PREFACE

This Ph.D.-thesis is submitted for evaluation at the Faculty of Health Sciences at the University of Copenhagen. The project is a part of the program “Air Pollution in a Life-time Health Perspective” (AIRPOLIFE), a Danish Centre of Excellence devoted to the study and prevention of health effects of air pollution. The study is based on DNA from blood samples and questionnaire data from the Danish prospective cohort “Diet, Cancer and Health” at the Danish Cancer Society, Copenhagen. The laboratory work was carried out at The National Research Center for Working Environment (Copenhagen) and at The National Food Institute (Mørkhøj), while the statistical work was performed at the Danish Cancer Society (Copenhagen).

Main supervisor of this thesis was Professor Steffen Loft (Department of Environmental and Occupational Medicine, University of Copenhagen), and co-supervisors were Head of Research Programme Environment & Cancer Ole Raaschou-Nielsen (Institute of Cancer Epidemiology, Danish Cancer Society, Copenhagen) and Senior Researcher Ulla Birgitte Vogel and Professor Håkan Wallin (Molecular Biology and Aerosol Science, The National Research Center for Working Environment, Copenhagen).

This thesis consists of a general overview of colorectal etiology and the DNA repair mechanisms base excision repair and nucleotide excision repair in defence of bulky DNA adducts and oxidative DNA damages including key-results from my own work, represented by the manuscripts I-IV. Results from association studies of other types of cancer and polymorphisms in the genes studied in the work of this thesis are described, including results on lung cancer from manuscript V-VII.


Manuscript IV Polymorphisms in ASE-1, RAI and ERCC1 and the effects of tobacco smoking and alcohol consumption on risk of colorectal cancer: A Danish prospective case-cohort study. Hansen, R.D. et al., (2007) BMC Cancer in review (accepted with revisions)


ACKNOWLEDGEMENTS

First and foremost, I want especially to express my gratitude to my daily supervisors Ulla Vogel and Ole Raaschou-Nielsen for your enthusiasm, all the helpful discussions, your good advises and valuable input along the way. I thank Steffen Loft for your valuable feedback during the writing process. And I want to thank all of my supervisors Ulla Vogel, Håkan Wallin, Ole Raaschou-Nielsen and Steffen Loft for the opportunity to do this Ph.D.

During my Ph.D.-study, I have had the opportunity of working in two dynamic research institutes: The National Research Centre for Working Environment and The Institute of Cancer Epidemiology (Danish Cancer Society) surrounded by fabulous colleagues. Thank you all for the great laughs and discussions. My Ph.D.-thesis was completed with the support of many people, and I am very grateful for the help I have received. My special thanks go to Anne-Karin Jensen and Lourdes Pedersen for helping me getting started with the laboratory analyses, and to Mette Sørensen and Kirsten Frederiksen for your support when I lost track of the statistical analyses or the SAS programming.

Elin Kure, Camilla Furu-Skjelbred and Mona Sabo, Norway: Thank you for giving me the opportunity to participate in the work with your KAM study and for a very nice co-operation.

I thank Lars Ove Dragsted and Britta Naimi Krath for making it possible for me to include a short stay at the National Food Institute during my Ph.D-study. Thank you for all the help with the analyses.

And I thank Sana Ebdah and Xu Jin from Roskilde University Center, Denmark, for genotyping the \textit{GPX1} 5’UTR and \textit{RHOA} 3’UTR polymorphisms.

Finally, I want to thank the Danish Research Council, the Danish Graduate School in Public Health Science and the Faculty of Medicine at University of Copenhagen for providing the financial support.

Copenhagen, July 2007

\begin{center}
\textit{Rikke Dalgaard Hansen}
\end{center}
SUMMARY

Colorectal cancer is the third most common cancer and the leading cause of cancer deaths in Western industrialised countries. Migrant studies and the large international variation in incidence rates indicate that lifestyle factors, including dietary, are associated with risk of colorectal cancer, but traditional epidemiological studies based on lifestyle questionnaires and outcome have mostly failed in identifying the exact risk and beneficial factors. Our current knowledge of colorectal carcinogenesis indicates a multifactorial and multi-step process that involves various genetic alterations and several biological pathways. An understanding of differences in individual susceptibility and better exposure assessment may be crucial in identifying lifestyle risk factors and possible interactions between susceptibility and exposures in relation to risk of colorectal cancer.

Various DNA alterations can be caused by exposure to environmental and endogenous carcinogens through direct binding of metabolites (adduct formation) or through oxidative stress. If not repaired the DNA lesions may lead to genetic instability, mutagenesis and cell death. Common occurring single nucleotide polymorphisms (SNPs) in the genes involved in defence of oxidative DNA damages and DNA repair may possibly contribute to the variation in the capacity of these mechanisms. Hence, these SNPs may be important biomarkers of susceptibility to cancer.

This Ph.D.-thesis presents the molecular and cellular mechanisms leading to colorectal cancer. A systematically review of the literature are conducted on associations between SNPs in genes involved in defence of oxidative DNA damages, nucleotide excision repair and apoptosis and risk of colorectal adenomas and colorectal cancer. The review is focused on SNPs, and interaction between the polymorphisms and various lifestyle factors, in the following genes: XPD, XPC, XPA, ERCC1, OGG1, GPX1, RHOA, ASE-1 and RAI. The polymorphisms, except for RHOA, are previously observed associated with risk of other types of cancer. In addition, association studies of the polymorphisms are examined on various other types of cancer.

The present review of colorectal cancer studies includes 17 studies on 25 different SNPs. The results were generally inconsistent or too few to compare to highlight any trend and no strong associations were observed for risk of colorectal adenomas or colorectal cancer. Overall, the role of genetic variants as SNPs in genes involved in defence of oxidative DNA damages, nucleotide excision repair and apoptosis is not satisfactorily clarified at present. It is possible that some of the SNPs may contribute to development of adenomas or colorectal cancer only in concomitance with certain dietary and lifestyle factors. Furthermore, it may be only the joint effect of multiple polymorphisms that will provide us with information about genetic susceptibility for colorectal cancer. Larger carefully designed studies with stratified/adjusted analyses of gene-gene and gene-environment interactions may be required in the future to achieve convincing statistically significant results on factors involved in colorectal carcinogenesis.
DANSK RESUME


Adskillige DNA-ændringer er forårsaget ved eksponering for exogene- og endogene carcinogener via direkte binding af metabolitter (dannelsen af addukter) eller ved oxidativ stress. Hvis det skadede DNA ikke repareres kan det lede til genetisk ustabilitet, mutagenese og celledød. Almindeligt forekommende enkelt-nukleotid polymorfismer (SNPs) i gener involveret i forsvar mod oxidative DNA-skader og i DNA-reparation kan muligvis være medvirkende til variationen i disse mekanismers kapacitet. Ergo, er disse SNPs mulige biomarkører for følsomhed for kræft.

Denne Ph.D.-afhandling præsenterer de molekylære og cellulære mekanismer der leder til kolorektal cancer. En systematisk gennemgang af litteraturen er udført omhandlende sammenhænge mellem SNPs i gener involveret i forsvar mod oxidative DNA-skader, nucleotid excision repair og programmeret celledød og risiko for adenomas i tarmen og kolorektal cancer. Der fokuseres på SNPs i følgende gener: XPD, XPC, XPA, ERCC1, OGG1, GPX1, RH0A, ASE-1 og RAI og en eventuel vekselvirkning mellem polymorfismerne og adskillige livsstilsfaktorer i relation til risiko for kolorektal cancer. Polymorfismerne er tidligere fundet associerede med risiko for andre kræftformer, RH0A undtaget. Derudover gennemgåes associations-studier for andre kræftformer.

# INTRODUCTION

Hypotheses and Aims

# COLORECTAL CANCER

## Molecular Epidemiology of Colorectal Cancer

## DNA Adducts

## Oxidative DNA Damages

## Life Style Factors and DNA Damages

## Morphology and Histology of Colon and Rectum

## Morphology and Histology of Polyps in Colon and Rectum

## The Adenoma-Carcinoma Sequence

# DNA REPAIR

## Base Excision Repair

## Nucleotide Excision Repair

## Double-strand Break Repair

## Mismatch Repair

# DISCUSSION

# CONCLUSION AND PERSPECTIVES

# REFERENCE LIST

# APPENDIX I – THE COHORTS

# APPENDIX II – LABORATORY ANALYSIS

# APPENDIX III – STATISTICAL ANALYSIS

# ABBREVIATIONS

# MANUSCRIPT I-VII
INTRODUCTION

Colorectal cancer is the endpoint of interest in this thesis. Colorectal cancer ranks among the three most common cancers in terms of both cancer incidence and cancer-related deaths in most Western industrialised countries. Thus, every year nearly one million people worldwide develop colorectal cancer. Lifetime risk of colorectal cancer may reach 6% of the population in the Western industrialised countries [2]. In Denmark 2471 new cases of colon cancer and 1155 new cases of rectal cancer were diagnosed in 2001 and the relative 5-year survival rate was approximately 50% according to the Danish National Board of Health. The age-specific incidence of colorectal cancer rises sharply after 35 years of age, with approximately 90% of cancers occurring in persons older than 50 years [3]. The mean age at time for diagnosis in Danish colorectal cancer patients is approximately 70 years for men and 72 years for women [4]. The disease develops either sporadically, as a part of a hereditary cancer syndrome, or induced by inflammatory bowel disease [3]. Ten to fifteen percent of colorectal cancer cases are caused by hereditary syndromes [3].

Migrant studies, where populations migrate from low-risk to high-risk areas, have demonstrated that the colorectal cancer incidence among the immigrants quickly (within one generation) approach the incidence of the native population of the host country [3] with the largest increase occurring in risk of cancer in the distal colon [5]. The large international variation in incidence rates and the shift in sub-site distribution (proximal or distal segment of colon) after migration, indicate the importance of environmental factors and life style factors as a part of colorectal carcinogenesis.

Although cross cultural and migrant studies suggest that the majority of colorectal cancer is related to life style, including diet, traditional epidemiological studies of associations between exposures assessed e.g. based at questionnaires and outcome have mostly failed in identifying the exact environmental risk or beneficial factors. Understanding of differences in individual susceptibility and better exposure assessment might be crucial in identifying life style risk factors and possible interactions between susceptibility and exposures in relation to risk for colorectal cancer. This may be achieved by the use of biomarkers in molecular epidemiology as originally proposed by Perera and Weinstein in 1982 [6]. For colorectal cancer there is now extensive understanding of the molecular changes in crucial genes and the relevance of mutations, especially in hereditary syndromes.
Environmental factors are likely to cause damage to DNA through direct binding of metabolites (adduct formation) or oxidative stress, whereas repair of such lesions and defence against oxidative stress could be crucial. Single nucleotide polymorphisms result in substantial variation in the capacity of these mechanisms and may be important biomarkers of susceptibility to cancer.

**Hypotheses and Aims**

The aims of the work underlying this Ph.D.-thesis was to evaluate whether single nucleotide polymorphisms in genes involved in defence of oxidative DNA damages (GPX and OGG1), repair of DNA adduct lesions (XPD, XPA, XPC, and ERCC1), and a previous identified high risk haplotype (encompassing polymorphisms in ERCC1, RAI, and ASE-1) were associated to risk of colorectal cancer, and furthermore to assess if the polymorphisms modify the association between various life style factors and risk of colorectal cancer development. I chose to address the following three hypotheses:

- Polymorphisms in genes involved in defence of oxidative DNA damages are associated with risk of colorectal cancer (manuscript I and II). Any association between life style factors and colorectal cancer development may be modified by the genotypes (manuscript II)
- Polymorphisms in genes involved in repair of DNA adduct lesions are associated with risk of colorectal cancer. Any association between life style factors and colorectal cancer development may be modified by the genotypes (manuscript III)
- A haplotype encompassing polymorphisms in the genes RAI, ASE-1 and ERCC1 are associated with risk of colorectal cancer. Any association between life style factors and colorectal cancer development may be modified by the haplotype or the single genotypes (manuscript IV)

Each of the hypotheses is addressed in the four manuscripts mentioned. Three studies investigate the risk of colorectal cancer among participants in the Danish prospective “Diet, Cancer and Health” cohort study (manuscript II-IV), while one study investigates the risk of adenomas and colorectal cancer among participants in the Norwegian case-control study “Kolorektal cancer, Arv og Miljø” (manuscript I). Similar hypotheses related to lung cancer among participants in the “Diet, Cancer and Health” cohort are addressed in manuscripts V-VII.

This Ph.D.-thesis will examine the molecular and cellular mechanisms leading to colorectal cancer and will review the current literature on polymorphisms in the genes studied in manuscript I-IV on their association with risk of colorectal adenoma and colorectal cancer. Association with various other types of cancer is examined briefly, including results from manuscript V-VII.
COLORECTAL CANCER

Molecular Epidemiology of Colorectal Cancer

Molecular epidemiology uses the same paradigm as traditional epidemiology in addition to using biological markers of exposure or susceptibility. The International Agency for Research on Cancer (IARC) has defined a biomarker as “any substance, structure, or process that can be measured in the body or its products and may influence or predict the incidence or outcome of disease” [7]. Biomarkers can be classified into markers of exposure, effects and susceptibility: Biomarkers of exposure include measures of internal or biological effective dose of a compound related to a certain exposure. Biomarkers of effect include measures of biochemical alterations within the organism, and biomarkers of susceptibility include measures of indicators of an organism’s sensitivity towards exposure [8]. Biomarkers of susceptibility include polymorphisms in genes involved in metabolism, cell cycle and DNA repair [8]. A genetic polymorphism is defined as a variation in the nucleotide sequence present in at least 1% of the population.

The relationship between the different biomarker categories are illustrated in Figure 1.

![Figure 1: The causal and mechanistic pathway from exposure to disease described by biomarkers of exposure, biological effects and susceptibility. Examples of biomarkers are indicated by a bullet. Adapted from [9]](image-url)

There is a continuous transition from biomarkers of exposure to biomarkers of effect, while susceptibility factors, such as polymorphisms in genes involved in e.g. metabolism, may affect biomarkers of both exposure and effects.
DNA adducts is an important biomarker for exposure of genotoxic carcinogens, as it gives the biologically effective dose of the carcinogen that has reached the DNA. Additionally, the level of DNA adducts are suggested to be indicative of the risk of cancer associated with the exposure [10;11].

Reactive oxygen species (ROS) are constantly generated endogenously as by-products from cell metabolism and in response to external factors from diet and life style. If ROS are formed in amounts that exceed the capacity of the antioxidant defence system, oxidative stress is said to occur in the cell, which may result in lipid peroxidations, oxidative protein damages and DNA lesions. Experimental animal studies and in vitro studies suggest that oxidative DNA damage is important in carcinogenesis [12;13], but the association are not firmly established for the carcinogenesis in humans.

In the present thesis, single nucleotide polymorphisms (SNPs) in genes involved in defence of oxidative DNA damages and repair of DNA adduct lesions (and induction of apoptosis) are evaluated as possible predictive biomarkers of susceptibility for colorectal cancer. Additionally, gene-environment interactions between the SNPs and possible life style risk and beneficial factors are studied in relation to development of colorectal cancer.

**DNA Adducts**

Extensive research has examined the association between exposure of N-nitroso compounds (NOCs), polycyclic aromatic hydrocarbons (PAHs) and heterocyclic aromatic amines (HCAs), assessed by means of any biomarker, and the risk of cancer in humans.

N-nitroso compounds (NOCs) are alkylating agents able to react with DNA and form adducts. More than 85% of 300 NOCs tested for carcinogenicity in experimental animals were observed to be carcinogenic [14], but epidemiologic studies have been inconclusive in finding association between the exposure of NOCs and risk of various cancer forms in humans [15-17], although an increased endogenous production of NOCs, suggested primarily by bacterial catalysis, are proposed associated to the etiology of colorectal cancer [18]. NOCs are present in tobacco smoke and in nitrate- or nitrite-treated meats [19;20].

Polycyclic aromatic hydrocarbons (PAHs) and heterocyclic aromatic amines (HCAs) constitute a major class of chemical carcinogens present in the environment. When metabolically activated, these compounds act as mutagens and carcinogens in animal models [21-23] and are able to form bulky DNA adducts in humans ([24] and [25]). Many PAHs and HCAs are found tumourigenic in humans or
experimental animals [26]. Cooking meat at high temperatures and certain preservation and processing procedures leads to the formation of PAHs and HCAs [27;28]. PAHs are ubiquitous environmental contaminants formed by incomplete combustion of organic matter. They are one of several classes of carcinogenic chemicals present in tobacco smoke [29;30]. PAH compounds may not only be formed by high cooking temperatures but are also found in uncooked food, like sea food and plants, due to contamination of the aquatic environment [31] or via atmospheric exposure [28].

The nucleotide excision repair (NER) pathway is the primary mechanism for removal of bulky adducts from DNA. Some of the contributors are the proteins xeroderma pigmentosum complementation group A, C, and D (XPA, XPC and XPD), and excision repair cross complementary group 1 (ERCC1). The NER pathway and the biological function of the four proteins are described in detail in the chapter “Nucleotide Excision Repair”.

Oxidative DNA Damages
Oxidative DNA lesions is one of the most diverse classes of biomarkers of oxidative damage, with nearly 100 different damages identified ranging from modified bases, formation of DNA adducts to double strand breaks [32;33].

An increased load of ROS may cause higher levels of 8-oxo-7, 8-dihydroguanine (8-oxoG) in human colorectal carcinoma compared with non-malignant tissue [34]. 8-oxoG is a strongly mutagenic lesion due to a mispair with adenine during DNA replication leading to G:C to T:A mutations, and is the most widely used biomarker of oxidative DNA damage because of it’s relatively easy detectability [35;36]. The protein 8-oxoguanine glycosylase 1 (OGG1) is a bifunctional glycosylase involved in the DNA base excision repair (BER) pathway that specifically removes 8-oxoG paired with cytosine from the DNA backbone [37-39]. Most recently another DNA glycosylase able to repair 8-oxoG has been identified and called OGG2. This enzyme is able to remove the oxidised guanine only in 8-oxoG:A mispairs [40].

Antioxidant enzyme systems are part of the first line defence against ROS in all cellular compartments and extracellularly. Some of the most important of these enzymes are glutathione peroxidases (GPX). GPX are involved in the defence against oxidative DNA damages by reduction of ROS in concert with the enzymes superoxide dismutases (SOD), catalases (CAT) and glutathione reductase (GR). GPX is a selenium-dependent antioxidant enzyme that reduces $H_2O_2$ and lipid peroxides/hydroperoxides by
oxidizing glutathione. Four isotypes have been characterized: GPX1-GPX4, of which GPX1 and GPX2 are expressed in colon and rectum [41].

Mice with disrupted GPX1 and GPX2 genes are more susceptible to colon cancer induced by inflammation caused by bacterial colonization [42] than are wild type mice. And OGG1 knock out mice have higher 8-oxoG content in the DNA [43;44] and higher rates of G:C to T:A transversions than wild type mice [43;45;46]. This suggests the two genes, OGG1 and GPX1, to play an important part in the defence of oxidative stress and the related oxidative DNA damages.

Life Style Factors and DNA Damages

Several life style factors and dietary components are suggested to be associated with risk of colorectal cancer, listed in Table 1. The associations may possibly be due to an increasing level of DNA adducts and oxidative DNA damages.

Table 1: Possible environmental risk and beneficial factors of colorectal cancer and their possible association with oxidative DNA damages and DNA adduct formation. Arrows indicate positive (↑), no (→) or negative (↓) association with risk of colorectal cancer, oxidative DNA lesions or DNA adduct formation.

<table>
<thead>
<tr>
<th>Life style factor</th>
<th>Risk of CRC</th>
<th>Oxidative DNA lesions</th>
<th>DNA adduct formation</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Air pollution</td>
<td>↑</td>
<td>↑</td>
<td>↑</td>
<td>[11-15]</td>
</tr>
<tr>
<td>Tobacco</td>
<td>↑</td>
<td>↑</td>
<td>↑</td>
<td>[11,44]</td>
</tr>
<tr>
<td>Alcohol</td>
<td>↑</td>
<td>↑</td>
<td>↑</td>
<td>[42]</td>
</tr>
<tr>
<td>Red meat</td>
<td>↑</td>
<td>↑</td>
<td>↑</td>
<td>[45,51]</td>
</tr>
<tr>
<td>Processed meat</td>
<td>↑</td>
<td>↑</td>
<td>↑</td>
<td>[46,50]</td>
</tr>
<tr>
<td>Protein</td>
<td>↑</td>
<td>↑</td>
<td>↑</td>
<td>[47]</td>
</tr>
<tr>
<td>Fat</td>
<td>↑</td>
<td>↑</td>
<td>↑</td>
<td>[48,53]</td>
</tr>
<tr>
<td>Poultry</td>
<td>→</td>
<td>↑</td>
<td>↑</td>
<td>[49]</td>
</tr>
<tr>
<td>Fish</td>
<td>↓</td>
<td>↓</td>
<td>↓</td>
<td>[48,50]</td>
</tr>
<tr>
<td>Vegetables</td>
<td>↓</td>
<td>↓</td>
<td>↓</td>
<td>[49,52,53]</td>
</tr>
<tr>
<td>Fruit</td>
<td>↓</td>
<td>↓</td>
<td>↓</td>
<td>[49,52,53]</td>
</tr>
<tr>
<td>Dietary Fiber</td>
<td>↓</td>
<td>↑</td>
<td>↑</td>
<td>[49,53]</td>
</tr>
<tr>
<td>HRT use</td>
<td>↓</td>
<td>↑</td>
<td>↑</td>
<td>[52,58]</td>
</tr>
<tr>
<td>NSAID use</td>
<td>↓</td>
<td>↓</td>
<td>↓</td>
<td>[53,60]</td>
</tr>
<tr>
<td>Physical activity</td>
<td>↓</td>
<td>↑</td>
<td>↑</td>
<td>[49,63]</td>
</tr>
</tbody>
</table>

Air pollution is not an established risk factor for development of colorectal cancer in humans, although several studies have shown higher risk among workers exposed to diesel exhaust [47]. Some studies have found an association between ambient air pollution and DNA adduct levels [48-53], whereas others failed to find such an association [54,55]. DNA adduct levels are increased following occupational exposure
among foundry and coke oven workers and among workers exposed to diesel exhaust [56-63], while among fire-fighters [64], traffic exposed policemen [65] and aluminium workers [66], no associations between occupational exposures and DNA adducts have been found. Exposure to ambient air particles and benzene has consistently been associated with oxidative DNA damages, e.g. high levels of 8-oxoG in lymphocytes [67-70].

Tobacco smoking may possibly be a risk factor for development of adenomas [71], but an association between tobacco smoking and risk of colorectal cancer has not been established. Following tobacco smoking, adducts formed by metabolites of NOCs and PAHs are not only located in airway tissue, but are also found in bladder and cervical tissue from smokers [29;30]. Higher levels of 8-oxoG and other oxidative bases or strand breaks has been observed in leukocyte DNA from smokers compared with nonsmokers, although this observation is far from consistent [72-75].

A growing body of evidence supports that avoidance of alcohol is recommended to prevent colorectal cancer [76]. Acetaldehyde is the primary oxidative metabolite of ethanol. Acetaldehyde and malondialdehyde, the end-product of lipid peroxidation by reactive oxygen species, can combine to form the malondialdehyde-acetaldehyde adduct, which is very reactive and avidly binds to DNA [77]. The level of acetaldehyde DNA adducts in white blood cell DNA in alcohol abusers have been measured up to 13-fold higher than in subjects from the non-drinking control group [78].

There is some evidence for adverse associations of intake of red and processed meat with risk of colorectal cancer [79-81]. The elevated risk may be due to an increased endogenous production of NOC, which may enhance the colonic formation of the DNA adduct O6-carboxymethyl guanine [18;82]. Cooking meat at high temperatures leads to the formation of polycyclic aromatic hydrocarbons (PAHs) and heterocyclic amines (HCAs) [27]. Additionally, intake of charbroiled or smoked meat may be associated with increased levels of DNA adducts [83-86], due to HCAs and PAHs [87-90]. The levels of some HCAs and PAHs are comparable for red meat, fish and poultry smoked or cooked at high temperatures [91;92]. Intake of red meat, but not of fish and poultry, increases the luminal contents of N-nitrosocompounds (NOCs) in colon [18;82]. The increase in endogenous N-nitrosation can be attributed to heme iron [93], which is 10-fold higher in red meat than in white meat [94]. An increase in the ratio of the consumption of red meat to consumption of fish/chicken was associated with an increase in colorectal polyp risk [95]. Colorectal cancer risk may be negatively associated with fish intake [81]. Intake of fish are reported to be negatively associated with DNA adduct levels [96;97], although another study
found no effect [98;99]. The protective effect of fish intake are suggested to be due to the content of n-3 poly-unsaturated fatty acids in fish [79].

High intake of dietary fat has been associated to an increased risk of proximal cancers [100;101], while high intake of protein has been associated with an increased incidence of distal cancers [101].

There is limited evidence for a preventive effect of intake of fruit and vegetables for cancer in colon and rectum [102]. Intake of fruit, vegetables or antioxidant vitamins have been shown to be negatively associated with DNA adduct levels [103-106], although some studies found no effect [107;108] and one study found an effect of increased vitamin intake only in females [109]. The significance of dietary fibres as a protective factor against colorectal cancer remains controversial. However, a large study of European populations (the EPIC study) including 519,978 individuals have confirmed intake of dietary fibres to be protective [110]. To my knowledge no studies has been published concerning intake of fibres and the level of oxidative DNA lesions or DNA adduct formation.

Meta-analyses studies showed that the risk of colorectal cancer, colon cancer risk in particular, was lower among recent postmenopausal users of hormonal replacement therapy (HRT) [111;112]. Regular use of aspirin or non-steroid anti-inflammatory drugs (NSAIDs) appeared to reduce the risk of colonic adenoma [113] and colorectal cancer [113;114], especially if used in high doses for more than 10 years. The beneficial effect of NSAIDs has been ascribed to the inhibition of cyclooxygenase-2 (COX-2), the enzyme responsible for the production of various inflammatory prostaglandins [115]. COX-2 reaction may cause DNA oxidation [116].

A high level of physical activity has been associated to reduced risk of proximal, but not distal colon cancer among Norwegian men [117]. In a recent study of the DCH cohort no association of physical activity with risk of colon cancer was observed [118]. Exercise may modulate oxidative DNA damage; strenous activity may increase the damage level [119;120], whereas moderate daily exercise are found to reduce the level of 8-oxoG in leucocytes [121]. Consistent body of evidence from prospective studies indicate that overweight and obesity increase risk of colon cancer [79].

Lately, focus has turned from single risk factor analyses towards gene-environment interactions in cancer development. Gene-environment interaction can be defined as a different effect of an environmental exposure on risk of disease in people with different genotypes, or a different effect of a genotype on risk
Interaction applies when one stratum (ex. carrier of high risk allele) responds differently to an exposure (ex. a dietary component) than another stratum (ex. carrier of low risk allele). The present study of gene-environment interactions in relation to risk of colorectal cancer are focused on life style factors as diet, tobacco smoking and alcohol consumption, and single nucleotide polymorphisms (SNPs) in genes involved in defence of oxidative DNA damages (GPX and OGG1), repair of DNA adduct lesions (XPD, XPA, and XPC), and a previous identified haplotype with previous observed gene-environment interactions in relation to risk of cancer (the haplotype encompassing polymorphisms in ERCC1, RAI and ASE-f).

**Morphology and Histology of Colon and Rectum**

Approximately 60% of colorectal cancer cases arise in the distal part of colon (including splenic flexure, descending colon, sigmoid and rectosigmoid colon and rectum) in countries where colonic cancer incidence is high, whereas proximal (including cecum, ascending colon, hepatic flexure and transverse colon) cases predominate in countries with low incidence [3;123]. The anatomy of colon and rectum are illustrated in Figure 2.

In the Danish population the anatomical distribution of colorectal cancer diagnosed in 2001-2005 was comparable with the rude estimate, with 30.6%, 35.2% and 34.2% of the diagnoses in the right segment of colon, the left segment of colon and in rectum, respectively [124]. Tumours in the hereditary Lynch syndrome occur predominantly in the proximal segment of colon, while the hereditary syndrome Familial adenomatous polyposis (FAP) occur predominantly in the distal segment of colon [125].

It has been suggested that the risk of colorectal cancer conferred by various environmental (and genetic) factors is different for proximal and distal tumours. Various physiological and histological differences exist between the proximal and distal part of a normal colon, which may predispose tumours originating at these sites to develop along different pathways. It may be convenient to categorize colorectal cancers into either proximal or distal location, but it is important to note that this is a simplification of colorectal
carcinogenesis, and that underlying molecular features are responsible for determining tumour phenotype. These features may very likely show considerable overlap between right- and left-sided colorectal cancers.

The principal functions of colon are recovery of water and propulsion of solid faeces to the rectum prior to defaecation. The luminal surface of the intestine are composed of a columnar epithelial mucosa, with finger-like projections (villi) and glandular invaginations (crypts). Mucosa consists mainly of two cells types: the absorptive cells recovering water and some salts from the liquid residue of the contents of the small intestine, and the mucus-secreting goblet cells lubricating the passage of faeces. Goblet cells predominate at the base of the villi, whereas the luminal surface is almost entirely lined by columnar absorptive cells. The cells of the intestinal epithelium are progressively more differentiated as they age and pass along the crypt–villus axis. The rectum is the short dilated terminal portion of colon. The rectal mucosa is similar to that of colon except from more numerous goblet cells.

The proximal colon originates from the embryonic midgut and is perfused by the superior mesenteric artery, surrounded by a multilayered capillary network, whereas the distal colon derives from the hindgut and is served by the inferior mesenteric artery, surrounded by a single-layered capillary network. The multilayered capillary network in proximal colon is possibly related to the greater water absorption and electrolyte transport capacity [126;127]. The average villi length is greater in the distal colon than in the proximal colon [128]. The apoptotic index is lower in the right colon compared to the left colonic mucosa [129].

Gastrointestinal stem cells undergo multi-potent division to produce the entire specialised cell repertoire of the gastrointestinal tract. The numbers and location of stem cells in the intestinal crypts and gastric glands have never been conclusively proven, and, consequently, the clonal origins of these structures under normal circumstances and in neoplasia are clouded issues. Intestinal stem cells are primitive cells located in a specialised compartment consisting of epithelial and mesenchymal cells and extra-cellular substrates that lack expression of any definitive markers of lineage commitment and are therefore difficult to define and to characterise morphologically. It is believed that the surrounding mesenchymal cells regulate stem cell behaviour through paracrine secretion of growth factors and cytokines [130]. The number of stem cells within the compartment is believed to be between four and six [131;132], but the exact number has never been conclusively proven and, consequently, is the topic of debate still. It has been postulated that stem cell number may fluctuate throughout the crypt cycle [133] and that the stem cell number varies throughout different regions of the gastrointestinal tract [134]. Monoclonal intestinal
crypts have been demonstrated following irradiation, showing that a single multipotent surviving stem cell can regenerate an entire crypt, thus confirming the hypothesis, that the epithelial cell lineages of the gastrointestinal tract are clonal populations derived from a single stem cell [135;136], albeit in damaged mucosa [137]. No evidence of any crypts with a mixed phenotype was observed in 2260 crypts located at the periphery of a patch, indicating that colonic crypts are indeed monoclonally derived, which is consistent with results obtained previously [138]. However, conflicting data have emerged from different studies, and the pathways and mechanisms of gastrointestinal neoplasia are thus far uncertain.

The turnover of cells in the gastrointestinal tract is high throughout life with the differentiating cells shed into the lumen and replaced every 2–7 days under normal circumstances. Thus, lifespan of the cells are not sufficient to accumulate the mutations necessary for malignant change, why the perpetual stem cell is widely believed to be the target of mutational changes [139-141]. A stem cell division can produce one stem cell and one daughter cell (asymmetric division), two stem cells by self-replication (symmetric division) or a stem cell loss, where both daughter cells go on to differentiate (symmetric division) [142]. The majority of divisions are thought to be asymmetric [143]. According to the so-called immortal strand hypothesis there may be a retention of the template DNA strand within the stem cell located in the niche [143], which allows any DNA replication errors to pass into the differentiating, shortlived daughter cell affording a mechanism of stem cell genome protection [144]. If indeed stem cells are the original targets for the mutation(s) required to initiate a neoplasm, then whether such a cell acts alone or in cooperation with other mutated stem cells becomes important.

The stem cell compartment is believed to be at the origin of the crypt–villus axis (reviewed in [145]). However, as mentioned before the location of the gastrointestinal stem cells is debated. Studies by Wright have suggested a location in the mid crypt of the ascending colon and in the base of the crypt of the descending colon [146], whereas different observations have been made in other studies.

It has been suggested that a crypt would be incited to go into fission when it reached a threshold size. However, the stem cell number is now thought to be the important factor [146].
Morphology and Histology of Polyps in Colon and Rectum

A polyp is defined as a mass that protrudes into the lumen of the colon. Polyps may be non-neoplastic or neoplastic. The non-neoplastic polyps are hyperplastic, inflammatory, juvenile or hamartomatous and lack dysplastic features. Adenomatous polyps are benign neoplasms that, by definition, display some dysplasia. The degree of dysplasia may be graded into mild, moderate and severe on the basis of cytological and structural features. Adenomatous polyps are generally believed to be precursors of most colorectal adenocarcinomas, which is supported by epidemiological, genetic and pathological studies [147]. Patients with adenomatous polyps have a higher risk of colon cancer over the general population and the risk increases if the polyps are multiple [148].

Neoplastic polyps are histological divided into three sub-groups: tubular adenomas, villous adenomas and mixed or tubulo-villous adenomas. The risk of malignant transformation is low in tubular adenomas (2-3%) and high in pure villous adenomas (15-25%), while the mixed adenomas have an intermediate risk of malignant transformation [149]. The risk of developing subsequent cancer is generally believed to be higher in patients with polyps larger than 1cm in diameter [150-152]. The initiated polyp may be present and proliferate for 10-15 years before undergoing malignant transformation [153;154].

The earliest and smallest recognizable histopathological entity may be an aberrant crypt focus (ACF). Two types of ACFs have been observed in humans: The common one called the hyperplastic or non-dysplastic crypt being a hypercellular crypt with normal individual cells which is unlikely to lead to clinically significant lesions, and the less common one called dysplastic ACFs, which are believed to be the precursors of the adenomas and carcinomas [154-156].

There are currently two proposed morphological pathways of spontaneous development of adenomas, the “bottom up” and the “top down” pathways (illustrated in Figure 3). The gastrointestinal stem cells are important players in each of them.

In the “bottom-up model” a stem cell situated in the base of the crypt acquires mutations in the tumour-suppressor gene adenomatous polyposis coli (APC), which thereby impairs the function of the APC protein (a). The mutated cell proliferates and produces neoplastic daughter cells, which migrate upwards to colonise the entire crypt (b) and form a monocryptal adenoma [157]. Further expansion is achieved by crypt fission (c) [158], where crypts undergo bifurcation (division into two) followed by longitudinal division, with the ultimate formation of two daughter crypts. Thus, this model involves monocryptal
adenomas, where the dysplastic cells occupy an entire single crypt. These lesions are observed to be common in FAP [159].

The “top-down model” is based on observations of dysplastic cells only located at the luminal surface of the crypts (d) [1;160;161], along with migration of adenomatous cells from the surface to the base of the crypt (e) [161]. In this model an initial stem cell mutation is proposed to occur in the epithelial mucosa situated in the intra-cryptal zone, between two crypt orifices, with subsequent stem cell division producing a mutant clone which expands laterally and downwards into the crypt, and thereby displacing the normal epithelial cells (f) [1]. Analysis of four single-nucleotide polymorphisms (SNP) within the APC gene in tissue from sporadic adenomas showed loss of heterogeneity (LOH) of APC in cells in the upper portion of the crypts, while no LOH was observed in the histological normal crypt bases [1]. Cells towards the top of the crypt display high proliferation activity [1;162]. These observations led to two hypotheses for a top-down model instead of the conventional bottom-up model: The stem cell could be located in the intra-cryptal zone [1], or if located in the base of the crypt the APC mutation in the stem cell would prevent it from a terminal differentiation and alter the cell’s migration dynamics, migrate to the luminal surface and allowing it to remain in the mucosa before expanding laterally and downwards [163].

Both models (“top-down” and “bottom-up”) may possibly occur. However, the bulk of evidence indicates, that the gastrointestinal stem cells are located in the base of the crypt [146], with no indication of a stem cell population in the intra-cryptal zone, and so the modified top-down hypothesis is proposed; that a stem cell in the crypt base acquires a mutation and subsequently migrates to the intra-cryptal zone, whereupon it undergoes neoplastic expansion [1].

A crypt cycle, the time from a crypt “born” by crypt fission until they divide by crypt fission themselves, takes approximately 9-18 years in the human colon [164;165]. Studies on the methylation patterns of adjacent crypts showed significant inter-crypt variation, both in adjacent crypts and in those up to 15 cm
apart, which may be a consequence of the time taken for crypts to divide, allowing neighbouring crypts to develop different methylation patterns during the process [166].

Identification of the origins, location, and molecular regulators of the intestinal stem cell will provide a clearer understanding of the genetic pathways and cell signalling involved in the neoplastic changes in colorectal carcinogenesis. The stepwise pattern of mutational activation of oncogenes and inactivation of tumour suppressor genes that causes adenomas to develop to adenocarcinoma are called the adenoma–carcinoma sequence [147;167].

The Adenoma-Carcinoma Sequence

The progression of normal tissue through dysplasia to tumour tissue involves numerous steps. It is estimated that a typical colorectal tumour contains at least 11,000 genomic alterations [168]. Two distinct pathways have been suggested in colorectal carcinogenesis. One involves chromosomal instability, which is characterized by allelic losses in chromosome 5q (\textit{APC}), 17p (\textit{p53}) and 18q (\textit{DCC/SMAD4}), and the other involves microsatellite instability (MSI).

The initial mutations in most of the cases occur at the \textit{APC} tumour-suppressor gene locus (5 q21- q22). Loss of \textit{APC} tumour suppressor gene function is thought to be one of the first genetic changes in colorectal adenoma development. \textit{APC} encodes a large multifunctional cytoplasmic protein [169], which is an essential component of a “destruction complex” in the Wnt pathway involved in the binding and down-regulation of beta-catenin and thereby preventing excessive cell proliferation. Additionally, \textit{APC} are involved in regulation of apoptosis, cell-cycle progression and chromosomal stability (reviewed in [170-173]). Hence, the importance of the \textit{APC} protein in a number of different regulatory functions in cells in colon means, that mutation in the \textit{APC} gene alone may be sufficient to provide a stem cell with a selective growth advantage [174] by allowing unregulated activation of Wnt signalling. Hundreds of specific \textit{APC} mutations have been characterised, and the position of the mutation appears to dictate the severity and onset of the hereditary syndrome FAP [175]. Patients with FAP have an autosomal dominant inherited germline mutation of \textit{APC} and are therefore susceptible to mutation of the remaining wild-type \textit{APC} allele [176]. FAP is characterized by the presence of hundreds of polyps in the large bowel. These arise first in the rectum and distal colon before extending to more proximal segments. Close to
100% of FAP individuals will develop colorectal cancer in the distal colon. The mutations and genetic occurrences in the adenoma-carcinoma sequence are summarized and illustrated in Figure 4.

Mutations in APC are found in 63% of sporadic adenomas [177] and up to 80% of sporadic colorectal tumours [175;178]. Mutations in beta-catenin, that prevents the breakdown of the protein, can also promote adenoma initiation; however, small adenomas with beta-catenin mutations alone do not progress to larger adenomas or carcinomas as frequently as adenomas with APC mutations [179]. The P53 gene, located on chromosome 17p, is a tumour suppressor gene and is frequently lost in colorectal malignancy. The gene encodes for a DNA-binding phosphoprotein that prevents progress past the G1-phase of the cell cycle if DNA damage has occurred [180;181]. It is also characterized as a transcription factor, activating and promoting expression of genes involved in growth inhibition. The protein p53 is involved in several essential cell functions including control of the cell cycle, DNA repair and apoptosis, and thus is called the “guardian of the genome”. The half-life of wild type p53 protein and mutant p53 protein is approximately 20 minutes and 24 hours, respectively. The extended half-life of mutant p53 allows it to accumulate in the nucleus and be over-expressed in tumours [182]. Mutations of P53 are found in more than 50% of all human cancers and in more than 75% of colorectal adenocarcinomas [167].

It is debated whether the gene “deleted in colorectal carcinogenesis” (DCC) is a candidate tumour-suppressor gene. The DCC gene is deleted in more than 70% of colorectal carcinomas [183;184]. A second candidate tumour-suppressor gene, DPC4/Smad4, located in the same region on 18q21, is deleted in up to a third of the cases [185]. The protein family SMAD are intracellular proteins that mediate the effects of signaling from extracellular transforming growth factor beta (TGF-β) and TGF-β-related factors [186].

Microsatellite instability (MSI) is explained by defects in DNA mismatch repair (MMR) genes, encoding proteins involved in recognition and repair of single base lesions and larger strand slippage mismatches in
DNA replication. In sporadic colorectal cancer MSI usually arises due to epigenetic silencing of the DNA mismatch repair gene MutL homologue 1 MLH1 [187] by methylation of cytosine and guanine residues in CpG-rich promoter regions [188;189], which prevents the gene-regions from being transcribed. MSI causes the Lynch syndrome primarily by a germline mutation in the mismatch repair genes MutS homologue 2 (MSH2) and MLH1 [167]. The life time risk of developing colorectal cancer is up to 75% higher in children with Lynch syndrome compared with the general population [190;191]. Approximately 70% of large bowel tumours in patients with Lynch syndrome arise in the right/proximal colon [192].

The two pathways in the adenoma-carcinoma sequence, involving the chromosomal instability and microsatellite instability, seems well characterized. However, recent molecular studies have shown that colorectal carcinogenesis is not necessarily clearly divided into these two pathways, and may include other routes like the transforming growth factor beta (TGF-β)/SMAD-pathway, the serrated pathway and the epigenetic pathway. The TGF-β family are known inhibitors of gastrointestinal epithelial cell proliferation. Under normal circumstances TGF-β are involved in phosphorylation of two cytoplasmic proteins, Smad2 and Smad3, following a formation of a heteromeric complex with Smad4. This complex translocates to the nucleus where it induces TGF-β target gene transcription [193]. Disruption of the TGF-β/Smad signalling pathway causes up-regulation of epithelial cell proliferation which may lead to tumorigenesis. Smad2 and Smad4 are frequently inactivated in human cancers confirming their function as tumour suppressor genes [194]. The serrated pathway is characterized by early involvement of oncogenic mutations in the BRAF or KRAS genes and excess CpG island methylation [195]. K-Ras and B-Raf are participants in a pathway regulating cell growth, differentiation and apoptosis (the MAPK-ERK pathway) [196].

Recently, a wealth of studies has implicated alterations in the epigenome, as also being important in cancer formation [197-199]. Epigenetics refers to heritable modifications to DNA that regulate gene expression without involvement of change in the DNA sequence. These modifications are amendments or chemical modifications to the DNA that includes global hypomethylation at repetitive sequences in satellite or pericentromeric regions, focal hypermethylation at CpG islands, histone modifications by deacetylation and methylation of amino acids in the histone tails (reviewed in [200]) and DNA alkylation by methylation of guanine [201]. A new aspect of recent studies of epigenetic alterations in cancer is the observation that some genes that are involved in DNA repair (mismatch repair) are commonly found to be aberrantly methylated in the early stages of tumours [202].
DNA REPAIR

DNA is constantly attacked by exogenous and endogenous agents causing DNA modifications or damages. If these DNA lesions are left un-repaired, they may contribute to mutagenesis and oncogenesis. Thus, DNA repair constitutes a first line of defense against cancer. Subtle variations in DNA repair capacity may be caused by commonly occurring polymorphisms in the DNA repair genes. The polymorphisms may thereby have an impact on individual genetic susceptibility to cancer. The two repair mechanisms base excision repair (BER) and nucleotide excision repair (NER) will be introduced in the following two chapters and the published literature on the SNPs investigated in the work underlying this Ph.D.-thesis will be summarised. Besides BER and NER there are two other well defined repair pathways, recombination repair and mismatch repair, which will be presented shortly.

Base Excision Repair

Base excision repair (BER) is the major repair pathway involved in removal of small lesions on DNA, like fragmented or non-bulky adducts, alkylation/methylation or oxidation of bases. This repair pathway can be subdivided into five steps: 1. Base removal by a specific DNA glycosylase, 2. Incision at the abasic site by an AP-endonuclease, 3. Processing of the produced blocked termini, 4. Resynthesis to fill in the gap, and 5. Resealing of the previous damaged DNA strand.

The first step of BER involves DNA glycosylases removing the damaged base. Three or four DNA polymerases are suggested to be involved in the BER pathways: Pol β, δ, ε and possibly Pol λ. The major BER polymerase is Pol β [203-205]. The glycosylases are classified as monofunctional and bifunctional: Monofunctional polymerases, like uracil-DNA glycosylase, excise the damage base from the DNA base stack by hydrolysing the N-glycosyl bond between the damaged base and the sugar moiety. This leads to formation of an abasic or apurinic/apyrimidinic (AP) site [206], which is substrate for the following action by an AP-endonuclease. Bifunctional polymerases, like 8-oxoguanine-DNA glycosylase (OGG1), have an associated β-lyase activity, which enables them to not only excise the damaged base, but also incise the DNA backbone 3´ to the abasic/AP site. The resulting single strand break are converted to be harboring a 3´-hydroxyl group prior to polymerization and/or ligation. Next, AP endonuclease 1 (APE1) recognizes the AP site, cuts the phosphodiester backbone 5´ to the AP site, and thereby leaving a 3´-hydroxyl group and a 5´deoxyribose phosphate (dRp) group at the borders of the nucleotide gap.
Further repair proceeds by two subpathways, illustrated in Figure 5: short-patch BER, that replaces one nucleotide, or long-patch BER, that may fill the repair gap with up to 6 nucleotides [208]. Both pathways are initiated by Pol β. Pol β’s binding to the AP site are facilitated by interaction with APE1 [209]. During short-patch BER one nucleotide are added into the repair gap and the 5’-dRp moiety are removed by Pol β. The remaining nick is sealed by a complex of X-ray cross-complementing 1 protein (XRCC1) and DNA ligase 3α. During long-patch BER the 5’-dRp moiety are removed by Pol β, and the first nucleotide is added to the repair gap [210], but further DNA synthesis is suggested to be conducted by Pol δ or ε [203;208;211] requiring proliferating cellular nuclear antigen (PCNA), flap endonuclease 1 (FEN1), DNA ligase I and possibly replication factor C (RF-C) [212]. Pol δ or ε adds several downstreams nucleotides to the 3’ end of the first added nucleotide, generating a flap containing the 5’-sugar phosphate. FEN1 cleaves the displaced oligonucleotide (the generated flap). FEN1 interacts with PCNA [213;214], that interacts with APE1 [215], why APE1 may be the factor to recruit the two proteins to the repair site. The assembling of PCNA around the DNA requires RF-C [216]. Resealing of DNA is induced by influence of APE1, enhancing the enzymatic activity of FEN1/PCNA and DNA ligase I [217].

Recent studies indicate that several more proteins are involved in BER. Poly (ADP-ribose) polymerase-1 (PARP-1) is observed to bind to the incised AP site at the very early stages of single strand break repair [218;219]. Following PARP-1 binding and dissociation, repair of single strand breaks is suggested to always be followed by a specific protein that is required to progress the particular lesion to the next stage of repair. An example of specific protein being polynucleotide kinase (PNK), which is required for the
initiation of repair of 3´-phosphate containing single strand breaks [220]. Although no stable complexes have been identified, immunoprecipitation experiments of BER protein-complexes have shown variable results: Pol β/DNA ligase I/Uracil DNA glycosylase (UNG) [204], UNG/APE1/Pol β/PoI δ/XRCC1/DNA ligase/PCNA [221] and DNA ligase III/XRCC1/PNK/Pol β [222], indicating that the highly coordinated interactions between the proteins may be more complex than in the model described above.

The studies underlying this Ph.D.-thesis includes the polymorphisms in genes involved in defence of oxidative DNA damages: OGG1 Ser326Cys and GPX1 Pro198Leu. Additionally, a polymorphism positioned in 5´UTR of GPX1 and a polymorphism in the 3´UTR of the gene RHOA positioned in close vicinity to the GPX1 gene was analysed. We observed strong linkage disequilibrium between the GPX1 Pro198Leu polymorphism and the GPX1 5´UTR and RHOA 3´UTR polymorphisms, so the effect of one polymorphism was not discernible from the other. Thus, the analyses were focused on the well studied GPX1 Pro198Leu polymorphism.

Carriers of the variant allele of the GPX1 Pro198Leu polymorphism had a lower enzyme activity than homozygous carriers of the wild type allele (manuscript II). Similar findings were previously reported from a breast cancer study also nested within the “Diet, Cancer and Health” (DCH) cohort [223]. This suggests that the polymorphism modulates the GPX1 activity. In the two above mentioned studies, prospectively measured erythrocyte GPX activity did not affect the risk for cancer. Small case-control studies involving cancer patients have made following observations: The GPX activity was observed lower in erythrocytes from patients with gastrointestinal [224;225], oesophageal [225], prostate [226] or cervical cancer [227] compared to healthy individuals, while a higher activity level was observed in colorectal cancer tissue [228;229] and breast cancer tissue [230] compared to tissue from healthy individuals or adjacent healthy tissue, respectively.

Overexpression of OGG1 were found to suppress more than 95% of G:C to T:A transversions in the lung cancer cell line H1299 [44]. A lowered DNA expression of the OGG1 gene were observed in colorectal adenoma tissue compared to adjacent normal tissue [231]. The level of expression was comparable in adenoma tissue and carcinoma tissue, suggesting the increased gene expression is an early event in the colorectal carcinogenesis. The OGG1 Cys326Cys had a lower capacity than OGG1 Ser326Ser to prevent G:C to T:A transversions in a human lung cell line [44;232] or in vitro to repair oxidative DNA damages in human erythrocytes [233] and thereby a lower capacity to prevent mutagenesis by 8-oxoG.
However, the *OGG1* Ser326Cys polymorphism were observed not to modify the 8-oxoG specific lyase activity of *OGG1* *in vitro* in human colorectal carcinoma tissue [34] and lymphocytes [234]. A study by Luna and colleagues [235] showed that the *OGG1* was localized in the nucleoli during the S-phase and associated with condensed chromosomes during mitosis of the cell. They observed the *OGG1* Ser326Cys polymorphism to have an affect on the nucleolar localization of the protein, the *OGG1*-326Cys protein being excluded from the nucleoli in the S-phase, and co-localization to condensed chromosomes being altered during mitosis.

Overall, even though the studies are few and at times with contradictory results, the above mentioned studies of the *OGG1* Ser326Cys and *GPX1* Pro198Leu polymorphisms indicate that the polymorphisms may modulate the defence of oxidative DNA damages and may thereby possibly be associated with development of cancer.

Studies of genes involved in defence of oxidative DNA damage and risk of colorectal cancer are few. A search on the PubMed database of the National Center for Biotechnology Information (NCBI) on July 16th 2007 on the MeSH terms “polymorphism, single nucleotide AND colorectal neoplasms” resulted in 148 hits of which only two studies included polymorphisms in *OGG1* and no studies of polymorphism in *GPX1*. Combined with a search on the PubMed database of NCBI by different combinations of the words: “GPX OGG1 polymorphism colorectal colon rectum” 5 studies of SNPs in *GPX1* and *OGG1* in relation to risk of colorectal cancer or pre-stages to colorectal cancer were identified. The studies are listed in Table 2, including the results from manuscript I and II.

In the Norwegian case-control study [236] and the Danish case-cohort study, manuscript I and II, we observed no association between the *GPX1* Pro198Leu polymorphism and the risk of colorectal cancer or pre-stages of colorectal cancer. However, in the Danish study we observed an interaction between the polymorphism and alcohol consumption (P=0.02) with an incidence rate ratio (IRR) of 1.45 (95% confidence interval (CI): 1.16-1.81) for colorectal cancer per 10g alcohol intake per day among homozygous *GPX1* 198Leu carriers. Similarly, the results showed a borderline significant interaction (P=0.06) with smoking intensity, with an IRR=1.67 (CI: 1.06-2.65), for risk of colorectal cancer per 10g tobacco smoking per day among homozygous carriers of the variant allele. Additionally, an interaction with vitamin C intake was observed (P=0.04) with a lower risk of colorectal cancer, IRR=0.57 (CI: 0.34-0.95), per 100mg intake per day only among homozygous carriers of the wild type. To my knowledge, no other studies have been published on association of *GPX* polymorphisms with risk of colorectal cancer.
In our Norwegian study, manuscript I, carriers of the OGG1 326Cys allele had a lowered risk of adenocarcinomas, with an OR of 0.56 (0.33-0.95), compared to homozygous carriers of the wild type allele. No association was found for risk of adenomas [236]. The protective effect of the OGG1 polymorphism on risk of carcinomas was previously observed in an American study of colon cancer among a mixed group of Caucasian and African-American men, with an OR of 0.68 (CI: 0.45-1.02) and 0.41 (CI: 0.14-1.20) among heterozygous and homozygous carriers of the variant allele, respectively, compared to carriers of the homozygous wild type [237]. A contradictory result was obtained in a larger study by Moreno and colleagues (377 cases of colorectal adenocarcinoma and 329 cancer-free comparison individuals): The risk of colorectal cancer was higher among the youngest homozygous carriers of the variant allele, with an OR of 2.31 (CI: 1.05-5.09), compared to homozygous carriers of the wild type allele[238]. In the Danish prospective study, manuscript II, including 397 cases and 800 members of the sub-cohort, we observed no association of the OGG1 polymorphism with risk of colorectal cancer, neither before nor after stratifying the analysis by gender or age. Additionally, a small Korean study with 125 cases of colon cancer and 247 cancer-free comparison individuals did not find an association of the polymorphism with risk of colon [239]. However, meat intake and smoking increased the risk of colon cancer only among the homozygous OGG1 326Cys carriers, with OR of 4.31 (CI: 1.64-11.48) and 2.75 (CI: 1.07-7.53), respectively. In the Danish study we observed no interaction between the polymorphism and various life style factors, including intake of meat, in relation to risk of colorectal cancer, manuscript II. A large study of a rare polymorphism, OGG1 Arg154His, has been made in Korea by Kim and colleagues. They observed a higher risk of colorectal cancer among carriers of the OGG1 154His allele, with an OR of 3.59 (CI: 0.98-13.11), compared to homozygous carriers of the wild type allele, but the number in this group was limited to 10 cases, why the result may be a chance finding [240].

Several association studies of the GPX1 Pro198Leu polymorphism have been carried out on various other types of cancer. Two studies have found higher risk of breast cancer among the carriers of the variant allele [223;241] with ORs of 1.9 (P<0.05) and 1.43 (CI: 1.07-1.92), respectively. But several other studies suggest no association of the polymorphism with breast cancer risk [242-246]. The association between the polymorphism and risk of lung cancer have been analysed in three studies, with
Table 2: Studies of possible associations between polymorphisms in *OGG1* and *GPX1* and risk of colorectal cancer

<table>
<thead>
<tr>
<th>Reference</th>
<th>Polymorphism</th>
<th>Endpoint</th>
<th>Study design</th>
<th>Cases</th>
<th>Controls</th>
<th>Ethnicity</th>
<th>DNA source</th>
<th>Associations (main results)</th>
<th>Interactions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Goodman <em>et al.</em>, [237]</td>
<td><em>OGG1</em> Ser326Cys</td>
<td>Colon cancer</td>
<td>Case-control</td>
<td>216 men with carcinomas</td>
<td>255 hospitalized men, not cancer, HBV, HIV or HCV</td>
<td>Caucasian, African American (USA)</td>
<td>Primarily blood samples, otherwise colon tissue (some cases)</td>
<td>↓ risk of carcinoma among carriers of the <em>OGG1</em> 326Cys allele compared to homozygous carriers of the wild type allele</td>
<td>No SNP-SNP interaction between the <em>OGG1</em> polymorphism and other BER polymorphisms</td>
</tr>
<tr>
<td>Moreno <em>et al.</em>, [238]</td>
<td><em>OGG1</em> Ser326Cys</td>
<td>Colorectal adenocarcinoma</td>
<td>Case-control</td>
<td>377 with carcinomas</td>
<td>329 hospitalized, not cancer</td>
<td>Caucasian (Spain)</td>
<td>Not mentioned</td>
<td>↑ risk of adenocarcinomas among homozygous carriers of the variant allele compared to homozygous carriers of the wild type allele</td>
<td>↑ risk of adenocarcinomas among carriers of the <em>OGG1</em> 326Cys allele below 54 years of age when diagnosed</td>
</tr>
<tr>
<td>Kim <em>et al.</em>, [239]</td>
<td><em>OGG1</em> Ser326Cys</td>
<td>Colon cancer</td>
<td>Case-control</td>
<td>125 with colon cancer</td>
<td>247 cancer-free</td>
<td>Asian (South Korea)</td>
<td>Not mentioned</td>
<td>No association</td>
<td>↑ risk of colon cancer among smokers carrying the homozygous variant allele</td>
</tr>
<tr>
<td>Kim <em>et al.</em>, [240]</td>
<td><em>OGG1</em> Arg154His</td>
<td>Colorectal cancer</td>
<td>Case-control</td>
<td>490 with sporadic colorectal cancer</td>
<td>524 healthy</td>
<td>Asian (South Korea)</td>
<td>Blood or tumour tissue?</td>
<td>↑ risk of colorectal cancer among carriers of the <em>OGG1</em> 154His allele (only 10 cases, rare polymorphism) compared to homozygous carriers of the wild type allele</td>
<td>-</td>
</tr>
<tr>
<td>Hansen <em>et al.</em>, [236] manuscript I</td>
<td><em>OGG1</em> Ser326Cys <em>GPX1</em> Pro198Leu</td>
<td>Carcinomas and adenomas (high and low-risk)</td>
<td>Case-control</td>
<td>166 with carcinomas, 974 with adenomas (227 high-risk / 756 low-risk)</td>
<td>397 negative flexible sigmoidscopy screening</td>
<td>Caucasian (Norway)</td>
<td>Blood samples</td>
<td>↓ risk of carcinoma among carriers of the <em>OGG1</em> 326Cys allele compared to homozygous carriers of the wild type allele</td>
<td>-</td>
</tr>
<tr>
<td>Hansen <em>et al.</em>, manuscript II</td>
<td><em>OGG1</em> Ser326Cys <em>GPX1</em> Pro198Leu</td>
<td>Colorectal cancer</td>
<td>Pros. case-cohort</td>
<td>397 with colorectal cancer</td>
<td>800 randomly selected from the cohort (30 with colorectal cancer)</td>
<td>Caucasian (Denmark)</td>
<td>Buffy coat</td>
<td>No association of single SNPs</td>
<td>Alcohol consumption associated with ↑ risk among homozygous carriers of the <em>GPX1</em> 198Leu allele</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Intake of vitamin C associated with ↑ risk among homozygous carriers of the <em>GPX1</em> 198Leu allele, and ↓ risk among homozygous carriers of the <em>GPX1</em> Pro198 allele</td>
<td>Tendency for smoking to be associated with ↑ risk among homozygous carriers of the <em>GPX1</em> 198Leu allele</td>
</tr>
</tbody>
</table>
three different results: A higher risk of lung cancer was observed among carriers of the \textit{GPX1} 198Leu allele, with an OR of 1.8 (CI: 1.2-2.8) and an OR of 2.3 (CI: 1.3-3.8) for heterozygous and homozygous carriers, respectively, compared to carriers of the homozygous wild type allele [247], while in a pooled analysis of three Danish and Norwegian studies no association was observed with non-small cell lung cancer at age 59 or below, and no interaction with smoking habits was detected [248]. In our recent study of lung cancer, manuscript V [249], we found a protective effect of the polymorphism among homozygous carriers of the \textit{GPX1} 198Leu allele, with an IRR of 0.60 (CI: 0.35-1.05) compared to homozygous carriers of the wild type allele. Smoking and alcohol consumption had a more adverse effect among the carriers of the homozygous variant allele than among carriers of the two other genotypes. A study from Japan observed higher risk of bladder cancer, with an OR of 2.63 (CI: 1.45–4.75), among heterozygous carriers of the variant allele compared to homozygous carriers of the wild type allele. The Leu/Leu genotype was not found in any subjects in the current study [250]. No association was observed between the \textit{GPX1} Pro198Leu polymorphism and risk of basal cell carcinoma [251], Non-Hodgkin lymphoma among women [252] or prostate cancer [253].

In previous studies, the \textit{OGG1} Ser326Cys polymorphism has been associated with risk of orolaryngeal cancer among smokers carrying the variant allele, with ORs of 1.6 (CI: 1.04-2.6) and OR=4.1 (CI: 1.3-13.0), for heterozygous and homozygous carriers, respectively [254]. A meta-analysis of seven Caucasian studies, including 3253 cases and 3371 controls, found increased lung cancer risk among subjects carrying the \textit{OGG1} Cys/Cys genotype, with an OR of 1.24 (CI: 1.01-1.53) [255]. Studies from Japan and Norway have reported similar findings [256,257]. However, a reduced risk of lung cancer has been observed among non-smokers and light smokers carrying the heterozygous genotype, with an OR of 0.51, P=0.033 [258], while other studies, amongst them our own (manuscript VI), found no significant association between the polymorphism and risk of lung cancer [259-262]. We did, however, observe an interaction between the polymorphism and intake of vegetables, with a reduced risk (IRR=0.46, CI: 0.25-0.84) of lung cancer per 50% increase in vegetable intake among homozygous carriers of the variant allele [262]. The homozygous \textit{OGG1} 326Cys genotype was observed more frequent among bladder cancer patients than in the cancer-free control group (OR=2.41, CI: 1.36-4.25) [263]. No association was observed between the \textit{OGG1} Ser326Cys polymorphism and risk of breast cancer [264-266], cervical cancer [267], Non-Hodgkin lymphoma [268] or basal cell carcinoma [269].

In general, the studies of the \textit{GPX1} Pro198Leu and \textit{OGG1} Ser326Cys polymorphisms and risk of colorectal cancer and studies of interaction between the polymorphisms and life style factors in relation to
risk of colorectal cancer are few and with contradicting results preventing any firm conclusions. Additionally, measurement of GPX enzyme activity in different tissues (cancer tissue or erythrocytes) was conducted at different points in time in relation to the cancer diagnosis, which may explain the inconsistent results. Regarding risk of various types of cancer, the OGG1 Ser326Cys polymorphism seems to be associated with lung cancer risk, with the possibility of interaction with smoking and alcohol consumption.

**Nucleotide Excision Repair**

The nucleotide excision repair (NER) pathway is the primary mechanism for removal of helix-distorting damages from DNA, including bulky adducts and UV-induced photolesions. The mechanism of NER includes five steps: 1. Damage recognition, 2. Assembly of the repair factors at the site of damage, 3. Dual incisions and excision of the damage-containing oligomers, 4. Resynthesis to fill in the gap, and 5. Ligation of the strands. All these steps involve more than 20 proteins, like recognition factors, replication protein, transcription factor, helicases, endonucleases and polymerases. Step 1 and 2 are illustrated in Figure 6.

There are two sub-pathways of NER, termed the global genome NER (GG-NER), which corrects lesions in the entire genome including the non-transcribed strands of active genes, and transcription-coupled NER (TC-NER), that only repair lesions in transcribed strands in active genes. The major differences of the two pathways are the damage recognition step: In GG-NER the proteins Xeroderma Pigmentosum complementation group A and C (XPA/XPC) makes the recognition complex [270-273], while in TC-NER a stalled RNA polymerase II (blocked by a lesion) and Cockayne syndrome proteins have this function to act as a signal to recruit NER proteins [270;274].

In global genomic NER the XPA and XPC enzymes are involved in the damage recognition-complex of NER. Several studies have shown the XPC-hHR23B complex to function at a very early stage of DNA damage recognition [270-273]. The hHR23B (also called Rad23) NER factor co-purifies with XPC [275] and is essential for high XPC activity in NER [276;277]. XPC-hHR23B complex exhibit a very strong affinity for damaged DNA [276;278;279], why it is thought to be the initiator in GG-NER. By interaction with the XPC complex XPA and the transcription factor II H (TFIIH) may be recruited
Figure 6: A proposed molecular mechanism of damage recognition process in the early stage of global genome nucleotide excision repair. Transient steps are indicated with 

To the damaged DNA site [272,273]. TFIIH is a nine sub-unit protein complex required for opening the DNA helix at the vicinity of the lesion [280-282]. Biochemical studies have generated conflicting results with regard to association between the XPC-hHR23B complex, XPA and TFIIH. Some have found recruitment of TFIIH to the site of DNA damage to be dependent on XPC [273,283], while others have found XPA to be interacting with TFIIH [284]. Undoubtedly, both XPC and XPA are vital factors in the very early steps of GG-NER, but exactly when XPA enters the site of damage is not clear. XPA physically interacts with replication factor A (RPA) and is essential to efficient NER [285] by stabilizing the interaction between XPA and the damaged DNA. XPA is capable of binding to the XPF-ERCC1 complex with very high affinity [286]. The XPF-ERCC1 is a specific 5’ endonuclease complex, and thus must be located near the site of 5’ incision [287]. XPG, a 3’ endonuclease, seems to be the next factor recruited to the site, and is probably positioned at the 3’ incision site [271]. Previous studies have observed XPG to copurify with TFIIH, like XPC, and that XPG exclude XPC when binding to TFIIH [288,289], which may suggest that the binding of XPG to the NER complex displaces XPC. Hence, XPA is thought to be crucial to the subsequent positioning of the involved NER enzymes by binding to XPF-
ERCC1 complex and possibly recruit XPG to the site of DNA damage. XPD and XPB are helicases and parts of the large TFIIH complex. They participate in the unwinding of helix in opposite directions of the region of damaged DNA [271;280]. When the DNA around the DNA lesion is unwound, the endonucleases XPG and XPF-ERCC1 complex excises an oligonucleotide of 24-32 bases including the damaged site [290]. The two endonucleases requires an opening of approximately 5-8 bases [291;292]. The final steps of NER are resynthesis of the strand to fill in the gap and ligation of the new strand with the remaining strand. In mammals the synthesis requires the DNA polymerases δ and/or ε [293;294], the replication protein A (RPA) and replication factor C (RFC) [295] and proliferating cell nuclear antigen (PCNA) [296]. The XPF-ERCC1 5´ incision leaves a hydroxyl-group at the 3´ terminus of the gap. This terminus may act as a DNA primer for DNA polymerases [297]. RPA is required for the gap-filling DNA synthesis [295], possibly to protect the template strand against nucleases, and RFC and PCNA as a complex that facilitates the assembly of the polymerases [296]. The new fragment of DNA are synthesized and the final step are ligation of the new patch to the original sequence, which possibly may be performed by DNA ligase I [298].

Combining of neighbouring SNPs into haplotypes may increase the association with disease. Nexø & Vogel have together with colleagues previously identified a haplotype at chromosome 19q13.2-3 encompassing three SNPs in the genes ERCC1 (excision repair cross complementary group 1), ASE-1 (antisense ERCC1) and RAI (RelA-associated inhibitor), which was strongly associated with risk of post-menopausal breast cancer [299] and lung cancer [300;301]. Thus, the haplotype seems to be involved in a mechanism affecting various cancer forms. RAI encodes a specific inhibitor of the RelA subunit (p65) in the nuclear factor-kappa B (NF-κB) transcription factor [302], which participates in inflammatory responses [303] and as an regulator of apoptosis [304]. ASE-1 encodes a nucleolar protein, and is positioned in an anti-sense orientation to and overlaps with the gene for ERCC1. It is possibly involved with the RNA polymerase I transcription complex [305].

The NER polymorphisms studied in the work underlying this Ph.D.-thesis, manuscript III, include the polymorphisms: XPD Lys751Gln, XPD Asp312Asn, XPA G23A and XPC Lys939Gln. The haplotype studied, manuscript IV, encompasses the polymorphisms RAI IVS1 A4364G, ERCC1 Asn118Asn, and ASE-1 G-21A.

The variant alleles of XPA G23A [306], XPD Asp312Asn and XPD Lys751Gln [307;308] polymorphisms and a polymorphism in XPC [308], in full linkage disequilibrium with the XPC Lys939Gln polymorphism
[309], have been associated with a lowered DNA repair capacity compared to the wild type allele. Furthermore, the variant alleles of the polymorphisms XPD Asp312Asn and XPD Lys751Gln have been associated with higher DNA adduct levels [310-312] than among homozygous carriers of the wild type allele.

For colorectal and lung cancer patients the mRNA levels and the expression of the ERCC1 protein are associated with the response to platinum-based chemotherapeutic drugs with direct impact on cancer patient survival. The level of response may be related to the ERCC1 Asn118Asn polymorphism [313-315], and the ERCC1 C8092A polymorphism as well, for lung cancer patients [316]. A recent study by Sæbø and colleagues made some interesting observations: The expression levels of ERCC1 and RAI being comparable in tissue from adenomas and adenocarcinomas, but when the expression levels was compared in normal tissue RAI expression was higher in cases with adenomas than in cases with adenocarcinoma [231]. RAI expression was increased in both adenoma and adenocarcinoma tissue compared to tissue from normal colonic mucosa from the same person [231]. The findings indicate that high RAI expression occurs very early in the malignant transformation. However, contradictory results have been published as to whether increased expression levels of RAI inhibit or promote apoptosis [317;318].

Mutations in the NER gene XPD are associated with the rare, autosomal-recessive inherited disorder Xeroderma Pigmentosum, where patients suffer from severe photosensitivity and actinic changes leading to early onset of skin cancers induced by sunlight [319]. Recently the first case of human inherited ERCC1 deficiency was reported [320]. Cells from the patient showed moderate hypersensitivity to ultraviolet rays, but the clinical features were very severe and compatible with a diagnosis of cerebro-oculo-facio-skeletal syndrome. This discovery represents a novel complementation group of patients with defective NER and suggests novel functions for ERCC1.

Overall, the above mentioned studies of the polymorphisms in the genes involved in NER, XPD Lys751Gln, XPD Asp312Asn, XPA G23A, XPC Lys939Gln, and ERCC1 Asn118Asn, indicate that the polymorphisms may modulate DNA repair capacity and may thereby possibly be associated with development of cancer. To my knowledge no studies have been published on association of RAI polymorphisms with induction of apoptosis, and the exact role of ASE-1 and whether the overlapping between ERCC1 and ASE-1 gene has biological relevance are still unknown.
There are limited numbers of studies of NER genes or the high risk haplotype in relation to risk of colorectal cancer. A search on the PubMed database of NCBI on June 10th 2007 on the MeSH terms “polymorphism, single nucleotide AND colorectal neoplasms” resulted in 148 hits of which seven studies included polymorphisms in \textit{XPD, XPA, XPC, ERCC1, RAI} and \textit{ASE-1}. In combination with a new search on the PubMed database of NCBI by using different combinations of the words: “\textit{XPD XPA XPC ERCC1 RAI ASE-1 polymorphism colorectal colon rectum}” 12 studies of SNPs in the six genes in relation to risk of colorectal cancer or prestages to colorectal cancer were identified. The studies are listed in Table 3, including the results from manuscript III and IV.

The \textit{XPD Lys751Gln} polymorphism is the most frequently studied of the NER polymorphisms in association with risk of cancer. In our Danish prospective study, manuscript III, we observed no association of the \textit{XPD Lys751Gln} and \textit{XPD Asp312Asn} polymorphisms with risk of colorectal cancer [321]. Previously, several studies had similar findings of no effect of the \textit{XPD Lys751Gln} [238;322-326] and the \textit{XPD Asp312Asn} [237;238;322;323] polymorphisms on risk for colorectal cancer. Additionally, Bigler and colleagues found no association of the two polymorphisms with development of adenomas [327]. However, they detected a higher risk of colorectal adenomas among individuals with at least two variant alleles of the \textit{XPD} polymorphisms, with an OR of 1.57 (CI: 1.04-2.38). When stratifying by age the association of the two polymorphisms with risk of adenomatous polyps was restricted to the individuals younger than 60 years when diagnosed (OR=3.77, CI: 1.94-7.35). The risk of adenomatous polyps was higher among smokers carrying the homozygous \textit{XPD} variant alleles (OR=3.93, OR: 1.68-9.21) compared with non-smokers carrying the homozygous wild type. A similar finding could not be detected on risk of hyperplastic polyps. Goodman \textit{et al.}, did not detect any SNP-SNP interaction
<table>
<thead>
<tr>
<th>Reference</th>
<th>Polymorphism</th>
<th>Endpoint</th>
<th>Study design</th>
<th>Cases</th>
<th>Controls</th>
<th>Ethnicity</th>
<th>DNA source</th>
<th>Associations (main results)</th>
<th>Interactions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yeh et al., [328]</td>
<td>XPD Lys751Gln</td>
<td>Colorectal cancer</td>
<td>Case-control</td>
<td>727</td>
<td>736</td>
<td>Asian (Taiwan)</td>
<td>Carcinomas/blood samples</td>
<td>Tendency of XPD 751Gln ↑ risk of CRC among men (69 cases/55 controls)</td>
<td>No GE-interactions.</td>
</tr>
<tr>
<td>Hansen et al., [321] Manuscript III</td>
<td>XPD Lys751Gln XPD Asp312Asn XPD A23G XPC Lys939Gln</td>
<td>Colorectal cancer</td>
<td>Prosp. case-cohort</td>
<td>397</td>
<td>800</td>
<td>Caucasian (Denmark)</td>
<td>Buffy coat</td>
<td>No association of single SNPs</td>
<td>GE-interaction between XPC polymorphism and intake of red meat</td>
</tr>
<tr>
<td>Skjelbred et al., [326]</td>
<td>XPD Lys751Gln</td>
<td>Carcinomas and adenomas (high and low-risk)</td>
<td>Case-control</td>
<td>157</td>
<td>983</td>
<td>Caucasian (Norway)</td>
<td>Blood samples</td>
<td>↑ risk for low-risk adenomas among carriers of the XPD 751Gln allele compared to homozygous carriers of the wild type allele</td>
<td>No GE-interactions with cigarette smoking</td>
</tr>
<tr>
<td>Mort et al., [324]</td>
<td>XPD exon 6 XPD exon 22 XPD exon 23 ERCC1 exon 4</td>
<td>Colorectal cancer</td>
<td>Case-control</td>
<td>45</td>
<td>71</td>
<td>Caucasian? (England)</td>
<td>Carcinomas/blood samples</td>
<td>No association of single SNPs</td>
<td>-</td>
</tr>
<tr>
<td>Goodman et al., [327]</td>
<td>XPD Asp312Asn</td>
<td>Colon cancer</td>
<td>Case-control</td>
<td>216</td>
<td>255</td>
<td>Caucasian, African American (USA)</td>
<td>Primarily blood samples, otherwise colon tissue (some cases)</td>
<td>No association of SNP</td>
<td>No SNP-SNP interaction between the XPD polymorphism and other NER polymorphisms</td>
</tr>
<tr>
<td>Bigler et al., [327]</td>
<td>XPD Lys751Gln XPD Asp312Asn</td>
<td>Adenomatous and hyperplastic polyps</td>
<td>Case-control</td>
<td>694</td>
<td>621</td>
<td>Caucasian and Afroamerican (USA)</td>
<td>Peripheral WBC</td>
<td>No association of single SNPs</td>
<td>↑ risk of adenomatous polyps among heavy smokers carrying homozygous XPD variant compared with nonsmokers who were homozygous wild type.</td>
</tr>
<tr>
<td>Stern et al., [329]</td>
<td>XPD Lys751Gln</td>
<td>Adenomas</td>
<td>Case-control</td>
<td>740</td>
<td>789</td>
<td>Caucasian American, Latinos, Asian Pacific Islander (USA)</td>
<td>Peripheral blood lymphocytes</td>
<td>↑ risk of adenomas among homozygous carriers of the variant allele compared with carriers of the wild type allele</td>
<td>GE-interaction between XPD polymorphism and alcohol consumption</td>
</tr>
<tr>
<td>Starinsky et al., [325]</td>
<td>XPD Lys751Gln</td>
<td>Colorectal cancer</td>
<td>Case-control</td>
<td>456</td>
<td>87</td>
<td>Jewish (64% Ashkenazi among cases) (Israel)</td>
<td>Peripheral blood leukocytes</td>
<td>No association of SNP</td>
<td>↑ risk of colorectal cancer among Ashkenazi jews age below 50 years when diagnosed, carrying the variant allele</td>
</tr>
</tbody>
</table>
Table 3 (continued)

<table>
<thead>
<tr>
<th>Reference</th>
<th>Polymorphism</th>
<th>Endpoint</th>
<th>Study design</th>
<th>Cases</th>
<th>Controls</th>
<th>Ethnicity</th>
<th>DNA source</th>
<th>Associations (main results)</th>
<th>Interactions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Moreno et al., [238]</td>
<td>XPD Lys751Gln, XPD Asp312Asn, ERCC1 19716 G→C, ERCC1 19007 T→C, ERCC1 17677 A→C, ERCC1 15310 G→C, ERCC1 8092 G→A</td>
<td>Colorectal adenocarcinoma</td>
<td>Case-control</td>
<td>377 with carcinomas</td>
<td>329 hospitalized, not cancer</td>
<td>Caucasian (Spain)</td>
<td>Not mentioned</td>
<td>↑ risk of colorectal cancer among carriers of an ERCC1 haplotype (19716C, 19007C, 17677C)</td>
<td>No interactions with age</td>
</tr>
<tr>
<td>Huang et al., [322]</td>
<td>XPD Lys751Gln, XPD Asp312Asn, XPC Arg492His, XPC Ala499Val, XPC Lys939Gln</td>
<td>Adenomas</td>
<td>Case-control</td>
<td>772 with advanced adenomas in the distal colon</td>
<td>777 negative colonoscopy screening, no family history of CRC</td>
<td>Mixed (USA)</td>
<td>Blood samples</td>
<td>No association of single SNPs</td>
<td>↑ risk of advanced adenomas among current or recent smokers carrying XPC haplotype (492Arg, 499Ala, 939Gln)</td>
</tr>
<tr>
<td>Berndt et al., [323]</td>
<td>XPD Lys751Gln, XPD Asp312Asn, XPD IVS19-70, XPC Arg492His, XPC Gln687Arg, XPC Lys939Gln, XPA 3´UTR 327C→G, ERCC1 IVS74 G→C, ASE-1 Gln504Lys</td>
<td>Colorectal cancer</td>
<td>Case-cohort</td>
<td>250 with carcinomas</td>
<td>2224 (no colorectal cancer diagnosis)</td>
<td>Mixed (98% caucasian among sub-cohort and full cohort) (USA)</td>
<td>Blood samples</td>
<td>Borderline significant ↑ risk of proximal colon cancer among homozygous carriers of the variant XPC Lys939Gln allele compared to homozygous carriers of the wild type allele</td>
<td>No interactions with age, gender, smoking, intake of red meat or folate, or BMI</td>
</tr>
<tr>
<td>Skjelbred et al., [330]</td>
<td>ERCC1 Asn118Asn, ASE-1 G-21A, RAI IVS1 A4364G</td>
<td>Carcinomas and adenomas (high and low-risk)</td>
<td>Case-control</td>
<td>156 with carcinomas, 981 with adenomas (227 high-risk/ 754 low-risk)</td>
<td>399 negative flexible sigmoidoscopy screening</td>
<td>Caucasian (Norway)</td>
<td>Blood samples</td>
<td>↑ risk of adenomas among women carrying the ASE-1 G-21A variant allele compared to homozygous carriers of the wild type</td>
<td>No interactions with smoking or alcohol consumption</td>
</tr>
<tr>
<td>Hansen et al., manuscript IV</td>
<td>ERCC1 Asn118Asn, ASE-1 G-21A, R-IF IVS1 A4364G</td>
<td>Colorectal cancer</td>
<td>Prosp. case-cohort</td>
<td>394 with colorectal cancer</td>
<td>791 randomly selected from the cohort (10 with colorectal cancer)</td>
<td>Caucasian (Denmark)</td>
<td>Buffy coat</td>
<td>No association of single SNPs</td>
<td>No association of a haplotype encompassing the polymorphisms in ERCC1, ASE-1 and the R-IF IVS1 A4364G</td>
</tr>
</tbody>
</table>
between the XPD Asp312Asn polymorphism and other NER polymorphisms[237]. Skjelbred and colleagues detected an association between the XPD Lys751Gln polymorphism and development of colorectal adenomas, with an OR of 1.40 (CI: 1.08-1.81), among carriers of the variant allele compared to carriers of the homozygous wild type allele [326]. The statistical significance was limited to the low-risk adenoma group (OR: 1.46, CI: 1.11-1.90). The results were contradicted by a large study by Stern et al., including 740 cases with adenomas and 789 controls, where a lower risk of adenomas was observed (OR=0.7, CI: 0.4-1.0) among homozygous carriers of the XPD 751Gln allele [329]. The result was not stratified for ethnicity (Caucasian, African-American, Latinos, Asian-Pacific Islander). When excluding the 1 case and the 17 controls of Latinos, the OR increases to 0.9 (confidence interval informations are not reported). An interaction between the XPD Lys751Gln polymorphism and alcohol consumption was observed (P=0.04), with higher risk of adenomas among ever-drinkers carrying the XPD 751 Gln/Gln genotype (OR=2.5, CI: 1.2-5.2) compared with never-drinkers carrying the same genotype. A higher risk of colorectal cancer has been observed among Ashkenazi Jews below 50 years of age when diagnosed [325]. The risk was higher among carriers of the XPD 751Gln allele, but it may be a chance finding due to low number of cases (only 15 cases were diagnosed before their 50 years birthday). Furthermore, the Ashkenazi population is known to have particular genetic characteristics, why the result may not be generalized to other populations. A large study from Taiwan observed a non-significant tendency for higher risk of colorectal cancer among men carrying the XPD 751Gln allele (OR=1.5, CI: 0.9-2.3), while no association was observed for women (OR=0.9, CI: 0.6-1.5) [328]. In our study, manuscript III, we detected a gender specific effect of the XPD Lys751Gln polymorphism (P=0.01), with lower risk of colorectal cancer among women carrying the variant allele of XPD Lys751Gln with an IRR of 0.62 (CI: 0.40-0.96) and 0.45 (CI: 0.23-0.87), respectively, compared to women carrying the wild type allele. No association was found among men. The gender differences could hypothetically be caused by a hormonal interaction. However, we observed no interaction between the use of hormone replacement therapy among women and the polymorphism. Thus, we did not find the hypothesis plausible. The result in our study may be a chance finding.

In our study, manuscript III, the XPC Lys939Gln polymorphism was not associated with risk of colorectal cancer [321]. However, we did observe an interaction between the polymorphism and intake of red meat, with an IRR of 3.70 (CI: 1.70-8.04) for colorectal cancer per 100g red meat intake per day among homozygous carriers of the XPC Lys939Gln variant allele. In the light of the sample size and the multiple comparisons being made, this result may be a chance finding. The association was not statistically significant after a Bonferroni correction. In a large American study by Huang three polymorphisms in
XPC was studied, including the XPC Lys939Gln polymorphism. No association was found between the XPC Lys939Gln polymorphism and risk of adenomas [322]. However, higher risk for development of adenomas was observed among current or recent smokers carrying the XPC 939Gln allele (OR=2.0, CI: 1.3-3.0) or a XPC haplotype encompassing three linked SNPs in XPC (Arg492His, Ala499Val, Lys939Gln) compared with never-smokers carrying the homozygous wild type allele. In a small study by Berndt et al., a tendency for higher risk of proximal colon cancer was observed among homozygous carriers of the variant XPC Lys939Gln allele, with an OR of 1.74 (CI: 0.98-3.08) [323]. The result may possibly be a chance finding due to sample size and multiple testing. Three other SNPs in the XPC gene were not associated with colorectal cancer risk.

To my knowledge, only two studies have been published on the association of polymorphisms in the XPA gene with risk of colorectal cancer: Our study in manuscript III and the study by Berndt. No association was observed of the XPA G23A polymorphism [321] or a polymorphism in the 3’ untranslated region of XPA [323] with risk of colorectal cancer.

The results in the Danish (manuscript IV) and the Norwegian study [330] of the ERCC1 Asn118Asn, ASE-1 G-21A and RAI IVS1 A4364G polymorphisms suggest neither the single polymorphisms nor the haplotype to be associated with risk of colorectal cancer. But a tendency of higher risk of colorectal cancer, with an OR of 2.19 (CI: 0.95-5.04), was observed among Norwegian female homozygous carriers of the haplotype, compared with non-homozygous carriers of the haplotype [330]. In the Norwegian study the ASE-1 polymorphisms was associated with higher risk of adenomas among women (OR=1.66, CI: 1.15-2.39). Moreno et al. examined 5 polymorphisms in the ERCC1 gene. A haplotype containing the minor allele of three of the ERCC1 polymorphisms (see Table 3) was associated with a higher risk of colorectal cancer (OR=2.3, CI: 1.0-5.3) compared with carriers of the most frequent haplotype [238]. Two other SNPs in the ERCC1 gene were not associated with risk of colorectal cancer [323;324].

Numerous association studies of polymorphisms in genes involved in NER are reported on various types of cancer, with the majority of studies focused on the XPD Lys751Gln and XPD Asp312Asn polymorphisms. A meta-analysis of lung cancer by Kiyohara et al. (with 1913 cases and 1882 controls of different ethnicities) [331] suggested among other studies [332-334], that carriers of the variant alleles of either of the two XPD polymorphisms were found to be at higher risk of lung cancer, while a number of other studies did not observe any association of the two polymorphisms with lung cancer risk [258;335-337]. Two large meta-analyses (with 3725 cases and 4152 controls) included identical nine case-control
studies but made two dissimilar conclusions: The XPD Lys751Gln and XPD Asp312Asn polymorphisms are associated with risk of lung cancer [333] or no clear association was found [338]. Some studies suggest an interaction between the two XPD polymorphisms and smoking in relation to risk of lung cancer [258;332;337]. Combinations of the XPD, XPC and XPA genotypes, variant alleles, is suggested to be associated with higher risk of lung cancer [335]. This may be plausible but in the light of multiple testing and the low number of cases this may be a chance finding. The largest breast cancer studies by the number of individuals, 1053 cases/1102 controls [339] and 1830 cases/1262 controls [340] observed modest associations of the XPD polymorphisms with breast cancer risk. Carriers of the variant XPD Lys751Gln allele was associated with a 20% higher risk (OR=1.21, CI: 1.01-1.44) compared with homozygous carriers of the wild type allele. The risk seemed limited to those with a PAH-DNA adduct level above the median, with an OR of 1.61 (CI: 0.99-2.63) among homozygous carriers of the XPD 751Gln allele [339]. Several other studies observed no association of the XPD Lys751Gln polymorphism [340-345] or the XPD Asp312Asn polymorphism [344;346] to risk of breast cancer. However, higher risk has been detected among ever smoking women carrying the XPD 751Gln allele (OR=2.52, CI: 1.27-5.03) compared to ever smoking women carrying the homozygous wild type allele [347]. Association with breast cancer risk has been detected when the homozygous variant XPD Lys751Gln allele and the homozygous variant XPD Asp312Asn allele segregated together, with OR=1.5 (p<0.05) and OR=3.69 (CI: 1.76-7.74), respectively [340;348]. A large study including 2485 cases with single primary melanoma and 1238 cases with second or higher order primary melanomas detected higher melanoma risk among homozygous carriers of the variant XPD Lys751Gln allele (OR=1.4, CI: 1.1-1.7) or the variant XPD Asp312Asn allele (OR=1.5, CI: 1.2-1.9), respectively [349]. Similar results were obtained in a study by Li et al. [350], while another study observed the inverse association for both polymorphisms [351]. When stratifying by age Baccarelli et al. observed an association of the two XPD polymorphisms to risk of melanoma only among the individuals older than 50 years when diagnosed [352]. The XPD Lys751Gln [353] and the XPD Asp312Asn polymorphism [354] have been associated with risk of bladder cancer. An interaction are suggested between the XPD Lys751Gln polymorphism and smoking in relation to bladder cancer risk [353;355;356]. Individuals carrying both the variant XPD alleles were more susceptible to development of bladder cancer [353;354] than carriers of wild type alleles. The XPD Lys751Gln and XPD Asp312Asn polymorphisms have not been associated to risk of basal cell carcinoma [351;357-359], endometrial cancer [360], prostate cancer [361] or gastric cancer [362].

A small study suggest that the variant allele of the polymorphism XPC Lys939Gln are associated with higher risk of bladder cancer (OR=1.49, CI:1.16-1.92) [363]. No association are observed between the
polymorphism and risk of lung cancer [335;337;364] but a haplotype encompassing more polymorphisms in XPC may contribute to a higher risk of lung cancer [335;337;364]: Individuals with both putative genotypes of XPC Lys939Gln and XPC Ala499Val polymorphisms are observed with a 2.4-fold (OR=2.37, CI: 1.33-4.21) higher risk of lung cancer compared with individuals with both wild type genotypes [335;337;364], with the highest risk observed among smokers. Polymorphisms in XPC have not been associated to risk of basal cell carcinoma [358;365], cutaneous melanoma [366;367] or breast cancer [344-346]. A lower risk of endometrial cancer may be associated with carriage of at least one variant allele for both XPC Lys939Gln and XPC Ala499Val polymorphisms [368].

In a Korean population carriers of the wild type allele (G/G or A/G) in the XPA G23A polymorphism were reported to have a lower risk of lung cancer compared to carriers of the A/A genotype, with an OR of 0.56 (CI:0.35-0.90) [369]. Similar results were obtained in studies on lung cancer risk in Caucasians and Mexican-Americans [306;335] [336], while a Norwegian study observed the inverse effect with a 1.6-fold higher risk (OR=1.59, CI:1.12-2.27) of lung cancer among carriers of the G/G genotype compared with carriers of the A-allele [257]. When stratifying by smoking status the protective effect for lung cancer was only observed among ever smokers [306] or current smokers [369] carrying at least one G-allele or the G/G genotype, respectively. A tendency for lower risk of basal cell carcinoma has been observed among carriers of the variant G-allele, with an OR of 0.82 (CI: 0.66-1.01) and an OR of 0.74 (CI: 0.53-1.03) for homozygous and heterozygous carriers, respectively [370]. The same tendency was observed for risk of squamous cell carcinoma [370]. Carriage of at least one A-allele for XPA G23A was associated with decreased risk of endometrial cancer, OR=0.47 (CI:0.25-0.82) compared with carriers of the G/G genotype, but only among women with a history of using oral contraceptives [360].

The polymorphism RAI IVS1 A4364G is strongly associated with risk of basal cell carcinoma [371;372], with a considerably lower risk among young (<50 years when diagnosed) carriers of the variant G-allele, with an OR of 0.08 (CI: 0.02-0.36), compared with the homozygous carriers of the A/A genotype [371]. The ERCC1 Asn118Asn, ASE-1 G-21A and RAI IVS1 A4364G polymorphisms are not associated with testicular cancer [373]. Furthermore, no association has been observed for the ERCC1 Asn118Asn polymorphism to risk of endometrial cancer [360;374], ovarian cancer [374] and adult glioma [375].

In studies of Caucasians the haplotype encompassing the three SNPs: ERCC1 Asn118Asn, ASE-1 G-21A and RAI IVS1 A4364G is shown strongly associated to risk of post-menopausal breast cancer [299] and lung cancer [300] among women less than 55 years of age when diagnosed with their first cancer. Female
homozygous carriers of the haplotype were at a 7.02-fold (CI: 1.88-26.18) higher risk of lung cancer compared to female non-carriers of the homozygous haplotype. The haplotype was associated to risk of breast cancer with an IRR of 9.50 (CI: 2.21-40.79) among female homozygous carriers of the haplotype, compared to women not carrying the homozygous haplotype. Even though the association was very strong among young carriers of the high risk haplotype, an association was not detected in the older age groups (55-60 years and > 60 years when diagnosed with breast cancer) [299]. Besides the study in manuscript IV and in the Norwegian KAM-study [330] to my knowledge only one other study has been reported on gene-environment interactions with the predefined haplotype, our own in manuscript VII [301]. We observed a borderline significant interaction (P=0.06) between high smoking intensity (>20g tobacco/day) and the haplotype, with a 2.03-fold increased risk of lung cancer per additional 5g tobacco/day among women carrying the haplotype with a high smoking intensity, while no association were observed among female non-carriers of the haplotype with a similar smoking intensity, manuscript VII [301]. An interaction between alcohol consumption and the haplotype was found (P=0.003), male carriers of the haplotype having a higher risk of lung cancer, with an IRR of 1.40 per 10g alcohol intake/day, whereas no association was found among men who was not carrying the homozygous haplotype.

In general, the studies suggest that the XPD Lys751Gln and XPD Asp312Asn polymorphisms may be associated with risk of colorectal adenomas with the possibility of interaction with smoking and alcohol consumption. The reported studies of polymorphisms in XPC, XPA and of the pre-defined haplotype in relation to risk of colorectal cancer are few, but the results are relatively consistent: In general, no association of the polymorphisms in the genes involed in NER (XPD, XPC, and XPA) or the pre-defined haplotype (ERCC1, RAI, and ASE-1) was observed with risk of colorectal cancer. The two XPD polymorphisms at amino acid position 312 or 751, the XPD Lys751Gln in particular, may be associated with risk of cancer in the lung, breast and bladder and seems to modify the effect of smoking on risk of the three cancer forms. The XPC Lys939Gln polymorphism may possibly be associated with risk of bladder cancer, and the XPA G23A polymorphism may be associated with risk of skin cancer (basal cell carcinoma), endometrial cancer and lung cancer. However, the studies are few and the results are inconsistent. The RAI IVS1 A4364G polymorphism seems to be strongly associated with risk of basal cell carcinoma, and the haplotype encompassing the polymorphisms ERCC1 Asn118Asn, ASE-1 G-21A and RAI IVS1 A4364G may possibly be associated with risk of breast cancer and lung cancer, in particular among young individuals. The haplotype may possibly modify the effect of smoking and alcohol consumption on risk of lung cancer.
Double-strand Break Repair

Translocations are genetic aberrations that occur when a broken fragment of a chromosome is erroneously rejoined to another chromosome. The initial event in the creation of a translocation is the formation of a double-strand break (DSB), which may result in the disruption of genes or an alteration of the normal gene expression due to juxtaposition of certain genetic elements with oncogenic characteristics that disturbs this expression. DSBs may be caused by both exogenous agents, like ionizing radiation and certain chemicals, as well as endogenous agents, like the oxygen free radicals from cellular metabolism [376-378] or may arise spontaneously in the S-phase, when a single-strand break (SSB) in a parental strand is passed by a replication fork [379]. Several pathways exist to repair this rather severe DNA lesion. In the following I will focus on the two major DSB repair pathways in mammals: Homologous recombination and non-homologous end-joining. Both can be divided in a number of subpathways that are presented in a review by Agarwal [380] and will not be discussed here. Homologous recombination repair of DSBs are favoured during the S- and G2-phases using the undamaged sister chromatid as a template for repair of the broken sister chromatid by sister chromatid exchange. The repair is carried out by the Rad52 epistasis group of proteins. The group includes the MRN complex (Mre11/Rad50/Nbs1) [381], that initiates the damage response together with Ataxia telangiectasia mutated protein in humans, and RAD51 that polymerizes the 3´ tails to create a nucleoprotein filament able to create an intermediate, the D-loop, joining the intact and the broken sister chromatids for subsequent sister chromatid exchange. Resolution of the joint homologous recombination factors requires structure-specific endonucleases, which is suggested to be the XRCC3 and Rad51C proteins [382]. Homologous recombination repair may among others be regulated by the breast cancer susceptibility protein BRCA2 [377;378], presumably by interrupting the binding of Rad51 to DNA. If the DSB occurs before DNA replication, in the G0- and G1-phases [383] or during V(D)J recombination in mature B cells (at the clonal expansion) [384], non-homologous end-joining repair is chosen, where no template sister chromatid are required. The pathway involves simply a religation of the two ends of the broken strands. Consequently, this pathway has a much higher error-rate than the homologous recombination repair pathway. The Ku70/Ku80 heterodimer binds the DNA ends, which facilitates the recruitment and phosphorylation of DNA dependent protein kinase catalytic subunit at the site of the DSB. The ligase IV-XRCC4 complex ligates the juxtaposed DNA ends [385;386]. Murine studies on the repair functions of the two pathways suggest, that they have potential to overlap in various tissues (reviewed in [387]).

An under expression of the Ku subunits have been observed in human colon cancer cells [380]. The expression pattern of both Ataxia telangiectasia mutated protein (ATM) and BRCA1 are observed to
predict survival in colorectal cancer patients [388]. Low expression of ATM and BRCA1 was associated with loss of MLH1 or MSH2 expression, which indicate a possible link between expression of DNA mismatch repair and DNA DSB repair proteins in sporadic colorectal cancer.

Mismatch Repair

DNA mismatch repair (MMR) corrects mismatches generated during DNA replication, such as base-base mismatches and shorter insertion/deletion loops, and mismatches that have escaped proofreading by DNA polymerase, particularly in microsatellites. Polymerase δ has 3’ to 5’ proofreading activity and corrects approximately 99% of the replication errors [389]. The mechanism of MMR includes four steps: 1. Damage recognition, 2. Assembly of the repair factors at the site of damage, 3. Excision of the incorrect sequence, 4. Re-synthesis. In humans the proteins responsible for the recognition of a mismatch are called human MutS Homologs (hMSH). MutSα is a heterodimer composed of hMSH2 and hMSH6. This predominant form of MutS recognizes base/base mismatches and insertion/deletion mispairs in which one strand contains one or two unpaired nucleotides [390-393]. MutSβ, consisting of hMSH2 and hMSH3, recognizes larger insertion/deletion loops with mispairs in up to 10 nucleotides [392;393]. The MutS proteins bind the double stranded DNA at the site of a mismatch, disengages and move laterally along the DNA [394;395] to identify the damaged strand. The MutL complex, consisting of hMLH1 and hPMS2, is recruited to the site, where it interacts with MutS and displaces the DNA polymerase and PCNA from the nascent daughter strand and recruits exonuclease 1 [389]. The mismatch is excised and DNA re-synthesis is carried out by Pol δ, primarily, and possibly Pol α and ε [396].

Germline mutations in mismatch repair genes, predominantly MSH2 and MLH1, have been found to underlie the Lynch syndrome, which accounts for 2-5% of all colon cancer cases [397].
DISCUSSION

Defence mechanism against development of cancer involves a series of genes encoding proteins involved in metabolism and reduction of potentially toxic/carcinogenic compounds and to repair subtle DNA lesions. A large number of molecular epidemiologic studies have been conducted with the purpose of identifying these genes and to assess their role in the etiology of cancer. This thesis systematically reviews the literature on associations between polymorphisms in genes involved in defence of oxidative DNA damages, in nucleotide excision repair and in a previous identified haplotype and risk of colorectal adenomas and colorectal cancer. Although the present review includes 17 studies on 25 different SNPs the results were generally inconsistent, and no strong associations were observed between the single polymorphisms and risk of colorectal adenomas or colorectal cancer. In addition, comparisons between studies to highlight any trends were not feasible, as only two of the polymorphisms, XPD Lys751Gln and OGG1 Ser326Cys, were analysed in more than two studies with similar study endpoints and ethnic genetic backgrounds in the study populations.

Several issues are important to consider when evaluating polymorphisms as biomarkers for susceptibility in molecular epidemiological studies like: Are there reason to consider the polymorphism in relation to the endpoint from a biological point of view, does the polymorphism modify the function of the encoding protein, are the allele frequency high enough to measure in the study population, and are there correspondence between the genotype and the phenotype.

As previously described in the chapters of BER and NER the majority of polymorphisms studied in manuscripts I-VII in the present Ph.D.-thesis have previously been observed to modify the repair capacity or enzyme activity of the proteins encoded. However, further studies are needed to make firm conclusions on the phenotypic effects of the polymorphisms on the functions of the proteins. The genes are involved in important cellular defence mechanisms against a large variety of structural unrelated DNA lesions. If these DNA lesions are left unrepairs, they may contribute to mutagenesis and oncogenesis. Additionally, previous studies have detected association between the polymorphisms or the haplotype and risk of various types of cancer. Thus, the polymorphisms chosen are biologically relevant to study in relation to risk of colorectal cancer.
The allelic frequencies of all the variant alleles of the polymorphisms reported were generally above 0.25 among members of the sub-cohort, except for a few polymorphisms with variant allele frequencies below 0.1 [238;240;326;398]. Low allele frequencies require large study populations for detecting reliable effects of low penetrance genes. The study by Kim et al. [240], Table 2, that included 490 cases of sporadic colorectal cancer and 524 controls, are an example of this particular limitation inherent from the study design. They examined the polymorphism OGG1 Arg154His, which was very rare in the study population: only 10 of the cases carried the OGG1 154His allele. Hence, the observed borderline significant association with risk of colorectal cancer (OR=3.59, CI: 0.98-13.11) among the carriers of the variant allele may very likely be a chance finding due to low statistical power. Although they published a study with a reasonable size of the study population, their choice of investigating a polymorphism with very low allele frequency makes the finding of an association between genotype and colorectal cancer risk statistically very questionable. Several of the studies on colorectal cancer listed in Table 2 and 3 include small size study populations. Thus, a possible explanation for the inconsistent results may be limitations inherent from the design of the studies concerning the number of individuals in the study population in relation to the allele frequency of the studied genotypes.

Most of the studies analyze only single polymorphisms in genes with modest effect in relation to risk of cancer. The modest associations, or borderline significant associations, observed between the polymorphisms and risk of colorectal adenomas or colorectal cancer may be secondary to linkage disequilibrium with a yet unidentified, but tightly linked, colorectal cancer risk locus. Alternatively, one SNP may not significantly affect the susceptibility to cancer, but by analyzing combinations of less favourable polymorphisms from multiple genes within the same pathway or of haplotypes the associations may be amplified and provide more realistic information on the association between the polymorphisms and risk of cancer. However, analysis of SNPs in combination or as haplotype reduces the number of observations and thus decreases the statistical power of the studies. Thus, larger studies are needed for evident haplotype or multiple gene-gene studies.

Several studies have not reported whether the association of the genotypes (or the life style factors) on risk of adenomas or colorectal cancer were different between genders. Three studies did observe differences: The association of the XPD Lys751Gln polymorphism with risk of colorectal cancer was different between genders in our study (manuscript III) [321] with a protective effect among women carrying one or two copies of the variant allele compared to women carrying the wild type allele, while no association of the polymorphism was observed among men. Similarly, in the study by Yeh et al. [328] the
association of the XPD polymorphism was gender-specific with a tendency for higher risk of colorectal cancer only among men carrying the XPD 751Gln allele. Skjelbred et al. [330] observed association between the ASE-1 G21A polymorphism and risk of colorectal adenomas only among women. And the haplotype was associated with risk of lung cancer only among women, manuscript VII [301]. Additionally, a number of studies detected different susceptibility to cancer according to age; significant associations of the XPD Lys751Gln [325], OGG1 Ser326Cys [238] polymorphisms and the haplotype [299] and risk of colorectal cancer or lung and breast cancer, respectively, seemed to be evident more frequently in younger individuals. However, in the light of the size of the study populations the results may be chance findings due to low number of individuals after stratification by gender or age or to multiple testing.

Unfortunately, the majority of molecular epidemiologic studies on genetic (and environmental) risk factors do not report separate data for topological subgroups (distal/proximal colon or colon/rectal) of the cancer, presumably due to no available information on tumour location or low statistical power when stratifications are made. Hence, very little information is obtained whether environmental or genetic risk or beneficial factors have different effect in the topological subgroups. If the topology distribution influences the outcome, this factor should be considered when comparisons are made between studies.

The colorectal cancer studies mentioned in Table 2 and 3 are focused on different ethnic groups, which preclude comparison of the results between studies. Six of the nineteen studies are even based on heterogenous populations [237;322;323;325;327;329]. The studies by Berndt, Goodman, Huang and Stern included ethnicity in their analysis of the mixed study populations [237;322;323;329]. If the inclusion of different ethnic groups is disproportional in the groups of cases and controls and/or if the ethnicity is not considered in the statistical analyses, ethnicity may have influence on the outcomes due to intrinsic differences in genetic background and life style. Furthermore, comparisons between studies with mixed populations must be done cautiously if ethnicity and life style factors are not included in the analyses of risk of cancer.

The use of genotype rather than phenotype may lead to misclassification, as the actual activity of the encoded enzyme may be influenced by a number of exogenous and endogenous factors and not just the genetic component, as observed for GPX activity in manuscript II and in the study by Ravn-Haren et al. [223]. Thus, when validating biomarkers of susceptibility the assessment of interaction between gene and life style factors in relation to the disease is important. Additionally, it is likely that some of the genes/polymorphisms may contribute to cancer only in concomitance of certain life style factors.
(including diet). In the present review, several colorectal cancer studies included stratification for main modifying factors (smoking, alcohol consumption, age and gender) while considerably less studies included analyses for dietary habits. Thereby, the capability to adjust the risk estimates for potential confounders and to detect effect modification by life style factors was not existing.

Data on life style factors are most often obtained by questionnaires, by which exposure assessment is based on answers concerning exposure in the recent year(s). For risk estimation on disease assumptions are generally made of a conservative life style and dietary habits decades before (before the presumed cancer onset) the interview data are obtained, usually at inclusion into the study. Thus, there is a risk of misclassification of the exposure variables at the critical time for carcinogenesis. While for genotyping the methods are quite accurate, a determination of more exact environmental exposure is both laborious and expensive. Measurements of biomarkers of internal dose or biological effective dose (preferently years before the cancer diagnosis and cancer onset) may possibly contribute to more accurate information on environmental association or gene-environment interactions related with risk of cancer.

The case-control studies in Table 2 and 3 recruited colonoscopically negative individuals, hospitalized cancer-free individuals (not screened) and healthy individuals (not screened) as members of the control group. The use of control groups not screened does not prevent inclusion of individuals with undetected cancer or prestages of cancer, while studies including only colonoscopically negative individuals may not be representative of the general population. Recruitment of hospitalized cancer-free patients or healthy individuals as control group may involve directed questionnaires, why the reliability of answers should be considered. Matching of case and controls (e.g. for age, gender and ethnicity) are not always reported.

In the prospective “Diet, Cancer and Health” (DCH) case-cohort studies of colorectal cancer, described in detail in appendix I and manuscript II-IV (and manuscript V-VII for lung cancer studies), a very detailed questionnaire on life style factors are included, which enabled us to adjust the risk estimates for potential confounders and to investigate gene-environment interactions in relation to risk of colorectal cancer. The participants were answering the questionnaires on life style factors before cancer diagnosis, which minimize the risk of recall bias. The average follow-up time was relatively short, with a median of less than 7 years. The latency period between exposure to a tumorigenic factor and development of sporadic colorectal cancer may presumably be 2-3 times longer than that. We assumed, however, that the life style and dietary habits of the individuals are quite conservative, and that the interview data also reflect the exposures 20-30 years before inclusion into the cohort. Still, there is a risk of misclassification of the
exposure variables at the critical time for carcinogenesis. I expected that only a low proportion of participants in the DCH study had experienced high occupationally exposure to DNA damaging carcinogens and therefore made no efforts to obtain such informations. Originally, as this Ph.D.-study is a part of the Centre of Excellence “Air Pollution in a Life-time Health Perspective”, it was the intention to include modelled data on individually exposure to air pollution in the analyses of colorectal cancer risk. Unfortunately, the development of the models by The Danish National Environmental Research Institute is delayed. NERI has not yet (July 1007) been able to deliver the calculations for the present study. The various samples of biomaterial (blood, urine, fat tissue and toenail clippings) gathered from the participants, makes the cohort very valuable in the service of research on association between biomarkers and risk of cancer on several levels of the biological pathways. The DCH study are population based including an ethnic homogenous group of Danish origin. The sub-cohort was selected as a stratified random sample from the same cohort, in which the colorectal cancer cases were identified. Thus, selection bias is unlikely. However, we do not know whether the individuals included in the cohort had an undiagnosed cancer at baseline. We do not know whether this aspect have influence on our results. Sociological differences (e.g. education and marital status) are observed between participants and non-participants, why a generalization to the Danish population of the results obtained must be made with caution. The allelic frequencies of the variant alleles were above 0.30 for the XPA, XPC, XPD, ERCC1, GPX1 (RHOA), and OGG1 polymorphisms among members of the sub-cohort, while the allelic frequency of the variant alleles was 0.17 for ASE-1 G-21A and 0.19 for RAI IVS1 A4364G. Thus, the inclusion of 405 colorectal cases and a comparison group of 810 individuals provided a relatively high statistical power. However, an even larger sample size may be required to achieve convincing statistically significant results for stratified analyses for gender, age and topology on SNP or haplotype association studies and possible gene-environment interactions in relation to risk of colorectal cancer. The association of the “Diet, Cancer and Health” cohort study to the “European Prospective Investigation into Cancer and Nutrition” study, including approximately 520,000 men and women, is a good opportunity to investigate whether interesting findings in the Danish cohort are reproducible in a large European co-study.

In the Norwegian “Kolorektal cancer, Arv og Miljø” (KAM) case-control study, described in detail in appendix I and manuscript I, biomaterial and questionnaire data are available from patients with either pre-stages of cancer or the endpoint adenocarcinomas. This is rarely seen, and makes the study very valuable in the research of the processes during colorectal carcinogenesis. The study is based on an ethnic homogenous group of Norwegian origin. The control individuals are polyp free, diagnosed by
sigmoidoscopy, and therefore not representative of the general Norwegian population. Cases and control individuals are recruited from the same cohort. The number of cases with diagnosed adenocarcinomas is low (170 cases) and provides low statistical power in the analysis of adenocarcinoma risk, while the number of cases with diagnosed adenomas, 1044 cases, are one of the largest colorectal adenoma studies published to date, July 2007. The KAM study is still collecting biomaterial and questionnaires from incident adenocarcinoma cases. For adenocarcinomas, we had an 80\% chance of detecting an OR of two among carriers of the variant allele at a significance level of 5\% assuming an allele frequency of 0.30. The allelic frequencies of the variant alleles were 0.30 for the GPX1 polymorphism and 0.27 for the OGG1 polymorphism among the control group. The enrolled individuals were acquainted with the endpoint of the study when answering the questionnaires due to a previous operation or screening for adenomas and colorectal cancer, which increases the risk of recall bias on the questionnaire data. At the time of the analyses for manuscript I the data from the questionnaires were not yet obtained. Hence, all risk estimates are crude and we were not able to investigate gene-environment interactions in relation with risk of adenomas or adenocarcinomas. However, in the Danish study of association between GPX1 and OGG1 and colorectal cancer risk adjustment for smoking, alcohol consumption and dietary factors did not change the estimate, why the crude estimates in the Norwegian study and the Danish study may be compared, under the assumption that the two Scandinavian study groups are having comparable life style habits. However, the Danes tend to have a higher percentage of smokers and adolescents consuming alcohol compared to other European populations [399], and the Norwegian study includes rural and urban living individuals while the Danish study only includes urban living individuals, why the life style and thereby the exposures may not be comparable in the two study populations. Thus, comparison between the two studies should be made cautiously.

A possible interpretation of the results may be, that the polymorphisms in the genes XPA, XPC, XPD, ERCC1, GPX1 (and RHOA), OGG1, ASE-1 and RAI are not of major importance in colorectal cancer carcinogenesis, which point towards that lowered defence of oxidative DNA damages and lowered repair capacity of the NER pathway may not be a risk factor for development of colorectal cancer. A high cell turnover of epithelial tissue in colon and rectum and/or a high apoptotic rate could be the explanation for not accumulating DNA damage leading to tumourigenese. Alternatively, a lowered defence or repair capacity in colonocytes could possibly result in excessive DNA lesions leading to activation of the apoptosis pathway in the cell and thereby avoiding malignant transformation of the cell. As previously described the dysplastic process in colorectal adenoma formation is proposed to be a “top-down” process that originates from the top of the villi, where exposure to luminal contents would not occur, rather than
originating from stem cells resided at the base of the crypts, “bottom-up”. Thus, a possible irritant capacity/effect of the luminal content of dietary compounds might together with their possible mutagenicity drive colon carcinogenesis. Such an effect may not be affected by the NER pathway, or the defence mechanisms against oxidative DNA damages.

**CONCLUSION AND PERSPECTIVES**

In summary, this review, limited by the bias against publication of null findings, highlights the complexities inherent in epidemiological research and, particularly, in molecular epidemiological research on colorectal cancer. Studies on possible associations between SNPs in genes involved in defence of oxidative DNA damages and in nucleotide excision repair and risk of colorectal cancer have not obtained consistent results, why the issue of whether the SNPs are possible biomarkers of susceptibility for colorectal cancer is not satisfactorially clarified at present.

Sample size coupled with allele frequency may have influenced the validity of the results. Differences in the study design, like distribution of gender, age, topology, ethnicity and criterias for recruitment of comparison individuals may have contributed to the dissimilar findings. The application of large, well-designed association studies of the polymorphisms will make it statistical reasonable to make stratified analyses for obtaining information on risk factors in sub-groups and will generally decrease the risk of chance findings. Furthermore, studies including both cases with prestages of colorectal cancer and cancer cases will contribute with valuable information of the processes during colorectal carcinogenesis.

Most of the studies analyze individual polymorphisms in genes with modest effect in relation to risk of cancer. Cancer is a complex multigenic and multistage disease involving the interplay of many genetic and environmental factors. Hence, it is unlikely that a single genetic polymorphism in low-penetrance genes would have a dramatic effect on cancer risk. More information may be obtained from haplotyping multiple polymorphisms within genes or from combining multiple polymorphisms within pathways. The continued advances in SNP maps and in high-throughput genotyping methods will facilitate these analyses. Defining haplotypes and whole genome association studies may yield information on un-explored regions of the genome that have impact on colorectal cancer risk and development.
Colorectal cancer are probably caused by a complex interaction between many genetic and environmental factors over time. More and large studies with informations on life style factors are required to assess these very possible gene-environment interactions.

Most environmental carcinogens require metabolic activation before they are able to form DNA damages. These activated forms may be detoxified or induce DNA repair or apoptosis. Thus, genetically determined susceptibility to colorectal cancer may depend on the balance among enzymes involved in metabolism and detoxification of carcinogens and on the balance between induction of DNA repair or apoptosis. Further investigations of the combined effects of polymorphisms between genes involved in these four mechanisms may help to clarify the influence of genetic variation in the carcinogenic process and may shed light on the complexities of the many pathways involved in colorectal cancer development, providing hypotheses for future functional studies.
REFERENCE LIST


APPENDIX I – THE COHORTS

A description of the collection of data and biomaterial to the Danish “Diet, Cancer and Health” cohort study and the “Norwegian Colorectal Cancer Prevention” case-control study is given here, as a supplement to the information presented in the manuscripts I-VII.

The “Diet, Cancer and Health study”

Between December 1993 and May 1997 160,725 individuals, 79,729 women and 80,996 men, were invited to participate in the “Diet, Cancer and Health” study (DCH), the inclusion criteria being age 50-64 years, born in Denmark, and with no prior diagnosis of cancer registered in the Danish Cancer Registry. The individuals invited were living in Copenhagen municipality, Frederiksberg municipality, Aarhus municipality, Hinnerup or Horning municipalities in Aarhus County and nearly all in Copenhagen County and were retrieved from the Civil Registration System (CPR). Invitations to the study were sent out containing information on the study. If the receiver had not responded within three weeks, a reminder was sent. If no respond in the following three weeks a second reminder was sent. Afterwards no further effort was made to include the individual. By this procedure only participants who were instantly motivated for participation was included in the study, which possibly secured a higher participation rate at the planned follow-up five years after baseline.

The clinical examinations of the participants were made at the study clinics in Aarhus and Copenhagen at the time of inclusion into the cohort. The time interval from invitation to the inclusion was between one week and two months. At the clinics the participants filled in a questionnaire about life style factors that are known or suspected to be risk factors for cancer development. They handed in toenail clippings, and a detailed 192-item food frequency questionnaire that was previously mailed to them to answer at home before visiting the clinic. In the food frequency questionnaire the participants were asked to report their average intake of various food and beverage items over the past 12 months within 12 possible categories ranging from never to eight times or more per day. Daily intake of specific foods and nutrients were calculated for each participant using the software program FoodCalc [1]. Development and validation of the food frequency questionnaire has previously been described [2-5]. The questionnaires were scanned immediately and checked for unclear or missing information, so that it could be clarified with the participant as soon as possible, preferably before he/she left the clinic, by an interview performed by a lab
technician. Furthermore, at the time of inclusion, clinical measurements were obtained and samples of blood, urine and fat tissue from the buttock were collected and stored at -180° C at the Danish Cancer Society for future research. Toenail clippings are stored at room temperature (20° C) at the Danish Cancer Society too.

All the participants living in Denmark approximately five years after inclusion into the DCH cohort received a follow-up questionnaire, including a new food frequency questionnaire, updated life style questions and new questions on social network and self-rated health. The data from follow-up are not a part of the analysis in this Ph.D.-thesis.

In total, 57,053 participants, 29,875 women (37% of those invited) and 27,179 men (33% of those invited), were recruited. 547 of these participants (326 women and 221 men) were excluded from the cohort due to cancer diagnosis prior to the inclusion, either by a delay in the notification of their cancer to the Danish Cancer Registry or they had been diagnosed with cancer between the invitation and the visit to the study clinic. Among the cohort members, 405 cases of colorectal cancer were identified, 184 women and 221 men, in the files of the nationwide Danish Cancer Registry [6], diagnosed between 1994 and 2003. Within the cohort we defined a sub-cohort consisting of 368 women and 442 men, who were randomly selected, matched on gender. Blood samples were available for 397 cases and 800 members of the sub-cohort for the studies in manuscripts II-IV.

Differences between participants and non-participants have been analysed [7]. In brief, participation was higher among people with longer formal education and a positive association was seen between years in the workforce during the preceding 16 years and participation. Marital status was also associated to participation with a low participation rate among both single men, in particular, and women. The sociological differences between participants and non-participants may not have a crucial effect on the gene-environment analyses of this Ph.D.-thesis, since it is a nested case-cohort study. But a generalization of the results obtained to the Danish population must be made cautiously.

In the nested case-cohort study of this Ph.D.-thesis the anatomical distribution among colorectal cancer cases was not entirely similar with the distribution in the Danish population (table A): A smaller percentage was diagnosed with cancer in the proximal segment of colon (20.8%), while a larger proportion of the study group was diagnosed with cancer in the distal segment of colon (40.1%). The percentage of diagnoses of cancer in rectum was similar to the Danish population (36.0%).
Table A: The anatomical distribution of colorectal cancer in Denmark diagnosed in the period 2001-2005 reported by the Danish Colorectal Cancer Group in 2005 [8], and among the participants of the prospective Danish "Diet, Cancer and Health" cohort study diagnosed after enrolment till July 2003.

<table>
<thead>
<tr>
<th>Topology</th>
<th>Danish population 2001-2005</th>
<th>Diet, Cancer and Health enrolment-2003</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cecum (incl. appendix)</td>
<td>2080 (13.3%)</td>
<td>44 (10.9%)</td>
</tr>
<tr>
<td>Ascending colon</td>
<td>1133 (7.2%)</td>
<td>23 (5.7%)</td>
</tr>
<tr>
<td>Hepatic flexure</td>
<td>708 (4.5%)</td>
<td>3 (0.7%)</td>
</tr>
<tr>
<td>Transverse colon</td>
<td>872 (5.6%)</td>
<td>14 (3.5%)</td>
</tr>
<tr>
<td>Splenic flexure</td>
<td>508 (3.2%)</td>
<td>6 (1.5%)</td>
</tr>
<tr>
<td>Descending colon</td>
<td>433 (2.8%)</td>
<td>14 (3.5%)</td>
</tr>
<tr>
<td>Sigmoid colon</td>
<td>4573 (29.2%)</td>
<td>104 (25.7%)</td>
</tr>
<tr>
<td>Rectosigmoid junction</td>
<td>-</td>
<td>38 (9.4%)</td>
</tr>
<tr>
<td>Rectum (incl. anal canal)</td>
<td>5358 (34.2%)</td>
<td>146 (36.0%)</td>
</tr>
<tr>
<td>More than one segment</td>
<td>11 (0.1%)</td>
<td>-</td>
</tr>
<tr>
<td>Unknown</td>
<td>-</td>
<td>13 (3.2%)</td>
</tr>
<tr>
<td>Total</td>
<td>15676</td>
<td>405</td>
</tr>
</tbody>
</table>

The “Diet, Cancer and Health” study (DCH) is associated to the European Prospective Investigation in Cancer and Nutrition (EPIC), a co-work consisting of 23 study centers in 10 European countries. EPIC includes approximately 520,000 men and women, with data collections in the time period 1992 to 2000 [9].

The studies of lung cancer in manuscript V-VII include 431 cases of lung cancer (201 women and 230 men). Among the cohort members cases were identified in the Danish Cancer Registry, diagnosed between 1994 and 2003. For comparison group 365 women and 431 men were selected randomly among the cohort members. Blood samples were available for all individuals for the lung cancer studies.

The “Norwegian Colorectal Cancer Prevention Study”

The “Kolorektal cancer, Arv og Miljø” (KAM) cohort, studied in manuscript I, is partly based on the screening group of the “Norwegian Colorectal Cancer Prevention Study” (NORCCAP) in Telemark County.

Between January 1999 and December 2000 a total of 20,780 men and women (1:1), age distribution 50–64 years, were invited to have a flexible sigmoidoscopy screening examination with or without (1:1) an additional faecal occult blood test (FOBT) and thereby to participate in the “Norwegian Colorectal Cancer Prevention Study” (NORCCAP). The invited were drawn randomly from the population registries.
of the City of Oslo (urban) and Telemark County (mixed urban and rural) population. Seven hundred and seventy-seven individuals were excluded according to exclusion criteria, leaving 20,003 participants eligible for screening. The exclusion criteria for screening were: previous open colorectal surgery, ongoing cytotoxic or radiation therapy for malignant disease, severe chronic cardiopulmonary disease, life-long anticoagulant therapy, a coronary episode or cerebrovascular accident during the previous 3 months, disabled, inability to give written informed consent, resident abroad, unknown address or deceased. The remaining relevant age cohorts in the screening areas constituted the control group, not invited to any screening. Exclusion criteria were applicable only to the screening groups, since the control groups were not contacted. Control groups were not informed about their status as controls. Invitations to the study were sent out 6-7 weeks prior to a suggested day for an appointment free of charge at a local screening centre. If no respond three weeks before the suggested day of appointment a reminding letter was sent. Afterwards no further effort was made to include the individual in the study.

The clinical examinations of the participants and collection of biological samples were made at two centres for sigmoidoscopy and colonoscopy in Oslo and Telemark. Biological material was collected during the screening examination, including tissue samples from pathological material and normal colonic mucosa, blood samples, fresh frozen stool samples as well as FOBT slides [10]. The biological material was stored for future research. The histological examination of material from the Telemark population was performed at the Department of Pathology at the Norwegian Radium Hospital, Oslo and from the Oslo population at the Department of Pathology, Ulleval Hospital in Oslo. The tumor histology of the carcinomas and adenomas was examined independently by two specialist histopathologists in order to determine the tumour stage. All of the participants filled in a detailed questionnaire on dietary habits for the year before inclusion and on personal history, including smoking habits and family history of cancer.

In total 12,960 participants (65% of those invited) were recruited to the NORCCAP study. Differences between participants and non-participants have been analysed [11]. Participation in the NORCCAP study was slightly higher among women than among men (6661 (66%) versus 6299 (64%)), and the attendance rates were increasing by age. The participation rate was higher in the County of Telemark than in the City of Oslo (7224 (71%) versus 5736 (58%)).

During a limited period of time, after the screening study of NORCCAP was well established, controls (polyp free by sigmoidoscopy) were invited to participate in the KAM study. Written consent was obtained from all the participants. The KAM cohort is based on an ethnic homogeneous group of Norwegian origin. Cases and controls had no known personal history of genetic predisposition to cancer.
Ultimo year 2004 the KAM study consisted of 170 colorectal cancer cases (still collecting samples), 1044 cases with adenomas, and 400 controls. Adenoma cases and controls were all drawn from the NORCCAP study. All of the participants completed a questionnaire on demographics, health status, dietary, and smoking habits, alcohol consumption, physical exercise and occupation. The questionnaire contained information on a family history of adenomas and carcinomas, and the included cases and controls had no known personal history of genetic predisposition. The colorectal cancer cases of the KAM study consist of patients operated on at Telemark Hospital in Skien and Ulleval University Hospital in Oslo. The adenocarcinomas were collected prior to chemo- or radiotherapy treatment. Two specialist histopathologists examined the histology of the adenomas independently in order to determine the tumor stage as mild, moderate or severe dysplasia. They reached the same conclusion in all cases. In the case-control study presented in manuscript I of this Ph.D.-thesis, DNA samples were available for 166 cases with adenocarcinomas, 974 cases with adenomas and 397 controls.

Reference List


APPENDIX II - LABORATORY ANALYSIS

A more profound description of the laboratory analysis used is given as a supplement to the information presented in the manuscripts I-VII.

DNA extraction

Genomic DNA was extracted from buffy coat samples of nucleated cells from the DCH study according to a standard salting out procedure [1]. Briefly, the buffy coats were resuspended in a lysis buffer and centrifuged, where after the cell lysates were digested overnight at 37°C with a sodium dodecyl sulphate (SDS) and proteinase K solution. SDS is a detergent for lipid membranes and proteinase K digest proteins including active DNases in the cell lysates. Protein precipitation was carried out by adding saturated NaCl to the lysates. After centrifugation and transfer of the supernatant containing the DNA, absolute ethanol was added to the supernatant, whereby DNA precipitated. DNA was stored at -80°C until use. The DNA extraction was performed at the National Research Centre for Working Environment, Copenhagen.

DNA from the KAM study was extracted by Camilla Furu-Skjelbred and Mona Sæbo, Telemark University College and Telemark Hospital, Norway. The DNA extraction was performed as described for the DCH study with minor modifications in buffer contents and incubation time. The procedure and the modifications are presented briefly in manuscript I.

Real Time-Polymerase Chain Reaction, Endpoint Reading

The TaqMan assay makes use of fluorescence resonance energy transfer (FRET) between two dyes situated at each end of a probe. On the 3’ end of the probe a “quencher” (a low-energy molecule) is situated, which changes the wavelength of the light emitted from the “reporter” dye (high-energy dye) situated at the 5’ end of the probe. When the probe is intact, the close proximity of the quencher greatly reduces the fluorescence emitted by the reporter dye by FRET. The probes are included in the PCR amplification reaction together with the two primers, forward and reverse, and MasterMix® (Applied Biosystems, Nærum, Denmark), including AmpliTaq Gold® DNA polymerase, the passive reference dye ROX, Mg-ions, and the four dNTPs, and sample DNA of course. The probe anneals to the target sequence. If the probe is hybridized it will be cleaved by 5’-3’exonuclease activity of the AmpliTaq Gold®
DNA polymerase, which separates the reporter and the quencher and thereby increases the fluorescent signal from the reporter dye and are displayed by the software. The reporter dye is different for the two allele-specific probes in each assay, emitting fluorescence at different wavelengths. Thus, it is possible to distinguish between the genotypes: the homozygote for each of the two alleles emitting a high signal at the wavelength of only one of the reporter dyes, and the heterozygote, carrying one of each allele, emitting an intermediate signal at both wavelengths.

In RT-PCR the increase in fluorescence associated with the exponential growth of PCR product is recorded by the software for each cycle. In the studies of this Ph.D.-thesis all genotypes were determined by endpoint reading on a Sequence Detection System ABI 7500 (Applied Biosystems, Nærum, Denmark), measuring the amount of accumulated PCR product at the end of the PCR process.

Figure 1A to the right shows the display for allelic discrimination plot for a 96-well reaction plate. The software plots the results of the allelic discrimination run on a scatter plot of allele X versus allele Y. Delta Rn are the magnitude of the signal generated by a set of PCR conditions (delta Rn = Rn − baseline). The clustering of points can vary along the horizontal axis (allele X), vertical axis (allele Y), or diagonal (allele X/allele Y). This variation is due to differences in the extent of reporter dye fluorescent intensity after PCR amplification. The genotype are assigned manually for each sample/cluster: Blue dots – homozygous allele Y, Red dots – homozygous allele X, Green dots – heterozygous alleles X and Y, Grey squares – no template control. Figure 1B shows the display of the linear amplification plot (Rn vs. cycle-plot) during RT-PCR. The Rn vs. cycle-plot displays normalized reporter (Rn) dye fluorescence as a function of cycle. The two displays in figure 1B shows the fluorescent signal from each of the two reporter dyes of allele X and Y.

The primers and probes used were already designed and optimized with respect to optimal concentrations at The National Research Centre for Working Environment at previous studies. Primers and probes were produced by Applied Biosystems, Nærum, Denmark, and TAGCopenhagen, Copenhagen, Denmark.
Measurement of Enzyme Activity

The GPX activity was determined spectrophotometrically in erythrocyte lysates on an Automated Hitachi 912 Analyzer (Boehringer Mannheim). The GPX activity was determined according to an established method [2] with a minor modification: t-butyl hydroperoxide, tBHQ, was used as substrate instead of \( \text{H}_2\text{O}_2 \). When samples and reagents were placed in the Automated Hitachi 912 Analyzer, the addition of sample and reagents, mixing, spectrophotometric analyses and calculations were performed automatically according to the programmed instructions. The procedure is presented in details in manuscript II.

Briefly, GPX catalyses the oxidation of glutathione (GSH) in presence of tBHQ into water, alcohol and GSSG. The GSH oxidation is coupled to the reaction, where NADPH is oxidized into NADP\(^+\) by the enzyme glutathione reductase (GR). The conversion of NADPH to NADP\(^+\) was measured. The spectrophotometric analyses were performed at 340 nm for 5 minutes after the addition of tBHQ. The assay condition was 37 °C. A commercially available kit, Drapkin’s Reagent (Randox, UK, cat.no.HG 980), was used for determination of the haemoglobin, Hb, content in the blood samples. Drapkin’s Reagent was added in double concentration to the blood samples to inhibit other peroxidases and to reduce any GSSG present in the hemolysate. Absorbance was measured at 540 nm for 5 minutes. The calculations of GPX activity related to the amount of Hb in the blood sample were performed automatically according to the programmed instructions of the Automated Hitachi 912 Analyzer.

Quality Control

We investigated primarily polymorphisms for which the allelic frequencies of the variant alleles were above 0.30 to provide a sufficient statistical power, but for the polymorphisms in RAI and ASE-1 the allelic frequencies of the variant alleles was lower (0.19 and 0.17 respectively). The case-control status of the samples was blinded when analysing the genotypes in the laboratory. Samples with known genotypes were included in each run, and repeated genotyping of a random 10% subset yielded 100% identical genotypes. RT-PCR data and data from endpoint readings were assured to give 100% identical genotype results for the first 2 plates in every polymorphism assay before the rest of the samples were determined by endpoint readings alone. The genotype distributions were in Hardy-Weinberg equilibrium in the control group or sub-cohort for all the polymorphisms studied.
For measurement of GPX enzyme activity control samples pooled from 5 healthy donors were included twice in each batch: one in the beginning and one in the end of each batch containing a maximum of 48 samples. Measurement of 48 samples lasted approximately 1½ hour. By including the control samples we were able to observe whether there was changes in the enzyme activity measured with 1½ hour intervals. For the possibility that the temperature and precipitation of the sample in that time period influenced the results obtained.

Reference List


APPENDIX III - STATISTICAL ANALYSIS

All statistical analyses for the Norwegian study in manuscript I were conducted by Mona Sæbø and Per Christian Hagen in Oslo, Norway, and Bjørn Andersen Nexø in Aarhus, Denmark. All estimates are crude. Potential confounders may be similar as to the Danish study, but at the time of the analyses the data from the questionnaires were not yet obtained.

All statistical analyses for the Danish lung cancer studies in manuscript V-VII were conducted by Mette Sørensen and Ole Raaschou-Nielsen in Copenhagen, Denmark.

Statistical methods

In the Danish study, manuscript II-IV, cases and sub-cohort were matched on gender. Incidence rate ratios were estimated by the Cox proportional hazards model stratified according to sex, with age as the underlying time axis, which ensured that the estimation procedure was based on comparison of individuals at the same age. The analyses were corrected for delayed entry, such that persons were considered under risk only from the age at enrolment in the cohort. We calculated two-sided 95% confidence intervals (CI) and p-values based on robust estimates of the variance-covariance matrix [1] and Wald’s test of the Cox regression parameter, that is, on the log rate ratio scale.

All quantitative variables were entered linearly in the Cox proportional hazards model. This is biologically more reasonable than the step functions corresponding to categorization, and furthermore, provides a higher power [2]. For each of the life style factors, the hypothesis of a linear association with risk of colorectal cancer was evaluated using a linear spline with three boundaries, placed at the quartiles among cases, as covariates in the Cox model [3]. The linearity was evaluated graphically and by a numerical test using the likelihood ratio test statistic to compare the model assuming linearity with the linear spline model. All variables were found linearly associated with colorectal cancer, and showed no signs of threshold values.

When analyzing interactions between the genotypes and the life style factors smoking intensity and alcohol consumption only exposed individuals were included. This was done by inclusion of a dichotomous indicator for ever-smoking (yes/no) and an indicator for alcohol drinking (yes/no) in the model.
The effects of genotype and environmental factors on GPX activity (manuscript II) was determined by using a multiple linear regression model. We investigated for correlation between the variables. We included the variables gender, genotype, smoking intensity, alcohol consumption, intake of dietary fibres, fruit and vegetables, selenium, and vitamin E and C, in the full model and removed one variable at the time by the backward stepwise approach, until only the variables with a p-value below 0.20, were included.

We used the procedure PHREG in the statistical software program SAS (version 8.2) for analyses of risk for colorectal cancer (the logistic regression analyses), whereas the procedure GLM was used for the analyses of predictors for GPX enzyme activity (the univariate and multiple linear regression analyses).

**Potential confounders**

A set of potential confounders were selected based on the available litterature: Smoking intensity, alcohol consumption, intake of red and processed meat, fruits and vegetables and dietary fibres, use of hormonal replacement therapy (HRT) or of non-steroid anti-inflammatory drugs (NSAID), physical activity and BMI. The potential confounders are previously observed as factors affecting risk of development of colorectal cancer. However, they were not statistically significant associated to risk of colorectal cancer in the present DCH case-cohort study, table B.

We chose to constrain the number of potential confounders to the most important based on a litterature-based assessment and our own calculations of association of the factors to risk of colorectal cancer in the present study, and thereby to limit the loss of statistical power: The variable average smoking intensity was included in preference to the many other smoking variables from the questionnaires, e.g. smoking duration and pack-years, because of the strongest association, although not statistically significant, to risk of colorectal cancer in the studies of this Ph.D.-thesis. Intake of fruit and vegetables were pooled into one variable (fruitveg), and intake of fish and poultry were also pooled (white meat).
Table B: Characteristics of the study group nested within the Danish cohort Diet, Cancer and Health

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Cases</th>
<th>Sub-cohort</th>
<th>Crude IRR^c</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Total</strong></td>
<td>405 (100)</td>
<td>810 (100)</td>
<td>-</td>
</tr>
<tr>
<td><strong>Gender, men/women</strong></td>
<td>221/184</td>
<td>434/366</td>
<td>-</td>
</tr>
<tr>
<td><strong>BMI</strong></td>
<td>26 (21-34)</td>
<td>26 (20-33)</td>
<td>1.01 (0.92-1.12)</td>
</tr>
<tr>
<td><strong>Food intake</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fruit, g/day</td>
<td>153 (25-464)</td>
<td>175 (28-548)</td>
<td>1.07 (0.89-1.27)</td>
</tr>
<tr>
<td>Vegetables, g/day</td>
<td>152 (46-345)</td>
<td>165 (48-364)</td>
<td>1.32 (0.99-1.75)</td>
</tr>
<tr>
<td>Fruit and Vegetables (FruVeg)</td>
<td></td>
<td></td>
<td>1.10 (0.96-1.26)</td>
</tr>
<tr>
<td>Poultry, g/day</td>
<td>17 (3-51)</td>
<td>18 (4-64)</td>
<td>0.96 (0.78-1.19)</td>
</tr>
<tr>
<td>Fish, g/day</td>
<td>38 (10-89)</td>
<td>39 (10-92)</td>
<td>1.04 (0.94-1.16)</td>
</tr>
<tr>
<td>Fish and Poultry (white meat)</td>
<td></td>
<td></td>
<td>1.19 (0.48-2.92)</td>
</tr>
<tr>
<td>Red meat, g/day</td>
<td>82 (38-170)</td>
<td>81 (31-176)</td>
<td>0.82 (0.37-1.83)</td>
</tr>
<tr>
<td>Processed meat, g/day</td>
<td>26 (6-79)</td>
<td>23 (4-76)</td>
<td>0.87 (0.74-1.03)</td>
</tr>
<tr>
<td>Dietary fibres, g/day</td>
<td>19 (9-30)</td>
<td>20 (11-34)</td>
<td>1.01 (0.60-1.70)</td>
</tr>
<tr>
<td><strong>Alcohol consumption</strong>, g/day</td>
<td>14 (1-68)</td>
<td>12 (1-62)</td>
<td>1.10 (0.99-1.24)</td>
</tr>
<tr>
<td><strong>Smoking status at inclusion</strong>, ever/never</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>282 (69%) / 123 (31%)</td>
<td>538 (66%) / 272 (34%)</td>
<td>0.79 (0.40-1.53)</td>
</tr>
<tr>
<td>Former</td>
<td>123 (30%)</td>
<td>272 (34%)</td>
<td></td>
</tr>
<tr>
<td>Present</td>
<td>122 (30%)</td>
<td>252 (31%)</td>
<td></td>
</tr>
<tr>
<td>Average smoking intensity among ever smokers, g tobacco/day</td>
<td>16 (5-37)</td>
<td>15 (4-35)</td>
<td>1.21 (0.71-2.05)</td>
</tr>
<tr>
<td><strong>Smoking duration among ever smokers</strong>, years</td>
<td>34 (8-46)</td>
<td>32 (7-46)</td>
<td>0.86 (0.64-1.15)</td>
</tr>
<tr>
<td><strong>HRT use among women</strong>, ever/never</td>
<td>75 (19%) / 330 (81%)</td>
<td>173 (21%) / 637 (79%)</td>
<td>1.31 (0.50-3.43)</td>
</tr>
<tr>
<td>NSAID use, ever/never</td>
<td>126 (32%) / 274 (68%)</td>
<td>247 (31%) / 554 (69%)</td>
<td>0.98 (0.49-1.95)</td>
</tr>
<tr>
<td>Missing information</td>
<td>5</td>
<td>9</td>
<td></td>
</tr>
</tbody>
</table>

a) Number (%)  
b) Median (5-95% percentiles)  
c) IRR: Incidence rate ratio

I chose not to include use of NSAID and level of physical activity in the analyses. Previously, no association has been observed between the level of physical activity and risk of colon cancer in the DCH cohort study [4] and the results obtained from the literature are inconsistent whether physical activity modify risk of colorectal cancer. A regular use of NSAIDs in high doses and for more than 10 years are suggested to have an effect on risk of colorectal cancer. The NSAID information I used was a dichotomous indicator for ever/never use of NSAID (yes/no). This variable do not contain informations of duration, doses or type of NSAID used (aspirin/COX-2 inhibitors/others). Adjustment for use of NSAID (ever/never) and physical activity did not change the estimates when analysing the possible
association between the various genotypes with risk of colorectal cancer, thus the two life style factors were excluded from the analyses. However, I have recently become aware of another variable in the DCH study, daily use of NSAID, that may be correct to use in the model.

Reference List


