Supplementary Materials

Body Mass Index and Molecular Subtypes of Colorectal Cancer

Murphy et al.

Supplementary Methods

Description of included studies

Colon Cancer Family Registry (CCFR)

The CCFR (www.coloncfr.org) is a National Cancer Institute-supported consortium consisting of six centers[1]. The CCFR includes data from approximately 42,500 total subjects (10,500 case probands and 26,900 unaffected and affected relatives, 4,280 unrelated population-based controls, and 920 spouse controls). The study recruited cases and unaffected controls (age 20 to 74 years) beginning in 1998. All participants self-completed a standardized questionnaire that included questions about established and suspected risk factors for colorectal cancer, including questions on medical history and medication use, reproductive history (for female participants), family history, physical activity, demographics, alcohol and tobacco use, and dietary factors. Colorectal case and population-based control participants from three of the six participating centers (Seattle-SCCFR, Australia-ACCFR, Ontario-OFCCR) were included in this study.

Cancer Prevention Study-II (CPS-II)

The CPS-II Nutrition cohort (established in 1992) is a prospective study of cancer incidence and mortality in the United States[2,3]. All participants filled out a self-administered questionnaire that included information on demographical, medical, dietary, and lifestyle factors. Biennial follow-up questionnaires have been sent out since 1997 in order to collect continuous information about current exposures and new cancer diagnoses. All reported cancers are verified through medical records, state cancer registry linkage, or death certificates. Controls were matched on race, gender, and age. The Emory University Institutional Review Board approves all aspects of the CPS-II Nutrition Cohort.

Darmkrebs: Chancen der Verhütung durch Screening (DACHS)

DACHS is a large German population-based case-control study started in 2003 in the Rhine-Neckar-Odenwald region (southwest region of Germany)[4,5]. The purpose of DACHS was to assess the potential of endoscopic screening for reduction of colorectal cancer risk and to investigate etiologic determinants of the disease, particularly lifestyle/environmental factors and genetic factors. Briefly, cases with a first diagnosis of invasive colorectal cancer (ICD-10 codes C18-C20) who were at least 30 years of age, German speaking, resident in the study region, and mentally and physically able to participate in a one-hour interview, were recruited by their treating physicians either in the hospital a few days after surgery, or by mail after hospital discharge. Cases were confirmed by histologic reports and hospital discharge letters following diagnosis of colorectal cancer. All hospitals treating colorectal cancer patients in the study region participated. Community based controls were randomly selected from population registries, employing age frequency matching (5-year groups), sex, and county of residence. Controls without a history of colorectal cancer were contacted by mail and follow-up calls. During an in-person interview, data on demographics, medical history, family history of colorectal cancer, and various lifestyle factors were collected. Participants also donated blood and mouthwash samples.

Diet, Activity, and Lifestyle Study (DALS)

DALS, which has been described in more detail elsewhere[6], was a population-based, case–control study of colon cancer. Participants were recruited between 1991 and 1994 from 3 locations: the Kaiser Permanente Medical Care Program of Northern California, an 8-county area in Utah, and the metropolitan Twin Cities area of Minnesota. Eligibility criteria for cases included age at diagnosis

between 30 and 79 years, diagnosis with first primary colon cancer (International Classification of Disease for Oncology, Second Edition, 18.0 and 18.2–18.9) between October 1, 1991, and September 30, 1994, English speaking, and competency to complete the interview. Individuals with cancer of the rectosigmoid junction or rectum were excluded, as were those with a pathology report noting familial adenomatous polyposis, Crohn's disease, or ulcerative colitis. A rapid-reporting system was used to identify all incident cases of colon cancer, resulting in the majority of cases being interviewed within 4 months of diagnosis. Controls from the Kaiser Permanente Medical Care Program were selected randomly from membership lists. In Utah, controls younger than 65 years of age were selected randomly through random-digit dialing and driver license lists. Controls 65 years of age and older were selected randomly from Health Care Financing Administration lists. In Minnesota, controls were identified from Minnesota drivers license or state identification lists. Patients with available tumor molecular characterization were included in this study.

Early Detection Research Network (EDRN)

The aim of the EDRN initiative is to develop and sustain a biorepository for support of translational research[7]. High-quality biospecimens from colorectal cancer patients ages 18 years or above were accrued and annotated with pertinent clinical, epidemiologic, molecular and genomic information. Information on molecular markers were abstracted from patient medical records and colorectal cancer with available MSI, CIMP, *KRAS* mutation, or *BRAF* mutation characterization were included in this study.

European Prospective Investigation into Cancer (EPIC) – Sweden

EPIC is an on-going multicenter prospective cohort study designed to investigate the associations between diet, lifestyle, genetic and environmental factors and various types of cancer[8]. Briefly, 521,448 participants (~70% women) mostly aged 35 years or above were recruited between 1992 and 2000. Participants were recruited from 23 study centers in ten European countries. All study participants provided written informed consent, and ethical approval for the EPIC study was obtained from the review boards of IARC and local participating centers. The current study included participants from the northern Swedish EPIC-Umeå site, which is the Västerbotten Intervention Study (VIP). Colorectal cancer cases were identified by linkage with the Cancer Registry of Northern Sweden, which reports to the Swedish Cancer Registry, and were verified by a gastrointestinal pathologist. Controls were selected from the full cohort of individuals who were alive and free of cancer (except non-melanoma skin cancer) at the time of case diagnosis.

Health Professionals Follow-up Study (HPFS)

The HPFS was started in 1986 with the purpose of evaluating underlying etiologies of cardiovascular disease and cancer[9]. It originally included 51,529 male health professionals currently residing in the United States who all completed a detailed questionnaire on health and diet. The all-male study was designed to complement the all-female Nurses' Health Study, which examines similar hypotheses. Colorectal cancer and other outcomes were reported by participants or next-of-kin and were followed up through review of the medical and pathology record by physicians. Overall, more than 97% of self-reported colorectal cancers were confirmed by medical record review. Information was abstracted on histology and primary anatomical location of the tumor. Follow-up evaluation has been excellent, with 94% of the men responding to date. Patients with available tumor molecular characterization were included in this study.

Melbourne Collaborative Cohort Study (MCCS)

The MCCS is a prospective study, run between 1990 and 1994, that recruited 41,514 healthy adult participants aged between 27 and 76 years (99% aged 40-69) from the Melbourne metropolitan area[10]. The goal of this study was to examine the role of lifestyle factors in the risk of cancer and

heart disease. Incident cases of colorectal cancer were identified through linkage to population-based cancer registries in Australia. Cases included participants with a histopathological diagnosis of invasive colorectal adenocarcinoma diagnosed after baseline. Participants provided informed consent and sufficient FFPE material for somatic testing. Study protocols were approved by the Human Research Ethics Committee at the Cancer Council Victoria.

Newfoundland Familial Colorectal Cancer Registry (NFCCR)

The NFCCR is a case-control study that includes pathology confirmed colorectal cancer cases less than 75 years of age diagnosed between January 1999 and December 2003, as identified from the Newfoundland Cancer Registry[11]. The Newfoundland Cancer Registry registers all cases of invasive cancer diagnosed among residents of the province of Newfoundland and Labrador in Canada. Consenting patients received a family history questionnaire and were asked to provide a blood sample and to permit access to tumor tissue and medical records. If a patient was deceased, they sought participation of a close relative for the purposes of obtaining the family history and permission to access tissue blocks and medical records. Population-based controls were identified by random digit dialing from the residents of the province and matched to the cases on sex and five-year age groups. Patients with available tumor molecular characterization were included in this study.

Nurses' Health Study (NHS)

The NHS cohort, initiated in 1976, originally included information on health related exposures from 121,700 married female registered nurses aged 30-55[12]. Since 1976, follow-up questionnaires have been mailed every 2 years. Colorectal cancer and other outcomes were reported by participants or next-of-kin and followed up through review of the medical and pathology record by physicians. Overall, more than 97% of self-reported colorectal cancers were confirmed by medical-record review. Information was abstracted on histology and primary anatomical location of the tumor. The rate of follow-up evaluation has been high: as a proportion of the total possible follow-up time, follow-up evaluation has been more than 92%. Colorectal cancer cases were ascertained through June 1, 2008.

Northern Sweden Health and Disease Study (NSHDS)

The NSHDS is a population based study including residents of Västerbotten county in Northern Sweden[13]. It includes more than 110,000 participants, of which approximately one third have repeated samples, from three population-based cohorts: the Västerbotten Intervention Project (VIP), the Northern Sweden WHO Monitoring of Trends and Cardiovascular Disease (MONICA) Study, and the local Mammography Screening Project (MSP). In the VIP cohort, which makes up approximately 85% of the NSHDS, aims to invite all residents of Västerbotten County to a health examination upon turning 30 (some years), 40, 50 and 60 years of age. It was established in 1985 and continues to recruit participants. In both the VIP and MONICA cohorts, extensive measured and self-reported health and lifestyle data were collected, whereas data in the MSP are more limited. Blood samples for research purposes are collected in all three cohorts. The NSHDS is a part of EPIC, and the selection of colorectal cases and controls were as described for EPIC-Sweden.

Harmonization of Colorectal Tumor Marker Data

Testing for microsatellite Instability (MSI), mutations in the *BRAF* gene, mutations in the *KRAS* gene, and CpG island methylator phenotype (CIMP) status was conducted by each study and according to individual study protocols. The harmonisation procedures have been previously described[14,15].

Microsatellite Instability (MSI) Status

Testing primarily consisted of polymerase chain reaction (PCR) based assessment of microsatellite status except for NSHDS and EPIC Sweden[13,16], which utilized immunohistochemical (IHC) detection of deficiency for mismatch repair (MMR) gene proteins MLH1, MSH2, MSH6, and PMS2 using standard procedures. Additionally, IHC was used for a subset of EDRN, MCCS[17-19], and CCFR[1,20] samples without PCR-based MSI characterization. For classification using IHC, tumors lacking nuclear staining in tumor cells for at least one of these proteins were considered to have a positive MSI screening status and MSI negative screens were considered microsatellite stable (MSS). See Table S2 for specific markers assessed using PCR-based methods. Tumor classification was based on > 4 interpretable markers for CCFR[1,20], NFCCR[21,22], MCCS[17], >5 interpretable markers for CPS-II (unless all four markers were unstable in which case the tumor was classified as MSI), and >7 interpretable markers for NHS and HPFS[23]. For these studies, tumors were classified as MSI-high (MSI-H) if 30% or more of the markers showed instability, and non-MSI-H if < 30% and > 0% showed instability, and if no marker exhibited instability. DALS, which carried out MSI testing prior to development of the Bethesda Consensus Panel[24], determined MSI based on the mononucleotides BAT26 and TGFβRII and a panel of 10 tetranucleotide repeats[25-27]. These have been shown to correlate highly with the Bethesda Panel[28]. A tumor classification of unstable was given if the panel of 10 tetranucleotides, BAT26, or TGFβRII were determined as unstable. Tumoral and normal DNA were PCR amplified with these 12 primer sets, and MSI was defined as > 1 new PCR products either smaller or larger than those produced from normal DNA. Specifically, for BAT26, the PCR product from tumor had to be >4 base pairs smaller than that from germline. A tumor classification of MSI from the tetranucleotide repeat panel was based on > 30% markers showing instability and MSS if <30% of repeats were unstable, with > 6 interpretable markers of the 10 evaluated. DACHS[29] determined MSI status using a mononucleotide marker panel[30] that has high concordance with the National Cancer Institute Bethesda Consensus Panel[24,31]. For EDRN, tumor markers were abstracted from medical records, and included both IHC and PCR methods to determine instability. As IHC with MLH1, MSH2, MSH6, and PMS2 has been shown to be as sensitive as MSI testing[32], we combined IHC and MSI results for the full EDRN case set. PCR-based testing included a 7 marker panel and IHC testing included testing of MLH1, MHS2, MSH6, and PMS2. For participants that had both IHC and MSI testing, concordance for the determination of stable and unstable using these two methods was very high with only one of 147 individuals identified as MSI with no protein expression changes identified in the MMR genes. In this case, we selected the most deleterious outcome of MSI.

BRAF and *KRAS* Mutation Status

Studies used PCR, sequencing, and IHC techniques to assess *BRAF* and *KRAS* mutations. Most studies evaluated V600E mutations in exon 15 and mutations in codons 12 and 13, though a few evaluated additional loci. In analyses, we included any mutation identified by a study.

CCFR tested for *BRAF* V600E mutations using a fluorescent allele-specific PCR (AS-PCR) assay[33] and used Sanger sequencing to assess mutations in *KRAS* codons 12 and 13[34,35]. NFCCR tested for *BRAF* V600E mutations using AS-PCR, followed by direct automatic sequencing to verify mutations[36], and did not evaluate *KRAS* mutations. MCCS used a fluorescent real-time AS-PCR assay[33] to test for the *BRAF* V600E mutation and a real-time PCR with high resolution melting (HRM) analysis followed by

direct Sanger sequencing for positive cases to identify KRAS mutations in codons 12 and 13[37]. CPS-II used PCR to assess *BRAF* V600E mutations and KRAS codon 12, 13, and 14 mutations.

DACHS[5] used both Sanger sequencing and IHC analysis of V600E expression to determine *BRAF* mutation status. For sequencing, they amplified exon 15 of *BRAF* using FideliTaq polymerase and sequenced using the BigDye Terminator v1.1 Cycle Sequencing Kit on an ABI 3500 Genetic Analyzer. DACHS determined KRAS mutation status by a single stranded conformational polymorphism technique (SSCP) or by Sanger sequencing, as reported previously[5]. NSHDS and EPIC Sweden[13] used real-time PCR using an allelic discrimination assay as described by Benlloch et. al.[38] to detect *BRAF* V600E mutations and BigDye v.3.1 sequencing to detect mutations in KRAS codons 12 and 13[39]. DALS evaluated *BRAF* mutations by amplifying exon 15 using Applied Biosystems AmpliTaq Gold and sequencing[40], and evaluated KRAS mutations by amplifying codons 12 and 13 using Taq FS DNA polymerase and sequencing using prism BigDye terminators and cycle sequencing on an ABI prism 377 automated sequencer[41].

EDRN tested for *BRAF* V600E mutation status primarily using real time PCR, though 5 samples were tested using DNA sequencing. EDRN tested for *KRAS* codons 12, 13, and 61 primarily using DNA sequencing with one sample tested using PCR. HPFS and NHS performed PCR and pyrosequencing to identify *BRAF* codon 600 mutations[42,43]. HPFS and NHS used real-time PCR and pyrosequencing to identify *KRAS* mutations in codons 12, 13, 61, and 146[42,44].

CpG Island Methylator Phenotype Status

Studies used methylation analysis to determine CIMP status. See Table S3 for specific genes assessed to determine CIMP status. Like the harmonization of MSI status, we created two CIMP categories for downstream analyses, CIMP-high and CIMP-low/negative. HPFS, NHS[45,46], CPS-II, NSHDS, EPIC Sweden[13,16], CCFR[47,48], MCCS[49] used the MethyLight[50]method to determine CIMP status. HPFS, NHS, CPS-II, NSHDS, and EPIC Sweden used a panel of eight genes, and CCFR and MCCS used a panel of five genes. The percent of methylated reference (PMR) value was calculated and, for CCFR, NSHDS, and EPIC Sweden a gene was considered positive for methylation when the PMR>10. CPS-II, HPFS, and NHS used a PMR cutoff value of >4 for CDKN2A, MLH1, CACNA1G, NEUROG1, RUNX3, SOCS1, and a PMR of >6 for CRABP1 and IGF2. HPFS, NHS, CPS-II, and NSHDS classified tumors with ≥5 methylated markers as CIMP-high, 1-4 markers as CIMP-low/negative, and no markers as CIMP-low/negative. CCFR and MCCS classified tumors with ≥3 methylated markers as CIMP-high and, otherwise, as CIMP-low/negative. DACHS[51] determined CIMP status using a panel of five genes, and methods described by Warth et. al.[52]. They determined methylation status from the methylation-specific PCR based on the presence or absence of amplified product, and classified tumors with ≥3 methylated markers as CIMP-high. DALS[40] determined CIMP status using a classic panel of CpG islands[53,54]. Tumors with ≥3 methylated markers were classified as CIMP-high and no methylated markers were classified as CIMP-low/negative, with three or more loci successfully evaluated.

Marker Combinations

Tumor subtypes were defined as follows, consistent with previously suggested classifications[55,56]: Type 1 (MSI-high, CIMP-high, *BRAF*-mutated, *KRAS*-wildtype), Type 2 (non-MSI-high, CIMP-high, *BRAF*-mutated, *KRAS*-wildtype), Type 3 (non-MSI-high, CIMP-low/negative, *BRAF*-wildtype, *KRAS*-mutated), Type 4 (non-MSI-high, CIMP-low/negative, *BRAF*-wildtype, *KRAS*-wildtype), and Type 5 (MSI-high, CIMP-low/negative, *BRAF*-wildtype, *KRAS*-wildtype).

Supplementary Table 1. Description of participating studies

Study name	Abbreviation	Design	Country of origin	Matching factors	Timing of body size measurement	N CRC cases	N controls
Colon Cancer Family Registry	CCFR_Australia	Case-control	Australia	Age, sex	Weight (kg) 2 years prior to enrolment. Height (cm) at enrolment	1,555	176
Colon Cancer Family Registry	CCFR_Ontario	Case-control	Canada	Age, sex	Weight (kg) 2 years prior to enrolment. Height (cm) at enrolment	1,706	1,259
Colon Cancer Family Registry	CCFR_Seattle	Case-control	United States	Age, sex	Weight (kg) 2 years prior to enrolment. Height (cm) at enrolment	1,812	745
Cancer Prevention Study II	CPSII	Cohort	United States	Age, sex, race, date of blood draw	Weight (kg) and height (cm) at enrolment	790	929
Diet Activity and Lifestyle Study	DALS	Case–control	United States	Age, sex	Weight (kg) 2 years prior to enrolment. Height (cm) at enrolment	1,083	1,148
Darmkrebs: Chancen der Verhütung durch Screenin	DACHS	Case–control	Germany	Age, sex, county of residence	Weight (kg) 5-14 years prior to enrolment. Height (cm) at enrolment	1,966	2,744
Early Detection Research Network	EDRN	Case–control	United States	Age, sex	Weight (kg) and height (cm) at enrolment	188	329
European Prospective Investigation into Cancer and Nutrition_Sweden	EPIC_Sweden	Cohort	Sweden	Age, sex, study center, follow-up time, time of day of blood collection, fasting status, menopausal status, phase of menstrual cycle at blood collection	Weight (kg) and height (cm) at enrolment	145	381
Health Professionals Follow-up Study	HPFS	Cohort	United States	Age, month/year of blood sampling	Weight (kg) and height (cm) at enrolment	585	591

Melbourne Collaborative Cohort Study	MCCS	Cohort	Australia	Sex, country of birth, year of baseline attendance	Weight (kg) and height (cm) at enrolment	490	670
Newfoundland Familial Colorectal Cancer Registries	NFCCR	Case-control	Canada	Age, sex	Weight (kg) 2 years prior to enrolment. Height (cm) at enrolment	489	461
Nurses' Health Study	NHS	Cohort	United States	Age, month/year of blood sampling,	Weight (kg) and height (cm) at enrolment	764	1,197
Northern Sweden Health and Disease Study	NSHDS	Cohort	Sweden	Subcohort, age, sex, age and year of blood sampling, fasting status	Weight (kg) and height (cm) at enrolment	299	383

Supplementary Table 2. Summary of study specific assessment of microsatellite instability (MSI) status

Study	Markers*/ Proteins	Threshold for Interpretability	Definitions
CCFR	BAT25, BAT26, BAT40, BAT34C4, D5S346, D17S250, ACTC, D18S55, D10S197, MYCL	>4 interpretable markers	* MSI-high if >30% markers showed instability
CPSII	BAT25, BAT26, BAT40, BAT34C4,ACTC, D10S197, D17S250, D18S55, D5S346, MYCL	>5 interpretable markers (unless 4 markers were unstable)	* MSI- high if >30% markers showed instability
DACHS	BAT25, BAT26, CAT25	All 3 markers interpretable	* MSI- high if >1 marker showed instability
DALS	BAT26, TGFBRII	>6 of 10 markers be interpretable from tetranucleotide repeat panel	* MSI: Instability in BAT26, TGFBRII, or 10 tetranucleotide marker panel. - 10 marker panel: ≥30% unstable repeats.
EDRN	BAT-25, BAT-26, CAT25	-	-
EPIC_Sweden	MLH1, MSH2, MSH6, and PMS2	Immunohistochemistry	Immunohistochemical detection of deficiency for selected mismatch repair proteins was used to determine MSI status.
HPFS	BAT25, BAT26, BAT40, D18S55, D18S56, D18S67, D18S487, D2S123, D5S346, D17S250	>7 interpretable markers	* MSI-high if >30% markers showed instability
MCCS	BAT25, BAT26, BAT40, BAT34C4, D5S346, D17S250, ACTC, D18S55, D10S197, MYCL	>4 interpretable markers	* MSI-high if >30% markers showed instability

NFCCR	BAT-25, BAT-26, BAT-40, BAT-34C4, D5S346, D17S250, ACTC, D18S55, D10S197, MYCL	>4 interpretable markers	* MSI-high if >30% markers showed instability
NHS	BAT25, BAT26, BAT40, D18S55, D18S56, D18S67, D18S487, D2S123, D5S346, D17S250	>7 interpretable markers	* MSI-high if >30% markers showed instability
NSHDS	MLH1, MSH2, MSH6, and PMS2	Immunohistochemistry	Immunohistochemical detection of deficiency for selected mismatch repair proteins was used to determine MSI status.

*EPIC_Sweden and NSHDS, and an EDRN subset used immunohistochemical detection of deficiency for mismatch repair gene proteins MLH1, MSH2, MSH6, and PMS2, and not PCR based assessment of microsatellite status.

Supplementary Table 3. Summary of study specific assessment of CpG island methylator phenotype (CIMP) status

Study	Panel genes	Marker positive definition	CIMP-high
CCFR	CACNA1G, IGF2, NEUROG1, RUNX3, and SOCS1	PMR > 10	>3 methylated markers
CPSII	CDKN2A, MLH1, CACNA1G, NEUROG1, RUNX3, SOCS1, IGF2, CRABP1	PMR >4 (>6 for CRABP1, IGF2)	≥5/8 methylated markers
DACHS	MGMT, MLH1, MINT1, MINT2, MINT31	N/A	≥3/5 methylated markers
DALS	MINT1, MINT2, MINT31, CDKN2A9, and hMLH1	N/A	≥3/5 methylated markers
EDRN*	N/A	N/A	N/A
EPIC_Sweden	CDKN2A, MLH1, CACNA1G, NEUROG1, RUNX3, SOCS1, IGF2, CRABP1	PMR > 10	≥5/8 methylated markers
HPFS	CDKN2A, MLH1, CACNA1G, NEUROG1, RUNX3, SOCS1, IGF2, CRABP1	PMR > 4 for CDKN2A, MLH1, CACNA1G, NEUROG1, RUNX3, SOCS1. PMR > 6 for CRABP1, IGF2	≥5/8 methylated markers
MCCS	CACNA1G, IGF2, NEUROG1, RUNX3, and SOCS1	PMR > 10	≥3/5 methylated markers
NFCCR*	N/A	N/A	N/A
NHS	CDKN2A, MLH1, CACNA1G, NEUROG1, RUNX3, SOCS1, IGF2, CRABP1	PMR > 4 for CDKN2A, MLH1, CACNA1G, NEUROG1, RUNX3, SOCS1. PMR > 6 for CRABP1, IGF2	≥5/8 methylated markers
NSHDS	CDKN2A, MLH1, CACNA1G, NEUROG1, RUNX3, SOCS1, IGF2, CRABP1	PMR > 10	≥5/8 methylated markers

*CIMP status was not assessed in EDRN and NFCCR.

Supplementary Table 4. Baseline characteristics according to contributing study

Characteristics	CCFR Australia	CCFR Ontario	CCFR Seattle	CPSII	DACHS	DALS	EDRN	EPIC Sweden	HPFS	MCCS	NFCCR	NHS	NSHDS
N	1731	2965	2557	1719	4710	2231	517	526	1176	1160	950	1961	682
Case-control status													
Cases	1555 (0.9)	1706 (0.6)	1812 (0.7)	790 (0.4)	1966 (0.4)	1083 (0.5)	188 (0.4)	145 (0.3)	585 (0.5)	490 (0.4)	489 (0.5)	764 (0.4)	299 (0.4)
Controls	176 (0.1)	1259 (0.4)	745 (0.3)	929 (0.6)	2744 (0.6)	1148 (0.5)	329 (0.6)	381 (0.7)	591 (0.5)	670 (0.6)	461 (0.5)	1197 (0.6)	383 (0.6)
Age, mean (SD)	49.6 (11.3)	58.1 (10.9)	54.9 (12.0)	67.6 (5.8)	68.6 (10.5)	64.1 (9.8)	60.5 (11.3)	54.7 (7.1)	69.1 (9.0)	59.5 (7.6)	60.0 (9.1)	65.5 (8.8)	55.8 (7.7)
Sex (%)													
Men	819 (47.3)	1570 (53.0)	1309 (51.2)	876 (51.0)	2831 (60.1)	1232 (55.2)	270 (52.2)	288 (54.8)	1176 (100.0)	610 (52.6)	578 (60.8)	0	267 (39.1)
Women	912 (52.7)	1395 (47.0)	1248 (48.8)	843 (49.0)	1879 (39.9)	999 (44.8)	247 (47.8)	238 (45.2)	0	550 (47.4)	372 (39.2)	1961 (100.0)	415 (60.9)
Body mass index (%)													
Normal weight (18.5-<25 kg/m^2)	705 (40.7)	1134 (38.2)	863 (33.8)	730 (42.5)	1688 (35.8)	810 (36.3)	188 (36.4)	243 (46.2)	519 (44.1)	349 (30.1)	278 (29.3)	872 (44.5)	285 (41.8)
Overweight (25-<30 kg/m^2)	643 (37.1)	1244 (42.0)	996 (39.0)	699 (40.7)	2223 (47.2)	942 (42.2)	162 (31.3)	224 (42.6)	543 (46.2)	556 (47.9)	421 (44.3)	682 (34.8)	306 (44.9)
Obese (≥30 kg/m^2)	383 (22.1)	587 (19.8)	698 (27.3)	290 (16.9)	799 (17.0)	479 (21.5)	167 (32.3)	59 (11.2)	114 (9.7)	255 (22.0	251 (26.4)	407 (20.8)	91 (13.3)
Tobacco smoking (%)													
Never	814 (47.0)	1218 (41.0)	1051 (41.1)	737 (42.9)	2279 (48.3)	1030 (46.2)	336 (65.0)	227 (43.2)	469 (40.0)	604 (52.1)	296 (31.1)	830 (42.3)	212 (31.1)
Past or current	780 (45.1)	1729 (58.3)	1212 (47.4)	982 (57.1)	2422 (51.4)	1198 (53.7)	168 (32.5)	201 (38.2)	646 (54.9)	556 (47.9)	607 (63.8)	1125 (57.4)	284 (41.6)
Unknown	137 (7.9)	18 (0.7)	294 (11.5)	0	9 (0.3)	3 (0.1)	13 (2.5)	98 (18.6)	61 (5.1)	0	47 (5.1)	6 (0.3)	186 (27.3)
Dietary intake													
Red meat, servings/day, mean (SD)	0.7 (0.7)	0.6 (0.6)	0.5 (0.5)	0.7 (0.6)	0.7 (0.4)	1.0 (0.8)	-	0.2 (0.1)	1.1 (0.8)	1.2 (0.8)	0.5 (0.5)	0.7 (0.5)	1.1 (1.0)
Education level													
Less high school graduate	559 (32.3)	724 (24.4)	173 (6.8)	92 (5.4)	793 (16.8)	304 (13.6)	17 (3.3)	227 (43.2)	0	710 (61.2)	346 (36.4)	0	213 (31.2)
High school grad	329 (19.0)	520 (17.5)	583 (22.8)	412 (24.0)	2447 (52.0)	622 (27.9)	105 (20.3)	74 (14.1)	0	109 (9.4)	142 (14.9)	0	134 (19.6)
Vocational or technical school/some college/university	430 (24.8)	980 (33.1)	886 (34.6)	502 (29.2)	533 (11.3)	730 (32.7)	52 (10.1)	156 (29.7)	0	121 (10.4)	306 (32.2)	525 (26.8)	60 (8.8)
Undergraduate or graduate degree	412 (23.8)	724 (24.4)	915 (35.8)	709 (41.2)	920 (19.5)	574 (25.7)	230 (44.5)	66 (12.5)	1176 (100.0)	220 (19.0)	110 (11.6)	1308 (66.7)	93 (13.6)
Missing	1 (0.1)	17 (0.6)	0	4 (0.2)	17 (0.4)	1 (0.1)	113 (21.9)	3 (0.6)	0	0	46 (4.8)	128 (6.5)	182 (26.7)

N (%) shown unless otherwise indicated. CCFR = Colon Cancer Family Registry; CPSII = Cancer Prevention Study II; DACHS = Darmkrebs: Chancen der Verhutung durch Screening Study; DALS = Diet Activity and Lifestyle Study; EDRN = Early Detection Research Network; EPIC = European Prospective Investigation into Cancer and Nutrition; HPFS = Health Professionals Follow-up Study; MCCS = Melbourne Collaborative Cohort Study; NFCCR = Newfoundland Familial Colorectal Cancer Study; NHS = Nurses' Health Study; NSHDS = Northern Sweden Health and Disease Study.

Supplementary Table 5. Association between body mass index and molecular subtypes of colorectal cancer according to study

	Microsatellite instability			CpG island methylator phenotype			*BRAF*			*KRAS*		
		Per 5 kg/m^2			Per 5 kg/m^2			Per 5 kg/m^2			Per 5 kg/m^2	
Study	Marker	OR (95% CI)	P-diff.	Marker	OR (95% CI)	P-diff.	Marker	OR (95% CI)	P-diff.	Marker	OR (95% CI)	P-diff.
CCFR_Australia	MSS/MSI-L	1.13 (0.95-1.35)	0.04	CIMP-low/negative	1.09 (0.92-1.31)	0.02	*BRAF*-wildtype	1.08 (0.91-1.28)	0.34	*KRAS*-wildtype	1.14 (0.92-1.42)	0.13
	MSI-H	1.00 (0.83-1.21)		CIMP-high	1.44 (1.08-1.91)		*BRAF*-mutated	0.99 (0.79-1.24)		*KRAS*-mutated	0.95 (0.73-1.24)	
CCFR_Ontario	MSS/MSI-L	1.05 (0.96-1.15)	0.79	CIMP-low/negative	1.00 (0.90-1.11)	0.05	*BRAF*-wildtype	1.04 (0.94-1.14)	0.74	*KRAS*-wildtype	1.05 (0.95-1.16)	0.47
	MSI-H	1.03 (0.89-1.20)		CIMP-high	1.16 (0.94-1.42)		*BRAF*-mutated	1.04 (0.88-1.24)		*KRAS*-mutated	1.01 (0.90-1.14)	
CCFR_Seattle	MSS/MSI-L	1.18 (1.08-1.28)	0.48	CIMP-low/negative	1.17 (1.06-1.29)	0.8	*BRAF*-wildtype	1.19 (1.09-1.29)	0.46	*KRAS*-wildtype	1.18 (1.08-1.28)	0.45
	MSI-H	1.23 (1.08-1.39)		CIMP-high	1.19 (1.03-1.38)		*BRAF*-mutated	1.23 (1.07-1.42)		*KRAS*-mutated	1.22 (1.10-1.35)	
CPSII	MSS/MSI-L	1.37 (1.19-1.57)	0.51	CIMP-low/negative	1.31 (1.15-1.49)	0.98	*BRAF*-wildtype	1.35 (1.18-1.55)	0.77	*KRAS*-wildtype	1.28 (1.10-1.49)	0.06
	MSI-H	1.26 (0.99-1.60)		CIMP-high	1.33 (1.10-1.62)		*BRAF*-mutated	1.43 (1.13-1.82)		*KRAS*-mutated	1.57 (1.30-1.89)	
DACHS	MSS/MSI-L	1.21 (1.11-1.31)	0.03	CIMP-low/negative	1.20 (1.11-1.30)	0.0004	*BRAF*-wildtype	1.25 (1.15-1.35)	0.1	*KRAS*-wildtype	1.32 (1.21-1.44)	0.03
	MSI-H	1.46 (1.23-1.74)		CIMP-high	1.54 (1.34-1.77)		*BRAF*-mutated	1.47 (1.20-1.80)		*KRAS*-mutated	1.15 (1.02-1.29)	
DALS	MSS/MSI-L	1.30 (1.19-1.44)	0.69	CIMP-low/negative	1.34 (1.21-1.48)	0.64	*BRAF*-wildtype	1.31 (1.19-1.45)	0.99	*KRAS*-wildtype	1.30 (1.17-1.44)	0.45
	MSI-H	1.33 (1.13-1.56)		CIMP-high	1.25 (1.03-1.51)		*BRAF*-mutated	1.29 (1.04-1.61)		*KRAS*-mutated	1.36 (1.19-1.54)	
EDRN	MSS/MSI-L	1.55 (1.27-1.89)	0.33	CIMP-low/negative	N/A	N/A	*BRAF*-wildtype	1.48 (1.14-1.90)	0.15	*KRAS*-wildtype	1.30 (0.98-1.72)	0.5
	MSI-H	1.27 (0.88-1.81)		CIMP-high	N/A	N/A	*BRAF*-mutated	0.80 (0.43-1.49)		*KRAS*-mutated	1.41 (1.05-1.90)	
EPIC_Sweden	MSS/MSI-L	1.00 (0.74-1.35)	0.12	CIMP-low/negative	0.93 (0.68-1.28)	0.02	*BRAF*-wildtype	1.08 (0.79-1.47)	0.11	*KRAS*-wildtype	1.14 (0.84-1.55)	0.88
	MSI-H	1.53 (0.89-2.60)		CIMP-high	1.78 (1.09-2.93)		*BRAF*-mutated	1.60 (1.02-2.51)		*KRAS*-mutated	1.19 (0.74-1.90)	
MCCS	MSS/MSI-L	1.16 (0.99-1.37)	0.39	CIMP-low/negative	1.19 (1.01-1.40)	0.33	*BRAF*-wildtype	1.16 (0.98-1.36)	0.6	KRAS-wildtype	1.19 (1.00-1.41)	0.34
	MSI-H	0.97 (0.69-1.35)		CIMP-high	0.91 (0.62-1.34)		*BRAF*-mutated	1.03 (0.73-1.44)		*KRAS*-mutated	1.05 (0.82-1.35)	
NFCCR	MSS/MSI-L	1.15 (0.99-1.34)	0.1	CIMP-low/negative	N/A	N/A	*BRAF*-wildtype	1.16 (0.99-1.35)	0.64	*KRAS*-wildtype	N/A	N/A
	MSI-H	0.86 (0.61-1.21)		CIMP-high	N/A	N/A	*BRAF*-mutated	1.07 (0.75-1.52)		*KRAS*-mutated	N/A	N/A

NSHDS	MSS/MSI-L	1.01 (0.80-1.27)	0.82	CIMP-low/negative	1.00 (0.79-1.27)	0.92	*BRAF*-wildtype	0.96 (0.75-1.22)	0.33	*KRAS*-wildtype	1.10 (0.87-1.39)	0.02
	MSI-H	1.11 (0.71-1.72)		CIMP-high	1.05 (0.71-1.55)		*BRAF*-mutated	1.15 (0.81-1.65)		*KRAS*-mutated	0.72 (0.49-1.05)	
HPFS	MSS/MSI-L	1.19 (0.98-1.44)	0.02	CIMP-low/negative	1.25 (1.03-1.51)	0.38	*BRAF*-wildtype	1.25 (1.04-1.51)	0.38	*KRAS*-wildtype	1.39 (1.13-1.72)	0.1
	MSI-H	1.87 (1.30-2.70)		CIMP-high	1.09 (0.76-1.56)		*BRAF*-mutated	1.54 (0.99-2.37)		*KRAS*-mutated	1.13 (0.89-1.44)	
NHS	MSS/MSI-L	1.12 (1.00-1.25)	0.85	CIMP-low/negative	1.11 (0.99-1.24)	0.84	*BRAF*-wildtype	1.12 (1.01-1.25)	0.67	*KRAS*-wildtype	1.13 (1.01-1.27)	0.68
	MSI-H	1.10 (0.92-1.32)		CIMP-high	1.06 (0.90-1.25)		*BRAF*-mutated	1.14 (0.96-1.34)		*KRAS*-mutated	1.09 (0.95-1.26)	

CI = confidence interval; CIMP = CpG island methylator phenotype; CRC = colorectal cancer; MSI = microsatellite instability; OR = odds ratio; p-diff. = p-difference. CCFR = Colon Cancer Family Registry; CPSII = Cancer Prevention Study II; DACHS = Darmkrebs: Chancen der Verhutung durch Screening Study; DALS = Diet Activity and Lifestyle Study; EDRN = Early Detection Research Network; EPIC = European Prospective Investigation into Cancer and Nutrition; HPFS = Health Professionals Follow-up Study; MCCS = Melbourne Collaborative Cohort Study; NFCCR = Newfoundland Familial Colorectal Cancer Study; NHS = Nurses' Health Study; NSHDS = Northern Sweden Health and Disease Study. Controls are used as reference for all odds ratios. Odds ratios are adjusted for study, age, sex, smoking status, education, and red meat intake. Case-only analyses used to calculate p-difference.

Supplementary Table 6. Association between body mass index and molecular subtypes of colorectal cancer from meta-analyzing (random effect models) individual study estimates

Exposure		Microsatellite instability		CpG island methylator phenotype		*BRAF*		*KRAS*	
	CRC	MSS/MSI-L	MSI-H	CIMP-low/negative	CIMP-high	*BRAF*-wildtype	*BRAF*-mutated	*KRAS*-wildtype	*KRAS*-mutated
	OR (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)
Both sexes									
N cases	11,872	8,967	1,809	7,160	1,386	9,423	1,297	6,011	2,961
18.5-<25 kg/m^2	1 (reference)	1 (reference)	1 (reference)	1 (reference)	1 (reference)	1 (reference)	1 (reference)	1 (reference)	1 (reference)
25-<30 kg/m^2	1.19 (1.12-1.27)	1.22 (1.14-1.31)	1.35 (1.16-1.58)	1.24 (1.14-1.35)	1.26 (1.05-1.52)	1.24 (1.16-1.32)	1.24 (1.09-1.40)	1.28 (1.20-1.37)	1.20 (1.10-1.31)
≥30 kg/m^2	1.47 (1.32-1.64)	1.54 (1.34-1.75)	1.74 (1.39-2.17)	1.46 (1.30-1.64)	2.00 (1.55-2.60)	1.56 (1.38-1.75)	1.80 (1.42-2.28)	1.62 (1.43-1.83)	1.50 (1.25-1.80)
p-trend	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
Per 5 kg/m^2	1.18 (1.12-1.25)	1.20 (1.13-1.27)	1.23 (1.11-1.36)	1.17 (1.10-1.24)	1.26 (1.14-1.40)	1.19 (1.13-1.26)	1.24 (1.13-1.35)	1.21 (1.15-1.28)	1.18 (1.08-1.30)
p-difference		0.93		0.14		0.32		0.27	
Men									
N cases	6,209	4,923	764	3,954	528	5,118	474	3,144	1,563
18.5-<25 kg/m^2	1 (reference)	1 (reference)	1 (reference)	1 (reference)	1 (reference)	1 (reference)	1 (reference)	1 (reference)	1 (reference)
25-<30 kg/m^2	1.16 (1.01-1.33)	1.31 (1.17-1.48)	1.34 (1.02-1.76)	1.32 (1.12-1.55)	1.34 (1.10-1.64)	1.28 (1.15-1.42)	1.84 (1.36-2.49)	1.36 (1.20-1.55)	1.23 (1.09-1.39)
≥30 kg/m^2	1.51 (1.34-1.70)	1.61 (1.45-1.78)	1.85 (1.45-2.38)	1.57 (1.38-1.78)	2.25 (1.41-3.60)	1.62 (1.43-1.84)	2.72 (2.17-3.42)	1.63 (1.38-1.92)	1.58 (1.34-1.87)
p-trend	<0.0001	<0.0001	0.0003	<0.0001	<0.0001	<0.0001	0.0002	<0.0001	<0.0001
Per 5 kg/m^2	1.23 (1.17-1.30)	1.25 (1.19-1.32)	1.25 (1.08-1.45)	1.23 (1.14-1.32)	1.27 (1.15-1.40)	1.25 (1.17-1.33)	1.34 (1.16-1.54)	1.25 (1.17-1.33)	1.24 (1.13-1.37)
p-difference		0.39		0.98		0.98		0.63	
Women									
N cases	5,663	4,044	1,045	3,206	858	4,305	823	2,867	1,398
18.5-<25 kg/m^2	1 (reference)	1 (reference)	1 (reference)	1 (reference)	1 (reference)	1 (reference)	1 (reference)	1 (reference)	1 (reference)
25-<30 kg/m^2	1.19 (1.08-1.30)	1.24 (1.12-1.38)	1.57 (1.31-1.87)	1.26 (1.15-1.38)	1.39 (1.05-1.84)	1.38 (1.14-1.66)	1.40 (1.17-1.69)	1.32 (1.17-1.48)	1.28 (1.14-1.45)
≥30 kg/m^2	1.43 (1.18-1.75)	1.56 (1.16-2.12)	1.77 (1.14-2.73)	1.48 (1.27-1.71)	2.27 (1.65-3.14)	1.63 (1.08-2.47)	2.02 (1.46-2.81)	1.71 (1.44-2.04)	1.71 (1.25-2.32)
p-trend	<0.0001	0.0001	0.008	<0.0001	0.0003	<0.0001	0.001	<0.0001	0.0004
Per 5 kg/m^2	1.15 (1.06-1.25)	1.18 (1.08-1.30)	1.21 (1.06-1.38)	1.14 (1.06-1.24)	1.29 (1.12-1.48)	1.19 (1.09-1.31)	1.24 (1.09-1.41)	1.21 (1.12-1.31)	1.19 (1.05-1.36)
p-difference		0.73		0.1		0.15		0.34	

CI = confidence interval; CIMP = CpG island methylator phenotype; CRC = colorectal cancer; MSI = microsatellite instability; OR = odds ratio. Controls are used as reference for all odds ratios. Odds ratios are adjusted for study, age, sex, smoking status, education, and red meat intake. Logistic regression models were used to calculate study specific ORs which were then pooled using random effects meta-analysis models.

Supplementary Table 7. Association between body mass index and Jass group defined subtypes of colorectal cancer

Jass group	N cases	Body mass index Normal weight 18.5-<25 kg/m^2	Overweight 25-<30 kg/m^2	Obese ≥30 kg/m^2	P-trend	Per 5 kg/m^2	P-diff.
Both sexes							
Type 1 (MSI-H, CIMP-high, *BRAF*-mutated, *KRAS*-wildtype), OR (95% CI)	453	1 (ref.)	1.12 (0.90-1.41)	1.60 (1.24-2.07)	0.002	1.24 (1.12-1.36)	0.12
Type 2 (MSS or MSI-L, CIMP-high, *BRAF*-mutated, *KRAS*-wildtype), OR (95% CI)	207	1 (ref.)	1.27 (0.91-1.76)	1.94 (1.35-2.79)	0.0004	1.33 (1.17-1.52)	0.05
Type 3 (MSS or MSI-L, CIMP-low/negative, *BRAF*-wildtype, *KRAS*-mutated), OR (95% CI)	1915	1 (ref.)	1.12 (1.00-1.26)	1.34 (1.17-1.54)	<0.0001	1.15 (1.09-1.22)	0.49
Type 4 (MSS or MSI-L, CIMP-low/negative, *BRAF*-wildtype, *KRAS*-wildtype), OR (95% CI)	3292	1 (ref.)	1.24 (1.13-1.37)	1.42 (1.26-1.59)	<0.0001	1.18 (1.13-1.24)	ref.
Type 5 (MSI-H, CIMP-low/negative, *BRAF*-wildtype, *KRAS*-wildtype), OR (95% CI)	234	1 (ref.)	1.03 (0.76-1.40)	1.26 (0.88-1.81)	0.23	1.04 (0.90-1.20)	0.08
Men							
Type 1 (MSI-H, CIMP-high, *BRAF*-mutated, *KRAS*-wildtype), OR (95% CI)	116	1 (ref.)	1.30 (0.84-2.01)	1.62 (0.92-2.86)	0.09	1.36 (1.07-1.74)	0.64
Type 2 (MSS or MSI-L, CIMP-high, *BRAF*-mutated, *KRAS*-wildtype), OR (95% CI)	74	1 (ref.)	1.24 (0.71-2.17)	1.78 (0.91-3.49)	0.1	1.35 (1.02-1.77)	0.53
Type 3 (MSS or MSI-L, CIMP-low/negative, *BRAF*-wildtype, *KRAS*-mutated), OR (95% CI)	987	1 (ref.)	1.10 (0.93-1.29)	1.37 (1.12-1.68)	0.004	1.20 (1.10-1.32)	0.50
Type 4 (MSS or MSI-L, CIMP-low/negative, *BRAF*-wildtype, *KRAS*-wildtype), OR (95% CI)	1902	1 (ref.)	1.27 (1.11-1.44)	1.43 (1.21-1.69)	<0.0001	1.24 (1.15-1.33)	ref.
Type 5 (MSI-H, CIMP-low/negative, *BRAF*-wildtype, *KRAS*-wildtype), OR (95% CI)	119	1 (ref.)	0.75 (0.50-1.14)	1.04 (0.62-1.74)	0.86	0.96 (0.75-1.22)	0.13
Women							
Type 1 (MSI-H, CIMP-high, *BRAF*-mutated, *KRAS*-wildtype), OR (95% CI)	337	1 (ref.)	1.06 (0.82-1.38)	1.62 (1.21-2.16)	0.003	1.21 (1.08-1.35)	0.14
Type 2 (MSS or MSI-L, CIMP-high, *BRAF*-mutated, *KRAS*-wildtype), OR (95% CI)	133	1 (ref.)	1.24 (0.82-1.88)	2.01 (1.30-3.12)	0.002	1.32 (1.13-1.54)	0.04
Type 3 (MSS or MSI-L, CIMP-low/negative, *BRAF*-wildtype, *KRAS*-mutated), OR (95% CI)	928	1 (ref.)	1.15 (0.98-1.36)	1.34 (1.11-1.63)	0.002	1.13 (1.06-1.21)	0.57
Type 4 (MSS or MSI-L, CIMP-low/negative, *BRAF*-wildtype, *KRAS*-wildtype), OR (95% CI)	1390	1 (ref.)	1.23 (1.07-1.42)	1.44 (1.22-1.70)	<0.0001	1.16 (1.09-1.23)	ref.
Type 5 (MSI-H, CIMP-low/negative, *BRAF*-wildtype, *KRAS*-wildtype), OR (95% CI)	115	1 (ref.)	1.47 (0.96-2.25)	1.50 (0.91-2.47)	0.08	1.08 (0.90-1.29)	0.44

CI = confidence interval; ref. = reference; CIMP = CpG island methylator phenotype; MSI = microsatellite instability; mut = mutated; N = number; OR = odds ratio; wild = wild type. Controls are used as reference for all odds ratios. Odds ratios are adjusted for study, age, sex, smoking status, education, and red meat intake. Multinomial logistic regression was used to compare each molecular subtype to the reference group (Type 4; p-difference).

Supplementary Table 8. Association between body mass index and Jass group defined subtypes of colorectal cancer according to study design (case-control or cohort)

Jass group	Per 5 kg/m^2
Case-control (CCFR_Australia, CCFR_Ontario, CCFR_Seattle, DACHS, DALS)	
Type 1 (MSI-H, CIMP-high, *BRAF*-mutated, *KRAS*-wildtype), OR (95% CI)	1.24 (1.09-1.41)
Type 2 (MSS or MSI-L, CIMP-high, *BRAF*-mutated, *KRAS*-wildtype), OR (95% CI)	1.46 (1.25-1.70)
Type 3 (MSS or MSI-L, CIMP-low/negative, *BRAF*-wildtype, *KRAS*-mutated), OR (95% CI)	1.17 (1.09-1.25)
Type 4 (MSS or MSI-L, CIMP-low/negative, *BRAF*-wildtype, *KRAS*-wildtype), OR (95% CI)	1.19 (1.12-1.25)
Type 5 (MSI-H, CIMP-low/negative, *BRAF*-wildtype, *KRAS*-wildtype), OR (95% CI)	1.02 (0.87-1.21)
Cohort (CPSII, EPIC_Sweden, HPFS, NHS, MCCS, NSHDS)	
Type 1 (MSI-H, CIMP-high, *BRAF*-mutated, *KRAS*-wildtype), OR (95% CI)	1.24 (1.06-1.44)
Type 2 (MSS or MSI-L, CIMP-high, *BRAF*-mutated, *KRAS*-wildtype), OR (95% CI)	1.07 (0.83-1.37)
Type 3 (MSS or MSI-L, CIMP-low/negative, *BRAF*-wildtype, *KRAS*-mutated), OR (95% CI)	1.12 (1.02-1.24)
Type 4 (MSS or MSI-L, CIMP-low/negative, *BRAF*-wildtype, *KRAS*-wildtype), OR (95% CI)	1.18 (1.09-1.28)
Type 5 (MSI-H, CIMP-low/negative, *BRAF*-wildtype, *KRAS*-wildtype), OR (95% CI)	1.08 (0.79-1.47)

CI = confidence interval; CIMP = CpG island methylator phenotype; MSI = microsatellite instability; OR = odds ratio; wild = wild type. Controls are used as reference for all odds ratios. Odds ratios are adjusted for study, age, sex, smoking status, education, and red meat intake. CCFR = Colon Cancer Family Registry; CPSII = Cancer Prevention Study II; DACHS = Darmkrebs: Chancen der Verhutung durch Screening Study; DALS = Diet Activity and Lifestyle Study; EDRN = Early Detection Research Network; EPIC = European Prospective Investigation into Cancer and Nutrition; HPFS = Health Professionals Follow-up Study; MCCS = Melbourne Collaborative Cohort Study; NFCCR = Newfoundland Familial Colorectal Cancer Study; NHS = Nurses' Health Study; NSHDS = Northern Sweden Health and Disease Study.

Supplementary Table 9. Association between body mass index and Jass group defined subtypes of colorectal cancer from meta-analyzing (random effect models) individual study estimates

Jass group	N cases	Per 5 kg/m^2
Both sexes		
Type 1 (MSI-H, CIMP-positive, *BRAF*-mutated, *KRAS*-wildtype), OR (95% CI)	453	1.30 (1.08-1.55)
Type 2 (MSS or MSI-L, CIMP-positive, *BRAF*-mutated, *KRAS*-wildtype), OR (95% CI)	207	1.36 (1.12-1.66)
Type 3 (MSS or MSI-L, CIMP-negative, *BRAF*-wildtype, *KRAS*-mutated), OR (95% CI)	1915	1.14 (1.02-1.27)
Type 4 (MSS or MSI-L, CIMP-negative, *BRAF*-wildtype, *KRAS*-wildtype), OR (95% CI)	3292	1.18 (1.13-1.24)
Type 5 (MSI-H, CIMP-negative, *BRAF*-wildtype, *KRAS*-wildtype), OR (95% CI)	234	1.03 (0.89-1.20)

CI = confidence interval; CIMP = CpG island methylator phenotype; MSI = microsatellite instability; mut = mutated; N = number; OR = odds ratio; wild = wild type. Controls are used as reference for all odds ratios. Odds ratios are adjusted for study, age, sex, smoking status, education, and red meat intake. Logistic regression models were used to calculate study specific ORs which were then pooled using random effects meta-analysis models.

References

1. Newcomb PA, Baron J, Cotterchio M, et al. Colon Cancer Family Registry: an international resource for studies of the genetic epidemiology of colon cancer. *Cancer Epidemiol Biomarkers Prev.* 2007;16(11):2331-2343.
2. Calle EE, Rodriguez C, Jacobs EJ, et al. The American Cancer Society Cancer Prevention Study II Nutrition Cohort. *Cancer.* 2002;94(9):2490-2501.
3. Campbell PT, Deka A, Briggs P, et al. Establishment of the cancer prevention study II nutrition cohort colorectal tissue repository. *Cancer Epidemiol Biomarkers Prev.* 2014;23(12):2694-2702.
4. Brenner H, Chang-Claude J, Jansen L, Knebel P, Stock C, Hoffmeister M. Reduced risk of colorectal cancer up to 10 years after screening, surveillance, or diagnostic colonoscopy. *Gastroenterology.* 2014;146(3):709-717.
5. Jia M, Jansen L, Walter V, et al. No association of CpG island methylator phenotype and colorectal cancer survival: population-based study. *Br J Cancer.* 2016;115(11):1359-1366.
6. Slattery ML. Physical activity and colorectal cancer. *Sports Med.* 2004;34(4):239-252.
7. Amin W, Singh H, Dzubinski LA, Schoen RE, Parwani AV. Design and utilization of the colorectal and pancreatic neoplasm virtual biorepository: An early detection research network initiative. *Journal of pathology informatics.* 2010;1:22.
8. Riboli E, Kaaks R. The EPIC Project: rationale and study design. European Prospective Investigation into Cancer and Nutrition. *International Journal of Epidemiology.* 1997;26(suppl 1):S6-14.
9. Giovannucci E, Ascherio A, Rimm EB, Colditz GA, Stampfer MJ, Willett WC. Physical activity, obesity, and risk for colon cancer and adenoma in men. *Annals of internal medicine.* 1995;122(5):327-334.
10. Giles GG, English DR. The Melbourne Collaborative Cohort Study. *IARC scientific publications.* 2002;156:69-70.
11. Green RC, Green JS, Buehler SK, et al. Very high incidence of familial colorectal cancer in Newfoundland: a comparison with Ontario and 13 other population-based studies. *Familial cancer.* 2007;6(1):53-62.
12. Belanger CF, Hennekens CH, Rosner B, Speizer FE. The nurses' health study. *The American journal of nursing.* 1978;78(6):1039-1040.
13. Dahlin AM, Palmqvist R, Henriksson ML, et al. The role of the CpG island methylator phenotype in colorectal cancer prognosis depends on microsatellite instability screening status. *Clinical cancer research : an official journal of the American Association for Cancer Research.* 2010;16(6):1845-1855.
14. Labadie JD, Harrison TA, Banbury B, et al. Postmenopausal Hormone Therapy and Colorectal Cancer Risk by Molecularly Defined Subtypes and Tumor Location. *JNCI cancer spectrum.* 2020;4(5):pkaa042.
15. Hidaka A, Harrison TA, Cao Y, et al. Intake of Dietary Fruit, Vegetables, and Fiber and Risk of Colorectal Cancer According to Molecular Subtypes: A Pooled Analysis of 9 Studies. *Cancer Res.* 2020;80(20):4578-4590.
16. Van Guelpen B, Dahlin AM, Hultdin J, et al. One-carbon metabolism and CpG island methylator phenotype status in incident colorectal cancer: a nested case-referent study. *Cancer causes & control : CCC.* 2010;21(4):557-566.
17. Buchanan DD, Clendenning M, Rosty C, et al. Tumor testing to identify lynch syndrome in two Australian colorectal cancer cohorts. *Journal of gastroenterology and hepatology.* 2017;32(2):427-438.
18. Cicek MS, Lindor NM, Gallinger S, et al. Quality assessment and correlation of microsatellite instability and immunohistochemical markers among population- and clinic-based colorectal tumors results from the Colon Cancer Family Registry. *J Mol Diagn.* 2011;13(3):271-281.

19. Walsh MD, Buchanan DD, Cummings MC, et al. Lynch syndrome-associated breast cancers: clinicopathologic characteristics of a case series from the colon cancer family registry. *Clinical cancer research : an official journal of the American Association for Cancer Research.* 2010;16(7):2214-2224.
20. Lindor NM, Burgart LJ, Leontovich O, et al. Immunohistochemistry versus microsatellite instability testing in phenotyping colorectal tumors. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology.* 2002;20(4):1043-1048.
21. Woods MO, Hyde AJ, Curtis FK, et al. High frequency of hereditary colorectal cancer in Newfoundland likely involves novel susceptibility genes. *Clinical cancer research : an official journal of the American Association for Cancer Research.* 2005;11(19 Pt 1):6853-6861.
22. Raptis S, Mrkonjic M, Green RC, et al. MLH1 -93G>A promoter polymorphism and the risk of microsatellite-unstable colorectal cancer. *Journal of the National Cancer Institute.* 2007;99(6):463-474.
23. Ogino S, Brahmandam M, Cantor M, et al. Distinct molecular features of colorectal carcinoma with signet ring cell component and colorectal carcinoma with mucinous component. *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc.* 2006;19(1):59-68.
24. Boland CR, Thibodeau SN, Hamilton SR, et al. A National Cancer Institute Workshop on Microsatellite Instability for cancer detection and familial predisposition: development of international criteria for the determination of microsatellite instability in colorectal cancer. *Cancer Res.* 1998;58(22):5248-5257.
25. Samowitz WS, Slattery ML, Kerber RA. Microsatellite instability in human colonic cancer is not a useful clinical indicator of familial colorectal cancer. *Gastroenterology.* 1995;109(6):1765-1771.
26. Samowitz WS, Slattery ML. Transforming growth factor-beta receptor type 2 mutations and microsatellite instability in sporadic colorectal adenomas and carcinomas. *The American journal of pathology.* 1997;151(1):33-35.
27. Samowitz WS, Slattery ML, Potter JD, Leppert MF. BAT-26 and BAT-40 instability in colorectal adenomas and carcinomas and germline polymorphisms. *The American journal of pathology.* 1999;154(6):1637-1641.
28. Slattery ML, Curtin K, Anderson K, et al. Associations between cigarette smoking, lifestyle factors, and microsatellite instability in colon tumors. *Journal of the National Cancer Institute.* 2000;92(22):1831-1836.
29. Hoffmeister M, Bläker H, Kloor M, et al. Body mass index and microsatellite instability in colorectal cancer: a population-based study. *Cancer Epidemiol Biomarkers Prev.* 2013;22(12):2303-2311.
30. Findeisen P, Kloor M, Merx S, et al. T25 repeat in the 3' untranslated region of the CASP2 gene: a sensitive and specific marker for microsatellite instability in colorectal cancer. *Cancer Res.* 2005;65(18):8072-8078.
31. Boland CR, Goel A. Microsatellite Instability in Colorectal Cancer. *Gastroenterology*

Colon Cancer: An Update and Future Directions. 2010;138(6):2073-2087.
32. Shia J. Immunohistochemistry versus microsatellite instability testing for screening colorectal cancer patients at risk for hereditary nonpolyposis colorectal cancer syndrome. Part I. The utility of immunohistochemistry. *J Mol Diagn.* 2008;10(4):293-300.
33. Buchanan DD, Sweet K, Drini M, et al. Risk factors for colorectal cancer in patients with multiple serrated polyps: a cross-sectional case series from genetics clinics. *PLoS One.* 2010;5(7):e11636.
34. Stewart CJ, Leung Y, Walsh MD, Walters RJ, Young JP, Buchanan DD. KRAS mutations in ovarian low-grade endometrioid adenocarcinoma: association with concurrent endometriosis. *Human pathology.* 2012;43(8):1177-1183.

35. Alsop K, Mead L, Smith LD, et al. Low somatic K-ras mutation frequency in colorectal cancer diagnosed under the age of 45 years. *European journal of cancer (Oxford, England : 1990).* 2006;42(10):1357-1361.
36. Loughrey MB, Waring PM, Tan A, et al. Incorporation of somatic BRAF mutation testing into an algorithm for the investigation of hereditary non-polyposis colorectal cancer. *Familial cancer.* 2007;6(3):301-310.
37. Rosty C, Buchanan DD, Walsh MD, et al. Phenotype and polyp landscape in serrated polyposis syndrome: a series of 100 patients from genetics clinics. *Am J Surg Pathol.* 2012;36(6):876-882.
38. Benlloch S, Payá A, Alenda C, et al. Detection of BRAF V600E mutation in colorectal cancer: comparison of automatic sequencing and real-time chemistry methodology. *J Mol Diagn.* 2006;8(5):540-543.
39. Myte R, Gylling B, Häggström J, et al. One-carbon metabolism biomarkers and genetic variants in relation to colorectal cancer risk by KRAS and BRAF mutation status. *PLoS One.* 2018;13(4):e0196233.
40. Samowitz WS, Albertsen H, Herrick J, et al. Evaluation of a large, population-based sample supports a CpG island methylator phenotype in colon cancer. *Gastroenterology.* 2005;129(3):837-845.
41. Samowitz WS, Curtin K, Schaffer D, Robertson M, Leppert M, Slattery ML. Relationship of Ki-ras mutations in colon cancers to tumor location, stage, and survival: a population-based study. *Cancer Epidemiol Biomarkers Prev.* 2000;9(11):1193-1197.
42. Ogino S, Kawasaki T, Brahmandam M, et al. Sensitive sequencing method for KRAS mutation detection by Pyrosequencing. *J Mol Diagn.* 2005;7(3):413-421.
43. Ogino S, Meyerhardt JA, Cantor M, et al. Molecular alterations in tumors and response to combination chemotherapy with gefitinib for advanced colorectal cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research.* 2005;11(18):6650-6656.
44. Imamura Y, Lochhead P, Yamauchi M, et al. Analyses of clinicopathological, molecular, and prognostic associations of KRAS codon 61 and codon 146 mutations in colorectal cancer: cohort study and literature review. *Molecular cancer.* 2014;13:135.
45. Ogino S, Kawasaki T, Kirkner GJ, Kraft P, Loda M, Fuchs CS. Evaluation of markers for CpG island methylator phenotype (CIMP) in colorectal cancer by a large population-based sample. *J Mol Diagn.* 2007;9(3):305-314.
46. Ogino S, Kawasaki T, Kirkner GJ, Loda M, Fuchs CS. CpG island methylator phenotype-low (CIMP-low) in colorectal cancer: possible associations with male sex and KRAS mutations. *J Mol Diagn.* 2006;8(5):582-588.
47. Weisenberger DJ, Levine AJ, Long TI, et al. Association of the colorectal CpG island methylator phenotype with molecular features, risk factors, and family history. *Cancer Epidemiol Biomarkers Prev.* 2015;24(3):512-519.
48. Weisenberger DJ, Siegmund KD, Campan M, et al. CpG island methylator phenotype underlies sporadic microsatellite instability and is tightly associated with BRAF mutation in colorectal cancer. *Nat Genet.* 2006;38(7):787-793.
49. English DR, Young JP, Simpson JA, et al. Ethnicity and risk for colorectal cancers showing somatic BRAF V600E mutation or CpG island methylator phenotype. *Cancer Epidemiol Biomarkers Prev.* 2008;17(7):1774-1780.
50. Eads CA, Danenberg KD, Kawakami K, et al. MethyLight: a high-throughput assay to measure DNA methylation. *Nucleic Acids Res.* 2000;28(8):E32.
51. Carr PR, Jansen L, Bienert S, et al. Associations of red and processed meat intake with major molecular pathological features of colorectal cancer. *Eur J Epidemiol.* 2017;32(5):409-418.

52. Warth A, Kloor M, Schirmacher P, Bläker H. Genetics and epigenetics of small bowel adenocarcinoma: the interactions of CIN, MSI, and CIMP. *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc.* 2011;24(4):564-570.
53. Park SJ, Rashid A, Lee JH, Kim SG, Hamilton SR, Wu TT. Frequent CpG island methylation in serrated adenomas of the colorectum. *The American journal of pathology.* 2003;162(3):815-822.
54. Rashid A, Shen L, Morris JS, Issa JP, Hamilton SR. CpG island methylation in colorectal adenomas. *The American journal of pathology.* 2001;159(3):1129-1135.
55. Jass JR. Classification of colorectal cancer based on correlation of clinical, morphological and molecular features. *Histopathology.* 2007;50(1):113-130.
56. Phipps AI, Limburg PJ, Baron JA, et al. Association between molecular subtypes of colorectal cancer and patient survival. *Gastroenterology.* 2015;148(1):77-87.e72.