Genetics and tumor genomics in familial colorectal cancer

Middeldorp, J.W.

Concluding Remarks and Future Perspectives

Chapter 8

The aim of the work described in this thesis was to identify novel genetic risk factors for colorectal cancer (CRC). We applied several approaches to identify such novel CRC risk factors. Our approaches can broadly be divided into germ-line genetic analyses and somatic genomic analyses.

Using the germ-line approach, linkage analysis in seven large familial CRC families provided supportive evidence for region on 3q, which has previously been linked to CRC susceptibility. However, no novel regions of linkage were identified. Study of low-risk CRC susceptibility loci revealed an enrichment of risk alleles in familial CRC patients as compared to sporadic CRC patients. In solitary patients with an early age at onset of disease no such enrichment of risk alleles was observed. These results suggest that clustering of low-risk factors explains part of the excess risk observed in CRC families.

Somatic genomic analysis showed that carcinomas from *MUTYH*-associated polyposis (MAP) patients have a profile of aberrations that is distinct from sporadic CRC and from Lynch syndrome carcinomas. The most distinguishing factor is the high frequency of LOH in the absence of copy number alterations (copy-neutral LOH; see also page 15) in MAP carcinomas. In analogy to the distinct genomic profiles of Lynch carcinomas and MAP carcinomas, we studied the genomic profile of mismatch repair (MMR) proficient familial CRC. The profiles of these MMR proficient familial carcinomas show an increased frequency of 20q gain and an increased frequency of genome-wide cnLOH compared to sporadic CRC, while the overall profile largely resembles the profile of sporadic CRC. These results suggest an important role for 20q in tumor progression in familial CRC.

Germ-line genetic analyses

Familial CRC of which the underlying genetics are currently unknown is likely to represent a heterogeneous group, including both cases with a strong familial clustering of CRC which are likely to have an inherited basis as well as cases with a more sporadic form of CRC that aggregated in the family as a result of shared environment and lifestyles or simply by chance.[1,2] Using germ-line approaches, we searched for rare high penetrance risk factors and we studied the role of common low penetrance risk variants in CRC families.

Linkage analysis in seven large DNA mismatch repair (MMR) proficient CRC families did not provide a novel region of significant linkage that could harbor a high penetrance risk factor (chapter 3). However, our results support linkage to 3q21-q24, a region that has previously been identified as a CRC susceptibility locus.[3,4] Other regions that were previously reported to be linked to CRC susceptibility using linkage analysis, including 7q31, 9q22.2-31.2, 11q23.2 and 11q13.4, 14q24.2, and 22q12.1, were not supported by our linkage results.[5-9] This might be explained by differences in family ascertainment, since the other studies analyzed nuclear families or sib-pairs.

Three independent studies, including our study, now reported linkage of the 3q region to CRC susceptibility, with a smallest region of overlap encompassing 3q22.1-q22.3.[3,4,10] Together, these studies provide evidence for a novel CRC risk factor on 3q22. However, none of the

studies found strong evidence of linkage, suggesting that the risk of this locus is moderate. Mutation analyses of 46 genes in this region in previous studies did not identify pathogenic mutations.[3,4,10] The genes were screened for pathogenic mutations using a common approach by analyzing all exonic sequences, intron/exon boundaries, 5'- and 3'-untranslated regions (UTRs) and the promoter sequences of genes. The type of mutations that is generally searched for is truncating mutations. However, less obvious types of genetic alterations can be responsible for CRC predisposition. A recent study in CRC families, for example, showed that a heterozygous germ-line deletion of the last exons of *TACSTD1*, upstream of *MSH2*, causes epigenetic inactivation of *MSH2* in *TACSTD1*-expressing tissues and thereby predisposes these families to colorectal cancer.[10] These results demonstrate the potential profit of broader screening approaches for mutations, insertions, and deletions (see also future perspectives section below).

We performed association analyses to study the role of low-risk variants in familial CRC (chapter 4). Six loci that were identified in genome wide association studies (GWAS) were analyzed and the association with CRC risk could be replicated for five of these loci (located on chromosomes 8q23.3, 8q24.1, 11q23.1, 15q13.3, and 18q21.2). The odds ratios for these loci were increased, although not significantly, as compared to initial GWAS. This is likely a result of the familial nature of our cohort. The association between rs10795668 (10p14) and CRC risk was not observed in our Dutch familial CRC cohort, possibly due to either a lack of power or a population difference between our cohort and the English cohort in the initial GWAS.
Interestingly, we observed a significant increase in the number of risk alleles in cases compared with controls and an increase in odds ratio with increasing numbers of risk alleles. These results are in line with previous studies in unselected CRC patients.[12] Moreover, we observed an increased number of risk alleles in the patients with a family history of CRC as compared to solitary cases with an early age of onset of CRC, where no enrichment of risk alleles was observed. This shows that although low-risk alleles initially were thought to play a role in 'sporadic' CRC, they also play a role in FCC families. Therefore, clustering of low-risk variants may explain part of the excess risk in CRC families. Our results were recently confirmed by a study of Finnish familial CRC patients, that also observed an increased number of risk alleles in familial CRC patients compared to a group of sporadic CRC patients.[11] In solitary cases, however, other genetic models are likely to play a role; rare recessive high-risk variants might provide an explanation for their increased CRC risk, as reflected by a lack of affected first-degree relatives and their early onset of disease.
A further example of a role for common low-risk variants in familial CRC was provided by analysis of the seven large CRC families that we also studied with linkage analysis. We detected a significant association between rs16892766 and rs12953717 and CRC within these families (chapter 3). Moreover, two of the risk variants (rs16892766 and rs3802842) appeared to have a modifier role in Lynch syndrome families.[13] In line with these results, a meta-analysis of two large genome-wide association studies estimated that the ten low-risk variants that are

currently known, together account for 6% of the excess familial risk on basis of an additive model.[14]

Overall, our linkage analysis results and those results of others do not support a model in which a single highly penetrant gene explains the excess risk in familial colorectal cancer. All linkage analysis efforts over the last decade did not yet provide a novel high-risk factor, suggesting that the underlying genetics of the remaining familial CRC may be more complex. To further analyze the possibility of a high penetrance factor following a recessive model of inheritance, we analyzed the seven MMR proficient CRC families by homozygosity mapping. Very recently, preliminary results provided a candidate region on chromosome 5 that could harbor a CRC susceptibility locus. In ongoing and future studies, we will analyze this region in further detail.

Our association studies showed that low-risk variants may explain part of the excess risk in CRC families. Moderate risk factors might explain part of the excess risk in the remaining CRC families, but these are very difficult to identify with the current methods because of their assumed relatively low population frequency and their moderate penetrance. With the appearance of novel sequencing methods the identification of rare variants involved in CRC predisposition becomes feasible, for example using exome-sequencing or whole genome sequencing in highly selected cases (further discussed below in the future perspectives section).

However, while it is tempting to explain the excess CRC risk by genetic factors, the observed aggregation of CRC in large families could also be partly explained by environmental factors that are shared by family members. Moreover, the families could represent sampling artifacts and thus show aggregation of CRC largely by chance, although based on the cancer burden in these families, this seems less likely.

Also for other complex diseases, it was observed that much of the estimated heritability is not explained by the low-risk variants identified through genome wide association studies (reviewed by [15]). Several explanations for this so-called “missing heritability” were proposed. The missing heritability might be explained by risk variants with a low minor allele frequency (below 5%) which are not captured in genome-wide association studies and exert risks that are too low to be detected in linkage analyses. Structural variation like inversions or translocations might also partly explain the missing heritability. In addition, the estimated heritability could be an overestimation of the actual heritability of disease.[15] All these aspects could also apply to colorectal cancer.

The spectrum of CRC, including sporadic, hereditary, and familial CRC is summarized in Figure 1.

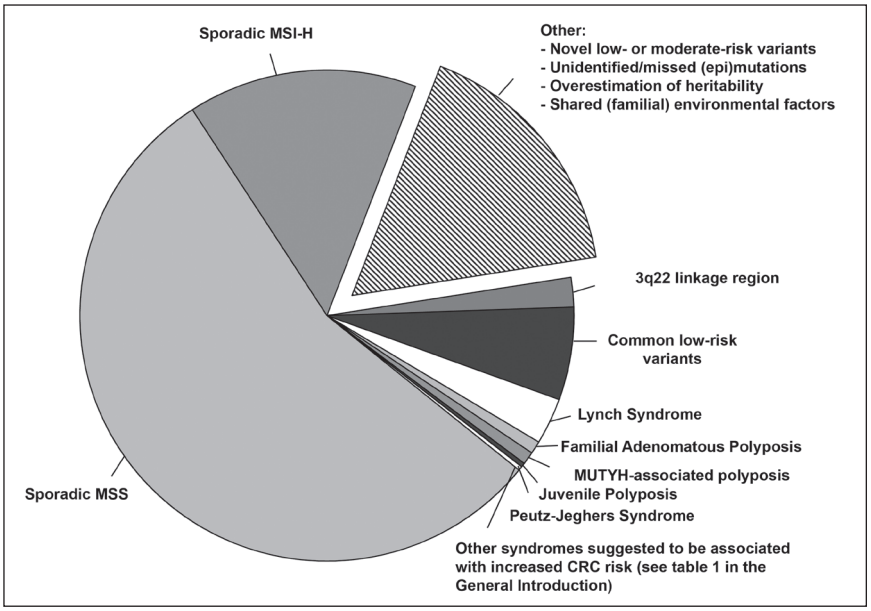

Figure 1. Spectrum of CRC.

In up to 35% of all CRC, hereditary factors play a role. About 5% of all CRC can be explained by known CRC syndromes. Common low risk variants account for 6% of the excess familial risk and linkage analyses identified 3q22 as a candidate region for a CRC susceptibility locus. In the remaining familial CRC cases other factors probably underlie the CRC susceptibility.

Several other approaches were applied by other groups to further study the nature of CRC susceptibility. Two recent studies analyzed homozygosity levels in colorectal cancer patients to study the role of consanguinity in CRC predisposition. Up till now, MUTYH-associated polyposis is the only known recessively inherited colorectal cancer syndrome. The results of both studies are inconclusive, since contradicting results were obtained. On one hand, the study of Bacolod et al. observed more and longer homozygosity regions in 74 colorectal cancer patients as compared to 264 controls, suggesting a role for recessively inherited CRC predisposition.[12] The number of homozygous regions (>4 Mb) per patient was low. The regions differed among the patients and were spread throughout the genome. On the other hand, a large study of Spain et al. did not provide evidence for increased homozygosity in CRC patients in a cohort comprising 921 cases and 929 controls.[13] The latter results show that the larger part of genetic CRC susceptibility likely follows a dominant mode of inheritance and that a minor part is explained by recessively inherited factors. A study of regions of homozygosity in specific subgroups of CRC patients in which a recessive inheritance mode is expected, such as solitary CRC patients with an early age of onset, might, however, be a good strategy to identify novel recessively inherited CRC risk factors.

The loci discovered in the genome-wide association studies inform on novel genes and/or pathways involved in CRC and might point to common molecular mechanisms involved in cancers.[14] Unraveling the biological mechanisms explaining the association between the identified risk loci and CRC will provide important novel insights in the etiology of colorectal cancer. In silico analyses to investigate the causality showed that many associated variants are in linkage disequilibrium (LD) with DNA sequence changes that influence gene expression rather than with nonsynonymous sequence changes that lead to altered proteins.[14] This might relate to the high population frequency and the low-risk these variants confer; different levels of expression likely exert more subtle effects than altered proteins.

Interestingly, five out of the ten identified low-risk factors are SNPs that are in linkage disequilibrium (LD) with genes of the TGF-β superfamily signaling pathway, including the genes *SMAD7*, *GREM1*, *BMP2*, *BMP4*, and *RHPN2*.[15] This further underlines the important role that the TGF-β pathway plays in CRC susceptibility; germ-line mutations in *SMAD4* and *BMPR2* were already known to be involved in Juvenile Polyposis and several members of the TGF-β pathway, including *SMAD4* and *TGFBR2*, are targeted by somatic mutations in colorectal tumors.[16,17]

An alternative approach to perform association studies as compared to using tagging SNPs, was adopted by Webb et al. who studied associations with gene-centric SNPs.[18] These gene-centric SNPs included 7000 genome-wide nonsynonymous SNPs, which alter the encoded amino-acid sequence. However, this study did not yield any significant association between CRC risk and any of the nonsynonymous SNPs, which a priori are more likely to have functional impact than synonymous SNPs. An explanation for the absence of associations could be that natural selection on alleles in coding regions has rendered the risk alleles rather rare.[18]

Somatic genomic analyses

In this thesis, we applied single nucleotide polymorphism (SNP) arrays to study colorectal tumors for genome-wide chromosomal copy number aberrations and loss of heterozygosity (LOH). Our aim was to generate a profile of genomic aberrations in MMR proficient familial CRC that might provide further insight in the biological basis of the increased CRC susceptibility in these families. Furthermore, these profiles might identify a candidate region that could harbor a CRC susceptibility factor.

Although many comparative genomic hybridization (CGH) studies of CRC have been performed over the last years (an overview is provided in chapter 1, table 2), most studies analyzed sporadic CRC, whereas few studies analyzed hereditary or familial CRC. Moreover, the majority of the studies used metaphase-based CGH or arrayCGH which only provides information on copy number aberrations. Genome-wide SNP arrays were used to a much lesser extent, even though these arrays provide information on both copy number and genome-wide LOH in the absence of copy number aberrations (copy-neutral LOH or cnLOH). In addition, we showed that, using SNP arrays, the genomic profiling of tumors can be further improved by analy-

zing flow-sorted tumor cells and incorporation of the DNA index in the analysis. We developed a novel algorithm, the lesser allele intensity ratio (LAIR), which can accurately determine the allelic state of all chromosomes. Upon incorporation of the DNA index of the tumors, LOH, cnLOH, balanced amplifications, and allelic imbalances can be distinguished (chapter 7). In addition to the assessment of allelic states, this method can address tumor heterogeneity.

The great value of genome-wide cnLOH analysis is illustrated by our study of MAP carcinomas (chapter 5). The main characteristic of these tumors was the high frequency of cnLOH; whereas physical loss occurred to a much lesser extent (Figure 2). This is in contrast to sporadic CRC, in which physical loss is frequent and few regions of cnLOH are observed.[19] The tumors from Lynch syndrome patients, which we studied previously, showed a characteristic profile, lacking gross chromosomal aberrations but only exhibiting a small region of copy neutral LOH around the locus of the mutated mismatch repair gene (Figure 2).[20] Also for many other cancers regions of cnLOH have described, including basal cell carcinoma and retinoblastoma (reviewed by [21]).

Compared to the unique and distinct profiles of genomic aberrations that were observed in MAP carcinomas and Lynch syndrome carcinomas, the group of MMR proficient familial colorectal carcinomas that was studied showed resemblance to sporadic CRC but with an increased frequency of 20q gain and genome-wide cnLOH (chapter 6). The most frequent aberrations in MMR proficient familial CRC included gains of chromosome 7, 8q, 13q, 20p and 20q, physical losses of 17p, 18p, and 18q. Remarkably, an increased frequency of 20q gain (77%) was observed as compared to the frequency in sporadic CRC (30-50%), in which it is considered to be an early event during tumorigenesis.[23,26] Moreover, an increased frequency of genome-wide cnLOH was observed at the expense of the frequency of physical losses in MMR proficient familial CRCs as compared to sporadic CRC. The observed high frequency of 20q gain in familial CRC confirmed a previous report on chromosomal aberrations in a Finnish cohort of familial CRC.[22] In this study 99 familial CRCs were compared to 186 sporadic CRCs using genome-wide allelotyping with microsatellite markers and copy number analysis using CGH on a subset of tumors. They observed gain of 20q in 85% of familial CRCs. They could, however, not confirm their results in a series of 67 familial and 96 sporadic CRCs from the UK.

Further analysis of chromosome 20q seems valuable and could include somatic sequence analysis of candidate genes or sequence analysis of the entire region. Gene expression analysis could provide information on differentially expressed genes. In a recent study, gene expression levels of genes located on 20q were compared between sporadic colorectal adenomas and carcinomas to identify oncogenes involved in adenoma to carcinoma progression. Several genes were found to be differentially expressed in carcinomas with gain of 20q as compared to carcinomas without such gain.[23] These genes could be also involved in familial CRCs, although the genetic targets could be very well be different from sporadic CRC. Moreover, profiles of imprinting on 20q could be studied, since for example *GNAS* (located on 20q13.32) is known to be regulated by complex tissue-specific imprinting patterns.[24]

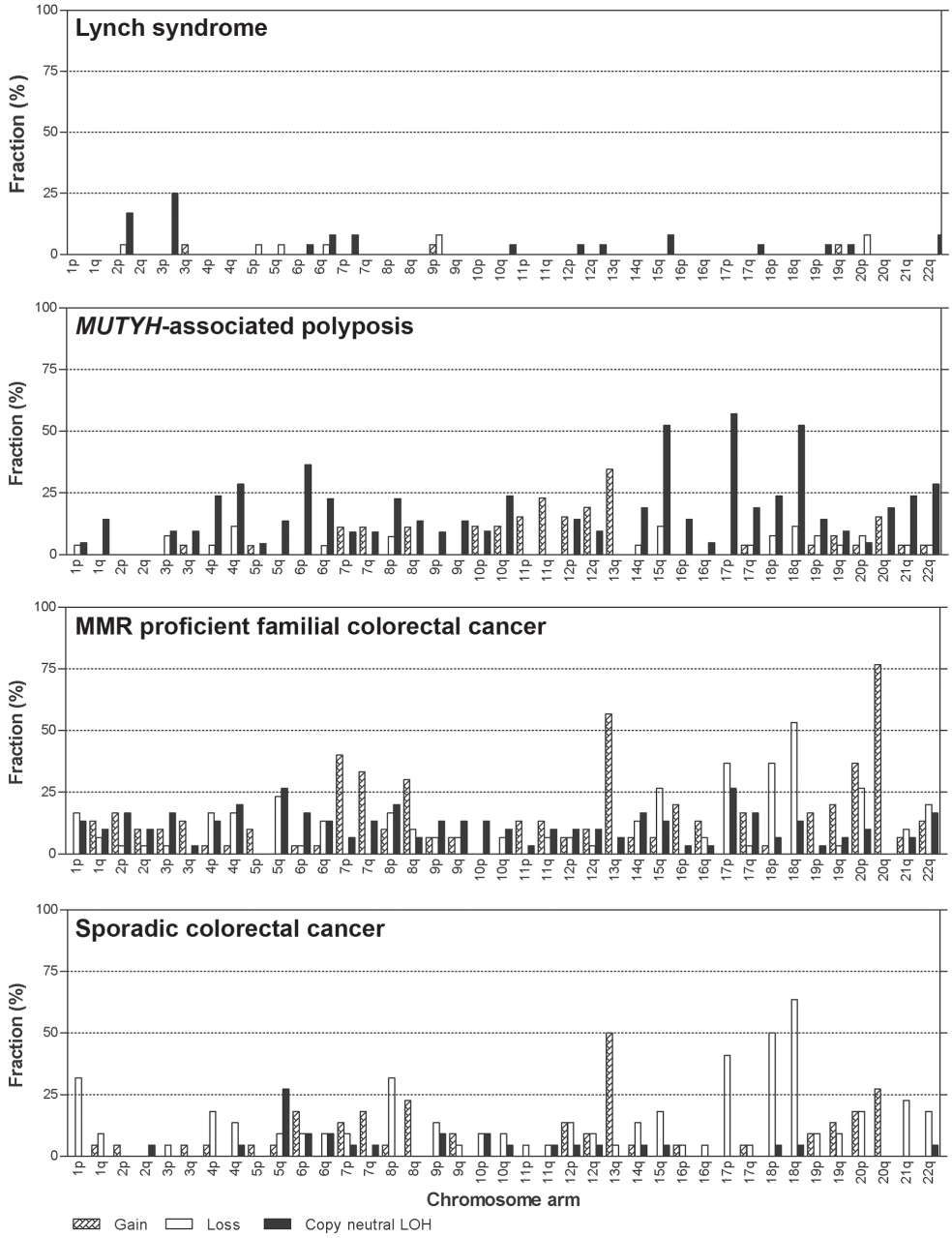

Figure 2. Profile of genomic copy number aberrations and cnLOH in colorectal cancer.

The bars indicate the percentage of carcinomas that exhibit an event of gain, loss or cnLOH of a chromosome. This percentage has been calculated for the respective chromosome arms. White bars, chromosomal gains; checked bars, physical losses of chromosomes; black bars, cnLOH. The profile of Lynch syndrome carcinomas (upper panel) is derived from Van Puijenbroek et al.[20]

Even though the genetic analysis of familial CRC did provide few distinct features for this group of MMR proficient familial CRC, tumor analysis provides a valuable approach to investigate CRC susceptibility. A successful example concerns the identification of mutations in *MUTYH* as the cause of polyposis (now termed *MUTYH*-associated polyposis), which came from the observation that tumors exhibited specific G>T DNA sequence transversions.[25] This provides a clear example of the value of tumor analysis of familial CRC. Additionally, before the identification of *BRCA2* as a breast cancer susceptibility gene, 13q deletions were already observed to be more frequent in familial breast cancer patients.[26]

Future perspectives

Future studies to identify novel genetic factors involved in CRC susceptibility, will require novel approaches because, despite of all efforts that were made over the last decades, a large part of the heredity remains unexplained. First, the search for novel genetic colorectal cancer risk factors has so far been focused mainly on genotypic variants. The study of copy number variants (CNV) may also yield novel CRC risk factors. As already discussed above, deletion of the last exons of *TACSTD1* is associated with CRC predisposition, because this deletion causes epigenetic inactivation of *MSH2* in *TACSTD1*-expressing tissues.[10] In addition, genomic (micro-)deletions showed already to be instrumental for the identification of cancer predisposing genes in for example retinoblastoma, Von Hippel-Lindau disease, and Wilms' tumor associated with WAGR syndrome (Wilms tumor, Aniridia, Genitourinary anomalies, and mental Retardation syndrome).[32-35] More recently, analysis of CNVs by array comparative genomic hybridization has led to the identification of a causative gene for example for CHARGE syndrome (Coloboma, Heart anomaly, choanal Atresia, Retardation, Genital and Ear anomalies syndrome).[36] In addition, association studies could be performed for CNVs that are polymorphic in the general population. Several associations between DNA copy number variants and common complex diseases have already been described (reviewed by [27]).

For example, a significant association was found between a low number of copies of a polymorphism in the human beta-defensin gene *HBD-2* (or *DEFB4*) and Crohn's disease, with a corresponding odds ratio of 3.06 (95% CI 1.46–6.45).[38] Other CNV associations were reported with HIV/AIDS susceptibility, rheumatoid arthritis, systemic autoimmune disease, systemic lupus erythematosus, psoriasis, and asthma, with odds ratios ranging from 1.34-5.27.[27]
Secondly, a useful next strategy could be to sequence the "exome" of colorectal cancer patients, including all protein encoding regions of the genome, or even to sequence the whole genome of CRC patients. Exome sequencing has already been applied successfully by Ng et al. on the exomes of 12 individuals.[28] Ng et al. studied the exomes of eight HapMap individuals and four unrelated individuals affected with Freeman-Sheldon syndrome, a rare autosomal dominant disease caused by mutations in *MYH3*. Filters were applied to identify the possible deleterious variant among all identified variants. Non-causal variants were removed by excluding variants that were not observed in one or more of the affected individuals.

Presumably common variants were removed by removing dbSNP catalogued variants and by removing the variants identified in the eight HapMap individuals. After application of these filters, *MYH3* was the only gene that was left on the candidate list.[28] The genetic homogeneity of the affected individuals and the availability of the HapMap individuals were important factors in the successful identification of *MYH3*. When applying this strategy for the identification of novel CRC susceptibility factors, probable genetic heterogeneity will have a significant impact on the performance and larger sample sizes will be required. Several genetic variants are likely to remain after filtering of the identified variants. To determine the significance of identified mutations, functional analyses could be performed.

Thirdly, in addition to a gene-centric approach, microRNA (miRNA) sequences could be studied for alterations. MiRNAs are small non-coding RNA sequences involved in post-transcriptional regulation of gene expression. Evidence for aberrant expression of miRNAs in human cancers is growing, indicating that they are involved in tumorigenesis (reviewed by [40]). However, both a global increase of miRNA levels in prostate cancer as global inhibition of miRNA processing have been described in cancer, suggesting a complex relation between miRNAs and tumorigenesis. A list of miRNAs involved in colorectal cancer has already been described.[29] Whether miRNAs are involved in CRC susceptibility still needs to be studied.

Fourthly, epigenetic changes or susceptibility to epigenetic changes might be involved in CRC predisposition and analysis of the epigenome could therefore be a fruitful approach. For example, germ-line methylation of the *MLH1* promoter region has been described in colorectal cancer patients.[42,43] However, the mode of inheritance of such epigenetic mutations remains unclear. Few examples of apparent inheritance of epigenetic states exist and it is generally believed that epigenetic modification are reset in germ cells.[30]

Finally, the role of the recently identified low-risk variants in tumor initiation (and progression) should be determined in future studies. This will provide important novel insights in the etiology of colorectal cancer and insight in the biological mechanisms involved in CRC susceptibility. Additional analyses need to be performed to identify the causal variants at the different loci and to unravel the biological mechanisms that cause the increased CRC risk. This will be a challenging task, since several of the identified risk alleles are located in regions that are without known genes (so-called gene deserts). Resequencing of the locus on 18q21.1 has identified Novel 1 to be the causal variant on this locus.[31] Variant rs6983267 is likely to be itself the causal variant, as determined by resequencing and linkage disequilibrium analysis in this region.[45] Further analyses of the region on 18q21.1, already showed that its causal allele (Novel 1) is associated with a reduced expression of *SMAD7* in a Xenopus laevis model.[31] The variant rs6983267 (8q24.1) is located in a transcriptional enhancer region that is bound by TCF7L2 (also referred to as TCF4), a transcriptional effector of the Wnt signaling pathway. The alleles of rs6983267 were found to differentially bind the transcription factor TCF7L2. It was shown that the region around rs6983267 physically interacts with *MYC*, but no robust association could be detected between rs6983267 and *MYC* mRNA expression.[46,47] However, a role for *MYC* is still likely, since *MYC* is a known target of TCF7L2 and is an important

oncogene in colorectal tumorigenesis. A strong indication for a role in somatic tumor evolution for rs6983267 was found by Tuupanen et al., who observed that in case of allelic imbalance at 8q24, the risk allele was favored in about two-thirds of the tumors.[32] Similar analyses for the other low-risk variants that were identified in recent genome-wide association studies will provide more and probably new insight into mechanisms of CRC initiation and progression.

In conclusion, the results described in this thesis suggest that it is unlikely that the excess risk in many of the MMR proficient familial CRC cases is explained by dominant high-risk genetic factors. Single young patients without a family history of CRC might be explained by a recessive origin of disease. In MMR proficient CRC families, one or more moderate risk factors might play a role. Research should therefore be directed more towards identifying novel factors conferring a moderate risk, even though these factors are more difficult to find. Recent advances in sequencing technology as well as novel knowledge of mechanisms involved in CRC development provided by the recently identified low-risk factors might facilitate the identification of novel moderate CRC risk factors.

Reference List

1. Castells A, Castellvi-Bel S, Balaguer F. Concepts in familial colorectal cancer: where do we stand and what is the future? Gastroenterology (2009), 137: 404-409.
2. Lichtenstein P, Holm NV, Verkasalo PK et al. Environmental and heritable factors in the causation of cancer--analyses of cohorts of twins from Sweden, Denmark, and Finland. N Engl J Med (2000), 343: 78-85.
3. Kemp Z, Carvajal-Carmona L, Spain S et al. Evidence for a colorectal cancer susceptibility locus on chromosome 3q21-q24 from a high-density SNP genome-wide linkage scan. Hum Mol Genet (2006), 15: 2903-2910.
4. Papaemmanuil E, Carvajal-Carmona L, Sellick GS et al. Deciphering the genetics of hereditary non-syndromic colorectal cancer. Eur J Hum Genet (2008), 16: 1477-1486.
5. Djureinovic T, Skoglund J, Vandrovcova J et al. A genome wide linkage analysis in Swedish families with hereditary non-familial adenomatous polyposis/non-hereditary non-polyposis colorectal cancer. Gut (2006), 55: 362-366.
6. Kemp ZE, Carvajal-Carmona LG, Barclay E et al. Evidence of linkage to chromosome 9q22.33 in colorectal cancer kindreds from the United kingdom. Cancer Res (2006), 66: 5003-5006.
7. Neklason DW, Kerber RA, Nilson DB et al. Common familial colorectal cancer linked to chromosome 7q31: a genome-wide analysis. Cancer Res (2008), 68: 8993-8997.
8. Skoglund J, Djureinovic T, Zhou XL et al. 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 (2006), 43: e7.
9. Wiesner GL, Daley D, Lewis S et al. A subset of familial colorectal neoplasia kindreds linked to chromosome 9q22.2-31.2. Proc Natl Acad Sci U S A (2003), 100: 12961-12965.
10. Ligtenberg MJ, Kuiper RP, Chan TL et al. Heritable somatic methylation and inactivation of MSH2 in families with Lynch syndrome due to deletion of the 3' exons of TACSTD1. Nat Genet (2009), 41: 112-117.
11. Niittymaki I, Kaasinen E, Tuupanen S et al. Low-penetrance susceptibility variants in familial colorectal cancer. Cancer Epidemiol Biomarkers Prev (2010), 19: 1478-1483.
12. Bacolod MD, Schemmann GS, Wang S et al. The signatures of autozygosity among patients with colorectal cancer. Cancer Res (2008), 68: 2610-2621.
13. Spain SL, Cazier JB, Houlston R et al. Colorectal cancer risk is not associated with increased levels of homozygosity in a population from the United Kingdom. Cancer Res (2009), 69: 7422-7429.
14. Kingsmore SF, Lindquist IE, Mudge J et al. Genome-wide association studies: progress and potential for drug discovery and development. Nat Rev Drug Discov (2008), 7: 221-230.
15. Tenesa A, Dunlop MG. New insights into the aetiology of colorectal cancer from genome-wide association studies. Nat Rev Genet (2009).
16. Howe JR, Roth S, Ringold JC et al. Mutations in the SMAD4/DPC4 gene in juvenile polyposis. Science (1998), 280: 1086-1088.
17. Howe JR, Bair JL, Sayed MG et al. Germline mutations of the gene encoding bone morphogenetic protein receptor 1A in juvenile polyposis. Nat Genet (2001), 28: 184-187.
18. Webb E, Broderick P, Lubbe S et al. A genome-wide scan of 10 000 gene-centric variants and colorectal cancer risk. Eur J Hum Genet (2009).
19. Lips EH, van Eijk R, de Graaf EJ et al. Progression and tumor heterogeneity analysis in early rectal cancer. Clin Cancer Res (2008), 14: 772-781.
20. van Puijenbroek M, Middeldorp A, Tops CM et al. Genome-wide copy neutral LOH is infrequent in familial and sporadic microsatellite unstable carcinomas. Fam Cancer (2008), 7: 319-330.
21. Tuna M, Knuutila S, Mills GB. Uniparental disomy in cancer. Trends Mol Med (2009), 15: 120-128.
22. Laiho P, Hienonen T, Karhu A et al. Genome-wide allelotyping of 104 Finnish colorectal cancers reveals an excess of allelic imbalance in chromosome 20q in familial cases. Oncogene (2003), 22:

2206-2214.
23. Carvalho B, Postma C, Mongera S et al. Multiple putative oncogenes at the chromosome 20q amplicon contribute to colorectal adenoma to carcinoma progression. Gut (2009), 58: 79-89.
24. Kelsey G. Epigenetics and imprinted genes: insights from the imprinted Gnas locus. Horm Res (2009), 71 Suppl 2: 22-29.
25. Al Tassan N, Chmiel NH, Maynard J et al. Inherited variants of MYH associated with somatic G:C-->T:A mutations in colorectal tumors. Nat Genet (2002), 30: 227-232.
26. Thorlacius S, Jonasdottir O, Eyfjord JE. Loss of heterozygosity at selective sites on chromosomes 13 and 17 in human breast carcinoma. Anticancer Res (1991), 11: 1501-1507.
27. Ionita-Laza I, Rogers AJ, Lange C et al. Genetic association analysis of copy-number variation (CNV) in human disease pathogenesis. Genomics (2009), 93: 22-26.
28. Ng SB, Turner EH, Robertson PD et al. Targeted capture and massively parallel sequencing of 12 human exomes. Nature (2009), 461: 272-276.
29. Cummins JM, He Y, Leary RJ et al. The colorectal microRNAome. Proc Natl Acad Sci U S A (2006), 103: 3687-3692.
30. Cropley JE, Martin DI, Suter CM. Germline epimutation in humans. Pharmacogenomics (2008), 9: 1861-1868.
31. Pittman AM, Naranjo S, Webb E et al. The colorectal cancer risk at 18q21 is caused by a novel variant altering SMAD7 expression. Genome Res (2009), 19: 987-993.
32. Tuupanen S, Niittymaki I, Nousiainen K et al. Allelic imbalance at rs6983267 suggests selection of the risk allele in somatic colorectal tumor evolution. Cancer Res (2008), 68: 14-17.