GENETIC FACTORS CORRELATED WITH SURVIVAL IN COLORECTAL CANCER: VALIDATION STUDIES IN PATIENTS FROM NEWFOUNDLAND AND LABRADOR

AMIT A. NEGANDHI

Genetic factors correlated with survival in colorectal cancer: validation studies in patients from Newfoundland and Labrador

by

Amit A. Negandhi

Submitted to the School of Graduate Studies

in partial fulfillment of the requirements for the degree of

Master of Science in Medicine (Human Genetics)

Faculty of Medicine, Discipline of Genetics

Memorial University of Newfoundland

St. John's, Newfoundland and Labrador, Canada

Abstract

Colorectal cancer is a significant health concern in the province of Newfoundland and Labrador (NL) which has the highest age-standardized incidence and mortality rates for colorectal cancer in Canada. Several studies have attempted to identify inherited genetic variants which can serve as independent prognostic markers in colorectal cancer patients. We have conducted such a study in two colorectal cancer patient cohorts (discovery and validation sets) from Newfoundland. We investigated 27 genetic polymorphisms in the discovery cohort and attempted to replicate the positive correlations in the validation cohort. Our results showed that the *MTHFR*_Glu429Ala polymorphism was associated with worse overall survival in two cohorts albeit with an apparently different pattern of inheritance. An association of the heterozygote genotype of this polymorphism with shorter overall survival was also detected in male patients from both cohorts. Another polymorphism, *ERCC5*_His46His, was also found to be associated with disease-free survival in these cohorts. Further studies on these polymorphisms may facilitate understanding of the mechanisms behind prognostic differences among colorectal cancer patients and aid in better prediction of clinical outcomes.

Contents

Acknowledgements

This graduate program has been an enriching experience in many ways. I started the program aspiring to be an exceptional researcher and this graduate program has groomed me in a wonderful way to fulfill that aspiration. Foremost, I thank my thesis supervisor Dr. Sevtap Savas for her constant guidance throughout the duration of this graduate program. The amount of time, attention and efforts she committed to my progress is truly remarkable. I am grateful to her for giving me the opportunity to work on this project.

I was fortunate to have a great team of committee members, Dr. Roger Green and Dr. William Pollett. I thank them for their valuable comments and time. I also thank the CIHR Interdisciplinary Health Research Team in Colorectal Cancer for awarding me a fellowship.

I thank Jessica Squires for technical assistance and Asan Haja Mohideen for his company in the lab. It has been a pleasure to have interacted with numerous great people in the Discipline of Genetics. Dr. Mike Woods and his lab members have always been very kind and helpful. I also thank Deborah Quinlan for all her help.

Lastly, staying away from my family for the first time, I have realized the importance of a good upbringing. I thank my parents for educating me and raising me with good values.

List of Figures

List of Tables

Abbreviations

2R: 2-repeats

3R: 3-repeats

4R: 4-repeats

AB: Alberta

AD: allelic discrimination

AGTC: Analytical Genetics Technology Centre

AJCC: American Joint Committee on Cancer

ALB: albumin

APC: adenomatous polyposis coli

AQ: absolute quantification

BC: British Columbia

bp: base pairs

BRAF1: v-raf murine sarcoma viral oncogene homolog B1

CAP: College of American Pathologists

CCND1: cyclin D1

CDKI: cyclin-dependent kinase inhibitor

CEA: carcinoembryonic antigen

CI: confidence interval

CIMP: CpG island methylator phenotype

CIN: chromosomal instability

CNV: copy number variation

CRM: circumferential resection margin

DCC: deleted in colorectal carcinoma

DFS: disease-free survival

DNA: deoxyribonucleic acid

EGFR: epidermal growth factor receptor

ERCC5: excision repair cross-complementing rodent repair deficiency, complementation group 5

EXO1: exonuclease 1

FAP: familial adenomatous polyposis

FAS: Fas (TNF receptor superfamily, member 6)

FCCTX: familial colorectal cancer type X

FFPE: formalin-fixed paraffin-embedded

FGFR: fibroblast growth factor receptor

G1: growth 1

GSTM1: glutathione S-transferase mu-1

GSTP1: glutathione S-transferase pi-1

GSTT1: glutathione S-transferase theta-1

GWAS: genome wide association study

HIC: Human Investigation Committee

HNPCC: hereditary non-polyposis colorectal cancer

HR: hazard ratio

HWE: Hardy-Weinberg Equilibrium

ID: identifier

IL6: interleukin 6 (interferon, beta-2)

LD: linkage disequilibrium

LOH: loss of heterozygosity

M: molar

mAF: minor allele frequency

MAP: mutYH-associated polyposis

MB: Manitoba

MMP: matrix metallopeptidase

MMR: mismatch repair

mRNA: messenger ribonucleic acid

MSI: microsatellite instability

MSI-H: microsatellite instability high

MSI-L: microsatellite instability low

MSS: microsatellite stable

MTHFR: methylene tetrahydrofolate reductase

MUTYH: mutY homolog (E. coli)

NB: New Brunswick

NFCCR: Newfoundland Colorectal Cancer Registry

NL: Newfoundland and Labrador

NS: Nova Scotia

NTC: non-template control

OGG1: 8-oxoguanine DNA glycosylase

ON: Ontario

OS: overall survival

PCR: polymerase chain reaction

PE: Prince Edward Island

PFS: progression-free survival

PTGS2: prostaglandin-endoperoxide synthase 2

QC: Quebec

S: synthesis

SDS: sequence detection system

SK: Saskatchewan

SNP: single nucleotide polymorphism

TBE: tris-borate-EDTA

TNM: tumor node metastasis

TYMS: thymidylate synthetase

UHN: University Health Network

UTR: untranslated region

UV: ultraviolet

V: volts

VNTR: variable number of tandem repeats

VEGF: vascular endothelial growth factor

WHO: World Health Organization

XRCC3: X-ray repair complementing defective repair in Chinese hamster cells 3

Chapter 1. Colorectal cancer

1.1 Cancer

Cancer is a disease characterized by uncontrollable division of certain abnormal cells which can develop into a tumor that can invade tissues or spread to distant organs (1). Over one hundred types of cancers have been identified based on the cell types in which they develop (1). Instability of the genome making the cell's deoxyribonucleic acid (DNA) hyper-mutable as well as increased inflammation that can favor carcinogenesis are recognized as the two primary reasons which can enable normal cells to acquire cancerous properties (2). Through the course of development of cancer cells, distinct proliferative abilities are acquired in a successive manner. Hanahan and Weinberg described these unique attributes of cancer cells as 'hallmarks of cancer' (2). Cancer cells have prolonged cellular growth signaling for proliferation which can be due to self-production of growth factors, induction of growth factor production in the surrounding normal cells, high sensitivity to growth factors due to changes in receptor structure or continually triggered pathways downstream of receptors (2). Normal cell proliferation is also controlled by the action of tumor suppressor genes which inhibit proliferation and growth in unfavorable conditions and can also induce cell senescence and death. Cancer cells escape the suppressive action of these genes to continue proliferating. A dysfunctional contact inhibition mechanism, which prevents excessive proliferation of cells under normal conditions, also contributes to continued proliferation in cancer cells (2). Normal cells have a way of regulating cell proliferation through induction of apoptosis which causes death of highly stressed and abnormal cells such as cancer cells. But cancer cells evade apoptosis via multiple mechanisms like loss of function or reduced activity of apoptotic factors and up-regulation of counter-apoptotic factors (2). In addition to such prolific properties, cancer cells have an added ability to be immortal, likely due to the maintenance of telomere lengths at the end of chromatids after each replication (2). This ability to replicate endlessly enables formation of a fully grown macroscopic tumor from microscopic cancerous cells. And like all tissues in the body, the growing tumor also requires a constant supply of blood and nutrients. This is facilitated by formation of new tumor vasculature by up-regulating pro-angiogenic factors early in neoplastic development (2). With advancing growth, the tumor cells begin to penetrate the surrounding normal tissues and vasculature, then spread to distant organs via blood and/or lymph vessels and develop into micrometastases and eventually grow into metastatic tumors. Cancer cells may also have the ability to modify cell metabolic processes in a way to favor tumorigenesis as well as evade destruction by the immune system (2). Evidently, cancer is a highly complex disease involving aberrations in multiple genes operating in multiple pathways, the accumulation of which can lead to initiation of cancer which can then grow into lethal forms by modifying cellular functions to suit its survival.

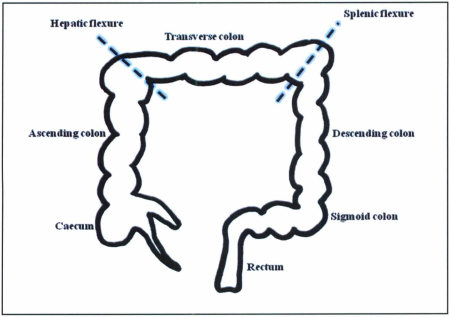
1.2 Structure and functions of the colon and rectum

The colon, also known as the large intestine or large bowel, is approximately 1.5 meters long (3). The colon begins as the caecum and progresses into the ascending colon, transverse colon, descending colon and sigmoid colon. The colon terminates in the rectum which opens exteriorly into the anal canal (**Figure 1**). A sharp curve at the level of the liver is known as the hepatic flexure and one at the level of spleen is known as the splenic flexure (3). Histologically, the colon and rectum are lined by 4 basic membranes. Beginning outwards, they are (3):

1) Visceral peritoneum: The outermost serous membrane.
2) Muscle layers: They are arranged as longitudinal and circular fibres.
3) Submucosa: This layer contains networks of nerves, blood vessels, lymph vessels and lymphoid tissue. For defence against microbial infections, the submucosa in colon has greater amount of lymphoid tissue compared to other parts of the alimentary canal.
4) Mucosa: This is composed of three layers of tissue. Starting inwards, they are:
    i. Mucous membrane-innermost layer of columnar epithelial cells responsible for absorption, secretion and protection.
    ii. Lamina propria-loose connective tissue layer responsible for support and protection.
    iii. Muscularis mucosa-provides involutions to the mucous layer.

The primary function of colon is to absorb water from the matter that arrives from the small intestines (3). This results in the formation of the fecal matter. The fecal matter then moves along the colon and to the rectum where it propelled by muscle movements to the anal canal for expulsion. The colon also expels swallowed air and gases produced by

Figure 1. Structure of colon and rectum



Adapted from 'Principles of Anatomy and Physiology' (4) bacterial action on unabsorbed food matter. The large amount of lymphoid tissue in the colonic submucosa protects the colon from microbial infections as the fecal matter is rich in microbes (3).

1.3 Colorectal cancer: Molecular mechanisms and pathology

Cancer of the colonic tissue is called 'colon cancer' while that of the rectal tissue is called 'rectal cancer' and they are referred together as colorectal cancer (5). Development and growth of colorectal cancer involve multiple and sequential changes in the genome such as destabilizing the genome by mutations that inactivate chromosome stabilizing genes, defects in DNA repair machinery, epigenetic silencing by DNA methylation, deactivation of tumor suppressor genes and activation of proto-oncogenes to oncogenes (6,7). This series of changes eventually manifests pathologically as colorectal cancer. According to the inheritance patterns, there are two forms of colorectal cancer:

i. Familial and inherited forms of colorectal cancers with familial clustering. In the case of inherited forms, there is a strong hereditary predisposition.

ii. Sporadic forms without a strong hereditary predisposition.

The familial and inherited forms comprise approximately 15-25% of all colorectal cancer syndromes while the sporadic forms comprise the majority with 70-85% of the cases (8-12). The inherited and sporadic forms may involve different genetic and molecular mechanisms. Inherited forms are due to high-penetrant mutations in critical genes (8). Examples of inherited forms include:

1) Lynch syndrome (previously known as Hereditary Non-polyposis Colorectal Cancer (HNPCC)) is characterized by germline mutations in the mismatch repair genes (MMR) such as *MLH1, MSH2, MSH6* and *PMS2* (8), leading to the microsatellite instability (MSI) phenotype in tumors.
2) Familial adenomatous polyposis (FAP) is an autosomal dominant form of colorectal cancer caused by the germline mutations in the adenomatous polyposis coli (*APC*) gene (13).
3) mutY homolog (E. coli) (*MUTYH*)-associated polyposis (MAP) is an autosomal recessive disease where mutations in *MUTYH* gene predispose the individual to colorectal cancer (11).
4) Examples of other rare forms of colorectal cancer syndromes are Juvenile Polyposis, Peutz-Jeghers Syndrome, Cowden disease and Bannayan-Ruvalcaba-Riley Syndrome (8).

The incompletely understood Familial colorectal cancer type X (FCCTX) is a form with a strong familial clustering of colorectal cancer but no well-defined hereditary predisposition or molecular mechanism (11,14,15). This form is distinct from the Lynch syndrome in terms of age of onset, tumor histology, tumor grade and absence of deficient MMR (16). Recent developments suggest that molecular mechanisms involved in chromosomal instability may be involved in development of FCCTX (16).

In sporadic colorectal cancer cases, a strong genetic predisposition may not exist. Rather, interaction of several low susceptibility alleles and environmental factors are proposed to results in carcinogenesis. Genome-wide association studies (GWAS) have identified at least 14 such low-susceptibility genetic variants that increase the risk of developing colorectal cancer (17).

Molecular mechanisms involved in sporadic forms of colorectal cancer are:

1) Chromosomal instability (CIN): Characterized by numerical or structural abnormalities in the chromosomes causing damage to tumor suppressor genes or oncogenes (18).
2) Defective MMR system leading to MSI: In sporadic cases, MSI is due to hypermethylation of the promoter of the mismatch repair gene *MLH1* leading to its silencing (19).
3) CpG island methylator phenotype (CIMP): In CIMP, the CpG islands are methylated causing inactivation of certain genes (20).

Histological types of colorectal cancer: Pathologically, at least eight different histological types of epithelial tumors have been defined by the World Health Organization (WHO) (21). Adenoma is the early benign tumor. Adenocarcinoma is the malignant type, shows moderate differentiation and can be either mucinous or non-mucinous (22). It is the most commonly observed histological type of colorectal cancer (~90-95%) (21,23). Mucinous adenocarcinoma, in which the tumor cells secrete mucin (> 50% of tumor mass is due to mucin) is found in up to 17% of tumors while the majority of adenocarcinomas are non-mucinous (21-23). Other rarer pathological forms are signet-cell carcinoma, squamous cell carcinoma, adenosquamous carcinoma, small cell carcinoma and medullary carcinoma (21,24).

1.4 Colorectal cancer incidence and mortality statistics

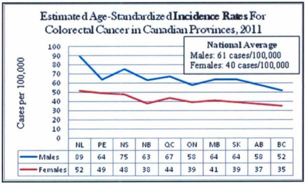
1.4.1 Worldwide incidence and mortality: According to WHO's report "The global burden of disease. 2004 Update" (25), colorectal cancer was responsible for approximately 639,000 deaths worldwide with 336,000 male deaths and 303,000 female deaths. On the list of lethal cancers in terms of number of cancer deaths, colorectal cancer was the 4th major global killer in the year 2004 (25). Of all the cancers worldwide, colorectal cancer ranks the 4th in men and the 3rd in women in terms of incidence (26). The general trend observed worldwide is high incidence of this disease predominantly in the western world such as North America, Australia and European countries and low incidence in South American, Asian and African populations (26).

1.4.2 Colorectal cancer in Canada: Among all cancers (excluding non-melanoma skin cancers), the incidence of colorectal cancer across Canada was expected to be the 4th highest with 22,200 estimated new cases in 2011 (27). In 2011, the mortality due to colorectal cancer was expected to be the 2nd highest among all cancers with 8,900 patients estimated to die because of it (27). Relative survival rate of colorectal cancer patients (survival of colorectal cancer patients compared to that of the general population from the same region) over a 5-year period is 63-64% (27). It is reported that the Atlantic Provinces in Canada have higher colorectal cancer incidence and mortality rates when compared to western provinces like Alberta (AB) and British Columbia (BC) (27). Multiple factors such as lifestyle factors (exercise, diet), family history, intensity of screening programs, differential participation as well as quality and availability of healthcare and diagnostic services may account for this inter-provincial variation in colorectal cancer incidence and mortality rates (27).

Figures 2 and 3 show the inter-provincial variation and the east-west gradient in incidence and mortality rates of colorectal cancer across Canada (27). NL shows the highest age-standardized incidence and mortality rates for both men and women. Other Atlantic provinces such as Prince Edward Island (PE), Nova Scotia (NS) and New Brunswick (NB) as well as Quebec (QC) have higher incidence and mortality rates compared to the western provinces of AB and BC.

1.4.3 Colorectal cancer in Newfoundland and Labrador (NL): When Canadian provinces are compared, the age-standardized incidence rate is the highest for both males and females from NL (27). Eighty nine cases per 100,000 new male colorectal cancer patients were expected in NL in 2011 while the national expected rate was 61/100,000. For females, fifty two new cases per 100,000 were expected in NL while the national average of incidence for females was 40/100,000 (**Figure 2**). Also, according to the Canadian Cancer Statistics 2011, men and women patients from NL have the highest age-standardized colorectal cancer mortality rates across Canada (27). Forty-five deaths per 100,000 men were expected in NL in 2011 while the national average was 25 deaths/100,000 males. For females, twenty-three deaths per 100,000 are expected in NL while the national expected number of deaths is 15 deaths/100,000 (**Figure 3**). These statistics show the relatively greater burden of colorectal cancer in NL when compared to other Canadian provinces.

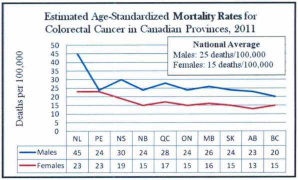
Figure 2. Estimated age-standardized incidence rates for colorectal cancer in Canadian provinces, 2011



	NL	PE	NS	NB	QC	ON	MB	SK	AB	BC
Males	89	64	75	63	67	58	64	64	58	52
Females	52	49	48	38	44	39	41	39	37	35

NL-Newfoundland & Labrador, PE-Prince Edward Island, NS-Nova Scotia, NB-New Brunswick, QC-Quebec, ON-Ontario, MB-Manitoba, SK-Saskatchewan, AB-Alberta, BC-British Columbia
Data for the figure obtained from Canadian Cancer Statistics 2011 (27)

Figure 3. Estimated age-standardized mortality rates for colorectal cancer in Canadian provinces, 2011



	NL	PE	NS	NB	QC	ON	MB	SK	AB	BC
Males	45	24	30	24	28	24	26	24	23	20
Females	23	23	19	15	17	15	16	15	13	15

NL-Newfoundland & Labrador, PE-Prince Edward Island, NS-Nova Scotia, NB-New Brunswick, QC-Quebec, ON-Ontario, MB-Manitoba, SK-Saskatchewan, AB-Alberta, BC-British Columbia.
Data for the figure obtained from Canadian Cancer Statistics 2011 (27)

1.5 Prognosis

Prognosis is the prediction of the course of a disease leading to specific health conditions, known as clinical outcomes, after diagnosis of the disease (28). The US National Library of Medicine defines clinical outcome as *"a measure of how a patient (or study subject) feels, functions, or survives; or a clinical measurement of the incidence or severity of a disease (e.g., diagnosis of disease)"* (29). Clinical outcomes in cancer include recurrence of cancer, metastasis or death. Two of the commonly used measures of clinical outcome, which are also the end-points analyzed in this thesis project are overall survival (OS) and disease free survival (DFS). While their definitions may change from one study to other, we refer to OS and DFS in this study as defined below.

i. **OS:** It is the survival period of the patient from the time of diagnosis until his/her death from any cause. OS rate, usually expressed as a 5-year survival rate, is the proportion of patients alive five years after diagnosis of the disease.

ii. **DFS:** DFS is the survival of patients after diagnosis without relapse (i.e. recurrence or metastasis) or death from any cause.

1.5.1 Factors affecting prognosis in colorectal cancer patients: Prognosis and clinical outcomes in cancer patients are highly variable and dependent on multiple factors. Currently, the tumor-node-metastasis (TNM) staging is the standard tool for prognostication in colorectal cancer patients (30). The TNM stage is a measure of the extent of tissue invasion by the tumor (T) and metastasis to lymph nodes (N) or distant organs (M). The TNM staging published by American Joint Committee on Cancer (AJCC) is the widely accepted standard for staging of colorectal cancer (30). The latest classification (published 2010) is depicted in **Table 1**.

In addition, there are a large number of acknowledged prognostic factors but their use in clinical practice is limited. In 1999, the College of American Pathologists (CAP) convened a consensus statement (31) categorizing the prognostic factors in colorectal cancer into five categories:

Category I: It includes factors which are conclusively established to have prognostic value based on the results of multiple trials considered statistically robust. These factors are routinely used in the clinic for patient management. This category includes depth of tumor invasion (T of TNM staging), metastasis to regional lymph nodes (N of TNM staging), lymphatic or vascular invasion, presence of residual tumor after surgical removal and levels of pre-operative carcinoembryonic antigen (CEA) in the serum.

Category IIA: This category includes factors which are considered important for inclusion in pathology reports and have repeatedly shown prognostic relevance. However, they await validation in large studies. This category includes tumor grade, circumferential resection margins (CRM) and tumor staging after neoadjuvant therapy.

Category IIB: This category includes factors which show prognostic relevance in multiple studies but further studies are needed for inclusion in category I or IIA. It includes tumor histology, MSI status in tumor cells, loss of heterozygosity (LOH) at 18q, allelic loss of *DCC* gene and the configuration of tumor border.

Table 1. Stage grouping for colorectal cancer

Stage Designation	TNM Characteristics
stage 0	Tis, N0, M0
stage I	T1, N0, M0 **or**
	T2, N0, M0
stage IIA	T3, N0, M0
stage IIB	T4a, N0, M0
stage IIC	T4b, N0, M0
stage IIIA	T1-T2, N1/N1c, M0 **or**
	T1, N2a, M0
stage IIIB	T3-T4a, N1/N1c, M0 **or**
	T2-T3, N2a, M0 **or**
	T1-T2, N2b, M0
stage IIIC	T4a, N2a, M0 **or**
	T3-T4a, N2b, M0 **or**
	T4b, N1-N2, M0
stage IVA	any T, any N, M1a
stage IVB	any T, any N, M1b

Tis=carcinoma in situ limited to lamina propria or basement membrane, T1=submucosal layer invaded by tumor cells, T2=tumor penetrated deeper into muscularis propria, T3=tumor penetrated into sub-serosa or tissues surrounding colon/rectum, T4a=direct penetration through the peritoneum, T4b=direct penetration into or attachment to other organs. N0=no metastasis of tumor cells into regional lymph nodes, N1=1-3 lymph nodes affected, N1a=1 lymph node affected, N1b=2-3 lymph nodes affected, N1c=no metastasis into regional lymph nodes but tumor deposit(s) present, N2a=4-6 lymph nodes affected, N2b=7 or more lymph nodes affected, M0=distant metastasis not observed. M1a=distant metastasis to a single organ/site, M1b=distant metastasis to multiple organs/sites.

Adapted from AJCC Cancer Staging Handbook, 7th Edition (2010) (30)

Category III: This category includes factors which have not been well-studied for their prognostic relevance. It includes DNA content, a large set of putative molecular markers including genes and proteins which may have prognostic roles due to altered function or abnormal expression (tumor suppressor genes affected due to LOH at 1p/p53, 8p, 1p, 5q, oncogenes *(KRAS, MYC),* apoptotic and cell suicide-related genes *(BCL2, BAX),* genes involved in DNA synthesis, growth factor-related genes *(TGF, EGFR),* cyclin-dependent kinase inhibitor genes *(CDKIs)*, genes involved in angiogenesis *(VEGF),* glycoprotein genes and adhesion molecules (E-cadherin, sialo-Tn antigen, CD44), matrix metalloproteases (MMPs) and inhibitors of MMPs, genes that suppress metastasis *(NME1)*) and other features such as perineural invasion, microvessel density, cell proteins and carbohydrates, peritumoral fibrosis, neuroendocrine differentiation foci, nucleolar organizing regions and proliferation indices.

Category IV: This category includes factors for which absence of prognostic relevance has been well established. It includes tumor size and gross tumor configuration.

A decade later, the 7th edition of AJCC cancer staging manual published in 2010 includes updates and recommendations for improved prognostication based on scientific evidence (30). TNM staging system still remains the most powerful prognostic tool. Stage-independent factors that are used on a general basis include tumor histology, tumor grade, presence/absence of residual tumor after surgical removal, serum CEA levels, serum cytokine levels, extramural venous invasion and vascular invasion into submucosa. However, they are not a part of an objective prognostic tool such as TNM staging. AJCC also recommends collection of eight parameters due to their prognostic significance (30). These are pre-operative serum CEA level, number of tumor deposits detached from primary tumor, tumor regression grade following neoadjuvant therapy to assess response to therapy (grades 0-3, grade 0 indicates total response to therapy and grade 3 indicates worst response), CRM measured from the tumor boundary to the closest margin of surgical removal, MSI status in tumor cells, perineural invasion (i.e invasion around local nerves by tumor cells), mutation status in codons 12 or 13 of *KRAS* gene in tumor cells, especially in advanced stage patients since mutations in these codons are strongly correlated with absence of response to monoclonal antibodies directed against epidermal growth factor receptor (EGFR) and 18q LOH status in tumor cells. Although these factors are not currently a part of a clinical prognostication system such as the TNM staging system, further studies may lead to their incorporation in future editions (30). Hence, the collection of data on these factors in pathology reports is strongly recommended by AJCC. Apart from these molecular and pathological factors, demographic factors such as gender, age and ethnicity may also play a strong role in the variable prognosis in colorectal cancer patients (30). For this thesis project, we used data on ten demographic, clinico-pathological and molecular variables for analysis.

1.5.2 Clinicopathological and molecular variables included in this thesis project

Ten demographic, clinico-pathological and molecular variables included for analysis in this thesis project are briefly described below. The data on these variables were available to us and many of them have been acknowledged by AJCC to have possible prognostic roles in colorectal cancer (30). These variables were included in the study to test their association with patient survival in our cohorts and for adjustment in the multivariate analyses to account for their effects in the model.

a) **Stage:** Stage is the only well-established and routinely used prognostic factor in colorectal cancer patients. The generally observed trend is that patient prognosis worsens with increasing disease stage (30).

b) **Tumor grade:** Based on the apparent differentiation of tumor cells, four tumor grades have been defined: G1 for a well differentiated tumor to G4 for a virtually undifferentiated tumor (30). The AJCC (30) as well as CAP consensus statement (31) recommend a two-tiered classification with low grade (G1 and G2) and high grade (G3 and G4) colorectal tumors. In this project, we have classified patients according to this two-tiered system for analyses. Low grade tumors generally have a low cell proliferation rate and metastatic potential while high grade tumors have a high cell proliferation rate and metastatic potential (32) which has been demonstrated to have a stage-independent adverse prognostic correlation in multiple studies (24). However, since grading is a subjective criterion, designation of a tumor grade varies from one observer to another (31). Due to lack of a widely accepted grading protocol, accurate use of tumor grade in prognostication is difficult and hence limited (24,31).

c) **Vascular/lymphatic invasion:** Presence of vascular or lymphatic invasion has been documented to be associated with unfavorable prognosis (24,30,31) and is routinely included in pathology reports. AJCC recommends inclusion of this information as a part of V and L staging classification (30). However, its objective use as a prognostic marker is limited by several factors. CAP recommends examination in at least 3 tumor blocks (ideally 5 tumor blocks) to conclusively establish presence or absence of invasion (31). This makes the process cumbersome, time-consuming and costly. Moreover, there is no standard protocol for assessing invasion adding undesirable inter-observer variability to the judgement, especially in cases of small and large vessel invasions (24). Due to these reasons, vascular/lymphatic invasion data are not included in an objective prognostication system. In this study, we have included vascular/lymphatic invasion as an exploratory variable in our analyses.

d) **Tumor histology:** After non-mucinous adenocarcinoma, mucinous adenocarcinoma is the next most common histological type of colorectal cancer (21-23). The prognostic significance of mucinous tumor type is undecided due to several conflicting reports (24,31).

e) **MSI status:** Mismatch repair proteins are responsible for correcting wrongly inserted nucleotide bases after DNA replication. Defects in mismatch-repair proteins (MLH1, MSH2, MSH6, PMS2) due to germline mutations can lead to increase or decrease in length of microsatellites which are repeating units of nucleotides, (commonly dinucleotides of cytosine and adenine (CA)), present in thousands of locations in the genome (9). This is termed "microsatellite instability" (MSI) (9). In a large meta-analysis conducted by Popat et al (33) including over 7,500 patients from 32 different studies, it was shown that patients with MSI-high (MSI-H) status have a significantly longer survival when compared with patients with microsatellite stable (MSS) or MSI-low (MSI-L). The 7th edition of AJCC cancer staging manual published in 2010 recommends the collection of MSI-status of patients for prognostic purposes (30).

f) **Tumor location:** Literature reports have consistently suggested that patients with rectal cancers have a worse survival compared to patients with colon cancer (34). However, tumor location is not clinically used as a prognostic factor nor is it considered in the guidelines and recommendations by CAP and AJCC.

g) **Familial risk status:** Familial risk status was assigned to the patients in the Newfoundland Colorectal Cancer Registry (NFCCR) previously as described by Green et al (35). Literature reports on association of familial risk status with prognosis are deficient. Therefore its role in prognosis of patients is not known.

h) ***BRAF1*_Val600Glu mutation status:** v-raf murine sarcoma viral oncogene homolog B1 *(BRAF1)* is a proto-oncogene and is a part of a signal transduction pathway (*Ras/Raf/MEK/MAP* pathway) (36). Activation of this pathway leads to cell proliferation. The somatic Val600Glu mis-sense mutation in *BRAF1* makes it oncogenic. As a result, the gene is continuously activated which causes cell proliferation and inhibited apoptosis (36). The correlation of this mutation with unfavorable prognosis has also emerged in the literature (37-39). For patients in NFCCR, the data on this mutation in tumor samples was collected for a previous study by Wish et al (40).

i) **Age:** It is acknowledged by AJCC that age may play a strong role in prognosis in colorectal cancer patients (30) although it is not a part of a clinical prognostication system yet. Since OS is our primary end-point for analysis, age may be a significant factor since the chances of survival are expected to be reduced with increasing age.

j) **Sex:** Gender is also acknowledged by AJCC to play an important role in variable prognoses in colorectal cancer patients (30) although further studies are required before it can be objectively used for prognostication.

1.5.3 Survival end-points analyzed in this thesis project

Two end-points were analyzed in this thesis project. The primary end-point was OS for which OS status and OS time are required for analysis. OS was our primary end-point since the selected 27 genetic polymorphisms for analysis in this study were associated with OS in at least one study in the literature. The secondary end-point was DFS for which DFS status and DFS time are required for analysis.

a) **OS status:** It indicates if the patient was alive or dead at the time of last follow up. The death of the patient could be due to any cause.

b) **OS time:** It is the time in years from diagnosis of colorectal cancer until death from any cause

c) **DFS status:** It indicates if the patient had recurrence of cancer, metastasis or died from any cause during the follow-up period. In the discovery cohort, recurrence and metastasis were identified using the information from the response to follow-up questionnaires and pathology reports. In the validation cohort, recurrence and metastasis were identified from surgical reports, pathological reports, imaging data and cancer clinic charts.

d) **DFS time:** It is the time in years from diagnosis of colorectal cancer until the first occurrence of the event (recurrence, metastasis or death).

1.6 Genetic variations and genetic prognostic research

Genetic variations can range from large scale structural or numerical karyotypic abnormalities affecting entire chromosomes to changes in single nucleotides (41). Chromosomal aberrations can be either structural where chromosomes have unrepaired or mis-repaired breaks; or numerical where there are more or less than the normal number of chromosomes causing polyploidy or aneuploidy (42). Single Nucleotide Polymorphisms (SNPs) are alterations in a single base in the DNA sequence and it is estimated that there are more than 10 million SNPs in the human genome (43). SNPs can occur within a gene and may alter a coding sequence. A SNP is silent when the substitution in the codon does not change the encoded amino acid, missense when the substituted codon encodes a different amino acid, or nonsense when it creates a stop codon producing truncated protein. SNPs can also occur in the untranslated regions (UTRs), in promoter regions or in splice sites (42). Copy number variations (CNVs) are variations in number of large segments of the DNA arising due to deletion or duplication events, and range from 1 kilobase to several megabases (42). CNVs may include a gene(s) or its parts. Insertion-deletion (indel) polymorphisms involve insertion or deletion of one or few nucleotides to large number of nucleotides in the DNA sequence. Inversion is another type of polymorphism where a sequence is present in an inverted manner in the DNA (42).

Genetic variations can be either germline or somatic. Somatic variations are tissue specific and non-inheritable. An example is the Val600Glu missense mutation in the *BRAF1* gene in tumor cells, such as in colorectal cancer (see **section 1.5.2**). Germline variations are inherited variations and occur in all cell types (44).

A large number of studies have been conducted in the past decade to find polymorphisms associated with prognosis in colorectal cancer. Currently, the identified polymorphisms are not used in the clinical setting as further studies in the field are required (44).

Recently, the commercial Oncotype DX® Colon Cancer Assay was developed by Webber and colleagues using tumor gene expression data for 12-genes in stage-II patients to predict risk of recurrence (45). On similar lines, ColoPrint® prognostic index was developed by Salazar and others and validated using gene expression profiles of 18 genes in colorectal tumor samples (46). If prognostic relevance of a germline variation is established, similar prognostic indices using germline variations may be valuable since germline DNA can easily be obtained from blood.

Of the large number of common germline polymorphisms investigated for their prognostic relevance, the 27 polymorphisms which are a part of this thesis project are discussed in the following section (**section 1.7**). The selection of these polymorphisms is described in **section 3.1** and the literature findings described below are based on the curations posted in the dbCPCO database as of late 2011 (47). These studies are not entirely homogenous in terms of study design, cohort characteristics, treatment regimen and statistical analyses. Hence, it is not surprising to find that several results reported in different studies are conflicting. In addition, study power issues and potential confounders not accounted for in different studies can yield different results.

1.7 Genetic polymorphisms investigated in this study and previous literature findings in colorectal cancer cohorts

1) **rs9344, NG_007375.1:g.12038G>A, Pro241Pro A/G synonymous polymorphism in cyclin D1 *(CCND1)* gene.** The activity of CCND1 protein is required for transition of the cell cycle from growth 1 (G1) phase to the synthesis (S) phase (48). The G allele for this synonymous polymorphism, located in the splice donor site following exon 4, produces an isoform of *CCND1* messenger ribonucleic acid (mRNA) by facilitating alternative splicing (49). In one study, young male patients from Singapore with GG genotype for this polymorphism had shorter cancer-specific survival following surgery in univariate survival analysis (50). In another study, advanced colorectal cancer patients with AA genotype (from a mixed population) treated with the monoclonal antibody cetuximab had poorer OS in univariate survival analysis when compared to patients with GA or GG genotypes (51). Thus the two results were not entire comparable, possibly due to different treatment characteristics and outcomes analyzed. In four other studies, no correlation was observed between this polymorphism and OS in colorectal cancer (52-55).

2) **rs2229080, NG_013341.1:g.571061C>G, Arg201Gly C/G mis-sense polymorphism in the deleted in colorectal carcinoma (*DCC)* gene.** *DCC* is a tumor suppressor gene (56). Schmitt et al (57) reported that the G allele (Gly) of this polymorphism was associated with lowered expression of the *DCC* gene. In a Swedish cohort, colorectal cancer patients homozygous for the C allele (Arg/Arg)

were reported to have better OS when compared to patients having C/G (Arg/Gly) or G/G (Gly/Gly) genotypes in univariate analysis (58), but multivariate analysis was not performed in this study. In another study in an Asian cohort, no correlation was observed with OS in colorectal cancer in univariate analysis (55).

3) **rs2227983, NG_007726.1:g.147531G>A, Arg521Lys G/A in the epidermal growth factor receptor *(EGFR)* gene.** EGFR is a transmembrane protein which upon binding to the EGF, initiates a signaling cascade which leads to cell proliferation (59). Functional characterization of *EGFR*_Arg521Lys polymorphism performed in Chinese hamster ovary cells is suggestive of impaired ligand binding to extracellular domain of EGFR and the reduced ability of EGFR to induce cell growth (60). Patients with metastatic colorectal cancer with an allele encoding lysine amino acid (A/A or G/A genotypes) were reported to have better progression-free survival (PFS) and OS in a French cohort (univariate analysis) (61), favorable OS in cohort of male patients from mixed population (univariate analysis) (62), and better OS in an Asian cohort (multivariate analysis) (63). Thus all these studies reported favorable survival in the presence of the allele encoding lysine. In five other studies, no association was observed between this polymorphism and OS in colorectal cancer (51,52,64-66).

4) **rs11615, NG_015839.1:g.8525T>C, Asn118Asn C/T synonymous polymorphism in excision repair cross-complementing rodent repair deficiency, complementation group 1 *(ERCC1)* gene.** ERCC1 repairs the abnormal lesions in

the DNA by nucleotide excision repair (67). The presence of T allele in Asn118Asn is associated with reduced gene expression by altering codon usage (68). Previously, T allele (CT and TT genotypes) was correlated with worse OS in Asian cohorts in multivariate (69) and univariate analysis (70,71) and TT genotype was correlated with worse PFS in an Italian cohort in univariate analysis (72). In mixed population cohorts, similar correlations were reported; i.e. patients with CC genotypes had better OS in univariate (73) as well as in multivariate analysis (74). A contradictory result was reported in a Spanish cohort (75) where the C allele (CC and CT genotypes) was correlated with worse OS in multivariate analysis. In three other similar studies, no association was observed between rs11615 and OS in colorectal cancer (76-78).

5) **rs13181, NG_007067.2:g.23927A>C, Lys751Gln G/T in *ERCC2* gene.** ERCC2 protein is involved in DNA repair machinery by nucleotide excision repair (79). Cells expressing Lys variant have inefficient DNA repair and abnormalities in chromatids, such as breaks in the DNA strand or damaged unrepaired bases (80). Poor OS in colorectal cancer patients (mixed population) carrying T allele (Gln/Gln and Lys/Gln) was previously found using univariate analysis (74,81). The genotypes for Lys/Lys and Gln/Lys werealso associated with poor PFS compared to patients with genotype for Gln/Gln in the Italian cohort treated with 5-fluorouracil (5-FU), leucovorin and oxaliplatin in multivariate analysis (72). However, in a Chinese patient cohort, also treated with 5-FU, leucovorin and oxaliplatin, homozygotes for lysine (Lys/Lys) had better OS and PFS compared to heterozygotes in multivariate analysis, presumably due to enhanced efficacy of oxaliplatin in patients with poor DNA repair function of

ERCC2 (due to Lys751Gln) (82). Also, in a Turkish cohort of metastatic colorectal cancer patients, Gln/Gln homozygotes had a shorter OS compared to Lys/Lys homozygotes (83). Six other studies reported no significant correlation between this polymorphism and OS in colorectal cancer (71,75,76,78,84,85).

6) **rs1047768, NG_007146.1:g.11344T>C, His46His C/T in *ERCC5* gene.** ERCC5 is also a DNA repair protein functioning in the nucleotide excision repair pathway (86). The functional impact of this synonymous polymorphism is not clearly established yet. Earlier, patients with the CC genotype for this synonymous polymorphism were reported to have a better OS in univariate analysis (84) and PFS in multivariate analysis (87) while one study reported no statistically significant correlation with OS in colorectal cancer (75).

7) **rs9350, NC_000001.10:g.242048674C>T, Pro757Leu C/T in exonuclease 1 *(EXO1)* gene.** EXO1 has a 5'→3' double stranded DNA exonuclease activity and functions in the DNA mis-match repair mechanism to remove the mis-matched DNA bases (88). The functional impact of this polymorphism is yet to be established. In a Japanese cohort, the patients with the Leu/Leu genotype were found to have worse OS relative to other genotypes in univariate analysis (55).

8) **rs1800682, NG_011541.1:g.6185T>C, c-24+733T>C in Fas (TNF receptor superfamily, member 6) *(FAS)* gene.** FAS is a cell membrane receptor and has a fundamental role in inducing cell death (apoptosis) upon binding to its ligand (89). The functional impact of this polymorphism has not been conclusively established.

Previously, patients with CC genotype were reported to have significantly worse OS in univariate analysis when compared to patients with TT or TC genotypes in a study by Hofmann and others (90).

9) **rs351855, NG_012067.1:g.11323G>A, Gly388Arg A/G in fibroblast growth factor receptor 4 *(FGFR4)* gene.** The receptors belonging to FGFR family activate a cascade of signals which induce cell division and differentiation but the exact function of this particular member of the family is currently unknown (91). In a study using breast cancer cells, the cells having an allele for Arg (GG or AG genotypes) were reported to have greater motility *in vitro* and potential for progression (92). The same study also reported univariate analysis results where colorectal cancer patients having the FGFR4 with Arg variant had a significantly worse OS compared to homozygotes for Gly in the early months after diagnosis (92). One study reported no correlation of this polymorphism with OS in colorectal cancer (93).

10) **Glutathione S-transferase mu-1 *(GSTM1)* gene deletion.** The primary function of GSTM1 enzyme is to detoxify the electrophilic xenobiotics including drugs by conjugating them with glutathione (94). A homozygous deletion of the gene would cause a total loss of enzyme. In one study published by Csejtei et al. (95), Hungarian Dukes' stage B colorectal cancer patients with homozygous deletion of *GSTM1* gene had significantly poorer OS in univariate analysis when compared to patients with at least one copy of the gene. Five other studies reported no significant correlation of *GSTM1* gene deletion with OS in colorectal cancer (74,77,78,96,97).

11) **rs1695, NG_012075.1:g.6624A>G, Ile105Val A/G in glutathione S-transferase pi 1 *(GSTP1)* gene.** GSTP1 enzyme, like other members of the GST family of enzymes, is also involved in metabolism of xenobiotics (98). The GSTP1 enzyme with the valine residue at amino acid position 105 has been reported to have a reduced activity (99). In a Dutch cohort of colorectal cancer patients, patients with the valine variant (Ile/Val+Val/Val) treated with capecitabine and irinotecan were found to have better PFS than patients with Ile/Ile genotype in multivariate analysis, likely because of reduced metabolism of irinotecan by GSTP1 due to this polymorphism, as authors suggested (100). A similar result was observed in a mixed population cohort of metastatic colorectal cancer patients treated with 5-FU and oxaliplatin where carriers of an allele for valine (Ile/Val and Val/Val) had significantly better OS than the Ile/Ile homozygotes in univariate analysis (101). A similar association with favorable OS in univariate analysis and favorable PFS in multivariate analysis was found in another study of Caucasian patients (102). In addition, in two studies with Chinese subjects, patients homozygotes for the allele coding for valine were detected to have better OS (univariate analysis) (103) and the carriers of the same allele were detected to have favorable PFS (univariate analysis) and OS (multivariate analysis) (104). Contrary to these reports, the carriers of the valine variant were reported to have a worse OS in a Swedish-Caucasian colorectal cancer patient cohort in multivariate analysis (105) and homozygosity for valine was correlated with poor PFS in univariate analysis in a Korean colorectal cancer patient cohort (106). Also, six studies reported no significant correlation of this polymorphism with OS in colorectal cancer

(71,76,77,96,97,107).

12) **Glutathione S-transferase theta-1 *(GSTT1)* gene deletion.** GSTT1 enzyme metabolizes the electrophilic and hydrophobic xenobiotics by conjugating them with glutathione (108). Homozygous deletion of this gene results in loss of enzyme. In an age-stratified analysis, Rajagopal et al. (109) reported that young colorectal cancer patients with the homozygous deletion of this gene have a significantly favorable OS in univariate analyses while older patients with the gene deletion have poorer OS. Four other studies reported no significant association of *GSTT1* gene deletion with OS in colorectal cancer (74,78,95,101).

13) **rs1800795, NG_011640.1:g.4880C>G, -174G/C in promoter in interleukin 6 (interferon, beta 2) *(IL6)* gene.** IL6 is a cytokine and is involved in a wide range of inflammatory responses (110). *In vitro* analysis of this polymorphism performed using the HeLa cells has shown that the C allele reduced the gene expression (111). In one study of a Swedish colorectal cancer patient cohort, patients with the CC genotype showed better OS compared to heterozygotes after univariate analysis (112).

14) **rs1799977, NG_007109.1:g.23590A>G, Ile291Val A/G in mutL homolog 1, colon cancer, nonpolyposis type 2 (E. coli) *(MLH1)* gene.** MLH1 protein plays a role in the MMR machinery which repairs the mis-matched bases in the DNA (113). The definite functional impact of this polymorphism is not known. The GG and AG genotypes were reported to be correlated with a favorable OS in multivariate analysis

in a Spanish cohort of sporadic colorectal cancer patients (114). In another large study of Caucasian colorectal cancer patients, no association was observed between this polymorphism and OS in colorectal cancer (115).

15) **rs1799750, NG_011740.1:g.3471delG, -1607 indel G in promoter of matrix metallopeptidase 1 (interstitial collagenase) *(MMP1)* gene.** This protein belongs to the MMP family of enzymes. The primary function of these enzymes is to catalyze the breakdown of the extracellular matrix during events like embryonic development, tissue remodeling and reproduction and MMP1 particularly breaks down interstitial collagen types I, II and III (116). They are also found to play a role in diseases such as arthritis and metastasis of cancer cells (116). Functionally, insertion of G (insG) has been reported to enhance the transcription of *MMP1* gene by facilitating an extra binding site for the transcription factor v-ets erythroblastosis virus E26 oncogene homolog 1 (avian) (117). In a study conducted in colorectal cancer patients from Australia, patients homozygous for insG had significantly better OS compared to other genotypes (insG/insG vs insG/delG+delG/delG) in a multivariate analysis (118). In a contradictory report, patients in a French study homozygous for insG (insG/insG) had significantly worse cancer-specific survival, OS and DFS in multivariate analysis when compared to the deletion homozygotes (delG/delG) (119).

16) **rs243865, NG_008989.1:g.3726C>T, -1306 C/T in promoter region of matrix metallopeptidase 2 (gelatinase A, 72kDa gelatinase, 72kDa type IV collagenase) *(MMP2)* gene.** MMP2 is involved in the degradation of type IV collagen found in the

basement membranes, regulates inflammatory response and vascularization and is involved in endometrial breakdown (120). For this polymorphism, the presence of T allele has been reported to abolish an SP1 binding site in the promoter of *MMP1* lowering its gene expression (121). A Dutch study of 215 colorectal cancer patients previously showed that the C allele (CC and CT genotypes) was associated with favorable OS in multivariate analyses (122). In another study by Hettiaratchi et al (118), no correlation for this polymorphism was observed with OS in colorectal cancer.

17) **rs1801133, NG_013351.1:g.14783C>T, Ala222Val C/T missense polymorphism in methylene tetrahydrofolate reductase (NAD(P)H) *(MTHFR)* gene.** The role of this enzyme is the conversion of 5,10-methylenetetrahydrofolate (5,10-MTHF) to 5-methyltetrahydrofolate (5-MTHF) (123). 5-MTHF acts as a co-substrate in synthesis of methionine from homocysteine (123). For this polymorphism, studies have reported that presence of T allele (CT or TT genotypes) is associated with reduced amount of MTHFR enzyme (124) and reduced enzymatic activity (125). In one study, Caucasian colorectal cancer patients homozygous for C allele had better OS and cancer specific survival in multivariate analysis (124). A similar association was observed in stage III patients in a Swedish cohort (multivariate analysis) (126). However, in a Mexican cohort of colorectal cancer patients, a conflicting result was obtained where patients homozygous for the C allele had a significantly worse OS in univariate analysis (127). Nine other studies reported no significant correlation of this polymorphism with OS in colorectal cancer (55,128-135).

18) **rs1801131, NG_013351.1:g.16685A>C, Glu429Ala A/C missense polymorphism in *MTHFR* gene.** The C allele for this polymorphism is reported to reduce the activity of MTHFR enzyme (136). In metastatic colorectal cancer patients from a mixed population, a sex-specific association was observed where females homozygous for A allele had a favorable OS relative to other genotypes in univariate analysis (133). Similar association was also observed in a Spanish colorectal cancer patient cohort where patients homozygous for A allele showed favorable OS in the multivariate analysis (137). In six other studies, no correlation was observed between rs1801131 and OS in colorectal cancer (76,78,128,130,134,135).

19) **rs1052133, NG_012106.1:g.12146C>G, Ser326Cys C/G in 8-oxoguanine DNA glycosylase *(OGG1)* gene.** OGG1 enzyme excises the abnormal 8-oxoguanine base formed due to exposure of guanine to reactive oxygen (138). OGG1 enzyme with cysteine at amino acid position 326 instead of serine has been reported to have a reduced DNA-binding ability and reduced ability to repair damaged DNA (139). A correlation of this polymorphism was observed with both OS and PFS in univariate analysis in a Dutch cohort treated with capecitabine and oxaliplatin (140). However, whether the prognosis was favorable or worse was not described by these authors. In another study, no significant correlation was observed between rs1052133 and OS (75).

20) **rs4648298, NC_000001.10:g.186641682T>C, c.3618A/G in 3'-UTR of prostaglandin-endoperoxide synthase 2 (prostaglandin G/H synthase and**

cyclooxygenase) *(PTGS2)* gene. PTGS2 is an essential enzyme in prostaglandin synthesis during inflammatory responses (141). The functional consequence of this polymorphism is currently unknown. Previously, in a Spanish colorectal cancer patient cohort, the G allele was correlated with a favorable OS in multivariate analysis (142).

21) **rs1799889, NG_013213.1:g.4332_4333insA, -675 indel 4G/5G in promoter of serpin peptidase inhibitor, clade E (nexin, plasminogen activator inhibitor type 1), member 1 *(SERPINE1)* gene.** This protein inhibits fibrinolysis by inhibiting tissue plasminogen activator (tPA) and urokinase plasminogen activator (uPA) (high amounts of this protein are associated with formation of blood clots) (143). In a study which assessed the functional impact of this polymorphism, the insertion allele (insG) was linked to lower transcriptional activity (144). In a Swedish cohort, in univariate analysis in Dukes' stage A/B colorectal cancer patients, the patients with insG/insG genotype were detected to have better OS compared to patients with delG/delG or delG/insG genotypes (145).

22) **rs34743033, NC_000018.9:g.657730(28 base pairs (bp))2/3/4, 2/3 repeats of 28bp in 5'-UTR in thymidylate synthetase *(TYMS)* gene.** TYMS enzyme, together with 5,10-MTHF, converts deoxyuridylate to deoxythymidylate which is used for DNA replication and repair. The drug 5-FU exerts its anti-neoplastic effect primarily by inhibiting this enzyme (146). There is a variable number of tandem repeat (VNTR) of 28 bp sequence in 5'-UTR of *TYMS* gene. Reportedly, the three repeat allele (3R) has

an enhanced translational efficiency (147). In a Hungarian study, colorectal cancer patients homozygous for the 2-repeats (2R) allele had a worse OS compared to patients with 3R allele in univariate analysis (148). Similarly, in another multi-center study with patients from across Europe and Australia, patients treated with the drugs pemetrexed/irinotecan and homozygous for 3R showed a significantly favorable PFS in multivariate analysis (149). However, in another study with patients treated with 5-FU, leucovorin and oxaliplatin, those with 2R homozygotes and heterozygotes (2R/3R) had a favorable PFS in multivariate analysis (78). In a Dutch cohort, it was reported that patients younger than 60 years and homozygous for 2R had a favorable OS in univariate analysis (150). In a Spanish cohort of rectal cancer patients, 3R homozygotes showed favorable PFS and OS following multivariate analysis (151). Contradictorily, stage III colorectal cancer patients from Asia homozygous for 3R were found to have worse OS in univariate analysis (152). In at least 21 other studies, no correlation was observed for this polymorphism with OS in colorectal cancer (71,74,76-78,104,129,130,135,153-164).

23) **rs16430, NC_000018.9:g.673444delTinsTTAAAG, indel 6 bp in 3'-UTR of *TYMS* gene.** In an *in vitro* study using the human embryonic kidney cell line, the allele with deletion of the 6 bp sequence was linked to lowered stability of TYMS mRNA (165). The same study reported reduced gene expression in the presence of 6 bp deletion in tumor cells obtained from metastatic colorectal cancer patients. In a Spanish colorectal cancer patient cohort treated with 5-FU based chemotherapy regimen, it was observed that patients with homozygous deletion of 6 bp had favorable OS in the

multivariate analysis (155). On the contrary, in a French cohort, patients homozygous for the 6 bp insertion had favorable OS compared to heterozygotes after univariate analysis (157). At least thirteen other studies did not find an association between rs16430 and OS in colorectal cancer (74,76,78,130,135,148,151,153,154,162-164,166).

24) **rs2010963, NG_008732.1:g.5398C>G, -634G/C polymorphism in 5'-UTR in vascular endothelial growth factor A *(VEGFA)* gene.** The VEGFA protein targets endothelial cells and induces angiogenesis, increased vascular permeability, cell migration and inhibition of apoptosis (167). In a Greek study to understand the functional impact of polymorphisms in *VEGFA* gene, tumors from patients with non-small cell lung cancer were used. This study reported that tumor cells homozygous for the G allele had low *VEGFA* expression level as well as low tumor vascularization (168). In another Greek cohort of colorectal cancer patients, those patients with the genotype CC of this polymorphism had a significantly worse OS relative to those with GG genotype in multivariate analysis (169). No association with OS in colorectal cancer was observed in three other studies (170-172).

25) **rs3025039, NG_008732.1:g.19584C>T, +936C/T polymorphism in 3'-UTR in *VEGFA* gene.** A study conducted in healthy post-menopausal women from Austria showed that homozygotes for the C allele had higher levels of plasma VEGF protein levels than those carrying the T allele (CT+TT combined) (173). In a study investigating Greek colorectal cancer patients, it was reported that patients

homozygous for the T allele had worse OS compared to homozygotes for C allele (169). Three other studies did not find a significant correlation between rs3025039 and OS in colorectal cancer (51,171,172).

26) **rs25487, NC_000018.9:g.44055726T>C, Arg399Gln G/A in X-ray repair complementing defective repair in Chinese hamster cells 1 *(XRCC1)* gene.** XRCC1 protein repairs single strand breaks in the DNA caused by alkylating agents and ionizing radiations via base excision repair mechanism (174). Wang et al (175) reported that cells homozygous for A allele (Gln/Gln) had a relatively greater number of breaks in the chromosome per cell than other genotypes, indicative of an impaired function of *XRCC1* gene. In a Spanish cohort, patients homozygous for the A allele (Gln/Gln) had a significantly favorable OS after univariate analysis (75). However, contradictory associations were observed in Korean, Chinese and Turkish cohorts after univariate analysis: in their analyses, patients homozygous for the A allele (Gln/Gln) showed worse OS compared to homozygotes for G allele (71,83,131). In six other reports, this polymorphism was not associated with OS in colorectal cancer (74,76,77,176-178).

27) **rs861539, NG_011516.1:g.21071C>T, Thr241Met C/T in *XRCC3* gene.** XRCC3 is involved in the homologous recombination, maintenance of the stability of chromosome as well as DNA damage repair (179). Cells expressing XRCC3 protein with the methionine variant have been reported to have a defective DNA repair mechanism leading to abnormalities in chromosomal structure (180). The allele

coding for the amino acid methionine was correlated with significantly favorable prognosis after univariate analysis in a Spanish cohort of colorectal cancer patients (75). In another study by Grimminger et al (178), a statistically significant correlation between this polymorphism and OS in colorectal cancer was not observed.

Evident from the literature, for a given polymorphism, conflicting results in relation to prognosis do exist. This is in fact a common observation. The cohorts described in these studies may be heterogenous in terms of size, patient characteristics (ethnicity, age, stage), study design, treatment regimen as well as the definition of endpoints and statistical approaches, and different results for same polymorphisms may be obtained due to these differences (181,182). In addition, it is possible that the associations reported might be false positives or false negatives. Hence before genetic markers can find application into clinical patient management, large and well designed studies in homogenous patient cohorts are required to validate the correlations of genetic markers with outcome (183).

Chapter 2. Thesis project

2.1 Research Objectives

This thesis project has two main objectives:

1) To test the associations of 27 genetic variations with prognosis using a large cohort of colorectal cancer patients from Newfoundland (the discovery set). These polymorphisms were previously reported to be associated with prognosis in colorectal cancer.

2) To replicate the findings obtained in the discovery cohort in an independent colorectal cancer cohort from Newfoundland (the validation set).

To achieve these objectives, genotypes for the 27 genetic variations were first obtained in the discovery set. These data were analyzed together with the clinicopathological, molecular and prognostic data of the patients using statistical analyses. The variables which were found to be correlated with survival in the discovery set were then chosen for replication in the validation set.

2.2 Hypothesis

Many genetic polymorphisms have been reported to be correlated with measures of prognosis such as OS and DFS in colorectal cancer patient cohorts from around the world (see **section 1.7**). We have selected a total of 27 such polymorphisms and hypothesized that these polymorphisms are also correlated with prognosis in colorectal cancer patients from Newfoundland. After this first phase of the study, we also hypothesized that the significant correlations detected can also be replicated in an additional cohort of colorectal cancer patients from Newfoundland.

2.3 Patient cohorts

We investigated two independent cohorts of colorectal cancer patients from Newfoundland in our study.

The discovery set includes colorectal cancer patients from the NFCCR who were diagnosed over a period of 5 years from January 1999-December 2003 (35). Patients age under 75 years at diagnosis, with colorectal cancer confirmed pathologically, with available tumor tissue and informed consent obtained from either the patient or the next-of-kin (proxies) were included in the registry (35). Patients with familial colorectal cancer syndromes were also included in this registry. Patients having recurrent cancer, showing presence of carcinoma in situ (stage 0 colorectal cancer) and carcinoid tumor were excluded from the registry and/or analysis (35). Out of a total of 1983 colorectal cancer cases diagnosed with colorectal cancer in Newfoundland in the 5-year recruitment period, over 730 patients meeting these criteria were included in NFCCR (35). Molecular and genetic characteristics of this cohort have been described in detail by Woods et al (184). Out of these 730, DNA and prognostic data were available for 537 patients. Four patients having stage 0 colon cancer (carcinoma in-situ) were excluded from analysis. Two of the patients belonged to the same family and one was excluded randomly to have the cohort consisting of unrelated individuals. Thus in the end, 532 patients from NFCCR were included in the discovery set. Patients' clinical and vital status data was collected until April 2010. Their baseline characteristics are shown in **Table 2**.

In the discovery set, the median age of diagnosis is 61.4 years and the median follow-up time is 6.4 years. Females, stage IV patients, patients with mucinous tumor histology, lymphatic and vascular invasion, rectal cancer, and poorly differentiated tumor grade are each present in a minority of the cohort. The cohort also has a low proportion of patients with tumors having MSI-H status (10.50%) and *BRAF1* Val600Glu mutation (9.20%). One-third of the patients (33.3%) died during the follow up. The age-adjusted survival curve of the discovery cohort is depicted in **Figure 4**. The median survival time of the patients in the discovery cohort is ~9.5 years and the 5-year survival rate is ~79%. The median survival time of the entire NFCCR cohort is ~7 years and the 5-year survival rate is ~62% **(see Fig.A1** in **appendix)**. The percentage of stage IV patients in the discovery cohort is low (9.80%) and this can account for the high survival characteristics of this cohort. In fact, the entire NFCCR cohort has a higher proportion of stage IV patients (20.9%) when compared to patients included in this study (9.80%), and this difference is statistically significant ($p<0.001$). This may be because of the fact that the terminally-ill stage IV patients are more likely to have died before their blood samples were collected, indicating a selection bias in our study. The discovery cohort is thus biased toward early stage colorectal cancer patients and is not representative of the entire NFCCR cohort.

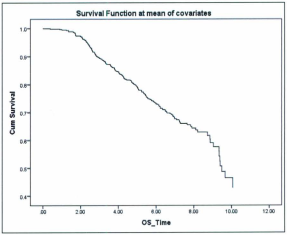
Table 2. Baseline characteristics of 532 patients in the discovery set

Variable	Number of patients	%
Sex		
male	327	61.50%
female	205	38.50%
Median age	61.36 years (20.7-75)	
Histology		
non-mucinous	471	88.50%
mucinous	61	11.50%
Location		
colon	353	66.40%
rectum	179	33.60%
Stage		
I	99	18.60%
II	206	38.70%
III	175	32.90%
IV	52	9.80%
Grade		
well diff/moderately diff	489	91.90%
poorly diff/undiff	39	7.30%
unknown	4	0.80%
Vascular invasion		
-	326	61.30%
+	166	31.20%
unknown	40	7.50%
Lymphatic invasion		
-	315	59.20%
+	174	32.70%
unknown	43	8.10%
OS status		
dead	177	33.30%
alive	354	66.60%
unknown	1	0.10%
Median OS follow-up time (range)	6.36 years (0.38-10.88)	
DFS status		
event	208	39.10%
no event	323	60.71%
unknown	1	0.19%
Median DFS follow-up time (range)	6 years (0.22-10.88)	

Familial risk		
low	256	48.10%
high/intermediate	276	51.90%
MSI Status		
MSI-H	56	10.50%
MSI-L/MSS	455	85.50%
unknown	21	4%
***BRAF1* mutation status**		
+	49	9.20%
-	435	81.80%
unknown	48	9%
Ethnicity		
Caucasian	486	91.35%
non-Caucasian	12	2.26%
unknown	34	6.39%
Treatment		
5-FU based	330	62.03%
other/no chemotherapy	199	37.41%
unknown	3	0.56%

diff: differentiated, MSI: microsatellite instability, 5-FU: 5-fluorouracil, ethnicity is based on the ethnicities of all four grandparents of the patients as reported by the patients.

Figure 4. Age-adjusted survival curve of discovery cohort



Median survival time is ~9.5 years

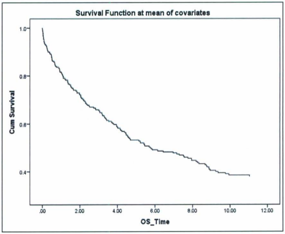
5-year survival rate is ~79%

The validation set: The discovery set was used to validate, in the Newfoundland population, the genetic polymorphisms correlated with outcome in other populations. To confirm the validity of the significant correlations detected in the discovery set, we also studied a second Newfoundland colorectal cancer cohort. All patients in this validation set were from Avalon Peninsula of Newfoundland and were diagnosed with primary colorectal cancer between January 1, 1997 and December 31, 1998. An eligibility criterion was presence of carcinoma in the polyp with invasion into the stalk. On the contrary to NFCCR, the age of diagnosis was not a criterion for inclusion. Exclusion criteria were recurrence of an earlier colorectal cancer, secondary colorectal cancer which is due to metastasis from a primary cancer elsewhere in the body, carcinoma in situ, mucosal carcinoma or carcinoid tumors and patients with FAP. Currently, the data and the biological specimen of these patients are preserved at the NFCCR. Although consent was not obtained from the patients or their proxies, collection of patient data, and the use of these data and biospecimen for research purposes were approved by the Regional Health Boards and Human Investigation Committee (HIC) (now known as Health Research Ethics Authority) of Memorial University of Newfoundland as long as the data were handled and analyzed anonymously. In this study, genotypes were obtained for 252 out of the total 280 patients who were included in our analyses. The baseline characteristics of this cohort are shown in **Table 3**. In this cohort, the median age of diagnosis is 68.7 years. Majority of patients (61.51%) had died till the time of last follow-up. The age-adjusted survival curve of the validation cohort is depicted in **Figure 5**. The median survival time of the patients in the validation set is ~6 years and the 5-year

Table 3. Baseline characteristics of 252 patients in the validation set.

Variable	Number of patients (n)	%
Sex		
male	133	52.78%
female	119	47.22%
Median age	68.7 years (25.3-91.6)	
Histology		
non-mucinous	211	83.73%
mucinous	41	16.27%
Location		
colon	202	80.16%
rectum	50	19.84%
Stage		
I	48	19.05%
II	88	34.92%
III	68	26.98%
IV	41	16.27%
unknown	7	2.78%
Grade		
well diff/moderately diff	211	83.73%
poorly diff/undiff	37	14.68%
unknown	4	1.59%
Lymphatic invasion		
-	64	25.40%
+	101	40.08%
unknown	87	34.52%
OS status		
dead	155	61.51%
alive	97	38.49%
Median OS follow-up time (range)	5.43 years (0-12.48)	
Median DFS follow-up time (range)	3.25 years (0-12.48)	
DFS status		
event	167	66.27%
no event	85	33.73%
MSI status		
MSI-H	24	9.52%
MSI-L/MSS	228	90.48%
Treatment		
5-FU based	88	34.92%
other/no chemotherapy	148	58.73%
unknown	16	6.35%

Figure 5. Age-adjusted survival curve of the validation cohort



Median survival time is ~6 years

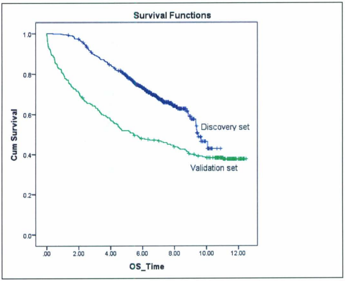
Median 5-year survival rate is ~55%

survival rate is ~55%. The validation cohort is not significantly different from the entire cohort (n=280) in terms of distribution of clinicopathological/molecular variables (**see Fig.A2** in **appendix**). It is assumed that most of the patients in the validation cohort are Caucasians since there was very low ethnic diversity in the Avalon Peninsula during the patient recruitment period (1997-98). Similar to the discovery set, females, stage IV patients, patients with mucinous tumor histology, lymphatic invasion, rectal cancer, and poorly differentiated tumor grade are each present in lower proportion in this cohort.

The Kaplan-Meier plots comparing the survival of the discovery and validation cohorts without age-adjustment is depicted in **Figure 6**. Without age-adjustment, the discovery cohort patients had a median survival time of ~9 years in contrast to ~9.5 years after age adjustment, although the 5-year survival rates are similar (~80% without age-adjustment and ~79% with age-adjustment) (**Figure 4**). For the validation cohort, the median survival time is ~5.2 years compared to ~6 years after age-adjustment and the 5-year survival rate is ~50% compared to ~55% after age-adjustment (**Figure 5**). The differences indicate the affect of age on OS, as generally, OS is expected to reduce with increasing age. For further comparisons between the discovery and validation cohorts, see **section 4.3.3**.

Ten clinicopathological and molecular variables were used in this study for analyses. These include stage, tumor grade, vascular or lymphatic invasion, tumor histology, MSI status, tumor location, familial risk status, *BRAF1*_Val600Glu mutation status, age and sex (see **section 1.5.2**). The discovery set was used for analysis of 27 genetic

Figure 6. Kaplan-Meier curve comparing the survival of discovery (n=532) and validation (n=252) sets



<u>Discovery set:</u> Median survival time is ~9 years. 5-year survival rate is ~80%

<u>Validation set:</u> Median survival time is ~5.2 years. 5-year survival rate is ~50%

polymorphisms (see **section 3.1**). The variables which were correlated with OS in the discovery set after multivariate analysis, including genotypes of 4 polymorphisms, were also analyzed in the validation set. Of the variables present in the final multivariate model for DFS in the discovery set, two polymorphisms for which the genotypes were available in the validation set were also analyzed for validating the results obtained in the discovery set.

Chapter 3. Methods

Ethics approval

This study was approved by HIC of Memorial University of Newfoundland (HIC Reference # 10.117).

Contributions and credits

Amit Negandhi: Performed TaqMan® SNP genotyping assays for rs1799889 (*SERPINE1* gene) and rs1799750 (*MMP1* gene) in the discovery set, rs1801131 (*MTHFR* gene), rs1047768 (*ERCC5* gene) and rs1799889 (*SERPINE1* gene) in the validation set. Performed PCR reaction and gel electrophoresis for *GSTT1* and *GSTM1* gene deletions and genotyping of the VNTR in *TYMS* gene in the discovery set and for *GSTM1* gene deletion in the validation set. Performed coding of the genotype data and statistical analyses described in the thesis document. Performed literature research to interpret and discuss the results obtained.

Michelle Simms: Prepared stock DNA plates of NFCCR and validation set samples.

Jessica Squires: Performed dilution of stock DNA samples and provided technical assistance in the lab.

Angela Hyde: Provided clinicopathological and prognostic data of the validation set samples.

Dr. Roger Green: Provided DNA samples from NFCCR and the validation set samples.

Dr. Sevtap Savas: Processed the raw clinicopathological, ethnicity, prognostic, and other data for the NFCCR samples and coded them for the statistical analyses, provided the Kaplan Meier survival curves for the entire NFCCR cohort and entire validation cohort and provided the baseline characteristics tables for the entire NFCCR samples, entire validation cohort samples and the validation set samples.

Dr. Patrick Parfrey, Dr. Wei Xu and Dr. Michelle Liu provided assistance with study design and statistical methods.

Funding agencies

This study was funded by the Memorial University of Newfoundland and The Medical Research Foundation of Faculty of Medicine-Cox Award (2010), Memorial University.

3.1 Selection of polymorphisms

For this thesis project, 30 polymorphisms were selected which were previously found to be correlated with survival in colorectal cancer patients from populations other than Newfoundland. The polymorphisms were selected based on the information collected and posted in the dbCPCO database (47) as of September 2010. The selection was based on the following order of priorities:

1) The polymorphisms which showed statistically significant correlations with

overall survival and/or disease specific survival in at least one study.

2) The polymorphisms which can be genotyped by methods available to us i.e. Sequenom MassArray®, TaqMan® SNP Genotyping assays and gel electrophoresis of polymerase chain reaction (PCR) products.

The polymorphisms selected for inclusion in this study are listed in **Table 4**. An attempt was made to genotype twenty-five polymorphisms using the Sequenom MassArray® technology. Among these, the TP53_rs1042522, PTGS2_rs20417, and EGF_rs4444903 polymorphisms failed to be genotyped by this method. The two gene deletions (*GSTM1* and *GSTT1* gene deletions) and the VNTR in *TYMS* gene were genotyped by gel electrophoresis of PCR products. *SERPINE1*_-675 indelG and *MMP1*_-1607 indelG polymorphisms were genotyped using the TaqMan® SNP genotyping assays. Therefore a total of 27 polymorphisms were genotyped in the discovery set using the MassArray, TaqMan®, and PCR and gel electrophoresis methods (**Table 4**).

3.2 Plates containing DNA samples

Discovery set: Patients recruited to the NFCCR and with available prognostic data and genomic DNA were included in this study. DNA samples were provided by Dr. Roger Green and were previously extracted from the blood samples of colorectal cancer patients. The stock DNA plates contained 541 DNA samples (10ng/μl in water) distributed over seven 96-well plates. For the purpose of genotyping by Sequenom MassArray system, the same concentration of DNA was used. For performing TaqMan® assays, the stock solutions were aliquoted to seven other plates and diluted to 4ng/μl

Table 4. Genetic polymorphisms selected for inclusion in this thesis project.

Gene symbol	Polymorphism	rs number	Type	Genotyping methodology
CCND1	Pro241Pro A/G NG_007375.1:g.12038G>A	rs9344	SNP	Sequenom MassArray®
DCC	Arg201Gly C/G NG_013341.1:g.571061C>G	rs2229080	SNP	Sequenom MassArray®
**EGF*	A61G in 5'-UTR NG_011441.1:g.5071A>G	rs4444903	SNP	Sequenom MassArray® (failed genotyping)
EGFR	Arg521Lys G/A NG_007726.1:g.147531G>A	rs2227983**	SNP	Sequenom MassArray®
ERCC1	Asn118Asn C/T NG_015839.1:g.8525T>C	rs11615	SNP	Sequenom MassArray®
ERCC2	Lys751Gln G/T NG_007067.2:g.23927A>C	rs13181	SNP	Sequenom MassArray®
ERCC5	His46His C/T NG_007146.1:g.11344T>C	rs1047768	SNP	Sequenom MassArray® in discovery set
				TaqMan® assay in validation set
EXO1	Pro757Leu C/T NC_000001.10:g.242048674C>T	rs9350	SNP	Sequenom MassArray®
FAS	c.-24+733T>C NG_011541.1:g.6185T>C	rs1800682	SNP	Sequenom MassArray®
FGFR4	Gly388Arg A/G NG_012067.1:g.11323G>A	rs351855	SNP	Sequenom MassArray®
GSTM1	Gene deletion	n/a	gene deletion	PCR and agarose gel electrophoresis

GSTP1	Ile105Val A/G NG_012075.1:g.6624A>G	rs1695	SNP	Sequenom MassArray®
GSTT1	Gene deletion	n/a	gene deletion	PCR and agarose gel electrophoresis
IL6	-174G/C in promoter NG_011640.1:g.4880C>G	rs1800795	SNP	Sequenom MassArray®
MLH1	Ile219Val A/G NG_007109.1:g.23590A>G	rs1799977	SNP	Sequenom MassArray®
MMP1	-1607 indel G in promoter NG_011740.1:g.3471delG	rs1799750	Indel	TaqMan® SNP genotyping assay
MMP2	-1306C/T in promoter NG_008989.1:g.3726C>T	rs243865	SNP	Sequenom MassArray®
MTHFR	Ala222Val C/T NG_013351.1:g.14783C>T	rs1801133	SNP	Sequenom MassArray®
MTHFR	Glu429Ala A/C NG_013351.1:g.16685A>C	rs1801131	SNP	Sequenom MassArray® in discovery set
				TaqMan® assay in validation set
OGG1	Ser326Cys C/G NG_012106.1:g.12146C>G	rs1052133	SNP	Sequenom MassArray®
**PTGS2*	-765G/C in promoter NC_000001.10:g.186650321C>G	rs20417	SNP	Sequenom MassArray® (failed genotyping)
PTGS2	c.3618A/G in 3'-UTR NC_000001.10:g.186641682T>C	rs4648298	SNP	Sequenom MassArray®
SERPINE1	-675 indelG in promoter NG_013213.1:g.4332_4333insA	rs1799889	Indel	TaqMan® SNP genotyping assay
**TP53*	Arg72Pro C/G NG_017013.1:g.16392C>G	rs1042522	SNP	Sequenom MassArray® (failed genotyping)

TYMS	2/3 repeats of 28bp in 5’-UTR NC_000018.9:g.657646_(28bp)/2/3/4	rs34743033	VNTR	PCR and agarose gel electrophoresis
TYMS	indel 6 bp in 3’-UTR NC_00018.9:g.673444delTins6bp	rs16430	Indel	Sequenom MassArray®
VEGFA	-634G/C in 5’-UTR NG_008732.1:g.5398C>G	rs2010963	SNP	Sequenom MassArray®
VEGFA	+936C/T in 3’-UTR NG_008732.1:g.19584C>T	rs3025039	SNP	Sequenom MassArray®
XRCC1	Arg399Gln G/A NC_000019.9:g.44055726T>C	rs25487	SNP	Sequenom MassArray®
XRCC3	Thr241Met C/T NG_011516.1:g.21071C>T	rs861539	SNP	Sequenom MassArray®

[1] Aimed to be designed in Sequenom MassArray® multiplex reactions. *These polymorphisms failed to be genotyped by Sequenom MassArray® method and were excluded from this project. VNTR: variable number of tandem repeats. **SNP is also designated as rs11543848.

concentration with water. In these DNA plates, the last column of each plate (column 12) contained 3 non-template controls (NTCs) and 5 duplicate DNA samples to test for PCR contamination and concordance of genotyping reactions, respectively.

Validation set: An additional set of 280 colorectal cancer patients constituted the validation set. DNAs that were previously extracted from blood (3ng/µl) or formalin-fixed paraffin-embedded (FFPE) non-tumor tissue (5ng/µl) were used to genotype the *MTHFR*_Glu429Ala, *ERCC5*_His46His, *SERPINE1*_-675indelG polymorphisms and *GSTM1* gene deletion.

3.3. Solutions

1) 5X Tris-borate-EDTA (TBE) Buffer

   Made by mixing 54 grams (gms) OmniPur® Tris-Hydrochloride (Tris-HCl) (Product code 9310, EMD Chemicals Inc. NJ, USA), 27.5 gms Boric acid (Product code BX0865, EMD Chemicals Inc. NJ, USA), 20 milliliters (ml) 0.5 Molar (M) EDTA (pH=8±0.1) (Catalog number (cat. #) 46-034-Cl, Mediatech Inc, VA, USA) in one liter of deionized (dH_20). The buffer solution was autoclaved, pH was adjusted to 8.3 with sodium hydroxide (Product code SX0590, EMD Chemicals Inc. NJ, USA) and solution was stored at room temperature.

2) 1X TBE buffer

   This solution was prepared by diluting 5X TBE solution in dH_20.

3) 1X Tris-EDTA (TE) buffer

Made by mixing the following chemicals in sterile dH_2O in a total volume of 200ml: 0.3152 gms Tris-HCl (Product code 9310, EMD Chemicals Inc. NJ, USA) equivalent to 10 millimoles (mM) Tris-HCl and 0.4 ml of 0.5M stock solution of EDTA (pH: 8±0.1) (cat. # 46-034-Cl, Mediatech Inc, VA, USA) equivalent to 1 mM EDTA.

3.4 Obtaining the genotype data

Discovery set:

3.4.1 Using Sequenom MassArray® technique

The Sequenom MassArray® system was the first choice for genotyping. This multiplex reaction system facilitates simultaneous genotyping of multiple polymorphisms in a reasonably short time and is cost-effective. The genotyping reactions were outsourced to the Analytical Genetics Technology Centre (AGTC) facility at University Health Network (UHN), Toronto. Seven DNA plates containing 541 DNA samples and duplicate samples (10ng/μl) were sent for genotyping. The DNA sample identifiers (IDs) were re-coded prior to sending to the facility. Initially we aimed for genotyping of 27 polymorphisms (**Table 4**). However, assays for only 25 polymorphisms (except MMP1_rs1799750 and SERPINE_rs1799889) could be designed by the facility. An additional 3 polymorphisms (TP53_rs1042522, PTGS2_rs20417 and EGF_rs4444903) failed genotyping by these assays after the reactions were run. Thus genotypes for a total of 22 polymorphisms were obtained by Sequenom MassArray® technology.

3.4.2 Design of primers and probes for Custom TaqMan® SNP Genotyping Assays

Polymorphisms in *SERPINE1* (rs1799889) and in *MMP1* (rs1799750) could not be incorporated in the MassArray multiplex reactions. Therefore we used the TaqMan® SNP genotyping assays to obtain genotypes for these SNPs. The primers and probes for these polymorphisms were custom designed using the 'Custom TaqMan® Assay Design Tool' available online (185). The sequences flanking these polymorphisms were obtained from the dbSNP database (186) (**Table 5**). These assays were used in genotyping of 541 samples in the discovery set.

3.4.3 Pre-designed TaqMan® SNP Genotyping Assays

The predesigned TaqMan® SNP genotyping assays for *MTHFR*_Glu429Ala (assay ID C_850486_20) and *ERCC5*_His46His (assay ID C_1891769_20) were obtained from the Applied Biosystems (187) (primer and probe sequence information for these assays are proprietary of Applied Biosystems and thus were not provided to us). Assays for these SNPs were performed for samples in the validation cohort.

TaqMan® SNP Genotyping assay procedure: Upon arrival, 40X TaqMan® assay mix (Applied Biosystems, CA, USA) containing the primers and probes for the TaqMan® genotyping reactions was diluted to 20X with 1X TE buffer, aliquoted and stored at -20°C. For a 96 well plate, the reaction mix was prepared by adding 525µl 2X TaqMan® Universal PCR Master mix (part. # 4304437, Roche, NJ, USA), 26.25µl 20X TaqMan®

Table 5. Primer and probe information for SNPs in *MMP1* and *SERPINE1* genes

SNP	*MMP1*_rs1799750	*SERPINE1*_rs1799889
*Assay ID	AHVI4S6	AHWR2ZE
Forward primer Seq.	ACATGTTATGCCACTTAGATGAGGAAA	AGACAAGGTTGTTGACACAAGAGA
Reverse primer Seq.	CGTCAAGACTGATATCTTACTCATAAACAATACTTC	GGCCGCCTCCGATGATAC
**Probe 1 Seq.	TGAGATAAGTCATATCCTTTC	ACGGCTGACTCCCCCAC
***Probe 2 Seq.	TGAGATAAGTCATATCTTTC	CGGCTGACTCCCCAC

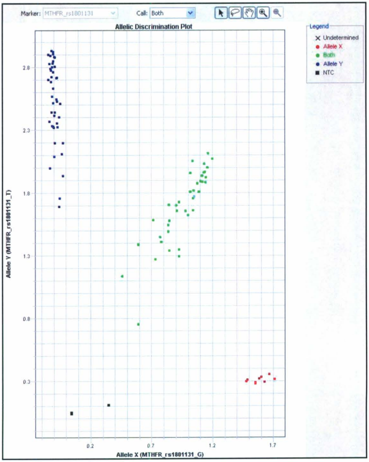
*ID by Applied Biosystems (USA). **Reporter 1 dye is VIC, which recognizes the G allele. ***Reporter 2 dye is FAM, which recognizes the deletion of G allele. Underlined are the sequences on probes that are complementary to the polymorphic sequences. Seq: sequence.

assay mix (Applied Biosystems, CA, USA) for the particular polymorphism and 393.75µl of sterile water. 114µl of the reaction mix was transferred to each well of an 8-well strip tube using a single channel pipette. 9µl of the reaction mix from each well of the strip tube was subsequently transferred to the wells of the MicroAmp® Fast Optical 96-well reaction plate with barcode (0.1 ml) (part. # 4346906, Applied Biosystems, CA, USA) using a multi channel pipette. These plates are custom-made for use in the 7900HT Fast Real Time PCR System (part. # 4330966, Applied Biosystems, CA, USA). 1µl of DNA extracted from blood with a concentration of 4ng/µl for *SERPINE1*_rs1799889 and *MMP1*_rs1799750 in the discovery set samples and either 3ng/µl (extracted from blood) or 5ng/µl (extracted from FFPE) DNA for *MTHFR*_rs1801131 and *ERCC5*_rs1047768 in the validation set samples was added to the plate containing the reaction mix. The final reaction volume was 10µl. A PCR-compatible optical adhesive cover (part. # 4360954, Applied Biosystems, CA, USA) was applied over the plate, sealed tightly, and the plate was centrifuged at 1000 revolutions per minute (rpm) for ~5-10 seconds in a bench top centrifuge (cat. # 75004367, Sorvall Legend T+ Centrifuge, ThermoFisher Scientific, MA, USA) prior to the PCR amplification.

The ABI 7900HT Sequence Detection Systems (SDS) software, version 2.4 accompanies the 7900HT Fast Real Time PCR System. For SNP genotyping assays, the allelic discrimination (AD) and absolute quantification (AQ) files were created using the SDS software following the instructions in 'Applied Biosystems 7900HT Fast Real-time PCR System Allelic Discrimination Getting Started Guide' (part. # 4364015, Applied Biosystems, CA, USA). The AD file contains information about the detector which is composed of a pair of fluorescent probes to detect the particular alleles, sample information in the plate and enables analysis of the fluorescence data after the PCR run is completed. The AQ file contains data for the real-time PCR run such as the thermocycling conditions. These files are essential for performing the PCR run and for calling the genotypes based on the fluorescence information generated. These files were initially prepared in a desktop computer, transferred to a USB drive and copied on the computer adjoining the 7900HT Fast Real Time PCR System. For the PCR amplification, the reaction plate was inserted in the machine and a pre-read procedure was performed using the AD file prepared for the plate. The pre-read is performed to record background fluorescence which is used as a reference against which the fluorescence recorded after amplification is compared to give the genotype in each well. After performing the pre-read, PCR amplification of the DNA samples using the AQ file was performed in the 7900HT Fast Real Time PCR System (part. # 4330966 Applied Biosystems, CA, USA). The PCR thermocycling conditions are as follows: 50°C for 2 minutes (*activation of AmpErase® UNG in TaqMan® Universal PCR Master Mix)*, 95°C for 10 minutes *(activation of AmpliTaq Gold® DNA Polymerase in TaqMan® Universal PCR Master Mix)* and 40 cycles of 95°C for 15 seconds *(melting DNA)* and 60°C for 1 minute *(annealing/extension of primer)*.

After the completion of the PCR run, a post-read was performed using the AD file. Pre-read and the post-read data in the AD file were automatically analyzed by the software and genotypes were called (**Figure** 7). The plots were also manually examined by an independent researcher (Dr. Savas) to confirm the genotype callings. In case of a

Figure 7. AD plot for TaqMan assay for *MTHFR*_rs1801131



Marker: MTHFR_rs1801131

Call: Both

Allelic Discrimination Plot
Legend
X Undetermined
● Allele X
● Both
● Allele Y
■ NTC
Allele Y (MTHFR_rs1801131_T)
2.8
2.3
1.8
1.3
0.8
0.3
Allele X (MTHFR_rs1801131_G)
0.2
0.7
1.2
1.7

A sample AD plot for TaqMan® SNP genotyping assay

Each dot represents a sample. The black squares at the bottom left show no amplification, which are the NTCs. The blue dots are homozygotes for the T allele while the red dots are homozygotes for the G allele. The green dots in the center are heterozygotes.

discrepancy between the visual inspection of the plots and the automatic genotype calling, the genotyping reaction was repeated. For failed samples up to three repeat attempts were made to obtain genotypes, whenever the DNA was available. The finalized genotyping data was exported into an excel sheet and organized for data analysis.

3.4.4 Genotyping for *GSTT1* and *GSTM1* gene deletions

To detect *GSTT1* and *GSTM1* gene deletions, we performed a multiplex PCR reaction followed by gel electrophoresis as previously described by Arand et al (188). This PCR reaction is a triplex reaction including the forward and reverse primers for amplification of three genes: *GSTT1, GSTM1* and albumin gene *(ALB). ALB* gene serves as a positive control for successful PCR amplification. *ALB* gene yields a PCR product which is 350 bp long, *GSTT1* gene product is 480 bp long and the *GSTM1* gene product is 215 bp long. The primer sequences for the three genes are shown in **Table 6**.

PCR method:

For a 96 well plate, reaction mix was prepared by adding 525µl 2X AmpliTaq Gold® 360 Master Mix (product. # 4398790, kit part. # 4398881, Applied Biosystems, CA, USA), 26.25µl GC enhancer (product # 4398799, kit part. # 4398881, Applied Biosystems, CA, USA), 288.75µl of sterile water and 105µl primers (Integrated DNA Technologies, Iowa, USA) containing 10µM of each primer (forward and reverse) for all three genes. The reaction mix was then equally distributed in wells of a 8-well strip tube using a single channel pipette. 9µl of the reaction mix was subsequently transferred to each well of the MicroAmp® Fast Optical 96-well reaction plate with barcode (0.1 ml) (part. # 4346906,

Table 6. Primer sequences for PCR amplification of *GSTT1, GSTM1, ALB* gene fragments and VNTR in *TYMS* gene

	Primer Sequence **5` to 3`**	**Reference**
GSTT1	F: TTCCTTACTGGTCCTCACATCTC R: TCACCGGATCATGGCCAGCA	(188)
GSTM1	F: GAACTCCCTGAAAAGCTAAAGC R: GTTGGGCTCAAATATACGGTGG	(188)
ALB	F: GCCCTCTGCTAACAAGTCCTAC R: GCCCTAAAAA GAAAATCGCCAATC	(188)
TYMS	F: GTGGCTCCTGCGTTTCCCCC R: TCCGAGCCGGCCACAGGCAT	(189)

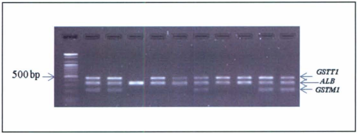
F=forward, R=reverse

Applied Biosystems, CA, USA) from the strip tube using a multichannel pipette. 1µl of DNA solution (4ng/µl) was added to the reaction mix in the reaction plate. Optical adhesive cover (part. # 4360954, Applied Biosystems, CA, USA) was applied, sealed tightly, and the reaction plate was spun at 1000 rpm for ~5-10 seconds in a bench-top centrifuge (cat. # 75004367, Sorvall Legend T+ Centrifuge, ThermoFisher Scientific, MA, USA). An AQ file was set up for each plate and the PCR runs were performed on the 7900HT Fast Real Time PCR System (part. # 4330966, Applied Biosystems, CA, USA) with the following thermocycling conditions: 95°C for 10 minutes *(primary denaturation and activation of AmpliTaq Gold 360 DNA polymerase in AmpliTaq Gold® 360 Master Mix)*, <u>34 cycles</u> of 95°C for 30 seconds *(denaturation)*, 64°C for 30 seconds *(primer annealing)* and 72°C for 1 minute *(primer extension)* and a <u>final cycle</u> of 72°C for 7 minutes *(final elongation)* followed by a hold at 4 °C until plate removed from the thermocycler. The plate was spun again at 1000 rpm for ~5 seconds after the completion of the PCR run and PCR products were then analyzed using the agarose gel electrophoresis.

<u>Agarose gel electrophoresis to genotype *GSTT1 and GSTM1* gene deletions:</u>

A 1.5% maxi gel was prepared by dissolving and melting 3.75 grams OmniPur® Agarose PCR Plus (Product code 2010, EMD Chemicals Inc. NJ, USA) in 250 ml 1X TBE buffer in a microwave. Eighteen µl of 10,000X SYBR® Safe DNA gel stain (cat. # S33102, Invitrogen, Oregon, USA) was added to the molten agar solution and mixed by gentle swirling. The mixture was then poured into the gel apparatus and allowed to solidify. The gel cast had two combs of 20 wells each and one comb of 17 wells. After solidification, the combs and rubber edges were removed and the gel was placed in the electrophoresis tank filled with 1X TBE buffer. 15μl of 6X DNA Gel loading buffer (cat. # AC10097, Omega Bio-tek, GA, USA) was added to each well of an 8-well strip tube. Approximately ~1μl of loading buffer was mixed with 10 μl of PCR products by pipetting up and down 2-3 times. The mixture was then loaded into the wells of the gel using a multichannel pipettor. The first well of each row in the agarose gel was loaded with ~3-4μl of 135ng/μl 100 bp DNA ladder (cat. # D-1030, Bioneer, Korea). The gel was then run at 70 volts (V) and images were taken at 45 and 65 minutes under ultraviolet (UV) transillumination in an AlphaImager® EP (Alpha Innotech, CA, USA). A filter transmitting UV light of wavelength 302 nanometer (nm) was used for visualizing SYBR® Safe DNA stained gels on the AlphaImager® EP. An example of the image is shown in **Figure 8**. Individuals with the absence of the topmost band have *GSTT1* gene deletion while those with the absence of the bottommost band have *GSTM1* gene deletion. One agarose gel can accommodate a total of 48 samples. Hence, two gels were used to analyze samples from one 96-well PCR plate. For the first gel, samples from columns 7-12 were loaded since column 12 contains the NTCs in plate wells F12, G12 and H12. If DNA contamination is observed in any of these wells, the PCR products were discarded and PCR reactions were repeated. If contamination was not observed, then the electrophoresis of PCR products from columns 1-6 was also performed.

Figure 8. Gel image for detection of *GSTT1* and *GSTM1* gene deletions



The first sample is a 100 bp DNA ladder. Individuals with absence of topmost band have deletion of *GSTT1* gene. Individuals with the absence of bottommost band have deletion of *GSTM1* gene.

3.4.5 Genotyping for 2/3 repeats of 28 bp in 5'-untranslated region (5'-UTR) of *TYMS* gene (rs34743033)

Primer sequences for region flanking rs34743033 in 5'-UTR in *TYMS* gene were obtained from the literature (189) and are shown in **Table 6**.

PCR Reaction

For a 96 well plate, reaction mix was prepared by adding 525µl 2X AmpliTaq Gold® 360 Master Mix (product. # 4398790, kit part. # 4398881, Applied Biosystems, CA, USA), 52.5µl GC enhancer (product. # 4398799, kit part. # 4398881 Applied Biosystems, CA, USA), 157.5µl of sterile water and 210µl primer solutions (Integrated DNA Technologies, Iowa, USA) containing 10µM forward and reverse primers. The reaction mix was equally distributed across an 8-well strip tube using a single channel pipette. 9µl of reaction mix was then transferred to each well of a MicroAmp® Fast Optical 96-well reaction plate with barcode (0.1 ml) (part. # 4346906, Applied Biosystems, CA, USA) using a multichannel pipette. 1µl DNA solution (4ng/µl) was then added into the reaction mix and the plate was sealed with VWR™ adhesive foil for microplates (cat. # 60941-072, VWR, PA, USA). The reaction was run in a Veriti 96-well fast thermal cycler (part. # 4375305, Applied Biosystems, CA, USA) with the following thermocycling conditions:

95°C for 10 minutes *(primary denaturation and activation of AmpliTaq Gold 360 DNA polymerase in AmpliTaq Gold® 360 Master Mix)*, 35 cycles of 95°C for 30 seconds *(denaturation)*, 70°C for 30 seconds *(annealing)* and 72°C for 1 minute *(extension)* and a final cycle of 72°C for 7 minutes *(final elongation)* followed by a hold at 4 °C until removal of the plate from the thermocycler.

Agarose gel electrophoresis for 2/3 repeats of 28 bp in 5'-UTR of *TYMS* gene

The method for electrophoresis is similar to that of gene deletions for *GSTT1* and *GSTM1* genes with the following changes:

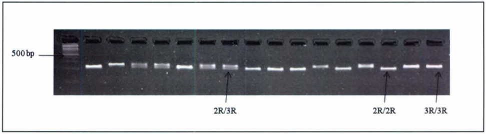
a) A 4% maxi gel was prepared by dissolving and melting 10 grams OmniPur® Agarose PCR Plus (Product code 2010, EMD Chemicals Inc. NJ, USA) in 250 ml 1X TBE buffer in a microwave.
b) The gel images were taken under UV transillumination in AlphaImager® EP (Alpha Innotech, CA, USA) at 45, 75 and 95 minutes.

PCR products with 2 repeats of 28bp VNTR (2R) of *TYMS* gene migrate faster and form the bottommost band while those with 3 repeats (3R) migrate slower and form the topmost band (**Figure 9**). We very rarely also observed a 4-repeat allele in our samples. These samples were confirmed by re-amplifications run on 4.5% agarose gels.

3.5 Data analysis

The genotype data was organized in an Excel sheet and combined with the clinicopathological, demographic, molecular, and prognostic data of the patients obtained from NFCCR and processed by Dr. Sevtap Savas. The prognostic data contained the clinicopathological and molecular variables described in detail in **section 1.5.2**.

Figure 9. Gel image for detection of 2/3 repeats of 28bp in *TYMS* gene



The first well is a 100 bp DNA ladder. Individuals with two bands are heterozygotes for 2 and 3 repeats (2R/3R). Individuals with only the topmost band have 3R. Individuals with only the bottommost band have 2R.

The minor allele frequencies (mAFs) of the polymorphisms in colorectal cancer patients in both the discovery and validation sets were separately calculated in an Excel document. The mAFs in other Caucasian populations were obtained from the dbSNP (186) database or published reports and compared to the mAFs in our cohorts. Duplicate genotypes were checked for concordance. In the case of MassArray®, if a discordant genotype was obtained in duplicate samples, these genotypes were excluded from the analysis. For the TaqMan® SNP genotyping and the PCR-gel electrophoresis techniques, the discordant samples were repeated to obtain the final genotype data. For VNTR in *TYMS* gene, the 2R allele has been shown to have lower transcription activity than the 3R allele (190). Therefore, we combined the rarely observed 4R alleles with the 3R alleles for data analysis since it is likely that both 3R and 4R alleles have activities greater than 2R allele.

Hardy-Weinberg equilibrium (HWE) calculations were performed in both the patient cohorts separately to observe deviations of genotype frequencies in the cohort using an online tool (191) and were confirmed by manual calculations. In case of any discrepancy, the manual calculations were repeated and noted.

Statistical tests were performed using the PASW Statistics 18 software Release 18.0.2 (April 2010) assuming three models of inheritance: co-dominant, dominant and recessive. In the co-dominant model, the survival times of patients with minor allele homozygotes and heterozygotes were separately compared with the survival times of patients with major allele homozygotes. In the dominant model, the survival times of patients with minor allele homozygotes+heterozygotes were compared with the survival times of patients with major allele homozygotes. In the recessive model, the survival times of patients with minor allele homozygotes were compared with the survival times of patients with major allele homozygotes+heterozygotes.

To illustrate the three models, let us consider a polymorphism with the major allele A and the minor allele G. In the co-dominant model, AA is the reference category and patients with genotypes AG and GG are separately compared to patients with AA genotypes. In the dominant model, AA genotypes are compared to AG+GG genotypes. In the recessive model, AA+AG genotypes are compared to GG genotypes (192).

For clinicopathological and molecular variables, the categorical variables included were sex (males vs females), tumor histology (mucinous vs non-mucinous), tumor location (rectum vs colon), stage (stages II, III and IV individually vs stage I), tumor grade (poorly differentiated/undifferentiated vs well differentiated/moderately differentiated), vascular invasion (presence vs absence), familial risk status (high/intermediate vs low), MSI status (MSI-H vs MSI-L/MSS) and BRAF1_Val600Glu mutation status (presence vs absence). Age was analyzed as a continuous variable. The vascular invasion data for the validation set were not available, but the lymphatic invasion data were. In the discovery set, it was observed that vascular invasion and lymphatic invasion were highly correlated with each other (see **section 4.2.5**) i.e. almost all tumors having vascular invasion had lymphatic invasion too. Thus we compared the vascular invasion data in the discovery set with the lymphatic invasion data in the validation set to test for significant differences between the two cohorts in terms of invasion.

Genotype data was available for 532 patients in the discovery set and 252 patients in the validation set. Following coding the data, univariate, multivariate, Chi-square and other analyses were performed as explained in the next sections.

3.5.1 Univariate survival analysis

Univariate analysis tests for one-to-one correlation of a particular variable with a time-dependent outcome. In univariate survival analysis, OS (the primary end-point) was analyzed using OS status and OS time (the time from diagnosis of colorectal cancer until death from any cause). DFS (the secondary end-point) was analyzed using DFS status and DFS time (the time from diagnosis of colorectal cancer until the first occurrence of recurrence, metastasis or death from any cause). The genotype, demographic, molecular and clinicopathological data and prognostic data collected in an Excel document were fed to PASW software. Analyses were performed to explore correlations between genotypes and other variables and OS and DFS.

Cox-regression and Kaplan-Meier survival analyses were performed for each variable separately. These analyses were also repeated separately for OS and DFS. Cox-regression analysis gave the p-value and the hazard-ratio with 95% confidence intervals while Kaplan-Meier analysis was used to construct survival curves. We have used Cox-regression analysis for construction of multivariate models as well as for univariate analyses. Cox-regression analysis is a proportional hazard regression method for analysis of time to event outcomes. This method has two main assumptions (193). Firstly, the patients who do not experience outcome at the time of last follow-up are censored. This allows the patients who did not experience the event to be included in the analysis. Secondly, the proportionality assumption states that the relative hazard of an event for persons in one group is constant over time and does not change over the course of the follow-up period. One way to check the proportionality assumption is to check the Kaplan-Meier curves for intersection. Two-sided p-values less than 0.05 were considered statistically significant. In these analyses, the patients who were alive at the time of last follow-up were censored. The statistical results obtained were exported from PASW and organized in an Excel document.

3.5.2 Chi-square test and Mann-Whitney U-test

Chi-square test was performed to check for multicollinearity between the variables included in this study (genotypes, clinicopathological, and molecular variables). If two variables were highly correlated, only one would be included in the multivariate model to reduce redundancy (e.g. vascular and lymphatic invasion, **section 4.2.5**). The Chi-square test was performed using the PASW statistical package using crosstabs analysis under descriptive statistics. The results obtained were exported from the PASW and organized in an Excel document. Chi-square test was also performed to test for significant differences between the discovery set (n=532) and entire NFCCR cohort (n=735); validation set (n=252) and entire validation cohort (n=280); and between the discovery set (n=532) and validation set (n=252) to check the comparability of the cohorts. Age, which is a continuous variable, were not normally distributed in either cohorts. Hence we used the non-parametric Mann-Whitney U-test to compare median age between the cohorts.

3.5.3 Multivariate survival analysis

Multivariate Cox-regression analysis results show the independent predictive potential of each variable in the final model. To obtain reliable results in multivariate analyses, it is desirable to have at least 10 outcomes for each variable (193). In the discovery set, for selection of variables to be entered in the final multivariate model, all the variables (genotypes, demographic, clinicopathological and prognostic data) were entered together in Cox-regression analysis and backward selection method (likelihood ratio (LR)) was performed. Backward selection method sequentially eliminates the statistically insignificant variables and provides the list of selected variables with highest statistical significance (194). This method selectively reduces the large number of variables to a small group of the most relevant variables. It is worth noting that in our analysis, using this selection method, variables with well-known prognostic significance (such as sex, age, stage, MSI-status) remained in the final model. These selected variables were then entered into the multivariate Cox-regression analysis to obtain the final multivariate analysis result.

In the validation set, our aim was to test the validity of the variables that were found to be independently correlated with OS in the discovery set. Therefore, these variables were entered together in multivariate Cox-regression analysis for OS. For DFS analysis in the validation set, only the variables with available data were entered in the multivariate analysis. The discovery and the validation sets were also pooled together and Cox-regression analysis was repeated for both OS and DFS. Although the two cohorts were found to be dissimilar in many aspects (see **section 4.3.3**), for exploratory purposes we combined the cohorts (i.e. the pooled cohort) and repeated the analysis to observe the associations of polymorphisms in a larger sample set. We also analyzed the multivariate model for OS in the discovery, validation and pooled sets in the male and female patients separately. In this study, we did not perform correction for multiple testing.

3.6 Construction of linkage disequilibrium (LD) maps

For *MTHFR*_Glu429Ala and *ERCC5*_His46His polymorphisms, the LD maps were created using the Haploview 4.2 software (195). For this purpose, the SNP genotype data for a 100kb region containing the gene of interest for Caucasian population was downloaded from the International HapMap Project website using the data in HapMap Genome Browser release #28 (Phases 1, 2 & 3-merged genotypes and frequencies) (196).

Chapter 4. Results

4.1 Genotype data

The quality control measures in terms of successful duplication rate and successful genotyping and the missing genotype data for 27 polymorphisms in the discovery set and 4 polymorphisms in the validation set are enlisted in **Table 7**.

To verify the genotypes obtained, at least 5.9% of the genotypes were successfully duplicated for each polymorphism with at least 99.7% concordance rate. The minimum successful genotyping rate was 97.4% in the discovery cohort and 94.4% in the validation cohort. The mAF of polymorphisms in the discovery and validation cohorts were also similar to those described in dbSNP (186) or to literature reports for Caucasian populations and are shown in **Table 8**.

Hardy-Weinberg Equilibrium (HWE) Calculations

In the discovery cohort, four polymorphisms deviated from HWE: *ERCC2*_Lys751Gln, *OGG1*_Ser326Cys, *VEGFA*_-634 G/C and *XRCC3*_Thr241Met. The remaining polymorphisms in the discovery set and the four polymorphisms analyzed in the validation set were in HWE. Reasons for deviation of genotype frequencies from HWE can be many such as errors in genotyping, founder effect, genetic drift, assorted mating or reproductive benefit for heterozygotes over wild-type homozygotes (197). However, it is suggested in the literature that deviation of genotype frequencies from HWE should not be a critical parameter for inclusion or exclusion of a polymorphism in the analysis (197).

Table 7. Genotype data quality measures

Gene	Polymorphism	SNP ID	Successful genotype duplication rate	Missing genotypes (n)	% successful genotyping
Discovery set					
ERCC2	Lys751Gln G/T	rs13181	6.43%	8	98.5
GSTP1	Ile105Val A/G	rs1695	6.23%	7	98.7
MTHFR	Glu429Ala A/C	rs1801131	6.40%	6	98.9
MTHFR	Ala222Val C/T	rs1801133	6.24%	8 (1 discordant)	98.5
VEGFA	-634G/C in 5’-UTR	rs2010963	6.43%	8	98.5
XRCC1	Arg399Gln G/A	rs25487	6.12%	14 (1 discordant)	97.4
ERCC5	His46His C/T	rs1047768	6.36%	2	99.6
OGG1	Ser326Cys C/G	rs1052133	6.34%	1	99.8
ERCC1	Asn118Asn C/T	rs11615	6.34%	1	99.8
TYMS	indel 6 bp in 3’-UTR	rs16430	6.16%	6	98.9
MLH1	Ile219Val A/G	rs1799977	6.34%	1	99.8
FAS	c.24+733T>C	rs1800682	6.36%	2	99.6
IL6	-174G/C in promoter	rs1800795	6.17%	2	99.6
EGFR	Arg521Lys G/A	rs2227983	6.36%	2	99.6
DCC	Arg201Gly C/G	rs2229080	6.36%	2	99.6
MMP2	-1306C/T in promoter	rs243865	6.36%	2	99.6
VEGFA	+936C/T in 3’-UTR	rs3025039	6.34%	1	99.8
FGFR4	Gly388Arg A/G	rs351855	6.34%	1	99.8
PTGS2	c.3618A/G in 3’UTR	rs4648298	6.26%	10	98.1
XRCC3	Thr241Met C/T	rs861539	6.34%	1	99.8

CCND1	Pro241Pro A/G	rs9344	6.17%	2	99.6
EXO1	Pro757Leu C/T	rs9350	6.34%	1	99.8
MMP1	-1607indelG in promoter	rs1799750	7.08%	0	100
SERPINE1	-675 indelG in promoter	rs1799889	7.45%	0	100
GSTT1	gene deletion	-	6.90%	0	100
GSTM1	gene deletion	-	6.30%	0	100
TYMS	2/3 repeats of 28 bp	rs34743033	7.09%	1	98.7
Validation set					
MTHFR	Glu429Ala A/C	rs1801131	8.80%	2	99.2
ERCC5	His46His	rs1047768	13.22%	10	96.0
SERPINE1	-675indelG in promoter	rs1799889	8.98%	7	97.2
GSTM1	Gene deletion	-	5.90%	14	94.4

Successful genotype duplication rate is the ratio of the number of samples successfully genotyped more than once to the total number of samples successfully genotyped. % successful genotyping is the percentage of samples successfully genotyped. Concordance for the duplicate genotypes obtained from UHN is 99.73% whereas in TaqMan® assays it was 100%. Concordance is the percentage of duplicated genotypes yielding concordant results. The discordant genotypes obtained in the duplicated samples using Sequenom MassArray® were not included in the analyses.

Table 8. Minor allele frequencies (mAF) of the polymorphisms studied

Gene Symbol	Polymorphism	mAF Caucasian	mAF in discovery (validation) set
CCND1	Pro241Pro A/G	48-63%	45.28%
DCC	Arg201Gly C/G	33-42%	36.98%
EGFR	Arg521Lys G/A	22-30%	26.89%
ERCC1	Asn118Asn C/T	33-45%	37.57%
ERCC2	Lys751Gln G/T	27-42%	35.69%
ERCC5	His46His C/T	32-51%	41.13% (42.15%)
EXO1	Pro757Leu C/T	15-27%	14.60%
FAS	-670A/G in promoter	39-50%	44.91%
FGFR4	Gly388Arg A/G	26-31%	31.26%
GSTM1	gene deletion	*38-62%	45.10% (44.54%)
GSTP1	Ile105Val A/G	29-42%	36.67%
GSTT1	gene deletion	*15-20%	17%
IL6	-174G/C in promoter	50-57%	44.25%
MLH1	Ile219Val A/G	0-35%	28.63%
MMP1	-1607 indelG in promoter	43.30%	46.90%
MMP2	-1306C/T in promoter	18-25%	22.92%
MTHFR	Glu429Ala A/C	33-38%	30.61% (30.00%)
MTHFR	Ala222Val C/T	21-37%	31.77%
OGG1	Ser326Cys C/G	15-22%	23.54%
PTGS2	c.3618A/G in 3'-UTR	1.7-1.8%	1.63%
SERPINE1	-675 indelG in promoter	54.30%	46.71% (46.53%)
TYMS	indel 6 bp in 3'-UTR	37.00%	34.13%
TYMS	2/3 repeats of 28-bp	44.60%	46.60%
VEGFA	-634G/C in 5'-UTR	20-43%	29.10%
VEGFA	+936C/T in 3'-UTR	10-22%	10.73%
XRCC1	Arg399Gln G/A	37-58%	34.36%
XRCC3	Thr241Met C/T	37-65%	39.74%

*mAFs obtained from a published report (198). mAF information for other variations were retrieved from the dbSNP database (186).

In addition, founder effect is prominent in the Newfoundland population (199) and therefore HWE assumptions may not be fulfilled. Hence, we included the four polymorphisms which deviated from HWE in our analyses. For HWE calculations for the polymorphisms included in this study, see **Table A1** in **appendix**. The polymorphisms with χ^2 value greater than 3.84 were considered to be deviating from HWE (191). Of special note, none of the polymorphisms that deviated from HWE were in the multivariate analysis models described in this thesis. Therefore, their inclusion into our analysis did not alter our main results.

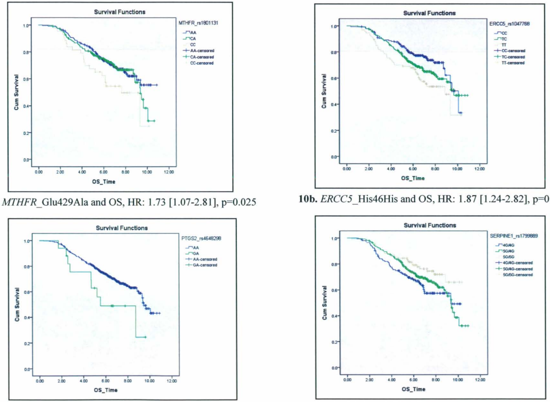
4.2 Univariate analysis

4.2.1 Polymorphisms correlated with OS

For exploratory purposes, univariate Cox-regression analysis was performed and Kaplan-Meier survival plots were obtained for each polymorphism. Since we observed that the co-dominant model gives a more robust result compared to the recessive and dominant models, statistically significant correlations ($p<0.05$) in only the co-dominant inheritance model are discussed here. For complete tabulated results of the analyses for co-dominant, recessive and dominant models, refer to **Tables A2, A3 and A4** in the **appendix**.

Six polymorphisms showed statistically significant correlations with OS in univariate analysis, assuming a co-dominant inheritance model (**Figure 10**).

Figure 10a-10f. Kaplan-Meier survival plots for polymorphisms and OS in the discovery set (co-dominant model)

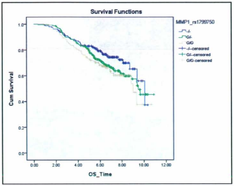


10a. *MTHFR*_Glu429Ala and OS, HR: 1.73 [1.07-2.81], p=0.025

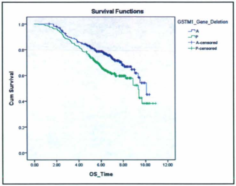
10b. *ERCC5*_His46His and OS, HR: 1.87 [1.24-2.82], p=0.003

10c. *PTGS2*_c.3618 A/G and OS, HR: 2.02 [1.03-3.95], p=0.041

10d. *SERPINE1*_-675 indelG and OS, HR: 0.56 [0.35-0.89], p=0.013



10e. *MMP1*_-1607 indelG and OS, HR: 1.54 [1.01-2.34], p=0.044



10f. *GSTM1* gene deletion and OS. HR: 1.48 [1.10-1.99], p=0.009
A=absence of gene and P=presence of gene

1) ***MTHFR*_Glu429Ala (NG_013351.1:g.16685A>C) (rs1801131).** Patients homozygous for alanine (CC) have a worse OS compared to patients homozygous for glutamate (AA) (p=0.025, HR=1.733, 95% CI: [1.070-2.807]) (**Figure 10a**).

2) ***ERCC5*_His46His (NG_007146.1:g.11344T>C) (rs1047768).** Patients homozygous for T allele have a worse OS compared to patients homozygous for C allele (p=0.003, HR=1.87, 95% CI: [1.238-2.824]) (**Figure 10b**).

3) ***PTGS2*_c.3618 A/G in 3'-UTR (NC_000001.10:g.186641682T>C) (rs4648298).** Heterozygotes (GA) have a worse OS compared to patients homozygous for A allele (p=0.041, HR=2.016, 95% CI: [1.030-3.946]). The mAF for this polymorphism is very low (1.63%). Hence we excluded this polymorphism from mulitivariate analysis to prevent obtaining unreliable statistical results (193) (**Figure 10c**).

4) ***SERPINE1*_-675 indelG (NG_013213.1:g.4332_4333insA) (rs1799889).** Patients homozygous for insG allele had a favorable OS compared to patients homozygous for delG (p=0.013, HR=0.557, 95% CI: [0.351-0.885]) (**Figure 10d**).

5) ***MMP1*_-1607 indelG (NG_011740.1:g.3471delG) (rs1799750).** Patients homozygous for insG allele had a worse OS compared to patients homozygous for delG (p=0.044, HR=1.539, 95% CI: [1.012-2.339]) (**Figure 10e**).

6) ***GSTM1* gene deletion.** Patients having at least one copy of the gene had a worse OS when compared to patients homozygous for deletion of the gene (p=0.009, HR=1.484, 95% CI: [1.104-1.994]) (**Figure 10f**).

The results on polymorphisms without statistically significant associations with OS are shown in **Table A2** in **appendix**.

4.2.2 Clinicopathological features correlated with OS

We also performed univariate Cox-regression analysis and constructed Kaplan-Meier survival plots to test correlation of clinicopathological variables with OS. The results are depicted in **Table 9**.

Sex, higher stages, vascular invasion, lymphatic invasion and MSI status were correlated with OS (**Figure 11**).

1) **Sex:** Males had approximately 50% greater hazard of death when compared to females (p=0.012, HR=1.501, 95% CI: [1.09-2.06]) (**Figure 11a**).

2) **Stage:** Stage III (p=0.005, HR=2.151, 95% CI=1.26-3.68) and stage IV (p<0.001, HR=10.211, 95% CI: [5.80-17.98]) patients had a greater hazard of death when compared to stage I patients (**Figure 11b**).

3) **Vascular invasion:** Patients with tumor vascular invasion had ~67% greater hazard of death when compared to patients without tumor vascular invasion (p=0.001, HR=1.674, 95% CI: [1.23-2.28]) (**Figure 11c**).

4) **Lymphatic invasion:** Patients with lymphatic invasion had an approximately 54% greater hazard of death when compared to patients without lymphatic invasion (p=0.006, HR=1.535, 95% CI: [1.13-2.08]) (**Figure 11d**)

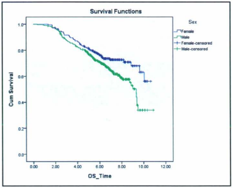
Table 9. Clinicopathological features correlated with OS in univariate analysis

(discovery set)

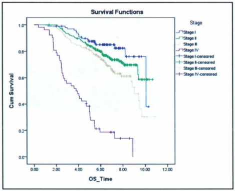
Variable	p-value	HR	95% CI	n
Sex (males vs females)	**.012**	**1.501**	**1.09-2.06**	**531**
Age at diagnosis	.230	1.010	0.99-1.03	531
Histology (mucinous vs non-mucinous)	.990	0.997	0.63-1.59	531
Location (rectum vs colon)	.129	1.264	0.93-1.71	531
Stage	**<.001**			**531**
II vs I	.182	1.449	0.84-2.50	
III vs I	**.005**	**2.151**	**1.26-3.68**	
IV vs I	**<.001**	**10.211**	**5.80-17.98**	
Grade (poorly/undifferentiated vs well/moderately differentiated)	.735	0.900	0.49-1.66	527
Vascular invasion (+ vs -)	**.001**	**1.674**	**1.23-2.28**	491
Lymphatic invasion (+ vs -)	**.006**	**1.535**	**1.13-2.08**	**488**
Familial risk (high/intermediate vs low)	.751	1.049	0.78-1.41	531
MSI status (MSI-H vs MSI-L/S)	**<.001**	**0.156**	**0.06-0.42**	**510**
BRAF1-Val600Glu mutation status (+ vs -)	.447	0.813	0.48-1.39	483

HR: hazard ratio, CI: confidence interval, n: number of patients, HR>1 implies increased hazard of death, HR<1 implies reduced hazard of death.

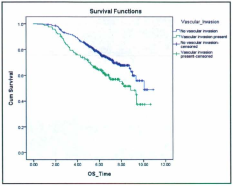
Figures 11a-11e. Kaplan-Meier survival plots for clinicopathological features and OS in the discovery set



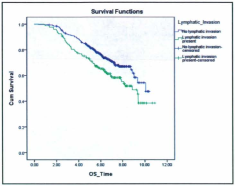
11a. Sex and OS
HR: 1.50 [1.09-2.06], p=0.012



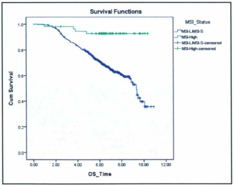
11b. Stage and OS
stage II vs stage I, HR: 1.45 [0.84-2.50], p=0.182
stage III vs stage, I HR: 2.15 [1.26-3.68], p=0.005
stage IV vs stage I, HR: 10.21 [5.80-17.98], p<0.001



11c. Vascular invasion and OS
HR: 1.67 [1.23-2.28], p=0.001



11d. Lymphatic invasion and OS
HR: 1.54 [1.13-2.08], p=0.006



11e. MSI status and OS
HR: 0.16 [0.06-0.42], p<0.001

5) **MSI status:** Patients with MSI-H tumors had a survival advantage compared to patients with MSI-L/MSS tumors: they had ~85% reduced hazard of death (p<0.001, HR=0.156, 95% CI: [0.06-0.42]) (**Figure 11e**).

4.2.3 Polymorphisms correlated with DFS

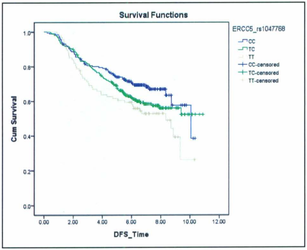
In univariate analysis assuming co-dominant inheritance model, two polymorphisms were significantly correlated with DFS (**Figures 12**).

1) ***ERCC5*_His46His (NG_007146.1:g.11344T>C) (rs1047768).** Patients homozygous for T allele had a worse DFS compared to patients homozygous for C allele (p=0.01, HR=1.647, 95% CI: [1.124-2.414]) (**Figure 12a**).

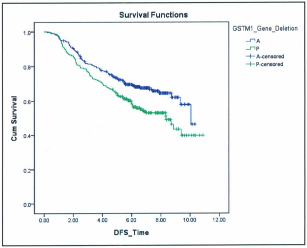
2) ***GSTM1* gene deletion.** Patients with at least one copy of the gene had a worse DFS when compared to patients homozygous for gene deletion (p=0.004, HR=1.489, 95% CI: [1.133-1.957]) (**Figure 12b**).

Both these polyporphisms were also associated with OS in the discovery cohort in univariate analysis (**section 4.2.1**). The results on polymorphisms without statistically significant associations with DFS are shown in **Table A5** in the **appendix**. Results for recessive and dominant models are shown in **Tables A6 and A7** in **appendix**.

Figures 12a-12b. Kaplan-Meier survival plots for polymorphisms and DFS in the discovery set (co-dominant model)



12a. *ERCC5*_His46His and DFS, HR: 1.65 [1.12-2.41], p=0.01



12b. *GSTM1* gene deletion and DFS, HR: 1.49 [1.13-1.96], p=0.004
A=absence of gene and P=presence of gene

4.2.4 Clinicopathological features correlated with DFS

The results for univariate Cox-regression analysis for correlation between clinicopathological features and DFS are shown in **Table 10**. Six clinicopathological features were correlated with DFS in univariate Cox-regression analysis (**Figure 13**).

1) **Sex:** Males had an approximate 47% greater hazard of event compared to females (p=0.01, HR=1.471, 95% CI: [1.097-1.973]) (**Figure 13a**).

2) **Location:** Patients with rectal cancer had ~40% greater hazard of event when compared to colon cancer patients (p=0.017, HR=1.403, 95% CI: [1.062-1.854]) (**Figure 13b**).

3) **Stage:** Stage III patients have ~100% greater hazard of event (p=0.002, HR=2.096, 95% CI: [1.314-3.345]) while stage IV patients have ~478% greater hazard of event (p<0.001, HR=5.778, 95% CI: [3.476-9.604]) when compared to stage I patients (**Figure 13c**).

4) **Vascular invasion:** Patients with tumor vascular invasion have ~60% greater hazard of event when compared to patients without tumor vascular invasion (p=0.001, HR=1.604. 95% CI: [1.206-2.134]) (**Figure 13d**).

5) **Lymphatic invasion:** Patients with lymphatic invasion have ~50% greater hazard of event when compared to patients without lymphatic invasion (p=0.005, HR=1.498, 95% CI: [1.129-1.988]) (**Figure 13e**).

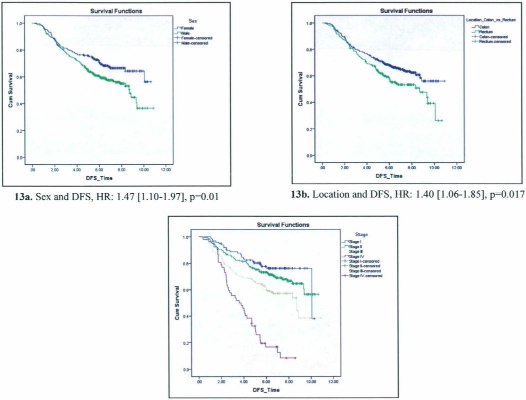
6) **MSI status:** Patients with MSI-H tumors had favorable survival when compared to

Table 10. Clinicopathological features correlated with DFS in univariate analysis (discovery set)

Variable	p-value	HR	95% CI		n
Sex (male vs female)	**0.01**	**1.471**	**1.097**	**1.973**	**530**
Age at diagnosis	0.62	1.004	0.989	1.019	530
Histology (mucinous vs non-mucinous)	0.861	0.962	0.624	1.484	530
Location (rectum vs colon)	**0.017**	**1.403**	**1.062**	**1.854**	**530**
Stage	**<0.001**				**530**
II vs I	0.248	1.324	0.823	2.131	
III vs I	**0.002**	**2.096**	**1.314**	**3.345**	
IV vs I	**<0.001**	**5.778**	**3.476**	**9.604**	
Grade (poorly diff/undiff vs well diff/moderately diff)	0.534	0.831	0.464	1.489	526
Vascular invasion (+ vs -)	**0.001**	**1.604**	**1.206**	**2.134**	**490**
Lymphatic invasion (+ vs -)	**0.005**	**1.498**	**1.129**	**1.988**	**487**
Familial risk (high/moderate vs low)	0.33	1.146	0.871	1.506	530
MSI status (MSI-H vs MSI-L/MSS)	**<0.001**	**0.279**	**0.137**	**0.566**	**509**
BRAF1 Val600Glu mutation (+ vs -)	0.714	0.915	0.57	1.47	483

HR: hazard ratio, CI: confidence interval, n: number of patients, diff: differentiated, HR>1 implies increased hazard of event, HR<1 implies reduced hazard of event.

Figure 13a-13f. Kaplan-Meier survival plots for clinicopathological features and DFS in the discovery set

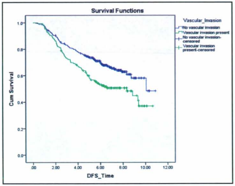


13a. Sex and DFS, HR: 1.47 [1.10-1.97], p=0.01

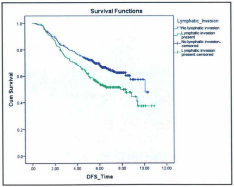
13b. Location and DFS, HR: 1.40 [1.06-1.85], p=0.017

13c. Stage and DFS

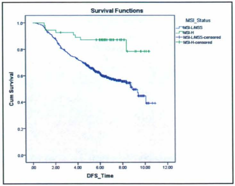
stage II vs stage I, HR: 1.32 [0.82-2.13], p=0.248; stage III vs stage I, HR: 2.10 [1.31-3.35], p=0.002; stage IV vs stage I, HR: 5.78 [3.48-9.60], p<0.001



13d. Vascular invasion and DFS, HR: 1.60 [1.21-2.13], p=0.001



13e. Lymphatic invasion and DFS, HR: 1.50 [1.13-2.00], p=0.005



13f. MSI status and DFS, HR: 0.28 [0.14-0.57], p<0.001

patients with MSI-L/MSS tumors with an approximately 72% reduction of hazard for the event ($p<0.001$, HR=0.279, 95% CI: [0.137-0.566]) (**Figure 13f**).

4.2.5 Chi-square test results for correlation between clinicopathological and molecular variables

We performed this analysis in the discovery set to test for association amongst clinicopathological and molecular variables. The statistically significant correlations ($p<0.05$) are depicted in (**Table 11**).

Female sex was correlated with colonic location, MSI-H tumors and presence of *BRAF1*_Val600Glu mutation in the tumors. Majority of the mucinous tumors were found in the colon and were also correlated with MSI-H and *BRAF1*_Val600Glu mutation in the tumors. *BRAF1*_Val600Glu mutation was also correlated with MSI-H tumors and mucinous histology. MSI-H tumors were mostly found in the colon, had high grade and were mostly found in early stage (stage I and II) patients. Presence of vascular and lymphatic invasions was correlated with increasing stage and high grade tumors. Vascular and lymphatic invasions were highly correlated with one another ($p=2.68\times10^{-100}$). Hence for multivariate analyses, only the data on vascular invasion status was included into the survival analysis to reduce redundancy.

Table 11. Correlation between clinicopathological and molecular variables (discovery set)

Variables	p-value	Correlation between	n
Location and **Sex**	0.039	females and colon	532
MSI and **Sex**	0.01	females and MSI-H	511
***BRAF1* mutation** and **Sex**	<0.001	females and mutation	484
Histology and **Location**	0.014	mucinous and colon	532
MSI and **Location**	<0.001	MSI-H and colon	511
***BRAF1*_Val600Glu** and **Location**	<0.001	mutation and colon	484
Stage and **Histology**	0.027	stage I and non-mucinous, stage II and mucinous	532
MSI and **Histology**	0.038	MSI-H and mucinous	511
***BRAF1* mutation** and **Histology**	0.048	mutation and mucinous	484
Vascular invasion and **Stage**	<0.001	invasion and stage	492
Lymphatic invasion and **Stage**	<0.001	invasion and stage	489
MSI and **Stage**	0.037	MSI-H and stages I & II	511
Vascular invasion and **Grade**	0.014	invasion and poorly differentiated	488
Lymphatic invasion and **Grade**	0.041	invasion and poorly differentiated	485
MSI and **Grade**	0.01	MSI-H and poorly diff/undiff	507
Lymphatic and **Vascular invasions**	<0.001	presence of invasion	486
***BRAF1*_Val600Glu** and **MSI**	<0.001	mutation and MSI-H	477

Only statistically significant associations are shown in the table, n: number of patients.

4.3 Multivariate analysis for OS

4.3.1 Multivariate analysis for OS in the discovery set (co-dominant model)

Multivariate analysis is performed to test for independent predictive value of a variable when adjusted for other variables in the model. The variables were selected for entry into multivariate analysis as explained in **section 3.5.3**. **Table 12** shows the multivariate analysis result for OS assuming co-dominant inheritance in the discovery set. For all the polymorphisms associated in the multivariate analysis in discovery cohort, the proportionality assumption was met in the univariate analysis (**Figure 10**).

In multivariate analysis, four polymorphisms showed an independent prognostic potential when adjusted for sex, age, stage and MSI status. For *MTHFR*_Glu429Ala (NG_013351.1:g.16685A>C), patients homozygous for the alanine variant had ~72% greater hazard of death when compared to patients homozygous for glutamate (p=0.036, HR=1.715, 95% CI: [1.036-2.839]). For *ERCC5*_His46His (NG_007146.1:g.11344T>C), patients homozygous for T had significantly worse OS with ~78% greater hazard of death when compared to patients homozygous for C (p=0.01, HR=1.782, 95% CI: [1.150-2.763]). For *SERPINE1*_-675 indelG (NG_013213.1:g.4332_4333insA), patients homozygous for insG had favorable OS with ~48% reduced hazard of death when compared to patients homozygous for delG allele (p=0.008, HR=0.517, 95% CI: [0.319-0.840]). In case of *GSTM1* gene deletion, patients with at least one copy of the gene had worse OS (~40% increased hazard) compared to patients homozygous for deletion of the gene (p=0.033, HR=1.404, 95% CI: [1.027-1.919]). Male sex, increasing age and stages

Table 12. Multivariate analysis result for OS in the discovery set (n=504)

(co-dominant model)

Variable	p-value	HR (95% CI)	n
*MTHFR*_rs1801131	0.105		
CA vs AA	0.342	1.175 (0.842-1.639)	230 vs 232
CC vs AA	**0.036**	**1.715 (1.036-2.839)**	**42 vs 232**
***ERCC5*_rs1047768**	**0.034**		
TC vs CC	0.098	1.365 (0.944-1.973)	240 vs 173
TT vs CC	**0.01**	**1.782 (1.15-2.763)**	**91 vs 173**
***SERPINE1*_rs1799889**	**0.029**		
G/- vs -/-	0.238	0.809 (0.569-1.15)	258 vs 141
GG vs -/-	**0.008**	**0.517 (0.319-0.84)**	**105 vs 141**
***GSTM1* gene deletion (+ vs -)**	**0.033**	**1.404 (1.027-1.919)**	**228 vs 276**
Sex (male vs female)	**0.031**	**1.456 (1.036-2.047)**	**313 vs 191**
Age at diagnosis	**0.046**	**1.018 (1-1.036)**	
Stage	**<0.001**		
II vs I	0.18	1.473 (0.836-2.594)	194 vs 95
III vs I	**0.01**	**2.084 (1.194-3.637)**	**165 vs 95**
IV vs I	**<0.001**	**11.685 (6.454-21.158)**	**50 vs 95**
MSI status (MSI-H/ MSI-L-MSS)	**0.004**	**0.233 (0.086-0.635)**	**56 vs 448**

*MTHFR*_rs1801131 is *MTHFR*_Glu429Ala, *ERCC5*_rs1047768 is *ERCC5*_His46His, *SERPINE1*_rs1799889 is *SERPINE1*_-675 indelG, G allele for *SERPINE1*_-675 indelG is referred to as insG allele and – allele is referred to as delG allele in the text, HR: hazard ratio, CI: confidence interval, n: number of patients, HR>1 implies increased hazard of death, HR<1 implies reduced hazard of death.

III and IV had an increased hazard of death while patients having tumors with MSI-H status had a significantly favorable OS.

After obtaining these results, we aimed to replicate them in another independent colorectal cancer patient cohort also from Newfoundland (the validation set). For this purpose, we obtained their genotypes for four polymorphisms (*MTHFR*_Glu429Ala, *ERCC5*_His46His, *SERPINE1*_-675 indelG and *GSTM1* gene deletion) correlated with OS in the multivariate analysis in the discovery set, and the multivariate analysis was repeated.

4.3.2 Multivariate analysis for OS in the validation set (co-dominant model)

In the validation set, only the *MTHFR*_Glu429Ala polymorphism showed independent prognostic value when adjusted for age, stage and MSI status (**Table 13**). Interestingly, while we had found the association of Ala/Ala homozygotes with worse OS in the discovery set, in the validation set, heterozygotes (Glu/Ala) had ~71% increased hazard of death when compared with Glu/Glu homozygotes (p=0.005, HR=1.713, 95% CI: [1.181-2.487]). Thus the same polymorphism (*MTHFR*_Glu429Ala) was correlated with worse OS in the discovery and validation sets, although with different patterns (homozygosity for alanine in the discovery set and heterozygosity in the validation set). In order to explore more, we also performed separate multivariate analysis with *MTHFR*_Glu429Ala genotypes assuming recessive and dominant models, together with the other clinicoptahological variables in the model (sex, age, stage and MSI status). Again we have found that the CC (Ala/Ala) genotype was associated with worse OS in

Table 13. Multivariate analysis result for OS in the validation set (n=224)

(co-dominant model)

Variable	p-value	HR (95% CI)	n
***MTHFR*_rs1801131**	**.010**		
AC vs AA	**.005**	**1.713 (1.181-2.487)**	**92 vs 112**
CC vs AA	.730	0.889 (0.454-1.738)	20 vs 112
*ERCC5*_rs1047768	.609		
TC vs CC	.387	1.197 (0.796-1.8)	112 vs 76
TT vs CC	.398	1.261 (0.737-2.159)	36 vs 76
*SERPINE1*_rs1799889	.716		
G/- vs -/-	.420	1.187 (0.782-1.802)	103 vs 69
GG vs -/-	.766	1.075 (0.669-1.727)	52 vs 69
GSTM1 gene deletion (+ vs -)	.261	1.234 (0.855-1.780)	99 vs 125
Sex (males vs females)	.175	1.282 (0.895-1.837)	118 vs 106
Age at diagnosis	**<.001**	**1.051 (1.034-1.069)**	
Stage	**<.001**		
II vs I	.662	1.144 (0.626-2.092)	80 vs 44
III vs I	**.001**	**2.609 (1.446-4.707)**	**64 vs 44**
IV vs I	**<.001**	**11.324 (5.918-21.669)**	**36 vs 44**
MSI status (MSI-H vs MSI-L/MSS)	**.002**	**0.257 (0.108-0.609)**	**21 vs 203**

*MTHFR*_rs1801131 is *MTHFR*_Glu429Ala, *ERCC5*_rs1047768 is *ERCC5*_His46His, *SERPINE1*_rs1799889 is *SERPINE1*_-675 indelG, G allele for *SERPINE1*_-675 indelG is referred to as insG allele and – allele is referred to as delG allele in the text, HR: hazard ratio, CI: confidence interval, n: number of patients, HR>1 implies increased hazard of death, HR<1 implies reduced hazard of death.

the discovery set when compared to AA+AC genotypes (i.e. recessive inheritance pattern). However, in the validation set, AC+CC (Ala/Glu and Ala/Ala) genotypes were associated with worse OS when compared to AA (Glu/Glu) genotype (data not shown) (dominant inheritance pattern).

4.3.3 Differences between discovery and validation sets

Three of the four the polymorphisms correlated with OS in the discovery set were not correlated in the validation set i.e. *ERCC5*_His46His, *SERPINE1*_-675 indelG and *GSTM1* gene deletion. However, the *MTHFR*_Glu429Ala Ala variant was associated with shorter OS in both sets (homozygosity for alanine in the discovery set and heterozygosity for alanine in the validation set correlated with shorter OS). We sought to understand these results by first looking at the differences between the discovery and validation sets in terms of their important clinicopathological and prognostic characteristics. Apart from a large difference in the sample size (discovery set has more than twice the number of patients in the validation set), the cohorts also differed in other features. To test if these differences were significant, we performed chi-square tests and Mann-Whitney U tests (**Table 14**). We observed that the validation set had a significantly higher median age (68.7 years compared to 61.36 years in the discovery set, p<0.001). This is expected since patients were recruited in the validation set regardless of their age and in the discovery set below 75 years of age. The validation set also had a greater proportion of deaths (61.51% compared to 33.3% in the discovery set, p<0.001) and greater proportion of events (recurrence/metastasis/death) (66.27% compared to 39.1% in the discovery set, p<0.001) which may be due to the longer follow-up times for patients

Table 14. Differences between the discovery and validation sets

	Discovery (n=532)	Validation (n=252)	χ^2/Mann-Whitney U test		Discovery (n=532)	Validation (n=252)	χ2/Mann-Whitney U test
Sex	**n (%)**	**n (%)**		**OS status**	**n (%)**	**n (%)**	
male	327 (61.50%)	133 (52.78%)		dead	177 (33.30%)	155 (61.51%)	
female	205 (38.50%)	119 (47.22%)	p=**0.021**	alive	354 (66.60%)	97 (38.49%)	
Median age (yrs)	**61.36 (20.7-75)**	**68.7 (25.3-91.6)**	**p<0.001**	unknown	1 (0.10%)	-	**p<0.001**
Histology				**DFS status**			
non-mucinous	471 (88.50%)	211 (83.73%)		no event	323 (60.71%)	85 (33.73%)	
mucinous	61 (11.50%)	41 (16.27%)	p=0.062	event**	208 (39.1%)	167 (66.27%)	
Location				unknown	1 (0.19%)	-	**p<0.001**
colon	353 (66.40%)	202 (80.16%)		**MSI Status**			
rectum	179 (33.60%)	50 (19.84%)	**p<0.001**	MSI-H	56 (10.50%)	24 (9.52%)	
Stage				MSS/MSI-L	455 (85.50%)	228 (90.48%)	
I	99 (18.60%)	48 (19.05%)		unknown	21 (4%)	-	p=0.543
II	206 (38.70%)	88 (34.92%)		***Vascular/Lymphatic invasion**			
III	175 (32.90%)	68 (26.98%)		-	326 (61.30%)	64 (25.40%)	
IV	52 (9.80%)	41 (16.27%)		+	166 (31.20%)	101 (40.08%)	
unknown	-	7 (2.78%)	p=**0.034**	unknown	40 (7.50%)	87 (34.52%)	**p<0.001**
Grade				**5-FU based treatment**			
well diff/moderately diff	489 (91.90%)	211 (83.73%)		5-FU treated	330 (62.03%)	88 (34.92%)	
poorlydiff/undiff	39 (7.30%)	37 (14.68%)		other/no chemotherapy	199 (37.41%)	148 (58.73%)	
unknown	4 (0.80%)	4 (1.59%)	p=**0.001**	unknown	3 (0.56%)	16 (6.35%)	**p<0.001**

*Vascular invasion in the discovery set and lymphatic invasion in the validation set were compared. Familial risk status and *BRAF1*_Val600Glu mutation status data were not available for the validation set samples and hence were not compared.**event refers to the first occurrence of recurrence, metastasis or death.

in the validation set. Even in the age-adjusted survival curves (**section 2.3**, **Figure 4** and **Figure 5**), the difference in survival times of the two cohorts remained significant. In addition, the proportion of rectal cancer patients was greater in the discovery set (33.6% compared to 19.84% in the validation set, p<0.001). Also, the proportion of patients without lymphatic/vascular invasion in the validation set was low (25.4% compared to 61.3% in the discovery set, p<0.001). There were also treatment related differences between the two cohorts. A large portion of patients in the discovery set received 5-FU based chemotherapy (~62%) compared to those in the validation set (~35%) and the difference was statistically significant (p<0.001). Additionally, the validation cohort had significantly greater proportion of female patients (p=0.021), stage IV patients (p=0.034) and patients with poorly differentiated or undifferentiated tumor grade (p=0.001) than the discovery cohort Thus a large number of differences between the two cohorts might be a likely reason for inconsistent results. These differences may partly account for the differences in correlations observed in the discovery and validation sets and are discussed in **section 5.3**.

4.3.4 Multivariate analysis for OS in the pooled set (co-dominant model)

We then combined the discovery and validation sets and performed the analysis again in this pooled sample set since it has a larger sample size and greater power for detection of correlations (**Table 15**). In the pooled set, when adjusted for age, stage and MSI status, *MTHFR*_Glu429Ala, *ERCC5*_His46His and *GSTM1* gene deletion show independent predictive potential for overall survival. For *MTHFR*_Glu429Ala, similar to the results in

Table 15. Multivariate analysis results for OS in the pooled sample set (n=728)

(co-dominant model)

Variable	**p-value**	**HR (95% CI)**	**n**
*MTHFR*_rs1801131	.106		
AC vs AA	**.035**	**1.298 (1.018-1.654)**	**322 vs 344**
CC vs AA	.660	1.094 (0.732-1.636)	62 vs 344
***ERCC5*_rs1047768**	**.007**		
TC vs CC	**.016**	**1.390 (1.064-1.816)**	**352 vs 249**
TT vs CC	**.003**	**1.652 (1.185-2.303)**	**127 vs 249**
*SERPINE1*_rs1799889	.381		
G/- vs -/-	.500	0.913 (0.700-1.190)	361 vs 210
GG vs -/-	.165	0.790 (0.566-1.102)	157 vs 210
***GSTM1* gene deletion (+ vs -)**	**.040**	**1.273 (1.011-1.604)**	**327 vs 401**
Sex (males vs females)	.146	1.197 (0.939-1.526)	431 vs 297
Age at diagnosis	**<.001**	**1.046 (1.034-1.059)**	
Stage	**<.001**		
II vs I	.091	1.419 (0.946-2.127)	274 vs 139
III vs I	**<.001**	**2.377 (1.592-3.550)**	**229 vs 139**
IV vs I	**<.001**	**10.735 (6.993-16.481)**	**86 vs 139**
MSI status (MSI-H vs MSI-L/MSS)	**<.001**	**0.269 (0.142-0.510)**	**77 vs 651**

*MTHFR*_rs1801131 is *MTHFR*_Glu429Ala, *ERCC5*_rs1047768 is *ERCC5*_His46His, *SERPINE1*_rs1799889 is *SERPINE1*_ -675 indelG, G allele for *SERPINE1*_-675 indelG is referred to as insG allele and – allele is referred to as delG allele in the text, HR: hazard ratio, CI: confidence intervals, n: number of patients, HR>1 implies increased hazard of death, HR<1 implies reduced hazard of death.

the validation set, the heterozygotes had worse survival when compared to homozygotes for the allele coding for the amino acid glutamate (Glu/Glu) with ~30% increased hazard of death (p=0.035, HR=1.30, 95% CI: [1.02-1.65]). For *ERCC5*_His46His, the heterozygotes (p=0.016, HR=1.39, 95% CI: [1.06-1.82]) and homozygotes for T allele (p=0.003, HR=1.65, 95% CI: [1.19-2.30]) had worse survival when compared to homozygotes for C allele. Patients having at least one copy of *GSTM1* gene had ~27% increased hazard of death when compared to patients with null allele (p=0.04, HR=1.27, 95% CI: [1.01-1.60]). Increasing age and stages III and IV were also correlated with poor OS. MSI-H status of tumor, as expected, was predictive of favorable prognosis.

4.3.5 Summary of results of multivariate analyses for OS

The results of multivariate analysis in the discovery set, validation set and pooled set are shown together in **Table 16**.

Because of the biological role of the MTHFR enzyme in 5-FU function (the main chemotherapeutic agent used in treatment of patients in the discovery and validation cohorts), we also attempted to replicate the multivariate model in those patients treated with 5-FU. This analysis, however, did not find association of this polymorphism in the discovery, validation or pooled set (data not shown).

Table 16. Summary of multivariate analysis results for OS in the discovery set (n=504), validation set (n=224) and pooled sample set (n=728) (co-dominant model)

	Discovery set (n=504, deaths=168)		**Validation set (n=224, deaths=134)**		**Pooled set (n=728, deaths=302)**	
Variable	**p-value**	**HR (95% CI)**	**p-value**	**HR (95% CI)**	**p-value**	**HR (95% CI)**
*MTHFR*_rs1801131	0.105		**0.01**		0.106	
CA vs AA	0.342	1.175 (0.842-1.639)	**0.005**	**1.713 (1.181-2.487)**	**0.035**	**1.298 (1.018-1.654)**
CC vs AA	**0.036**	**1.715 (1.036-2.839)**	0.73	0.889 (0.454-1.738)	0.66	1.094 (0.732-1.636)
***ERCC5*_rs1047768**	**0.034**		0.609		**0.007**	
TC vs CC	0.098	1.365 (0.944-1.973)	0.387	1.197 (0.796-1.80)	**0.016**	**1.390 (1.064-1.816)**
TT vs CC	**0.01**	**1.782 (1.15-2.763)**	0.398	1.261 (0.737-2.159)	**0.003**	**1.652 (1.185-2.303)**
***SERPINE1*_rs1799889**	**0.029**		0.716		0.381	
G/- vs -/-	0.238	0.809 (0.569-1.15)	0.42	1.187 (0.782-1.802)	0.5	0.913 (0.700-1.190)
GG vs -/-	**0.008**	**0.517 (0.319-0.84)**	0.766	1.075 (0.669-1.727)	0.165	0.790 (0.566-1.102)
***GSTM1* gene deletion (+ vs -)**	**0.033**	**1.404 (1.027-1.919)**	0.261	1.234 (0.855-1.780)	**0.04**	**1.273 (1.011-1.604)**
Sex (male vs female)	**0.031**	**1.456 (1.036-2.047)**	0.175	1.282 (0.895-1.837)	0.146	1.197 (0.939-1.526)
Age at diagnosis	**0.046**	**1.018 (1-1.036)**	**<0.001**	**1.051 (1.034-1.069)**	**<0.001**	**1.046 (1.034-1.059)**
Stage	**<0.001**		**<0.001**		**<0.001**	
stage II vs I	0.18	1.473 (0.836-2.594)	0.662	1.144 (0.626-2.092)	0.091	1.419 (0.946-2.127)
stage III vs I	**0.01**	**2.084 (1.194-3.637)**	**0.001**	**2.609 (1.446-4.707)**	**<0.001**	**2.377 (1.592-3.550)**
stage IV vs I	**<0.001**	**11.685 (6.454-21.158)**	**<0.001**	**11.324 (5.918-21.669)**	**<0.001**	**10.735 (6.993-16.481)**
MSI status (MSI-H/ MSI-L-MSS)	**0.004**	**0.233 (0.086-0.635)**	**0.002**	**0.257 (0.108-0.609)**	**<0.001**	**0.269 (0.142-0.510)**

*MTHFR*_rs1801131 is *MTHFR*_Glu429Ala, *ERCC5*_rs1047768 is *ERCC5*_His46His, *SERPINE1*_rs1799889 is *SERPINE1*_-675 indelG, G allele for *SERPINE1*_-675 indelG is referred to as insG allele and – allele is referred to as delG allele in the text, HR: hazard ratio, CI: confidence interval, n=number of patients, HR>1 implies increased hazard of death, HR<1 implies reduced hazard of death.

4.3.6 Multivariate analysis for OS in sex-stratified patients

To test for sex-specific differences in associations, we tested the applicability of the multivariate analysis model in males and females separately in the discovery, validation and pooled sample sets. The results of analysis in female and male patients are summarized in **Table 17** and **Table 18**, respectively. In the case of female patients, none of the polymorphisms was associated with OS in the discovery or validation sets. *ERCC5*_His46His polymorphism was correlated in the pooled set where the heterozygotes had ~78% increased hazard of death compared to CC homozygotes.

Interestingly, in male patients, all four polymorphisms were correlated with OS in the discovery set. For *MTHFR*_Glu429Ala, both the heterozygotes and Ala/Ala homozygotes had worse survival when compared to Glu/Glu homozygotes. The heterozygotes had ~52% increased hazard of death when compared to Glu/Glu homozygotes. In the validation set, the heterozygotes had ~116% increased hazard of death compared to Glu/Glu homozygotes. Thus correlation of heterozygotes with shorter OS in male patients was confirmed in the validation set. This suggests a sex-specific correlation of this polymorphism with OS. This observation may also be a reflection of the greater study power in the males than in females since males are present in a larger proportion than females in both the cohorts. In the pooled set, heterozygotes were correlated with worse OS with ~59% increased hazard of death. The *ERCC5*_His46His, *SERPINE1*_-675indelG polymorphisms and *GSTM1* gene deletion were correlated with OS in the discovery set but not in the validation set. Their correlation with OS was also observed in the pooled set.

Table 17. Multivariate analysis for OS in female patients (co-dominant model)

	Discovery set (n=191, deaths=54)		**Validation set (n=106, deaths=54)**		**Pooled sample set (n=297, deaths=108)**	
Variables	**p-value**	**HR (95% CI)**	**p-value**	**HR (95% CI)**	**p-value**	**HR (95% CI)**
*MTHFR*_rs1801131	0.586		0.1		0.673	
AC vs AA	0.332	0.744 (0.409-1.353)	0.163	1.524 (0.843-2.754)	0.643	1.100 (0.734-1.651)
CC vs AA	0.92	1.052(0.389-2.846)	0.223	0.388 (0.085-1.779)	0.542	0.774 (0.340-1.762)
*ERCC5*_rs1047768	0.15		0.53		0.057	
TC vs CC	0.051	1.968 (0.996-3.891)	0.268	1.452 (0.750-2.812)	**0.019**	**1.731 (1.094-2.737)**
TT vs CC	0.26	1.686 (0.679-4.19)	0.786	1.130 (0.468-2.728)	0.105	1.656 (0.900-3.046)
*SERPINE1*_rs1799889	0.91		0.15		0.503	
G/- vs -/-	0.934	1.028 (0.542-1.947)	0.073	1.987 (0.937-4.215)	0.352	1.251 (0.781-2.004)
GG vs -/-	0.72	0.806 (0.248-2.617)	0.088	2.109 (0.895-4.965)	0.273	1.418 (0.759-2.648)
GSTM1 gene deletion (+ vs -)	0.455	1.241 (0.705-2.184)	0.871	0.950 (0.511-1.765)	0.939	.985 (0.663-1.462)
Age at diagnosis	0.547	1.01 (0.978-1.043)	**0.003**	**1.040 (1.014-1.067)**	**<0.001**	**1.040 (1.021-1.060)**
Stage	**<0.001**		**<0.001**		**<0.001**	
II vs I	0.234	1.957 (0.648-5.911)	0.213	1.963 (0.679-5.675)	**0.034**	**2.254 (1.064-4.773)**
III vs I	0.074	2.82 (0.905-8.794)	**0.002**	**5.264 (1.817-15.250)**	**0.001**	**3.846 (1.781-8.305)**
IV vs I	**<0.001**	**13.373 (4.2-42.584)**	**<0.001**	**28.262 (9.192-86.895)**	**<0.001**	**22.335 (10.257-48.635)**
MSI status (MSI-H vs MSS/MSI-L)	**0.027**	**0.193 (0.045-0.829)**	**0.013**	**0.228 (0.071-0.728)**	**0.001**	**0.245 (0.105-0.568)**

*MTHFR*_rs1801131 is *MTHFR*_Glu429Ala, *ERCC5*_rs1047768 is *ERCC5*_His46His, *SERPINE1*_rs1799889 is *SERPINE1*_-675 indelG, G allele for *SERPINE1*_-675 indelG is referred to as insG allele and – allele is referred to as delG allele in the text, HR: hazard ratio, CI: confidence interval,n: number of patients, HR>1 implies increased hazard of death, HR<1 implies reduced hazard of death.

Table 18. Multivariate analysis for OS in male patients (co-dominant model)

	Discovery Set **(n=313, deaths=114)**		**Validation set** **(n=118, deaths=80)**		**Pooled set** **(n=431, deaths=194)**	
Variable	**p-value**	**HR (95% CI)**	**p-value**	**HR (95% CI)**	**p-value**	**HR (95% CI)**
***MTHFR*_rs1801131**	**0.024**		**.013**		**.015**	
AC vs AA	**0.048**	**1.516 (1.004-2.288)**	**.004**	**2.168 (1.284-3.660)**	**.004**	**1.592 (1.161-2.183)**
CC vs AA	**0.014**	**2.144 (1.168-3.937)**	.951	1.025 (0.459-2.293)	.267	1.317 (0.810-2.142)
*ERCC5*_rs1047768	0.085		.685		.070	
TC vs CC	0.661	1.106 (0.705-1.734)	.457	1.236 (0.707-2.162)	.171	1.263 (0.904-1.766)
TT vs CC	**0.037**	**1.72 (1.033-2.866)**	.439	1.334 (0.643-2.769)	.022	1.599 (1.071-2.387)
***SERPINE1*_rs1799889**	**0.019**		.279		**.033**	
G/- vs -/-	0.125	0.71 (0.458-1.1)	.438	.807 (0.470-1.387)	.054	0.717 (0.511-1.006)
GG vs -/-	**0.005**	**0.458 (0.265-0.789)**	.110	.601 (0.322-1.123)	**.013**	**0.601 (0.403-0.897)**
***GSTM1* gene deletion (+ vs -)**	**0.044**	**1.481 (1.01-2.17)**	.222	1.352 (0.834-2.192)	**.018**	**1.422 (1.061-1.904)**
Age at diagnosis	0.129	1.017 (0.995-1.039)	**<0.001**	**1.068 (1.043-1.095)**	**<0.001**	**1.055 (1.037-1.072)**
Stage	**<0.001**		**<0.001**		**<0.001**	
II vs I	0.461	1.288 (0.657-2.527)	.885	.947 (0.451-1.988)	.474	1.195 (0.734-1.944)
III vs I	**0.033**	**2.024 (1.06-3.867)**	**.032**	**2.263 (1.072-4.778)**	**.002**	**2.122 (1.318-3.416)**
IV vs I	**<0.001**	**11.808 (5.744-24.276)**	**. <0.001**	**6.717 (2.860-15.776)**	**<0.001**	**7.366 (4.333-12.522)**
MSI status (MSI-H vs MSS/MSI-L)	0.062	0.26 (0.063-1.067)	**.036**	**.206 (0.047-0.900)**	**.012**	**0.279 (0.103-0.758)**

*MTHFR*_rs1801131 is *MTHFR*-Glu429Ala, *ERCC5*_rs1047768 is *ERCC5*-His46His, *SERPINE1*_rs1799889 is *SERPINE1* -675 indelG, G allele for *SERPINE1*_-675 indelG is referred to as insG allele and – allele is referred to as delG allele in the text, HR=hazard ratio, CI=confidence interval, n: number of patients, HR>1 implies increased hazard of death, HR<1 implies reduced hazard of death.

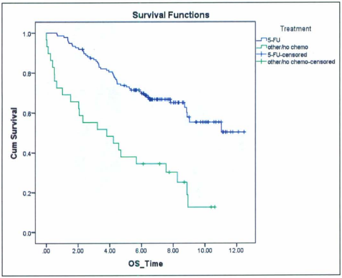
4.4 Treatment with 5-FU and survival in stage III colon cancer patients

5-FU alone or in combination with other drugs is the most widely used chemotherapeutic agent in treatment of stage III colon cancer (200,201). To see the effect of 5-FU treatment on patient survival, we compared survival times of stage III colon cancer patients in the pooled set treated with 5-FU (n=134) and those who received no chemotherapy or were treated with other drugs (n=29). Stage III colon cancer patients were chosen since most of these patients are treated with 5-FU (~82%). In addition, the analysis was performed in the pooled set to have a large sample size.

In the univariate analysis, as expected (200,201), patients treated with 5-FU had longer survival times (p<0.001) when compared to other patients (**Figure 14**). A multivariate analysis including MSI-H status and age also showed that 5-FU treatment is a MSI status independent prognostic factor and patients who received other chemotherapy or no chemotherapy had ~235% increased hazard of death (p<0.001, HR=3.348, 95% CI: [2.034-5.511]). These results confirm that 5-FU treatment improves survival in treated patients.

Since MTHFR enzyme is indirectly involved in the mechanism of action of 5-FU (201), we also analyzed the *MTHFR*_Glu429Ala polymorphism (which reduces MTHFR activity) with survival in stage III colon cancer patients treated with 5-FU from the pooled set (n=134). The polymorphism was not correlated with survival in both univariate and multivariate analyses, suggesting that the *MTHFR*_Glu429Ala polymorphism does not affect survival in 5-FU treated stage III colon cancer patients,

Figure 14. Kaplan-Meier plot for stage III colon cancer patients based on treatment characteristics (pooled set, OS)



although these results could have also been obtained due to small sample size. This analysis was not performed in patients who received other chemotherapy or no chemotherapy due to small sample size (n=29).

As an exploratory analysis, we also wanted to analyze the combined effect of polymorphisms in *MTHFR* and *TYMS* genes on survival in 5-FU treated stage III colon cancer patients since these proteins are involved in 5-FU pathway. The genotypes for the four polymorphisms *MTHFR* (Glu429Ala and Ala222Val) and *TYMS* (2R/3R VNTR and indel6bp) were available only for the samples in the discovery set (n=106). This small sample size made it imposible to perform the statistical analyses intended above (data not shown). Thus currently it is not known whether these polymorphisms in *MTHFR* and *TYMS* genes affect survival in our 5-FU treated patients.

4.5 Multivariate analysis for DFS

DFS was our secondary end-point for analysis. In the discovery set, similar to analysis with OS, we selected the variables using the backward elimination LR method and performed multivariate Cox-regression analysis. The results for analyses in the discovery, validation and pooled sets are shown in **Table 19**. In the discovery set, polymorphisms in ERCC5 and OGG1 genes were correlated with DFS after adjustment for stage and MSI status. For the *ERCC5*_His46His polymorphism, patients homozygous for the T allele

Table 19. Multivariate analysis for DFS in the discovery set (n=504), validation set (n=227) and pooled sample set (n=734) (co-dominant model)

	Discovery set **(n=504, events=198)**		**Validation set** **(n=227, events=148)**		**Pooled sample set** **(n=734, deaths=348)**	
Variable	p-value	HR (95% CI)	p-value	HR (95% CI)	p-value	HR (95% CI)
***ERCC5*_rs1047768**	0.098		**0.036**		**0.007**	
TC vs CC	0.211	1.235 (0.887-1.72)	**0.041**	**1.483 (1.015-2.167)**	**0.035**	**1.304 (1.018-1.670)**
TT vs CC	**0.032**	**1.54 (1.039-2.288)**	**0.018**	**1.805 (1.107-2.943)**	**0.002**	**1.611 (1.190-2.182)**
*OGG1*_rs1052133	0.082					
GC vs CC	0.59	1.088 (0.801-1.477)				
GG vs CC	**0.025**	**1.81 (1.075-3.038)**				
*ERCC1*_rs11615	0.152					
TC vs TT	0.281	1.193 (0.866-1.643)				
CC vs TT	0.054	1.477 (0.993-2.196)				
*TYMS*_rs16430	0.171					
6 bp/- vs 6 bp/6 bp	0.235	0.831 (0.611-1.128)				
-/- vs 6 bp/6 bp	0.325	1.252 (0.8-1.96)				
GSTM1 gene deletion (+ vs -)	0.09	1.278 (0.962-1.698)	0.366	1.167 (0.835-1.632)	0.125	1.179 (0.955-1.456)
Location (rectum vs colon)	0.055	1.334 (0.994-1.789)	0.743	1.070 (0.714-1.604)	0.386	1.107 (0.88-1.392)
Stage	**<0.001**		**<0.001**		**<0.001**	
II vs I	0.099	1.512 (0.925-2.472)	**0.036**	**1.821 (1.041-3.187)**	**0.013**	**1.588 (1.101-2.292)**
III vs I	**0.003**	**2.09 (1.281-3.407)**	**<0.001**	**3.144 (1.793-5.513)**	**<0.001**	**2.321 (1.614-3.339)**
IV vs I	**<0.001**	**6.24 (3.692-10.533)**	**<0.001**	**130.162 (52.48-322.83)**	**<0.001**	**7.721 (5.224-11.414)**
MSI status (MSI-H vs MSI-L/MSS)	**0.004**	**0.35 (0.168-0.71)**	**0.007**	**0.366 (0.176-0.758)**	**<0.001**	**0.373 (0.225-0.621)**

*ERCC5*_rs1047768 is *ERCC5*_His46His, *OGG1*_rs1052133 is *OGG1*_Ser326Cys, *ERCC1*_rs11615 is *ERCC1*_Asn118Asn, *TYMS*_rs16430 is *TYMS*_indel 6 bp in 3'-UTR, 6 bp in *TYMS*_rs16430 refers to the sequence CTTTAA, HR: hazard ratio, CI: confidence interval, n: number of patients, HR>1 implies increased hazard of event, HR<1 implies reduced hazard of event.

had shorter DFS (~54% increased hazard of event) compared to patients homozygous for the C allele (p=0.032, HR: 1.542, 95% CI: [1.039-2.288]). For the *OGG1*_Ser326Cys polymorphism, patients homozygous for cysteine had significantly reduced DFS (~81% increased hazard) compared to patients homozygous for serine (p=0.025, HR:1.808, 95% CI: [1.075-3.038]). The proportionality assumption was fulfilled for associations of *ERCC5*_His46His and *OGG1*_Ser326Cys polymorphisms with DFS in the univariate analysis. In addition, tumor stages III and IV were correlated with significantly worse DFS when compared to stage I, and MSI-H status of tumor was correlated with a favorable DFS.

For analysis in the validation set, genotypes for *OGG1*_rs1052133, *ERCC1*_rs11615 and *TYMS*_rs16430 polymorphisms were not available. On analyzing the available variables (*ERCC5*_His46His, *GSTM1* gene deletion, location, stage and MSI status), both the heterozygotes and minor allele homozygotes for *ERCC5*_His46His C/T were correlated with worse DFS when adjusted for stage and MSI status. T allele homozygotes had ~81% increased hazard of event when compared to C allele homozygotes (p=0.018, HR: 1.805, 95% CI: [1.107-2.943]). Heterozygotes had ~48% increased hazard of the event (p=0.041, HR: 1.483, 95% CI: [1.015-2.167]). Thus the results suggest the association of *ERCC5*_His46His with poor DFS in colorectal cancer patients.

In the pooled set, *ERCC5*_His46His was again correlated with worse DFS when adjusted for stage and MSI status. Both the heterozygotes (~30% increased hazard) (p=0.035, HR=1.304, 95% CI: [1.018-1.67]) and homozygotes for T allele (~61% increased hazard of event) (p=0.002, HR=1.611, 95% CI: [1.190-2.182]) had significantly worse DFS when compared to homozygotes for C allele.

When the multivariate analysis results for DFS and OS were compared, *ERCC5*_His46His polymorphism, which was associated with DFS in the discovery, validation and pooled sets, was also associated with OS in the discovery and pooled cohorts (**Table 16** and **Table 19**). *MTHFR*_Glu429Ala polymorphism was associated with OS in discovery, validation and pooled cohorts, but did not remain in the multivariate model of DFS. Two other polymorphisms associated with OS in the discovery cohort namely *SERPINE1*_-675indelG and *GSTM1* gene deletion, were not associated with DFS in multivariate analysis. In the case of clinicopathological and demographic variables, sex, age, stage and MSI status were significantly associated with OS in the discovery cohort while only stage and MSI status were found to be significantly associated with DFS in multivariate analysis in all three cohorts.

Chapter 5. Discussion

Colorectal cancer is a critical health concern in Newfoundland since it has the highest age-standardized incidence and mortality rates in Canada (27). In recent years, there has been an upsurge in genetic prognostic studies performed in various colorectal cancer patient cohorts in an attempt to identify independent genetic prognostic markers. Identification of genetic prognostic markers may not only help in clinical prognostication of the patients but will also help us to understand the underlying mechanisms of variable prognosis in patients. For this thesis project, we have performed genetic prognostic research in two independent colorectal cancer patient cohorts from Newfoundland. The survival end-points analyzed were OS (primary end-point) and DFS (secondary end-point).

In the first stage of the project, 27 genetic polymorphisms were analyzed in relation to OS and DFS in a discovery cohort of 532 patients from the NFCCR. The second stage of the project was for the replication of results obtained in the first stage in a validation set comprising an additional 252 colorectal cancer patients, also from Newfoundland. For OS, a sex-stratified analysis was also performed in the discovery and validation sets.

Compared to most other genetic prognostic studies in colorectal cancer, this retrospective cohort study has certain unique strengths. This is the first such study conducted in the Newfoundland population and amongst the few in Canada. In addition to external validation of previously reported correlations, we have performed an internal validation in which we tried to replicate the initial findings in another cohort from Newfoundland.

Such internal validation studies are rarely found in the literature. Both cohorts have a significantly large number of patients followed-up for a significant duration (up to over 10 years), a resource which only a few research groups have.

5.1 Univariate analysis results for OS in the discovery set

In univariate analysis, six polymorphisms were correlated with OS in the discovery set in the co-dominant model: *MTHFR*_Glu429Ala, *ERCC5*_His46His, *PTGS2*_c.3618A/G in 3`-UTR, *SERPINE1*_-675 indelG, *MMP1*_-1607 indelG and *GSTM1* gene deletion. *PTGS2*_3618A/G was excluded from multivariate analysis because of its low minor allele frequency (1.63%) in order to prevent unreliable statistical results (193). Correlations with the remaining 21 polymorphisms were not detected in this cohort.

5.2 Multivariate model for OS in the discovery set

The multivariate analysis model for the discovery set assuming codominant inheritance includes eight variables, each of which had independent predictive value for OS when adjusted for other variables in the model. Male sex, increasing age, tumors with advanced stage (III and IV) and MSI-L/MSS were predictive of poor survival. Along with these clinicopathological variables, four genetic polymorphisms showed independent predictive value for OS: *MTHFR*_Glu429Ala, *ERCC5*_His46His, *SERPINE1*_-675 indelG and *GSTM1* gene deletion.

For *ERCC5*_His46His, our finding suggests worse OS (~78% increased hazard) for patients homozygous for T allele (TT) when compared to patients homozygous for C allele (CC). This result is similar to two other studies in which patients homozygous for T allele had a worse OS and PFS (84,87). Two other studies did not find a correlation of this polymorphism with OS (75,87). *ERCC5*_His46His is a non-splice site synonymous polymorphism whose functional impact is not clearly known and its potential biological role in prognosis of cancer patients remains to be elucidated.

In case of *SERPINE1*_-675 indelG, the insG allele has been linked to lower transcriptional activity of the gene (144). The functional role of SERPINE1 in cancer prognosis is ambiguous. For example, it has been shown to reduce tumor angiogenesis at high concentration while at low concentration it has been shown to induce tumor angiogenesis and metastasis (202,203). On the other hand, studies in animal models as well as *in vitro* experiments suggest that the worse prognosis of high SERPINE1 expression due to delG allele may be due to its pro-metastatic and pro-angiogenic effect via multiple mechanisms such as altering cell migration and adhesion properties (203). In our study, patients homozygous for the insG allele, which is associated with decreased transcription of the gene, had ~ 48% reduced hazard of death compared to the patients homozygous for delG allele, which may be due to the reduced pro-angiogenic and pro-metastatic abilities of the protein. Our finding is concordant with that in a Swedish colorectal cancer patient cohort in which insG homozygotes had a favorable prognosis compared to heterozygotes and delG homozygotes (145).

In the case of a *GSTM1* gene deletion, patients with at least one copy of the *GSTM1* gene showed ~40% increased hazard of death when compared to patients with null genotype. Most patients in the discovery set were treated with 5-FU based chemotherapy and/or radiotherapy and it is known that part of the mechanism of these therapies is through generation of reactive oxygen species (ROS) which cause oxidative damage to the tumor cells (204,205). A possible explanation for our finding could be the enhanced efficacy of these therapies in patients with *GSTM1* null genotypes leading to favorable prognosis. This result contrasts with the findings in a small Hungarian cohort of colorectal cancer patients in which Dukes' stage B colorectal cancer patients (n=34) with homozygous deletion of *GSTM1* gene had worse OS when compared to patients with at least one copy of the gene (95). This discrepancy between our results and Csejtei et al. (95) study may be due to differences in patient cohort size and stage (95). However, several other studies also did not find a correlation of this gene deletion with OS (77,78,96,97,101).

For *MTHFR*_Glu429Ala, patients homozygous for the amino acid alanine (Ala/Ala) had ~72% increased hazard of death when compared to patients homozygous for the amino acid glutamate (Glu/Glu). This correlation of alanine variant with poor survival is concordant with another study in a Spanish colorectal cancer patient cohort in which patients homozygous for the amino acid glutamate (Glu/Glu) had favorable OS (137). In another study, a result discordant with ours was reported. Female colorectal cancer patients (mixed ethnicities) homozygous for amino acid glutamate (Glu/Glu) were reported to have favorable OS relative to other genotypes (Ala/Ala and Glu/Ala) (133). However, several other studies did not find a correlation with this polymorphism

(76,78,128,130,134,135). This polymorphism and its relation to prognosis are discussed in detail in the later sections.

5.3 Multivariate analysis for OS in the validation set

We next aimed to replicate the multivariate model in the discovery set in the validation set (consisting of 252 patients from Newfoundland) including sex, age, stage, MSI-status, *MTHFR*_Glu429Ala, *ERCC5*_His46His, *SERPINE1*_-675 indelG and *GSTM1* gene deletion genotypes. In the validation set, the correlations of age, stage and MSI-status, but not sex were replicated. Similar to the results in the discovery set, increasing age, advanced stages (III and IV) and MSI-L/MSS were significantly correlated with worse OS in the validation set. In the case of genetic polymorphisms, *ERCC5*_His46His, *SERPINE1*_-675 indelG polymorphisms and *GSTM1* gene deletion were not correlated with OS in the validation set. Therefore, their results in the discovery set were not replicated in the validation set. However, interestingly, *MTHFR*_Glu429Ala polymorphism was correlated with OS, although this time, the heterozygotes (Glu/Ala) had worse prognosis compared to homozygotes for glutamate (Glu/Glu). This association is different than that in the discovery set where homozygotes for alanine (Ala/Ala) had poor prognosis.

5.4 Possible reasons for differences in results obtained in the discovery and validation sets

Our validation study did not validate the associations of *ERCC5*_His46His, *SERPINE1*_-675 indelG polymorphisms and *GSTM1* gene deletion with OS. However, we found the association of two different genotypes with OS in the case of *MTHFR*_Glu429Ala polymorphism. While these two genotypes (CC homozygous genotype coding for the alanine variant in the discovery set and AC heterozygous genotype coding for both alanine and glutamate variants in the validation set) were different from each other, nevertheless, they contained the same allele (C allele coding for alanine variant). The possible reasons for such an observation could be:

i) Chance of correlations being false positives or false negatives

ii) Differences in study power in two cohorts

iii) Differences between the two cohorts

iv) Sex-specific effects

v) Other polymorphisms in linkage disequilibrium with *MTHFR*_Glu429Ala

i) **Chance of correlations being false positives or false negatives:** It is possible that the correlations observed in the discovery set, which were not replicated in the validation set (*ERCC5*_His46His, *SERPINE1*_-675indelG, *GSTM1* gene deletion) are false positives, particularly in case of the *SERPINE1*_-675 indelG, *ERCC5*_His46His polymorphisms and the *GSTM1* gene deletion. Alternatively, it is possible that the results obtained in the validation set are false negatives. Considering the small sample size of the validation set, it may not have enough power to detect a similar effect (see below).

ii) **Differences in study power in two cohorts:** Our analysis showed that heterozygotes generally have more study power to detect a correlation since they are in greater numbers compared to minor allele homozygotes (data not shown). Correlations of minor allele homozygotes with OS were observed in the discovery set for *MTHFR*_Glu429Ala, *ERCC5*_His46His, *SERPINE1*_-675 indelG polymorphisms and *GSTM1* gene deletion, but not in the validation set. This might be due to the insufficient power because of the small cohort size in the validation set (less than half the size of discovery set) and the lower number of minor allele homozygotes when compared to the discovery set. Alternatively, the observation may also be due to smaller effect-size of the polymorphisms in the validation cohort than the discovery cohort, which might have remained undetected.

iii) **Differences between the two cohorts:** The validation set is not fully comparable to the discovery set in terms of cohort size, number of events and a few variables (e.g. age). Specifically, the validation set has a greater percentage (or earlier occurrence) of deaths than the discovery cohort (62% compared to 33% in discovery set, $p<0.001$) and this cohort is characterized by patients with a statistically significantly higher median age compared to the discovery set ($p<0.001$). It is also likely that medical care might have been different for the discovery and validation cohort patients, since they were recruited at different time periods.

It is also likely that inter-patient variability in folate intake or bioavailability can modify the prognosis of the patients. Additionally, folate pathway involves a number of other genes which may be polymorphic (206). These variations may also modify the effect of *MTHFR*_Glu429Ala in colorectal cancer prognosis. It is known that older individuals have an impaired ability to absorb dietary folate (207). Therefore the age difference between the cohorts may also explain why we detected an association with different patterns (homozygosity in discovery set and heterozygosity in the validation set) of *MTHFR*_Glu429Ala with OS in these two cohorts. Possible differences between young and old colorectal cancer patients in terms of folate pathway are discussed in detail in **section 5.4.2**.

Additionally, a significantly greater proportion of patients in the discovery set were treated with 5-FU compared to patients in the validation set. This difference may account for the higher OS rate of the discovery set patients compared to the validation set patients, even after age-adjustment (**section 2.3**).

iv) **Sex-specific effect:** It is also possible that the differences in associations in the two cohorts may be due to sub-group effects. For example, in females, none of the polymorphisms were correlated with OS in the multivariate analysis in either patient set. But in males, the association of heterozygotes for *MTHFR*_Glu429Ala with OS was detected in both sets. This result suggests that prognostic mechanisms may differ between male and female colorectal cancer patients and it is discussed in detail in

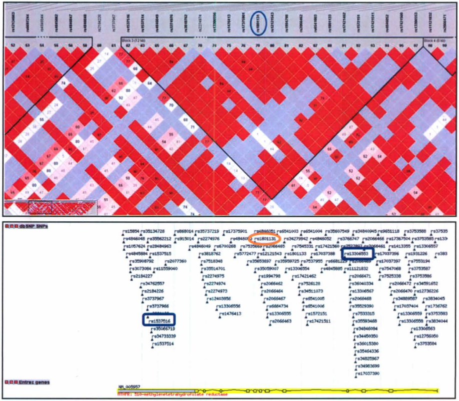
section 5.5.

v) **Other polymorphisms in LD with *MTHFR*_Glu429Ala:** *MTHFR*_Glu429Ala lies in a 12 kb long LD block which has a number of other known SNPs **(Figure 15)**. It is possible that the true prognostic marker, if it indeed exists, may be a SNP in close proximity to *MTHFR*_Glu429Ala in this LD block with a high (but not complete) correlation with it. For example, another polymorphism *MTHFR*_Ala222Val is in the same LD block as *MTHFR*_Glu429Ala but these two SNPs are not correlated with each other (data not shown). *MTHFR*_Ala222Val results in a thermolabile enzyme and causes a more significant reduction in the MTHFR enzyme activity than *MTHFR*_Glu429Ala (136,208). It is also reported that MTHFR activity is further reduced if these two polymorphisms are present together (136,208,209). This polymorphism was included in our study too. However, it was not associated with OS in the discovery set. Further studies on other SNPs in this LD block and their correlations with prognosis are warranted.

5.5 Folate pathway, *MTHFR*_Glu429Ala polymorphism and their possible relation to cancer prognosis

Although the patterns of associations differ, *MTHFR*_Glu429Ala polymorphism was associated with OS in both the discovery and validation sets. In the discovery set, Ala/Ala homozygotes had ~72% increased hazard of death compared to Glu/Glu homozygotes

Figure 15. LD block of *MTHFR*_Glu429Ala (rs1801131)



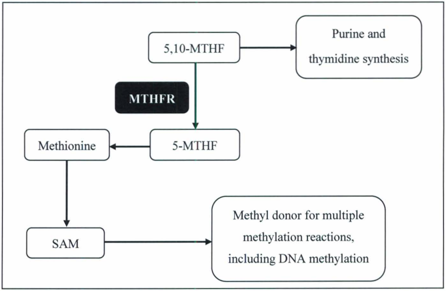
The black triangle shows the LD block in which *MTHFR*_Glu429Ala (rs1801131) is located (circled). Below the LD map, other known polymorphisms in this block are shown. rs1537516 and rs13306553 (which are shown in boxes) are the first and last SNPs respectively of the LD block. rs1801131 is circled to help demonstrate the relative position of this polymorphism within this LD block.

(p=0.036, HR: 1.715, 95% CI: [1.04-2.84]) while in the validation set, the heterozygotes (Glu/Ala) had ~71% increased hazard of death compared to Glu/Glu homozygotes (p=0.005, HR: 1.713, 95% CI: [1.181-2.487]).

Both genotypes (CC, AC) are known to lead to reduced MTHFR enzyme activity (136,208). The role of the *MTHFR*_Glu429Ala polymorphism in colorectal cancer outcome seems to be complex and currently not well understood. Based on the current literature findings about this variant and its function, the following mechanisms by which *MTHFR* variants leads to poor outcome can be suggested.

Folate, also known as vitamin B_9, is an essential molecule for one-carbon transfer reactions. MTHFR is involved in folate metabolism where it converts 5,10-methylene tetrahydrofolate (5,10-MTHF) to 5-methyl tetrahydrofolate (5-MTHF), which is the circulatory form of folate (206). Both forms of folate mediate one-carbon transfer reactions although for different purposes. 5,10-MTHF is predominantly used for the *de novo* synthesis of thymidine and purines which are used by the replicating cells for DNA synthesis whereas 5-MTHF is predominantly used for synthesis of methionine from homocysteine, which is then used for synthesis of S-adenosyl methionine (SAM) (206). SAM serves as a methyl donor for a large number of biological reactions, including methylation of DNA (206) (**Figure 16**). MTHFR enzyme has two domains, a catalytic domain and a regulatory domain and the Glu429Ala polymorphism lies in the regulatory domain of the protein (210). Studies in human lymphocytes have reported reduced MTHFR enzyme activity in alanine variant (136,208). Although both heterozygotes and

Figure 16. Folate pathway with normal MTHFR activity

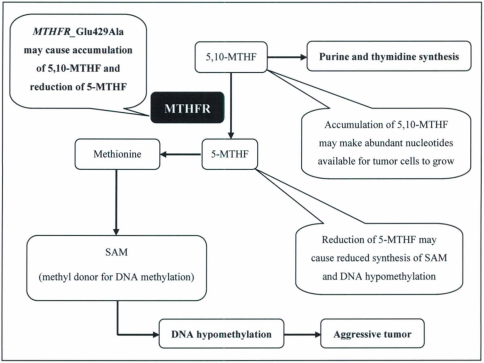

5,10-MTHF
Purine and thymidine synthesis
MTHFR
Methionine
5-MTHF
SAM
Methyl donor for multiple methylation reactions, including DNA methylation

5,10-MTHF: 5,10-methylene tetrahydrofolate, 5-MTHF: 5-methyl tetrahydrofolate, MTHFR: methylene tetrahydrofolate reductase, SAM: S-adenosyl methionine homozygotes have reduced enzyme activity, the effect is more severe in alanine homozygotes which have ~60% of the normal enzymatic activity, while the heterozygotes have ~80% of the normal enzymatic activity (136,208). Reduced MTHFR activity can thus result in the accumulation of 5,10-MTHF and a concurrent reduction of 5-MTHF since the former is not efficiently converted to the latter. We believe that the accumulation of 5,10-MTHF and concurrent reduced availability of 5-MTHF may lead to poor prognosis in patients. From clinical trials and animal studies, the role of folate supplementation in prevention of colorectal cancer has been established (211). However, reports have recently emerged which suggest different roles of folate supplementation in different scenarios, i.e. folate supplementation indeed prevents development of colorectal adenoma but once a colorectal adenoma has developed, high folate intake in fact aids its growth and progression (211-214). In rat models of colorectal cancer, folate supplementation has been associated with progression of already developed cancer (213). Also, in the Aspirin/Folate Polyp Prevention Study, folate supplementation was associated with higher risk of advanced adenomas as well as increased number of adenomas in patients with previously established colorectal adenomas (211,213). It is believed that with folate supplementation, the greater availability of nucleotide precursors is used by the rapidly dividing tumor cells which favor tumor progression (206,212,213,215). In the case of our study, it is likely that for patients with reduced MTHFR enzyme activity (Ala/Ala homozygotes and Glu/Ala heterozygotes for *MTHFR*_Glu429Ala polymorphism), the accumulation of 5,10-MTHF, which is predominantly used for nucleotide synthesis, may make nucleotide precursors available to tumor cells in abundance. This may have assisted tumor growth and progression eventually leading to poor prognosis (**Figure 17**). In a study using knockout mice with heterozygous or homozygous deletions of the *MTHFR* gene, it was observed that the amount of SAM as well as the extent of DNA methylation were significantly reduced, suggesting that reduced MTHFR activity (in our case, due to Glu429Ala polymorphism) may lead to similar, although less severe observation (216). A Harvard group also reported that global DNA hypomethylation in colon tumor cells was correlated with worse cancer-specific survival as well as OS in two independent cohorts with over 600 samples (217). Thus, reduced activity of MTHFR due to *MTHFR*_Glu429Ala may have led to reducedsynthesis of SAM, and this may have led to DNA hypomethylation which in turn could have led to poor prognosis in our patients. DNA hypomethylation is known to induce carcinogenesis by mechanisms such as rendering the DNA hypermutable and inducing strand breaks, destabilizing the chromatin's conformation, deregulating gene transcription or even triggering inflammatory pathways (215,217). These mechanisms may increase tumor aggression as well and lead to poor prognosis (217). These hypotheses and possible explanations are based on literature findings, often ambiguous, and hence need to be further evaluated.

5.5.1 Correlation of Glu/Ala heterozygotes with worse OS in the validation set

In the validation set, the heterozygotes for *MTHFR*_Glu429Ala had a worse OS compared to Glu/Glu homozygotes while in the discovery set, Ala/Ala homozygotes had poor OS. This difference in associations may be due to the age-specific differences in the

Figure 17. Hypothesized changes in folate pathway with reduced MTHFR activity due to *MTHFR*_Glu429Ala polymorphism


*MTHFR*_Glu429Ala may cause accumulation of 5,10-MTHF and reduction of 5-MTHF
5,10-MTHF
Purine and thymidine synthesis
MTHFR
Accumulation of 5,10-MTHF may make abundant nucleotides available for tumor cells to grow
Methionine
5-MTHF
Reduction of 5-MTHF may cause reduced synthesis of SAM and DNA hypomethylation
SAM
(methyl donor for DNA methylation)
DNA hypomethylation
Aggressive tumor

5,10-MTHF: 5,10-methylene tetrahydrofolate, 5-MTHF: 5-methyl tetrahydrofolate, MTHFR: methylene tetrahydrofolate reductase, SAM: S-adenosyl methionine folate pathway. The validation set has a significantly higher median age compared to the discovery set ($p<0.001$). It is also known that older individuals have an inherent reduced ability to absorb dietary folate (207). We hypothesize that although the low availability of folate may not provide ample amount of nucleotide precursors for tumor progression, reduced absorption of folate coupled with reduced MTHFR activity may lead to a severe deficiency of available 5-MTHF in aged individuals. This may have caused severe deficiency of SAM and subsequent DNA hypomethylation. Hence this association may be age-specific in older individuals and heterozygosity of the polymorphism may be sufficient to cause worse prognosis (**Figure 17**). In this case, we would also expect to find association of the Ala/Ala homozygotes with OS as well. This possible association might have been missed because of the low number of homozygotes in this cohort (i.e. because of insufficient power).

5.6 Validation of correlation of *MTHFR*_Glu429Ala polymorphism with OS in male patients (co-dominant model)

In the sub-set of male patients, correlation of *MTHFR*_Glu429Ala polymorphism was replicated in the validation set. In both the discovery and validation sets, the hetereozygotes (Glu/Ala) had a worse OS when compared to Glu/Glu homozygotes. The Ala/Ala homozygotes were also associated with worse OS in the male patients of the discovery set. However, in female patients, none of the polymorphisms were correlated with OS either in the discovery set or validation set. Although this may be due to lack power (i.e. false negative findings), these data suggest a gender-specific correlation of this polymorphism with OS.

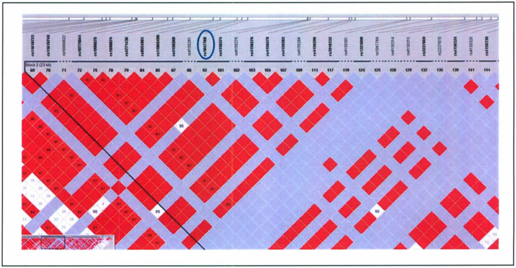
Reports on gender-specific differences for *MTHFR*_Glu429Ala or in the folate pathway are limited. In one study in healthy Singaporean Chinese individuals, males had a significantly greater extent of methylation of the *MTHFR* gene compared to females (218). If this does cause an inherent reduction in *MTHFR* gene expression in men, then the lower amount of MTHFR coupled with the Glu429Ala polymorphism may have led to increased 5,10-MTHF and reduced 5-MTHF in males compared to females. This increase in 5,10-MTHF and concurrent decrease in 5-MTHF may have led to worse prognosis in males via increased availability of nucleotide precursors for tumor cells and increased DNA hypomethylation respectively. This male-specific correlation with survival in our study is in conflict with a previous study in a cohort of 141 metastatic colorectal cancer patients in which female patients homozygous for glutamate (Glu/Glu) had a longer OS compared to female patients homozygous for alanine (Ala/Ala) or heterozygotes (Glu/Ala) after univariate analysis (133). However, all the patients in that study were stage IV patients (metastatic colorectal cancer) and these authors did not perform a multivariate analysis. Our study predominantly contains early stage patients and includes multivariate analysis. Therefore their results are not directly comparable to ours.

5.7 Validation of correlation of *ERCC5*_His46His polymorphism with DFS in the validation set (co-dominant model)

DFS was our secondary end-point for analysis and included the patients who experienced recurrence or metastasis in addition to those included in OS analysis. In the discovery set, the *ERCC5*_His46His and *OGG1*_Ser326Cys polymorphisms along with stage and MSI status were correlated with DFS. For *ERCC5*_His46His (C>T), patients homozygous for the T allele had worse DFS compared to homozygotes for C allele (p=0.032, HR=1.54, 95% CI= [1.04-2.29],) while for *OGG1*_Ser326Cys, patients homozygous for cysteine had worse DFS (p=0.025, HR=1.81, 95% CI: [1.2-3.72]).

In the validation set, only the genotypes for the *ERCC5*_His46His polymorphism but not *OGG1*_Ser326Cys, were available for analysis. In this set too, patients homozygous for the T allele had a worse DFS with ~81% increased hazard of event when compared to patients homozygous for the C allele (p=0.018, HR=1.805, 95% CI: [1.107-2.943]). The functional consequences of this polymorphism are not yet known. One possibility is that the true correlation could be due to another polymorphism in LD with *ERCC5*_His46His (**Figure 18**). ERCC5 is a DNA repair protein and the *ERCC5*_His46His polymorphism has been reported to be associated with reduced risk of developing lung cancer in individuals homozygous for the variant allele (TT) in a Norwegian case-control study (219). In other studies, LOH at 13q33 which encompasses the *ERCC5* gene is observed in prostate cancer, head and neck cancer and ovarian cancer cells (220-223). However, LOH of the *ERCC5* gene is less frequently observed in colon cancer cells when compared to

Figure 18. LD block of *ERCC5*_His46His (rs1047768)



Only the beginning of the LD block is shown due to space limitations. Location of *ERCC5*_His46His polymorphism (rs1047768) in the block is circled.

other cancers (224). LOH of *ERCC5*, as well as its down regulation were associated with a favorable PFS in ovarian cancer patients treated with platinum-based chemotherapy, presumably due to increased efficacy of the drugs (225). However, the role of *ERCC5* and the *ERCC5*_His46His polymorphism in recurrence or metastasis in colorectal cancer patients is yet to be investigated. Therefore this polymorphism or other genetic variations closely linked to it are interesting candidates as disease-progression markers in colorectal cancer and further studies are warranted.

5.8 Absence of correlations of 22 polymorphisms in the discovery set

In the discovery set, only four out of the 27 chosen polymorphisms were correlated with OS. Thus correlations of 22 polymorphisms (*PTGS2*_3618A/G in 3'-UTR was excluded from analysis due to its low mAF) with survival were not detected. All 27 polymorphisms were reported to be correlated with survival in at least one study in the literature (**section 1.7**) which was the primary reason for selection of these polymorphisms for inclusion in this project. It is likely that the absence of correlations of these 21 polymorphisms (*PTGS2*_c.3618A/G excluded) in our study is due to differences in cohort characteristics between our study and previous studies, a situation commonly observed in literature (181,182). These differences between the cohorts may be in terms of ethnicity, treatment characteristics, variable follow-up times and variable clinical characteristics. The discovery cohort is one of the largest colorectal cancer cohorts in which such a study has been performed. This cohort is predominantly composed of early stage Caucasian patients followed up to over 10 years, a large percentage of which were treated with 5-FU-based chemotherapy. These characteristics may not be shared by other cohorts and we suggest that this may be a reason why these 22 polymorphisms were not correlated in this cohort.

Our study has certain drawbacks. Firstly, the validation cohort has less than half the number of patients compared to discovery cohort. Secondly, the discovery and validation cohorts have dissimilarities in terms of patient and tumor characteristics. Thirdly, the discovery cohort is biased toward early-stage patients relative to the validation cohort. These differences between the two cohorts may have limited the validation of associations observed in the discovery cohort.

Genetic prognostic research is an emerging field and it currently faces certain challenges. Multiple studies performed on the same genetic marker may not always give the same results due to differences in cohort characteristics, treatment characteristics, study design and statistical methods used. Hence larger studies, including meta-analysis or large prospective studies may be necessary to establish the prognostic relevance of genetic markers.

5.9 Conclusion

This is the first study in NL and one of the few studies in Canada to investigate the potential for using inherited variants as prognostic markers in colorectal cancer. It is also one of the few studies in the world that attempts to validate the results obtained in an additional patient cohort in colorectal cancer. We suggest that larger studies on the *MTHFR*_Glu429Ala and *ERCC5*_His46His polymorphisms, as well as other variants in linkage disequlibrium with these polymorphisms, should be performed. In the case of *MTHFR*_Glu429Ala, sex-specific functional studies are also warranted. Eventually these studies may help to better predict the outcome of patients and to enable personalized treatment based on a patient's genetic profile.

References

(1) National Cancer Institute, US National Institutes of Health. What is Cancer? 2010; Available at: http://www.cancer.gov/cancertopics/cancerlibrary/what-is-cancer. Accessed 07/29, 2010.

(2) Hanahan D, Weinberg R. Hallmarks of Cancer: The Next Generation. Cell 2011 3/4;144(5):646-674.

(3) Waugh A. GA. The digestive system. 9th ed. United Kingdom: Churchill Livingstone (Elsevier Science); 2001.

(4) Tortora GJ DB. Principles of Anatomy and Physiology. 11th Edition ed. United States of America: John Wiley & Sons, Inc; 2006.

(5) National Cancer Institute, US National Institutes of Health. Colon and Rectal Cancer, National Cancer Institute, US National Institutes of Health. 2010; Available at: http://www.cancer.gov/cancertopics/types/colon-and-rectal. Accessed June/25, 2010.

(6) Markowitz SD, Bertagnolli MM. Molecular Basis of Colorectal Cancer. N Engl J Med 2009 December 17;361(25):2449-2460.

(7) Barber TD, McManus K, Yuen KWY, Reis M, Parmigiani G, Shen D, et al. Chromatid cohesion defects may underlie chromosome instability in human colorectal cancers. Proceedings of the National Academy of Sciences 2008 March 04;105(9):3443-3448.

(8) Benito M DE. Molecular biology in colorectal cancer. Clin Transl Oncol 2006;8(6):391-8.

(9) Söreide K, Janssen EA, Söiland H, Körner H, Baak JP. Microsatellite instability in colorectal cancer. Br J Surg 2006 April;93(4):395-406.

(10) M.A.Hayat. Introduction: Colorectal Cancer. Methods of Cancer Diagnosis, Therapy and Prognosis: Springer Netherlands; 2009. p. 3-4,5,6,7,8,9.

(11) Castells A, Castellví–Bel S, Balaguer F. Concepts in Familial Colorectal Cancer: Where Do We Stand and What Is the Future? Gastroenterology 2009 8;137(2):404-409.

(12) Kitisin K, Mishra L. Molecular Biology of Colorectal Cancer: New Targets. Semin Oncol 2006 12;33(Supplement 11):14-23.

(13) Half E, Bercovich D, Rozen P. Familial adenomatous polyposis. Orphanet J Rare

Dis 2009 Oct;4(22):1-23.

(14) Lindor NM. Familial Colorectal Cancer Type X: The Other Half of Hereditary Nonpolyposis Colon Cancer Syndrome. Surg Oncol Clin N Am 2009 10;18(4):637-645.

(15) Lindor NM, Rabe K, Petersen GM, Haile R, Casey G, Baron J, et al. Lower Cancer Incidence in Amsterdam-I Criteria Families Without Mismatch Repair Deficiency: Familial Colorectal Cancer Type X. JAMA 2005 April 27;293(16):1979-1985.

(16) Francisco I, Albuquerque C, Lage P, Belo H, Vitoriano I, Filipe B, Claro I, Ferreira S, Rodrigues P, Chaves P, Leitão CN, Pereira AD. Familial colorectal cancer type X syndrome: two distinct molecular entities? Fam Cancer 2011.

(17) Peters U et al. Meta-analysis of new genome-wide association studies of colorectal cancer risk. Hum Genet 2011.

(18) Pino MS CD. The Chromosomal Instability Pathway in Colon Cancer. Gastroenterology 2010 Jun;138(6):2059-2072.

(19) Iacopetta B, Grieu F, Amanuel B. Microsatellite instability in colorectal cancer. Asia Pac J Clin 2010;6(4):260--269.

(20) Grady WM, Carethers JM. Genomic and Epigenetic Instability in Colorectal Cancer Pathogenesis. Gastroenterology 2008 10;135(4):1079-1099.

(21) Jass JR SL. Histological typing of intestinal tumors. . 2nd ed. Berlin-New York: Springer-Verlag; 1989.

(22) Labianca R, Beretta GD, Kildani B, Milesi L, Merlin F, Mosconi S, et al. Colon cancer. Crit Rev Oncol 2010 5;74(2):106-133.

(23) Treanor D, Quirke P. Pathology of Colorectal Cancer. Clin Oncol 2007 12;19(10):769-776.

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

(25) World Health Organization. The global burden of disease-2004 Update. 2008:1--160.

(26) Center MM, Jemal A, Smith RA, Ward E. Worldwide variations in colorectal cancer. CA Cancer J Clin 2009;59(6):366-378.

(27) Canadian Cancer Society's Steering Committee on Cancer Statistics. Canadian

Cancer Statistics 2011. May 2011.

(28) National Cancer Institute, US National Institutes of Health. Dictionary of Cancer Terms. 2010; Available at: http://www.cancer.gov/dictionary/?CdrID=45849. Accessed 7/7, 2010.

(29) National Center for Biotechnology Information, National Library of Medicine. National Library of Medicine, National Center for Biotechnology Information. 2010; Available at: http://www.ncbi.nlm.nih.gov/bookshelf/br.fcgi?book=hstechrev&part=A29976. Accessed June, 2010.

(30) Edge SB, Byrd DR, Compton CC, Fritz AG, Greene FL, Trotti A. AJCC (American Joint Committee on Cancer) Cancer Staging Handbook, 7th Edition. . 7th ed. New York: Springer; 2010. p. 192--206.

(31) Compton CC, Fielding LP, Burgart LJ, Conley B, Cooper HS, Hamilton SR, et al. Prognostic Factors in Colorectal Cancer. Arch Pathol Lab Med 2000 06/01;124(7):979-994.

(32) 2010 Canadian Cancer Society. Staging and Grading for colorectal cancer. 2009; Available at: http://www.cancer.ca/canada-wide/about%20cancer/types%20of%20cancer/staging%20and%20grading%20for%20colorectal%20cancer.aspx?sc_lang=en. Accessed June/25, 2010.

(33) Popat S, Hubner R, Houlston RS. Systematic Review of Microsatellite Instability and Colorectal Cancer Prognosis. Journal of Clinical Oncology 2005 January 20;23(3):609-618.

(34) Li FY LM. Colorectal cancer, one entity or three. J Zhejiang Univ Sci B 2009;10(3):219-229.

(35) Green RC, Green JS, Buehler SK, Robb JD, Daftary D, Gallinger S, McLaughlin JR, Parfrey PS, Younghusband HB. 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.

(36) Davies H, Bignell GR, Cox C, Stephens P, Edkins S, Clegg S, Teague J, Woffendin H, Garnett MJ, Bottomley W, Davis N, Dicks E, Ewing R, Floyd Y, Gray K, Hall S, Hawes R, Hughes J, Kosmidou V, Menzies A, Mould C, Parker A, Stevens C, Watt S, Hooper S, Wilson R, Jayatilake H, Gusterson BA, Cooper C, Shipley J, Hargrave D, Pritchard-Jones K, Maitland N, Chenevix-Trench G, Riggins GJ, Bigner DD, Palmieri G, Cossu A, Flanagan A, Nicholson A, Ho JW, Leung SY, Yuen ST, Weber BL, Seigler HF, Darrow TL, Paterson H, Marais R, Marshall CJ, Wooster R, Stratton MR, Futreal PA.

Mutations of the BRAF gene in human cancers. Nature 2002;417(6892):949--954.

(37) Fariña-Sarasqueta A, van Lijnschoten G, Moerland E, Creemers G-, Lemmens VEPP, Rutten HJT, et al. The BRAF V600E mutation is an independent prognostic factor for survival in stage II and stage III colon cancer patients. Annals of Oncology 2010 December 01;21(12):2396-2402.

(38) Saridaki Z, Papadatos-Pastos D, Tzardi M, Mavroudis D, Bairaktari E, Arvanity H, Stathopoulos E, Georgoulias V, Souglakos J. BRAF mutations, microsatellite instability status and cyclin D1 expression predict metastatic colorectal patient's outcome. Br J Cancer 2010;102(12):1762--1768.

(39) Samowitz WS, Sweeney C, Herrick J, Albertsen H, Levin TR, Murtaugh MA, et al. Poor survival associated with the BRAF V600E mutation in microsatellite-stable colon cancers. Cancer Research 2005 July 15;65(14):6063-6069.

(40) Wish TA, Hyde AJ, Parfrey PS, Green JS, Younghusband HB, Simms MI, et al. Increased Cancer Predisposition in Family Members of Colorectal Cancer Patients Harboring the p.V600E BRAF Mutation: a Population-Based Study. Cancer Epidemiology Biomarkers & Prevention 2010 July 01;19(7):1831-1839.

(41) Kidd JM, Cooper GM, Donahue WF, Hayden HS, Sampas N, Graves T, Hansen N, Teague B, Alkan C, Antonacci F, Haugen E, Zerr T, Yamada NA, Tsang P, Newman TL, Tüzün E, Cheng Z, Ebling HM, Tusneem N, David R, Gillett W, Phelps KA, Weaver M, Saranga D, Brand A, Tao W, Gustafson E, McKernan K, Chen L, Malig M, Smith JD, Korn JM, McCarroll SA, Altshuler DA, Peiffer DA, Dorschner M, Stamatoyannopoulos J, Schwartz D, Nickerson DA, Mullikin JC, Wilson RK, Bruhn L, Olson MV, Kaul R, Smith DR, Eichler EE. Mapping and sequencing of structural variation from eight human genomes. Nature 2008;453(7191):56--64.

(42) Strachan T RA. Instability of the human genome: mutation and DNA repair. Human Molecular Genetics. 2nd ed. USA and Canada: Wiley-Liss; 1999. p. 209--240.

(43) Miller RD, Phillips MS, Jo I, Donaldson MA, Studebaker JF, Addleman N, et al. High-density single-nucleotide polymorphism maps of the human genome. Genomics 2005 8;86(2):117-126.

(44) Coate L, Cuffe S, Horgan A, Hung RJ, Christiani D, Liu G. Germline Genetic Variation, Cancer Outcome, and Pharmacogenetics. Journal of Clinical Oncology 2010 September 10;28(26):4029-4037.

(45) Webber EM, Lin JS, Evelyn P Whitlock. Oncotype DX tumor gene expression profiling in stage II colon cancer. Application: prognostic, risk prediction. PLoS curr 2010;2(RRN1177).

(46) Salazar R, Roepman P, Capella G, Moreno V, Simon I, Dreezen C, et al. Gene Expression Signature to Improve Prognosis Prediction of Stage II and III Colorectal Cancer. Journal of Clinical Oncology 2011 January 01;29(1):17-24.

(47) Savas S, Younghusband HB. dbCPCO: a database of genetic markers tested for their predictive and prognostic value in colorectal cancer. Hum Mutat 2010;31(8):901--7.

(48) US National Library of Medicine. CCND1 cyclin D1 [Homo sapiens]. 2011; Available at: http://www.ncbi.nlm.nih.gov/gene/595, 2011.

(49) Betticher DC, Thatcher N, Altermatt HJ, Hoban P, Ryder WD, Heighway J. Alternate splicing produces a novel cyclin D1 transcript. Oncogene 1995;11(5):1005.

(50) Hong Y, Eu KW, Seow-Choen F, Fook-Chong S, Cheah PY. GG genotype of cyclin D1 G870A polymorphism is associated with increased risk and advanced colorectal cancer in patients in Singapore. Eur J Cancer 2005 5;41(7):1037-1044.

(51) Zhang W, Gordon M, Press OA, Rhodes K, Vallböhmer D, Yang DY, Park D, Fazzone W, Schultheis A, Sherrod AE, Iqbal S, Groshen S, Lenz HJ. Cyclin D1 and epidermal growth factor polymorphisms associated with survival in patients with advanced colorectal cancer treated with Cetuximab. Phar 2006;16(7):475--83.

(52) Graziano F, Ruzzo A, Loupakis F, Canestrari E, Santini D, Catalano V, et al. Pharmacogenetic Profiling for Cetuximab Plus Irinotecan Therapy in Patients With Refractory Advanced Colorectal Cancer. Journal of Clinical Oncology 2008 March 20;26(9):1427-1434.

(53) Ho-Pun-Cheung A, Assenat E, Thezenas S, Bibeau F, Rouanet P, Azria D, et al. Cyclin D1 Gene G870A Polymorphism Predicts Response to Neoadjuvant Radiotherapy and Prognosis in Rectal Cancer. International Journal of Radiation Oncology*Biology*Physics 2007 7/15;68(4):1094-1101.

(54) McKay JA, Douglas JJ, Ross VG, Curran S, Murray GI, Cassidy J, McLeod HL. Cyclin D1 protein expression and gene polymorphism in colorectal cancer. Aberdeen Colorectal Initiative. Int J Cancer 2000;88(1):77--81.

(55) Yoshiya G, Takahata T, Hanada N, Suzuki K, Ishiguro A, Saito M, Sasaki M, Fukuda S. Influence of cancer-related gene polymorphisms on clinicopathological features in colorectal cancer. J Gastroenterol Hepatol 2008;23(6):948--53.

(56) US National Library of Medicine. DCC deleted in colorectal carcinoma [Homo sapiens]. 2011; Available at: http://www.ncbi.nlm.nih.gov/gene/1630, 2011.

(57) Schmitt CA, Thaler KR, Wittig BM, Kaulen H, Meyer zum Büschenfelde KH,

Dippold WG. Detection of the DCC gene product in normal and malignant colorectal tissues and its relation to a codon 201 mutation. Br J Cancer 1998;77(4):588-594.

(58) Zhang H, Arbman G, Sun X. Codon 201 polymorphism of DCC gene is a prognostic factor in patients with colorectal cancer. Cancer Detect Prev 2003;27(3):216-221.

(59) US National Library of Medicine. EGFR epidermal growth factor receptor [Homo sapiens]. 2011; Available at: http://www.ncbi.nlm.nih.gov/gene/1956, 2011.

(60) Moriai T, Kobrin MS, Hope C, Speck L, Korc M. A variant epidermal growth factor receptor exhibits altered type alpha transforming growth factor binding and transmembrane signaling. Proceedings of the National Academy of Sciences 1994 October 11;91(21):10217-10221.

(61) Goncalves A, Esteyries S, Taylor-Smedra B, Lagarde A, Ayadi M, Monges G, et al. A polymorphism of EGFR extracellular domain is associated with progression free-survival in metastatic colorectal cancer patients receiving cetuximab-based treatment. BMC Cancer 2008;8(1):169.

(62) Press OA, Zhang W, Gordon MA, Yang D, Lurje G, Iqbal S, et al. Gender-Related Survival Differences Associated with EGFR Polymorphisms in Metastatic Colon Cancer. Cancer Research 2008 April 15;68(8):3037-3042.

(63) Wang W, Chen P, Chiou T, Liu J, Lin J, Lin T, et al. Epidermal Growth Factor Receptor R497K Polymorphism Is a Favorable Prognostic Factor for Patients with Colorectal Carcinoma. Clinical Cancer Research 2007 June 15;13(12):3597-3604.

(64) Lurje G, Zhang W, Schultheis AM, Yang D, Groshen S, Hendifar AE, et al. Polymorphisms in VEGF and IL-8 predict tumor recurrence in stage III colon cancer. Annals of Oncology 2008 October 01;19(10):1734-1741.

(65) Zhang W, Stoehlmacher J, Park DJ, Yang D, Borchard E, Gil J, Tsao-Wei DD, Yun J, Gordon M, Press OA, Rhodes K, Groshen S, Lenz HJ. Gene polymorphisms of epidermal growth factor receptor and its downstream effector, interleukin-8, predict oxaliplatin efficacy in patients with advanced colorectal cancer. Clin Colorectal Cancer 2005;5(2):124--131.

(66) Zhang W, Azuma M, Lurje G, Gordon MA, Yang D, Pohl A, Ning Y, Bohanes P, Gerger A, Winder T, Hollywood E, Danenberg KD, Saltz L, Lenz HJ. Molecular predictors of combination targeted therapies (cetuximab, bevacizumab) in irinotecan-refractory colorectal cancer (BOND-2 study). Anticancer Res 2010;30(10):4209--4217.

(67) US National Library of Medicine. ERCC1 excision repair cross-complementing rodent repair deficiency, complementation group 1 (includes overlapping antisense

sequence) [Homo sapiens]. 2011; Available at: http://www.ncbi.nlm.nih.gov/gene/2067, 2011.

(68) Yu JJ, Mu C, Lee KB, Okamoto A, Reed EL, Bostick-Bruton F, Mitchell KC, Reed E.
A nucleotide polymorphism in ERCC1 in human ovarian cancer cell lines and tumor tissues. Mutat Res 1997;382(1-2):13-20.

(69) Chang PM, Tzeng CH, Chen PM, Lin JK, Lin TC, Chen WS, Jiang JK, Wang HS, Wang WS. ERCC1 codon 118 C→T polymorphism associated with ERCC1 expression and outcome of FOLFOX-4 treatment in Asian patients with metastatic colorectal carcinoma. Cancer Sci 2009;100(2):278--83.

(70) Liang J, Jiang T, Yao RY, Liu ZM, Lv HY, Qi WW. The combination of ERCC1 and XRCC1 gene polymorphisms better predicts clinical outcome to oxaliplatin-based chemotherapy in metastatic colorectal cancer. Cancer Chemother Pharmacol 2010;66(3):493--500.

(71) Huang MY, Huang ML, Chen MJ, Lu CY, Chen CF, Tsai PC, Chuang SC, Hou MF, Lin SR, Wang JY. Multiple genetic polymorphisms in the prediction of clinical outcome of metastatic colorectal cancer patients treated with first-line FOLFOX-4 chemotherapy. Pharmacogenet Genomics 2011;21(1):18--25.

(72) Ruzzo A, Graziano F, Loupakis F, Rulli E, Canestrari E, Santini D, et al. Pharmacogenetic Profiling in Patients With Advanced Colorectal Cancer Treated With First-Line FOLFOX-4 Chemotherapy. Journal of Clinical Oncology 2007 April 01;25(10):1247-1254.

(73) Park DJ, Zhang W, Stoehlmacher J, Tsao-Wei D, Groshen S, Gil J, Yun J, Sones E, Mallik N, Lenz HJ. ERCC1 gene polymorphism as a predictor for clinical outcome in advanced colorectal cancer patients treated with platinum-based chemotherapy. Clin Adv Hematol 2003;1(3).

(74) Stoehlmacher J, Park DJ, Zhang W, Yang D, Groshen S, Zahedy S, Lenz HJ. A multivariate analysis of genomic polymorphisms: prediction of clinical outcome to 5-FU/oxaliplatin combination chemotherapy in refractory colorectal cancer. Br J Cancer 2004;91(2):344--354.

(75) Moreno V, Gemignani F, Landi S, Gioia-Patricola L, Chabrier A, Blanco I, et al. Polymorphisms in Genes of Nucleotide and Base Excision Repair: Risk and Prognosis of Colorectal Cancer. Clinical Cancer Research 2006 April 01;12(7):2101-2108.

(76) Etienne-Grimaldi MC, Milano G, Maindrault-Goebel F, Chibaudel B, Formento JL, Francoual M, Lledo G, André T, Mabro M, Mineur L, Flesch M, Carola E, de Gramont

A. Methylenetetrahydrofolate reductase (MTHFR) gene polymorphisms and FOLFOX response in colorectal cancer patients. Br J Clin Pharmacol 2010;69(1):58-66.

(77) McLeod HL, Sargent DJ, Marsh S, Green EM, King CR, Fuchs CS, et al. Pharmacogenetic Predictors of Adverse Events and Response to Chemotherapy in Metastatic Colorectal Cancer: Results From North American Gastrointestinal Intergroup Trial N9741. Journal of Clinical Oncology 2010 July 10;28(20):3227-3233.

(78) Boige V, Mendiboure J, Pignon J, Loriot M, Castaing M, Barrois M, et al. Pharmacogenetic assessment of toxicity and outcome in patients with metastatic colorectal cancer treated with LV5FU2, FOLFOX, and FOLFIRI: FFCD 2000-05. Journal of Clinical Oncology 2010 May 20;28(15):2556-2564.

(79) US National Library of Medicine. ERCC2 excision repair cross-complementing rodent repair deficiency, complementation group 2 [Homo sapiens]. 2011; Available at: http://www.ncbi.nlm.nih.gov/gene/2068, 2011.

(80) Lunn RM, Helzlsouer KJ, Parshad R, Umbach DM, Harris EL, Sanford KK, et al. XPD polymorphisms: effects on DNA repair proficiency. Carcinogenesis 2000 April 01;21(4):551-555.

(81) Park DJ, Stoehlmacher J, Zhang W, Tsao-Wei DD, Groshen S, Lenz H. A Xeroderma Pigmentosum Group D Gene Polymorphism Predicts Clinical Outcome to Platinum-based Chemotherapy in Patients with Advanced Colorectal Cancer. Cancer Research 2001 December 15;61(24):8654-8658.

(82) Lai JI, Tzeng CH, Chen PM, Lin JK, Lin TC, Chen WS, Jiang JK, Wang HS, Wang WS. Very low prevalence of XPD K751Q polymorphism and its association with XPD expression and outcomes of FOLFOX-4 treatment in Asian patients with colorectal carcinoma. Cancer Sci 2009;100(7):1261--1266.

(83) Artac M, Bozcuk H, Pehlivan S, Akcan S, Pehlivan M, Sever T, Ozdogan M, Savas B. The value of XPD and XRCC1 genotype polymorphisms to predict clinical outcome in metastatic colorectal carcinoma patients with irinotecan-based regimens. J Cancer Res 2010;136(6):803--809.

(84) Monzo M, Moreno I, Navarro A, Ibeas R, Artells R, Gel B, Martinez F, Moreno J, Hernandez R, Navarro-Vigo M. Source. Single nucleotide polymorphisms in nucleotide excision repair genes XPA, XPD, XPG and ERCC1 in advanced colorectal cancer patients treated with first-line oxaliplatin/fluoropyrimidine. Oncology 2007;72(5-6):364--370.

(85) Lamas MJ, Duran G, Balboa E, Bernardez B, Touris M, Vidal Y, Gallardo E, Lopez R, Carracedo A, Barros F. Use of a comprehensive panel of biomarkers to predict

response to a fluorouracil-oxaliplatin regimen in patients with metastatic colorectal cancer. Pharmacogenomics 2011;12(3):433--442.

(86) US National Library of Medicine. ERCC5 excision repair cross-complementing rodent repair deficiency, complementation group 5 [Homo sapiens]. 2011; Available at: http://www.ncbi.nlm.nih.gov/gene/2073, 2011.

(87) Kweekel DM, Antonini NF, Nortier JW, Punt CJ, Gelderblom H, Guchelaar HJ. Explorative study to identify novel candidate genes related to oxaliplatin efficacy and toxicity using a DNA repair array. Br J Cancer 2009;101(2):357-362.

(88) US National Library of Medicine. EXO1 exonuclease 1 [Homo sapiens]. 2011; Available at: http://www.ncbi.nlm.nih.gov/gene/9156, 2011.

(89) US National Library of Medicine. FAS Fas (TNF receptor superfamily, member 6) [Homo sapiens]. 2011; Available at: http://www.ncbi.nlm.nih.gov/gene/355, 2011.

(90) Hofmann G, Langsenlehner U, Langsenlehner T, Yazdani-Biuki B, Clar H, Gerger A, Fuerst F, Samonigg H, Krippl P, Renner W. A common hereditary single-nucleotide polymorphism in the gene of FAS and colorectal cancer survival. J Cel Mol Med 2009;13(9B):3699--702.

(91) US National Library of Medicine. FGFR4 fibroblast growth factor receptor 4 [Homo sapiens]. 2011; Available at: http://www.ncbi.nlm.nih.gov/gene/2264, 2011.

(92) Bange J, Prechtl D, Cheburkin Y, Specht K, Harbeck N, Schmitt M, et al. Cancer progression and tumor cell motility are associated with the FGFR4 Arg388 allele. Cancer Research 2002 February 01;62(3):840-847.

(93) Spinola M, Leoni VP, Tanuma J, Pettinicchio A, Frattini M, Signoroni S, Agresti R, Giovanazzi R, Pilotti S, Bertario L, Ravagnani F, Dragani TA. FGFR4 Gly388Arg polymorphism and prognosis of breast and colorectal cancer. Oncol Rep 2005;14(2):415--419.

(94) US National Library of Medicine. GSTM1 glutathione S-transferase mu 1 [Homo sapiens]. 2011; Available at: http://www.ncbi.nlm.nih.gov/gene/2944, 2011.

(95) Csejtei A, Tibold A, Varga Z, Koltai K, Ember A, Orsos, Zsuzsa, Feher, Gergely, et al. GSTM, GSTT and p53 Polymorphisms as Modifiers of Clinical Outcome in Colorectal Cancer. Anticancer Research May-June 2008 May-June 2008;28(3B):1917-1922.

(96) Holley SL, Rajagopal R, Hoban PR, Deakin M, Fawole AS, Elder JB, Elder J, Smith V, Strange RC, Fryer AA. Polymorphisms in the glutathione S-transferase mu cluster are

associated with tumor progression and patient outcome in colorectal cancer. Int J Oncol 2006;28(1):231--236.

(97) Funke S, Timofeeva M, Risch A, Hoffmeister M, Stegmaier C, Seiler CM, Brenner H, Chang-Claude J. Genetic polymorphisms in GST genes and survival of colorectal cancer patients treated with chemotherapy. Pharmacogenomics 2010;11(1):33--41.

(98) US National Library of Medicine. GSTP1 glutathione S-transferase pi 1 [Homo sapiens]. 2011; Available at: http://www.ncbi.nlm.nih.gov/gene/2950, 2011.

(99) Watson MA, Stewart RK, Smith GB, Massey TE, Bell DA. Human glutathione S-transferase P1 polymorphisms: relationship to lung tissue enzyme activity and population frequency distribution. Carcinogenesis 1998 February 01;19(2):275-280.

(100) Kweekel DM, Koopman M, Antonini NF, Van der Straaten T, Nortier JW, Gelderblom H, Punt CJ, Guchelaar HJ. GSTP1 Ile105Val polymorphism correlates with progression-free survival in MCRC patients treated with or without irinotecan: a study of the Dutch Colorectal Cancer Group. Br J Cancer 2008;99(8):1316--1321.

(101) Stoehlmacher J, Park DJ, Zhang W, Groshen S, Tsao-Wei DD, Yu MC, et al. Association Between Glutathione S-Transferase P1, T1, and M1 Genetic Polymorphism and Survival of Patients With Metastatic Colorectal Cancer. Journal of the National Cancer Institute 2002 June 19;94(12):936-942.

(102) Zarate R, Rodríguez J, Bandres E, Patiño-Garcia A, Ponz-Sarvise M, Viudez A, Ramirez N, Bitarte N, Chopitea A, Gacía-Foncillas J. Oxaliplatin, irinotecan and capecitabine as first-line therapy in metastatic colorectal cancer (mCRC): a dose-finding study and pharmacogenomic analysis. Br J Cancer 2010;102(6):987-994.

(103) Jun L, Haiping Z, Beibei Y. Genetic polymorphisms of GSTP1 related to response to 5-FU-oxaliplatin-based chemotherapy and clinical outcome in advanced colorectal cancer patients. Swiss Med Wkly 2009;139(49-50):724--728.

(104) Chen YC, Tzeng CH, Chen PM, Lin JK, Lin TC, Chen WS, Jiang JK, Wang HS, Wang WS. Influence of GSTP1 I105V polymorphism on cumulative neuropathy and outcome of FOLFOX-4 treatment in Asian patients with colorectal carcinoma. Cancer 2010;101(2):530--535.

(105) Sun XF, Ahmadi A, Arbman G, Wallin A, Asklid D, Zhang H. Polymorphisms in sulfotransferase 1A1 and glutathione S-transferase P1 genes in relation to colorectal cancer risk and patients' survival. World J Gastroenterol 2005;11(43):6875--6879.

(106) Hong J, Han SW, Ham HS, Kim TY, Choi IS, Kim BS, Oh DY, Im SA, Kang GH, Bang YJ, Kim TY. Phase II study of biweekly S-1 and oxaliplatin combination

chemotherapy in metastatic colorectal cancer and pharmacogenetic analysis. Cancer Chemother Pharmacol 2011;67(6):1323--1331.

(107) Kweekel DM, Gelderblom H, Antonini NF, Van der Straaten T, Nortier JWR, Punt CJA, et al. Glutathione-S-transferase pi (GSTP1) codon 105 polymorphism is not associated with oxaliplatin efficacy or toxicity in advanced colorectal cancer patients. Eur J Cancer 2009 3;45(4):572-578.

(108) US National Library of Medicine. GSTT1 glutathione S-transferase theta 1 [Homo sapiens]. 2011; Available at: http://www.ncbi.nlm.nih.gov/gene/2952, 2011.

(109) Rajagopal R, Deakin M, Fawole AS, Elder JB, Elder J, Smith V, et al. Glutathione S-transferase T1 polymorphisms are associated with outcome in colorectal cancer. Carcinogenesis December 2005 December 2005;26(12):2157-2163.

(110) US National Library of Medicine. IL6 interleukin 6 (interferon, beta 2) [Homo sapiens]. 2011; Available at: http://www.ncbi.nlm.nih.gov/gene/3569, 2011.

(111) Fishman D, Faulds G, Jeffery R, Mohamed-Ali V, Yudkin JS, Humphries S, Woo P. The effect of novel polymorphisms in the interleukin-6 (IL-6) gene on IL-6 transcription and plasma IL-6 levels, and an association with systemic-onset juvenile chronic arthritis. J Clin Invest 1998;102(7):1369--76.

(112) Wilkening S, Tavelin B, Canzian F, Enquist K, Palmqvist R, Altieri A, et al. Interleukin promoter polymorphisms and prognosis in colorectal cancer. Carcinogenesis 2008 June 01;29(6):1202-1206.

(113) US National Library of Medicine. MLH1 mutL homolog 1, colon cancer, nonpolyposis type 2 (E. coli) [Homo sapiens]. 2011; Available at: http://www.ncbi.nlm.nih.gov/gene/4292, 2011.

(114) Nejda N, Iglesias D, Moreno Azcoita M, Medina Arana V, González-Aguilera JJ, Fernández-Peralta AM. A MLH1 polymorphism that increases cancer risk is associated with better outcome in sporadic colorectal cancer. Cancer Genet Cytogenet 2009 9;193(2):71-77.

(115) Koessler T, Azzato EM, Perkins B, Macinnis RJ, Greenberg D, Easton DF, Pharoah PD. Common germline variation in mismatch repair genes and survival after a diagnosis of colorectal cancer. Int J Cancer 2009;24(8):1887--1891.

(116) US National Library of Medicine. MMP1 matrix metallopeptidase 1 (interstitial collagenase) [Homo sapiens]. 2011; Available at: http://www.ncbi.nlm.nih.gov/gene/4312, 2011.

(117) Rutter JL, Mitchell TI, Butticè G, Meyers J, Gusella JF, Ozelius LJ, et al. A Single Nucleotide Polymorphism in the Matrix Metalloproteinase-1 Promoter Creates an Ets Binding Site and Augments Transcription. Cancer Research 1998 December 01;58(23):5321-5325.

(118) Hettiaratchi A, Hawkins NJ, McKenzie G, Ward RL, Hunt JE, Wakefield D, Di Girolamo N. The collagenase-1 (MMP-1) gene promoter polymorphism - 1607/2G is associated with favourable prognosis in patients with colorectal cancer. Br J Cancer 2007;95(5):783--92.

(119) Zinzindohoué F, Lecomte T, Ferraz JM, Houllier AM, Cugnenc PH, Berger A, Blons H, Laurent-Puig P. Prognostic significance of MMP-1 and MMP-3 functional promoter polymorphisms in colorectal cancer. Clin 2005;11(2 (pt 1)):594--599.

(120) US National Library of Medicine. MMP2 matrix metallopeptidase 2 (gelatinase A, 72kDa gelatinase, 72kDa type IV collagenase) [Homo sapiens]. 2011; Available at: http://www.ncbi.nlm.nih.gov/gene/4313, 2011.

(121) Price SJ, Greaves DR, Watkins H. Identification of Novel, Functional Genetic Variants in the Human Matrix Metalloproteinase-2 Gene. Journal of Biological Chemistry 2001 March 09;276(10):7549-7558.

(122) Langers AM, Sier CF, Hawinkels LJ, Kubben FJ, van Duijn W, van der Reijden JJ, Lamers CB, Hommes DW, Verspaget HW. MMP-2 geno-phenotype is prognostic for colorectal cancer survival, whereas MMP-9 is not. Br J Cancer 2008;98(11):1820--1823.

(123) US National Library of Medicine. MTHFR methylenetetrahydrofolate reductase (NAD(P)H) [Homo sapiens]. 2011; Available at: http://www.ncbi.nlm.nih.gov/gene/4524, 2011.

(124) Odin E, Wettergren Y, Carlsson G, Danenberg PV, Termini A, Willén R, Gustavsson B. Expression and clinical significance of methylenetetrahydrofolate reductase in patients with colorectal cancer. Clin Colorectal Cancer 2006;5(5):344-349.

(125) van der Put NM, van den Heuvel LP, Steegers-Theunissen RP, Trijbels FJ, Eskes TK, Mariman EC, den Heyer M, Blom HJ. Decreased methylene tetrahydrofolate reductase activity due to the 677C-->T mutation in families with spina bifida offspring. . J Mol Med (Ber) 1996;74(11):691--694.

(126) Derwinger K, Wettergren Y, Odin E, Carlsson G, Gustavsson B. A study of the MTHFR gene polymorphism C677T in colorectal cancer. Clin Colorectal Cancer 2009;8(1):43--48.

(127) Castillo-Fernández O, Santibáñez M, Bauza A, Calderillo G, Castro C, Herrera R,

et al. Methylenetetrahydrofolate Reductase Polymorphism (677 C>T) Predicts Long Time to Progression in Metastatic Colon Cancer Treated with 5-Fluorouracil and Folinic Acid. Arch Med Res 2010 8;41(6):430-435.

(128) Marcuello E, Altés A, Menoyo A, Rio ED, Baiget M. Methylenetetrahydrofolate reductase gene polymorphisms: genomic predictors of clinical response to fluoropyrimidine-based chemotherapy? Cancer Chemother Pharmacol 2006;57(6):835-840.

(129) Massacesi C, Terrazzino S, Marcucci F, Rocchi MB, Lippe P, Bisonni R, Lombardo M, Pilone A, Mattioli R, Leon A. Uridine diphosphate glucuronosyl transferase 1A1 promoter polymorphism predicts the risk of gastrointestinal toxicity and fatigue induced by irinotecan-based chemotherapy. Cancer 2006;106(5):1007--1016.

(130) Sharma R, Hoskins JM, Rivory LP, Zucknick M, London R, Liddle C, et al. Thymidylate synthase and methylenetetrahydrofolate reductase gene polymorphisms and toxicity to capecitabine in advanced colorectal cancer patients. Clinical Cancer Research 2008 February 01;14(3):817-825.

(131) Suh KW, Kim JH, Kim do Y, Kim YB, Lee C, Choi S. Which gene is a dominant predictor of response during FOLFOX chemotherapy for the treatment of metastatic colorectal cancer, the MTHFR or XRCC1 gene? Ann surg oncol 2006;13(11):1379--1385.

(132) Wisotzkey JD, Toman J, Bell T, Monk JS, Jones D. MTHFR (C677T) polymorphisms and stage III colon cancer: response to therapy. Mol Diagn 1999;4(2):95--99.

(133) Zhang W, Press OA, Haiman CA, Yang DY, Gordon MA, Fazzone W, et al. Association of Methylenetetrahydrofolate Reductase Gene Polymorphisms and Sex-Specific Survival in Patients With Metastatic Colon Cancer. Journal of Clinical Oncology 2007 August 20;25(24):3726-3731.

(134) Afzal S, Jensen SA, Vainer B, Vogel U, Matsen JP, Sørensen JB, et al. MTHFR polymorphisms and 5-FU-based adjuvant chemotherapy in colorectal cancer. Annals of Oncology 2009 October 01;20(10):1660-1666.

(135) Gusella M, Frigo AC, Bolzonella C, Marinelli R, Barile C, Bononi A, Crepaldi G, Menon D, Stievano L, Toso S, Pasini F, Ferrazzi E, Padrini R. Predictors of survival and toxicity in patients on adjuvant therapy with 5-fluorouracil for colorectal cancer. Br J Cancer 2009;100(10):1549-1557.

(136) van der Put NMJ, Gabreëls F, Stevens EMB, Smeitink JAM, Trijbels FJM, Eskes TKAB, et al. A second common mutation in the methylenetetrahydrofolate reductase

gene: An additional risk factor for neural-tube defects? The American Journal of Human Genetics 1998 5;62(5):1044-1051.

(137) Fernández-Peralta AM, Daimiel L, Nejda N, Iglesias D, Medina Arana V, González-Aguilera JJ. Association of polymorphisms MTHFR C677T and A1298C with risk of colorectal cancer, genetic and epigenetic characteristic of tumors, and response to chemotherapy. Int J Colorectal Dis 2010;25(2):141-151.

(138) US National Library of Medicine. OGG1 8-oxoguanine DNA glycosylase [Homo sapiens]. 2011; Available at: http://www.ncbi.nlm.nih.gov/gene/4968, 2011.

(139) Hill JW, Evans MK. Dimerization and opposite base-dependent catalytic impairment of polymorphic S326C OGG1 glycosylase. Nucleic Acids Research ;34(5):1620-1632.

(140) Kweekel DM, Antonini NF, Nortier JW, Punt CJ, Gelderblom H, Guchelaar HJ. Explorative study to identify novel candidate genes related to oxaliplatin efficacy and toxicity using a DNA repair array.
. Br J Cancer 2009;101(2):357--362.

(141) US National Library of Medicine. PTGS2 prostaglandin-endoperoxide synthase 2 (prostaglandin G/H synthase and cyclooxygenase) [Homo sapiens]. 2011; Available at: http://www.ncbi.nlm.nih.gov/gene/5743, 2011.

(142) Iglesias D, Nejda N, Azcoita MM, Schwartz S Jr, González-Aguilera JJ, Fernández-Peralta AM. Effect of COX2 -765G>C and c.3618A>G polymorphisms on the risk and survival of sporadic colorectal cancer. Cancer Causes Control 2009;20(8).

(143) US National Library of Medicine. SERPINE1 serpin peptidase inhibitor, clade E (nexin, plasminogen activator inhibitor type 1), member 1 [Homo sapiens]. 2011; Available at: http://www.ncbi.nlm.nih.gov/gene/5054, 2011.

(144) Eriksson P, Kallin B, van 't Hooft FM, Båvenholm P, Hamsten A. Allele-specific increase in basal transcription of the plasminogen-activator inhibitor 1 gene is associated with myocardial infarction. Proceedings of the National Academy of Sciences 1995 March 14;92(6):1851-1855.

(145) Försti A, Lei H, Tavelin B, Enquist K, Palmqvist R, Altieri A, et al. Polymorphisms in the genes of the urokinase plasminogen activation system in relation to colorectal cancer. Annals of Oncology 2007 December 01;18(12):1990-1994.

(146) US National Library of Medicine. TYMS thymidylate synthetase [Homo sapiens]. 2011; Available at: http://www.ncbi.nlm.nih.gov/gene/7298, 2011.

(147) Kaneda S, Takeishi K, Ayusawa D, Shimizu K, Seno T, Altman S. Role in translation of a triple tandemly repeated sequence in the 5'-untranslated region of human thymidylate synthase mRNA. Nucleic Acids Research 1987 February 11;15(3):1259-1270.

(148) Hitre E, Budai B, Adleff V, Czeglédi F, Horváth Z, Gyergyay F, Lövey J, Kovács T, Orosz Z, Láng I, Kásler M, Kralovánszky J.
Influence of thymidylate synthase gene polymorphisms on the survival of colorectal cancer patients receiving adjuvant 5-fluorouracil. Pharmacogenet Genomics 2005;15(10):723--730.

(149) Underhill C, Goldstein D, Gorbounova VA, Biakhov MY, Bazin IS, Granov DA, Hossain AM, Blatter J, Kaiser C, Ma D. A randomized phase II trial of pemetrexed plus irinotecan (ALIRI) versus leucovorin-modulated 5-FU plus irinotecan (FOLFIRI) in first-line treatment of locally advanced ormetastatic colorectal cancer. Oncology 2007;73(1-2):9--20.

(150) Fariña-Sarasqueta A, Gosens MJ, Moerland E, van Lijnschoten I, Lemmens VE, Slooter GD, Rutten HJ, van den Brule AJ. TS gene polymorphisms are not good markers of response to 5-FU therapy in stage III colon cancer patients. Cell Oncol (Dordr) 2011.

(151) Páez D, Paré L, Altés A, Sancho-Poch FJ, Petriz L, Garriga J, Monill JM, Salazar J, del Rio E, Barnadas A, Marcuello E, Baiget M. Thymidylate synthase germline polymorphisms in rectal cancer patients treated with neoadjuvant chemoradiotherapy based on 5-fluorouracil. J Cancer Res Clin Oncol 2010;136(11):1681--1689.

(152) Suh KW, Kim JH, Kim YB, Kim J, Jeong S. Thymidylate Synthase Gene Polymorphism as a Prognostic Factor for Colon Cancer. Journal of Gastrointestinal Surgery 2005 3/1;9(3):336-342.

(153) Chen J, Hunter DJ, Stampfer MJ, Kyte C, Chan W, Wetmur JG, et al. Polymorphism in the Thymidylate Synthase Promoter Enhancer Region Modifies the Risk and Survival of Colorectal Cancer. Cancer Epidemiology Biomarkers & Prevention 2003 October 01;12(10):958-962.

(154) Curtin K, Ulrich CM, Samowitz WS, Bigler J, Caan B, Potter JD, Slattery ML. Thymidylate synthase polymorphisms and colon cancer: associations with tumor stage, tumor characteristics and survival. Int J Cancer 2007;120(10):2226-2232.

(155) Dotor E, Cuatrecases M, Martínez-Iniesta M, Navarro M, Vilardell F, Guinó E, et al. Tumor Thymidylate Synthase 1494del6 Genotype As a Prognostic Factor in Colorectal Cancer Patients Receiving Fluorouracil-Based Adjuvant Treatment. Journal of Clinical Oncology 2006 April 01;24(10):1603-1611.

(156) Fernández-Contreras ME, Sánchez-Prudencio S, Sánchez-Hernández JJ, García de Paredes ML, Gisbert JP, Roda-Navarro P, Gamallo C. Thymidylate synthase expression pattern, expression level and single nucleotide polymorphism are predictors for disease-free survival in patients of colorectal cancer treated with 5-fluorouracil. Int J Oncol 2006;28(5):1303--1310.

(157) Lecomte T, Ferraz J, Zinzindohoué F, Loriot M, Tregouet D, Landi B, et al. Thymidylate Synthase Gene Polymorphism Predicts Toxicity in Colorectal Cancer Patients Receiving 5-Fluorouracil-based Chemotherapy. Clinical Cancer Research 2004 September 01;10(17):5880-5888.

(158) Matsui T, Omura K, Kawakami K, Morita S, Sakamoto J. Genotype of thymidylate synthase likely to affect efficacy of adjuvant 5-FU based chemotherapy in colon cancer. Oncol Rep 2006;16(5):1111--1115.

(159) Morganti M, Ciantelli M, Giglioni B, Putignano AL, Nobili S, Papi L, et al. Relationships between promoter polymorphisms in the thymidylate synthase gene and mRNA levels in colorectal cancers. Eur J Cancer 2005 9;41(14):2176-2183.

(160) Pullarkat ST, Stoehlmacher J, Ghaderi V, Xiong YP, Ingles SA, Sherrod A, Warren R, Tsao-Wei D, Groshen S, Lenz HJ. Thymidylate synthase gene polymorphism determines response and toxicity of 5-FU chemotherapy. Pharmacogenomics J 2001;1(1):65--70.

(161) Villafranca E, Okruzhnov Y, Dominguez MA, García-Foncillas J, Azinovic I, Martínez E, et al. Polymorphisms of the Repeated Sequences in the Enhancer Region of the Thymidylate Synthase Gene Promoter May Predict Downstaging After Preoperative Chemoradiation in Rectal Cancer. Journal of Clinical Oncology 2001 March 15;19(6):1779-1786.

(162) Fernández-Contreras ME, Sánchez-Hernández JJ, González E, Herráez B, Domínguez I, Lozano M, García De Paredes ML, Muñoz A, Gamallo C. Combination of polymorphisms within 5' and 3' untranslated regions of thymidylate synthase gene modulates survival in 5 fluorouracil-treated colorectal cancer patients. Int J Oncol 2009;34(1):219-229.

(163) Schwarzenbach H, Goekkurt E, Pantel K, Aust DE, Stoehlmacher J. Molecular analysis of the polymorphisms of thymidylate synthase on cell-free circulating DNA in blood of patients with advanced colorectal carcinoma. Int J Cancer 2010;127(4):881--888.

(164) Martinez-Balibrea E, Abad A, Martínez-Cardús A, Ginés A, Valladares M, Navarro M, Aranda E, Marcuello E, Benavides M, Massutí B, Carrato A, Layos L, Manzano JL, Moreno V. UGT1A and TYMS genetic variants predict toxicity and response of

colorectal cancer patients treated with first-line irinotecan and fluorouracil combination therapy. Br J Cancer 2010;103(4):581--589.

(165) Mandola MV, Stoehlmacher J, Zhang W, Groshen S, Yu MC, Iqbal S, Lenz HJ, Ladner RD. A 6 bp polymorphism in the thymidylate synthase gene causes message instability and is associated with decreased intratumoral TS mRNA levels. Pharmacogenetics 2004;14(5):319--327.

(166) Vignoli M, Nobili S, Napoli C, Putignano AL, Morganti M, Papi L, et al. Thymidylate synthase expression and genotype have no major impact on the clinical outcome of colorectal cancer patients treated with 5-fluorouracil. Pharmacological Research 2011 9;64(3):242-248.

(167) US National Library of Medicine. VEGFA vascular endothelial growth factor A [Homo sapiens]. 2011; Available at: http://www.ncbi.nlm.nih.gov/gene/7422, 2011.

(168) Koukourakis MI, Papazoglou D, Giatromanolaki A, Bougioukas G, Maltezos E, Siviridis E. VEGF gene sequence variation defines VEGF gene expression status and angiogenic activity in non-small cell lung cancer. Lung Cancer 2004 12;46(3):293-298.

(169) Dassoulas K, Gazouli M, Rizos S, Theodoropoulos G, Christoni Z, Nikiteas N, Karakitsos P. Common polymorphisms in the vascular endothelial growth factor gene and colorectal cancer development, prognosis, and survival. Mol Carcinog 2009;48(6):563--569.

(170) Formica V, Palmirotta R, Del Monte G, Savonarola A, Ludovici G, De Marchis ML, Grenga I, Schirru M, Guadagni F, Roselli M. Predictive value of VEGF gene polymorphisms for metastatic colorectal cancer patients receiving first-line treatment including fluorouracil, irinotecan, and bevacizumab. Int J Colorectal Dis 2011;26(2):143--151.

(171) Hansen TF, Garm Spindler KL, Andersen RF, Lindebjerg J, Brandslund I, Jakobsen A. The predictive value of genetic variation in the vascular endothelial growth factor A gene in metastatic colorectal cancer. Pharmacogenomics J 2011;11(1):53--60.

(172) Loupakis F, Ruzzo A, Salvatore L, Cremolini C, Masi G, Frumento P, Schirripa M, Catalano V, Galluccio N, Canestrari E, Vincenzi B, Santini D, Bencardino K, Ricci V, Manzoni M, Danova M, Tonini G, Magnani M, Falcone A, Graziano F. Retrospective exploratory analysis of VEGF polymorphisms in the prediction of benefit from first-line FOLFIRI plus bevacizumab in metastatic colorectal cancer. BMC Cancer 2011.

(173) Krippl P, Langsenlehner U, Renner W, Yazdani-Biuki B, Wolf G, Wascher TC, Paulweber B, Haas J, Samonigg H. A common 936 C/T gene polymorphism of vascular endothelial growth factor is associated with decreased breast cancer risk. Int J Cancer

2003;103(4):468--471.

(174) US National Library of Medicine. XRCC1 X-ray repair complementing defective repair in Chinese hamster cells 1 [Homo sapiens]. 2011; Available at: http://www.ncbi.nlm.nih.gov/gene/7515, 2011.

(175) Wang Y, Spitz MR, Zhu Y, Dong Q, Shete S, Wu X. From genotype to phenotype: correlating XRCC1 polymorphisms with mutagen sensitivity. DNA Repair 2003 8/12;2(8):901-908.

(176) Stoehlmacher J, Ghaderi V, Iobal S, Groshen S, Tsao-Wei D, Park D, Lenz HJ. A polymorphism of the XRCC1 gene predicts for response to platinum based treatment in advanced colorectal cancer. Anticancer Res 2001;21(4B):3075--3079.

(177) Kim JG, Chae YS, Sohn SK, Moon JH, Kang BW, Park JY, et al. IVS10+12A>G polymorphism in hMSH2 gene associated with prognosis for patients with colorectal cancer. Annals of Oncology 2010 March 01;21(3):525-529.

(178) Grimminger PP, Brabender J, Warnecke-Eberz U, Narumiya K, Wandhöfer C, Drebber U, et al. XRCC1 Gene Polymorphism for Prediction of Response and Prognosis in the Multimodality Therapy of Patients with Locally Advanced Rectal Cancer. J Surg Res 2010 11;164(1):e61-e66.

(179) US National Library of Medicine. XRCC3 X-ray repair complementing defective repair in Chinese hamster cells 3 [Homo sapiens]. 2011; Available at: http://www.ncbi.nlm.nih.gov/gene/7517, 2011.

(180) Yoshihara T, Ishida M, Kinomura A, Katsura M, Tsuruga T, Tashiro S, Asahara T, Miyagawa K. XRCC3 deficiency results in a defect in recombination and increased endoreduplication in human cells.
. EMBO J 2004;23(3):670-680.

(181) McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM, et al. Reporting Recommendations for Tumor Marker Prognostic Studies (REMARK). Journal of the National Cancer Institute 17 August 2005 17 August 2005;97(16):1180-1184.

(182) Hopkins J, Cescon DW, Tse D, Bradbury P, Xu W, Ma C, et al. Genetic polymorphisms and head and neck cancer outcomes: A review. Cancer Epidemiology Biomarkers & Prevention 2008 March 01;17(3):490-499.

(183) Roukos DH, Murray S, Briasoulis E. Molecular genetic tools shape a roadmap towards a more accurate prognostic prediction and personalized management of cancer. Cancer Biol Ther 2007;6(3):308--312.

(184) 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 October 01;59(10):1369-1377.

(185) Applied Biosystems by Life Technologies. Custom TaqMan(R) SNP Genotyping Assays. 2010; Available at: https://products.appliedbiosystems.com/ab/en/US/adirect/ab;jsessionid=SRpFN4hWTggFJVvJP2QjVTt0wSB1R28yhSCkMDFYmnr15vLq1jDT!1226338385?cmd=catNavigate2&catID=601279. Accessed November, 2010.

(186) National Center for Biotechnology Information. dbSNP. Available at: http://www.ncbi.nlm.nih.gov/projects/SNP/.

(187) Applied Biosystems by Life Technologies. Applied Biosystems Products. Available at: https://products.appliedbiosystems.com/ab/en/US/adirect/ab.

(188) Arand M, Mühlbauer R, Hengstler J, Jäger E, Fuchs J, Winkler L, et al. A Multiplex Polymerase Chain Reaction Protocol for the Simultaneous Analysis of the GlutathioneS-Transferase GSTM1 and GSTT1 Polymorphisms. Anal Biochem 1996 4/5;236(1):184-186.

(189) Carlini LE, Meropol NJ, Bever J, Andria ML, Hill T, Gold P, et al. UGT1A7 and UGT1A9 Polymorphisms Predict Response and Toxicity in Colorectal Cancer Patients Treated with Capecitabine/Irinotecan. Clinical Cancer Research 2005 February 01;11(3):1226-1236.

(190) Horie N, Aiba H, Oguro K, Hojo H, Takeishi K. Functional analysis and DNA polymorphism of the tandemly repeated sequences in the 5'-terminal regulatory region of the human gene for thymidylate synthase. Cell Struct Funct 1995;20(3):191--197.

(191) Rodriguez S, Gaunt TR, Day INM. Hardy-Weinberg Equilibrium Testing of Biological Ascertainment for Mendelian Randomization Studies. American Journal of Epidemiology 2009 February 15;169(4):505-514.

(192) Lewis CM. Genetic association studies: Design, analysis and interpretation. Briefings in Bioinformatics 2002 June 01;3(2):146-153.

(193) Katz MH. Multivariable Analysis: A Primer for Readers of Medical Research. Annals of Internal Medicine 2003 April 15;138(8):644-650.

(194) Andy Field. Discovering Statistics Using SPSS. : SAGE Publications Ltd.; 2009.

(195) Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 2005 January 15;21(2):263-265.

(196) International HapMap 3 Consortium, Altshuler DM, Gibbs RA, Peltonen L, Altshuler DM, Gibbs RA, Peltonen L, Dermitzakis E, Schaffner SF, Yu F, Peltonen L, Dermitzakis E, Bonnen PE, Altshuler DM, Gibbs RA, de Bakker PI, Deloukas P, Gabriel SB, Gwilliam R, Hunt S, Inouye M, Jia X, Palotie A, Parkin M, Whittaker P, Yu F, Chang K, Hawes A, Lewis LR, Ren Y, Wheeler D, Gibbs RA, Muzny DM, Barnes C, Darvishi K, Hurles M, Korn JM, Kristiansson K, Lee C, McCarrol SA, Nemesh J, Dermitzakis E, Keinan A, Montgomery SB, Pollack S, Price AL, Soranzo N, Bonnen PE, Gibbs RA, Gonzaga-Jauregui C, Keinan A, Price AL, Yu F, Anttila V, Brodeur W, Daly MJ, Leslie S, McVean G, Moutsianas L, Nguyen H, Schaffner SF, Zhang Q, Ghori MJ, McGinnis R, McLaren W, Pollack S, Price AL, Schaffner SF, Takeuchi F, Grossman SR, Shlyakhter I, Hostetter EB, Sabeti PC, Adebamowo CA, Foster MW, Gordon DR, Licinio J, Manca MC, Marshall PA, Matsuda I, Ngare D, Wang VO, Reddy D, Rotimi CN, Royal CD, Sharp RR, Zeng C, Brooks LD, McEwen JE. Integrating common and rare genetic variation in diverse human populations. Nature 2010;467(7311):52--58.

(197) Salanti G, Sanderson S, Higgins JP. Obstacles and opportunities in meta-analysis of genetic association studies. Genet Med 2005;7(1):13--20.

(198) Lin HJ, Han C, Bernstein DA, Hsiao W, Lin BK, Hardy S. Ethnic distribution of the glutathione transferase Mu 1-1 (GSTM1) null genotype in 1473 individuals and application to bladder cancer susceptibifity. Carcinogenesis 1994 May 01;15(5):1077-1081.

(199) Rahman P, Jones A, Curtis J, Bartlett S, Peddle L, Fernandez BA, et al. The Newfoundland population: a unique resource for genetic investigation of complex diseases. Hum Mol Genet 2003 October 15;12(suppl_2):R167-172.

(200) Diaz-Canton EA PR. Adjuvant medical therapy for colorectal cancer. Surg Clin North Am 1997;77(1):211-228.

(201) Longley DB, Harkin DP, Johnston PG. 5-fluorouracil: mechanisms of action and clinical strategies. Nature reviews. Cancer 2003;3(5):330--338.

(202) Jankun J SE. Yin and yang of the plasminogen activator inhibitor. Pol Arch Med Wewn 2009;119(6):410--417.

(203) Binder BR, Mihaly J. The plasminogen activator inhibitor “paradox” in cancer. Immunol Lett 2008 6/30;118(2):116-124.

(204) Afzal S, Jensen SA, Sørensen JB, Henriksen T, Weimann A, Poulsen HE. Oxidative damage to guanine nucleosides following combination chemotherapy with 5-fluorouracil and oxaliplatin. Cancer Chemother Pharmacol 2011.

(205) Ambrosone CB, Sweeney C, Coles BF, Thompson PA, McClure GY, Korourian S,

et al. Polymorphisms in Glutathione S-Transferases (GSTM1 and GSTT1) and Survival after Treatment for Breast Cancer. Cancer Research 2001 October 01;61(19):7130-7135.

(206) Kim YI. Folate and colorectal cancer: an evidence based critical review. Mol Nutr Food Res 2007;51(3):267-292.

(207) Crott JW, Choi S, Ordovas JM, Ditelberg JS, Mason JB. Effects of dietary folate and aging on gene expression in the colonic mucosa of rats: implications for carcinogenesis. Carcinogenesis 2004 January 01;25(1):69-76.

(208) Weisberg I, Tran P, Christensen B, Sibani S, Rozen R. A second genetic polymorphism in methylenetetrahydrofolate reductase (MTHFR) associated with decreased enzyme activity. Mol Genet Metab 1998 7;64(3):169-172.

(209) Ulvik A, Ueland PM, Fredriksen A, Meyer K, Vollset SE, Hoff G, Schneede J. Functional inference of the methylenetetrahydrofolate reductase 677C>T and 1298A>C polymorphisms from a large-scale epidemiological study. Hum Genet 2007;121(1):57--64.

(210) Yamada K, Chen Z, Rozen R, Matthews RG. Effects of common polymorphisms on the properties of recombinant human methylenetetrahydrofolate reductase. Proceedings of the National Academy of Sciences 2001 December 18;98(26):14853-14858.

(211) Ulrich CM, Potter JD. Folate and Cancer—Timing Is Everything. JAMA: The Journal of the American Medical Association 2007 June 06;297(21):2408-2409.

(212) Holmes RS, Zheng Y, Baron JA, Li L, McKeown-Eyssen G, Newcomb PA, et al. Use of folic acid–containing supplements after a diagnosis of colorectal cancer in the colon cancer family registry. Cancer Epidemiology Biomarkers & Prevention 2010 August 01;19(8):2023-2034.

(213) Kim Y. Folate: a magic bullet or a double edged sword for colorectal cancer prevention? Gut 2006 October 01;55(10):1387-1389.

(214) Duthie SJ. Folate and cancer: how DNA damage, repair and methylation impact on colon carcinogenesis. J Inherit Metab Dis 2011;34(1):101-109.

(215) Ryan BM, Weir DG. Relevance of folate metabolism in the pathogenesis of colorectal cancer. J Lab Clin Med 2001 9;138(3):164-176.

(216) Chen Z, Karaplis AC, Ackerman SL, Pogribny IP, Melnyk S, Lussier-Cacan S, et al. Mice deficient in methylenetetrahydrofolate reductase exhibit hyperhomocysteinemia and decreased methylation capacity, with neuropathology and aortic lipid deposition.

Human Molecular Genetics 2001 March 01;10(5):433-443.

(217) Ogino S, Nosho K, Kirkner GJ, Kawasaki T, Chan AT, Schernhammer ES, et al. A cohort study of tumoral LINE-1 hypomethylation and prognosis in colon cancer. Journal of the National Cancer Institute 2008 December 03;100(23):1734-1738.

(218) Sarter B, Long TI, Tsong WH, Koh WP, Yu MC, Laird PW. Sex differential in methylation patterns of selected genes in Singapore Chinese. Hum Genet 2005;117(4):402--403.

(219) Zienolddiny S, Campa D, Lind H, Ryberg D, Skaug V, Stangeland L, et al. Polymorphisms of DNA repair genes and risk of non-small cell lung cancer. Carcinogenesis March 2006 March 2006;27(3):560-567.

(220) Hyytinen ER, Frierson HF Jr, Boyd JC, Chung LW, Dong JT. Three distinct regions of allelic loss at 13q14, 13q21-22, and 13q33 in prostate cancer. Genes Chromosomes Cancer 1999;25(2):108--114.

(221) Hyytinen ER, Frierson HF Jr, Sipe TW, Li CL, Degeorges A, Sikes RA, Chung LW, Dong JT. Loss of heterozygosity and lack of mutations of the XPG/ERCC5 DNA repair gene at 13q33 in prostate cancer. Prostate 1999;41(3):190--195.

(222) Maestro R, Piccinin S, Doglioni C, Gasparotto D, Vukosavljevic T, Sulfaro S, et al. Chromosome 13q Deletion Mapping in Head and Neck Squamous Cell Carcinomas: Identification of Two Distinct Regions of Preferential Loss. Cancer Research 1996 March 01;56(5):1146-1150.

(223) Yang-Feng TL, Li S, Han H, Schwartz PE. Frequent loss of heterozygosity on chromosomes Xp and 13Q in human ovarian cancer. Int J Cancer 1992;52(4):575--580.

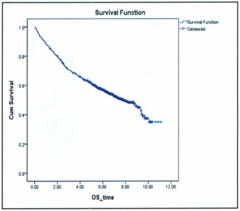
(224) Takebayashi Y, Nakayama K, Kanzaki A, Miyashita H, Ogura O, Mori S, et al. Loss of heterozygosity of nucleotide excision repair factors in sporadic ovarian, colon and lung carcinomas: implication for their roles of carcinogenesis in human solid tumors. Cancer Lett 2001 12/28;174(2):115-125.

(225) Walsh CS, Ogawa S, Karahashi H, Scoles DR, Pavelka JC, Tran H, et al. ERCC5 is a novel biomarker of ovarian cancer prognosis. Journal of Clinical Oncology June 20, 2008 June 20, 2008;26(18):2952-2958.

Appendix

List of figures and tables

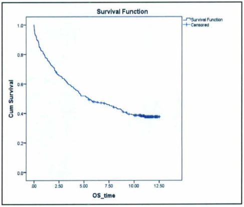
Fig A1. OS plot of NFCCR cohort (n=735)



OS plot of NFCCR cohort (n=735).

5-year OS rate ~62%.

Fig A2. OS plot of entire validation cohort (n=280)



OS plot of entire validation cohort (n=280).

5-year OS rate ~50%.

Table A1. Hardy-Weinberg Equilibrium (HWE) calculations

Gene Symbol	Polymorphism	n	χ^2 value	p ≤ 0.05	Genotypes in HWE
Discovery set					
CCND1	rs9344_Pro241Pro A/G	530	0.01	no	yes
DCC	rs2229080_Arg201Gly C/G	530	0.7	no	yes
EGFR	rs2227983_Arg521Lys G/A	530	2.61	no	yes
ERCC1	rs11615_Asn118Asn C/T	531	3.46	no	yes
ERCC2	**rs13181_Lys751Gln G/T**	**524**	**4.6**	**yes**	**no**
ERCC5	rs1047768_His46His C/T	530	0.6	no	yes
EXO1	rs9350_Pro757Leu C/T	531	0.01	no	yes
FAS	rs1800682_c-24+733T>C	530	0.81	no	yes
FGFR4	rs351855_Gly388Arg A/G	531	2.68	no	yes
**GSTM1*	gene deletion	n/a	n/a	n/a	n/a
GSTP1	rs1695_Ile105Val A/G	525	0.01	no	yes
**GSTT1*	gene deletion	n,a	n/a	n/a	n/a
IL6	rs1800795_-174G/C in promoter	530	0.1	no	yes
MLH1	rs1799977_Ile219Val A/G	531	0.1	no	yes
MMP1	rs1799750_-1607 indel G in promoter	532	0.76	no	yes
MMP2	rs243865_-1306C/T in promoter	530	2.07	no	yes
MTHFR	rs1801133_Ala222Val C/T	524	0.15	no	yes
MTHFR	rs1801131_Glu429Ala A/C	526	1.66	no	yes
OGG1	**rs1052133_Ser326Cys C/G**	**531**	**4.32**	**yes**	**no**
PTGS2	rs4648298_c.3618A/G in 3'-UTR	522	0.14	no	yes
SERPINE1	rs1799889_-675 indelG in promoter	532	1.12	no	yes
TYMS	rs34743033_2/3 repeats of 28bp	532	1.28	no	yes
TYMS	rs16430_indel 6 bp in 3'-UTR	526	0.02	no	yes
VEGFA	**rs2010963_-634G/C in 5'-UTR**	**524**	**9.58**	**yes**	**no**

VEGFA	rs3025039_+936C/T in 3'-UTR	531	0.5	no	yes
XRCC1	rs25487_Arg399Gln G/A	518	0.05	no	yes
XRCC3	**rs861539_Thr241Met C/T**	**531**	**5.42**	**yes**	**no**
Validation set					
MTHFR	rs1801131_Glu429Ala A/C	250	0.02	no	yes
ERCC5	rs1047768_His46His C/T	242	0.28	no	yes
SERPINE1	rs1799889_-675 indelG in promoter	245	1.62	no	yes
**GSTM1*	gene deletion	n/a	n/a	n/a	n/a

n=number of samples genotyped. n/a= not applicable. Polymorphisms with χ^2 value greater than 3.84 were considered to be deviating from HWE with statistical significance (Rodriguez S et al. American Journal of Epidemiology, 2009). Polymorphisms deviated from HWE are shown in bold. *For these deletions, the methods applied did not detect heterozygotes.

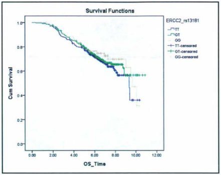
Table A2. Univariate Cox-regression analysis for 27 polymorphisms with OS (co-dominant model)

Variable	p-value	HR	95% CI	n
*ERCC2*_rs13181	0.488			
GT vs TT	0.315	0.848	0.615-1.169	523
GG vs TT	0.343	0.804	0.513-1.261	
*GSTP1*_rs1695	0.66			
GA vs AA	0.415	1.145	0.827-1.584	524
GG vs AA	0.483	1.175	0.749-1.843	
*MTHFR*_rs1801131	0.079			
CA vs AA	0.654	1.075	0.784-1.474	525
CC vs AA	0.025	1.733	1.070-2.807	
*MTHFR*_rs1801133	0.932			
TC vs CC	0.738	1.055	0.771-1.443	523
TT vs CC	0.949	0.983	0.582-1.660	
*VEGFA*_rs2010963	0.369			
GC vs GG	0.705	1.063	0.774-1.461	523
CC vs GG	0.218	0.71	0.412-1.224	
*XRCC1*_rs25487	0.442			
AG vs GG	0.202	1.23	0.895-1.691	517
AA vs GG	0.701	1.105	0.663-1.841	
*ERCC5*_rs1047768	0.012			
TC vs CC	0.097	1.347	0.948-1.914	529
TT vs CC	0.003	1.87	1.238-2.824	
*OGG1*_rs1052133	0.868			
GC vs CC	0.71	1.062	0.772-1.462	530
GG vs CC	0.655	1.141	0.641-2.030	
*ERCC1*_rs11615	0.705			
TC vs TT	0.958	1.009	0.727-1.399	530
CC vs TT	0.434	1.183	0.776-1.802	
*TYMS*_rs16430	0.549			
6 bp/- vs 6 bp/6 bp	0.313	0.85	0.619-1.166	525
-/- vs 6 bp/6 bp	0.482	0.836	0.507-1.378	
*MLH1*_rs1799977	0.72			
GA vs AA	0.701	1.062	0.782-1.443	530
GG vs AA	0.55	0.832	0.454-1.522	
*FAS*_rs1800682	0.478			
TC vs TT	0.848	0.967	0.686-1.362	529
CC vs TT	0.348	1.214	0.810-1.820	
*IL6*_rs1800795	0.146			

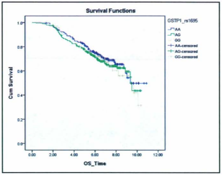
GC vs GG	0.079	1.361	0.965-1.918	
CC vs GG	0.892	1.032	0.654-1.628	529
*EGFR*_rs2227983	0.209			
GA vs GG	0.522	1.106	0.813-1.504	
AA vs GG	0.079	1.662	0.944-2.926	529
*DCC*_rs2229080	0.829			
CG vs CC	0.783	1.045	0.762-1.434	
GG vs CC	0.68	0.9	0.546-1.483	529
*MMP2*_rs243865	0.736			
CT vs CC	0.939	1.012	0.742-1.380	
TT vs CC	0.435	1.313	0.663-2.598	529
*VEGFA*_rs3025039	0.373			
CT vs CC	0.304	1.205	0.844-1.722	
TT vs CC	0.305	1.826	0.578-5.769	530
*FGFR4*_rs351855	0.257			
CT vs CC	0.103	1.298	0.949-1.775	
TT vs CC	0.439	1.215	0.742-1.991	530
*PTGS2*_rs4648298	0.041	2.016	1.030-3.946	521
*XRCC3*_rs861539	0.394			
TC vs CC	0.209	1.234	0.889-1.714	
TT vs CC	0.961	1.012	0.618-1.658	530
*CCND1*_rs9344	0.191			
GA vs GG	0.237	0.813	0.577-1.146	
AA vs GG	0.548	1.132	0.755-1.697	529
*EXO1*_rs9350	0.483			
CT vs CC	0.329	1.177	0.849-1.632	
TT vs CC	0.532	0.694	0.221-2.182	530
*SERPINE1*_rs1799889	0.046			
G/- vs -/-	0.252	0.823	0.589-1.149	
GG vs -/-	0.013	0.557	0.351-0.885	531
*MMP1*_rs1799750	0.126			
G/- vs -/-	0.153	1.31	0.904-1.897	
GG vs -/-	0.044	1.539	1.012-2.339	531
*GSTT1*_gene_deletion	0.585	0.894	0.597-1.339	531
*GSTM1*_gene_deletion	0.009	1.484	1.104-1.994	531
*TYMS*_rs34743033	0.829			
2R/3R vs 3R/3R	0.886	1.026	0.723-1.455	
2R/2R vs 3R/3R	0.562	1.129	0.749-1.702	530

n=no. of samples available for analysis, HR=hazard ratio, CI=confidence interval, 6 bp in *TYMS*_rs16430 refers to the sequence CTTTAA, HR>1 implies increased hazard of death, HR<1 implies reduced hazard of death.

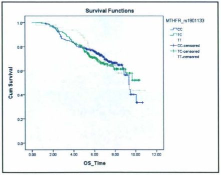
Figures A3.1-A3.21. Kaplan-Meier survival plots for OS in the discovery set (codominant model)



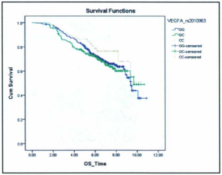
A3.1 *ERCC2*_rs13181 and OS



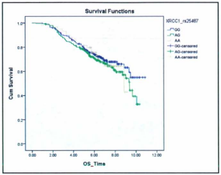
A3.2 *GSTP1*_rs1695 and OS



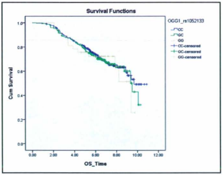
A3.3 *MTHFR*_rs1801133 and OS



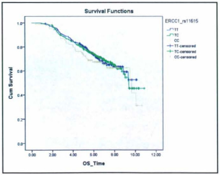
A3.4 *VEGFA*_rs2010963 and OS



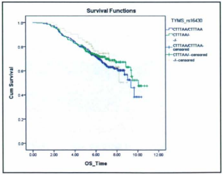
A3.5 *XRCC1*_rs25487 and OS



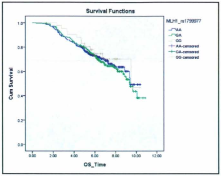
A3.6 *OGG1*_rs1052133 and OS



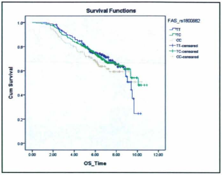
A3.7 *ERCC1*_rs11615 and OS



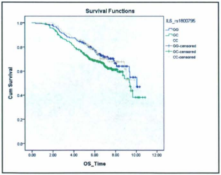
A3.8 *TYMS*_rs16430 and OS



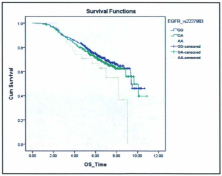
A3.9 *MLH1*_rs1799977 and OS



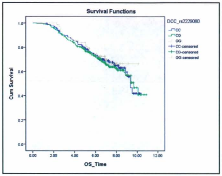
A3.10 *FAS*_rs1800682 and OS



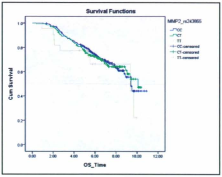
A3.11 *IL6*_rs1800795 and OS



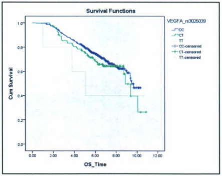
A3.12 *EGFR*_rs2227983 and OS



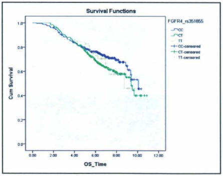
A3.13 *DCC*_rs2229080 and OS



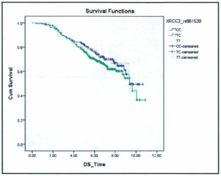
A3.14 *MMP2*_rs243865 and OS



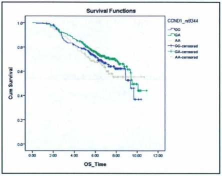
A3.15 *VEGFA*_rs3025039 and OS



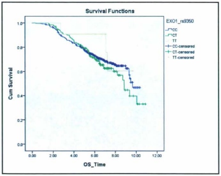
A3.16 *FGFR4*_rs351855 and OS



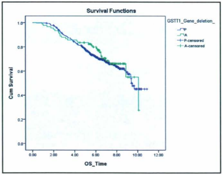
A3.17 *XRCC3*_rs861539 and OS



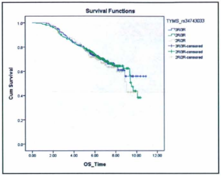
A3.18 *CCND1*_rs9344 and OS



A3.19 *EXO1*_rs9350 and OS



A3.20 *GSTT1* gene deletion and OS



A3.21 *TYMS*_rs34743033 and OS

Table A3. Univariate Cox-regression analysis for OS in discovery set

(recessive model)

Variable	Genotypes	p-value	HR	95% CI	n
*ERCC2*_rs13181	GG vs GT+TT	0.524	0.871	0.57-1.332	523
*GSTP1*_rs1695	GG vs AG+AA	0.676	1.092	0.723-1.648	524
*MTHFR*_rs1801131	CC vs CA+AA	0.027	1.673	1.060-2.641	525
*MTHFR*_rs1801133	TT vs TC+CC	0.865	0.957	0.580-1.580	523
*VEGFA*_rs2010963	CC vs GC+GG	0.174	0.693	0.408-1.177	523
*XRCC1*_rs25487	AA vs AG+GG	0.965	0.989	0.613-1.596	517
*ERCC5*_rs1047768	TT vs TC+CC	0.012	1.564	1.105-2.213	529
*OGG1*_rs1052133	GG vs GC+CC	0.702	1.116	0.635-1.964	530
*ERCC1*_rs11615	CC vs TC+TT	0.404	1.177	0.803-1.727	530
*TYMS*_rs16430	-/- vs 6 bp/- + 6 bp/6 bp	0.68	0.904	0.561-1.459	525
*MLH1*_rs1799977	GG vs GA+AA	0.476	0.808	0.450-1.452	530
*FAS*_rs1800682	CC vs TC+TT	0.23	1.239	0.873-1.757	529
*IL6*_rs1800795	CC vs GC+GG	0.415	0.849	0.573-1.259	529
*EGFR*_rs2227983	AA vs GA+GG	0.098	1.588	0.918-2.744	529
*DCC*_rs2229080	GG vs CG+CC	0.585	0.878	0.551-1.4	529
*MMP2*_rs243865	TT vs CT+CC	0.436	1.306	0.667-2.557	529
*VEGFA*_rs3025039	TT vs CT+CC	0.335	1.757	0.558-5.537	530
*FGFR4*_rs351855	TT vs CT+CC	0.776	1.07	0.671-1.706	530
*XRCC3*_rs861539	TT vs TC+CC	0.61	0.89	0.569-1.392	530
*CCND1*_rs9344	AA vs GA+GG	0.159	1.286	0.906-1.825	529
*EXO1*_rs9350	TT vs CT+CC	0.481	0.663	0.212-2.077	530
*SERPINE1*_rs1799889	GG vs G/- + -/-	0.03	0.634	0.421-0.956	531
*MMP1*_rs1799750	GG vs G/- + -/-	0.135	1.29	0.924-1.803	531
*TYMS*_rs34743033	2R/2R vs 2R/3R+3R/3R	0.551	1.111	0.785-1.572	530

n=number of samples available for analysis, HR=hazard ratio, CI=confidence interval, 6 bp in *TYMS*_rs16430 refers to the sequence CTTTAA, *GSTT1* and *GSTM1* gene deletions as well as *PTGS2*_rs4648298 are not a part of the recessive model, HR>1 implies increased hazard of death, HR<1 implies reduced hazard of death

Table A4. Univariate Cox-regression analysis for OS in discovery set (dominant model)

Polymorphism	Genotype	p-value	HR	95% CI	n
*ERCC2*_rs13181	GG+GT vs TT	0.239	0.836	0.621-1.126	523
*GSTP1*_rs1695	GG+GA vs AA	0.366	1.152	0.848-1.565	524
*MTHFR*_rs1801131	CC+CA vs AA	0.299	1.171	0.869-1.578	525
*MTHFR*_rs1801133	TT+TC vs CC	0.791	1.041	0.772-1.404	523
*VEGFA*_rs2010963	CC+GC vs GG	0.85	0.972	0.720-1.310	523
*XRCC1*_rs25487	AA+AG vs GG	0.23	1.206	0.888-1.636	517
*ERCC5*_rs1047768	TT+TC vs CC	0.019	1.483	1.067-2.062	529
*OGG1*_rs1052133	GG+GC vs CC	0.634	1.076	0.797-1.452	530
*ERCC1*_rs11615	CC+TC vs TT	0.733	1.054	0.778-1.429	530
*TYMS*_rs16430	-/-+6 bp/- vs 6 bp/6 bp	0.275	0.847	0.628-1.141	525
*MLH1*_rs1799977	GG+GA vs AA	0.872	1.025	0.762-1.377	530
*FAS*_rs1800682	CC+TC vs TT	0.819	1.038	0.755-1.427	529
*IL6*_rs1800795	CC+GC vs GG	0.159	1.267	0.911-1.763	529
*EGFR*_rs2227983	AA+GA vs GG	0.307	1.166	0.868-1.566	529
*DCC*_rs2229080	GG+CG vs CC	0.922	1.015	0.750-1.374	529
*MMP2*_rs243865	TT+CT vs CC	0.794	0.961	0.713-1.296	529
*VEGFA*_rs3025039	TT+CT vs CC	0.232	1.235	0.874-1.747	530
*FGFR4*_rs351855	TT+CT vs CC	0.104	1.281	0.950-1.725	530
*XRCC3*_rs861539	TT+TC vs CC	0.29	1.187	0.864-1.630	530
*CCND1*_rs9344	AA+GA vs GG	0.51	0.899	0.653-1.236	529
*EXO1*_rs9350	TT+CT vs CC	0.445	1.133	0.822-1.560	530
*SERPINE1*_rs1799889	GG + G/- vs -/-	0.072	0.745	0.541-1.026	531
*MMP1*_rs1799750	GG + G/- vs -/-	0.07	1.381	0.974-1.959	531
*TYMS*_rs34743033	2R/3R+2R/2R vs 3R/3R	0.736	1.058	0.763-1.468	530

n=number of samples available for analysis, HR=hazard ratio, CI=confidence interval, 6 bp in *TYMS*_rs16430 refers to the sequence CTTTAA, *GSTT1* and *GSTM1* gene deletions as well as *PTGS2*_rs4648298 are not a part of the dominant model, HR>1 implies increased hazard of death, HR<1 implies reduced hazard of death.

Table A5. Univariate Cox-regression analysis for DFS in discovery set

(co-dominant model)

Polymorphism	p-value	HR	95% CI	n
*ERCC2*_rs13181	0.713			
GT vs TT	0.415	0.884	0.657-1.189	
GG vs TT	0.707	0.924	0.612-1.395	522
*GSTP1*_rs1695	0.286			
AG vs AA	0.349	1.155	0.855-1.560	
GG vs AA	0.122	1.381	0.917-2.078	523
*MTHFR*_rs1801131	0.394			
CA vs AA	0.581	1.085	0.812-1.449	
CC vs AA	0.174	1.389	0.864-2.231	524
*MTHFR*_rs1801133	0.906			
TC vs CC	0.994	1.001	0.750-1.336	
TT vs CC	0.672	0.899	0.549-1.472	522
*VEGFA*_rs2010963	0.905			
GC vs GG	0.656	1.07	0.795-1.439	
CC vs GG	0.94	1.018	0.643-1.611	522
*XRCC1*_rs25487	0.794			
AG vs GG	0.892	1.02	0.763-1.364	
AA vs GG	0.555	0.864	0.531-1.404	516
*ERCC5*_rs1047768	0.037			
TC vs CC	0.131	1.28	0.929-1.763	
TT vs CC	0.01	1.647	1.124-2.414	528
*OGG1*_rs1052133	0.215			
GC vs CC	0.74	1.052	0.781-1.415	
GG vs CC	0.08	1.558	0.949-2.559	529
*ERCC1*_rs11615	0.234			
TC vs TT	0.307	1.172	0.864-1.590	
CC vs TT	0.094	1.392	0.945-2.050	529
*TYMS*_rs16430	0.559			
6 bp/- vs 6 bp/6 bp	0.494	0.903	0.673-1.211	
-/- vs 6 bp/6 bp	0.573	1.134	0.733-1.754	525
*MLH1*_rs1799977	0.83			
GA vs AA	0.927	1.013	0.763-1.346	
GG vs AA	0.574	0.856	0.498-1.472	529
*FAS*_rs1800682	0.566			
TC vs TT	0.769	0.954	0.695-1.309	
CC vs TT	0.46	1.152	0.791-1.680	528
*IL6*_rs1800795	0.155			
GC vs GG	0.203	1.225	0.896-1.676	
CC vs GG	0.515	0.869	0.571-1.325	528

*EGFR*_rs2227983	0.389			
GA vs GG	0.952	0.991	0.746-1.318	
AA vs GG	0.187	1.44	0.838-2.476	528
*DCC*_rs2229080	0.819			
CG vs CC	0.742	1.05	0.784-1.407	
GG vs CC	0.701	0.914	0.579-1.445	528
*MMP2*_rs243865	0.884			
CT vs CC	0.827	1.032	0.776-1.373	
TT vs CC	0.634	1.179	0.599-2.322	528
*VEGFA*_rs3025039	0.397			
CT vs CC	0.234	1.219	0.880-1.688	
TT vs CC	0.462	1.538	0.489-4.840	529
*FGFR4*_rs351855	0.274			
CT vs CC	0.107	1.268	0.950-1.694	
TT vs CC	0.603	1.129	0.714-1.786	529
*PTGS2*_rs4648298 (GA vs AA)	0.027	1.985	1.080-3.646	521
*XRCC3*_rs861539	0.465			
TC vs CC	0.236	1.201	0.887-1.627	
TT vs CC	0.854	1.044	0.663-1.643	529
*CCND1*_rs9344	0.444			
GA vs GG	0.949	0.989	0.718-1.364	
AA vs GG	0.294	1.229	0.836-1.808	528
*EXO1*_rs9350	0.483			
CT vs CC	0.464	1.121	0.826-1.520	
TT vs CC	0.367	0.591	0.188-1.854	529
*SERPINE1*_rs1799889	0.533			
G/- vs -/-	0.383	0.869	0.633-1.192	
GG vs -/-	0.294	0.807	0.541-1.204	530
*MMP1*_rs1799750	0.149			
G/- vs -/-	0.221	1.235	0.880-1.733	
GG vs -/-	0.051	1.464	0.998-2.147	530
*GSTT1*_Gene_deletion (A vs P)	0.161	0.758	0.515-1.117	530
*GSTM1*_Gene_Deletion (P vs A)	0.004	1.489	1.133-1.957	530
*TYMS*_rs34743033	0.918			
2R/3R vs 3R/3R	0.846	0.969	0.705-1.331	
2R/2R vs 3R/3R	0.679	0.922	0.628-1.354	529

n=number of patients available for analysis, HR=hazard ratio, CI=confidence interval, 6 bp in *TYMS*_rs16430 refers to the sequence CTTTAA, HR>1 implies increased hazard of event, HR<1 implies reduced hazard of event.

Table A6. Univariate Cox-regression analysis for DFS in discovery set

(recessive model)

Polymorphism	p-value	HR	95% CI	n
*ERCC2*_rs13181 (GG vs GT+TT)	0.925	0.982	.666-1.446	522
*GSTP1*_rs1695 (GG vs AG+AA)	0.198	1.278	.880-1.855	523
*MTHFR*_rs1801131 (CC vs CA+AA)	0.21	1.335	.850-2.096	524
*MTHFR*_rs1801133 (TT vs TC+CC)	0.657	0.899	.560-1.441	522
*VEGFA*_rs2010963 (CC vs GC+GG)	0.967	0.991	.636-1.543	522
*XRCC1*_rs25487 (AA vs AG+GG)	0.506	0.855	.539-1.357	516
*ERCC5*_rs1047768 (TT vs CC+TC)	0.034	1.422	1.027-1.970	528
*OGG1*_rs1052133 (GG vs CC+GC)	0.085	1.531	.943-2.484	529
*ERCC1*_rs11615 (CC vs TC+TT)	0.167	1.279	.902-1.812	529
*TYMS*_rs16430 (-/- vs 6 bp/- + 6 bp/6 bp)	0.401	1.193	.790-1.802	525
*MLH1*_rs1799977 (GG vs GA+AA)	0.546	0.851	.503-1.439	529
*FAS*_rs1800682 (CC vs TC+TT)	0.304	1.186	.857-1.642	528
*IL6*_rs1800795 (CC vs GC+GG)	0.154	0.765	.529-1.105	528
*EGFR*_rs2227983 (AA vs GA+GG)	0.17	1.446	.854-2.448	528
*DCC*_rs2229080 (GG vs CG+CC)	0.59	0.889	.581-1.362	528
*MMP2*_rs243865 (TT vs CT+CC)	0.655	1.164	.597-2.272	528
*VEGFA*_rs3025039 (TT vs CT+CC)	0.507	1.473	.469-4.626	529
*FGFR4*_rs351855 (TT vs CT+CC)	0.973	1.008	.653-1.555	529
*XRCC3*_rs861539 (TT vs TC+CC)	0.739	0.933	.618-1.407	529
*CCND1*_rs9344 (AA vs GA+GG)	0.203	1.237	.891-1.718	528
*EXO1*_rs9350 (TT vs CT+CC)	0.339	0.573	.183-1.792	529
*SERPINE1*_rs1799889 (GG vs G/- + -/-)	0.489	0.885	.627-1.250	530
*MMP1*_rs1799750 (GG vs G/- + -/-)	0.12	1.277	.938-1.739	530
*TYMS*_rs34743033 (2R/2R vs 2R/3R+3R/3R)	0.716	0.94	.675-1.310	529

n=number of patients available for analysis, HR=hazard ratio, CI=confidence interval, 6 bp in *TYMS*_rs16430 refers to the sequence CTTTAA, *GSTT1* and *GSTM1* gene deletion as well as *PTGS2*_rs4648298 are not included in the recessive model, HR>1 implies increased hazard of event, HR<1 implies reduced hazard of event.

Table A7. Univariate Cox-regression analysis for DFS in the discovery set (dominant model)

Polymorphism	p-value	HR	95% CI	n
*ERCC2*_rs13181 (GG+GT vs TT)	0.426	0.894	.679-1.178	522
*GSTP1*_rs1695 (AG+GG vs AA)	0.197	1.205	.908-1.600	523
*MTHFR*_rs1801131 (CA+CC vs AA)	0.381	1.131	.859-1.490	524
*MTHFR*_rs1801133 (TC+TT vs CC)	0.896	0.982	.745-1.293	522
*VEGFA*_rs2010963 (GC+CC vs GG)	0.692	1.057	.803-1.393	522
*XRCC1*_rs25487 (AG+AA vs GG)	0.939	0.989	.749-1.306	516
*ERCC5*_rs1047768 (TC+TT vs CC)	0.036	1.378	1.020-1.861	528
*OGG1*_rs1052133 (GC+GG vs CC)	0.393	1.128	.856-1.488	529
*ERCC1*_rs11615 (TC+CC vs TT)	0.153	1.23	.926-1.633	529
*TYMS*_rs16430 (-/- + 6 bp/- vs 6 bp/6 bp)	0.7	0.947	.719-1.248	525
*MLH1*_rs1799977 (GA+GG vs AA)	0.927	0.987	.752-1.297	529
*FAS*_rs1800682 (TC+CC vs TT)	0.942	1.011	.753-1.358	528
*IL6*_rs1800795 (GC+CC vs GG)	0.461	1.12	.829-1.512	528
*EGFR*_rs2227983 (GA+AA vs GG)	0.779	1.04	.792-1.366	528
*DCC*_rs2229080 (CG+GG vs CC)	0.88	1.022	.772-1.353	528
*MMP2*_rs243865 (CT+TT vs CC)	0.751	1.046	.793-1.378	528
*VEGFA*_rs3025039 (CT+TT vs CC)	0.196	1.234	.897-1.697	529
*FGFR4*_rs351855 (CT+TT vs CC)	0.128	1.238	.941-1.630	529
*XRCC3*_rs861539 (TC+TT vs CC)	0.297	1.169	.872-1.566	529
*CCND1*_rs9344 (GA+AA vs GG)	0.735	1.053	.779-1.425	528
*EXO1*_rs9350 (CT+TT vs CC)	0.644	1.073	.796-1.447	529
*SERPINE1*_rs1799889 (G/- + GG vs -/-)	0.293	0.851	.630-1.149	530
*MMP1*_rs1799750 (G/- + GG vs -/-)	0.099	1.307	.951-1.797	530
*TYMS*_rs34743033 (2R/3R+2R/2R vs 3R/3R)	0.755	0.954	.708-1.285	529

n=number of patients available for analyis, HR=hazard ratio, CI=confidence interval, 6 bp in *TYMS*_rs16430 refers to the sequence CTTTAA, *GSTT1* and *GSTM1* gene deletions as well as *PTGS2*_rs4648298 are not included in the dominant model, HR>1 implies increased hazard of event, HR<1 implies reduced hazard of event.

Table A8. Multivariate analysis for OS in the discovery set (recessive model)

Variable	p-value	HR	95% CI for HR	
*MTHFR*_rs1801131 (CC vs CA+AA)	0.03	1.693	1.052	2.723
*ERCC5*_rs1047768 (TT vs CC+TC)	0.009	1.647	1.13	2.4
*OGG1*_rs1052133 (GG vs GC+CC)	0.228	1.444	0.794	2.624
*IL6*_rs1800795 (CC vs GC+GG)	0.05	0.66	0.435	1.001
*EGFR*_rs2227983 (AA vs GA+GG)	0.019	1.963	1.118	3.444
*SERPINE1*_rs1799889 (GG vs G/- + -/-)	0.037	0.634	0.414	0.972
Age at diagnosis	0.016	1.021	1.004	1.039
Stage	<0.001			
II vs I	0.174	1.48	0.841	2.604
III vs I	0.005	2.223	1.274	3.879
IV vs I	<0.001	13.194	7.213	24.135
MSI status (MSI-H vs MSI-L/MSS)	0.002	0.21	0.077	0.57

n=503. *GSTM1* and *GSTT1* gene deletions were not included in the recessive model, HR: hazard ratio, CI: confidence interval, HR>1 implies increased hazard of death, HR<1 implies reduced hazard of death.

Table A9. Multivariate analysis for OS in the discovery set (dominant model)

Variable	p-value	HR	95% CI for HR	
*MTHFR*_rs1801131 (CA+CC vs AA)	0.199	1.224	0.899	1.666
*ERCC5*_rs1047768 (TC+TT vs CC)	0.013	1.544	1.095	2.177
Age at diagnosis	0.013	1.022	1.005	1.039
Stage	<0.001			
II vs I	0.102	1.597	0.911	2.801
III vs I	0.002	2.385	1.371	4.15
IV vs I	<0.001	11.365	6.302	20.498
MSI status (MSI-H vs MSI-L/MSS)	0.001	0.19	0.07	0.516

n=504. *GSTM1* and *GSTT1* gene deletions are not included in the dominant model, HR: hazard ratio, CI: confidence interval, HR>1 implies increased hazard of death, HR<1 implies reduced hazard of death.

Table A10. Multivariate analysis for DFS in the discovery set (recessive model)

Variable	p-value	HR	95% CI for HR	
*MTHFR*_rs1801131 (CC vs CA+AA)	0.067	1.564	0.97	2.523
*ERCC5*_rs1047768 (TT vs CC+CT)	0.069	1.379	0.976	1.95
*OGG1*_rs1052133 (GG vs CC+GC)	0.035	1.727	1.04	2.869
*TYMS*_rs16430 (-/- vs 6 bp/6 bp + 6 bp/-)	0.039	1.586	1.023	2.459
*DCC*_rs2229080 (GG vs CG+CC)	0.128	0.708	0.454	1.104
*XRCC3*_rs861539 (TT vs TC+CC)	0.292	0.79	0.51	1.225
Location (rectum vs colon)	0.006	1.552	1.137	2.117
Stage	<0.001			
II vs I	0.299	1.308	0.788	2.169
III vs I	0.009	1.951	1.185	3.212
IV vs I	<0.001	5.469	3.19	9.376
MSI status (MSI-H vs MSI-L/MSS)	0.002	0.274	0.121	0.62
*BRAF1*_Val600Glu mutation status (+ vs -)	0.022	1.87	1.095	3.193

n=466. *TYMS*_rs16430 is referred as the indel 6 bp polymorphism, 6 bp in *TYMS*_rs16430 refers to the sequence CTTTAA, *GSTT1* and *GSTM1* gene deletions were not included in the recessive model, HR: hazard ratio, CI: confidence interval, HR>1 implies increased hazard of event, HR<1 implies reduced hazard of event.

Table A11. Multivariate analysis for DFS in the discovery set (dominant model)

Variable	p-value	HR	95% CI for HR	
*ERCC5*_rs1047768 (TC+TT vs CC)	0.08	1.318	0.967	1.795
*ERCC1*_rs11615 (TC+CC vs TT)	0.126	1.256	0.938	1.683
Location (rectum vs colon)	0.054	1.328	0.995	1.772
Stage	<0.001			
II vs I	0.101	1.505	0.924	2.453
III vs I	0.002	2.139	1.322	3.46
IV vs I	<0.001	5.941	3.527	10.006
MSI status (MSI-H vs MSI-L/MSS)	0.004	0.346	0.169	0.712

n=507. *GSTT1* and *GSTM1* gene deletions are not included in the dominant model, HR: hazard ratio, CI: confidence interval, HR>1 implies increased hazard of event, HR<1 implies reduced hazard of event.

Table A12. Chi-square test results between polymorphisms and clinicopathological & molecular variables (recessive model)

Polymorphism	Variable	p-value	n
*CCND1*_rs9344	Histology	0.03	530
*CCND1*_rs9344	Stage	0.016	530
*FAS*_rs1800682	Histology	0.001	530
*IL6*_rs1800795	Sex	0.009	530
*MMP1*_rs1799750	Vascular invasion	0.04	492
*SERPINE1*_rs1799889	Sex	0.039	532
*VEGFA*_rs2010963	MSI status	0.003	503
**VEGFA*_rs2010963	Grade	0.03	521
*XRCC3*_rs861539	*BRAF1*_Val600Glu mutation status	0.027	483

*By Fisher's exact test. Only statistically significant correlations are shown. n: number of patients

Table A13. Chi-square test results between polymorphisms and clinicopathological & molecular variables (dominant model)

Polymorphism	Variable	p-value	n
*CCND1*_rs9344	Histology	0.02	530
*ERCC1*_rs11615	Stage	0.031	531
*FAS*_rs1800682	Location	0.046	530
*FAS*_rs1800682	Familial risk	0.027	530
*IL6*_rs1800795	Grade	0.031	526
*XRCC1*_rs25487	Vascular invasion	0.023	479
*XRCC1*_rs25487	Lymphatic invasion	0.028	476
*XRCC1*_rs25487	MSI status	0.017	499
*XRCC3*_rs861539	*BRAF1*_Val600Glu mutation status	0.03	483
*TYMS*_rs34743033	Sex	0.006	531

Only statistically significant correlations are shown. n: number of patients

Table A14. Chi-square test results between polymorphisms and clinicopathological & molecular variables (co-dominant model)

Polymorphism	Variable	p-value	n
*CCND1*_rs9344	Histology	0.022	530
*CCND1*_rs9344	Stage	0.017	530
*FAS*_rs1800682	Location	0.014	530
*FAS*_rs1800682	Histology	0.003	530
*FGFR4*_rs351855	Location	0.032	531
*IL6*_rs1800795	Sex	0.029	530
*MMP2*_rs243865	Histology	0.029	530
*VEGFA*_rs2010963	MSI status	0.012	503
*XRCC1*_rs25487	Vascular invasion	0.046	479
*XRCC1*_rs25487	Lymphatic invasion	0.041	476
*XRCC1*_rs25487	MSI status	0.047	499
*XRCC3*_rs861539	*BRAF1*_Val600Glu mutation status	0.024	483
*TYMS*_rs34743033	Sex	0.018	531

Only statistically significant correlations are shown. n: number of patients