From the Department of Molecular Medicine and Surgery
Karolinska Institutet, Stockholm, Sweden

GENETIC ASSOCIATION- AND LINKAGE STUDIES IN COLORECTAL CANCER

Susanna von Holst

Karolinska
Institutet

Stockholm 2014

All previously published papers were reproduced with permission from the publisher.
Published by Karolinska Institutet.
Printed by Universitetsservice US-AB
© Susanna von Holst, 2014
ISBN 978-91-7549-677-1

Genetic association- and linkage studies in colorectal cancer

THESIS FOR DOCTORAL DEGREE (Ph.D.)

By

Susanna von Holst

Principal Supervisor:
Annika Lindblom, Professor
Karolinska Institutet
Department of Molecular Medicine and Surgery

Co-supervisor(s):
Simone Picelli
Karolinska Institutet
Ludwig Institute for Cancer Research

Vinaykumar Kontham
Karolinska Institutet
Department of Molecular Medicine and Surgery

Johanna Lundin
Karolinska Institutet
Department of Women´s and Children´s Health

Opponent:
Claes Wadelius, Professor
Uppsala University
Department of Immunology, Genetics and Pathology

Examination Board:
Ingrid Kockum, Assoc professor
Karolinska Institutet
Department of Clinical Neuroscience

Per Eriksson, Professor
Karolinska Institutet
Department of Medicine

Tobias Sjöblom, Assoc professor
Uppsala University
Department of Immunology, Genetics and Pathology

ABSTRACT

Colorectal cancer (CRC) is the third most common cancer type in the Western world. Over one million patients are diagnosed worldwide yearly. A family history of CRC is a major risk factor for CRC. The total genetic contribution to disease development is estimated to be 35%. High-risk syndromes caused by known genes such as familial adenomatous polyposis (FAP) and Lynch Syndrome (LS) explain less than 5% of that number. Recently, several genome-wide association studies (GWAS) have independently found numerous loci at which common single-nucleotide polymorphisms (SNPs) modestly influence the risk of developing colorectal cancer. In total, germline mutations in known genes and moderate- and low risk variants are today suggested to explain 10-15% of the total genetic burden. Hence, predisposed genetic factor are still left to be found.

The aim of **paper I** was to investigate if 11 published loci reported to be associated with an increased or decreased risk of colorectal cancer could be confirmed in a Swedish-based cohort. The cohort was composed of 1786 cases and 1749 controls that were genotyped and analyzed statistically. Genotype–phenotype analysis, for all 11 SNPs and sex, age of onset, family history of CRC and tumor location, was performed. Of 11 loci, 5 showed statistically significant odds ratios similar to previously published findings. Most of the remaining loci showed similar OR to previous publications. Four statistically significant genotype–phenotype associations were reported.

The aim of **paper II** was to further study these 11 SNPs and their possible correlation with morphological features in tumors. We analyzed 15 histological features in 1572 CRC cases. Five SNPs showed statistically significant associations with morphological parameters. The parameters were poor differentiation, mucin production, decreased frequency of Crohn-like peritumoral reaction and desmoplastic response.

The aim of **paper III** was to identify new CRC loci using a genome wide linkage analysis. We used 121 non-FAP/LS colorectal cancer families and genotyped 600 subjects using SNP array chips. No statistically significant result was found. However, suggestive linkage was found in the parametric analysis. This was observed in a recessive model for high-risk families, at locus 9q31.1 (HLOD=2.2) and for moderate-risk families, at locus Xp22.33 (LOD=2.2 and HLOD=2.5). Using families with early-onset, recessive analysis suggested one locus on 4p16.3 (LOD=2.2) and one on 17p13.2 (LOD/HLOD=2.0). Our linkage study adds support for the previously suggested region on chromosome 9 and suggests three additional loci to be involved in colorectal cancer risk.

It is debated whether CRC is a single entity or two different entities, colon- and rectal cancer. Studies have recognized their molecular differences. The aim of **paper IV** was to identify novel colon- and rectal loci. We performed a genome wide linkage analysis using 32 colon- and 56 rectal cancer families. No LOD or HLOD score above three was observed. However, results close to three could be demonstrated. A maximum HLOD= 2.49 at locus 6p21.1-p12.1 and HLOD= 2.55 at locus 18p11.2 was observed for the colon- and rectal cancer families respectively. Exome sequencing was done, on colon and rectal patients, in these regions of interest. We report 25 variants mutated in family members on chromosome 6 and 27 variants on chromosome 18. Further studies are ongoing to elucidate the importance of these variants.

Keywords: Colorectal cancer, SNP (single nucleotide polymorphism), association studies, linkage analysis, exome sequencing.

LIST OF SCIENTIFIC PAPERS

I. **von Holst S**, Picelli S, Edler D, Lenander C, Dalén J, Hjern F, Lundqvist N, Lindforss U, Påhlman L, Smedh K, Törnqvist A, Holm J, Janson M, Andersson M, Ekelund S, Olsson L, Ghazi S, Papadogiannakis N, Tenesa A, Farrington SM, Campbell H, Dunlop MG, Lindblom A.
Association studies on 11 published colorectal cancer risk loci.
Br J Cancer. 2010 Aug 10;103(4):575-80. doi: 10.1038/sj.bjc.6605774. Epub 2010 Jul 20.

II. Ghazi S, **von Holst S**, Picelli S, Lindforss U, Tenesa A, Farrington SM, Campbell H, Dunlop MG, Papadogiannakis N, Lindblom A; Low-Risk Colorectal Cancer Study Group.
Colorectal cancer susceptibility loci in a population-based study: Associations with morphological parameters.
Am J Pathol. 2010 Dec;177(6):2688-93. doi: 10.2353/ajpath.2010.100298.

III. Kontham V*, **von Holst S***, Lindblom A.
Linkage analysis in familial non-Lynch syndrome colorectal cancer families from Sweden.
PLoS One. 2013 Dec 11;8(12):e83936. doi: 10.1371/journal.pone.0083936. eCollection 2013.
**equal contribution*

IV. **von Holst S**, Kontham V, Thutkawkorapin J, Nilsson D and Lindblom A.
Linkage analysis in familial colon- and rectal cancer
Manuscript

Related publications not included in this thesis:

Dunlop MG, Tenesa A, Farrington SM, Ballereau S, Brewster DH, Koessler T, Pharoah P, Schafmayer C, Hampe J, Völzke H, Chang-Claude J, Hoffmeister M, Brenner H, **von Holst S**, Picelli S, Lindblom A, Jenkins MA, Hopper JL, Casey G, Duggan D, Newcomb PA, Abulí A, Bessa X, Ruiz-Ponte C, Castellví-Bel S, Niittymäki I, Tuupanen S, Karhu A, Aaltonen L, Zanke B, Hudson T, Gallinger S, Barclay E, Martin L, Gorman M, Carvajal-Carmona L, Walther A, Kerr D, Lubbe S, Broderick P, Chandler I, Pittman A, Penegar S, Campbell H, Tomlinson I, Houlston RS. **Cumulative impact of common genetic variants and other risk factors on colorectal cancer risk in 42,103 individuals.** Gut. 2013 Jun;62(6):871-81. doi: 10.1136/gutjnl-2011-300537. Epub 2012 Apr 5.

Stacey SN, Manolescu A, Sulem P, Rafnar T, Gudmundsson J, Gudjonsson SA, Masson G, Jakobsdottir M, Thorlacius S, Helgason A, Aben KK, Strobbe LJ, Albers-Akkers MT, Swinkels DW, Henderson BE, Kolonel LN, Le Marchand L, Millastre E, Andres R, Godino J, Garcia-Prats MD, Polo E, Tres A, Mouy M, Saemundsdottir J, Backman VM, Gudmundsson L, Kristjansson K, Bergthorsson JT, Kostic J, Frigge ML, Geller F, Gudbjartsson D, Sigurdsson H, Jonsdottir T, Hrafnkelsson J, Johannsson J, Sveinsson T, Myrdal G, Grimsson HN, Jonsson T, **von Holst S**, Werelius B, Margolin S, Lindblom A, Mayordomo JI, Haiman CA, Kiemeney LA, Johannsson OT, Gulcher JR, Thorsteinsdottir U, Kong A, Stefansson K. **Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer**. Nat Genet. 2007 Jul;39(7):865-9. Epub 2007 May 27.

LIST OF ABBREVIATIONS

AFAP	Attenuated FAP
APC	Adenomatous polyposis coli gene
APCDD1	Adenomatosis polyposis down-regulated 1
BAX	BCL2-associated protein X
BER	Base excision repair
BMP	Bone morphogenetic protein
BRAF	v-Raf murine sarcoma viral oncogene homolog B
BMPR1A	Bone morphogenetic protein receptor, type IA
CCND2	Cyclin D2
CIMP	CpG island methylator phenotype
CIN	Chromosomal instability pathway
cM	centimorgan
CLR	Crohn-like peritumoral lymphocytic reaction
CRAC1	Colorectal Adenoma and Carcinoma 1
CRC	Colorectal cancer
CTAGE1	Cutaneous T-cell lymphoma-associated antigen 1
CUL7/CUL9	Cullin 7/ Cullin 9
DCC	DCC netrin 1 receptor
DNA	Deoxyribonucleic acid
EIF3H	Eukaryotic translation initiation factor 3, subunit H
FAP	Familial adenomatous polyposis
FCCTX	Familial colorectal cancer type X
GALNT12	Polypeptide N-Acetylgalactosaminyltransferase 12
GREM1	Gremlin 1
GWAS	Genome wide association studies
HCRC	Hereditary colorectal cancer
HNPCC	Hereditary non polyposis colorectal cancer
JPS	Juvenile polyposis syndrome
KRAS	Kirsten rat sarcoma viral oncogene homolog
LAMC1	Laminin, gamma 1
LD	Linkage disequilibrium
LOD	Logarithm of odds
LOH	Loss of heterozygosity
LS	Lynch Syndrome
LRFN2	Leucine rich repeat and fibronectin type III domain containing 2
MAF	Minor allele frequency
MAP	MUTYH-associated polyposis
Mb	Megabase
MSH3	Mut S homolog 3
MSH2	Mut S homolog 2
MLH1	Mut L homolog 1
MSH6	Mut S homolog 6
MSI	Microsatellite instability
MMR	Mismatch repair
MUTYH	Mut Y homolog
MYC	V-myc avian myelocytomatosis viral oncogene homolog
NGS	Next generation sequencing
NPL	Non-parametric linkage
OR	Odds ratio
PCR	Polymerase chain reaction
PJS	Peutz-Jeghers syndrome
PMS2	Postmeiotic segregation 2
POLD1	Polymerase (DNA directed), delta 1, catalytic subunit
POLE	Polymerase epsilon catalytic subunit
POTEC	POTE ankyrin domain family, member C
PTEN	Phosphatase and tensin homolog
PTK7	Protein tyrosine kinase 7
ROCK1	Rho-associated, coiled-coil containing protein kinase 1
SMAD	Mothers against decapentaplegic homolog
SNP	Single nucleotide polymorphism
SNV	Single nucleotide variant
STK11	Serine/threonine kinase
TGFβ	Transforming growth factor β
TGFβR1	Tumor growth factor β receptor I
TGFβR2	Tumor growth factor β receptor II
TCF4	Transcription factor 4
TCR	Two close relatives
TILs	Tumor infiltrating lymphocytes
TP53	Tumor protein 53
WES	Whole exome-sequencing

CONTENTS

1 Introduction 5
1.1 Colorectal cancer 5
1.1.1 Incidence....5
1.1.2 Risk factors....5
1.1.3 Pathological features in tumors....6
1.1.4 Age, gender and location 7
1.2 Pathways to colorectal cancer 8
1.2.1 Chromosomal instability (CIN) pathway 9
1.2.2 Microsatellite instability (MSI) pathway....10
1.2.3 Epigenetic alterations 10
1.3 Colorectal cancer predisposition....11
1.3.1 Colorectal cancer syndromes 11
1.3.2 Low-risk genetic variants....15
1.4 Evidence for additional colorectal cancer susceptibility genes....18
1.4.1 Non- FAP/LS familial cancer 18
1.4.2 Colorectal cancer susceptibility loci 19
1.5 Strategies for finding colorectal cancer predisposing genes 20
1.5.1 Linkage analysis....20
1.5.2 Association studies....21
1.5.3 Next generation sequencing....22
1.5.4 Additional approaches....23
2 Aims of the thesis 24
3 Materials and methods 25
3.1 Patients....25
3.2 Methods 25
3.2.1 Genotyping 25
3.2.2 Statistical analysis 26
3.2.3 Linkage analysis....27
3.2.4 Exome sequencing 27
4 Results and discussion....29
4.1 Paper I....29
4.2 Paper II....30
4.3 Paper III 32
4.4 Paper IV 33
5 Conclusion and future perspectives 35
6 Sammanfattning på svenska....37
7 Acknowledgements....39
8 References 41
Papers I-IV

1 INTRODUCTION

1.1 COLORECTAL CANCER

1.1.1 Incidence

Colorectal cancer (CRC) is the third most common cancer type in the western world and more than one million new cases are diagnosed yearly. The annual mortality rate is almost 400,000 and CRC is ranked as fourth among cancer related deaths. It affects men and women equally, and represents 9.4% and 10.1% of all cancers respectively (Haggar & Boushey, 2009). In Sweden, over 6,000 cases are diagnosed with colon- or rectal cancer every year and the mortality rate is over 2,600 yearly. CRC is counted as the third most common cancer type among men and women (Johansson, 2013; Socialstyrelsen, 2012).

The highest incidence rates are observed in Australia, New Zealand, Canada, United States and Western Europe whilst the lowest are seen in parts of Asia, Africa and South America. Interestingly, when moving from a low-risk to a high-risk country, it has been proposed that the incidence number for off-springs tend to increase to that of the new host county. This demonstrate the importance of lifestyle and environmental factors among different countries (Haggar & Boushey, 2009).

1.1.2 Risk factors

CRC can be distinguished between familial and sporadic forms, where familial represents the minority. However, having a family history of the disease is one major risk factor for developing the disease. The risk increases more than twofold when having a first degree relative with CRC and more than four times with more than one relative with CRC (Johns & Houlston, 2001). A large twin study comparing the genetic and environmental contribution to cancer development showed that 35% are due to hereditable factors (Lichtenstein et al., 2000).

A personal history of Inflammatory Bowel Disease (IBD); ulcerative colitis (inflammation of the mucosa of colon and rectum) or Crohn disease (inflammation of the bowel wall) increases the risk of developing CRC. It has been proposed that the increased relative risk of colorectal cancer in IBD patients would be 4 to 20-fold (Haggar & Boushey, 2009).

Smoking is highly correlated to lung cancer but it is also estimated that up to 12 % of CRC cases are attributable to smoking (Haggar & Boushey, 2009). Also, the role of alcohol consumption and CRC are debated, and pooled data shows an increased risk in developing CRC for individuals drinking more than 45g alcohol/day (Cho et al., 2004).

Studies have shown that dietary habits affect the risks of developing CRC. A high consumption of red meat and fat, mainly animal fat, increases the risk of developing CRC whereas higher intake of dietary fibre, vegetables and fruits lower the risks. Obesity has shown association to CRC whereas physical activity have a protective effect against CRC (Haggar & Boushey, 2009).

Additionally, it has been proposed that aspirin and NSAIDs (non-steroidal anti-inflammatory drugs) have a protective effect on CRC. Even a minimum aspirin intake of 75mg /day reduces the risk of disease development (Din et al., 2010). A recent study concluded that a minimum of five years use of prophylactic aspirin, 75-325 mg/day, would be favorable and reduce the cancer risk (Cuzick et al., 2014). Estrogen intake is known to have a protective effect against microsatellite instability (MSI) (described 1.2.2), whereas lack of estrogen, in older women, escalates the risk of instability. This pinpoints the protective effect of hormone replace therapy (Slattery et al., 2001).

1.1.3 Pathological features in tumors

Each tumor is unique when it comes to histopathological parameters. Pathologists use standardized protocols when categorizing tumors (paper II).

TNM classification is a system to divide the tumor into a level/stage to evaluate the spread. T is referring to the spread of the primary tumor, N to the spread in lymph nodes and M indicates the occurrence of distant metastases.

Medullary carcinoma is a rare CRC variant characterized by sheets of malignant cells with vesicular nuclei, prominent nucleoli and abundant pink cytoplasm exhibiting prominent infiltration by intraepithelial lymphocytes. It has been associated with MSI-H (Jessurun, Romero-Guadarrama, & Manivel, 1999).

Tumor grading is used to state the grade of differentiation. A well differentiated tumor is described as low-grade while a poorly differentiated tumor is high graded. A high graded tumor is an adverse prognostic factor (Compton, 2003).

A tumor with *mucin production*, i.e. more than 50% of pools of extracellular mucin containing malignant epithelium is referred to as a mucinous adenocarcinoma. Mucinous differentiation has been reported with an association to mismatch repair protein deficiency (Langner et al., 2012).

Tumor infiltrating lymphocytes (TILs) are T-lymphocytes (intraepithelial and primarily cytotoxic) that exist in tumor tissue. The existence of TILs is associated with a better prognosis due to their implication in killing the tumor cells (Zhang et al., 2003).

Crohn-like peritumoral lymphocytic reaction (CLR) represents an immune response towards the tumor and is defined as the occurrence of nodular aggregates of B-lymphocytes (mostly) deep to the advancing tumor front, in muscularis propria or pericolic adipose tissue. CLR is associated with improved survival, intense TILs at the tumor edge, lower incidence of nodal metastasis and right sided carcinomas (Graham & Appelman, 1990).

Desmoplasia is often seen in CRC and is characterized by an intense fibrous reaction around infiltrating tumor tissue. It is demonstrated that desmoplasia may inhibit cancer invasiveness by building a barrier against tumor diffusion and thus act as a protecting factor (Caporale et al., 2005).

Necrosis is when cells are injured and follows by premature cell death. Tumor necrosis is usually seen in CRC.

Perineural invasion describes tumor cells that infiltrates underneath the perineurium at the margin of the tumor or deep to it. It has shown independently to be associated with poor CRC prognosis (Compton, 2003).

Vascular invasion is an independent prognostic factor and has two components; blood vessel invasion and lymphatic vessel invasion. Both are shown to be associated with a decreased survival in CRC (Minsky & Mies, 1989).

Budding is defined as presence of individual cells or small clusters of tumor cells at the invading front of the tumor. This is an independent prognostic factor but associated with lymph node metastasis and also with more advanced TNM stage tumors (Grizzi, Celesti, Basso, & Laghi, 2012).

Tumor margin configuration has been described to have prognostic significance that is independent of stage. A tumor with a circumscribed smooth-pushing pattern is a better prognostic factor as compared to an infiltrative irregular pattern of growth (Compton, 2003; Jass, Love, & Northover, 1987).

1.1.4 Age, gender and location

Studies have shown that cancer appears later in life in women compared to men and that the survival rate is higher for women under 50 compared to men. This figure is reversed for older women, most likely due to the decreased level of estrogen later in life (Koo & Leong, 2010). The CRC risk increases with age and for a non-familial case the average age at diagnosis is older than 65 years (Weiss et al., 2011). Most CRC cases, 90%, are diagnosed after the age of 50. However, familial cases are present earlier in life (Haggar & Boushey, 2009). Some studies argue that younger cases (<50) generally are present with a later cancer stage and poorly differentiated tumor at diagnosis. This might be due to more aggressive tumors in young patients or the lack of screening or awareness of having a predisposed increased risk for CRC in this group (Fairley et al., 2006). A Swedish study by Ghazi *et al.* concluded that younger cases have higher degree of perineural invasion and infiltrative tumor margin as well as later stage tumors at diagnosis (Ghazi et al., 2012).

It has been demonstrated that gender differences in tumor location occurs, where female seem to have more proximal colon cancers and males more distal colon- and rectal cancers (Koo & Leong, 2010). It has also been described that female patients more commonly has tumors with TILs (tumor infiltrating lymphocytes) and tumors of medullary type whereas males has a higher frequency of infiltrative tumor margin (Ghazi et al., 2012).

So far, no one has managed to biologically answer the question about the differences on cancer location. Bulfill *et al.* published a theory about the molecular mechanism and site where he proposed that proximal tumors seem to be equal to hereditary non-polyposis

colorectal cancer (HNPCC) when it comes to the molecular mechanism. On the contrary, distal tumors are more like polyposis-associated colorectal cancer syndromes (Bufill, 1990). Another study presented survival differences for adenocarcinoma patients. The median survival was 78 and 89 months for right- and left-sided cancers respectively. Correcting for confounding variables, the mortality risk was still higher for patients with right- compared to left sided tumors (Meguid, Slidell, Wolfgang, Chang, & Ahuja, 2008). In Swedish cases, right sided tumors were larger, later stage, often more poorly differentiated, had more mucin production and more peritumoral lymphocytic infiltrate and higher level of TILs compared to left colon and rectal tumors (Ghazi et al., 2012).

The following sections will discuss genetic components underlying CRC.

1.2 PATHWAYS TO COLORECTAL CANCER

Colorectal cancer (CRC) is characterized by uncontrolled cell growth in colon, rectum or the appendix. The tumor cells have a capacity to invade neighboring tissue and metastasize. The colorectal cancer development process is one of the most well characterized pathways and the availability of biopsies from the different histopathological steps has led to the molecular pathogenesis of CRC. The development from normal colonic epithelium to adenoma and further to carcinoma is due to progressive accumulations of genetic and epigenetic alterations (Fearon & Vogelstein, 1990; Kinzler & Vogelstein, 1996). Three types of genes are involved in the carcinogenesis; tumor suppressor genes, oncogenes and DNA repair genes.

Tumor suppressor genes protect cells from progressing to cancer. When mutated, the gene loses the function of controlling the cell cycle, apoptosis, growth factors and signaling pathways. Three main tumor suppressor genes occurs in CRC development; *APC*, *DCC* and *P53* (*TP53*) (Fearon & Vogelstein, 1990). Mutant tumor suppressor genes are usually recessive, and follow the “two hit hypothesis”, which means that both alleles coding for a protein must be damaged for an affect to occur (Knudson, 1971). Usually, mutation in only one allele does not affect protein production. However, exceptions occur; some mutations in the *P53* gene only require “one hit” to lose its function (Baker, Markowitz, Fearon, Willson, & Vogelstein, 1990).

Proto-oncogenes/oncogenes have a dominant effect; only one allele has to be mutated for the gene to be over activated and thus trigger the cancer development. *KRAS* is the primary oncogene in CRC tumorigenesis (Fearon & Vogelstein, 1990).

Errors, such as insertions or deletions, which evolved during DNA replication, are repaired by *DNA repair genes*. Colorectal cancer development is strongly associated with mutations in the mismatch repair (MMR) pathway. Mutations in MMR genes, *MLH1*, *MSH2*, *MSH6* or *PMS2* will lead to accumulation of mutations in microsatellites and further to tumor development (Shibata, Peinado, Ionov, Malkhosyan, & Perucho, 1994).

1.2.1 Chromosomal instability (CIN) pathway

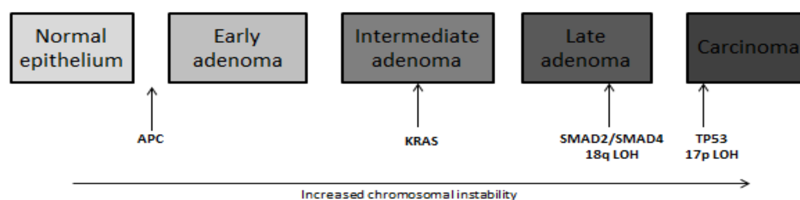
In 1990, Fearon and Vogelstein proposed that colorectal cancer is a result from mutational activation of oncogenes together with an inactivation of tumor suppressor genes (Fearon & Vogelstein, 1990). This multistep model (figure 1), adenoma-carcinoma sequence, suggests how the genetic alterations relate in order during tumor progression. The mutation accumulation is characterized by events of chromosomal aberrations in the chromosomal instability (CIN) pathway.

The first genetic event in the adenoma-carcinoma sequence is a mutation in the tumor suppressor gene *APC*, where a second inactivating mutation of the other allele leads to activation of the WNT-signaling pathway (Bodmer, Bailey, Bodmer, Bussey, Ellis, Gorman, Lucibello, Murday, Rider, Scambler, Sheer, et al., 1987). Mutations in *APC* are seen in up to 80% of all colorectal adenomas and carcinomas (Miyoshi et al., 1992). Germline mutations in *APC* are found in patients with familial adenomatous polyposis (FAP), an autosomal dominant condition (Groden et al., 1991), which will be discussed in chapter 1.3.1.1.

An activating mutation in the *KRAS* oncogene occurs early in this multistep process. *KRAS* is involved in normal tissue signaling (Kranenburg, 2005). A mutation in *KRAS* leads to adenoma growth and progression and is found in about 50% of adenomas and carcinomas (Bos, 1989).

Moreover, allelic losses of chromosome 17p and 18q are extensively well-known late in the process (Fearon & Vogelstein, 1990) and leads to tumor development which is seen in up to 75% and 73% in colorectal carcinomas respectively. Most tumors show allelic loss on both 17p and 18q, rather than loss on either of them alone (Vogelstein et al., 1988). For the region on 18q, allelic loss mainly occurs later in the adenoma-carcinoma process. The most prominent genes to be involved are *SMAD2* and *SMAD4*. Both encode key signaling molecules in the transforming growth factor β (TGFβ) pathway and are involved in cell growth, differentiations, matrix production and apoptosis. These genes are also mutated in CRC cases (Eppert et al., 1996; Leslie, Carey, Pratt, & Steele, 2002). The region on chromosome 17p contains the gene *P53*, which has the ability to, in the presence of DNA damage, prevent cell proliferation, stimulate DNA repair but also promote apoptosis if repair is insufficient. Alterations of *P53* or allelic loss at 17q exist at a low frequency early in the process and in up to 75% of carcinomas, pinpointing that the functional inactivation of *P53* plays a role late in the adenoma-carcinoma sequence (Baker et al., 1989; Leslie et al., 2002).

Figure 1. Schematic view over genetic events in the CIN pathway.



Adapted and modified from Fearon and Vogelstein1990.

1.2.2 Microsatellite instability (MSI) pathway

Microsatellites are short repetitive DNA sequences of 1-6 base pairs in length, distributed throughout the genome. The microsatellite instability (MSI) pathway is characterized by a deficiency in the mismatch repair (MMR) system. The MMR genes are responsible for recognizing and repairing mismatched bases in the DNA strand during replication. A deficient MMR system leads to erroneous insertions or deletions in microsatellites which consequently leads to alleles with different sizes and frameshift mutations, which can end up in a truncated protein (Bogaert & Prenen, 2014; Peltomaki, 2001).

The MSI pathway is present in about 15% of sporadic CRC cases and in Lynch syndrome (Aaltonen et al., 1993; Peltomaki, 2001). MMR gene deficiency can be due to a germline mutation (alone) and/or a second hit (mutation mainly in *MLH1*, *MSH2*, *MSH6* or *PMS2*) as in LS tumors (which will be discussed in 1.3.1.2). In sporadic cases the *MLH1* inactivation is due to hypermethylation of the promotor (Kuismanen, Holmberg, Salovaara, de la Chapelle, & Peltomaki, 2000). Moreover, sporadic MSI-H tumors is characterized by widespread DNA hypermethylation (Cunningham et al., 1998).

MSI is graded by the amount of instability in the tumor; MSI-H (high), MSI-L (low) and MSS (stable). Tumors with a high level of MSI have at least more than a hundred times higher mutation rate than normal cells (Parsons et al., 1993). Phenotypic tumor characteristics are right-sided location, mucinous cell type and manifestation of tumor infiltrating lymphocytes (Bogaert & Prenen, 2014).

Additional MSI associated mutations are seen in genes as *TGFβR2*, *MSH3*, *MSH6*, *TCF4* and *BAX* (Peltomaki, 2001).

1.2.3 Epigenetic alterations

It has been suggested that epigenetics have a considerable effect on CRC development, both independently and in interaction with genetic events.

The CpG island methylator phenotype (CIMP) is characterized by high rate of methylated CpG islands in genes. CpG islands are approximately 1kb of GC-rich DNA sequence often located near the promotor of genes. Methylation, addition of a methyl group (CH_3) to the CpG site cysteine in the promotor in tumor suppressor genes like *APC* and *MLH1,* is associated with CRC (Jones & Laird, 1999). This interference event leads through transcriptional silence to gene inactivation.

It is proposed that CIMP CRCs develop via the serrated neoplasia[1] pathway (Jass et al., 2000). The serrated neoplasia pathway has been reported to be associated, not only with DNA

---

[1] Criteria's for serrated polyposis are: two larger (>1cm in diameter) polyps mainly in proximal colon, a first degree relative with serrated polyps and also having >30 polyps spread in colon.

methylation and MSI, but also with mutations in *BRAF* and *KRAS* (Sweetser, Smyrk, & Sugumar, 2011). It has been suggested that CIMP positive tumors can be classified into subgroups: CIMP-H (high degree of methylated genes) + MSI + *BRAF* mutations (rare *KRAS* + *P53*); CIMP-L (low degree of methylated genes) + *KRAS* mutations (rare MSI, *BRAF* and *P53*); and the non-CIMP group + *P53* mutations (rare MSI, *BRAF* and *KRAS*). CIMP-H and CIMP-L seem to be found primarily in proximal colon whereas non-CIMP tumors mainly exist in distal colon (Shen et al., 2007). Similar classifications are done but with addition of a fourth group; non-CIMP + low frequency of gene mutations, which mainly associates with rectal tumors (Bogaert & Prenen, 2014).

1.3 COLORECTAL CANCER PREDISPOSITION

It has been shown that having a first-degree relative with CRC increases the risk of disease to approximately twofold and more than fourfold when having more than one relative with the disease. Also, the risk increases up to fourfold if the relative is diagnosed at an early age (<45) indicative of a genetic contribution when the CRC development occurs earlier in life (Johns & Houlston, 2001).

In a large twin study, Lichtenstein *et al.* estimated the genetic contribution to CRC development to 35%, and the remainder would be due to environmental factors (Lichtenstein et al., 2000). Today, approximately 10-15% can be explained by genetic factors (Whiffin & Houlston, 2014). Further investigation to learn about the missing genetics is needed to be able to explain the total genetic burden. The remaining genetic contribution is likely to be attributable to a mixture of low- to moderate penetrant genetic variants acting together.

1.3.1 Colorectal cancer syndromes

Two to three decades ago, the genetic research focused mainly on large families and the localization and identification of CRC susceptibility high-penetrance genes were successfully reported by linkage and positional cloning studies (table 1).

Table 1. Known genetic predisposing CRC syndromes

Gene(s)	**Syndrome**	**Risk in mutation carriers**	**Genetic model**	**Reference**
APC	Familial adenomatous polyposis	90% by age 45	Dominant	Bodmer *et al.* 1987
MMR (*MLH1, MSH2, MSH6, PMS2*)	Lynch Syndrome	40-80% by age 75	Dominant	Peltomaki *et al.* 1993, Lindblom *et a.l* 1993, Miyaki *et al.* 1997, Hendriks *et al*. 2006
SMAD4/BMPR1A	Juvenile polyposis syndrome	17-68% by age 60	Dominant	Howe *et al.* 1998, Howe *et al.* 2001
STK11	Peutz-Jeghers syndrome	39% by age 70	Dominant	Hemminki *et al.* 1998
MUTYH	MYH-associated polyposis	35-53%	Recessive	Al-Tassan *et al.* 2002, Lubbe *et al.* 2009
POLD1/POLE	Oligopolyposis		Dominant	Palles *et al.* 2013

Adapted and modified from (Whiffin & Houlston, 2014).

1.3.1.1 Familial adenomatous polyposis

Familial adenomatous polyposis (FAP) is caused by mutations in the tumor suppressor gene *APC*. Herrera *et al.* were the first to demonstrate a deletion of chromosome 5q in a CRC patient (Herrera, Kakati, Gibas, Pietrzak, & Sandberg, 1986). This was followed by two independent studies reporting linkage of CRC to chromosome 5q21 (Bodmer, Bailey, Bodmer, Bussey, Ellis, Gorman, Lucibello, Murday, Rider, Scambler, & et al., 1987; Leppert et al., 1987). Germline mutations in *APC* were, in 1991, demonstrated to cause FAP (Groden et al., 1991). The incidence of FAP is estimated to around 1:10,000 – 30,000 and accounts for approximately 1% of all CRC cases (Bogaert & Prenen, 2014). Generally, patients with FAP develops hundreds to thousands of adenomatous polyps in colon and rectum in late childhood and adolescence and if left untreated they will develop into cancer (Kinzler & Vogelstein, 1996). Other features reported in FAP are upper gastrointestinal tract polyps, congenital hypertrophy of retinal pigment epithelium, desmoid tumors as well as other extracolonic malignancies (Galiatsatos & Foulkes, 2006).

FAP is a autosomal dominant disease and for the majority of patients a germline mutation occurrs in *APC*. De novo mutations are estimated to account for 25% of the mutations (Bogaert & Prenen, 2014). In sporadic tumors, a mutation in *APC* occurs early in the chromosomal instability (CIN) pathway (Kinzler & Vogelstein, 1996; Powell et al., 1992) described in 1.2.1. Individuals carrying a de novo germline *APC* mutation are predisposed to polyps and subsequent colorectal cancer. When the wild-type alleles in cells become inactivated by a second-hit mutation, adenomatous polyps start to develop. This event supports the two-hit-hypothesis proposed by Knudson in 1971 (Knudson, 1971).

The *APC* gene plays a role in the WNT-signaling pathway by regulating the transcription factor β-catenin. *APC* prevents accumulation of β-catenin. An inactivating mutation (e.g. insertions or deletions) in the gene leads to a truncated protein which follows by β-catenin accumulation and the control of the cell proliferation is lost. Most germline mutations are within exon 15 (at the 5´end, position 1061 and 1309) (Galiatsatos & Foulkes, 2006). The APC protein is composed of 2843 amino acids.

Today, around one thousand *APC* mutations are known to cause FAP (Bogaert & Prenen, 2014). Due to the different types of mutations, phenotypic variants of FAP exists. One variant is attenuated FAP (AFAP), a milder form of FAP with lesser number of adenomatous polyps (10-100), later clinical presentation and also with a lower risk of cancer development (Bogaert & Prenen, 2014). The cumulative CRC risk for a AFAP patient at the age of 80 is up to 75% in the proximal colon. Gardner syndrome (GS) is a variant characterized by osteomas, dental anomalies, epidermal cysts, and soft tissue tumors. Turcot syndrome (TS) is an association between CRC and medulloblastoma, a cancerogenic tumor in the brain (Galiatsatos & Foulkes, 2006). Additionally, a mutation in *APC*, I1307K, is found to be associated with CRC in the Ashkenazi Jewish population with a frequency of 10% of Ashkenazi CRC patients compared to 6% in controls (Laken et al., 1997).

1.3.1.2 Lynch Syndrome

Over hundred years ago, a pathologist Aldred S. Warthin, described a family with different cancer types: the "family G" (Warthin, 1985). Later, in 1966, Henry Lynch and colleagues observed families with an early onset and dominant genetic predisposition to colorectal cancer but with a different phenotype than FAP due to absence of multiple colonic polyps (Lynch, Shaw, Magnuson, Larsen, & Krush, 1966). This syndrome was called Lynch Syndrome (LS) and later also hereditary nonpolyposis colorectal cancer (HNPCC). It is characterized by rapid growth from adenoma to carcinoma, right sided tumors and a mean age at diagnosis of 45 years. Also other cancer types like endometrium, stomach, bowel, pancreas, urinary, brain and ovaries occurs in this syndrome. Up to 3% of CRCs are attributed to LS (Lynch & de la Chapelle, 2003). The lifetime risk of developing CRC for a LS patient is 25-75% with an increased risk in *MLH1* and *MSH2* mutation carriers compared to *MSH6* and *PMS2* carriers who in addition have an increased risk for endometrial cancer (de Vos tot Nederveen Cappel et al., 2013).

As described earlier, MMR genes are responsible to repair errors during DNA replication. LS develop due to germline mutations in the MMR genes where *MLH1* and *MSH2* account for 90% of the mutations and *MSH6* approximately 7-10%. It is estimated that *PMS2* is responsible for 1-2% of MMR mutations (Hendriks et al., 2006; Lynch & de la Chapelle, 2003; Peltomaki & Vasen, 2004).

In 1993, HNPCC was mapped to chromosome 2p16 (Peltomaki et al., 1993) and 3p21 (Lindblom, Tannergard, Werelius, & Nordenskjold, 1993; Peltomaki et al., 1993). Both these loci where identified by linkage analysis in two and three large families respectively. The disease causing genes were recognized as *MSH2* and *MLH1* (Bronner et al., 1994; Fishel et al., 1993; Leach et al., 1993). *MSH6*, was identified a little later and localized close to 2p16 (within 1 Mb from *MSH2*) (Akiyama et al., 1997; Drummond, Li, Longley, & Modrich, 1995; Miyaki et al., 1997) and *PMS2* were localized to chromosome 7p22 (Hendriks et al., 2006; Nicolaides et al., 1994).

Diagnosis classification criteria of LS, the Amsterdam criteria I, were proposed 1991 (Vasen, Mecklin, Khan, & Lynch, 1991). This criteria was revised in 1999 to be less stringent and to also include extracolonic cancers, Amsterdam criteria II, (table 2) (Vasen, Watson, Mecklin, & Lynch, 1999). Later on, the Bethesda criteria (table 3) were established to be used for identifying patients for MSI screening (Rodriguez-Bigas et al., 1997; Umar et al., 2004). The MSI test, using five microsatellite markers (BAT25, BAT26, D2S123, D5S346 and D17S250), is performed on tumor and matching normal DNA. MSI at two or more markers is defined as MSI-H, at one marker or no markers as MSI-L and MSS, respectively.

Table 2. Amsterdam Criteria II
There should be at least three relatives with an LS-associated cancer* and:
One should be a first-degree relative of the other two
At least two successive generations should be affected
At least one cancer should be diagnosed before age 50
FAP should be excluded in the CRC case
Tumors should be verified by pathological examination

*LS-associated cancer: CRC, cancers of endometrium, small bowel, ureter or renal pelvis.

Table 3. The Bethesda Criteria (revised)
Tumors from individuals should be tested for MSI in the following situations:
CRC diagnosed in patients who is less than 50 years of age
Presence of synchronous, metachronous CRC or other LS-associated tumor*, regardless of age
CRC with the MSI-H histology diagnosed in a patient who is less than 60 years of age
CRC in a patient with one or more first-degree relatives with an LS-related tumor, with one of the cancers being diagnosed under age of 50 years
CRC in a patient with two or more first- or second degree relatives with LS-related tumors, regardless of age

* LS-associated tumor: colorectal, endometrial, stomach, ovarian, pancreas, ureter and renal pelvis, biliary tract, brain tumors, sebaceous gland adenomas and keratoacanthomas, and carcinoma of the small bowel. **MSI-H (high), pathological associated features include infiltrating lymphocytes, Crohn´s like lymphocytic reaction, mucinous/signet-ring differentiation, or medullary growth pattern.

1.3.1.3 Additional colorectal cancer syndromes

MUTYH-associated polyposis (MAP) has a recessive mode of inheritance and is associated with biallelic germline mutations in the *MUTYH* gene, mainly the variants Y179C and G396D. *MUTYH* has a role in the base excision repair (BER) system. The syndrome was described after the investigation of a FAP phenotype-like family but without an identifiable *APC* mutation. The estimated risk of developing CRC for these patients is 80%. No extra-intestinal manifestations are described for the patients but polyps in the upper gastrointestinal tract are seen in about 30% of them (Al-Tassan et al., 2002). It has been discussed whether it also exists an association between MAP and monoallelic mutation, but this has been ruled out by at least one study done by Lubbe *et al*. (Lubbe, Di Bernardo, Chandler, & Houlston, 2009). Somatic mutations are reported in *APC* and *KRAS*. Furthermore, MAP seems to proceed through both CIN and MSI pathway (Lipton et al., 2003).

Peutz-Jeghers syndrome (PJS) patients carry a mutation in *STK11* (Hemminki et al., 1998). The incidence is estimated to 1:120,000. It is a rare, dominantly inherited condition characterized by association to gastrointestinal polyposis (primarily in the small intestine but also in stomach, large bowel, and extraintestinal sites), mucocutaneous pigmentation and an increased risk of colorectal, gastric, pancreatic, breast, and ovarian cancers (McGarrity, Kulin, & Zaino, 2000).

Juvenile polyposis syndrome (JPS) is caused by mutations in the TGF-β signaling pathway genes, *SMAD4* or *BMPR1A*. The syndrome is characterized by hamartomatous polyps and an increased risk of gastrointestinal cancer (Howe et al., 2001; Howe et al., 1998). It has an autosomal dominant mode of inheritance but with a reduced penetrance. The risk alleles are rare but still carrying them gives more than a tenfold risk of developing CRC (Whiffin & Houlston, 2014).

Recently, high penetrant mutations in CRC families have been reported in the proof reading domains of *POLD1* and *POLE*. A germline mutation in *POLE* L424V predisposes to multiple colorectal adenoma and carcinoma where some cases have phenotypic similarities to MAP and LS. *POLD1* S478N predispose to not only colorectal tumors but also endometrial cancer and possibly also to brain tumors (Palles et al., 2013). Interestingly, a recent study reports *POLE* mutation in a large CRC family from Sweden. The variant *POLE* A363L predispose to a broad spectrum of tumors, such as ovarian, endometrial and brain tumors (Rohlin et al., 2014). Furthermore, it seems like different phenotypes exists in polymerase proof reading-associated polyposis and, just like LS, caused by different mutations.

Familial colorectal cancer type X (FCCTX) is described as families who meet the clinical criteria's Amsterdam-I but have no evidence of an abnormality in any of the DNA mismatch repair genes. Additionally, in contrast to LS, FCCTX patients may not have an increased risk for other cancers; their relatives show a lower CRC incidence and, the mean age is higher, 60.7 (48.7 for LS). Little is known about the molecular mechanism for FCCTX, but it has been suggested to aggregate by chance alone, in combination with lifestyle factors or be due to some genetic components yet to be found (Lindor et al., 2005).

The mixed polyposis syndrome is demonstrated to be caused by duplication and by elevated expression of the gene *GREM1* which likely reduces the BMP pathway activity which drive tumorigenesis (Jaeger et al., 2012). Moreover, CRC associated syndromes such as Cowden- and Bannayan-Ruvalcaba-Riley syndrome have been described with mutations in *PTEN* (Marsh et al., 1998) and Muir-Torre syndrome with deficiency in the MMR genes (Schwartz & Torre, 1995). The serrated polyp's syndrome and FCCTX (mentioned above) cannot fully be explained yet. This supports the hypothesis that there exist additional risk genes and genetic variants to be found and explained.

Genetic counselling, pre-symptomatic testing and early detection are of major clinical importance. For individuals with an increased risk of developing the disease, surveillance programs including mutational screening, regular colonoscopy and removal of precancerous polyps can reduce morbidity and mortality.

1.3.2 Low-risk genetic variants

Finding genes by linkage analysis have met little success lately. Recently, efforts to identify genetic variants with a modest individual risk have been more successful. These variants seldom segregate within one family and therefore linkage analysis is not efficient for identification. Instead, genome wide association studies (GWAS) (will be explained in 1.5.2) have been used to identify numerous single nucleotide polymorphisms (SNPs). The SNPs are common low penetrance alleles with a minor allele frequency (MAF) > 10% and each with a small individual relative risk: odds ratio (OR) < 1.5. They may, however, additively increase an individual's risk of developing CRC (figure 2).

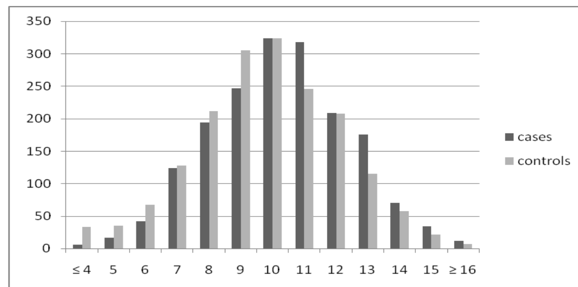


Figure 2. Polygenic model of 11 CRC-related SNPs. Distribution of risk alleles among cases and controls: black, cases; grey, controls.
Adopted from (von Holst et al., 2010), (***paper I***).

Since the first GWAS in 2007, over 20 SNPs have been proposed as CRC susceptibility loci in the European population (table 4) (Broderick et al., 2007; Dunlop et al., 2012; Haiman et al., 2007; Houlston et al., 2010; Houlston et al., 2008; Jaeger et al., 2008; Poynter et al., 2007; Tenesa et al., 2008; I. P. Tomlinson et al., 2011; I. P. Tomlinson et al., 2008; I. Tomlinson et al., 2007; Whiffin et al., 2014; Zanke et al., 2007). Since GWAS are used to capture variations in the genome, the knowledge of each SNP´s actual effect is limited. Using tag-SNPs in linkage disequilibrium (LD) makes it even more difficult to know which the strongest candidate variant is. Also, since most of these CRC-associated SNPs are within non-coding regions of the genome, the influence on the gene regulation and disease development is difficult to explain.

Table 4. CRC susceptibility loci in the European population.

Locus	Closest gene(s)	SNP (tagged)	Reference
1q25.3	LAMC1	Rs10911251	Whiffin et al. 2014
1q41	DUSP10	Rs6691170	Houlston et al. 2010
3q26.2	TERC, MYNN	Rs10936599	Houlston et al. 2010
6p21.2	CDKN1A	Rs1321311	Dunlop et al. 2012
8q23.3	EIF3H	Rs16892766	Tomlinson et al. 2008
8q24.21	MYC	Rs6983267	Zanke et al. 2007, Tomlinson et al. 2007
9p24.1		Rs719725	Poynter et al. 2007
10p14	GATA3	Rs10795668	Tomlinson et al. 2008
10q24.2		Rs1035209	Whiffin et al. 2014
11q13.4	POLD3	Rs3824999	Dunlop et al. 2012
11q23.1	FLJ45803	Rs3802842	Tenesa et al. 2008
12p13.32	CCND2	Rs3217810	Whiffin et al. 2014
12q13	DIP2B, ATF1	Rs11169552	Houlston et al. 2010
14q22.2	BMP4	Rs4444235 Rs1957636	Houlston et al. 2008 Tomlinson et al. 2011
15q13.3	SCG5, GREM1	Rs4779584 Rs16969681 Rs11632715	Jaeger et al. 2008 Tomlinson et al. 2011
16q22.1	CDH1	Rs9929218	Houlston et al. 2008
18q21.2	SMAD7	Rs4939827	Broderick et al. 2007
19q13.11	RHPH2, GPATCH1	Rs10411210	Houlston et al. 2008
20p12.3	BMP2	Rs961253 Rs4813802	Houlston et al. 2008 Tomlinson et al. 2011
20q13.33	LAMA5	Rs4925386	Houlston et al. 2010
Xp22.2	SHROOM2	Rs5934683	Dunlop et al. 2012

Today, little is known about the role of these variants but some of them map close to genes that may influence cancer development. A few variants are close to, or within, genes acting in the TGF-β signaling pathway, which is already known to predispose to CRC. The CRC associated variant at 18q21.1, in *SMAD7*, acts as an antagonist in the TGF-β signaling pathway. One study reported, after resequencing a large LD block, an association between the G allele of Novel 1 (rs58920878) and a reduced expression of *SMAD7* in colorectal tumors (Pittman et al., 2009). Furthermore, the closest genes in the regions 14q22.2, 20p12.3 and 15q13.3 are *BMP4*, *BMP2* and *GREM1*, also members of the TGF-β family (I. P. Tomlinson et al., 2011).

The region on 8q24.21 is associated with an increased risk of developing not only CRC but also other tumors, such as breast-, bladder-, ovarian- and prostate cancer (Whiffin & Houlston, 2014). For example, the CRC specific SNP rs6983267 has been reported to be associated with prostate cancer. The risk allele (G) of rs6983267 affects the last nucleotide in a transcription factor-binding site by stronger binding to the WNT-regulated transcription factor 4 (TCF4) (Tuupanen et al., 2009), which has been reported to interact with the *MYC* proto-oncogene (Pomerantz et al., 2009). *MYC* is known to be over-expressed in multiple cancer types. Interestingly, the risk allele has been proposed to up-regulate *MYC* and thus worsen the CRC prognosis (Takatsuno et al., 2013).

The region on 8q23.3 has been further analyzed by sequencing, mapping LD block and functional studies. The variant rs16888589, in strong LD with rs16892766, has been proposed to explain the association to CRC by the influence of the transcriptional regulation of *EIF3H*. It is shown that higher expression of *EIF3H* increases CRC progression (Pittman et al., 2010). Another study reported that the variants on 8q23.3 and 11q23.1 are associated with an increased CRC risk among MMR gene mutation carriers, which indicates that SNPs also can act as modifiers (Wijnen et al., 2009).

In an attempt to replicate the published CRC risk loci in Swedish patients, we performed a case-control study on 11 SNPs reported at the time. We were able to confirm most of the SNPs and furthermore suggest some unique genotype-phenotype correlations (***paper I***, chapter 4.1). We continued to study these SNPs of interest by doing an association study with morphological features. We present some data that might, if replicated in a larger material, help to link the genetic contribution with tumor morphology, (***paper II***, chapter 4.2).

Moreover, a very recent study identified a novel CRC susceptibility locus at 10q24.2 (rs1035209). Computer analysis with combined data sets also showed significant associations at 12p13.32 (rs3217810, *CCND2*) and 1q25.3 (rs10911251, *LAMC1*). *CCND2* plays a role in the cell cycle control and *LAMC1* have earlier been reported with mutation in CRC (Whiffin et al., 2014).

More studies are needed to elucidate the functional role of the SNPs described here and thereby explaining the effect on the disease. More SNPs reported to be associated with CRC

were not brought up here since they are population specific and not statistically significant or validated in the European population.

Noteworthy, rare mutations and variation changes in genes such as *TGFβR1*, *AXIN2*, *CDH1*, *CHEK2*, *CTNNB1*, *GALNT12, EPHB2, PTPRJ*, *HRAS1* as well as additional MMR genes (not mentioned before) such as *EPCAM*, *MLH3* and *MSH3* are also reported to be associated with CRC but will not be discussed further (de Jong et al., 2002; Fearnhead, Winney, & Bodmer, 2005; Guda et al., 2009; Kilpivaara, Alhopuro, Vahteristo, Aaltonen, & Nevanlinna, 2006; Richards et al., 1999; Valle et al., 2008; Venkatachalam et al., 2010; Zogopoulos et al., 2008).

1.4 EVIDENCE FOR ADDITIONAL COLORECTAL CANCER SUSCEPTIBILITY GENES

1.4.1 Non- FAP/LS familial cancer

Around 10% of Swedish CRC patients have a family history of the disease but without an identified disease causing gene mutation. A Swedish study concluded and categorized these patients into sub-groups (Olsson & Lindblom, 2003). Hereditary colorectal cancer (HCRC) patients, with a dominant mode of inheritance in the family, having ≥ 3 relatives with CRC account for almost 2% of these patients. Patients having two close relatives (TCR) with CRC represent more than 8% of these and the SNPs from GWAS studies are estimated to explain 6% (figure 3) (Olsson & Lindblom, 2003; Picelli, Von Holst, & Wessendorf, 2009). The observed CRC lifetime risk is similar between HCRC and LS patients, even though the former group is diagnosed later in life. TCR patients show a higher number of adenomas compared to LS and HCRC patients. Still, HCRC is considered to be of high-risk whereas a TCR patients has a low- to moderate risk of developing CRC (Lindgren, Liljegren, Jaramillo, Rubio, & Lindblom, 2002). This indicates that there are both high-risk genes and low-risk variants yet to be found.

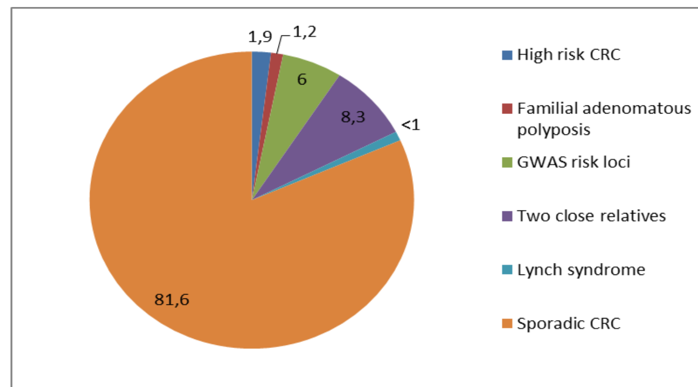


Figure 3. Distribution based on the genetic background in Swedish CRC cases. Adopted and modified from (Olsson & Lindblom, 2003; Picelli et al., 2009).

1.4.2 Colorectal cancer susceptibility loci

A Swedish linkage analysis (will be described in 1.5.1) study on 18 CRC families by Djureinovic *et al.* found suggestive linkage to regions on chromosome 11q13.2-13.4, 11q22.1-23.1, 14q23.1-24.1 and 22q12.1 (Djureinovic et al., 2006). As a follow up to further investigate those regions, Picelli *et al.* pooled this data by adding 12 families, which interestingly gave support for a novel locus on 3q (Picelli et al., 2008). Another study, using 69 families from the United Kingdom, showed proof of linkage (HLOD> 3) to this region on 3q21-24 (Z. Kemp et al., 2006).

A region on 9q has been described as a candidate gene region in many studies. Wiesner *et al*. were the first to report this region at 9q22.2-31.2 (Wiesner et al., 2003). Skoglund *et al.* confirmed this region, using Swedish families, and defined the region to be within the previously reported region with a maximum LODscore of 2.4 at 9q22.32-31.1 for one large family. Further fine-mapping narrowed down this region to 7.9 centimorgan (cM) (Skoglund et al., 2006). Once again Kemp *et al.* managed to find suggestive linkage also within this region and defined the region to 9q22.33 (Z. E. Kemp et al., 2006).

Saunders *et al*. have reported evidence for linkage to regions on chromosome 9q33.3-q34.3, 10p15.3-p15.1 and 14q24.3-q31.1 (Saunders et al., 2012) and Roberts *et al*., using serrated polyposis cases, published linkage to chromosome 2q32.2-q33.3 (Roberts et al., 2011).

Furthermore, performing linkage analysis using an Ashkenazi family resulted in linkage to CRC at locus 15q14-22 (*CRAC1*) (I. Tomlinson et al., 1999).Within this region, the gene *GREM1* is now known to be disease causing for hereditary mixed polyposis syndrome (Jaeger et al., 2012).

Recently, an American study by Cicek *et al.* proposed CRC linkage to four regions: 4q21, 8q13, 12q24 and 15q22 (Cicek et al., 2012). Using the same approach as Cicek *et al.*, who performed subgroup analysis such as: all families, high risk families, moderate risk families and early onset families, our group tried to confirm these findings using 121 Swedish families. None of these regions were replicated. However, suggestive linkage to loci 9q31.1 was found, which once again indicate a disease-causing gene in this region yet to be found. Additionally, suggestive linkage to 4p16.3, 17p13.2 and Xp22.33 was reported (Kontham, von Holst, & Lindblom, 2013) (***paper III****,* chapter 4.3).

In recent times our group performed a linkage analysis on colon- and rectal cancer families which identified linkage to the region 6p21.1-p12.1 in the colon cancer families and to 18p11.2 in the rectal cancer families (von Holst et al, unpublished data, ***paper IV****,* chapter 4.4).

1.5 STRATEGIES FOR FINDING COLORECTAL CANCER PREDISPOSING GENES

1.5.1 Linkage analysis

Linkage analysis has been an effective tool when identifying monogenic disease causing colorectal cancer genes like *APC*, and mis-match repair genes (MMR) like *MLH1* and *MSH2* (Bodmer, Bailey, Bodmer, Bussey, Ellis, Gorman, Lucibello, Murday, Rider, Scambler, & et al., 1987; Lindblom et al., 1993; Peltomaki et al., 1993).

The basic idea of linkage analysis is to identify shared genomic regions in family members who segregate the disease (figure 4). Family members are genotyped for numerous of genetic markers, microsatellites or SNPs, spread in the genome. These polymorphic markers have known genomic positions from public marker maps. Genotypic data is used to find linked regions; i.e. regions that are likely to segregate with disease more often than expected by chance.

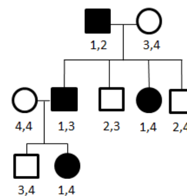


Figure 4. Simplified illustration of linkage analysis; nr 1 segregates with the disease.

During meiosis, crossing-over between loci on two homolog chromosomes occurs, i.e. recombination. Recombination is more likely to occur if two loci are distant and thus more unlikely the closer the loci are. The likelihood for a recombination to occur, is described by the recombination fraction theta (θ) and ranges from zero to 0.5. Two loci are completely linked when θ=0; no recombination has occurred and they segregate together to the next generation. Unlinked loci, far in distance on the same or a different chromosome, corresponds to θ=0.5. The genetic distance between two loci could be measured by theta; 1% of recombination corresponds approximately to 1 cM of genetic distance.

There are two main variants of linkage analysis: parametric analysis and non-parametric analysis. Parametric analysis is also known as model-based analysis since it requires defining the genetic model and providing information such as mode of inheritance, penetrance rate and disease allele frequency. The probability of linkage is defined as the logarithm of odds (LOD) score where the odds of linkage represent the ratio of two hypotheses; the alternative hypothesis that the loci are linked (θ=0) and the null hypothesis that the loci are not linked (θ=0.5) (Morton, 1955). A LOD score of $\geq$3 (the odds of 1000:1 in favor of linkage) indicate linkage and thus that the marker and disease are linked and not separated by recombination. Linkage $\leq$ -2 indicates no linkage, conferring that the disease are not linked to the marker.

Everything in between ≥3 and ≤ -2 are considered to be suggestive linkage and requires further studies. When several families are studied the LOD scores are summed up.

Non-parametric linkage (NPL) is an alternative to be used when the mode of inheritance, dominant or recessive, is unknown (Kruglyak, Daly, ReeveDaly, & Lander, 1996). NPL is thus a model-free method and is suitable when analyzing a complex disease influenced by numerous genes, environmental factors and their interaction. This method was originally used for sib-pair analysis, NPL_{pairs}, and calculates the probability for affected relative pairs to share identical-by-decent (IBD) alleles at a locus close to the disease gene more often than expected by random segregation. To fulfill the IBD criterion these alleles have to be inherited from the same ancestor. According to Mendelian second law of inheritance the probability for a sib-pair to share 0, 1 or 2 alleles are 25%, 50% and 25%. If the pair would share allele's more than random assumptions, linkage would be indicated. Later this method was further developed and applied to general pedigrees, NPL_{all}, where not only sib-pair but the alleles of all affected relatives are compared (Whittemore & Halpern, 1994).

No further biological or physiological information is needed to perform the linkage analysis, which is an advantage with this tool. One disadvantage with linkage analysis is the lack of consideration of genetic heterogeneity; where the disease trait is caused by mutations at several different loci. One way to come across this is to map one family at a time, informative enough, as was done for *MLH1* and *MSH2* genes (Lindblom et al., 1993; Peltomaki et al., 1993). However, many families and thus many genotyped affected individuals increase the power to detect linkage, but it is important to be selective when collecting families; they need to segregate the same phonotype. When a linked region is identified, additional studies are needed to find the harboring gene.

1.5.2 Association studies

Genome-wide association studies were introduced in the beginning of the 21st century as a complementary tool to be used in the continuous search for colorectal cancer susceptibility loci. Numerous genetic variants are analyzed and their frequencies compared between a case cohort (affected individuals) and a control cohort (unaffected individuals), preferably in a large population. Common single nucleotide polymorphisms (SNPs), variation at a single DNA base, are used. If a SNP is more common among the cases compared to the controls, the SNP is associated with the disease.

A GWAS is often preceded by at least two steps. First cases and controls are genotyped for a large number of SNPs. The SNPs with the lowest p-value are then replicated in an additional independent cohort of cases and controls to verify which SNPs of interest remain, thus avoid reporting false-positive findings. For a SNP to be considered significant the p-value must be low (10^{-8} or below), when correction for multiple testing has been done. The Bonferroni correction is generally applied even though it tends to be a bit strict because it considers all SNPs to be independent from each other and not correlated as known due to linkage disequilibrium (LD). This could lead to loss of potentially interesting SNPs (Pearson &

Manolio, 2008). LD refers to a number of alleles or genetic markers occurring more or less frequently, as haplotypes, than would be likely by chance.

The allele frequencies could vary between population subgroups, due to ethnicity or geographical background and could confound the results. To avoid this, adjustment for population stratification can be made (Pearson & Manolio, 2008). A way to avoid population stratification is to perform family-based association studies where the affected offspring are the cases and the unaffected parents the controls. The frequencies of transmitted and non-transmitted alleles to the affected child are calculated for the markers.

Association studies could be performed using different methods and genetic models. Odds Ratio (OR) is calculated by comparing frequencies in cases and controls. It is significantly better to compare not only genotypes for the homozygotes but also combining homozygotes and heterozygotes for the risk- or wild-type allele to clarify any dominant and recessive effect.

Due to the rapid development of high throughput genotyping technologies, assays with over a million variants/SNPs are used nowadays. Until today, over 20 SNPs have been reported to be associated with colorectal cancer in the European population (Broderick et al., 2007; Dunlop et al., 2012; Haiman et al., 2007; Houlston et al., 2010; Houlston et al., 2008; Jaeger et al., 2008; Poynter et al., 2007; Tenesa et al., 2008; I. P. Tomlinson et al., 2011; I. P. Tomlinson et al., 2008; I. Tomlinson et al., 2007; Whiffin et al., 2014; Zanke et al., 2007).

1.5.3 Next generation sequencing

Using next generation sequencing (NGS), also called massively parallel sequencing, makes it possible to sequence millions of DNA fragments at the same time. The sequenced fragments are aligned with a reference genome with the help of bioinformatics tools. Annotations of the variants convey their position in a gene and frequency information (from databases) and categorize them by their functional class.

Compared to conventional Sanger sequencing, the most beneficial effect of NGS is the speed and the possibility to sequence one entire human genome in few days as a cost effective alternative. Still in its infancy, NGS has problem such as mis-incorporation of bases during PCR amplification and difficulty to map sequences correctly (Meldrum, Doyle, & Tothill, 2011). Despite this, NGS is most likely as accurate as Sanger sequencing nowadays (Esteban-Jurado, Garre, et al., 2014).

It is optional to do whole-genome sequencing (WGS) or capture and sequence only specific regions of interest or the coding parts, the exons. Whole-exome sequencing (WES) is cheaper than WGS and, since the majority of mutations in Mendelian disorders are within coding regions, it is a suitable alternative. The cost reduction allows sequencing more patients or gives a higher coverage (average time a nucleotide has been sequenced), which is needed for a variant to be accurately called and retained in the filtering process. When looking at candidate variants one should consider the mode of inheritance, penetrance and frequency of

the disease. Variants shared among affected and not shared among unaffected are to be chosen. Additionally, genes in pathways or interacting genes already known to be involved in the disease development could at first be stronger candidates. Candidate variants should preferably be tested in additional patients.

NGS will hopefully add more knowledge of the functional variation responsible for Mendelian and common diseases. Recently, WES identified a number of rare pathogenic CRC predisposing variants (Esteban-Jurado, Vila-Casadesus, et al., 2014).

1.5.4 Additional approaches

The candidate gene approach is a method to test a potential gene contributing to the disease development. This can be done in disease linked/GWAS regions by investigating genes that might have a pathogenic role. Also, functional genes that may have a potential role in disease mechanism, like genes acting in disease pathways or other genes known from previously published data, could be validated. Candidate gene studies have, during the last two decades, identified rare moderately-penetrant risk alleles with a MAF < 2% and with a relative risk (RR) < 2.0. These rare variants are hypothesized to give a small and independent increased risk for disease development (Whiffin & Houlston, 2014). For example, the variant -93G>A of *MLH1* is reported as a low-penetrance variant for CRC, specifically in relation to MSI-H, and the influence has been shown to be independent of CRC loci identified by GWAS (Whiffin et al., 2011). Furthermore, the variant *APC* I1307K found in 6% of the general Ashkenazi Jews population and in 28% with a family history of CRC increases the risk approximately twofold (Laken et al., 1997; Whiffin & Houlston, 2014).

Somatic chromosomal alterations are commonly found in tumors. Deleted regions in tumor cells but not in normal cells may pinpoint that tumor suppressor genes are located within these regions. LOH of the wild-type allele at 19p and significant results from a linkage study supporting the same region was seen when identifying Peutz-Jegher syndrome (Hemminki et al., 1997). Even though constitutional chromosomal changes are rare, they can point toward a susceptibility gene which was done when localizing *APC* (Bodmer, Bailey, Bodmer, Bussey, Ellis, Gorman, Lucibello, Murday, Rider, Scambler, & et al., 1987).

2 AIMS OF THE THESIS

The overall aim of this thesis was to investigate predisposing genetic loci, both low- and high risk variants, underlying the development of colorectal cancer.

The specific aims were:

1. To confirm published CRC susceptibility loci in a Swedish-based cohort and study genotype-phenotype correlations.
2. To investigate correlations between CRC susceptibility loci and morphological features.
3. To identify novel colorectal cancer loci by genome wide linkage analysis in CRC families.
4. To identify novel colon and rectal cancer loci by linkage analysis followed by exome sequencing in regions of interest.

3 MATERIALS AND METHODS

3.1 PATIENTS

The case cohort in *paper I* and *paper II* were CRC patients of Swedish origin recruited through the Swedish Low-Risk CRC Study Group from 14 different hospitals from central Sweden during 2004–2006. *Paper I* was composed of 1786 consecutive CRC cases. The mean age (at diagnosis) was 68.6 years (range 28–95 years), 53% were men and 47% were women and 22% had a family history of CRC among first- or second-degree relatives. The control cohort was composed of 1749 individuals as follows: 1319 blood donors from the general population between the age of 18 and 65 years and 430 unaffected spouses of CRC patients with the mean age of 66.3 (25–92) years, which were cancer-free (at the time of blood collection) and did not have a family history of any type of cancer. *Paper II* included 1572 cases in the analysis (after exclusion FAP and LS cases according to Amsterdam criteria). The cases were evaluated and discussed by experienced gastrointestinal pathologists using established consensus criteria regarding morphological parameters.

The families used in *paper III* and *paper IV* were ascertained through the clinical genetics department at the Karolinska University Hospital in Stockholm, Sweden between 1990 and 2005. FAP was excluded using medical records from affected individuals and Lynch syndrome was excluded using our current clinical protocol (Lagerstedt Robinson et al., 2007). Furthermore, families with at least two affected relatives informative for linkage analysis, thus at least a sib-pair, were included and individuals with colorectal cancer or a polyp with high degree dysplasia were coded as affected. In *paper III*, 121 CRC families were used. In *paper IV*, CRC families were divided into specifically colon and rectal families respectively. A family was considered a colon- or rectal cancer family depending on what tumor site was most prominent in the family. Families with a less clear predominance of one of the two locations were not included and thus 88 (32 colon and 56 rectal) families were used in this study.

3.2 METHODS

3.2.1 Genotyping

3.2.1.1 Papers I and II

We genotyped one SNP from each locus among those published; rs16892766 on 8q23.3, rs6983267 on 8q24.21, rs719725 on 9p24, rs10795668 on 10p14, rs3802842 on 11q23.1, rs4444235 on 14q22.2, rs4779584 on 15q13.3, rs9929218 on 16q22.1, rs4939827 on 18q21.1, rs10411210 on 19q13.1, rs961253 on 20p12.3 and excluded the following from the analysis: rs355527 on 20p12.3 (tagged by rs961253) and rs7259371 on 19q13.1 (tagged by rs10411210). Six of the SNPs (rs9929218, rs719725, rs4444235, rs4779584, rs10411210 and rs961253) were genotyped using TaqMan SNP Genotyping Assay (Applied Biosystems,

Foster City, CA, USA). Genotyping and first-quality check of the remaining five SNPs (rs6983267, rs16892766, rs10795668, rs4939827 and rs3802842) were performed, using a technology developed by Nanogen, at deCode Genetics, Reykjavik, Iceland.

A Taqman assay includes two allele-specific probes containing the complementary base of the SNP alleles analyzed as well as sequence specific primers. The probes comprise, for each allele separate fluorescent (VIC or FAM) dyes, and together with the PCR primer, they pair to detect the specific target SNP. The assay align with the genome to observe the allele of interest.

3.2.1.2 Papers III and IV

Genotyping was done, at the SNP Technology Platform in Uppsala, separately in two different sets of family material. Genotyping of 548 patients with 6090 markers was performed by the Illumina Infinium assay using the Illumina HumanLinkage-12 DNA analysis bead chip. Additionally, 52 subjects were genotyped using the Illumina Golden Gate assay and the Illumina Linkage Panel IVb (6008 markers). Arrays were processed according to manufactures protocol. Two marker files were merged and 7256 markers were used in the analysis.

3.2.2 Statistical analysis

3.2.2.1 Test for association, Paper I

Deviations of the genotype frequencies in cases and controls from those expected under Hardy–Weinberg equilibrium were calculated by χ^2-tests (one degree of freedom). Allelic frequencies of the SNPs in the case and control groups were compared using a χ^2-test (allele 1 (common) vs allele 2 (minor)) and odds ratio (OR) was calculated. Analyses using different genetic models were done: comparison of homozygotes (genotype 11 vs 22), dominant (11 vs (12+22)), recessive ((11+12) vs 22) and allele frequency ((1) vs (2)), using DeFinetti programme. Level for significance and confidence intervals (CIs) was set to $p<0.05$ and 95% respectively and the results were validated using Statistica 7.0 (StatSoft Inc., Tulsa, OK, USA).

For the genotype–phenotype analysis, we studied sex, age of onset (early vs late, >60 years), family history of CRC (any case of CRC among first- or second-degree relatives), location, colon vs rectum and right vs left (proximal and distal to the splenic flexure). Cross-tabulation analysis, pearson χ^2-test was done using Statistica to calculate the p-value (significance was set to 0.05).

3.2.2.2 Test for association, Paper II

Histopathological assessment included in *paper II* was done according to a standardized protocol (by pathologist Sam Ghazi). The micromorphological parameters assessed were tumour grade, stage, medullary features, mucin production, mucin type, Crohn-like peritumoral lymphocytic reaction (CLR), tumor infiltrating lymphocytes (TILs), desmoplasia,

tumor necrosis, vascular invasion, perineural growth, co-existing polyps, budding and type of tumor margin (for more details see chapter 1.1.3 and material and methods paper II).

The cross tabulation between SNP data and morphology was done using Statistica version 8, (Statsoft Inc., Tulsa, OK). Pearson χ^2-test was used for calculating the p-value. Genotype–phenotype analyses were done using the DeFinetti program. χ^2-tests (one degree of freedom), OR, $p< 0.05$ and 95% CI was used. The results were validated using Statistica version 8.

3.2.3 Linkage analysis

3.2.3.1 Papers III and IV

The principles of linkage are discussed in 1.5.1. Pedcheck was used to check for Mendelian errors. The family based genetic model was used for parametric linkage analysis for all chromosomes, including chromosome X. As a supplement, non-parametric analysis using Whittemore and Halpern NPL statistics was made. LOD scores as well as heterogeneity LOD scores were computed using MERLIN (version 1.1.2). Both dominant and recessive mode of inheritance was analyzed and the disease allele frequency was set to 0.0001. The penetrance rates for the dominant and recessive mode of inheritance for homozygous normal, heterozygous, and homozygous affected were set to 0.05, 0.80, 0.80 and 0.001, 0.001, 1.0 respectively.

In *paper III*, four different analyses were performed using different sets of families: all families, all families with at least three cases (high-risk), all families with CRC among sibs (moderate-risk) and families with a mean age below 50 (early-onset). In *paper IV*, two sub-sets of analyses were run using colon and rectal families respectively.

3.2.4 Exome sequencing

3.2.4.1 Paper IV

Twelve patients from six families, three colon- (110, 301, 350) and three rectal (8, 918, 1213) cancer families respectively, were selected for exome sequencing. In four families two affected sibs were sequenced. In one family a single patient was sequenced and in the last family three sibs were chosen. These families were chosen based on their contribution to the HLODscores in the linkage regions.

Exome capture was performed at SciLifeLab Uppsala using TruSeq Exome Enrichment kit (Illumina). For one patient, exome capture was performend at SciLifeLab Stockholm using SureSelect XT Human All Exon V5 (Agilent, CA, USA). Libraries were sequenced on Illumina HiSeq2500 systems (Illumina, CA, USA) according to manufacturer's instructions. Base calling was performed using RTA (Illumina), filtering and demultiplexing by CASAVA (Illumina).

The raw reads from the sequencing centers at Uppsala and Stockholm were mapped to the hg19 GRCh37 reference genome sequence using bwa (0.5.9), variants called using GATK (v

1.0.5974) after realignment and recalibration. After quality check, the variant call format file (vcf) for all individuals was provided from the sequencing center in Uppsala. ANNOVAR (released fri 23 aug. 2013) was used to annotate and tabulate our linked regions of interest, which added information such as gene name, position, function and amino acid change. 1000genomes was used to verify the minor allele frequency (MAF) and the rsID was taken from dbSNP (ver. 138). Additionally, we filtered the vcf file to keep non-synonymous variants and indels. We ended up analysing all exonic non-synonymous variants and table_annovar was used to summarize the variants common among our affected individuals.

4 RESULTS AND DISCUSSION

4.1 PAPER I

Association studies on 11 published colorectal cancer risk loci

Genome-wide association studies have recently been performed on different diseases with a great success in discovering disease associated loci. At the time this study was performed, 11 CRC loci had been reported to be associated with the disease. In order to replicate those findings in a Swedish sample set a case-control study was done for those 11 loci.

Significant association between SNPs and Swedish CRC patients were seen for five of the 11 SNPs: rs16892766, rs6983267, rs10795668, rs4779584 and rs4939827, thus confirming previous publications (Broderick et al., 2007; Jaeger et al., 2008; I. P. Tomlinson et al., 2008; I. Tomlinson et al., 2007).

For the SNP rs16892766 on 8q23.3, an increased risk of CRC was identified (P=0.002 for all analyses except under the recessive model) with the highest OR=1.34 (1.13–1.60) for the heterozygous. The risk suggested for the variant rs6983267 on 8q24.21 was confirmed in all the analyses, with the highest OR=1.37 (1.13–1.67) for the homozygous state. rs4779584 on 15q13.3 could be confirmed with an increased risk for heterozygous individuals, OR=1.18 (1.02–1.36). The protecting effects suggested for rs10795668 on 10p14 and rs4939827 on 18q21.1 were both confirmed for homozygous and heterozygous carriers with an OR=0.66 (0.52–0.83) and OR=0.82 (0.70–0.96), respectively.

The rest of the SNPs were not statistically significant but rs3802842 on 11q23.1 indicate a CRC risk showing an OR=1.27 for homozygous. rs9929218 on 16q22.1, rs10411210 on 19q13.1 and rs961253 on 20p12.3 showed weaker results in the same direction as previous publications. Finally, the two SNPs rs719725 on 9p24 and rs4444235 on 14q22.2 were not confirmed.

We could confirm five out of the 11 CRC-associated SNPs using a Swedish case cohort. However, most of our results showed similar OR as previous publications but without a significant p-value. Our data supports that these variants are associated with a little increased risk for disease development.

A polygenic model of distribution of risk alleles between cases and controls in the Swedish population shows a clear shift with a higher number of alleles in affected individuals compared to controls. This pinpoints towards the additive effect the number of risk loci may have on CRC patients.

Genotype–phenotype analysis was performed for all 11 loci and for sex, age, family history and tumor location. Four associations were seen, three for age and one for family history. Being homozygous for the risk allele G for rs6983267 (8q24.21) showed association with

older age (P=0.0014). In contrast, for rs1075668 (10p14) the risk allele G was associated with younger age (P=0.035) and sporadic cases (P=0.047). The T allele of rs10411210 (19q13.1) was associated with younger age (P=0.045) in homozygotes. Replications are needed to verify these genotype-phenotype associations.

Other studies done at that time have performed genotype–phenotype analysis for only one of the loci, 8q24.21, and sex, tumor site, age at diagnosis and family history. Our observed association between rs6983267 (8q24.21) and age, was not seen in other studies. Furthermore, the other studies did not show any significant association with sex or tumor site. However, Tuupanen *et al.* reported an association between this locus at 8q24.21 and family history (Haiman et al., 2007; Poynter et al., 2007; Tuupanen et al., 2008).

At the time for our study, 11 CRC risk loci were published. Short after, two publications reported new loci associated with an increased risk for CRC (Houlston et al., 2010; I. P. Tomlinson et al., 2011). Since then, our Swedish material has been included in the additional GWAS and meta-analysis performed in the European material. One study reported three new CRC susceptibility loci: 6p21 (rs1321311 near *CDKN1A*), 11q13.4 (rs3824999 intronic to *POLD3*) and Xp22.2 (rs5934683, near *SHROOM2*) (Dunlop et al., 2012). Recently, three additional loci were presented associated with an increased CRC risk at locus 10q24.2 (rs1035209), 12p13.32 (rs3217810, intronic in *CCND2*) and 1q25.3 (near *LAMC1*) (Whiffin et al., 2014).

4.2 PAPER II

Colorectal Cancer Susceptibility Loci in a Population-Based Study - Associations with Morphological Parameters

We performed association analysis of CRC associated SNPs and morphological parameters. Of the 11 tested SNPs, six showed significant correlations to ten different tumor features. After OR- and p-value calculations five genotype-phenotype correlations remained statistically significant.

Heterozygous carriers of the T allele of rs6983267 (8q24.21) and homozygous carriers of the C allele of rs4444235 (14q22.2) were both associated with a decreased Crohn-like peritumoral lymphocytic reaction (p=0.021 and p=0.024, respectively). rs10795668 (10p14) showed association between heterozygous carriers and poor differentiation (p=0.015). The T allele of rs10411210 (19q13.1) was negatively associated with desmoplastic response (p=0.004) for homozygous. For the variant rs961253 (20p12.3), the A allele showed association with mucin-producing tumors for both homozygous and heterozygous carriers (p=0.010 and p=0.009). Furthermore, homozygous carriers of the A allele presented more frequently tumors with circumscribed margins (p=0.034) but opposite results were seen for heterozygote carriers.

The susceptibility locus at 8q24.21 (rs6983267) has previously been associated with an elevated risk of adenoma development and an increased risk for prostate cancer (I. Tomlinson

et al., 2007; Yeager et al., 2007). Interestingly, significant correlation between this SNP and MMR status (normal protein expression for MLH1, MSH2, MSH6 as well as MSI-L/MSS), tumor site (distal) and later tumor stage has been reported (Cicek et al., 2009). In our study, being heterozygote for the T allele (protective allele) was associated to Crohn-Like peritumoral lymphocytic infiltration (CRL), an immune response which has been linked to improved survival (Graham & Appelman, 1990). On the contrary, the risk allele on locus on 14q22.2 (rs4444235) showed associated with CRL.

The SNP on 19q13.11 (rs10411210) was negatively associated with desmoplastic response for homozygote carriers of the variant allele (T). Desmoplasia may have a protective effect though it inhibits cancer invasiveness by building barrier against tumor diffusion (Caporale et al., 2005).

The region on 20p12.3 harbors a risk allele (A) associated with mucin-producing tumors. Mucinous differentiation may be used as an indicator of MMR deficiency (Langner et al., 2012). The same SNP showed conflicting association with tumor margin; homozygote carriers showed association to circumscribed margin whereas heterozygote carriers showed and opposite effect. This makes it difficult to interpret this result.

Heterozygosity for the variant allele (A) at 10p14 (rs10795668), reported to be a protective allele, was correlated to poor differentiation.

Furthermore, associations between rs719725 and necrosis, budding as well as desmoplastic reaction showed inconsistent results in homo- and heterozygous carriers indicating that these results are probably false positive. The same conclusion was drawn for the association between rs961253 and infiltrative tumor margin.

It is difficult to interpret some of these correlations because the pathogenicity is still unknown for most of the variants. However, the locus 8q24.21 has been reported to affect the transcription factor-binding site for TCF4 and thereby up-regulate the oncogene *MYC,* which might explain some of the increased CRC risk observed in carriers of the G allele (Tuupanen et al., 2009). Additionally, the closest genes in the regions of 20p12.3 and 14q22.2 are *BMP2* and *BMP4*, both belonging to the TGFβR-family, a morphogenic factor in CRC (I. P. Tomlinson et al., 2011). *BMP4* is overexpressed by human colon cancer cells with mutant *APC* genes. Interestingly, this supports the hypothesis that WNT and BMP pathways interact in cancer pathogenesis (J. S. Kim et al., 2002). The region 19q13.1 is in LD with the gene *RHPN2,* member of the RhoGTPase family. Increased expression of RhoC is correlated with decreased survival time in CRC patients (Bellovin et al., 2006). Further studies are needed to validate our findings in additional cohort.

4.3 PAPER III

Linkage Analysis in Familial Non-Lynch Syndrome Colorectal Cancer Families from Sweden

The estimated lifetime risk of CRC is 5% in the general population. Having a family history of the disease increases this risk. Only a small number of the genetics of CRC can be explained by a known gene mutation. In order to identify novel candidate genes predisposing to CRC, we performed a genome wide linkage analysis on 121 Swedish CRC families. We divided the families into four sub-groups: all families, high-risk families, moderate-risk families and families with an early age of onset of CRC. By using this approach we tried to replicate the beforehand reported findings from Cicek *et al.* showing linkage to 12q24 in all CRC families, 4q21 in early onset 15q22.31 in high-risk families and to 8q13.2 in moderate risk families (Cicek et al., 2012).

No evidence of linkage was found in our study and we failed to confirm the previously reported regions by Cicek *et al.* However, we could report suggestive linkage in the recessive analysis to two regions with a HLOD= 2.2 (rs1338121) for high-risk families, at locus 9q31.1 and for moderate-risk families, at locus Xp22.33 the HLOD=2.5 (rs2306737). Furthermore, for the families with early-onset, one locus on 4p16.3, LOD=2.2 (rs920683) came up as the most prominent one. Also, for the same sub-set group one more locus came up on 17p13.2 with a LOD/HLOD=2.0 (rs884250). We could not see any NPL score above two.

We could not show a statistical significant result from this linkage analysis. However, our region on chromosome 9q31.1 falls within the region reported before (Z. E. Kemp et al., 2006; Skoglund et al., 2006; Wiesner et al., 2003). It has been shown that germline allele-specific expression in the gene *TGFBR1* occurs in 10-20% patients compared to only in 1-3% of controls (Valle et al., 2008). Another study show little support for this gene, with pooled data results, with an OR=1.20 (95% CI, 0.64-2.24) for homozygosity, 1.11 (95% CI, 0.96-1.29) for heterozygosity, and1.13 (95% CI, 0.98-1.30) for carriers of *TGFBR1*6A* (Skoglund et al., 2007). Also, the gene *GALNT12*, in this regions, has been demonstrated with truncating somatic and germline mutations and defects in the O-glycosylation pathway which partly underlie aberrant glycosylation that would influence the risk of developing CRC (Guda et al., 2009).

Four genome wide linkage studies, including our, have been published recently. All using SNP array chips (Cicek et al., 2012; Kontham et al., 2013; Roberts et al., 2011; Saunders et al., 2012). Interestingly, none of the studies succeeded to replicate each other's regions. It might be due to the difference in sample set used or ethnicity in the families.

Further studies are needed to evaluate all candidate regions and genes before one could conclude the importance of our results. We aim to look at exome-data for our families of interest at our four regions.

4.4 PAPER IV

Linkage analysis in familial colon- and rectal cancer

By subdividing the colorectal cancer families into colon and rectal families, we hope that it could result in new specific phenotypic loci. Previous studies show that tumor location is different in FAP compare to HNPCC and that different tumor sites give different genetic alteration and allelic loss at 5p, 17q and 18 (Bufill, 1990; Delattre et al., 1989). Also, it has been shown that gene and protein expressions and mutation rates differs among colon and rectal cancer (Frattini, Balestra, Pilotti, Bertario, & Pierotti, 2003; Kapiteijn et al., 2001).

Linkage analysis showed no LOD or HLOD score above three for our 56 rectal- or 32 colon cancer families (306 and 169 individuals). However, results close to three could be demonstrated for colon- as well as rectal families with a HLOD=2.49, for the colon families on locus 6p21.1-p12.1 and a HLOD=2.55 for the rectal families on locus 18p11.2. These regions, are around 6 Mb and 10 Mb respectively, and have not yet, to our knowledge, been reported by other linkage studies. This might be due to the fact that no previous linkage analysis has been performed studying colon- and rectal families separately.

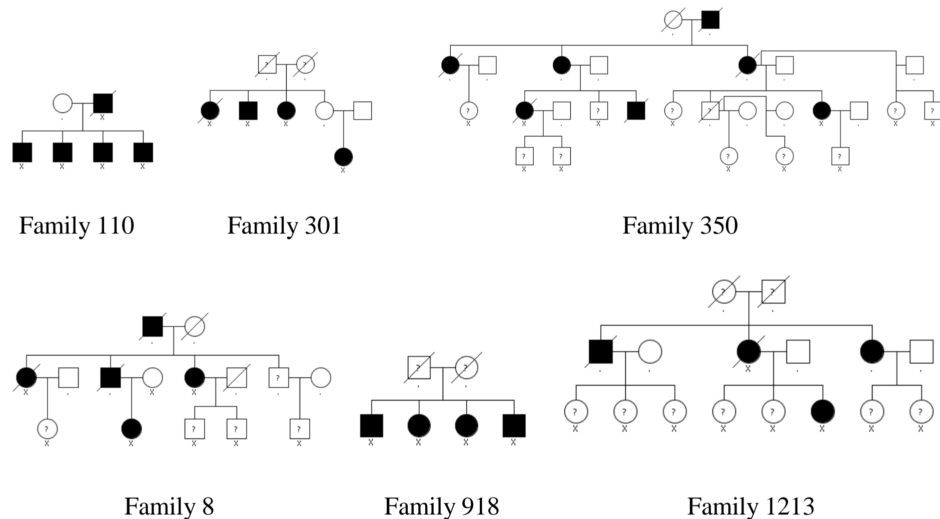
Our 88 families were quite small which explains why only a few families (110, 301 and 350 in the colon region and 8, 918 and 1213 in the rectal region), (figure 5), contribute to each max HLOD. In total, 12 patients were investigated using whole exome sequencing to further study the candidate region on chromosome 6 and 18. In total, 25 colon and 27 rectal variants remained interesting after filtering all but exonic and splice-site variants with a MAF< 20%. We consider variants shared among family members and furthermore, variants shared among more than one family as stronger candidates. Altogether, 49 synonymous SNVs, one frameshift insertion and two non-frameshift deletions were shortlisted. Of these, 27 variants non-synonymous single nucleotide variants (SNVs) were kept among the rectal patients and 25 variants among the colon patients where 10 were non-synonymous SNVs, one a frameshift insertion and two non-frameshift deletions.

Studying the table of variants, some of them, in the genes *LRFN2*, *CUL7*, *CUL9* and *PTK7* on chromosome 6, are more likely to be involved in cancer development. Even though population specific, SNPs located in initial region of *LRFN2* is shown to be in strong linkage disequilibrium with a SNP (rs2494938 at 6p21.1) associated with multiple cancers in the Chinese population (Jin et al., 2012). Interestingly, *CUL7* has been suggested as an oncogene. Kim *et al.* showed that *CUL7* can bind directly to *P53* which can prevent cells from *P53*-dependent apoptosis. Also, they proposed that *CUL7* can interact with *MYC* to promote anchorage-independent cell growth (S. S. Kim et al., 2007). Furthermore, within the same protein family, *CUL9* is as a tumor suppressor gene that promotes *P53*-dependent apoptosis (Pei et al., 2011). *PTK7*, presented in two families, is said to be up-regulated in cancers, including CRC and might be involved in cell proliferation and apoptosis even though the function remains unclear (Meng et al., 2010).

Of the 27 variants in the ten candidate genes on chromosome 18, four are particularly interesting candidates. The variant in the gene *APCDD1* is shared between family members in two families. *APCDD1* is suggested to be regulated by the *β-catenin/Tcf* complex involved in colorectal tumorigenesis (Takahashi et al., 2002). Another interesting variant, present in one of our families, are within the gene *ROCK1,* which is part of the Rho-kinase family and even though the exact role needs to be elucidated, it has been proposed to be involved in WNT-signaling pathway. *ROCK1* has shown to be involved in the CRC development (Sari et al., 2013). The gene *CTAGE1* is described as a cancer antigen for T-cell lymphoma and other malignancies (Usener, Schadendorf, Koch, Dubel, & Eichmuller, 2003). The *POTEC* gene had one variant shared in two families and other variants in single families. *POTE-18* is a member of the *POTE* family, which consists of homologous genes located on chromosomes 2, 8, 13, 14, 15, 18, 21, and 22. The gene on chromosome 21 is expressed in prostate cancer and other cancer types (Bera et al., 2002).

In conclusion, we propose two new linkage regions for colon- and rectal cancer. Within those linked region there are interesting candidate genes likely to predispose to the CRC development. Further studies are needed to support or exclude these to be the disease causing variants.

Figure 5. Pedigrees of three colon (110, 301, 350) and three rectal (8, 918, 1213) cancer families.



Black=affected, white=unaffected, ?=unknown, small polyps or other cancers. Crossed=dead. X=DNA available and used in linkage.

5 CONCLUSION AND FUTURE PERSPECTIVES

The overall aim with these studies was to investigate genetic predisposing factors underlying the development of colorectal cancer. More specifically, finding and confirming both high-risk genetic factors and low-risk variants.

The conclusions made from the thesis are:

Published CRC risk loci could be confirmed in the Swedish population. Even though all eleven were not significantly confirmed, most of them showed similar OR as previous publications. We also showed some significant genotype-phenotype correlations, which need to be replicated in a larger dataset. Furthermore, looking at a polygenic model among the Swedish patients, cases share higher number of risk alleles, which may indicate the collective effect of low-risk variants. At present, an international low-risk study with more than 500,000 SNPs is being genotyped in around 45,000 CRC patients. More than 3,000 Swedish patients are included in the study.

Investigating correlations between the CRC risk loci and morphological parameters resulted in some new unique results. Five out of 11 CRC loci showed correlation to specific morphological features. Our data may be of help linking the genetic contribution and tumor morphology. These findings need to be replicated in other populations and in a larger patient-set, such as the ones from the ongoing GWAS on the low-risk study.

Linkage analysis in 121 CRC families from Sweden resulted in suggestive linkage to chromosome 9q in our high-risk families. This region has been published before in several studies. Genes in this region such as *TGFBR1* and *GALNT12* have been reported to have some impact on CRC development. The region includes several other genes and further studied are ongoing.

By sub-dividing our CRC families into colon- and rectal families specifically, we report two chromosomal regions with suggestive linkage to disease. A region on chromosome 6 was suggested for the colon families and one region on chromosome 18 for the rectal families. Exome sequencing of these regions helped us to identify common mutated variants within the families. Some of these variants have a potential role in the cancer development. A few interesting candidates are being tested in a larger material at the moment.

It has been difficult lately to map genes responsible for Mendelian diseases. It is more likely that what is left to find are rather low-risk variants giving a slightly increased disease risk alone or variants that act as modifiers who actually influence genes already known to be part of the CRC pathways.

Low-risk variants, SNPs, associated with CRC have been found through GWAS and meta-analysis. These variants give a very little relative risk alone but acting together they may have

an impact on disease development. However, even though there is a statistical significant correlation between these loci and CRC, the biological mechanism and functional characterization of the SNPs are mostly unknown. Until the overall contribution of low-risk SNPs is known, they cannot be used in diagnostics and many more loci and risk SNPs are expected from ongoing studies and meta-analysis.

During the time working with this thesis a lot has happened in the field of genetic research. It is now possible to use high throughput techniques such as NGS, which make it possible to sequence an entire human genome within short time. Still in its infancy, the techniques will be more accurate and the mapping of genetic causes for disease development will continue and hopefully result in additional knowledge to be used in diagnostics and disease prevention.

6 SAMMANFATTNING PÅ SVENSKA

Tarmcancer är den tredje vanligaste cancerformen i västvärlden. Män och kvinnor drabbas i samma omfattning. I Sverige drabbas årligen över 6000 personer av tarmcancer. Det finns kända livsstilsfaktorer som ökar risken att drabbas av tarmcancer, såsom övervikt, en kost rik på rött kött och animaliskt fett, rökning och en för hög alkoholkonsumtion men även förekomst av inflammatoriska tarmsjukdomar medför en ökad risk. En kost rik på frukt och fibrer samt fysisk aktivitet har en skyddande effekt mot cancerutveckling. En av de största riskerna att drabbas av tarmcancer är förekomsten av sjukdom hos nära släktingar. Att ha en nära släkting med sjukdomen fördubblar risken att insjukna medan två nära släktingar med tarmcancer ger en fyrdubbelt ökad risk.

Majoriteten av de som drabbas av tarmcancer är sporadiska patienter, dvs de har inga nära släktingar med sjukdomen. Studier har visat att upptill 35% av risken för att insjukna i tarmcancer antas bero på genetiska faktorer. Idag kan man förklara 10-15% av alla cancerfall med hjälp av kartlagda genetiska faktorer. Ca 5% kan förklaras av högrisk gener som orsakar ärfliga cancersyndrom som familjär colonpolypos (FAP) och Lynch Syndrom (LS). FAP kännetecknas av tidigt insjuknade pga hundtals polyper i tarmen. Även LS kännetecknas av tidigt insjuknande. Både FAP och LS patienter har ökad risk för andra tumörformer. De gener som orsakar FAP och LS är kända och genetisk testning och preventionsprogram utgör en viktig komponent för dessa patienter och deras släktingar. De resterande procenten kan härledas till gener som utgör en måttlig risk samt till s.k. lågrisk varianter som är associerade med en något ökad risk för sjukdomen. Flertalet familjer som har en till synes ärftlig form av tarmcancer bär inte på någon mutation i någon av de kartlagda generna. Det är därför viktigt att fortsatta arbetet med att lokalisera gener som är sjukdomsorsakande i dessa familjer.

Denna avhandling syftar till att genom associations- och kopplingsstudier identifiera nya genetiska sjukdomsorsakande faktorer.

I ***Studie I*** ville vi konfirmera publicerade risk-varianter (SNPs) som medför en liten riskökning för tarmcancer. Vid tiden för studien hade 11 riskloci rapporterats. Vi genotypade och jämförde dessa varianter hos 1786 svenska patienter och 1749 friska kontrollpersoner. Vi kunde bekräfta att fem av dessa 11 varianter är associerade med en ökad eller minskad risk för tarmcancer. De resterade varianterna påvisade likande resultat som tidigare rapporterats men utan statistiskt säkerställande. Vi gjorde genotyp-fenotyp korrelationsanalyser för kön, ålder vid insjuknande, familjehistoria av cancer och tumörens lokalisation. Vissa intressanta korrelationer kunde påvisas.

I ***Studie II*** undersökte vi om det finns någon korrelation mellan de genetiska varinterna ovan (samma som studie I) och morfologiska parametrar i tumörer. Vi analyserade 15 histopatologiska parameterar i 1572 patienter. Fem genetiska varianter kunde påvisas med statistiskt signifikant association till morfologiska parameterar.

I ***Studie III*** gjorde vi en hel-genomvid kopplingsanalys, för att identifiera gemensamma kromosomala regioner som nedärvs hos sjukdomsdrabbade individer. Vi använde 121 familjer med ärftlig tarmcancer, där FAP och LS hade uteslutits. Vi fann stöd för koppling till sjukdom i en region på kromosom 9. Denna region har tidigare identifierats som kandidatlocus och vår studie stödjer alltså teorin om sjukdomsorsakande gen/gener i denna region.

I ***Studie IV*** utfördes kopplingsanalys i två grupper: 32 familjer med ärftlig tjocktarmscancer och 66 familjer med ärftlig ändtarmscancer. Studier har tidigare visat att det kan vara olika genetiska och molekylära mekanismer beroende på om cancern uppkommer i tjocktarmen eller ändtarmen. Vi identifierade två kromosomala regioner med koppling till sjukdom, en region för tjocktarmsfamiljerna och en för ändtarmsfamiljerna. Vi sekvenserade exonen (som är de kodande delarna av gener) i dessa regioner i utvalda patienter. Flera intressanta kandidatgener identifierades.

Sammanfattningsvis har vi bekräftat tidigare publicerade genetiska varianter och loci associerade med tarmcancer hos svenska patienter. Vi har identifierat ett samband mellan genotyp-fenotyp och genotyp-morfologiska faktorer. Genom kopplingsanalys har vi identifierat nya kromosomala regioner som kan spela en roll i tarmcancerutveckling. Kandidatgener i dessa regioner studeras förnärvarande.

Det är viktigt att identifiera individer med förhöjd risk för cancer. Identifiering av nya riskgener leder till bättre diagnostik och prevention av tarmcancer.

7 ACKNOWLEDGEMENTS

This work has been carried out at the department of Molecular Medicine and Surgery, at Karolinska Institutet. I would like to express my sincere gratitude and appreciation to all of you who have helped and supported me during these years.

Thanks especially to,
Professor Annika Lindblom, my supervisor, thank you for giving me the opportunity to do a PhD. Your passion for science and your care for the patients are truly admirable. Thank you for always having time for my questions and discussions. I have learned a lot during these years thanks to you.

Simone Picelli, my first co-supervisor: first of all thank you for teaching me everything in the lab and for your patience with all my questions. Thank you also for all the "Italian cooking classes" and for all delicious food you cooked for me. For the good times we had in the lab but also outside CMM. I am so happy that research is going so well for you, out of anyone – you deserve that. Good luck in your academic career and with everything else!

Vinaykumar Kontham, my co-supervisor, thank you for being a great person to share a project with, for the valuable discussions and specially our linkage project that finally comes to an end. All my best to you and your family, I hope you will stay in Sweden.

Johanna Lundin, my co-supervisor, thank you for the nice introduction the very first day I came to the lab, for making me feel so welcome. Thank you for all valuable support in the projects and life in general.

Professor Magnus Nordenskjöld, for providing a very nice and qualified research climate at CMM and for inspiring discussions.

My colleagues in the cancer genetic group: **Barbro Werelius**, for teaching me how to extract the DNA, and our perfect companionship on all the sending's. Thank you for all your support and nice chats about everything. **Johanna Rantala**, for all your support and valuable talks, for being my friend both in- and outside work. **Jenny von Salome**, for all our interesting talks about life. Good luck with your research. **Rajeshwari Marikkannu**, for teaching me about Indian culture. I wish you all well and luck. **Tao Liu**, for being helpful, I wish you all luck at the clinic. **Jessada Thutkawkorapin**, for being so friendly. Good luck with you PhD! Thank you for your assistance with the exome data. **Petra Wessendorf**, for the good times we shared in the lab and outside work. **Christos Aravides**, for nice company. **Daniel Nilsson**, for your valuable input in NGS. **Xiang Jiao**, welcome to our group.

Anna Forsberg, thank you for nice collaboration, for all nice discussions and talks whenever we meet. **Sam Ghazi**, for nice companionship in the projects of ours. I am now looking forward reading your other book (not only your thesis).

The extended cancer genetic group: **Kicki, Erik, Peter, Emma, Bodil, Norma, Eva** for nice get-togethers.

All people at Klinisk Genetik, for always being helpful.

My friends in the writing room: **Izabella**, **Sofia**, **Anna S**, **Ellen**: thank you for your support and talks about everything. It has always been fun coming to work meeting you! **Clara**, **Said**, **Tianling**, **Florian**, **Stanley**, **Jia**, **Michaela**, **Fredrik**, **Husameldin**, **Maria**, for being so nice, creating a pleasant atmosphere in our writing room.

Other people at CMM:02: **Anna Bremer, Johanna Winberg, Josephine Wincent, Marie Meeths, Miriam Entesarian, Josefin Edner** and **Anna Aminoff**, for the good times we shared, at CMM and outside. **Ameli Norling, Malin Kvarnung, Agneta Nordenskjöld, Mårten Winge, Vasilios**

Zachariadis, Bianca Tesi, Wolfgang Hofmeister, Agne Liden, Nina Jäntti, Ellika Sahlin, Ingegerd Öfverholm, Alexandra Löfstedt and **Annika Sääf,** for sharing nice lunches, journal clubs and seminars.

Christina Nyström, **Anna-Lena Kastman, Sigrid Sahlén**, **Anki Thelander**, **Selim Sengul** for all your valuable help and support. Christina and Anna-Lena, for nice after works!

Lennart Helleday, **Jan-Erik Kaare** and the **IT department**, for handling computer problems. **Ann-Britt Wikström** and **Kerstin Florell**, for continuous support with all the admin.

My colleagues at the Unit for Bioentrepreneurship: I am so happy to call all of you friends. Thank you for always giving me energy, a smile or a hug. **Hanna, Lena, Madelen, Jesper, Therese, Linda, Ulrika, Jessica** and **Max** it has been so fun and rewarding working with you all! **Charlotta** and **Danielle,** for friendship. **Patrik B, Liisa, Francesca, Patrik** and **Pauline,** for having you around. **Bosse**, for teaching us tango! **Carl-Johan**, for letting me be a part of the UBE-team and for being such an amazing source of inspiration!

My dear friends outside science: **Karin**, **Mia**, **Jenny I**, **Jenny F**, **Lina, Sandra L**, **Sandra G**, **Ingrid** and **Per**, for your friendship. **Mira**, my soul-mate, for sharing the best research trip ever!

My Svärdsjö family and my Jämtland family, you are important and thank you for caring about us.

Mamma & **Pappa**, for all your love and support and for always being there when I really need you. Pappa, for being a role-model in many ways. Mamma, for giving so much of yourself! Brother **Johan**, for introducing me to science and KTH. For going you own way and for letting me visit you wherever in the world! **Maria**, for being so generous with your love and care, you are the best sister one can have!

Anders, for believing in me and for your never fading support. For your excellent excel- and writing scripts skills. But most of all, thank you for your love. **Arvid** och **Axel**, mina älskade busfrön, ni är mina hjärtan och jag är så stolt över att vara er mamma!

Finally, my deep appreciation to all patients and their families for participating in the studies.

This work has been supported by Karolinska doctoral funding, Nilsson-Ehle Donationerna, Anders Otto Swärds Stiftelse/Ulrika Eklunds Stiftelse, Cancerfonden, Vetenskapsrådet and the Stockholm County Council (ALF project).

8 REFERENCES

Aaltonen, L. A., Peltomaki, P., Leach, F. S., Sistonen, P., Pylkkanen, L., Mecklin, J. P., . . . Delachapelle, A. (1993). Clues to the Pathogenesis of Familial Colorectal-Cancer. *Science, 260*(5109), 812-816. doi: DOI 10.1126/science.8484121

Akiyama, Y., Sato, H., Yamada, T., Nagasaki, H., Tsuchiya, A., Abe, R., & Yuasa, Y. (1997). Germ-line mutation of the hMSH6/GTBP gene in an atypical hereditary nonpolyposis colorectal cancer kindred. *Cancer Res, 57*(18), 3920-3923.

Al-Tassan, N., Chmiel, N. H., Maynard, J., Fleming, N., Livingston, A. L., Williams, G. T., . . . Cheadle, J. P. (2002). Inherited variants of MYH associated with somatic G:C-->T:A mutations in colorectal tumors. *Nat Genet, 30*(2), 227-232. doi: 10.1038/ng828

Baker, S. J., Fearon, E. R., Nigro, J. M., Hamilton, S. R., Preisinger, A. C., Jessup, J. M., . . . Vogelstein, B. (1989). Chromosome 17 deletions and p53 gene mutations in colorectal carcinomas. *Science, 244*(4901), 217-221.

Baker, S. J., Markowitz, S., Fearon, E. R., Willson, J. K. V., & Vogelstein, B. (1990). Suppression of Human Colorectal-Carcinoma Cell-Growth by Wild-Type-P53. *Science, 249*(4971), 912-915. doi: DOI 10.1126/science.2144057

Bellovin, D. I., Simpson, K. J., Danilov, T., Maynard, E., Rimm, D. L., Oettgen, P., & Mercurio, A. M. (2006). Reciprocal regulation of RhoA and RhoC characterizes the EMT and identifies RhoC as a prognostic marker of colon carcinoma. *Oncogene, 25*(52), 6959-6967. doi: DOI 10.1038/sj.onc.1209682

Bera, T. K., Zimonjic, D. B., Popescu, N. C., Sathyanarayana, B. K., Kumar, V., Lee, B., & Pastan, I. (2002). POTE, a highly homologous gene family located on numerous chromosomes and expressed in prostate, ovary, testis, placenta, and prostate cancer. *Proc Natl Acad Sci U S A, 99*(26), 16975-16980. doi: 10.1073/pnas.262655399

Bodmer, W. F., Bailey, C. J., Bodmer, J., Bussey, H. J., Ellis, A., Gorman, P., . . . et al. (1987). Localization of the gene for familial adenomatous polyposis on chromosome 5. *Nature, 328*(6131), 614-616. doi: 10.1038/328614a0

Bodmer, W. F., Bailey, C. J., Bodmer, J., Bussey, H. J. R., Ellis, A., Gorman, P., . . . Spurr, N. K. (1987). Localization of the Gene for Familial Adenomatous Polyposis on Chromosome-5. *Nature, 328*(6131), 614-616. doi: Doi 10.1038/328614a0

Bogaert, J., & Prenen, H. (2014). Molecular genetics of colorectal cancer. *Ann Gastroenterol, 27*(1), 9-14.

Bos, J. L. (1989). ras oncogenes in human cancer: a review. *Cancer Res, 49*(17), 4682-4689.

Broderick, P., Carvajal-Carmona, L., Pittman, A. M., Webb, E., Howarth, K., Rowan, A., . . . Houlston, R. S. (2007). A genome-wide association study shows that common alleles of SMAD7 influence colorectal cancer risk. *Nat Genet, 39*(11), 1315-1317. doi: 10.1038/ng.2007.18

Bronner, C. E., Baker, S. M., Morrison, P. T., Warren, G., Smith, L. G., Lescoe, M. K., . . . et al. (1994). Mutation in the DNA mismatch repair gene homologue hMLH1 is associated with hereditary non-polyposis colon cancer. *Nature, 368*(6468), 258-261. doi: 10.1038/368258a0

Bufill, J. A. (1990). Colorectal cancer: evidence for distinct genetic categories based on proximal or distal tumor location. *Ann Intern Med, 113*(10), 779-788.

Caporale, A., Vestri, A. R., Benvenuto, E., Mariotti, M., Cosenza, U. M., Scarpini, M., . . . Angelico, F. (2005). Is desmoplasia a protective factor for survival in patients with colorectal carcinoma? *Clinical Gastroenterology and Hepatology, 3*(4), 370-375. doi: Doi 10.1053/S1542-3565(04)00674-3

Cho, E., Smith-Warner, S. A., Ritz, J., van den Brandt, P. A., Colditz, G. A., Folsom, A. R., . . . Hunter, D. J. (2004). Alcohol intake and colorectal cancer: a pooled analysis of 8 cohort studies. *Ann Intern Med, 140*(8), 603-613.

Cicek, M. S., Cunningham, J. M., Fridley, B. L., Serie, D. J., Bamlet, W. R., Diergaarde, B., . . . Colon, C. F. R. (2012). Colorectal cancer linkage on chromosomes 4q21, 8q13, 12q24, and 15q22. *PLoS One, 7*(5), e38175. doi: 10.1371/journal.pone.0038175

Cicek, M. S., Slager, S. L., Achenbach, S. J., French, A. J., Blair, H. E., Fink, S. R., . . . Thibodeau, S. N. (2009). Functional and clinical significance of variants localized to 8q24 in colon cancer. *Cancer Epidemiol Biomarkers Prev, 18*(9), 2492-2500. doi: 10.1158/1055-9965.EPI-09-0362

Compton, C. C. (2003). Colorectal carcinoma: diagnostic, prognostic, and molecular features. *Mod Pathol, 16*(4), 376-388. doi: 10.1097/01.MP.0000062859.46942.93

Cunningham, J. M., Christensen, E. R., Tester, D. J., Kim, C. Y., Roche, P. C., Burgart, L. J., & Thibodeau, S. N. (1998). Hypermethylation of the hMLH1 promoter in colon cancer with microsatellite instability. *Cancer Research, 58*(15), 3455-3460.

Cuzick, J., Thorat, M. A., Bosetti, C., Brown, P. H., Burn, J., Cook, N. R., . . . Umar, A. (2014). Estimates of benefits and harms of prophylactic use of aspirin in the general population. *Ann Oncol*. doi: 10.1093/annonc/mdu225

de Jong, M. M., Nolte, I. M., Meerman, G. J. T., van der Graaf, W. T. A., de Vries, E. G. E., Sijmons, R. H., . . . Kleibeuker, J. H. (2002). Low-penetrance genes and their involvement in colorectal cancer susceptibility. *Cancer Epidemiology Biomarkers & Prevention, 11*(11), 1332-1352.

de Vos tot Nederveen Cappel, W. H., Jarvinen, H. J., Lynch, P. M., Engel, C., Mecklin, J. P., & Vasen, H. F. (2013). Colorectal surveillance in Lynch syndrome families. *Fam Cancer, 12*(2), 261-265. doi: 10.1007/s10689-013-9631-1

Delattre, O., Olschwang, S., Law, D. J., Melot, T., Remvikos, Y., Salmon, R. J., . . . Thomas, G. (1989). Multiple genetic alterations in distal and proximal colorectal cancer. *Lancet, 2*(8659), 353-356.

Din, F. V., Theodoratou, E., Farrington, S. M., Tenesa, A., Barnetson, R. A., Cetnarskyj, R., . . . Dunlop, M. G. (2010). Effect of aspirin and NSAIDs on risk and survival from colorectal cancer. *Gut, 59*(12), 1670-1679. doi: 10.1136/gut.2009.203000

Djureinovic, T., Skoglund, J., Vandrovcova, J., Zhou, X. L., Kalushkova, A., Iselius, L., & Lindblom, A. (2006). A genome wide linkage analysis in Swedish families with hereditary non-familial adenomatous polyposis/non-hereditary non-polyposis colorectal cancer. *Gut, 55*(3), 362-366. doi: 10.1136/gut.2005.075333

Drummond, J. T., Li, G. M., Longley, M. J., & Modrich, P. (1995). Isolation of an hMSH2-p160 heterodimer that restores DNA mismatch repair to tumor cells. *Science, 268*(5219), 1909-1912.

Dunlop, M. G., Dobbins, S. E., Farrington, S. M., Jones, A. M., Palles, C., Whiffin, N., . . . Houlston, R. S. (2012). Common variation near CDKN1A, POLD3 and SHROOM2 influences colorectal cancer risk. *Nat Genet, 44*(7), 770-776. doi: 10.1038/ng.2293

Eppert, K., Scherer, S. W., Ozcelik, H., Pirone, R., Hoodless, P., Kim, H., . . . Attisano, L. (1996). MADR2 maps to 18q21 and encodes a TGFbeta-regulated MAD-related protein that is functionally mutated in colorectal carcinoma. *Cell, 86*(4), 543-552.

Esteban-Jurado, C., Garre, P., Vila, M., Lozano, J. J., Pristoupilova, A., Beltran, S., . . . Castellvi-Bel, S. (2014). New genes emerging for colorectal cancer predisposition. *World J Gastroenterol, 20*(8), 1961-1971. doi: 10.3748/wjg.v20.i8.1961

Esteban-Jurado, C., Vila-Casadesus, M., Garre, P., Lozano, J. J., Pristoupilova, A., Beltran, S., . . . Castellvi-Bel, S. (2014). Whole-exome sequencing identifies rare pathogenic variants in new predisposition genes for familial colorectal cancer. *Genet Med*. doi: 10.1038/gim.2014.89

Fairley, T. L., Cardinez, C. J., Martin, J., Alley, L., Friedman, C., Edwards, B., & Jamison, P. (2006). Colorectal cancer in US adults younger than 50 years of age, 1998-2001. *Cancer, 107*(5), 1153-1161. doi: Doi 10.1002/Cncr.22012

Fearnhead, N. S., Winney, B., & Bodmer, W. F. (2005). Rare variant hypothesis for multifactorial inheritance: susceptibility to colorectal adenomas as a model. *Cell Cycle, 4*(4), 521-525.

Fearon, E. R., & Vogelstein, B. (1990). A Genetic Model for Colorectal Tumorigenesis. *Cell, 61*(5), 759-767. doi: Doi 10.1016/0092-8674(90)90186-I

Fishel, R., Lescoe, M. K., Rao, M. R., Copeland, N. G., Jenkins, N. A., Garber, J., . . . Kolodner, R. (1993). The human mutator gene homolog MSH2 and its association with hereditary nonpolyposis colon cancer. *Cell, 75*(5), 1027-1038.

Frattini, M., Balestra, D., Pilotti, S., Bertario, L., & Pierotti, M. A. (2003). Tumor location and detection of k-ras mutations in stool from colorectal cancer patients. *J Natl Cancer Inst, 95*(1), 72-73; author reply 73.

Galiatsatos, P., & Foulkes, W. D. (2006). Familial adenomatous polyposis. *Am J Gastroenterol, 101*(2), 385-398. doi: 10.1111/j.1572-0241.2006.00375.x

Ghazi, S., Lindforss, U., Lindberg, G., Berg, E., Lindblom, A., Papadogiannakis, N., & Low-Risk Colorectal Cancer Study, Group. (2012). Analysis of colorectal cancer morphology in relation to sex, age, location, and family history. *J Gastroenterol, 47*(6), 619-634. doi: 10.1007/s00535-011-0520-9

Graham, D. M., & Appelman, H. D. (1990). Crohn's-like lymphoid reaction and colorectal carcinoma: a potential histologic prognosticator. *Mod Pathol, 3*(3), 332-335.

Grizzi, F., Celesti, G., Basso, G., & Laghi, L. (2012). Tumor budding as a potential histopathological biomarker in colorectal cancer: Hype or hope? *World Journal of Gastroenterology, 18*(45), 6532-6536. doi: DOI 10.3748/wjg.v18.i45.6532

Groden, J., Thliveris, A., Samowitz, W., Carlson, M., Gelbert, L., Albertsen, H., . . . et al. (1991). Identification and characterization of the familial adenomatous polyposis coli gene. *Cell, 66*(3), 589-600.

Guda, K., Moinova, H., He, J., Jamison, O., Ravi, L., Natale, L., . . . Markowitz, S. D. (2009). Inactivating germ-line and somatic mutations in polypeptide N-acetylgalactosaminyltransferase 12 in human colon cancers. *Proc Natl Acad Sci U S A, 106*(31), 12921-12925. doi: 10.1073/pnas.0901454106

Haggar, F. A., & Boushey, R. P. (2009). Colorectal cancer epidemiology: incidence, mortality, survival, and risk factors. *Clin Colon Rectal Surg, 22*(4), 191-197. doi: 10.1055/s-0029-1242458

Haiman, C. A., Le Marchand, L., Yamamato, J., Stram, D. O., Sheng, X., Kolonel, L. N., . . . Henderson, B. E. (2007). A common genetic risk factor for colorectal and prostate cancer. *Nat Genet, 39*(8), 954-956. doi: 10.1038/ng2098

Hemminki, A., Markie, D., Tomlinson, I., Avizienyte, E., Roth, S., Loukola, A., . . . Aaltonen, L. A. (1998). A serine/threonine kinase gene defective in Peutz-Jeghers syndrome. *Nature, 391*(6663), 184-187. doi: 10.1038/34432

Hemminki, A., Tomlinson, I., Markie, D., Jarvinen, H., Sistonen, P., Bjorkqvist, A. M., . . . Aaltonen, L. A. (1997). Localization of a susceptibility locus for Peutz-Jeghers syndrome to 19p using comparative genomic hybridization and targeted linkage analysis. *Nature Genetics, 15*(1), 87-90. doi: Doi 10.1038/Ng0197-87

Hendriks, Y. M., Jagmohan-Changur, S., van der Klift, H. M., Morreau, H., van Puijenbroek, M., Tops, C., . . . Wijnen, J. T. (2006). Heterozygous mutations in PMS2 cause hereditary nonpolyposis colorectal carcinoma (Lynch syndrome). *Gastroenterology, 130*(2), 312-322. doi: 10.1053/j.gastro.2005.10.052

Herrera, L., Kakati, S., Gibas, L., Pietrzak, E., & Sandberg, A. A. (1986). Gardner syndrome in a man with an interstitial deletion of 5q. *Am J Med Genet, 25*(3), 473-476. doi: 10.1002/ajmg.1320250309

Houlston, R. S., Cheadle, J., Dobbins, S. E., Tenesa, A., Jones, A. M., Howarth, K., . . . Tomlinson, I. P. (2010). Meta-analysis of three genome-wide association studies identifies susceptibility loci for colorectal cancer at 1q41, 3q26.2, 12q13.13 and 20q13.33. *Nat Genet, 42*(11), 973-977. doi: 10.1038/ng.670

Houlston, R. S., Webb, E., Broderick, P., Pittman, A. M., Di Bernardo, M. C., Lubbe, S., . . . Dunlop, M. G. (2008). Meta-analysis of genome-wide association data identifies four new susceptibility loci for colorectal cancer. *Nat Genet, 40*(12), 1426-1435. doi: 10.1038/ng.262

Howe, J. R., Bair, J. L., Sayed, M. G., Anderson, M. E., Mitros, F. A., Petersen, G. M., . . . Vogelstein, B. (2001). Germline mutations of the gene encoding bone morphogenetic protein receptor 1A in juvenile polyposis. *Nat Genet, 28*(2), 184-187. doi: 10.1038/88919

Howe, J. R., Roth, S., Ringold, J. C., Summers, R. W., Jarvinen, H. J., Sistonen, P., . . . Aaltonen, L. A. (1998). Mutations in the SMAD4/DPC4 gene in juvenile polyposis. *Science, 280*(5366), 1086-1088.

Jaeger, E., Leedham, S., Lewis, A., Segditsas, S., Becker, M., Cuadrado, P. R., . . . Tomlinson, I. (2012). Hereditary mixed polyposis syndrome is caused by a 40-kb upstream duplication that leads to increased and ectopic expression of the BMP antagonist GREM1. *Nature Genetics, 44*(6), 699-U127. doi: Doi 10.1038/Ng.2263

Jaeger, E., Webb, E., Howarth, K., Carvajal-Carmona, L., Rowan, A., Broderick, P., . . . Tomlinson, I. (2008). Common genetic variants at the CRAC1 (HMPS) locus on chromosome 15q13.3 influence colorectal cancer risk. *Nat Genet, 40*(1), 26-28. doi: 10.1038/ng.2007.41

Jass, J. R., Iino, H., Ruszkiewicz, A., Painter, D., Solomon, M. J., Koorey, D. J., . . . Leggett, B. A. (2000). Neoplastic progression occurs through mutator pathways in hyperplastic polyposis of the colorectum. *Gut, 47*(1), 43-49. doi: Doi 10.1136/Gut.47.1.43

Jass, J. R., Love, S. B., & Northover, J. M. A. (1987). A New Prognostic Classification of Rectal-Cancer. *Lancet, 1*(8545), 1303-1306.

Jessurun, J., Romero-Guadarrama, M., & Manivel, J. C. (1999). Medullary adenocarcinoma of the colon: Clinicopathologic study of 11 cases. *Human Pathology, 30*(7), 843-848. doi: Doi 10.1016/S0046-8177(99)90146-6

Jin, G. F., Ma, H. X., Wu, C., Dai, J. C., Zhang, R. Y., Shi, Y. Y., . . . Shen, H. B. (2012). Genetic Variants at 6p21.1 and 7p15.3 Are Associated with Risk of Multiple Cancers in Han Chinese. *American Journal of Human Genetics, 91*(5), 928-934. doi: DOI 10.1016/j.ajhg.2012.09.009

Johansson, Elizabeth. (2013). *Cancer i siffror 2013 : populärvetenskapliga fakta om cancer*. Stockholm: Cancerfonden :.

Johns, L. E., & Houlston, R. S. (2001). A systematic review and meta-analysis of familial colorectal cancer risk. *Am J Gastroenterol, 96*(10), 2992-3003. doi: 10.1111/j.1572-0241.2001.04677.x

Jones, P. A., & Laird, P. W. (1999). Cancer epigenetics comes of age. *Nat Genet, 21*(2), 163-167. doi: 10.1038/5947

Kapiteijn, E., Liefers, G. J., Los, L. C., Kranenbarg, E. K., Hermans, J., Tollenaar, R. A., . . . van Krieken, J. H. (2001). Mechanisms of oncogenesis in colon versus rectal cancer. *J Pathol, 195*(2), 171-178. doi: 10.1002/path.918

Kemp, Z., Carvajal-Carmona, L., Spain, S., Barclay, E., Gorman, M., Martin, L., . . . ColoRectal tumour Gene Identification Study, Consortium. (2006). Evidence for a colorectal cancer susceptibility locus on chromosome 3q21-q24 from a high-density SNP genome-wide linkage scan. *Hum Mol Genet, 15*(19), 2903-2910. doi: 10.1093/hmg/ddl231

Kemp, Z. E., Carvajal-Carmona, L. G., Barclay, E., Gorman, M., Martin, L., Wood, W., . . . Colorectal Tumour Gene Identification Study, Consortium. (2006). Evidence of linkage to chromosome 9q22.33 in colorectal cancer kindreds from the United Kingdom. *Cancer Res, 66*(10), 5003-5006. doi: 10.1158/0008-5472.CAN-05-4074

Kilpivaara, O., Alhopuro, P., Vahteristo, P., Aaltonen, L. A., & Nevanlinna, H. (2006). CHEK2 I157T associates with familial and sporadic colorectal cancer. *J Med Genet, 43*(7), e34. doi: 10.1136/jmg.2005.038331

Kim, J. S., Crooks, H., Dracheva, T., Nishanian, T. G., Singh, B., Jen, J., & Waldman, T. (2002). Oncogenic beta-catenin is required for bone morphogenetic protein 4 expression in human cancer cells. *Cancer Research, 62*(10), 2744-2748.

Kim, S. S., Shago, M., Kaustov, L., Boutros, P. C., Clendening, J. W., Sheng, Y., . . . Penn, L. Z. (2007). CUL7 is a novel antiapoptotic oncogene. *Cancer Research, 67*(20), 9616-9622. doi: Doi 10.1158/0008-5472.Can-07-0644

Kinzler, K. W., & Vogelstein, B. (1996). Lessons from hereditary colorectal cancer. *Cell, 87*(2), 159-170.

Knudson, A. G. (1971). Mutation and Cancer - Statistical Study of Retinoblastoma. *Proceedings of the National Academy of Sciences of the United States of America, 68*(4), 820-&. doi: DOI 10.1073/pnas.68.4.820

Kontham, V., von Holst, S., & Lindblom, A. (2013). Linkage analysis in familial non-lynch syndrome colorectal cancer families from sweden. *PLoS One, 8*(12), e83936. doi: 10.1371/journal.pone.0083936

Koo, J. H., & Leong, R. W. (2010). Sex differences in epidemiological, clinical and pathological characteristics of colorectal cancer. *J Gastroenterol Hepatol, 25*(1), 33-42. doi: 10.1111/j.1440-1746.2009.05992.x

Kranenburg, O. (2005). The KRAS oncogene: past, present, and future. *Biochim Biophys Acta, 1756*(2), 81-82. doi: 10.1016/j.bbcan.2005.10.001

Kruglyak, L., Daly, M. J., ReeveDaly, M. P., & Lander, E. S. (1996). Parametric and nonparametric linkage analysis: A unified multipoint approach. *American Journal of Human Genetics, 58*(6), 1347-1363.

Kuismanen, S. A., Holmberg, M. T., Salovaara, R., de la Chapelle, A., & Peltomaki, P. (2000). Genetic and epigenetic modification of MLH1 accounts for a major share of microsatellite-unstable colorectal cancers. *Am J Pathol, 156*(5), 1773-1779. doi: 10.1016/S0002-9440(10)65048-1

Lagerstedt Robinson, K., Liu, T., Vandrovcova, J., Halvarsson, B., Clendenning, M., Frebourg, T., . . . Lindblom, A. (2007). Lynch syndrome (hereditary nonpolyposis colorectal cancer) diagnostics. *J Natl Cancer Inst, 99*(4), 291-299. doi: 10.1093/jnci/djk051

Laken, S. J., Petersen, G. M., Gruber, S. B., Oddoux, C., Ostrer, H., Giardiello, F. M., . . . Vogelstein, B. (1997). Familial colorectal cancer in Ashkenazim due to a hypermutable tract in APC. *Nat Genet, 17*(1), 79-83. doi: 10.1038/ng0997-79

Langner, C., Harbaum, L., Pollheimer, M. J., Kornprat, P., Lindtner, R. A., Schlemmer, A., . . . Rehak, P. (2012). Mucinous differentiation in colorectal cancer--indicator of poor prognosis? *Histopathology, 60*(7), 1060-1072. doi: 10.1111/j.1365-2559.2011.04155.x

Leach, F. S., Nicolaides, N. C., Papadopoulos, N., Liu, B., Jen, J., Parsons, R., . . . et al. (1993). Mutations of a mutS homolog in hereditary nonpolyposis colorectal cancer. *Cell, 75*(6), 1215-1225.

Leppert, M., Dobbs, M., Scambler, P., O'Connell, P., Nakamura, Y., Stauffer, D., . . . et al. (1987). The gene for familial polyposis coli maps to the long arm of chromosome 5. *Science, 238*(4832), 1411-1413.

Leslie, A., Carey, F. A., Pratt, N. R., & Steele, R. J. (2002). The colorectal adenoma-carcinoma sequence. *Br J Surg, 89*(7), 845-860. doi: 10.1046/j.1365-2168.2002.02120.x

Lichtenstein, P., Holm, N. V., Verkasalo, P. K., Iliadou, A., Kaprio, J., Koskenvuo, M., . . . Hemminki, K. (2000). Environmental and heritable factors in the causation of cancer - Analyses of cohorts of twins from Sweden, Denmark, and Finland. *New England Journal of Medicine, 343*(2), 78-85. doi: Doi 10.1056/Nejm200007133430201

Lindblom, A., Tannergard, P., Werelius, B., & Nordenskjold, M. (1993). Genetic mapping of a second locus predisposing to hereditary non-polyposis colon cancer. *Nat Genet, 5*(3), 279-282. doi: 10.1038/ng1193-279

Lindgren, G., Liljegren, A., Jaramillo, E., Rubio, C., & Lindblom, A. (2002). Adenoma prevalence and cancer risk in familial non-polyposis colorectal cancer. *Gut, 50*(2), 228-234.

Lindor, N. M., Rabe, K., Petersen, G. M., Haile, R., Casey, G., Baron, J., . . . Seminara, D. (2005). Lower cancer incidence in Amsterdam-I criteria families without mismatch repair deficiency: familial colorectal cancer type X. *JAMA, 293*(16), 1979-1985. doi: 10.1001/jama.293.16.1979

Lipton, L., Halford, S. E., Johnson, V., Novelli, M. R., Jones, A., Cummings, C., . . . Tomlinson, I. P. (2003). Carcinogenesis in MYH-associated polyposis follows a distinct genetic pathway. *Cancer Res, 63*(22), 7595-7599.

Lubbe, S. J., Di Bernardo, M. C., Chandler, I. P., & Houlston, R. S. (2009). Clinical implications of the colorectal cancer risk associated with MUTYH mutation. *J Clin Oncol, 27*(24), 3975-3980. doi: 10.1200/JCO.2008.21.6853

Lynch, H. T., & de la Chapelle, A. (2003). Hereditary colorectal cancer. *N Engl J Med, 348*(10), 919-932. doi: 10.1056/NEJMra012242

Lynch, H. T., Shaw, M. W., Magnuson, C. W., Larsen, A. L., & Krush, A. J. (1966). Hereditary factors in cancer. Study of two large midwestern kindreds. *Arch Intern Med, 117*(2), 206-212.

Marsh, D. J., Coulon, V., Lunetta, K. L., Rocca-Serra, P., Dahia, P. L., Zheng, Z., . . . et al. (1998). Mutation spectrum and genotype-phenotype analyses in Cowden disease and Bannayan-Zonana syndrome, two hamartoma syndromes with germline PTEN mutation. *Hum Mol Genet, 7*(3), 507-515.

McGarrity, T. J., Kulin, H. E., & Zaino, R. J. (2000). Peutz-Jeghers syndrome. *Am J Gastroenterol, 95*(3), 596-604. doi: 10.1111/j.1572-0241.2000.01831.x

Meguid, R. A., Slidell, M. B., Wolfgang, C. L., Chang, D. C., & Ahuja, N. (2008). Is there a difference in survival between right- versus left-sided colon cancers? *Ann Surg Oncol, 15*(9), 2388-2394. doi: 10.1245/s10434-008-0015-y

Meldrum, C., Doyle, M. A., & Tothill, R. W. (2011). Next-generation sequencing for cancer diagnostics: a practical perspective. *Clin Biochem Rev, 32*(4), 177-195.

Meng, L., Sefah, K., O'Donoghue, M. B., Zhu, G., Shangguan, D., Noorali, A., . . . Tan, W. (2010). Silencing of PTK7 in colon cancer cells: caspase-10-dependent apoptosis via mitochondrial pathway. *PLoS One, 5*(11), e14018. doi: 10.1371/journal.pone.0014018

Minsky, B., & Mies, C. (1989). The clinical significance of vascular invasion in colorectal cancer. *Dis Colon Rectum, 32*(9), 794-803.

Miyaki, M., Konishi, M., Tanaka, K., Kikuchi-Yanoshita, R., Muraoka, M., Yasuno, M., . . . Mori, T. (1997). Germline mutation of MSH6 as the cause of hereditary nonpolyposis colorectal cancer. *Nat Genet, 17*(3), 271-272. doi: 10.1038/ng1197-271

Miyoshi, Y., Nagase, H., Ando, H., Horii, A., Ichii, S., Nakatsuru, S., . . . Nakamura, Y. (1992). Somatic mutations of the APC gene in colorectal tumors: mutation cluster region in the APC gene. *Hum Mol Genet, 1*(4), 229-233.

Morton, N. E. (1955). Sequential tests for the detection of linkage. *Am J Hum Genet, 7*(3), 277-318.

Nicolaides, N. C., Papadopoulos, N., Liu, B., Wei, Y. F., Carter, K. C., Ruben, S. M., . . . et al. (1994). Mutations of two PMS homologues in hereditary nonpolyposis colon cancer. *Nature, 371*(6492), 75-80. doi: 10.1038/371075a0

Olsson, L., & Lindblom, A. (2003). Family history of colorectal cancer in a Sweden county. *Fam Cancer, 2*(2), 87-93.

Palles, C., Cazier, J. B., Howarth, K. M., Domingo, E., Jones, A. M., Broderick, P., . . . Tomlinson, I. (2013). Germline mutations affecting the proofreading domains of POLE and POLD1 predispose to colorectal adenomas and carcinomas. *Nat Genet, 45*(2), 136-144. doi: 10.1038/ng.2503

Parsons, R., Li, G. M., Longley, M. J., Fang, W. H., Papadopoulos, N., Jen, J., . . . Modrich, P. (1993). Hypermutability and mismatch repair deficiency in RER+ tumor cells. *Cell, 75*(6), 1227-1236.

Pearson, T. A., & Manolio, T. A. (2008). How to interpret a genome-wide association study. *Jama-Journal of the American Medical Association, 299*(11), 1335-1344. doi: DOI 10.1001/jama.299.11.1335

Pei, X. H., Bai, F., Li, Z. J., Smith, M. D., Whitewolf, G., Jin, R., & Xiong, Y. (2011). Cytoplasmic CUL9/PARC Ubiquitin Ligase Is a Tumor Suppressor and Promotes p53-Dependent Apoptosis. *Cancer Research, 71*(8), 2969-2977. doi: Doi 10.1158/0008-5472.Can-10-4300

Peltomaki, P. (2001). Deficient DNA mismatch repair: a common etiologic factor for colon cancer. *Hum Mol Genet, 10*(7), 735-740.

Peltomaki, P., Aaltonen, L. A., Sistonen, P., Pylkkanen, L., Mecklin, J. P., Jarvinen, H., . . . et al. (1993). Genetic mapping of a locus predisposing to human colorectal cancer. *Science, 260*(5109), 810-812.

Peltomaki, P., & Vasen, H. (2004). Mutations associated with HNPCC predisposition -- Update of ICG-HNPCC/INSiGHT mutation database. *Dis Markers, 20*(4-5), 269-276.

Picelli, S., Vandrovcova, J., Jones, S., Djureinovic, T., Skoglund, J., Zhou, X. L., . . . Lindblom, A. (2008). Genome-wide linkage scan for colorectal cancer susceptibility genes supports linkage to chromosome 3q. *BMC Cancer, 8*, 87. doi: 10.1186/1471-2407-8-87

Picelli, S., Von Holst, S., & Wessendorf, P. (2009). The continuing search for predisposing colorectal cancer variants. *Cancer Genomics Proteomics, 6*(6), 305-316.

Pittman, A. M., Naranjo, S., Jalava, S. E., Twiss, P., Ma, Y., Olver, B., . . . Houlston, R. S. (2010). Allelic Variation at the 8q23.3 Colorectal Cancer Risk Locus Functions as a Cis-Acting Regulator of EIF3H. *Plos Genetics, 6*(9). doi: ARTN e1001126

DOI 10.1371/journal.pgen.1001126

Pittman, A. M., Naranjo, S., Webb, E., Broderick, P., Lips, E. H., van Wezel, T., . . . Houlston, R. S. (2009). The colorectal cancer risk at 18q21 is caused by a novel variant altering SMAD7 expression. *Genome Res, 19*(6), 987-993. doi: 10.1101/gr.092668.109

Pomerantz, M. M., Ahmadiyeh, N., Jia, L., Herman, P., Verzi, M. P., Doddapaneni, H., . . . Freedman, M. L. (2009). The 8q24 cancer risk variant rs6983267 shows long-range interaction with MYC in colorectal cancer. *Nat Genet, 41*(8), 882-884. doi: 10.1038/ng.403

Powell, S. M., Zilz, N., Beazer-Barclay, Y., Bryan, T. M., Hamilton, S. R., Thibodeau, S. N., . . . Kinzler, K. W. (1992). APC mutations occur early during colorectal tumorigenesis. *Nature, 359*(6392), 235-237. doi: 10.1038/359235a0

Poynter, J. N., Figueiredo, J. C., Conti, D. V., Kennedy, K., Gallinger, S., Siegmund, K. D., . . . Colon, C. F. R. (2007). Variants on 9p24 and 8q24 are associated with risk of colorectal cancer: results from the Colon Cancer Family Registry. *Cancer Res, 67*(23), 11128-11132. doi: 10.1158/0008-5472.CAN-07-3239

Richards, F. M., McKee, S. A., Rajpar, M. H., Cole, T. R., Evans, D. G., Jankowski, J. A., . . . Maher, E. R. (1999). Germline E-cadherin gene (CDH1) mutations predispose to familial gastric cancer and colorectal cancer. *Hum Mol Genet, 8*(4), 607-610.

Roberts, A., Nancarrow, D., Clendenning, M., Buchanan, D. D., Jenkins, M. A., Duggan, D., . . . Young, J. P. (2011). Linkage to chromosome 2q32.2-q33.3 in familial serrated neoplasia (Jass syndrome). *Fam Cancer, 10*(2), 245-254. doi: 10.1007/s10689-010-9408-8

Rodriguez-Bigas, M. A., Boland, C. R., Hamilton, S. R., Henson, D. E., Jass, J. R., Khan, P. M., . . . Srivastava, S. (1997). A National Cancer Institute Workshop on Hereditary Nonpolyposis Colorectal Cancer Syndrome: meeting highlights and Bethesda guidelines. *J Natl Cancer Inst, 89*(23), 1758-1762.

Rohlin, A., Zagoras, T., Nilsson, S., Lundstam, U., Wahlstrom, J., Hulten, L., . . . Nordling, M. (2014). A mutation in POLE predisposing to a multi-tumour phenotype. *Int J Oncol, 45*(1), 77-81. doi: 10.3892/ijo.2014.2410

Sari, I., Berberoglu, B., Ozkara, E., Oztuzcu, S., Camci, C., & Demiryurek, A. T. (2013). Role of rho-kinase gene polymorphisms and protein expressions in colorectal cancer development. *Pathobiology, 80*(3), 138-145. doi: 10.1159/000341395

Saunders, I. W., Ross, J., Macrae, F., Young, G. P., Blanco, I., Brohede, J., . . . Hannan, G. N. (2012). Evidence of linkage to chromosomes 10p15.3-p15.1, 14q24.3-q31.1 and 9q33.3-q34.3 in non-syndromic colorectal cancer families. *Eur J Hum Genet, 20*(1), 91-96. doi: 10.1038/ejhg.2011.149

Schwartz, R. A., & Torre, D. P. (1995). The Muir-Torre syndrome: a 25-year retrospect. *J Am Acad Dermatol, 33*(1), 90-104.

Shen, L., Toyota, M., Kondo, Y., Lin, E., Zhang, L., Guo, Y., . . . Issa, J. P. (2007). Integrated genetic and epigenetic analysis identifies three different subclasses of colon cancer. *Proc Natl Acad Sci U S A, 104*(47), 18654-18659. doi: 10.1073/pnas.0704652104

Shibata, D., Peinado, M. A., Ionov, Y., Malkhosyan, S., & Perucho, M. (1994). Genomic instability in repeated sequences is an early somatic event in colorectal tumorigenesis that persists after transformation. *Nat Genet, 6*(3), 273-281. doi: 10.1038/ng0394-273

Skoglund, J., Djureinovic, T., Zhou, X. L., Vandrovcova, J., Renkonen, E., Iselius, L., . . . Lindblom, A. (2006). Linkage analysis in a large Swedish family supports the presence of a susceptibility locus for adenoma and colorectal cancer on chromosome 9q22.32-31.1. *J Med Genet, 43*(2), e7. doi: 10.1136/jmg.2005.033928

Skoglund, J., Song, B., Dalen, J., Dedorson, S., Edler, D., Hjern, F., . . . Lindblom, A. (2007). Lack of an association between the TGFBR1*6A variant and colorectal cancer risk. *Clin Cancer Res, 13*(12), 3748-3752. doi: 10.1158/1078-0432.CCR-06-2865

Slattery, M. L., Potter, J. D., Curtin, K., Edwards, S., Ma, K. N., Anderson, K., . . . Samowitz, W. S. (2001). Estrogens reduce and withdrawal of estrogens increase risk of microsatellite instability-positive colon cancer. *Cancer Res, 61*(1), 126-130.

Socialstyrelsen. (2012). *Cancer incidence in Sweden 2011 = Cancerförekomst i Sverige 2011*. Stockholm: Socialstyrelsen.

Sweetser, S., Smyrk, T. C., & Sugumar, A. (2011). Serrated polyps: critical precursors to colorectal cancer. *Expert Rev Gastroenterol Hepatol, 5*(5), 627-635. doi: 10.1586/egh.11.67

Takahashi, M., Fujita, M., Furukawa, Y., Hamamoto, R., Shimokawa, T., Miwa, N., . . . Nakamura, Y. (2002). Isolation of a novel human gene, APCDD1, as a direct target of the beta-Catenin/T-cell factor 4 complex with probable involvement in colorectal carcinogenesis. *Cancer Res, 62*(20), 5651-5656.

Takatsuno, Y., Mimori, K., Yamamoto, K., Sato, T., Niida, A., Inoue, H., . . . Mori, M. (2013). The rs6983267 SNP is associated with MYC transcription efficiency, which promotes progression and worsens prognosis of colorectal cancer. *Ann Surg Oncol, 20*(4), 1395-1402. doi: 10.1245/s10434-012-2657-z

Tenesa, A., Farrington, S. M., Prendergast, J. G., Porteous, M. E., Walker, M., Haq, N., . . . Dunlop, M. G. (2008). Genome-wide association scan identifies a colorectal cancer susceptibility locus on 11q23 and replicates risk loci at 8q24 and 18q21. *Nat Genet, 40*(5), 631-637. doi: 10.1038/ng.133

Tomlinson, I. P., Carvajal-Carmona, L. G., Dobbins, S. E., Tenesa, A., Jones, A. M., Howarth, K., . . . Dunlop, M. G. (2011). Multiple common susceptibility variants near BMP pathway loci GREM1, BMP4, and BMP2 explain part of the missing heritability of colorectal cancer. *PLoS Genet, 7*(6), e1002105. doi: 10.1371/journal.pgen.1002105

Tomlinson, I. P., Webb, E., Carvajal-Carmona, L., Broderick, P., Howarth, K., Pittman, A. M., . . . Houlston, R. S. (2008). A genome-wide association study identifies colorectal cancer susceptibility loci on chromosomes 10p14 and 8q23.3. *Nat Genet, 40*(5), 623-630. doi: 10.1038/ng.111

Tomlinson, I., Rahman, N., Frayling, I., Mangion, J., Barfoot, R., Hamoudi, R., . . . Stratton, M. R. (1999). Inherited susceptibility to colorectal adenomas and carcinomas: evidence for a new predisposition gene on 15q14-q22. *Gastroenterology, 116*(4), 789-795.

Tomlinson, I., Webb, E., Carvajal-Carmona, L., Broderick, P., Kemp, Z., Spain, S., . . . Houlston, R. (2007). A genome-wide association scan of tag SNPs identifies a susceptibility variant for colorectal cancer at 8q24.21. *Nat Genet, 39*(8), 984-988. doi: 10.1038/ng2085

Tuupanen, S., Niittymaki, I., Nousiainen, K., Vanharanta, S., Mecklin, J. P., Nuorva, K., . . . Aaltonen, L. A. (2008). Allelic imbalance at rs6983267 suggests selection of the risk allele in somatic colorectal tumor evolution. *Cancer Research, 68*(1), 14-17. doi: Doi 10.1158/0008-5472.Can-07-5766

Tuupanen, S., Turunen, M., Lehtonen, R., Hallikas, O., Vanharanta, S., Kivioja, T., . . . Aaltonen, L. A. (2009). The common colorectal cancer predisposition SNP rs6983267 at chromosome 8q24 confers potential to enhanced Wnt signaling. *Nat Genet, 41*(8), 885-890. doi: 10.1038/ng.406

Umar, A., Boland, C. R., Terdiman, J. P., Syngal, S., de la Chapelle, A., Ruschoff, J., . . . Srivastava, S. (2004). Revised Bethesda Guidelines for hereditary nonpolyposis colorectal cancer (Lynch syndrome) and microsatellite instability. *J Natl Cancer Inst, 96*(4), 261-268.

Usener, D., Schadendorf, D., Koch, J., Dubel, S., & Eichmuller, S. (2003). cTAGE: a cutaneous T cell lymphoma associated antigen family with tumor-specific splicing. *J Invest Dermatol, 121*(1), 198-206. doi: 10.1046/j.1523-1747.2003.12318.x

Valle, L., Serena-Acedo, T., Liyanarachchi, S., Hampel, H., Comeras, I., Li, Z., . . . de la Chapelle, A. (2008). Germline allele-specific expression of TGFBR1 confers an increased risk of colorectal cancer. *Science, 321*(5894), 1361-1365. doi: 10.1126/science.1159397

Warthin, A. S. (1985). Classics in Oncology - Heredity with Reference to Carcinoma as Shown by the Study of the Cases Examined in the Pathological Laboratory of the University-of-Michigan, 1895-1913 (Reprinted from Archives of Internal Medicine, Vol 12, Pg 546-555, 1913). *Ca-a Cancer Journal for Clinicians, 35*(6), 348-359. doi: DOI 10.3322/canjclin.35.6.348

Vasen, H. F., Mecklin, J. P., Khan, P. M., & Lynch, H. T. (1991). The International Collaborative Group on Hereditary Non-Polyposis Colorectal Cancer (ICG-HNPCC). *Dis Colon Rectum, 34*(5), 424-425.

Vasen, H. F., Watson, P., Mecklin, J. P., & Lynch, H. T. (1999). New clinical criteria for hereditary nonpolyposis colorectal cancer (HNPCC, Lynch syndrome) proposed by the International Collaborative group on HNPCC. *Gastroenterology, 116*(6), 1453-1456.

Weiss, J. M., Pfau, P. R., O'Connor, E. S., King, J., LoConte, N., Kennedy, G., & Smith, M. A. (2011). Mortality by stage for right- versus left-sided colon cancer: analysis of surveillance, epidemiology, and end results--Medicare data. *J Clin Oncol, 29*(33), 4401-4409. doi: 10.1200/JCO.2011.36.4414

Venkatachalam, R., Ligtenberg, M. J., Hoogerbrugge, N., Schackert, H. K., Gorgens, H., Hahn, M. M., . . . Kuiper, R. P. (2010). Germline epigenetic silencing of the tumor suppressor gene PTPRJ in early-onset familial colorectal cancer. *Gastroenterology, 139*(6), 2221-2224. doi: 10.1053/j.gastro.2010.08.063

Whiffin, N., Broderick, P., Lubbe, S. J., Pittman, A. M., Penegar, S., Chandler, I., & Houlston, R. S. (2011). MLH1-93G > A is a risk factor for MSI colorectal cancer. *Carcinogenesis, 32*(8), 1157-1161. doi: DOI 10.1093/carcin/bgr089

Whiffin, N., Hosking, F. J., Farrington, S. M., Palles, C., Dobbins, S. E., Zgaga, L., . . . Dunlop, M. G. (2014). Identification of susceptibility loci for colorectal cancer in a genome-wide meta-analysis. *Hum Mol Genet, 23*(17), 4729-4737. doi: 10.1093/hmg/ddu177

Whiffin, N., & Houlston, R. S. (2014). Architecture of inherited susceptibility to colorectal cancer: a voyage of discovery. *Genes (Basel), 5*(2), 270-284. doi: 10.3390/genes5020270

Whittemore, A. S., & Halpern, J. (1994). A Class of Tests for Linkage Using Affected Pedigree Members. *Biometrics, 50*(1), 118-127. doi: Doi 10.2307/2533202

Wiesner, G. L., Daley, D., Lewis, S., Ticknor, C., Platzer, P., Lutterbaugh, J., . . . Markowitz, S. D. (2003). A subset of familial colorectal neoplasia kindreds linked to chromosome

9q22.2-31.2. *Proc Natl Acad Sci U S A, 100*(22), 12961-12965. doi: 10.1073/pnas.2132286100

Wijnen, J. T., Brohet, R. M., van Eijk, R., Jagmohan-Changur, S., Middeldorp, A., Tops, C. M., . . . Vasen, H. F. (2009). Chromosome 8q23.3 and 11q23.1 variants modify colorectal cancer risk in Lynch syndrome. *Gastroenterology, 136*(1), 131-137. doi: 10.1053/j.gastro.2008.09.033

Vogelstein, B., Fearon, E. R., Hamilton, S. R., Kern, S. E., Preisinger, A. C., Leppert, M., . . . Bos, J. L. (1988). Genetic Alterations during Colorectal-Tumor Development. *New England Journal of Medicine, 319*(9), 525-532. doi: Doi 10.1056/Nejm198809013190901

von Holst, S., Picelli, S., Edler, D., Lenander, C., Dalen, J., Hjern, F., . . . Lindblom, A. (2010). Association studies on 11 published colorectal cancer risk loci. *Br J Cancer, 103*(4), 575-580. doi: 10.1038/sj.bjc.6605774

Yeager, M., Orr, N., Hayes, R. B., Jacobs, K. B., Kraft, P., Wacholder, S., . . . Thomas, G. (2007). Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. *Nature Genetics, 39*(5), 645-649. doi: Doi 10.1038/Ng2022

Zanke, B. W., Greenwood, C. M., Rangrej, J., Kustra, R., Tenesa, A., Farrington, S. M., . . . Dunlop, M. G. (2007). Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24. *Nat Genet, 39*(8), 989-994. doi: 10.1038/ng2089

Zhang, L., Conejo-Garcia, J. R., Katsaros, D., Gimotty, P. A., Massobrio, M., Regnani, G., . . . Coukos, G. (2003). Intratumoral T cells, recurrence, and survival in epithelial ovarian cancer. *N Engl J Med, 348*(3), 203-213. doi: 10.1056/NEJMoa020177

Zogopoulos, G., Jorgensen, C., Bacani, J., Montpetit, A., Lepage, P., Ferretti, V., . . . Gallinger, S. (2008). Germline EPHB2 receptor variants in familial colorectal cancer. *PLoS One, 3*(8), e2885. doi: 10.1371/journal.pone.0002885