2006

Gene expression profiles and clinical parameters for survival prediction in stage II and III colorectal cancer

Mubeena Begum
University of South Florida

Gene Expression Profiles and Clinical Parameters for Survival Prediction in Stage II and III Colorectal Cancer Patients

by

Mubeena Begum, M.D

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science in Public Health
Department of Epidemiology and Biostatistics
College of Public Health
University of South Florida

Major Professor: Thomas J. Mason, Ph.D.
James F. Helm, M.D, Ph.D.
Heather G. Stockwell, Ph.D.
Tao Wang, Ph.D

Date of Approval:

February 1st, 2006

Keywords: Survival time, Prognostic factors, Molecular markers, SAM, mRNA.

Dedication

I dedicate this research to my parents: Amartur Maqhbool Ahmed and Basheera Khatoon for being an inspiration to me all my life.

Acknowledgements

I would like to thank my advisor, Dr. Thomas J. Mason, for his tremendous support and invaluable guidance during the time of this thesis work, not only in the research projects, but also in my life beyond science. It has been a great pleasure and honor to work and study under his mentorship

I am grateful to Dr. James Helm for his guidance throughout the preparation of my thesis. My many thanks to my thesis committee members Dr. Heather G. Stockwell and Dr. Tao Wang for their valuable suggestions and for taking time to serve on my thesis. I would like to extend my appreciation to our collaborators at Moffiitt Cancer Hospital and Research Institute: Dr. Thimothy Yeatman, Dr. Erin Siegel, Dr. Alan B. Cantor and Steve Eschrich. Helen Lewis in Moffitt Cancer Registry for providing me with patient information.

Sincere thanks to Dr. Noreen Poor, Paul Tate, Hillary Strayer, Marilyn Williams, Silvia Calderon and Dr. Connie Mizak for their constant encouragement during my study at USF. A special thank you to my racquetball coach and friend Tom Vann, for always being there for me, for his moral support and confidence in me.

Finally, I would like to thank my parents and family members for their love, encouragement and support. It is because of them that I moved forward.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ACS	American Cancer Society
BMI	Body Mass Index
CEA	Carcinoembriyonic Antigen
CI	Confidence Interval
CIN	Chromosomal Instability
CRC	Colorectal Cancer
FAP	Familial Adenomatous Polyposis
FDR	False Discovery Rate
GEP	Gene Expression Profiles
HNPCC	Hereditary Nonpolyposis Colorectal Cancer
LN	Lymph Node
MSI	Microsatellite Instability
NSAIDS	Non Steroidal Anti Inflammatory Drugs
RMA	Robust Microarray
SAM	Significance Analysis of Microarrays
SE	Standard Error
VEGF	Vascular En dothelial Growth Factor

Gene Expression Profiles and Clinical Parameters for Survival Prediction in Stage II and III Colorectal Cancer Patients

Mubeena Begum

ABSTRACT

Prediction of outcome in colorectal cancer (CRC) is currently based on the TNM staging classification; however, histopathological classification alone is insufficient for accurately predicting survival in stage II and III patients. Studies indicate that microarray gene expression profiles can predict survival in CRC. We hypothesize that tumor gene expression in combination with clinical parameters, is a better predictor of outcome in stage II and III colorectal cancers than the TNM stage classification alone.

Clinical records and follow-up data were retrospectively reviewed for 58 Stage II and Stage III patients with primary colorectal cancer, who did not receive any neoadjuvant therapy preoperatively and whose samples had been previously analyzed for gene expression profiles using the Affymetrix U 133a Gene chip. For molecular classification of patients as being at high or low risk for poor survival, samples were divided into two clusters by hierarchical cluster analysis of genes selected by SAM. Univariate and multivariate analyses

using Cox proportional hazard models were done to identify significant prognostic factors

The 3-year and 5-year survival estimates were 72.41% (SE=5.8%) and 55.17% (SE=6.7%), respectively, for all 58 patients. Univariate analysis showed that advanced stage, older age, high-risk molecular classification, positive lymph nodes were the statistically significant prognostic factors of poor survival ($p<0.05$), while gender, preoperative CEA level, and family history of CRC in first degree relatives were not statistically significant. In multivariate analysis molecular classification, age and body mass index were independent significant prognostic factors. In Cox proportional hazard model, the estimated hazard ratios for Stage III vs II was 2.45 (95%CI: 0.85-7.04), for high vs low molecular risk was 3.83 (95%CI: 1.22-12.06) and old vs young age was 3.72 (95%CI: 1.2-11.49). Model containing clinical stage in conjunction with molecular risk, body mass index, and age was a stronger indicator of clinical outcome ($p= 0.0056$) than model with clinical stage alone.

Gene expression profiles predict survival independent of clinical parameters, and the addition of gene expression profiles to stage is more predictive of survival than stage alone. Further analysis needs to be done to validate the molecular classification on an independent dataset.

CHAPTER ONE

INTRODUCTION

1.1 Background

Colorectal cancer (CRC), cancer of colon and rectum, the third most common cancer worldwide and the second leading cause of cancer-related mortality in the US [1,2]. It is the second most site-specific cancer affecting both men and women (lung cancer is first, affecting both men and women, breast is the leader in women and prostate in men) [2]. The lifetime probability of developing colon cancer in men is 1 in 17 and in women it is 1 in 18 [2]. A study conducted by Parkin, DM et al, 1999 [34] discusses that approximately, 6% of the American population will eventually develop invasive CRC and over 6 million Americans who are alive today will die of the disease. 75% of patients with CRC have sporadic disease, with no evidence of having inherited the disorder and 25% have a family history of CRC suggesting a genetic contribution.

Majority of the colorectal cancers arise due to malignant transformation of an adenomatous polyp. The malignant tumor arises from colonic epithelial cells that line the mucosa. Transition from normal epithelium to adenoma and to carcinoma is due to acquired molecular events [1], that is, 85% of CRC are due to events which lead to chromosomal instability (CIN) and 15% are due to microsatellite instability (MSI). These events alter chromosomes 5q (APC), 18q (DCC) and 17p (TP53) involved in DNA repair.

1.2 Colon and Rectum: Structure and function

Colon refers to the upper six feet of the large intestine and rectum to the last five to six inches. Together colon and rectum make up the large intestine. The colon is made of four sections: ascending colon, transverse colon, descending colon and the sigmoid colon. Cancer can develop in any of the four sections of the colon and in the rectum. The distributions of colorectal cancer in the large intestine are: ascending colon and cecum-25%, transverse colon-15%, descending colon-5%, sigmoid colon-25%, rectum-25% and rectosigmoid junction-10% [43].

Tumors on the right side of the colon near the cecum usually grow large enough to be painful and cause bleeding. As a result they commonly present with anemia from chronic blood loss. Polyps commonly appear on the left side of the colon. Cancer on the left colon usually grows around the colon wall and encircles it. Common symptoms of a tumor on the left side include constipation and change in bowel habits. Cancer can grow inward toward the hollow part of the colon or rectum, and /or outward through the wall of the colon or rectum. In untreated cases the cancerous cells break away from the primary site and spread to distant organs through bloodstream or lymphatic system. This process is called metastases. 95% of CRC are carcinomas, and 95% of these are adenocarcinomas.

1.3 Classification of Disease:

The currently used staging system for CRC is UICC-AJCC TNM staging system. The AJCC TNM staging system is considered to be more useful for clinical decision-making, due to its precise stratification. It consists of three independent prognostic variables: the depth of tumor invasion into the bowel wall (T), the presence or absence of lymph node involvement (N), and the presence or absence of distant metastases (M) [21, 53]. The pathologic staging is assigned after the resection of the primary tumor, removal and examination of regional lymph nodes and analysis of the surgical specimen. The AJCC/ UICC and Dukes' classification system is shown in table 1. The survival rates for the different staging systems are summarized in table 2.

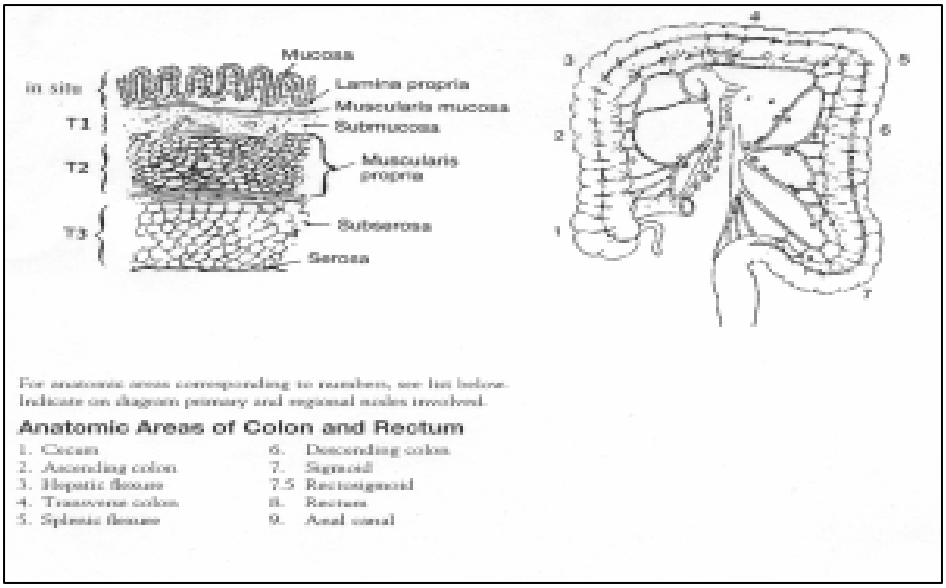


Fig1. Anatomic Division and Tumor Penetration (AJCC 6th Edition System)

Table1. AJCC/UICC Staging Sytem for Colon and Rectal Cancer

AJCC/UICC	Description of AJCC Staging System	Dukes'	MAC
Stage 0 and Tis	Carcinoma in situ: intraepithelial or invasion of the lamina propria.		
Stage I (T1 N0 M0)	Tumor invades through the submucosa. No metastases to regional nodes or distant metastases.	Dukes' A	A
Stage I (T2 N0 M0)	Tumor invades into the muscularis propria. No metastases to regional nodes or distant metastases.	Dukes' A	B1
Stage IIA (T3 N0 M0)	Tumor invades through the muscularis propria into the subserosa, or into non-peritonealized pericolic or perirectal tissues, no metastases to regional nodes or distant metastases.	Dukes' B	B2
Stage IIB (T4 N0 M0)	Tumor directly invades other organs or structures, and/or perforates visceral peritoneum, no metastases to regional nodes or distant metastases.	Dukes' B	B3
Stage IIIA (T1-2 N1 M0)	Tumor invades submucosa or muscularis propria. Metastases in 1 to 3 regional lymph nodes. No distant metastases.	Dukes' C	C1
Stage IIIB (T3-4 N1 M0)	Tumor invades through the muscularis propria into the subserosa, or into non-peritonealized pericolic or perirectal tissues or tumor directly invades other organs or structures, and/or perforates visceral peritoneum. Metastases in 1 to 3 regional lymph nodes. No distant metastases.	Dukes' C	C2/C3
Stage IIIC (Any T, N2, M0)	Any extent of tumor invasion. Metastasis in 4 or more regional. No distant metastases.	Dukes' C	C1/C2/ C3
Stage IV	Any extent of tumor invasion or number of metastases to regional nodes. Distant metastases present.	Dukes' D	D

1.4 Statement of Problem

Prediction of outcome is an important aspect in cancer research. In patients with colorectal cancer limited to the mucosa (TNM Stage I) and in patients with distant metastases (TNM Stage IV), the 5-year survival rates are 90-95% and <10%, respectively [2, 32]. In these patients histopathologic criteria (TNM staging) are good predictors of survival. However, in patients with invasion of the colon wall or adjacent structures (stage II) and those who have regional nodal metastases (stage III), the probability of survival is about 70-85% and 40-80% respectively based on the TNM staging [32]. Moreover, patients with same tumor stages may show different prognosis indicating that conventional staging procedures may be unable to precisely predict cancer risk. Thus, it is in stage II and III patients that a better predictor of survival is needed. As an alternative to clinical staging, recently developed microarray technology has permitted the development of multiorgan cancer classifier, identification of tumor subclass, discovery of progression markers, and prediction of disease outcome in many types of cancer. Unlike clinicopathological staging, molecular staging is able to better predict the long-term outcome of an individual based on the gene expression profile of the tumor at diagnosis [11].

Preliminary studies indicate that microarray gene expression profiles have been most accurate to date in predicting the overall survival in CRC [11]. However, most research examining gene expression profile has not taken clinical parameters (stage, age, sex, grade, preoperative CEA levels) into consideration.

Table 2: 30 months and 60 months relative survival rate by AJCC sixth Edition System [32].

Stage	30 months (%)	60 months % (5-year relative survival rate)
Stage I	96.1	93.2
Stage IIA	91.0	84.7
Stage IIB	80.2	72.2
Stage IIIA	91.4	83.4
Stage IIIB	77.3	64.1
Stage IIIC	59.1	44.3
Stage IV	17.3	8.1

The 5-year survival rate refers to the percentage of people who live at least 5 years after their cancer is diagnosed.

1.5 Study Purpose

The purpose of the study was to investigate whether clinicopathological (TNM staging) based outcome (survival) prediction can be improved by combining microarray gene expression profiles together with other clinical predictors (age, sex, grade, preoperative CEA level).

Stage ----------→ Survival

Stage+ GEP+ age + gender + body mass index (BMI)+ family history

+ location of tumor+ grade+ preoperative CEA level ---------- → Survival

+ total resected lymph nodes

1.6 Research Question:

1. What is the predictive value of traditional clinical predictors- age, sex, race, stage and grade in determining prognosis (survival) in Stage II and III CRC patients?
2. What is the predictive value of gene expression profiles (GEP) by itself in this sample?
3. What does the addition of GEP together with clinical parameters (age, sex, family history, BMI, grade, location of tumor, preoperative CEA levels) contribute to the usual clinical predictiveness of outcome (survival) in stage II and III colorectal cancer?

1.7 Hypothesis

Gene expression profiles (GEP) and clinicopathological factors (age, sex, BMI, family history, grade, preoperative CEA level) add to the predictive value of staging in predicting the postoperative outcome (survival) in stage II and III CRC.

CHAPTER TWO

REVIEW OF LITERATURE

2.1 Descriptive Characteristics

Currently CRC constitutes 10% of new cancer cases in men and 11% new cancer cases in women. Estimated new cases and deaths from CRC in the US in 2005 are 104,950 new cases and 56,290 deaths [2]. SEER data for 1998-2002 show the overall incidence of CRC is higher in men (62.1/100,000) than in women (24.8/100,000), and this holds true for mortality rates, in men (46.2/100,000) and women (17.4/100,000). The median age at diagnosis in the United States, during this period is 70 for men and 74 for women.

Incidence is higher among African American men and women (62.4%) compared to White men and women (52.5%), and so are the mortality rates in African Americans (20%) to Whites (27.9%). The risk of CRC increases after the age of 50-55 years and continues to rise exponentially with increasing age. Between 1998 and 2001, the incidence rate has declined by 2.9% and 5-year survival rate has increased by 7.3%, which could be due to advances in detection and screening and the increasing use of combination therapies [40].

Although the exact cause of CRC cancer is unknown, several factors play a crucial role in the development and prognosis of CRC survival, which can be classified as prognostic factors (tumor related, host related and environmental related factors) and risk factors.

Definition of Prognostic Factor: In epidemiological literature prognostic factor refers to the probability of future event in patients who currently have a disease. It implies prediction of an event that will occur in the future. It can be considered in the context of probability of cure or prolongation of survival. Knowledge of prognostic factors helps us to understand the progress of the disease [19].

Definition of a Risk Factor: "A clearly defined occurrence or characteristic that has been associated with the increased rate of a subsequently occurring disease". It is limited to those who don't have a disease [19].

2.2 Prognostic Factors

2.2.1 *Tumor related prognostic factors*

- ***Pathological staging*:** Is the most important predictor of outcome in patients with newly diagnosed colorectal cancer [39, 52], which depends on the degree of penetration of tumor through the bowel wall, presence of or absence of nodal involvement and presence or absence of distant metastasis. Majority of the CRC are adenocarcinomas.
- ***Histological grade*:** Tumor prognosis correlates with histological grade: poor differentiation has a worse prognosis than a high degree of differentiation [39, 45]. Large studies have shown that histological grade correlates with survival and recurrence, with low-grade tumors having better survival [21]. Venous, lymphatic and perineural invasion have also shown to decrease survival and increase the risk of local recurrence [21, 52].

- ***Surgical margins and radial margins***: The presence of positive surgical or radial margin is a poor prognostic factor, with local failure rate in all stages increasing from 3%-85% [37].
- ***Molecular markers*:** The use of molecular markers that have prognostic significance aids in identifying high-risk patients who can benefit from adjuvant chemotherapy and avoiding those who have low risk from the toxicities of adjuvant chemotherapy. Presence of microsatellite instability has shown to improve prognosis in sporadic colorectal cancers [20, 22]. Three studies have independently shown unfavorable prognosis of patients with the loss of 18q in stage II and III CRC [24, 33, 41].
- ***Carcinoembryonic antigen* (CEA):** CEA levels are used to monitor the course of colorectal cancer. Elevated pre-operative levels of CEA at diagnosis, has shown to be an independent prognostic factor for survival and recurrence in CRC patients [21].
- ***Lymph nodes examined:*** AJCC and NCI recommend that at least 12 lymph nodes should be examined in patients with colorectal cancer to confirm the absence of nodal involvement by tumor [9]. It is a reflection of the aggressiveness of lymphovascular mesenteric dissection during surgery and pathological identification of nodes in the specimen. Retrospective studies have demonstrated that the number of LN examined in CR surgery may be associated with patient's outcome [47]. A study by Berger. A. C et al., 2005, have demonstrated that lymph node resection is a statistically important prognostic factor for determining overall survival and disease free survival.

2.2.2 ***Patient-related prognostic factors:***

- ***Genetic Syndromes*:** Patients with Hereditary Nonpolyposis colorectal cancer

(HNPCC- is a familial syndrome in which individuals develop CRC before 50 years of age), chronic ulcerative colitis (UC- inflammatory condition of the large intestine) and Familial adenomatous polyposis (FAP- an autosomal dominant condition characterized by multiple polyps with high potential to progress to cancer) are at increased risk of colorectal cancer [51].

Hereditary Nonpolyposis Colorectal Cancer (HNPCC): Also called Lynch Syndrome. It is an autosomal dominant condition caused by mutation in the DNA mismatch repair genes (hMSH2, hMLH1). HNPCC accounts for about 3-5% of all CRC [2]. They have increased risk of developing adenomas at an early age, the average age of CRC diagnosis in HNPCC syndrome patients is around 44 years [2]. They also have an increased risk of developing other cancers such as endometrial, ovarian, small intestine, pancreatic, renal pelvis and brain tumors.

Familial Adenomatous Polyposis (FAP): is an autosomal dominant condition with a prevalence of 1/8,000 [2]. FAP is due to mutation in the APC gene on chromosome 5q21. The disease is characterized by hundreds of polyps in the colon and rectum, which develop after the first decade of life. By the age of 20 and 30 the probability of developing colonic adenoma increases by 75% and 90%, respectively [2].

- ***Familial Colon Cancer:*** A positive family history is an important risk factor developing CRC. A two to three-fold increase in risk is seen if an individual has a first degree relative with CRC and the risk increases if more relatives are affected [40]. Family history of CRC is seen in 10-15 % of persons with CRC. This increases the person's risk to develop CRC by 2-6 fold. With a family history, the risk of CRC increases earlier in life (less than 45 years) than later.

- ***Age*:** 3 % of the CRC arise before age of 30 years, and 11% have predisposing conditions such as FAP and UC. The risk of CRC increases with age and it is most common in men and women above 55 years. Studies have reported poor prognosis in CRC patients who are less than 40 years [1]. Tumors with microsatellite instability (MSI) have better prognosis irrespective of age [2].
- ***Racial difference:*** Racial differences in the overall survival were observed in few studies although some studies have shown that co-morbid conditions play a role in the survival outcome in different patient population. Jews of European decent have a higher rate of CRC due to genetic mutation.
- ***Weight:*** Many studies have reported an increased risk of colorectal cancer with increasing body mass index [10]. Having excess fat in the waist are (intra-abdominal fat) increases the risk more than having the same amount of fat distributed in other areas (thighs, hips). Obesity placed men more than women at increased risk for colon cancer. This association in men is due to greater waist circumference in men and the protective effect of estrogen in women, which decreases CRC risk. A study by Giovanucci et al., 2001 [16] found a correlation between colon cancer and type-2 diabetes. Obesity predisposes a premenopausal woman to the same risk as does for men in general [12].
- ***Geography:*** rates of colorectal cancer vary geographically. The disease is more common among industrialized nations –USA, Western Europe, Australia and uncommon in Asia, Africa and South America [48].

Table: 3 Literature Review of Clinical and Pathological Factors Assessed by Cox Multivariate Analysis

Author, year	Study Features	Factors Analyzed	Findings and Strengths/Weakness
Bertucci, F et al. 2004 [5]	Prospective study Single institution, France Sample size: 50 samples (26 patients) between 1990-1998. All tumor sections and Medical records were reviewed prior to analysis. Unsupervised hierchical clustering was used to investigate relationship between samples and genes.	Gender, age, site of tumor, grade, tumor penetration, LN involvement, vascular invasion, stage, surgery,	Significance of prognostic classification made by AJCC stage and the obtained gene set was compared. Classification based on AJCC stage was significant but less than that made by gene expression profiles. prognostic impact of gene set persisted when applied to patients without metastasis at diagnosis and patients without metastasis and LN involvement. Strengths: Accuracy of prediction of "molecular metastatic signature' was estimated by leave one out procedure. DNA microarray was able to identify clinically relevant tumor subgrous based on the gene cluster. Weakness: Multivariate analysis was not done to determine significant clinical and pathological factors affecting survival.
Barrier, A. et al., 2005 [3]	Prospective study Sample size: 18patients Stage II and II CRC patients Follow-up: Every 3 years-1st postoperative thereafter. year and every 6 months	Gender, stage, grade and location. Prognostic prediction was build based on microarray gene expression measure for stage II and II CRC patients. For each dataset a total of 150 prognosis predictors were considered and Performance was assessed using six-fold cross-validation.	70 gene predictor was built, 35 were overexpressed in patients who developed a recurrence and 35 in patients who were disease free for 5 years. Strengths: Double cross validation was done by splitting the data into testing and training set. The results of the study suggest the ability to build a prognosis predictor for both stage II and III CRC, based on either T or NM gene expression profiles. The accuracy of the 70-gene NM based predictor was greater than that of 30-gene based predictor (83 vs 78%). Weakness: The study did not analyze the effect of Clinico pathological factor in association with gene Expression profiles in predicting the survival.

Author, year	Study Features	Factors Analyzed	Findings and Strengths/Weakness
	syndrome, cancer with IBD, rectal cancers were excluded because their molecular feature, recurrence rate, and overall differs from sporadic colon cancers.	Dukes' stage, grade, microsatellite status, resected nodes, postoperative complications, recurrence. molecular and structural markers- P53, P27, VEGF, microvessel count	regulation and Dukes' stage, LN status. VEGF overexpression correlated significantly with Dukes' stage. Strengths: Study compared levels of p53, p27, VEGF and MVC, which are involved in cycle regulation, apoptosis, and tumor neoangiogenesis, for normal and colon cancer cells with clinopathological variables. In conjuction with clinicopathological staging, molecular Expression markers p27, p53, VEGF, provide a stronger Indication of clinical outcome than with staging alone and help better select therapeutic option in colorectal cancer patients.
Murphy T K et al. 2000 [31]	Large prospective study Population based, 50 States Sample size: 1,184,659 Follow –up: 12 years 1616 final sample size followinginclusion and exclusion criteria	Age, sex, smoking and alcohol. history, dietary history, exercise, estrogen replacement therapy and asprin use. BMI was categorized for age and gender.	The study was done to examine the association between BMI and colon cancer mortality in both men and women. The findings in the study were that BMI was an independent risk factor for colon cancer death in both sexes and the relationship is stronger and more linear in men than in women. This could be due to central obesity causing hyperinsulinemia and increased glycemic load causing tumor growth. Alcohol intake significantly modified the association between BMI and colon cancer mortality. Strengths: Large prospective study, generalizeable to the Population. Weakness: lack of screening data, self reported measurements. Did not look into other clinicopathological factors which affect survival in CRC.

Author, year	Study Features	Factors Analyzed	Findings and Strengths/Weakness
Ponz de Leon et al, 1992 [35]	Population based study Sample size: 132 Follow-up: 5 years Post operatively Equal number of male and female patient's	Age, site of tumor, family history, interval between diagnosis and surgery, interval between symptoms and diagnosis, stage, pattern of growth, extent of fibrosis.	Factors which were significant in univariate analysis analysis were: age, pattern of growth and extent of fibrosis. However, the only significant factor related to prognosis in multivariate analysis was stage. Staging being the factor significant in multivariate analysis confirms the importance of stage in predicting survival in CRC.
Prall, F et al. 2004 [36]	Retrospective study with review of clinical charts and medical records. Sample size: 184 All stages of CRC Follow-up history was obtained for 5 years postoperatively.	Clinical and immunohistochemical tumor marker levels were estimated (p53, p27, p21 levels).	In Cox multivariate analysis growth pattern of tumor, lymphohistiocytic response, lymphatic permeation, and extramural venous were found to be significant when tested against UICC stage. Mitotic index added to the prognostic information to TNM stage in multivariate analysis. Weakness of study: Effect of other prognostic markers in combination with tumor markers (p53, p27, p21) was not done on CRC patients in predicting survival.
Rene A. C. et al. 2004 [39]	Retrospective study Sample size: 96 patients From 1950 to 1990. Follow-up: 3 years All stages of CRC	Age, Gender, Karnofsky performance at admission, site of tumor, type of surgery, preoperative albumin level, number of resected organs/ structure, hospital stay, grade, stage, lymph node status, lymphatic invasion, perineural invasion, tumor margins, clinical presentation.	Study determine better survival (5-year) rate for patients undergoing curative surgery (58.3%) than palliative surgeries (0%). Multivariate analysis showed Karnofsky performance status was strongly related to the risk of postoperative complications and postoperative deaths. Factors which were related to better prognosis for CRC, were grade I and II, non metastatic LN, absence of vascular, lymphatic, perineural invasion, Poor prognostic factors were lympn node status and adjacent organ infiltration. Strengths: A significant decrease of postoperative deaths and complications from 1950 to 1990 could be due to improvement in staging methods, pre and post operative and

Author, year	Study Features	Factors Analyzed	Findings and Strengths/Weakness
Ratto, Carlo et al, 1998 [38]	Prospective study Sample size: 853 Male and female patients With CRC of all stages Follow-up: Every 3months-1st year Every 6 months in 2nd-3rd year, once per year thereafter	Gender, Age (</> 60 years) Location of primary history of bowel obstruction Tumor size, Stage, Grade, LN involvement, Metastasis, Preoperative CEA, tumor ploidy and vascular invasion.	Factors significant and independently influencing outcome were gender, lymph node involvement, history of metastasis and bowel obstruction. Strengths: Large sample size and good power, including tumor markers similar results were observed in other studies too. Weakness: Did not look into molecular markers on survival prediction.
Wang et al. 2004 [50]	Retrospective study with review of clinical charts. Sample size: 74 Dukes' B patients Follow-up history for 3 years	Clinicopathological information was collected. The sample was divided into 2 groups –testing and training set to select gene markers using the training set and build a prognostic signature and validate it on the testing set.	The study demonstrates the potential of DNA microarray based gene expression pattern for the prediction of patient's outcome in colon cancer. This is likely to have an impact on the current clinical practice for the eligibility of adjuvant chemotherapy on treatment of Dukes' B colon cancer patients. The study identified 23-gene signature that predicts recurrence in Dukes' B patients. The signature was validated in 36 independent patients. The overall accuracy was 78%.

CHAPTER THREE

METHODS

3.1 IRB Approval

Prior to the initiation of the research, the study was approved by the Institutional Review Board (IRB) of the University of South Florida.

3.2 Study Design and Study Population

Retrospective cohort study with the review of clinical charts of 116 CRC patients; who underwent curative surgery from 1/5/1993 to 5/1/2002, at Moffitt Cancer Center and had a follow-up history up to the date of last contact or death.

Initial study started with the selection of 400 frozen tumor specimens of patients with any of the four colorectal cancer tumor staging from Moffitt Cancer Center Tumor Bank (Tampa, Florida); such that all patients had postoperative follow-up for at least 36 months (because majority of patients who would die of CRC, will have done so by then) determined by Moffitt Cancer Registry. The sample size was reduced to 116 as gene expression profiling using mRNA technique was done to only 116 samples of the 400 CRC tumor samples.

Retrospective review of inpatient charts, including operative and pathologic reports of the 116 stages II and III CRC was done to obtain clinicopathological data. Data was

collected on patient's demographics, clinical, pathological and survival data. A copy of the data entry form is provided in Appendix 1. Once the data was collected, in order to maintain the patient confidentiality each patient was given a unique identification number generated by SAS Randomisation. These data were entered on a standardized data entry form and entered into a database. Based on the exclusion criteria the final sample size was reduced to 58 patients consisting of confirmed stage II or stage III primary colorectal cancer only.

3.2.1 ***Inclusion Criteria:***

Confirmed Stage II and Stage III primary colorectal cancer patients, who did not receive any neo-adjuvant therapy preoperatively and, who had a follow-up of at least 36 months and also had gene expression profiling done using the mRNA technique.

3.2.2 ***Exclusion criteria:***

As the study addressed only primary colorectal cancer 41 samples were excluded from the 116, who were Stage I or Stage IV colorectal cancer patients and also those for whom the cancer site were not colorectal such as (abdominal wall, periaortic lymph nodes, mesenteric lymph node, lung, liver, bladder, retroperitoneal lymph nodes, kidney, small intestine) after reviewing the histopthological reports. Five patients had multiple primaries at the time of surgery and were assessed by the histological grade to determine the earliest primary and only the earliest primary was considered in the sample size and the others were excluded. Nine patients who received preoperative neoadjuvant therapy such as radiated rectal cancers were dropped from the study, as preoperative neoadjuvant therapy would affect gene expression profiles, hence cannot be used as predictor. Histopathological information could not be reconfirmed for three patients and were excluded from the sample size. Clinical and histoathological characteristics of patients and their tumors are summarized in tables 6 and 7.

3.3 Gene expression profiles (GEP) and Molecular classification:

The Gastrointestinal Tumor Program, Bioinformatics and Biostatistics Core of Moffitt Cancer Center and Research Institute performed the GEP of the 116 colorectal cancer specimens.

The microarray data was analyzed using Significance Analysis of Microarrays program (SAM) with censored survival data [11]. SAM identified genes most correlated with survival time (3-year survival) and used permutation analysis to estimate the False Discovery Rate (FDR). The first analysis was done for stage II and III CRC patients only. The data was preprocessed using Robust microarray (RMA) and the censored survival time was calculated for each sample. In this work, censoring occurred at time of last follow-up or for death in which there was no evidence of disease. SAM was used to calculate genes that correlated to survival time. A threshold yielding a 10% median FDR was selected, which resulted in 53 overexpressed genes. The analysis was repeated for cases of all stages (stage I, II, III, IV). The data was again preprocessed using RMA and survival time was calculated. SAM was again used to select genes. A median FDR of 2.5% (the minimum FDR possible) was selected, resulting in 30 genes.

The gene expression for all 30 genes was extracted and the data was clustered. The Gene Cluster 3.0 program for clustering and Java Treeview for visualizing the heatmap was used. The genes were median centered and normalized prior to clustering. Hierarchical clustering with the un-centered correlation similarity metric and complete linkage was used. The resulting dendrogram can be seen below. This resulted in two main groupings (Clusters) of the sample (columns) in the data, which were chosen as the “prognostic” groups for this work. These groups were listed as Cluster 1(Low risk) and Cluster 2 (High risk) based on 30-

gene cluster. Cluster Analysis of 30 SAM selected genes were performed. Red color represents over expressed genes relative to green, underexpressed genes.

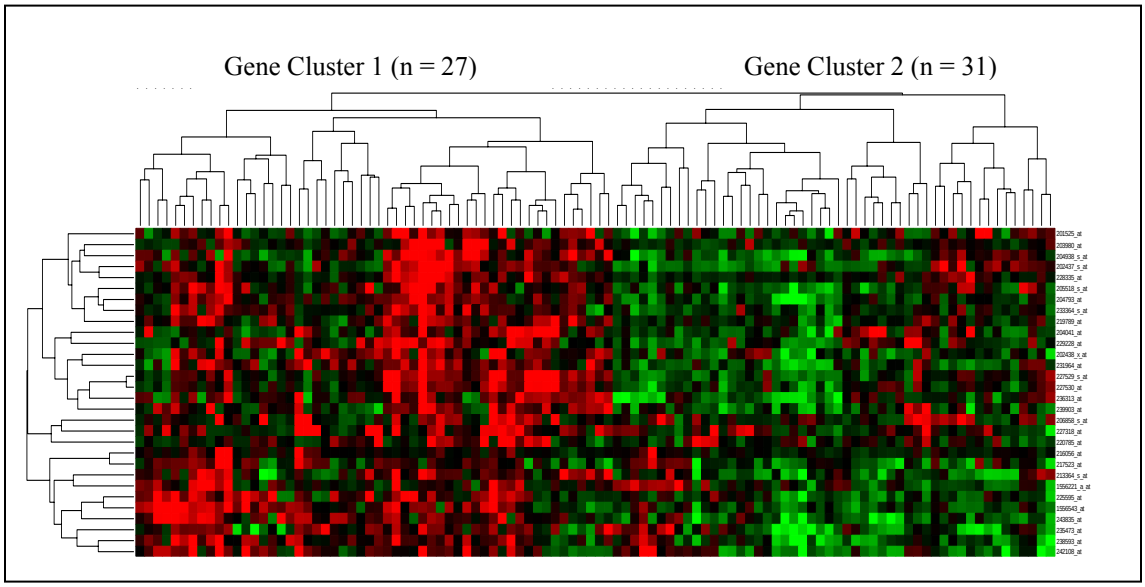


Fig 2. Dendrogram

3.4 Censoring

Patients were considered for censoring when incomplete information was available about their survival time that is if they were lost to follow-up, alive with no evidence of CRC at the date of last contact or who died but not due to CRC. Patients who did not die of CRC were considered censored and patients who died of colorectal cancer at the end of the study were considered as having experienced the outcome of interest (death) and were not censored.

3.5 Statistical Analysis

Univariate analysis and Kaplan-Meier procedure was done to find predictive value of each prognostic variable. Multivariate analysis using stepwise selection procedure was done to determine the effect of GEP in the presence of other clinical predictors. Final model selection was based on factors which met proportional hazard assumption and/or had biological or clinical significance.

CHAPTER FOUR

DATA ANALYSIS

4.1 Descriptive Analysis:

A total of 58 patients matched the inclusion criteria for the study and all patients had undergone curative surgery for cancer. Preliminary descriptive analyses were performed to characterize the sample. Frequency and percentages for each variable were obtained accounting for the missing values. Table 6 and 7 summarizes the clinical and pathological characteristics of patients and their tumors.

Age at diagnosis was calculated as date of surgery minus the date of birth. Mean, median and age range at diagnosis were determined. Age at diagnosis was divided into 3 groups: less than 50 years, 50-70 years and above 70 years. To further determine the influence of advanced age on survival, age at diagnosis was divided into tertiles of upper 1/3 rd age group and lower 2/3 rd age group.

Body mass index was calculated based on the height and weight at the time of surgery of index primary colorectal cancer and divided into 4 groups: less than 18.5 (underweight), 18.5-24.5 (normal), 25-29.9 (overweight) and above 30 (obese). To further assess the effect increased BMI on overall survival, BMI was divided into 2 groups: ≤ 25 and >25.

Location of primary cancer was merged into three groups: proximal colon (cecum, ascending colon, hepatic flexure and transverse colon), distal colon (splenic flexure,

descending colon and sigmoid colon) and rectum and rectosigmoid junction.

Tumor stage at the time of primary cancer was regrouped into two groups with Stage IIA and IIB grouped as stage II, and stage IIIA, IIIB and IIIC grouped as stage III, because there were very few patients in stage IIB and III A (table 6).

Mean and range of Lymph nodes taken at the time of resection of primary tumor were determined and also grouped into ≤12 lymph nodes and greater than 12 lymph nodes.

Survival differences for regional lymph node (LN) involvement were assessed for: No LN involvement, 1-3 group of LN's, ≥ 4 group of LN. Further analysis was done by collapsing the regional lymph node involvement into: lymph node positive group and lymph node negative group.

Histological differentiation was categorized based on the grading: well-differentiated, moderately differentiated and poor differentiation. Since, the number of patients in the well-differentiated group were few, these patients were merged with patients with moderate differentiation to compare survival with poor differentiated tumors. For preoperative CEA level was divided into < 5ng/ml and >5 ng/ml.

Information on the vital status and cancer status at the date of last contact was obtained through the Moffitt cancer registry and summarized in table 5.

4.2 Statistical Analysis

Once the descriptive analysis was performed on the study population, the research questions were analyzed using SAS 9.0 software.

Survival distribution were estimated using the Kaplan-Meier procedure to assess the influence of individual predictors (clinical parameters and GEP) on the overall survival of patients in the study, to determine which parameters met the proportional hazard assumption

and if there was significant difference between the strata for each parameter (predictor).

Univariate analysis was done to determine statistically significant prognostic factors, the hazard ratios, confidence interval and p-values. Log-rank test was used to compare survival estimates for each stratified variable. Log-rank test was employed to evaluate the null hypothesis being tested that no overall survival difference exits between the strata for each variable. Hypotheses were tested using p-values of 0.05 for statistical significance.

Multivariable analysis using stepwise Cox regression analysis was used to determine independent significant prognostic variables. The likelihood ratio test based on maximum partial likelihood estimates was used to eliminate confounding variables from the model. Variables were considered eligible for removal if the likelihood ratio test significance level was >0.05.

The final model contained factors that met the proportional hazard assumption and had a biological or clinical significance in predicting colorectal cancer survival. Survival rates were estimated by Cox proportional hazard model. Stepwise regression methods were used to build statistical model for the association of prognostic factors with overall survival. Time dependent hazard ratios were estimated. By observing HR, 95% CI for each factor, and the change in the log likelihood statistic, it was ascertained which variables should remain in the final model.

CHAPTER FIVE

RESULTS

5.1 Descriptive Analysis

5.1.1 Demographic data

Table 6 displays the demographic (clinical) characteristics of the 58 CRC subjects in the study.

The median survival time for the study sample was 75.36 years and the 3-year and 5-year overall survival estimates were 72.41% (SE=5.8%) and 55.17% (SE=6.7%), respectively (Fig 3).

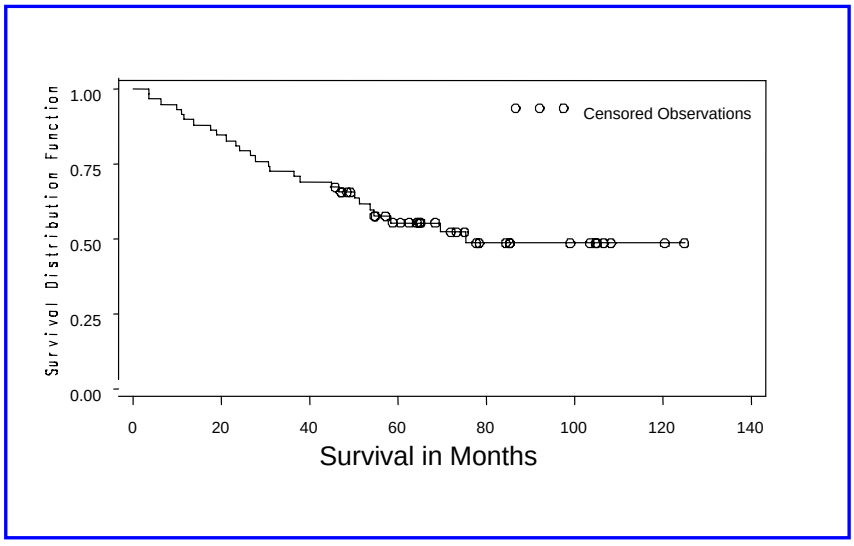


Fig 3 Kaplan Meier Curve for Overall Survival of the Study Population

Female patients had higher incidence rate of CRC (53%) than males (47%). In the study sample, CRC was more commonly seen among Caucasians (81%), followed by African Americans (4%) and Hispanics (2%). BMI of majority of the patients (58%) was above normal (>25) at surgery. History of weight loss at the time of surgery was reported by 45% of CRC patients. Positive family history for CRC in first degree relative was present in only 8 of the 58 patients.

Based on age, 10% were in the <50 years age group, 38% were 50-70 years age group and 52 % were >70 years, at diagnosis. The median survival time for <50 years, 50-70 years and >70 years at surgery was 4.8 years, 5.12 years and 3 years, respectively. Majority of the individuals diagnosed with CRC were of the between 70-79 years of age (35%).

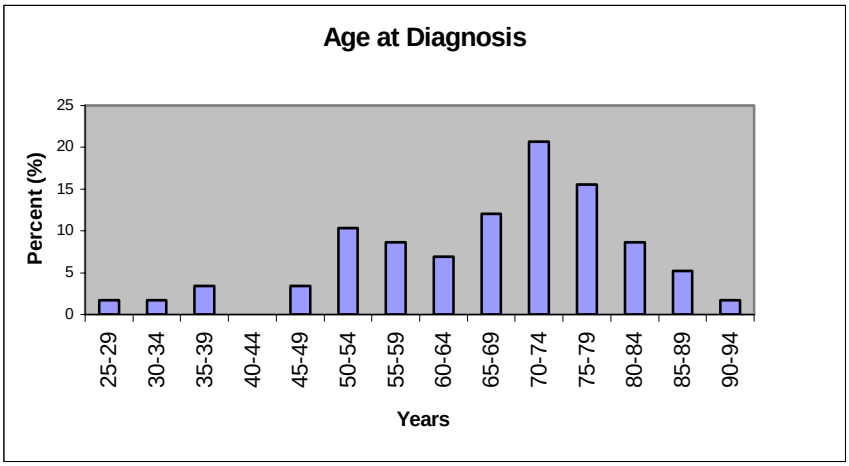


Fig.4 Age distribution of the study population

At the end of 3 year follow-up 42 patients (72%) were alive and 16 patients were dead (28%). At the date of last contact 29 patients (50%) were alive with no evidence of cancer, 18 patients (31%) were dead with evidence of cancer, 4 patients (7%) died but with no evidence of cancer and cancer status could not be ascertained for 7 patients (12%).

Table 4 Survival Status at the end of 3-year follow-up

Vital Status	n	%
Alive	42	72
Dead	16	28

Table 5 Cancer Status at the Date of Last Contact

	Cancer Status at Last Contact		
Vital Status at Last Contact	**No evidence of Cancer**	**Evidence of Cancer**	**Unknown**
Alive (n=31)	29	0	2
Dead (n=27)	4	18	5

Table 6 Clinical Characteristics of Patients

Characteristics		No. of Patients	(%)
Total Sample Size (n)		58	
Age at Diagnosis, years			
Mean	67.17 years		
Median	70.47 years		
Age Range	26.97 - 92.41		
≤ 50 years		6	(10)
50-70 years		22	(38)
≥ 70 years		30	(52)
Upper 1/3rd Age group (≥ 74.75 years)		19	(33)
Lower 2/3rd Age group (< 74.75 years)		39	(67)
Gender			
Female		31	(53)
Male		27	(47)
Race			
Caucasian		47	(81)
African American		4	(7)
Hispanic		2	(3)
Asian		1	(2)
Other/Unknown		4	(7)
Body Mass Index			
<18.5 (underweight)		1	(2)
18.5-24.9 (normal)		18	(31)
25.9-29.9 (overweight)		21	(36)
≥ 30 (obese)		13	(22)
Unknown		5	(9)
History of Smoking			
Ever		25	(43)
Never		28	(48)
Unknown		5	(9)
Ever Female (n=31)		15	(48)
Ever Male (n=27)		10	(37)
Family History of Cancer			
Present		34	(58)
Absent		12	(21)
Unknown		12	(21)
Family History of Colorectal Cancer			
Present		8	(14)
Absent		38	(65)
Unknown		12	(21)

5.1.2 Histopathological Characteristics:

Table 7 displays the histopathologic characteristics of the sample population as described below.

The three most common site descriptions of CRC were sigmoid colon (27%), ascending colon (25%) and rectosigmoid junction (17%). Based on the AJCC classification-6th edition, in the study sample, 26% were stage IIA, 3% were stage IIB, 2% were stage IIIB, 14% were stage IIIB and 13% were stage IIIC colorectal cancers. Stage IIA and IIB when combined together constituted 46% had Stage II CRC and stage IIIA, IIIB, IIIC when combined together constituted 56% of colorectal cancers in the study. Of the patients observed, 47% had no regional lymph node involvement, 29% had 1-3 group of regional node positivity and 21% had more than 4 group of lymph node involvement.

Moderate differentiation of the tumor (76%) was the most common histological grade in the sample population followed by poor differentiation (16%) and well-differentiated tumor (7%). Pretreatment CEA levels were unknown for 59% of the patients. 17% had preoperative CEA level more than 5 ng/ml and 24% had levels < 5ng/ml. Based on the gene cluster analysis 47% were classified as low-risk group and 53% were of the high-risk group.Other histological features observed were vascular invasion (13 patients), lymphatic invasion (7 patients) and perineural invasion (3 patients). In addition to adenocarcinomatous histology, mucinous histology of tumor was observed in 5 patients. Perineural, lymphatic and vascular invasion was seen in 3 patients, 7 patients and 13 patients, respectively. Resected margins were not clear of cancer after surgery with curative intent in 4 patients only.

Table 7 Histopathological Characteristics of Patients

Characteristics	No. of Patients	(%)
Location of Primary		
Proximal colon	21	(36)
Cecum & Ileocecal Valve	2	
Acending colon	15	
Hepatic flexure	0	
Transverse colon	4	
Distal Colon	20	(34)
Splenic flexure	0	
Descending colon	4	
Sigmoid Colon	16	
Rectum and rectosigmoid junction	17	(29)
Rectosigmoid junction	10	
Rectum	7	
Stage		
IIA (T3 N0 M0)	26	(45)
IIB (T4 N0 M0)	3	(5)
IIIA (T1-2 N1 M0)	2	(3)
IIIB (T3-4 N1 M0)	14	(24)
IIIC (Any T N2 M0)	13	(23)
Regional Lymph Node Metastasis		
No Lymph node involvement (N0)	27	(47)
1-3 Lymph node involvement (N1)	17	(29)
≥ 4 Lymph node involvement (N2)	12	(21)
Could not be assessed (Nx)	2	(3)
Total Lymph Nodes Resected		
Mean Lymph nodes resected	13.63	
Range of Lymph nodes examined	2 - 35	
≤ 12	28	(48)
>12	30	(52)
Grade/Differentiation		
Well	4	(7)
Moderately	44	(76)
Poor	9	(16)
Unknown	1	(2)
Preoperative CEA* level		
CEA ≥ 5.0 ng/ml	10	(17)
CEA < 5.0 ng/ml	14	(24)
Unknown	34	(59)
Molecular Classification		
Low risk (Cluster 1)	27	(47)
High risk (Cluster 2)	31	(53)
Other Histological Features		
Mucinous Histology	5	
Signet Ring Histology	0	
Perineural Invasion	3	
Lymphatic Invasion	7	
Vascular Invasion	13	
Resected Margins Not Clear of Cancer	4	

* CEA: Carcinoembriyonic Antigen

5.2 Statistical Analysis

The research questions are restated in this section to facilitate coherence and readability.

The first two research questions are stated as follows:

1. What is the predictive value of traditional clinical predictors- age, sex, race, stage and grade in determining prognosis (survival) in Stage II and III CRC patients?
2. What is the predictive value of gene expression profiles (GEP) by itself in this sample?

5.2.1 Univariate Analysis:

Table 8 lists the prognostic variables, their hazard ratios, 95% CI and p-values for the comparisons of interest. 3-year and 5-year survival rates of the study sample are listed in table 9.

General Hypothesis:

H_0: No difference between the survival distribution for Group1 and Group 2, (S_1 (t) =S_2 (t) for all t >0)

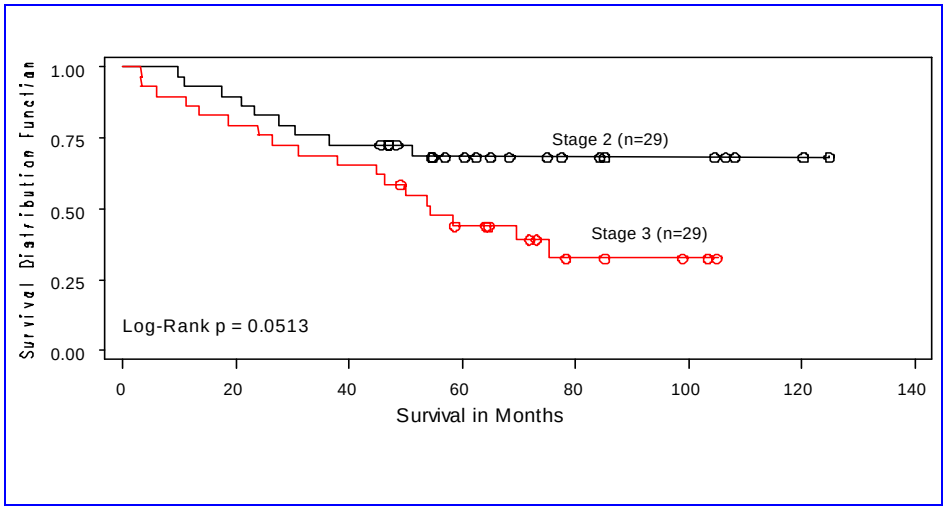
H_A: There is difference in survival distribution between Group1and Group 2, (S_1 (t) ≠ S_2 (t) for some t >0)

Where, Group1: Survival function S_1 (t)
Group 2: Survival function S_2 (t)

If the p-value < alpha (0.05) then H_0 is rejected and concluded that group1 and group2 have different survival distributions.

If p-value ≥ alpha, H_0 is retained and conclude that there no sufficient evidence in the data suggesting the opposite is true.

A) TNM staging:



	3-year rate (%)	5-year rate(%)	β	HR	p-value
Stage III	68.97	43.97	0.77	2.17	0.0513
Stage II	75.86	67.65			

Fig 5 Kaplan Meier Survival Curves for TNM Staging of Tumor

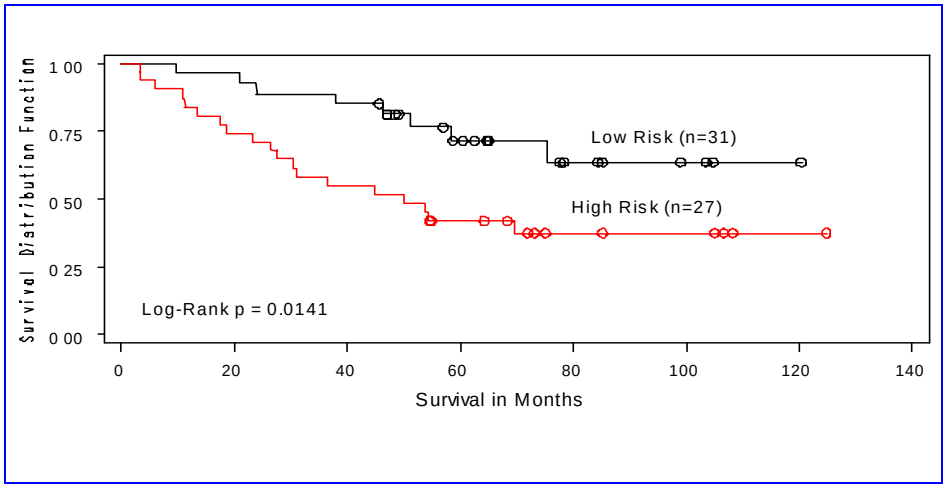
H_0: No difference between the survival distribution for stage III and stage II CRC, (S_1 (t) =S_2 (t) for all t >0).

H_A: There is difference in survival distribution between stage III and stage II CRC, (S_1 (t) ≠ S_2 (t) for some t >0).

Patients with stage II CRC had better 3-year and 5-year survival rates compared to patients with stage III CRC. The p-value < alpha (0.05) hence, H_0 is rejected and concluded that stage III and stage II patients have different survival distributions.

The hazard ratio for patients with stage III CRC was twice as high than that for patients with stage II cancers.

B) Molecular Risk:



	3-year rate (%)	5-year rate(%)	β	HR	p-value
High Risk	58.06	41.29	0.987	2.68	0.0141
Low Risk	88.89	71.46			

Fig 6 Kaplan Meier Survival Curves for Molecular Risk

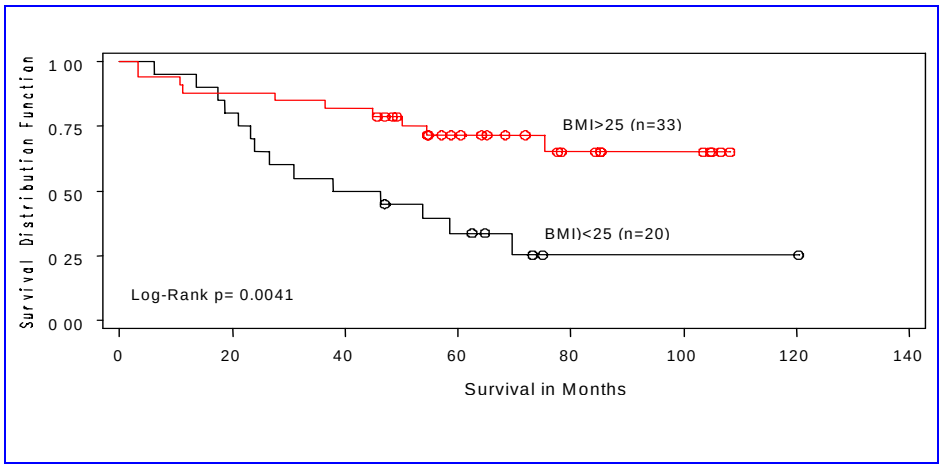
H_0: No difference between the survival distribution for high molecular risk patients and low molecular risk patients, (S_1 (t) =S_2 (t) for all t >0).

H_A: There is difference in survival distribution between high and low Molecular risk patients, (S_1 (t) ≠ S_2 (t) for some t >0).

Patients of low molecular risk had better 3-year and 5-year survival rates compared to patients of the high molecular risk cluster. The p-value < alpha (0.05) hence, H_0 is rejected and concluded that low molecular risk and high molecular risk patients have different survival distributions.

According to this univariate analysis, the hazard for death for patients in the high-risk gene expression group was double than that for patients in the low-risk gene expression group and it is statistically significant.

C) BMI:



	3-year rate (%)	5-year rate(%)	β	HR	p-value
BMI >25	84.85	71.46	-1.15	0.317	0.00041
BMI<25	55	33.39			

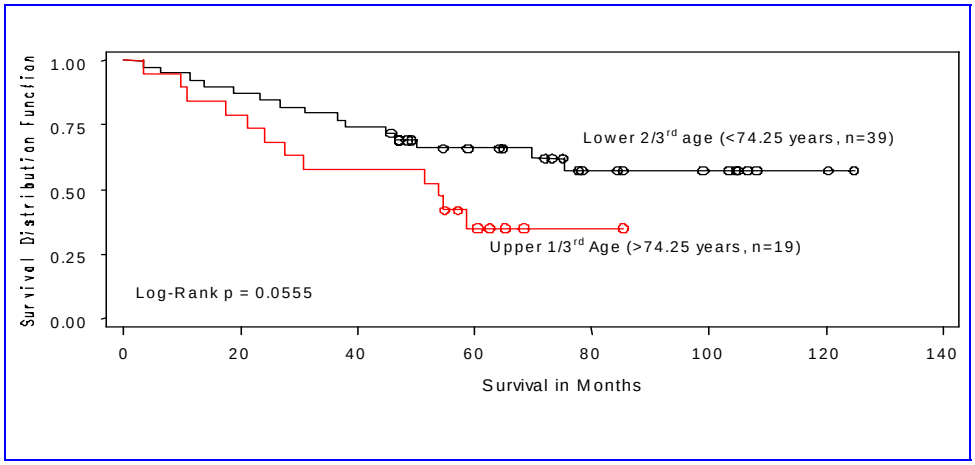
Fig. 7 Kaplan Meier Survival Curves for BMI of Patients

H_0: No difference between the survival distribution between patients with body mass index >25 and <25, (S_1 (t) =S_2 (t) for all t >0).

H_A: There is difference in survival distribution between between patients with body mass index >25 and <25, (S_1 (t) ≠ S_2 (t) for some t >0).

Patients with BMI > 25 were found to have better survival and lower hazard for death than patients with BMI <25 and an inverse relationship were seen between BMI and mortality. The hazard for death for patients in the BMI >25 group was 32% lower than that for patients in the BMI<25 group

D) Age



	3-year rate (%)	5-year rate(%)	β	HR	p-value
Upper 1/3rd	57.89	34.74	0.74	2.09	0.0555
Lower 2/3rd	79.49	65.59			

Fig. 8 Kaplan Meier Survival Curves for Age of Patients

Patients in the lower 2/3rd age group had better 3-year and 5-year survival rates and lower hazard for death than patients who were in the upper 1/3rd age group. The p-value < alpha (0.05) hence, H_0 is rejected and concluded that older age group patients (upper 1/3 rd tertile) have different survival distributions compared to younger patients (lower 2/3rd tertile).

The hazard for death for patients in the upper 1/3rd age group was double than that for patients in the lower 2/3rd age group.

Gender (HR: 1.58, p=0.2407), race (HR: 0.83, p=0.6932), family history of CRC (HR: 0.79, p=0.6568), grade of tumor (HR: 1.42, p=0.4816), preoperative CEA levels (HR: 2.3, p=0.1461), location of tumor (HR: 0.73, p=0.2092), and total lymph nodes examined (HR: 0.84, p=0.6607), were not statistically significantly associated with survival (Table 8).

Table 9 summarizes the 3-year and 5-year survival estimates of the study population following surgery with curative intent based on the prognostic variable. Decreased survival rates were observed for patients in the younger age group (<50 years) and the older age group (>70 years) as compared to those in the middle age (50-70 years), with approximately 40-50 %, 5-year survival rate for patients in the young and old age group following diagnosis of CRC.

Patients who were overweight and obese (BMI>25) at the time of surgery had better survival rates compared to patients who were normal or underweight.

Higher survival rates at year 3 and year 5 were observed for patients in the low molecular risk group as compared to those in the high risk group, which were statistically significant. Specifically Stage 2 with low molecular risk and stage 3 with low molecular risk.

Patients with stage II CRC had better survival rates than Stage III patients and as cancer progressed, the survival rates showed a trend with decreased survival with time. Similar results were observed for lymph node involvement, with better survival rates for patients with no nodal involvement as compared to <3 group and > 3 group of nodes positive patients and survival rate decreased with time.

Table 8 Univariate Analysis

Variable	Parameter Estimate	Chi-Square	Hazard Ratio (95% CI)	p-value
Stage (Stage 3 vs Stage 2)	0.776	3.61	2.17 (0.98-4.8)	0.06
Molecular Risk –GEP (High risk vs Low risk)	0.987	5.46	2.68 (1.2-6.15)	0.02
Age (Upper 1/3rd vs Lower 2/3rd age group)	0.740	3.51	2.09 (0.9-4.5)	0.06
Gender (Female vs Male)	0.456	1.35	1.58 (0.7-3.4)	0.24
Race (Caucasian vs Other)	-0.183	0.16	0.83 (0.3-2.1)	0.64
BMI (Overweight and obese vs Normal)	-1.150	7.45	0.317 (0.1-0.7)	0.006
Family History of Colorectal Cancer	-0.276	0.20	0.79 (0.2-2.6)	0.66
Grade/Differentiation (Poor vs moderate and well)	0.352	0.49	1.42 (0.5-3.8)	0.48
Regional Lymph Node Involvement (N1-N2 vs N0)	0.784	3.40	2.09 (0.9-5.0)	0.06
Location of Primary Tumor	-0.309	1.57	0.73 (0.5-1.2)	0.21
Preoperative CEA Level (>5ng/ml vs ≤5 ng/ml)	0.834	1.99	2.3 (0.7-7.3)	0.16

CEA: Carcinoembriyonic antigen, GEP: gene expression profiles.

Table 9: 3-year and 5-year Survival Estimates of the Study Population Following Curative Surgery Based on the Prognostic Variable

Variable	No. of Patients	(%)	3-year survival rate (%)	5-year survival rate (%)
Age				
≤ 50 years	6	10	66.67	50
50-70 years	22	38	86.36	75.4
≥ 70 years	30	52	63.33	42.04
BMI				
<18.5 (underweight)	1	2	50	-
18.5-24.9 (normal)	18	34	55.56	31.33
25.9-29.9 (overweight)	21	40	90	72.73
≥ 30 (obese)	13	25	76.92	68.83
>25	34	64	84.85	71.08
<25	19	36	55	33.39
Family History of CRC				
Yes	8	17	75	-
No	38	82	76.32	56
Molecular Risk				
Low	27	47	88.89	71.46
High	31	53	58.06	41.29
Stage				
Stage II	29	50	75.86	67.65
Stage III	29	50	68.97	43.97
IIA (T3 N0 M0)	26	45	76.92	67.57
IIB (T4 N0 M0)	3	5	66.67	66.67
IIIA (T1-2 N1 M0)	2	3	50	50
IIIB (T3-4 N1 M0)	14	24	78.57	48.21
IIIC (Any T N2 M0)	13	23	61.54	38.46
Lymph Node Involvement				
N0	27	48	77.78	68.87
N1	17	30	76.47	51.73
N2	12	21	58.33	33.33
Grade				
Well	4	7	75	NA*
Moderately	44	77	77.27	55.24
Poor	9	16	55.56	44.44
Total LN Resected				
>12	28	48	73.33	57
<12	30	52	71.43	50

NA*: Censored, LN: Lymph Node, CRC: colorectal cancer

5.2.2 Multivariate analysis (Stepwise Model Selection):

Research Question

3. What does the addition of molecular classification (GEP) together with clinical parameters (age, sex, family history, BMI, grade, location of tumor, preoperative CEA levels) contribute to the usual clinical predictiveness of outcome (survival) in stage II and III colorectal cancer?

General Hypothesis:

H_0: $\beta i = 0$ (no difference of covariates)

H_A: $\beta i \neq 0$ (covariates influence survival)

where, βi is parameter estimate for multiple variables.

If p-value >0.05, then H_0 is not rejected.

Is based on the Cox proportional hazard model: $h(t,x) = h(t_0) \exp(\beta x)$.

This model allowed the estimation of the effect of each covariate in the presence of the others. The hazard ratio for each variable is adjusted for the effects of all of the other variables in the multivariate model.

The aim of the study was to establish whether GEP with clinipathological variables provides better prognostic information for patients with stage II and III CRC in addition to that afforded by staging alone.

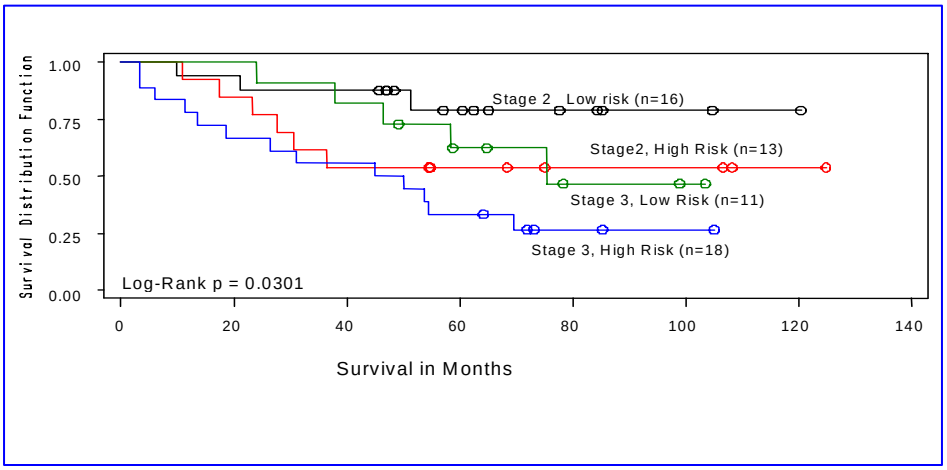
In order to address this multivariate analysis using stepwise Cox Regression method was used, with threshold of 0.05. Factors that were found to be significant in univariate analysis were tested for independent statistical significance in multivariate analysis. Three

prognostic variables: molecular risk, BMI and age were found to be independent predictors of overall survival (Table 10). Staging of tumor was found to be statistically not significant on multivariate model; however, it was retained in the final model in order to determine the association of other factors in the presence of clinical staging of tumor. None of the other variables were found to be significant. The likelihood ratio test based on maximum partial likelihood estimates was used for eliminating confounding. The results of the multivariate analysis for overall patient survival are shown in table 10.

Table 10: Significant Prognostic Risk Factors for Mortality (Overall Patient Survival) Determined by Multivariate Analysis

Variable	Parameter Estimate	Chi-Square	Hazard Ratio (95% CI)	p-value
Stage (Stage 3 vs Stage 2)	0.897	2.78	2.45 (0.85-7.04)	0.09
Molecular Risk –GEP (High risk vs Low risk)	1.343	5.26	3.83 (1.22-12.06)	0.02
Age (Upper 1/3rd vs Lower 2/3rd age group)	1.313	5.21	3.72(1.2-11.49)	0.02
BMI (overweight and obese vs normal)	-1.22	5.78	0.29 (0.11-0.79)	0.016

E. Multivariate analysis of Molecular Risk and Clinical Stage:



	3-year rate (%)	5-year rate (%)
Stage II, Low Mol. Risk (n=16)	86.67	77.04
Stage II, High Mol. Risk (n=13)	58.33	50
Stage III, Low Mol. Risk (n=11)	89.34	65.63
Stage III, High Mol. Risk (n=18)	57.89	36.84

Fig. 9 Kaplan Meier Survival Curves for Clinical Stage and Molecular risk combined.

According to fig 9, that there is difference in overall survival when molecular risk and clinical staging were combined in multivariate analysis, and patients of the low molecular risk group had better survival outcome than patients with high molecular risk group within stage II and stage III clinical staging. Combining clinical stage with molecular risk classification we were able to differentiate patients into different strata.

5.2.3 Predictor Model

Factors, which were prognostically significant (biological or clinical) and met the proportional hazard model by the assessment of log-minus log survival plot were included in the final model. Stepwise procedure was used and based on the assessment of log likelihood statistic final predictor model was selected.

Hypothesis:

Ho: Model with more variables (predicted model) is similar to the model with fewer variables (basic model).

HA: The two models are different.

Basic Model: h (t,x) = $h(t_0)$ exp [(β1 (stage)

Predicted model: h (t,x) = $h(t_0)$ exp [(β1 (stage)+ β2 (molecular risk)+ β3 (age)+ β4(body mass index)].

The Cox proportional hazard model: h (t,x) = $h(t_0)$ exp (β x), where, $h(t_0)$ is base line hazard function, h (t,x) is hazard at time t, β is parameter estimate for the variable and x is value for each variable denoted by "1" and "0" for presence and absence of the prognostic factor.

Based on the Cox Proportional Hazard Model the hazard ratios observed for stage, molecular risk, age and body mass are:

h (t,x) = $h(t_0)$ exp [(β1 (stage)+ β2 (molecular risk)+ β3 (age)+ β 4(body mass index)].

h (t, stage III) = exp [(0.897* (1))+ (1.343* (0))+ (1.313*(0))+ (-1.22* (0))]= 2.4522;

h (t, stage III & high molecular risk) = exp[(0.897* (1))+ (1.343* (1))+ (1.313*(0))+ (-1.22* (0))]= 9.39;

h (t, stageIII & high mol risk & old age& BMI >25) = exp [(0.897* (1))+ (1.343* (1))+ (1.313*(1))+ (-1.22* (1))]= 10.27;

The hazard ratio for stage alone= 2.452. And the addition of molecular risk to stage increased the hazard to 9.39 and for the final model containing stage, molecular risk, age and

BMI was 10.27. So, there is evidence that the advanced tumor stage (stage III), high molecular risk, old age and body mass index >25 result in higher the hazard ratio and decreased survival.

Predictor model with clinical stage, molecular risk, age and BMI was compared with the model containing clinical stage alone. The test statistic is equal to the difference between the –2LogL value in the model fit statistic for both models. The test statistic = 125.256-112.659= 12.597, which is greater than Chi-square, 3 degree of freedom and at 95% significance = 7.815.

By observing HR, 95% CI for each factor, and the change in the log likelihood statistic, it was ascertained that clinical stage, molecular risk, age and BMI should remain in the final model. Model containing clinical stage in conjunction with molecular risk, BMI, and age was a stronger indicator of clinical outcome (p= 0.0056) than model with clinical stage alone.

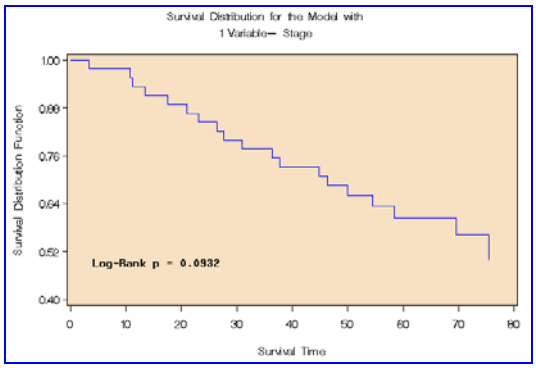


Fig.10 Survival Distribution for Prediction Model with Stage

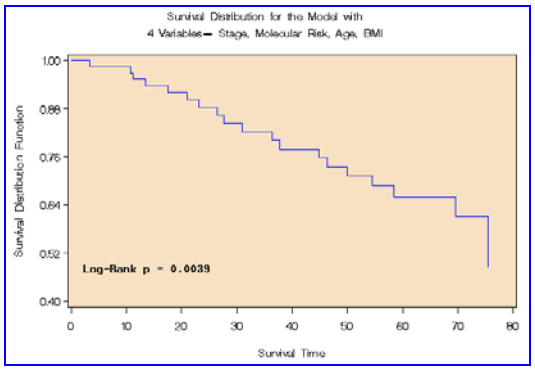


Fig.11 Survival Distribution for Prediction Model Containing Stage, Molecular Risk, Age and BMI

CHAPTER SIX

DISCUSSION

6.1 Important Findings

The mean age at diagnosis was 67.17 years (SE= 14.32%). The median age was observed to be 70.47 years, and the age range at diagnosis was 26.97 to 92.41 years. 33% of patients were in the upper 1/3rd age tertile and 67% were in the lower 2/3rd age tertile. The median survival after surgery for Stage IIA and IIB was 4.6 and 5 years, respectively. And for stage III A, IIIB and IIIC the median survival was 3.7, 4.5 and 4.2 years, respectively. The median survival in years following surgery for the low molecular risk group was 4.9 years and 4.1 years for the high risk group.

The risk of mortality increased with advance in tumor stage and with age. A progressive decrease in 5-year survival rates was evident as cancer progressed from TNM stage II A to stage IIIC (table 9). Low molecular risk patients had better survival outcome and 5-year survival rates compared to high molecular risk patient.

The inverse relationship seen with between BMI >25 and mortality could be due to the fact that most patients were overweight before cancer diagnosis and had significant weight loss from the time of diagnosis to surgery at which time BMI was recorded and were less healthy due to cancer. Hence, BMI <25 indicates significant weight loss in these patients and poor survival rates due to advance in cancer, co-morbid condition and poor

immune status. Patients with BMI >25 at surgery had 32% lower hazard for death than patients with BMI<25.

Although, family history of CRC did not show to be significantly associated with poor outcome, information on history of CRC in the family and the size of the patient cohort may be insufficient to identify positive family history of CRC as a prognostic factor in this study. A larger patient cohort will be required to definitely determine whether family history of CRC at diagnosis improves the accuracy of outcome prediction.

Significant difference in survival outcome was not observed based on the race, because majority of the patients in the study were Caucasians and could not be generalized to the population. Grade, an important factor in survival prediction was not found to be a statistically significant prognostic factor, as has been reported in previous studies. In some studies significant difference in survival was seen for preoperative CEA level, lymph node resection, regional lymph node involvement, however, in this study no significant difference in survival was observed.

Univariate analysis of the study showed 4 out of 10 variables were statistically significantly associated with overall survival (table 8). Among the clinical variables analyzed, gene expression profiling analysis distinguished two groups with significantly different survival outcomes (high risk vs low risk, HR: 2.68, p=0.0151, fig6). Borderline statistical significance was observed for TNM staging of tumor (stages III vs II, HR: 2.17, p=0.0513, fig5). The strong measure of association indicates that the significance is not due to chance alone.

Being older (>74.27 years) doubled the hazard for death compared to those younger (<74.25 years). Patients with BMI > 25 were found to have better survival and lower hazard

for death than patients with BMI <25 and an inverse relationship were seen between BMI and mortality. The hazard for death for patients in the BMI >25 group was 32% lower than that for patients in the BMI<25 group (Table 8). The hazard ratio for patients with lymph node involvement was two times higher than patients without nodal involvement.

The purpose of the study was to investigate the hypothesis that combined assessment of clinicopathological factors and gene expression profiles will allow increased accuracy of survival prediction for patients with CRC over the use of staging alone. A significant difference in survival was observed between the starta when staging and molecular risk were combined (p=0.0301, fig 9).

In this study it was found that GEP (molecular risk) are powerful predictors of survival independent of clinicopathological predictors. Stage, molecular risk, age and body mass index, have been identified as significant prognosticators of survival in this study. Combining molecular classification, BMI and age of patients to clinical staging, was a better indicator of clinical outcome than clinical staging alone.

Four prognostic variables: TNM staging, molecular risk, BMI and age were found to be independent predictors of overall survival in multivariate analysis.

6.2 Strengths and Weakness of the study:

Strength: Study had highly specific inclusion and exclusion criteria and limited to only primary colorectal cancer of stage II and III patients. All surgical and pathology reports for the study population were individually verified. To minimize error due to case abstraction, majority of the patient's charts were reviewed by same individual and for the few patients whose history was abstracted by other abstractors, the information collected by them was

cross validated. Certain risk factors that could affect the survival were controlled for in this study in multivariate analysis by stratifying the variables based on presence or absence of the risk factor (postmenopausal hormone use, smoking history, alcohol history). In the study only 10% of the patients were lost to follow-up. This type of study is useful to obtain information on colorectal cancer disease patterns over time and associate such patterns to the distribution of time to death among patients diagnosed with colorectal cancer. A study of this nature can be applied to public health practice by identifying factors that are important in cancer control and intervention, which can minimize mortality and morbidity by screening procedures and early diagnosis and treatment among the general population.

The study was able to determine the contribution of gene expression profile as a predictor along with other parameters in survival prediction in colorectal cancer patients. Weakness: The study is a retrospective cohort study from which information on patient's demographics and histopathological characteristic were gathered, which could contribute to misclassification. Due to the small sample size the study lacked power and hence significant differences in survival outcomes for certain prognostic variables such as preoperative CEA levels, grade and smoking and alcohol history, could not be determined. Cancer status at the end of follow up was ascertained from Moffitt Cancer Registry. Since 7 patients were lost to follow up and this is related to outcome (survival) and this could have caused attrition bias in the study. It is possible that differential misclassification of tumor stage and cause of death could have occurred either by recording deaths from other causes as deaths from colorectal cancer or vice versa.

Since most patients were Caucasians the results of this study cannot be generalized to the general population, as the study sample does not represent the population in general. The

influence of confounding factors such as smoking and alcohol consumption could not be determined in this study, as detailed information could not be obtained from clinical charts. Other factors, which could modify the risk of developing CRC such as diet, postmenopausal hormone use and physical activity, were not taken into consideration because of inadequate information. Certain factors that could influence survival outcome such as body mass index at the baseline (before cancer diagnosis) and the effect of treatment following surgery were not taken into consideration during data analysis.

6.3 Consistency with literature:

As in this study; similar results were seen in other studies that identified stage, molecular risk and age as significant prognosticators of survival. Data from literature and present study suggest clinical parameters, particularly stage [9, 39, 52], molecular risk [11,20, 22], age [1,2, 31] and body mass index [12,16,27] are related to patient survival rate and are most reliable prognostic factors.

Prognostic factors that were found to be significant in other studies but were not significant in this study were family history [2, 40], histological grade of tumor [21, 39, 52], and preoperative CEA levels [21, 30]. Differences between our study and others are mostly due to differences in the number of patients, lack of power, length of follow-up, and grouping of continuous variables also because the influence of comorbid conditions on survival which could not be analyzed in this study, family history, personal history of cancers, dietary history, multivitamin use, physical activity, screening history were not taken into consideration because of inadequate information.

6.4 Conclusion

Because of the frequency of the disease, the ability to identify high risk groups, better survival of patients with early-stage lesion, and the relative simplicity and accuracy of the screening tests, screening for colon cancer should be part of routine care for all adults at the age of 50 years, especially those with family history of colorectal cancer. Periodic evaluation following treatment of CRC helps to identify and manage recurrent disease.

6.5 Public Health Importance:

Colorectal cancer presents a major health problem with an annual estimated incidence rate of 106,680 new cases of colon cancer and 41,930 rectal cancers in 2006 and together they will cause 55,170 deaths in the US. Due to the long natural history of cancer there is time for early diagnosis and treatment before it reaches an advanced and incurable stage.

Since CRC is highly treatable and detected early, effective preventive approaches help reduce the morbidity and mortality associated with due the disease. Removal of premalignant lesion (adenoma) at the time of screening (colonoscopy) may be an effective form of primary prevention.

Certain behavioral factors which modify the risk of developing CRC such as dietary habits, physical activity, alcohol and cigarette smoking represent potential means of prevention.

According to the ACS a guideline listed in Appendix 2, screening helps in early detection and reduces the risk of dying from CRC. People who have no identifiable risk factors (other than age) should begin regular screening at the age of 50 years. Those who have a family history or other risk factors for CRC polyps or cancer should start screening at

a younger age and more frequently.

Following diagnosis and treatment for persons with CRC, periodic evaluation by monitoring preoperative CEA levels, aids in earlier identification and management of recurrent disease and cancer progression.

6.6 Future directions

For future studies it will be important to validate the gene expression cluster obtained in this study and the predictor model on an independent dataset. To conduct a prospective study of a larger sample size and include people of differences in ethnicity, study the prognostic ability of GEP and other factors such as family history, comorbid conditions, treatment following surgery, histopathological features (p27, p53, VEGF), vascular, perineural and lymphatic invasion and the influence of comorbid conditions on outcome prediction. Analyze the interaction between clinicopathological factors and GEP and other outcome measures such as time to relapse and treatment response. To determine if differences exists in CRC survival on the basis of screening between those persons who underwent regular screening and those who did not undergo regular screening as recommended by the American Cancer Society.

List of References

1. National Cancer Institute (NCI), 2005

2. American Cancer Society: Cancer Facts and Figures 2005. Atlanta, GA., 2005

3. Barrier, A. and S. Dutoit (2005). "Colon cancer prognosis prediction by gene expression profiling." Oncogene **24**: 6155-6164.

4. Berger, A. C. and e. al (2005). "Colon Cancer Survival is associated with decreasing ratio of metastatic to examined lymph nodes." J Clin Oncol **23**(34): 8706-12.

5. Bertucci, F., S. Salas, et al. (2004). "Gene expression profiling of colon cancer by DNA microarrays and correlation with histoclinical parameters." Oncogene **23**: 1377-1391.

6. Bostcik, R., J. Potter, et al. (1993). "Reduced risk of colon cancer with high intake of vitamin E: the Iowa womens' Health Study." Cancer Epidemiol Biomarkers Prev **5**(11): 897-900.

7. Charlson, M. E. and R. C. MacKenzie (1987). "A new method of classifying prognostic comorbidity in longitudinal studies: development and validation." J Chron Dis **40**(5): 373-383.

8. Concato, J. and A. R. Feinstein (1995). "Importance of events per independent variable in proportional hazards analysis. Background, goals, and general strategy." J Clin Epidemiology **48**(12): 1495-1501.

9. Compton CC, Greene FL (2004): The staging of colorectal cancer: 2004 and beyond. CA Cancer J Clin 54 (6): 295-308.

10. Doria-Rosa, V. P. et. al (2006). "Body mass index and the risk of death following the diagnosis of colorectal cancer in postmenopausal women (United States)." Cancer Causes and Control **17**: 63-70.

11. Eschrich, S. and T. J. Yeatman (2005). "Molecular staging for survival prediction of colorectal cancer patients." Journal of Clinical Oncology **23**(15): 3526-3535.

12. Frezza, E. E., M. S. Wachtel, et al. (2006). "Influence of obesity on the risk of developing colon cancer." Gut **55**: 285-291.

13. Friedenreich, C. (2001). "Physical activity and cancer prevention: from observational to intervention research." Cancer Epidemiol Biomarkers Prev **10**(4): 287-301.

14. Fuchs, C., E. Giovannucci, et al. (1999). "Dietary fiber and the risk of colorectal cancer and adenoma in women." N Engl J Med **340**(3): 169-76.

15. Galizia, G., E. Lieto, et al. (2004). "Determination of molecular marker expression can predict clinical outcome in colon carcinomas." Clinical Cancer Research **10**: 3490-3499.

16. Giovannucci, E. (2001). "Insulin, Insulin -like growth factors and colon cancer: a review of the evidence." J. Nutr **131**: 3109-20s.

17. Giovannucci, E., D. Nikolova, et al. (1998). "Multivitamin use, folate and colon cancer in women in the Nurses' Health Study." Ann Intern Med **129**(7): 517-524.

18. Giovannucci, E., E. Rimm, et al. (1994). "Aspirin use and the risk for colorectal cacner and adenoma in male health professionals." Ann Intern Med **121**(4): 241-246.

19. Gospodarowicz, M. K., D. E. Henson, et al. (2001). Prognostic Factors in Cancer. New York, A John Wiley & Sins, inc.

20. Gryfe, R. and H. Kim (2000). "Tumor microsatellite instability and clinical outcome in young patients with colorectal cancer." N Engl J Med **342**: 69-70.

21. Guerra, A. and F. J. Jimenez (1998). "Multivariate analysis of prognostic factors in resected colorectal cancer: A new prognostic index." Eur J Gast Hepatol **10**: 51-8.

22. Halling, K. and A. French (1999). "microsatellite instability and 8p allelic imbalance in stage B2 and C colorectal caners." J. NAtl Cancer Inst **91**: 1295-1303.

23. Iwatsuki, S. and T. E. Starzi (1999). "Hepatic resection for metastatic colorectal adenocarcinoma: a proposal of a prognostic scoring system." Journal of American College of Surgeons **189**(3): 291-299.

24. Jen, J. and H. Kim (1994). "Allelic loss of chromosome 18q and prognosis in colorectal cancer." N Engl J Med **331**: 213-21.

25. Johnston, P. G. (2004). "Of what value genomics in colorectal cancer? Opportunities and challenges." Journal of Clinical Oncology **22**(9): 1538-1539.

26. Kampman, E. and E. Giovannucci (1994). "Calcium, Vitamin D, dairy foods and the occurrence of colorectal adenoma among men and women in two prospective studies." Am J Epidemiology **139**(1): 16-29.

27. Kune, G., S. Kune, et al. (1990). "Body weight and physical activity as predictors of colorectal cancer risk." Nutr Cancer **13**(1-2): 9-17.

28. Longley, D., D. Harkin, et al. (2003). "5-Fluorouracil: Mechanism of action and clinical strategies." Nat Rev Cancer **3**: 330-338.

29. Mackillop, W. J. and C. F. Quirt (1997). "Measuring the accuracy of prognostic judgement in oncology." J Clin Epidemiology **50**(1): 21-29.

30. Moertel, C. and V. Go (1986). "The preoperative carcinoembriyonic antigen test in the diagnosis, staging and prognosis of colorectal cancer." cancer **58**(603).

31. Murphy, T. K. and e. al (2000). "Body mass index and colon cancer mortality in large prospective study." Am J Epidemiology **152**(9): 847-854.

32. O'Connell, J. B., M. A. Maggard, et al. (2004). "Colon cancer survival rates with the new American Joint Committee on Cancer sixth edition staging." Journal of the National Cancer Institute **96**(19): 1420-1425.

33. Ogunbiyi, O. and P. Goodfellow (1998). "Confirmation that 18q allelic loss in colon cancer is a prognostic indicator." J Clin Oncol **16**: 427-33.

34. Parkin, D., P. Pisani, et al. (1999). "Global cancer statistics." CA Cancer J Clin **49**(33).

35. Ponz de Leon et al. (1992). "Clinical and pathological prognostic Indicators in Colorectal Cancer". Cancer **69**: 626-635.

36. Prall, F. and e. al (2004). "Expression Profiling of Colorectal Carcinomas Using Tissue Microarray: Cell Cycle Regulatory Proteins p21, p27, and p53 as Immunohistochemical Prognostic Markers in Univariate and Multivariate Analysis." Applied Immunohistochemistry and Molecular Morphology **12**(2): 111-121.

37. Quirke, P. and P. Durdy (1986). "Local recurrence of rectal adenocarcinoma due to inadequate surgical resection. Histopathological study of lateral tumor spread and surgical excision." Lancet **1**: 996-999.

38. Ratto, C. and F. Crucitti (1998). "Prognostic factors in colorectal cancer. Literature review for clinical application." Dis Colon Rectum **41**(8): 1033-49.

39. Rene A.C. et al (2004). "Prognostic factors in locally advanced colon cancer treated by extended resection". Rev. Hosp. Clin. Med. S. Paulo **59**(6): 361-368.

40. Sandler, R. S. (1996). "Epidemiology and risk factors for colorectal cancer." Gastroenterology Clinics of North America **25**(4): 717-734.

41. Shibata, D. and M. Reale (1996). "The DCC protein and prognosis in colorectal cancer." N Engl J Med **335**: 1727-32.

42. Slattery, M., M. Schumacher, et al. (1988). "Physical activity , diet, and risk of colon cancer in Utah." Am J Epidemiology **128**(5): 989-99.

43. Sleisenger, M., M. Feldman, et al. (2004). Gastrointestinal and Liver diseases, Pathophysiology/ diagnosis and management, Saunders.

44. Society, A. C. (2005). "Cancer Facts and Fiures."

45. Steinberg, S. and K. Barwick (1986). "Importance of tumor pathology and morphology in patients with surgically resected colorectal cancer." cancer **58**: 1340.

46. Strumer, T., R. Glynn, et al. (1998). "Aspirin use and colorectal cancer: post-trial follow-up data from the Physicians' Health Study." Ann Intern Med **128**(9): 713-20.

47. Tepper JE, O'Connell MJ, Niedzwiecki D, et al.(2001): Impact of number of nodes retrieved on outcome in patients with rectal cancer. J Clin Oncol 19 (1): 157-63.

48. Trichopoulos, D. and e. al (1997). Epideimiology of Cancer. Philadelphia, Lippincott-Raven.

49. Vieira, R. A. C. and C. A. Melo (2004). "Prognostic factors in locally advanced colon cancer treated by extended resection." Rev. Hosp.Clin. Fac. Med **59**(6): 361-368.

50. Wang, Y., T. Jatkoe, et al. (2004). "Gene expression profiles and molecular markers to predict recurrence of Dukes B colon cancer." Journal of Clinical Oncology **22**(9): 1564-1571.

51. Watson, P. and K. Lin (1998). "Colorectal carcinoma survival among hereditary nonpolyposis colorectal carcinoma family members." Cancer **83**: 259.

52. Wolmark, N. and B. Fisher (1986). "the prognostic value of the modification of the Dukes' C class of colorectal cancer." Ann Surgery **203**: 115.

53. Wolters, U. and H. W. Keller (1996). "Colorectal cancer- A multivariate analysis of prognostic factors." Eur J Surg Oncol. **22**: 592-7

APPENDIX 1

Colorectal Cancer Prevention

Colorectal cancer occurs as a result of complex interaction between a person's inherited susceptibility (genetic) and environmental factors. Epidemiological and clinical investigations suggest that diet high in fat, calories, protein, alcohol and meat and low in calcium and folate are associated with increased with increased incidence of CRC.

Modifiable factors:

- **Diet:** Diet high in fat, meat (both red and white), alcohol and low in calcium and folate are associated with increased incidence of CRC. Evidence on whether diet high in fiber exerts a protective role in reducing the incidence of CRC is mixed. A high-fiber diet is thought to be protective, because it accelerates the rate at which fats pass through the bowel and reducing the exposure/contact of the large intestine to carcinogens [26, 40]. However, conflicting results are seen in some studies.
- **NSAIDS:** Some studies reported a reduction in colon cancer incidence with the use of aspirin, with a 30% overall reduction in colorectal cancer, including a 50% reduction in advanced cases [18]. However, in a follow-up study there was no association between the use of aspirin and the incidence of CRC [46]. The use of NSAIDS as a primary prevention measure is being considered and will depend on the dose and duration of intake.

- **Cigarette smoking:** Cigarette smoking is associated with an increased tendency to form adenomas and develop into colorectal cancer [2]. Most case control studies of cigarette exposure and adenomas have found an elevated risk for smokers. In the Cancer Prevention study II, a large nationwide cohort study, multivariate-adjusted colorectal cancer mortality rates were highest among current smokers, intermediate among former smokers, and lowest in never smokers, with increased risk after 20 or more years of smoking history among both men and women. Based on this study data, it was estimated that 12% of colorectal cancer deaths in the US population in 1997 were attributable to smoking (Chao, A et al, 2000). A positive relationship between alcohol intake and large bowel cancers was seen in some studies [10].
- **Vitamins:** An inverse association was found between the risk of CRC and intake of vitamins E; the RR for the highest compared to the lowest quartile was 0.3 (95% CI, 0.19-0.54) [6]. A similar association was seen for Vitamin D and folic acid intake and risk of CRC [17].
- **Calcium:** Several studies have observed an inverse relationship between calcium intake and cancer risk. Orally ingested calcium binds with bile acids and fatty acids released into the intestine following a high fat diet, to form insoluble compounds which are not harmful to the colonic mucosa and thereby reduces the exposure to the toxic effects of bile acids [43].
- **Post menopausal female hormones:** Epidemiologic Studies have suggested a decreased risk of colon cancer among users of postmenopausal female hormone supplements [10]

- **Physical activity:** An inverse relationship is seen between level of physical activity and colon cancer incidence. A sedentary lifestyle has been associated with an increased risk of colorectal cancer in some studies [13, 44] but not all [27].
- **Colonscopy:** Colonscopy with the removal of adenomatous polyps helps in reducing the risk of CRC.

APPENDIX 2

ACS Screening Guidelines

Beginning at age 50, men and women should have 1 of the 5 screening option below:

1. A fecal occult blood test (FOBT) or fecal immunochemical test (FIT) every year
2. Flexible sigmoidoscopy every 5 years
3. FOBT/FIT every year plus sigmoidoscopy every 5 years
4. Double contrast barium enema every 5 years or
5. Colonoscopy every 10 years.

APPENDIX 3

CLINICAL CHART REVIEW DATA ENTRY FORM

Note: Dates are entered in MM/DD/YYYY format.
If month or day is unknown, specify value as 01.
If entire date is unknown, specify value as 01/01/1111

2.	Moffitt medical record number	
3-4.	Tissue for study: Note: This tissue should be the earliest colorectal primary; if colorectal primary tissue not available, then earliest available other site	1 First colorectal primary 2 Second colorectal primary 3 Local recurrence 4 Appendiceal adenocarcinoma 5 Hepatic metastasis 6 Pulmonary metastasis 7 Other metastasis: Site _____ (4)
5.	Date of collection	/ /
6.	Surgical Accession Number	S - _____
7.	Gender	0 Male 1 Female
8.	Race/ethnicity	1 Caucasian 2 African American 3 Hispanic 4 Asian 5 Other/unknown
9.	Date of birth (MM/DD/YYYY)	/ /

10-18.	Clinical presentation with index primary colorectal cancer: None recorded Asymptomatic Melena Hematochezia Change in bowel habits Abdominal pain Weight loss Clinically "Obstructed" Hct at presentation	0 No 1 Yes (10) 0 No 1 Yes (11) 0 No 1 Yes (12) 0 No 1 Yes (13) 0 No 1 Yes (14) 0 No 1 Yes (15) 0 No 1 Yes (16) 0 No 1 Yes (17) _____ (18)
19.	Weight loss in lbs prior to presentation with index primary colorectal cancer (use 999 if unknown)	
20.	Weight in lbs at presentation with index primary colorectal cancer (use 999 if unknown)	
21.	Height in inches (use 999 if unknown)	
22-40.	<u>Charlson Index comorbidities</u>: Myocardial infarct Congestive heart disease Peripheral vascular disease Cerebrovascular disease Dementia Chronic pulmonary disease Connective tissue disease Peptic ulcer disease Mild liver disease Diabetes Hemiplegia Moderate/severe renal disease Diabetes w/ end organ disease Non-metastatic cancer, other than colon cancer Metastatic cancer, other than colon cancer	0 No/unknown 1 Yes (22) 0 No/unknown 1 Yes (23) 0 No/unknown 1 Yes (24) 0 No/unknown 1 Yes (25) 0 No/unknown 1 Yes (26) 0 No/unknown 1 Yes (27) 0 No/unknown 1 Yes (28) 0 No/unknown 1 Yes (29) 0 No/unknown 1 Yes (30) 0 No/unknown 1 Yes (31) 0 No/unknown 1 Yes (32) 0 No/unknown 1 Yes (33) 0 No/unknown 1 Yes (34) 0 No/unknown 1 Yes (35) 0 No/unknown 1 Yes (36) 0 No/unknown 1 Yes (37)

	Leukemia Lymphoma Moderate/severe liver disease AIDS	0 No/unknown 1 Yes (38) 0 No/unknown 1 Yes (39) 0 No/unknown 1 Yes (40)
41-44.	Smoking history Packs per day Years Pack-years If any quantity unknown, specify "999." Give only numbers from chart.	0 No 1 Yes 2 Unknown (41) _____ (42) _____ (43) _____ (44)
45.	History of alcohol use	0 No 1 Yes 2 Unknown
46.	History of hormone replacement therapy	0 No 1 Yes 2 Unknown
47 - 55.	Personal history of cancer other than the index primary colorectal cancer	Cancer #1 Cancer #2 Cancer #3 _____ (47) _____ (48) _____ (49) CODING KEY - additional primary cancers 0 Primary of unknown origin 1 2nd Colorectal 2 Stomach 3 Small Intestine 4 Ampulla 5 Urothelium (renal pelvis, ureter, bladder) 6 Endometrium 7 Ovary 8 Brain 9 Other 99 None Cancer #1 _____

	If cancer is "9 Other," specify primary: Did the additional primary cancer coexist with the index primary colorectal cancer? Cancer #1 Cancer #2 Cancer #3	(50) Cancer #2 _____ (51) Cancer #3 _____ (52) 0 No 1 Yes 2 Unknown (53) 0 No 1 Yes 2 Unknown (54) 0 No 1 Yes 2 Unknown (55)
56-92.	Family history recorded in chart? Family history of cancer: Family member # 1 # 2 # 3 # 4 # 5 # 6 # 7 # 8 # 9	0 No 1 Yes (56) Relation to Patient #2 / 2nd Degree? / Primary Cancer #1 / Primary Cancer _____ (57) _____ (58) _____ (59) _____ (60) _____ (61) _____ (62) _____ (63) _____ (64) _____ (65) _____ (66) _____ (67) _____ (68) _____ (69) _____ (70) _____ (71) _____ (72) _____ (73) _____ (74) _____ (75) _____ (76) _____ (77) _____ (78) _____ (79) _____ (80) _____ (81) _____ (82) _____ (83) _____ (84) _____ (85) _____ (86) _____ (87) _____ (88) _____ (89) _____ (90) _____ (91) _____ (92) CODING KEY – Relation to Patient 0 Unknown 7 Grandmother 1 Mother 8 Grandfather 2 Father 9 Aunt 3 Sister 10 Uncle 4 Brother 99 No cancer

		5 Daughter 6 Son CODING KEY - 2nd Degree? 0 1st degree or unknown 1 maternal 2nd degree relative 2 paternal 2nd degree relative Note: 2nd degree relatives are grandparents, aunts and uncles CODING KEY - Primary Cancer 0 Primary of unknown origin 1 Colorectum 2 Stomach 3 Small Intestine 4 Ampulla 5 Urothelium (renal pelvis, ureter, bladder) 6 Endometrium 7 Ovary 8 Brain 9 Other
93.	Location of index primary colorectal adenocarcinoma in patient	1 Cecum & Ileocecal Valve 2 Appendix 3 Ascending colon (Right colon) 4 Hepatic flexure of colon 5 Transverse colon 6 Splenic flexure of colon 7 Descending colon (Left colon) 8 Sigmoid colon 9 Colon, NOS 10 Rectosigmoid junction (Rectosigmoid colon) 11 Rectum
94-97.	TNM tumor stage at time of surgery for index primary colorectal cancer	Primary Tumor (T) (94) 0 **T0** – No evidence of primary tumor 1 **T1** – Tumor invades submucosa 2 **T2** – Tumor invades muscularis propria 3 **T3** – Tumor invades through the muscularis propria into the subserosa, or into non-peritonealized pericolic or perirectal tissues

	TNM staging (continued)	4 **T4** – Tumor directly invades other organs or structures, and/or perforates visceral peritoneum Note 1: Direct invasion in T4 includes invasion of other segments of the colorectum by way of the serosa; for example, invasion of the sigmoid colon by a carcinoma of the cecum Note 2: Tumor that is adherent to other organs or structures, macroscopically, is classified as T4 5 **TX** – Primary tumor cannot be assessed 6 **Tis** – Carcinoma in situ; intraepithelial or invasion of lamina propria Note: Tis includes cancer cells confined within the glandular basement membrane (intraepithelial) or lamina propria (intramucosal) with no extension through the muscularis mucosa into the submucosa Regional Lymph Nodes (N) (95) 0 **N0** – No regional lymph node metastasis 1 **N1** – Metastasis in 1 to 3 regional lymph nodes 2 **N2** – Metastasis in 4 or more regional lymph nodes 3 **NX** – Regional lymph nodes cannot be assessed Distant Metastasis (M) (96) 0 **M0** – No distant metastasis 1 **M1** – Distant metastasis 2 **MX** – Distant metastasis cannot be assessed

		Radial Margins and Residual Tumor (R) (97) 0 **R0** – Complete resection, margins histologically negative, no residual tumor left after resection 1 **R1** – Incomplete resection, margins histologically involved, microscopic tumor remains after resection of gross disease 2 **R2** – Incomplete resection, margins involved or gross disease remains after resection 3 Resection margins cannot be assessed
98.	Tumor stage at time of surgery for index primary colorectal cancer (use TNM classification to determine stage)	0 **Stage 0** Tis N0 M0 1 **Stage I** T1 N0 M0 T2 N0 M0 2 **Stage IIA** T3 N0 M0 3 **Stage IIB** T4 N0 M0 4 **Stage IIIA** T1-T2 N1 M0 5 **Stage IIIB** T3-T4 N1 M0 6 **Stage IIIC** Any T N2 M0 7 **Stage IV** Any T Any N M1

		8 Tumor stage cannot be assessed
99.	Histologic grading of index primary adenocarcinoma	1 **G1** – Well differentiated 2 **G2** – Moderately well differentiated 3 **G3** – Poorly differentiated 4 **G4** – Undifferentiated 9 **GX** – Grade not determined or unspecified
100.	Histology – Mucinous component?	0 No mucinous component noted 1 Yes
101.	Histology – Signet ring cells?	0 No signet ring cells noted 1 Yes
102.	Histology – Perineural invasion?	0 No perineural invasion noted 1 Yes
103.	Histology – Lymphatic invasion?	0 No lymphatic invasion noted 1 **L** – Microscopic lymphatic invasion
104.	Vascular invasion?	0 No vascular invasion noted 1 **V1** – Microscopic vascular invasion 2 **V2** – Macroscopic vascular invasion
105-108.	Metastases present at time of surgery for index primary colorectal cancer: Liver Lung Metastasis other than liver or lung	 0 No 1 Yes (105) 0 No 1 Yes (106) 0 No 1 Yes (107) Site(s): _____ (108)
109.	Date of surgery for index primary colorectal cancer (MM/DD/YYYY) If month or day is unknown, specify value as 01. If entire date is unknown, specify value	/ /

	as 01/01/1111.	
110.	Type of surgery for index primary colorectal cancer	0 Not resected 1 Subtotal colectomy 2 Right hemicolectomy 3 Transverse colectomy 4 Left hemicolectomy 5 Sigmoid resection 6 Low anterior resection (rectal cancer) 7 Abdominoperineal resection (rectal cancer)
111.	Were resection margins clear on index primary colorectal cancer?	0 No 1 Yes 2 Primary not resected 3 Unknown
112-113.	Size of index primary colorectal cancer from pathology report	Length along axis of colon _____ cm (112) Maximum diameter _____ cm (113)
114-115.	Number of nodes taken at resection of index primary colorectal cancer Number of these nodes that were positive for cancer	_____ (114) _____ (115)
116-123.	Treatment for hepatic metastasis present at time of surgery for index primary colorectal cancer: Surgical treatment Wedge resection Segmentectomy Lobectomy Radio frequency ablation Infusion pump Timing of this treatment for hepatic metastasis present at surgery for index primary colorectal cancer:	0 No 1 Yes (116) 0 No 1 Yes (117) 0 No 1 Yes (118) 0 No 1 Yes (119) 0 No 1 Yes (120) 0 No 1 Yes (121) (122) 0 None of the above treatments done for hepatic metastasis at time of surgery or later 1 Treatment done at time of surgery

		2 Treatment delayed until a later time Date _____ / _____ / _____ (123)
124-126.	Treatment for pulmonary metastases at time of surgery for index primary colorectal cancer: Surgical treatment Wedge resection Lobectomy	0 No 1 Yes (124) 0 No 1 Yes (125) 0 No 1 Yes (126)
127.	Preoperative treatment for index primary colorectal cancer	0 None 1 Chemotherapy 2 Radiation therapy 3 Both
128.	Intra-operative radiation therapy at time of treating index primary?	0 No 1 Yes
129.	Postoperative treatment for index primary colorectal cancer	0 None 1 Chemotherapy 2 Radiation therapy 3 Both
130.	Recurrence of index primary colorectal cancer (radiographic)	0 No 1 Yes 2 Unknown
131-135.	Sites of recurrent colorectal cancer: Local Liver Lung Sites other than liver or lung	0 No 1 Yes (131) 0 No 1 Yes (132) 0 No 1 Yes (133) 0 No 1 Yes (134) Site(s): _____ (135)
136.	Date recurrence discovered	/ /

	(MM/DD/YYYY) If month or day is unknown, specify value as 01. If entire date is unknown, specify value as 01/01/1111.	
137-139.	Treatment of recurrent disease: Surgical resection Radio frequency ablation (RFA) Chemotherapy	0 No 1 Yes 2 Unknown (137) 0 No 1 Yes 2 Unknown (138) 0 No 1 Yes 2 Unknown (139)
140.	Date of treatment of recurrence by resection or RFA. If month or day is unknown, specify value as 01. If entire date is unknown, specify value as 01/01/1111.	/ /
141.	Date of last recorded contact with patient in chart If month or day is unknown, specify value as 01. If entire date is unknown, specify value as 01/01/1111.	/ /
142-209.	CEA:	Date drawn Level _____ /_____ / _____ (142) _____ (143) _____ /_____ / _____ (144) _____ (145) _____ /_____ / _____ (146) _____ (147) _____ /_____ / _____ (148) _____ (149)

	CEA (continued)	_____ /_____ / _____ (150) _____ (151) _____ /_____ / _____ (152) _____ (153) _____ /_____ / _____ (154) _____ (155) _____ /_____ / _____ (156) _____ (157) _____ /_____ / _____ (158) _____ (159) _____ /_____ / _____ (160) _____ (161) _____ /_____ / _____ (162) _____ (163) _____ /_____ / _____ (164) _____ (165) _____ /_____ / _____ (166) _____ (167) _____ /_____ / _____ (168) _____ (169) Date drawn Level _____ /_____ / _____ (170) _____ (171) _____ /_____ / _____ (172) _____ (173) _____ /_____ / _____ (174) _____ (175) _____ /_____ / _____ (176) _____ (177) _____ /_____ / _____ (178) _____ (179) _____ /_____ / _____ (180) _____ (181) _____ /_____ / _____ (182) _____ (183) _____ /_____ / _____ (184) _____ (185) _____ /_____ / _____ (186) _____ (187) _____ /_____ / _____ (188) _____ (189) _____ /_____ / _____ (190) _____ (191) _____ /_____ / _____ (192) _____ (193)

		_____ /_____ / _____ (194) _____ (195) _____ /_____ / _____ (196) _____ (197) _____ /_____ / _____ (198) _____ (199) _____ /_____ / _____ (200) _____ (201) _____ /_____ / _____ (202) _____ (203) _____ /_____ / _____ (204) _____ (205) _____ /_____ / _____ (206) _____ (207) _____ /_____ / _____ (208) _____ (209)
210.	Second tissue for study: Note: This tissue cannot be a first colorectal primary; if second tissue available, then it should be the next earliest available site	1 none 2 Second colorectal primary 3 Local recurrence 4 Appendiceal adenocarcinoma 5 Hepatic metastasis 6 Pulmonary metastasis 7 Other metastasis: Site _____ (211)
212.	Date of collection, second tissue	/ /
213.	Surgical Accession Number, second tissue	S - _____
214.	Third tissue for study: Note: This tissue cannot be a first colorectal primary; if second tissue available, then it should be the next earliest available site	1 none 2 Second colorectal primary 3 Local recurrence 4 Appendiceal adenocarcinoma 5 Hepatic metastasis 6 Pulmonary metastasis 7 Other metastasis: Site _____ (215)
216.	Date of collection, third tissue	/ /

217.	Surgical Accession Number, third tissue	S - _____