Gastroenterology 2015;148:77–87

Association Between Molecular Subtypes of Colorectal Cancer and Patient Survival

Amanda I. Phipps,[1,2] Paul J. Limburg,[3] John A. Baron,[4] Andrea N. Burnett-Hartman,[1,2] Daniel J. Weisenberger,[5] Peter W. Laird,[5] Frank A. Sinicrope,[3] Christophe Rosty,[6,7] Daniel D. Buchanan,[6] John D. Potter,[1,2,8] and Polly A. Newcomb[1,2]

[1]*Epidemiology Department, University of Washington, Seattle, Washington;* [2]*Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, Washington;* [3]*Division of Gastroenterology and Hepatology, Mayo Clinic, Rochester, Minnesota;* [4]*Department of Medicine, University of North Carolina, Chapel Hill, North Carolina;* [5]*USC Epigenome Center, USC/Norris Comprehensive Cancer Center, University of Southern California, Los Angeles, California;* [6]*Cancer and Population Studies Group, Queensland Institute of Medical Research, Herston, QLD, Australia;* [7]*Department of Molecular and Cellular Pathology, University of Queensland School of Medicine, Herston, QLD, Australia; and* [8]*Centre for Public Health Research, Massey University, Wellington, New Zealand*

See Covering the Cover synopsis on page 1; see editorial on page 10.

BACKGROUND AND AIMS: Colorectal cancer (CRC) is a heterogeneous disease that can develop via several pathways. Different CRC subtypes, identified based on tumor markers, have been proposed to reflect these pathways. We evaluated the significance of these previously proposed classifications to survival. **METHODS:** Participants in the population-based Seattle Colon Cancer Family Registry were diagnosed with invasive CRC from 1998 through 2007 in western Washington State (N = 2706), and followed for survival through 2012. Tumor samples were collected from 2050 participants and classified into 5 subtypes based on combinations of tumor markers: type 1 (microsatellite instability [MSI]–high, CpG island methylator phenotype [CIMP] –positive, positive for *BRAF* mutation, negative for *KRAS* mutation); type 2 (microsatellite stable [MSS] or MSI-low, CIMP-positive, positive for *BRAF* mutation, negative for *KRAS* mutation); type 3 (MSS or MSI low, non-CIMP, negative for *BRAF* mutation, positive for *KRAS* mutation); type 4 (MSS or MSI-low, non-CIMP, negative for mutations in *BRAF* and *KRAS*); and type 5 (MSI-high, non-CIMP, negative for mutations in *BRAF* and *KRAS*). Multiple imputation was used to impute tumor markers for those missing data on 1–3 markers. We used Cox regression to estimate hazard ratios (HR) and 95% confidence intervals (CI) for associations of subtypes with disease-specific and overall mortality, adjusting for age, sex, body mass, diagnosis year, and smoking history. **RESULTS:** Compared with participants with type 4 tumors (the most predominant), participants with type 2 tumors had the highest disease-specific mortality (HR = 2.20, 95% CI: 1.47–3.31); subjects with type 3 tumors also had higher disease-specific mortality (HR = 1.32, 95% CI: 1.07–1.63). Subjects with type 5 tumors had the lowest disease-specific mortality (HR = 0.30, 95% CI: 0.14–0.66). Associations with overall mortality were similar to those with disease-specific mortality. **CONCLUSIONS:** Based on a large, population-based study, CRC subtypes, defined by proposed etiologic pathways, are associated with marked differences in survival. These findings indicate the clinical importance of studies into the molecular heterogeneity of CRC.

Keywords: Oncogene; Methylation; Serrated Colorectal Cancer; Prognostic Factor.

Increasing evidence indicates that colorectal cancer (CRC) is a biologically heterogeneous disease that can develop via a number of distinct pathways involving different combinations of genetic and epigenetic changes.[1,2] Proposed subtype classifications for CRC, based on the presence of microsatellite instability (MSI), the CpG island methylator phenotype (CIMP), and somatic mutations in *BRAF* and *KRAS*, are thought to approximate these distinct pathways.[1,2] In particular, CRC reflective of the "traditional" adenoma-carcinoma pathway has been described as typically demonstrating absent (microsatellite stable [MSS]) to low-level MSI (MSI-low) without CIMP and without somatic *BRAF* or *KRAS* mutations; CRC resulting from a "serrated" pathway has been described as frequently *BRAF* mutated and CIMP positive; and an additional pathway has been suggested for *KRAS*-mutated CRC that is MSS/MSI-low and CIMP-low.[2,3]

The biologic distinctions between CRC subtypes resulting from different etiologic pathways can plausibly translate to differences in survival. As tumor markers that can reflect such different pathways, MSI, CIMP, *BRAF*-mutation, and *KRAS*-mutation status have each been studied extensively, with evidence of differences in the distribution of tumor site, sex, age and stage at diagnosis, and survival.[4–22] However, the significance of subtype classifications based on combinations of these 4 tumor markers with respect to survival has been minimally described.[3,23] In the only prior study to evaluate differences in survival across CRC subtypes defined by these 4

Abbreviations used in this paper: **CI, confidence interval; CRC, colorectal cancer; CIMP, CpG island methylator phenotype; HR, hazard ratio; MSI, microsatellite instability; MSS, microsatellite stable; PMR, percentage of methylated reference; SCCFR, Seattle Colon Cancer Family Registry; SEER, Surveillance, Epidemiology, and End Results.**

0016-5085/$36.00
http://dx.doi.org/10.1053/j.gastro.2014.09.038

tumor markers in combination, Samadder et al[3] suggested that CRC with a *BRAF*-mutated/CIMP-high phenotype, suggestive of the serrated pathway, was associated with modestly worse survival than CRC with a MSS/CIMP-negative/*BRAF*-mutation–negative/*KRAS*-mutation–negative phenotype, suggestive of the traditional pathway.

Using data from the population-based Seattle Colon Cancer Family Registry (SCCFR) and the Postmenopausal Hormones Supplemental Study to the SCCFR,[24,25] we further explored the relationship between CRC molecular subtypes, defined by common tumor marker combinations, and survival.

Methods

Study Population

A description of the study populations has been published elsewhere.[24,25] Briefly, SCCFR study participants included persons diagnosed with incident invasive CRC between January 1998 and June 2002 who, at the time of diagnosis, were aged 20–74 years and resided in King, Pierce, or Snohomish counties of Washington State (Supplementary Table 1). During this same period, women aged 50–74 years at CRC diagnosis and residing in 10 surrounding counties were also recruited for participation in the Postmenopausal Hormones Supplemental Study to the SCCFR. During a second SCCFR recruitment phase (diagnosis dates April 2002 to July 2007), eligible participants were identified as individuals diagnosed at ages 18–49 years with invasive CRC within the combined 13-county region. All cases were identified through the population-based Surveillance, Epidemiology, and End Results (SEER) cancer registry serving western Washington State. Eligibility was limited to English speakers with publicly available telephone numbers. Of 3525 eligible individuals contacted, 302 (9%) were deceased, 401 (11%) refused participation, 92 (3%) were lost to follow-up before interview, and 24 (1%) completed only a partial interview. Among participants who completed the interview (N = 2706), adequate tumor specimens were available for 77% (n = 2080). Participants for whom tumor specimens were not obtained were excluded from this analysis.

This study was approved by the Institutional Review Board of the Fred Hutchinson Cancer Research Center in accordance with assurances filed with and approved by the US Department of Health and Human Services.

Tumor Characteristics

DNA extracted from paraffin-embedded formalin-fixed diagnostic tumor tissue specimens was used in tumor marker testing. Testing for MSI was based on a 10-gene panel in DNA from tumor and normal surrounding tissue (BAT25, BAT26, BAT40, MYCL, D5S346, D17S250, ACTC, D18S55, D10S197, and BAT34C4) for the majority of cases (n = 1430):[24,26] tumors were classified as MSI-high if instability was observed for ≥30% of markers, and MSS/MSI-low if instability was observed in <30% of markers. For other cases (n = 534), MSI status was based on immunohistochemistry testing of 4 markers (MLH1, MSH2, MSH6, PMS2): cases for which tissue exhibited positive staining for all markers were considered MSS/MSI-low, and cases negative for the expression of at least one marker were considered MSI-high.[27,28] Tumor DNA was tested for the p.V600E *BRAF* mutation (n = 1948) using a fluorescent allele-specific polymerase chain reaction assay, as described previously;[29] this mutation accounts for approximately 90% of *BRAF* mutations in CRC.[30] Mutations in *KRAS* codons 12 and 13 were identified through forward and reverse sequencing of amplified tumor DNA (n = 1894)[8,31]; mutations in this hotspot region account for approximately 80% of *KRAS* mutations in CRC.[32,33] CIMP testing was completed for a large subset of cases (n = 1508) based on a validated quantitative DNA methylation assay using a 5-gene panel (*CACNA1G*, *IGF2*, *NEUROG1*, *RUNX3*, and *SOCS1*).[34–36] As described elsewhere,[34] tumors were classified as CIMP positive if the percentage of methylated reference (PMR) ratio was ≥10 for at least 3 of 5 markers and as non-CIMP if the PMR ratio was ≥10 for <3 markers. PMR is calculated as the amount of methylated tumor DNA at a specific locus (normalized to input bisulfite DNA amount measured at *ALU* repetitive elements) divided by the *ALU*-normalized amount in a methylated reference sample, multiplied by 100. Tumor site and stage information were available from SEER.

Subtype Classifications

Tumor subtypes were defined as follows, consistent with previously suggested classifications[1,2]: type 1 (ie, MSI-high, CIMP-positive, *BRAF*-mutated, *KRAS*-mutation–negative); type 2 (ie, MSS/MSI-low, CIMP-positive, *BRAF*-mutated, *KRAS*-mutation–negative); type 3 (ie, MSS/MSI-low, non-CIMP, *BRAF*-mutation–negative, *KRAS*-mutated); type 4 (ie, MSS/MSI-low, non-CIMP, *BRAF*-mutation–negative, *KRAS*-mutation–negative); and type 5 (ie, MSI-high, non-CIMP, *BRAF*-mutation–negative, *KRAS*-mutation–negative). Other marker combinations were grouped together as an "other" category for tabulations. In sensitivity analyses, we explored changes to the type 3 subtype classification for comparison with previous reports,[3] removing cases for whom all methylation markers had a PMR ratio <10 from this subgroup.

Of the 2080 cases for which tumor tissue was available, 30 were excluded due to insufficient tissue or uninformative assays. Multiple imputation was used to approximate tumor marker status for cases with 1 (n = 564), 2 (n = 104), or 3 missing markers (n = 38)[37,38]: the imputation model included variables for MSI, *BRAF*- and *KRAS*-mutation status, methylation status for the 5 genes used in classifying CIMP, stage, histology, sex, age at diagnosis, diagnosis year, body mass index, height, smoking history, use of nonsteroidal anti-inflammatory drugs at diagnosis, history of endoscopic screening before diagnosis, education, race, first line of therapy, time from diagnosis to interview, censoring indicators, and analysis time. Iterative rounds of imputation (n = 25) were performed using the *mi* command in STATA SE version 13.1 (Stata Corp, College Station, TX). Tumor subtype classifications were determined on the basis of assayed and, as necessary, imputed tumor markers. In addition to analyses utilizing these imputed data, we conducted sensitivity analyses using a complete-case approach, wherein only cases with complete tumor marker data were included.

Outcome Information

Vital status, death date, and cause of death were determined through linkage to SEER and the National Death Index. CRC-specific deaths included those with an underlying cause