DNA methylation as a prognostic marker in clear cell Renal Cell Carcinoma

Emma Andersson Evelönn

Department of Medical Biosciences
Department of Surgical and Perioperative Sciences
Umeå 2020
Framgångens hemlighet är att aldrig släppa målet ur sikte.

Till Hugo och Hanna, ni är det bästa i mitt liv
# Table of Contents

Abstract ..................................................................................................................................... iii

Populärvetenskaplig sammanfattning ................................................................. iv

Abbreviations ......................................................................................................................... vi

## Introduction .......................................................................................................................... 1

Clear Cell Renal Cell Carcinoma ......................................................................................... 1
Genomic traits in ccRCC ........................................................................................................ 4
   Genomic aberrations ........................................................................................................... 4
   Common mutations in ccRCC ............................................................................................. 4
Epigenetics ............................................................................................................................. 5
   Histone modifications ........................................................................................................ 7
   RNA epigenetic modifications ............................................................................................ 8
   DNA methylation ............................................................................................................... 9
   DNA methylation in cancer ................................................................................................. 10
   DNA methylation in ccRCC ............................................................................................... 12
   Methylation biomarkers for ccRCC .................................................................................. 12

## Aims ...................................................................................................................................... 14

Paper I ..................................................................................................................................... 14
Paper II ................................................................................................................................... 14
Paper III .................................................................................................................................. 14

## Materials and Methods ...................................................................................................... 15

Workflow describing patients and methods included in this thesis .................................... 15
Patients and tissue samples .................................................................................................... 15
   Tissue samples included in paper I .................................................................................. 16
   Tissue samples in papers II and III .................................................................................. 16
   Tumor samples from the Tumor Cancer Genome Atlas (Paper II) .................................... 16
   Blood analysis .................................................................................................................... 16
DNA extraction ...................................................................................................................... 16
Methylation analysis ............................................................................................................. 17
   Bisulfite conversion (Papers I-III) .................................................................................... 17
   Methylation array analysis (Papers I-III) ....................................................................... 17
   Data preprocessing .......................................................................................................... 18
   Array data analysis .......................................................................................................... 18

The building of prognostic classifiers ..................................................................................... 18
   Promoter Methylation Classification (PMC) Panel (Paper II) ......................................... 19
   Triple Classifier (Paper III) ............................................................................................. 19
Genetic analysis ..................................................................................................................... 20
   Genetic aberrations (Papers I and II) .............................................................................. 20
   VHL status (Paper I) ........................................................................................................ 20
Gene expression analysis ....................................................................................................... 20
   RNA preparation (Paper II) ............................................................................................. 20
   Gene expression arrays (Paper II) ................................................................................... 20
Statistical analysis........................................................................................................................................21

Results .................................................................................................................................................. 22
  Demographic overview of the ccRCC cohorts......................................................................................22
  Loss of chromosome arm 3p and VHL genetic and epigenetic status (Paper I)......................23
    Methylation analysis of the VHL promoter (Paper I)...................................................................23
  Genomic aberrations and ccRCC disease progression (Paper II)..............................................24
  DNA methylation classification and its correlation to clinicopathological parameters
    (Papers I and II)..........................................................................................................................25
    Methylation alterations associated with tumor progression (Paper I).................................27
    A stepwise increased DNA methylation level during tumor progression (Papers I and II).....28
    Inter- and intra- heterogeneity in DNA methylation profiles (Paper II)...............................28
  Methylation and genetic aberration and their correlation with biological and mitotic
    age (Papers I and II)....................................................................................................................29
  New prognostic classifiers in ccRCC (Papers II-III)................................................................30
    Promoter Methylation Classifier (PMC) panel (Paper II).....................................................30
    Genes represented in the PMC panel..........................................................................................31
    Triple Classifier (Paper III)..............................................................................................................32

Discussion ......................................................................................................................................... 34
  General conclusions and future perspectives..............................................................................40

Acknowledgment ..........................................................................................................................42

References ..........................................................................................................................................45
Abstract

Clear cell renal cell carcinoma (ccRCC) is the most common type of renal cell carcinoma worldwide. Metastatic ccRCC is correlated to poor prognosis whereas non-metastatic disease has a 5-year survival rate up to 90%. Due to increased accessibility to different types of diagnostic imaging the frequency of metastatic ccRCC at diagnosis has decreased since the beginning of the 21st century. This has led to an earlier detection of primary tumors before patients present symptoms. However, 20-30% of the non-metastatic patients at diagnosis will progress and metastasize within five years of primary nephrectomy. Identifying patients at high risk of tumor progression at an early stage after diagnosis is of importance to improve outcome and survival. Currently, in Sweden, the Mayo scoring system is used to divide tumors into low, intermediate or high risk for tumor progression.

DNA methylation has been associated with tumor development and progression in different malignancies. In this thesis, Illumina Infinium HumanMeth27 BeadChip Arrays and Human Meth450K BeadChip Arrays have been used to evaluate the relationship between methylation and clinicopathological variables as well as ccRCC outcome in 45 and 115 patients.

Our studies identified an association between higher level of promoter-associated DNA methylation and clinicopathological variables in ccRCC. There was a significant stepwise increase of average methylation from tumor-free tissue, via non-metastatic tumors to metastatic disease. Cluster analysis divided patients into two distinct groups that differed in average methylation levels, TNM stage, Fuhrman nuclear grade, tumor size, survival and tumor progression. We also presented two prognostic classifiers for non-metastatic tumors; the promoter methylation classifier (PMC) panel and the triple classifier. The PMC panel divided tumors depending on the methylation level, PMC low or PMC high, with significantly worse prognosis in the PMC high group. This data was verified in an independent, publically available cohort. The triple classifier was created using a combination of clinicopathological variables, previously identified CpGs biomarkers and a novel cluster analysis approach (Directed Cluster Analysis). The triple classifier had a higher specificity compared to the clinically used Mayo scoring system and predicted tumor progression with higher accuracy at a fixed sensitivity.

The identification of two epigenetic classifiers that predicted outcome in non-metastatic ccRCC further establishes the role of DNA methylation as a prognostic marker. This knowledge can contribute to identification of patients with a high risk of tumor progression and can be of importance in the decision regarding adjuvant treatment post-nephrectomy.
Populärvetenskaplig sammanfattning

Klarcellig njurcancer är den vanligaste form av njurcancer. I Sverige diagnostiseras ca 1 000 individer årligen med sjukdomen. Idag upptäcks ofta njurcancer när patienter undersöks med bilddiagnostik av buken av andra anledningar, exempelvis vid oklar buksmärta och vid trauma. Detta gör att tumörerna upptäcks innan de hunnit ge symtom och andelen patienter med spridd sjukdom vid diagnos är under 20%, jämfört med över 30% i början av milleniet.

Den enda botande behandling som finns för njurcancer är total eller partiell nefrektomi, vilket innebär att hela njuren eller delen av njuren med tumör opereras bort. Om sjukdomen upptäcks tidigt, innan den hunnit sprida sig till kringliggande organ, är prognosen god och 90% av patienterna lever minst fem år efter diagnos. År sjukdomen spridd vid diagnos finns det idag behandlingar som förlänger överlevnadstiden, och ny immunbehandling som också kan bota sjukdomen.

Även om klarcellig njurcancer som är begränsad till njuren botas när tumören avlägsnas är det ungefär en tredjedel av dessa patienter som drabbas av en ny njurtumör eller spridning till andra organ inom fem år efter diagnos. Behovet av att identifiera dessa patienter är stort när de är i behov av tilläggsbehandling. Idag utgår klinikerna från hur tumören ser ut när de bestämmer hur patienterna ska följas efter kirurgi, vid karakteristika som talar för en hög risk för spridning följs patienterna tätare och under längre tid.

Genom att analysera hur ccRCC tumörer ser ut på cellnivå har skillnader i genuttryck och DNA-sekvenser kunna identifieras. Att utnyttja dessa skillnader är viktiga för att identifiera tumörer med ökad risk för spridning.

I min avhandling har vi analyserat DNA metylering i klarcellig njurcancer vid diagnos. DNA metylering är en epigenetisk förändring, en förändring på DNA nivå som inte påverkar DNA sekvensen men kan påverka vilka gener som uttrycks i cellerna. Nivåerna av DNA metylering skiller sig mellan olika prognostiska grupper ccRCC. De tumörer som vid diagnos redan spridit sig till kringliggande organ har en högre grad av metylering jämfört med de tumörer som växer endast i njuren. Vi har även kunnat visa på skillnader mellan de lokala tumörer som senare sprids. Detta gör att DNA metylering kan användas som prognostisk markör i ccRCC.
Genom att utnyttja skillnaderna i DNA metylering har vi kunnat identifiera två olika sätt att förutspå risk för spridning av lokal tumörsjukdom vid diagnos. Först utnyttjade vi likheterna i metylering mellan de tumörer som är spridda vid diagnos och de som spröds med tiden. Vi skapade med detta en ‘Promoter Methylator Classifier (PMC) panel’. Denna metod kan användas för att bedöma risk för spridning och vi kunde verifiera resultaten i ett stort ccRCC tumörmatrial från USA.

Vår andra diagnostiska modell byggde på en kombination av DNA metylering och kliniska variabler (analys av blodprov och kunskap om tumörens utseende). Genom att kombinera skillnader i metylering tillsammans med de kliniska variabler, som är kända att öka risken för tumörspridning, kunde vi ytterligare förbättra klassificeringen, den metoden kallades för ‘the triple classifier’.

Båda modellerna kunde identifiera 85% av tumörerna som spreds inom fem år efter diagnos, det är identiskt med den metod som används vid klinisk klassificering av ccRCC, Mayo system. Största skillnaden mellan våra båda kliniska modeller och Mayo låg i antalet som bedömdes ha en ökad risk för spridning men sedan inte drabbas. För Mayo låg denna på 50%, för PMC panelen på 45% och slutligen för the triple classifier på 36%. Detta indikerar att vår triple classifier var bäst på att både identifiera patienter med ökad risk för tumörspridning utan att därför fånga in en alltför stor andel ”friska” som bedömdes ha en ökad risk.

Sammanfattningsvis så har vi i denna avhandling visat att DNA metylering är en stabil epigenetisk förändring som kan korreleras till risk för spridning vid klarcelllig njurcancer. Genom att utnyttja DNA metylering med eller utan kliniska variabler har vi kunnat bygga två prognostiska modeller för att bedöma spridningsrisken av lokalisert njurcancer vid tidpunkten för diagnos. Ett framtida mål, är att vidareutveckla dessa modeller för att ingå i vårdprogrammet för njurcancer och förbättra livskvaliteten hos dessa patienter.
Abbreviations

CIMP – CpG Island Methylator Phenotype
CIP – Cumulative Incidence of Progress
CpG – Cytosine-phosphate-Guanine
CSS – Cancer-Specific Survival
ccRCC – clear cell Renal Cell Carcinoma
CNV – Copy Number Variations
DCA – Directed Cluster Analysis
HRP – High Risk for Progress
LRP – Low Risk for Progress
Mo – Non-metastatic tumor at diagnosis
Mo-PF – Non-metastatic tumor at diagnosis that will not progress, ‘true’ Mo.
Mo-P – Non-metastatic tumor at diagnosis that later will progress.
M1 – Metastatic tumor at diagnosis
PCA – Principal Component Analysis
PI-CpGs – CpGs previously shown to be associated with ccRCC
PMC – Promoter Methylation Classifier
PFS – Progression-Free Survival
RCC – Renal Cell Carcinoma
T-tissue – tumor tissue
TF-tissue – tumor-free tissue
Original Papers

List of original papers included in this thesis:

I. DNA methylation status defines clinicopathological parameters including survival for patients with clear cell renal cell carcinoma (ccRCC).

**Evelönn EA, Degerman S, Köhn L, Landfors M, Ljungberg B, Roos G**

*Tumor Biol. (2016) 37:10219-10228*

II. DNA methylation associates with survival in non-metastatic clear cell renal cell carcinoma.


*BMC cancer (2019) 19:65*

III. Combining epigenetic and clinicopathological variables improves prognostic prediction in clear cell Renal Cell Carcinoma


*Manuscript*

Reprints were made with permission from the respective publishers.

**Contributed equally**
Published articles not included in the thesis:

- Wilms’ tumor 1 can suppress hTERT gene expression and telomerase activity in clear cell renal cell carcinoma via multiple pathways.
  
  
  *Br J Cancer*. **2010 Oct 12;103**(8):1255-62

- Immortalization of T-cells is accompanied by gradual changes in CpG methylation resulting in a profile resembling a subset of T-cell leukaemia.
  
  
Introduction

Clear Cell Renal Cell Carcinoma

Kidney cancer is the 14th most common type of cancer worldwide and 8th most common in the western world (1). Renal cell carcinoma (RCC) accounts for approximately 80% of kidney cancers and the most common subtype is clear cell RCC (ccRCC). In Sweden, approximately 1 200 individuals are yearly diagnosed with kidney cancer (2). Over 80% of patients diagnosed with ccRCC have localized disease at diagnosis and therefore have a good prognosis. Out of these, 30% will later be burdened with tumor progression. One of the challenges today is to identify these patients and give them the best outcome and quality of life.

The incidence of RCC differs between countries, the highest incidence is seen in the Czech Republic and the lowest in China, Thailand, and African countries (2,19 vs. <2 per 100,000, age-standardized rate for males) (3). However, differences between parts of countries and between ethnicities within countries are also seen (1,3). RCC is more common in men than in women (3).

In a review by Scelo and Larose (2018) a list of risk factors for kidney cancer were evaluated (3). Tobacco smoking, excess body weight, hypertension, and chronic kidney disease are the most common and well-defined risk factors (3). Hereditary traits have also been observed for kidney cancer, with 3-5% of cases occurring within families (4). For ccRCC, there are at least six different syndromes associated with the disease, with von Hippel Lindau (VHL) disease as the most common (1,4,5).

The classic triad of symptoms associated with kidney cancer, including ccRCC, are flank pain, haematuria, and a palpable abdominal mass. However, these symptoms often occur late in tumor development. Other diffuse symptoms of ccRCC are paraneoplastic syndromes with hypercalcemia, unexplained fever, erythrocytosis and/or Stauffer’s syndrome with cholestasis unrelated to tumor infiltration of the liver. Today, the diffuse symptoms associated with kidney cancer in combination with more extensive use of diagnostic imaging in the clinic have led to over 50% incidentally detection of kidney cancer. Patients that seek medical attention due to abdominal symptoms often get examined with computed tomography (CT), ultrasound or magnetic resonance imaging (MRI) and small tumor masses on the kidneys can then be detected even though they are not the cause of the problem.
Early detection of cancer is of importance since the opportunity to treat with curative intention decreases with tumor progression. In ccRCC small tumors located in the kidney without spread to regional lymph nodes, vena cava and/or distant organs have the best prognosis. Localized ccRCC has a 5-year survival rate of up to 90% whereas if metastasized the prognosis is poor with a survival rate of 10%. The high frequency of incidentally diagnosed tumors has led to a decrease in metastatic disease, from 30% at the beginning of the 21st century to 18% around the year 2010 in Sweden. At diagnosis, all patients are treated according to a standardized process described in the Swedish Health Care Program for kidney cancer (2).

The diagnosis is set for each patient, based on results from CT scans and later from histological examination of the dissected tumor. From CT scans tumor diameter is measured, the localization and growth pattern of the tumor is determined and involvement of lymph nodes, signs of tumor growth into the renal veins and signs of distant metastasis are also analyzed. This information is then used to stage tumors according to the TNM classification; Tumor (T), Node (N) Metastasis (M) (Figure 1) (6).

Non-metastatic disease is treated with either partial or total nephrectomy. Small (<7 cm in diameter) tumors can be treated with partial nephrectomy (PN) to save the non-affected kidney tissue and thereby spare kidney function. If the tumor is locally advanced or if the tumor is disadvantageously located in the kidney, radical nephrectomy (RN) is performed. Local lymph nodes and the adrenal gland are vacated if there are signs of involvement from CT scans or during the clinical evaluation in surgery. The adrenal gland is saved if there is no sign of tumor overgrowth (2).

Figure 1. Simplified representation of Tumor, Node, Metastasis (TNM) staging in ccRCC. **TNM I:** Small tumor (T), <7 cm in diameter, limited to the kidney; **TNM II:** Larger tumor, > 7 cm in greatest diameter, limited to the kidney; **TNM III:** Tumor of any size that involve major veins and/or adrenal glands, limited by Gerota’s fascia surrounding the kidney and might involve one lymph node (N); and **TNM IV:** Tumor growing beyond Gerota’s fascia or involve more than one lymph node or have distant metastasis (M).
If the patient has metastatic disease at diagnosis, the best possible treatment plan is finally determined by a team composed of a urologist, oncologist, radiologist and sometimes a pathologist. Some patients benefit from nephrectomy even though it is not with a curable intention but to prolong survival and quality of life. These patients might also benefit from systematic treatment and/or surgical removal of metastasis (2).

After nephrectomy, tumors are analyzed by pathologists to determine the type of kidney cancer the patient is burdened by. In the case of RCC, there is further subclassification into clear cell (70-80%), papillary (10-15%), chromophobic (approx. 5%) or one of the more uncommon collecting duct carcinoma or medullary carcinoma that constitute for 0.5% of RCC cases each. If evaluated as a ccRCC the tumor is staged according to Fuhrman grade (7), level of sarcoïd tissue, presence of necrosis and vessel infiltration. The pathologist also evaluates the tumor size and determine if there are abnormalities in the surrounding tissue to verify that the tumor has been radically removed.

Even if a patient has non-metastatic disease at diagnosis, approximately 30% of patients will later develop metastasis. Tumor progression is most common within three to five years post-nephrectomy with 80-90% tumor progression events occurring within this time period. Therefore, these patients are scored according to the Mayo Scoring System that includes information of the primary tumor (i.e. T-stage), tumor size, involvement of regional lymph nodes (i.e. N-stage), Fuhrman Grade, and the presence/absence of necrosis (8). Depending on the Mayo score, patients are divided into three groups with different risk of tumor progression; low, intermediate or high risk. Post-surgical follow up differs between these risk groups, with more frequent and longer follow up in the intermediate and high-risk groups in comparison to the low-risk group. At the moment, only one study indicates that adjuvant treatment is beneficial in non-metastatic ccRCC before tumor progression. It is therefore, not recommended to enroll patients without progress in adjuvant treatment (4,9,10).

In contrast to non-metastatic ccRCC, treatment for metastatic ccRCC has shown to be beneficial throughout the years with prolonged overall survival when combining nephrectomy and/or systematic treatment. Radiation and chemotherapy have low or no effect in treating ccRCC whereas the development of systematic treatment with tyrosine kinase inhibitors (TKI), mTOR-inhibitors and immunological treatment have been of importance to prolonged survival (11-13).
Genomic traits in ccRCC

Genomic aberrations and mutations are present in ccRCC tumors in a heterogeneous pattern and at least 90% of tumors show at least one genetic alteration (14-16). Even though several different copy number variations and genetic mutations are published, only a small number of them are common within the ccRCC population. Moreover, few studies have shown a correlation between genetic aberrations and aggressiveness or outcome in a reproducible matter (17).

Studies investigating intratumoral heterogeneity have shown that the genetic landscape can differ widely depending on where in the tumor the sample is taken from (17,18).

Genomic aberrations

The most common genomic aberration in ccRCC is the allelic loss of the chromosome arm 3p, which is present in approximately 90% of ccRCC tumors (15,19). Other genetics aberrations are presented in variable frequencies; loss of the whole chromosome 4, 9, 19, 20 and 22, as well as losses of regions within 1p, 4p, 4q, 9p, 9q, 13q, and 14q, have been reported. Chromosomal gains are reported in less quantity with a gain of chromosome arms 1q, 3q, 5q, 7q, 8q, and 20q, as well as gain of whole chromosome 5, as the most commonly amplified regions (17,20-25).

Köhn et al., (2014) analyzed a subset of ccRCC patient samples included in this thesis for genomic aberrations and identified twelve genomic regions altered in at least 20% of samples (22). Loss of chromosome 3p was the most common alteration present in 88% of the ccRCC samples. The gain of a region on chromosome arm 7q and a loss of regions on chromosome arms 9p and 9q were all correlated to poor outcome (22).

Common mutations in ccRCC

The most common genetic mutations present in ccRCC are that of VHL, PBRM1, SETD2, BAP1, KMD5C, TERT, PTEN, mTOR, and TP53 genes, which are present in a range from 2% (PTEN and TP53) to 90% (VHL) of cases (14,16).

The VHL, PBRM1, SETD2, and BAP1 genes are all located on chromosome 3p. One allele of the VHL gene is lost as an early event during tumor development, whereas the silencing of the other allele occurs later in tumor progression via either a frameshift mutation or epigenetic alterations (14). Correlation between VHL gene inactivation and clinical outcome has not been reported (5).
The *PBRM1* gene encodes for a subunit in the nucleosome remodeling complex and its mutation rate is reported in up to 50% of ccRCC tumors (26). Mutations in *PBRM1* can result in abnormal/malfunctioning gene product resulting in unchecked cell growth (27). *SETD2* encodes for a histone H3K36 lysine trimethyltransferase (28) and has a mutation rate of up to 30% in ccRCC (26). Inactivation of *SETD2* is correlated to poor overall survival and increased risk of tumor progression in the TCGA-KIRC cohort (29).

*BAP-1* mutations are present in up to 15% of tumors and it is known that BAP1 is involved in cellular proliferation and regulation of DNA damage repair. Mutations have been reported to be correlated to a worse prognosis (26).

The telomerase reverse transcriptase (*TERT*) gene plays an important role in telomere lengthening and is mutated in up to 14% of ccRCC tumors (26). Relative telomere length in tumor tissue has been positively correlated to larger tumor diameter but not to the clinical outcome (30).

The PI3K-mTOR pathway plays an important role in proliferation, apoptosis, and migration in ccRCC (26). The *PTEN* and *mTOR* genes are key regulators of this pathway and are mutated in 2-12% of ccRCC tumors. Mutations in the *TP53* gene occurs in 2-9% of ccRCC tumors, and mutations within this gene are correlated to metastatic disease and decreased cancer-specific survival (26).

**Epigenetics**

Epigenetics was introduced as a biological term in the early 1940s and defined by Waddington as “the branch of biology which studies the causal interactions between genes and their products which bring the phenotype into being”. Epigenetic research has evolved since then and is now known to regulate gene expression without altering the DNA sequence (31). In the human body, there are over two hundred different cell types and all cells contain, more or less, the same genomic sequence. Gene expression depends on epigenetic regulation amongst other regulatory events. Epigenetics also play an important role in the stabilization and packing of DNA, silencing of centromeres, telomeres, and transposable elements (32).
The key epigenetic regulations within a cell are histone modifications, DNA methylation and RNA modifications (Figure 2). The interplay between epigenetic modifications is tightly regulated and is a complex machinery with writers, readers, and erasers. Epigenetic writers are effector proteins that modify DNA and/or histones. Readers recognize these modifications leading to transcriptional activation or repression of said region. Since the epigenetic modifications are dynamic reversible processes, epigenetic erasers exists that can remove the modifications set by writers and thereby altering transcription (32,33).

Figure 2. A schematic and simplified presentation of the three most common epigenetic modifications, i.e. DNA methylation, histone modifications, and RNA modifications.
**Histone modifications**

The DNA strand is wrapped around an octamer of histones i.e. two units of H2A, H2B, H3, and H4, respectively giving rise to a nucleosome with histone H1 as a link. Nucleosomes then coil into chromatin fibers that are further condensed and forms the chromosome (Figure 2). Each histone constitutes of a core and a histone tail which is accessible for modifications such as acetylation, methylation, phosphorylation, ubiquitination, and sumoylation (34).

Acetylation, methylation, and phosphorylation are the most common histone modifications. The combination of modifications affects how tightly the DNA is wrapped around the histones making genes more or less accessible for transcription (34). Acetylation of the tails of histone H3 and H4 are done by histone acetyltransferases (HATs). Acetylation of H3K9/14/18/23, as well as H4K5/8/12/16 are all known to activate transcription by inducing euchromatin formation and making DNA accessible to transcription factors. Acetylation is a reversible modification and is reversed by histone deacetylases (HDACs). HDACs can also deacetylate other proteins and in that way affect cellular functions (34,35).

Histone tail methylation can both activate and repress transcription. Methylation of H3K4/36/79 are all related to active transcription whereas, H3K9/20/27 causes gene silencing. Histone methylation is mediated by histone methyltransferases (HMTs) that bind to specific DNA sequences and interacts with Trithorax group (Trx) proteins, the polycomb group (PcG) proteins, and/or RNA-interference (RNAi) to mediate histone methylation. One example of this is the EZH2 protein that binds to methylated DNA and attracts polycomb target (PcG) proteins by binding SUZ12 and EED. EZH2, SUZ12, and EED form the polycomb repressive complex 2 (PRC2) along with associated proteins, and cause methylation of H3K27 inducing a repressive chromatin state (34).

Phosphorylation of histones plays a crucial role during mitosis and meiosis and also interacts with other histone modifications affecting gene transcription. All histones can be affected by phosphorylation at different sites at their tails. Phosphorylation of H3S10 in combination with H3S28 are important in both mitosis and meiosis when they play a role in chromosome condensation (36).
RNA epigenetic modifications

Both coding and non-coding RNAs (ncRNAs) play an important role in epigenetic regulation of gene expression and can also be subjected to epigenetic alterations themselves. The most common and well-characterized RNA epigenetic modification is methylation on adenosines (m6A). The writers that catalyze m6A methylation methyltransferase-like 3 and 14 (METTL3/14) with the help of s-adenosylmethionine (SAM). m6A methylation can alter mRNA splicing, export, decay, stabilization, and translation (37). Epigenetic alterations of DNA accessibility might also affect transcription of ncRNA.

Non-coding RNA (ncRNA) constitutes 80% of transcripts from the DNA and do not form functional proteins. There are different kinds of ncRNA with different functions, but they can all play a role in gene function and epigenetics.

The role of Xist, a long ncRNA (lncRNA), is important in X chromosome inactivation in females. Xist binds to the X chromosome and recruits PRC2 to induce inactivation (38).

Micro RNAs (miRNA) are single-stranded RNAs that can interact with RNA before translation and induce its degradation. miRNA can also change the chromatin state by altering remodeling enzyme activity (39) and by modifying histone proteins (40). The activity of EZH2 and the formation of heterochromatin can be regulated by miRNA and is correlated to tumor progression (41).

Piwi-interacting RNA (piRNA) binds to Piwi, a protein important in the recruitment and binding of PcG proteins, and plays a role in euchromatin formation (39).

RNA-induced silencing (RISC) complex constitutes small RNA and argonaute protein and can inhibit the translation of RNA via degradation and/or non-degradative mechanisms. The RISC complex is also involved in heterochromatinization of pericentromeric regions when it recruits important components, i.e. SUV39H1 and SUV39H2, to these regions (42).
**DNA methylation**

The most well studied epigenetic alteration is DNA methylation. It was first described by Avery et al., in 1944 (43). Methylation occurs at the 5\textsuperscript{th} carbon position of cytosines (C), most often on cytosines located next to a guanine (G) which forms a so-called CpG site (44). In embryonic stem cells, 25% of methylated cytosines are located in the non-CpG context and seem to be important for maintaining cellular pluripotency. They are demethylated during cell differentiation (45). Distribution of CpGs throughout the genome varies and depending on the amount of CG-content areas are denoted CpG Islands (CGIs), CpG Shores, CpG Shelves, or open sea. CGIs have a higher level of CpGs then expected by random and about 60-70% of genes have a CGI in their promoter (46). Non-coding regions of the DNA harbors around 80% of the CpGs and are located in intergenic regions, satellite DNA, and transposable repetitive elements. DNA methylation in these regions acts to stabilize DNA by protecting against transposon activity (47).

DNA methylation pattern can either be inherited or can occur \textit{de novo}, and the pattern is established by DNA methyltransferases (DNMTs). DNMT1 ensures the maintenance of DNA methylation through cellular division and DNMT3A and DNMT3B implement \textit{de novo} methylation after the recruitment of DNMT3L (48,49). As other epigenetic alterations, DNA methylation is a dynamic process where demethylation is mediated by ten-eleven translocation protein (TET1-3) via hydroxy-methylation steps (50).

The epigenetic alterations regarding both histone modifications and \textit{de novo} methylation are tightly related to each other. As reviewed by Ceder and Bergman (2009) the interplay between histone modifiers and DNA methyltransferases is tightly regulated (51). During embryogenesis RNA polymerase II binds to promoter regions and initiates H3K4 methylation, which in turn inhibits DNMT3L binding, causing CpGs to remain unmethylated (52). In pericentromeric satellite repeats, the DNA forms heterochromatin due to H3K9 methylation that also attracts DNMT3B to further stabilize the heterochromatin (53). Methylated DNA attracts methylation binding protein (MBP) that can repress transcription by interfering with the binding of transcription factors as well as induce H3K9 methylation to maintain a repressive chromatin state (54). Histone acetylation is also correlated to DNA methylation. Methylated DNA is wrapped around non-acetylated H3 and H4 while unmethylated DNA is associated with euchromatin and acetylated histones (42).
The role of DNA methylation in gene transcription is yet to fully be established. A high proportion of genes have CGI in their promoter, which is most often unmethylated making the gene accessible for transcription. Methylation of a CGI located in a gene promoter regulates gene silencing by inhibition of transcription factor binding, initiation of an inactive heterochromatin state, and/or by attracting the Kaiso-like family (KLF) proteins involved in gene silencing (47).

De novo methylation is affected by environmental, stochastic and/or genetic events and might also be involved in complex diseases (55-58). DNA methylation patterns change over time and are associated with the mitotic age of cells as well as the chronological age of an individual. Different bioinformatics tools have been developed to calculate the biological and mitotic age of tissues (59-61).

Some DNA methylation positions in the genome seem to be associated with ethnicity and gender. These positions are called methylation quantitative trait loci (mQTL) and have been identified by analyzing DNA methylation in blood samples from several time points in large population-based cohorts (62). mQTLs show a consistent pattern with fewer mQTLs at birth compared to childhood but then the number decreases with increased age (62). The regulatory effect of mQTLs is yet to be investigated but there is evidence of contributions of mQTLs in complex diseases i.e. Crohn’s disease, hypertension, rheumatoid arthritis, and infectious diseases (62,63).

DNA methylation in cancer

DNA methylation and its correlation to cancer was first described in 1983 when Feinberg and Vogelstein identified a reduction of methylation of specific genes in colon cancer (64). Since then the research area concerning methylation and cancer progression has increased drastically. A number of methods to analyze DNA methylation alterations have been developed, e.g. DNA methylation arrays, bisulfite sequencing, and methylation-specific PCR. Tumor cells have a different DNA methylation landscape compared to normal cells and the alterations correlated to tumorigenesis include global hypomethylation and CGI hypermethylation (65,66).

De novo hypomethylation that occurs in methylated and inactivated regions leads to a higher risk of mutagenesis by induced genomic instability. It has been shown that DNA hypomethylation can activate oncogenes and drive tumor progression (67). Hypomethylation has also been reported to affect ribosomal DNA which alters the expression of different subunits of the ribosomes and thereby nucleolar function within cells (68).
Methylation alterations are known to be an early event in tumorigenesis. Arai et al., (2012) identified differences in methylation patterns between histological normal tissue from a tumor-bearing kidney compared to the kidney cortex from a healthy kidney (69). Altered methylation has also been identified in histological normal colon polyps adjacent to colorectal cancer (70) and Sato et al., (2013) described DNA methylation patterns that differed between histological normal lung tissue from a tumor-free lung compared to tissue from a tumor-bearing lung (71).

Hypermethylation of unmethylated CGI located in gene promoters might affect gene expression. Even though the genes affected by methylation alterations differ between tumors, the gene ontologies for effected genes often are involved in cell cycle control, apoptosis, DNA damage, immune recognition, and cell self-renewal (33,66,72). The polycomb gene EZH2 binds to a wide set of target promoters and these target genes are often affected by aberrant DNA methylation in different types of tumors. Overexpression of EZH2 has been reported in prostate, bladder, gastric, lung, hepatocellular cancer, and breast cancer and are correlated to poor outcome (41).

The correlation between DNA methylation patterns and clinical outcome has been studied in several types of cancer and has been evaluated in relation to tumor progression, clinical outcome, and response to treatment. Locke et al., (2019) have summarized commercially available screening methods of “liquid biopsies” that test for methylation of circulating tumor DNA (ctDNA) for several tumor types (73). Methods have been described for bladder cancer, breast cancer, cervical cancer, colorectal cancer, glioblastomas, liver cancer, lung cancer, and prostate cancer with the goal to detect primary tumors, tumor progression after treatment, and response to therapy (73). DNA methylation-based method can also be used for the classification of Central Nervous System (CNS) tumors (74). Capper et al., (2018) have presented a method that classifies unusual, non-specific or non-representative histological CNS tumors (75).

The CpG Methylator Phenotype (CIMP) was first described for colorectal cancer (CRC) in 1999 by Toyota et al., (76). CRC tumors were divided into CIMP low and CIMP high depending on their CGI methylation levels (76). Since then, the CIMP classification method has been used in several different tumor types, including gastric, lung, liver, ovarian, kidney, and leukemias (77,78). CIMP panels differ between tumor types, with different genes and/or CpGs included in the different panels, and with a wide range in the number of included CpGs (77,78).
**DNA methylation in ccRCC**

DNA methylation plays an important role in tumorigenesis in ccRCC. More than seventy genes, summarized by Ricketts et al., (2014), are thought to be affected by DNA methylation (79). These epigenetically affected genes are involved in pathways associated with RCC carcinogenesis, i.e cell cycle regulation, apoptosis, and cell metabolism amongst others (80).

The VHL-HIF (hypoxia-inducible factor) pathway plays an important role in renal carcinogenesis since it affects angiogenesis, proliferation, apoptosis, survival, metabolism, and pH regulation. Important mediators in this pathway i.e VHL, TIMP3, and GREM1 amongst others, are affected by DNA methylation leading to its dysregulation (80).

The WNT-β-catenin signaling pathway regulates DNA-binding transcription factors and is tightly regulated. In RCC, WNT antagonists are commonly inactivated by DNA methylation, leading to activation of the WNT-β-catenin pathway. These events are associated with increased levels of ongogenic target genes and tumorigenesis (14,80).

Epithelial-mesenchymal transition (EMT) plays an important role in embryogenesis and also in pathological processes like wound healing, tissue fibrosis and cancer progression (81,82). In carcinomas, EMT is thought to be necessary for the invasion-metastasis cascade. Histone modifications, DNA promoter alterations, and post-translational activity are important for EMT by changing the levels of transcription factors (i.e. SNAIL, SLUG, and Twist), cell surface proteins, cytoskeleton structure, and miRNA expression (81,82). The best-studied gene important for EMT is CDH1. The expression of CDH1 can be inactivated by promoter methylation, inducing EMT as a result. Correlation of hypermethylation of CDH1 and metastatic disease has been shown in breast, bladder, lung, liver, gastric, and prostate cancer (82).

**Methylation biomarkers for ccRCC**

As established above, DNA methylation plays an important role in ccRCC tumor progression. Several differently methylated genes and/or combinations of these genes have been shown to be involved in ccRCC tumor progression and related to clinical outcome (69,79,83-86).

Arai et al., (2012) published a CIMP panel that predicted outcome in ccRCC. Tumor-free and tumor tissue were analyzed using the 27K arrays, differently methylated CpGs were identified and cluster analysis was used to separate tumors into two distinct groups. Sixteen CpGs with the highest difference in methylation patterns between the two clusters were included in the CIMP panel (69). Later, Tian et al., (2014) evolved this panel by using the MassARRAY system and
identified 23 units (more than one unit per gene) including 32 CpGs representing seven of the primary identified CIMP genes (83).

Wei et al., (2015) created a five CpG methylation-based assay to predict overall survival in ccRCC (84). They used two different data sets to build their classifier and could later verify the data in three independent cohorts. This classifier was also investigated for intratumor heterogeneity (ITH) with accurate classification of all pieces from the same tumor in 21 of 23 patients (84).

A four-gene methylation profile was described by van Vladrop et al., (2016) that was associated with ccRCC disease and/or poor outcome. The panel; constituted of the GREM1, LAD1, NEFH, and NEURL genes was created using subsequent nested multiplex methylation-specific PCR in two different cohorts (n=150 and 185 respectively). The panel was also verified in the TCGA-KIRC data set using one single CpG site for each of the four included genes, indicating its robustness as a prognostic marker (86).

Ricketts et al., (2014) summarized and analyzed DNA methylation of 77 genes known to be differently methylated in kidney cancer (79). They used both 27K and 450K arrays to analyze CpGs represented by these genes. By both methods, methylation of SFRP1 and BNC1 were shown to be associated with poor survival in ccRCC (79).

Joosten et al., (2017) performed a systematic review of genes methylated in RCC (85). By this approach, nine genes were identified that correlated to progression-free survival, disease-specific survival and/or overall survival. The conclusion of this review highlighted the importance of using a combination of several genes to predict outcome in ccRCC (85).

Today, the clinical use of DNA methylation as a prognostic marker for ccRCC is not established. The importance of epigenetic alterations, including DNA methylation, is of big interest in ccRCC where there is strong evidence of its importance in tumorigenesis.

In this study, we identified two different methylation-based classifiers that predict tumor progression in non-metastatic ccRCC.
Aims

The frequency of metastatic ccRCC at diagnosis has decreased since the beginning of the 21st century but still, 20-30% of non-metastatic tumors progress into metastatic disease. There is a need for better risk stratification tools to be able to identify the non-metastatic tumors with high risk for tumor progression and treat them accordingly to increase the quality of life for these patients.

The aim of this thesis was to evaluate the DNA methylation alterations at diagnosis and to correlate these epigenetic modifications with clinicopathological parameters and clinical outcome in ccRCC.

Paper I

The aim of paper I was to determine if DNA methylation can be used as a potential prognostic marker in ccRCC.

Paper II

The aim of paper II was to evaluate the prognostic relevance of DNA methylation in relation to clinical characteristics in ccRCC, with a special focus on non-metastatic patients at diagnosis.

Paper III

The aim of paper III was to define a diagnostic model based on a combination of DNA methylation profiling and clinicopathological variables for predicting tumor progression after nephrectomy in non-metastatic ccRCC.
Materials and Methods

Workflow describing patients and methods included in this thesis

Patients and tissue samples

All tissue samples from clear cell Renal Cell Carcinoma (ccRCC) patients were obtained after written, informed consent. Ethical approval for the studies was given by the regional ethical review board in Umeå (Dnr 2011-156-31M, 20110523).

The ccRCC patients included in this thesis (summarized in Table 1) were diagnosed between 2001 and 2009. Patients were treated with partial or total nephrectomy without neoadjuvant treatment and followed according to the Swedish health care program (2). Tissue samples were collected after surgery and snap-frozen in fluid nitrogen and stored at -80°C. Both tumor (T) tissue and histological normal tumor-free (TF) tissue from the tumor-bearing kidney were collected. Tumors were staged according to the 2002 TNM classification system (6) and the nuclear grade was determined according to Fuhrman et al., (1982) (7).

Survival data from all included patients were extracted in March 2014 (paper I) and in August 2017 (paper II and III)
**Tissue samples included in paper I**

Paper I included 45 patients; 19 non-metastatic (M0) and 26 metastatic (M1) tumors from which T samples were collected. From six M0 and six M1 patients, TF tissue was also collected.

Blood samples from 24 M0 patients were collected and stored as buffy coats at -80°C. Thirteen ccRCC patients were analyzed for methylation in both T tissue and blood samples.

**Tissue samples in papers II and III**

Papers II and III included 115 patients; 87 non-metastatic (M0) and 28 metastatic (M1); and 12 TF tissue samples from eight M0 and four M1 patients. Out of the 87 non-metastatic patients, 23 developed metastasis after nephrectomy (M0-P) whereas 59 remained ‘true’ non-metastatic (M0-PF).

Thirty-nine patients included in paper I were also included in papers II and III.

For analyzing intratumor heterogeneity (paper II), multiple samples (two or three) from the same individual were analyzed with 450K arrays (paper II).

**Tumor samples from the Tumor Cancer Genome Atlas (Paper II)**

In Paper II, the publically available TCGA-KIRC dataset was used as a validation cohort. Information from 230 non-metastatic ccRCC patients including methylation data from 450K arrays and clinical information were downloaded from the Broad Institute’s Genome Data Analysis Center Firehose (http://gdac.broadinstitute.org/). All included patients were treated with partial or total nephrectomy without neoadjuvant or adjuvant treatment.

**Blood analysis**

Paper III included blood samples taken within a week prior to nephrectomy. The blood samples were analyzed for hemoglobin levels (Hb), thrombocyte particle count (TPC), calcium, albumin, alkaline phosphatase (ALP), gamma-glutamic-transferase (GGT), and creatinine concentrations.

**DNA extraction**

DNA was extracted from T, TF tissue and blood samples as described by Svenson et al., (2009) (30) using the MagAttract technology (Qiagen).
Methylation analysis

Bisulfite conversion (Papers I-III)

Purified DNA from different tissues were bisulfite treated prior to DNA methylation analysis using the EZ DNA Methylation Kit (Zymo Research, Irvine, USA). The efficiency of the bisulfite conversion was verified by MethyLight analysis of the ALU gene (87).

Methylation array analysis (Papers I-III)

Two hundred nanogram of bisulfite converted DNA was analyzed on either the Illumina Infinium HumanMeth27K Bead Arrays (27K arrays, Paper I) or HumanMeth450K BeadChip Arrays (450K arrays, papers II and III) according to manufacturer’s protocol (Illumina In., San Diego, USA).

The 27K arrays were scanned and the fluorescence was measured using a Bead Array Reader (Illumina) and methylation levels were analyzed using the Genome Studio Software (version 2011.1 and the methylation analysis module (version 1.9.0). The 450K arrays were scanned on the HiScan array reader (Illumina) and fluorescence intensities were extracted using the Genome Studio software (V2011.1) and the Methylation module (1.9.0).

The fluorescence intensities from methylated (M) vs. unmethylated (U) probes were used to calculate a methylation level (β-value) \( \beta = \frac{\text{Max}(M, 0)}{\text{Max}(M, 0) + \text{Max}(U, 0) + 100} \) for each CpG site ranging from 0 (completely unmethylated) to 1 (completely methylated). \( \beta \)-values from the 450K arrays were normalized using the BMIQ method (88).

Technical reproducibility was analyzed in the 450K arrays by including replicate samples, two or three samples from nine individuals. Duplicate and triplicate samples showed a high correlation with \( R^2 \) values that ranged from 0.995 to 0.997.
Data preprocessing
Data filtration was performed for both array types. CpGs lacking observations in any sample and CpGs with a detection \( p \)-value greater than 0.05 were omitted. To avoid gender bias CpGs located on chromosomes X and Y were omitted. For the 450K arrays, CpGs aligned to multiple loci or located within 3 bp from a known single nucleotide polymorphism (SNP) were also omitted (89) as well as CpGs located outside the promoter region. In paper III CpGs known to be located in a methylation quantitative trait loci (mQTLs) were excluded from the analysis (62).

In paper II, mQTLs were excluded from the Promoter Methylation Classification (PMC) panel.

Array data analysis
Methylation array data analysis was done using R (v2.15.0 paper I, v3.4.1 papers II-III and v3.4.3 paper III).

Cluster analysis was performed using Ward’s method with Euclidean distance metric (90), including all promoter-associated CpGs; 21,930 (paper I) and 155,931 (paper II).

Biological epigenetic age was calculated for T and TF tissue samples as well as blood samples by the method presented by Horvath (2013) for patients in paper I (60). The EpiTOC model was used in paper II to determine the mitotic age of T and TF tissue samples (61).

Differently methylated CpGs (DM-CpGs) were identified by comparing T samples to TF samples \( (\Delta \beta = \beta_T - \bar{\beta}_{TF}) \). Average methylation \( (\bar{\beta}_{TF}) \) for included TF samples was calculated for each study, i.e. papers I and II, and compared to each individual T sample. A CpG site was established as DM if it had an absolute \( \Delta \beta \geq 0.35 \) (Paper I) or an absolute \( \Delta \beta \geq 0.2 \) (paper II). DM-CpGs with a positive \( \Delta \beta \) was defined as hypermethylated and to DM-CpGs with a negative \( \Delta \beta \) was defined as hypomethylated.

The building of prognostic classifiers
Two different classifiers that can be used as tools for the prediction of tumor progression in non-metastatic patients are presented in this thesis.
**Promoter Methylation Classification (PMC) Panel (Paper II)**

DM-CpGs present in at least 70% of samples within each group of samples (i.e. Mo-PF, Mo-P, and M1) were selected for further analysis. DM-CpGs common for Mo-P and M1 (n=172) but not DM in Mo-PF were defined as a Promoter Methylation Classification (PMC) panel (Figure 10).

The average methylation for all 172 CpGs, with hypomethylated CpGs mirrored (1 – average beta), was calculated for all Mo tumors. A ROC curve was created using the average methylation as a test variable and tumor progression as a state variable. From this curve, a Youden index was calculated to determine a cut-off (average beta=0.688) for PMC high and low classification. The TCGA-KIRC data set was used to evaluate the PMC panel, missing values after preprocessing were imputed using the k-nearest neighbors’ method.

**Triple Classifier (Paper III)**

Classifiers for tumor progression were built by using a combination of clinicopathological variables (clinical), CpGs previously shown to be associated with ccRCC prognosis (PI-CpGs) and by directed cluster analysis (DCA).

The CpGs previously associated with ccRCC progression were chosen from five original publications and one review (69,79,84-86,91). In total 64 CpGs were chosen to be included and denoted as PI-CpGs.

The DCA method identified cluster of CpGs with potential relevance for prognosis. Each individual CpG site was separated into two groups using 2-means. CpGs with a skew deviation (less than 10% of samples in the smaller group) and with an absolute mean difference (Δβ) between groups less than 0.2 were omitted from the analysis. The group labels of the remaining CpGs were adjusted such that the groups with the highest mean methylation were all labeled as 1 and the other group as 0. The group belongings (0, 1) were then clustered using 40-means, giving rise to 40 DCA clusters. For each cluster, a consensus variable was calculated as the average methylation of all included CpGs.

In total, six different classifiers were built for Mo patients; clinical, PI-CpGs, DCA, clinical+PI-CpGs, clinical+DCA, and clinical+PI-CpGs+DCA; by using logistic regression on the five first Principal Components. Inclusion criteria for patients were at least five year follow up time, n=78; 58 Mo-PF and 20 Mo-P. To compare the classifiers the sensitivity was set to a fixed value of 85%.
Genetic analysis

Genetic aberrations (Papers I and II)
To evaluate overall genetic aberrations within ccRCC both CytoSNP-12 arrays (CytoSNP-12, Illumina) and the 450K arrays were used.

In paper I information from the CytoSNP-12, as described by Köhn et al., 2014 (22), was used to evaluate the genetic aberration of the VHL gene located on chromosome 3p and DNA methylation in the VHL gene promoter. In paper II, the twelve most common genetic aberrations described by Köhn et al., 2014 were analyzed using raw data signals from the 450K arrays and the Conumee package (v1.9.0) in R (v3.4.1). The limitations in the detection of gains and deletions within the set regions were done through visual inspection and was set for each individual sample.

VHL status (Paper I)
Information concerning the mutation of the VHL gene was derived from Köhn et al., 2014 were Sanger sequencing of the VHL gene was performed (22).

Gene expression analysis

RNA preparation (Paper II)
RNA was extracted from T tissue as described in detail in paper II using the MagAttract RNA Universal Tissue M48 Kit (Qiagen, Hilden, Germany). The quality of RNA was analyzed using the 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). cRNA was produced using the Illumina TotalPrep RNA amplification kit (Ambion Inc., St. Austin, TX, USA) and the quality was evaluated using the RNA 6000 pico kit (Agilent Technology).

Gene expression arrays (Paper II)
Genome-wide gene expression analysis was performed using Human HT12 Illumina Beadchip gene expression array (Illumina) according to the manufacturer's protocol and the arrays were scanned with the Illumina Bead Array Reader (Illumina). The Illumina BeadStudio V2011.1 software was used to normalize the data by the cubic spline algorithm, and expression levels for three selected genes i.e. MX2 (ILMN_2231928), SMAD6 (ILMN_1767068) and SOCS3 (ILMN_1781001) were extracted.
**Statistical analysis**

Statistical analyses were performed using the Statistical Package for the Social Sciences (SPSS Inc., Chicago, IL, USA) software version 22 (Paper I) and version 24 (papers II-III). In paper I R (v2.12.0) was used, in paper II R (v3.4.1) and in paper III both R (v3.4.1) and R (v3.4.3) were used.

To compare differences between patient groups Chi-square test was used for categorical variables. Mann-Whitney U test was used to compare continuous variables between two groups and the Kruskal Wallis test when comparing three groups. Correlations between continuous variables were analyzed using the Pearson correlation coefficient.

Survival analysis was performed using Kaplan-Meier and the Log rank test. Five year cancer-specific survival (CSS$_{5yr}$) and cumulative incidence of progress (CIP$_{5yr}$) were obtained from Kaplan-Meier survival tables. Multivariate Cox Regression analysis was used to analyze the combination of several predictor variables.

Genes associated with DM-CpGs (Paper I) and CpGs included in the PMC panel (paper II) were analyzed for protein function using the Metacore software (GeneGO, Inc.) in Paper I and WebGestalt (92) in paper II.

Similarities within and between tumors were determined using Principal Component Analysis (PCA) using the average beta values for 156K CpGs from the 450K arrays. PCA was performed using SIMCA version 14 (Umetrics, Umeå, Sweden).

Twelve common regions identified as genetic aberrations by the CytoSNP-12 arrays were also analyzed using the 450K methylation arrays. Results from the two methods were compared using Cohen’s kappa test.

Distribution of DM-CpGs was analyzed against frequent gain/loss regions in ccRCC defined by Köhn et al., (2014) (22) to evaluate if any overrepresentation of DM-CpGs in affected regions. The correlation was analyzed using the Chi-square test.

Identification confirmation of multiple samples from the same individual was verified by the 65 built-in SNP probes in the 450K arrays.
Results

Demographic overview of the ccRCC cohorts.
Forty-five ccRCC patients (paper I) and 115 ccRCC patients (papers II-III) were included in these studies. Clinicopathologic variables were collected for all included patients and are briefly summarized in Table 1 (for details see each individual paper). Differences in tumor progression frequencies of ccRCC patients between papers II and III are due to different criteria for follow up. In paper III, patients with follow up time less than five years or patients that died within five years after diagnosis, were censored. (Table 1).

**Table 1. ccRCC patient overview for paper I-III.**

<table>
<thead>
<tr>
<th>Patients</th>
<th>Paper I (n=45)</th>
<th>Paper II (n=115)</th>
<th>Paper III1,2 (n=115)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (mean ± SD)</td>
<td>64.4 ± 10.1</td>
<td>65.0 ± 11.5</td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>30</td>
<td>66</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>15</td>
<td>49</td>
<td></td>
</tr>
<tr>
<td>Tumor diameter (mm, mean ± SD)</td>
<td>95.4 ± 40.1</td>
<td>70.0 ± 41.6</td>
<td></td>
</tr>
<tr>
<td>Morphological Grade</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G1</td>
<td>4</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>G2</td>
<td>13</td>
<td>46</td>
<td></td>
</tr>
<tr>
<td>G3</td>
<td>15</td>
<td>34</td>
<td></td>
</tr>
<tr>
<td>G4</td>
<td>13</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>T-stage</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>6</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>T2</td>
<td>14</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>T3</td>
<td>24</td>
<td>39</td>
<td></td>
</tr>
<tr>
<td>T4</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>M-stage</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mo</td>
<td>19</td>
<td>87</td>
<td></td>
</tr>
<tr>
<td>M1</td>
<td>26</td>
<td>28</td>
<td></td>
</tr>
<tr>
<td>TNM Stage</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>5</td>
<td>49</td>
<td></td>
</tr>
<tr>
<td>II</td>
<td>8</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>III</td>
<td>6</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>IV</td>
<td>26</td>
<td>28</td>
<td></td>
</tr>
<tr>
<td>Follow Up status</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alive</td>
<td>9</td>
<td>39</td>
<td></td>
</tr>
<tr>
<td>Alive, disease</td>
<td>1</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Dead, ccRCC</td>
<td>31</td>
<td>45</td>
<td></td>
</tr>
<tr>
<td>Dead</td>
<td>4</td>
<td>26</td>
<td></td>
</tr>
<tr>
<td>Tumor Progression</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No (M0-PF)</td>
<td>64</td>
<td>58</td>
<td></td>
</tr>
<tr>
<td>Yes (M0-P)</td>
<td>23</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>Mo-censored</td>
<td>9</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1 – 5-year follow up status
2 – 5-year tumor progression status
Loss of chromosome arm 3p and VHL genetic and epigenetic status (Paper I)

Genomic aberrations are commonly observed in ccRCC tumor tissue and show considerable heterogeneity. The tumors included in this thesis were screened for the most common genomic aberrations presented by Köhn et al., (2014) (22).

Loss of one chromosome arm, 3p, is present in approximately 90% of ccRCC tumors and inactivation of the VHL gene due to a frame-shift mutation, located on 3p, is a common finding in ccRCC (19).

A subset of tumor (T) samples (n=38 out of 45) within the ccRCC cohort in paper I, was previously screened for VHL mutation in the study by Köhn et al., (2014) (22). Genomic aberration in the VHL gene was present in 85% of analyzed T samples, 28% had an allelic loss of one VHL copy and 56% had both one allele loss and a frame-shift mutation in the other copy of the VHL gene (Figure 4).

Methylation analysis of the VHL promoter (Paper I).

Methylation alterations in the VHL gene have been reported at different frequencies (14-16). Our ccRCC cohort was characterized based on VHL promoter methylation status. Five CpGs in the VHL gene promoter were analyzed using the Illumina Infinium HumanMeth27K BeadChip Arrays (27K arrays). The VHL promoter methylation pattern was homogenus with two CpGs (cg27226214 and cg16869108) hypermethylated in more than 97% of tumor samples and two CpGs (cg03509024 and cg22782492) hypomethylated in more than 88% of samples compared to tumor-free (TF) tissue samples. One CpG site (cg24092914) was hypomethylated in 20% of samples and was significantly correlated to metastatic disease (p=0.004). There was no association between the presence of genetic and epigenetic alterations in the VHL gene (Figure 4).

**Figure 4.** Percentage of DM-CpGs in the VHL promoter depending on VHL status, wild type (wt), VHL loss or VHL loss and frame-shift mutation (VHL loss+mut).
Genomic aberrations and ccRCC disease progression (Paper II)

Twelve frequent genomic aberrations in ccRCC were previously described by Köhn et al., (2014) in a subset of this cohort (Figure 5) (22). These genomic regions were analyzed in 115 ccRCC patients by using the raw data signal from the HumanMethy450K BeadChip Arrays (450K arrays) to identify genomic gains or losses. The most common aberration (84% of samples) was a loss of 3p, encompassing the VHL gene. The other 11 analyzed aberrations present in 15 to 53% of samples (Figure 5). The distribution of genetic aberrations was compared between the groups (M0-PF, M0-P and M1) and five regions showed a significant difference in distribution between the groups. A gain of 7p and 7q were more common in M1 samples compared to M0-P (p=0.013 and 0.024 respectively) and loss of 9p, 9q, and 14q were more common in M1 compared to M0-PF (p=0.001, p<0.001, and 0.042 respectively) (Figure 5). The total number of genetic aberrations was significantly higher in M1 samples compared to both M0-PF and M0-P samples (p=0.024 and 0.050 respectively). The number of genetic aberrations showed heterogeneity within all prognostic groups. All groups included tumors with both high and low numbers of genetic aberrations.

![Figure 5](image.png)

**Figure 5.** Allocation of common genomic aberrations across different prognostic groups. White color indicates no aberrations and black color indicates gain or loss of specific region.
DNA methylation classification and its correlation to clinicopathological parameters (Papers I and II)

DNA methylation arrays were used to evaluate methylation alterations for the prognostic relevance of ccRCC progression. In paper I DNA methylation was analyzed using the 27K arrays and in paper II the 450K arrays were used.

Hierarchical clustering analysis of genome-wide promoter-associated CpG sites was used to classify ccRCC and tumor-free (TF) tissue samples into groups depending on similarities in methylation patterns/profiles. In papers I and II cluster analysis was performed using 21,930 (paper I) and 155,931 (paper II) promoter-associated CpGs. In both studies, samples were divided into two clusters, cluster A and B, and the TF tissue samples clustered together or in close proximity of cluster A. There was a significant difference in overall methylation between the two clusters with higher average methylation in cluster B. There was also a significantly higher number of hypermethylated CpGs in cluster B but no difference in the number of hypomethylated CpGs (Figure 6A-B).

Even though the number of patients (n=45 in paper I and n=115 in paper II) and distribution of samples in cluster A/B (58% M1 in paper I and 24% M1 in paper II) differed between the studies the patterns of sample allocation between clinicopathological variables were similar. Cluster B was significantly associated with larger tumors and metastatic disease (Figure 6A-B) in both studies. However, neither study showed a correlation of cluster status with age or gender.

Survival analysis comparing cancer-specific survival (CSS) between cluster A and B classified tumors showed a significant difference in survival with poorer outcome for patients belonging to cluster B in both studies. In paper II tumors could be stratified depending on metastasis status (M0 and M1). The difference in CSS between clusters A and B was significant in M0 tumors but not in M1 were both groups of patients had a poor outcome (p<0.001 and p=0.840) (Figure 7A). Investigation of the cumulative incidence of progress (CIP) in non-metastatic tumors showed that tumors belonging to cluster B were more prone to progress (Figure 7B).
Figure 6. Cluster A/B grouping based on hierarchical clustering of promoter-associated DNA methylation in (A) 21,930 CpGs and (B) 155,931 CpGs. A gradient of average methylation (β) and tumor diameter (mm) are shown as well as the number of hyper- and hypomethylated CpGs. Non-metastatic (M0) and metastatic (M1) tumors are shown in white (M0) and black (M1). In (B) tumor progression is shown with 'true' non-metastatic (M0-PF) in white, non-metastatic that later will progress (M0-P) in light grey and metastatic disease (M1) in black.
**Figure 7.** Survival analysis based on a combination of Mo/M1-status and cluster belonging at diagnosis. (A) Kaplan Meier cancer-specific survival (pCSS5yr) in 115 ccRCC patients depending on metastatic status and cluster belonging. (B) cumulative incidence of progress (pCIP5yr) analysis of 87 non-metastatic ccRCC tumors in relation to cluster status. Log rank p-values are presented.

**Methylation alterations associated with tumor progression (Paper I).**

To identify CpGs associated with ccRCC tumor progression, Δβ values for each CpG were calculated by comparing each T sample to pooled TF tissue samples, as described in the material and method section. CpGs with an absolute Δβ ≥ 0.35 were defined as differently methylated CpGs (DM-CpGs). DM-CpGs present in at least 20% of T tissue samples were chosen for further analysis to determine protein function and its relation to polycomb target genes.

In total, 959 CpGs were defined as DM in at least 20% of the T samples representing 840 coding genes. There was an accumulation of genes involved in developmental processes, morphogenesis, differentiation, and cell-cell signaling. Twenty-five percentage of genes represented by DM-CpGs were significantly enriched by polycomb target genes, as defined by Lee et al., (2006) which were significantly higher than expected (p<0.001) (93).
A stepwise increased DNA methylation level during tumor progression (Papers I and II).

After establishing the correlation of differential methylation landscape with clinicopathological variables, the correlation between methylation levels and tumor progression was investigated. A significant stepwise increase in average promoter methylation; from TF-tissue, via non-metastatic tumors (M0) to metastatic tumors (M1) was shown in both papers I and II (Figure 8). Overall promoter methylation levels were higher in samples analyzed by the 450K arrays compared to the 27K arrays.

Inter- and intra- heterogeneity in DNA methylation profiles (Paper II).

As shown above, methylation levels between samples in the same group (i.e. M0 and M1) showed high variation (Figure 8). Using principal component analysis (PCA) based on methylation levels of promoter-associated CpGs, similarities between groups of samples were analyzed. TF-samples showed a homogenous methylation pattern and the same for T-samples belonging to the same cluster (i.e. cluster A/B). On the other hand distribution of T-samples belonging to the same group (i.e. M0-PF, M0-P, and M1) showed overlapping methylation patterns (Figure 9A).

In order to evaluate DNA methylation within tumors, two or three samples from the same tumor in six patients were analyzed using the 450K arrays. PCA analysis of these samples showed a homogenous methylation pattern (Figure 9B).
**Figure 9.** Principal component analysis (PCA) on (A) 115 ccRCC and 12 tumor-free (TF) tissue samples. Samples are highlighted as Tumor-Free tissue (blue) non-metastatic tumors that without progress (Mo-PF, green) and with progress (Mo-P, orange), and ccRCC with metastasis (M1, red), and (B) two or three samples from the same tumor were collected from five patients. Each patient is represented by an individual color and each sample with a dot. PCA was performed using the two first principal components of the average methylation of promoter-associated CpGs.

**Methylation and genetic aberration and their correlation with biological and mitotic age (Papers I and II)**

The biological and mitotic age of patient tissue samples were estimated by different epigenetically based age predictors (60,61). In paper I the method described by Horvath (2013) was used to calculate biological age (60) and mitotic age was calculated by using the epiTOC model described by Yang et al., (2016) in paper II (61).

In paper I, the biological age of blood from patients with ccRCC, TF and T tissue were calculated and compared with the chronological age of the individual patient. There was a significant correlation between biological and chronological age in blood and T-tissue. However, the estimated biological age in T-tissue was estimated significantly higher than the chronological age of the patient.

In paper II, the mitotic age of T tissue was calculated and related to the number of genetic and epigenetic alterations. There was a significant correlation between the number of genomic aberrations and of hypermethylated CpGs, that were also correlated to predicted mitotic age. In contrast, there was no correlation between either the number of hypomethylated CpGs or the number of genomic aberrations to mitotic age.
New prognostic classifiers in ccRCC (Papers II-III).
Whereas methylation alterations showed correlation to tumor progression we aimed to define differences in methylation patterns that could identify non-metastatic tumors at diagnosis that later progress (Mo-P).

Promoter Methylation Classifier (PMC) panel (Paper II).
The heterogenous methylation pattern seen between tumors within the same group (i.e. Mo-PF, Mo-P, and M1) made it difficult to use an average methylation level for each group to identify CpGs of importance for tumor progression. Instead, DM-CpGs were identified at individual level as described in the material and method section. DM-CpGs in at least 70% of samples in each group were selected for further analysis (Figure 10). Combining DM-CpGs from each group in a Venn-diagram showed that Mo tumors had a similar methylation pattern as M1 with only 12% and 17% unique DM-CpGs for Mo-PF and Mo-P respectively whereas M1 tumors had 56% unique DM-CpGs.

A Promoter Methylation Classifier (PMC) panel was defined based on the 172 DM-CpGs common for Mo-P and M1 (Figure 10). A receiver operating characteristic (ROC) curve was used to set the PMC high and PMC low threshold.

Figure 10. Workflow showing the identification of the Promoter Methylator Classifier (PMC) panel.

Non-metastatic tumors were classified as PMC low or high based on the cut-off derived from the ROC curves. The PMC status was correlated to tumor progression, where the PMC high patients showed a significantly higher frequency of tumor progression (PMC low pCIP 5yr 8% vs. PMC high pCIP 5yr 38%, p=0.001). This result was verified in an independent cohort from the TCGA-KIRC data set (PMC low pCIP 5yr 16% vs. PMC high pCIP 5yr 39%, p<0.001) (Figure 11).
**Figure 11.** Kaplan-Meier cumulative incidence of progress (CIP) for the promoter methylation classification (PMC) panel (A) in our cohort involving 87 non-metastatic ccRCC and (B) in the TCGA-KIRC validation cohort including 230 non-metastatic ccRCC. The difference in CIP was between groups was analyzed using the Log Rank test.

**Genes represented in the PMC panel**

The 172 CpGs included in the PMC panel represented 160 unique genes of which 109 genes were hypermethylated and 51 hypomethylated in relation to TF tissue. The hypermethylated genes were enriched for gene ontology terms associated with SMAD protein complex assembly, RNA polymerase II regulation and response to pH, whereas hypomethylated genes were associated with several defense mechanisms, immune response, and response to external stimuli.

In total, 17 genes were represented by two or more DM-CpGs in the PMC panel. One of these genes (SMAD family member 6 (SMAD6)) showed significantly higher average methylation in the M0-P group compared to Mo-PF whereas suppressor of cytokine signaling 3 (SOCS3) and MX dynamin-like GTPase 2 (MX2) showed a significantly lower level of methylation. To evaluate if methylation in these gene promoters was associated with mRNA gene expression levels data were extracted from 28 analyzed patients on the Illumina Beadchip gene expression arrays. There were significantly higher mRNA levels for SOCS3 and MX2 expressions in both M0-P and M1 compared with Mo-PF whereas no significant difference was seen for SMAD6.
**Triple Classifier (Paper III).**

In the clinic, the Mayo scoring system is currently used to stratify patients into risk groups for tumor progression (8). The risk for tumor progression differs between Mayo score based risk groups and these risk groups are followed up by different procedures. In our study, the Mayo risk score was set as ‘golden standard’ when creating new classifiers for tumor progression. Since Mayo Intermediate and High-risk groups have identical follow up protocols, we pooled these patients together to be able to calculate sensitivity and specificity for this classifier. In our cohort, the Mayo scoring system had 85% sensitivity and 50% specificity for 5-year risk for tumor progression.

A combination of clinicopathological variables and methylation profiles were used to build risk classifiers for tumor progression in non-metastatic ccRCC. In total, twelve clinicopathological variables; age, gender, morphological grade, TNM, tumor diameter and blood analysis (albumin, alkaline phosphatase, calcium, creatinine, gamma glutamyltransferase, hemoglobin, thrombocyte particle count) were included, as well as 64 previously identified CpG sites (PI-CpGs) of relevance in ccRCC and 40 variables identified by Directed Cluster Analysis (DCA).

The categories described above, i.e. clinical, PI-CpGs and DCA were used individually or in combination to build six different classifiers; clinical, PI-CpGs, DCA, clinical+PI-CpGs, clinical+DCA and clinical+PI-CpGs+DCA (denoted “triple classifier”). The classifiers were built by logistic regression with the five first principal components were used, to be able to compare the new classifiers to the clinically used Mayo scoring system the sensitivity was set at 85% and the specificity was compared.

Each patient was classified as low risk for progress (LRP) or high risk for progress (HRP) by each classifier and specificity was evaluated at 85% sensitivity. The specificity for the six classifiers ranged from 43% to 64%, with the lowest specificity was seen for DCA alone and the highest for the triple classifier.

The triple classifier showed higher specificity (64%) compared to Mayo (50%) at 85% sensitivity in our ccRCC cohort. Both Mayo and the triple classifier showed a significant difference in the cumulative incidence of progress within five years (pCIP$_{5yr}$ p=0.005 and <0.001, respectively). The Triple classifier was better to prognosticate progress (CIP$_{5yr}$ low risk$_{Mayo}$ 9.4% vs 7.5% LRP$_{triple classifier}$ and CIP$_{5yr}$ intermediate/high risk$_{Mayo}$ 37% vs 45% HRP$_{triple classifier}$) (Figure 12).
Figure 12. Cumulative incidence of progress (pCIP$_{5yr}$). 87 non-metastatic ccRCC tumors were classified according to (A) the Mayo Scoring System and (B) the triple classifier (clinical+PI-CpGs+DCA) at diagnosis. The pCIP$_{5yr}$ was compared between risk groups and Log Rank p-values are presented.
Discussion

There is evidence that histological normal tissue taken from a tumor-bearing kidney contains differently methylated CpGs compared to tissue from a tumor-free kidney (69). It has also been shown that DNA methylation is an early event in several different cancer types and therefore, made us investigate the role of DNA methylation in ccRCCs at diagnosis and to evaluate if there is a correlation to tumor progression and clinical outcome.

This project started with a cohort of 45 ccRCCs that were analyzed for methylation using the 27K arrays with a promoter-directed approach. The conclusion drawn from paper I that, methylation classification has a potential role as a prognostic marker in ccRCC led us to expand the study. In paper II and III, the study cohort included 115 patients, and methylation was analyzed using the 450K arrays. The 27K array includes 27,758 CpG dinucleotides located in gene promoter regions. The 450K array is an evolution of the 27K array including 485,577 CpGs scattered throughout the whole genome. The decision to use the 450K arrays in paper II and III became natural when the updated array not only analyzed promoter-associated CpGs, but also CpGs within the whole genome as well as intergenic regions. The total number of CpGs associated with each gene promoter also increased from the 27K arrays to 450K arrays. The distribution of CpGs per gene is biased on both types of methylation arrays, more “interesting/well-characterized genes” are represented by a higher number of CpGs. The decision to focus this thesis on promoter-associated CpGs was a result of the conclusions drawn in paper I concerning methylation alterations. We wanted to verify these patterns in a larger more population-based cohort.

The 450K arrays can not only be used for methylation analysis but also to analyze genetic gains and/or losses. Therefore, the twelve genetic aberrations previously identified with SNP arrays and described as common in a subset of the ccRCC cohort were analyzed using methylation arrays as well and the agreement between both methods was good (22). Using a predefined region when analyzing gains and losses with the 450K arrays might have led to an underestimation of genetic aberrations in the cohort. Other regions that might be of interest were not analyzed and important information might have been missed, ccRCC biomarkers are summarized by Gulati et al., (2014) (17).
Previously identified genetic aberrations show heterogeneity between studies, and their correlation to clinical outcome might be cohort dependent due to the diversity in identified regions between different studies. Chromosomal losses on 4p, 9p, 9q, and 14q are correlated to survival time in several studies indicating that these aberrations are of relevance for ccRCC tumor progression (17). In our study, the correlation between genetic aberrations and clinical groups of patients (Mo-PF, Mo-P, and M1) was analyzed. Out of the twelve regions, gain of chromosome arm 7p and loss of chromosomal arms 9p and 9q showed a significant difference when all three groups were compared. Only the loss of chromosome arm 9q was significantly more common in Mo-P compared to Mo-PF, whereas loss of chromosome arms 9p, 9q, and 14q were significantly more common in metastatic disease compared to Mo-P. Interestingly, the region lost on 9q was enriched for DM-CpGs indicating an epigenetic regulation in this region. This contributes to the previous finding that chromosome 7, 9 and chromosome arm 14q constitute genomic regions of importance for ccRCC tumor progression.

In both papers I and II, overall methylation alterations were analyzed. As paper II was an expansion of the study in paper I some analyzes are identical for the two studies. In paper II, the number of patients was 2.5 times higher than in paper I (115 in paper II compared to 45 in paper I), the cohort was more selectively chosen in paper I compared to a more population-based cohort in paper II (58% M1 in paper I compared to 24% M1 in paper II), and the number of analyzed CpGs, in paper II, was more than seven times higher.

Hierarchical cluster analysis on promoter-associated CpGs in both studies divided the tumors into two distinct clusters, denoted A and B. Cluster B was associated with a higher level of average methylation and a higher number of hypermethylated CpGs. In both studies patients belonging to cluster B had poorer prognosis and shorter cancer-specific survival. When the analysis was restricted to non-metastatic tumors only (paper II) there was a significant difference in the cumulative incidence of tumor progression. This indicates that when analyzing a high number of CpGs, the ccRCCs can be divided into groups correlated to clinical outcome of patients. Our studies on ccRCC have identified similar DNA methylation patterns as described in previous studies, confirming that general hypermethylation correlated to poor prognosis (69,94-96).
The TF-tissue samples clustered together with tumors in cluster A in paper II and in close proximity to cluster A in paper I. This indicates that the methylation pattern is altered even though the TF-tissue is histologically normal. This is in line with previous results showing different methylation patterns in normal tissue from a tumor-free organ, compared to histological normal tissue from a tumor-bearing organ in both kidney and lung (69,71), and what has been presented as TINT (tumor-indicating normal tissue) in prostate cancer (97). In colorectal cancer, the difference in methylation between cancer adjacent polyps (CAP) and cancer-free polyps (CFP) has been analyzed (70). A significantly higher average methylation level in analyzed CpGs was seen in CAP compared to CFP, which might be due to the close proximity to a tumor in CAP and that close proximity to a tumor can affect methylation patterns (70).

Today, patients with small ccRCCs undergo partial nephrectomy to spare kidney function post-operatively. When comparing outcome in T1 tumors depending on radical (RN) or partial nephrectomy (PN) no difference in time to progression or overall survival could be established. One important benefit of PN compared to RN is the reduced risk of chronic kidney disease post-operative. Therefore, larger tumors have also been enrolled to be treated with PN, the risk of tumor progression in this group is yet to be evaluated (98).

The correlation between higher levels of methylation and tumor development in several types of cancers has led to the development of DNA-demethylating agents for treatment. The demethylation agents used today, i.e. 5-azacytidine and 5-aza-2’-deoxycytidine, have been approved for the treatment of myelodysplastic syndrome (MDS) and acute myeloid leukemia (AML) and have improved the outcome in these diseases. These agents have not been approved or found suitable for treatment of solid tumors, that depend on the instability of the molecule, its short half-life and above all, the lack of clinical response (99).

In paper I, genes represented by DM-CpGs were analyzed in relation to a list of 1,893 genes presented as polycomb repressor complex 2 (PRC2) target by Lee et al., (2006) and one-fourth of the DM genes was a target for PRC2 (93) indicating epigenetic regulation of these genes. The PRC2 complex can both attract DNMTs to induce DNA methylation (100) and the PRC2 subunit AEBP2 favors binding methylated DNA and induce H3K27 methylation (101). This indicates that DNA methylation of PRC2 target genes can both be induced by PRC2 binding as well as induce PRC2 recruitment. EZH2 is a catalytic factor in the PRC2 complex that methylates H3K27, which in turn induces heterochromatin formation and gene silencing. EZH2 overexpression is associated with prostate, bladder, breast, and gastric cancer as well as hepatocellular and renal cell carcinoma (41).
Degerman et al., (2014) identified an accumulation of hypermethylated CpGs with the number of cell divisions in immortalized T-cells (102). This made us hypothesize that metastasized tumors were biologically older than non-metastatic tumors. The method described by Horvath (2013) to calculate biological age was created for several different tissues and when analyzing different tumors they were predicted both older and younger than normal tissue (60,103). Interestingly, in paper I there was a trend towards lower biological age in tumors belonging to cluster B compared to cluster A even though the number of hypermethylated CpGs was significantly higher. In paper II, the EpiTOC model was used to calculate the mitotic age in tumor tissues (61). There was an accumulation of both number hypermethylated CpGs and the total number of genetic aberrations associated with increased mitotic age (61). Metastatic tumors harbor significantly higher numbers of hypermethylated CpGs and genetic aberrations compared to non-metastatic tumors and also have a significantly higher mitotic age. This made us speculate that metastatic tumors have a longer and/or more accelerated proliferative history than non-metastatic tumors.

One challenge in ccRCC is to identify the 30% of non-metastatic tumors that later will progress. In Sweden today, non-metastatic tumors are scored using the Mayo scoring system with different follow-up strategies depending on the risk group, with the goal of early detection of tumor progression. None of the identified genetic and/or epigenetic biomarkers for ccRCC are established in clinical use as the reproducibility between different cohorts is scarce. Joosten et al., (2017) drew the conclusion that a combination of several CpGs and/or genes might be beneficial in creating biomarkers for ccRCC (85). This made us investigate different approaches to predict outcome in non-metastatic ccRCC.

The knowledge that hierarchical cluster analysis on promoter-associated CpGs could separate tumors with significant different methylation patterns and outcome was the starting point for building our prognostic classifiers. We wanted to create a prognostic classifier that could predict outcome in a single ccRCC sample after the classifier was built. Even though a centroid point for each cluster can be calculated and a new sample can be related to that we decided against that method. Instead, we analyzed the data in three different ways and identified two classifiers for ccRCC tumor progression.

Firstly, in paper II, principal component analysis (PCA) was used to investigate if prognostic groups showed different methylation patterns. Tumors belonging to the same cluster showed similar methylation patterns whereas prognostic groups had overlapping methylation patterns and the hypothesis was therefore discarded.
Secondly, in paper II, CpGs identified as DM in Mo-P and M1, but not in Mo-PF, were used to create a promoter methylator classifier (PMC). The PMC panel divided non-metastatic tumors into PMC high and low with a strong association with the risk of tumor progression. The clinical relevance remained significant in a multivariate analysis including PMC status, TNM, morphological grade, age, and gender and was also validated using the TCGA-KIRC cohort. The prognostic classifier presented by Wei et al., (2015) predicting overall survival was also applied to our cohort to elucidate if it could predict progress in non-metastatic tumors (84). Almost 50% of Mo-P tumors were classified as low risk and there was no significant difference in CIP between low- and high-risk tumors indicating that the Wei Risk score can not be used as a prognostic classifier for progress in non-metastatic ccRCC.

Thirdly, in paper III, an unsupervised approach combining clinicopathological variables with previously identified CpGs of prognostic relevance and DCA clusters was used. In total, six classifiers were built with a fixed sensitivity at 85% and compared to the clinically used Mayo scoring system (8). The high fixed sensitivity might lead to a reduced specificity. The requirement of the specificity for a prognostic marker depends on the treatment program for high-risk patients, i.e. close monitoring or if to enroll them in adjuvant treatment. Adjuvant treatment in non-metastatic ccRCC has been reported as beneficial in the S-TRAC study, with longer progression-free survival time in the sunitinib arm, but no data are yet available for overall survival (10). There was no survival advantage in none of the other adjuvant TKI studies in non-metastatic ccRCC, e.g., but a high proportion of adverse side-effects and adjuvant treatment for non-metastatic tumors prior to progression is not recommended (9). New immunotherapies have been proven beneficial in the treatment of metastatic ccRCC. These therapies block the inhibitory T-cell receptor RD-1 or cytotoxic T-lymphocyte associated antigen 4 (CTLA-4) signaling to restore tumor-specific T-cell immunity and are shown to beneficial for patients with metastatic disease (13). Due to the high risk of adverse side-effects with adjuvant treatment, there is a need for high specificity of risk classifier markers for tumor progression before enrolment of patients with non-metastatic disease in adjuvant treatment program (9,13,104).

The classifier with the best specificity, the triple classifier, combined all included variables and had a specificity of 64%. The PMC panel had a specificity of 55% which was lower than the triple classifier but still higher compared to the Mayo scoring system which had a specificity of 50% in our cohort. It has previously been shown that the combination of methylation and gene-expression profiles with clinical variables can predict overall survival in ccRCC (105). This knowledge supports our result that a combination of molecular biomarkers and clinical variables improves risk-classification in non-metastatic ccRCC.
Previous analysis concerning heterogeneity in ccRCC have identified differences in the mutational rate and metabolic patterns from different samples from the same tumor (18,106). ClearCode34 is a ccRCC classifier that uses gene expression levels of 34 genes to separate non-metastatic tumors into ccA or ccB with poorer outcome in the ccB subgroup (107). ClearCode34 is a “slimmed-down version” of the 110 gene classifier presented by Brannon et al., (2010) (108). Heterogeneity analysis with several samples from the same tumor, and the 110 gene panel, identified eight out of ten tumors classified as both ccA and ccB (17). Even though, there was a difference in methylation patterns between tumors in the same prognostic group we could identify a homogenous methylation pattern within the individual tumor. This is in line with the analysis presented by Wei et al., (2015) whose risk classifier could correctly classify over 90% of tumor samples analyzed for intra-tumoral heterogeneity (84). This indicates that methylation alterations are more stable within an individual tumor and therefore might be better suited in developing prognostic markers for ccRCC. Another advantage with a DNA based marker is that DNA is more stable than, eg. RNA, and is routinely extracted in most clinics.
General conclusions and future perspectives

In this thesis, DNA methylation classification was identified as a prognostic marker for tumor progression in ccRCC. We identified a stepwise increase in average methylation and the number of hypermethylated CpGs at increasing tumor stages. A high number of methylation alterations at diagnosis correlated to shorter cancer-specific survival and tumors more prone to progress. We have created two different classifiers that used DNA methylation profiles, with or without combination with clinical variables to predict tumor progression in non-metastatic tumors.

Our analysis was solely focused on CpGs located within gene promoter regions. The decision not to analyze DNA methylation in non-coding regions, in samples analyzed with the 450K arrays, might have led to a miss in information correlated to tumor progression. Demethylation of non-coding regions can contribute to genomic instability and it would be interesting to correlate genetic aberrations to DM-CpGs in intrinsic regions and correlate them, if possible, to ccRCC tumor progression. However, this analysis would preferably be performed on whole genome bisulfite-sequenced DNA.

Even though the methylation analysis developed by Illumina Inc. easily can be implemented in the laboratory work, there are still some challenges to over-build. At many diagnostic hospitals in Sweden, including Norrlands Universitetssjukhus, Umeå, analysis of genetic aberrations using the SNP arrays from Illumina Inc. are already implemented. However, the amount of information derived from the methylation arrays demands bioinformatic knowledge to pre-process and adapt the data before applying it to an available script to calculate the risk for tumor progression. Although bioinformatics pipelines have been set up at many hospitals, this is still a limiting factor for the integration of our predicted DNA methylation-based risk classifier for ccRCC.

The epigenetic regulation of gene expression is a complex machinery, with the interaction between the different regulatory mechanisms. The consequence of specific DNA methylation alterations for gene expression has not been analyzed in detail in this thesis. It would be of utmost interest to further investigate the functional consequence of specific epigenetic alterations for gene expression. Functional analysis might contribute to deepened knowledge of the role of DNA methylation in ccRCC progression and could be beneficial in identifying targets for treatment. Methylation analysis can also be used to select patients for adjuvant treatments. Hopefully, patients identified as high risk for recurrence by DNA methylation might be helped by the new immunotherapies at hand.
The long term goal of this project is to further evolve the prognostic marker to include the best combination of methylation markers and clinical variables, to verify it in additional ccRCC cohorts and lastly implement it in clinical use.
Acknowledgment

I never thought this day would come, for me to defend my thesis. Since acceptance as a Ph.D. student, I have had two wonderful kids, medical school, internship and now finally this day. It would not have been possible without my supervisors, collaborators, friends, and family.

To my supervisor Sofie Degerman, we started off as fellow Ph.D.-students and now a lector and (hopefully soon) a Ph.D. A role model in laboratory work and organization, a hard-working and dedicated supervisor that can keep a lot of projects and work in her head at once. Always ready to lend a helping hand, all day of the week. Through thick and thin (and a lot of chocolate cakes), we made it!

To my co-supervisor Göran Roos for sticking by my side through the years, first as a supervisor and later as a co-supervisor. Thank you for giving me the opportunity to become a part of your research group. It took a bit longer than expected but kids and medical school came in between. If some reason I do have an urge to answer with a “schooo” every time you ask me something.

Börje Ljungberg, my co-supervisor since the beginning. Your clinical knowledge has been so important to this thesis and a big help along the way.

Magnus Hultdin, my newest co-supervisor. Intelligent, funny and always with a big smile. Your contributions to me, and the group as a whole, have made me want to work a bit harder each day. We will see, maybe I one day will become a pathologist as well.

“The Group”: Zahra my Pakistani friend and co-worker. Intelligent and hard-working, you are a true inspiration and your Swedish skills are top-notch. Always up for “äventyr”? Mattias the statistician that has had to endure all my million questions concerning everything. All emails with the subject “Heeeeeeej”, without you this thesis had been nothing. Oliva the new PhD-student, funny and open-minded, you will be great! Pia both former and current member of the group. Your technical skills are amazing. Former members; Ulrika went from study mates at the beginning of the 21st century to Ph.D. students in the same group. You inspire me with your intelligence and your sewing skills. Magnus B, you are one of the smartest and funniest men I know, thanks for the years as co-working Ph.D-students. Linda, you took care of me when I started as a “stipendiat” in Irina’s lab. Bright and cool, I’ve learned a lot from you. Helene, we never worked with the same project but you are a fun and inspiring woman in so many ways. Ravi, the other kidney cancer Ph.D. in the group, thanks for inspiring discussions through the beginning of my Ph.D. studies.

Thanks to the collaborating group of Patrik Rydén, finally we finished the last manuscript. A special thanks to Linda Vidman that did not give up on me and all my questions about the statistical stuff that “MatStat” contributed to the last manuscript and this thesis.
Fikamänniskorna, the once that drinks coffee (and tea) and have lunch at 6M 2nd floor. The once that like chocolate and have taken care of me when I have been in and out from the lab. Smart, bright and funny (and good looking) folks that make even the hard days a bit funnier.

Ida, Terry, and Carina the administrative staff that have made this journey easier for me. Percentage of employment, registration of courses, salary, parental leave, and everything in between.

And to all other employees at 6M for making all these years as a Ph.D. fun, interesting and worthwhile with interesting seminars and discussions.

My dear friends/Mina kära vänner:

Pumorna, ni som funnits där genom alla år. Ni som alltid har tid att ses när vi är på samma breddgrader. Ni som känner mig inifrån och ut oavsett hur lång tid det går mellan gångerna vi ses och hörs är allt alltid "som vanligt". Ni som älskar mig innerst inne och ytterst ute, jag älskar er på samma sätt!

Ida, du får stå här också. Så många gånger du har fått mig att skratta så jag fått träningsvärk, så många gånger jag suttit på ditt skrivbord. Min eldfluga på 6M.

Angelica tack för alla felspringningar och kaffedejter. Så himla skönt att bara kunna dyka upp hemma hos dig och din familj, dricka kaffe och kolla på barn som leker och känna lugnet.


Elin, utan dig (och din familj) vet jag inte hur mitt liv skulle se ut. Mina barn har fått en storasyster och en lille-/storebror. Du är en trygghet att ha som vän, du finns alltid där, ställer alltid upp och du är bland de bästa som finns. Att bara kunna andas hemma hos dig, lyssna på musik eller dra på äventyr. You see the whole of the moon.

Team Tallberg ni vet vilka ni är, Göteborgsvarv, plump och underbara släkthelger med världens bästa Mormor som mittpunkt. Farfar, farbröder, fastrar och kusiner alltid lika underbart att komma "hem" och träffa er.

Övriga vänner, ni som där runt omkring mig. Ni som peppar, stöttar och finns där för att göra mitt och barnens liv ännu lite bättre.

Sist men inte minst...


Det bästa jag gjort i livet: Hugo och Hanna, ni gör livet värt att leva. Älskar er av hela mitt hjärta och varje dag med er gör mig lycklig.
References


64. Feinberg AP, Vogelstein B. Hypomethylation distinguishes genes of some human cancers from their normal counterparts. Nature 1983;301(5895):89-92 doi 10.1038/301089a0.


