Metabolic Risk Factors and Molecular Subtypes of Colorectal Cancer

Robin Myte
Table of Contents

Abstract ........................................................................................................... i
Abbreviations .................................................................................................. ii
Original Papers ............................................................................................... iii
Related Publications ......................................................................................... iv
Overview of Papers ............................................................................................ v
Populärvetenskaplig sammanfattning ............................................................... vi

Background ...................................................................................................... 1
Colorectal cancer ............................................................................................... 1
  Incidence and Mortality .................................................................................. 1
  Prognosis and treatment ............................................................................... 2
  Risk and preventive factors ........................................................................... 3
  Screening ....................................................................................................... 4
  Colorectal carcinogenesis and molecular subtypes ........................................... 5
Energy metabolism .......................................................................................... 7
  Energy metabolism and body fatness ............................................................. 7
  Energy metabolism and CRC ......................................................................... 8
One-carbon metabolism ................................................................................... 10
  The folate and methionine cycles ................................................................ 10
  One-carbon metabolism and CRC .............................................................. 12
Heterogeneity in CRC etiology ....................................................................... 14
  Body fatness ................................................................................................ 14
  Adult height .................................................................................................. 14
  Smoking ....................................................................................................... 14
  Alcohol ........................................................................................................ 15
  Red and processed meats ............................................................................. 15
  Physical activity ............................................................................................ 15
  Hormone replacement therapy .................................................................... 15
  Aspirin use .................................................................................................... 15
  Other factors, including one-carbon metabolism ......................................... 15
Summary and knowledge gap ......................................................................... 16

Aims ................................................................................................................. 18
  Paper I .......................................................................................................... 18
  Paper II ......................................................................................................... 18
  Paper III ....................................................................................................... 18
  Paper IV ....................................................................................................... 18

Materials and Methods .................................................................................. 19
  Study population ......................................................................................... 19
  Study cohort ................................................................................................ 19
  The Northern Sweden Health and Disease Study ........................................ 19
  Follow-up ..................................................................................................... 19
  Study design and participants ...................................................................... 20
Abstract

**Background:** Colorectal cancer (CRC) is a heterogeneous disease developing from distinct pathways, resulting in tumor subtypes with large differences in clinical and molecular characteristics. Molecular characteristics are increasingly being used clinically to guide therapy. However, whether molecular subtypes of CRC differ in etiology or risk factors is not clear. Clarifying such potential differences may lead to an improved understanding of CRC etiology, with implications for CRC prevention and screening.

**Aim:** The aim of this thesis was to investigate whether risk factors related to energy metabolism, such as body fatness, and one-carbon metabolism, such as circulating B-vitamin status, are associated with specific subtypes of CRC defined by molecular characteristics of the tumor.

**Methods:** These prospective studies are based on data and blood samples from cohorts within the population-based Northern Sweden Health and Disease Study (NSHDS). Prospective CRC cases with available archived tumor tissue were analyzed for key molecular features (*KRAS* and *BRAF* mutation status, Microsatellite instability (MSI) status, and CpG Island Methylator Phenotype (CIMP) status). Paper I was a cohort study of metabolic factors related to the metabolic syndrome (117 687 participants). Paper II was a nested-case control study on circulating insulin resistance-markers and adipokines (1010 cases and 1010 matched controls). Papers III and IV were nested case-control studies of one-carbon metabolism biomarkers and genetic variants (613 cases and 1190 matched controls).

**Results:** In paper I, we observed associations between metabolic factors, such as BMI, blood pressure, and blood lipids, and CRC risk consistent with previous studies. These associations were similar regardless of tumor *KRAS* and *BRAF* mutation status. In paper II, circulating biomarkers of insulin resistance and adipokines were not associated with the risk of CRC or specific molecular subtypes of CRC defined by *KRAS* and *BRAF* mutation or MSI status. In paper III, higher circulating levels of metabolites involved in the methionine cycle (namely, betaine and methionine) were associated with a lower CRC risk. In paper IV, we found no support for clear subtype-specific roles of any circulating one-carbon metabolism biomarker or genetic variants in CRC development.

**Conclusions:** The result of these prospective studies suggests that metabolic factors related to energy metabolism and one-carbon metabolism are generally associated with the risk of CRC, regardless of major subtypes defined by key molecular tumor features. If causal, metabolic risk factors likely influence the risk of colorectal cancer through more than one carcinogenic pathway.
**Abbreviations**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIC</td>
<td>Akaike information criterion</td>
</tr>
<tr>
<td>BHMT</td>
<td>Betaine-homocysteine S-methyltransferase</td>
</tr>
<tr>
<td>BMI</td>
<td>Body mass index</td>
</tr>
<tr>
<td>CBS</td>
<td>Cystathionine-β-synthase</td>
</tr>
<tr>
<td>CH-THF</td>
<td>Methylenetetrahydrofolate</td>
</tr>
<tr>
<td>CH2-THF</td>
<td>5,10-methylenetetrahydrofolate</td>
</tr>
<tr>
<td>CH3</td>
<td>Methyl group</td>
</tr>
<tr>
<td>CH3-THF</td>
<td>5-methylenetetrahydrofolate</td>
</tr>
<tr>
<td>CI</td>
<td>Confidence interval</td>
</tr>
<tr>
<td>CIMP</td>
<td>CpG Island Methylator phenotype</td>
</tr>
<tr>
<td>CIN</td>
<td>Chromosomal instability</td>
</tr>
<tr>
<td>CMS</td>
<td>Consensus Molecular Subtypes</td>
</tr>
<tr>
<td>COX-2</td>
<td>Cyclooxygenase-2</td>
</tr>
<tr>
<td>CRC</td>
<td>Colorectal cancer</td>
</tr>
<tr>
<td>CTH</td>
<td>Cystathionine γ-lyase</td>
</tr>
<tr>
<td>CUP</td>
<td>World Cancer Research Fund/American Institute for Cancer Research Continuous Update Project</td>
</tr>
<tr>
<td>CV</td>
<td>Coefficient of variation</td>
</tr>
<tr>
<td>DMG</td>
<td>Dimethylglycine</td>
</tr>
<tr>
<td>dTMP</td>
<td>Deoxythymidine 5′-monophosphate</td>
</tr>
<tr>
<td>eGFR</td>
<td>Estimated glomerular filtration rate</td>
</tr>
<tr>
<td>EPIC</td>
<td>European Prospective Investigation into Cancer and Nutrition</td>
</tr>
<tr>
<td>GECCO</td>
<td>Genetics and Epidemiology of Colorectal Cancer Consortium</td>
</tr>
<tr>
<td>GWAS</td>
<td>Genome-wide association study</td>
</tr>
<tr>
<td>H₂S</td>
<td>Hydrogen sulfide</td>
</tr>
<tr>
<td>IBD</td>
<td>Inflammatory bowel disease</td>
</tr>
<tr>
<td>ICC</td>
<td>Intra-class correlation</td>
</tr>
<tr>
<td>MONICA</td>
<td>Multinational MONitoring of Trends and Determinants in CArdiovascular Disease</td>
</tr>
<tr>
<td>MPE</td>
<td>Molecular pathological epidemiology</td>
</tr>
<tr>
<td>MSI</td>
<td>Microsatellite instability</td>
</tr>
<tr>
<td>MSP</td>
<td>Mammography Screening Project</td>
</tr>
<tr>
<td>MSS</td>
<td>Microsatellite stability</td>
</tr>
<tr>
<td>MTHFR</td>
<td>5,10-methylenetetrahydrofolate reductase</td>
</tr>
<tr>
<td>MTR</td>
<td>Methionine synthase</td>
</tr>
<tr>
<td>NSHDS</td>
<td>Northern Sweden Health and Disease Study</td>
</tr>
<tr>
<td>OR</td>
<td>Odds ratio</td>
</tr>
<tr>
<td>RD</td>
<td>Risk difference</td>
</tr>
<tr>
<td>ROR</td>
<td>Ratio of odds ratio</td>
</tr>
<tr>
<td>RR</td>
<td>Relative risk</td>
</tr>
<tr>
<td>SAH</td>
<td>S-adenosylhomocysteine</td>
</tr>
<tr>
<td>SAM</td>
<td>S-adenosylmethionine</td>
</tr>
<tr>
<td>SD</td>
<td>Standard deviation</td>
</tr>
<tr>
<td>SHMT</td>
<td>Serine hydroxymethyltransferase</td>
</tr>
<tr>
<td>SNP</td>
<td>Single nucleotide polymorphism</td>
</tr>
<tr>
<td>TCGA</td>
<td>The Cancer Genome Atlas</td>
</tr>
<tr>
<td>THF</td>
<td>Tetrahydrofolate</td>
</tr>
<tr>
<td>TYMS</td>
<td>Thymidine synthase</td>
</tr>
<tr>
<td>VIP</td>
<td>Västerbotten Intervention Programme</td>
</tr>
<tr>
<td>Vitamin B12</td>
<td>Cobalamin</td>
</tr>
<tr>
<td>Vitamin B2</td>
<td>Riboflavin</td>
</tr>
<tr>
<td>Vitamin B6</td>
<td>Pyridoxal 5′-phosphate</td>
</tr>
</tbody>
</table>
Original Papers


The original papers were reprinted with the permission from the publishers.
Related Publications

(Not included in thesis)


* These authors contributed equally to this work
### Overview of Papers

<table>
<thead>
<tr>
<th>Paper</th>
<th>NSHDS cohorts / study design</th>
<th>Main exposures</th>
<th>Main outcomes</th>
<th>CRC cases</th>
<th>Main findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>VIP, MONICA 1986-2016 / cohort study</td>
<td>7 metabolic factors</td>
<td>CRC risk by KRAS and BRAF mutation status</td>
<td>1250 (766 with complete tumor data)</td>
<td>BMI, blood lipids, and blood pressure were associated with CRC risk, regardless of KRAS and BRAF mutation status.</td>
</tr>
<tr>
<td>II</td>
<td>VIP, MONICA 1986-2016 / nested case-control study</td>
<td>4 metabolic biomarkers</td>
<td>CRC risk by KRAS and BRAF mutation status, and MSI status</td>
<td>1010 (704/708 with complete tumor data)</td>
<td>Circulating biomarkers of insulin resistance and adipokines were not associated with the risk of CRC or specific molecular subtypes.</td>
</tr>
<tr>
<td>III</td>
<td>VIP, MSP 1986-2009 / nested case-control study</td>
<td>5 one-carbon metabolism biomarkers</td>
<td>CRC risk</td>
<td>613</td>
<td>Higher circulating methionine and betaine were associated with a lower risk of CRC.</td>
</tr>
<tr>
<td>IV</td>
<td>VIP, MSP 1986-2009 / nested case-control study</td>
<td>14 one-carbon metabolism biomarkers and 17 genetic variants</td>
<td>CRC risk by KRAS and BRAF mutation status</td>
<td>488 with complete tumor data</td>
<td>One-carbon metabolism biomarkers were associated with CRC risk, regardless of KRAS and BRAF mutation status.</td>
</tr>
</tbody>
</table>

Populärvetenskaplig sammanfattning

Tjock- och ändtarmscancer är en av de vanligaste cancerformerna och ledande orsakerna för cancerdöd i världen. Livsstilen har stor betydelse för risken att utveckla tjock- och ändtarmscancer. Ingen enskild riskfaktor förklarar en stor del av alla fall, men det finns ett antal etablerade riskfaktorer som tillsammans har stor påverkan på den totala incidensen i befolkningen. Dessa inkluderar bland annat övervikt, rökning, fysisk inaktivitet och konsumtion av rött eller processat kött.


Syftet med denna avhandling var att undersöka om olika molekylära subtyper av tjock- och ändtarmscancer är associerade med olika riskfaktorer. Den ena halvan av avhandlingen fokuserade på metaboliska riskfaktorer så som kroppsmassindex (body mass index, BMI), blodsocker, blodfetter, blodtryck och nivåer av olika proteiner i blodet som indikerar på dålig blodsockerkontroll eller hög fettmassa. Tidigare studier har visat att några sådana metaboliska faktorer är kopplade till risken att utveckla tjock- och ändtarmscancer. Dock har det inte utförligt studerats om specifika subtyper har en starkare koppling till metaboliska faktorer än andra.


Sammantaget visar våra resultat att avvikande värden på metabola faktorer, så som övervikt, höga blodfetter, högt blodtryck eller avvikande blodnivåer av B-vitaminer, är kopplade till risken att utveckla tjock- och ändtarmscancer generellt, oavsett subtyp av sjukdomen.
Background

Colorectal cancer

Incidence and Mortality
Cancer, a group of diseases characterized by uncontrolled cell division, is a leading cause of death worldwide.\(^1\) Colorectal cancer (CRC) originates from epithelial cells in the colon or rectum and is currently the third most common cancer and fourth leading cause of cancer death in the world. Approximately 1.4 million new cases and 693,900 deaths were reported in 2012.\(^1\) About 2/3 of the cancers are located in the colon and 1/3 in the rectum. The incidence is low before age 50, with a mean age at diagnosis of about 70, and men have a higher risk compared to women (Figure 1).

![Cumulative incidence of CRC estimated in Sweden between 1996-2015 from NORDCAN© 2017 Association of the Nordic Cancer Registries, IARC (Assessed 180612).](-)

Figure 1 Cumulative incidence of CRC estimated in Sweden between 1996-2015 from NORDCAN© 2017 Association of the Nordic Cancer Registries, IARC (Assessed 180612).\(^2\)

There is a sizeable geographical variation in age-standardized incidence rates (Figure 2). The geographical patterns are similar in men and women. The highest rates are found in western high-income countries, such as Australia and European countries (38.4 and 29.5 per 100,000, respectively, in 2012), and lowest in Western Africa (4.1 per 100,000).\(^1\) Incidence rates in previous low-incidence countries, such as Latin American, Asian, and Central European countries, have gradually increased\(^3\) – likely due to increases in risk factors such as excess caloric intake, smoking, and physical inactivity.\(^4\) At the same time, incidence rates in some high-incidence countries, such as the United States, have started to decline - likely due to population screening.\(^3\) Mortality rates have
declined in most countries, likely due to screening and improved treatment, with the potential exception of some countries in Latin America and Eastern Europe. In Sweden, CRC incidence rates are comparable with those in high-incidence countries (29.2 per 100 000) and show no clear signs of decreasing. Swedish CRC mortality rates are comparable to those of other high-income countries and decreasing (10.9 per 100 000).

**Figure 2** Age-standardized incidence rates (per 100 000) of colorectal cancer in both sexes. Data from GLOBOCAN® 2012, IARC (Assessed 180612). Countries in grey lack data.

**Prognosis and treatment**

The 5-year relative survival of patients with CRC ranges from approximately 65% in high-income countries, including Sweden, to below 50% in low-income countries. The prognosis is highly dependent on tumor stage at diagnosis, a combined measure based on three variables: degree of local invasion (T stage), spread to lymph nodes (N stage), and presence of distant metastases in other organs (M stage). In Sweden, the average 5-year survival is good for localized disease (stage I: 95% and stage II 89%), and gradually worsens with regionally spread (stage III: 68%) and distantly spread metastatic disease (stage IV: 15%, Swedish National Quality Registry for Colorectal Cancer data 2007-2016).

The main curative treatment of CRC is surgery. However, the overall treatment regimen depends on tumor site, tumor stage, and molecular markers. Patients with stage I tumors generally require no treatment beyond surgery. More advanced or lower situated rectal tumors receive preoperative neoadjuvant radiotherapy to reduce the risk of local recurrence. More advanced colon tumors receive adjuvant chemotherapy. In metastatic disease, stage IV, treatment
consists of a combination of surgical removal of primary tumors or metastases (most often in the liver or lung), chemotherapy, and targeted therapies.\textsuperscript{7} The effect of targeted therapies depends on molecular markers. Targeted therapies can include epidermal growth factor receptor-inhibitors for KRAS-wild type tumors,\textsuperscript{8,9} and immune checkpoint inhibitors for hypermutated tumors.\textsuperscript{10,11}

**Risk and preventive factors**

The risk of CRC is highly influenced by environmental factors. In the largest twin study to date, utilizing the twin registries in the Nordic countries,\textsuperscript{12} the proportion of the variation of cancer risk explained by hereditary factors was one of the lowest for CRC, only 15\% for colon and 14\% for rectal cancer, comparable to the heritability of lung cancer (18\%). Additional evidence for a strong environmental influence on CRC risk can be found in trends of geographical differences in incidence over time. Countries with large economic growth in the last decade have experienced a rapid increase in CRC incidence (e.g., Latin American, Asian, and Central European countries).\textsuperscript{3} Furthermore, studies in the United States and Canada have shown that Asian immigrant populations with an initially lower CRC incidence approach the incidence of the native population over time.\textsuperscript{13-15}

There are a number of established risk factors for CRC beyond age and male sex, each with a small contribution to the overall risk. Modifiable risk or preventive factors with convincing evidence according to the World Cancer Research Fund/American Institute for Cancer Research Continuous Update Project (CUP),\textsuperscript{16} are summarized in **Table 1**.

<table>
<thead>
<tr>
<th>Factor</th>
<th>Meta-analysis relative risk</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>INCREASE RISK</strong></td>
<td></td>
</tr>
<tr>
<td>Body fatness\textsuperscript{a,17}</td>
<td>1.06 per 5 kg/m(^2) increase</td>
</tr>
<tr>
<td>Adult height\textsuperscript{17}</td>
<td>1.04 per 5 cm increase</td>
</tr>
<tr>
<td>Smoking\textsuperscript{18}</td>
<td>1.17 current vs never smokers</td>
</tr>
<tr>
<td>High alcohol consumption\textsuperscript{19}</td>
<td>1.07 per 10 grams/day increase</td>
</tr>
<tr>
<td>High intake of red and processed meat\textsuperscript{19}</td>
<td>1.12 per 100 grams/day increase</td>
</tr>
<tr>
<td><strong>DECREASE RISK</strong></td>
<td></td>
</tr>
<tr>
<td>Physical activity\textsuperscript{b,20}</td>
<td>0.73 most vs least physically active</td>
</tr>
<tr>
<td>Hormone replacement therapy\textsuperscript{21,22}</td>
<td>0.74-0.88 use vs no use</td>
</tr>
<tr>
<td>Regular aspirin use\textsuperscript{23-25}</td>
<td>0.73 use vs no use</td>
</tr>
</tbody>
</table>

\(a\) Measured as high body mass index (BMI), waist circumference, or waist-to-hip ratio.

\(b\) Mainly related to colon cancer.
An established non-modifiable, but clinically relevant, risk factor is family history of CRC not due to known hereditary forms (such as familial adenomatous polyposis or Lynch syndrome). The risk of CRC for individuals with a first-degree relative with CRC is about two times that of individuals with no affected relatives, and is increased with multiple affected relatives and relatives affected at younger ages.

Other potential risk or preventive factors with less consistent evidence include intake of whole grains or dairy products, and inflammatory bowel disease (IBD). The increased risk of CRC in individuals with IBD has generally been thought to arise from a causal tumor-promoting effect of colonic inflammation, rather than shared risk factors. However, associations between anti-inflammatory treatment of IBD and CRC risk have been inconsistent and there is no randomized evidence.

**Screening**

As CRC develops slowly and has effective curative treatments for low-stage disease, there is a large potential benefit from screening. Randomized controlled trials of screening with fecal occult blood tests report an on average 16% reduced CRC mortality in screened participants compared to non-screened, yet with no difference in overall mortality. Screening trials with flexible sigmoidoscopy report reductions in both CRC incidence and mortality, with the possible exception of older women (>60 years). The impact of screening with colonoscopy on CRC incidence and mortality has not been assessed in a trial but shows great potential in observational studies, and trials are underway. Currently, many countries have ongoing pilots or implemented screening programs, including most of Europe and Sweden. Opportunistic screening with colonoscopy is common for 50-75 year-olds in the United States and Germany. Efforts are being made to create better and non-invasive multi-marker screening tests, for instance, by using circulating tumor DNA or protein biomarkers. These are, however, so far not competitive enough with regards to diagnostic accuracy and cost-effectiveness. The cost and efficacy of screening may also be improved by a more individualized approach to determine the starting age and frequency of screening. A recent study reported that information on genetic and environmental risk factors could predict CRC with greater accuracy than family history, and that the recommended starting age at screening based on the prediction model differed by approximately 10 years between low and high-risk individuals.
**Colorectal carcinogenesis and molecular subtypes**

Cancer is the result of normal cells which acquire the ability to divide uncontrollably, creating a tumor with the ability to invade surrounding tissues or spread to distant organs. The cause of this abnormal cell behavior is somatic genetic or epigenetic alterations. These alterations, primarily mutations, occur in genes involved in cell proliferation, differentiation, programmed cell death, and DNA-repair and differ by cancer type. Colorectal tumors develop from precancerous adenomas to carcinomas in a multi-step process known as the adenoma-carcinoma sequence, first proposed by Fearon and Vogelstein in 1990. The process is typically slow, taking up to more than 20 years. According to the model, around 80% of tumors develop through the chromosomal instability (CIN) pathway (Figure 3). CIN tumors, characterized by widespread changes in the number and structure of chromosomes, acquire early mutations in oncogenes such as the *APC* and/or *KRAS* gene, and in tumor suppressor genes such as the *TP53*. The remaining tumors develop through completely different pathways. These pathways include tumors characterized by the loss of DNA mismatch repair, around 15% of tumors, causing numerous mutations throughout the genome, in particular at microsatellites. These, so-called, microsatellite instable (MSI) tumors often acquire mutations in the *BRAF* oncogene early and are further associated with the CpG Island Methylator phenotype (CIMP). CIMP-positive (low or high) tumors are characterized by the silencing of tumor suppressor genes through DNA methylation at CpG Islands in the promoter.

![Colorectal cancer](image)

<table>
<thead>
<tr>
<th>Non-CIN pathways</th>
<th>CIN pathway</th>
</tr>
</thead>
<tbody>
<tr>
<td>~20% of all CRC</td>
<td>~80% of all CRC</td>
</tr>
<tr>
<td><em>APC</em> mutation</td>
<td><em>APC</em> mutation</td>
</tr>
<tr>
<td><strong>BRAF</strong> mutation</td>
<td><strong>KRAS</strong> mutation</td>
</tr>
<tr>
<td>MSI</td>
<td>MSS</td>
</tr>
<tr>
<td>CIMP-high</td>
<td>CIMP-low</td>
</tr>
<tr>
<td>Origin in serrated polyps</td>
<td>Origin in adenomas or serrated polyps</td>
</tr>
<tr>
<td>~20%</td>
<td>~20%</td>
</tr>
</tbody>
</table>

**Figure 3.** Major pathways of colorectal carcinogenesis and common molecular characteristics and precursor lesions related to them.
The complex and heterogeneous pathogenesis of CRC results in molecular subtypes with large differences in clinical and molecular characteristics. Molecular subtypes of CRC based on mutations in the KRAS and BRAF oncogenes, MSI status, and CIMP status may be used to approximate CRC developmental pathways (Figure 3).45-48 Mutations in KRAS and BRAF, essentially mutually exclusive mutations present in 20-50% and 4-18% of CRC tumors, respectively,49-51 are particularly interesting subtyping markers. The KRAS and BRAF proteins are both involved in the cell proliferation and division regulator mitogen-activated protein kinase pathway. This pathway becomes constitutively activated by mutations in KRAS and BRAF, leading to uncontrolled cell proliferation and resistance to apoptosis.52 Mutations in KRAS and BRAF occur early in the carcinogenic process and are associated with distinct clinical and molecular characteristics.49,50 BRAF-mutated CRC are often MSI, CIMP-high, and originate from flat, so-called, serrated polyps.45,48 KRAS-mutated tumors are generally CIN tumors that originate from serrated polyps or adenomas, and are characterized by microsatellite stability (MSS) and a lower degree of methylated markers (CIMP-low).45-48 KRAS and BRAF wild type tumors are generally CIN tumors originating from adenomas. Clinically, mutated KRAS indicates resistance to anti-epidermal growth factor receptor therapy,8,9 and mutated BRAF is associated with a poor prognosis.53 Mutated BRAF is more prevalent in women, older patients, right-sided tumors,45,49,50 and in tumors with high infiltration of certain immune cells.54,55

Instead of using a few molecular markers to subtype CRC, the CRC Subtyping Consortium combined results from gene expression-based subtyping algorithms to form four, so-called, Consensus Molecular Subtypes (CMS).56 The subtypes with distinct gene-expression and molecular profiles were: CMS1 – MSI immune (14% of CRC), CMS2 – Canonical (37%), CMS3 – Metabolic (13%), and CMS4 – Mesenchymal (23%). The CMS1 subtype represents the hypermutated microsatellite unstable tumors, often BRAF-mutated, CIMP-high, and with a high degree of immune activation. The remaining CMS2-4 represents the CIN pathway. The CMS3 subtype is characterized by metabolic dysregulation, and are often KRAS-mutated, CIMP-low, and sometimes MSI. The CMS2 and CMS4 subtypes are both characterized by a large number of somatic copy number alterations, but while CMS2 has WNT and MYC signaling activation, CMS4 has TGF-β activation, stromal infiltration, and increased angiogenesis.
Energy metabolism

Energy metabolism and body fatness

Energy metabolism includes all reactions in the body that produce and store energy from food intake. Food enters the human digestive system and is digested in the mouth, stomach, small intestine, and colon. Most chemical digestion occurs in the small intestine, where also most nutrients are absorbed into the bloodstream. Carbohydrates, fats, and proteins are broken down and enter various steps in the cellular respiration pathways within cells to generate energy in the form of adenosine triphosphate.\(^5\) Carbohydrates, which are broken down into sugars like glucose, are the major energy source, accounting for \(~45\text{-}70\%\) of the total energy intake and expenditure.\(^6\) Blood glucose levels are regulated by the hormone insulin.\(^7\) Insulin is produced by beta cells in the pancreas as a response to increased blood glucose levels. Insulin lowers glucose levels by stimulating the uptake of glucose into cells and inhibiting glucose production. Insulin also stimulates the storage of glucose in glycogen in the liver and muscles and inhibits breakdown of lipids as an energy source.\(^8\) In insulin resistance, cells have become less responsive to insulin, leading to higher blood glucose levels which, in turn, cause beta cells to produce more insulin, leading to higher blood insulin levels. This causes a vicious cycle, and when beta cells cannot keep up with increasing blood sugar concentrations this may cause type II diabetes.\(^9\)

Long-term excess energy intake leads to the accumulation of excess adipose tissue (i.e., body fat), resulting in weight gain.\(^10\) Body fatness, often measured as high BMI (weight in kg/[height in m]\(^2\)), is caused by a combination of genetic susceptibility and environmental influences (heritability of BMI \(40\text{-}70\%\)), and leads to several adverse health outcomes and increased mortality.\(^11\) With the exception of parts of Eastern Europe, South Asia, and Central Africa, the average BMI has increased worldwide since 1980.\(^12\) During the same time period, the prevalence of obesity (defined as BMI\(\geq30\)) has doubled in more than 70 countries.\(^13\) Body fatness is a major global health problem. Higher BMI was estimated to account for 4 million deaths in 2015, most of which were related to cardiovascular disease and of which 40\% occurred in non-obese individuals.\(^14\)

In addition to storing energy, adipose tissue also regulates metabolism by releasing hormones, signaling molecules, and inflammatory cytokines.\(^15\) Body fatness leads to profound changes in these processes. For instance, body fatness increases the release of inflammatory cytokines, such as tumor-necrosis factor-\(\alpha\) and interleukin-6, which likely promotes insulin resistance.\(^16\) Body fatness also alters the production of adipocyte-derived hormones, so called adipokines, such as adiponectin and leptin. Adiponectin, involved in energy balance and an anti-inflammatory agent, is lowered in obesity while leptin, involved in appetite
regulation, is increased.\textsuperscript{69} Furthermore, body fatness is related to increased blood lipids (specifically triglycerides and cholesterol), as well as increased blood pressure.\textsuperscript{61} Body fatness, insulin resistance, hyperlipidemia, and hypertension together form an intertwined cluster of metabolic factors known as the metabolic syndrome (\textbf{Figure 4}).\textsuperscript{70} These are strong risk factors for cardiovascular disease.\textsuperscript{70} Metabolic aberrations related to the metabolic syndrome are, however, not always co-existent with, or preceded by, overweight or obesity.\textsuperscript{71,72} For instance, only about 50\% of obese individuals are insulin resistant.\textsuperscript{73}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure4.png}
\caption{Metabolic factors related to the metabolic syndrome.}
\end{figure}

\textbf{Energy metabolism and CRC}

Body fatness is associated with an increased risk of several cancers,\textsuperscript{74} including CRC.\textsuperscript{17} In addition to the bulk of observational studies, there is strong support of a causal relationship between body fatness and CRC risk from Mendelian randomization studies.\textsuperscript{75-77} Mendelian randomization studies use genetic variants related to an exposure as instrumental variables in studies of disease.\textsuperscript{78} As genotypes are randomly assigned from parents to offspring during meiosis, the genetic instrument should not be associated with confounding factors. Given some assumptions, these studies can thus infer causality of an association in the same manner as a randomized trial.\textsuperscript{78} For other metabolic factors related to the metabolic syndrome, elevated blood lipids and blood pressure have been associated with an increased CRC risk, independently of BMI.\textsuperscript{79,80} For total cholesterol, a Mendelian randomization study also found evidence for a causal effect independent of BMI.\textsuperscript{81}
In longitudinal or retrospective observational studies, adult weight gain has been associated with an increased risk of CRC. Furthermore, a longitudinal study of women showed that a longer duration of adulthood overweight and obesity was associated with a higher risk of CRC. These results and the strong case for a causal relationship between body fatness and CRC development, suggest a large potential for CRC prevention by weight management. Intentional weight loss by surgical methods, such as bariatric surgery in obese individuals, has been associated with a reduced risk of CRC in some, but not all, observational studies. Small randomized controlled trials of diet-induced weight loss have shown reductions in biomarkers implicated in CRC development, such as colorectal tissue markers of apoptosis, proliferation, and inflammation.

Several biological mechanisms may explain the relationship between body fatness and CRC development. Plausible contributing mechanisms include increased low-grade inflammation, insulin resistance, altered adipokine production, and potentially altered gut microbiota. Low-grade chronic inflammation produces proinflammatory cytokines and reactive oxygen species, which can initiate cancer by causing DNA damage, or by creating a favorable microenvironment that promotes the growth and invasion of malignant cells. Insulin signaling has been linked to increased inflammation and cell proliferation. The adipokine adiponectin inhibits cell proliferation, inhibits inflammation, and induces apoptosis, while leptin may promote proliferation and angiogenesis or inhibit apoptosis. Body fatness is also associated with alterations in the gut microbiome composition, which can cause inflammation or produce metabolites that increase proliferation or reduce apoptosis. Prospective observational studies of circulating biomarkers related to the proposed mechanisms and CRC risk have also been made. Markers of low-grade inflammation, such as C-reactive protein, are weakly associated with CRC risk, mainly in men and with significant heterogeneity between studies. Markers of insulin resistance, including insulin, C-peptide, or insulin-like growth factor I, have been associated with CRC risk in most studies. Elevated plasma glucose or type II diabetes may be associated with an increased risk of CRC. However, there is considerable heterogeneity among studies. For adipokines, lower adiponectin has been associated with an increased risk of CRC, but with significant heterogeneity among studies, while results for leptin have been inconsistent. In addition, a mediation analysis reported biomarkers of insulin resistance and adipokine-related markers as the primary mediators of the association between adult weight gain and colon cancer risk.
One-carbon metabolism

The folate and methionine cycles
One-carbon metabolism, a network of intra-cellular reactions centered around the folate and methionine cycles and involving the transfer one-carbon units (i.e., methyl groups, CH₃), is central for nucleotide and protein synthesis and methylation (Figure 5).¹⁰⁷,¹⁰⁸


The main contributors of methyl groups into one-carbon metabolism are serine and choline.¹⁰⁷ Folic acid, the synthetic form of the B-vitamin folate, is available through supplementation or folic acid-fortified foods. Folic acid fortification of foods is done in several countries to reduce the incidence of neural tube defect in new-borns.¹⁰⁹ Folic acid is transported into cells, reduced to tetrahydrofolate (THF) by several enzymes, and finally enters the folate cycle.¹⁰⁷ In the folate cycle, THF is metabolized to 5,10-methylenetetrahydrofolate (CH₂-THF) by serine hydroxymethyltransferase (SHMT) with vitamin B6 as a co-factor. CH₂-THF can
take one of three paths, it can 1) be converted back to THF by thymidine synthase (TYMS), creating the nucleotide deoxythymidine 5′-monophosphate (dTMP) in the process; 2) be converted to methenyltetrahydrofolate (CH-THF) and successively back to THF creating purines for DNA or RNA synthesis, or 3) be reduced to 5-methyltetrahydrofolate (CH₃-THF) by 5,10-methylenetetrahydrofolate reductase (MTHFR). CH₃-THF is demethylated back to THF by methionine synthase (MTR) with vitamin B12 as a cofactor, donating its methyl group into the methionine cycle converting homocysteine to methionine. In the methionine cycle, homocysteine can take an alternative path to methionine by receiving a methyl group from the demethylation of the choline-product betaine to dimethylglycine (DMG), catalyzed by betaine-homocysteine methyltransferase (BHMT). Methionine is converted to the universal methyl group donor S-adenosylmethionine (SAM). SAM donates its methyl group creating methylated products as it converts to S-adenosylhomocysteine (SAH). SAH then completes the methionine cycle by being hydrolyzed back into homocysteine. Homocysteine can also be metabolized into cysteine via cystathionine through B6-dependant reactions with the cystathionine-β-synthase (CBS) and cystathionine γ-lyase (CTH, also abbreviated CSE) enzymes in the transsulfuration pathway.

The overall B-vitamin and one-carbon metabolism status in individuals are determined by both genetic and environmental factors. Circulating folate, homocysteine, and vitamin B12 (cobalamin) levels have a heritability of 56-66%, with the rest of the total variance being explained by unique environmental factors (e.g., diet or vitamin supplements). A common functional single nucleotide polymorphism (SNP) in the MTHFR gene that reduces enzyme activity, the rs1801133 (sometimes referred to as the MTHFR 677C>T), is associated with decreased circulating folate and increased circulating homocysteine.

Balanced one-carbon metabolism is crucial for genome stability and repair. Inhibiting the folate cycle blocks cell proliferation and low folate has been associated with decreases in global DNA methylation. Similar associations with methylation have been observed for components in the methionine cycle. For instance, dietary restriction of methionine reduces SAM levels, which reduce histone methylation and affects gene expression, and a high choline intake preserves SAM and DNA methylation in folate-compromised men. Therefore, through its role in both genetic and epigenetic management, one-carbon metabolism has a plausible role in cancer development.
One-carbon metabolism and CRC

The study of one-carbon metabolism in cancer can be traced back to 1948, when Sidney Farber and his colleagues discovered that inhibiting one-carbon metabolism caused remission in children with leukemia. Today, anti-folates constitute a major class of first-line chemotherapies for several cancers. This includes the colorectal cancer agent fluorouracil, which hinders the production of nucleotides in tumor cells by inhibiting TYMS in the folate cycle.

The role of vegetables and fruits, major natural food sources of folate, in CRC development has also been extensively studied. Today, the evidence for a protective role in CRC is less convincing for vegetables and limited for fruits. The potential protective role of these factors has generally been attributable to dietary fiber, a probable protective factor for CRC. Yet, the earlier observational studies and trials showed conflicting results, and mainly dietary fiber from whole grains (also a probable CRC protective factor), and not from vegetables or fruits, displayed an association. This lead researchers to turn their attention to other components of vegetables and fruits, such as folate.

Numerous observational studies have investigated whether dietary folate, folic acid supplementation, and/or blood concentrations of folate are related to the risk of cancer, and in particular, CRC. Low dietary folate intake is associated with an increased risk of colorectal cancer in most studies, while the results for folic acid supplements and blood concentrations are inconclusive. Some studies of blood concentrations (including two studies in the Northern Sweden Health and Disease Study, NSHDS, the study cohort on which this thesis is based) observed an increased risk in individuals with higher circulating folate. Most randomized clinical trials of folic acid supplementation have shown no significant effects on cancer endpoints (e.g., colorectal polyp recurrence). However, in some trials on predisposed individuals, supplementation increased the risk of recurring advanced colorectal adenomas. Post-hoc analysis of this trial and another study suggest that the positive association was due to higher circulating CH2-THF, and not a direct effect from high levels of unmetabolized circulating folic acid, which is associated with other potentially cancer-promoting mechanisms such as reduced natural killer cell cytotoxicity. Therefore, the general consensus to date, which is further backed by animal studies, is that balanced folate metabolism likely prevents tumor initiation, whereas imbalances may promote the growth of established tumors or precancerous lesions.

The potentially harmful effect of high folate levels in the presence of precancerous lesions fueled concerns over the safety of food folic acid fortification. However, despite the implementation of mandatory folic acid fortification of flour and cereal in the United States, CRC incidence has decreased over time. If higher folate status prevents tumor initiation in normal tissue but facilitates the
progression of precancerous lesions, then the combination of folic acid fortification with colonoscopy screening including the removal of precancerous lesions may have contributed to the reduction in CRC incidence in the United States. This hypothesis is also in line with the lack of decrease in CRC incidence rates in Canada, where mandatory folic acid fortification was introduced at about the same time as the United States but national screening programs do not include colonoscopy.

To further understand the role of one-carbon metabolism in cancer, studies of other components involved in one-carbon metabolism have been made. Studies of dietary intakes of one-carbon metabolism micronutrients other than folate have been inconclusive. A reduced risk of CRC has generally been observed in individuals with higher circulating levels of vitamins B2 (riboflavin or flavin mononucleotide), B6 (mostly pyridoxal 5'-phosphate), and B12, also in the NSHDS, a study population with low folate status and where high folate was associated with an increased CRC risk. Some randomized clinical trials of folic acid and cancer incidence also included treatment arms with vitamin B6 or B12 supplementation, with no apparent effect. Higher homocysteine levels have been associated with an increased risk of CRC, but with inconsistencies between studies (e.g., null association in European study populations). For other metabolites in the transsulfuration pathway, results have been inconclusive. Metabolites in the methionine cycle, such as methionine and metabolites in the choline oxidation pathway, have not been extensively studied, but have been associated with a lower risk of CRC or colorectal adenoma in some studies.

Common SNPs in genes encoding for enzymes involved in one-carbon metabolism have also been extensively studied as independent risk factors or effect modifiers of dietary intakes or circulating levels of one-carbon metabolism components. Most studies have focused on the functional rs1801133 SNP in the MTHFR gene. Candidate gene studies of rs1801133 show a slight decreased risk in variant allele carriers, but the SNP has not been associated in large genome-wide association studies (GWAS) of CRC. No other SNP in genes involved in one-carbon metabolism has been associated in GWAS of CRC, and associations of SNPs in candidate gene studies have generally been weak. The link between one-carbon metabolism-related genes and CRC is further complicated by the potential interaction between circulating levels of one-carbon metabolites found in some studies. For example, the association between MTHFR rs1801133 and CRC risk may be limited to populations with a high folate status.
Heterogeneity in CRC etiology
CRC develops through distinct pathways resulting in molecular subtypes with large differences in clinical and molecular characteristics. CRC may, therefore, more accurately be described as a group of diseases with potential differences in etiology and risk factors. Risk factors may either directly cause specific molecular alterations related to specific developmental pathways of CRC, or contribute to a colorectal microenvironment favorable to the development of tumors with specific molecular features. By studying risk factors in relation to subtypes of CRC with homogenous pathogenesis, valuable insights about the mechanisms behind the relationship can be made. Furthermore, if the relationship of a risk factor differs by CRC subtypes, such information may be used to improve the overall validity and precision of that risk factor. These potential gains lead to the initiation of molecular pathological epidemiology (MPE), an interdisciplinary research field using recent advances in molecular classification of disease to study environmental and genetic risk factors for disease subtypes with homogenous pathogenesis. The main findings from MPE studies of established and potential CRC risk and preventive factors to date are summarized here.

Body fatness
MPE studies of BMI have not found conclusive support for subtype-specific associations for molecular subtypes defined by KRAS and BRAF mutation status, MSI status, CIMP status, or degree of tumor-infiltrating lymphocytes. In two studies on the American Nurses’ Health Study and Health Professionals Follow-up Study, higher BMI was associated with an increased risk of CRC without, and not with, poorly differentiated foci, and CRC without fatty acid synthase expression in women. Another study in that study population observed a subtype-specific association with KRAS-mutated CRC risk for low plasma levels of adiponectin. Whether other metabolic factors related to the metabolic syndrome are associated with specific molecular subtypes of CRC has, to our knowledge, not been investigated.

Adult height
Adult height has not been extensively studied in an MPE setting, but was mainly associated with an increased risk of BRAF-mutated or MSI CRC in one study.

Smoking
Smoking likely increases the risk of MSI, BRAF-mutated, or CIMP-positive CRC, and not CRC developing through the traditional CIN pathway. In line with these results, smoking has also been related to an increased risk of serrated colorectal adenomas.
**Alcohol**
Alcohol intake has not been conclusively associated with specific molecular subtypes, of which most studies focused on MSI or KRAS and BRAF mutation status subtypes.\(^{174,175,193-200}\)

**Red and processed meats**
Similar to alcohol intake, there is no conclusive evidence for a subtype-specific relationship between intake of red and/or processed meat and CRC.\(^{171,175,195-197,201-203}\)

**Physical activity**
Physical activity may primarily be associated with a reduced risk of KRAS-mutated CRC,\(^{170,171}\) while for subtypes defined by MSI or CIMP status, there are no clear differences.\(^{174,176,199}\)

**Hormone replacement therapy**
Hormone replacement therapy in women is mainly associated with a reduced risk of MSS, and not MSI, CRC,\(^{175}\) with no clear difference by KRAS-mutation status.\(^{204}\)

**Aspirin use**
MPE studies of regular aspirin use and CRC risk have reported subtype-specific associations for CRC without BRAF-mutations,\(^{205}\) or with a low degree of immune infiltration,\(^{206}\) but no difference by MSI status.\(^{175}\) Furthermore, aspirin use appears to associate mainly with a reduced risk of colorectal tumors that overexpress cyclooxygenase-2 (COX-2).\(^{207}\) The COX-2 enzyme promotes inflammation and cell proliferation and is directly inhibited by aspirin.\(^{208}\) The subtype-specific association between aspirin and CRC overexpressing COX-2 therefore provides a likely mechanism for the aspirin-CRC link. Moreover, postdiagnostic aspirin use has been associated with an improved survival in patients with tumors overexpressing COX-2 or harboring mutations in genes in the PI3K pathway (which cause COX-2 overexpression).\(^{209}\) The potential benefit of adjuvant aspirin in patients diagnosed with CRC and mutations in the PI3K pathway is currently being evaluated in a Swedish randomized controlled trial (the ALASCCA study, clinicaltrials.gov identifier: NCT02647099).

**Other factors, including one-carbon metabolism**
Studies of various dietary factors in relation to molecular subtypes of CRC have not provided conclusive evidence for subtype-specific associations.\(^{166}\)
Studies of dietary intakes of one-carbon metabolism nutrients, mostly focused on dietary folate, in relation to molecular subtypes of CRC (e.g., defined by KRAS and BRAF mutations, MSI, and CIMP status) have been inconsistent. A study in the low-folate status NHSDS population reported a stronger association between low circulating folate and a decreased risk of CIMP-positive, compared to CIMP-negative, CRC. Other biomarkers of one-carbon metabolism have, to our knowledge, not been studied in relation to CRC subtypes.

Summary and knowledge gap
CRC is one of the most common cancer diagnoses and a leading cause of cancer death worldwide. The risk of developing CRC is highly influenced by environmental risk factors. Therefore, as the prevalence of risk factors related to the so-called western lifestyle increase, so does CRC incidence. This causes a significant amount of suffering and economic cost, which may be prevented by lifestyle changes or preventive medical treatment. Although several risk factors for CRC have been established, such as body fatness, smoking, and physical inactivity, other risk factors remain to be discovered or require more research. Furthermore, the etiology of CRC, i.e., the manner of how risk factors affect colorectal carcinogenesis, is not fully understood.

Etiological studies of CRC are complex due to the heterogeneity in colorectal carcinogenesis. CRC develops through distinct pathways leading to molecular subtypes with substantial differences in molecular and clinical characteristics. Differences in etiology between such subtypes are likely, but have not been extensively studied, probably because of the lack of study populations for which both prospective exposure data and molecular tumor data are available. Such MPE studies of risk factors and molecular subtypes of CRC may provide insights into the mechanisms behind the associations, identify new subtype-specific risk factors which could have been masked in the analysis of overall CRC, and improve the precision of risk factors used in, for instance, prediction models for individualized screening. Currently, the strongest support for etiological heterogeneity in CRC exists for smoking, which specifically associates with an increased risk of CRC developing through non-CIN pathways, such as MSI. Research on whether other risk factors differ by molecular subtypes of CRC have been less clear or not extensively studied.

Several studies have investigated whether the association between BMI and CRC risk differ by molecular subtypes. One of the more studied relationships is the association between BMI and CRC risk by MSI status, where there probably is no large difference. For other subtypes, such as subtypes by KRAS and BRAF mutation status, however, results are inconclusive. Some smaller studies observed a potential subtype-specific increased risk of KRAS-mutated CRC in
women with higher BMI, or men and women with low adiponectin, a consequence of overweight and the metabolic syndrome. This could imply that certain aspects of the metabolic syndrome, which does not always coincide with overweight and are independently associated with the risk of overall CRC, relate to a specific molecular subtype of CRC. Another possibility is that specific mechanisms of the body fatness-CRC relationship, such as insulin resistance or alterations in adipokine concentrations, increase the risk of specific CRC subtypes. Such subtype-specific associations may explain the moderate to large variations in the results of observational studies of biomarkers related to these mechanisms and overall CRC risk. To answer these questions, prospective studies investigating wider aspects of body fatness or metabolic health in relation to the risk CRC by molecular subtypes are needed.

Balanced folate-mediated one-carbon metabolism, important for genetic stability, may prevent cancer initiation, whereas imbalances may support the progression of established tumors or precancerous lesions. While folate is involved in both DNA synthesis and methylation, components within the methionine cycle are involved in methylation. Studies of such components, in particular in populations such as the NSHDS with a low folate status, can provide insights into how the two main outputs of one-carbon metabolism, DNA synthesis and methylation, relate to cancer development. Furthermore, the role of one-carbon metabolism in different developmental pathways of CRC has not been extensively studied. For instance, do folate or components mainly involved in methylation specifically relate to hypermethylated CRC subtypes (i.e., BRAF-mutated or CIMP-positive tumors)? Suggestive evidence for such a subtype-specific association for folate has been observed in the NSHDS. But replication is required, and the role of other one-carbon metabolism factors in different pathways of colorectal carcinogenesis is unclear.

In summary, factors related to energy and one-carbon metabolism have plausible roles in CRC development. Whether their roles are specific to certain pathways of CRC development is not clear. Clarifying potential differences could help clarify the inconsistent findings of some factors, and lead to an improved understanding of CRC etiology, with implications for prevention and screening.
Aims

The general aim of this thesis was to investigate whether factors related to energy metabolism and one-carbon metabolism are associated with the risk of specific molecular subtypes of CRC.

**Paper I**
- To investigate metabolic traits related to the metabolic syndrome, including BMI, blood glucose, blood lipids, and blood pressure, in relation to the risk of molecular subtypes of CRC defined by tumor KRAS and BRAF mutation status.
- A secondary aim was to investigate associations with CRC subtypes defined by MSI status.

**Paper II**
- To investigate circulating metabolic biomarkers of insulin resistance (insulin and C-peptide) and the adipokines adiponectin and leptin in relation to the risk of CRC and molecular subtypes of CRC defined by tumor KRAS and BRAF mutation status and MSI status.

**Paper III**
- To investigate associations between biomarkers of one-carbon metabolism mainly involved in the methionine cycle, namely choline, betaine, dimethylglycine, and sarcosine, in relation to CRC risk in a study population with low folate status.

**Paper IV**
- To investigate biomarkers and genetic variants, covering a wide-range of one-carbon metabolism aspects, in relation to the risk of molecular subtypes of CRC defined by tumor KRAS and BRAF mutation status.
- A secondary aim was to investigate associations with CRC subtypes defined by MSI and CIMP status.
Materials and Methods

Study population
The studies in this thesis are based on individuals residing in Västerbotten County in northern Sweden. Västerbotten County has approximately 270,000 inhabitants, most of whom live in the coastal cities of Umeå (45%) and Skellefteå (28%). The rest live in smaller towns and rural areas across the county. The county is the second largest geographically, and 14th largest by population. The population density is higher, and the average age lower, in Umeå, mainly due to the 31,000 students at Umeå University. The smaller towns and rural areas are characterized by a lower population density and older population.

Study cohort

The Northern Sweden Health and Disease Study
All papers in this thesis are based on the prospective population-based Northern Sweden Health and Disease Study (NSHDS). From 1986 until 2016, the NSHDS contained 125,381 participants with 225,157 observations (i.e., visits) within three cohorts: the Västerbotten intervention programme (VIP, approximately 75% of all observations), the Multinational MONitoring of Trends and Determinants in CArdiovascular Disease study (MONICA, 7%), and the Mammography Screening cohort (MSP, 18%). The ongoing VIP is a health screening intervention in Västerbotten County. In 10 year intervals (starting at age 40), all residents of Västerbotten are invited to a general health exam. Participants fill out an extensive questionnaire on health and lifestyle, and may also donate a blood sample for research. The participation rate in the VIP is on average 70% (50% in earlier years), and no major selection has been found. The MONICA consists of randomly selected 25-74 year-olds living in the Northern Swedish Västerbotten and Norrbotten Counties, who were invited to participate in six cross-sectional health surveys between 1986-2014 (average participation rate 74%). The MSP, established in 1995 and concluded in 2006, invited women residing in Västerbotten County, approximately 50-70 years of age, who attended mammography screening to donate a blood sample and limited lifestyle data (participation in screening 85%, of which 33% donated a blood sample).

Follow-up
Participants were followed up from cohort entry (baseline) using the essentially complete Swedish national registries for the first event of: cancer diagnosis other than non-melanoma skin cancer (Swedish Cancer Registry), death (Swedish Cause of Death Registry), or migration (Swedish Registry of Total Population and
Population Changes). The completeness of the Swedish Cancer Registry is ensured due to mandatory reporting by law, which minimizes the risk of a selected cancer case population. Participants diagnosed with colorectal adenocarcinoma were identified using ICD-10 C18.0 and C18.2–18.9 for colon, C19.9 and C20.9 for rectum. CRC diagnoses were verified and data on tumor stage and anatomical site collected by linkage to the Swedish National Quality Registry for Colorectal Cancer or by medical record assessment by a single gastrointestinal pathologist. In paper I and II, participants were followed until May 31, 2016. In paper III and IV, participants were followed until March 31, 2009. The estimated cumulative incidence at age 85 in the NSHDS was 6.9% for men (95% CI: 6.0, 7.8%), and 5.7% for women (95% CI: 5.2, 6.2%), which was almost identical to the average cumulative incidence in the Swedish population (7.3% for men and 5.5% for women, Figure 6).

![Figure 6](image)

Figure 6. Cumulative incidence of CRC estimated in the NSHDS 1986–2016 and in Sweden between 1996-2015 from NORDCAN© 2017 Association of the Nordic Cancer Registries, IARC (Assessed 180612).

**Study design and participants**

Paper I and II were based on participants in the VIP and MONICA cohorts between 1986-2016 (Figure 7). Paper I was designed as a cohort study of all participants in VIP and MONICA. At the final date of study entrance (January 19, 2016), these cohorts included 119 738 participants with 183 699 observations. Participants diagnosed with cancer other than non-melanoma skin cancer before study entrance were excluded, as were observations entered <1 year before cancer diagnosis, with missing data on height or weight, or with implausible anthropometric/metabolic factor measurements (height <130 or >210 cm, weight <35kg, BMI<15 or >70 kg/m², fasting plasma glucose <1 mmol/l, glucose tolerance <1 or >35 mmol/l, total cholesterol <0.5 or >15 mmol/l, triglycerides...
<0.15 or >20 mmol/l, systolic blood pressure <20 or >300 mmHg, diastolic blood pressure <20 or >250 mmHg). For participants with more than one observation, we used the observation with the longest follow-up. After exclusions, the NSHDS included 117 687 participants (92% VIP). After up to 30.5 years of follow-up time (1.7 million person-years total follow-up), we identified 1250 verified CRC cases (median follow-up until CRC case diagnosis: 12.4 years, 52% men, 1122 in VIP and 128 in MONICA). A total of 24 cases (2%) lacked data on tumor site, and 102 lacked data on tumor stage (8%, mainly due to incomplete clinical staging).

Figure 7. Study design in paper I and II. *With no previous cancer diagnosis other than non-melanoma skin cancer.

Paper II was a nested case-control study based on prospective CRC cases within the VIP and MONICA cohorts diagnosed from 1986 to 2016 (Figure 7). A total of 1013 CRC cases had an available prediagnostic blood sample in the biobank. After excluding cases with different blood sampling and questionnaire dates (n=3), 1010 cases remained. The median follow-up time between baseline and case diagnosis was 12.3 years. A total of 8 cases (1%) had unknown tumor site, and 64 cases (6%) lacked data on stage. For each case, one control was randomly
selected, matched by sex, age at and year of blood sampling and data collection, fasting status, and cohort. All controls had to be alive and with no diagnosed cancer other than non-melanoma skin cancer at the time of diagnosis of their index cases.

Papers III and IV were based on a case-control data set which was available at the start of the PhD work included in this thesis. The data set consisted of participants in VIP and MSP with follow-up 1986-2009 (Figure 8), whose blood samples had been analyzed for a panel of factors and SNPs involved in one-carbon metabolism. At the final date of case identification for these papers (March 31, 2009), the VIP and MSP included 92,733 participants with 165,883 observations. After excluding five cases due to high methionine sulfoxide (indicating sample degradation), there were 613 available prospective CRC cases. The median follow-up time between baseline and case diagnosis was 8.2 years. Two controls were matched on the same variables as in paper II, of which some were excluded due to insufficient blood (n=8) or prioritized to other studies (n=18), resulting in 1190 available controls.

Figure 8. Study design in paper III and IV. *With no previous cancer diagnosis other than non-melanoma skin cancer.
Blood analyses

Venous blood samples in the NSDHS were aliquoted and frozen at -80°C and biobanked within one hour of collection, or at -20°C for at most one week prior to storage at -80°C. In the VIP, blood samples are collected in the morning following an overnight fast. In the MONICA, blood samples were collected after fasting for a minimum of 4 hours up until 1992, and 8 hours after 1992. In the MSP cohort, sampling was spread out during the day. In paper I and II (VIP and MONICA), approximately 80% of the study participants had fasted for more than 8 hours and 3% less than 4 hours. In paper III and IV (VIP and MSP), approximately 60% had fasted more than 8 hours, and 23% less than 4 hours. All samples were analyzed keeping case-control sets together, with random positioning of the case. All investigators and laboratory staff were blinded to case and control status until statistical analyses.

In paper II, plasma concentrations of insulin, C-peptide, adiponectin, and leptin were measured in pre-coated 96-well plates with a custom designed multiplex immunoassay (7-spot Prototype Human Metabolic 5-plex) from Meso Scale Discovery (Rockville, MD, United States). Adiponectin was measured separately with the Human Adiponectin Kit, also from Meso Scale Discovery (Rockville, MD, United States). All reagents were from the manufacturer, and all assays were run according to their instructions. Inter- and intra-assay coefficients of variation (CVs), estimated based on internal control samples, were low for all biomarkers (inter/intra-assay CV (%): insulin (3.2/0.7), C-peptide (2.7/1.4), adiponectin (1.4/0.5), leptin (1.5/0.4). For insulin and leptin, there was an indication of laboratory drift manifested as somewhat higher plasma concentrations in the internal control samples for the first five plates compared to the last 25 plates. Therefore, these biomarkers were normalized to the last plate (plate 30) by multiplying measurements with the control sample ratio according to the formula:  

$$ Y'_{i,b} = Y_{i,b} \times \frac{\bar{C}_b}{\bar{C}_{30}} $$

where $Y'_{i,b}$ are the adjusted and $Y_{i,b}$ the unadjusted biomarker measurements, respectively, for individual $i$ in batch $b$, and $\bar{C}_b$ and $\bar{C}_{30}$ are the mean of the control sample measurements in plate $b$ and 30, respectively. Biomarker measurements below the curve range (0 to 2% of samples per biomarker) were assumed to be low, and were therefore replaced with the plate-specific minimum concentration. In repeated measures in controls taken less than 10 years apart (n=28), we observed good to excellent intra-class correlations for all biomarkers (ICCs, insulin: 0.63, C-peptide: 0.69, adiponectin: 0.91, leptin: 0.82).

For paper III and IV, all biomarker analyses on biobanked EDTA plasma were performed at Bevital AS (Bergen, Norway, http://www.bevital.no/). The included biomarkers were chosen based on previous studies of one-carbon metabolism and CRC development, and to represent the many aspects of one-carbon
metabolism while maintaining adequate biomarker stability and reproducibility.\textsuperscript{223,224} Plasma concentrations of cystathionine, vitamin B2 (riboflavin), vitamin B6 (pyridoxal 5'-phosphate), methionine, choline, betaine, dimethylglycine, creatinine, and neopterin were measured with liquid chromatography mass spectrometry methods (between-day CV: 3-13 \%, within-day CV: 2-8\%).\textsuperscript{225} Plasma concentrations of total homocysteine, total cysteine, serine, glycine, and sarcosine were measured using an isotope dilution gas chromatography mass spectrometry method (between-day CV: 2-9 \%, within-day CV: 1-2\%).\textsuperscript{226} Folate and vitamin B12 (cobalamin) concentrations were determined with a microbiological method using \textit{Lactobacillus casei} and \textit{Lactobacillus leichmannii}, respectively, which was adapted to a microtiter plate format and carried out by a robotic workstation (between-day CV: 5 \%, within-day CV: 4\%).\textsuperscript{227,228} To adjust biomarker associations for genetic confounding and assess associations of genetically determined one-carbon metabolism status, we also included a panel of SNPs in genes involved in one-carbon metabolism, some of which had previously been studied in relation to CRC.\textsuperscript{159} SNPs were determined using MALDI-TOF mass spectrometry.\textsuperscript{229} A total of 17 SNPs in 13 genes were included: \textit{BHMT} (rs3733890), \textit{CBS} (rs234706, 844ins68), \textit{CTH} (rs1021737), \textit{DHFR} (rs70991108), \textit{FOLR1} (rs2071010), \textit{MTHFD1} (rs2236225), \textit{MTHFR} (rs1801131, rs1801133), \textit{MTR} (rs1805087), \textit{MTRR} (rs1532268, rs1801394), \textit{SHMT} (rs1979277), \textit{SLC19A1} (rs1051266), \textit{TCN2} (rs1801198, rs9606756), and \textit{TYMS} (rs34489327).

\textbf{Tumor tissue analyses}

DNA was extracted and purified from formalin-fixed, paraffin-embedded tumor tissue collected during routine clinical practice at the Department of Clinical Pathology, Umeå University Hospital, Umeå, Sweden. In paper I, 916 of the included 1250 cases (73\%) had available tumor tissue. In paper II, 841 of 1010 cases with available prediagnostic blood (83\%) had tumor tissue. In paper IV, 570 of the 613 (93\%) had available tumor tissue.

\textit{KRAS} was analyzed by sequencing the activating mutations in codon 12 and 13 using Big Dye v. 3.1, according to the manufacture protocol (Applied Biosystems, Life Technologies, Foster City, CA, United States).\textsuperscript{230} The mutation status of \textit{BRAF}\textsuperscript{V600E} was analyzed using TaqMan allelic discrimination assay (reagents from Applied Biosystems)\textsuperscript{231} or digital droplet PCR in later years (reagents from Bio-Rad Laboratories, Hercules, CA, United States). As \textit{KRAS} and \textit{BRAF} are considered mutually exclusive within cell clones, CRC cases were classified as \textit{KRAS}-mutated, \textit{BRAF}-mutated, or \textit{KRAS/BRAF} wild-type. Cases with mutations in both \textit{KRAS} and \textit{BRAF} were censored/removed in the analyses (1-2\%, in the complete data sets).
Microsatellite instability (MSI) status was assessed by immunohistochemical analysis of mismatch repair proteins MLH1, MSH2, MSH6, and PMS2 for cases diagnosed up until 2009 (with MSI defined as samples lacking tumor cells with nuclear staining for any of the proteins). For cases diagnosed 2009-2016 a DNA-based method was used (Promega MSI Analysis System, Version 1.2, Madison, WI, United States). A set of cases (n=70) were analyzed with both methods, with 100% concordance.

For paper IV, CIMP status was determined using MethyLight real-time PCR on bisulfate treated and purified DNA (0.5 μl DNA, EZ or EZ-96 DNA methylation kit, Zymo Research, Orange, CA, USA). An established 8-gene panel (CDKN2A, MLH1, CACNA1G, NEUROG1, RUNX3, SOCS1, IGF2, and CRABP1) were amplified on a TaqMan 7900 HT sequence detection system (Applied Biosystems, Life Technologies, Foster City, CA, United States), and a percentage of methylated reference (PMR) value calculated. Cases were classified as CIMP-negative if there was no promoter methylation in any of the eight genes defined as PMR value above 10%, CIMP-low if there was promoter methylation in one to five genes, and CIMP-high if there was promoter methylation in six to eight genes. CIMP-positive tumors were defined as either CIMP-low or high.

Some cases had inconclusive KRAS or BRAF mutation status (paper I: n=144, paper II: n=133, and paper IV: n=82), MSI status (paper I: n=145, paper II: n=132, and paper IV: n=91), or CIMP status (paper IV: n=96), mostly due to insufficient DNA in the tissue sample. Thus, in total, 766 cases had complete KRAS/BRAF data in paper I (Figure 7), 704 cases had complete KRAS/BRAF data and 708 complete MSI data in paper II (Figure 7), and 488 cases had complete KRAS/BRAF data in paper IV (Figure 8).

**Variables**

The lifestyle data in the VIP and MONICA consist of measurements made by a health professional, as well as self-reported questionnaire data. The lifestyle data in MSP are more limited, in these studies limited to only height and weight.

Height and weight were measured in light clothing without shoes in all cohorts. In the VIP, a capillary blood sample was drawn after an overnight fast, and again 2h after a 75g oral glucose load to measure glucose tolerance. Plasma glucose concentrations were analyzed with a Reflotron bench-top analyzer (Roche Diagnostics) until 2004, and after that with Hemocuse bench-top analyzer (Quest Diagnostics). Blood lipids, total cholesterol and triglyceride levels, were analyzed with a Reflotron bench-top analyzer until 2009, and from 2009 with an enzymatic method at a clinical chemistry laboratory at the nearest hospital. Systolic and diastolic blood pressure was measured once with a mercury
sphygmomanometer after a 5-minute rest in the supine position until 2009, and from 2009 in the sitting position. Total cholesterol, triglyceride, and blood pressure measurements were adjusted to make measurements before and after 2009 comparable using formulas developed by the cohort estimated from individuals measured with both methods at the same sampling occasion. Blood glucose, blood lipid, and blood pressure measurements in the MONICA were measured with the above-mentioned most recent methods throughout all study years. We adjusted blood lipid and blood pressure measurements for lipid-lowering or antihypertensive medication usage by adding estimated constants (total cholesterol + 1.347 mmol/l, and triglyceride levels + 0.208 mmol/l, systolic blood pressure + 15 mmHg, diastolic blood pressure +10 mmHg).

Self-reported questionnaire data in VIP and MONICA used in the studies included: educational status (elementary school, junior secondary school, upper secondary school, or university education), smoking status (never-, ex-, or current smoker), recreational physical activity (on a scale of 1-5, where 1=never, 2=every now and then - not regularly, 3=1-2 times/week, 4=2-3 times/week, 5=more than 3 times/week), occupational physical activity (on a scale of 1-5, where 1=sedentary or standing work, 2=light but partly physically active, 3=light and physically active, 4=sometimes physically straining, 5=physically straining most of the time), and alcohol intake from a validated food frequency questionnaire (zero intake, and above/below sex and questionnaire version-specific median of self-reported grams/day intake).

In paper III and IV, we estimated glomerular filtration rate (eGFR) as a marker for kidney function calculated by the Cockcroft-Gault formula (based on plasma creatinine levels, age, sex, and body weight).

### Statistical analyses

#### General
All computations were conducted in R (v.3.0.2 to 3.5.0, R Foundation for Statistical Computing, Vienna, Austria). All statistical tests were 2-sided when applicable. All code can be provided on request to the author.

Associations between continuous variables were estimated using Spearman’s correlations or linear regression. In paper I and II, partial correlations between metabolic factors and biomarkers adjusted for age, sex, and BMI were calculated by correlating residuals from linear regressions including the adjustment factors. Associations between continuous and categorical variables were assessed with linear regression or tested with Wilcoxon’s nonparametric tests. Associations between categorical variables were tested using Chi-square tests.
**Associations between exposures and CRC and CRC subtype risk**

In paper I, associations between metabolic factors and CRC risk were evaluated by estimating relative risks (RRs) per 1 standard deviation (SD) increase in metabolic traits using Cox proportional hazards regression, with age as the time scale. The metabolic factors were included as continuous z-transformed (scaled to mean 0 and SD 1) variables in the models. Z-transformations were made separately by sex and cohort for BMI and blood pressure, and by sex, cohort, and fasting status (above or below 8 hours) for glucose, glucose tolerance, cholesterol, and triglycerides. Because of a skewed distribution, the glucose, glucose tolerance, and triglycerides variables were log-transformed before z-transformation. RRs were adjusted for sex, cohort, smoking, recreational and occupational physical activity, alcohol intake, and BMI. The proportional hazards assumption was checked by graphically examining time-dependent log(RRs) and in Schoenfeld residual-based tests. No violations were observed. To assess linearity in the associations, continuous variables were modeled using restricted cubic splines (with knots at the 5th, 50th and 95th percentiles). Nonlinearity was tested with a likelihood ratio test comparing the spline model to a linear model.

In papers II, III, and IV, associations between biomarkers and SNPs and CRC risk were evaluated by estimating odds ratios (ORs) using conditional logistic regression models conditioned on the matched case sets. For biomarkers, we estimated ORs per 1 SD increase (by modeling z-transformed log-transformed biomarker concentrations) or by tertiles of the biomarker distributions (with cut-offs based on the distribution of the controls). For the SNPs in paper IV, we estimated ORs per allele. Besides the matching variables, we evaluated adjustment by smoking, recreational and occupational physical activity, alcohol intake, and BMI. In studies of one-carbon metabolism biomarkers and SNPs (paper III and IV), we also evaluated adjustment by eGFR, plasma neopterin concentrations as a marker of immune system activation (tertiles), plasma concentrations of B-vitamins (B2, B6, and B12) and homocysteine, metabolic factors, and one-carbon metabolism SNPs. In paper II, biomarkers were also modeled using restricted cubic splines in conditional logistic regression models (with knots at the 5th, 50th and 95th percentiles) to check for nonlinear associations.

In paper III, to improve the validity and interpretability of results, we estimated absolute marginal risk differences (RDs). RDs were estimated with a weighted maximum likelihood estimator, using cumulative incidence at end of follow-up from the study cohort at large, and within groups defined by sampling year, age, sex, and cohort (cumulative incidence of CRC in the study cohort was 830 per 100,000 over the period 1987-2009).
To evaluate whether the associations between the exposures and CRC risk differed by anatomical site or molecular subtypes, we estimated subtype-specific RRs or ORs for the exposures in Cox regression models (paper I) or conditional logistic regression models (papers II and IV) estimated using a competing risks approach called the duplication method. Heterogeneity was tested with a likelihood ratio test, comparing a model in which the risk association could vary across subtypes to a model in which all associations were held constant. A low P-value indicates that RRs or ORs for CRC differ by subtypes in the population.

**Bayesian network learning (paper IV)**

Given the complexity of one-carbon metabolism, univariate modeling of single variables may miss higher-order interactions and mediations. Therefore, in paper IV, we modeled all variables in relation to CRC risk by KRAS and BRAF mutation status simultaneously using multivariate Bayesian network learning. Bayesian network learning is a data-driven method to estimate independent associations between all modeled variables, both exposures and outcomes, and presents these graphically as a network. The method reduces the risk of chance findings due to multiple testing, as all associations are estimated simultaneously, while still adjusting for complex relations between many variables. This approach has previously been applied to similar research questions. Other methods that estimate associations between many related exposures and an outcome, such as random forests, do not provide estimates of independent associations between exposures. Bayesian network learning, therefore, paints a more complete picture of the complex relationships in a data set. In this thesis, the Bayesian networks were estimated using the bnlearn R-package. In 1000 bootstrap samples, networks were estimated on discrete data with the Hill-climbing algorithm using the Akaike information criterion (AIC) score. Briefly, this algorithm finds the network model that best fit the data according to AIC by attempting all possible edge additions and removals until convergence is reached. The continuous biomarkers were included as categorized tertile variables, with cut-offs based on the distribution of the controls. SNPs were included as dichotomous variables (common or variant genotype). The final network was obtained by averaging over the 1000 bootstrap networks. An edge, indicating an independent association between two variables, was included if the edge confidence, defined as the proportion of times an edge was present among the 1000 bootstrap networks, was above an estimated significance threshold. Edge confidence was also used to measure the strength of the association with CRC subtype risk for each one-carbon metabolism exposure.
Methods for handling missing data

Missing data on metabolic traits and biomarkers and potential confounders in paper I and II were assumed to depend on observable characteristics, so called missing-at-random. Therefore, these data were imputed using multiple imputation by chained equations with the mice R-package.\textsuperscript{244} Data were imputed in a number of data sets (5 in paper I, 10 in paper II), in a number of iterations (15-25), with a predictive mean matching model. The models included metabolic factors/biomarkers, and age at and year of blood sampling and/or data collection, sex, fasting status, cohort, educational status, smoking, recreational and occupational physical activity, alcohol intake, event status, and follow-up time as predictors. We graphically checked for convergence of imputed values across iterations, and compared distributions of imputed and observed values as a plausibility check. All statistical analyses were run separately on the imputed data sets and then aggregated using Rubin’s rules.\textsuperscript{245} In papers III and IV, missing data on exposures was uncommon and assumed to be missing-completely-at-random, and therefore omitted separately in each analysis. Missing values for potential confounders were in these papers were assigned to a separate “missing” category.

Tumor data in papers I, II, and IV were unavailable for a portion of the included CRC cases. The probability of missing tumor data depends on clinical characteristics such as tumor site and stage, and potentially other observable characteristics (i.e., missing-at-random data). In this thesis, cases with missing tumor data were more often diagnosed in recent years, distal tumors, and advanced stage tumors. To account for potential selection bias, we applied two different methods. In paper I and II, we used the same multiple imputation procedure as described above. Missing data on tumor stage, site, \textit{KRAS} and \textit{BRAF} mutation status, and MSI status were imputed separately in cases in models including the tumor variables as well as age at and year of diagnosis, sex, cohort, educational status, smoking, recreational and occupational physical activity, alcohol intake, and BMI. In paper IV we used inverse probability weighting.\textsuperscript{246} We first fitted a conditional logistic regression model in all cases with tumor data availability as the outcome. Included predictors were tumor stage, tumor site, age at and year of diagnosis, cohort, and sex. Cases without data on tumor stage or site were excluded from these analyses (n=29). Then, for each case \(i\), the model was used to estimate fitted probabilities, \(p_i\). These fitted probabilities were then used in selection bias-adjusted weighted conditional logistic regression models for CRC subtype risk, fitted as in the complete case analysis. Weights were set to 1 for controls, \(\frac{1}{p_i}\) for cases with available tumor data, and 0 for cases without available tumor data (i.e., not included in the estimation). To make sure the choice of missing tumor data method did not affect our results, we also ran analyzes in paper I and II using the inverse probability weighting method used in paper IV, with very similar results.
The Cancer Genome Atlas replication analysis (paper IV)
To externally validate a potential SNP finding in paper IV, we utilized The Cancer Genome Atlas (TCGA) data generated by the TCGA Research network (https://cancergenome.nih.gov/). Case-case analyses were made for 533 colorectal cancer patients with data available for both somatic mutation calls and germline Affymetrix Genome-Wide Human SNP6.0 array data that passed the quality control. KRAS (activating mutations in codon 12 or 13) and BRAF (BRAF\textsuperscript{V600E}) mutation status for TCGA COAD + READ individuals was obtained from the United States National Cancer Institute’s Genomic Data Commons repository (https://gdc.cancer.gov). Germline SNP genotypes were obtained by imputation to the Haplotype Reference Consortium. Raw Affymetrix Genome-Wide Human SNP 6.0 array data were processed using a standardized GWAS quality control pipeline. Genotypes called with the Birdseed genotype-calling algorithm were set to missing if their confidence score was >0.1. Samples with genotype missing call rate (<98%), and samples with mismatches between genotypic and reported sex were excluded. We estimated haplotype phase using SHAPEIT\textsuperscript{v2} and imputed to the Haplotype Reference Consortium panel using the SNP6 array data as imputation target. Ratio of odds ratios (RORs) per allele of the SNP was calculated for KRAS-mutated and BRAF-mutated tumors vs KRAS/BRAF wild type tumors using multinomial logistic regression. Heterogeneity in the SNP-CRC KRAS/BRAF subtype association was tested with a likelihood ratio test. To facilitate comparisons, identical case-case analyses were made on the NSHDS data.

Ethical considerations
All participants gave a written informed consent for all collection made for research purposes. The papers in this thesis are observational studies based on collected data, blood samples, and tumor tissue samples. Therefore, there was no risk of further physical harm or discomfort by conducting the studies. The findings were expected to be of importance on a population, and not individual, level. As such, the results were not anticipated to be of any relevance to the study participants. All data handling conducted outside the university or hospital servers were conducted on anonymized data and only aggregated results on large groups are presented, ensuring the personal integrity of the study participants. Overall, the risks for the study participants were judged to be minimal, and the potential knowledge gains from the studies, large.

The study protocols for all papers in this thesis was approved by the Regional Ethical Review Board in Umeå, Sweden (Dnr 2015/243-32 and 2017/172-32M for papers I and II, Dnr 2003-186 for paper III, and Dnr 2015/167-32 for paper IV). The Swedish Data Protection Authority approved all data handling procedures.
Results and Discussion

Baseline and case characteristics

Papers I and II

Characteristics of participants in paper I and II are displayed in Table 2. Both study cohorts consisted of roughly equal parts men and women. Baseline age was higher for paper II participants due to the case-control study design. The median BMI for paper I participants was 25.3 kg/m² and 21% were current smokers. In the paper II, the median BMI was 25.7 kg/m² and 24% were current smokers. Clinical characteristics of CRC cases were similar in the two papers, as almost all cases in paper II were included in paper I. The median age at diagnosis was 66.7 in paper I and 66.4 in paper II, the relatively young age reflecting the recruitment pool and follow-up period in the VIP. About two-thirds of the tumors were situated in the colon in both papers. Tumors were equally distributed across early and more advanced stages at diagnosis (51% stage I and II, 49% stage III and IV).

Table 2. Characteristics of NSHDS participants in paper I and II (VIP and MONICA 1986-2016). Median (quartiles) or n (%).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Paper I</th>
<th>Paper II</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Baseline characteristics</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Participants</td>
<td>117,687 participants</td>
<td>1,010 cases, 1,010 controls</td>
</tr>
<tr>
<td>Age at baseline, years</td>
<td>41.6 (40.0-41.6)ᵃ</td>
<td>56.3 (49.9-56.3)</td>
</tr>
<tr>
<td>Cohort</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VIP</td>
<td>108,107 (92%)</td>
<td>1,854 (92%)</td>
</tr>
<tr>
<td>MONICA</td>
<td>9589 (8%)</td>
<td>166 (8%)</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Men</td>
<td>58,493 (50)</td>
<td>1,050 (52)</td>
</tr>
<tr>
<td>Women</td>
<td>59,194 (50)</td>
<td>970 (48)</td>
</tr>
<tr>
<td>BMI, kg/m²</td>
<td>25.3 (23.0-25.3)</td>
<td>25.7 (23.5-25.7)</td>
</tr>
<tr>
<td>Current smokers</td>
<td>24,884 (21)</td>
<td>465 (24)</td>
</tr>
<tr>
<td><strong>Case characteristics</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of prospective cases, n</td>
<td>1250</td>
<td>1,010</td>
</tr>
<tr>
<td>with KRAS/BRAF mutation data</td>
<td>766</td>
<td>704</td>
</tr>
<tr>
<td>with MSI data</td>
<td>759</td>
<td>708</td>
</tr>
<tr>
<td>Age at diagnosis, years</td>
<td>66.7 (60.8-66.7)</td>
<td>66.4 (60.5-66.4)</td>
</tr>
<tr>
<td>Tumor siteᵇ</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Right-sided colon</td>
<td>388 (32)</td>
<td>318 (32)</td>
</tr>
<tr>
<td>Left-sided colon</td>
<td>361 (29)</td>
<td>296 (30)</td>
</tr>
<tr>
<td>Rectum</td>
<td>477 (39)</td>
<td>388 (39)</td>
</tr>
<tr>
<td>Tumor stageᶜ</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stage I and II</td>
<td>582 (51)</td>
<td>478 (51)</td>
</tr>
<tr>
<td>Stage III and IV</td>
<td>566 (49)</td>
<td>468 (49)</td>
</tr>
</tbody>
</table>

ᵃ Age at baseline 56.6 (49.9-56.6) years in CRC cases.
ᵇ 24 cases lacked data on tumor site in paper I, and 8 lacked data in paper II.
ᶜ 102 cases lacked data on tumor stage in paper I, and 64 lacked data in paper II.
Correlations between exposures in paper I and II are presented in Figure 9. BMI was correlated with most metabolic factors and biomarkers (r=0.2 to 0.6). BMI-independent relationships included correlations among insulin resistance-related factors (insulin, C-peptide, glucose, and glucose tolerance, r=0.3 to 0.7), blood lipids (cholesterol and triglycerides, r=0.3), and blood pressure (systolic and diastolic blood pressure, r=0.7). Triglycerides were also correlated with insulin and C-peptide (r=0.2 and 0.3), and adiponectin was correlated with C-peptide (r=-0.3).

**Figure 9. Correlation network of metabolic factors and biomarkers included in paper I and II.** Correlations larger than 0.2 are displayed. Calculated in paper II controls (n=1010), adjusted for age, sex, and BMI. BP: Blood pressure.

**Table 3** displays CRC case characteristics by KRAS and BRAF mutation status in cases included in paper I. BRAF mutations were more common in women, in patients who were older at diagnosis, in right-sided colon tumors, and in MSI tumors. KRAS-mutated tumors were approximately equally distributed across tumor sites, whereas double-wildtype tumors were more often situated in the rectum. These associations were the same for the subgroup of cases included in paper II.
Table 3. Colorectal cancer case characteristics by KRAS and BRAF mutation status in paper I. Median (quartiles) or n (%)

<table>
<thead>
<tr>
<th>Variable</th>
<th>BRAF-mutated (n=169)</th>
<th>KRAS-mutated (n=184)</th>
<th>Both wildtype (n=484)</th>
<th>P (a)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age at diagnosis, years</td>
<td>69.9 (45.2-85.9)</td>
<td>67.4 (37.7-85.4)</td>
<td>66.0 (41.0-89.6)</td>
<td>1.9(\times10^-5)</td>
</tr>
<tr>
<td>Sex, women</td>
<td>109 (64)</td>
<td>96 (52)</td>
<td>176 (43)</td>
<td>7.7(\times10^-6)</td>
</tr>
<tr>
<td>Tumor site</td>
<td></td>
<td></td>
<td></td>
<td>1.1(\times10^-29)</td>
</tr>
<tr>
<td>Right colon</td>
<td>123 (73)</td>
<td>63 (34)</td>
<td>86 (21)</td>
<td></td>
</tr>
<tr>
<td>Left colon</td>
<td>26 (15)</td>
<td>55 (30)</td>
<td>139 (34)</td>
<td></td>
</tr>
<tr>
<td>Rectum</td>
<td>20 (12)</td>
<td>65 (36)</td>
<td>185 (45)</td>
<td></td>
</tr>
<tr>
<td>Tumor stage</td>
<td></td>
<td></td>
<td></td>
<td>0.37</td>
</tr>
<tr>
<td>Stage I&amp;II</td>
<td>80 (49)</td>
<td>95 (56)</td>
<td>215 (55)</td>
<td></td>
</tr>
<tr>
<td>Stage III&amp;IV</td>
<td>84 (51)</td>
<td>76 (44)</td>
<td>178 (45)</td>
<td></td>
</tr>
<tr>
<td>MSI status</td>
<td></td>
<td></td>
<td></td>
<td>7.8(\times10^-4)</td>
</tr>
<tr>
<td>MSI</td>
<td>72 (49)</td>
<td>3 (2)</td>
<td>21 (6)</td>
<td></td>
</tr>
<tr>
<td>MSS</td>
<td>74 (51)</td>
<td>148 (98)</td>
<td>345 (94)</td>
<td></td>
</tr>
</tbody>
</table>

\(^a\) P-values from linear regression tests of equal distributions across subtypes for age at diagnosis or chi-square tests of equal distributions for categorical variables.

**Papers III and IV**

Table 4 displays characteristics of paper III and IV case-control study participants. There was a higher proportion of women among participants in these papers compared to papers I and II due to the all-female MSP cohort (59%). The median age at baseline was 59.7, the median BMI was 25.6 kg/m\(^2\), and 20% were current smokers. Folate levels were relatively low, \(249\) median 7.2 nmol/L. The median age at diagnosis for cases was 65.2 years, and cases were approximately equally distributed across tumor sites and early/more advanced tumor stages.
<table>
<thead>
<tr>
<th>Variable</th>
<th>Papers III and IV</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Baseline characteristics</strong></td>
<td></td>
</tr>
<tr>
<td>Participants</td>
<td>613 cases, 1190 controls</td>
</tr>
<tr>
<td>Age at baseline, years</td>
<td>59.7 (50.1-59.7)</td>
</tr>
<tr>
<td>Cohort</td>
<td></td>
</tr>
<tr>
<td>VIP</td>
<td>1410 (78)</td>
</tr>
<tr>
<td>MSP</td>
<td>393 (22)</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
</tr>
<tr>
<td>Men</td>
<td>740 (41)</td>
</tr>
<tr>
<td>Women</td>
<td>1063 (59)</td>
</tr>
<tr>
<td>BMI, kg/m²</td>
<td>25.6 (23.4-25.6)</td>
</tr>
<tr>
<td>Current smokers</td>
<td>353 (20)</td>
</tr>
<tr>
<td>Folate, nmol/L</td>
<td>7.2 (4.7-7.2)</td>
</tr>
<tr>
<td><strong>Case characteristics</strong></td>
<td></td>
</tr>
<tr>
<td>Number of prospective cases, n</td>
<td>613</td>
</tr>
<tr>
<td>with KRAS/BRAF mutation data</td>
<td>488&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>with MSI data</td>
<td>392&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>with CIMP data</td>
<td>397&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>Age at diagnosis, years</td>
<td>65.2 (59.3-70.2)</td>
</tr>
<tr>
<td>Tumor site&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>Right-sided colon</td>
<td>183 (30)</td>
</tr>
<tr>
<td>Left-sided colon</td>
<td>215 (35)</td>
</tr>
<tr>
<td>Rectum</td>
<td>214 (35)</td>
</tr>
<tr>
<td>Tumor stage&lt;sup&gt;c&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>Stage I and II</td>
<td>308 (53)</td>
</tr>
<tr>
<td>Stage III and IV</td>
<td>276 (47)</td>
</tr>
</tbody>
</table>

<sup>a</sup> These 488 cases and their 947 matched controls were included in paper IV.
<sup>b</sup> Out of the 488 cases in paper IV.
<sup>c</sup> 29 cases lacked data on tumor site and 1 case lacked data on tumor stage.

CRC subtypes defined by KRAS and BRAF mutation status displayed expected clinical and molecular characteristics (Table 5). In addition to the associations mentioned for case participants in paper I and II, BRAF-mutated cases were more often CIMP-high. and KRAS-mutated cases more often CIMP-low, compared to the other subtypes.
Table 5. Colorectal cancer case characteristics by KRAS and BRAF mutation status in paper IV. Median (quartiles) or n (%).

<table>
<thead>
<tr>
<th>Variable</th>
<th>BRAF-mutated (n=117)</th>
<th>KRAS-mutated (n=125)</th>
<th>Both wildtype (n=246)</th>
<th>P*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age at diagnosis, years</td>
<td>67.2 (62.7-72.1)</td>
<td>65.8 (58.0-71.8)</td>
<td>64.2 (58.3-69.2)</td>
<td>0.002</td>
</tr>
<tr>
<td>Sex, women</td>
<td>76 (61)</td>
<td>86 (74)</td>
<td>131 (53)</td>
<td>0.001</td>
</tr>
<tr>
<td>Tumor site</td>
<td></td>
<td></td>
<td></td>
<td>4.1*10^-44</td>
</tr>
<tr>
<td>Right colon</td>
<td>70 (60)</td>
<td>40 (32)</td>
<td>47 (19)</td>
<td>0.16</td>
</tr>
<tr>
<td>Left colon</td>
<td>35 (30)</td>
<td>41 (33)</td>
<td>90 (37)</td>
<td>0.16</td>
</tr>
<tr>
<td>Rectum</td>
<td>12 (10)</td>
<td>44 (35)</td>
<td>108 (44)</td>
<td>0.16</td>
</tr>
<tr>
<td>Tumor stage</td>
<td></td>
<td></td>
<td></td>
<td>5.4*10^-21</td>
</tr>
<tr>
<td>Stage I&amp;II</td>
<td>54 (47)</td>
<td>69 (59)</td>
<td>134 (56)</td>
<td>5.4*10^-21</td>
</tr>
<tr>
<td>Stage III&amp;IV</td>
<td>60 (53)</td>
<td>47 (41)</td>
<td>107 (44)</td>
<td>5.4*10^-21</td>
</tr>
<tr>
<td>MSI status</td>
<td></td>
<td></td>
<td></td>
<td>2.5*10^-34</td>
</tr>
<tr>
<td>MSI</td>
<td>46 (48)</td>
<td>2 (2)</td>
<td>18 (9)</td>
<td>2.5*10^-34</td>
</tr>
<tr>
<td>MSS</td>
<td>49 (52)</td>
<td>97 (98)</td>
<td>185 (91)</td>
<td>2.5*10^-34</td>
</tr>
<tr>
<td>CIMP status</td>
<td></td>
<td></td>
<td></td>
<td>2.5*10^-34</td>
</tr>
<tr>
<td>High</td>
<td>54 (57)</td>
<td>2 (2)</td>
<td>10 (5)</td>
<td>2.5*10^-34</td>
</tr>
<tr>
<td>Low</td>
<td>31 (33)</td>
<td>42 (43)</td>
<td>63 (32)</td>
<td>2.5*10^-34</td>
</tr>
<tr>
<td>Negative</td>
<td>9 (10)</td>
<td>54 (55)</td>
<td>127 (63)</td>
<td>2.5*10^-34</td>
</tr>
</tbody>
</table>

a P-values from linear regression tests of equal distributions across subtypes for age at diagnosis or chi-square tests of equal distributions for categorical variables.

Paper I results

Main results

In this cohort study of metabolic factors related to the metabolic syndrome, we confirmed previously reported linear associations between higher BMI, total cholesterol, triglycerides, and diastolic blood pressure and an increased risk of overall CRC (RRs per 1 SD increase: 1.06-1.12, all Pnonlinearity>0.01, Figure 10). Associations were similar in men and women, and slightly stronger for colon cancer compared to rectal cancer for most factors, yet, there was no significant heterogeneity (all Pheterogeneity>0.2).
Associations between metabolic factors and the risk of CRC subtypes by KRAS and BRAF mutation status are presented in Figure 11A. All associations between metabolic factors and CRC risk were similar regardless of subtype, and no new subtype-specific associations appeared. RRs for certain CRC subtypes were higher for some factors, such as the higher RR for BRAF-mutated CRC risk per 1 SD increase in triglycerides (RRs: 1.32 for BRAF-mutated, 1.06 for KRAS-mutated, and 1.08 for both wild type CRC risk). However, the test for heterogeneity implied no difference in the underlying population (P$_{\text{heterogeneity}}=0.34$). No other test for heterogeneity indicated differences by KRAS and BRAF mutation status in associations between metabolic factors and CRC risk (all P$_{\text{heterogeneity}}>0.4$). The same was true in secondary analyses of subtypes defined by MSI status (all P$_{\text{heterogeneity}}>0.5$, Figure 11B). The results were similar in men and women, and in separate analyses of colon and rectal cancer.
A.  

<table>
<thead>
<tr>
<th>Factor</th>
<th>P</th>
<th>P_{het}</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Triglycerides</td>
<td>0.007</td>
<td>0.34</td>
<td>0.03 - 0.09</td>
</tr>
<tr>
<td>Diastolic BP</td>
<td>0.12</td>
<td>1.00</td>
<td>0.03 - 0.16</td>
</tr>
<tr>
<td>Cholesterol</td>
<td>0.07</td>
<td>0.76</td>
<td>0.10 - 0.52</td>
</tr>
<tr>
<td>BMI</td>
<td>0.88</td>
<td>0.42</td>
<td>0.09 - 0.46</td>
</tr>
<tr>
<td>Systolic BP</td>
<td>0.11</td>
<td>0.45</td>
<td>0.59 - 0.11</td>
</tr>
<tr>
<td>Glucose tolerance</td>
<td>0.60</td>
<td>0.79</td>
<td>0.59 - 0.68</td>
</tr>
<tr>
<td>Glucose</td>
<td>0.86</td>
<td>0.97</td>
<td>0.83 - 0.88</td>
</tr>
</tbody>
</table>

B.  

<table>
<thead>
<tr>
<th>Factor</th>
<th>P</th>
<th>P_{het}</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Triglycerides</td>
<td>0.08</td>
<td>0.59</td>
<td>0.03 - 0.08</td>
</tr>
<tr>
<td>Diastolic BP</td>
<td>0.16</td>
<td>0.80</td>
<td>0.04 - 0.16</td>
</tr>
<tr>
<td>Cholesterol</td>
<td>0.30</td>
<td>0.77</td>
<td>0.03 - 0.30</td>
</tr>
<tr>
<td>BMI</td>
<td>0.93</td>
<td>0.60</td>
<td>0.03 - 0.93</td>
</tr>
<tr>
<td>Systolic BP</td>
<td>0.26</td>
<td>0.66</td>
<td>0.36 - 0.26</td>
</tr>
<tr>
<td>Glucose tolerance</td>
<td>0.55</td>
<td>0.62</td>
<td>0.43 - 0.55</td>
</tr>
<tr>
<td>Glucose</td>
<td>0.89</td>
<td>0.85</td>
<td>0.77 - 0.89</td>
</tr>
</tbody>
</table>

BRAF-‐mutated (n=169)  
KRAS-‐mutated (n=184)  
Both wild type (n=413)  
MSI (n=104)  
MSS (n=655)

**Figure 11. Relative risks (RRs) for CRC subtypes by metabolic factors.** RRs (95% CIs) per 1 SD increase in metabolic factors for CRC subtypes defined by (A) KRAS/BRAF mutation status and (B) MSI status were calculated with Cox regression using age as time scale, further adjusted for sex, cohort, smoking, recreational and occupational physical activity, alcohol intake, and BMI. Numbers (n) within subtypes represent CRC cases with complete molecular data, data for the remaining cases were imputed. Factors are presented in descending order according to RRs of overall CRC risk.

**Interpretation**

In line with previous studies,\textsuperscript{79,80} we observed linear associations between metabolic factors related to the metabolic syndrome, namely higher BMI, triglycerides, total cholesterol, and diastolic blood pressure, and an increased risk of CRC. There was no association between glucose variables and CRC risk, which is in line with the results in larger studies and the moderate heterogeneous association reported in the most recent meta-analysis.\textsuperscript{101} Previous MPE studies of BMI have not reported differences in the association with CRC risk by major molecular subtypes, including subtypes defined by KRAS and/or BRAF mutation status.\textsuperscript{166,170-174} Furthermore, BMI appears to be associated with the risk of both conventional and serrated colorectal adenomas.\textsuperscript{190-192} The previous results on BMI and CRC risk by KRAS and BRAF mutation status are based on three case-control studies and two cohort studies, with n=491 to 1451 colorectal, colon, or rectal cancer cases. One study, Brändstedt et al (n=491),\textsuperscript{172} included data on both KRAS and BRAF mutation status, the remaining studies analyzed one or the other. The sample size in our cohort study, n=766 cases with complete data and n=1250 including imputed cases, was therefore the largest to date studying associations between BMI and CRC subtypes defined by both KRAS and BRAF mutation status, and one of the larger studying subtypes by either KRAS or BRAF.
mutations. Even so, our sample size was limited by the requirement of both prediagnostic exposure data and molecular tumor data. Thus, even though no large differences in the association between metabolic factors and CRC subtypes were present, smaller differences may still exist.

Similar to the previous results for BMI, associations between metabolic factors and CRC risk did not differ by molecular subtypes defined by KRAS and BRAF mutation status in our paper. CRC subtypes defined by KRAS and BRAF mutations, both early mutations in colorectal carcinogenesis, are associated with distinct clinical and molecular characteristics and approximate major CRC developmental pathways. Therefore, the general associations between metabolic factors and overall CRC in our paper suggests that metabolic factors probably do not influence colorectal carcinogenesis through a specific CRC developmental pathway. This may explain the consistent and strong evidence for body fatness and metabolic factors as CRC risk factors.

To facilitate clinical use, the metabolic syndrome is typically defined as abnormal measures on a majority of individual components according to pre-defined cut-offs. Individual metabolic syndrome components (i.e., body fatness, insulin resistance, hyperlipidemia, and hypertension), are related but do not necessarily overlap. Therefore, the metabolic syndrome as a single pathophysiological entity has been questioned, and with it, the added value of a combined syndrome for predicting disease. Another issue when studying metabolic factors arises from using predefined cut-offs to define metabolic disturbances. Such cut-offs rarely represent true biological thresholds of effect, causing measurement error and misclassification. Associations between metabolic factors within the metabolic syndrome and CRC risk appear to be linear, with no added predictive value of a combined metabolic syndrome beyond the individual components. The use of metabolic factors to, for instance, classify high-risk individuals will therefore be more efficient by using continuous risk scores or CRC prediction models based on continuous measurements of several factors, instead of a combined syndrome based on predefined cut-offs.

Tumors with KRAS and BRAF mutations in our paper displayed expected clinical and molecular characteristics. This was despite the notably higher BRAF mutation frequency of 22% in the NSHDS compared to the typically observed 4-18%. Another Swedish population-based study of metastatic CRC observed a similar BRAF mutation frequency of 21%, likely caused by the unselected study population. The same explanation might hold true for studies in the population-based NSHDS, due to the high participation rate, low selection, and use of the essentially complete Swedish Cancer Registry for CRC diagnosis follow-up. Taken together, this supports the generalizability of our findings.
Paper II results

Main results
In this study of 1010 CRC cases and 1010 matched controls, higher circulating levels of the insulin resistance marker C-peptide and lower levels of the adipokine adiponectin were imprecisely associated with an increased risk of CRC (Model 1 ORs per 1 SD increase (95% CI): 1.11 (1.01, 1.23) and 0.91 (0.83, 1.00), respectively, Figure 12A). Adjusting for smoking, physical activity variables, and alcohol intake had little effect on the risk estimates (Model 2). Further adjusting for BMI attenuated risk estimates for both biomarkers (Model 3 ORs per 1 SD increase (95% CI) in: C-peptide 1.07 (0.96, 1.19) and adiponectin 0.93 (0.84, 1.03)). Insulin and leptin were not associated with CRC risk in any model. There were no indications of non-linearity for any biomarker (all \( P_{\text{nonlinearity}} > 0.2 \)). Associations for C-peptide were present in women even adjusting for BMI (OR\(_{men}\): 0.96 (0.82, 1.12) and OR\(_{women}\): 1.19 (1.01, 1.40), \( P_{\text{heterogeneity}} = 0.06 \), Figure 12B). There were no clear differences in analyses stratified by follow-up time between blood sampling and diagnosis or tumor sites (all \( P_{\text{heterogeneity}} > 0.1 \)).

![Figure 12](image-url)

**Figure 12. Odds ratios (ORs) for CRC by metabolic biomarkers.** ORs (95% CIs) for CRC risk per 1 SD increase in metabolic biomarkers in (A) all participants (\( n = 1010 \)) cases and (B) by sex. Model 1 was adjusted for the matching variables. Model 2 was additionally adjusted for smoking, recreational and occupational physical activity, and alcohol intake. Model 3 was additionally adjusted for BMI. Sex-specific estimates were adjusted for the covariates included in Model 3. Heterogeneity by sex was tested with Wald’s test. Mean (SD): log(insulin): -1.46 (0.81), log(C-peptide): 0.27 (0.50), log(adiponectin): 2.99 (0.57), log(leptin): 1.18 (1.16).

Next, we evaluated associations between the biomarkers and molecular subtypes of CRC (Figure 13). The OR per 1 SD increase in circulating adiponectin was lower for KRAS-mutated CRC compared to the other subtypes (0.82 for KRAS-mutated, 0.92 for BRAF-mutated, and 0.99 for double wild type, Figure 13A).
However, no subtype-specific estimate significantly differed from 1, and there was no significant difference in risk estimates ($P_{\text{heterogeneity}}=0.67$). All other biomarkers had similar risk estimates for CRC regardless of $KRAS/BRAF$ subtype ($P_{\text{heterogeneity}} = 0.74$ to 0.95, Figure 13A). No risk estimates differed by subtypes defined by MSI status ($P_{\text{heterogeneity}} = 0.33$ to 0.74, Figure 13B). Separate analysis by sex, although limited in power, displayed similar patterns for all biomarkers.

![Figure 13](image)

**Figure 13. Odds ratios (ORs) for CRC subtypes by metabolic biomarkers.** ORs (95% CIs) per 1 SD increase in metabolic biomarkers for CRC subtypes defined by (A) $KRAS/BRAF$ mutation status and (B) MSI status. Adjusted for matching variables, and smoking, recreational and occupational physical activity, alcohol intake, and BMI (Model 3). Numbers (n) within subtypes represent CRC cases with complete molecular data, data for the remaining cases were imputed.

**Interpretation**

Experimental studies implicate insulin resistance and altered adipokine production as primary mediators in body fatness-induced CRC development.\textsuperscript{256} While meta-analyses of circulating insulin and C-peptide or adiponectin have reported associations with CRC risk, results vary significantly among studies,\textsuperscript{101,104} and the results for leptin are inconclusive.\textsuperscript{105} The previous studies on these biomarkers and CRC risk have been of case-control or case-cohort design including less than 550 CRC cases, with the exception of studies on adiponectin and leptin in the European Prospective Investigation into Cancer and Nutrition (EPIC) cohort, including approximately 1200 cases.\textsuperscript{257,258} Our paper, including 1010 CRC cases from the population-based NSHDS, is therefore the largest individual study of circulating insulin and C-peptide, and one of the largest investigating adiponectin and leptin, and CRC risk. We observed imprecise associations between higher circulating levels of the C-peptide, and lower levels of adiponectin, and an increased CRC risk. Yet, the associations were not independent of BMI, with the possible exception of C-peptide in women.
Circulating insulin and leptin concentrations were not associated with CRC risk. Given the varying results on these biomarkers and CRC risk in previous studies, these inconclusive results are not implausible. Furthermore, the inconclusive results on adipokines were in line with the results in the larger EPIC studies.\textsuperscript{257,258}

In a case-control study nested within the American Nurses’ Health Study and Health Professional Follow-up Study (n=307 CRC cases), lower plasma adiponectin was associated with the risk of KRAS-mutated, and not KRAS wild type, CRC.\textsuperscript{180} In our paper including 1010 CRC cases, although risk estimates for adiponectin were slightly larger for KRAS-mutated CRC, there was no significant heterogeneity. The result is similar to results for BMI,\textsuperscript{170-174} and other metabolic factors (paper I). One established CRC risk factor has been linked to the KRAS-mutated subtype in a few MPE studies, namely physical activity.\textsuperscript{170,171} Physical activity, which does not necessarily reduce body fatness,\textsuperscript{259} has been demonstrated to suppress tumor growth by improving immune function in various tumor mouse models.\textsuperscript{260} As inflammation is a driver of KRAS-mutated carcinogenesis,\textsuperscript{261} an association between physical activity and a reduced risk of KRAS-mutated CRC is plausible. Adiponectin also has anti-inflammatory properties.\textsuperscript{261} Yet, associations between circulating adiponectin and inflammatory markers have been inconsistent.\textsuperscript{262-264} The discordant results between the American study and NSHDS may also be explained by study populations differences. Besides sample size, notable differences include the truly population-based nature of the NSHDS,\textsuperscript{49,50} follow-up time (median 12.3 years in this paper, 7.8 years in the American study), and the higher BRAF-mutation frequency in the NSHDS likely caused by the unselected case population (21% in the NSHDS vs 8% in the American study). The KRAS-mutation frequency also differed (24% in the NSHDS vs 44% in the American study), but the absolute number of KRAS-mutated cases was comparable (n=167 complete KRAS-mutated cases vs n=136 in the American study). To our knowledge, no previous study has investigated leptin in relation to molecular subtypes of CRC. Taken together, results to date provide no conclusive evidence for a role of low circulating adiponectin or high leptin in a specific carcinogenic pathway of CRC.

To our knowledge, this was the first MPE study of CRC and insulin resistance biomarkers. The insulin resistance marker C-peptide is produced by beta cells in equal amounts as insulin.\textsuperscript{265} Circulating C-peptide has a relatively long half-life,\textsuperscript{266} and is a better marker for long-term insulin production compared to circulating fasting insulin. This was also evident in our study, in which circulating C-peptide displayed a higher within-individual correlation compared to insulin in repeated samples taken approximately 10 years apart (ICC insulin: 0.60 and C-peptide: 0.74). This likely explains the inconclusive results for circulating insulin in CRC risk in this, and other studies.\textsuperscript{101} Similar to the results for adiponectin, we observed no clear association between C-peptide and the risk of
CRC molecular subtypes. Insulin may be linked to CRC through several tumor-promoting processes, including increased inflammation, cell proliferation (e.g., through increasing the growth factor IGF-1 or other hormones), or cell survival. Given this wide range of tumor-promoting mechanisms, if a link between insulin resistance and CRC risk exist, a more general association with overall CRC, unspecific to a certain developmental pathway, seems plausible.

As insulin resistance can be efficiently medically treated, the potential link with cancer provides an opportunity for chemoprevention. Randomized controlled trials of metformin, a common drug for reducing insulin resistance, glycogenesis, and insulin levels in overweight individuals, have not consistently reduced the risk of cancer in patients with diabetes, despite promising observational studies. However, a trial of non-diabetic patients with previous colorectal adenoma or polyps reported both reduced prevalence and number of polyps in participants treated with low dose metformin for one year. If this is replicated, metformin could be used for chemoprevention of CRC by reducing insulin resistance in overweight individuals. The results in our study suggest that such chemoprevention may be effective for all major subtypes of CRC.

**Paper III results**

**Main results**

OR and RD estimates for overall CRC risk by tertiles of circulating one-carbon metabolism components choline, betaine, DMG, sarcosine, and methionine in the 613 CRC cases and 1190 controls are presented in Figure 14. Higher levels of betaine were associated with a lower risk of CRC (OR highest vs lowest tertile: 0.76, 95%CI: 0.59, 0.99). The same was true for higher levels of methionine (OR highest vs lowest tertile: 0.72, 95%CI: 0.55, 0.97). Neither choline, DMG, nor sarcosine had a clear association with CRC risk. Adjusting for various potential confounders had little effect on the risk estimates. In the RD analyses, the highest tertiles of circulating betaine and methionine were associated with an estimated average absolute CRC risk reduction of 215 (95% CI: 9, 401) and 201 (95% CI: 26, 372) cases per 100 000, respectively, compared to the lowest tertiles.
Figure 14. Odds ratios (ORs) and risk differences (RD) for CRC by circulating levels of methyl group donors. ORs were only adjusted for matching variables since adjustment had no effect on estimates. RDs were adjusted for the matching variables, folate, vitamin B2 and vitamin B12, methionine, BMI, occupational physical activity, and smoking status.

In analyses of biologically plausible interactions, we observed a reduced colorectal cancer risk for high plasma methionine concentrations in subjects with low plasma folate concentrations (highest/lowest versus lowest/lowest methionine/folate tertiles OR: 0.39, 95% CI: 0.24, 0.64, \( P_{\text{interaction}} = 0.06 \)).

**Interpretation**

In this paper, we studied associations with CRC risk and circulating methyl group donors involved in one-carbon metabolism, namely metabolites in the choline oxidation pathway (choline, betaine, DMG, and sarcosine), and methionine. Choline is one of the main providers of methyl group input into one-carbon metabolism.\(^{107}\) The choline-product betaine offers a methyl group for the alternative conversion of homocysteine to methionine, which is the precursor of the universal methyl group donor SAM\(^{113}\)

Higher circulating levels of betaine and methionine were associated with a lower risk of CRC. Notably, this was consistent with two previous case-control studies in the EPIC (n=1367 cases) and Women's Health Initiative (n=835 cases) cohorts,\(^{155,156}\) despite a much lower folate status in our study population due to the absence of mandatory folic acid fortification of foods and low supplement use.\(^{129,269}\) A third study nested in the Alpha-Tocopherol, Beta-Carotene Cancer Prevention study (n=644 cases) found no link between circulating betaine and CRC risk, and instead an increased risk in participants with higher choline,\(^{158}\) but
this may have been caused by the selected study population of male smokers. In line with a protective role of methyl-group donors in colorectal carcinogenesis, a lower risk of colorectal adenomas has also been observed in individuals with higher circulating betaine or methionine, or SAM status.

By estimating absolute risk differences using incidence data from the cohort, we got an assessment of the potential public health impact of changed circulating levels of methyl donors, assuming causality. For instance, if all participants in the NSHDS had a betaine status corresponding to the highest tertile in our study (≥33.8 μmol/l) and this association was causal, around 200 CRC cases per 100,000 individuals would be prevented according to our estimates. However, the assumption of causality is strong, and although we adjusted for several important potential confounders, our results may still have been affected by residual confounding. For instance, we were unable to adjust for aspirin use and data on genetic confounders were limited.

Folate, as well as methionine-cycle components choline, betaine, and methionine, all contribute to the availability of methyl groups in one-carbon metabolism. While folate is important for both major outputs of one-carbon metabolism, nucleotide synthesis and methylation, choline, betaine, and methionine mainly have a role in methylation reactions. The importance of these components in methylation reactions may increase with lower folate status. Yet, betaine is a weak determinant of homocysteine status and does not determine folate status. The null association between homocysteine status and CRC in the NSHDS and other European study populations and positive association between folate status and CRC in the NSHDS is therefore in line with the findings in this paper. Taken together, our results support facilitated nucleotide synthesis, rather than DNA methylation, as the primary mechanism behind the plausible tumor-promoting effect of folate. The interaction between folate and methionine suggests that such a promoting effect might require an adequate methionine status.
**Paper IV results**

**Main results**
Independent associations between all studied one-carbon metabolism biomarkers, SNPs, and lifestyle or background variables and CRC subtypes estimated with Bayesian network learning are presented in Figure 15A. In the network, an edge between two variables represents an association independent of all other variables. We observed plausible relationships between variables. These include, for instance, associations between components within the methionine and folate cycles, the rs1801133 in MTHFR and circulating folate, and smoking and inflammatory marker neopterin. In this multivariate analysis, no variable had a significant independent association to CRC risk by KRAS and BRAF mutation status (i.e., edge confidence above the estimated significance threshold of 49%, Figure 15C). The variable with the strongest association to CRC subtype risk compared to other variables was a SNP in the CTH gene, rs1021737 (edge confidence: 34%). Although not evident in the estimated Bayesian network, median plasma cystathionine levels were approximately 7% higher per variant allele of the rs1021737 SNP (P = 0.0009).

ORs for CRC subtypes per 1 SD increase in biomarker levels and per allele of SNPs from conditional logistic regression models are displayed in Figure 15B. Biomarker estimates were adjusted for the matching variables, smoking, recreational and occupational physical activity, alcohol intake, BMI, and plasma neopterin, while the SNP estimates were only adjusted for the matching variables. Participants with the variant CTH rs1021737 genotype had an increased risk of BRAF-mutated CRC (OR per allele = 1.56, 95% CI = 1.07, 2.30), and decreased risk of KRAS-mutated CRC (OR per allele = 0.72, 95% CI = 0.50, 1.05), but with weak evidence for heterogeneity (P_{heterogeneity} = 0.01). This potential subtype-specific association was, however, not replicated in case-case analysis within the TCGA (P=0.85, Table 6). Two other variables displayed nominally significant heterogeneity in univariate models, vitamin B2 (P_{heterogeneity} = 0.02) and the rs3733890 SNP in the BHMT gene (P_{heterogeneity} = 0.03). Yet, associations strengths between these variables and CRC subtypes in the multivariate analysis were low (15.3 and 26.7%, respectively), suggesting that these associations were either confounded or mediated by other factors, or chance findings.
Figure 15. Multivariate and univariate analysis of one-carbon metabolism biomarkers and SNPs in relation to CRC risk by KRAS and BRAF mutation status. (A) Bayesian network estimated with the Hill-climbing algorithm showing independent associations between all included biomarkers, SNPs, and other variables. Edge thickness corresponds to association strength (measured as edge confidence, proportion of times an edge was present in 1000 bootstrap sample networks), and node size corresponds to number of connections. (B) Univariate ORs for CRC risk by KRAS and BRAF mutation status per 1 SD increase in biomarker levels or per allele of SNPs. (C) Multivariate association strength with CRC subtypes in the network, measured by edge confidence.
Table 6. Replication of the CTH rs1021737 association in The Cancer Genome Atlas (TCGA).

<table>
<thead>
<tr>
<th></th>
<th>NSHDS</th>
<th></th>
<th>TCGA</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(n=420)</td>
<td></td>
<td>(n=553)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>n</td>
<td>ROR (95% CI)(^\text{a})</td>
<td>Pb(^\text{b})</td>
<td>n</td>
</tr>
<tr>
<td>BRAF-mutated</td>
<td>115</td>
<td>1.33 (0.93, 1.89)</td>
<td>0.04</td>
<td>47</td>
</tr>
<tr>
<td>KRAS-mutated</td>
<td>124</td>
<td>0.76 (0.52, 1.11)</td>
<td>0.04</td>
<td>169</td>
</tr>
<tr>
<td>Both wild type</td>
<td>242</td>
<td>1 (ref)</td>
<td></td>
<td>337</td>
</tr>
</tbody>
</table>

ROR: Ratio of odds ratio, CI: Confidence interval.
\(^{a}\) RORs per allele calculated in multinomial logistic regression models.
\(^{b}\) Likelihood ratio test of heterogeneity, testing a common association across subtypes.

To adjust for potential bias from missing tumor data, we fitted inverse probability weighted conditional logistic regression, weighted by the estimated inverse probability of tumor KRAS and BRAF mutation data availability. This analysis yielded very similar risk estimates (correlation between log(ORs): 0.98).

In secondary analyses, no one-carbon metabolism variable clearly differed with respect to associations with CRC risk by either MSI or CIMP status, including folate (OR per 1 SD increase: 1.13 (0.89, 1.42) for CIMP-positive, and 1.07 (0.87, 1.31) for CIMP-negative, CRC risk, \(P_{\text{heterogeneity}}=0.73\)).

**Interpretation**

Balanced one-carbon metabolism, important for genome stability and repair,\(^{107}\) is believed to decrease the risk of CRC, while over-activation of the metabolic pathways may promote the proliferation and growth of precancerous or cancerous cells.\(^{107,108}\) The latter relationship is used by anti-folate chemotherapy agents to treat cancer. Results from prospective studies of one-carbon metabolism and CRC development in the NSHDS include an increased risk of overall CRC in individuals with higher circulating folate,\(^{129,131}\) lower levels of other B-vitamins (namely vitamins B2,\(^{146}\) B6,\(^{147}\) and B12\(^{151}\)), and lower betaine and methionine (paper III in this thesis). With the possible exception of folate, for which associations likely depends on the timing of exposure,\(^{140}\) the results in NSHDS are in line with most other studies. In this paper, the previously reported associations between one-carbon metabolism factors and CRC were similar regardless of tumor KRAS and BRAF mutation status, as well as MSI and CIMP status. To our knowledge, no previous studies of circulating components of one-carbon metabolism have investigated differences by tumor KRAS and BRAF status. Studies of dietary intakes of one-carbon metabolism macronutrients (including folate, other B-vitamins, and methyl donors) have not reported differences by major CRC subtypes defined by KRAS and BRAF mutation, MSI, or CIMP status.\(^{174,195-197,201,210-216}\)
A previous study in the NSHSD observed a suggestive stronger association between folate and the risk of CIMP-positive CRC.\textsuperscript{217} No such differences by CIMP-status was found in this larger study. Circulating folate status or folic acid supplementation has been associated with aberrant global DNA methylation patterns, but results vary, and the causality of the relationship is uncertain.\textsuperscript{121} Moreover, it is unknown whether folate-associated DNA methylation is spread out randomly across the genome, or if specific regions are affected.\textsuperscript{121} Consequently, even if a higher folate status causes an increase in genome-wide DNA methylation, this may not affect the core set of genes commonly hypermethylated in CIMP tumors.\textsuperscript{272} Thus, in combination with the inverse associations between the methylation-related markers in paper III of this thesis, the lack of a subtype-specific association between folate and CIMP-positive CRC risk provides further support for DNA synthesis, rather than methylation, as the primary mechanism behind the putative tumor-promoting effect of folate.

The \textit{CTH} gene encodes for the CTH enzyme (also abbreviated CSE), required in the vitamin B6-dependant conversion of cystathionine to cysteine in the transsulfuration pathway of one-carbon metabolism (Figure 5).\textsuperscript{114} Variant genotypes of a non-synonymous SNP in the gene, rs1021737, were associated with lower circulating cystathionine in this paper, and increased circulating homocysteine in another study.\textsuperscript{273} These results suggest a decreased enzyme activity with increased variant T alleles of the SNP. Neither the SNP nor circulating transsulfuration pathway-metabolites homocysteine, cystathionine, or cysteine was associated with the risk of overall CRC in the NSHSD.\textsuperscript{129,131,146} In this paper, while transsulfuration metabolites were unrelated to CRC subtypes, the rs1021737 SNP had opposite associations with CRC risk depending on \textit{KRAS} and \textit{BRAF} mutation status. This subtype-specific association was ultimately not replicated in the TCGA data set, providing no conclusive evidence for a role of rs1021737 in specific carcinogenic pathways of CRC. Despite the inconsistent result for the \textit{CTH} SNP in this paper, recent experimental studies support a role in CRC development for both CTH and the related CBS enzyme through production of the gasotransmitter hydrogen sulfide (H\textsubscript{2}S).\textsuperscript{274-277} Interestingly, if CTH rs1021737 variant carriers do have decreased enzyme activity, then a role for H\textsubscript{2}S in preventing \textit{KRAS}-mutated CRC and promoting \textit{BRAF}-mutated CRC would be consistent with our observations for the rs1021737 SNP in the NSHDS. Further studies of the role of CTH and the risk of CRC subtypes might, therefore, be warranted.
Conclusions

The overall conclusion of this thesis was that factors related to energy and one-carbon metabolism generally associate with CRC risk, and not specifically with subtypes defined by key molecular features (KRAS and BRAF mutation status, MSI, or CIMP status). Conclusions were:

**Paper I**
- Metabolic factors related to the metabolic syndrome, namely high BMI, total cholesterol, triglyceride levels, and blood pressure, were associated with an increased risk of CRC.
- No metabolic factor was associated with the risk of specific molecular CRC subtypes defined by KRAS and BRAF mutation status or MSI status.

**Paper II**
- Circulating levels of the insulin resistance markers insulin and C-peptide and the adipokines adiponectin and leptins were not associated with the risk of CRC or specific molecular subtypes of CRC defined by KRAS and BRAF mutation or MSI status.

**Paper III**
- Higher circulating levels of one-carbon metabolites primarily involved in the methionine cycle, methionine and betaine, were associated with a lower risk of CRC in a population with low folate status.
- The association between methionine and CRC risk may be modified by folate status; individuals with high methionine and low folate levels had the lowest risk of CRC.

**Paper IV**
- One-carbon metabolism biomarkers and SNPs were not associated with the risk of specific molecular subtypes of CRC defined by KRAS and BRAF mutation status, MSI status, or CIMP status.
- A non-synonymous SNP in the CTH gene, rs1021737, displayed diametrically opposite associations with CRC subtypes defined by KRAS and BRAF mutation status, but this difference was not replicated in an independent data set.
Implications

Our results suggest that body fatness and poor metabolic health, and potentially factors related to one-carbon metabolism, are universal risk mechanisms for CRC acting across multiple developmental pathways. This implies that, for these factors, CRC risk prediction to allow for individualized screening probably would not be improved by molecular subtype-specific models. Furthermore, the effect of public health strategies or interventions targeting body weight will likely be effective in reducing the incidence or recurrence of all major types of CRC, unlike, for instance, the potential subtype-specific effect of aspirin on CRC with PIK3 pathway mutations.
Future directions

Colorectal carcinogenesis is heterogeneous, resulting in clinically and molecularly distinct tumor subtypes. MPE studies on whether CRC etiology differs by molecular subtypes have focused on subtypes defined by key molecular characteristics such as KRAS and BRAF mutation, MSI, or CIMP status.165,166 While subtypes defined by these molecular markers may represent major CRC pathways,45,48 they are unlikely to capture all between-tumor heterogeneity in CRC.56 Etiological studies of more finely defined molecular subtypes, such as the gene expression-based CMS subtypes,56 is therefore an interesting field of future research. As more detailed subtyping leads to higher economic costs and issues of statistical power, such studies will require collaborative efforts involving many study cohorts. Large consortia combining many study cohorts to study CRC epidemiology have already been initiated, notably the Genetics and Epidemiology of Colorectal Cancer Consortium (GECCO).278 The GECCO, in which the NSHDS is participating, now consists of more than 30,000 CRC cases from over 30 study cohorts. Although the main purpose of GECCO is to discover new genetic variants related to CRC development, the process of pooling and harmonizing tumor data for large MPE studies of CRC in GECCO has been initiated. In the future, MPE studies of CMS or CRC subtypes defined by other molecular characteristics in GECCO are possible. In this setting, as at least some CRC risk factors appear subtype-specific, it is also of interest to investigate whether the overall accuracy of CRC prediction models can be improved by molecular subtype-specific models.

Associations from observational studies do not imply causality. A major reason for this is confounding bias, i.e., spurious relationships caused by the effect of a confounding variable on both exposure and outcome.279 Randomized controlled trials avoid confounding by randomizing exposure, and thus balancing both known and unknown confounders between the exposed and unexposed. However, randomized controlled trials are not feasible in most etiological studies of cancer for ethical reasons, and the economic costs and time such trials would require. An alternative approach to infer causal relationships in observational studies is Mendelian randomization. Mendelian randomization studies use genetic instruments for exposures to estimate relationships with disease less prone to confounding bias.78 For this to be successful, robust and specific genetic instruments are required, which have been successfully discovered through GWAS for many cancer risk factors.78 Using Mendelian randomization to evaluate risk factors specific for molecular subtypes of CRC can provide important insights into the causality of observed etiological heterogeneities, and is an important field of future research. Yet, as effect sizes of common genetic variants related to exposures are generally small, such future research will require large sample sizes, for instance by utilizing consortia like GECCO.
Overweight and obesity are growing global health problems. Among other adverse health outcomes, increased body fatness will lead to increased CRC incidence rates worldwide. Although CRC treatments are improving, an increased CRC incidence will cause a considerable amount of individual suffering. Furthermore, treatment costs will be substantial if no effort for prevention is made. A future challenge for research lies in finding successful and cost-effective methods for weight management.
Acknowledgments

I would like to express my warmest thanks to all participants in the NSHDS cohorts. Your time and involvement is instrumental for medical research.

The work was supported financially by the Swedish Cancer Foundation and the Cancer Research Foundation in Northern Sweden. Thanks to all who donated to cancer research.

To my main supervisor, Associate Professor Bethany van Guelpen: I cannot express how grateful I am for this unexpected journey into the world of medical science. It has been a privilege to work with and learn from such a skilled and enthusiastic scientist and clinician. Thank you for encouraging me to try new things, for always making time for me in your hectic schedule, and for making me believe in my abilities. If I someday become half the scientist you are, I will be happy.

To my co-supervisor, Professor Richard Palmqvist: I admire your knowledge, your ability to twist and turn ideas, and your overall positive attitude and witty ways. Thank you for all you have taught me and for creating a research atmosphere where everyone feels appreciated.

To my co-supervisor, Associate Professor Jenny Häggström: It has been absolutely instrumental to have access to your amazing statistics brain during all this. Thank you for always being there when I need it and for being a calm voice of reason when I am worrying about stuff.

Special thanks to Björn Gylling, colleague, semi-supervisor, and friend. You are a double-edged sword of kindness and humor on one side, and intellect on the other. Thanks for showing me the ropes at NUS and for teaching me biological concepts without making me feel like a dummy variable.

“As iron sharpens iron, so one person sharpens another.” (Proverbs 27:17, the Bible). I do not believe in God, but I do believe in the power of the Colorectal Cancer Network in Umeå. To all former and current members, Sofia, Agnes, Ida, Mona, Anna, Carl, Pär, Roger, Kerstin, Åsa, Xingru, Martin, Maria, Vincy, Jörgen, Åke, Malin, Sophia, Stina, Johannes, Anneli, and others: Thank you for making me feel at home in a new place and for all you have taught me, for all help, and all good times throughout the years.

To the staff at the Departments of Biobank Research and Nutritional Research, Åsa, Robert, Veronica, Anette, Gerd, and the all others in the biobank crew.
Your work makes it all possible. Thank you for putting up with me in the corridors and lunch room all these years. You guys are great!

To Professor **Göran Hallmans**, your tireless enthusiasm for science is impressive. All that you have done, from founding the biobank to key scientific projects in the NSHDS, is invaluable. A day when Göran pops his head into your office with some new idea is a good day. Thank you!

To Professor **Ingegerd Johansson**, for providing the funding for the plasma analyses of one-carbon metabolism and setting up the diet data base. Your work has been instrumental for this thesis. Thank you!

Thanks to the colleagues at the Department of Radiation Sciences. Special thanks to **Maria Wing**, probably the best administrator in the world.

To Associate Professor **Jörn Schneede** and the Norwegians, Professor **Per-Magne Ueland**, **Arve Ulvik**, **Klaus Meyer**, **Øivind Midttun**, and others: It has been a privilege and a pleasure working with and learning from you all. I hope to do it again soon.

Thanks to all co-authors on papers included and not included in the thesis for your important contributions and for fruitful collaboration. It has been incredibly educational and fun working with you all.

To my roommates at the “Shungin Institute”, **Christel Häggström** and **Dmitry Shungin**: Thank you for first class scientific and non-scientific discussions.

To my friends, **Sebastian, Adam**, Kristianstad friends, members of the PiP foundation (Shine on!), members of the New Year’s group, friends at IKSU Kampsport, SINGS buddies, 6M social corner buddies, members of the Umeå pancake group **Carl Larsson, Bore, Wilmer**, and **Gabriel**, and others: You are all irreplaceable. Thank you for sharing your time and energy with me.

Tack till **Maggan** för att du alltid får mig på gott humör.

Till min familj, Mamma **May**, Pappa **Jan** och mina syskon **Anton** och **Lina**: Tack för all villkorslös kärlek och för att ni alltid finns där för mig. Ni betyder allt. Till mormor **Anna**, resten av Myte/Hindorf-släkten och till minnet av morfar **Elvir**, farmor **May** och farfar **Rune**: Tack för all kärlek, trygghet och glädje.

Slutligen, tack till **Clara**. Ingenting hade varit möjligt utan ditt oändliga stöd, din kärlek, och den lycka du ger mig varje dag. Din för evigt.
References


Lin KJ, Cheung WY, Lai JY, Giovannucci EL. The effect of estrogen vs. combined estrogen-progestogen therapy on the risk of colorectal cancer. *Int J Cancer* 2012;130: 419-30.


49 Rosty C, Young JP, Walsh MD et al. Colorectal carcinomas with KRAS mutation are associated with distinctive morphological and molecular features. Mod Pathol 2013;26: 825-34.


Bray MS, Loos RJ, McCaffery JM et al. NIH working group report-using genomic information to guide weight management: From universal to precision treatment. *Obesity (Silver Spring)* 2016;24:14-22.


Hanyuda A, Ogino S, Qian ZR et al. Body mass index and risk of colorectal cancer according to tumor lymphocytic infiltrate. Int J Cancer 2016;139: 854-68.


Weijenberg MP, Aardening PW, de Kok TM, de Goeij AF, van den Brandt PA. Cigarette smoking and KRAS oncogene mutations in sporadic colorectal cancer: results from the Netherlands Cohort Study. *Mutat Res* 2008;652: 54-64.


Jayasekara H, MacInnis RJ, Williamson EJ et al. Lifetime alcohol intake is associated with an increased risk of KRAS plus and BRAF-/KRAS- but not BRAF plus colorectal cancer. *Int J Cancer* 2017;140: 1485-93.


Midttun O, Townsend MK, Nygard O et al. Most blood biomarkers related to vitamin status, one-carbon metabolism, and the kynurenine pathway show adequate preanalytical stability and within-person reproducibility to allow assessment of exposure or nutritional status in healthy women and cardiovascular patients. *J Nutr* 2014;144: 784-90.


Eussen SJ, Nilsen RM, Midttun O et al. North-south gradients in plasma concentrations of B-vitamins and other components of one-carbon metabolism

Alberti KG, Eckel RH, Grundy SM *et al.* Harmonizing the metabolic syndrome: a joint interim statement of the International Diabetes Federation Task Force on Epidemiology and Prevention; National Heart, Lung, and Blood Institute; American Heart Association; World Heart Federation; International Atherosclerosis Society; and International Association for the Study of Obesity. *Circulation* 2009;**120**: 1640-5.


Bardou M, Barkun AN, Martel M. Obesity and colorectal cancer. *Gut* 2013;**62**: 933-47.


Niculescu MD, Zeisel SH. Diet, methyl donors and DNA methylation: interactions between dietary folate, methionine and choline. *J Nutr* 2002;132: 2333S-5S.


