


RESEARCH ARTICLE

Open Access



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

Emma Andersson Evelönn<sup>1</sup>, Mattias Landfors<sup>1</sup>, Zahra Haider<sup>1</sup>, Linda Köhn<sup>2</sup>, Börje Ljungberg<sup>3</sup>, Göran Roos<sup>1</sup> and Sofie Degerman<sup>1\*</sup> 

## Abstract

**Background:** Clear cell renal cell carcinoma (ccRCC) is the most common subtype among renal cancer and is associated with poor prognosis if metastasized. Up to one third of patients with local disease at diagnosis will develop metastasis after nephrectomy, and there is a need for new molecular markers to identify patients with high risk of tumor progression. In the present study, we performed genome-wide promoter DNA methylation analysis at diagnosis to identify DNA methylation profiles associated with risk for progress.

**Method:** Diagnostic tissue samples from 115 ccRCC patients were analysed by Illumina HumanMethylation450K arrays and methylation status of 155,931 promoter associated CpGs were related to genetic aberrations, gene expression and clinicopathological parameters.

**Results:** The ccRCC samples separated into two clusters (cluster A/B) based on genome-wide promoter methylation status. The samples in these clusters differed in tumor diameter ( $p < 0.001$ ), TNM stage ( $p < 0.001$ ), morphological grade ( $p < 0.001$ ), and patients outcome (5 year cancer specific survival (pCSS<sub>5yr</sub>)  $p < 0.001$  and cumulative incidence of progress (pCIP<sub>5yr</sub>)  $p < 0.001$ ). An integrated genomic and epigenomic analysis in the ccRCCs, revealed significant correlations between the total number of genetic aberrations and total number of hypermethylated CpGs ( $R = 0.435$ ,  $p < 0.001$ ), and predicted mitotic age ( $R = 0.407$ ,  $p < 0.001$ ). We identified a promoter methylation classifier (PMC) panel consisting of 172 differently methylated CpGs accompanying progress of disease. Classifying non-metastatic patients using the PMC panel showed that PMC high tumors had a worse prognosis compared with the PMC low tumors (pCIP<sub>5yr</sub> 38% vs. 8%,  $p = 0.001$ ), which was confirmed in non-metastatic ccRCCs in the publically available TCGA-KIRC dataset (pCIP<sub>5yr</sub> 39% vs. 16%,  $p < 0.001$ ).

**Conclusion:** DNA methylation analysis at diagnosis in ccRCC has the potential to improve outcome-prediction in non-metastatic patients at diagnosis.

**Keywords:** Clear cell renal cell carcinoma, DNA methylation, Prognosis, Genetic

## Background

Clear cell renal cell carcinoma (ccRCC) is the most common histological subtype accounting for 80–90% of all RCCs. Clear cell RCC is associated with few clinical symptoms; i.e. flank pain, hematuria or a palpable abdominal mass, but is nowadays mostly discovered incidentally due to extensive use of computed tomography

(CT), ultrasound, and magnetic resonance tomography (MRT) [1]. This has led to earlier discovery of tumors, and the number of patients with metastases at diagnosis has decreased to 18% in Sweden [2]. If metastases are present at diagnosis, the probability of 5-year survival (pOS<sub>5yr</sub>) may be as low as 10–15% [3]. Among patients with local disease, the prognosis is better (pOS<sub>5yr</sub> up to 90% with modern protocols), but still 20–30% of patients with local disease at diagnosis will develop metastasis after nephrectomy [3]. In addition to TNM stage and morphological grade [4, 5], there is a need for new

\* Correspondence: [sofie.degerman@umu.se](mailto:sofie.degerman@umu.se)

<sup>1</sup>Department of Medical Biosciences, Umeå University, NUS, Blg 6M, 2nd floor, SE-90185 Umeå, Sweden

Full list of author information is available at the end of the article



molecular markers to identify patients with high risk of progress.

Genetic alterations in ccRCC show considerable heterogeneity. Loss of chromosome 3p and inactivation of the *VHL* (von Hippel–Lindau) gene are frequently observed [6]. Gain of chromosomes 1q, 3q, 7q, 8q, 20q; loss of chromosomes 1p, 4p, 4q, 9p, 9q, 13q, 14q and loss of whole chromosomes 4, 9, 19, 20 and 22, have all been reported in ccRCC [7–12].

Several studies have aimed at identifying molecular markers that predict survival in ccRCC. Gene expression alterations have been associated with prognosis [13–26], but none of these genes are currently clinically used.

DNA methylation has emerged as an important regulator of gene expression, and has been implicated in both cancer development and progression. DNA methylation on Cytosine-phosphate-Guanine (CpG) sites in promoter regions may alter the affinity of transcription factors for their binding sites, and may also, in combination with chromatin modifications, contribute to silencing of genomic regions [27]. Altered DNA methylation has been identified as a prognostic marker, as well as a potential target for therapy, in several malignancies [27, 28]. De novo methylated CpGs in ccRCC assumed to be of relevance for RCC tumorigenesis have been identified, but their clinical value requires further validation [29, 30]. Arai et al., (2012) and Tian et al., (2014) identified CpG island methylator phenotype (CIMP) panels using the Infinium Human Methylation27K array and MassARRAY, respectively, that predicted cancer-free survival and overall survival [31, 32]. In 2015, Wei et al. presented a CpG-methylation-based assay using the Illumina HumanMethylation450K array, calculating a risk score that predicted overall survival independently of clinicopathological parameters in ccRCC [33].

We have previously shown that genome-wide promoter methylation status can predict survival in ccRCC [34]. Using Illumina HumanMethylation27K arrays, we found a stepwise increase in methylation with TNM stage and morphological grade. In the present study, we increased the number of patients and performed a detailed analysis of promoter associated CpGs by Illumina HumanMethylation450K arrays. Thereby, we further investigated the prognostic value of alterations associated with tumor progression. Identifying methylation patterns at diagnosis unique for non-metastatic patients with high risk of later progress is important since these patients may need adjuvant treatment and/or more frequent follow up to improve survival.

## Methods

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

## Patients and tissue samples

The study cohort consisted of 115 ccRCC patients, primary treated with radical or partial nephrectomy between 2001 and 2009, and diagnosed at the University hospital in Umeå, Sweden. None of the patients received neoadjuvant or adjuvant therapy. Eighty-seven patients were metastasis-free (M0), while 28 had metastases (M1) at diagnosis. Tumor free (TF) tissue samples were obtained from 12 surgically removed tumor bearing kidneys and were considered histologically normal by a pathologist. The tumor and TF tissue samples obtained were snap-frozen in liquid nitrogen, and stored in  $-80^{\circ}\text{C}$  until analysis.

Patients were followed-up at least yearly by routine clinical and radiological examination in accordance with a scheduled follow-up program. Clinical follow-up data were extracted in August 2017. All patients have given informed consent and the study was approved by the regional ethical review board in Umeå (Dnr 2011–156-31 M, 20110523).

The publically available TCGA-KIRC dataset was used as a validation cohort and clinical information was downloaded from the Broad Institute's Genome Data Analysis Center Firehose (<http://gdac.broadinstitute.org/>). Only unique non-metastatic (M0) ccRCC samples (technical replicates excluded) analysed with Illumina HumanMethylation450K array were included in the analysis ( $n = 230$ ). All patients were treated with radical or partial nephrectomy and patients receiving neoadjuvant and/or adjuvant therapy were excluded.

## Methylation array analysis

DNA was extracted from the tissue samples as described previously [35] and was subjected to bisulfite conversion (500 ng of each sample) using the EZ DNA Methylation Kit (Zymo Research, Irvine, USA) according to the manufacturer's protocol. Bisulfite DNA conversion was verified by MethyLight analysis of the *ALU* gene with the ALU-C4 primer/probe set as described [36]. Genome-wide assessment of DNA methylation was performed using HumanMethylation450K BeadChip arrays (Illumina, San Diego, CA, USA) according to manufacturer's protocol. To each array, 200 ng of bisulfite-converted DNA was applied, and the arrays were scanned with a HiScan array reader (Illumina). The fluorescence intensities were extracted using the Methylation module (1.9.0) in the Genome Studio software (V2011.1). Pre-filtering and normalization steps are shown in Additional file 1: Table S1 and were performed as previously described [37, 38], which excluded the X and Y chromosomes, CpGs with detection  $p$ -value  $> 0.05$  and CpG probes that aligned to multiple loci in the genome or were located less than 3 bp from a known single nucleotide polymorphism [39]. The methylation levels (i.e., the

$\beta$  value) of each CpG sites ranges from 0, corresponding to completely unmethylated DNA, to 1, representing fully methylated DNA. The technical reproducibility of methylation array analysis was monitored by including a replicate sample on each array and the  $R^2$ -values ranged from 0.995 to 0.997.

The analysis was restricted to the  $n = 155,931$  CpGs located in gene promoter regions, i.e. located within TSS1500, TSS200 and 5'UTR, remaining after the initial filtration steps (Additional file 1: Table S1). Twelve TF tissue samples were included as reference samples. The TF-samples showed a high similarity in promoter methylation, the  $R^2$ -values ranged from 0.96 to 0.99. A CpG site was determined as differently methylated (DM-CpG) if the absolute value of the difference in beta value between tumor sample and the mean of TF samples ( $\Delta\beta$ ) was greater than or equal to 0.2. The DM-CpGs in the M0-PF (non-metastatic progression free), M0-P (non-metastatic with later progress) and M1 (metastasis at diagnosis) groups were analysed against known methylation quantitative trait loci (mQTLs) [40] based on middle-aged individuals to evaluate genetic variants versus cancer specific alterations, and these sites were excluded from further analysis.

The epigenetic mitotic clock described by Yang et al., 2016 was used to estimate mitotic age [41]. A prognostic Risk score was calculated for 114 out of 115 tumor samples using the CpG-methylation-based assay previously presented by Wei et al., (2015) [33]. Patients with a Risk score higher than  $-0.1$  were defined as high risk and lower than  $-0.1$  were defined as low risk, as previously stated [33].

DM-CpGs within the M0-PF, M0-P and M1 groups were selected for further analysis if differently methylated in at least 70% of samples (in Fig. 7). Hypermethylated CpGs showed increased methylation compared to TF whereas hypomethylated CpGs were less methylated.

The commonly DM-CpGs ( $n = 172$ ) in the M0-P and M1 samples were defined as a Promoter Methylation Classification (PMC) panel. The hypomethylated CpGs ( $n = 51$ ) were mirrored ( $1 - \text{average beta}$ ) and thereafter an average beta of the 172 DM-CpGs were calculated for each sample. A ROC-curve was constructed with PMC average beta as test variable and tumor progression within five years as state variable. Youden index was used to determine the cut off for PMC groups (PMC high/low), and was set to PMC average beta 0.688 (specificity 0.85 and sensitivity 0.55).

The prognostic relevance of PMC classification was confirmed in a separate ccRCC cohort ( $n = 230$ ) within the KIRC project of TCGA [42]. Clinical information along with methylation raw data (Illumina HumMeth450Karrays) were downloaded from the Broad Institute's Genome Data Analysis Center Firehose (<http://gdac.broadinstitute.org/>).

Beta values were constructed using the genome studio definition and was normalized for different bead types using BMIQ. Missing values among the 172 CpG-sites in the PMC panel were imputed using the k-nearest neighbours method. Deaths without prior progression were counted as non-events and were censored.

The distribution of hyper- and hypomethylated CpGs within twelve genomic regions with frequent gain/loss in ccRCC (as defined by Köhn et al. 2014 [9] and listed in Additional file 2: Table S3) was analysed to identify potential overrepresentation of hyper- and/or hypomethylated CpGs within these regions.

#### Heterogeneity analysis

In six individuals, multiple tumor samples were taken from different locations within the same kidney to study intratumoral heterogeneity. To confirm that the collected samples originated from the same individual, the methylation levels of the 65 built in SNP probes (in the methylation array) were analysed. One of the 6 patients (number 3) was excluded at this step due to signs of contaminated DNA (Additional file 3: Figure S1).

#### Genomic aberrations

The genetic aberrations were profiled using the total intensity signals of the raw data from the HumanMethylation450K arrays [43, 44]. Briefly, copy number variation (CNV) analysis was performed in R (v3.4.1) using the Conumee package (v1.9.0) [45] with data imported through minfi (v1.18) [46]. Parameters and limits for calling deletion and gain were set for each sample individually through visual inspection.

The validity of this method to identify CNV was confirmed by comparing the methylation results with genetic aberrations identified by HumanCytoSNP-12 v2.1 arrays (Illumina) in a subset of samples ( $n = 57$  ccRCC) [9]. The CNV analysis was performed in Genome Studio v1.8 using the Genotyping Module (Illumina). Cohen's kappa test was used to compare the results from the two array types, which were significantly overlapping ( $p \leq 0.001$  for all analysed aberrations), with quality of agreement moderate or good (Additional file 4: Table S2).

The genomic aberrations for all ccRCCs were summarized by investigating the minimal overlapping regions of commonly occurring CNVs in ccRCC [9]. The CNVs are presented as percentage of samples where alterations across the entire regions were found (Additional file 5: Figure S2 and Additional file 2: Table S3).

#### RNA preparation and gene expression array analysis

RNA was extracted from 28 tumors using MagAttract RNA Universal Tissue M48 Kit (Qiagen, Hilden, Germany) according to manufacturer's protocol using BioRobot M48. RNA concentrations were determined by

spectrophotometry (NanoDrop, Thermo Scientific, Wilmington, DE, USA) and quality was analysed using the 2100 Bioanalyzer (Agilent technologies, Santa Clara, CA, USA).

Two hundred ng of total RNA from each sample was used for cRNA production by the Illumina TotalPrep RNA amplification kit (Ambion Inc., St. Austin, TX, USA) according to the provided protocol. The quality of purified cRNA was evaluated using the RNA 6000 p kit (Agilent Technologies) in the Agilent 2100 Bioanalyzer (Agilent Technologies). A total of 750 ng biotinylated cRNA was hybridized to the human HT12 Illumina Beadchip gene expression array (Illumina, San Diego, CA, USA) according to manufacturer's protocol and scanned using the Illumina Bead Array Reader (Illumina). Expression array data was analysed using the Illumina BeadStudio V2011.1 software and samples were normalized using the cubic spline algorithm. Gene expression levels of the *MX2* (ILMN\_2231928), *SMAD6* (ILMN\_1767068) and *SOCS3* (ILMN\_1781001) genes were extracted from the arrays.

#### Statistical and bioinformatical analysis

For statistical analysis, the Statistical Package for the Social Sciences (SPSS Inc., Chicago, IL) software version 24, was used. The chi-square/Fisher's exact test was used to compare differences between subgroups among categorical variables and the Mann–Whitney U test was used for continuous variables. Kruskal Wallis test was used for continuous variables when comparing three groups of samples. Mann–Whitney U test with Bonferroni correction was used in Additional file 6: Figure S4.

Estimates of 5-year cancer specific survival ( $pCSS_{5yr}$ ) rates and Cumulative incidence of Progress ( $CIP_{5yr}$ ) in subgroups of ccRCC were obtained from Kaplan–Meier survival tables, and the equality of survival distributions for the groups was compared using the log rank test. The significance level used in all tests was 0.05. In CIP analysis local, regional or distant metastatic progression was the endpoint. In the CSS analysis, ccRCC specific death was the endpoint.

Hierarchical clustering was performed using the Ward's method and a Euclidean distance metric for clustering samples.

Principal Component Analysis (PCA) was performed in SIMCA version 14 (Umetrics, Umeå, Sweden) after centering the average beta values of 155,931 promoter associated CpG sites.

WebGestalt [47] was used to analyse functional enrichment of the genes associated with progress of disease (specified in Fig. 7a), using all genes represented by the 155,931 promoter associated CpGs as the background list. The gene functions of potential relevance for ccRCC

pathology were selected among the 20 most significant GO Terms (Additional file 7: Table S8).

## Results

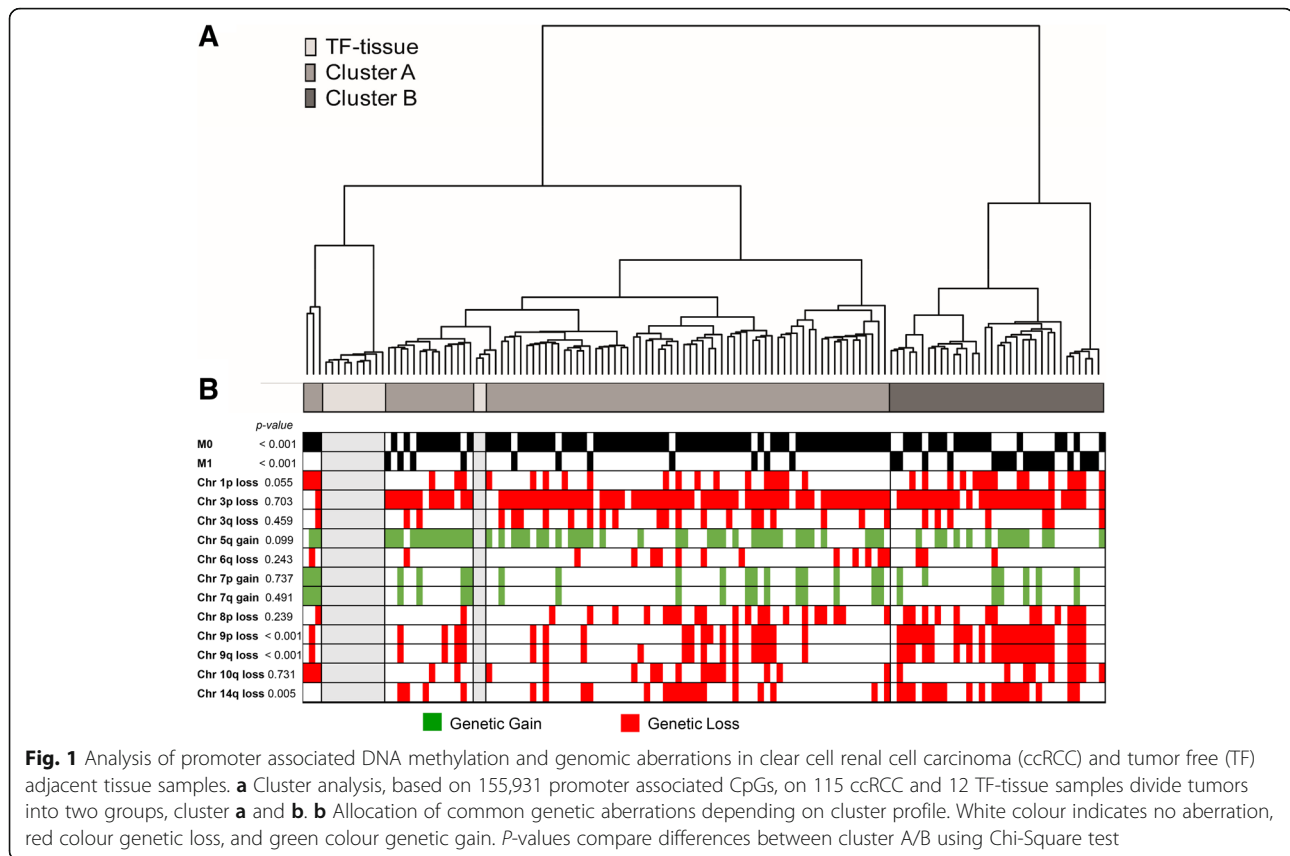
### DNA methylation classification holds prognostic relevance in ccRCC patients

One hundred and fifteen tumor tissue samples from ccRCC patients along with 12 adjacent tumor free (TF) kidney-cortex tissue samples were analysed using the genome-wide HumanMethylation450K arrays.

We have previously shown that genome-wide promoter methylation status based on Illumina Human Methylation 27 K arrays can predict survival in ccRCC [34]. To further examine the prognostic relevance of promoter methylation status with regard to cancer specific survival and progress free survival we focused the analysis to 155,931 promoter associated CpGs on the HumanMethylation450K array, in metastatic and non-metastatic ccRCCs. Hierarchical clustering of the 155,931 promoter associated CpGs divided the ccRCCs ( $n = 115$ ) and the TF samples ( $n = 12$ ), into two clusters; A ( $n = 81$  tumor tissue samples and  $n = 12$  TF samples) and B ( $n = 34$  tumor tissue samples) (Fig. 1). Cluster A contained a high fraction of samples from patients with non-metastatic disease at diagnosis (M0), as well as the TF tissue samples, whereas samples from patients with metastasis at diagnosis (M1) were enriched in cluster B ( $p < 0.001$ ). Cluster B tumors had larger tumor diameter ( $p < 0.001$ ), higher morphological grade ( $p < 0.001$ ), higher TNM stage ( $p < 0.001$ ) and poorer outcome (Table 1, Fig. 2). There was no difference in gender or age at diagnosis between cluster A and B (Table 1).

Cancer specific survival ( $pCSS$ ) analysis confirmed the prognostic relevance of methylation cluster classification in ccRCC ( $pCSS_{5yr}$  80% for cluster A vs 27% for cluster B,  $p < 0.001$ ) (Fig. 2a). As previously known, the difference in CSS is large between patients with distant metastasis (M1,  $n = 28$ ) at diagnosis and patients with non-metastatic disease (M0,  $n = 87$ ) ( $pCSS_{5yr}$  7% vs. 84%,  $p < 0.001$ ) (Fig. 2b). We found no difference in  $pCSS_{5yr}$  in the M1 patients regardless of cluster status ( $pCSS_{5yr}$  9% vs 6%,  $p = 0.840$ ) (Fig. 2c), whereas cluster status clearly separated survival in the M0 patients ( $pCSS_{5yr}$  92% vs 50%;  $p < 0.001$ ) (Fig. 2c). Importantly, the cumulative incidence of progress (CIP) in M0 patients showed a significant difference between cluster A and B ( $p < 0.001$ ), with lower incidence in cluster A patients ( $pCIP_{5yr}$  15%) compared to cluster B patients ( $pCIP_{5yr}$  63%) (Fig. 2d). However, methylation cluster status did not remain as an independent prognostic marker for neither CSS ( $n = 115$ ) nor progression free survival (PFS) ( $n = 87$ ) in multivariate Cox regression analysis including cluster status, TNM-stage, morphological grade, age and gender. In this analysis, only TNM stage,





grade, and gender remained as independent markers (Additional file 8: Table S4 and Additional file 9: Table S5).

To relate cluster A/B status to previously defined prognostic panels in ccRCC, we classified the tumor samples based on the five CpG-site Risk classification panel described by Wei et al., (2015) [33]. Risk score could be calculated for 114 out of 115 tumor samples and were classified as Low ( $n = 63$ ) or High risk ( $n = 51$ ). There was no significant correlation between cluster A/B status and Wei Risk score ( $p = 0.745$ , Table 1). Risk classification according to Wei predicted CSS (pCSS<sub>5yr</sub> 72% for Low risk vs 55% for High risk,  $p = 0.027$ ) (Additional file 10: Figure S3A), but not as strong as genome-wide promoter methylation cluster classification (Fig. 2a). When stratifying patients into non-metastatic ( $n = 86$ ) and metastatic ( $n = 28$ ) groups according to Wei there was neither any statistical difference in CSS (pCSS<sub>5yr</sub> 88% for Low risk vs 78% for High risk  $p = 0.113$ ) in non-metastatic tumors nor between Low risk vs High risk in metastatic tumors (pCSS<sub>5yr</sub> 0% vs 12%,  $p = 0.288$ , Additional file 10: Figure S3B). Further there was no difference in CIP between non-metastatic tumors with low risk versus high risk (pCIP<sub>5yr</sub> 20 and 31% respectively,  $p = 0.099$ ) (Additional file 10: Figure S3C).

#### DNA methylation and genetic aberrations

The frequency of 12 common genetic aberrations in ccRCC as defined by Köhn et al. 2014 [9] (listed in Additional file 2: Table S3 and Additional file 5: Figure S2), were related to methylation status in the 115 ccRCCs. Among the 12 genomic regions, aberrations were found in 15 to 84% of the ccRCCs (Fig. 1, Additional file 2: Table S3, Additional file 5: Figure S2).

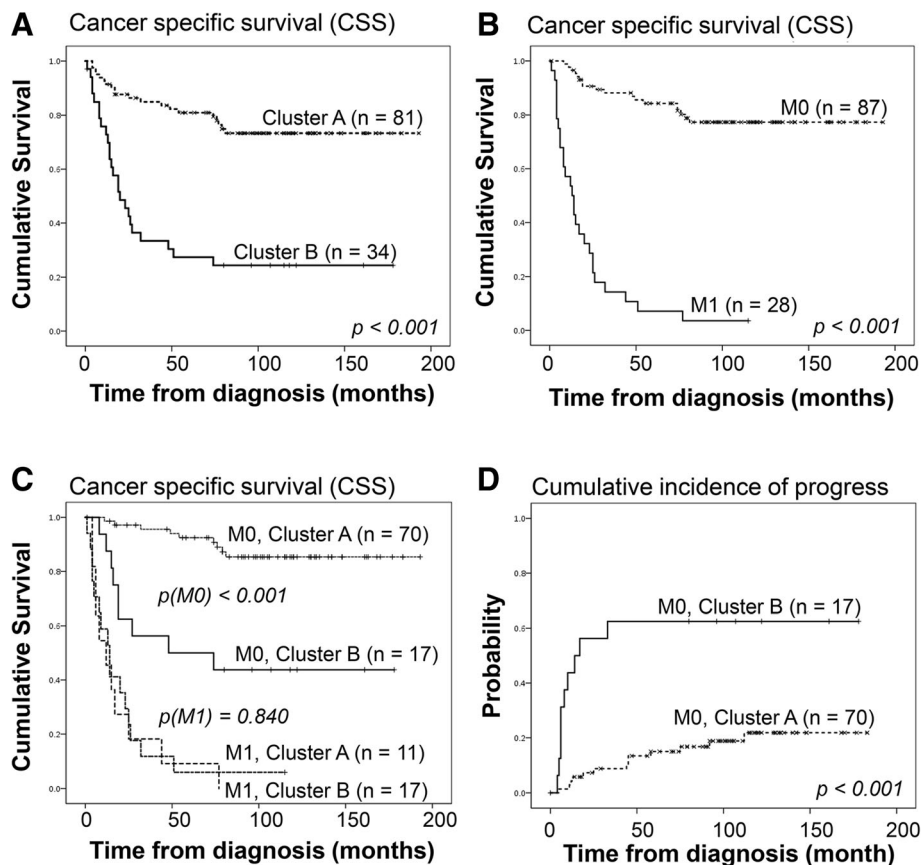
The most common genetic aberration in ccRCC is loss of chromosome 3p (including the VHL gene) [48]. In our cohort, 84% of the ccRCCs showed a loss of 3p, and deletions were similarly distributed in clusters A and B (Fig. 1, Additional file 2: Table S3). Loss of 9p ( $p < 0.001$ ), 9q ( $p < 0.001$ ) and 14q ( $p = 0.005$ ) were more frequently observed in cluster B samples, whereas the other aberrations were similarly distributed among the clusters (Fig. 1, Table 1).

The chromosomal distribution of hyper- and hypomethylated CpGs was analysed in relation to the genomic aberrations in the 12 previously defined regions (Additional file 2: Table S3 and Additional file 6: Figure S4). Although genetic aberrations on chromosome 7p and 9q were associated with overrepresentation of hyper- and/or hypomethylated CpGs, there were no general enrichment of DM-CpGs within genomic regions

**Table 1** Clinicopathological parameters, methylation and genetic alterations and its relation to cluster status

	All Samples (n = 115)	Cluster A (n = 81)	Cluster B (n = 34)	p-value
Age (median ± SD)	65.0 ± 11.6	67.0 ± 11.5	63.0 ± 11.2	0.103 <sup>1</sup>
Gender				
Male	66	42	24	0.064 <sup>2</sup>
Female	49	39	10	
Tumor diameter (mm, median ± SD)	70.0 ± 41.6	55.0 ± 36.1	95.0 ± 38.5	< 0.001 <sup>1</sup>
Morphological grade				
G1	16	16	0	< 0.001 <sup>2</sup>
G2	46	40	6	
G3	34	20	14	
G4	18	5	13	
T Stage				
T1	50	47	3	< 0.001 <sup>2</sup>
T2	25	17	8	
T3	39	17	22	
T4	1	0	1	
M stage				
M0	87	70	17	< 0.001 <sup>2</sup>
M1	28	11	17	
TNM				
I	49	46	3	< 0.001 <sup>2</sup>
II	14	12	2	
III	24	12	12	
IV	28	11	17	
Progress (n = 87)				
No	64	57	7	0.001 <sup>2</sup>
Yes	23	13	10	
Follow Up Status				
Living	39	35	4	< 0.001 <sup>2</sup>
Living with disease	5	3	2	
Dead in ccRCC	45	20	25	
Dead	26	23	3	
Average methylation (median ± SD)	0.3215 ± 0.0136	0.3196 ± 0.0105	0.3344 ± 0.0127	< 0.001 <sup>1</sup>
Loss Chr 9p				
WT	73	62	11	< 0.001 <sup>2</sup>
Loss	42	19	23	
Loss Chr 9q				
WT	75	62	13	< 0.001 <sup>2</sup>
Loss	40	19	21	
Loss Chr 14q				
WT	76	60	16	0.005 <sup>2</sup>
Loss	39	21	18	
Risk score (n = 114) [33]				
Low Risk	63	45	18	0.745 <sup>2</sup>
High Risk	51	35	16	

Differences between groups are compared using <sup>1</sup> Mann-Whitney Test and <sup>2</sup>  $\chi^2$ -test



**Fig. 2** Survival analysis based on M0/M1 and methylation cluster status at diagnosis. Kaplan-Meier cancer specific survival ( $p_{CSS_{5yr}}$ ) analysis in 115 patients with ccRCC, in relation to **(a)** cluster status; **(b)** presence of distant metastasis at diagnosis and **(c)** combination of cluster status and presence of distant metastasis at diagnosis. **d** Cumulative incidence of progress (CIP) analysis in 87 non-metastatic (M0) ccRCC in relation to cluster status. Log-rank  $p$ -values are presented

harboring aberrations (Additional file 6: Figure S4). Importantly, loss of 3p which is frequently observed in ccRCC was not associated with significant accumulation of neither hyper- nor hypomethylated CpGs (Additional file 6: Figure S4).

We observed an intra-individual heterogeneity in the number of aberrations within the M0-PF (progress-free), M0-P (progress) and M1 (distant metastasis) group of patients (Fig. 3). M1 patients had generally more genomic aberrations compared with the M0-PF and M0-P patients ( $p = 0.024$  and  $p = 0.050$ , respectively) (Fig. 3 and Table 2). The only genetic aberration that was more frequent in the M0-P compared to M0-PF patients, was the loss of 9q (43% vs. 20%,  $p = 0.031$ ), which was even more frequent in the M1 patients (61%) (Table 2).

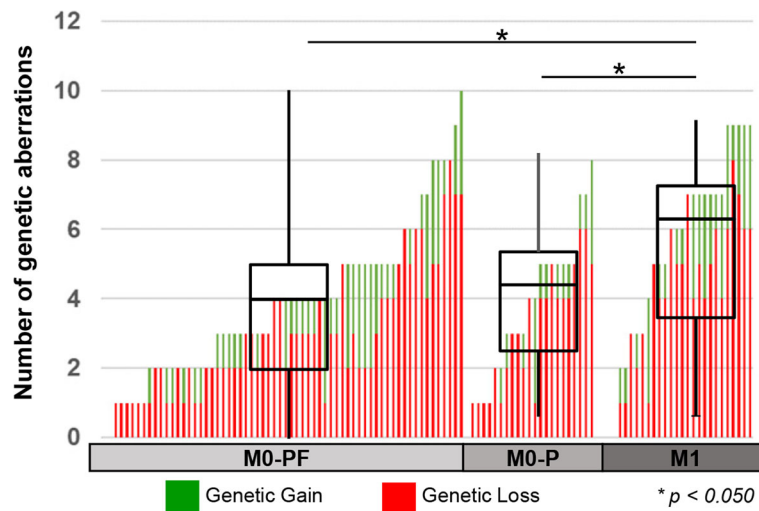
Genetic and epigenetic aberrations may accumulate during extended replication. In order to evaluate such possible associations in ccRCC, we analysed the number of hyper- and hypomethylated CpG sites in relation to number of genetic alterations and estimated mitotic age of the tumor samples. The number of genetic aberrations

were positively correlated to number of hypermethylated CpGs ( $p < 0.001$ ), but not to the number of hypomethylated CpGs in promoters (Fig. 4a, Additional file 11: Figure S5A). The number of genetic aberrations and number of hypermethylated CpGs were both strongly positively associated with estimated mitotic age [41] ( $R = 0.407$ , and  $R = 0.641$ , respectively,  $p < 0.001$ ) (Fig. 4b-c), but there was no correlation between number of hypomethylated CpGs and mitotic age (Additional file 11: Figure S5B).

#### Methylation profiles associated with progression of disease

Since cluster status was of prognostic relevance in M0 patients, we focused the analysis on identifying methylation alterations associated with clinical progress.

M0-PF, M0-P and M1 patients showed gradually larger median tumor diameter ( $p < 0.001$ ), higher TNM stage ( $p < 0.001$ ), higher morphological grade ( $p < 0.001$ ), increased number of hypermethylated CpGs ( $p = 0.015$ ), higher mitotic age ( $p = 0.047$ ), and a higher correlation



**Fig. 3** Distribution of genetic aberrations in ccRCC subgroups. Bar-plot showing number of genetic aberrations in individual samples. Red colour indicates genetic loss, and green colour genetic gain. Inserted boxplots show the median number of genetic aberrations and variations, in the M0-PF, M0-P, and M1 subgroups

to cluster B status ( $p < 0.001$ ) (Table 2). The M0-PF and M0-P groups differed significantly in all these clinico-pathological parameters (except for mitotic age) (Table 2). In contrast, the M0-P and M1 groups differed in TNM ( $p < 0.001$ ) and gender distribution only ( $p = 0.046$ ) (Table 2).

With the aim to distinguish TF, M0-PF, M0-P and M1 subgroups of patients based on the promoter associated methylation profiles, we used principal component analysis (PCA) (Fig. 5). All ccRCC and TF samples were included in the PCA, and different sample categories were highlighted to identify similarities and/or differences. TF tissue samples showed a homogenous methylation pattern (Fig. 5a-b), and cluster A and cluster B patients could be separated (Fig. 5a). However, M0-PF, M0-P and M1 patients could not be clearly separated by PCA since their methylation patterns were partly overlapping (Fig. 5b).

Instead, we aimed to identify specific CpGs associated with progress after primary treatment. DM-CpGs were identified in each tumor sample using the average methylation for TF tissue samples as a reference. Large inter-individual variations in the number of hyper/hypomethylated DM-CpGs was observed in the tumor samples (Fig. 6a). The number of hypermethylated DM-CpGs were significantly higher in the M0-P and M1 patients compared with M0-PF ( $p = 0.025$  and  $0.001$ , respectively) (Fig. 6a, left chart). The number of hypomethylated DM-CpGs did not differ significantly between the groups of patients (Fig. 6a, right chart).

To further dissect possible variations in methylation profiles at intra-tumoral level, we analysed multiple samples taken from various parts of the tumor in five individuals (Additional file 12: Table S6). We found that

samples from the same tumor generally had similar methylation profiles ( $R^2$  correlations 0.97 to 0.99;  $p \leq 0.001$ ). Accordingly, the number of hyper- and hypomethylated CpGs were similar (Fig. 6b), and the individuals' samples clustered together in PCA-analysis (Fig. 6c).

DNA methylation alterations have been described as early events in tumor progression. Therefore, we focused on common DM-CpGs in the M0-P and M1 samples, with potential mQTLs excluded [40], which might represent epigenetic alterations of relevance for the formation of metastasis. CpGs that showed either hyper- or hypomethylation in at least 70% of samples in the M0-PF, M0-P and M1 groups of patients were selected and combined in a Venn diagram (Fig. 7a). Most of the DM-CpGs in the M0-PF patients were also present in the M0-P and M1 patients, indicating that methylation alterations accumulated during progression and was not dominated by stochastic events. Only a low percentage of DM-CpGs were unique for M0-PF and M0-P (13 and 17% respectively) whereas M1 had 56% unique DM-CpGs. 172 DM-CpGs were common for the M0-P/M1 patient groups (121 hyper- and 51 hypo-methylated) (Fig. 7a). These 172 DM-CpGs were defined as a Promoter Methylation Classifier (PMC) panel, and was used to classify samples as PMC low and high as described in the materials and methods section. In the non-metastatic tumors of our cohort, 40 samples were classified as PMC low and 47 samples as PMC high. The  $pCIP_{5yr}$  survival analysis showed poorer outcome in the PMC high subgroup (PMC low  $pCIP_{5yr}$  8% vs. PMC high  $pCIP_{5yr}$  38%,  $p = 0.001$ ; Fig. 7b). The PMC panel (PMC high Hazard ratio (HR) 4.4 (1.3–15.8) and TNM stage (TNM III HR 3.9 (1.4–11.0)) remained significant prognostic



**Table 2** Clinicopathological parameters, methylation and genetic alternations and its relation to ccRCC progression

		M0-PF (n = 64)	M0-P (n = 23)	M1 (n = 28)	p-value all groups of patients	p-value M0-PF vs. M0-P	p-value M0-PF vs. M1	p-value M0-P vs. M1
Age	(median ± SD)	68.5 ± 11.2	64.0 ± 12.7	63.5 ± 11.0	0.103 <sup>1</sup>	0.165	0.093 <sup>3</sup>	0.857 <sup>3</sup>
Gender	Male	34	11	21	0.087 <sup>2</sup>	0.663 <sup>2</sup>	0.049 <sup>2</sup>	0.046 <sup>2</sup>
	Famale	30	12	7				
Tumor diameter (mm)	(median ± SD)	50.0 ± 30.7	90.0 ± 45.7	100.0 ± 35.7	< 0.001 <sup>1</sup>	0.001 <sup>3</sup>	< 0.001 <sup>3</sup>	0.107 <sup>3</sup>
Morphological grade	G1	14	2	0	< 0.001 <sup>2</sup>	0.006 <sup>2</sup>	< 0.001 <sup>2</sup>	0.199 <sup>2</sup>
	G2	31	8	7				
	G3	18	8	8				
	G4	1	5	12				
T Stage	T1	42	7	1	< 0.001 <sup>2</sup>	0.010 <sup>2</sup>	< 0.001 <sup>2</sup>	0.054 <sup>2</sup>
	T2	11	6	8				
	T3	11	10	18				
	T4	0	0	1				
TNM	I	42	7	0	< 0.001 <sup>2</sup>	0.001 <sup>2</sup>	< 0.001 <sup>2</sup>	< 0.001 <sup>2</sup>
	II	11	3	0				
	III	11	13	0				
	IV	0	0	28				
Follow Up Status	Alive	38	1	0	< 0.001 <sup>2</sup>	< 0.001 <sup>2</sup>	< 0.001 <sup>2</sup>	0.126 <sup>2</sup>
	Alive, disease	0	4	1				
	Dead, ccRCC	0	18	27				
	Dead	26	0	0				
Cluster	A	57	13	11	< 0.001 <sup>2</sup>	0.001 <sup>2</sup>	< 0.001 <sup>2</sup>	0.220 <sup>2</sup>
	B	7	10	17				
PMC status	Low	36	4	–	–	0.001	–	–
	High	28	19					
DM-CpGs, tot. Number	(median ± SD)	9191 ± 4779	10,460 ± 3698	12,996 ± 4448	0.001 <sup>1</sup>	0.227 <sup>3</sup>	0.005 <sup>3</sup>	0.130 <sup>3</sup>
Hypermethylated CpGs	(median ± SD)	5826 ± 2820	7617 ± 2990	8233 ± 448	0.015 <sup>1</sup>	0.025 <sup>3</sup>	0.001 <sup>3</sup>	0.150 <sup>3</sup>
Hypomethylated CpGs	(median ± SD)	2810 ± 3173	2843 ± 2990	2688 ± 4249	0.970 <sup>1</sup>	0.810 <sup>3</sup>	0.905 <sup>3</sup>	0.910 <sup>3</sup>
Mitotic Age	(median ± SD)	0.16 ± 0.05	0.18 ± 0.05	0.18 ± 0.06	0.047 <sup>1</sup>	0.132 <sup>3</sup>	0.021 <sup>3</sup>	0.538 <sup>3</sup>
Loss 1p	WT	42	19	18	0.272 <sup>2</sup>	0.127 <sup>2</sup>	0.901 <sup>2</sup>	0.145 <sup>2</sup>
	Loss	22	4	10				
Loss 3p	WT	11	3	4	0.288 <sup>2</sup>	0.643 <sup>2</sup>	0.729 <sup>2</sup>	0.898 <sup>2</sup>
	Loss	53	20	24				
Loss 3q	WT	45	20	21	0.995 <sup>2</sup>	0.115 <sup>2</sup>	0.646 <sup>2</sup>	0.285 <sup>2</sup>
	Loss	19	3	7				
Gain 5q	WT	30	11	13	0.054 <sup>2</sup>	0.938 <sup>2</sup>	0.969 <sup>2</sup>	0.921 <sup>2</sup>
	Gain	34	12	15				
Loss 6q	WT	50	22	26	0.027 <sup>2</sup>	0.056 <sup>2</sup>	0.086 <sup>2</sup>	0.673 <sup>2</sup>
	Loss	14	1	2				
Gain 7p	WT	51	21	17	0.061 <sup>2</sup>	0.206 <sup>2</sup>	0.057 <sup>2</sup>	0.013 <sup>2</sup>
	Gain	13	2	11				
Gain 7q	WT	51	21	18	0.598 <sup>2</sup>	0.206 <sup>2</sup>	0.116 <sup>2</sup>	0.024 <sup>2</sup>

**Table 2** Clinicopathological parameters, methylation and genetic alternations and its relation to ccRCC progression (Continued)

		M0-PF (n = 64)	M0-P (n = 23)	M1 (n = 28)	p-value all groups of patients	p-value M0-PF vs. M0-P	p-value M0-PF vs. M1	p-value M0-P vs. M1
	Gain	13	2	10				
Loss 8p	WT	47	15	18	0.002 <sup>2</sup>	0.455 <sup>2</sup>	0.375 <sup>2</sup>	0.945 <sup>2</sup>
	Loss	17	8	10				
Loss 9p	WT	49	13	11	0.002 <sup>2</sup>	0.069 <sup>2</sup>	0.001 <sup>2</sup>	0.220 <sup>2</sup>
	Loss	15	10	17				
Loss 9q	WT	51	13	11	0.001 <sup>2</sup>	0.031 <sup>2</sup>	< 0.001 <sup>2</sup>	0.220 <sup>2</sup>
	Loss	13	10	17				
Loss 10q	WT	47	20	20	0.361 <sup>2</sup>	0.186 <sup>2</sup>	0.842 <sup>2</sup>	0.180 <sup>2</sup>
	Loss	17	3	8				
Loss 14q	WT	48	13	15	0.076 <sup>2</sup>	0.097 <sup>2</sup>	0.042 <sup>2</sup>	0.833 <sup>2</sup>
	Loss	16	10	13				
Tot. number of alterations	(median ± SD)	4 ± 2	4 ± 2	6 ± 3	0.157 <sup>1</sup>	0.934 <sup>3</sup>	0.024 <sup>3</sup>	0.050 <sup>3</sup>

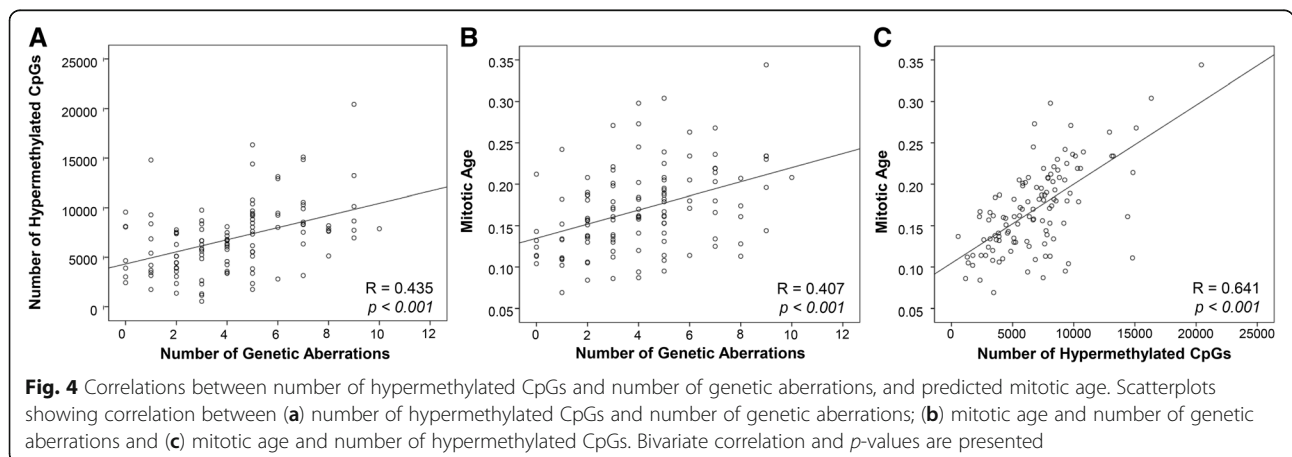
Differences between groups are compared using <sup>1</sup>Kruskall-Wallis Test, <sup>2</sup>χ<sup>2</sup>-test and <sup>3</sup>Mann-Whitney Test

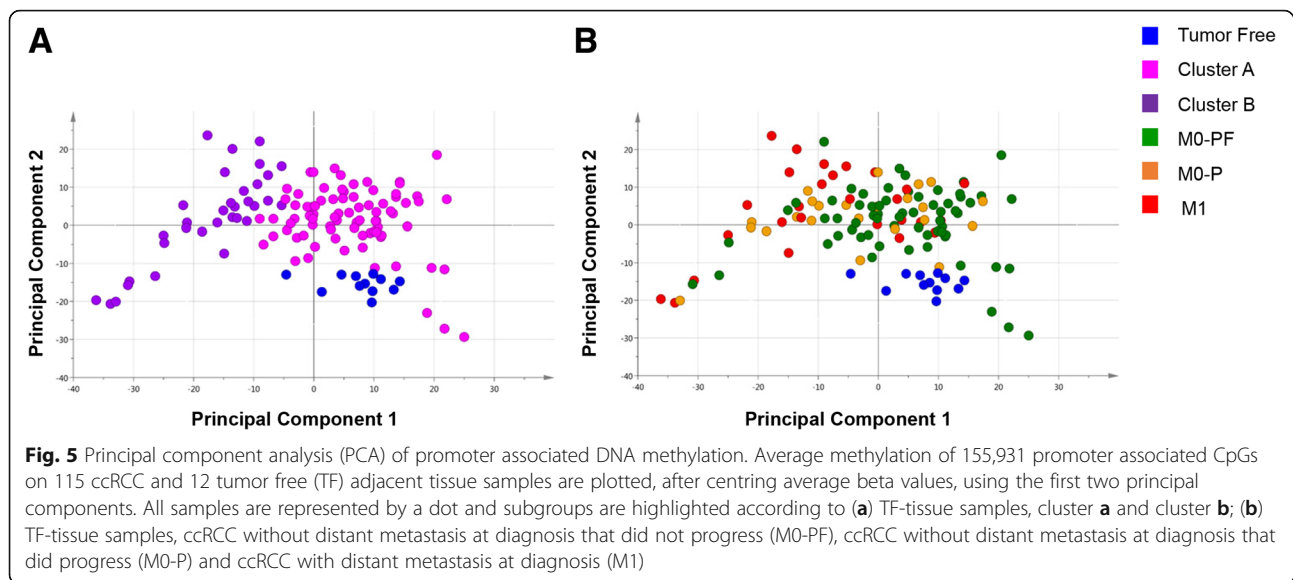
markers for PFS in a Cox regression analysis including PMC status, TNM, morphological grade, age and gender (Additional file 13: Table S7). The prognostic relevance of PMC was confirmed in 230 ccRCC tumor samples from the TCGA-KIRC data set (PMC low (n = 144) pCIP<sub>5yr</sub> 16% vs. PMC high (n = 86) pCIP<sub>5yr</sub> 39%, p < 0.001; Fig. 7c).

The functional relevance of the PMC associated genes were analysed for the most significant gene ontology categories (GO terms) (Fig. 7a, Additional file 7: Table S8). The commonly hypermethylated genes were enriched for GO terms including SMAD protein complex assembly, RNA polymerase II regulation and response to pH (Fig. 7a). The hypomethylated genes were enriched for GO terms associated with external stimulus-, immune- and defence- response (Fig. 7a). Three genes were of special interest with regard to ccRCC progress; i.e.

SMAD family member 6 (SMAD6, CpG cg10402698 and cg12476188), Suppressor of cytokine signaling 3 (SOCS3, CpG cg10279487 and cg27637521) and MX dynamin like GTPase 2 (MX2, CpG cg05656374 and cg15281283). These genes showed significant altered methylation levels between TF/M0-PF and the M0-P/M1 samples in more than one promoter associated CpG site (Fig. 7d-f).

Methylation alterations in these genes were analysed in relation to mRNA gene expression alterations in 28 samples with available RNA (M0-PF = 13; M0-P = 5 and M1 = 10). A significant increased gene expression was observed in M0-P and M1 samples compared to M0-PF samples in the SOCS3 (ILMN\_1781001; p = 0.019 and 0.001) and MX2 (ILMN\_2231928; p = 0.014 and 0.002) genes, whereas no significant difference in expression for SMAD6 (ILMN\_1767068) was seen (Fig. 7g-i).





## Discussion

The aim with this study was to evaluate if genome-wide promoter associated methylation classification can be used as a prognostic tool to identify patients with non-metastatic ccRCC at risk for disease progression. These patients may benefit from alternative therapy approaches such as adjuvant therapy and/or more intense follow-up.

Cluster analysis of 155,931 promoter associated CpG sites divided the 115 ccRCCs into two groups (clusters A and B), where the less methylated ccRCC samples in cluster A clustered together with TF samples. Cluster B status associated with higher average promoter methylation, and this group contained a high fraction of metastasized tumors and patients with local disease who later progressed. This confirmed our previous finding of poor prognosis associated with higher promoter methylation status in ccRCC [34]. Importantly, our current study showed that the prognostic relevance of DNA methylation was limited to M0 patients since the prognosis was very poor regardless of methylation status for patients with metastasized ccRCC at diagnosis.

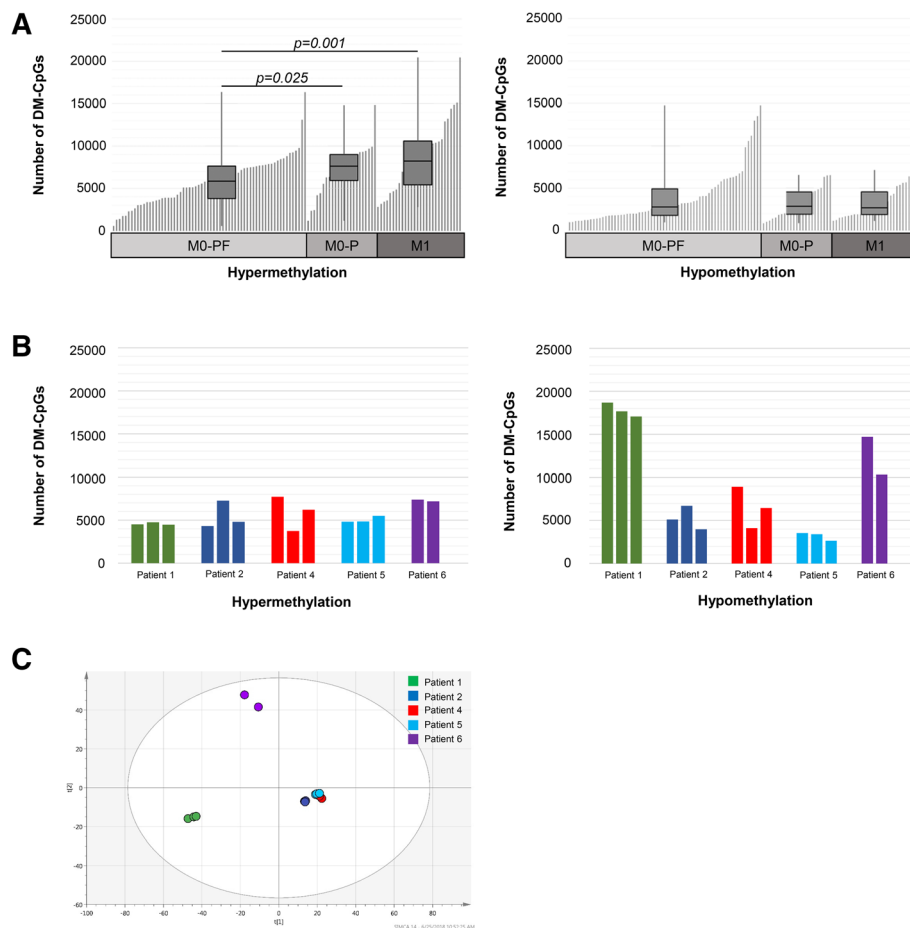
In order to relate our data with previously described DNA methylation panels for risk stratification in ccRCC, we applied the five CpG-site risk panel defined by Wei et al., (2015) on our cohort [33]. We found no significant association between cluster status and Risk group classification. The Wei Risk score could neither separate the survival nor the progress prognosis in non-metastatic tumors, but was predictive for CSS when including the metastatic tumors.

To predict incidence of progress in non-metastatic patients, genome-wide promoter associated cluster analysis seems to be more efficient than risk stratification according to the Wei Risk score restricted to five CpG sites.

The previously described CIMP panel by Arai et al., (2012) could not be applied on our cohort since not all CpGs in the CIMP panel were present in the Infinium HumanMethylation450K array used in our study [31]. However, the Arai CIMP classification has previously shown that high methylation status was associated with poor prognosis, which is in line with our data.

Genomic aberrations are commonly observed in ccRCC and we performed an integrated analysis of methylation status and genomic aberrations in order to identify potential correlations. We used raw data from the methylation arrays to determine genomic aberrations in twelve regions previously defined as harboring common changes (loss or gain) in a subset of our cohort of ccRCC patients [9]. Using methylation arrays to identify genomic aberrations might introduce systematic bias due to segmentation at the ends of several chromosomes. However, the correlation between the identified genetic aberrations by methylation array and SNP-array analysis revealed comparable results in the subset of samples analysed by both techniques. The number of identified genetic alterations were likely underestimated since we focused the analysis on previously defined regions with aberrations [9]. Also, the number of DM-CpGs might be underestimated since we used histological normal tumor-adjacent tissue from a kidney with ccRCC as reference samples. Methylation differences have been reported between histological normal tumor-adjacent tissue and tissue taken from healthy individuals [31].

There was a significantly higher frequency of deletions of 9p, 9q and 14q in cluster B tumors, and these three genetic aberrations have been associated with poorer outcome in ccRCC [7–11, 49, 50]. Importantly, we did not find a general increased number of DM-CpGs in



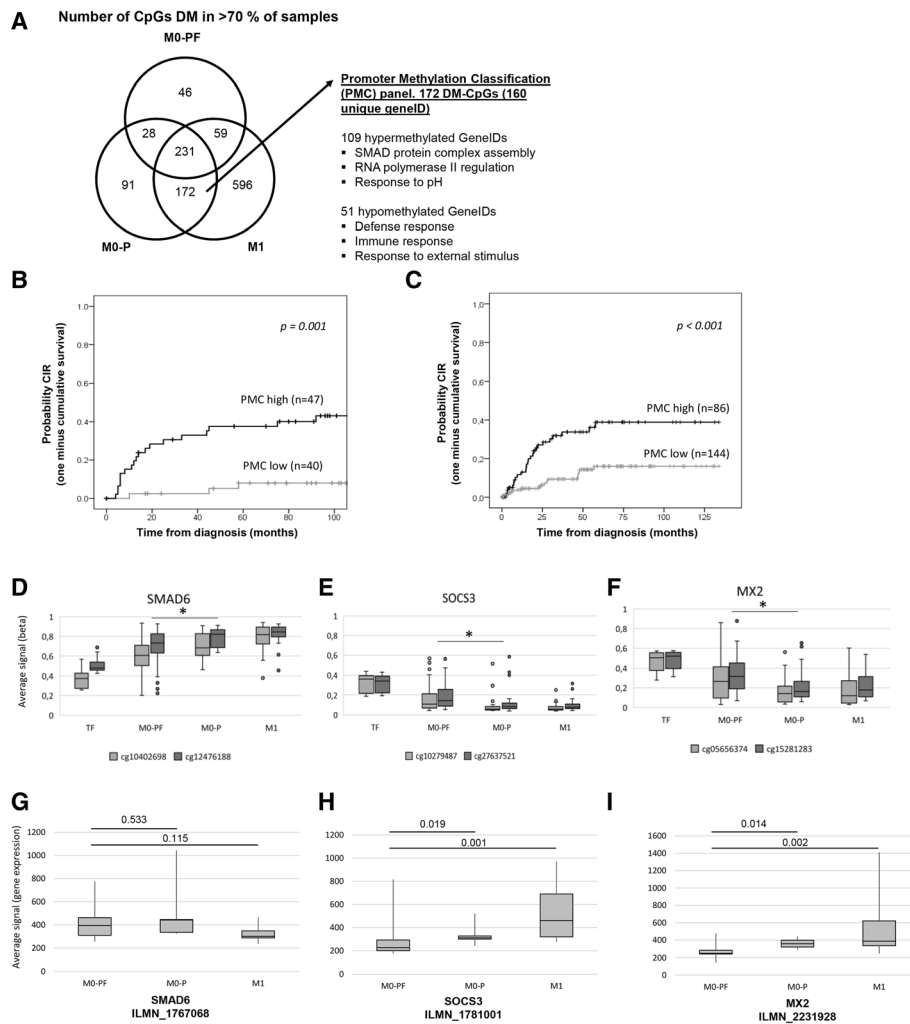
**Fig. 6** Differently methylated CpG sites (CpGs) and its distribution over clear cell renal cell carcinoma **(a)** Bar-plot of number of hyper- (left panel) and hypomethylated (right panel) CpGs, within promoter regions of individual samples in the M0-PF, M0-P and M1 ccRCC groups. Inserted boxplots show number of hyper- and hypomethylated CpGs and its variation within the subgroups. Differences between groups were analysed using Mann-Whitney U-test. **b** Bar-plot for five ccRCC patients (number 1 through 6) represented by two to three pieces from the same tumor, showing the number of hypermethylated (left panel) and hypomethylated (right panel) promoter associated CpGs; **c** PCA analysis of the first two principal components of the average methylation of promoter associated CpGs in multiple samples from five ccRCC patients. All samples are represented by a dot and each individual is highlighted according to legend in figure

regions frequently lost or gained. Loss of chromosome 9q was the only genetic aberration that was more common in M0-P (and M1) patients compared to M0-PF patients. Previous studies have shown significant correlation of loss of chromosome 9q to both histological grade and TNM stage [9, 49, 50] as well as to poor outcome [9]. Patients with 9q loss also showed an enriched number of DM-CpGs within this region. The possible contribution of epigenetic alterations within this region, to poor prognosis, has to be further evaluated.

However, we observed a significant positive correlation of total number of hypermethylated DM-CpGs and total number of genetic aberrations, supporting previous results as shown in a review by Arai and Kanai [51]. In that study a correlation between number of clones with CNVs and total number of DM-CpGs was shown in a subset of ccRCC samples with high methylation levels.

The correlation between number of hypermethylated CpGs, number of genetic aberrations, and predicted mitotic age indicates an accumulation of alterations associated with number of cell divisions. Correlation between genetic and epigenetic alterations was shown previously in both chronic lymphatic leukemia [52] and in breast cancer cell lines [53], but less is known about correlations with mitotic age.

The fact that genome-wide promoter methylation cluster status separated the survival time of patients with non-metastatic disease at diagnosis, made us focus on identifying the specific methylation profiles associated with progress. Initially, PCA was used to investigate whether patient groups with different outcome could be separated based on general promoter methylation patterns. This analysis showed overlapping and heterogeneous methylation patterns within the M0-PF, M0-P and



**Fig. 7** Identification of CpGs associated with progress/metastasis in ccRCC. **a** Venn-diagram showing overlap of CpGs that were differently methylated compared with TF tissue, in more than 70% of samples in each ccRCC group; M0-PF; M0-P and M1. The 172 commonly DM-CpGs in the M0-P and M1 groups were defined as a Promoter Methylation Classification (PMC) panel. Selected GO-terms associated with hyper- and hypomethylated genes in PMC panel are listed. **b-c** Kaplan-Meier cumulative incidence of progress (CIP) **(b)** in 87 non-metastatic ccRCC patients **(c)** in 230 non-metastatic ccRCC patients in the validation TCGA-KIRC cohort. **d-f** Box-plots of selected DM-CpGs in promoter regions of genes associated with progress **(d)** SMAD6 **(e)** SOCS3 and **(f)** MX2. \* $p < 0.05$ . **g-i** Box-plots of mRNA levels for **(g)** SMAD6, **(h)** SOCS3 and **(i)** MX2. Differences in methylation and mRNA levels between groups were analysed using Mann-Whitney U-test

M1 groups, in contrast to homogenous methylation patterns within the TF samples. Also, patients with similar outcome showed large inter-individual variations in methylation patterns. This cannot be explained by intra-individual tumor heterogeneity since multiple samples taken within the same tumor showed similar methylation profiles.

By recognizing the commonly DM-CpGs in the M0-P and M1 samples, we identified a Promoter Methylation Classifier (PMC) consisting of 172 CpGs associated with progress. Classification of non-metastatic patients in PMC high/low subgroups showed strong prognostic relevance for progress in our cohort as well as in the validation TCGA-KIRC cohort. Importantly, the PMC

panel remained a significant prognostic marker for progression free survival in a Cox regression analysis including PMC status, TNM, morphological grade, age and gender.

The genes in the PMC panel were associated with cellular processes including SMAD protein complex assembly and immune response. Genes of special interest previously associated with various types of cancer where the SMAD6 and SOCS3 genes coupled to aggressive kidney cancer [54–56], and the MX2 gene with suggested role in melanoma pathogenesis [57]. Interestingly these genes were differently methylated in the M0-P group, compared with TF and M0-PF samples and a significant difference in mRNA levels was observed for the SOCS3



and MX2 genes. These findings indicate a functional relevance of the methylation alterations in ccRCC pathogenesis but needs to be confirmed in larger samples cohorts.

In a systematic review by Joosten et al., 2017, a number of DNA methylation studies in ccRCC were summarized [30] and the need for validation of identified prognostic markers was claimed. We could not confirm the prognostic relevance of the previously defined Wei risk score (based on five CpG sites) in our cohort. Instead, we could confirm the prognostic relevance of genome-wide promoter methylation cluster analysis (Cluster A/B, >150 K CpGs), suggesting that larger panels are probably more robust. However, in contrast to genome-wide clustering, a defined set of CpGs (or genes) is likely more clinically suitable. In this study, we defined a PMC panel consisting of 172 CpGs, which was a strong prognostic marker for non-metastatic patients in both our cohort and in the validation cohort. Modern bioinformatics tools that combines DNA methylation classification with clinical prognostic markers is an important next step to implement epigenetic analysis in clinical practice.

## Conclusion

Genome-wide promoter-associated DNA methylation associated significantly with genetic aberrations, cellular mitotic age, and clinical parameters, including follow up status. We defined a Promoter Methylation Classification (PMC) panel, including genes of potential relevance for tumor progression. The PMC panel predicted progression free survival in non-metastatic ccRCC patients, which was confirmed in the independent TCGA-KIRC cohort. DNA methylation status has the potential to identify non-metastatic patients with high risk of recurrence already at diagnosis. These high-recurrence risk patients may benefit from alternative therapy approaches and more intense follow-up.

## Additional files

**Additional file 1: Table S1.** Filtration steps in the HumanMeth450K arrays. (PDF 79 kb)

**Additional file 2: Table S3.** Analyzed genomic regions for CNV in 115 ccRCC samples. (PDF 74 kb)

**Additional file 3: Figure S1.** Single nucleotide polymorphism (SNP) analysis. SNP analysis of 65 genotyping probes on the HumanMethylation450K array to confirm patient identity of multiple samples taken from the same tumor. (PDF 256 kb)

**Additional file 4: Table S2.** Comparison of CNV results gained from HumanMethylation450K and HumanCytoSNP-12 arrays in 57 ccRCC samples. (PDF 102 kb)

**Additional file 5: Figure S2.** Copy number variation (CNV) analysis. CNV analysis of twelve chromosome regions previously identified to be altered in ccRCC. The analyzed region is marked in grey. (PDF 205 kb)

**Additional file 6: Figure S4.** Distribution of hyper- and hypomethylated CpGs in patients with or without specific genomic aberrations.

Percentage of (A) hypermethylated and (B) hypomethylated CpGs in the genomic aberration regions associated with ccRCC defined in Additional file 2: Table S3. \* = Bonferroni adjusted  $p$ -value < 0.05. (PDF 75 kb)

**Additional file 7: Table S8.** Top 20 most significant GO Terms for GeneIDs either hyper- or hypomethylated in M0-P and M1 tumor samples. (PDF 54 kb)

**Additional file 8: Table S4.** Cox's proportional hazard regression analysis for cancer specific survival (CSS) in 115 ccRCC samples. (PDF 108 kb)

**Additional file 9: Table S5.** Cox's proportional hazard regression analysis for progress free survival (PFS) in 87 M0 ccRCC samples. (PDF 108 kb)

**Additional file 10: Figure S3.** Survival analysis based on Risk Score at diagnosis. Kaplan-Meier cancer specific survival analysis (pCSS) in 114 ccRCC patients in relation to (A) Wei Risk Score at diagnosis (B) a combination of Wei Risk Score and presence of distant metastasis at diagnosis. (C) Cumulative incidence of progress (CIP) analysis in 86 non-metastatic (M0) ccRCC patients in relation to Wei Risk Score at diagnosis. Log-rank  $p$ -values are presented. (PDF 79 kb)

**Additional file 11: Figure S5.** Correlations between number of hypomethylated CpGs and number of genetic aberrations and predicted mitotic age. Scatterplots showing correlation between (A) number of hypomethylated CpGs and number of genetic aberrations; (B) mitotic age and number of hypomethylated CpGs. Bivariate correlation and  $p$ -values are presented. (PDF 74 kb)

**Additional file 12: Table S6.** Clinicopathological parameters for tumors included in the heterogeneity analysis. (PDF 65 kb)

**Additional file 13 Table S7.** Cox's proportional hazard regression analysis for progress free survival (PFS) in 87 M0 ccRCC samples. (PDF 108 kb)

## Abbreviations

ccRCC: clear cell renal cell carcinoma; CpG: cytosine-phosphate-guanine; DM-CpG: differently methylated CpG site; DMG: differently methylated gene; M0-P: non-metastatic ccRCC at diagnosis that later progress; M0-PF: non-metastatic ccRCC at diagnosis and progress-free at follow up; M1: metastatic ccRCC at diagnosis; PMC: promoter methylation classifier; TF: tumor free; TNM: tumor-node-metastasis

## Acknowledgements

The authors thanks dr Ulrika Svenson for preparation of DNA from the ccRCC tissue samples.

## Funding

The study was supported by grants from the Medical Faculty of Umeå University, Lion's Cancer Research Foundation, Umeå, the Kempe research foundation, Uppsala-Umeå Comprehensive Cancer Consortium and the Swedish cancer foundation. Financial support was provided through a regional agreement between Umeå University and Västerbotten County Council on cooperation in the field of Medicine, Odontology and Health. The funding bodies supported the laboratory costs, salaries, and publication costs, but did not influence the design of the study, analysis and interpretation of data, or in writing the manuscript.

## Availability of data and materials

The datasets generated and/or analysed during the current study are available in the NCBI Gene Expression Omnibus (GEO) repository [GSE113501].

## Authors' contributions

EAE, BL, GR, and SD conceived and designed the experiments. EAE and LK performed the experiments. EAE, ML, ZH, LK and SD analysed the data. EAE and SD wrote the first draft of the manuscript with contribution from all co-authors. All authors read and approved the final manuscript.

## Ethics approval and consent to participate

All patients have given informed and written consent, and the study was approved by the regional ethical review board in Umeå (Dnr 2011–156-31 M, 20110523).

**Consent for publication**

Not applicable.

**Competing interests**

The authors declare that they have no competing interests.

**Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Author details**

<sup>1</sup>Department of Medical Biosciences, Umeå University, NUS, Blg 6M, 2nd floor, SE-90185 Umeå, Sweden. <sup>2</sup>Department of Radiation Sciences, Umeå University, Umeå, Sweden. <sup>3</sup>Department of Surgical and Perioperative Sciences, Urology and andrology, Umeå University, Umeå, Sweden.

Received: 25 September 2018 Accepted: 7 January 2019

Published online: 14 January 2019

**References**

- Siegel R, Ma J, Zou Z, Jemal A. Cancer statistics, 2014. *CA Cancer J Clin*. 2014;64(1):9–29. <https://doi.org/10.3322/caac.21208>.
- Dabestani S, Thorstenson A, Lindblad P, Harmenberg U, Ljungberg B, Lundstam S. Renal cell carcinoma recurrences and metastases in primary non-metastatic patients: a population-based study. *World J Urol*. 2016;34(8):1081–6. <https://doi.org/10.1007/s00345-016-1773-y>.
- Thorstenson A, Bergman M, Scherman-Plogell AH, Hosseinnia S, Ljungberg B, Adolfsson J, et al. Tumour characteristics and surgical treatment of renal cell carcinoma in Sweden 2005–2010: a population-based study from the National Swedish Kidney Cancer Register. *Scand J Urol*. 2014;48(3):231–8. <https://doi.org/10.3109/21681805.2013.864698>.
- Greene FL, Sobin LH. The TNM system: our language for cancer care. *J Surg Oncol*. 2002;80(3):119–20. <https://doi.org/10.1002/jso.10114>.
- Fuhrman SA, Lasky LC, Limas C. Prognostic significance of morphologic parameters in renal cell carcinoma. *Am J Surg Pathol*. 1982;6(7):655–63.
- Kaelin WG Jr. The von Hippel-Lindau tumor suppressor gene and kidney cancer. *Clin Cancer Res*. 2004;10(18 Pt 2):6290s–5s. <https://doi.org/10.1158/1078-0432.CCR-sup-040025>.
- Arai E, Ushijima S, Tsuda H, Fujimoto H, Hosoda F, Shibata T, et al. Genetic clustering of clear cell renal cell carcinoma based on array-comparative genomic hybridization: its association with DNA methylation alteration and patient outcome. *Clin Cancer Res*. 2008;14(17):5531–9. <https://doi.org/10.1158/1078-0432.ccr-08-0443>.
- Sanjmyatav J, Junker K, Matthes S, Muehr M, Sava D, Sternal M, et al. Identification of genomic alterations associated with metastasis and cancer specific survival in clear cell renal cell carcinoma. *J Urol*. 2011;186(5):2078–83. <https://doi.org/10.1016/j.juro.2011.06.050>.
- Kohn L, Svenson U, Ljungberg B, Roos G. Specific genomic aberrations predict survival, but low mutation rate in Cancer hot spots, in clear cell renal cell carcinoma. *Appl Immunohistochem Mol Morphol*. 2014. <https://doi.org/10.1097/pai.0000000000000087>.
- Monzon FA, Alvarez K, Peterson L, Truong L, Amato RJ, Hernandez-McClain J, et al. Chromosome 14q loss defines a molecular subtype of clear-cell renal cell carcinoma associated with poor prognosis. *Mod Pathol*. 2011; 24(11):1470–9. <https://doi.org/10.1038/modpathol.2011.107>.
- Klatte T, Rao PN, de Martino M, LaRochelle J, Shuch B, Zomorodian N, et al. Cytogenetic profile predicts prognosis of patients with clear cell renal cell carcinoma. *J Clin Oncol*. 2009;27(5):746–53. <https://doi.org/10.1200/jco.2007.15.8345>.
- Antonelli A, Arrighi N, Tardanico R, Balzarini P, Zanotelli T, Corti S, et al. Prognostic value of cytogenetic analysis in clear cell renal carcinoma: a study on 131 patients with long-term follow-up. *Anticancer Res*. 2010; 30(11):4705–9.
- Bai Q, Liu L, Xia Y, Long Q, Wang J, Xu J, et al. Prognostic significance of ST3GAL-1 expression in patients with clear cell renal cell carcinoma. *BMC Cancer*. 2015;15:880. <https://doi.org/10.1186/s12885-015-1906-5>.
- Fu H, Zhu Y, Wang Y, Liu Z, Zhang J, Wang Z, et al. High NUCB2 expression level represents an independent negative prognostic factor in Chinese cohorts of non-metastatic clear cell renal cell carcinoma patients. *Oncotarget*. 2016. <https://doi.org/10.18632/oncotarget.12961>.
- Fu Q, Chang Y, Zhou L, An H, Zhu Y, Xu L, et al. Positive intratumoral chemokine (C-C motif) receptor 8 expression predicts high recurrence risk of post-operation clear-cell renal cell carcinoma patients. *Oncotarget*. 2016; 7(7):8413–21. <https://doi.org/10.18632/oncotarget.6761>.
- Bai Q, Liu L, Xi W, Wang J, Xia Y, Qu Y, et al. Prognostic significance of ST6GalNAc-1 expression in patients with non-metastatic clear cell renal cell carcinoma. *Oncotarget*. 2016. <https://doi.org/10.18632/oncotarget.11258>.
- Wang J, Liu Y, Yang Y, Xu Z, Zhang G, Liu Z, et al. High expression of galectin-7 associates with poor overall survival in patients with non-metastatic clear-cell renal cell carcinoma. *Oncotarget*. 2016. <https://doi.org/10.18632/oncotarget.9749>.
- Wang Z, Xie H, Zhou L, Liu Z, Fu H, Zhu Y, et al. CCL2/CCR2 axis is associated with postoperative survival and recurrence of patients with non-metastatic clear-cell renal cell carcinoma. *Oncotarget*. 2016. <https://doi.org/10.18632/oncotarget.10492>.
- Ingels A, Hew M, Algaba F, de Boer OJ, van Moorselaar RJ, Horenblas S, et al. Vimentin over-expression and carbonic anhydrase IX under-expression are independent predictors of recurrence, specific and overall survival in non-metastatic clear-cell renal carcinoma: a validation study. *World J Urol*. 2016. <https://doi.org/10.1007/s00345-016-1854-y>.
- Xie H, Zhu Y, An H, Wang H, Fu H, Wang Z, et al. Increased B4GALT1 expression associates with adverse outcome in patients with non-metastatic clear cell renal cell carcinoma. *Oncotarget*. 2016;7(22):32723–30. <https://doi.org/10.18632/oncotarget.8737>.
- Liu W, Liu Y, Fu Q, Zhou L, Chang Y, Xu L, et al. Elevated expression of IFN-inducible CXCR3 ligands predicts poor prognosis in patients with non-metastatic clear-cell renal cell carcinoma. *Oncotarget*. 2016;7(12):13976–83. <https://doi.org/10.18632/oncotarget.7468>.
- An H, Xu L, Zhu Y, Lv T, Liu W, Liu Y, et al. High CXC chemokine receptor 4 expression is an adverse prognostic factor in patients with clear-cell renal cell carcinoma. *Br J Cancer*. 2014;110(9):2261–8. <https://doi.org/10.1038/bjc.2014.179>.
- Lyu X, Li H, Ma X, Li X, Gao Y, Ni D, et al. High-level S100A6 promotes metastasis and predicts the outcome of T1-T2 stage in clear cell renal cell carcinoma. *Cell Biochem Biophys*. 2015;71(1):279–90. <https://doi.org/10.1007/s12013-014-0196-x>.
- Qin C, Liu Z, Yuan Y, Zhang X, Li H, Zhang C, et al. Glycoprotein non-metastatic melanoma protein B as a predictive prognostic factor in clear-cell renal cell carcinoma following radical nephrectomy. *Mol Med Rep*. 2014; 9(3):851–6. <https://doi.org/10.3892/mmr.2014.1896>.
- Nishikawa M, Miyake H, Liu B, Fujisawa M. Expression pattern of autophagy-related markers in non-metastatic clear cell renal cell carcinoma: association with disease recurrence following radical nephrectomy. *J Cancer Res Clin Oncol*. 2015;141(9):1585–91. <https://doi.org/10.1007/s00432-015-1923-4>.
- Cai MY, Luo RZ, Li YH, Dong P, Zhang ZL, Zhou FJ, et al. High-expression of ZBP-89 correlates with distal metastasis and poor prognosis of patients in clear cell renal cell carcinoma. *Biochem Biophys Res Commun*. 2012;426(4): 636–42. <https://doi.org/10.1016/j.bbrc.2012.08.146>.
- Esteller M. Epigenetics in cancer. *N Engl J Med*. 2008;358(11):1148–59. <https://doi.org/10.1056/NEJMra072067>.
- Morris MR, Maher ER. Epigenetics of renal cell carcinoma: the path towards new diagnostics and therapeutics. *Genome Med*. 2010;2(9):59. <https://doi.org/10.1186/gm180>.
- Ricketts CJ, Hill VK, Linehan WM. Tumor-specific hypermethylation of epigenetic biomarkers, including SFRP1, predicts for poorer survival in patients from the TCGA kidney renal clear cell carcinoma (KIRC) project. *PLoS One*. 2014;9(1):e85621. <https://doi.org/10.1371/journal.pone.0085621>.
- Joosten SC, Deckers IA, Aarts MJ, Hoeven A, van Roermund JG, Smits KM, et al. Prognostic DNA methylation markers for renal cell carcinoma: a systematic review. *Epigenomics*. 2017;9(9):1243–57. <https://doi.org/10.2217/epi-2017-0040>.
- Arai E, Chiku S, Mori T, Gotoh M, Nakagawa T, Fujimoto H, et al. Single-CpG-resolution methylome analysis identifies clinicopathologically aggressive CpG island methylator phenotype clear cell renal cell carcinomas. *Carcinogenesis*. 2012;33(8):1487–93. <https://doi.org/10.1093/carcin/bgs177>.
- Tian Y, Arai E, Gotoh M, Komiyama M, Fujimoto H, Kanai Y. Prognostication of patients with clear cell renal cell carcinomas based on quantification of DNA methylation levels of CpG island methylator phenotype marker genes. *BMC Cancer*. 2014;14:772. <https://doi.org/10.1186/1471-2407-14-772>.
- Wei JH, Haddad A, Wu KJ, Zhao HW, Kapur P, Zhang ZL, et al. A CpG-methylation-based assay to predict survival in clear cell renal cell carcinoma. *Nat Commun*. 2015;6:8699. <https://doi.org/10.1038/ncomms9699>.

34. Evelönn EA, Degerman S, Kohn L, Landfors M, Ljungberg B, Roos G. DNA methylation status defines clinicopathological parameters including survival for patients with clear cell renal cell carcinoma (ccRCC). *Tumour Biol.* 2016; 37(8):10219–28. <https://doi.org/10.1007/s13277-016-4893-5>.
35. Svenson U, Ljungberg B, Roos G. Telomere length in peripheral blood predicts survival in clear cell renal cell carcinoma. *Cancer Res.* 2009;69(7): 2896–901. <https://doi.org/10.1158/0008-5472.can-08-3513>.
36. Weisenberger DJ, Campan M, Long TI, Kim M, Woods C, Fiala E, et al. Analysis of repetitive element DNA methylation by MethyLight. *Nucleic Acids Res.* 2005;33(21):6823–36. <https://doi.org/10.1093/nar/gki987>.
37. Degerman S, Landfors M, Siwicki JK, Revie J, Borssen M, Evelönn E, et al. Immortalization of T-cells is accompanied by gradual changes in CpG methylation resulting in a profile resembling a subset of T-cell leukemias. *Neoplasia.* 2014;16(7):606–15. <https://doi.org/10.1016/j.neo.2014.07.001>.
38. Teschendorff AE, Marabita F, Lechner M, Bartlett T, Tegner J, Gomez-Cabrero D, et al. A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. *Bioinformatics.* 2013;29(2):189–96. <https://doi.org/10.1093/bioinformatics/bts680>.
39. Nordlund J, Backlin CL, Wahlberg P, Busche S, Berglund EC, Eloranta ML, et al. Genome-wide signatures of differential DNA methylation in pediatric acute lymphoblastic leukemia. *Genome Biol.* 2013;14(9):r105. <https://doi.org/10.1186/gb-2013-14-9-r105>.
40. Gaunt TR, Shihab HA, Hemani G, Min JL, Woodward G, Lyttleton O, et al. Systematic identification of genetic influences on methylation across the human life course. *Genome Biol.* 2016;17:61. <https://doi.org/10.1186/s13059-016-0926-z>.
41. Yang Z, Wong A, Kuh D, Paul DS, Rakyen VK, Leslie RD, et al. Correlation of an epigenetic mitotic clock with cancer risk. *Genome Biol.* 2016;17(1):205. <https://doi.org/10.1186/s13059-016-1064-3>.
42. Comprehensive molecular characterization of clear cell renal cell carcinoma. *Nature.* 2013;499(7456):43–9. <https://doi.org/10.1038/nature12222>.
43. Kwee I, Rinaldi A, Rancoita P, Rossi D, Capello D, Forconi F, et al. Integrated DNA copy number and methylation profiling of lymphoid neoplasms using a single array. *Br J Haematol.* 2012;156(3):354–7. <https://doi.org/10.1111/j.1365-2141.2011.08946.x>.
44. Feber A, Guilhamon P, Lechner M, Fenton T, Wilson GA, Thirlwell C, et al. Using high-density DNA methylation arrays to profile copy number alterations. *Genome Biol.* 2014;15(2):R30. <https://doi.org/10.1186/gb-2014-15-2-r30>.
45. Hovestadt V, Zapatka M. Conumee: Enhanced copy-number variation analysis using Illumina DNA methylation arrays. 1.9.0 ed2017. <https://rdrr.io/bioc/conumee/>.
46. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics.* 2014; 30(10):1363–9. <https://doi.org/10.1093/bioinformatics/btu049>.
47. Wang J, Vasaikar S, Shi Z, Greer M, Zhang B. WebGestalt 2017: a more comprehensive, powerful, flexible and interactive gene set enrichment analysis toolkit. *Nucleic Acids Res.* 2017. <https://doi.org/10.1093/nar/gkx356>.
48. Ogawa O, Kakehi Y, Ogawa K, Koshiba M, Sugiyama T, Yoshida O. Allelic loss at chromosome 3p characterizes clear cell phenotype of renal cell carcinoma. *Cancer Res.* 1991;51(3):949–53.
49. Moore LE, Jaeger E, Nickerson ML, Brennan P, De Vries S, Roy R, et al. Genomic copy number alterations in clear cell renal carcinoma: associations with case characteristics and mechanisms of VHL gene inactivation. *Oncogene.* 2012;1:e14. <https://doi.org/10.1038/oncis.2012.14>.
50. Chen M, Ye Y, Yang H, Tamboli P, Matin S, Tannir NM, et al. Genome-wide profiling of chromosomal alterations in renal cell carcinoma using high-density single nucleotide polymorphism arrays. *Int J Cancer.* 2009;125(10): 2342–8. <https://doi.org/10.1002/ijc.24642>.
51. Arai E, Kanai Y. Genetic and epigenetic alterations during renal carcinogenesis. *Int J Clin Exp Pathol.* 2010;4(1):58–73.
52. Oakes CC, Claus R, Gu L, Assenov Y, Hullein J, Zucknick M, et al. Evolution of DNA methylation is linked to genetic aberrations in chronic lymphocytic leukemia. *Cancer Discov.* 2014;4(3):348–61. <https://doi.org/10.1158/2159-8290.cd-13-0349>.
53. Chae H, Lee S, Nephew KP, Kim S. Subtype-specific CpG island shore methylation and mutation patterns in 30 breast cancer cell lines. *BMC Syst Biol.* 2016;10(Suppl 4):116. <https://doi.org/10.1186/s12918-016-0356-2>.
54. Hu CY, Mohtat D, Yu Y, Ko YA, Shenoy N, Bhattacharya S, et al. Kidney cancer is characterized by aberrant methylation of tissue-specific enhancers that are prognostic for overall survival. *Clin Cancer Res.* 2014;20(16):4349–60. <https://doi.org/10.1158/1078-0432.ccr-14-0494>.
55. Urbschat A, Stumpf S, Hanze J, Paulus P, Maier TJ, Weipert C, et al. Expression of the anti-inflammatory suppressor of cytokine signaling 3 (SOCS3) in human clear cell renal cell carcinoma. *Tumour Biol.* 2016;37(7): 9649–56. <https://doi.org/10.1007/s13277-016-4857-9>.
56. Stofas A, Levidou G, Piperi C, Adamopoulos C, Dalagiorgou G, Bamias A, et al. The role of CXC-chemokine receptor CXCR2 and suppressor of cytokine signaling-3 (SOCS-3) in renal cell carcinoma. *BMC Cancer.* 2014;14:149. <https://doi.org/10.1186/1471-2407-14-149>.
57. Roos L, Sandling JK, Bell CG, Glass D, Mangino M, Spector TD, et al. Higher nevus count exhibits a distinct DNA methylation signature in healthy human skin: implications for melanoma. *J Invest Dermatol.* 2017;137(4):910–20. <https://doi.org/10.1016/j.jid.2016.11.029>.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

