Epigenetic alterations
in sporadic and familial cancers

Emmi Joensuu

Department of Medical and Clinical Genetics
University of Helsinki

Helsinki Graduate Program in Biotechnology and Molecular Biology
Integrative Life Science Doctoral Program

ACADEMIC DISSERTATION

To be publicly discussed,
with the permission of the Faculty of Medicine, University of Helsinki
in the Auditorium XIV, University main building, Fabianinkatu 33, Helsinki
on the 20th of November 2015, at 12 noon.
Supervisor  Professor Päivi Peltomäki, MD, PhD
Department of Medical and Clinical Genetics
University of Helsinki
Finland

Reviewers  Docent Anu Loukola, PhD
Department of Public Health
University of Helsinki
Finland

Docent Rainer Lehtonen, PhD
Research Programs Unit of the Faculty of Medicine
University of Helsinki
Finland

Opponent  Professor Anne Kallioniemi, MD, PhD
Department of Biomedical Sciences
University of Tampere
Finland

ISSN 2342-3161 (Print)
ISSN 2342-317X (Online)


Hansaprint Oy
Vantaa 2015
“I used to be an adventurer like you, but then I took an arrow in the knee.”
~Guard
# CONTENTS

## ABBREVIATIONS

## LIST OF ORIGINAL PUBLICATIONS

## ABSTRACT

## INTRODUCTION

## REVIEW OF THE LITERATURE

### 1 Epigenetics overview
- 1.1 DNA methylation
- 1.2 Post-translational histone modifications and chromatin remodeling
- 1.3 Noncoding RNAs and RNA interference
- 1.4 Interplay between different epigenetic factors
- 1.5 Epigenetic inheritance

### 2 Basic characteristics of cancer

### 3 Cancer genetics
- 3.1 Tumor suppressor genes
- 3.2 Oncogenes
- 3.3 Genomic instability
  - 3.3.1 Chromosomal instability and loss of heterozygosity
  - 3.3.2 Microsatellite instability
  - 3.3.2.1 Mismatch repair (MMR) pathway
- 3.4 Colorectal cancer
  - 3.4.1 Hereditary non-polyposis colorectal cancer syndrome (HNPCC)
  - 3.4.2 Lynch syndrome (LS)
  - 3.4.3 Familial colorectal cancer type X (FCCX)
- 3.5 Endometrial cancer
  - 3.5.1 Lynch syndrome associated endometrial cancer
  - 3.5.2 Familial site-specific endometrial cancer (FSSEC)
- 3.6 Gastric cancer

### 4 Cancer epigenetics
- 4.1 DNA hypermethylation in cancer
- 4.2 DNA hypomethylation in cancer
- 4.3 Other epigenetic events in cancer

## AIMS OF THE STUDY

## MATERIALS & METHODS

### 1 Patients and samples

### 2 Genomic mutation, microsatellite instability and loss of heterozygosity analyses

### 3 DNA methylation analyses

### 4 Protein expression analyses
5 Statistical analyses

RESULTS

1 DNA hypermethylation in cancer (I, II)
   1.1 Shared and tissue-specific tumor suppressor gene methylation patterns
   1.2 DNA methylation patterns in separate patient categories – dependence of methylation patterns on the MSI status and the familial background of cancer
   1.3 CpG island methylator phenotype as a driver of tumorigenesis?

2 Global genomic hypomethylation in cancer (III)

3 Changes in methyltransferase activity in cancer (IV)
   3.1 Aberrant DNA methylation – aberrant methyltransferase patterns?
   3.2 Looking for underlying causes for methyltransferase protein overexpression

DISCUSSION

1 DNA methylation in colorectal, endometrial and gastric cancers
   1.1 DNA methylation patterns as tumor classifiers
   1.2 Methodological aspects

2 Elucidating the background of methylation changes

3 Overall conclusions of the findings

CONCLUSIONS & FUTURE PROSPECTS

ACKNOWLEDGEMENTS

REFERENCES
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AC-1</td>
<td>Amsterdam criteria I</td>
</tr>
<tr>
<td>ADP</td>
<td>Adenosine diphosphate</td>
</tr>
<tr>
<td>bp</td>
<td>Base pair</td>
</tr>
<tr>
<td>BS</td>
<td>Bisulfite sequencing</td>
</tr>
<tr>
<td>CIMP</td>
<td>CpG island methylator phenotype</td>
</tr>
<tr>
<td>CIN</td>
<td>Chromosomal instability</td>
</tr>
<tr>
<td>COBRA</td>
<td>Combined bisulfite restriction analysis</td>
</tr>
<tr>
<td>CpG</td>
<td>Cytosine-guanine dinucleotide</td>
</tr>
<tr>
<td>CRC</td>
<td>Colorectal cancer / carcinoma</td>
</tr>
<tr>
<td>Dm</td>
<td>Methylation dosage ratio</td>
</tr>
<tr>
<td>EC</td>
<td>Endometrial cancer / carcinoma</td>
</tr>
<tr>
<td>FAP</td>
<td>Familial adenomatous polyposis</td>
</tr>
<tr>
<td>FCCX</td>
<td>Familial colorectal cancer type X</td>
</tr>
<tr>
<td>FFPE</td>
<td>Formalin-fixed paraffin embedded</td>
</tr>
<tr>
<td>FSSEC</td>
<td>Familial site-specific endometrial cancer</td>
</tr>
<tr>
<td>GC</td>
<td>Gastric cancer / carcinoma</td>
</tr>
<tr>
<td>HDGC</td>
<td>Hereditary diffuse gastric cancer</td>
</tr>
<tr>
<td>HNPCC</td>
<td>Hereditary non-polyposis colorectal cancer syndrome</td>
</tr>
<tr>
<td>H3K4</td>
<td>Histone 3 lysine 4</td>
</tr>
<tr>
<td>H3K27me3</td>
<td>Histone 3 lysine 27 trimethylation</td>
</tr>
<tr>
<td>IDL</td>
<td>Insertion-deletion loop</td>
</tr>
<tr>
<td>IHC</td>
<td>Immunohistochemistry</td>
</tr>
<tr>
<td>kb</td>
<td>Kilobase</td>
</tr>
<tr>
<td>LINE-1</td>
<td>Long interspersed nuclear element 1</td>
</tr>
<tr>
<td>lncRNA</td>
<td>Long noncoding RNA</td>
</tr>
<tr>
<td>LOH</td>
<td>Loss of heterozygosity</td>
</tr>
<tr>
<td>LS</td>
<td>Lynch syndrome</td>
</tr>
<tr>
<td>MINT</td>
<td>Methylated-in-tumors</td>
</tr>
<tr>
<td>miRNA</td>
<td>MicroRNA</td>
</tr>
<tr>
<td>MLPA</td>
<td>Multiplex ligation-dependent probe amplification</td>
</tr>
<tr>
<td>MMR</td>
<td>DNA mismatch repair</td>
</tr>
<tr>
<td>mRNA</td>
<td>Messenger RNA</td>
</tr>
<tr>
<td>MSI</td>
<td>Microsatellite instability</td>
</tr>
<tr>
<td>MS-MLPA</td>
<td>Methylation-specific MLPA</td>
</tr>
<tr>
<td>MSP</td>
<td>Methylation-specific PCR</td>
</tr>
<tr>
<td>MSS</td>
<td>Microsatellite stable</td>
</tr>
<tr>
<td>MT</td>
<td>Methyltransferase</td>
</tr>
<tr>
<td>MutLα/γ</td>
<td>MutL homolog α / γ</td>
</tr>
<tr>
<td>MutSα/β</td>
<td>MutS homolog α / β</td>
</tr>
<tr>
<td>MZ</td>
<td>Monozygotic</td>
</tr>
<tr>
<td>ncRNA</td>
<td>Noncoding RNA</td>
</tr>
<tr>
<td>nt</td>
<td>Nucleotide</td>
</tr>
<tr>
<td>PRC2</td>
<td>Polycomb repressive complex 2</td>
</tr>
<tr>
<td>RNAi</td>
<td>RNA interference</td>
</tr>
<tr>
<td>SAM</td>
<td>S-adenosyl-L-methionine</td>
</tr>
<tr>
<td>SI</td>
<td>Staining index</td>
</tr>
<tr>
<td>siRNA</td>
<td>Small interfering RNA</td>
</tr>
<tr>
<td>TNM</td>
<td>Tumor-nodes-metastasis</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>TSG</td>
<td>Tumor suppressor gene</td>
</tr>
<tr>
<td>5caC</td>
<td>5-carboxylycytosine</td>
</tr>
<tr>
<td>5fC</td>
<td>5-formylycytosine</td>
</tr>
<tr>
<td>5-FU</td>
<td>5-fluorouracil</td>
</tr>
<tr>
<td>5mC</td>
<td>5-methylcytosine</td>
</tr>
<tr>
<td>5hmC</td>
<td>5-hydroxymethylcytosine</td>
</tr>
<tr>
<td>-CH₃</td>
<td>Methyl group</td>
</tr>
</tbody>
</table>

**Gene/protein abbreviations**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Akt</td>
<td>Protein kinase B</td>
</tr>
<tr>
<td>APC</td>
<td>Adenomatous polyposis coli</td>
</tr>
<tr>
<td>ARID1A</td>
<td>AT rich interactive domain 1A</td>
</tr>
<tr>
<td>ATM</td>
<td>Ataxia telangiectasia mutated</td>
</tr>
<tr>
<td>ATR</td>
<td>Ataxia telangiectasia and Rad3-related protein</td>
</tr>
<tr>
<td>β-catenin</td>
<td>Cadherin-associated protein beta 1</td>
</tr>
<tr>
<td>BAX</td>
<td>BCL2-associated X protein</td>
</tr>
<tr>
<td>BMI1</td>
<td>Polycomb group RING finger protein 4</td>
</tr>
<tr>
<td>BMPR1A</td>
<td>Bone morphogenetic protein receptor type IA</td>
</tr>
<tr>
<td>BRAF</td>
<td>B-Raf proto-oncogene</td>
</tr>
<tr>
<td>BRCA1/2</td>
<td>Breast cancer 1/2, early onset</td>
</tr>
<tr>
<td>CADM1 (IGSF4)</td>
<td>Immunoglobulin superfamily member 4</td>
</tr>
<tr>
<td>CASP8</td>
<td>Caspase 8, apoptosis-related cysteine peptidase</td>
</tr>
<tr>
<td>CD44</td>
<td>CD44 antigen precursor</td>
</tr>
<tr>
<td>CDH1</td>
<td>Cadherin 1, type 1; E-cadherin</td>
</tr>
<tr>
<td>CDH13</td>
<td>Cadherin 13; H-cadherin</td>
</tr>
<tr>
<td>CDKN2A/B</td>
<td>Cyclin-dependent kinase inhibitor 2 A / B</td>
</tr>
<tr>
<td>CHFR</td>
<td>Checkpoint with forkhead and ring</td>
</tr>
<tr>
<td>CTNNB1</td>
<td>gene encoding β-catenin</td>
</tr>
<tr>
<td>DAPK1</td>
<td>Death-associated protein kinase 1</td>
</tr>
<tr>
<td>DNMT1</td>
<td>DNA methyltransferase 1</td>
</tr>
<tr>
<td>DNMT3A/B</td>
<td>DNA methyltransferase 3 isoform A / isoform B</td>
</tr>
<tr>
<td>EPCAM (TACSTD1)</td>
<td>Epithelial cell adhesion molecule</td>
</tr>
<tr>
<td>ERBB2 (HER2/neu)</td>
<td>Human epidermal growth factor receptor 2</td>
</tr>
<tr>
<td>ESR1</td>
<td>Estrogen receptor 1</td>
</tr>
<tr>
<td>EZH2</td>
<td>Enhancer of Zeste homolog 2</td>
</tr>
<tr>
<td>FGFR2</td>
<td>Fibroblast growth factor receptor 2</td>
</tr>
<tr>
<td>FHIT</td>
<td>Fragile histidine triad</td>
</tr>
<tr>
<td>GSTP1</td>
<td>Glutathione S-transferase pi 1</td>
</tr>
<tr>
<td>HAT (1)</td>
<td>Histone acetyltransferase (1)</td>
</tr>
<tr>
<td>HDAC (1-3)</td>
<td>Histone deacetylase (1-3)</td>
</tr>
<tr>
<td>HIC1</td>
<td>Hypermethylated in cancer 1</td>
</tr>
<tr>
<td>HOTAI R</td>
<td>HOX transcript antisense RNA</td>
</tr>
<tr>
<td>H2A/B</td>
<td>Histone 2A/B family</td>
</tr>
<tr>
<td>H3</td>
<td>Histone 3 family</td>
</tr>
<tr>
<td>H4</td>
<td>Histone 4 family</td>
</tr>
<tr>
<td>KDM1A (LSD1)</td>
<td>Lysine (K)-specific histone demethylase 1A</td>
</tr>
<tr>
<td>Ki-67</td>
<td>Antigen Ki-67</td>
</tr>
<tr>
<td>KRAS</td>
<td>Kirsten rat sarcoma viral oncogene homolog</td>
</tr>
<tr>
<td>MBD (1-4)</td>
<td>Methyl CpG binding domain protein (1-4)</td>
</tr>
<tr>
<td>MECP2</td>
<td>Methyl CpG binding protein 2</td>
</tr>
<tr>
<td>Gene</td>
<td>Description</td>
</tr>
<tr>
<td>-------</td>
<td>-------------</td>
</tr>
<tr>
<td>MEX3C</td>
<td>Mex-3 RNA binding family member C</td>
</tr>
<tr>
<td>MLH1/3</td>
<td>MutL homolog 1 / 3</td>
</tr>
<tr>
<td>MRE11</td>
<td>Meiotic recombination 11 homolog A</td>
</tr>
<tr>
<td>MSH2/3/6</td>
<td>MutS homolog 2 / 3 / 6</td>
</tr>
<tr>
<td>NME1</td>
<td>NME/NM23 nucleoside diphosphate kinase 1</td>
</tr>
<tr>
<td>PCNA</td>
<td>Proliferating cell nuclear antigen</td>
</tr>
<tr>
<td>PIK3CA</td>
<td>Phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha</td>
</tr>
<tr>
<td>PMS2/1</td>
<td>Postmeiotic segregation increased 2 / 1</td>
</tr>
<tr>
<td>PTEN</td>
<td>Phosphatase and tensin homolog</td>
</tr>
<tr>
<td>p16</td>
<td>Cyclin-dependent kinase inhibitor 2A, encoded by CDKN2A</td>
</tr>
<tr>
<td>p53</td>
<td>Tumor protein 53</td>
</tr>
<tr>
<td>RARB</td>
<td>Retinoic acid receptor beta</td>
</tr>
<tr>
<td>RASSF1</td>
<td>Ras association (RalGDS/AF-6) domain family 1</td>
</tr>
<tr>
<td>RB1</td>
<td>Retinoblastoma 1</td>
</tr>
<tr>
<td>REST</td>
<td>RE1-silencing transcription factor</td>
</tr>
<tr>
<td>RPS20</td>
<td>Ribosomal protein S20</td>
</tr>
<tr>
<td>SIRT1</td>
<td>Sirtuin 1, NAD-dependent deacetylase sirtuin-1</td>
</tr>
<tr>
<td>SMAD4</td>
<td>SMAD family member 4</td>
</tr>
<tr>
<td>TET</td>
<td>Ten-eleven translocation methylcytosine dioxygenase (family)</td>
</tr>
<tr>
<td>TGFBR2/II</td>
<td>Transforming growth factor beta receptor II</td>
</tr>
<tr>
<td>THBS1</td>
<td>Thrombospondin 1</td>
</tr>
<tr>
<td>TIMP3</td>
<td>Tissue inhibitor of metalloproteinases 3</td>
</tr>
<tr>
<td>TP53</td>
<td>gene encoding p53</td>
</tr>
<tr>
<td>TP73</td>
<td>Cellular tumor antigen p73</td>
</tr>
<tr>
<td>VHL</td>
<td>Von Hippel-Lindau disease tumor suppressor</td>
</tr>
<tr>
<td>XIST</td>
<td>X-inactive specific transcript (non-protein coding)</td>
</tr>
<tr>
<td>ZNF516</td>
<td>Zinc finger protein 516</td>
</tr>
</tbody>
</table>
LIST OF ORIGINAL PUBLICATIONS

This thesis is based on the following publications:


The publications are referred to in the text by their roman numerals (I-IV).

Publication II was also included in the thesis of PhD Annette Gylling (“Molecular mechanisms of cancer predisposition in HNPCC/Lynch syndrome”) in 2008.

All the original research articles in this book are reprinted with the permission from the publishers.
Cancer usually arises through mutational changes in the genome but also epigenetic changes can contribute to tumorigenesis. In this research we studied both sporadically occurring and familial colorectal, endometrial and gastric tumors. Sporadic tumors were divided into separate categories depending on the microsatellite instability status of the tumor. In addition to sporadic tumors we studied tumors from patients with different cancer syndromes: Lynch syndrome, Familial colorectal cancer type X and Familial site-specific endometrial cancer. Lynch syndrome patients have a predisposing germline mutation in one of the mismatch repair genes (MLH1, MSH2 or MSH6) and the tumors are typically microsatellite unstable. Despite the extensive research efforts, the genetic or epigenetic background of the other studied syndromes is not known and remains to be molecularly characterized. We therefore explored the possible epigenetic basis of cancer susceptibility in these syndromes.

First we studied the promoter methylation of 24 established tumor suppressor genes. Hypermethylation patterns were found to be characteristic of each tissue and diversely dependent on the microsatellite instability status of the tumor, or family category. The CpG island methylator phenotype (CIMP) in which multiple loci are silenced by promoter methylation, was most evident in sporadic microsatellite unstable tumors ($P < 0.001$) and was present in 38% of all of the studied colorectal, 19% of endometrial and 29% of gastric tumors. In these tumors the CIMP phenotype can contribute to the genomic instability and the progression of cancer. In addition, despite being microsatellite stable, 50% of Familial colorectal cancer type X tumors displayed the CIMP phenotype.

Our results of global hypomethylation confirm that tumors have significantly lower methylation levels compared to normal tissues in most of the studied patient groups ($P < 0.05$) and that the hypomethylation levels depend significantly on the microsatellite instability status of the tumors ($P = 0.042$ for colorectal and $P = 0.018$ for gastric tumors). The significant decrease in the methylation levels, observed especially in the normal tissues of Familial colorectal cancer type X patients, could function as a premalignant field defect, where a large area of tissue is affected by carcinogenic alteration, and hence promote cancer development by facilitating the accumulation of other lesions such as genetic mutations or other epigenetic changes in the affected areas.

After the characterization of different DNA methylation aberrations in distinct tumor categories, we studied the possible mechanisms behind the observed methylation changes. We evaluated the
association of the expression of DNA methyltransferases DNMT1 and DNMT3B and histone methyltransferase EZH2 with CIMP+ phenotype and global hypomethylation patterns. Compared to the normal tissues, all the studied methyltransferases were significantly overexpressed in colorectal tumors ($P < 0.001$) and DNMT3B also in endometrial tumors ($P < 0.001$). EZH2 overexpression was shown to associate with CIMP+ phenotype especially in sporadic colorectal tumors and the finding was statistically significant ($P = 0.003$).

The overall aim of this research was to elucidate epigenetic mechanisms in cancer, including cancers of different organs and also different familial cancers. Available information on the epigenetic events of cancers is increasing and although the topic is under continuous study, our understanding of it is still limited. New knowledge in the field can increase the understanding of the basic tumorigenic mechanisms and thereby facilitate more specific and earlier diagnosis and treatment of different types of cancer. Also the potential reversibility of epigenetic states offers interesting possibilities for drug development.
INTRODUCTION

Cancer is among the leading causes of death being responsible for about 15% of all deaths worldwide. Colorectal cancer is the third most common cancer and a significant cause of cancer-related deaths accounting for over 690,000 deaths annually. Gastric cancer is the fifth most common cancer globally and endometrial cancer the sixth most common cancer among women. In the Finnish population, over 3850 people are diagnosed with some of these cancers every year.

Cancer arises from the concurrent or sequential accumulation of mutations in oncogenes and tumor suppressor genes. These genes are essential for normal cell functions: they code for proteins that help to regulate cell growth and differentiation, proto-oncogenes by activating and tumor suppressor genes by limiting growth. The progress of a normal cell into a cancer cell is slow but accelerates when multiple different mutations cluster. In tumorigenesis, the function of at least one of the DNA repairing mechanisms will usually be lost allowing more mutations to occur. In tumorigenesis epigenetic changes can occur simultaneously with gene mutations.

Epigenetics refers to the regulation of gene expression in the absence of mutational changes in DNA sequence through certain chemical changes such as DNA methylation and various histone modifications or microRNA function. In mammalian DNA methylation, an additional methyl group is attached to the cytosine in cytosine-guanine dinucleotides (CpG) in DNA. Gene promoter methylation is a normal and widely used control mechanism in cells for gene expression regulation. DNA methylation controls the DNA transcription of a given gene usually by blocking the transcription. DNA methylation is essential, for example, in embryonic development and cell differentiation.

Many tumor suppressor genes are known to be inactivated by the hypermethylation of critical promoter region CpG sites in different cancers and other diseases. Tumor suppressor genes are usually not methylated in normal cells. Another DNA methylation abnormality in cancer is global hypomethylation at certain repetitive sequences. The underlying causes for abnormal function of DNA methylation pathway enzymes resulting in DNA hyper- or hypomethylation or errors in histone modification pathways are not well-known at present. In this research we investigated the roles of different DNA methylation changes in cancers of different origins and clarify the underlying mechanisms of the DNA methylation changes in these tumors.
REVIEW OF THE LITERATURE

1 Epigenetics overview

Genetic information of an organism is encoded in the DNA sequence. Epigenetics refers to the regulation of gene expression through certain chemical changes such as DNA methylation or histone modifications or the function of noncoding RNAs, without involving mutational changes in DNA sequence (Brazel and Vernimmen, 2015). The genetic material, the genome, is the same in every somatic cell of an individual. Epigenetic events are the switches that guide and regulate the expression of the genotype into various visible phenotypes and functions in different cells, tissues and individuals (Goldberg et al., 2007), also in monozygotic twins (Castillo-Fernandez et al., 2014).

The term epigenetics was first introduced in 1942 to combine the words epigenesis (‘epi’ + genesis = “above the development”) and genetics, and was defined as “the branch of biology which studies the causal interactions between genes and their products, which bring the phenotype into being” by Conrad Waddington (Waddington, 1942; Waddington, 1968).

Today, epigenetic control of gene expression is thought to consist of transcriptional gene activation or repression and post-transcriptional gene silencing (Rivera and Bennett, 2010). Transcriptional events include DNA methylation (e.g. in gene or transposon silencing and genomic imprinting), covalent histone modifications affecting gene activation or silencing and RNA-directed DNA methylation. Post-transcriptional gene silencing involves mRNA degradation by RNA interference (Murray et al., 2014).

Epigenetic events are usually reversible (Tompkins et al., 2012). The transcriptional control of the gene expression has to have the potential to be switched on and off when needed. During embryonic development and cell differentiation cells must maintain their epigenetic flexibility to ensure the possibility of different tissues to form (Goldberg 2007; Heyn et al., 2013). On the other hand, in order to retain a normal function, a fully developed and differentiated cell must maintain its tissue-specific epigenetic and genetic stability and cellular form (Murray et al., 2014).

Different epigenetic mechanisms e.g. DNA methylation and histone modifications also functionally interact with each other (Choi and Lee, 2013). The epigenetic identity of a cell is
transferred from mother to daughter cells through cell division (Goldberg et al., 2007) but the mechanisms still remain largely unknown.

The maintenance of epigenetic homeostasis is a sophisticated, complex and strictly controlled process and the malfunction in this process may predispose to cancer or other diseases (Choi and Lee, 2013).

1.1 DNA methylation

DNA methylation is a normal and widely used control mechanism in cells (Jones, 2012) and it is associated with the silencing of repetitive and centromeric sequences and transposable elements throughout the genome, as well as in genomic imprinting and X-chromosome inactivation (dosage compensation in human females) (Portela and Esteller, 2010). DNA methylation is also actively used in transcriptional gene silencing, chromatin compaction and it can function as a genomic defense mechanism against foreign (e.g. viral) DNA in cells (Brena et al., 2006). Most mammalian DNA methylation involves the covalent addition of a methyl group (-CH₃) to the fifth carbon of cytosine followed by guanine (CpG dinucleotides) to form 5-methylcytosine (5mC) in DNA (Fig. 1) (Rivera and Bennett, 2010). Up to 80% of all the mammalian CpGs are methylated (Ziller et al., 2013) in most cell types, except in primordial germ cells (Seisenberger et al., 2012) and in pre-implantation embryos (Smith et al., 2014) and approximately 20% of the autosomal CpGs participate in dynamic genomic regulation (Ziller et al., 2013).

The control of gene expression through DNA methylation primarily occurs in CpG islands, which by definition are at least 200 bp stretches of DNA with enriched C and G nucleotide content, usually greater than 50%, and an observed:expected ratio over 0.6 (Takai and Jones, 2002; Wang and Leung, 2004). CpG islands are usually located in the promoter areas upstream to the transcription start site and contain transcription factor binding sites and other control sequences such as enhancers (Ziller et al. 2013). DNA methylation regulates the expression of these genes most frequently by suppressing the transcription of a gene when the protein product of the given gene is or is not needed (Portela and Esteller, 2010). In some cases DNA methylation is coupled with the transcriptional activation when it occurs at gene bodies (Hellman and Chess, 2007).

More than 50% of the human genes contain CpG islands at their promoter regions (Vavouri and Lehner, 2012). Promoter CpG island methylation controls the DNA transcription of a given gene typically by blocking the transcription. Initiation of a gene transcription requires the binding of specific transcription factors to DNA (Rivera and Bennett, 2010). The attachment of methyl groups
to cytosines induces conformational changes in DNA and chromatin structure. As the chromatin becomes more condensed the transcription factors can not bind to the DNA which results in the suppression of gene transcription (Severin et al., 2011; Paska and Hudler, 2015). Tumor suppressor genes are essential for normal cell functioning (Sun and Yang, 2010). The protein products of these genes are needed continuously in the cell to maintain homeostasis of the tissue and promoters of these genes are usually not methylated. Aberrant CpG island hypermethylation in these genes is related to cancer as well as aging and other defects (Portela and Esteller, 2010).

DNA methylation is also observed to regulate gene expression by either inhibiting or facilitating DNA strand separation during gene transcription depending on the sequence and the level of methylation (Severin et al., 2011). In normal cells DNA methylation is an essential control mechanism, for example, during embryonic development and cell differentiation (Plongthongkum et al., 2014). Tissue-specific differential DNA methylation occurs mostly at the non-promoter CpG island shores which are located upstream (~2000 bp) of the promoter CpG islands and have a relatively low CG content (Irizarry et al., 2009).

DNA methyltransferases (DNMTs) are the enzymes responsible for DNA methylation (Wu and Zhang, 2014). DNA methyltransferases transfer a methyl group from S-adenosyl-L-methionine (SAM) to CpG cytosines (Fig. 1) in the DNA methylation pathway (Niculescu and Zeisel, 2002; Ulrey et al., 2005). DNMT3A and DNMT3B are de novo methyltransferases responsible for establishing new methylation patterns in CpG regions that need to be transcriptionally repressed (Okano et al., 1998; Okano et al., 1999). DNMT1 is a maintenance methyltransferase. DNMT1 methylates mainly hemimethylated DNA during DNA replication (Fig. 1) to ensure the methylation pattern established during the differentiation of a cell is retained (Hermann et al., 2004). To an extent, all DNMTs can also act as de novo and maintenance methyltransferases (Svedružić, 2011).

DNA methylation is a reversible event. Compelling evidence suggests that TET proteins demethylate cytosines by catalyzing the conversion of 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC) by iterative oxidation (Guo et al., 2011). TET proteins can also convert 5-hydroxymethylcytosine to 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) (Ito et al., 2011). Restoration of unmodified cytosines takes place by either replication-dependent dilution or DNA glycosylase-initiated base excision repair (Cortellino et al., 2011).
1.2 Post-translational histone modifications and chromatin remodeling

Chromatin is wrapped around the histone core complex which is an octamer of histone proteins (H2A, H2B, H3 and H4) forming a nucleosome that is the basic DNA packaging unit in eukaryotes (Luger et al., 1997). Histone modifications actively regulate gene expression. Histones undergo multiple covalent post-translational modifications in their N-terminal tails that modulate the nucleosome structure and function leading either to activation or repression of transcription depending on the amino acids involved and the number and the type of the modifications (Sharma et al., 2010; Rossetto et al., 2012). These modifications include lysine acetylation, lysine and
arginine methylation, serine and threonine phosphorylation, lysine ubiquitination and sumoylation and ADP-ribosylation (Strahl and Allis, 2000; Zhang and Reinberg, 2001).

Histone methylation and demethylation correlates notably with functional genomics by the means of suppressed or active promoters, enhancers and gene bodies (Ernst et al., 2011). Histone tails get methylated in different lysine or arginine residues which associate with their distinct biological functions. For example lysine 27 trimethylation on histone 3 (H3K27me3) functions as a repressive mark that leads to chromatin condensation and reduces the transcription of affected genes (Ferrari et al., 2014). H3K27 is trimethylated by histone methyltransferase EZH2 which is a member of the Polycomb repressive complex 2 (PRC2) (Deb et al., 2015). KDM1A is an example of a histone demethylase (Shi et al., 2004) and it functions as a member of the RE1-silencing transcription factor (REST) repressor complex, with the main function of repressing neuronal gene expression in non-neuronal cells (Ballas et al., 2005; Arnold et al., 2013). Removal of active histone marks such as H3K4me3 (Voigt et al., 2013) or various histone tail acetylations by the REST complex induces the recruitment of additional silencing machineries to ensure the suppression (Arnold et al., 2013; Zhou et al., 2013).

While histone methylation is associated with either repressive or activating functions depending on the modified amino acid residue and the level of acquired methylation (Ferrari et al., 2014), histone acetylation usually has an activating effect on chromatin. Histone acetyltransferases (HATs) and deacetylases (HDACs) modify histone acetylation patterns (Verdone et al., 2005; Lopez et al., 2015). Acetylation affects the stability of a nucleosome by transforming the chromatin into a more relaxed state making space for transcription factors to bind the DNA. Acetylation also creates docking sites for the binding regulatory proteins facilitating gene transcription (Verdone et al., 2005).

Histone variant exchange adds more complexity to this epigenetic gene regulation system. Each conventional histone has non-allelic variants differing by a few amino acids (Talbert and Henikoff, 2010). These are expressed at low levels and have a specific localization, expression and structural and functional properties (Kamakaka and Biggins, 2005; Santoro and Dulac, 2015).

1.3 Noncoding RNAs and RNA interference

Noncoding RNAs (ncRNA) are a variable class of RNA molecules that do not encode any protein. They control gene expression in higher eukaryotes and affect many developmental and physiological processes including disease (Holoch and Moazed, 2015). NcRNAs include e.g.
microRNAs (miRNA), small interfering RNAs (siRNA) and long noncoding RNAs (lncRNA). MiRNAs and siRNAs control the silencing of gene expression through mRNA degradation or through direct transcriptional or translational repression. LncRNAs regulate gene expression through interference with RNA polymerases or by targeting chromatin-modifying enzymes (Geisler and Coller, 2013).

Small ncRNAs that are 21-25 nt long are designated as miRNAs. They regulate temporal and tissue-specific gene regulation and development through RNA interference (RNAi) in several organisms including humans (Storz, 2002). RNAi is an epigenetic control mechanism by which RNA molecules inhibit gene expression (Fire et al., 1998). In this process, RNAs typically cause the destruction of specific mRNAs (Khanmi et al., 2015). MiRNAs are a large class of gene products that are transcribed as 60-70 nt hairpin RNA precursors and typically excised to functional length by Drosha (Lee et al., 2006) and Dicer (endoribonucleases that recognize target mRNAs via base-pairing interactions) (Cai et al., 2004) and Argonaute family members (Calin et al., 2002). SiRNAs and miRNAs also silence genes at the transcriptional level. Promoter directed siRNAs induce transcriptional gene silencing in humans (Morris et al., 2004; Castanotto et al., 2005). This silencing is associated with DNA methylation of the targeted sequence.

Long noncoding RNAs can be several tens of kb long and they are involved in gene silencing by different mechanisms (Geisler and Coller, 2013). Transcription of an lncRNA across promoter region of downstream of a gene may also interfere with transcription factor binding and repress the expression of the given gene (Martens et al., 2004). LncRNAs such as HOTAIR and XIST are known to interact with PRC2 (e.g. EZH2) (Zhao et al., 2008) and Trithorax group proteins (Sanchez-Elsner et al., 2006) by recruiting them to the target site to repress or activate the transcription of the target gene (Rinn et al., 2007).

1.4 Interplay between different epigenetic factors

Different epigenetics pathways are intertwined and this crosstalk between different pathways is significant (Choi and Lee, 2013). The entire epigenetic machinery co-operates together to ensure that an appropriate chromatin state and accessibility is achieved and maintained so that normal levels of gene expression is possible (Sandoval and Esteller, 2012).

Components of the epigenetic machineries themselves are regulated by other epigenetic factors such as miRNAs. These epi-miRNAs target for example DNMTs, HDACs and Polycomb genes.
MiRNAs can be regulated by CpG island methylation and changes in histone modifications (Valeri et al., 2009).

Genes from the X chromosome are only expressed from one parental copy in an individual. Males have only one X chromosome but in females the expression from only one of the two chromosomes is achieved by a mechanism called dosage compensation (Lyon, 1961). Dosage compensation in mammals is interplay between noncoding RNAs and chromatin leading to changes in chromatin structure and the repression of the other X chromosome in females (Heard, 2004).

DNA methylation is regulated by chromatin proteins and DNA methylation affects the interaction of chromatin proteins and DNA (Hoffmann et al., 2007). DNA methylation and chromatin repressive proteins act together in gene silencing and for example EZH2 possibly serves a recruitment platform for DNMTs (Viré et al., 2006).

1.5 Epigenetic inheritance

The term epigenetic inheritance is used in two meanings, the inheritance of the epigenetic pattern of a mother cell to the daughter cells during cell division (mitotic inheritance) or the inheritance of epigenetic marks from parent to offspring in different generations (transgenerational inheritance).

All different cell types have their unique epigenetic patterns. The epigenetic identity of a cell must be transferred from mother to daughter cells through cell division to maintain cellular identity (Goldberg et al., 2007). DNA methylation patterns are known to be copied to the newly synthesized DNA strand by DNMT1 DNA methyltransferase using the methylated parent strand as a template (Fig. 1) (Leonhardt et al., 1992; Smith et al., 1992; Long et al., 2013). Watson-Crick base-pairing dictates that C pairs with G in DNA double-strand, so CpG sequences align and both strands are methylated. Similarly to DNA methylation patterns, histone modification marks have to be re-established on the newly synthesized histones during replication. Histone chaperone proteins coordinate the assembly of the chromatin alongside with the histone modifying enzymes which seem to be the epigenetic factors that remain associated with the DNA through replication rather than the parental histones (Budhavarpu et al., 2013).

Conventionally it has been thought that epigenetic patterns are erased twice during the cell life cycle; firstly in embryonic cells in order to achieve pluripotency after fertilization and secondly in
primordial germ cells (Seisenberger et al., 2013). Pluripotent cells can give rise to all cell types in the bodies of developing embryos through epigenetic reprogramming (Lee et al., 2014). Now however, it has become evident that epigenetic marks can be inherited from parent to offspring at least at some scale. Two recent papers indicate that there is a massive loss of DNA methylation after fertilization in human embryos, but not all of the methylation marks are erased. In fact, demethylation of the paternal genome was found to be more remarkable than demethylation of maternal genome and conserved imprinted regions retained the methylation throughout the development (Guo et al., 2014; Smith et al., 2014). There is also evidence that some epigenetic modifications, hypermethylation of certain tumor suppressor genes, can also be inherited in germline from parent to offspring (Chan et al., 2006; Hitchins et al., 2007). The mechanisms of epigenetic inheritance are diverse and most of them are not well described at the present.

2 Basic characteristics of cancer

Cancer is among the leading causes of death being responsible for about 15% of all deaths worldwide. In 2012, 32.6 million people were living with cancer (WHO, World Cancer Report 2014; WHO, Fact Sheet 2015). The most commonly diagnosed cancers include lung, prostate, colon, stomach and liver cancer in men, and breast, colon, lung, cervix and stomach cancers among women (WHO, World Cancer Report 2014). Cancer incidence rates are not globally uniform and cancer is more common in developed than in developing countries. Cancers are divided into different types e.g. depending on the tissue of origin. Most common cancers are carcinomas transforming from epithelial tissue and sarcomas from mesenchymal origin. More than 100 different types of cancer in humans have been described (Hanahan and Weinberg, 2000).

Cancer is a genetic disease which requires alterations in the genome (Hanahan and Weinberg, 2000) and the causes of genetic defects leading to cancer initiation and development vary. Most cancer cases are due to behavioral and lifestyle-related factors, such as high body mass index, low vegetable and fruit intake, lack of physical activity, and tobacco or alcohol use (WHO, World Cancer Report 2014). Tobacco use is the single most important risk factor accounting for 20% of cancer deaths globally. In low- and middle income countries 20% of cancer deaths are due to viral infections e.g. hepatitis B or C and human papilloma viruses (de Martel et al., 2012). Cancer causing mutations are known to accumulate over the time (Hanahan and Weinberg, 2011). Thus aging increases the risk of developing cancer especially
in western countries where the life expectancy is high (WHO, World Cancer Report 2014; Finnish Cancer Registry; 2015).

Most cancers, approximately 90-95%, develop and progress as sporadic cancers with two major driving events being accumulation of different mutational events or other errors leading to genomic instability and copy number abnormalities in individuals (Ciriello et al., 2013; Zack et al., 2013). These cancers do not elevate the risk of developing a cancer in the relatives. However, in 5-10% of cancers there is a genetic hereditable component which multiplies the risk of developing cancer in family members (American Cancer Society, 2014).

In the Finnish population, 30% of males and 25% of females are diagnosed with a cancer by the age of 75 (Cancer Society of Finland, 2015; for the most prevalent cancers in Finland, see Table 1). The current 5-year relative survival rate for cancer patients is 69% in women and 66% in men in Finland (2013) and during the most recent follow-up period (2010-2012) the survival rates have improved by 4% (Finnish Cancer Registry, 2015). The prognosis, treatment and survival of the cancer patients however depend drastically on the affected tissue and the classification (e.g. grade and stage) of the tumor (Table 2). For example in Finland (2012) over 90% of the breast and prostate cancer patients were still alive after 5 years of the diagnosis, whereas the relative 5 year survival rate for pancreatic cancer was only approximately 5% (Cancer Society of Finland, 2015).

<table>
<thead>
<tr>
<th>Primary cancer</th>
<th>Male</th>
<th>Incidence* new cases/year</th>
<th>Primary cancer</th>
<th>Female</th>
<th>Incidence* new cases/year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prostate</td>
<td>1.</td>
<td>87.2</td>
<td>Breast</td>
<td>1.</td>
<td>92.6</td>
</tr>
<tr>
<td>Lung and trachea</td>
<td>2.</td>
<td>29.3</td>
<td>Colon</td>
<td>2.</td>
<td>14.4</td>
</tr>
<tr>
<td>Bladder and urinary tract</td>
<td>3.</td>
<td>15.6</td>
<td>Skin (non-melanoma)</td>
<td>3.</td>
<td>8.9</td>
</tr>
<tr>
<td>Colon</td>
<td>4.</td>
<td>15.7</td>
<td>Lung and trachea</td>
<td>4.</td>
<td>12.6</td>
</tr>
<tr>
<td>Skin (non-melanoma)</td>
<td>5.</td>
<td>13.3</td>
<td>Endometrium</td>
<td>5.</td>
<td>13.8</td>
</tr>
<tr>
<td>Gastric</td>
<td>12.</td>
<td>6.3</td>
<td>Gastric</td>
<td>15.</td>
<td>3.6</td>
</tr>
</tbody>
</table>

*Adjusted for age to the world standard population, 1/100 000 people (Finnish Cancer Registry, 2015, www.cancer.fi/syoparekisteri/en), bold; tumor tissues included in this research.
Table 2. Simplified classification of solid epithelial tumors: basis of grading and staging.

<table>
<thead>
<tr>
<th>Grade of differentiation</th>
<th>Developmental stage</th>
<th>Metastases on lymph nodes</th>
<th>Metastases</th>
</tr>
</thead>
<tbody>
<tr>
<td>well differentiated</td>
<td>Cancer found on very early stages. Tumors visible only on inner mucosa</td>
<td>no tumor cells in lymph nodes</td>
<td>cannot be evaluated</td>
</tr>
<tr>
<td>moderately differentiated</td>
<td>Tumor infiltrating muscle layer of the mucosa</td>
<td>regional metastasis present</td>
<td>no</td>
</tr>
<tr>
<td>poorly differentiated</td>
<td>Tumor penetrating outside the organ but not to surrounding tissues</td>
<td>regional metastasis present in several lymph nodes</td>
<td>yes</td>
</tr>
<tr>
<td>undifferentiated</td>
<td>Tumor penetrating surrounding tissues but not to rest of the body</td>
<td>tumor cells found more distant or numerous regional lymph nodes</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cancer metastasized to other organs OR primary tumor not completely removable</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The classification of a specific tumor depends on the tissue, and there are more detailed guidelines for each tissue type. Most systems stratify tumors into three to four grades (Compton, 1999) and the basis is similar for solid tissues regarding the developmental or metastatic stage (based on Dukes classification for colorectal cancer and TNM classification for gastric and endometrial cancers; Cancer Society of Finland, 2015). The table shows the main basis of the solid tumor classification. The more well differentiated the tumor is, the more it resembles normal healthy cells, grows more slowly than poorly differentiated tumors and is more unlikely to be invasive or send metastasis (Compton, 1999; Cancer Society of Finland, 2015; Diaz-Cano, 2015).

In the light of new research, the diagnostics and treatment methods of cancer are continuously improving, and it is reflected on the improving prognosis and survival of the cancer patients. Primary treatment in many types of solid tumors is surgery. Other treatments include radiation therapy and chemotherapeutics including cytotoxic drugs, hormones or interferons as well as palliative care all of which can be used in different combinations (Cancer Society of Finland, 2015). New knowledge of different cancer types provides the possibility of more personalized treatment depending on the molecular basis of that specific tumor.

Multiple acquired aberrations are required for cells to become malignant. Cancer cells have characteristic abnormal abilities allowing them to grow and transform into malignant tumors. These include sustaining proliferative signaling, evading growth suppressors, activating invasion and metastasis, immortal replicative potential, sustaining angiogenesis and resisting apoptotic signals (Hanahan and Weinberg, 2000; Hanahan and Weinberg, 2011). In addition to these, cancer cells are able to reprogram energy metabolism in the cells and can evade immune destruction (Hanahan and Weinberg, 2011). These capabilities evolve due to genomic instability, mutations and tumor-promoting inflammation. Cancer cells are tricky little bastards since they possess the ability to make their neighboring normal cells create a tumor microenvironment where the tumor cells can thrive (Hanahan and Weinberg, 2011).
Cancer genetics

Cancer is considered a genetic disease. The transformation of a normal cell into a tumor cell is a multistep process, from pre-cancerous lesions to malignant tumors and it is the result of different aberrations in genetic and epigenetic processes (Hanahan and Weinberg, 2000; Hanahan and Weinberg, 2011). Cancer arises from concurrent or sequential accumulation of mutations in different oncogenes and tumor suppressor genes. These genes are essential for normal cell functions: they code for proteins that help to regulate cell growth, proliferation and differentiation or cell death, proto-oncogenes by activating and tumor suppressor genes by limiting growth. Mutations can be classified as driver mutations which are responsible for cancer growth and metastasis and passenger mutations which do not affect the growth (Vogelstein et al., 2013; Marx, 2014). Tomasetti et al. (2015) have calculated that as few as three mutations in any of the approximately 140 known driver genes can be enough for cancer development (Vogelstein et al., 2013). Besides mutations in oncogenes and tumor suppressor genes the changes involving genomic instability (chromosomal or microsatellite instability) have been characterized in multiple cancers (Hanahan and Weinberg, 2011). In addition to genetic changes, simultaneously occurring epigenetic changes (Paska and Hudler, 2015) facilitate the aberrant expression of tumor suppressor genes and oncogenes and contribute to the genomic instability of cancer cells. These are more discussed in later chapters.

The progress of a normal cell into a cancer cell is slow but accelerates when multiple different mutations and epigenetic aberrations cluster (Tomasetti et al., 2015). Tumorigenesis is also faster if a critical cancer predisposing defect is inherited in birth and is thus in every cell of an individual. During tumorigenesis, the function of one or more of the DNA repairing mechanisms can be lost allowing more alterations to occur (Hanahan and Weinberg, 2011).

3.1 Tumor suppressor genes

Tumor suppressor genes (TSGs) control gene differentiation and growth by limiting e.g. proliferation, motility or invasion (Sun and Yang, 2010). TSGs function e.g. as intra- or intercellular signaling molecules, DNA repair proteins, checkpoint proteins or on the apoptotic cascade. TSGs regulate the cell cycle by arresting it for example in the case of DNA damage (Macleod, 2000). TSGs also direct cells to apoptotic pathways when the damage cannot be repaired and the cell needs to be eliminated (Sancar et al., 2004).
Tumor suppressor genes are often silenced in tumors, and TSG inactivation is a frequent component of many hereditary cancer syndromes. TSG inactivation occurs usually by deletions, mutations or promoter methylation (Boland and Goel, 2010). According to Knudson’s two-hit theory, both alleles of TSGs have to be inactivated in order to cause aberrant gene function (Knudson, 1971). This theory explains the relationship between hereditary and sporadic cancers. In their study Knudson explained how an individual will develop retinoblastoma if they either inherit one mutated RB1 gene allele from the parent and one allele copy is lost by random somatic mutation, or if they randomly acquire mutations separately in both alleles. According to this theory, only one intact allele is enough to maintain its function while the faulty allele does not induce cancer formation. The recessive nature is the functional basis of many of the TSGs (Fig. 2). However, mutations in tumor suppressor genes can also be dominant-negative preventing the function of the wild-type allele (Goh et al., 2011), or they can involve a gain-of-function or be haploinsufficient when the product of only one working allele is insufficient to fulfil the functionality needed in a cell (Ramdzan and Nepveu, 2014); in these cases only one hit may be sufficient for TSG inactivation.

Tumor suppressor genes are divided into gatekeepers, caretakers and landcsapers (Kinzler and Vogelstein, 1997; Kinzler and Vogelstein, 1998). Gatekeepers directly inhibit cell proliferation and tumor growth and promote cell death (e.g. APC gene associated with familial adenomatous polyposis). In each cell type there is only a few gatekeepers and individuals with a pathogenic hereditary mutation in these genes have a significantly higher risk (>10^3) of developing tumors than the rest of the population. The inactivation of caretaker genes leads to the disruption of the genomic integrity causing genetic instability through increased mutation rates in other genes which further promote neoplasia formation (e.g. MMR genes predisposing to Lynch syndrome; see chapter 3.3.2.1). In these cases the familial risk is usually 5–50-fold greater than in general population (Kinzler and Vogelstein, 1997). Landscaper defects change the microenvironment which facilitates tumorigenesis (e.g. SMAD4 and PTEN in juvenile polyposis syndrome or ulcerative colitis, where normal epithelial cells are at increased risk of neoplasia formation) (Kinzler and Vogelstein, 1998).
Figure 2. The two-hit theory of cancer causation. A) Normal cells have two undamaged chromosomes one inherited from each parent. The loss of function of tumor suppressor genes requires two ‘hits’ i.e. both alleles need to be inactivated. Individuals with hereditary susceptibility to cancer have inherited one mutated allele from either of the parents, so the first ‘hit’ is already present in each cell of that individual. In sporadic cases the first ‘hit’ is acquired by chance. In both cases, if cell acquires mutation to the remaining wild-type chromosome, the second ‘hit’, the cell can become tumorigenic: B) Here maternal (M) chromosome (pink) harbor one mutated allele of a gene (blue star). Paternal allele (green) is intact and the site is heterozygous. Silencing of paternal (P) allele can be obtained by multiple ways; 1-3) loss of heterozygosity by whole chromosome loss in mitosis, through mitotic recombination or gene conversion through which the mutated allele replaces the wild-type allele or by deletion. Loss of gene function can also be acquired by 4) somatic mutation (yellow star) or 5) by epigenetic silencing (red star) (Based on Knudson, 1971; Aittomäki and Peltomäki, 2006).

3.2 Oncogenes

Proto-oncogenes are genes that promote cell growth, proliferation and inhibit differentiation and cell death by encoding for example growth factors or growth factor receptors, cell cycle regulators or signaling molecules and are essential for regulating cell proliferation or survival and the maintenance of tissues in normal cells (Boland and Goel, 2010). Proto-oncogenes become cancer promoting oncogenes when they are inappropriately activated (Adamson, 1987; Weinstein and Joe, 2006). Proto-oncogenes can become oncogenes by point mutations resulting in hyperactive protein or promoter function leading to increased transcription (Croce, 2008). Oncogenes can also
be overexpressed due to gene or promoter region amplifications or through genomic rearrangements such as translocations. These changes result in copy-number changes and excess of the protein, fusion proteins with oncogenic activity, or through epigenetic mechanisms affecting promoter function. Proto-oncogenes are dominant in nature but are only rarely responsible for hereditary cancers (Hodgson, 2008).

3.3 Genomic instability

Genomic instability is an integral component of human neoplasia (Lengauer et al., 1997). All somatic cells in an individual contain a fixed number of chromosomes which harbor the genes. A cell has machineries to maintain its genomic integrity, the euploidy of the chromosomes and the nucleotide sequences in the DNA throughout cell divisions. Different factors can contribute to the loss of genomic integrity. Alterations in the genome may be due to external environmental or chemical stress, radiation, diet, reactive oxygen species (Sancar et al., 2004), defects in DNA repair machineries or errors during mitosis (chromosome segregation).

High fidelity DNA synthesis and repair is necessary to maintain genetic information from generation to generation and to avoid mutations that cause cancer and other diseases (Kunkel, 2004). The cell itself can also make mistakes in reading and copying DNA. The average baseline mutation rate in the normal somatic cell cycle is $10^{-9}$ mutations per nucleotide base pairs, per cellular generation (Albertini et al., 1990). The intrinsic spontaneous mutation rate is insufficient to account for the all the mutations required for tumorigenesis so cancer cells have to acquire genomic instability in order to increase the rate of new mutations (Loeb et al., 2003).

DNA repair mechanisms correct different aberrations in the genome of normal cells. DNA damage activates DNA repair complexes which recognize and eliminate the damage while DNA damage checkpoints arrest the cell cycle progression until the damage is repaired (Sancar et al., 2004). If the repair machineries do not detect the errors in the DNA the errors are copied during the DNA replication and passed to daughter cells during cell division.

Different kinds of DNA repair mechanisms repair different kinds of lesions or damage in DNA. Double-strand breaks are usually repaired by double-strand break repair such as homologous recombination or non-homologous end-joining mechanisms. Direct repair of replication errors is conducted by mismatch repair. Nucleotide excision repair and base excision repair including single-strand break repair act by removing damaged bases and $O^6$-methylguanine-DNA
methyltransferase (MGMT) removes O⁶-alkylation adducts and restores the guanine to its normal state (Sancar et al., 2004; Iyama and Wilson, 2013).

There are at least two different types of genomic instability: chromosomal instability (CIN) and microsatellite instability (MSI). Chromosomal instability is a dominant trait while microsatellite instability is recessive (both discussed more in chapters 3.3.1 and 3.3.2; Casares et al., 1995; Lengauer et al., 1997). It has been proposed (Stephens et al., 2011) that there is an additional mechanism affecting genome stability. The phenomenon called chromothripsis was first observed in chronic lymphocytic leukemia. It distorts the euploidic state of the cell by a massive catastrophic event in which a whole chromosome or a chromosome arm is shattered into pieces almost simultaneously and is reassembled randomly together by DNA repair mechanisms, creating deletions of some segments and complex rearrangements of the others. Chromothripsis has been seen to be present in 2-3% (Zhang et al., 2015a) of all cancers and it causes oncogene amplifications, tumor suppressor gene deletions and the heterogeneous loss of heterozygosity (Maher and Wilson, 2012). Chromothripsis has been found to be involved in colorectal as well as in other cancers (Forment et al., 2012; Kim et al., 2015). This model opposes the conventional theory of cancer progression through the accumulation of somatic mutations over a long period of time. There are differing opinions about the mechanism of chromothripsis, for example Sorzano et al. (2013) proposed that chromothripsis could result from repetitive breakage-fusion-bridges rather than from a single massive chromosome breaking event.

### 3.3.1 Chromosomal instability and loss of heterozygosity

Chromosomal instability (CIN) is a state of genomic instability where chromosomes are unstable. Either one or more chromosomes can be entirely or partially deleted or duplicated causing aneuploidy i.e. widespread imbalances in chromosome number in the cell (Lengauer et al., 1997; Pino and Chung, 2010). Aneuploidy can be caused by the unequal distribution of the chromosomes in the nucleus due to defects in chromosome segregation during mitosis. The CIN phenotype in tumors correlates with poor prognosis, metastatic potential and drug resistance (Thompson and Compton, 2011).

CIN is observed in most solid tumors. Aneuploidy can present as the loss or gain of a whole chromosome due to errors during mitosis in excess of 10² per chromosome per generation in tumors without microsatellite instability (Lengauer et al., 1997; Geigl et al., 2008). Partial aneuploidy arises from double-strand breaks in DNA which causes deletions or amplifications in
parts of the chromosome, or chromosomal rearrangements causing only a part of a chromosome to be inverted or translocated to another place in the same or different chromosome (Geigl et al., 2008). Aneuploidy is a common feature in tumor cells, but all tumors with aneuploidy do not display CIN. The distinguishing feature between CIN positive (CIN+) and negative (CIN-) tumors is that the CIN phenotype in tumors causes a wide variety of different chromosomal alterations, whereas aneuploidy without CIN causes more clonal aberrations (Bakhoum and Compton, 2012).

CIN can arise from defects in mitotic checkpoint (spindle assembly checkpoint) signaling (Pino and Chung, 2010). This results in no delays in the cell cycle before the onset of anaphase and hence the duplicated chromatids may not be properly aligned on the metaphase plate before their division. CIN-suppressor genes (e.g. PIGN, MEX3C and ZNF516 all located on chromosome 18q) may be deleted leading to the silencing of these genes which causes DNA replication stress, structural chromosome abnormalities and defects on chromatin segregation (Burrell et al., 2013). Missegregation can also arise through specific kinetochore-microtubule attachment errors (Thompson and Compton, 2011).

CIN can also be driven by telomere dysfunction or inactivating mutations in DNA damage response genes responsible for cell cycle arrest such as ATM, ATR, BRCA1/2, TP53 or MRE11 (O’Hagan et al., 2002; Pino and Chung, 2010). Without telomere end protection, chromosome ends enter breakage-fusion-bridge cycles which can lead to genome reorganization over multiple cell generations. In CIN tumors, a specific set of tumor suppressor genes and oncogenes critical for tumorigenesis are mutated. Whether these mutations drive CIN or alternatively, CIN drives the accumulation of these mutations is not clear.

CIN is characterized by and can be detected as loss of heterozygosity (LOH) (Pino and Chung, 2010). LOH means that a heterozygous locus, a whole chromosome or parts of it, is lost and the genetic material is present in only one copy. LOH can result from deletions, mitotic recombination errors or gene conversion events (Fig. 2). LOH can cause TSG inactivation but a silencing somatic alteration of the remaining allele is still required for tumorigenesis. LOH can be studied by comparing normal and affected tissue of the same individual for example by fragment analysis (Aittomäki and Peltomäki, 2006).

### 3.3.2 Microsatellite instability

Microsatellites are simple short repetitive nucleotide sequences, usually mono- or dinucleotide repeats that are abundant throughout the genome (Ellegren, 2004). Microsatellite sequences are
polymorphic in the population, but unique and uniform in each individual and hence they have been deployed in allele discrimination analyses, gene mapping as well as in forensics (Sharma et al., 2007). Microsatellite instability (MSI) is a hypermutable phenotype that is due to MMR deficiency (Boland and Goel, 2010). MSI is observable at the nucleotide level as deletions or insertions of a few nucleotides at repetitive sequences (Peltomäki, 2001).

MSI was described in 1993 when separate groups of scientists studied colorectal tumors (Ionov et al., 1993; Thibodeau et al., 1993) and was first thought to be characteristic of certain types of hereditary cancer syndromes. MSI was indeed the first marker to help identifying hereditary colon cancers (Boland and Goel, 2010).

A subset of tumors displays microsatellite instability. Today it is known that about 15% of all colorectal cancers display the MSI phenotype (Xiao and Freeman, 2015), of which approximately 3% are associated with Lynch syndrome, whereas the other 12% are of sporadic origin, most likely due to MLH1 promoter hypermethylation repressing MLH1 expression in the target tissues (Boland and Goel, 2010). Compared to cancers without MSI, colorectal cancers with MSI are more likely to arise in the proximal colon, present in younger patients, are poorly differentiated (Ionov et al., 1993; Thibodeau et al., 1993) and have better prognosis particularly in stage II and III tumors (Benatti et al., 2005). Hereditary and sporadic colorectal MSI cancers evolve through a similar pathway for developing cancer without the loss of heterozygosity (Aaltonen et al., 1993; Thibodeau et al., 1993).

Microsatellite instability is conventionally studied with the Bethesda panel of 5 mono- or dinucleotide markers, BAT25, BAT26, D2S123, D5S346 and D17S250 (Boland et al., 1998; Umar et al., 2004a). The MSI-high phenotype is defined by two or more unstable markers.

### 3.3.2.1 Mismatch repair (MMR) pathway

The DNA MMR pathway in humans contains a specific repair machinery for the repair of base-base mismatches or insertion-deletion loops (IDLs) caused by DNA replication errors acquired during the S-phase of the cell cycle (Iyama and Wilson, 2013). Mispairing of bases can also arise during recombination or DNA damage. The MMR machinery consists of a family of enzymes (MLH1, MLH3, MSH2, MSH3, MSH6 and PMS2) with the capability to recognize and repair these mismatches (Jiricny and Nyström-Lahti, 2000; Modrich, 2006).
During DNA replication DNA polymerase can make errors especially at the sites with long repetitive DNA sequences, such as microsatellites, which results in one or more misincorporated or missing nucleotides in the newly synthesized strand (Fig. 3; Jiricny, 2006). In normal cells the rate of single base substitution errors during DNA synthesis is in the range of $10^{-6}$ to $10^{-8}$ for replicative polymerases with intrinsic proofreading and exonuclease activities (Peltomäki, 2001; Kunkel, 2004).

Generally DNA polymerases proofread and correct errors during the DNA synthesis, but sometimes the newly synthesized DNA strand escapes the intrinsic polymerase proofreading. This is when the DNA MMR machinery is needed: MutSα complex (a protein duplex composed of MSH2 and MSH6) of the MMR machinery detects base-base mismatches and short polymerase slippage induced IDLs whereas MutSβ (MSH2-MSH3) recognizes larger IDLs based on the structural change in DNA caused by the mismatch (Fig. 3) (Jiricny and Nyström-Lahti, 2000; Boland and Goel 2010). Binding of the MutS complex to the DNA recruits MutLα (MLH1/PMS2) or MutLγ (MLH1/MLH3) heterodimer which interacts with the replication sliding clamp proliferating cell nuclear antigen (PCNA) and the DNA polymerase complex. PCNA directs the endonuclease activity of MutLα and DNA polymerase which removes the wrong nucleotide and attaches the correct one (Iyama and Wilson, 2013). If the mismatch stays unrepaired, the single base mismatches become point mutations and the IDLs results in frame-shift mutations leading to a premature stop codon and a truncated protein in the next cell generations.

### 3.4 Colorectal cancer

Colorectal cancer (CRC) is the fourth common cause of cancer related deaths in Finland among males and females with over 2000 patients diagnosed with CRC in Finland every year (Finnish Cancer Registry, 2015). Worldwide CRC is the third most common cancer and a leading cause of cancer-related death causing over 690.000 deaths yearly (WHO, World Cancer Report 2014).

The risk of developing colorectal carcinoma increases with age. Other risk factors include a western style low-fiber high-fat diet, smoking, high alcohol consumption, diabetes or a familial background of colorectal cancer (Weitz et al., 2005) The typical age of onset of sporadic CRCs is around 75 years, but with familial predisposition CRC appear at much earlier age (e.g. around 45 years in Lynch syndrome) (Lynch and de la Chapelle, 1999). Prognosis depends on the Dukes classification status of the tumor. The 5-year survival rate associated with Dukes A and B tumors is approximately 70%, Dukes C 52% and Dukes D 12% in Finnish population (Allemani et al., 2013).
Figure 3. DNA mismatch repair. The figure shows the MMR repair machinery which recognizes one base mismatches as well as insertion or deletion loops acquired during DNA replication. Lower strand in A) is the parental strand and the upper is the newly synthesized daughter strand. MutSα (MSH2/MSH6) recognizes single base substitutions or insertion-deletion loops (IDLs) in DNA. MutSα interacts with MutLα (MLH1/PMS2) and recruits the DNA polymerase complex and replication factors (not shown) to the mutation site. The DNA polymerase complex repairs the mismatch in the daughter strand and continues replication. IDLs can also be detected and corrected by MutSβ (MSH2/MSH3) complex with MutLα or MutLγ (MLH1/MLH3) complexes. Defective mismatch repair results in the accumulation of mutations B) in following cell generations and contributes to genomic instability (based on Jiricny and Nyström-Lahti, 2000; Jiricny, 2006).

The majority of CRCs are sporadic but the CRC predisposition can also be inherited. About 25% of CRC cases occur in individuals with a family history of cancer. Approximately 5% of all colorectal cancers are hereditary with identified inherited mutations (Gala and Chung, 2011). These patients harbour highly penetrant mutations in the germline predisposing them to tumor development. Lynch syndrome diagnosed by germline mutations in MMR genes is the most prevalent CRC syndrome accounting for 2-4% of all colon cancer cases (Gala and Chung 2011). Other CRC syndromes include e.g. familial adenomatous polyposis (FAP), MUTYH-associated polyposis, Peutz-Jeghers polyposis, juvenile polyposis syndrome and Cowden disease (de la Chapelle, 2004; Jasperson et al., 2010).
CRCs develop through different pathways and can be classified as CIN+/aneuploid (CIN pathway) or CIN-/MSI (MSI pathway) (Fig. 4) (Lengauer et al., 1998). Most of the CRC pathogenesis develops through the CIN pathway (Pino and Chung, 2010). First and most common model for the multistep tumorigenesis in CRC (Morson, 1974; Fearon and Vogelstein, 1990) follows the CIN pathway and is characterized by the inactivation of the APC gene which is associated with adenoma formation through the activated Wnt-signaling. Subsequent mutations activating KRAS are associated with adenoma growth through the activated RAS downstream signaling, and deletions (or other alterations) of genes in chromosome 18q affect adenoma growth and progression. The biallelic loss or inactivation of TP53 results in the activation of the adenoma-carcinoma transition. These changes are accompanied with widespread chromosomal imbalances and LOH in CIN tumors (Fearon and Vogelstein, 1990; Powell et al., 1992; Kinzler and Vogelstein, 1996; Vogelstein et al., 2000; Takayama et al., 2001). Subsequent mutations in TGFβR and PIK3CA drive the cancer formation (Markowitz et al., 1995; Samuels and Velculescu, 2004). FAP is an example of a hereditary syndrome typically exhibiting these types of genetic changes.

15-20% of sporadic CRCs and most Lynch syndrome tumors arise through the MSI pathway (Fig. 4) in which MMR deficiency causes the microsatellite instability at nucleotide level. These tumors frequently have mutations in TGFβR2 and BAX genes and in BRAF in sporadic cases and a have near diploid DNA content (Lengauer et al., 1997; Jass, 2004). The MSI pathway in sporadic CRCs arises likely due to MLH1 hypermethylation (Lynch et al., 2007).

An important difference between CIN and MSI CRC tumors concerns their sensitivity to different chemotherapeutics. 5-fluorouracil (5-FU) has been commonly used for treating colorectal cancer in combination with oxaliplatin and/or leucovorin, irinotecan and capecitabine (Moertel et al., 1995; Saltz et al., 2000; Andre et al., 2004; Twelves et al., 2005). Patients with MSI colorectal cancer may not benefit from 5-FU-based chemotherapy (des Guetz et al., 2007).
CRC is used as a model of tumorigenesis for other epithelial cancers while hereditary cancers are models for their sporadic counterparts since the mechanisms and pathways are similar. It has become evident that CRCs are genetically and epigenetically heterogeneous with distinct subgroups. Moreover, individual tumors from a given subgroup may significantly differ from each other, and even a single tumor may consist of multiple clones (Marisa et al., 2013). Knowledge about the molecular background of each tumor is essential for the right kind of treatment.

3.4.1 Hereditary non-polyposis colorectal cancer syndrome (HNPCC)

Hereditary non-polyposis colorectal cancer (HNPCC) is one of the most common cancer syndromes (Lynch et al., 2015). HNPCC diagnosis is based on the clinical criteria for HNPCC which are the Amsterdam Criteria I (AC-1), Amsterdam criteria II and the revised Bethesda guidelines (Table 3) (Vasen et al., 1991, Rodriguez-Bigas et al., 1997, Vasen et al., 1999, Umar et al., 2004b).
Patients fulfilling the clinical criteria for HNPCC are subjected to studies for MMR gene defects and microsatellite instability in order to verify whether they fulfill the criteria for Lynch syndrome (LS) and genetic counseling is recommended to the relatives.

Previously the term HNPCC was used in place of Lynch syndrome (Lynch et al., 2015). The finding of MMR genes and their contribution to cancer development in HNPCC revealed two distinct categories of HNPCC; Lynch syndrome (LS) and colorectal cancer type X (FCCX) (both discussed below in more detail). In Lynch syndrome patients one or more MMR genes are mutated and they exhibit loss of corresponding functional MMR protein. FCCX patients meet the clinical criteria for HNPCC, but without evidenced loss of MMR gene function (Lindor, 2009; Shiovitz et al., 2014).

Table 3. Amsterdam criteria I & II and the revised Bethesda guidelines for HNPCC

<table>
<thead>
<tr>
<th>Amsterdam criteria I ¹</th>
<th>At least 3 relatives with histologically confirmed CRC, at least one first-degree relative of the other two</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>At least 2 successive generations involved</td>
</tr>
<tr>
<td></td>
<td>At least 1 of the cancers diagnosed before age of 50</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Amsterdam criteria II ²</th>
<th>3 or more relatives with an associated cancer (CRC, endometrial, small intestine, ureter or renal pelvis)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2 or more successive generations affected</td>
</tr>
<tr>
<td></td>
<td>1 or more relatives diagnosed before the age of 50</td>
</tr>
<tr>
<td></td>
<td>1 should be a first-degree relative of the other two</td>
</tr>
<tr>
<td></td>
<td>Familial adenomatous polyposis (FAP) should be excluded in cases of CRC</td>
</tr>
<tr>
<td></td>
<td>Tumors should be verified by pathologic examination</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>The revised Bethesda guidelines³</th>
<th>CRC diagnosed in a patient before the age of 50</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Presence of synchronous or metachronous CRC or other Lynch syndrome-related tumors, regardless of age</td>
</tr>
<tr>
<td></td>
<td>CRC with MSI-high histology in a patient less than 60 years old</td>
</tr>
<tr>
<td></td>
<td>CRC diagnosed in one or more first-degree relative with Lynch syndrome-associated tumor, one of these diagnosed before the age of 50</td>
</tr>
<tr>
<td></td>
<td>CRC diagnosed in two or more first- or second-degree relatives with Lynch syndrome-associated tumor, regardless of age</td>
</tr>
</tbody>
</table>

*For the identification of individuals who should be subjected for genetic testing for MSI and Lynch syndrome (¹Vasen et al., 1991, ²Vasen et al., 1999 (each of the criteria must be fulfilled), ³Rodriguez-Bigas et al., 1997, ⁴Umar et al., 2004b).

3.4.2 Lynch syndrome (LS)

Lynch syndrome (LS, previously known as HNPCC) is the most common hereditary colon cancer syndrome accounting for 2-4% of all colon cancer cases (Peltomäki, 2001; Gala and Chung, 2011). LS is characterized by deficient DNA mismatch repair, early onset of cancer (approximately 20 years earlier than sporadic cancers) and with increased risk of multiple cancers of different organs.
including endometrial, ovarian, gastric, small bowel, upper urologic tract and pancreatic cancer (Familial Cancer Database, 2015). LS tumors are most likely located in the proximal colon, colorectal neoplasms are non-polypoid and numerous synchronous and metachronous tumors are present. LS is identified by molecular screening of patients fulfilling the Amsterdam criteria or the revised Bethesda panel for HNPCC (Watson and Lynch, 1994; Lin et al., 1998; Watson and Riley, 2005; Gylling et al., 2007; Gala and Chung, 2011; Lynch et al., 2015).

Inherited autosomal dominant germline mutations in the MMR genes predispose to Lynch syndrome (Jacob and Praz, 2002; Worthley et al., 2005; Barnetson et al., 2006). MMR protein deficiency causes the lack of functional DNA mismatch repair and microsatellite instability (see chapter 3.3.2). Affected genes in Lynch syndrome are MLH1 (42%), MSH2 (33%), MSH6 (18%) and PMS2 (7.5%) (Plazzer et al., 2013) in which one altered allele is inherited from either of the parents. Also, the germline deletion of 3’ exons of EPCAM can result in Lynch syndrome by causing somatic methylation and inactivation of MSH2 (Ligtenberg et al., 2009). In the Finnish population 32 different mutations in MLH1, 9 in MSH2, 5 in MLH6 and 2 in PMS2 have been described (InSiGHT database, 2015). The Finnish mutations include three MLH1 founder mutations originating from a single Finnish individual: 3.5 kb deletion of exon 16 and flanking introns (Mut1), a single base substitution at a splice acceptor site in exon 6 (c.454-1g>a) (Mut2) and a c.320T>G substitution at exon 4 causing the amino acid change I107R (Mut3) (Moisio et al., 1996; Nyström-Lahti et al., 1996). The mutated gene is present in every cell of these individuals and loss of functional MMR occurs when the wild-type allele is inactivated by a somatic mutation or an epigenetic event. Basically this will happen eventually in some cells. MMR deficiency contributes to microsatellite instability which silences tumor suppressor genes during tumorigenesis. In addition, a small proportion of hereditary mutations occur in other MMR genes such as PMS1 and MLH3 (Peltomäki, 2005; Lynch et al., 2015). Predisposed individuals have an over 80% lifetime risk of developing CRC (Dinh et al., 2011) which is significantly higher compared to the 2% risk of the general population. For extra-colonic tumors, such as endometrial tumors the risk is up to 71% while the ovarian cancer risk is 10-15% for the Lynch syndrome mutation carriers compared to 1.5% or <1% of the general population, respectively (Dinh et al., 2011; Ketabi et al., 2011).

3.4.3 Familial colorectal cancer type X (FCCX)

FCCX is by definition a familial colorectal cancer syndrome characterized by patients that meet the Amsterdam criteria I (similar to Lynch syndrome) but without identified mutations or functional
loss in the MMR genes. These cancers are MMR-proficient (tested either by IHC or MSI) and the patients may have no increased risk for extra-colonic tumors (Lindor et al., 2005; Llor et al., 2005). Approximately 50% of the CRC cases meeting AC-1 are classified as FCCX (Lynch and de la Chapelle, 2003; Lindor, 2009). Compared to LS, the mean age of onset is higher (50-60 years, compared to 40-50 years of LS).

FCCX tumors tends to lack the typical characteristics of CRC tumors in general such as MSI, CIN, active nuclear β-catenin and mutations in TP53 (Abdel-Rahman et al., 2005). FCCX tumors are less likely to be located in the proximal colon (compared to LS, no difference compared to sporadic CRC). Compared to LS, FCCX tumors show histopathologically more heterogeneous architecture, and are more commonly poorly differentiated than LS or non-AC-1 tumors are (Mueller-Koch et al., 2005; Shiovitz et al., 2014).

FCCX pedigrees show an autosomal-dominant inheritance pattern but to date with mostly unknown genetic basis (Shiovitz et al., 2014). When compared to the epidemiological or clinicopathological data, there are no clear features distinguishing FCCX from sporadic tumors (Shiovitz et al., 2014). It is still questionable whether FCCX occurs by chance due to environmental factors or if certain mutations are the driving force for cancer development in these patients (Lindor et al., 2005; Lindor, 2009) Recent genome-wide linkage scan and next-generation sequencing analyses have discovered germline mutations in BMPR1A and RPS20 genes in some FCCX families (Nieminen et al., 2011; Nieminen et al., 2014).

3.5 Endometrial cancer

Endometrial carcinoma (EC) is the sixth most common cancer in women globally (WHO, World Cancer Report 2014) and the fifth most common in the Finnish population. Approximately 850 patients in Finland are diagnosed with EC every year (Finnish Cancer Registry, 2015). EC is most typically an adenocarcinoma which initiates with the malignant transformation of cells on the inner lining of the uterus. More rarely EC begins from the muscle layer of the uterus and grows as an endometrial sarcoma.

The risk of developing endometrial cancer increases with age. 90% of the new ECs are diagnosed in patients over 50 years of age, and 50% in patients over 65 years of age. Other risk factors include obesity, diabetes, hormonal imbalance, hypertension, late menopause or exogenous estrogen use (Lynch et al., 1994; McPherson et al., 1996; Grénman and Leminen, 2013; Tiitinen, 2013). Most of
the EC patients are cured because EC is often diagnosed in its early stages. The survival rate after 5 years from diagnosis is approximately 80% (Cancer Society of Finland, 2015).

Approximately 90% of ECs are sporadic and 10% hereditary. EC in a first-degree relative increases ones risk of developing EC by 3-fold (Gruber and Thompson, 1996). Family history of EC also increases the risk of developing colon cancer. This suggests that genes involved in familial CRC may play an important role in EC tumorigenesis. There is a 50% risk of developing EC in Lynch syndrome mutation carriers. *MSH6* germline mutations are associated with high risk of EC and these cancers typically coexist with other Lynch syndrome associated cancers (Wijnen et al., 1999). Individuals with Lynch syndrome mutations are recommended to take yearly gynecological follow-up and later hysterectomy as preventive measure.

ECs are divided into two groups based on clinical and histopathological variables (Bokhman, 1983): estrogen-dependent endometrioid endometrial carcinomas (Type I; EECs) accounting for 70-80% of the EC cases and non-endometrioid endometrial carcinomas (Type II; NEECs). Type I tumors are histopathologically endometrioid, occur typically in younger menopausal and premenopausal women, are lower grade and more stable than Type II serous or clear cell tumors.

Molecular genetic evidence suggests that ECs develop through a multistep process similar to CRCs. ECs typically exhibit gain-of-function alterations in *KRAS*, *BRAF*, *HER2/neu*, β-catenin, Akt and *FGFR2* and loss-of-function of *PTEN* and *p53* (Okuda et al., 2010). The most frequent alterations in Type I tumors include the repression of *PTEN* and microsatellite instability. *KRAS* is mutated in approximately in 30-43%, *CTNNB1* in 20-45% and *ARID1A* in approximately 40% of Type I tumors. In contrast, most typical alterations in Type II tumors include *TP53* mutation or *HER2/neu* overexpression (O’Hara and Bell, 2012).

### 3.5.1 Lynch syndrome associated endometrial cancer

Endometrial cancer is the most common extra-colonic malignancy in Lynch syndrome accounting for approximately 2% of all ECs (see 3.4.2 Lynch syndrome; Koornstra et al., 2009). Individuals with pathogenic germline MMR gene mutation (*MLH1*, *MSH2* or *MSH6*) are prone to develop ECs in association with CRC and other tumors. The lifetime risk of MMR mutation carriers developing EC is up to 71% which means a 47 times higher risk compared to 1.5% of the general population (Strafford, 2012). Heterozygous mutations of *MSH6* are associated with 44-61% of risk of developing EC, while the risk of CRC is only 30% in female *MSH6* mutation carriers (Ramsoekh et al., 2009; Baglietto et al., 2010). The age of diagnosis of LS associated EC patients is approximately
38

20 years earlier than in the individuals without MMR deficiency (Marra and Boland, 1995; Schweizer et al., 2001).

### 3.5.2 Familial site-specific endometrial cancer (FSSEC)

There are families that show clustering of endometrial carcinoma without any other associated cancers. These ECs are called familial site-specific endometrial carcinoma (FSSEC) (Sandles et al., 1992). According to the study of Ollikainen et al. (2005) germline mutations in MMR genes were present in less than 10% of the families with familial clustering of EC only. In most cases of FSSEC it is not known whether there is a specific heritable molecular background behind these cancers or if they occur as part of another syndromes yet to be described.

### 3.6 Gastric cancer

Gastric cancer (GC) is a common reason for cancer-related deaths and it is the fifth most common cancer worldwide (Ferlay et al., 2015; WHO, World Cancer Report 2014). Approximately 1000 new patients are diagnosed with GC in Finland every year (Cancer Society of Finland, 2015). 95% of the diagnosed GCs are adenocarcinomas, the rest are lymphomas or other rare cancers (such as carcinoid tumors or gastrointestinal stromal (GIST) tumors) (Finnish Cancer Registry, 2015). The incidence of GC is lower than for example in colorectal or endometrial cancers, but the death rate of GC is relatively high.

GC is defined as intestinal, diffuse or non-classified based on its microscopic evaluation. Different GC types differ slightly by behavior. Diffuse GC appears typically at an earlier age than intestinal GC. Prognosis is better with intestinal GC, but the overall 5-year survival rate is worse. Only 30% of patients diagnosed with intestinal GC are alive after 5-year period (Laurén, 1965; Eto et al., 2006; Ristamäki et al., 2010). Intestinal GC is prone to metastasizing to the liver and diffuse GC to the lymphatic vessels. Risk for developing GC increases by age. Other risk factors include salted and smoked food, depletion of certain vitamins, pernicious anemia and inflammation induced by *Helicobacter pylori* infection, peptic ulcers, Ménétrier’s disease and smoking (Neugut et al., 1996; Jakszyn and Gonzalez, 2006; Tsugane and Sasazuki, 2007).

Gastric cancer is morphologically a very heterogeneous disease (Carneiro, 1997). Most cases of GC are sporadic in origin and familial clustering accounts for approximately 10% of the cases (Palli et al., 1994; Ekstrom et al., 2000). Chromosomal aberrations such as deletions, amplifications and
loss of heterozygosity affect multiple genes in GC. Most commonly affected genes in GC progression are APC, TP53, and NME1 all with over 30% of frequency and ERBB2 (also known as HER2/neu) (Gazvoda et al., 2007; Kaur and Dasanu, 2011). A single hereditary GC syndrome, Hereditary diffuse gastric cancer (HDGC), accounting for <3% of inherited GC cases is identified (Guilford et al., 1998). 40% of HDGC families harbor germline mutation in CDH1 gene and the rest of the cases fulfilling the clinical criteria for HDGC remain to be molecularly characterized. GC is also included in the Lynch syndrome tumor spectrum (Gylling et al., 2007) and LS patients have an increased risk for developing intestinal type GC. Other hereditary syndromes with increased risk of GC include Li-Fraumeni syndrome, Peutz-Jeghers syndrome and FAP (McLean and El-Omar, 2014).

4 Cancer epigenetics

Cancer research has traditionally focused on genetic defects but during the last decades epigenetic deregulation has been recognized as an important player in this field. Epigenetic regulation is involved in every step of tumorigenesis from tumor initiation through cancer development to expansion and metastasis (Sandoval and Esteller, 2012).

The role of DNA methylation (hypermethylation, hypomethylation and loss of imprinting) in tumorigenesis is well evidenced and the knowledge of different methylation events in cancer has accumulated during the past decades. The roles of other epigenetic mechanisms as well as the interplay between different epigenetic systems are an area of interest and under continuous study.

4.1 DNA hypermethylation in cancer

The function of tumor suppressors is often lost in cancers, enabling uncontrolled cell proliferation, division and growth. In addition to alterations in the genome tumor suppressor genes can also be silenced by promoter DNA hypermethylation (Stefansson and Esteller, 2013).

Many tumor suppressor genes are known to be methylated in the critical CpG sites in their promoter CpG island regions especially in sporadic cancers (Toyota et al., 1999; Suvá et al. 2013). Increased DNA methylation levels can be due to e.g. increased DNMT activity or carcinogen exposure. According to Knudson’s two-hit theory (Knudson, 1971) two hits are required for tumor suppressor genes to be silenced, one for both alleles. Promoter methylation can silence both
alleles or it can serve as either of these hits accompanying one mutated allele especially in the case of hereditary cancers (Peltomäki, 2012). Aberrant TSG CpG island hypermethylation is associated with transcriptional repression and loss of gene function (Esteller, 2008).

Several hundred to over a thousand CpG islands are found to be differentially methylated in different type of tumors (Kalari and Pfeifer, 2010). Tumors with extensive CpG island hypermethylation can be classified as displaying the CpG island methylator phenotype (CIMP; Toyota et al., 1999). The CIMP phenotype was originally defined by a set of DNA sequences known to display differential methylation in tumor and normal colorectal tissues (MINT1, MINT2, MINT12, MINT17, MINT25, MINT27 and MINT31) of which at least three had to be methylated in a tumor in order for it to be defined as CIMP positive (Toyota et al., 1999). CIMP+ tumors also have a high incidence of p16 (encoded by CDKN2A) and THBS1 methylation, and also MLH1 methylation is seen in sporadic MSI tumors. The definition of CIMP or a gene panel to test for it is not uniform and several panels exist for determining the CIMP phenotype of a tumor (e.g. Issa, 2004; Joensuu et al., 2008; Ogino et al., 2009). CIMP+ pathway is suggested to be the third pathway for genome-wide instability after CIN and MSI in cancers (Guan et al., 2013).

Epimutations, the heritable changes in gene expression through epigenetic aberrations, in MLH1 (Hitchins et al., 2007; Gylling et al., 2009) or MSH2 (Chan et al., 2006) leading to silencing of the affected allele have been described in some individuals with MSI tumors.

Also e.g. the deletion of 3’ exons of EPCAM is evidenced to cause epimutation and inactivation of MSH2 (Ligtenberg et al., 2009). Some of these epimutations seems to affect the germline, albeit frequently mosaic (Hitchins, 2015), and can possibly be passed through generations (Chan et al., 2006; Hitchins et al., 2007).

4.2 DNA hypomethylation in cancer

CpG hypomethylation was the first epigenetic abnormality recognized in human tumors, but the biological significance of tumor hypomethylation was not a particular area of interest for researchers who rather focused on hypermethylation events in tumor suppressor genes (Gama-Sosa et al., 1983; Feinberg and Vogelstein, 1983).

Global hypomethylation is a common feature in tumors. Loss of CpG methylation can activate oncogenes or when present globally, hypomethylation can promote tumorigenesis by inducing chromosomal instability (Feinberg and Vogelstein, 1983; Portela and Esteller, 2010).
Hypomethylation at repetitive DNA sequences or transposable elements accounts for most of the global hypomethylation observed in cancer. The most frequently observed hypomethylated sites in cancer are tandem centromeric satellite α, juxtacentromeric satellite 2, long interspersed nuclear elements (e.g. LINE-1) and interspersed Alu sequences (Ehrlich, 2009). Less frequently hypomethylation is seen in gene regions or CpG island promoters and mostly on oncogenes including transcription control sequences (Wasson et al., 2006, Wilson et al., 2007).

Global hypomethylation is associated with tumor progression but it can also be found in early carcinogenesis (Feinberg et al., 1988; Foy et al., 2015). Metastases are more susceptible to global DNA hypomethylation than primary tumors (Qu et al., 1999). Hypomethylation of repetitive DNA is almost always accompanied by DNA hypermethylation of tumor suppressor genes, both promoting aberrant tumor cell proliferation and for example the loss of imprinting (Jelinic and Shaw, 2007). Hypo- and hypermethylation events occur in the same tumors and seem not to be random, but the mechanisms behind these events are poorly understood.

4.3 Other epigenetic events in cancer

The interactions between epigenetics and genetics are altered in cancers. Aberrant histone modification patterns are common in different cancers and tumor types can be identified based on the histone modification patterns (Suvá et al., 2013). Genetic alterations in histone-modifying complexes are associated with altered epigenetic patterns and can induce global aberrant histone acetylation and methylation in cancer cells leading either to aberrant activation or repression depending on the modified site and the number of modifications (Ozdag et al., 2006). For example H3K27 targets loci for de novo methylation in cancer cells. Silencing of genes has been reported by trimethylated H3K27 alone without DNA methylation (Viré et al., 2006, Kondo et al., 2008). Also HDACs and Polycomb-group proteins EZH2 and BMI1 are known to be overexpressed in multiple human cancers and some lysine demethylases (e.g. KDM1A) have been associated with tumorigenesis (Berdasco et al., 2009; Bracken and Helin, 2009; Chase and Cross, 2011).

The expression of different miRNAs has been shown to have distinct profiles in tumor and normal tissues and between tumor types (He and Hannon, 2004; Lu et al., 2005) and certain epigenetically silenced or oncogenic miRNAs are associated with cancer progression and metastasis (Lujambio et al., 2008; O’Day and Lal, 2010).

Epigenetic mechanisms in disease, such as cancer, can be effectively studied with disease-discordant monozygotic (MZ) twins. MZ twins have the same genome, but as the epigenetic
patterns develop differently after fertilization (Castillo-Fernandez et al., 2014) MZ studies help the researchers to better understand the molecular mechanisms involved in disease and facilitate the search for new epigenetic markers.
AIMS OF THE STUDY

This study was conducted to identify epigenetic changes and associated factors underlying tumorigenesis in familial and sporadic cancers of different origins. The specific aims were the following:

1. To investigate gene promoter methylation changes of well-characterized tumor suppressor genes in colorectal, endometrial and gastric tumors and to identify differences between familial and sporadic cases (I, II)

2. To study global DNA hypomethylation in colorectal, endometrial and gastric tumors of different sporadic and familial backgrounds (III)

3. To elucidate the underlying causes and mechanisms behind aberrant DNA methylation in colorectal and endometrial tumors and endometrial hyperplasia (I-IV)
MATERIALS & METHODS

1 Patients and samples

This study was conducted on a number of cancer specimens with corresponding normal tissues and blood samples (for the number of patient material used in each study see Table 4). Commercially available cell lines were also used (description of the cell lines in Table 5). Patient samples had been partly characterized for MSI status, mutation screening for MMR genes (MLH1, MSH2, MSH6 and PMS2) and MMR protein expression in other studies (Holmberg et al., 1998; Kuismanen et al., 2000; Schweizer et al., 2001; Renkonen et al., 2003; Abdel-Rahman et al., 2005; Ollikainen et al., 2005). In this study, mutation and MSI analyses were conducted for gastric cancer patient samples (n=59 in study II) and immunohistochemical analyses of MMR proteins were carried out on gastric and sporadic colorectal patient samples (n=59 in study II and n=49 in study I, respectively). Additionally, the DNA methylation status of endometrial hyperplasias had been evaluated in a separate investigation (Nieminen et al., 2009). In the study series we divided the samples into different groups depending on the familial background of the patient (LS, FCCX and FSSEC) and sporadic cancers in two groups depending on their microsatellite instability status (MSS and MSI). 94% (17/18) of FCCX and 84% (20/24) of FSSEC cases were microsatellite stable (MSS), whereas most Lynch syndrome patients displayed MSI: 88-94% in LS CRC and 55-66% in LS EC (varying slightly between studies I, II and IV) and 100% LS GC (in both studies II and III).

Table 4. Number of the patient samples used in each study (I-IV).

<table>
<thead>
<tr>
<th></th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Colorectal cancer</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sporadic MSS</td>
<td>35</td>
<td>-</td>
<td>55</td>
<td>43</td>
</tr>
<tr>
<td>sporadic MSI</td>
<td>14</td>
<td>-</td>
<td>52</td>
<td>14</td>
</tr>
<tr>
<td>Lynch syndrome</td>
<td>33</td>
<td>-</td>
<td>43</td>
<td>42</td>
</tr>
<tr>
<td>FCCX</td>
<td>18</td>
<td>-</td>
<td>18</td>
<td>18</td>
</tr>
<tr>
<td><strong>Endometrial cancer</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lynch syndrome</td>
<td>38</td>
<td>-</td>
<td>50</td>
<td>48</td>
</tr>
<tr>
<td>FSSEC</td>
<td>24</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Lynch syndrome hyperplasia</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>38</td>
</tr>
<tr>
<td><strong>Gastric cancer</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sporadic MSS</td>
<td>-</td>
<td>36</td>
<td>34</td>
<td>-</td>
</tr>
<tr>
<td>sporadic MSI</td>
<td>-</td>
<td>10</td>
<td>11</td>
<td>-</td>
</tr>
<tr>
<td>Lynch syndrome</td>
<td>-</td>
<td>13</td>
<td>13</td>
<td>-</td>
</tr>
<tr>
<td><strong>Total number of the samples in the project</strong></td>
<td><strong>162</strong></td>
<td><strong>59</strong></td>
<td><strong>276</strong></td>
<td><strong>203</strong></td>
</tr>
</tbody>
</table>
Table 5. Description of cell lines used in this research.

<table>
<thead>
<tr>
<th>Cell line</th>
<th>Cancer type</th>
<th>Basis of MMR defect</th>
<th>Model for:</th>
</tr>
</thead>
<tbody>
<tr>
<td>HCA7</td>
<td>colon</td>
<td>biallelic MLH1 methylation</td>
<td>Sporadic CRC</td>
</tr>
<tr>
<td>HCT15</td>
<td>colon</td>
<td>homozygous MSH6 mutation</td>
<td>Lynch syndrome CRC</td>
</tr>
<tr>
<td>HCT116</td>
<td>colon</td>
<td>hemizygous MLH1 nonsense mutation</td>
<td>Lynch syndrome CRC</td>
</tr>
<tr>
<td>HEC59</td>
<td>endometrium</td>
<td>MSH2 mutation in both alleles</td>
<td>Lynch syndrome EC</td>
</tr>
<tr>
<td>KM12</td>
<td>colon</td>
<td>biallelic MLH1 methylation</td>
<td>Sporadic CRC</td>
</tr>
<tr>
<td>K562</td>
<td>leukemia</td>
<td>-</td>
<td>Leukemia</td>
</tr>
<tr>
<td>LIM1215</td>
<td>colon</td>
<td>hemizygous MLH1 mutation</td>
<td>Lynch syndrome CRC</td>
</tr>
<tr>
<td>LoVo</td>
<td>colon</td>
<td>homozygous MSH2 mutation</td>
<td>Sporadic CRC</td>
</tr>
<tr>
<td>RKO</td>
<td>colon</td>
<td>biallelic MLH1 methylation</td>
<td>Sporadic CRC</td>
</tr>
<tr>
<td>SW48</td>
<td>colon</td>
<td>biallelic MLH1 methylation</td>
<td>Sporadic CRC</td>
</tr>
</tbody>
</table>

(Adapted from: Andersson et al., 1979; Joensuu et al., 2008)

A more detailed description of the patient material and specifics of the methods, including the information of primers, probes and antibodies used in this study, can be found in the original publications (I-IV). For the summary of the methods used in each study see Table 6.

Table 6. Summary of the methods used in this study.

<table>
<thead>
<tr>
<th>Methods</th>
<th>Publications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bisulfite sequencing (BS)</td>
<td>I, II, III, IV</td>
</tr>
<tr>
<td>Combined bisulfite restriction analysis (COBRA)</td>
<td>III</td>
</tr>
<tr>
<td>DNA bisulfite modification</td>
<td>I, II, III</td>
</tr>
<tr>
<td>DNA extraction</td>
<td>I, II, III, IV</td>
</tr>
<tr>
<td>DNA fragment analysis</td>
<td>I, II, III, IV</td>
</tr>
<tr>
<td>DNA sequencing</td>
<td>I, II</td>
</tr>
<tr>
<td>Immunohistochemistry (IHC)</td>
<td>I, II, IV</td>
</tr>
<tr>
<td>Methylation-specific multiplex ligation-dependent probe amplification (MS-MLPA)</td>
<td>I, II, III, IV</td>
</tr>
<tr>
<td>Methylation-specific PCR (MSP)</td>
<td>I, II</td>
</tr>
<tr>
<td>Microsatellite instability (MSI) analyses</td>
<td>I, II</td>
</tr>
<tr>
<td>Multiplex ligation-dependent probe amplification (MLPA)</td>
<td>II</td>
</tr>
<tr>
<td>Single nucleotide primer extension (SNuPE)</td>
<td>II</td>
</tr>
<tr>
<td>Single-strand conformation polymorphism analysis (SSCP)</td>
<td>II</td>
</tr>
<tr>
<td>Statistical analyses</td>
<td>I, II, III, IV</td>
</tr>
</tbody>
</table>
2 Genomic mutation, microsatellite instability and loss of heterozygosity analyses

DNA was extracted from formalin-fixed paraffin-embedded (FFPE) tissue blocks (patient specimens) by a method modified from Isola et al. (1994) or cells (blood and cell lines) as described in Lahiri and Nurnberg (1991). The DNA was amplified using polymerase chain reaction (PCR) before further analyses.

Exon-specific genomic sequencing was used to search for point mutations in MLH1, MSH2, MSH6, KRAS, BRAF and CTNNB1 and to specifically test for the MMR founder mutations, as described previously (Chadwick et al., 2001; Vahteristo et al., 2001; Nyström-Lahti et al., 1995).

Single-strand conformation polymorphism (SSCP) was used to screen hot spot mutations in KRAS, CTNNB1 and BRAF (Deng et al., 2004; Abdel-Rahman et al., 2005). Amplified DNA was separated on a polyacrylamide gel and 1 x MDE Gel Solution (Cambrex BioScience Rockland Inc.) was used to improve sensitivity to conformational changes. Gels were stained using silver nitrate for DNA detection. If changes were observed in the SSCP gels, the DNA was sequenced to determine the nucleotide changes.

Microsatellite instability (MSI) was studied by DNA fragment analyses using microsatellite markers BAT25, BAT26, D2S123, D5S346 and D17S250 (Umar et al., 2004a; the revised Bethesda panel). The markers were run on an automatic ABI 3730 DNA sequencer and analyzed with GeneMapper 4.0 software (Applied Biosystems). Tumors without unstable markers were considered MSS, while tumors with two or more unstable markers were MSI.

Single nucleotide primer extension analysis (SNuPE) based on MLH1 polymorphism in exon 8 (I219V A/G polymorphism) was used to analyze LOH (Renkonen et al., 2003). SNuPE is based on the incorporation of a single selected dideoxy-nucleotide to produce differential extension when the primer is annealed next to the polymorphic sequence. LOH in APC was also studied by SNuPE using polymorphisms in exon 11 and flanking MSI markers D5S1965 and D5S346 (Renkonen et al., 2005). Ratios of peak areas were calculated in normal and tumor samples of the same patient and values <0.6 or >1.67 were considered to indicate strict LOH (Ollikainen et al., 2005).

In patients with a verified MMR germline mutation, multiplex ligation-dependent probe amplification (MLPA, SALSA MLPA kit P003, MRC-Holland; Schouten et al., 2002) was used to detect locus-restricted LOH following manufacturer’s standard protocols. SALSA MLPA P003 kit contains probes for all the exons of MLH1 and MSH2. In MLPA, DNA was first denatured after which the MLPA probes were hybridized to their target sequences. If a deletion was not present, the probe was able to hybridize with its target sequence. After hybridization the probes were
ligated and the ligated probes were amplified in a PCR reaction. PCR primers contain fluorescent labels and could be analyzed on an automatic sequencer and analyzed using GeneMapper 4.0 software (Applied Biosystems). Relative peak values were calculated for the normal and tumor samples of the same patient. For LOH analysis the dosage ratio of the tumor sample was divided by the corresponding ratio of the normal sample and values <0.6 or >1.67 were considered to indicate LOH. MLPA was also used in the study of large genomic deletions (MLH1 ex16 and MSH2 ex3-5).

3 DNA methylation analyses

Methylation-specific multiplex ligation-dependent probe amplification method (MS-MLPA, MRC-Holland; Nygren et al., 2005) was used as the primary method for DNA methylation studies in this research. The SALSA MS-MLPA kits ME001 and ME001B (MRC-Holland) were used for the TSG promoter methylation studies (I, II, III, for the genes studied see Table 7 in Results) and the SALSA MLPA P300 Human DNA reference-2 probemix with FAM-labeled SALSA MLPA EK1 reagent kit with custom designed probes for LINE-1 and methyltransferase gene promoter methylation studies (III, IV) (Pavicic et al. 2012; Joensuu et al., 2015). The basis and outline of the MS-MLPA protocol and the interpretation of the results are explained in Figures 5 and 6; also see description of MLPA outline above (chapter 2 in Materials and methods).

MS-MLPA results were validated in cell lines by bisulfite sequencing (BS), combined bisulfite restriction analysis (COBRA) or methylation-specific PCR (MSP). DNA bisulfite modification required for further analyses by BS, COBRA and MSP was conducted by CpGenome DNA Modification kit (Chemicon) following manufacturer’s instructions.

BS was done using specifically designed methylation-unbiased primers on the bisulfite modified DNA. Both, the methylated and unmethylated samples were amplified in the PCR reaction. The PCR products were then sequenced to determine methylation status.

COBRA was used to study LINE-1 promoter methylation by a modified COBRA PCR protocol (Chalitchagorn et al., 2004). The bisulfite modified DNA was amplified in a PCR reaction prior to digestion with Tacl or TasI in corresponding enzyme buffer (MBI Fermentas). The digested PCR products were run in 4.5% NuSieve 3.1 agarose gels (Cambrex). NIH Image Software was used to scan and score the intensities of digested and undigested bands, and the methylation level was calculated as a percentage of the intensity of TaqI divided by the sum of TaqI and TasI positive products.
MSP analyses were conducted using a HotStarTaq DNA polymerase (Qiagen) following manufacturer’s standard cycling protocol and the bisulfite modified DNA. The PCR was done in parallel with a different set of primers for methylated and unmethylated target sequences. PCR products were visualized on an agarose gel under UV illumination.

Figure 5. Outline of the MS-MLPA protocol. The probes are designed separately for each methylated target and one full length probe consists of two parts (right and left oligo probes) per target (black-blue-green; black is the universal primer, blue is the hybridization sequence of the target gene and green is a stuffer sequence). MS-MLPA is multiplexed and up to 40 genes can be studied in one sample run. The DNA is denatured and the probes are hybridized to the DNA. The probes, each consisting of two parts, are then ligated to obtain one full length probe per gene under study. Digestion is conducted simultaneously using methylation-sensitive endonuclease (HhaI) which recognizes GCGC sites but does not cut if the site is methylated (see the difference between methylated and unmethylated targets). Ligated uncut probes are amplified with PCR using one universal primer pair X and Y, yielding DNA fragments of different lengths. PCR products are then subjected to fragment analysis and sample comparison (Fig. 6) (modified from Nygren et al., 2005).
Figure 6. Detection of DNA methylation by MS-MLPA. The figure shows one normal unmethylated tissue sample as a reference, and one tumor sample under investigation. Ligated probes show as peaks in the electropherogram. Control genes (here 15) do not have the restriction site for HhaI and always generate a peak after ligation, digestion and PCR, but the target genes (here 26) only generate peaks when methylated. Three genes in this tumor sample show DNA methylation compared to normal tissue (see the red arrows). The relative peak area of the ligated and digested sample is compared to the peak area of the ligated sample from the same tissue sample. The methylation dosage ratio is obtained as follows:

\[
D_m = \frac{P_x}{P_{ctrl}^{Undig}} / \frac{P_x}{P_{ctrl}^{Dig}},
\]

where \( D_m \) is the methylation dosage ratio, \( P_x \) is the peak area of a given probe, \( P_{ctrl} \) is the sum of peak areas of all the probes, Dig is sample digested by methylation-sensitive HhaI endonuclease, and Undig is undigested (only ligated) sample. The obtained \( D_m \) ranges from 0 to 1, referring to 0 to 100% methylation. We used the cut-off value 0.15 for indicating DNA hypermethylation based on the mean \( D_m \) in normal tissue samples +1SD (Joensuu et al., 2008).
4 Protein expression analyses

Immunohistochemical staining was used to study the presence or the absence of specific proteins in the patient samples using FAM-labelled antibodies for desired proteins with Dako EnVision™+ System-HRP (DAB) detection kit (with mouse or rabbit primary antibodies when appropriate) by manufacturer’s standard protocol. Nuclear protein expression was determined in tumors and compared to the results of normal tissue samples. Protein expression was quantified by percentage of stained nuclei in the specimen for MLH1, MSH2, MSH6, β-catenin (cut-off >10% positive nuclei), p53 (cut-off >10% positive nuclei) and Ki-67 or by using the Staining Index (SI) value, determined by the percentage of stained nuclei (0=0, <10=1, 10-50=2, >50=3) multiplied by staining intensity (range 0-3) (similar to e.g. Fluge et al., 2009), ranging from 0-9 for DNMT1, DNMT3B and EZH2. Nuclear stainings of corresponding normal tissues were used as reference.

5 Statistical analyses

Statistical analyses were conducted by using PASW Statistics 18 software (SPSS Inc.) (III-IV) or VassarStats statistical computation website (I-II). Shapiro-Wilk test was used to test the normality of the data. Depending on the distribution of the data a parametric (for normally distributed data) or nonparametric (for non normal data) test was chosen. Wilcoxon signed-rank test (nonparametric) or t-test (parametric) was used for pairwise analyses of correlated samples (intragroup comparison). The significance of the difference between means of two independent groups were analyzed by Mann-Whitney U test (for nonparametric) or t-test (for parametric data). Categorical data with dichotomous variables was analyzed by Fisher’s exact test or by Chi-squared test. Comparison of multiple independent groups was done by using one-way ANOVA (for parametric data) or Welch test when the assumption of homogeneity of variances (analyzed by Levene’s test) was not met, followed by appropriate post hoc test (Games-Howell test) to determine the differing group. Correlations were tested by Pearson product-moment correlation coefficient (r) for parametric or by Spearman’s rank correlation coefficient (rho) for nonparametric data. Two-tailed $P$ values of <0.05 were considered significant.
RESULTS

In this research our aim was to clarify the epigenetic changes present in cancers of different origin, colon, endometrium and gastric, and to define the underlying causes for these changes. We also wanted to analyze the extent of concordance between the epigenetic patterns of each tissue type in sporadic and familial tumors and if there were epigenetic differences between microsatellite stable and unstable tumors.

In addition to sporadic and Lynch syndrome (LS) tumors we also had two interesting groups of familial cancers with HNPCC-like susceptibility i.e. Familial colorectal cancer type X (FCCX) and Familial site-specific endometrial cancer (FSSEC). FCCX patients have colorectal cancer meeting the clinical criteria for Lynch syndrome but without identified MMR gene mutations or accompanying extracolonic tumors. In FSSEC endometrial cancers occur in families without tumors of other tissues. In a previous study Ollikainen et al. (2005) found germline MMR gene mutations in only 8% of the families with familial clustering of EC only. MMR protein expression changes (MLH1, MSH2 or MSH6) were observed in 52% of the families, discordantly between family members. Fascinated by the observations that there could be an epigenetic basis for the cancer susceptibility (Suter et al., 2004; Chan et al., 2006) we studied whether certain epigenetic changes could be responsible for cancer initiation or progression in these familial tumors that remain molecularly unexplained, so far.

1 DNA hypermethylation in cancer (I-II)

We studied 221 carcinomas from sporadic and familial cases for promoter methylation of 24 tumor suppressor genes (TSGs; Table 7) using methylation-specific multiplex ligation-dependent probe amplification (MS-MLPA) method (Fig. 5 and Fig 6) (ME001B kit, MRC-Holland). These TSGs are known to be inactivated in cancer and the genes were selected based on their published relevance in carcinogenesis (Verma and Srivastava, 2002; Laird, 2003). Our study series consisted of 100 colorectal (CRC), 62 endometrial (EC) and 59 gastric (GC) tumor specimens with corresponding normal tissue samples.
Table 7. Tumor suppressor genes studied for promoter methylation.

<table>
<thead>
<tr>
<th>Gene symbol</th>
<th>Protein function</th>
</tr>
</thead>
<tbody>
<tr>
<td>APC</td>
<td>Participates in Wnt signaling, free beta-catenin level regulator, involved in cell adhesion and active cell migration</td>
</tr>
<tr>
<td>ATM</td>
<td>Protein kinase, DNA damage sensor and damage response regulator, checkpoint signaling activator at double-strand breaks, apoptosis and genotoxic stresses, involved in signal transduction and cell cycle control</td>
</tr>
<tr>
<td>BRCA1</td>
<td>Central role in DNA repair by facilitating cellular response to DNA repair, mediator of E2-dependent ubiquitination, lipid synthesis inhibitor</td>
</tr>
<tr>
<td>BRCA2</td>
<td>Involved in double-strand break repair and/or homologous recombination</td>
</tr>
<tr>
<td>CASP8</td>
<td>Protease in the activation cascade of caspases responsible for apoptosis</td>
</tr>
<tr>
<td>CD44</td>
<td>Receptor for hyaluronic acid, mediator of cell-cell and cell-matrix interactions, role in cell migration, tumor growth and progression</td>
</tr>
<tr>
<td>CDH13</td>
<td>Calcium dependent cell adhesion protein</td>
</tr>
<tr>
<td>CDKN1B</td>
<td>Involved in cell cycle arrest, cellular response to DNA damage, regulator of cell cycle, effector of signal transduction pathways that control cell differentiation</td>
</tr>
<tr>
<td>CDKN2A</td>
<td>Induces cell cycle arrest in G1 and G2 phases, negative regulator of the proliferation of normal cells</td>
</tr>
<tr>
<td>CDKN2B</td>
<td>Regulator of cell cycle, passage through the G1 checkpoint, tightly linked and highly homologous to CDKN2A</td>
</tr>
<tr>
<td>CHFR</td>
<td>E3 ubiquitin-protein ligase, cell arrest in early prophase</td>
</tr>
<tr>
<td>DAPK1</td>
<td>Calcium/calmodulin-dependent serine/threonine kinase, positive regulator of apoptosis</td>
</tr>
<tr>
<td>ESR1</td>
<td>Nuclear estrogen receptor α, involved in gene expression regulation, cellular proliferation and differentiation</td>
</tr>
<tr>
<td>FHIT</td>
<td>Diadenosine 5,5-P1,P3-triphosphate hydrolase, involved in purine metabolism</td>
</tr>
<tr>
<td>GSTP1</td>
<td>Involved in conjugation of reduced glutathione to a wide number of exogenous and endogenous hydrophobic electrophiles</td>
</tr>
<tr>
<td>HIC1</td>
<td>Transcriptional repressor</td>
</tr>
<tr>
<td>IGFS4 (=CADM1)</td>
<td>Cell adhesion molecule</td>
</tr>
<tr>
<td>MLH1</td>
<td>DNA mismatch repair protein</td>
</tr>
<tr>
<td>PTEN</td>
<td>Protein and lipid phosphatase, cell cycle progression and cell survival modulator</td>
</tr>
<tr>
<td>RARB</td>
<td>Receptor for retinoic acid, gene expression regulator</td>
</tr>
<tr>
<td>RASSF1(A)</td>
<td>Required for death receptor-dependent apoptosis Isoform A inhibits proliferation by negatively regulating cell cycle progression</td>
</tr>
<tr>
<td>TIMP3</td>
<td>Metalloproteinase inhibitor</td>
</tr>
<tr>
<td>TP73</td>
<td>Involved in apoptotic response to DNA damage</td>
</tr>
<tr>
<td>VHL</td>
<td>Transcriptional repressor, ubiquitin conjugating protein, cell cycle regulator</td>
</tr>
</tbody>
</table>

(Adapted from Joensuu et al., 2008; based on www.genecards.org, www.ensembl.org)

1.1 Shared and tissue-specific tumor suppressor gene methylation patterns

Of the 24 tumor suppressor gene promoter areas studied, a great proportion were found to display differential methylation between the tumor and corresponding normal tissue samples (Table 8). 11 genes out of 24 studied were commonly hypermethylated in at least 15% of the tumor samples per tissue category. In general, paired normal samples showed negligible methylation in these genes in normal colorectal, endometrial or gastric mucosae or blood (Table 8). If methylation was observable in the normal tissue samples, the methylation dosage ratio (Dm) was notably lower than in tumor samples. Out of these 11 commonly methylated genes, 10 genes showed frequent promoter hypermethylation in colorectal cancer (ESR1, CDH13, CHFR, RARB, APC, RASSF1(A), TIMP3, GSTP1, IGFS4 and MLH1), 6 in endometrial cancer (RASSF1(A), CDH13,
APC, GSTP1, CDKN2A and TIMP3) and 5 in gastric cancer (APC, ESR1, CHRF, TIMP3 and CDH13) (study I for CRC and EC; unpublished results for GC; Table 8). Overall, individual tumors displayed hypermethylation in 0–11 tumor suppressor genes, except one LS CRC tumor with 18 hypermethylated loci. The average hypermethylation ratio was significantly higher in CRC (4.2 methylated loci per tumor) compared to EC (2.7; \( P < 0.001 \)) and GC (2.9; \( P = 0.005 \)) (Table 9, below).

The promoters of three tumor suppressor genes, APC, CDH13 and TIMP3, were commonly hypermethylated in all of the studied tissues. APC was hypermethylated in 35% of CRC and 37% of EC but showed highest frequency of methylation in GC being present in 78% of the GC tumors. CDH13 was hypermethylated in approximately in half of the CRC (57%) and EC (44%) tumors and in 26% of GC tumors. TIMP3 was also commonly hypermethylated in all of the studied tissues (18–29%) (see Table 8).

Some TSGs were hypermethylated in a tissue-specific manner. ESR1, RARB, IGSF2 and MLH1 hypermethylation was characteristic for CRC, whereas CDKN2A characterized EC. RASSF1(A) was hypermethylated in both CRC and EC, but three times more often in EC (74%) than in CRC (24%). APC which was hypermethylated to some extent in all of the studied tissues, showed tissue-specificity to GC, with a much higher percentage than seen in other tissues. ESR1 and CHFR hypermethylation was present in 82% and 44% of CRC and 40% and 38% of GC tumors (respectively; Table 8), but not in EC, suggesting specificity to the gastrointestinal tract.

Table 8. Tumors with tumor suppressor gene hypermethylation shown as the percentage of each tissue category.

<table>
<thead>
<tr>
<th></th>
<th>Colon tumors (n= 100)</th>
<th>Endometrial tumors (n= 62)</th>
<th>Gastric tumors (n=59)</th>
</tr>
</thead>
<tbody>
<tr>
<td>APC</td>
<td>35 % (&lt; 0.001)</td>
<td>37 % (&lt; 0.001)</td>
<td>78 % (= 0.004)</td>
</tr>
<tr>
<td>CDH13</td>
<td>57 % (&lt; 0.001)</td>
<td>44 % (&lt; 0.001)</td>
<td>26 % (ns.)</td>
</tr>
<tr>
<td>CDKN2A</td>
<td>21 % (ns.)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CHFR</td>
<td>44 % (&lt; 0.001)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ESR1</td>
<td>82 % (&lt; 0.001)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GSTP1</td>
<td>22 % (ns.)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IGSF4</td>
<td>16 % (= 0.027)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MLH1</td>
<td>15 % (= 0.004)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RARB</td>
<td>42 % (&lt; 0.001)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RASSF1(A)</td>
<td>24 % (&lt; 0.001)</td>
<td>74 % (&lt; 0.001)</td>
<td></td>
</tr>
<tr>
<td>TIMP3</td>
<td>27 % (&lt; 0.001)</td>
<td>18 % (= 0.007)</td>
<td>29 % (=0.061)</td>
</tr>
</tbody>
</table>

Table displaying all the tumor suppressor genes which were at least 15% hypermethylated per tissue category and with methylation dosage ratio (Dm) ≥0.15, corresponding to ≥15% of methylation (based on the mean Dm in normal tissue samples +1SD), similarly to Rashid and Issa (2004). \( P \) values for differential methylation in tumor and normal samples given in parentheses; ns. = non-significant. Empty designates no methylation in the given gene promoter in that tissue.
RMA (robust multiarray average) pre-processed microarray mRNA data (HG-U133 Plus 2.0 array, Affymetrix) from another ongoing study (Abdel-Rahman et al.) was utilized to verify that the hypermethylation of the studied CpG sites was in fact associated with gene silencing in cell lines. The TSG promoter hypermethylation results were associated with reduced or absent mRNA expression, as best evident for TIMP3, MLH1, RARB, CHFR, IGSF4, CDH13, and GSTP1, but the statistical significance was not determined (Joensuu et al., 2008; Supplementary Table 6).

1.2 DNA methylation patterns in separate patient categories – dependence of methylation patterns on the MSI status and the familial background of cancer

Patients were divided into smaller groups depending on their familial background of cancer or the MSI status of the tumor (Table 9). In these subcategories, the highest amount of methylated TSGs was found in sporadic MSI CRC with an average of 7.3 methylated loci per tumor. Overall the sporadic MSI CRC group displayed significantly higher methylation ratios compared to other CRC groups \((P < 0.001)\). The average number of hypermethylated TSG promoter loci was higher in MSI than in MSS tumors in GC (Table 9) and the difference was statistically significant \((P < 0.001)\). In ECs no clear differences were found between groups. Even though Lynch syndrome CRC also display MSI (in 94% of all the LS CRC tumors), the methylation ratio was significantly lower compared to the corresponding sporadic MSI tumors \((P < 0.001)\) but not in GC \((P = 0.614)\).

Table 9. Average number of hypermethylated TSGs and the presence of CIMP+ phenotype in different patient categories.

<table>
<thead>
<tr>
<th>Patient category</th>
<th>(n)</th>
<th>Methylated loci* per tumor</th>
<th>CIMP+ (in % of the tumors per category)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Average</td>
<td>Range</td>
</tr>
<tr>
<td>Colorectal cancer</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sporadic MSS</td>
<td>35</td>
<td>3.5</td>
<td>1-8</td>
</tr>
<tr>
<td>sporadic MSI</td>
<td>14</td>
<td>7.3</td>
<td>3-10</td>
</tr>
<tr>
<td>Lynch syndrome</td>
<td>33</td>
<td>3.9</td>
<td>0-11**</td>
</tr>
<tr>
<td>FCCX</td>
<td>18</td>
<td>3.7</td>
<td>0-8</td>
</tr>
<tr>
<td>Endometrial cancer</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lynch syndrome</td>
<td>38</td>
<td>2.9</td>
<td>0-8</td>
</tr>
<tr>
<td>FSSEC</td>
<td>24</td>
<td>2.3</td>
<td>0-7</td>
</tr>
<tr>
<td>Gastric cancer</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sporadic MSS</td>
<td>36</td>
<td>2.0</td>
<td>0-7</td>
</tr>
<tr>
<td>sporadic MSI</td>
<td>10</td>
<td>4.6</td>
<td>1-8</td>
</tr>
<tr>
<td>Lynch syndrome</td>
<td>13</td>
<td>4.0</td>
<td>2-6</td>
</tr>
</tbody>
</table>

* out of 24 tumor suppressor gene loci studied. One locus was studied for each gene, except 2 loci for RASSF1A.
CIMP+, CpG island methylator phenotype (≥5/24 methylated tumor suppressor genes; see chapter 1.3 in Results).
** plus one individual tumor with 18 methylated loci. Bold = groups with statistically significant higher hypermethylation levels compared to other patient categories of the same tissue.
1.3 CpG island methylator phenotype as a driver of tumorigenesis?

CpG island methylator phenotype (CIMP, or tumor suppressor gene methylator phenotype) is an alternative mechanism causing genome-wide instability in tumors (Toyota et al., 1999). CIMP in our sample series was set as ≥5 methylated genes out of 24 tumor suppressors studied. 38% (38/100) of all CRC samples displayed the CIMP phenotype (CIMP+) and it was present also in 18% (11/62) of all EC and 29% (17/59) of all GC tumors. CIMP+ phenotype was much more common in the sporadic MSI groups (in CRC 79% and GC 73%) compared to the other patient categories (P < 0.001) (Table 9). In sporadic MSI cases, the CIMP+ phenotype was characterized by the presence of MLH1 promoter methylation in both CRC and GC, which could explain the MSI through defective MMR repair. Of all the tumors displaying MLH1 hypermethylation (23/29), 79% also displayed CIMP+ (P < 0.001). Clear dichotomy in CIMP status was observed in FCCX samples and the group could be divided into two categories with 50% of the group displaying CIMP+ whereas the other half did not.

Methylation patterns in sporadic MSI (but not MSS) cancers resembled those in LS cancers (CRC and GC) with the exception of MLH1 promoter methylation, which was usually not methylated in LS tumors. Distinct methylation patterns in sporadic MSS and MSI tumors suggest different developmental pathways of these cancers.

In our study series no transgenerational or germline defect in DNA methylation was observed. Generally, tumors from single representatives (usually the index persons) per family were investigated. In the occasional cases where tumors from several family members were available, family-specific methylation patterns were not identified. The constitutional tissues representing normal cells of different lineages showed negligible methylation. Furthermore, we took a look at nine LS patients with both CRC and EC tumor but the tumors from the same individuals did not show any concordance between promoter methylation profiles when tissue-specific differences were taken into account. Because of the small sample size statistical analyses could not be conducted.

2 Global genomic hypomethylation in cancer (III)

Global hypomethylation is often observed in cancers and it may cause genetic instability and oncogene activation. It has been suggested that DNA hypomethylation could be one of the first events in tumorigenesis (Feinberg et al., 1988; Foy et al., 2015). We studied CRC, EC and GC tumors (n=276) and their corresponding normal tissue samples for LINE-1 hypomethylation, which is a
transposable element present in over 500,000 copies in the genome (~17% of the whole genome; Xiao-Je et al., 2015). The FCCX group was of particular interest as no cancer predisposing molecular defects had previously been found in it.

Measured by LINE-1 methylation, we found a decrease in the global methylation levels of all tumors. Compared to the corresponding normal tissues the decrease was statistically significant in most cases (Table 10). The degree of LINE-1 methylation in normal mucosa showed little variation between patient groups (Dm 0.90–0.93) except for the FCCX group (Dm 0.84) which clearly differed from the remaining groups (P < 0.05). This interesting finding may suggest that global hypomethylation promotes cancer development as a field defect, a field of pre-malignant tissue harboring one potentially carcinogenic aberration vulnerable to accumulation of other carcinogenic defects (Bernstein et al., 2013), in the normal tissues of these patients. Inter-group comparisons revealed that the decreasing trend of LINE-1 methylation in sporadic CRC tumors with MSI to LS to FCCX was statistically significant between first and last group (P = 0.042; Dm 0.87 and 0.80, respectively). A similar finding was observed for GC between the sporadic MSI (Dm 0.88) and MSS (Dm 0.79) groups (P = 0.018).

### Table 10. Average LINE-1 methylation (Dm) in different patient categories.

<table>
<thead>
<tr>
<th>Patient category</th>
<th>(n)</th>
<th>Tumor</th>
<th>Normal mucosa</th>
<th>P value</th>
<th>Blood*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Colorectal cancer</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sporadic MSS</td>
<td>55</td>
<td>0.85</td>
<td>0.93</td>
<td>&lt;0.001</td>
<td></td>
</tr>
<tr>
<td>Sporadic MSI</td>
<td>52</td>
<td>0.87</td>
<td>0.91</td>
<td>&lt;0.001</td>
<td></td>
</tr>
<tr>
<td>Lynch syndrome</td>
<td>43 (27)</td>
<td>0.84 (0.84)</td>
<td>0.90 (0.91)</td>
<td>&lt;0.001</td>
<td>(0.95)</td>
</tr>
<tr>
<td>FCCX</td>
<td>18 (10)</td>
<td>0.80 (0.80)</td>
<td>0.84 (0.85)</td>
<td>&lt;0.05</td>
<td>(0.89)</td>
</tr>
<tr>
<td>Endometrial cancer</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lynch syndrome</td>
<td>50 (19)</td>
<td>0.88 (0.91)</td>
<td>0.90 (0.93)</td>
<td>ns.</td>
<td>(0.95)</td>
</tr>
<tr>
<td>Gastric cancer</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sporadic MSS</td>
<td>34</td>
<td>0.79</td>
<td>0.90</td>
<td>&lt;0.001</td>
<td></td>
</tr>
<tr>
<td>Sporadic MSI</td>
<td>11</td>
<td>0.88</td>
<td>0.90</td>
<td>ns.</td>
<td></td>
</tr>
<tr>
<td>Lynch syndrome</td>
<td>50</td>
<td>0.86</td>
<td>0.90</td>
<td>&lt;0.05</td>
<td></td>
</tr>
</tbody>
</table>

Dm = methylation dosage ratio, the values for samples with all three tissues available are marked in parentheses.

*Blood samples were available only for LS CRC, FCCX and LS EC, in which the differences between tumor and blood were all statistically significant (P < 0.001, P < 0.01 and P < 0.01, respectively). ns. = non-significant.

Tumor suppressor gene hypermethylation and global hypomethylation have been described as related events (Shi et al., 2007). Even though both hyper- and hypomethylation was observed in the same tumors, there was no clear correlation between TSG hypermethylation (I, II) and LINE-1 hypomethylation (III) in our sample series. Overall, the results suggest that the LINE-1 methylation status depends on the MSI status of the different patient categories and that hypomethylation
may function as a field defect (Bernstein et al., 2013) when seen already in the normal mucosa, as in the case of the FCCX patient samples studied.

3 Changes in methyltransferase activity in cancer (IV)

After discovering changes in epigenetic patterns of different carcinomas we wanted to elucidate the underlying causes for aberrant DNA methylation. DNA methyltransferases DNMT1 and DNMT3B and histone methyltransferase EZH2 are described to be overexpressed in colorectal and several other tumors (Eads et al., 1999; Vallbohmer et al., 2006; Zhu et al., 2007; Fluge et al., 2009). For example, Nosho et al. (2009) suggest that aberrant DNMT3B expression might contribute to the CIMP phenotype in CRC. The protein expression of three methyltransferases (MTs), EZH2, DNMT1 and DNMT3B, was investigated by immunohistochemical evaluation on 117 familial and sporadic CRC, 48 EC and 38 endometrial hyperplasia samples to determine if abnormal expression of some of these proteins contributes to tumor suppressor gene hypermethylation, CIMP phenotype or the global hypomethylation patterns observed in the samples.

3.1 Aberrant DNA methylation – aberrant methyltransferase patterns?

DNMT1, DNMT3B and EZH2 were expressed at significantly higher levels ($P < 0.001$) in CRC tumors compared to their paired normal samples (mean SI values ranging from 2.65 to 5.35 for DNMT1, 6.52–8.00 for DNMT3B, and 1.79–6.58 for EZH2 in tumors and 1.13–2.56, 0–2.11, and 0–1.33 in paired normal mucosa, respectively) in all studied patient categories except DNMT1 in FCCX patient category (not reaching statistical significance $P = 0.067$) (Table 11). In EC MT expressions were relatively low except for DNMT3B expression which was significantly higher in tumors than in normal endometrium ($P = 0.001$). DNMT1 and EZH2 expressions were discovered to be significantly higher in CRC than in EC tumors (mean 4.20 vs. 0.89 for EZH2 and mean 4.34 vs. 1.19 for DNMT1 in CRC and EC, respectively; $P < 0.001$ in both).
EZH2 expression was positively correlated with TSG hypermethylation (CIMP+, being ≥5 methylated TSG promoters) in sporadic CRC tumors ($P = 0.003$; rho .418), whereas the expression of DNMT1 did not display correlation with TSG methylation status in any of the studied patients groups. Neither EZH2 nor DNMT1 expression correlated with the LINE-1 hypomethylation (our unpublished data). DNMT3B was significantly overexpressed in all tumors when compared to the normal tissues ($P < 0.001$, with the group means ranging from SI 6.50 to 8.00) but because of the general lack of variation observed in DNMT3B expression correlations with DNA methylation levels could not be calculated for groups other than the LS CRC patient group which did not show any significance. Although heavily overexpressed, it still seems that DNMT3B overexpression does not affect DNA methylation patterns at least as much as EZH2, since DNMT3B was also overexpressed in tumors without TSG hypermethylation.

DNMT1 and EZH2 expression was positively correlated especially in colorectal tumors ($P = 0.001$; rho .317) suggesting co-operative functions in the epigenetic pathways of tumorigenesis. In EC the trend was similar ($P = 0.091$; rho .258). DNMT1 and EZH2 expression was higher in MSI than in MSS tumors as was Ki-67 expression which was studied as a proliferation marker. The difference between MSI and MSS tumors was statistically significant between LS CRC and FCCX for DNMT1 ($P = 0.002$), and between sporadic MSI CRC and sporadic MSS CRC for EZH2 ($P = 0.026$).

In endometrial hyperplasia series from Lynch syndrome patients an increasing trend in MT expression was observed (from simplex hyperplasia to complex hyperplasia to complex atypical hyperplasia) but the statistical significance could not be determined due the small sample sizes (Table 12). This result is in line with the finding of Nieminen et al. (2009) where TSG promoter methylation increases along with histological abnormality (observed in the same endometrial
hyperplasia samples as in this study). This suggests that the overexpression of MTs could be important in neoplastic formation.

Table 12. Mean methyltransferase protein expression (SI) in normal endometrium, different endometrial hyperplastic stages and endometrial carcinoma from Lynch syndrome patients.

<table>
<thead>
<tr>
<th></th>
<th>DNMT1</th>
<th>DNMT3B</th>
<th>EZH2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal endometrium</td>
<td>0.73</td>
<td>3.33</td>
<td>0.17</td>
</tr>
<tr>
<td>Simplex hyperplasia (SH)</td>
<td>1.71</td>
<td>4.89</td>
<td>0.44</td>
</tr>
<tr>
<td>Complex hyperplasia (CH)</td>
<td>3.00</td>
<td>5.22</td>
<td>0.44</td>
</tr>
<tr>
<td>Complex atypical hyperplasia (CAH)</td>
<td>3.84</td>
<td>7.68</td>
<td>0.85</td>
</tr>
<tr>
<td>Endometrial carcinoma (EC)</td>
<td>2.40</td>
<td>7.30</td>
<td>2.60</td>
</tr>
</tbody>
</table>

The given numbers are staining index values (SI), for description see Table 11.

3.2 Looking for underlying causes for methyltransferase protein overexpression

Since there was evidence that MT genes themselves can be regulated by promoter methylation (e.g. Huidobro et al., 2012) we designed a MS-MLPA probe set to study the promoter methylation of DNMT1, DNMT3B and EZH2 genes in our samples. The studied sites were selected from promoter areas or near to specific transcription factor binding sites. One of these sites, the DNMT3B distal promoter, showed statistically significant differential methylation between CRC tumors (Dm 0–0.46) and normal tissues (Dm 0–0.24; P < 0.001). DNMT3B promoter methylation was negatively correlated with DNMT3B protein expression in LS CRC group (the only group with variation in DNMT3B expression) with borderline significance (P = 0.079; rho -0.358).

We also investigated the promoter CpG island methylation of other key epigenetic pathway genes (DNMT3A, ΔDNMT3B, HAT1, HDAC1, HDAC2, HDAC3 and SIRT1) by bisulfite sequencing, but no differential methylation was found at these sites in cell lines (our unpublished data).

We also took advantage of methylation data previously generated on the same tumors regarding methylation-dependent miRNA silencing (Pavicic et al., 2011) and compared the miRNA loci (miR-152, miR-148a and miR-124a(1-3) known to target DNMT1, DNMT3B and EZH2, respectively) methylation results to the MT protein expressions. We did not find positive correlations between miRNA expressions in relation to the expression of their protein targets.
DISCUSSION

This study was designed to investigate the epigenetic changes in familial and sporadic colorectal, endometrial and gastric cancers. Our study series consisted of colorectal, endometrial and gastric tumors from Lynch syndrome patients, colorectal and gastric tumors from patients whose cancer occurred sporadically, and two groups of patients with familial clustering of either colorectal or endometrial tumors without specified genetic defects. At first we focused on DNA methylation patterns in separate patient groups and we then tried to elucidate the underlying causes behind the observed methylation changes.

1 DNA methylation in colorectal, endometrial and gastric cancers (I-III)

1.1 DNA methylation patterns as tumor classifiers

We examined the promoter hypermethylation of 24 established TSG loci by the MS-MLPA method. Methylation profiles showed tumor-specificity for a number of TSGs and tissue-specificity for many of those TSGs. Altogether 11 genes were differentially methylated in tumors and normal tissue samples. Three genes (APC, CDH13 and TIMP3) were commonly hypermethylated in colorectal, endometrial and gastric cancers indicating the important role of these genes in tumorigenesis in general. Otherwise methylation patterns differed and tumors from different origins could be identified by their hypermethylation patterns.

The CIMP phenotype has been associated with MSI in previously published studies (Esteller et al., 1998; Veigl et al., 1998; Goel et al., 2007). In these CIMP+/MSI cancers MSI is due to epigenetic instability inactivating key tumor suppressor genes such as MLH1 by promoter hypermethylation (Kane et al., 1997; Shigeyasu et al., 2015). We observed the TSG hypermethylation profiles to be concordant with the previously reported findings of sporadic MSI tumors displaying significantly more frequent tumor suppressor gene hypermethylation (and CIMP+ phenotype) compared to sporadic MSS tumors in CRC and GC. Our results of the CIMP and genomic instability patterns in CRC, EC and GC also comply with the classifications based on the comprehensive genetic and epigenetic characterization of the respective sporadic cancers by the Cancer Genome Atlas Research Network (2012; 2013; 2014).

Compared to sporadic tumors, tumor suppressor gene hypermethylation (and CIMP) has been studied much less extensively in familial cancers. Our results show that the hypermethylation of
tumor suppressor genes is characteristic dependent on the tumor type and family category (LS, FCCX or FSSEC) in familial cancers. The hypermethylation patterns in Lynch syndrome tumors resembled those in their sporadic MSI counterparts in CRC and GC but without MLH1 promoter hypermethylation that was often present in sporadic MSI tumors. Lynch syndrome tumors generally do not require MLH1 hypermethylation in order to induce microsatellite instability because of an inherited germline mutation in a MMR gene that is often combined with the somatic deletion of the wild-type allele (LOH) in the tumors (Hemminki et al., 1994). Only two CRC tumors from MLH1 germline mutation carriers showed low-level MLH1 methylation (Dm 0.18) which could serve as the second hit inactivating the wild-type allele (Lynch et al., 2015). The frequency (2/27) of somatic MLH1 promoter methylation in LS CRC complies with findings reported by other researchers (e.g. 3/19 in Moreira et al., 2015). Despite the fact that the CIMP+ phenotype was mostly observed in MSI tumors and with MLH1 promoter methylation in sporadic cases, we also observed CIMP+ in 50% of the microsatellite stable FCCX tumors without MLH1 methylation. This suggests that there has to be two distinct mechanisms for the development of these tumors with one following the MSS/CIMP+ pathway whereas the other one may involve some other mechanism (see the next paragraph).

Global hypomethylation can occur early in tumorigenesis and is often observed in cancers. Decrease in the genomic 5-methylcytosine content results in genetic instability and oncogene activation (Foy et al., 2015). LINE-1 is a transposable element present in over 500,000 copies in the genome (Xiao-Je et al., 2015), and it is widely used for determining the global methylation status of cells. LINE-1 methylation levels have been shown to decrease progressively from healthy tissue to adenoma to carcinoma in several studies (e.g. Park et al., 2009; Ibrahim et al., 2011). The global hypomethylation levels we observed were in agreement with previously published results and the result was statistically significant. Goel et al. (2010) reported lower methylation levels in CRCs from FCCX patients compared to other CRCs. Their report focused on tumor tissues only. To add to this, we show that interestingly, the FCCX patient group’s normal tissue samples already show a significant decrease in the genomic 5-methylcytosine content. Overall, our results confirm that the LINE-1 methylation status depends on the tissue and the MSI status and that hypomethylation may function as a field defect promoting tumorigenesis already in the normal mucosa as seen in FCCX patients. The idea that global hypomethylation may be linked to the hereditable predisposition to CRC is supported by epidemiological observations that patients with LINE-1 hypomethylation in CRC tend to have a family history of CRC (Ogino et al., 2013). Although the hypermethylation of specific tumor suppressor gene loci and the CIMP phenotype often co-existed with global hypomethylation in the same tumors, the hyper- and hypomethylation profiles
in our studies were not statistically correlated and it is unknown whether these two seemingly opposite patterns are regulated jointly or independently of each other.

1.2 Methodological aspects

Cytosine methylation can be studied by a variety of different mechanisms. Conventionally most of the applications to study DNA CpG methylation are based on either bisulfite treated DNA or methylation-sensitive restriction enzymes (Jorda and Peinado 2010) (e.g. BS, MSP and COBRA also used in validation purposes in this research; see Materials and methods) and are laborious and time-consuming. Genome-wide DNA methylome assays are expensive and also often require good quality bisulfite modified DNA.

We chose to use the MS-MLPA method for the CpG methylation studies because of the advantages of this method in our study and samples. MS-MLPA is based on a methylation-specific endonuclease, does not require DNA bisulfite conversion and works well with even short single-stranded DNA fragments since the probes only require 50–60 bp for hybridization and the ligation and digestion is performed on the MS-MLPA DNA-probe complex (Nygren et al., 2005). Using MS-MLPA, multiple target sites (up to 40) can be examined simultaneously with a single primer pair, which saves both sample and time. MS-MLPA allows the detection of specific methylated CpGs at site of interest and is ideal when you know which sites you want to study. MS-MLPA was the most suitable method for our sample collection in which DNA was mostly extracted from ancient FFPE archives since DNA extracted form paraffin blocks is fragile and usually fragmented. Each MS-MLPA probe detects the methylation status of one or two specific CpG sites only, so the CpG sites to be examined by MS-MLPA have to be chosen so that their methylation status represents that of the surrounding larger region. We used ME001B probe set (MRC Holland) for TSG methylation studies and designed new probe sets for the LINE-1 and MT promoter methylation studies and validated the MS-MLPA results in cell lines with parallel examination by BS and additionally MSP (Study I) or COBRA (Study III) of the same samples.

In some cases the samples sizes in our study were quite small so statistical significance could not be determined. In these cases however, our results provided valuable information about the trends which can be used when planning new studies and findings that can be further investigated in the future with larger sample size. Also the investigation of DNA methylation profiles with cell lines may give distorted results due to methylation changes, both global hypomethylation (~10-
30% less than in human primary cells) (Ziller et al. 2013) and increased hypermethylation of TGSs (Varley et al., 2013) observed in the long-term cultured cell lines.

2 Elucidating the background of methylation changes (I-IV)

It is not known what lies behind the DNA methylation changes in cancer. In the past it has been described as a random phenomenon resulting from the errors in maintaining methylation patterns during DNA replication and selection of tumor suppressor deficient cells in the future cell generations. But since DNA methylation seems to be a non-random event in cancer cells it is more likely that methylation errors are due to specific defects in DNA methylation or demethylation machinery resulting in certain types of aberrant methylation. We wanted to study the possible roles of different methyltransferases in relation to the observed aberrant DNA methylation. We studied the expression of DNA methyltransferases DNMT1 and DNMT3B and histone methyltransferase EZH2 in CRC and EC samples. All the studied methyltransferases were overexpressed in tumors, especially in MSI tumors, compared to their normal counterparts. Surprisingly the elevated expression of DNA methyltransferases in tumors did not show any particular correlation with TSG hypermethylation or the CIMP phenotype. However, the expression of EZH2 histone methyltransferase showed significant positive correlation with TSG hypermethylation in sporadic CRC suggesting that EZH2 overexpression can contribute to tumor development via aberrant DNA methylation in sporadic but not necessarily in familial CRC. Nosho et al. (2009) suggested that DNMT3B expression contributes to the CIMP phenotype in CRC. The recent study of Sarabi and Naghibalhossaini (2015) also shows the mRNA expression of DNMTs to correlate with global methylation levels but their study was done on CRC cell lines and the results cannot be directly applied to patient samples. In our sample series DNMT1 and DNMT3B expression levels were also higher in cancer than in normal tissues but the expression did not correlate statistically with DNA methylation changes in any of the studied patient groups.

Our data confirm previously published results in which DNMT expression changes correlated poorly with DNA methylation alterations in various tumor types (Eads et al., 1999; Ehrlich, 2006; Vallbohmer et al., 2006,) and DNA hypermethylation is mostly associated with EZH2 expression (Viré et al., 2006; Hoffmann et al., 2007). Hoffmann et al. (2007) reported that in prostate cancer, the expression changes of EZH2 are directly associated with DNA methylation changes, both hypermethylation of certain TSGs and hypomethylation of LINE-1, rather than DNMT1 or DNMT3B investigated in their study. Our findings support their finding concerning the EZH2 expression in
association to hypermethylation of tumor suppressor genes in sporadic CRC. However, our data
did not confirm any association between LINE-1 hypomethylation and methyltransferase
expressions, which might be partly due to the relatively low samples sizes. EZH2 was also found
to correlate positively with DNMT1 expression in tumors in concordance with the study of Viré et
al. (2006) where they proposed that these methyltransferases are co-operating in transcription
control systems where EZH2 serves as a recruitment platform for DNMTs.

TSG promoter methylation is shown to increase until an appropriate neoplastic state is achieved
(Nieminen et al. 2009; Valo et al., 2015). DNMT1 and DNMT3B protein expressions investigated in
this study followed the same pattern in the present endometrial hyperplasia series (from normal
endometrium to SH to CH to CAH). The DNMT expressions were systematically lower in EC than in
CAH which indicates the importance of DNA methyltransferase overexpression in the developing
neoplasm but not necessarily anymore in the developed or malignant tumor. Our findings
especially emphasize the role of EZH2 in tumorigenesis through associations with elevated TSG
hypermethylation, CIMP phenotype and DNMT1 expression. The overexpression of EZH2 and
other MTs can be one of the first events in neoplastic formation, tumor progression and
development into malignant cancer and could be used as prognostic markers in endometrial and
possibly other neoplasms. Targeted inhibition of EZH2 could offer a therapeutic target in cancer
cells for preventing the aberrant hypermethylation of other genes in the developing tumor
without altering the global methylation of the cells.

After observing elevated methyltransferase expression levels in tumors, we set out to elucidate
the causes for these findings. EZH2 (Fussbroich et al., 2011; Zhang et al., 2015b) and DNMT1 (Zhan
et al., 2015) overexpression is known to associate with cell proliferation in cancer, while the loss
of EZH2 is associated with cancer cell growth inhibition, decrease in invasion and metastasis and
inducing senescence and apoptosis. We examined the Ki-67 proliferation marker expression in the
samples and found that the expression changes of methyltransferases could not be explicitly
explained by faster proliferation in all the studied patients. Inspired by the fact that
methyltransferases themselves can be regulated by promoter methylation (Huidobro et al., 2012),
we designed a probe set targeted for DNMT1, DNMT3B and EZH2 CpG rich areas with specific
transcription factor binding sites for DNA methylation studies. Only one of these probes, at
DNMT3B distal promoter, displayed differential methylation between tumor and normal samples
and the methylation status was indeed associated with protein expression. Based on our findings,
DNMT1 and EZH2 are apparently not regulated by promoter methylation in human primary
samples. Several studies have shown that certain miRNAs affect the expression of the studied

64
methytransferases. MiR-152 targets DNMT1, miR-148a DNMT3B and miR-124a(1-3) EZH2 according to published literature (Braconi et al., 2010, Duursma et al., 2008, Zheng et al., 2012, respectively). We took advantage of our existing data on methylation-dependent miRNA silencing (Pavicic et al., 2011) but found no association between the miRNA methylation and the methyltransferase gene expression. However, several miRNAs target one gene so this only confirms that the studied miRNAs were not powerful enough to regulate the expression of these methyltransferases alone.

3 Overall conclusion of the findings (I-IV)

Overall, this research provided new information about the molecular epigenetic phenomena in cancers, especially between familial and sporadic cancers. Tissue type, MSI status and the familial background of the cancer determined the frequency of TSG hypermethylation mirroring different developmental pathways in different cancers. According to our results, TSG hypermethylation seems to serve as the cause of MSI in many cases of sporadic CRC and was statistically correlated with EZH2 overexpression. Cancer in LS cases was evidently due to the identified mutations of the patients, but the accompanying TSG hypermethylation may fasten the cancer development in these patients by affecting essential cellular pathways and in some cases serve as the second-hit alongside with the inherited mutation. APC (Sawa et al., 2015), CDH13 (Zhong et al., 2015) and TIMP3 (Guan et al., 2013) are well known to be inactivated in various cancers and are related to tumor growth, angiogenesis, invasion and metastasis. In our study, these genes were commonly hypermethylated regardless of the tissue emphasizing the importance of the silencing of these genes by promoter hypermethylation in the tumorigenesis of all the studied tissue types included in this research. Global hypomethylation was evident in tumors vs. normal tissue samples and blood in a tissue-specific manner. There are conflicting results regarding the role of global hypomethylation observed in peripheral blood (Brennan and Flanagan, 2012), but our results suggest that it could be one of the first visible evidence of cancer initiation as seen in the FCCX patients in our study series.

DNA methyltransferase DNMT3B was overexpressed in almost all of the tumors but direct correlations could not be seen between the level of expression and TSG hypermethylation status of the samples. On the contrary, histone methyltransferase EZH2 overexpression correlated with the CIMP+ phenotype suggesting its role in the aberrant tumor suppressor gene hypermethylation of sporadic tumors. Recent evidence show the potential effect of EZH2 inhibition in treating
human melanoma cell lines harbouring activating mutation in EZH2 (Tiffen et al., 2015) resulting in reactivation of repressed TSGs in these cells. DNMT1 expression correlated positively with EZH2 expression suggesting co-operative functions of these methyltransferases in cells. Unexpectedly the aberrant hypermethylation patterns in tumors did not follow the DNA methyltransferase expression levels. Future research should be directed to study more closely the function of the mechanisms directing DNMTs to DNA for target sequence methylation or marking the sites for de novo methylation. The activities of DNA demethylases and passive demethylating mechanisms should be studied more carefully in the context of global hypomethylation and cancer. Our results show that aberrant methyltransferase expression, notably EZH2, and DNA methylation changes can be detected already in early tumorigenesis in different tumors and may therefore have diagnostic potential.

Methylation and other epigenetic changes can be reversed, and certain genes have been reactivated in cancer using demethylating agents (De Carvalho et al., 2012; Wu and Zhang, 2014). At present, demethylating agents, 5-azacytidine (Falchook et al., 2013) and decitabine (Garrido-Laguna et al., 2013), are used to treat different cancers and there are at least 20 histone deacetylases in clinical testing (Yan et al., 2015) such as panobinostat (Bauer et al., 2014) and valproic acid (Falchook et al., 2013). Especially the targeted reactivation of TSGs commonly inactivated by promoter hypermethylation in various cancers, such as APC, CDH13 and TIMP3 also observed to be hypermethylated in our study, could be useful in treating different types of cancers. At present, there is no established therapeutics targeting the Wnt signaling pathway (Sawa et al., 2015). Demethylating treatments have been effective in some cancers, but it is difficult to target demethylating treatments to specific gene promoters only without affecting the whole genome methylation levels of the cell. Targeting the levels of methyltransferase proteins responsible for the aberrant hypermethylation might provide some more selectivity. Our study confirms that EZH2 is one possible target since the EZH2 protein levels were positively correlated with tumor suppressor gene hypermethylation in the present study and also with DNMT1 expression which can be responsible either for de novo methylation in the tumor cells or for loss of imprinting that promotes accumulation of methylation defects in cancer cells (Biniszkwiewicz et al., 2002; Jelinic and Shaw, 2007). Some EZH2 inhibitors (e.g. EPZ-6438) are already currently being clinically tested (Yan et al., 2015).

Moreover, it is important to distinguish between driver and passenger effects of different genes in cancer (De Carvalho et al., 2012; Marx, 2014). Currently oncogenes are the major molecular target for anti-cancer drug design and studies exist showing that for example the restoration of a
caretaker tumor suppressor gene function in a cancer cell does not affect its growth (Kinzler and Vogelstein 1998). That has to be kept in mind when planning new molecular therapies. Based on our results, for example patients with sporadic CRC could benefit from the combined treatment with a demethylating agent and an EZH2 inhibitor.

This study provided information on the epigenetic events, hyper- and hypomethylation in different types of cancers. Hypermethylation of certain genes can be used as a marker for the early detection of cancer. By studying prostate biopsies Paziewska et al. (2014) similarly noted that DNA methylation status of specific genes is a reliable marker of cancer detection and may be even more reliable than gene expression investigated by the most promising mRNA-based prostate cancer biomarkers. Tumors of different origin can be identified by their methylation profiles as is most evident for tumors displaying CpG island methylator phenotype. Alongside with genetic profiling, epigenetic profiling can be used to identify the tissue of the primary tumor in the case of a metastasis of an unknown origin and treatment plan can be designed accordingly. Observed global hypomethylation patterns suggest that hypomethylation in normal tissue can promote tumorigenesis, at least in mutation negative MMR proficient tumors and the decrease of methylation can be observed already in the blood and normal mucosa samples. This information may be used as a tool in cancer diagnosis. Molecular classification of tumors is essential for determining rational cancer treatment. Cancer is a heterogeneous disease and develops through multiple different pathways. Aberrations in the functions of these pathways can be detected by genetic and epigenetic profiling. When genetic and epigenetic profiles of a tumor are recognized the information can be utilized in counseling and planning of targeted treatment for each individual patient based on the defects of the tumor. Our FCCX tumor global hypomethylation results suggest that the treatment should be targeted to restoring normal methylation patterns to prevent genomic instability caused by DNA hypomethylation. Sporadic tumors with CIMP+ phenotype and MSI could benefit from treatment with EZH2 inhibitors and demethylating agents, while LS tumors could be treated using demethylating agents in order to restore TSG functions or MMR function in cases where MMR gene promoter hypermethylation provokes the loss-of-function of the wild-type allele. Molecular profiling can provide opportunities for the development of new screening methods allowing earlier diagnosis and thereby improved prognosis. Finally, the reversibility of epigenetic states is especially interesting from the perspective of drug development.
CONCLUSIONS & FUTURE PROSPECTS

Cancer is a very heterogeneous disease. Even though the genetic basis of Lynch syndrome is well established, the knowledge of the roles of different epigenetic phenomena in this syndrome is still lacking. Also little is known about the background of cancer initiation and the progression of microsatellite stable familial tumors (FCCX and FSSEC). These tumor types, as well as sporadic cancers, would benefit from a deeper molecular characterization of the tumors and the underlying processes in order to provide more adequate diagnostic tools and treatment.

In this research we focused on certain types of DNA methylation changes and the methyltransferases directly affecting DNA methylation in sporadic and familial cancers. Next it could be informative to investigate more closely all the genes on DNA methylation pathways in our sample collection for possible mutations in normal and tumor tissues and more intensively study the global methylation profiles in these tumors using different genome-wide methods. It would also be beneficial to investigate the degree of promoter methylation necessary for carcinogenesis more closely. Furthermore, the allelic origin of methylation should be determined using single-cell analysis methods to see if the methylation events affect both of the alleles or just a single allele of a given gene.

Methyl CpG binding domain proteins (MBDs) e.g. MECP2, MBD1, MBD2, MBD3 and MBD4 are capable of specifically binding methylated DNA. Inactivated MBDs can contribute to tumorigenesis by disturbing the transcriptional repression and allowing the expression of oncogenes (Du et al., 2015). For example MBD4 interacts with the DNA mismatch protein MLH1 with DNA repairing abilities (Grigera et al., 2013) and is associated with microsatellite instability and MMR defects in sporadic cancer. MBD4 deficiency can also cause global hypomethylation by increased C to T mutations (Chen et al., 2011). The possible MBD alterations and DNA methylation and transcription interactions would be interesting to study especially in MLH1 deficient Lynch syndrome patients and FSSEC cases of which 48% in our samples are MMR deficient without identified alterations in MMR genes (MLH1, MSH2 or MSH6).

The entity of FCCX still remains a mystery. The patients fulfil the criteria for HNPCC, but without identifiable MMR defects or microsatellite instability. The main aberrations found in this patient group are the CIMP+ phenotype without MLH1 methylation in 50% of the cases and a lower global methylation status. Germline mutations in BMPR1A (Nieminen et al., 2011) and RPS20 (Nieminen et al., 2014) were recently found in few families in the studies of our research group by genome-
wide linkage scan and next-generation sequencing analyses; methylation studies of *BMPR1A* and *RPS20* promoter and enhancer areas as well as related genes might shed light on the mechanisms of tumorigenesis in these patients.

Transcriptional gene regulation is a complex process and besides promoter CpG islands other sequences can affect transcription as well. The regulatory regions of a specific gene can be located thousands of kilobases up- or downstream of the designated promoter CpG island. Characterization of distal enhancers and silencers is another challenge in the investigation of the mechanisms and effects of different epigenetic events on mRNA or protein expression. Recent studies have also shown that histone modification-directed chromatin organization is closely related to mutation rates in CRC (Dominiguez-Valentin et al., 2015).

Overall, the emerging knowledge of different epigenetic marks by writers, presenters, readers and erasers provides an interesting field of study. A better knowledge of the epigenetic phenomena and interactions between different epigenetic pathways is necessary to understanding cancer development and progression. Epigenetic research has become a hot topic among scientists during the past ten years and the rate of new discoveries is increasing. Recent discoveries include e.g. the elucidation of the functional roles of DNA demethylation intermediates (Ito et al., 2011), 5-hydroxymethylcytosine (5hmC) (Severin et al., 2013) and 5-formylcytosine (5fC) (Bachman et al., 2015), in gene transcription control. 5hmC levels have been found to associate with stem cell pluripotency (Song and He, 2013) and can thus also be important in cancer in context of differentiation. The study of the mechanisms and functions of these bases is on the way to be extended to cancer research. The functions of different noncoding RNAs are still largely undescribed and we remain unaware of the functions of e.g. different histone modifications. These for example, are areas in which there is supposedly lot to be discovered in the field of cancer. These are at present under continuous study with new effective research methods emerging in the field. CpG dinucleotides are the most abundantly methylated and studied in mammalian genomes. Epigenetic cancer research is also expanding to investigations of regulatory capacities of other methylated dinucleotides (Oster et al., 2013; Kinde et al., 2015; Yamazaki et al., 2015). The research could be extended to study the possible functions of other DNA modifications that are still uncovered.

Epigenetic cancer research is now becoming easier by the development of new efficient methods by which epigenetic DNA modifications can be easily detected and separated. Also the functional properties and interactions of for example alterations of different histone marks can now be studied in living cells. While conventional direct genomic sequencing is still useful in detecting
mutations or DNA methylation after bisulfite conversion of candidate sequences, 2nd generation (or next-generation) amplification-based massively parallel high-throughput sequencing technologies are useful in whole genome, exome or methylome studies using bisulfite applications and contribute to genome-wide sequencing based personalized medicine. 3rd generation sequencing methods based on the sequencing of single-molecules (Bheda and Schneider, 2014), of which nanopore sequencing seems at the moment to be the most promising, can in addition to normal base detection, distinguish between methylated and hydroxymethylated cytosines and other modified or damaged bases and hence gives accurate results directly without error-prone amplification steps (Wang et al., 2015). In our study of TSG methylation we concentrated only on promoter areas of the selected genes. Whole-genome methylation analyses would give more information about the overall methylation changes in different genes and in important control regions such as enhancers as well as overlapping, intergenic or intragenic IncRNAs and miRNAs (Li et al., 2015). To study the clonal evolution of the tumors, single-cell analyses (Bheda and Schneider, 2014) could have been used but because it requires fresh tissue samples it was not possible in our samples.

Cancer processes can be studied and therapies tested using the CRISPR/Cas9 system which is an RNA guided site-specific genome-editing tool for studying massive functional interactions in living cells (Lu et al., 2015). CRISPR/Cas9 enables the quick modeling of different combinations of epigenetic and genetic cancer-related aberrations simultaneously or sequentially (Sander and Joung, 2014; Zentner and Henikoff, 2015). By this method the epigenome can be manipulated at individual loci and for example fusions of Cas9 to histone-modifying enzymes could allow the functional characterization of epigenome. The development of such new technologies and databases of functional elements in the human genome, such as The Encyclopedia of DNA Elements (ENCODE, 2015) can make the life of an epigenetics researcher much easier and epigenetics research much more powerful in the future.
ACKNOWLEDGEMENTS

First of all, I want to thank my supervisor Päivi Peltomäki who has been such an inspirational character on my academic and scientific journey. Thank you for this great opportunity to work as a part of your research team on the most interesting field of the cancer research ever, the epigenetics. Thank you for your time and you patiently supervising my research and my thesis.

This work was carried out at the Department of Medical and Clinical Genetics, Faculty of Medicine, at the University of Helsinki, as part of GPBM/ILS doctoral programs, which are thanked for financial and personal support (especially Erkki Raulo and Anita Tienhaara, thank you for all your help and guidance during these years), as well as The Medicine Fund of UH, Finnish Cancer Foundation, Orion Research Foundation and Oskar Öflunds Stiftelse which are acknowledged for funding this research project. I also want to thank the members of my thesis committee, Hannele Laivuori and Hannes Lohi for reviewing my progress during these years.

I also want to thank the reviewers of my thesis, Anu Loukola and Rainer Lehtonen, for your time and valuable comments and suggestions for improving my thesis. MKK is thanked for the all the help with the language revision.

Thanks to all the former and current members of our research group: Saila for the help, technical expertise and knowing everything in the lab. Rest of the people in the order of appearance: Annette, Miina, Wael, Taina, Sippy, Walter, Johanna, Anni, Satu, Oona and Noora it has been a pleasure to work with you guys, thanks for the help and of course for your contributions to my articles. And thank you very much for the encouragement during the last months, I really appreciate it.

Special thanks to the beautiful people I met during these years, Annette and Sippy, thank you for all the (+/- scientific) crazy people’s discussions in the office. Minttu K for your friendship but also for accompanying me on all the grad school courses and finally helping me to finalize this thesis, Laura V for always being there (from the academic day zero, love you cutie!), Jaana especially for the help in the statistics but also for discussions about life in general and Nina for being an inspiration and an example of scientific career building.

In addition to above mentioned (you also belong here), Isabel and Tiina P (my mental sisters), Elsa, Denis, Jukka and Maral thanks for all the superb conversations and sharing (and caring) all the science related (and other) pain and love (and beer). Aarno for your help especially in the magical
computer stuff but also other. And all my other fellow grad students sharing your thoughts and moments during the studies. Thank you all for your friendship and for encouraging me to do this job.

Thank you Minttu S (my partner in crime since the childhood), Pia and Miisa (for everything and love) and the team AMA for your stress-relieving properties during the hard times. Christian (for being my PT/PH), all the fellow trainer girls at the VC (no pain, no gain!), Emilia, Henna, Nata, Sara, for all the funny moments and caring. I got my thesis ready now! :) Jori, Olga and Iida, you are wonderful people and the best siblings I could have <3.

Äiti and Iskä, thanks for making me, I kind of like the genetic profile I got and I like to exist! :) And thanks for making my sisters and baby brother too, I like them existing as well. Mummi and Vaari, you are the best fairy people I’ve ever known. <3 Thank you for your everlasting encouragement and for understanding and accepting all the decisions I make.

Juho, thank you for understanding and being there, especially during the last months (years). I know I have been a pain in the ass with all my (important) stuff. My kids, Alvar, Jooseppi and Iivari are not to be thanked since they only got my thesis writing almost impossible at the times, but I love you, till the sun and back, as the little Pokémon you are!
REFERENCES


Cai X, Hagedorn CH, Cullen BR. 2004. Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. RNA 10:1957-1966.


76


ENCODE. 2015. www.encodeproject.org


Okano M, Bell DW, Haber DA, Li E. 1999. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell 99:247-257.


