.2	30	10.71		
HIF2A rs6712143	MassArray	100	5.2	29	10.36		
HIF2A rs7583392	MassArray	100	5.2	29	10.36		

SNPs successfully genotyped, the genotyping technique applied, duplication and concordance rates, missing genotype information, minor allele frequencies, and Hardy-Weinberg equilibrium calculations(**See **Section 3.7** and *See **Section 2.7**)*in phase I of this project.*

HWE: Hardy Weinberg Equilibrium, MAF: Minor allele frequency.

Phase II

While phase I was in progress, NFCCR obtained the genotypes of a large number of SNPs (>1 million) using the Illumina Human Omni1-Quad Bead Chip genotyping platform at an outsourced genomics facility (Centrillion Genomic Services, USA) for NFCCR patients. Using this data, in phase II we aimed to have a more comprehensive analysis of the key genes functioning in the hypoxia pathway. Thus, in addition to the genes in phase I, the genotypes for genetic variations in two additional HIF-coding genes, namely *HIF2B* and *HIF3A* were also included. TagSNP and r^2 values based on HapMap Caucasian samples were used to select the SNPs in *HIF2B* and *HIF3A* as well. As a result, genotypes for a total of 125 SNPs were available in the genomewide SNP genotyping data. Among these SNPs, 44 SNPs with a highly correlated proxy SNP (r^2 >0.8) were excluded to avoid data redundancy, leaving 81 SNPs. Four additional SNPs were excluded from the statistical analysis based on their small MAF. Therefore, a total of 77 SNPs were investigated in phase II (**Table 19**). These SNPs included in this study together with their proxy SNPs and r^2 values are shown in **Appendix 8** and the list of SNPs excluded is shown in **Appendix 9**.

Table 19: List of SNPs included in phase II.

Gene	SNP ID	Major homozygote (n)	Heterozygote (n)	Minor homozygote (n)	Missing genotype (n)	Total genotyped patients (n)	MAF
HIF1A	rs2301106	TT=405	TC=118	CC=13	0	536	0.13
HIF1A	rs12434438	AA=329	AG=176	GG=31	0	536	0.22
HIF1A	rs2057482	CC=391	CT=132	TT=13	0	536	0.15
HIF1B	rs10847	GG=281	AG=221	AA=34	0	536	0.27
HIF1B	rs2228099	GG=229	GC=245	CC=62	0	536	0.34
HIF1B	rs11204737	TT=186	TC=260	CC=90	0	536	0.41
HIF2A	rs1867785	GG=212	AG=259	AA=65	0	536	0.36
HIF2A	rs2121266	AA=203	AC=262	CC=70	1	535	0.38
HIF2A	rs17034950	GG=289	AG=218	AA=29	0	536	0.26
HIF2A	rs9973653	GG=279	TG=214	TT=43	0	536	0.28
HIF2A	rs4953342	AA=272	AG=208	GG=56	0	536	0.3
HIF2A	rs1868089	TT=139	TC=268	CC=129	0	536	0.49
HIF2A	rs6758592	TT=167	TC=263	CC=106	0	536	0.44
HIF2A	rs12614710	GG=149	TG=269	TT=118	0	536	0.47
HIF2A	rs4953352	TT=140	TC=267	CC=129	0	536	0.49
HIF2A	rs4953353	GG=230	TG=247	TT=59	0	536	0.34
HIF2A	rs2346175	TT=159	TC=264	CC=113	0	536	0.46
HIF2A	rs6756667	AA=144	AG=266	GG=126	0	536	0.48
HIF2A	rs1868086	GG=343	GT=171	TT=21	1	535	0.2
HIF2A	rs6712143	AA=292	AG=203	GG=41	0	536	0.27
HIF2A	rs10176396	CC=253	CT=228	TT=55	0	536	0.32
HIF2A	rs1374748	GG=407	GT=120	TT=9	0	536	0.13
HIF2A	rs3768728	TT=402	TC=123	CC=11	0	536	0.14

HIF2A	rs2346176	TT=207	TC=252	CC=77	0	536	0.38
HIF2A	rs10178633	GG=173	GA=257	AA=106	0	536	0.44
HIF2A	rs7594912	AA=165	AC=260	CC=111	0	536	0.45
HIF2A	rs7557402	GG=161	GC=251	CC=124	0	536	0.47
HIF2A	rs7571218	GG=205	GA=259	AA=72	0	536	0.38
HIF2B	rs12591286	GG=216	AG=260	AA=60	0	536	0.35
HIF2B	*rs8041826	AA=376	AG=138	GG=22	0	536	0.17
HIF2B	rs1446337	GG=306	GA=193	AA=37	0	536	0.25
HIF2B	rs12593988	GG=350	GA=164	AA=22	0	536	0.19
HIF2B	rs3848206	AA=132	GA=274	GG=129	1	535	0.5
HIF2B	rs3848207	GG=405	GA=117	AA=14	0	536	0.14
HIF2B	*rs7172914	CC=343	CT=160	TT=33	0	336	0.34
HIF2B	rs10431813	GG=291	AG=205	AA=38	2	534	0.26
HIF2B	rs3910982	GG=262	GT=220	TT=54	0	536	0.31
HIF2B	rs11635014	CC=187	TC=255	TT=94	0	536	0.41
HIF2B	*rs1020398	TT=278	TC=231	CC=27	0	536	0.27
HIF2B	*rs11633642	GG=246	GA=219	AA=71	0	536	0.34
HIF2B	rs7184010	CC=291	CT=212	TT=32	1	535	0.26
HIF2B	rs3848170	CC=426	CT=101	TT=9	0	536	0.11
HIF2B	rs4778791	GG=141	AG=276	AA=119	0	536	0.48
HIF2B	rs8034535	AA=388	AG=137	GG=11	0	536	0.15
HIF2B	rs895442	CC=429	CT=99	TT=8	0	536	0.11
HIF2B	rs1037124	GG=383	GA=138	AA=15	0	536	0.16
HIF2B	rs1374213	TT=160	TC=275	CC=100	1	535	0.44
HIF2B	rs3901896	CC=184	TC=274	TT=78	0	536	0.4
HIF2B	rs2278709	CC=270	CT=217	TT=49	0	536	0.29
HIF2B	rs8028295	CC=164	CT=264	TT=108	0	536	0.45
HIF2B	rs4609803	GG=294	GA=199	AA=43	0	536	0.27
HIF2B	rs4778800	GG=325	GT=184	TT=27	0	536	0.22

HIF2B	rs7178902	TT=173	TC=257	CC=106	0	536	0.44
HIF2B	rs4238521	GG=382	GA=143	AA=11	0	536	0.15
HIF2B	rs4331301	GG=208	AG=245	AA=83	0	536	0.38
HIF2B	*rs4778600	GG=392	GT=124	TT=19	1	535	0.15
HIF2B	rs11856676	CC=162	TC=259	TT=115	0	536	0.46
HIF2B	rs4238522	TT=197	TC=250	CC=89	0	536	0.4
HIF2B	rs4074666	CC=146	TC=276	TT=114	0	536	0.47
HIF2B	rs11635554	AA=283	AG=210	GG=43	0	536	0.28
HIF2B	rs4778818	AA=296	AG=205	GG=35	0	536	0.26
HIF2B	rs4778819	CC=275	TC=210	TT=51	0	536	0.29
HIF2B	rs7403706	TT=347	TC=167	CC=22	0	536	0.2
HIF2B	rs6495509	GG=332	GA=181	AA=23	0	536	0.21
HIF2B	rs8039725	AA=349	AG=165	GG=22	0	536	0.19
HIF2B	*rs8033706	TT=141	TC=207	CC=132	56	480	0.49
HIF2B	rs4301984	GG=328	GA=182	AA=26	0	536	0.22
HIF2B	rs4459508	GG=279	GA=212	AA=45	0	536	0.28
HIF3A	rs2072491	CC=424	CT=104	TT=8	0	536	0.11
HIF3A	rs757638	GG=388	GA=133	AA=15	0	536	0.15
HIF3A	*rs12461322	GG=427	AG=97	AA=12	0	536	0.11
HIF3A	rs887946	AA=255	AG=224	GG=57	0	536	0.32
HIF3A	rs3764610	CC=374	TC=140	TT=22	0	536	0.17
HIF3A	*rs11665853	AA=430	AG=95	GG=11	0	536	0.11
LOX	rs3792802	GG=366	GA=160	AA=9	1	535	0.17
LOX	rs1800449	GG=364	GA=157	AA=15	0	536	0.17
CXCL12	rs2839695	TT=345	TC=170	CC=21	0	536	0.2

*MAF: minor allele frequency. * genotypes of polymorphisms deviated from HWE.*

3.5.4) Quality control measures and statistical analyses

The genotype replication rates (see description in **Section 2.7**) for the 49 SNPs included in phase I of this project are shown in **Table 18**. The genotypes of at least 5.2% of the DNA samples were successfully replicated for each SNP. All replicated genotypes yielded identical genotypes. Replicated genotype data was not available for phase II SNPs.

Calculation of deviations from the Hardy Weinberg Equilibrium (HWE)

The Hardy-Weinberg principle states that both genotype and allele frequencies remain constant in a population unless the population is subjected to external influences such as migration into and out of the population, gene flow, genetic drift or selection (261). Therefore, deviation from the HWE implies either one or a combination of these external influences. Since the Newfoundland population is characterized by founder effects and genetic isolation, HWE may not be applicable to our cohorts. Deviation from HWE may also occur due to genotyping errors (261).

To ascertain whether the genotype frequencies of SNPs investigated in this study deviated from the HWE, the observed genotype frequencies were compared with the expected frequencies. This was checked using an online tool (262) (http://www.oege.org/software/hwe-mr-calc.shtml) and they were verified manually by computing in a Microsoft® Excel sheet. In brief, the genotypes were sorted in an Excel spread sheet into major homozygotes, minor homozygotes and heterozygotes. The major allele is designated as "A", minor allele is designated as "a" and their corresponding

allele frequencies are designated as p and q respectively. Minor and the major allele frequencies were computed using the equation p=1-q. According to the Hardy-Weinberg equation ($p^2+2pq+q^2$), genotype frequency of major homozygotes are designated as p^2 (AA), genotype frequency of minor heterozygotes are designated as 2pq (Aa) and genotype frequency of minor homozygotes (aa) are designated as q^2. The expected genotype frequencies are computed using these formulas:

Expected frequency f (AA) = p^2 n

Expected frequency f (Aa) =2 pqn

Expected frequency f (aa) = q^2 n

Where, n represents the number of genotypes that have been successfully genotyped for each polymorphism.

Once the expected frequencies are calculated, they are compared with observed genotype frequencies to see whether they deviate from the HWE using the Pearson Chi-square test. This equation is as follow:

$$x^2 = \sum (Observed - Expected)^2 / Expected$$

As evident from the equation above, the change in frequencies between observed and expected frequencies normalized to the expected frequencies were calculated for each of the three genotypes and summed up. For this analysis, the null hypothesis is that the

genotype frequencies are in HWE and the alternative hypothesis is that they are not in HWE. The resulting Chi-square values were then concurrently analyzed with the Chi-square table for 5% significance level and their respective degree of freedom in order to check whether it satisfies the null or alternative hypothesis. The degree of freedom is calculated using the formula:

Degree of freedom = number of genotypes - number of alleles

In this study, the number of genotypes is three and the number of alleles is two. Thus, the degree of freedom is one. For one degree of freedom, the 5% significance level for Chi-square value is 3.84. Therefore, when the Chi-square value is less than 3.84, the alternative hypothesis is rejected. When it is above 3.84, the null hypothesis is rejected. In other words, when the Chi-square value is less than 3.84 for a particular polymorphism it implies that it does not deviate from HWE, but if it is above 3.84 then it implies that it does deviated from HWE.

The HWE test results for the polymorphisms investigated in these projects are shown in **Table 18** for phase I and **Table 19** for phase II SNPs.

Genetic model used

In order to increase study power, we have analyzed the SNP genotypes assuming the dominant genetic model. This model assumes that patients with major homozygote genotypes (AA) have an increased (or decreased) risk when compared to heterozygous patients (Aa) and patients with minor homozygote genotypes (aa) [AA vs (Aa+aa)] (263).

Statistical analyses

The relationship between genotypes and outcome were analyzed using the statistical tests explained in **Section 2.7**. In addition to OS and DFS, in phase I, we also investigated the disease specific survival (DSS), which is the time interval from date of diagnosis to date of the last follow up or date of death from colorectal cancer. In the multivariate Cox regression analysis, variables that had a $p < 0.05$ in the univariate analysis were entered into the model. Additionally, the polymorphisms whose genotype data deviate from the HWE were investigated in the univariate analysis for exploratory purposes, but were excluded from the multivariate analyses.

3.6) Results

3.6.1) Phase I results

Description of the patient cohort

For the study samples (i.e. the pilot project study samples) investigated in phase I of this project, the median age at diagnosis was 68.4 years, the median OS and DSS (follow up) time was 5.3 years and the DSS median (follow up) time was 3.4 years. Other characteristics of this cohort are summarised in **Table 17**.

Overall survival analysis (phase I)

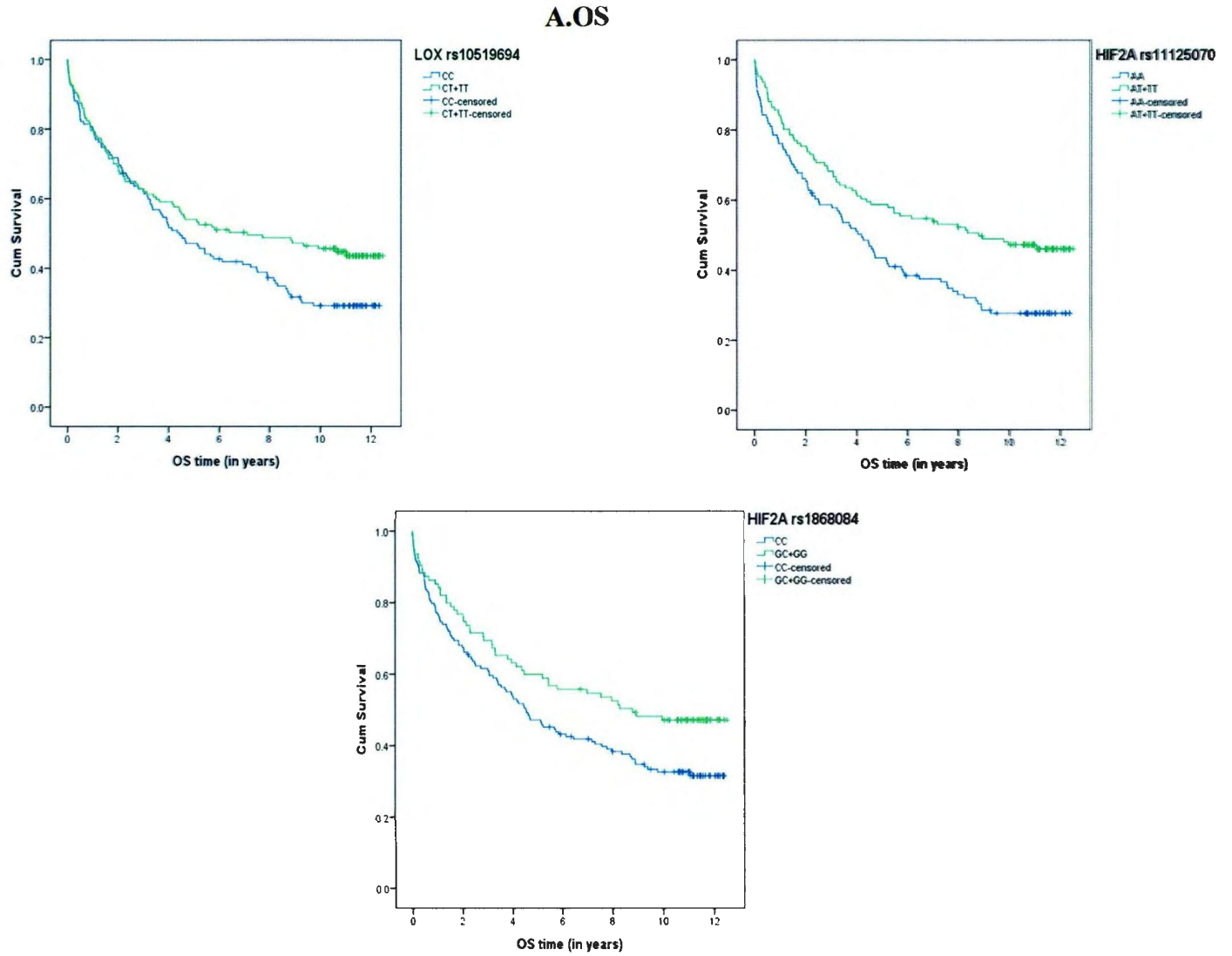
In univariate analysis, three of the 49 polymorphisms investigated (rs10519694 in *LOX*, rs11125070 and rs1868084 in *HIF2A*) were significantly associated with OS (**Table 20**). The genotype distributions of these three polymorphisms were in HWE. Their Kaplan Meier survival curves are shown in **Figure 13.A**. The univariate analysis results of other polymorphisms and clinicopathological features can be found in **Appendix 10**.

Table 20: Polymorphisms significantly associated with outcome in univariate analysis (phase I, overall survival)

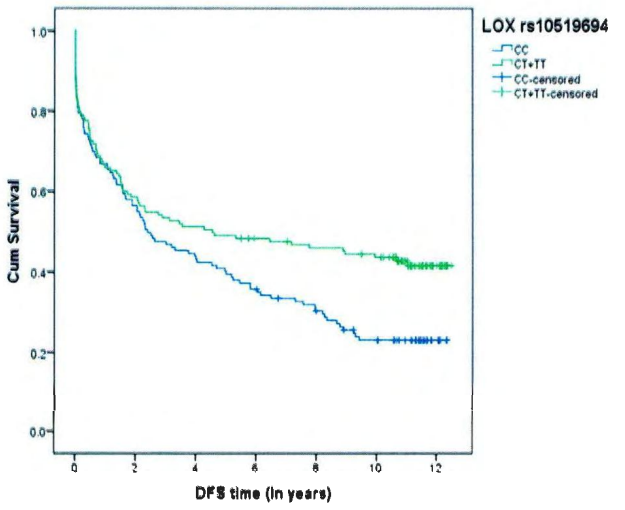
Variables	n	Number of patients in each category	p-value	HR	95% CI	
					Lower	Upper
***LOX* rs10519694 (CT +TT vs CC)**	272	(137 vs 135)	0.046	0.735	0.543	0.99
***HIF2A* rs11125070 (AT+TT vs AA)**	247	(126 vs 121)	0.003	0.616	0.447	0.85
***HIF2A* rs1868084 (GC+GG vs CC)**	249	(95 vs 154)	0.024	0.678	0.483	0.95

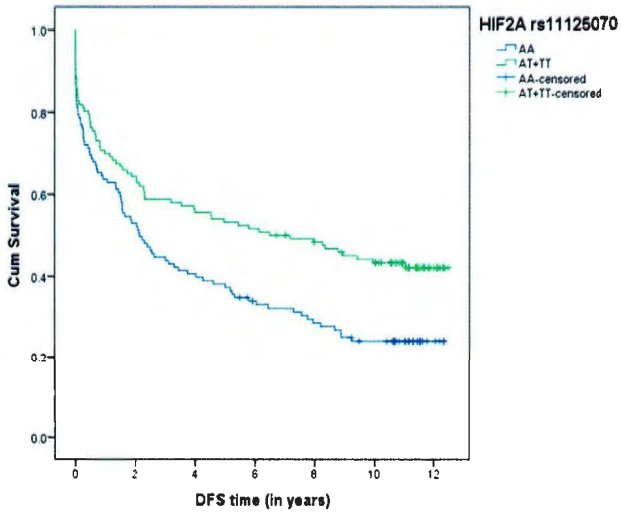
For the multivariate analysis, variables with >15% of missing data (i.e. the lymphatic invasion) were excluded. The results of the multivariate analysis demonstrated that while the increasing age, stage, grade, and MSS/MSI-L status were associated with shorter OS, the three polymorphisms that were significantly associated with OS in the univariate analysis had lost their statistical significance (**Table 21**).

Figure 13: Kaplan Meier survival plots for polymorphisms whose associations were significant in univariate analysis (phase I)



B. DFS





OS time is defined as the time interval from the time of diagnosis to the date of last follow up or the date of death. DFS time is defined as the time interval from the time of diagnosis to the date of last follow up or the date of recurrence or metastasis or death. Cumulative survival refers to the proportion of patients who are alive at a particular point of time. Survival plots were generated by Kaplan Meier method.

Table 21: Multivariable model for overall survival (phase I) (n=234)

Variables	Number of patients in each category	p-value	HR	95% CI	
				Lower	Upper
LOX rs10519694 (CT+TT vs CC)	(116 vs 118)	0.129	0.767	0.544	1.08
HIF2A rs11125070 (AT+TT vs AA)	(119 vs 115)	0.282	0.801	0.535	1.2
HIF2A rs1868084 (GC+GG vs CC)	(88 vs 146)	0.293	0.796	0.521	1.218
Age	-	**<0.001**	**1.045**	**1.029**	**1.06**
Grade (poorly differentiated/undifferentiated vs well/moderately differentiated)	(35 vs 199)	**<0.001**	**2.594**	**1.669**	**4.03**
Stage	45	**<0.001**			
Stage (II vs I)	(85 vs 45)	0.247	1.399	0.792	2.472
Stage (III vs I)	(65 vs 45)	**0.002**	**2.534**	**1.423**	**4.514**
Stage (IV vs I)	(39 vs 45)	**<0.001**	**14.4**	**7.617**	**27.213**
MSI-H status (Yes vs No)	(22vs 212)	**0.002**	**0.285**	**0.131**	**0.62**

CI: Confidence Interval, HR: Hazard Ratio, MSI-H: Microsatellite Instability-High, MSI-L: Microsatellite Instability-Low, MSS: Microsatellite Stable, n: number of samples entered into the multivariate model, vs: versus. Values that are bolded are statistically significant (p value <0.05).

Disease specific survival analysis (phase I)

In a univariate disease specific survival (DSS) analysis, none of the 49 polymorphisms in hypoxia related genes was significantly associated with the outcome. Univariate Cox regression analysis results for these 49 SNPs are shown in **Appendix 10**.

Disease free survival analysis (phase I)

In a univariate disease free survival (DFS) analysis (**Table 22**), two of the 49 polymorphisms investigated in this study were significantly associated with outcome. **Figure 13.B** shows the Kaplan Meier survival curves for these two SNPs. The univariate analysis results of other polymorphisms and clinicopathological features for DFS are shown in **Appendices 11 and 12**.

In a multivariate analysis of DFS (**Table 23**), the association of *HIF2A* polymorphism (rs11125070) remained statistically significant (p=0.004; HR=0.619; 95% CI: 0.446-0.859) when adjusted for other variables.

Table 22: Polymorphisms significantly associated with outcome in univariate (phase I) disease free survival analysis

Variables	n	Number of patients in each category	p-value	HR	95% CI	
					Lower	Upper
***LOX* rs10519694 (CT +TT vs CC)**	272	(137 vs 135)	**0.012**	**0.69**	**0.51**	**0.919**
***HIF2A* rs11125070 (AT+TT vs AA)**	247	(126 vs 121)	**0.003**	**0.63**	**0.461**	**0.858**

Table 23: Multivariable model for disease free survival (phase I) (n=236)

Variables	Number of patients in each category	p-value	HR	95% CI	
				Lower	Upper
LOX rs10519694 (CT +TT vs CC)	(117 vs 119)	0.113	0.767	0.552	1.065
***HIF2A* rs11125070 (AT+TT vs AA)**	(119 vs 117)	**0.004**	**0.619**	**0.446**	**0.859**
Age		**<.001**	**1.034**	**1.019**	**1.048**
Grade (poorly/ undifferentiated vs well/moderately differentiated)	(35 vs 201)	0.099	1.437	0.934	2.213
Stage		**<.001**			
Stage (II vs I)	(85 vs 46)	0.112	1.553	0.903	2.673
Stage (III vs I)	(65 vs 46)	**<.001**	**2.93**	**1.684**	**5.095**
Stage (IV vs I)	(40 vs 46)	**<.001**	**133.7**	**55.98**	**319.3**
MSI-H status (Yes vs No)	(22 vs 214)	**0.013**	**0.4**	**0.194**	**0.825**

CI: Confidence Interval, HR: Hazard Ratio, MSI-H: Microsatellite Instability-High, vs: versus. Values that are bolded are statistically significant (p value <0.05).

Patients with the AT or TT genotypes (containing the minor allele T) had 38% decreased risk of recurrence, metastasis, or death when compared to patients with the AA genotype. Also in this model, while increasing age and stage were also associated with higher risk of recurrence, metastasis, or death, the MSI-H status was associated with the decreased risk of these events (**Table 23**).

3.6.2) Phase II results

Description of the patient cohort

The patient cohort investigated in phase II (i.e. the NFCCR SNP genotyping cohort) is described previously in **Section 2.4.3**.

Overall survival analysis (phase II)

In phase II, 77 hypoxia pathway gene polymorphisms (**Section 3.6.**) were analyzed in univariate analyses and their OS results are shown in **Appendix 12**. Two polymorphisms were associated with overall survival time (**Table 24**). Presence of one or two C alleles for *HIF2A* rs4953352 polymorphism was associated with increased risk of death when compared to the patients with TT genotype. Presence of one or two A alleles for *HIF2B* rs12593988 polymorphism were associated with decreased risk of death when compared to the GG genotype. Genotype data for both of these polymorphisms were in HWE and Kaplan-Meier survival curves are displayed in **Figure 14.A**.

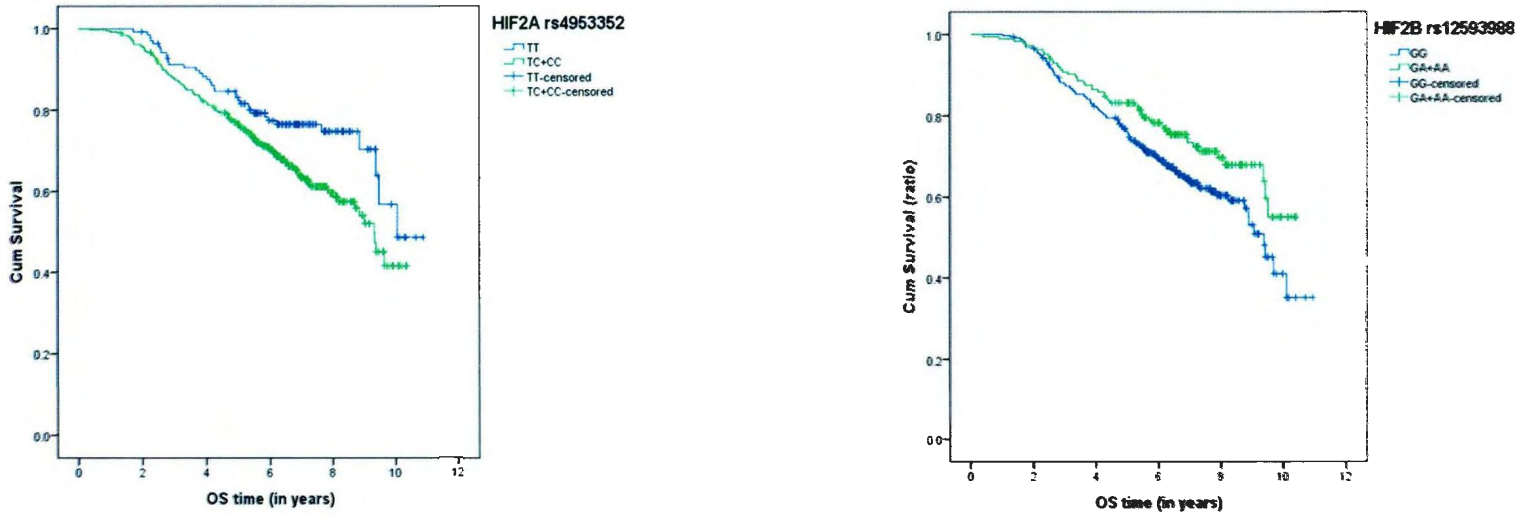
Table 24: Polymorphisms significantly associated with outcome in OS univariate analysis (phase II)

Variables	n	Number of patients in each category	p-value	HR	95% CI	
					Lower	Upper
***HIF2A* rs4953352 (TC+CC vs TT)**	535	(396 vs 139)	**0.013**	**1.59**	**1.103**	**2.292**
***HIF2B* rs12593988 (AG+AA vs GG)**	535	(185 vs 350)	**0.021**	**0.69**	**0.497**	**0.946**

These two polymorphisms (*HIF2A* rs4953352 and *HIF2B* rs12593988) when adjusted for sex, stage, vascular invasion and MSI status, were also significantly associated with overall survival in a multivariate model (**Table 25**). Specifically, the genotypes containing the minor allele of the *HIF2A* rs4953352 polymorphism (TC and CC genotypes) were associated with poor prognosis, whereas the genotypes containing the minor allele of the *HIF2B* rs12593988 polymorphism (GA and AA genotypes) were associated with better prognosis when compared to the major homozygote genotypes (TT in *HIF2A* rs4953352 and GG in *HIF2B* rs12593988). In addition, MSI and stage were also associated with outcome in this OS multivariate analysis (**Table 25**).

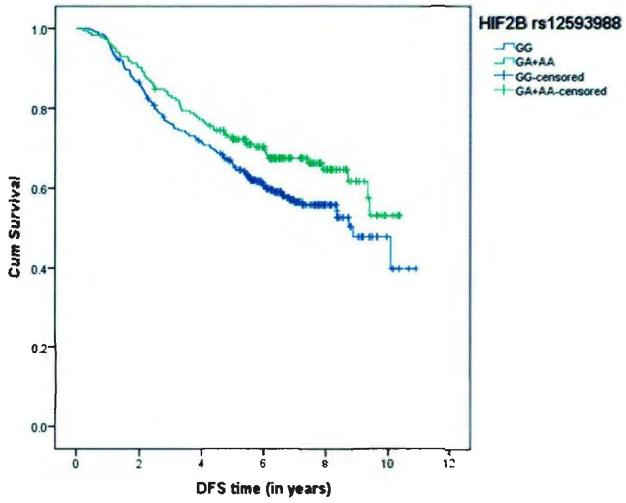
Figure 14: Kaplan Meier survival plots for polymorphisms whose associations were significant in the univariate analysis (phase II)

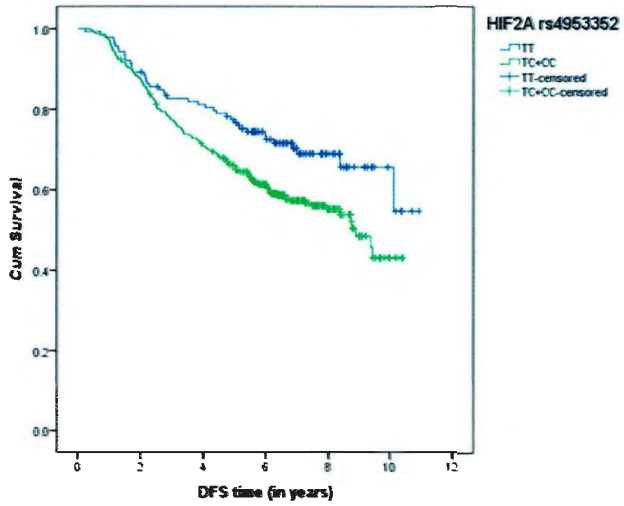
A) Overall survival



OS time is defined as the time interval from the time of diagnosis to the date of last follow up or the date of death. Cumulative survival refers to the proportion of patients who are alive at a particular point of time. Survival plots were generated by the Kaplan Meier method.

B) Disease free survival





DFS time is defined as the time interval from the time of diagnosis to the date of last follow up or the date of recurrence or metastasis or death.

Table 25: Multivariate model for overall survival (phase II) (n=478)

Variables	Number of patients in each category	p-value	HR	95% CI	
				Lower	Upper
***HIF2A* rs4953352 (TC+CC vs TT)**	(352 vs 126)	**<.001**	**2.184**	**1.464**	**3.257**
***HIF2B* rs12593988 (GA+AA vs GG)**	(165 vs 313)	**0.008**	**0.62**	**0.437**	**0.881**
Vascular invasion (+ vs -)	(164 vs 314)	0.162	1.268	0.909	1.769
Sex (male vs female)	(297 vs 181)	0.122	1.304	0.932	1.826
Stage		**<.001**			
Stage (II vs I)	(197 vs 90)	0.167	1.489	0.846	2.62
Stage (III vs I)	(154 vs 90)	**0.042**	**1.825**	**1.021**	**3.259**
Stage (IV vs I)	(47 vs 90)	**<.001**	9.704	5.264	17.89
MSI-H status (Yes vs No)	(56 vs 422)	**0.002**	**0.27**	**0.118**	**0.616**

CI: Confidence Interval, HR: Hazard Ratio, MSI-H: Microsatellite Instability-High, vs: versus. Values that are bolded are statistically significant (p <0.05).

Disease free survival analysis (phase II)

In the case of the DFS univariate analysis, significant associations of three out of seventy-seven polymorphisms were detected, namely the *HIF2A* rs4953352, the *HIF2B* rs12593988, and the *HIF2B* rs8033706 (**Table 26**). Out of these three polymorphisms, the genotype distribution of the *HIF2B* rs8033706 polymorphism deviated from HWE, which was excluded from the multivariate analysis. Univariate analysis for rs8033706 and other SNPs are shown in **Appendix 14**. The Kaplan-Meier survival curves for the remaining two polymorphisms are presented in **Figure 14.B**.

Table 26: Polymorphisms significantly associated with outcome in univariate disease free survival analysis

Variables	n	Number of patients in each category	p-value	HR	95% CI	
					Lower	Upper
***HIF2A* rs4953352 (TC+CCvsTT)**	534	(395 vs 139)	**0.01**	**1.6**	**1.12**	**2.2**
***HIF2B* rs12593988 (AG+AAvsGG)**	534	(185 vs 349)	**0.04**	**0.7**	**0.54**	**0.98**

The association of *HIF2A* rs4953352 and *HIF2B* rs12593988 polymorphisms remained significant in a multivariate model after adjusting for sex, location, stage, vascular invasion and MSI status (**Table 27**). The combined TC and CC genotypes of the *HIF2A* rs4953352 polymorphism were associated with a 2-fold increased risk of recurrence, metastasis or death ($p<0.001$; HR=1.96; 95% CI: 1.362-2.82), when compared to the TT genotype. The combined AA and AG genotypes of the *HIF2B* rs12593988 polymorphism were associated with a 33% decreased risk of recurrence, metastasis, or death ($p=0.014$; HR=0.671; 95% CI: 0.488-0.922) when compared to the GG genotype (**Table 27**).

Table 27: Multivariate model for disease free survival (phase II) (n=477)

Variables	Number of patients in each category	p-value	HR	95% CI	
				Lower	Upper
***HIF2A* rs4953352 (TC+CC vs TT)**	(351 vs 126)	**<0.001**	**1.96**	**1.362**	**2.82**
***HIF2B* rs12593988 (GA+AA vs GG)**	(165 vs 312)	**0.014**	**0.671**	**0.488**	**0.922**
Sex (male vs female)	(296 vs 181)	0.146	1.26	0.923	1.72
Location (rectum vs colon)	(151 vs 326)	0.111	1.284	0.944	1.746
Stage		**<0.001**			
Stage (II vs I)	(187 vs 89)	0.233	1.351	0.824	2.212
Stage (III vs I)	(154 vs 89)	**0.033**	**1.735**	**1.045**	**2.88**
Stage (IV vs I)	(47 vs 89)	**<0.001**	**5.58**	**3.178**	**9.8**
Vascular invasion (+ vs -)	(164 vs 313)	0.239	1.206	0.883	1.646
MSI-H status (Yes vs No)	(55 vs 422)	**0.019**	**0.455**	**0.237**	**0.876**

CI: Confidence Interval, HR: Hazard Ratio, MSI-H: Microsatellite Instability-High. Values that are bolded are statistically significant ($p < 0.05$).

3.6.3) Investigation of identical or highly correlated polymorphisms in both phases and the pooled cohort analyses

A total of 13 polymorphisms were investigated in both phase I and phase II (**Appendix 14**). Association of none of these 13 polymorphisms with outcome was detected in either phase I or phase II cohorts. A univariate analysis of the pooled cohort data did not find an association of these polymorphisms with OS and DFS. (**Appendices 15** and **16**).

There were 15 additional polymorphisms investigated in phase I whose genotypes were highly correlated ($r^2>0.8$) with the genotypes of 15 other polymorphisms

investigated in phase II (**Appendix 14**). These highly correlated polymorphisms can serve as surrogates for each other.

As discussed in **Section 3.8.1.**, *HIF2A* rs11125070 polymorphism was associated with DFS in a multivariate model in phase I patient cohort. There was a polymorphism (rs4953342) genotyped in phase II patient cohort, which was highly correlated (r^2=0.91) with *HIF2A* rs11125070. Therefore, association of rs4953342 with DFS was examined in phase II cohort. Association of this polymorphism was not detected in phase II patient cohort, although there was a trend for association (**Appendix 17**). In the pooled cohort, this association was not detected (**Appendix 19**).

On the other hand, in the phase II cohort, *HIF2A* rs4953352 was associated with OS and DFS in multivariate models (**Section 3.8.2**). However, association of *HIF2A* rs4953352 polymorphism, which is highly correlated with *HIF2A* rs6706003 (r^2=0.87), was not associated with OS or DFS in phase I cohort (in univariate analyses) (**Appendices 9-11**). In a univariate analysis of pooled cohort data, a trend towards association of this SNP with OS and DFS was observed, but it did not reach statistical significance (p=0.091, HR=1.242, 95%CI: 0.966-1.596 for OS; **Appendix 18** and p=0.058, HR=1.258, 95%CI: 0.992-1.596 for DFS; **Appendix 19**).

For the remaining identical or highly correlated polymorphisms (r^2>0.8) with outcome, no associations were detected in either OS or DFS in the univariate analyses of the pooled cohort (**Appendices 15-16** and **18-19**).

3.6.4) Differences between phase I and phase II cohorts

In order to test whether the failure to replicate the associations observed (**Table 21**, **Appendix 17, Tables 26-27 and Appendix 9** and **11**) might be explained by the differences in the baseline and prognostic characteristics of phase I and phase II cohorts, additional statistical analyses were performed. There were significant differences between the pilot project study and the NFCCR SNP genotyping cohorts for age ($p < 0.001$), sex ($p=0.036$), grade ($p=0.001$), lymphatic invasion ($p<0.001$), location ($p < 0.001$) and stage ($p=0.018$) (**Table 28**).

Table 28: Differences between phase I and phase II cohorts in terms of clinicopathological and prognostic characteristics

Variables compared	p-value	Comments
Age	**<.001**	
Sex	**.036**	There is a higher proportion of female patients in the pilot project study cohort (46.3%) when compared to the NFCCR SNP genotyping cohort (38.6%)
Histology	.102	
Location	**<0.001**	NFCCR SNP genotyping cohort has a higher proportion of patients with rectal tumours (33.6%) when compared to the pilot project study cohort (20.2%)
Stage	**.018**	Pilot project study cohort had 17.4% Stage IV patients, whereas the NFCCR SNP genotyping cohort had 9.9% stage IV patients
Grade	**.001**	14.9% patients had poorly differentiated tumours in pilot project study cohort, whereas in the NFCCR SNP genotyping cohort only 7.3% of patients had tumours with poor differentiation
Lymphatic invasion	**<0.001**	61.6% patients in the pilot project study cohort had tumours characterized by lymphatic invasion, whereas, only 36.2% patients in the NFCCR SNP genotyping cohort had tumours with lymphatic invasion
OS status	**<0.001**	62.5% patients experienced death in the pilot project study cohort, whereas only 34% of patients experienced death in the NFCCR SNP genotyping cohort
DFS status	**<0.001**	66.9% patients experienced DFS outcome in the pilot project study cohort, whereas only 39.8% patients experienced outcome in the NFCCR cohort
MSI status	.962	

Variables that are bolded have a p <0.05. p-values are based on the Chi-square and the Mann-Whitney U test results.

3.7) Discussion

Hypoxia is a common condition in solid tumours (167) and has various effects on cancer progression. First, hypoxia results in a metabolic shift from the oxidative phosphorylation pathway to aerobic glycolysis pathway. The acidic products produced as a result of glycolysis lowers the pH within the tumour (264). This promotes invasion (265) and metastasis (264). Second, hypoxia may promote resistance to conventional anti-cancer therapies (266). Thus, genetic variations in the genes that control the response to hypoxia and in other genes that are regulated under hypoxic conditions may be markers that can predict outcome. In this study, genetic polymorphisms within select genes of the hypoxia pathway were investigated for their potential prognostic significance in colorectal cancer.

This project was performed in two phases. Initially, the prognostic significance of 49 SNPs from six hypoxia-pathway genes (*HIF1A, HIF1B, HIF2A, CXCL12, LOX* and *MIF*) were tested in relation to outcome in a colorectal cancer patient cohort (phase I cohort, n=272). In phase II, 77 polymorphisms from seven genes were investigated (*HIF1A, HIF1B, HIF2A, HIF2B, HIF3A, LOX* and *CXCL12*) in a separate cohort (phase II cohort, n=536).

None of the 49 SNPs investigated in phase I were associated with DSS. Two polymorphisms (*LOX* rs10519694 and *HIF2A* rs11125070) were associated with DFS and three polymorphisms (*LOX* rs10519694, *HIF2A* rs11125070 and *HIF2A* rs1868084) were associated with OS in the univariate analyses. In multivariate analyses, after adjusting for

other clinicopathological variables, only one of these polymorphisms (*HIF2A* rs11125070) was significantly associated with outcome (DFS). However, association of this polymorphism with DFS was not detected in phase II cohort using the genotype data of a highly correlated polymorphism (univarate analysis). This result may be attributed to the differences between the two cohorts in terms of their characteristics (**Section 3.8.4, Table 28**). For example, in phase I cohort, there were more patients who had passed away or experienced recurrence or metastasis until the end of the follow up period when compared to phase II cohort (**Table 28**). In addition, phase II cohort had a lower proportion of stage IV patients than phase I cohort. There were also differences between the two cohorts in terms of age, sex and grade distributions (**Table 28**). Alternatively, the replication of association of this polymorphism with DFS may be due to the fact that its association detected in phase I was a false-positive detected by chance. Therefore, the main conclusion of phase I is that none of the 49 polymorphisms analyzed were associated with outcome in colorectal cancer patients. However, due to the small sample size of the cohort analyzed (n=272), it is also likely that this study may be underpowered to detect effects of these polymorphisms, as the statistical power to detect an association increases with increase in sample size (267).

In phase II, a cohort twice the size of phase I was tested (n=536). The potential prognostic associations of 77 polymorphisms in seven hypoxia pathway genes were investigated. In this phase of the project, the primary aim was to investigate the associations of polymorphisms studied in phase I in a larger colorectal cancer patient

cohort. However, genotype data for only 28 polymorphisms (either identical or with highly correlated genotypes (r^2>0.8)) were available in both cohorts (**Appendix 14)**.

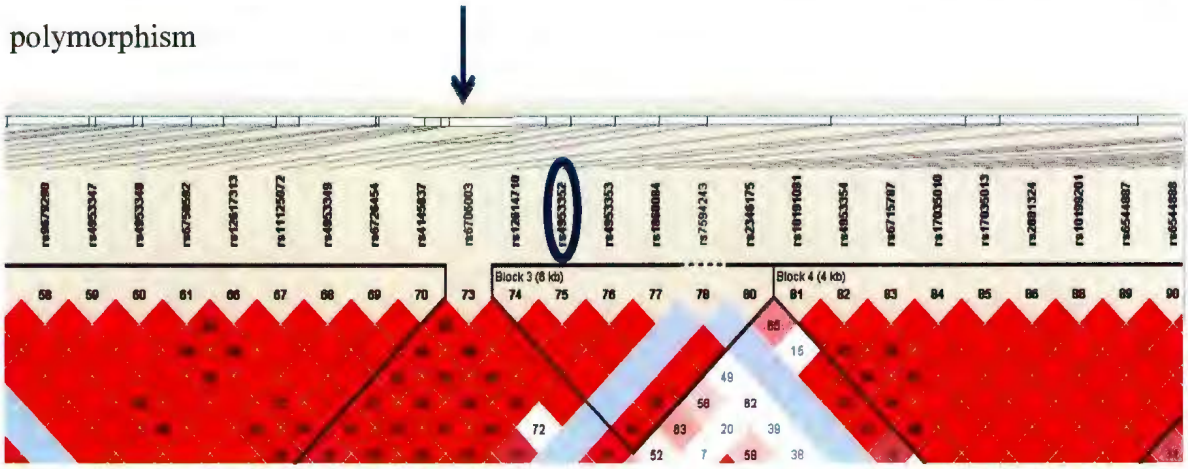
The main conclusion from phase II is that two of the 77 polymorphisms (*HIF2B* rs12593988 and *HIF2A* rs4953352**)** were associated with both OS and DFS in multivariable models when adjusted for other variables (**Table 25** and **Table 27**). The association of *HIF2A* rs4953352 with OS and DFS may be a false-positive, since no association was detected in phase I patient cohort using the genotype data of a highly correlated polymorphism (*HIF2A* rs6706003) in this cohort. However, it is also possible that an association of *HIF2A* rs6706003 with OS and DFS was missed in phase I patient cohort, which has a small sample size (n=272). Associations of *HIF2B* rs12593988 and *HIF2A* rs4953352/rs6806003 therefore should be tested in another colorectal cancer patient cohort.

Both the *HIF2B* rs12593988 and the *HIF2A* rs4953352 polymorphisms are located in introns and their minor allele frequencies in phase II patient cohort were 19% and 49%, respectively. The linkage disequilibrium (LD) maps of these two regions are shown in **Figure 15** and **Figure 16,** respectively.

As shown in **Figure 15**, according to the HapMap data (259), the *HIF2A* rs4953352 polymorphism is in a 6 kb LD block with five other SNPs in intron 1. *HIF2A* rs4953352 is correlated (r^2=0.87) with rs6706003, which is located in the same intron as rs4953352. Thus, if these associations are true, then either of these polymorphisms may be the causal variant. To our knowledge, no functional experiments have been performed to study the

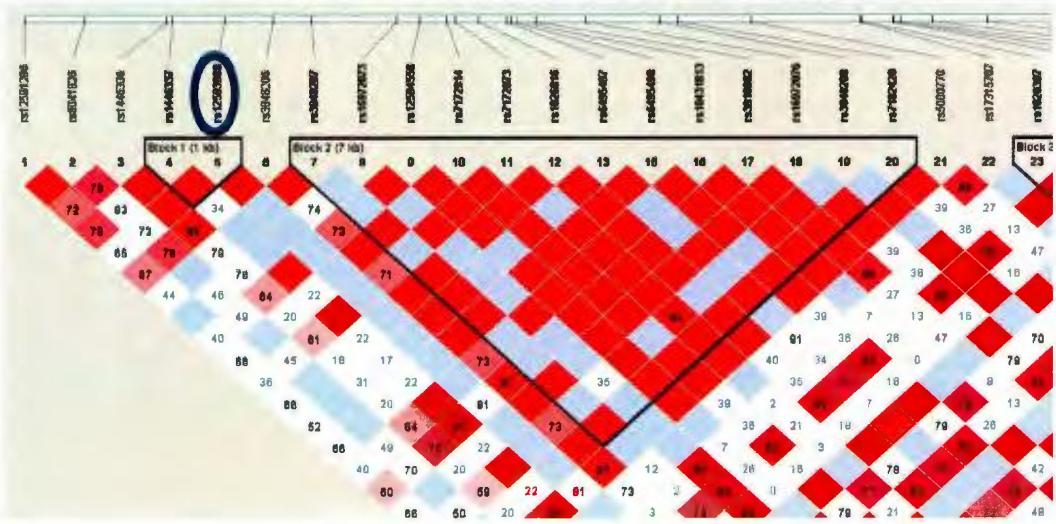
biological effects of these two polymorphisms. However, a regulatory role of *HIF2A* and *HIF2B* in influencing the expression of these genes cannot be ruled out. The *HIF2B* rs12593988 polymorphism is located in a one kb LD block (**Figure 16**). This polymorphism does not highly correlate with any other HapMap polymorphism and its functional consequence is unknown.

Figure 15: Linkage disequilibrium plot showing the position of the *HIF2A* rs4953352 polymorphism



The LD map is based on the HapMap (259) *data for the Caucasian population. The HIF2A rs4953352 polymorphism is circled. The arrow indicates the highly correlated polymorphism, rs6706003 (r^2=0.87).*

Figure 16: Linkage disequilibrium plot showing the position of the *HIF2B* rs12593988 polymorphism



The HIF2B rs12593988 polymorphism is circled. The LD map is based on the HapMap (259) *data for the Caucasian population.*

For the 28 polymorphisms investigated in both phase I and phase II, univariate analyses using the pooled patient data were also performed. This pooled cohort analysis was done to test whether the increased sample size (and thus the study power) would be able to detect the potential effects of these polymorphisms in patient prognosis. Between 719-807 patients were investigated. However, no significant associations between the SNPs and outcomes were detected (**Appendices 16-17** and **19-20**). These results suggest either these hypoxia-pathway polymorphisms are not associated with outcome in colorectal cancer or their effects are not large enough to be identified with this sample size.

This project generated novel scientific knowledge of a total of 97 different polymorphisms which were not previously studied in relation to prognosis in colorectal cancer. A search in the dbCPCO database (268) (which is a database of SNPs that have been tested for their associations with patient prognosis in colorectal cancer) revealed that among the 98 SNPs investigated in this study, only the *HIF1B* rs2228099 polymorphism had been investigated by other researchers previously (190). This polymorphism was not associated with prognosis which is concordant with the results obtained in this study.

Every study has its own strengths and limitations, and this study is not an exception. A correction for multiple testing was not performed to minimize the false negative findings. Because of the large number of statistical tests performed, it is possible that the associations detected for the *HIF2A* rs4953352 and *HIF2B* rs12593988 in phase II are false-positives. These results therefore should be considered as hypothesis-generating and be interpreted with caution. The patient cohort studied in phase I had a relatively small sample size (n=272). In addition, phase II patient cohort which is a subset of the NFCCR patient cohort was inclined towards earlier stages and was not representative of the NFCCR cohort. This was because DNA samples (extracted from blood) were not available for all stage IV patients recruited to the NFCCR. In other words, this is attributed to the fact many stage IV patients were not alive or unable to give blood samples. Therefore, stage IV patients were underrepresented in the SNP genotyping cohort and this may be one possible reason for missing potential associations between majority of the polymorphisms in the hypoxia pathway genes and outcome in the phase-II cohort. To our knowledge, phase II cohort (n=536) is also one of the largest to be

investigated in such a genetic prognostic study of colorectal cancer. Both of these cohorts also had long follow up periods allowing the detection of clinical events of interest over time, which increased the study power. This study did not include all the genes (~450) (269) functioning in the hypoxia pathway; five genes with critical biological roles in cellular response to hypoxia (*HIF1A*, *HIF1B*, *HIF2A*, *HIF2B*, and *HIF3A*) and three biologically well-characterized genes regulated under hypoxic conditions (*MIF*, *CXCL12*, and *LOX*) were included in this study. In addition, the polymorphisms included in this study were selected based on the tagSNP (or correlation) data, reducing the redundancy in experiments and statistical analysis. We should also note that the correlations between polymorphisms were calculated based on the genotype data in the HapMap database, which does not contain all the polymorphism in the genes studied. For example, 1000 Genome project generated a more comprehensive catalogue of human genetic variations (270). Therefore, a detailed analysis of the hypoxia genes investigated using 1000 Genome Project data may be helpful in studying the prognostic associations of other polymorphisms not included in this project.

In conclusion, the results obtained in this study suggest no evidence of association of the majority of the polymorphisms investigated with outcome in colorectal cancer. Association of *HIF2A* rs4953352 and *HIF2B* rs12593988 polymorphisms with outcome need to be studied in larger colorectal cancer cohorts to either confirm their association with outcome or to rule them out as false positive associations.

CHAPTER 4: GENERAL CONCLUSIONS

This is one of the first genetic prognostic studies involving polymorphisms conducted in colorectal cancer in Canada (271,272).

Colorectal cancer is a common disease with a significant impact on mortality rates in Canada, especially in Newfoundland (3). In addition to commonly used clinicopathological prognostic markers such as age and stage, genetic markers can also be used to predict outcome. The main objective of this study was to identify genetic markers that can predict outcome in colorectal cancer patients.

Hypoxic conditions by resulting in aggressive tumour phenotype and resistance to anti-cancer therapies, facilitates cancer progression and could affect patient outcome (173,266). Similarly, mitochondrion is involved in energy production, cellular metabolism and apoptosis. Mitochondrial dysfunction may result in cancer progression (113). Thus, in this study it is hypothesized that genetic variations in hypoxia pathway genes and mitochondrial DNA may predict outcome in colorectal cancer. To test these hypotheses, six mtDNA SNPs and the mtDNA copy number change in colorectal tumours were investigated. In addition, in the hypoxia pathway project, a total of 98 different SNPs in selected hypoxia genes were investigated in two different cohorts.

Neither the six SNPs in the mitochondrial DNA nor the mtDNA copy number change (in tumours compared to non-tumour colorectal tissues) were significantly associated with outcome in colorectal cancer patients. However, detected associations of

two SNPs namely *HIF2A* (rs4953352) and *HIF2B* (rs12593988) in the hypoxia pathway genes with outcome in phase II patient cohort may be promising. Association of one of these SNP (*HIF2A* rs4953352) failed to be replicated in a smaller colorectal cancer cohort (phase I cohort). Whether this is due to the small sample size of the phase I cohort or it was a false-positive association in phase II cohort can be confirmed by testing the association of this SNP in another large cohort. Additionally, association of the *HIF2B* rs12593988 polymorphism with outcome detected in this thesis project has to be replicated in another cohort to confirm this result.

5) Appendices

Appendix 1: Solutions used in the project described in Chapter 2

5X Tris-Borate-EDTA (TBE) adjusted to pH 8.3

Before preparing the 5X TBE solution the material safety data sheet (MSDS) was referred to and precautionary measures such as personal protective equipment (PPE) were used as recommended in the MSDS.

54 gr Tris-HCl (OmniPur® Tris-Hydrochloride (Tris-HCl), Product code 9310, EMD Chemicals Inc. NJ, USA) and 27.5 gr Boric acid (Product code BX0865, EMD Chemicals Inc. NJ, USA) were weighed in an analytical balance (P/P.I – 214, Denver instruments, USA) and transferred to a one-litre glass beaker. Approximately 850 ml of deionised water (dH_2O) was added and the beaker was placed on a magnetic stirrer/heater (AP 195025, Thermo Scientific, USA). The beaker contents were dissolved with the help of a magnetic stirrer. 20 ml 0.5M EDTA (pH.8; catalogue number: 46-034-Cl, Mediatech Inc, VA, USA) was then added to the solution using a 10 ml serological pipette. When the contents got mixed, the solution was transferred to a one-litre measuring cylinder and the volume was adjusted to one litre with dH_2O. Then, the solution was transferred back to the beaker and pH of the solution was adjusted as follows: the pH of the solution was approximately 4. To increase the pH value to 8.3, sodium hydroxide pellets (NaOH: Product code SX0590, EMD Chemicals Inc. NJ, USA) were added in small quantities to this solution and the pH was checked each time after the pellets dissolved completely.

The process was continued until the pH of the solution reached the value of 8.3. Then, the solution was transferred into a Workplace Hazardous Materials Information System (WHIMS) labeled glass bottle using a funnel, autoclaved, and stored at room temperature (RT).

1X TBE buffer

This solution is prepared by diluting the 5X TBE solution using sterile dH_2O for the desired volume using the equation $V_1X_1=V_2X_2$, where V_1 and X_1 are the initial volume and concentration and V_2 and X_2 are the desired volume and concentration, respectively. In short, to prepare 1X TBE, a 1:5 dilution of 5X TBE stock solution with dH_2O was performed. The same precautionary safety measures (such as using the lab coats and safety glasses) that were used during the preparation of 5X TBE solution was followed during the preparation of 1X TBE solution. After dilution, the 1X TBE solution was transferred to a WHIMS labeled glass bottle and stored at RT.

1X Tris-EDTA (TE) buffer

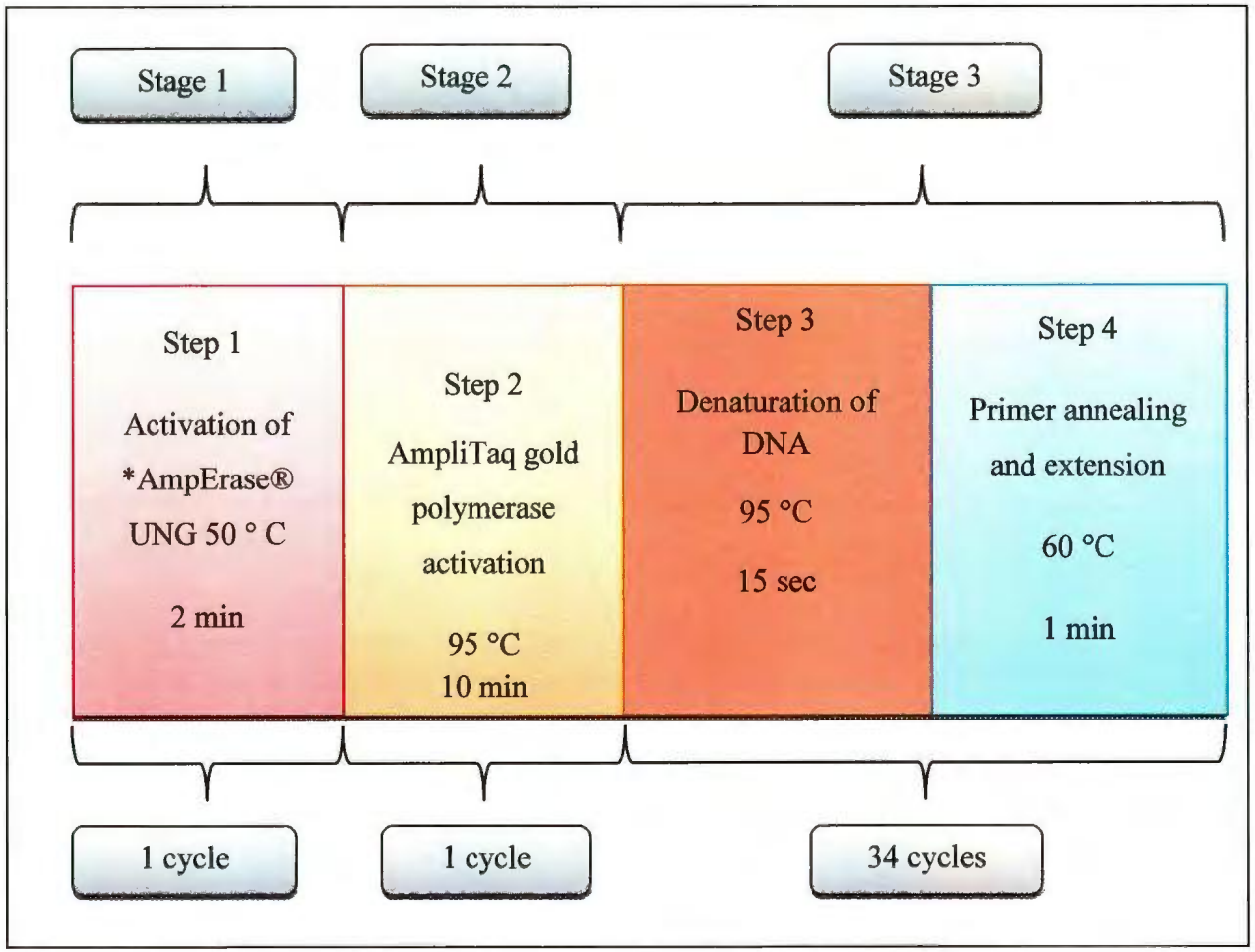
0.3152 gr of Tris-HCl (Product code 9310, EMD Chemicals Inc. NJ, USA) was weighed in an analytical balance (P/P.I – 214 , Denver instruments, USA) and dissolved in 200 ml of dH_2O. To this mixture, 0.4 ml of 0.5M stock solution of EDTA (pH: 8±0.1; catalogue number: 46-034-Cl, Mediatech Inc, VA, USA) was added using a serological pipette. The solution was then mixed, autoclaved, and stored in a WHIMS labeled bottle at RT.

Appendix 2: List of excluded mtDNA polymorphisms

SNP ID	Number of missing genotypes	Number of samples with major allele	Number of samples with minor allele	Number of samples that are heteroplasmic	MAF (%)
MitoC1050T	1	536	0	0	0
MitoG3667A	1	536	0	0	0
MitoC3971T	1	536	0	0	0
MitoT6681C	1	536	0	0	0
MitoT7176C	1	536	0	0	0
MitoT14179C	1	536	0	0	0
MitoC16149T	8	529	0	0	0
MitoG7522A	1	535	1	0	0.19
MitoT1191C	1	513	23	0	4.3
MitoG4770A	1	532	4	0	0.75
MitoG4821A	1	531	4	1	0.84
MitoA9668G	1	524	12	0	2.24
MitoA12309G	1	452	84	0	15.7
MitoA14234G	1	501	35	0	6.54
MitoT9951C	1	535	1	0	0.19
MitoA10045G	1	534	2	0	0.37
MitoG10590A	1	534	2	0	0.37
MitoA14583G	1	531	5	0	0.93
MitoG15258A	1	535	1	0	0.19
MitoA16164G	29	507	1	0	0.21
MitoC16184A	4	525	8	0	1.51

1) SNPs in black fonts are monoallelic, 2) SNPs in green font had sequence homology with nDNA and 3) SNPs in blue font had MAFs less than 4%. Heteroplasmic: patients with both alleles for a particular genotype, MAF: minor allele frequency

Appendix 3: The PCR thermocycling conditions for both the Veriti and Applied Biosystems 7900HT Fast Real-time PCR Systems



**AmpErase® UNG is used to prevent the carry-over of PCR products from previous reactions and thus help prevent contamination.*

Appendix 4: Additional steps taken to make sure the selected qPCR primer and probe binding sequences did not have known sequence alteartions that can affect the qPCR results

FASLG target sequence (nuclear gene): Our first step was to make sure that there was no somatic mutation in colorectal cancer tumours or SNPs interfering with the primer and probe binding regions of the *FASLG* gene described by Cossarizza and colleagues (134). We spotted the location of primer and probes on the *FASLG* gene sequence by blasting the sequences of *FASLG* gene with the primer and probe sequences using the BLAST program (273). The somatic point mutations in colorectal cancer cell lines for this particular gene were also checked in the catalogue of somatic mutations in cancer (COSMIC) genomic browser (274). According to this database, there were no non-sense or mis-sense substitutions or deletions in the *FASLG* gene in tumour samples obtained from the large intestine samples. However, there were three mutations after the amino acid position 190 in other tissue samples. To check whether these probe and primer binding regions fell under the coding region of the first 190 amino acids, we blasted the mRNA sequence with the entire *FASLG* gene sequence and found that both the probes and primers binding sites were located within the mutation free region. Additionally, we did not find any reported copy number variation for this gene using the Database of Genomic Variants in October 2011 (275). We then checked if there were any SNPs interfering with the primer and probe binding site using the information in dbSNP database (138). None of the SNPs located in primer/probe binding sites were deemed to

affect our reactions (i.e. MAFs >0.1%). Therefore, these investigations showed that the selected primers and probes in this gene (134) were suitable for our qPCR design.

***ND-2* target sequence (mtDNA gene):** First, there is a 4977 bp common deletion in mitochondrial DNA encompassing positions 8469 and 13447 (112). We found that the *ND-2* primers and probe were not located within this deletion. Second, the exact location of the primer and probe binding regions were identified by blasting the entire mtDNA sequence with the primer and probe sequences and we found that there were no somatic mutations in the primer and probe binding sites using the information from the Mitomap database (276). Next, we analysed whether there were any SNPs in the primer and probe binding sites using the information in the mtdb database (131). According to this database, there was a SNP at position 4655 located within the probe binding site. This SNP was present in 48 samples out of the 2704 samples analysed, but among the 48 samples that had this SNP only one was a European sample. Since the SNP is a rare variant (frequency less than 0.001) in Caucasians, we came to a conclusion that this SNP was not going to confound our quantification reactions substantially. Altogether, we concluded that the *ND-2* gene was suitable to be used in the qPCR reaction.

Appendix 5: Assay IDs for SNPs genotyped by TaqMan® SNP genotyping technique in phase I

SNP ID	Assay ID
rs9679290	C_229862_10
rs4952818	C_2148920_20
rs2346175	C_11158065_10
rs6753127	C_29117756_10
rs7583554	C_29117759_10
rs9973653	C_30621213_10
rs3768730	C_2162965_10
rs11687512	C_2148916_10
rs10178633	C_2163036_10
rs2236533	C_15954640_10
rs11592974	C_17776531_10
rs10847	AHHSOC5
rs2121266	C_16104055_20
rs4953340	C_29557501_10

Appendix 6: Genes selected and information related to 53 polymorphisms included in phase I

Gene	SNP	*MAF of SNP
HIF1A	rs11158358	0.133
	rs2301106	0.1
	rs2301111	0.161
	rs2301113	0.172
ARNT (HIF1B)	rs10305724	0.125
	rs10847	0.225
	rs11204737	0.473
	rs2228099	0.35
	rs3738483	0.142
LOX	rs10040971	0.158
	rs10519694	0.258
	rs2288393	0.183
	rs2956540	0.431
MIF	rs2096525	0.12
CXCL12	rs11592974	0.138
	rs2236533	0.218
	rs2236534	0.224
	rs2839688	0.17
EPAS1 (HIF2A)	rs10178633	0.491
	rs10199201	0.2
	rs11125070	0.237
	rs11687512	0.1
	rs12614710	0.367
	rs13019414	0.424
	rs13412887	0.186
	rs1374748	0.15
	rs1562453	0.458
	rs1562453	0.458
	rs1867783	0.417
	rs1868084	0.175
	rs1868087	0.2

	rs1992846	0.225
	rs2034327	0.4
	rs2044456	0.292
	rs2121266	0.333
	rs2346175	0.458
	rs2346176	0.392
	rs3768728	0.169
	rs3768728	0.169
	rs3768728	0.169
	rs3768730	0.496
	rs4145836	0.167
	rs4953340	0.317
	rs4953344	0.417
	rs4953349	0.417
	rs4953353	0.408
	rs6706003	0.45
	rs6712143	0.319
	rs6753127	0.142
	rs7583392	0.466
	rs7583554	0.358
	rs9679290	0.5
	rs9973653	0.258

**MAFs are based on the genotype data of the Caucasian samples genotyped in the HapMap project* (259).

More detailed information about these SNPs can be found in **Appendix 7**.

Appendix 7: Genes selected in phase II and the information related to their polymorphisms

Gene	tagSNP	tagged SNP	r^2
HIF1A	rs11158358	rs2057482	0.929
	rs2301111	rs1951795	0.938
	rs2301111	rs12435848	0.938
	rs2301111	rs12434438	1
	rs2301111	rs10873142	0.935
	rs2301113	rs2301113	1
HIF1B	rs10305724	rs10305724	1
	rs10847	rs10847	1
	rs11204737	rs11204735	1
	rs11204737	rs11204737	1
	rs2228099	rs1889740	1
	rs2228099	rs10305714	0.892
	rs2228099	rs10305711	0.892
	rs2228099	rs3768016	0.892
	rs2228099	rs2256355	1
	rs2228099	rs3768015	0.894
	rs2228099	rs3768013	1
	rs2228099	rs7532045	1
	rs3738483	rs2134688	1
	rs3738483	rs3820541	0.918
LOX	rs10040971	rs3792801	0.838
	rs10040971	rs3792802	0.831
	rs10519694	rs10519694	1
	rs2288393	rs10059661	1
	rs2288393	rs1800449	0.944
	rs2956540	rs2956540	1
MIF	rs2096525	rs2096525	1
CXCL12	rs11592974	rs2839695	1
	rs11592974	rs2839692	1
	rs11592974	rs2839690	0.932
	rs11592974	rs2839689	1
	rs11592974	rs7092453	1
	rs2236533	rs2236533	1

	rs2236534	rs2236534	1
	rs2839688	rs2839688	1
HIF2A	rs10178633	rs1374749	0.898
	rs10199201	rs4953354	0.95
	rs10199201	rs10187368	1
	rs11125070	rs4953342	0.913
	rs11687512	rs11694197	1
	rs12614710	rs12614710	1
	rs13019414	rs3754556	0.807
	rs13019414	rs7565341	0.803
	rs13019414	rs3088359	0.803
	rs13019414	rs7571218	0.933
	rs13019414	rs11690950	1
	rs13019414	rs7590087	1
	rs13412887	rs13412887	1
	rs1374748	rs13409493	0.871
	rs1562453	rs10191091	0.846
	rs1562453	rs2881324	0.818
	rs1562453	rs6544887	0.875
	rs1562453	rs6544888	0.875
	rs1562453	rs6544889	0.875
	rs1562453	rs4953356	0.967
	rs1562453	rs6756667	0.967
	rs1562453	rs1562452	0.875
	rs1562453	rs6735812	0.967
	rs1562453	rs4953358	0.967
	rs1562453	rs4953359	0.966
	rs1562453	rs6743991	0.935
	rs1562453	rs6707241	0.967
	rs1562453	rs6740096	0.966
	rs1562453	rs1530631	0.967
	rs1562453	rs1530633	0.935
	rs1867783	rs2346417	1
	rs1867783	rs4952818	0.966
	rs1868084	rs1868084	1
	rs1868087	rs4953355	1
	rs1868087	rs1868085	0.948
	rs1868087	rs1868086	1

	rs1868087	rs1868088	0.898
	rs1992846	rs1992846	1
	rs2034327	rs2034327	1
	rs2044456	rs17034950	0.921
	rs2121266	rs1867787	0.963
	rs2121266	rs10211665	0.927
	rs2121266	rs11125068	0.963
	rs2121266	rs11894252	0.963
	rs2121266	rs11684885	0.963
	rs2121266	rs1867785	0.962
	rs2121266	rs7579899	0.891
	rs2121266	rs11689011	0.929
	rs2346175	rs2346175	1
	rs2346176	rs2346176	1
	rs3768728	rs2278754	0.825
	rs3768728	rs3768727	1
	rs3768730	rs7568285	0.871
	rs3768730	rs7594912	0.871
	rs4145836	rs4145836	1
	rs4953340	rs4953340	1
	rs4953344	rs4953344	1
	rs4953349	rs2121267	0.816
	rs4953349	rs4953345	0.963
	rs4953349	rs1868089	1
	rs4953349	rs12613526	1
	rs4953349	rs10208823	1
	rs4953349	rs12712973	1
	rs4953349	rs4953347	1
	rs4953349	rs6726454	0.86
	rs4953349	rs4145837	0.86
	rs4953353	rs4953353	1
	rs6706003	rs4953352	0.87
	rs6712143	rs6715787	0.849
	rs6712143	rs17035010	0.885
	rs6712143	rs17035013	0.885
	rs6712143	rs6743087	1
	rs6712143	rs1562451	1
	rs6712143	rs7589621	1

	rs6753127	rs6753127	1
	rs7583392	rs11675232	0.934
	rs7583392	rs11692911	0.934
	rs7583392	rs12467821	0.905
	rs7583392	rs11675441	0.966
	rs7583392	rs7606559	0.931
	rs7583392	rs7567582	1
	rs7583392	rs7583088	1
	rs7583392	rs11678465	0.967
	rs7583392	rs7557402	0.964
	rs7583392	rs11694193	0.967
	rs7583392	rs7594278	0.967
	rs7583392	rs7598371	0.966
	rs7583392	rs13006131	1
	rs7583392	rs11690951	0.935
	rs7583392	rs13019268	0.846
	rs7583554	rs2278753	1
	rs7583554	rs10176396	1
	rs7583554	rs4953360	0.962
	rs7583554	rs6755594	0.963
	rs7583554	rs1530632	0.895
	rs7583554	rs3768729	0.855
	rs9679290	rs4952819	0.934
	rs9679290	rs4953343	0.934
	rs9679290	rs4952820	0.841
	rs9679290	rs11125071	0.841
	rs9679290	rs11900910	0.848
	rs9679290	rs4953348	0.965
	rs9679290	rs6758592	0.967
	rs9679290	rs12617313	0.841
	rs9679290	rs11125072	0.841
	rs9973653	rs9973653	1
HIF2B	rs1020398	rs1020398	1
	rs1026016	rs7172914	1
	rs1026016	rs6495497	1
	rs1026016	rs6495498	1
	rs1026016	rs16972076	1
	rs1026016	rs7182420	1

	rs10431813	rs12594558	1
	rs10431813	rs7172073	0.95
	rs10431813	rs3848208	1
	rs10519287	rs7182553	1
	rs10519287	rs1037123	1
	rs10519287	rs1037124	1
	rs10519287	rs1037125	1
	rs10519287	rs3848173	1
	rs11072917	rs11072917	1
	rs1139651	rs7172733	1
	rs1139651	rs1139650	1
	rs1139651	rs4338755	1
	rs11629832	rs11635554	1
	rs11629832	rs10851935	0.915
	rs11633642	rs11633642	1
	rs11635014	rs1020397	0.805
	rs11635014	rs4778790	1
	rs12439449	rs12439449	1
	rs12591286	rs12591286	1
	rs12593988	rs12593988	1
	rs12908321	rs12908321	1
	rs1374213	rs10851934	1
	rs1374213	rs8036233	0.88
	rs1374213	rs3858942	1
	rs1374213	rs6495501	1
	rs1374213	rs4778793	1
	rs1374213	rs7168908	1
	rs1374213	rs11072918	1
	rs1446336	rs1446336	1
	rs1446337	rs1446337	1
	rs16972160	rs16972160	1
	rs1912	rs1912	1
	rs1979026	rs1125522	1
	rs1979026	rs3901896	1
	rs1979026	rs2305145	1
	rs1979026	rs2278706	1
	rs1979026	rs3848174	1
	rs1979026	rs4778595	1

	rs1979026	rs12908010	1
	rs1979026	rs12916797	1
	rs2278709	rs17788150	1
	rs2278709	rs3848175	1
	rs2278709	rs12591546	1
	rs3848170	rs3848170	1
	rs3848206	rs3848206	1
	rs3848207	rs3848207	1
	rs3910982	rs5000770	1
	rs4074666	rs4074666	1
	rs4238521	rs4238521	1
	rs4238522	rs8034361	0.808
	rs4238522	rs4778803	1
	rs4238522	rs4128219	1
	rs4238522	rs7178524	1
	rs4238522	rs7180938	1
	rs4238522	rs11856273	1
	rs4238522	rs7167664	0.884
	rs4238522	rs7169055	0.923
	rs4238522	rs7172094	1
	rs4238522	rs8027410	1
	rs4238522	rs12148133	0.884
	rs4238522	rs3924894	0.883
	rs4303445	rs12594226	0.821
	rs4303445	rs7178902	0.962
	rs4303445	rs11858186	0.819
	rs4423382	rs6495506	1
	rs4423382	rs4778810	0.936
	rs4423382	rs4778605	1
	rs4423382	rs4778811	1
	rs4459508	rs3935130	0.95
	rs4459508	rs7172548	1
	rs4459508	rs7172912	1
	rs4459508	rs6495511	1
	rs4459508	rs7484	1
	rs4609803	rs12594341	0.876
	rs4609803	rs4778597	0.838
	rs4609803	rs4778598	1

	rs4609803	rs4778795	0.838
	rs4609803	rs12905523	0.88
	rs4609803	rs4778796	1
	rs4609803	rs4274393	1
	rs4609803	rs3935990	0.838
	rs4609803	rs12902300	1
	rs4609803	rs4778799	1
	rs4778600	rs12439920	0.907
	rs4778610	rs4072568	0.841
	rs4778610	rs7178743	0.923
	rs4778610	rs4301984	0.923
	rs4778791	rs4778791	1
	rs4778792	rs895442	1
	rs4778792	rs6495502	1
	rs4778800	rs11072922	0.933
	rs4778818	rs4778818	1
	rs4778819	rs4778819	1
	rs6495509	rs6495509	1
	rs7162181	rs11072923	0.848
	rs7162181	rs11856676	1
	rs7162181	rs4238523	1
	rs7162181	rs4778604	1
	rs7162181	rs7168037	0.922
	rs7175825	rs3936145	0.813
	rs7175825	rs4331301	0.961
	rs7175825	rs11072921	0.885
	rs7180520	rs7180520	1
	rs7184010	rs7184010	1
	rs7403201	rs7403201	1
	rs7403418	rs7403418	1
	rs8028295	rs17788120	0.96
	rs8028295	rs8041887	0.96
	rs8033706	rs8033706	1
	rs8034535	rs4238517	0.825
	rs8034535	rs11637295	0.844
	rs8034535	rs4238518	0.838
	rs8039725	rs4778825	0.855
	rs8039725	rs7359233	0.855

	rs8039725	rs8041814	0.855
	rs8039725	rs6495508	0.855
	rs8039725	rs7403013	1
	rs8039725	rs4778829	1
	rs8041336	rs7403706	1
	rs8041826	rs8041826	1
HIF3A	rs10411556	rs10411556	1
	rs11665853	rs11672731	0.837
	rs16980445	rs12461322	1
	rs2072491	rs2072491	1
	rs3764610	rs3764609	0.835
	rs3764610	rs3764611	0.864
	rs757638	rs757638	1
	rs887946	rs887946	1

r^2: correlation coefficient based on the data from the HapMap database (259).

Please note that this table contains more detailed information on the genes and their SNPs listed in the previous Appendix. This table also contains the SNP information for two additional genes investigated in phase II (*HIF2B* and *HIF3A*).

Appendix 8: List of SNPs excluded from phase II because of their high correlation with other SNPs or low minor allele frequencies

Gene	SNP ID	Excluded based on correlations with other SNPs	Major homozygote (n)	Heterozygote (n)	Minor homozygote (n)	Missing genotype (n)	Total genotyped patients (n)	MAF
HIF1A	*rs1951795*	*yes*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
HIF1B	rs10305724	no	CC=473	CT=62	TT=0	1	535	**0.06**
HIF1B	*rs1889740*	*yes*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
HIF1B	*rs10305714*	*yes*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
HIF1B	rs3738483	no	GG=465	GA=69	AA=1	1	535	**0.07**
HIF1B	*rs3768016*	*yes*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
HIF1B	*rs3768013*	*yes*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
HIF1B	*rs2134688*	*yes*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
HIF1B	*rs7532045*	*yes*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
HIF1B	*rs11204735*	*yes*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
HIF2A	*rs10211665*	*yes*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
HIF2A	*rs4952820*	*yes*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
HIF2A	*rs10191091*	*yes*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
HIF2A	*rs17035010*	*yes*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
HIF2A	*rs6544887*	*yes*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
HIF2A	*rs1868088*	*yes*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
HIF2A	*rs4953359*	*yes*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
HIF2A	*rs6743991*	*yes*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
HIF2A	*rs6707241*	*yes*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
HIF2A	*rs1530633*	*yes*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
HIF2A	*rs1374749*	*yes*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
HIF2B	*rs3848208*	*yes*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*

HIF2B	*rs4238518*	*yes*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
HIF2B	*rs6495501*	*yes*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
HIF2B	*rs7168908*	*yes*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
HIF2B	rs16972160	no	TT=492	TC=44	CC=0	0	536	**0.04**
HIF2B	*rs4778795*	*yes*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
HIF2B	*rs12905523*	*yes*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
HIF2B	*rs12594226*	*yes*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
HIF2B	*rs4778799*	*yes*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
HIF2B	*rs3936145*	*yes*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
HIF2B	*rs7178524*	*yes*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
HIF2B	*rs11856273*	*yes*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
HIF2B	*rs7172094*	*yes*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
HIF2B	*rs12148133*	*yes*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
HIF2B	*rs10851935*	*yes*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
HIF2B	*rs4778825*	*yes*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
HIF2B	*rs6495508*	*yes*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
HIF2B	*rs4778829*	*yes*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
HIF2B	*rs4072568*	*yes*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
HIF2B	*rs7172733*	*yes*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
HIF2B	*rs7172548*	*yes*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
HIF2B	*rs6495511*	*yes*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
HIF2B	*rs1139650*	*yes*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
HIF2B	rs1139651	no	GG=471	GA=63	AA=2	0	536	**0.06**
HIF2B	*rs7484*	*yes*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
HIF3A	*rs3764609*	*yes*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
HIF3A	*rs3764611*	*yes*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*

List of SNPs excluded based on missing genotype or minor allele frequency. MAF: minor allele frequency.

Appendix 9: Results of the univariate analysis for overall survival (phase I)

Variables	n	Number of patients in each category	p-value	HR	95% CI	
					Lower	Upper
LOX rs2956540 (GC+GG vs CC)	271	(191 vs 80)	0.571	0.91	0.656	1.262
LOX rs2288393 (CG + CC vs GG)	271	(74 vs 197)	0.946	0.988	0.704	1.387
LOX rs10040971 (CT+CC vs TT)	272	(69 vs 203)	0.854	1.033	0.733	1.456
HIF1A rs2301106 (TC + CC vs TT)	271	(52 vs 219)	0.247	0.787	0.524	1.18
HIF1A rs2301111 (GC+GG vs CC)	272	(82 vs 190)	0.23	0.813	0.579	1.14
HIF1A rs2301113 (CA +CC vs AA)	272	(108 vs 164)	0.721	0.945	0.693	1.289
HIF1A rs11158358 (GC+GG vs CC)	272	(62 vs 210)	0.106	0.731	0.499	1.068
HIF1B rs2228099 (GC + CC vs GG)	272	(157 vs 115)	0.506	0.902	0.666	1.222
HIF1B rs3738483 (AG +AA vs GG)	272	(49 vs 223)	0.241	1.255	0.858	1.836
HIF1B rs11204737 (CT + CC vs TT)	271	(190 vs 81)	0.25	1.22	0.869	1.713
HIF1B rs10847 (AG +AA vs GG)	261	(110 vs 151)	0.775	1.046	0.767	1.427
CXCL12 rs2236534 (GT+TT vs GG)	271	(97 vs 174)	0.563	1.097	0.801	1.504
CXCL12 rs2839688 (CG + CC vs GG)	264	(60 vs 204)	0.741	1.064	0.738	1.534
CXCL12 rs2236533 (GA +AA vs GG)	261	(123 vs 138)	0.69	1.065	0.78	1.455
CXCL12 rs11592974 (TC + CC vs TT)	267	(102 vs 165)	0.531	0.904	0.661	1.238
MIF rs2096525 (TC + CC vs TT)	263	(78 vs 185)	0.77	0.951	0.681	1.329
HIF2A rs2121266 (AC +CC vs AA)	251	(163 vs 88)	0.312	1.191	0.849	1.672
HIF2A rs2346175 (CT + CC vs TT)	185	(125 vs 60)	0.565	0.889	0.595	1.329
HIF2A rs3768730 (GT +TT vs GG)	247	(178 vs 69)	0.92	1.018	0.713	1.454
HIF2A rs4952818 (TC + CC vs TT)	250	(180 vs 70)	0.941	0.987	0.692	1.407
HIF2A rs4953340 (GC + CC vs GG)	248	(158 vs 90)	0.722	0.942	0.677	1.31
HIF2A rs6753127 (CT + TT vs CC)	252	(35 vs 217)	0.208	0.729	0.446	1.193
HIF2A rs7583558 (CT + CC vs TT)	242	(125 vs 117)	0.324	0.85	0.615	1.174
HIF2A rs9679290 (CG + CC vs GG)	252	(174 vs 78)	0.137	0.777	0.557	1.083
HIF2A rs9973653 (GT +TT vs GG)	242	(132 vs 110)	0.731	0.945	0.684	1.306
HIF2A rs10178633 (GA +AA vs GG)	240	(152 vs 88)	0.86	1.031	0.736	1.444
HIF2A rs11687512 (GC+CC vs GG)	250	(19 vs 231)	0.252	0.688	0.362	1.306
HIF2A rs10199201 (CT + CC vs TT)	250	(73 vs 177)	0.849	0.967	0.681	1.371
HIF2A rs12614710 (GT +TT vs GG)	248	(167 vs 81)	0.532	1.115	0.793	1.566
HIF2A rs13019414 (GC + CC vs GG)	251	(156 vs 95)	0.254	0.83	0.603	1.143
HIF2A rs13412887 (CG+GG vs CC)	249	(89 vs 160)	0.101	0.752	0.534	1.057
HIF2A rs1374748 (GT +TT vs GG)	250	(66 vs 184)	0.563	1.11	0.78	1.578
HIF2A rs1562453 (CT + CC vs TT)	248	(68 vs 180)	0.919	1.019	0.714	1.452

HIF2A rs1868087 (AG+GG vs AA)	250	(86 vs 164)	0.823	0.962	0.688	1.346
HIF2A rs1992846 (CT +TT vs CC)	249	(92 vs 157)	0.424	1.141	0.826	1.578
HIF2A rs2034327 (GC + CC vs GG)	243	(177 vs 66)	0.875	0.971	0.676	1.396
HIF2A rs2044456 (GA+GG vs AA)	239	(138 vs 101)	0.612	0.919	0.663	1.273
HIF2A rs2346176 (TC + CC vs TT)	251	(147 vs 104)	0.304	1.185	0.857	1.637
HIF2A rs3768728 (TC + CC vs TT)	251	(62 vs 189)	0.398	1.166	0.816	1.667
HIF2A rs4145836 (AG +AA vs GG)	251	(62 vs 189)	0.414	0.855	0.586	1.246
HIF2A rs4953344 (CT + CC vs TT)	248	(63 vs 185)	0.061	1.391	0.985	1.965
HIF2A rs4953349 (GT +TT vs GG)	245	(175 vs 70)	0.856	1.033	0.726	1.471
HIF2A rs4953353 (GT +TT vs GG)	251	(139 vs 112)	0.402	0.874	0.637	1.198
HIF2A rs6706003 (CG + CC vs GG)	250	(181 vs 69)	0.899	1.023	0.721	1.451
HIF2A rs6712143 (GA+GG vs AA)	251	(115 vs 136)	0.948	0.99	0.722	1.357
HIF2A rs7583392 (AG +AA vs GG)	251	(176 vs 75)	0.766	0.95	0.676	1.335
Sex (male vs female)	280	(150 vs 130)	0.094	1.296	0.957	1.757
Age	280	-	**<.001**	**1.034**	**1.021**	**1.05**
Grade (poorly differentiated/ undifferentiated vs well/moderately differentiated)	276	(42 vs 234)	**0.001**	**1.949**	**1.325**	**2.87**
Histology (mucinous vs non-mucinous)	280	(43 vs 237)	0.603	1.113	0.743	1.669
Lymphatic invasion (+ vs -)	180	(110 vs 70)	**<.001**	**2.472**	**1.648**	**3.71**
Location (rectum vs colon)	280	(57 vs 223)	0.343	1.191	0.83	1.708
Stage	271		**<.001**			
Stage (II vs I)		(94 vs 54)	0.15	1.458	0.873	2.435
Stage (III vs I)		(76 vs 54)	**<.001**	**2.464**	**1.483**	**4.09**
Stage (IV vs I)		(47 vs 54)	**<.001**	**9.647**	**5.629**	**16.5**
MSI-H status (Yes vs No)	280	(34 vs 246)	**0.001**	**0.359**	**0.195**	**0.66**

(+): presence, (-): absence, CI: Confidence Interval, HR: Hazard Ratio, MSI-H: Microsatellite Instability-High, vs: versus. Values that are bolded are statistically significant (p value <0.05). Please note that the numbers might have been rounded to fit the table in this page.

Appendix 10: Results of the univariate analysis for disease specific survival (phase I)

Variables	n	Number of patients in each category	p-value	HR	95% CI	
					Lower	Upper
LOX rs2956540 (GC+GG vs CC)	271	(80 vs 191)	0.649	0.911	0.611	1.359
LOX rs2288393 (CG + CC vs GG)	271	(74 vs 197)	0.995	0.999	0.657	1.519
LOX rs10040971 (CT+CC vs TT)	272	(69 vs 203)	0.827	1.048	0.687	1.6
LOX rs10519694 (CT +TT vs CC)	272	(137 vs 135)	0.284	0.816	0.563	1.183
HIF1A rs2301106 (TC + CC vs TT)	271	(52 vs 219)	0.092	0.626	0.364	1.08
HIF1A rs2301111 (GC+GG vs CC)	272	(82 vs 190)	0.315	0.807	0.531	1.226
HIF1A rs2301113 (CA +CC vs AA)	272	(108 vs 164)	0.377	1.184	0.814	1.721
HIF1A rs11158358 (GC+GG vs CC)	272	(62 vs 210)	0.099	0.665	0.41	1.079
HIF1B rs2228099 (GC + CC vs GG)	272	(157 vs 115)	0.279	0.814	0.561	1.181
HIF1B rs3738483 (AG +AA vs GG)	272	(49 vs 223)	0.241	1.316	0.832	2.082
HIF1B rs11204737 (CT + CC vs TT)	271	(190 vs 81)	0.355	1.218	0.802	1.852
HIF1B rs10847 (AG +AA vs GG)	261	(110 vs 151)	0.8	1.05	0.719	1.534
CXCL12 rs2236534 (GT+TT vs GG)	271	(97 vs 174)	0.231	1.263	0.862	1.85
CXCL12 rs2839688 (CG + CC vs GG)	264	(60 vs 204)	0.949	0.985	0.622	1.561
CXCL12 rs2236533 (GA +AA vs GG)	261	(123 vs 138)	0.493	1.142	0.781	1.672
CXCL12 rs11592974 (TC + CC vs TT)	267	(102 vs 165)	0.11	0.722	0.485	1.077
MIF rs2096525 (TC + CC vs TT)	263	(78 vs 185)	0.962	0.99	0.656	1.493
HIF2A rs2121266 (AC +CC vs AA)	251	(163 vs 88)	0.711	0.927	0.62	1.385
HIF2A rs2346175 (CT + CC vs TT)	185	(125 vs 60)	0.275	0.767	0.476	1.235
HIF2A rs3768730 (GT +TT vs GG)	247	(178 vs 69)	0.638	0.902	0.588	1.385
HIF2A rs4952818 (TC + CC vs TT)	250	(180 vs 70)	0.226	0.773	0.51	1.172
HIF2A rs4953340 (GC + CC vs GG)	248	(158 vs 90)	0.218	0.78	0.525	1.158
HIF2A rs6753127 (CT + TT vs CC)	252	(35 vs 217)	0.743	0.91	0.518	1.6
HIF2A rs7583558 (CT + CC vs TT)	242	(125 vs 117)	0.127	0.735	0.495	1.092
HIF2A rs9679290 (CG + CC vs GG)	252	(174 vs 78)	0.275	0.796	0.529	1.198
HIF2A rs9973653 (GT +TT vs GG)	242	(132 vs 110)	0.092	0.712	0.48	1.057
HIF2A rs10178633 (GA +AA vs GG)	240	(152 vs 88)	0.689	0.92	0.612	1.383
HIF2A rs11687512 (GC+CC vs GG)	250	(19 vs 231)	0.687	0.862	0.419	1.775
HIF2A rs10199201 (CT + CC vs TT)	250	(73 vs 177)	0.737	0.928	0.6	1.435
HIF2A rs11125070 (AT+TT vs AA)	247	(121 vs 126)	0.128	0.737	0.497	1.092
HIF2A rs12614710 (GT +TT vs GG)	248	(81 vs 167)	0.724	1.078	0.711	1.634
HIF2A rs13019414 (GC + CC vs GG)	251	(95 vs 156)	0.153	0.752	0.508	1.111
HIF2A rs13412887 (CG+GG vs CC)	249	(89 vs 160)	0.486	0.864	0.572	1.304
HIF2A rs1374748 (GT +TT vs GG)	250	(66 vs 184)	0.574	1.132	0.735	1.742

HIF2A rs1562453 (CT + CC vs TT)	248	(80 vs 68)	0.675	0.913	0.596	1.398
HIF2A rs1868084 (GC+GG vs CC)	249	(95 vs 154)	0.392	0.838	0.559	1.257
HIF2A rs1868087 (AG+GG vs AA)	250	(86 vs 164)	0.927	1.019	0.677	1.534
HIF2A rs1992846 (CT +TT vs CC)	249	(92 vs 157)	0.184	0.751	0.492	1.146
HIF2A rs2034327 (GC + CC vs GG)	243	(177 vs 66)	0.768	0.935	0.6	1.458
HIF2A rs2044456 (GA+GG vs AA)	239	(138 vs 101)	0.088	0.707	0.475	1.053
HIF2A rs2346176 (TC + CC vs TT)	251	(147 vs 104)	0.896	0.974	0.658	1.442
HIF2A rs3768728 (TC + CC vs TT)	251	(62 vs 189)	0.543	1.146	0.738	1.78
HIF2A rs4145836 (AG +AA vs GG)	251	(62 vs 189)	0.94	1.017	0.652	1.588
HIF2A rs4953344 (CT + CC vs TT)	248	(63 vs 185)	0.836	1.049	0.667	1.649
HIF2A rs4953349 (GT +TT vs GG)	245	(70 vs 175)	0.946	1.015	0.66	1.562
HIF2A rs4953353 (GT +TT vs GG)	251	(139 vs 112)	0.381	0.84	0.57	1.24
HIF2A rs6706003 (CG + CC vs GG)	250	(181 vs 69)	0.918	1.023	0.665	1.573
HIF2A rs6712143 (GA+GG vs AA)	251	(115 vs 136)	0.287	0.807	0.544	1.198
HIF2A rs7583392 (AG +AA vs GG)	251	(176 vs 75)	0.356	0.825	0.548	1.242
Sex (male vs female)	280	(150 vs 130)	0.246	1.248	0.858	1.814
Age	280	-	**0.03**	**1.02**	**1.002**	**1.032**
Grade (poorly/ undifferentiated vs well/moderately differentiated)	276	(42 vs 234)	**<.001**	**2.33**	**1.493**	**3.647**
Histology (mucinous vs non-mucinous)	280	(43 vs 237)	0.774	0.926	0.545	1.571
Lymphatic invasion (+ vs -)	180	(110 vs 70)	**<.001**	**3.02**	**1.821**	**4.998**
Location (rectum vs colon)	280	(57 vs 223)	0.52	1.157	0.742	1.805
Stage	271		**<.001**			
Stage (II vs I)		(94 vs 54)	0.067	2.324	0.942	5.731
Stage (III vs I)		(76 vs 54)	**<.001**	**6.07**	**2.562**	**14.36**
Stage (IV vs I)		(47 vs 54)	**<.001**	**28.1**	**11.78**	**66.92**
MSI-H status (Yes vs No)	280	(34 vs 246)	**<.001**	**0.2**	**0.075**	**0.551**

(+): presence, (-): absence, CI: Confidence Interval, HR: Hazard Ratio, MSI-H: Microsatellite Instability-High, vs: versus. Values that are bolded are statistically significant (p value <0.05). Please note that the numbers might have been rounded to fit the table in this page.

Appendix 11: Results of the univariate analysis for disease free survival (phase I)

Variables	n	Number of patients in each category	p-value	HR	95% CI	
					Lower	Upper
LOX rs2956540 (GC+GG vs CC)	271	(191 vs 80)	0.225	0.824	0.603	1.126
LOX rs2288393 (CG + CC vs GG)	271	(74 vs 197)	0.982	1.004	0.725	1.39
LOX rs10040971 (CT+CC vs TT)	272	(69 vs 203)	0.995	1.001	0.718	1.395
HIF1A rs2301106 (TC + CC vs TT)	271	(52 vs 219)	0.588	0.902	0.621	1.31
HIF1A rs2301111 (GC+GG vs CC)	272	(82 vs 190)	0.334	0.853	0.618	1.177
HIF1A rs2301113 (CA +CC vs AA)	272	(108 vs 164)	0.646	0.932	0.691	1.258
HIF1A rs11158358 (GC+GG vs CC)	272	(62 vs 210)	0.291	0.826	0.58	1.177
HIF1B rs2228099 (GC + CC vs GG)	272	(157 vs 115)	0.402	0.882	0.658	1.183
HIF1B rs3738483 (AG +AA vs GG)	272	(49 vs 223)	0.253	1.24	0.857	1.794
HIF1B rs11204737 (CT + CC vs TT)	271	(190 vs 81)	0.261	1.206	0.87	1.671
HIF1B rs10847 (AG +AA vs GG)	261	(110 vs 151)	0.847	1.03	0.763	1.391
CXCL12 rs2236534 (GT+TT vs GG)	271	(97 vs 174)	0.553	1.096	0.809	1.485
CXCL12 rs2839688 (CG + CC vs GG)	264	(60 vs 204)	0.776	0.949	0.663	1.359
CXCL12 rs2236533 (GA +AA vs GG)	261	(123 vs 138)	0.52	1.104	0.817	1.49
CXCL12 rs11592974 (TC + CC vs TT)	267	(102 vs 165)	0.8	0.962	0.711	1.301
MIF rs2096525 (TC + CC vs TT)	263	(78 vs 185)	0.339	0.852	0.613	1.183
HIF2A rs2121266 (AC +CC vs AA)	251	(163 vs 88)	0.303	1.187	0.857	1.645
HIF2A rs2346175 (CT + CC vs TT)	185	(125 vs 60)	0.989	1.003	0.679	1.481
HIF2A rs3768730 (GT +TT vs GG)	247	(178 vs 69)	0.931	0.985	0.699	1.389
HIF2A rs4952818 (TC + CC vs TT)	250	(180 vs 70)	0.96	1.009	0.716	1.421
HIF2A rs4953340 (GC + CC vs GG)	248	(158 vs 90)	0.91	0.982	0.714	1.349
HIF2A rs6753127 (CT + TT vs CC)	252	(35 vs 217)	0.321	0.793	0.502	1.254
HIF2A rs7583558 (CT + CC vs TT)	242	(125 vs 117)	0.503	0.899	0.66	1.226
HIF2A rs9679290 (CG + CC vs GG)	252	(174 vs 78)	0.123	0.777	0.564	1.071
HIF2A rs9973653 (GT +TT vs GG)	242	(132 vs 110)	0.908	1.018	0.746	1.391
HIF2A rs10178633 (GA +AA vs GG)	240	(152 vs 88)	0.835	1.035	0.747	1.434
HIF2A rs11687512 (GC+CC vs GG)	250	(19 vs 231)	0.184	0.648	0.342	1.229
HIF2A rs10199201 (CT + CC vs TT)	250	(73 vs 177)	0.734	0.943	0.674	1.32
HIF2A rs12614710 (GT +TT vs GG)	248	(167 vs 81)	0.378	1.16	0.834	1.612
HIF2A rs13019414 (GC + CC vs GG)	251	(156 vs 95)	0.318	0.854	0.627	1.164
HIF2A rs13412887 (CG+GG vs CC)	249	(89 vs 160)	0.141	0.782	0.564	1.085
HIF2A rs1374748 (GT +TT vs GG)	250	(66 vs 184)	0.842	1.035	0.735	1.458
HIF2A rs1562453 (CT + CC vs TT)	248	(68 vs 180)	0.825	0.962	0.683	1.356
HIF2A rs1868084 (GC+GG vs CC)	249	(95 vs 154)	0.122	0.776	0.563	1.07

HIF2A rs1868087 (AG+GG vs AA)	250	(86 vs 164)	0.638	0.925	0.67	1.278
HIF2A rs1992846 (CT +TT vs CC)	249	(92 vs 157)	0.541	1.102	0.807	1.505
HIF2A rs2034327 (GC + CC vs GG)	243	(177 vs 66)	0.578	0.906	0.64	1.282
HIF2A rs2044456 (GA+GG vs AA)	239	(138 vs 101)	0.691	0.938	0.684	1.286
HIF2A rs2346176 (TC + CC vs TT)	251	(147 vs 104)	0.422	1.136	0.832	1.551
HIF2A rs3768728 (TC + CC vs TT)	251	(62 vs 189)	0.654	1.083	0.765	1.532
HIF2A rs4145836 (AG +AA vs GG)	251	(62 vs 189)	0.566	0.9	0.629	1.289
HIF2A rs4953344 (CT + CC vs TT)	248	(63 vs 185)	0.257	1.218	0.866	1.713
HIF2A rs4953349 (GT +TT vs GG)	245	(175 vs 70)	0.579	1.102	0.782	1.554
HIF2A rs4953353 (GT +TT vs GG)	251	(139 vs 112)	0.486	0.897	0.662	1.217
HIF2A rs6706003 (CG + CC vs GG)	250	(181 vs 69)	0.832	1.037	0.74	1.454
HIF2A rs6712143 (GA+GG vs AA)	251	(115 vs 136)	0.935	0.987	0.729	1.338
HIF2A rs7583392 (AG +AA vs GG)	251	(176 vs 75)	0.46	0.884	0.637	1.227
Sex (male vs female)	280	(150 vs 130)	0.23	1.196	0.893	1.601
Age	280	-	**<.001**	**1.03**	**1.012**	**1.037**
Grade (poorly/ undifferentiated vs well/moderately differentiated)	276	(42 vs 234)	**0.008**	**1.67**	**1.141**	**2.432**
Histology (mucinous vs non-mucinous)	280	(43 vs 237)	0.661	1.092	0.738	1.614
Lymphatic invasion (+ vs -)	180	(110 vs 70)	**<.001**	**2.08**	**1.414**	**3.051**
Location (rectum vs colon)	280	(57 vs 223)	0.273	1.213	0.859	1.711
Stage	271		**<.001**			
Stage (II vs I)		(94 vs 54)	**0.04**	**1.7**	**1.024**	**2.806**
Stage (III vs I)		(76 vs 54)	**<.001**	**2.88**	**1.737**	**4.763**
Stage (IV vs I)		(47 vs 54)	**<.001**	**106**	**48.63**	**232.9**
MSI-H status (Yes vs No)	280	(34 vs 246)	**<.001**	**0.35**	**0.195**	**0.63**

(+): presence, (-): absence, CI: Confidence Interval, HR: Hazard Ratio, MSI-H: Microsatellite Instability-High, vs: versus. Values that are bolded are statistically significant (p value <0.05). Please note that the numbers might have been rounded to fit the table in this page.

Appendix 12: Results of the univariate analysis for overall survival (phase II)

Variables	n	Number of patients in each category	p-value	HR	95% CI	
					Lower	Upper
HIF1B rs10847 (AG+AAvsGG)	535	(255 vs 280)	0.41	0.885	0.66	1.185
HIF1B rs2228099 (GC+CCvsGG)	535	(306 vs 229)	0.661	1.068	0.795	1.435
HIF1B rs11204737 (TC+CCvsTT)	535	(349 vs 186)	0.485	1.116	0.82	1.52
HIF2A rs1867785 (AG+AAvsGG)	535	(323 vs 212)	0.826	0.967	0.719	1.301
HIF2A rs2121266 (AC+CCvsAA)	534	(331 vs 203)	0.362	0.871	0.648	1.171
HIF2A rs17034950 (AG+AAvsGG)	535	(247 vs 288)	0.833	0.969	0.724	1.298
HIF2A rs9973653 (TG+TTvsGG)	535	(257 vs 278)	0.846	0.972	0.726	1.3
HIF2A rs4953342 (AG+GGvsAA)	535	(264 vs 271)	0.522	1.1	0.822	1.471
HIF2A rs1868089 (TC+CCvsTT)	535	(396 vs 139)	0.561	0.906	0.649	1.264
HIF2A rs6758592 (TC+CCvsTT)	535	(369 vs 166)	0.24	1.212	0.879	1.671
HIF2A rs12614710 (TG+TTvsGG)	535	(386 vs 149)	0.836	1.035	0.744	1.44
HIF2A rs4953353 (TG+TTvsGG)	535	(306 vs 229)	0.183	1.224	0.909	1.648
HIF2A rs2346175 (TC+CCvsTT)	535	(377 vs 158)	0.148	1.276	0.917	1.775
HIF2A rs6756667 (AG+GGvsAA)	535	(391 vs 144)	0.993	1.001	0.723	1.387
HIF2A rs1868086 (GT+TTvsGG)	534	(343 vs 191)	0.856	1.028	0.759	1.394
HIF2A rs6712143 (AG+GGvsAA)	535	(291 vs 244)	0.982	0.997	0.744	1.334
HIF2A rs10176396 (CT+TTvsCC)	535	(253 vs 282)	0.573	1.088	0.812	1.456
HIF2A rs1374748 (GT+TTvsGG)	535	(406 vs 129)	0.14	1.281	0.922	1.779
HIF2A rs3768728 (TC+CCvsTT)	535	(401 vs 134)	0.846	0.967	0.69	1.355
HIF2A rs2346176 (TC+CCvsTT)	535	(328 vs 207)	0.755	0.954	0.708	1.285
HIF2A rs10178633 (GA+AAvsGG)	535	(362 vs 173)	0.933	0.987	0.723	1.346
HIF2A rs7594912 (AC+CCvsAA)	535	(370 vs 165)	0.481	1.122	0.815	1.542
HIF2A rs7557402 (GC+CCvsGG)	535	(374 vs 161)	0.888	0.978	0.713	1.339
HIF2A rs7571218 (GA+AAvsGG)	535	(330 vs 205)	0.533	1.1	0.815	1.487
LOX rs3792802 (GA+AAvsGG)	534	(169 vs 365)	0.216	1.212	0.894	1.644
LOX rs1800449 (GA+AAvsGG)	535	(172 vs 363)	0.209	1.214	0.897	1.643
CXCL12 rs2839695 (TC+CCvsTT)	535	(191 vs 344)	0.136	0.789	0.577	1.078
HIF1A rs2301106 (TC+CCvsTT)	535	(130 vs 405)	0.966	0.993	0.709	1.391
HIF1A rs12434438 (AG+GGvsAA)	535	(206 vs 329)	0.29	1.172	0.874	1.572
HIF1A rs2057482 (CT+TTvsCC)	535	(144 vs 391)	0.831	0.965	0.695	1.339
HIF2B rs12591286 (AG+AAvsGG)	535	(320 vs 215)	0.272	0.848	0.633	1.137
HIF2B rs8041826 (AG+GGvsAA)	535	(160 vs 375)	0.848	1.031	0.754	1.411
HIF2B rs1446337 (AG+AAvsGG)	535	(229 vs 306)	0.062	0.751	0.557	1.014
HIF2B rs3848206 (AG+AAvsGG)	534	(402 vs 132)	0.659	1.08	0.767	1.521

HIF2B rs3848207 (AG+AAvsGG)	535	(131 vs 404)	0.306	1.187	0.855	1.647
HIF2B rs7172914 (CT+TTvsCC)	535	(193 vs 342)	0.134	1.255	0.932	1.688
HIF2B rs10431813 (GA+AAvsGG)	533	(242 vs 291)	0.28	0.85	0.632	1.142
HIF2B rs3910982 (TG+TTvsGG)	535	(274 vs 261)	0.159	0.811	0.606	1.085
HIF2B rs11635014 (CT+TTvsCC)	535	(348 vs 187)	0.76	1.049	0.773	1.424
HIF2B rs1020398 (TC+CCvsTT)	535	(257 vs 278)	0.487	0.901	0.673	1.207
HIF2B rs11633642 (GA+AAvsGG)	535	(289 vs 246)	0.27	0.849	0.635	1.136
HIF2B rs7184010 (CT+TTvsCC)	534	(244 vs 290)	0.394	1.135	0.848	1.52
HIF2B rs3848170 (CT+TTvsCC)	535	(110 vs 425)	0.648	1.087	0.76	1.555
HIF2B rs4778791 (GA+AAvsGG)	535	(394 vs 141)	0.203	0.813	0.592	1.118
HIF2B rs8034535 (AG+GGvsAA)	535	(147 vs 388)	0.515	1.112	0.808	1.531
HIF2B rs895442 (CT+TTvsCC)	535	(107 vs 428)	0.977	1.005	0.698	1.448
HIF2B rs1037124 (GA+AAvsGG)	535	(153 vs 382)	0.494	1.116	0.815	1.526
HIF2B rs1374213 (TC+CCvsTT)	534	(374 vs 160)	0.844	1.033	0.751	1.42
HIF2B rs3901896 (CT+TTvsCC)	535	(351 vs 184)	0.292	0.851	0.631	1.149
HIF2B rs2278709 (CT+TTvsCC)	535	(266 vs 269)	0.339	0.867	0.648	1.161
HIF2B rs8028295 (CT+TTvsCC)	535	(372 vs 163)	0.608	0.922	0.675	1.259
HIF2B rs4609803 (GA+AAvsGG)	535	(242 vs 293)	0.146	1.241	0.928	1.659
HIF2B rs4778800 (TG+TTvsGG)	535	(211 vs 324)	0.914	1.017	0.756	1.367
HIF2B rs7178902 (TC+CCvsTT)	535	(362 vs 173)	0.853	1.03	0.755	1.405
HIF2B rs4238521 (GA+AAvsGG)	535	(154 vs 381)	0.446	1.128	0.827	1.539
HIF2B rs4331301 (GA+AAvsGG)	535	(328 vs 207)	0.874	1.025	0.759	1.383
HIF2B rs4778600 (TG+TTvsGG)	534	(143 vs 391)	0.964	0.993	0.713	1.381
HIF2B rs11856676 (CT+TTvsCC)	535	(373 vs 162)	0.704	0.941	0.687	1.289
HIF2B rs4238522 (TC+CCvsTT)	535	(339 vs 196)	0.845	1.031	0.759	1.4
HIF2B rs4074666 (CT+TTvsCC)	535	(389 vs 146)	0.306	0.845	0.612	1.167
HIF2B rs11635554 (AG+GGvsAA)	535	(253 vs 282)	0.821	1.034	0.773	1.385
HIF2B rs4778818 (AG+GGvsAA)	535	(239 vs 296)	0.952	1.009	0.753	1.352
HIF2B rs4778819 (CT+TTvsCC)	535	(261 vs 274)	0.867	0.975	0.729	1.305
HIF2B rs7403706 (TC+CCvsTT)	535	(188 vs 347)	0.565	1.092	0.808	1.477
HIF2B rs6495509 (GA+AAvsGG)	535	(204 vs 331)	0.608	1.081	0.802	1.457
HIF2B rs8039725 (AG+GGvsAA)	535	(187 vs 348)	0.729	0.947	0.697	1.287
HIF2B rs8033706 (TC+CCvsTT)	479	(338 vs 141)	0.085	0.75	0.54	1.04
HIF2B rs4301984 (GA+AAvsGG)	535	(207 vs 328)	0.97	1.006	0.746	1.356
HIF2B rs4459508 (GA+AAvsGG)	535	(256 vs 279)	0.849	0.972	0.726	1.301
HIF3A rs2072491 (CT+TTvsCC)	535	(112 vs 423)	0.123	0.737	0.501	1.086
HIF3A rs757638 (GA+AAvsGG)	535	(148 vs 387)	0.482	0.888	0.636	1.238
HIF3A rs12461322 (GA+AAvsGG)	535	(108 vs 427)	0.336	1.188	0.836	1.687
HIF3A rs887946 (AG+GGvsAA)	535	(280 vs 255)	0.76	1.047	0.782	1.401

HIF3A rs3764610 (CT+TTvsCC)	535	(161 vs 374)	0.157	1.249	0.918	1.699
HIF3A rs11665853 (AG+GGvsAA)	535	(105 vs 430)	0.817	1.044	0.727	1.498
Lymphatic invasion (+ vs -)	493	(179 vs 314)	**0.003**	**1.58**	**1.169**	**2.131**
Sex (male vs female)	535	(329 vs 206)	**0.013**	**1.49**	**1.088**	**2.035**
Histology (mucinous vs non-mucinous)	535	(61 vs 474)	0.893	0.969	0.609	1.542
Location (rectum vs colon)	535	(179 vs 356)	0.184	1.225	0.908	1.652
Stage	535		**<0.001**			
Stage (II vs I)		(206 vs 97)	0.183	1.446	0.84	2.491
Stage (III vs I)		(179 vs 97)	**0.004**	**2.19**	**1.281**	**3.727**
Stage (IV vs I)		(53 vs 97)	**<.001**	**10.1**	**5.73**	**17.7**
Grade (poorly differentiated/ undifferentiated vs well/moderately differentiated)	531	(38 vs 493)	0.67	0.876	0.476	1.612
Vascular invasion (+ vs -)	496	(171 vs 325)	**<0.001**	**1.72**	**1.271**	**2.328**
Familial risk (high/moderate vs low)	535	(281 vs 254)	0.663	1.067	0.797	1.429
MSI-H status (Yes vs No)	513	(48 vs 455)	**<0.001**	**0.23**	**0.103**	**0.526**
BRAF mutation status (+ vs -)	484	(49 vs 435)	0.401	0.796	0.467	1.356
Age	535	-	0.429	1.007	0.99	1.023

(+): presence, (-): absence, CI: Confidence Interval, HR: Hazard Ratio, MSI-H: Microsatellite Instability-High. Values that are bolded are statistically significant (p value <0.05). Please note that the numbers might have been rounded to fit the table in this page.

Appendix 13: Results of the univariate analysis for disease free survival (phase II)

Variables	n	Number of patients in each category	p-value	HR	95% CI	
					Lower	Upper
HIF1B rs10847 (AG+AAvsGG)	534	(255 vs 279)	0.378	0.9	0.68	1.16
HIF1B rs2228099 (GC+CCvsGG)	534	(305 vs 229)	0.995	1	0.76	1.31
HIF1B rs11204737 (TC+CCvsTT)	534	(348 vs 186)	0.781	1	0.78	1.38
HIF2A rs1867785 (AG+AAvsGG)	534	(322 vs 212)	0.771	1	0.79	1.37
HIF2A rs2121266 (AC+CCvsAA)	533	(330 vs 203)	0.719	1	0.72	1.25
HIF2A rs17034950 (AG+AAvsGG)	534	(246 vs 288)	0.742	1.1	0.8	1.37
HIF2A rs9973653 (TG+TTvsGG)	534	(256 vs 278)	0.711	1.1	0.8	1.38
HIF2A rs4953342 (AG+GGvsAA)	534	(263 vs 271)	0.278	1.2	0.89	1.52
HIF2A rs1868089 (TC+CCvsTT)	534	(395 vs 139)	0.624	0.9	0.68	1.26
HIF2A rs6758592 (TC+CCvsTT)	534	(368 vs 166)	0.343	1.2	0.86	1.55
HIF2A rs12614710 (TG+TTvsGG)	534	(385 vs 149)	0.67	1.1	0.79	1.45
HIF2A rs4953353 (TG+TTvsGG)	534	(306 vs 228)	0.365	1.1	0.86	1.49
HIF2A rs2346175 (TC+CCvsTT)	534	(376 vs 158)	0.349	1.2	0.85	1.56
HIF2A rs6756667 (AG+GGvsAA)	534	(390 vs 144)	0.629	0.9	0.69	1.25
HIF2A rs1868086 (GT+TTvsGG)	533	(191 vs 342)	0.678	0.9	0.71	1.25
HIF2A rs6712143 (AG+GGvsAA)	534	(244 vs 290)	0.775	1	0.73	1.26
HIF2A rs10176396 (CT+TTvsCC)	534	(282 vs 252)	0.846	1	0.74	1.27
HIF2A rs1374748 (GT+TTvsGG)	534	(128 vs 406)	0.384	1.2	0.84	1.56
HIF2A rs3768728 (TC+CCvsTT)	534	(134 vs 400)	0.702	0.9	0.69	1.29
HIF2A rs2346176 (TC+CCvsTT)	534	(327 vs 207)	0.706	1	0.72	1.25
HIF2A rs10178633 (GA+AAvsGG)	534	(362 vs 172)	0.717	1	0.71	1.26
HIF2A rs7594912 (AC+CCvsAA)	534	(369 vs 165)	0.992	1	0.75	1.34
HIF2A rs7557402 (GC+CCvsGG)	534	(373 vs 161)	0.417	0.9	0.67	1.18
HIF2A rs7571218 (GA+AAvsGG)	534	(329 vs 205)	0.908	1	0.75	1.3
LOX rs3792802 (GA+AAvsGG)	533	(169 vs 364)	0.166	1.2	0.92	1.62
LOX rs1800449 (GA+AAvsGG)	534	(172 vs 362)	0.241	1.2	0.89	1.57
CXCL12 rs2839695 (TC+CCvsTT)	534	(191 vs 343)	0.268	0.9	0.64	1.13
HIF1A rs2301106 (TC+CCvsTT)	534	(130 vs 404)	0.628	1.1	0.79	1.47
HIF1A rs12434438 (AG+GGvsAA)	534	(205 vs 329)	0.066	1.3	0.98	1.69
HIF1A rs2057482 (CT+TTvsCC)	534	(144 vs 390)	0.708	1.1	0.79	1.43
HIF2B rs12591286 (AG+AAvsGG)	534	(320 vs 214)	0.339	0.9	0.67	1.15
HIF2B rs8041826 (AG+GGvsAA)	534	(160 vs 374)	0.358	1.1	0.86	1.53
HIF2B rs1446337 (AG+AAvsGG)	534	(229 vs 305)	0.108	0.8	0.61	1.05
HIF2B rs3848206 (AG+AAvsGG)	533	(401 vs 132)	0.596	1.1	0.79	1.5
HIF2B rs3848207 (AG+AAvsGG)	534	(131 vs 403)	0.244	1.2	0.88	1.62

HIF2B rs7172914 (CT+TTvsCC)	534	(192 vs 342)	0.127	1.2	0.94	1.63
HIF2B rs10431813 (GA+AAvsGG)	532	(241 vs 291)	0.511	0.9	0.69	1.2
HIF2B rs3910982 (TG+TTvsGG)	534	(274 vs 260)	0.274	0.9	0.66	1.13
HIF2B rs11635014 (CT+TTvsCC)	534	(347 vs 187)	0.217	1.2	0.9	1.6
HIF2B rs1020398 (TC+CCvsTT)	534	(256 vs 278)	0.974	1	0.77	1.32
HIF2B rs11633642 (GA+AAvsGG)	534	(289 vs 245)	0.126	0.8	0.62	1.06
HIF2B rs7184010 (CT+TTvsCC)	533	(243 vs 290)	0.666	0.9	0.72	1.24
HIF2B rs3848170 (CT+TTvsCC)	534	(110 vs 424)	0.956	1	0.72	1.41
HIF2B rs4778791 (GA+AAvsGG)	534	(394 vs 140)	0.244	0.8	0.62	1.13
HIF2B rs8034535 (AG+GGvsAA)	534	(146 vs 388)	0.34	1.2	0.86	1.56
HIF2B rs895442 (CT+TTvsCC)	534	(107 vs 427)	0.761	1.1	0.75	1.47
HIF2B rs1037124 (GA+AAvsGG)	534	(153 vs 381)	0.791	1	0.71	1.29
HIF2B rs1374213 (TC+CCvsTT)	533	(376 vs 166)	0.608	1.1	0.8	1.45
HIF2B rs3901896 (CT+TTvsCC)	534	(351 vs 183)	0.431	0.9	0.68	1.18
HIF2B rs2278709 (CT+TTvsCC)	534	(265 vs 269)	0.825	1	0.74	1.27
HIF2B rs8028295 (CT+TTvsCC)	534	(371 vs 163)	0.576	0.9	0.69	1.23
HIF2B rs4609803 (GA+AAvsGG)	534	(242 vs 292)	0.301	1.2	0.88	1.51
HIF2B rs4778800 (TG+TTvsGG)	534	(211 vs 323)	0.553	0.9	0.7	1.21
HIF2B rs7178902 (TC+CCvsTT)	534	(361 vs 173)	0.497	1.1	0.83	1.48
HIF2B rs4238521 (GA+AAvsGG)	534	(154 vs 386)	0.458	1.1	0.84	1.49
HIF2B rs4331301 (GA+AAvsGG)	534	(327 vs 207)	0.403	1.1	0.85	1.49
HIF2B rs4778600 (TG+TTvsGG)	533	(143 vs 390)	0.387	0.9	0.64	1.19
HIF2B rs11856676 (CT+TTvsCC)	534	(372 vs 162)	0.441	1.1	0.83	1.51
HIF2B rs4238522 (TC+CCvsTT)	534	(338 vs 196)	0.098	1.3	0.96	1.7
HIF2B rs4074666 (CT+TTvsCC)	534	(388 vs 146)	0.805	1	0.71	1.3
HIF2B rs11635554 (AG+GGvsAA)	534	(253 vs 281)	0.888	1	0.75	1.29
HIF2B rs4778818 (AG+GGvsAA)	534	(238 vs 296)	0.765	1	0.73	1.26
HIF2B rs4778819 (CT+TTvsCC)	534	(261 vs 273)	0.971	1	0.77	1.32
HIF2B rs7403706 (TC+CCvsTT)	478	(187 vs 347)	0.852	1	0.73	1.29
HIF2B rs6495509 (GA+AAvsGG)	534	(204 vs 330)	0.686	0.9	0.71	1.25
HIF2B rs8039725 (AG+GGvsAA)	534	(187 vs 347)	0.208	0.8	0.62	1.11
***HIF2B* rs8033706 (TC+CCvsTT)**	534	(338 vs 140)	**0.02**	**0.7**	**0.52**	**0.95**
HIF2B rs4301984 (GA+AAvsGG)	534	(207 vs 327)	0.443	0.9	0.68	1.18
HIF2B rs4459508 (GA+AAvsGG)	534	(256 vs 278)	0.303	0.9	0.66	1.14
HIF3A rs2072491 (CT+TTvsCC)	534	(111 vs 423)	0.092	0.7	0.51	1.05
HIF3A rs757638 (GA+AAvsGG)	534	(147 vs 387)	0.408	0.9	0.65	1.19
HIF3A rs12461322 (GA+AAvsGG)	534	(108 vs 426)	0.501	1.1	0.8	1.56
HIF3A rs887946 (AG+GGvsAA)	534	(279 vs 255)	0.942	1	0.77	1.32
HIF3A rs3764610 (CT+TTvsCC)	534	(160 vs 374)	0.221	1.2	0.9	1.6

HIF3A rs11665853 (AG+GGvsAA)	534	(105 vs 429)	0.979	1	0.71	1.4
Sex (male vs female)	534	(328 vs 206)	**0.01**	**1.5**	**1.1**	**1.96**
Histology (mucinous vs non-mucinous)	534	(61 vs 473)	0.778	0.9	0.61	1.45
Location (rectum vs colon)	534	(179 vs 355)	**0.03**	**1.4**	**1.04**	**1.8**
Stage	534		**<0.001**			
Stage (II vs I)		(206 vs 96)	0.258	1.3	0.82	2.11
Stage (III vs I)		(179 vs 96)	**0.002**	**2.1**	**1.32**	**3.35**
Stage (IV vs I)		(53 vs 96)	**<0.001**	**5.7**	**3.43**	**9.43**
Vascular invasion (+ vs -)	498	(171 vs 324)	**0.001**	**1.6**	**1.24**	**2.17**
Lymphatic invasion (+ vs -)	492	(179 vs 313)	**<0.001**	**1.5**	**1.16**	**2.02**
Familial risk (high/moderate vs low)	534	(280 vs 254)	0.302	1.2	0.88	1.51
MSI-H status (Yes vs No)	512	(57 vs 455)	**0.001**	**0.4**	**0.18**	**0.65**
BRAF mutation status (+ vs -)	484	(49 vs 435)	0.654	0.9	0.56	1.44
Age	534	-	0.882	1	0.99	1.02
Grade (poorly differentiated/ undifferentiated vs well/moderately differentiated)	530	(38 vs 492)	0.483	0.81	0.453	1.454

(+): presence, (-): absence, CI: Confidence Interval, HR: Hazard Ratio, MSI-H: Microsatellite Instability-High, vs: versus. Values that are bolded are statistically significant (p value <0.05). Please note that the numbers might have been rounded to fit the table in this page.

Appendix 14: List of phase I and phase II SNPs that are identical or highly correlated

phase I SNPs	phase I SNPs analyzed in phase II	SNP investigated in phase II that are highly correlated with SNPs in phase I	r^2
LOX rs10040971		rs3792802	0.831
LOX rs2288393		rs1800449	0.944
HIF1A rs2301106	rs2301106		
HIF1A rs2301111		rs12434438	1
HIF1A rs11158358		rs2057482	0.929
HIF1B rs2228099	rs2228099		
HIF1B rs11204737	rs11204737		
HIF1B rs10847	rs10847		
CXCL12 rs11592974		rs2839695	1
HIF2A rs2121266	rs2121266		
HIF2A rs2346175	rs2346175		
HIF2A rs3768730		rs7594912	0.871
HIF2A rs9679290		rs6758592	0.967
HIF2A rs9973653	rs9973653		
HIF2A rs10178633	rs10178633		
HIF2A rs11125070		rs4953342	0.913
HIF2A rs12614710	rs12614710		
HIF2A rs13019414		rs7571218	0.933
HIF2A rs1562453		rs6756667	0.967
HIF2A rs1868087		rs1868086	1
HIF2A rs2044456		rs17034950	0.921
HIF2A rs2346176	rs2346176		
HIF2A rs3768728	rs3768728		
HIF2A rs4953349		rs1868089	1
HIF2A rs4953353	rs4953353		
HIF2A rs6706003		rs4953352	0.87
HIF2A rs6712143	rs6712143		
HIF2A rs7583392		rs7557402	0.964

SNPs that are in green are those SNPs that are common between phase I and II. SNPs that are in red are SNPs that are investigated in phase II and are highly correlated with corresponding SNPs in phase I. r^2 values are the correlation coefficient between the corresponding SNPs.

Appendix 15: Pooled univariate analysis results for the polymorphisms that were common between phase I and phase II (overall survival)

Variables	Number of patients in each category	p-value	HR	95% CI for HR	
				Lower	Upper
HIF1A rs2301106	(182 vs 624)	0.238	0.856	0.662	1.108
HIF1B rs2228099	(463 vs 344)	0.944	0.992	0.803	1.226
HIF1B rs11204737	(539 vs 267)	0.126	1.194	0.951	1.499
HIF1B rs10847	(365 vs 431)	0.484	0.927	0.749	1.146
HIF2A rs2121266	(494 vs 291)	0.837	1.024	0.82	1.278
HIF2A rs2346175	(502 vs 218)	0.534	1.084	0.841	1.398
HIF2A rs9973653	(389 vs 388)	0.975	1.003	0.809	1.245
HIF2A rs10178633	(514 vs 261)	0.84	0.977	0.777	1.227
HIF2A rs12614710	(553 vs 230)	0.779	1.034	0.817	1.31
HIF2A rs2346176	(475 vs 311)	0.726	1.04	0.835	1.295
HIF2A rs3768728	(196 vs 590)	0.675	1.054	0.825	1.346
HIF2A rs6712143	(359 vs 427)	0.99	0.999	0.806	1.237
HIF2A rs4953353	(445 vs 341)	0.832	1.02	0.825	1.269
Sex (male vs female)	(479 vs 336)	**0.015**	**1.309**	**1.054**	**1.626**
Age	-	**<.001**	**1.03**	**1.024**	**1.045**
Grade (poorly differentiated/ undifferentiated vs well/moderately differentiated)	(80 vs 727)	**0.002**	**1.65**	**1.195**	**2.263**
Histology (mucinous vs non-mucinous)	(104 vs 711)	0.501	1.11	0.819	1.505
Lymphatic invasion (+ vs -)	(289 vs 384)	**<.001**	**2.1**	**1.663**	**2.652**
Location (rectum vs colon)	(236 vs 579)	0.447	1.092	0.871	1.37
Stage		**<.001**			
Stage (II vs I)	(300 vs 151)	0.061	1.426	0.983	2.067
Stage (III vs I)	(255 vs 151)	**<.001**	**2.19**	**1.516**	**3.155**
Stage (IV vs I)	(100 vs 151)	**<.001**	**9.61**	**6.528**	**14.136**
MSI-H status (Yes vs No)	(92 vs 701)	**<.001**	**0.32**	**0.194**	**0.514**

(+): presence, (-): absence, CI: Confidence Interval, HR: Hazard Ratio, MSI-H: Microsatellite Instability-Highvs: versus. Values that are bolded are statistically significant (p value <0.05). Please note that the numbers might have been rounded to fit the table in this page.

Appendix 16: Pooled univariate analysis results for the polymorphisms that were common between phase I and phase II (disease free survival)

Variables	Number of patients in each category	p-value	HR	95% CI for HR	
				Lower	Upper
HIF1A rs2301106	(182 vs 623)	0.766	0.965	0.761	1.223
HIF1B rs2228099	(462 vs 344)	0.651	0.955	0.782	1.166
HIF1B rs11204737	(538 vs 267)	0.197	1.151	0.93	1.425
HIF1B rs10847	(365 vs 430)	0.381	0.914	0.748	1.118
HIF2A rs2121266	(493 vs 291)	0.557	1.065	0.863	1.315
HIF2A rs2346175	(501 vs 218)	0.539	1.077	0.849	1.367
HIF2A rs9973653	(388 vs 388)	0.455	1.081	0.882	1.324
HIF2A rs10178633	(514 vs 260)	0.712	0.96	0.775	1.19
HIF2A rs12614710	(552 vs 230)	0.587	1.064	0.851	1.329
HIF2A rs2346176	(474 vs 311)	0.881	1.016	0.827	1.248
HIF2A rs3768728	(196 vs 589)	0.992	1.001	0.793	1.264
HIF2A rs6712143	(359 vs 426)	0.863	0.982	0.803	1.202
*HIF2A*rs4953353	(445 vs 340)	0.921	1.01	0.824	1.238
Sex (male vs female)	(478 vs 336)	**0.027**	**1.26**	**1.027**	**1.544**
Age		**<.001**	**1.03**	**1.016**	**1.036**
Grade (poorly differentiated/ undifferentiated vs well/moderately differentiated)	(80 vs 726)	**0.016**	**1.465**	**1.074**	**1.999**
Histology (mucinous vs non-mucinous)	(104 vs 710)	0.587	1.083	0.811	1.447
Lymphatic invasion (+ vs -)	(289 vs 383)	**<.001**	**1.918**	**1.54**	**2.388**
Location (rectum vs colon)	(236 vs 578)	0.139	1.173	0.949	1.45
Stage		**<.001**			
Stage (II vs I)	(300 vs 150)	**0.036**	**1.446**	**1.024**	**2.043**
Stage (III vs I)	(255 vs 150)	**<.001**	**2.253**	**1.601**	**3.169**
Stage (IV vs I)	(100 vs 150)	**<.001**	**8.301**	**5.762**	**11.957**
MSI-H status (Yes vs No)	(91 vs 701)	**<.001**	**0.362**	**0.236**	**0.557**

(+): presence, (-): absence, CI: Confidence Interval, HR: Hazard Ratio, MSI-H: Microsatellite Instability-Highn: number, vs: versus. Values that are bolded are statistically significant (p value <0.05). Please note that the numbers might have been rounded to fit the table in this page.

Appendix 17: Results of the multivariate analysis of the *HIF2A* rs4953342 polymorphism with respect to the disease free survival in the replication cohort

Variables	n	Number of patients in each category	p value	HR	95% CI	
					Lower	Upper
HIF2A rs4953342 (AG+GG vs AA)	509	(251 vs 258)	0.094	1.267	0.96	1.672
Age			0.5	1.005	0.99	1.021
Grade (poorly differentiated vs well/moderately differentiated)		(37 vs 472)	0.512	0.821	0.455	1.482
Stage			**<.001**			
Stage (II vs I)		(196 vs 93)	0.136	1.447	0.89	2.352
Stage (III vs I)		(169 vs 93)	**0.001**	**2.166**	**1.344**	**3.49**
Stage (IV vs I)		(51 vs 93)	**<.001**	**5.894**	**3.491**	**9.951**
MSI-H status (Yes vs No)			**0.01**	**0.428**	**0.223**	**0.819**

CI: Confidence Interval, HR: Hazard Ratio, MSI-H: Microsatellite Instability-High, vs-: versus. Values that are bolded are statistically significant (p <0.05). Please note that the numbers might have been rounded to fit the table in this page.

Appendix 18: Pooled univariate analysis results for the polymorphisms that were investigated in phase I and were highly correlated with the other SNPs in phase II (overall survival)

*Variables compared	Number of patients in each category	p-value	HR	95% CI for HR	
				Lower	Upper
LOX rs10040971or rs3792802	(238 vs 568)	0.596	1.063	0.848	1.334
LOX rs2288393 or rs1800449	(246 vs 560)	0.623	1.058	0.845	1.324
HIF1A rs2301111 or rs 12434438	(288 vs 519)	0.595	0.942	0.757	1.173
HIF1A rs11158358 or rs2057482	(206 vs 601)	0.14	0.83	0.648	1.063
CXCL12 rs11592974 or rs2839695	(293 vs 509)	0.235	0.875	0.701	1.091
HIF2A rs3768730 or rs7594912	(548 vs 234)	0.496	1.086	0.857	1.376
HIF2A rs9679290 or rs6758592	(543 vs 244)	0.839	0.976	0.776	1.229
HIF2A rs11125070 or rs4953342	(390 vs 392)	0.151	0.854	0.689	1.059
HIF2A rs13019414 or rs7571218	(486 vs 300)	0.715	0.96	0.771	1.195
HIF2A rs1562453 or rs6756667	(571 vs 212)	0.978	1.003	0.789	1.275
HIF2A rs1868087 or rs1868086	(277 vs 507)	0.865	0.981	0.783	1.228
HIF2A rs2044456 or rs17034950	(385 vs 389)	0.845	1.022	0.823	1.268
HIF2A rs4953349 or rs1868089	(571 vs 209)	0.7	0.953	0.749	1.214
HIF2A rs6706003 or rs4953352	(577 vs 208)	0.091	1.242	0.966	1.596
HIF2A rs7583392 or rs7557402	(550 vs 236)	0.731	0.96	0.762	1.21
Sex (male vs female)	(479 vs 336)	**0.015**	**1.31**	**1.054**	**1.626**
Age		**<.001**	**1.03**	**1.024**	**1.045**
Grade (poorly differentiated/undifferentiated vs well/moderately differentiated)	(80 vs 727)	**0.002**	**1.65**	**1.195**	**2.263**
Histology (mucinous vs non-mucinous)	(104 vs 711)	0.501	1.11	0.819	1.505
Lymphatic invasion (+ vs -)	(289 vs 384)	**<.001**	**2.1**	**1.663**	**2.652**
Location (rectum vs colon)	(236 vs 579)	0.447	1.092	0.871	1.37
Stage		**<.001**			
Stage (II vs I)	(300 vs 151)	**0.061**	**1.43**	**0.983**	**2.067**
Stage (III vs I)	(255 vs 151)	**<.001**	**2.19**	**1.516**	**3.155**
Stage (IV vs I)	(100 vs 151)	**<.001**	**9.61**	**6.528**	**14.14**
MSI-H status (Yes vs No)	(92 vs 701)	**<.001**	**0.32**	**0.194**	**0.514**

*(+): presence, (-): absence, CI: Confidence Interval, HR: Hazard Ratio, MSI-H: Microsatellite Instability-High, n: number. Values that are bolded are statistically significant (p value <0.05). *The first SNP is the SNP investigated in phase I and the second SNP is the highly correlated SNP investigated in phase-2. Please note that the numbers might have been rounded to fit the table in this page.*

Appendix 19: Pooled univariate analysis results for the polymorphisms that were investigated in phase I and were highly correlated with the other SNPs in phase II (disease free survival)

*Variables compared	Number of patients in each category	p-value	HR	95% CI for HR	
				Lower	Upper
LOX rs10040971 or rs3792802	(238 vs 567)	0.544	1.068	0.863	1.323
LOX rs2288393 or rs1800449	(246 vs 559)	0.573	1.063	0.86	1.314
HIF1A rs2301111 or rs 12434438	(287 vs 519)	0.765	1.032	0.841	1.266
HIF1A rs11158358 or rs2057482	(206 vs 600)	0.555	0.934	0.744	1.172
CXCL12 rs11592974 or rs2839695	(293 vs 508)	0.444	0.922	0.75	1.135
HIF2A rs3768730 or rs7594912	(547 vs 234)	0.84	1.023	0.82	1.276
HIF2A rs9679290 or rs6758592	(542 vs 244)	0.768	0.968	0.779	1.202
HIF2A rs11125070 or rs4953342	(389 vs 392)	0.334	0.905	0.739	1.108
HIF2A rs13019414 or rs7571218	(485 vs 300)	0.506	0.933	0.759	1.146
HIF2A rs1562453 or rs6756667	(570 vs 212)	0.638	0.948	0.757	1.186
HIF2A rs1868087 or rs1868086	(277 vs 506)	0.473	0.925	0.747	1.145
HIF2A rs2044456 or rs17034950	(384 vs 389)	0.487	1.075	0.877	1.318
HIF2A rs4953349 or rs1868089	(570 vs 209)	0.855	0.979	0.779	1.23
HIF2A rs6706003 or rs4953352	(576 vs 208)	0.058	1.258	0.992	1.596
HIF2A rs7583392 or rs7557402	(549 vs 236)	0.301	0.892	0.719	1.107
Sex (male vs female)	(478 vs 336)	**0.027**	**1.259**	**1.027**	**1.544**
Age	-	**<.001**	**1.026**	**1.016**	**1.036**
Grade (poorly differentiated/undifferentiated vs well/moderately differentiated)	(80 vs 726)	**0.016**	**1.465**	**1.074**	**1.999**
Histology (mucinous vs non-mucinous)	(104 vs 710)	0.587	1.083	0.811	1.447
Lymphatic invasion (+ vs -)	(289 vs 383)	**<.001**	**1.918**	**1.54**	**2.388**
Location (rectum vs colon)	(236 vs 578)	0.139	1.173	0.949	1.45
Stage		**<.001**			
Stage (II vs I)	(300 vs 150)	**0.036**	**1.446**	**1.024**	**2.043**
Stage (III vs I)	(255 vs 150)	**<.001**	**2.253**	**1.601**	**3.169**
Stage (IV vs I)	(100 vs 150)	**<.001**	**8.301**	**5.762**	**11.957**
MSI-H status (Yes vs No)	(91 vs 701)	**<.001**	**0.362**	**0.236**	**0.557**

*(+): presence, (-): absence, CI: Confidence Interval, HR: Hazard Ratio, MSI-H: Microsatellite Instability-High, vs: versus. Values that are bolded are statistically significant (p value <0.05). *The first SNP is the SNP investigated in phase I and the second SNP is the highly correlated SNP investigated in phase-2. Please note that the numbers might have been rounded to fit the table in this page.*

6) References

(1) Ma Y, Liang D, Liu J, Axcrona K, Kvalheim G, Stokke T, et al. Prostate cancer cell lines under hypoxia exhibit greater stem-like properties. PLoS One 2011;6(12):e29170.

(2) Boyle, P., Levin, B. World Cancer Report 2008. IARC Nonserial Publication 2008;ISBN-13 9789283204237.

(3) Canadian Caner Statistics-2012. 2012.

(4) Tom Strachan and Andrew P Read. Human Molecular Genetics. . 3rd ed. United States of America: Garland Publishing; 2004. p. 241-244.

(5) Modica-Napolitano JS, Kulawiec M, Singh KK. Mitochondria and human cancer. Curr Mol Med 2007 Feb;7(1):121-131.

(6) Lee HC, Wei YH. Mitochondrial DNA instability and metabolic shift in human cancers. Int J Mol Sci 2009 Feb;10(2):674-701.

(7) Ahmed FE. Gene-gene, gene-environment & multiple interactions in colorectal cancer. J Environ Sci Health C Environ Carcinog Ecotoxicol Rev 2006 Apr;24(1):1-101.

(8) Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. Int J Cancer 2010 Dec 15;127(12):2893-2917.

(9) Canadian Cancer Society's Steering Committee on Cancer Statistics. Canadian Cancer Statistics 2011. Toronto, ON: Canadian Cancer Society; 2011. .

(10) O'Connell JB, Maggard MA, Ko CY. Colon Cancer Survival Rates With the New American Joint Committee on Cancer Sixth Edition Staging. Journal of the National Cancer Institute 2004 October 06;96(19):1420-1425.

(11) Lichtenstein P, Holm NV, Verkasalo PK, Iliadou A, Kaprio J, Koskenvuo M, et al. Environmental and heritable factors in the causation of cancer--analyses of cohorts of twins from Sweden, Denmark, and Finland. N Engl J Med 2000 Jul 13;343(2):78-85.

(12) D'Emilia JC, Rodriguez-Bigas MA, Petrelli NJ. The clinical and genetic manifestations of hereditary nonpolyposis colorectal carcinoma. Am J Surg 1995 Mar;169(3):368-372.

(13) Vasen HF. Clinical diagnosis and management of hereditary colorectal cancer syndromes. J Clin Oncol 2000 Nov 1;18(21 Suppl):81S-92S.

(14) McShane LM, Altman DG, Sauerbrei W. Identification of Clinically Useful Cancer Prognostic Factors: What Are We Missing? Journal of the National Cancer Institute 20 July 2005 20 July 2005;97(14):1023-1025.

(15) Lynch HT, de la Chapelle A. Genetic susceptibility to non-polyposis colorectal cancer. J Med Genet 1999 Nov;36(11):801-818.

(16) Abdel-Rahman WM, Peltomaki P. Lynch syndrome and related familial colorectal cancers. Crit Rev Oncog 2008;14(1):1-22; discussion 23-31.

(17) Bradley BA, Evers BM. Molecular advances in the etiology and treatment of colorectal cancer. Surg Oncol 1997 Nov;6(3):143-156.

(18) Cunningham C, Dunlop MG. Molecular genetic basis of colorectal cancer susceptibility. Br J Surg 1996 Mar;83(3):321-329.

(19) Oshima M, Oshima H, Kitagawa K, Kobayashi M, Itakura C, Taketo M. Loss of Apc heterozygosity and abnormal tissue building in nascent intestinal polyps in mice carrying a truncated Apc gene. Proc Natl Acad Sci U S A 1995 May 9;92(10):4482-4486.

(20) Tsujii M, DuBois RN. Alterations in cellular adhesion and apoptosis in epithelial cells overexpressing prostaglandin endoperoxide synthase 2. Cell 1995 Nov 3;83(3):493-501.

(21) Arnold C,N., Goel ,Ajay, Blum H,E., Richard Boland ,C. Molecular pathogenesis of colorectal cancer.

(22) Half EE, Bresalier RS. Clinical management of hereditary colorectal cancer syndromes. Curr Opin Gastroenterol 2004 Jan;20(1):32-42.

(23) Bartsch H, Dally H, Popanda O, Risch A, Schmezer P. Genetic risk profiles for cancer susceptibility and therapy response. Recent Results Cancer Res 2007;174:19-36.

(24) Webb EL, Rudd MF, Sellick GS, El Galta R, Bethke L, Wood W, et al. Search for low penetrance alleles for colorectal cancer through a scan of 1467 non-synonymous SNPs in 2575 cases and 2707 controls with validation by kin-cohort analysis of 14 704 first-degree relatives. Human Molecular Genetics 2006 November 01;15(21):3263-3271.

(25) Peters U, Hutter CM, Hsu L, Schumacher FR, Conti DV, Carlson CS, et al. Meta-analysis of new genome-wide association studies of colorectal cancer risk. Hum Genet 2012 Feb;131(2):217-234.

(26) Houlston RS, Webb E, Broderick P, Pittman AM, Di Bernardo MC, Lubbe S, et al. Meta-analysis of genome-wide association data identifies four new susceptibility loci for colorectal cancer. Nat Genet 2008 Dec;40(12):1426-1435.

(27) Fearon ER, Vogelstein B. A genetic model for colorectal tumorigenesis. Cell 1990 Jun 1;61(5):759-767.

(28) Vogelstein B, Fearon ER, Hamilton SR, Kern SE, Preisinger AC, Leppert M, et al. Genetic alterations during colorectal-tumor development. N Engl J Med 1988 Sep 1;319(9):525-532.

(29) Harrison S, Benziger H. The molecular biology of colorectal carcinoma and its implications: a review. Surgeon 2011 Aug;9(4):200-210.

(30) Raptis S, Bapat B. Genetic instability in human tumors. 2006;96:303-320.

(31) Grady WM, Carethers JM. Genomic and epigenetic instability in colorectal cancer pathogenesis. Gastroenterology 2008 Oct;135(4):1079-1099.

(32) Boland CR, Thibodeau SN, Hamilton SR, Sidransky D, Eshleman JR, Burt RW, et al. 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 1998 11/15;58(22):5248-5257.

(33) Kuismanen SA, Holmberg MT, Salovaara R, de la Chapelle A, Peltomaki P. Genetic and epigenetic modification of MLH1 accounts for a major share of microsatellite-unstable colorectal cancers. Am J Pathol 2000 May;156(5):1773-1779.

(34) Chen D, Yu Z, Zhu Z, Lopez CD. The p53 pathway promotes efficient mitochondrial DNA base excision repair in colorectal cancer cells. Cancer Res 2006 Apr 1;66(7):3485-3494.

(35) Chen RZ, Pettersson U, Beard C, Jackson-Grusby L, Jaenisch R. DNA hypomethylation leads to elevated mutation rates. Nature 1998 Sep 3;395(6697):89-93.

(36) Ehrlich M. DNA hypomethylation in cancer cells. Epigenomics 2009 Dec;1(2):239-259.

(37) Shen L, Toyota M, Kondo Y, Lin E, Zhang L, Guo Y, et al. Integrated genetic and epigenetic analysis identifies three different subclasses of colon cancer. Proc Natl Acad Sci U S A 2007 Nov 20;104(47):18654-18659.

(38) Marzouk O, Schofield J. Review of Histopathological and Molecular Prognostic Features in Colorectal Cancer. Cancers 2011;3(2):2767-2810.

(39) Bianchi P, Laghi L, Delconte G, Malesci A. Prognostic Value of Colorectal Cancer Biomarkers. Cancers 2011;3(2):2080-2105.

(40) Eadens M, Grothey A. Curable Metastatic Colorectal Cancer. Curr Oncol Rep 2011;13(3):168-176.

(41) Wesselborg S, Engels IH, Rossmann E, Los M, Schulze-Osthoff K. Anticancer Drugs Induce Caspase-8/FLICE Activation and Apoptosis in the Absence of CD95 Receptor/Ligand Interaction. Blood 1999 May 01;93(9):3053-3063.

(42) Meyers M, Hwang A, Wagner MW, Boothman DA. Role of DNA mismatch repair in apoptotic responses to therapeutic agents. Environ Mol Mutagen 2004;44(4):249-264.

(43) T. Boulikas, A. Pantos, E. Bellis, and P. Christofis. Designing platinum compounds in cancer: structures and mechanisms. Cancer Therapy 2007;5:537–583.

(44) Pang SK, Yu CW, Au-Yeung SC, Ho YP. DNA damage induced by novel demethylcantharidin-integrated platinum anticancer complexes. Biochem Biophys Res Commun 2007 Nov 9;363(1):235-240.

(45) Szczepankiewicz BG, Liu G, Jae H, Tasker AS, Gunawardana IW, von Geldern TW, et al. New Antimitotic Agents with Activity in Multi-Drug-Resistant Cell Lines and in Vivo Efficacy in Murine Tumor Models. J Med Chem 2001 12/01;44(25):4416-4430.

(46) Grivicich I, Mans DR, Peters GJ, Schwartsmann G. Irinotecan and oxaliplatin: an overview of the novel chemotherapeutic options for the treatment of advanced colorectal cancer. Braz J Med Biol Res 2001 Sep;34(9):1087-1103.

(47) Jakobsen A, Pfeiffer P, Hansen F, Sandberg E, Aabo K. Dual modulation of UFT with leucovorin and hydroxyurea in metastatic colorectal cancer. Acta Oncol 2001;40(1):63-66.

(48) Marcuello E, Paez D, Pare L, Salazar J, Sebio A, del Rio E, et al. A genotype-directed phase I-IV dose-finding study of irinotecan in combination with fluorouracil/leucovorin as first-line treatment in advanced colorectal cancer. Br J Cancer 2011 06/28;105(1):53-57.

(49) Hubner RA, Houlston RS. Folate and colorectal cancer prevention. Br J Cancer 2009 Jan 27;100(2):233-239.

(50) Zampino MG, Lorizzo K, Rocca A, Locatelli M, Zorzino L, Manzoni S, et al. Oxaliplatin Combined with 5-Fluorouracil and Methotrexate in Advanced Colorectal Cancer. Anticancer Research May-June 2006 May-June 2006;26(3B):2425-2428.

(51) Levine J,S., Burakoff ,Robert. Chemoprophylaxis of colorectal cancer in inflammatory bowel disease: Current concepts.

(52) Ferrara N, Hillan KJ, Novotny W. Bevacizumab (Avastin), a humanized anti-VEGF monoclonal antibody for cancer therapy. Biochem Biophys Res Commun 2005 Jul 29;333(2):328-335.

(53) Grey MR. Death Foretold: Prophecy and Prognosis in Medical Care. JAMA: The Journal of the American Medical Association 2000 July 19;284(3):370-372.

(54) Altman DG. Prognostic models: a methodological framework and review of models for breast cancer. Cancer Invest 2009 Mar;27(3):235-243.

(55) Stewart MG, Neely JG, Hartman JM, Wallace MS, Forsen JW,Jr. Tutorials in clinical research: part V: outcomes research. Laryngoscope 2002 Feb;112(2):248-254.

(56) Vickers AJ, Jang K, Sargent D, Lilja H, Kattan MW. Systematic review of statistical methods used in molecular marker studies in cancer. Cancer 2008 Apr 15;112(8):1862-1868.

(57) Sanchez-Arago M, Chamorro M, Cuezva JM. Selection of cancer cells with repressed mitochondria triggers colon cancer progression. Carcinogenesis 2010 Apr;31(4):567-576.

(58) Jain L, Vargo CA, Danesi R, Sissung TM, Price DK, Venzon D, et al. The role of vascular endothelial growth factor SNPs as predictive and prognostic markers for major solid tumors. Mol Cancer Ther 2009 Sep;8(9):2496-2508.

(59) Wallner M, Herbst A, Behrens A, Crispin A, Stieber P, Göke B, et al. Methylation of Serum DNA Is an Independent Prognostic Marker in Colorectal Cancer. Clinical Cancer Research 2006 December 15;12(24):7347-7352.

(60) Smits KM, Cleven AH, Weijenberg MP, Hughes LA, Herman JG, de Bruine AP, et al. Pharmacoepigenomics in colorectal cancer: a step forward in predicting prognosis and treatment response. Pharmacogenomics 2008 Dec;9(12):1903-1916.

(61) Moons KG, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research: what, why, and how? BMJ 2009 Feb 23;338:b375.

(62) Verhulst J, Ferdinande L, Demetter P, Ceelen W. Mucinous subtype as prognostic factor in colorectal cancer: a systematic review and meta-analysis. J Clin Pathol 2012 May;65(5):381-388.

(63) Guastadisegni C, Colafranceschi M, Ottini L, Dogliotti E. Microsatellite instability as a marker of prognosis and response to therapy: a meta-analysis of colorectal cancer survival data. Eur J Cancer 2010 Oct;46(15):2788-2798.

(64) Hemingway H, Riley RD, Altman DG. Ten steps towards improving prognosis research. BMJ 2009 Dec 30;339:b4184.

(65) Mallett S, Royston P, Dutton S, Waters R, Altman DG. Reporting methods in studies developing prognostic models in cancer: a review. BMC Med 2010 Mar 30;8:20.

(66) Braitman LE, Davidoff F. Predicting Clinical States in Individual Patients. Annals of Internal Medicine 1996 September 01;125(5):406-412.

(67) Katz MH. Multivariable analysis: a primer for readers of medical research. Ann Intern Med 2003 Apr 15;138(8):644-650.

(68) Altman DG, Vergouwe Y, Royston P, Moons KG. Prognosis and prognostic research: validating a prognostic model. BMJ 2009 May 28;338:b605.

(69) Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. J Clin Epidemiol 1996 Dec;49(12):1373-1379.

(70) Whitley E, Ball J. Statistics review 4: sample size calculations. Crit Care 2002 Aug;6(4):335-341.

(71) Bewick V, Cheek L, Ball J. Statistics review 12: survival analysis. Crit Care 2004 Oct;8(5):389-394.

(72) Hess DR. Retrospective studies and chart reviews. Respir Care 2004 Oct;49(10):1171-1174.

(73) Manolio TA, Bailey-Wilson J, Collins FS. Genes, environment and the value of prospective cohort studies. Nat Rev Genet 2006 print;7(10):812-820.

(74) Clark TG, Bradburn MJ, Love SB, Altman DG. Survival analysis part I: basic concepts and first analyses. Br J Cancer 2003 Jul 21;89(2):232-238.

(75) Szczech LA, Coladonato JA, Owen WF,Jr. Key concepts in biostatistics: using statistics to answer the question "is there a difference?". Semin Dial 2002 Sep-Oct;15(5):347-351.

(76) Clark TG, Bradburn MJ, Love SB, Altman DG. Survival analysis part IV: further concepts and methods in survival analysis. Br J Cancer 2003 Sep 1;89(5):781-786.

(77) Compton CC, Fielding LP, Burgart LJ, Conley B, Cooper HS, Hamilton SR, et al. Prognostic factors in colorectal cancer. College of American Pathologists Consensus Statement 1999. Arch Pathol Lab Med 2000 Jul;124(7):979-994.

(78) O'Connell JB, Maggard MA, Ko CY. Colon cancer survival rates with the new American Joint Committee on Cancer sixth edition staging. J Natl Cancer Inst 2004 Oct 6;96(19):1420-1425.

(79) Zlobec I, Lugli A. Prognostic and predictive factors in colorectal cancer. Postgraduate Medical Journal 2008 August 01;84(994):403-411.

(80) Compton CC. Colorectal carcinoma: diagnostic, prognostic, and molecular features. Mod Pathol 2003 Apr;16(4):376-388.

(81) Talbot IC, Ritchie S, Leighton MH, Hughes AO, Bussey HJ, Morson BC. The clinical significance of invasion of veins by rectal cancer. Br J Surg 1980 Jun;67(6):439-442.

(82) Bentzen SM, Balslev I, Pedersen M, Teglbjaerg PS, Hanberg-Sorensen F, Bone J, et al. Time to loco-regional recurrence after resection of Dukes' B and C colorectal cancer with or without adjuvant postoperative radiotherapy. A multivariate regression analysis. Br J Cancer 1992 Jan;65(1):102-107.

(83) Harris EI, Lewin DN, Wang HL, Lauwers GY, Srivastava A, Shyr Y, et al. Lymphovascular invasion in colorectal cancer: an interobserver variability study. Am J Surg Pathol 2008 Dec;32(12):1816-1821.

(84) Saito Y, Ouchi A, Goto S, Nukada Y, Mizoi T, Matsuno S, et al. Relationship between venous invasion and hematogenous metastasis in the colorectal cancer patients. Nihon Geka Gakkai Zasshi 1992 Feb;93(2):133-138.

(85) Yue SQ, Yang YL, Dou KF, Li KZ. Expression of PCNA and CD44mRNA in colorectal cancer with venous invasion and its relationship to liver metastasis. World J Gastroenterol 2003 Dec;9(12):2863-2865.

(86) Sternberg A, Amar M, Alfici R, Groisman G. Conclusions from a study of venous invasion in stage IV colorectal adenocarcinoma. J Clin Pathol 2002 Jan;55(1):17-21.

(87) Plank M.J and Sleeman B.D. Tumour-induced Angiogenesis: A Review. Journal of Theoretical Medicine September–December 2003;Vol. 5 (3–4):pp. 137–153.

(88) Claij N, Riele Ht. Microsatellite Instability in Human Cancer: A Prognostic Marker for Chemotherapy? Exp Cell Res 1999;246(1):1.

(89) Deschoolmeester V, Baay M, Specenier P, Lardon F, Vermorken JB. A Review of the Most Promising Biomarkers in Colorectal Cancer: One Step Closer to Targeted Therapy. The Oncologist 2010 July 01;15(7):699-731.

(90) Emterling A, Wallin A, Arbman G, Sun XF. Clinicopathological significance of microsatellite instability and mutated RIZ in colorectal cancer. Ann Oncol 2004 Feb;15(2):242-246.

(91) Storojeva I, Boulay JL, Heinimann K, Ballabeni P, Terracciano L, Laffer U, et al. Prognostic and predictive relevance of microsatellite instability in colorectal cancer. Oncol Rep 2005 Jul;14(1):241-249.

(92) Lamberti C, Lundin S, Bogdanow M, Pagenstecher C, Friedrichs N, Buttner R, et al. Microsatellite instability did not predict individual survival of unselected patients with colorectal cancer. Int J Colorectal Dis 2007 Feb;22(2):145-152.

(93) Popat S, Houlston RS. A systematic review and meta-analysis of the relationship between chromosome 18q genotype, DCC status and colorectal cancer prognosis. Eur J Cancer 2005 Sep;41(14):2060-2070.

(94) Edward Chu. New Treatment Strategies for Metastatic Colorectal Cancer.

(95) Prall F, Ostwald C. High-degree tumor budding and podia-formation in sporadic colorectal carcinomas with K-ras gene mutations. Hum Pathol 2007 Nov;38(11):1696-1702.

(96) Oldenhuis CN, Oosting SF, Gietema JA, de Vries EG. Prognostic versus predictive value of biomarkers in oncology. Eur J Cancer 2008 May;44(7):946-953.

(97) Kluftinger AM, Robinson BW, Quenville NF, Finley RJ, Davis NL. Correlation of epidermal growth factor receptor and c-erbB2 oncogene product to known prognostic indicators of colorectal cancer. Surg Oncol 1992 Feb;1(1):97-105.

(98) Steele RJ, Kelly P, Ellul B, Eremin O. Epidermal growth factor receptor expression in colorectal cancer. Br J Surg 1990 Dec;77(12):1352-1354.

(99) Kountourakis P, Pavlakis K, Psyrri A, Rontogianni D, Xiros N, Patsouris E, et al. Clinicopathologic significance of EGFR and Her-2/neu in colorectal adenocarcinomas. Cancer J 2006 May-Jun;12(3):229-236.

(100) Koretz K, Schlag P, Moller P. Expression of epidermal growth factor receptor in normal colorectal mucosa, adenoma, and carcinoma. Virchows Arch A Pathol Anat Histopathol 1990;416(4):343-349.

(101) Cunningham MP, Essapen S, Thomas H, Green M, Lovell DP, Topham C, et al. Coexpression of the IGF-IR, EGFR and HER-2 is common in colorectal cancer patients. Int J Oncol 2006 Feb;28(2):329-335.

(102) McKay JA, Murray LJ, Curran S, Ross VG, Clark C, Murray GI, et al. Evaluation of the epidermal growth factor receptor (EGFR) in colorectal tumours and lymph node metastases. Eur J Cancer 2002 Nov;38(17):2258-2264.

(103) Walther A, Johnstone E, Swanton C, Midgley R, Tomlinson I, Kerr D. Genetic prognostic and predictive markers in colorectal cancer. Nat Rev Cancer 2009 Jul;9(7):489-499.

(104) Catalano V, Loupakis F, Graziano F, Bisonni R, Torresi U, Vincenzi B, et al. Prognosis of mucinous histology for patients with radically resected stage II and III colon cancer. Annals of Oncology 2011 April 29.

(105) Szkandera J, Knechtel G, Stotz M, Hofmann G, Langsenlehner U, Krippl P, et al. Association of hypoxia-inducible factor 1-alpha gene polymorphisms and colorectal cancer prognosis. Anticancer Res 2010 Jun;30(6):2393-2397.

(106) Chang SC, Lin PC, Yang SH, Wang HS, Li AF, Lin JK. SDF-1alpha G801A polymorphism predicts lymph node metastasis in stage T3 colorectal cancer. Ann Surg Oncol 2009 Aug;16(8):2323-2330.

(107) Wang J, Zhao Y, Jiang J, Gajalakshmi V, Kuriki K, Nakamura S, et al. Polymorphisms in DNA repair genes *XRCC1*, *XRCC3* and *XPD*, and colorectal cancer risk: a caseâ€"control study in an Indian population. J Cancer Res Clin Oncol 2010;136(10):1517-1525.

(108) Xing J, Wan S, Zhou F, Qu F, Li B, Myers RE, et al. Genetic polymorphisms in pre-microRNA genes as prognostic markers of colorectal cancer. Cancer Epidemiol Biomarkers Prev 2012 Jan;21(1):217-227.

(109) Chihara N, Amo T, Tokunaga A, Yuzuriha R, Wolf AM, Asoh S, et al. Mitochondrial DNA alterations in colorectal cancer cell lines. J Nippon Med Sch 2011;78(1):13-21.

(110) Fang DC, Fang L, Wang RQ, Yang SM. Nuclear and mitochondrial DNA microsatellite instability in hepatocellular carcinoma in Chinese. World J Gastroenterol 2004 Feb 1;10(3):371-375.

(111) Park CB, Larsson NG. Mitochondrial DNA mutations in disease and aging. J Cell Biol 2011 May 30;193(5):809-818.

(112) Chen T, He J, Shen L, Fang H, Nie H, Jin T, et al. The mitochondrial DNA 4,977-bp deletion and its implication in copy number alteration in colorectal cancer. BMC Med Genet 2011 Jan 13;12:8.

(113) Lee HC, Chang CM, Chi CW. Somatic mutations of mitochondrial DNA in aging and cancer progression. Ageing Res Rev 2010 Nov;9 Suppl 1:S47-58.

(114) Warburg OH. The Metabolism of Tumours: Investigations from the Kaiser Wilhelm Institute for Biology. 1930.

(115) Jakupciak JP, Wang W, Markowitz ME, Ally D, Coble M, Srivastava S, et al. Mitochondrial DNA as a cancer biomarker. J Mol Diagn 2005 May;7(2):258-267.

(116) Bi X, Lin Q, Foo TW, Joshi S, You T, Shen HM, et al. Proteomic analysis of colorectal cancer reveals alterations in metabolic pathways: mechanism of tumorigenesis. Mol Cell Proteomics 2006 Jun;5(6):1119-1130.

(117) Shen L, Fang H, Chen T, He J, Zhang M, Wei X, et al. Evaluating mitochondrial DNA in cancer occurrence and development. Ann N Y Acad Sci 2010 Jul;1201:26-33.

(118) Lin PC, Lin JK, Yang SH, Wang HS, Li AF, Chang SC. Expression of beta-F1-ATPase and mitochondrial transcription factor A and the change in mitochondrial DNA content in colorectal cancer: clinical data analysis and evidence from an in vitro study. Int J Colorectal Dis 2008 Dec;23(12):1223-1232.

(119) Lievre A, Chapusot C, Bouvier AM, Zinzindohoue F, Piard F, Roignot P, et al. Clinical value of mitochondrial mutations in colorectal cancer. J Clin Oncol 2005 May 20;23(15):3517-3525.

(120) Theodoratou E, Din FV, Farrington SM, Cetnarskyj R, Barnetson RA, Porteous ME, et al. Association between common mtDNA variants and all-cause or colorectal cancer mortality. Carcinogenesis 2010 Feb;31(2):296-301.

(121) Tamori A, Nishiguchi S, Nishikawa M, Kubo S, Koh N, Hirohashi K, et al. Correlation between clinical characteristics and mitochondrial D-loop DNA mutations in hepatocellular carcinoma. J Gastroenterol 2004 Nov;39(11):1063-1068.

(122) Lee HC, Yin PH, Lin JC, Wu CC, Chen CY, Wu CW, et al. Mitochondrial genome instability and mtDNA depletion in human cancers. Ann N Y Acad Sci 2005 May;1042:109-122.

(123) Woods MO, Younghusband HB, Parfrey PS, Gallinger S, McLaughlin J, Dicks E, et al. The genetic basis of colorectal cancer in a population-based incident cohort with a high rate of familial disease. Gut 2010 Oct;59(10):1369-1377.

(124) Green RC, Green JS, Buehler SK, Robb JD, Daftary D, Gallinger S, et al. Very high incidence of familial colorectal cancer in Newfoundland: a comparison with Ontario and 13 other population-based studies. Fam Cancer 2007;6(1):53-62.

(125) Liou CW, Lin TK, Chen JB, Tiao MM, Weng SW, Chen SD, et al. Association between a common mitochondrial DNA D-loop polycytosine variant and alteration of mitochondrial copy number in human peripheral blood cells. J Med Genet 2010 Nov;47(11):723-728.

(126) Bendall KE, Sykes BC. Length heteroplasmy in the first hypervariable segment of the human mtDNA control region. Am J Hum Genet 1995 Aug;57(2):248-256.

(127) Poulton J, Luan J, Macaulay V, Hennings S, Mitchell J, Wareham NJ. Type 2 diabetes is associated with a common mitochondrial variant: evidence from a population-based case–control study. Human Molecular Genetics 2002 June 15;11(13):1581-1583.

(128) Roberti M, Musicco C, Polosa PL, Milella F, Gadaleta MN, Cantatore P. Multiple protein-binding sites in the TAS-region of human and rat mitochondrial DNA. Biochem Biophys Res Commun 1998 Feb 4;243(1):36-40.

(129) Lin TK, Chen SD, Wang PW, Wei YH, Lee CF, Chen TL, Chuang YC, Tan TY, Chang KC, Liou CW. Increased Oxidative Damage with Altered Antioxidative Status in Type 2 Diabetic Patients Harboring the 16189 T to C Variant of Mitochondrial DNA.

(130) Kulawiec M, Owens KM, Singh KK. mtDNA G10398A variant in African-American women with breast cancer provides resistance to apoptosis and promotes metastasis in mice. J Hum Genet 2009 Nov;54(11):647-654.

(131) Ingman, M. & Gyllensten, U. mtDB: Human Mitochondrial Genome Database, a resource for population genetics and medical sciences. Nucleic Acids Res 34, D749-D751 (2006). .

(132) Juo SH, Lu MY, Bai RK, Liao YC, Trieu RB, Yu ML, et al. A common mitochondrial polymorphism 10398A>G is associated metabolic syndrome in a Chinese population. Mitochondrion 2010 Apr;10(3):294-299.

(133) CONAN online tool - Copy Number Analysis database of Sanger http://www.sanger.ac.uk/cgi-bin/genetics/CGP/conan/search.cgi?geneName=FASLG&location=1:170894777-170902637&stop=170902637&chr=1&action=DisplayFootprint&start=170894777.

(134) Cossarizza A, Riva A, Pinti M, Ammannato S, Fedeli P, Mussini C, et al. Increased mitochondrial DNA content in peripheral blood lymphocytes from HIV-infected patients with lipodystrophy. Antivir Ther 2003 Aug;8(4):315-321.

(135) Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods 2001 Dec;25(4):402-408.

(136) Isabel dos Santos Silva. Cancer Epidem iology: Principles and Methods. World Heal th Organization, International Agency For Research On Cancer; 1999.

(137) Streiner DL, Norman GR. Correction for multiple testing: is there a resolution? Chest 2011 Jul;140(1):16-18.

(138) Database of Single Nucleotide Polymorphisms (dbSNP database) http://www.ncbi.nlm.nih.gov/projects/SNP/snp_ref.cgi?locusId=356.

(139) Canter JA, Kallianpur AR, Parl FF, Millikan RC. Mitochondrial DNA G10398A polymorphism and invasive breast cancer in African-American women. Cancer Res 2005;65(17):8028-8033.

(140) Chinnery PF, Elliott HR, Patel S, Lambert C, Keers SM, Durham SE, et al. Role of the mitochondrial DNA 16184â€“16193 poly-C tract in type 2 diabetes. The Lancet 5; 2005/11;366(9497):1650-1651.

(141) Mueller EE, Eder W, Ebner S, Schwaiger E, Santic D, Kreindl T, et al. The mitochondrial T16189C polymorphism is associated with coronary artery disease in Middle European populations. PLoS One 2011 Jan 26;6(1):e16455.

(142) Theodoratou E, Din FVN, Farrington SM, Cetnarskyj R, Barnetson RA, Porteous ME, et al. Association between common mtDNA variants and all-cause or colorectal cancer mortality. Carcinogenesis 2010 February 01;31(2):296-301.

(143) Selvanayagam P, Rajaraman S. Detection of mitochondrial genome depletion by a novel cDNA in renal cell carcinoma. Lab Invest 1996 Mar;74(3):592-599.

(144) Heddi A, Faure-Vigny H, Wallace DC, Stepien G. Coordinate expression of nuclear and mitochondrial genes involved in energy production in carcinoma and oncocytoma. Biochim Biophys Acta 1996 Aug 23;1316(3):203-209.

(145) Kim MM, Clinger JD, Masayesva BG, Ha PK, Zahurak ML, Westra WH, et al. Mitochondrial DNA quantity increases with histopathologic grade in premalignant and malignant head and neck lesions. Clin Cancer Res 2004 Dec 15;10(24):8512-8515.

(146) Mambo E, Chatterjee A, Xing M, Tallini G, Haugen BR, Yeung SC, et al. Tumor-specific changes in mtDNA content in human cancer. Int J Cancer 2005 Oct 10;116(6):920-924.

(147) Wang Y, Liu VW, Xue WC, Tsang PC, Cheung AN, Ngan HY. The increase of mitochondrial DNA content in endometrial adenocarcinoma cells: a quantitative study using laser-captured microdissected tissues. Gynecol Oncol 2005 Jul;98(1):104-110.

(148) Wang Y, Liu VW, Xue WC, Cheung AN, Ngan HY. Association of decreased mitochondrial DNA content with ovarian cancer progression. Br J Cancer 2006 Oct 23;95(8):1087-1091.

(149) Mizumachi T, Muskhelishvili L, Naito A, Furusawa J, Fan CY, Siegel ER, et al. Increased distributional variance of mitochondrial DNA content associated with prostate cancer cells as compared with normal prostate cells. Prostate 2008 Mar 1;68(4):408-417.

(150) Meierhofer D, Mayr JA, Foetschl U, Berger A, Fink K, Schmeller N, et al. Decrease of mitochondrial DNA content and energy metabolism in renal cell carcinoma. Carcinogenesis 2004 Jun;25(6):1005-1010.

(151) Yin PH, Lee HC, Chau GY, Wu YT, Li SH, Lui WY, et al. Alteration of the copy number and deletion of mitochondrial DNA in human hepatocellular carcinoma. Br J Cancer 2004 Jun 14;90(12):2390-2396.

(152) Cuezva JM, Krajewska M, de Heredia ML, Krajewski S, Santamaria G, Kim H, et al. The bioenergetic signature of cancer: a marker of tumor progression. Cancer Res 2002 Nov 15;62(22):6674-6681.

(153) Yamada S, Nomoto S, Fujii T, Kaneko T, Takeda S, Inoue S, et al. Correlation between copy number of mitochondrial DNA and clinico-pathologic parameters of hepatocellular carcinoma. Eur J Surg Oncol 2006 Apr;32(3):303-307.

(154) Wu CW, Yin PH, Hung WY, Li AF, Li SH, Chi CW, et al. Mitochondrial DNA mutations and mitochondrial DNA depletion in gastric cancer. Genes Chromosomes Cancer 2005 Sep;44(1):19-28.

(155) Tseng LM, Yin PH, Chi CW, Hsu CY, Wu CW, Lee LM, et al. Mitochondrial DNA mutations and mitochondrial DNA depletion in breast cancer. Genes Chromosomes Cancer 2006 Jul;45(7):629-638.

(156) Yu M, Zhou Y, Shi Y, Ning L, Yang Y, Wei X, et al. Reduced mitochondrial DNA copy number is correlated with tumor progression and prognosis in Chinese breast cancer patients. IUBMB Life 2007 Jul;59(7):450-457.

(157) Yu M, Zhou Y, Shi Y, Ning L, Yang Y, Wei X, et al. Reduced mitochondrial DNA copy number is correlated with tumor progression and prognosis in Chinese breast cancer patients. IUBMB Life 2007;59(7):450-457.

(158) Chang SC, Lin PC, Yang SH, Wang HS, Liang WY, Lin JK. Mitochondrial D-loop mutation is a common event in colorectal cancers with p53 mutations. Int J Colorectal Dis 2009 Jun;24(6):623-628.

(159) Wang GL, Jiang BH, Rue EA, Semenza GL. Hypoxia-inducible factor 1 is a basic-helix-loop-helix-PAS heterodimer regulated by cellular O2 tension. Proc Natl Acad Sci U S A 1995 Jun 6;92(12):5510-5514.

(160) Baugh JA, Gantier M, Li L, Byrne A, Buckley A, Donnelly SC. Dual regulation of macrophage migration inhibitory factor (MIF) expression in hypoxia by CREB and HIF-1. Biochem Biophys Res Commun 2006 Sep 8;347(4):895-903.

(161) Mitchell RA, Bucala R. Tumor growth-promoting properties of macrophage migration inhibitory factor (MIF). Semin Cancer Biol 2000 Oct;10(5):359-366.

(162) Santiago B, Calonge E, Del Rey MJ, Gutierrez-Canas I, Izquierdo E, Usategui A, et al. CXCL12 gene expression is upregulated by hypoxia and growth arrest but not by inflammatory cytokines in rheumatoid synovial fibroblasts. Cytokine 2011 Feb;53(2):184-190.

(163) Ceradini DJ, Kulkarni AR, Callaghan MJ, Tepper OM, Bastidas N, Kleinman ME, et al. Progenitor cell trafficking is regulated by hypoxic gradients through HIF-1 induction of SDF-1. Nat Med 2004 Aug;10(8):858-864.

(164) Kim Y, Roh S, Park JY, Kim Y, Cho DH, Kim JC. Differential expression of the LOX family genes in human colorectal adenocarcinomas. Oncol Rep 2009 Oct;22(4):799-804.

(165) Anderson GR, Stoler DL, Scarcello LA. Normal fibroblasts responding to anoxia exhibit features of the malignant phenotype. J Biol Chem 1989 Sep 5;264(25):14885-14892.

(166) Zhong H, De Marzo AM, Laughner E, Lim M, Hilton DA, Zagzag D, et al. Overexpression of hypoxia-inducible factor 1alpha in common human cancers and their metastases. Cancer Res 1999 Nov 15;59(22):5830-5835.

(167) Young SD, Marshall RS, Hill RP. Hypoxia induces DNA overreplication and enhances metastatic potential of murine tumor cells. Proc Natl Acad Sci U S A 1988 Dec;85(24):9533-9537.

(168) Heacock CS, Sutherland RM. Enhanced synthesis of stress proteins caused by hypoxia and relation to altered cell growth and metabolism. Br J Cancer 1990 Aug;62(2):217-225.

(169) Blumenson LE, Bross ID. A possible mechanism for enhancement of increased production of tumor angiogenic factor. Growth 1976 Sep;40(3):205-209.

(170) Gray LH, Conger AD, Ebert M, Hornsey S, Scott OC. The concentration of oxygen dissolved in tissues at the time of irradiation as a factor in radiotherapy. Br J Radiol 1953 Dec;26(312):638-648.

(171) Deschner EE, Gray LH. Influence of oxygen tension on x-ray-induced chromosomal damage in Ehrlich ascites tumor cells irradiated in vitro and in vivo. Radiat Res 1959 Jul;11(1):115-146.

(172) Teicher BA. Hypoxia and drug resistance. Cancer Metastasis Rev 1994 Jun;13(2):139-168.

(173) Sullivan R, Graham CH. Hypoxia-driven selection of the metastatic phenotype. Cancer Metastasis Rev 2007 Jun;26(2):319-331.

(174) Krishnamachary B, Berg-Dixon S, Kelly B, Agani F, Feldser D, Ferreira G, et al. Regulation of colon carcinoma cell invasion by hypoxia-inducible factor 1. Cancer Res 2003 Mar 1;63(5):1138-1143.

(175) Kaidi A, Qualtrough D, Williams AC, Paraskeva C. Direct transcriptional up-regulation of cyclooxygenase-2 by hypoxia-inducible factor (HIF)-1 promotes colorectal tumor cell survival and enhances HIF-1 transcriptional activity during hypoxia. Cancer Res 2006 Jul 1;66(13):6683-6691.

(176) Dang DT, Chen F, Gardner LB, Cummins JM, Rago C, Bunz F, et al. Hypoxia-inducible factor-1alpha promotes nonhypoxia-mediated proliferation in colon cancer cells and xenografts. Cancer Res 2006 Feb 1;66(3):1684-1936.

(177) Ryan HE, Poloni M, McNulty W, Elson D, Gassmann M, Arbeit JM, et al. Hypoxia-inducible factor-1alpha is a positive factor in solid tumor growth. Cancer Res 2000 Aug 1;60(15):4010-4015.

(178) Carmeliet P, Dor Y, Herbert JM, Fukumura D, Brusselmans K, Dewerchin M, et al. Role of HIF-1alpha in hypoxia-mediated apoptosis, cell proliferation and tumour angiogenesis. Nature 1998 Jul 30;394(6692):485-490.

(179) Scortegagna M, Martin RJ, Kladney RD, Neumann RG, Arbeit JM. Hypoxia-inducible factor-1alpha suppresses squamous carcinogenic progression and epithelial-mesenchymal transition. Cancer Res 2009 Mar 15;69(6):2638-2646.

(180) Sowter HM, Ratcliffe PJ, Watson P, Greenberg AH, Harris AL. HIF-1-dependent regulation of hypoxic induction of the cell death factors BNIP3 and NIX in human tumors. Cancer Res 2001 Sep 15;61(18):6669-6673.

(181) Goda N, Ryan HE, Khadivi B, McNulty W, Rickert RC, Johnson RS. Hypoxia-inducible factor 1alpha is essential for cell cycle arrest during hypoxia. Mol Cell Biol 2003 Jan;23(1):359-369.

(182) Rasheed S, Harris AL, Tekkis PP, Turley H, Silver A, McDonald PJ, et al. Hypoxia-inducible factor-1alpha and -2alpha are expressed in most rectal cancers but only hypoxia-inducible factor-1alpha is associated with prognosis. Br J Cancer 2009 May 19;100(10):1666-1673.

(183) Baba Y, Nosho K, Shima K, Irahara N, Chan AT, Meyerhardt JA, et al. HIF1A overexpression is associated with poor prognosis in a cohort of 731 colorectal cancers. Am J Pathol 2010 May;176(5):2292-2301.

(184) Schmitz KJ, Muller CI, Reis H, Alakus H, Winde G, Baba HA, et al. Combined analysis of hypoxia-inducible factor 1 alpha and metallothionein indicates an aggressive subtype of colorectal carcinoma. Int J Colorectal Dis 2009 Nov;24(11):1287-1296.

(185) Rajaganeshan R, Prasad R, Guillou PJ, Scott N, Poston G, Jayne DG. Expression patterns of hypoxic markers at the invasive margin of colorectal cancers and liver metastases. Eur J Surg Oncol 2009 Dec;35(12):1286-1294.

(186) Kuwai T, Kitadai Y, Tanaka S, Onogawa S, Matsutani N, Kaio E, et al. Expression of hypoxia-inducible factor-1alpha is associated with tumor vascularization in human colorectal carcinoma. Int J Cancer 2003 Jun 10;105(2):176-181.

(187) Furlan D, Sahnane N, Carnevali I, Cerutti R, Bertoni F, Kwee I, et al. Up-regulation of the hypoxia-inducible factor-1 transcriptional pathway in colorectal carcinomas. Hum Pathol 2008 Oct;39(10):1483-1494.

(188) Tanimoto K, Yoshiga K, Eguchi H, Kaneyasu M, Ukon K, Kumazaki T, et al. Hypoxia-inducible factor-1alpha polymorphisms associated with enhanced transactivation capacity, implying clinical significance. Carcinogenesis 2003 Nov;24(11):1779-1783.

(189) Lee SJ, Kim JG, Sohn SK, Chae YS, Moon JH, Kang BW, et al. No association of the hypoxia-inducible factor-1alpha gene polymorphisms with survival in patients with colorectal cancer. Med Oncol 2011 Dec;28(4):1032-1037.

(190) Gerger A, El-Khoueiry A, Zhang W, Yang D, Singh H, Bohanes P, et al. Pharmacogenetic angiogenesis profiling for first-line Bevacizumab plus oxaliplatin-based chemotherapy in patients with metastatic colorectal cancer. Clin Cancer Res 2011 Sep 1;17(17):5783-5792.

(191) Arpiainen S, Lamsa V, Pelkonen O, Yim SH, Gonzalez FJ, Hakkola J. Aryl hydrocarbon receptor nuclear translocator and upstream stimulatory factor regulate Cytochrome P450 2a5 transcription through a common E-box site. J Mol Biol 2007 Jun 8;369(3):640-652.

(192) Huang LE, Gu J, Schau M, Bunn HF. Regulation of hypoxia-inducible factor 1alpha is mediated by an O2-dependent degradation domain via the ubiquitin-proteasome pathway. Proc Natl Acad Sci U S A 1998 Jul 7;95(14):7987-7992.

(193) Salceda S, Caro J. Hypoxia-inducible factor 1alpha (HIF-1alpha) protein is rapidly degraded by the ubiquitin-proteasome system under normoxic conditions. Its stabilization by hypoxia depends on redox-induced changes. J Biol Chem 1997 Sep 5;272(36):22642-22647.

(194) Jiang BH, Rue E, Wang GL, Roe R, Semenza GL. Dimerization, DNA binding, and transactivation properties of hypoxia-inducible factor 1. J Biol Chem 1996 Jul 26;271(30):17771-17778.

(195) Choi SM, Oh H, Park H. Microarray analyses of hypoxia-regulated genes in an aryl hydrocarbon receptor nuclear translocator (Arnt)-dependent manner. FEBS J 2008 Nov;275(22):5618-5634.

(196) Glover LE, Irizarry K, Scully M, Campbell EL, Bowers BE, Aherne CM, et al. IFN-gamma attenuates hypoxia-inducible factor (HIF) activity in intestinal epithelial cells through transcriptional repression of HIF-1beta. J Immunol 2011 Feb 1;186(3):1790-1798.

(197) Lim W, Park Y, Cho J, Park C, Park J, Park YK, et al. Estrogen receptor beta inhibits transcriptional activity of hypoxia inducible factor-1 through the downregulation of arylhydrocarbon receptor nuclear translocator. Breast Cancer Res 2011 Mar 24;13(2):R32.

(198) Partch CL, Gardner KH. Coactivators necessary for transcriptional output of the hypoxia inducible factor, HIF, are directly recruited by ARNT PAS-B. Proc Natl Acad Sci U S A 2011 May 10;108(19):7739-7744.

(199) Shi S, Yoon DY, Hodge-Bell KC, Bebenek IG, Whitekus MJ, Zhang R, et al. The aryl hydrocarbon receptor nuclear translocator (Arnt) is required for tumor initiation by benzo[a]pyrene. Carcinogenesis 2009 Nov;30(11):1957-1961.

(200) Chang KY, Shen MR, Lee MY, Wang WL, Su WC, Chang WC, et al. Epidermal growth factor-activated aryl hydrocarbon receptor nuclear translocator/HIF-1{beta} signal pathway up-regulates cyclooxygenase-2 gene expression associated with squamous cell carcinoma. J Biol Chem 2009 Apr 10;284(15):9908-9916.

(201) Huang WC, Chen ST, Chang WC, Chang KY, Chang WC, Chen BK. Involvement of aryl hydrocarbon receptor nuclear translocator in EGF-induced c-Jun/Sp1-mediated gene expression. Cell Mol Life Sci 2010 Oct;67(20):3523-3533.

(202) Gartel AL. P21(WAF1/CIP1) may be a tumor suppressor after all. Cancer Biol Ther 2007 Aug;6(8):1171-1172.

(203) Liang Y, Li W, Yang B, Tao Z, Sun H, Wang L, et al. Aryl hydrocarbon receptor nuclear translocator is associated with tumor growth and progression of hepatocellular carcinoma. International Journal of Cancer 2012;130(8):1745-1754.

(204) Holmquist-Mengelbier L, Fredlund E, Lofstedt T, Noguera R, Navarro S, Nilsson H, et al. Recruitment of HIF-1alpha and HIF-2alpha to common target genes is differentially regulated in neuroblastoma: HIF-2alpha promotes an aggressive phenotype. Cancer Cell 2006 Nov;10(5):413-423.

(205) Kondo K, Kim WY, Lechpammer M, Kaelin WG,Jr. Inhibition of HIF2alpha is sufficient to suppress pVHL-defective tumor growth. PLoS Biol 2003 Dec;1(3):E83.

(206) Raval RR, Lau KW, Tran MG, Sowter HM, Mandriota SJ, Li JL, et al. Contrasting properties of hypoxia-inducible factor 1 (HIF-1) and HIF-2 in von Hippel-Lindau-associated renal cell carcinoma. Mol Cell Biol 2005 Jul;25(13):5675-5686.

(207) Imamura T, Kikuchi H, Herraiz MT, Park DY, Mizukami Y, Mino-Kenduson M, et al. HIF-1alpha and HIF-2alpha have divergent roles in colon cancer. Int J Cancer 2009 Feb 15;124(4):763-771.

(208) Acker T, Diez-Juan A, Aragones J, Tjwa M, Brusselmans K, Moons L, et al. Genetic evidence for a tumor suppressor role of HIF-2alpha. Cancer Cell 2005 Aug;8(2):131-141.

(209) Favier J, Lapointe S, Maliba R, Sirois MG. HIF2 alpha reduces growth rate but promotes angiogenesis in a mouse model of neuroblastoma. BMC Cancer 2007 Jul 26;7:139.

(210) Yoshimura H, Dhar DK, Kohno H, Kubota H, Fujii T, Ueda S, et al. Prognostic impact of hypoxia-inducible factors 1alpha and 2alpha in colorectal cancer patients: correlation with tumor angiogenesis and cyclooxygenase-2 expression. Clin Cancer Res 2004 Dec 15;10(24):8554-8560.

(211) Sekine H, Mimura J, Yamamoto M, Fujii-Kuriyama Y. Unique and overlapping transcriptional roles of arylhydrocarbon receptor nuclear translocator (Arnt) and Arnt2 in xenobiotic and hypoxic responses. J Biol Chem 2006 Dec 8;281(49):37507-37516.

(212) Maltepe E, Keith B, Arsham AM, Brorson JR, Simon MC. The role of ARNT2 in tumor angiogenesis and the neural response to hypoxia. Biochem Biophys Res Commun 2000 Jun 24;273(1):231-238.

(213) Bloom BR, Bennett B. Mechanism of a reaction in vitro associated with delayed-type hypersensitivity. Science 1966 Jul 1;153(731):80-82.

(214) David JR. Delayed hypersensitivity in vitro: its mediation by cell-free substances formed by lymphoid cell-antigen interaction. Proc Natl Acad Sci U S A 1966 Jul;56(1):72-77.

(215) Calandra T, Bernhagen J, Metz CN, Spiegel LA, Bacher M, Donnelly T, et al. MIF as a glucocorticoid-induced modulator of cytokine production. Nature 1995 Sep 7;377(6544):68-71.

(216) Bernhagen J, Calandra T, Mitchell RA, Martin SB, Tracey KJ, Voelter W, et al. MIF is a pituitary-derived cytokine that potentiates lethal endotoxaemia. Nature 1993 Oct 21;365(6448):756-759.

(217) Churchill WH,Jr, Piessens WF, Sulis CA, David JR. Macrophages activated as suspension cultures with lymphocyte mediators devoid of antigen become cytotoxic for tumor cells. J Immunol 1975 Sep;115(3):781-786.

(218) Nathan CF, Remold HG, David JR. Characterization of a lymphocyte factor which alters macrophage functions. J Exp Med 1973 Feb 1;137(2):275-290.

(219) Nathan CF, Karnovsky ML, David JR. Alterations of macrophage functions by mediators from lymphocytes. J Exp Med 1971 Jun 1;133(6):1356-1376.

(220) Bacher M, Metz CN, Calandra T, Mayer K, Chesney J, Lohoff M, et al. An essential regulatory role for macrophage migration inhibitory factor in T-cell activation. Proc Natl Acad Sci U S A 1996 Jul 23;93(15):7849-7854.

(221) Yao K, Shida S, Selvakumaran M, Zimmerman R, Simon E, Schick J, et al. Macrophage migration inhibitory factor is a determinant of hypoxia-induced apoptosis in colon cancer cell lines. Clin Cancer Res 2005 Oct 15;11(20):7264-7272.

(222) Koong AC, Denko NC, Hudson KM, Schindler C, Swiersz L, Koch C, et al. Candidate genes for the hypoxic tumor phenotype. Cancer Res 2000 Feb 15;60(4):883-887.

(223) Welford SM, Bedogni B, Gradin K, Poellinger L, Broome Powell M, Giaccia AJ. HIF1alpha delays premature senescence through the activation of MIF. Genes Dev 2006 Dec 15;20(24):3366-3371.

(224) Winner M, Koong AC, Rendon BE, Zundel W, Mitchell RA. Amplification of tumor hypoxic responses by macrophage migration inhibitory factor-dependent hypoxia-inducible factor stabilization. Cancer Res 2007 Jan 1;67(1):186-193.

(225) Fingerle-Rowson G, Petrenko O, Metz CN, Forsthuber TG, Mitchell R, Huss R, et al. The p53-dependent effects of macrophage migration inhibitory factor revealed by gene targeting. Proc Natl Acad Sci U S A 2003 Aug 5;100(16):9354-9359.

(226) Hudson JD, Shoaibi MA, Maestro R, Carnero A, Hannon GJ, Beach DH. A proinflammatory cytokine inhibits p53 tumor suppressor activity. J Exp Med 1999 Nov 15;190(10):1375-1382.

(227) Raz A, Levine G, Khomiak Y. Acute local inflammation potentiates tumor growth in mice. Cancer Lett 2000 Feb 1;148(2):115-120.

(228) Bucala R, Donnelly SC. Macrophage migration inhibitory factor: a probable link between inflammation and cancer. Immunity 2007 Mar;26(3):281-285.

(229) Bini L, Magi B, Marzocchi B, Arcuri F, Tripodi S, Cintorino M, et al. Protein expression profiles in human breast ductal carcinoma and histologically normal tissue. Electrophoresis 1997 Dec;18(15):2832-2841.

(230) Kamimura ,Akira, Kamachi ,Masafumi, Nishihira ,Jun, Ogura ,Shigeaki, Isobe ,Hiroshi, Dosaka-Akita ,Hirotoshi, et al. Intracellular distribution of macrophage migration inhibitory factor predicts the prognosis of patients with adenocarcinoma of the lung.

(231) del Vecchio MT, Tripodi SA, Arcuri F, Pergola L, Hako L, Vatti R, et al. Macrophage migration inhibitory factor in prostatic adenocarcinoma: correlation with tumor grading and combination endocrine treatment-related changes. Prostate 2000 Sep 15;45(1):51-57.

(232) Meyer-Siegler KL, Bellino MA, Tannenbaum M. Macrophage migration inhibitory factor evaluation compared with prostate specific antigen as a biomarker in patients with prostate carcinoma. Cancer 2002 Mar 1;94(5):1449-1456.

(233) Ogawa H, Nishihira J, Sato Y, Kondo M, Takahashi N, Oshima T, et al. An antibody for macrophage migration inhibitory factor suppresses tumour growth and inhibits tumour-associated angiogenesis. Cytokine 2000 Apr;12(4):309-314.

(234) Ren Y, Chan HM, Li Z, Lin C, Nicholls J, Chen CF, et al. Upregulation of macrophage migration inhibitory factor contributes to induced N-Myc expression by the activation of ERK signaling pathway and increased expression of interleukin-8 and VEGF in neuroblastoma. Oncogene 2004 May 20;23(23):4146-4154.

(235) Ren Y, Law S, Huang X, Lee PY, Bacher M, Srivastava G, et al. Macrophage migration inhibitory factor stimulates angiogenic factor expression and correlates with differentiation and lymph node status in patients with esophageal squamous cell carcinoma. Ann Surg 2005 Jul;242(1):55-63.

(236) Tomiyasu M, Yoshino I, Suemitsu R, Okamoto T, Sugimachi K. Quantification of macrophage migration inhibitory factor mRNA expression in non-small cell lung cancer tissues and its clinical significance. Clin Cancer Res 2002 Dec;8(12):3755-3760.

(237) Suzuki F, Nakamaru Y, Oridate N, Homma A, Nagahashi T, Yamaguchi S, et al. Prognostic significance of cytoplasmic macrophage migration inhibitory factor expression in patients with squamous cell carcinoma of the head and neck treated with concurrent chemoradiotherapy. Oncol Rep 2005 Jan;13(1):59-64.

(238) Yasasever V, Camlica H, Duranyildiz D, Oguz H, Tas F, Dalay N. Macrophage migration inhibitory factor in cancer. Cancer Invest 2007 Dec;25(8):715-719.

(239) Shirozu M, Nakano T, Inazawa J, Tashiro K, Tada H, Shinohara T, et al. Structure and chromosomal localization of the human stromal cell-derived factor 1 (SDF1) gene. Genomics 1995 Aug 10;28(3):495-500.

(240) Fernandis AZ, Prasad A, Band H, Klosel R, Ganju RK. Regulation of CXCR4-mediated chemotaxis and chemoinvasion of breast cancer cells. Oncogene 2004 Jan 8;23(1):157-167.

(241) Nagasawa T, Tachibana K, Kishimoto T. A novel CXC chemokine PBSF/SDF-1 and its receptor CXCR4: their functions in development, hematopoiesis and HIV infection. Semin Immunol 1998 Jun;10(3):179-185.

(242) Salcedo R, Wasserman K, Young HA, Grimm MC, Howard OM, Anver MR, et al. Vascular endothelial growth factor and basic fibroblast growth factor induce expression

of CXCR4 on human endothelial cells: In vivo neovascularization induced by stromal-derived factor-1alpha. Am J Pathol 1999 Apr;154(4):1125-1135.

(243) Wendt MK, Cooper AN, Dwinell MB. Epigenetic silencing of CXCL12 increases the metastatic potential of mammary carcinoma cells. Oncogene 2007 09/03;27(10):1461-1471.

(244) Drury LJ, Wendt MK, Dwinell MB. CXCL12 chemokine expression and secretion regulates colorectal carcinoma cell anoikis through Bim-mediated intrinsic apoptosis. PLoS One 2010 Sep 22;5(9):e12895.

(245) Akishima-Fukasawa Y, Nakanishi Y, Ino Y, Moriya Y, Kanai Y, Hirohashi S. Prognostic significance of CXCL12 expression in patients with colorectal carcinoma. Am J Clin Pathol 2009 Aug;132(2):202-10; quiz 307.

(246) Brand S, Dambacher J, Beigel F, Olszak T, Diebold J, Otte JM, et al. CXCR4 and CXCL12 are inversely expressed in colorectal cancer cells and modulate cancer cell migration, invasion and MMP-9 activation. Exp Cell Res 2005 Oct 15;310(1):117-130.

(247) Wendt MK, Johanesen PA, Kang-Decker N, Binion DG, Shah V, Dwinell MB. Silencing of epithelial CXCL12 expression by DNA hypermethylation promotes colonic carcinoma metastasis. Oncogene 2006 Aug 17;25(36):4986-4997.

(248) Hidalgo-Pascual M, Galan JJ, Chaves-Conde M, Ramirez-Armengol JA, Moreno C, Calvo E, et al. Analysis of CXCL12 3'UTR G>A polymorphism in colorectal cancer. Oncol Rep 2007 Dec;18(6):1583-1587.

(249) Csiszar K, Fong SF, Ujfalusi A, Krawetz SA, Salvati EP, Mackenzie JW, et al. Somatic mutations of the lysyl oxidase gene on chromosome 5q23.1 in colorectal tumors. Int J Cancer 2002 Feb 10;97(5):636-642.

(250) Smith-Mungo LI, Kagan HM. Lysyl oxidase: properties, regulation and multiple functions in biology. Matrix Biol 1998 Feb;16(7):387-398.

(251) Denko NC, Fontana LA, Hudson KM, Sutphin PD, Raychaudhuri S, Altman R, et al. Investigating hypoxic tumor physiology through gene expression patterns. Oncogene 2003 Sep 1;22(37):5907-5914.

(252) Kenyon K, Contente S, Trackman PC, Tang J, Kagan HM, Friedman RM. Lysyl oxidase and rrg messenger RNA. Science 1991 Aug 16;253(5021):802.

(253) Contente S, Kenyon K, Rimoldi D, Friedman RM. Expression of gene rrg is associated with reversion of NIH 3T3 transformed by LTR-c-H-ras. Science 1990 Aug 17;249(4970):796-798.

(254) Csiszar ,Katalin, Fong S,F.T., Ujfalusi ,Aniko, Krawetz S,A., Salvati E,P., Mackenzie J,W., et al. Somatic mutations of the lysyl oxidase gene on chromosome 5q23.1 in colorectal tumors.

(255) Fong SF, Dietzsch E, Fong KS, Hollosi P, Asuncion L, He Q, et al. Lysyl oxidase-like 2 expression is increased in colon and esophageal tumors and associated with less differentiated colon tumors. Genes Chromosomes Cancer 2007 Jul;46(7):644-655.

(256) Kirschmann DA, Seftor EA, Fong SF, Nieva DR, Sullivan CM, Edwards EM, et al. A molecular role for lysyl oxidase in breast cancer invasion. Cancer Res 2002 Aug 1;62(15):4478-4483.

(257) Gorogh T, Weise JB, Holtmeier C, Rudolph P, Hedderich J, Gottschlich S, et al. Selective upregulation and amplification of the lysyl oxidase like-4 (LOXL4) gene in head and neck squamous cell carcinoma. J Pathol 2007 May;212(1):74-82.

(258) Kim Y, Boyd CD, Csiszar K. A new gene with sequence and structural similarity to the gene encoding human lysyl oxidase. J Biol Chem 1995 Mar 31;270(13):7176-7182.

(259) HapMap Data Rel24/PhaseII Nov08, on NCBI B36 assembly, dbSNP B126 .

(260) Haploview (ver 4.2) .

(261) Mayo O. A century of Hardy-Weinberg equilibrium. Twin Res Hum Genet 2008;Jun11(3):249-56.

(262) Santiago Rodriguez, Tom R. Gaunt and Ian N. M. Day.
Hardy-Weinberg Equilibrium Testing of Biological Ascertainment for Mendelian Randomization Studies.
American Journal of Epidemiology Advance Access published on January 6, 2009, DOI 10.1093/aje/kwn359. .

(263) Lewis CM. Genetic association studies: design, analysis and interpretation. Brief Bioinform 2002 Jun;3(2):146-153.

(264) Svastova E, Hulikova A, Rafajova M, Zat'ovicova M, Gibadulinova A, Casini A, et al. Hypoxia activates the capacity of tumor-associated carbonic anhydrase IX to acidify extracellular pH. FEBS Lett 2004 Nov 19;577(3):439-445.

(265) Weljie AM, Jirik FR. Hypoxia-induced metabolic shifts in cancer cells: moving beyond the Warburg effect. Int J Biochem Cell Biol 2011 Jul;43(7):981-989.

(266) Rohwer N, Cramer T. Hypoxia-mediated drug resistance: novel insights on the functional interaction of HIFs and cell death pathways. Drug Resist Updat 2011 Jun;14(3):191-201.

(267) Freedman KB, Back S, Bernstein J. Sample size and statistical power of randomised, controlled trials in orthopaedics. J Bone Joint Surg Br 2001 Apr;83(3):397-402.

(268) Savas S, Younghusband HB. dbCPCO: a database of genetic markers tested for their predictive and prognostic value in colorectal cancer. Hum Mutat 2010 Aug;31(8):901-907.

(269) Ortiz-Barahona A, Villar D, Pescador N, Amigo J, del Peso L. Genome-wide identification of hypoxia-inducible factor binding sites and target genes by a probabilistic model integrating transcription-profiling data and in silico binding site prediction. Nucleic Acids Res 2010 Apr;38(7):2332-2345.

(270) 1000 Genomes Project. Available at: http://www.1000genomes.org/data.

(271) Negandhi AA, Hyde A, Dicks E, Pollett W, Younghusband BH, Parfrey P, et al. MTHFR Glu429Ala and ERCC5 His46His polymorphisms are associated with prognosis in colorectal cancer patients: analysis of two independent cohorts from Newfoundland. PLoS One 2013 Apr 23;8(4):e61469.

(272) Savas S, Hyde A, Stuckless SN, Parfrey P, Younghusband HB, Green R. Serotonin transporter gene (SLC6A4) variations are associated with poor survival in colorectal cancer patients. PLoS One 2012;7(7):e38953.

(273) Nucleotide BLAST (blastn) http://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&BLAST_PROGRAMS=megaBlast&PAGE_TYPE=BlastSearch&SHOW_DEFAULTS=on&LINK_LOC=blasthome.

(274) Catalogue of Somatic Mutations in Cancer - COSMIC browser http://www.sanger.ac.uk/perl/genetics/CGP/cosmic?action=gene&ln=FASLG.

(275) Database of Genomic Variants http://projects.tcag.ca/variation/search.asp?keyword=FASLG&source=hg18&table=Locus&filter=&show=table&exactMatch=no

(276) Ruiz-Pesini, E., Lott, M.T., Procaccio, V., Poole, J., Brandon, M.C., Mishmar, D., Yi, C., Kreuziger, J., Baldi, P., and Wallace, D.C. 2007. An enhanced MITOMAP with a global mtDNA mutational phylogeny. Nucleic Acids Research 35 (Database issue):D823-D828.