Characterization of epigenetic changes and their connection to gene expression abnormalities in clear cell renal cell carcinoma

Pubudu Manoj Nawarathna Nawarathna Mudiyanselage

Department of Human Genetics

McGill University

Montreal, Quebec

July 2019

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Master of Science

DEDICATION

This thesis is dedicated to my parents, sister and brother for their unfailing support throughout my life.

Thank you so much!!!

ABSTRACT

Clear cell renal cell carcinoma (ccRCC) is the most common kidney cancer subtype comprising 70-75% of all adult kidney malignancies. Although the genetic mechanisms of ccRCC have been widely studied, the mechanisms underlying its epigenetic modifications are not yet well understood, despite the evidence on genome-wide epigenome abnormalities in ccRCC. Therefore, we identified and investigated the ccRCC-associated epigenetic alterations on cis-regulatory elements, including distal enhancers and proximal regulatory elements that are close to transcription start sites (TSSs), and their contribution to gene expression changes in ccRCC.

Using a machine-learning approach to combine ChIP-seq and whole genome bisulphite sequencing data of tumors and adjacent normal kidney tissues of ccRCC patients, we identified thousands of gained and lost *cis*-regulatory elements. Amongst them, ~71% are novel elements that had not been reported in ccRCC. Using a second machine-learning classifier, we showed that the dysregulated *cis*-regulatory elements are predictive of the genes with abnormal expression patterns in ccRCC (minimum area under ROC curve: 0.88, P-value $< 2.59 \times 10^{-73}$). As such, gained and lost enhancers are significantly associated with up-regulated and down-regulated genes in ccRCC, respectively.

Our analysis of the binding sites of 161 regulatory factors, including transcription factors and specific epigenome modifiers, revealed several regulatory factors whose binding sites were enriched in gained or lost enhancers. The target genes of these regulatory factors were enriched among up or down regulated genes in ccRCC consistent with the status of the associated enhancers (gained or lost). Among the top-ranked activated regulatory factors identified in our analysis, FOXM1, SPI1, IKZF1, JUNB, JUN, FOS, BCL11A and STAT3 are significantly enriched in gained enhancers, while EZH2, FOXA1, FOXA2, GATA3, ESR1 are significantly associated with lost enhancers (FDR < 0.05). The target genes of these activated regulatory factors are involved in biological pathways that are central to the biology and function of ccRCC cells. For example, HIF1α and HIF2α transcription factor network, VEGF and VEGFR signaling network and immune system-related pathways are enriched in up regulated target genes in ccRCC, associated with gained enhancers (FDR < 0.01).

Finally, we sought to examine potential involvement of von Hippel-Lindau (VHL), the most frequently mutated gene in ccRCC, in these epigenome alterations, given the recent reports on VHL function as a regulator of the epigenome. Our analysis of VHL-deficient

ccRCC cell lines and their wild-type VHL- reconstituted counterparts revealed that ~10% of gene expression changes in ccRCC can be recovered by VHL-driven epigenome alterations on enhancers. We uncovered several regulatory factors that could be regulated by VHL-mediated DNA methylation and that are associated with dysregulated enhancers, including BATF, BCL11A, IKZF1, JUN, SPI1, STAT3 and EZH2.

In conclusion, our study proposes a new method to identify active (gained) and inactive (lost) *cis*-regulatory elements by combining histone modifications and DNA methylation data through machine-learning. In addition, our results provide new insights into the functional consequences of dysregulation of cis-regulatory elements on gene expression patterns in ccRCC. These findings will improve our understanding of the epigenetic mechanisms underlying transcriptome aberrations in ccRCC, which may eventually lead to the development of new preventive or therapeutic interventions.

RÉSUMÉ

Le carcinome à cellules claires du rein (ccRCC) est le sous-type de cancer du rein le plus répandu, représentant 70 à 75% de toutes les tumeurs malignes du rein chez l'adulte. Bien que les mécanismes génétiques du ccRCC aient été étudiés considérablement et qu'on puisse observer des anomalies de l'épigénome à l'échelle du génome dans le ccRCC, les mécanismes responsables de ses modifications épigénétiques n'ont pas encore été établis. Par conséquent, nous avons identifié et étudié les altérations épigénétiques associées au ccRCC situées sur les éléments cis-régulateurs, y compris les amplificateurs (enhancers) distaux et les éléments régulateurs proximaux proches des sites d'initiation de la transcription (transcription start site, TSS), et leur contribution aux changements d'expression génique dans le ccRCC.

En utilisant une approche d'apprentissage automatique pour combiner les données de ChIP-seq et de séquençage bisulfite de l'ADN génomique provenant de tumeurs et de tissus rénaux normaux adjacents de patients atteints du ccRCC, nous avons identifié des milliers d'éléments cis-régulateurs acquis et perdus. Parmi ceux-ci, environ 71% sont des éléments nouveaux qui n'avaient pas été signalés dans le ccRCC auparavant. En utilisant un deuxième classificateur d'apprentissage automatique, nous avons démontré que les éléments cis-régulateurs peuvent prédire les gènes présentant des profils d'expression anormaux dans le ccRCC (aire minimale sous la courbe ROC : 0.88; $p < 2.59 \times 10^{-73}$). Ce faisant, les amplificateurs acquis et perdus sont associés de manière significative aux gènes régulés positivement et négativement dans le ccRCC, respectivement.

Notre analyse des sites de liaison de 161 facteurs de régulation, y compris des facteurs de transcription et des modificateurs de l'épigénome, en a révélé plusieurs dont les sites de liaison sont enrichis en amplificateurs acquis ou perdus. Les gènes cibles de ces facteurs de régulation sont enrichis parmi les gènes sur- ou sous-régulés dans le ccRCC, conformément au statut des amplificateurs associés (acquis ou perdus). Parmi les facteurs de régulation activés les mieux classés dans notre analyse, FOXM1, SPI1, IKZF1, JUNB, JUN, FOS, BCL11A et STAT3 sont enrichis de manière significative en amplificateurs acquis, tandis que EZH2, FOXA1, FOXA2, GATA3, ESR1 sont associés de manière significative aux amplificateurs perdus (FDR < 0.05). Les gènes cibles des facteurs de régulation activés sont impliqués dans des voies biologiques qui jouent un rôle central dans la biologie et la fonction des cellules du ccRCC. Par exemple, le réseau de facteurs de transcription HIF1α et HIF2α, le

réseau de signalisation VEGF et VEGFR et les voies liées au système immunitaire sont enrichis en gènes régulés à la hausse dans le ccRCC, qui sont associés à des amplificateurs acquis (FDR < 0.01).

Finalement, nous avons examiné l'implication potentielle de Von Hippel-Lindau (VHL), le gène le plus fréquemment muté dans le ccRCC, dans ces altérations de l'épigénome, étant donné les récents rapports sur la fonction de VHL en tant que régulateur de l'épigénome. Notre analyse des lignées cellulaires du ccRCC déficientes ou reconstituées en VHL a révélé qu'environ 10% des changements dans l'expression des gènes dans le ccRCC peuvent être renversés par des changements de l'épigénome induits par VHL sur des amplificateurs. Nous avons découvert plusieurs facteurs de régulation pouvant être régulés par la méthylation de l'ADN via VHL et associés à des amplificateurs dérégulés, notamment BATF, BCL11A, IKZF1, JUN, SPI1, STAT3 et EZH2.

En conclusion, notre étude propose une nouvelle méthode d'identification des éléments cis-actifs (acquis) et inactifs (perdus) qui combine des données de modifications d'histones et de méthylation de l'ADN par apprentissage automatique. De plus, nos résultats font lumière sur les conséquences fonctionnelles de la dérégulation des éléments cis-régulateurs sur l'expression génique dans le ccRCC. Ces découvertes amélioreront notre compréhension des mécanismes épigénétiques sous-jacents aux altérations du transcriptome dans le ccRCC, ce qui pourrait éventuellement mener à la mise au point de nouvelles approches cliniques et thérapeutiques.

TABLE OF CONTENTS

DEDICATION II

ABSTRACT III

RÉSUMÉ V

TABLE OF CONTENTS VII

LIST OF ABBREVIATIONS XI

LIST OF FIGURES XIV

LIST OF TABLES XV

ACKNOWLEDGEMENTS XVI

FORMAT OF THE THESIS XVII

CONTRIBUTION OF AUTHORS XVIII

CHAPTER 1. LITERATURE REVIEW, HYPOTHESIS AND OBJECTIVES 1

1.1 Introduction 1

1.2 Epigenetic regulation of gene expression 3

1.2.1 Histone modification 3

1.2.1.1 Histone acetylation 4

1.2.1.2 Histone methylation 4

1.2.2 Histone code and the cross talk between different histone modifications 5

1.2.3 Chromatin remodelling 7

1.2.4 DNA methylation 7

1.2.5 Interplay of different epigenetic modifications, TFs and gene expression 8

1.3 Regulatory elements and their connection with epigenetic modifications and gene expression 10

1.3.1 Promoters 10

1.3.2 Enhancers 11

1.3.3 TFs and gene expression 12

1.4 Epigenetic modifications in cancer....14
1.4.1 DNA methylation aberrations in cancer....14
1.4.1.1 DNA hypo-methylation in cancer....14
1.4.1.2 DNA hyper-methylation in cancer....15
1.4.2 Histone modification aberrations in cancer....16
1.4.2.1 Histone acetylation in cancer....16
1.4.2.2 Histone Methylation in cancer....16
1.4.3 Accumulations of epigenetic alterations at enhancers....17
1.4.4 Identification of regulatory elements....18
1.5 Epigenetic abnormalities in ccRCC....19
1.5.1 Loss of histone modifiers and chromatin remodeler genes....19
1.5.2 Other histone modification aberrations in ccRCC....20
1.5.3 DNA methylation aberrations in ccRCC....20
1.5.4 Functions of VHL and its role in epigenetic modifications in ccRCC....21
1.5.4.1 Characteristics and general function of VHL....21
1.5.4.2 HIF-VHL gene regulation....21
1.5.4.3 VHL-mediated DNA methylation....22
1.5.4.4 Epigenetic remodeling in VHL-deficient ccRCC....22
1.6 Transcriptome changes in ccRCC and their drivers....23
1.7 Mapping of epigenetic modifications and data generation....23
1.8 Applications of machine learning on epigenetic data analysis....24
1.9 Hypothesis....25
1.10 Objectives....26
CHAPTER 2. MATERIALS AND METHODS....27
2.1 Patient Information....27
2.2 Histone ChIP-seq Analysis....27
2.3 RNA-Seq differential gene expression Analysis....28

2.4 Defining proximal regulatory elements (PREs) 28
2.5 Defining enhancers 28
2.6 Average DHS signals at the centre of enhancers 28
2.7 Overlapping enhancers from tumor and normal samples with GenoSTAN enhancers 29
2.8 WGBS data analysis 29
2.9 Classification of gained and lost elements of ccRCC 29
2.10 GSEA pre-ranked test 30
2.11 450k methylation microarray data analysis 30
2.12 RCC4 cell culture and reagents 31
2.13 RNA isolation and RNA-Seq of RCC4 cells 31
2.14 The enrichment analysis of regulatory factor binding sites 31
CHAPTER 3. RESULTS 32
3.1 Epigenome aberrations on regulatory elements can be identified by histone modifications 32
3.2 Active distal and proximal regulatory regions in ccRCC 33
3.3 Differential DNA methylation correlates with differential activity of regulatory elements 34
3.4 Integrative analysis of histone modifications and DNA methylation uncovers the landscape of gain and loss of active regulatory elements in ccRCC 38
3.5 Gain and loss of active regulatory elements can predict the differential expression of associated genes 40
3.6 Gained and lost enhancers harbor binding sites for specific regulatory factors 41
3.7 Altered enhancers and associated gene expression changes can be partially reversed by VHL 45
CHAPTER 4. DISCUSSION 50
4.1 Discovery of gained and lost regulatory elements in ccRCC 50

4.2 Relationship between gain and loss of regulatory elements with gene expression changes in ccRCC 50

4.3 Regulatory factors-enhancer complexes drive activation of ccRCC pathways. 51

4.4 Changes of enhancer activity status affects expression of cancer-related genes in ccRCC 52

4.5 VHL-mediated DNA methylation reprogramming and its role in enhancer regulation in ccRCC 53

CHAPTER 5. CONCLUSIONS AND FUTURE DIRECTIONS 56

CHAPTER 6. REFERENCES 57

APPENDICES…… 77

Supplementary Materials 77

Supplementary methods 77

Histone broad peaks comparison with reference peaks 77

Visualizing the histone enrichment profiles surrounding TSS 77

Visualizing the differential ChIP-Seq signal for histone marks around TSS 77

Supplementary figures 78

Supplementary tables 86

Copyright clearance 89

Ethical approval for using patient samples 91

LIST OF ABBREVIATIONS

5hmC - 5-hydroxymethylcytosine

5mC - 5-methylcytosine

AHEAD - Human Epigenome and Disease

PRE – Proximal regulatory elements

ATP - Adenosine triphosphate

BER - Base excision repair

ccRCC – clear cell renal cell carcinoma

CGIs - CpG islands

CHD - Chromodomain helicase DNA-binding

ChIP-seq - Chromatin immunoprecipitation followed by sequencing

CIMP - CpG island methylator phenotype

CPDB - Consensus pathway database

CpGs - Cytosines followed by guanine residues

CRC - Chromatin Remodelling Complexes

DHS - DNase I hypersensitive site

DNA - Deoxyribonucleic acid

DNMTs - DNA methyltransferases

DPE - Downstream promoter element

ELMO3 - Engulfment and cell motility 3

ENCODE - ENCyclopedia Of DNA Elements

eRNA - Enhancer RNAs

EV – Empty vector

EZH2 - Enhancer of zeste homolog 2

FDR – False discovery rate

GEO - Gene Expression Omnibus

GSEA - Gene set enrichment analysis

HATs - Histone acetyltransferases

HDACs - Histone deacetylases

HDMTs - Histone demethylases

HEP - Human Epigenome Project

HIF - Hypoxia Inducible Factor

HMTs - Histone methyl transferases

HP1 – Heterochromatin protein 1

IHEC - International Human Epigenome Consortium

IHW – Independence hypothesis weighting

Inr - initiator

LSD1 - Lysine-specific demethylase 1

MBD - Methyl-CpG binding domain proteins

MLL- Mixed lineage leukemia

mRNA - Messenger RNA

NGS- Next Generation Sequencing

PRC2 - Polycomb repressive complex 2

RCC – Renal cell carcinoma

RFBSs – Regulatory factor binding sites

RFs - Regulatory factors

RNA - Ribonucleic acid

ROC - Receiver operating characteristic

SWI/SNF - Switch/sucrose non-fermentable

TBP - TATA-binding protein

TCGA - The Cancer Genome Atlas

TET - Ten-Eleven-Translocation

TFs – Transcription factors

TSG - Tumor suppressor gene

TSSs - Transcription Start Sites

VHL – Von Hippel-Lindau

WGBS - Whole genome bisulphite sequencing

LIST OF FIGURES

Figure 1: Multiple levels of chromatin folding:.....4
Figure 2: Histone modification:.....5
Figure 3: The histone code at different levels of genomic features:.....6
Figure 4: Interplay of different epigenetic modifications:.....10
Figure 5: Existing models for the function of enhancers.....12
Figure 6: A Schematic of a TF and its domains.....13
Figure 7: *Cis*-regulatory elements of the genome are characterized by histone modifications: 35
Figure 8: DNA methylation is a good indicator of gain and loss of regulatory elements: 37
Figure 9: Investigating the inner working of machine learning classifier:.. 40
Figure 10: Relationship of gene expression changes and gain or loss of active regulatory elements: 42
Figure 11: Regulatory factors whose binding sites are enriched in gained or lost enhancers: 44
Figure 12: VHL reconstitution partially recovers the changes occurred in ccRCC: 49
Supplementary Figure S1.....78
Supplementary Figure S2.....80
Supplementary Figure S3.....82
Supplementary Figure S4.....83
Supplementary Figure S5.....85
Supplementary Figure S6.....85

LIST OF TABLES

Table 1: Functions of frequently mutated epigenetic modifiers in ccRCC 19
Supplementary Table S1: Clinical information of patients; M: male, F: female.....86
Supplementary Table S2: ChIP-seq analysis statistics; T: tumor, N: Normal.....86
Supplementary Table S3: WGBS analysis statistics.....87
Supplementary Table S4: Significantly enriched pathways of up-regulated target genes of gained enhancer-RF pairs 87
Supplementary Table S5: Up-regulated target genes associated with HIF1α and HIF2α transcription factor network and RFs associated with these target genes.....88

ACKNOWLEDGEMENTS

Notes of sincere gratitude to:

Drs. Yasser Riazalhosseini and **Hamed Shateri Najafabadi**, at Department of Human Genetics, McGill University, Canada for their ceaseless support, patience, motivation, enthusiasm and vast knowledge, extended for the completion my M.Sc. research project. Their guidance helped me immensely during the research and even during the writing process of the thesis. I cannot forget how their insightful comments on the research project helped me to get through difficult situations during the research project.

My supervisory committee members **Profs. Nada Jabado, Guillaume Bourque** and **Michael Ohh** for the guidance given in the direction and scope of the project.

All my colleagues in the Riazalhosseini and Najafabadi labs, specially **Gabrielle, Rached, Rick, Ali Mehdi, Ali Nehme and Pouria** for many helps throughout the research.

Gabrielle, Rached and Shavindi Ediriarachchi for proof reading the thesis.

Xiaojian Shao, Francois Lefebvre, Drs. **Mathieu Bourgey** and **Tony Kwan** for supports with raw data obtaining and analysis.

Keheliya (ayya) and **Shabir** (my two amazing room mates) for wonderful moments we shared in our apartment, motivational speeches, guidance, advices and stress releasing "coffee talks".

Amanda, Chalani, Kalani, Nayani and **Udari** for being with me during the hard times and helped me in so many ways.

Piumi, Dilangani, Praneeth, Chathumal, Gasitha, Prabhath, **Yashoda** and **Supipi** for always motivating me.

My father, mother, sister and brother for their strength and continuous support. If it wasn't for their immense support I wouldn't be in the **McGill University** doing this research project in the first place.

Queen Elizabeth II Diamond Jubilee (QEII) Scholarship program for the continuous financial support throughout the degree.

FORMAT OF THE THESIS

This thesis was prepared in adherence to the traditional thesis format outlined by McGill University's Faculty of Graduate and Postdoctoral Studies. This thesis is composed of six chapters. Chapter 1 is a comprehensive review of the literature relevant to this thesis. Chapter 2 is a presentation of the materials and methods. Chapter 3 is a presentation of results, which is in preparation for submission to a peer-reviewed journal. Chapter 4 is a discussion of the results presented in this thesis, while Chapter 5 is a conclusion discussing future research directions. Chapter 6 includes all the references of the thesis. Finally, appendices include supplementary materials such as figures, tables, methods and copy right clearances.

CONTRIBUTION OF AUTHORS

This work was the result of a collaboration between the labs of Drs. Yasser Riazalhosseini (Y.R) and Hamed Shateri Najafabadi (H.S.N.)

Detailed contributions are as follows:

Pubudu Manoj Nawarathna (P.M.N.) analyzed all the data, developed figures and wrote the thesis.

P.M.N and H.S.N developed the computational and statistical methods.

H.S.N. and Y.R. conceived and directed the study, designed the experiments, and edited the thesis.

Pouria Jandaghi (P.J) performed cell culturing and RNA-Seq experiments.

CHAPTER 1. LITERATURE REVIEW, HYPOTHESIS AND OBJECTIVES

1.1 Introduction

Epigenetics is "the study of mitotically and/or meiotically heritable changes in gene function that cannot be explained by changes in DNA sequence"[1]. Epigenetic gene control is involved in modulation of gene expression without any modifications to DNA sequence itself but by remodelling the structure of the chromatin and altering chromatin accessibility through various mechanisms[2,3]. Moreover, tissue-specific gene expression is mainly influenced by epigenetic regulation, which is an essential feature in development, differentiation and maintenance[4].

Epigenetic regulation largely involves specific modifications that mainly include DNA methylation and post-translational histone modification, and play a central role in the regulation of global and local gene expression[5]. Alteration of epigenetic modifications is one of the main mechanisms that cause drastic deregulation of the normal gene expression programmes, which can lead to disruption of normal cellular processes[6]. Because these different epigenetic modifications often work together in order to regulate the gene expression, it is important to study the whole set of epigenetic modifications to correctly explain the aberrant gene expression observed in diseases like cancers[7].

The human genome approximately consists of ~20,000 protein-coding genes, which encode ~80,000 transcripts that can be translated into proteins[8]. However, protein-coding genes only account for less than 2% (less than 60 million bases of DNA) of the whole genome[9,10]. *Cis*-regulatory elements (regulatory elements), including promoters and enhancers, are non-coding DNA[11] that regulate the expression of the genes, and comprise a sizeable fraction of the non-coding genome. These regulatory elements are bound by different trans-acting DNA-binding proteins such as transcription factors (TFs) and are essential for proper spatiotemporal expression of nearby genes or those connected to them through long-range chromatin interactions[8,11]. These regulatory elements are the main functional targets of epigenetic modifications[3,12].

Cancer is a complex disorder in which cells undergo uncontrolled cell division and thus gain the ability of invading neighbouring tissues. The uncontrolled cell division accompanies various molecular abnormalities that imbalance the normal cellular homeostasis. Importantly,

epigenetic aberrations play a crucial role in initiation, progression and metastasis together with genetic alterations[13,14]. Recent studies support this notion by showing that genes coding for regulators of epigenetic pathways are often dysregulated not only by mutations but also by amplifications, deletions and rearrangement of DNA in cancer cells[15]. Therefore among different diseases, malfunction of epigenetic machinery is linked most unambiguously to cancer and is considered as a putative driver and hallmark of cancer[16]. These findings highlight the need to understand the role of epigenetic modifications in cancer. Given that epigenetic aberrations are reversible and they represent attractive candidates for development of therapeutic strategies.

Renal cell carcinoma (RCC) is one of the ten most common cancers accounting for 2.4% of all adult cancers and 90% of all kidney cancers[17]. The rates of diagnosing RCC and RCC-related deaths are increasing each year worldwide. Diverse tumor heterogeneity leads to several histological subtypes of RCC with varying clinical outcomes and diverse therapeutic responses. The most common sub-type is clear cell RCC (ccRCC), comprising 70-80% of all RCC cases[18]. Despite the fact that the relative 5-year survival rate of localized ccRCC (stages I or II) is 91% with radical or partial nephrectomy, 20-30% of the patients are in the metastasis state at the time of diagnosis[19]. Metastatic ccRCC is incurable due to the inherent resistance to chemo- and radio-therapies[20]. Therefore, it is essential to improve knowledge of the molecular biology of tumor initiation, progression and metastasis to develop novel diagnostic and therapeutic tools to better manage and treat ccRCC patients.

In our study, we focused on epigenetic modifications of regulatory elements specifically on enhancers and promoters in ccRCC, and their relationship with gene expression changes. The regulatory state of enhancers and promoters are determined by a combination of epigenetic marks[21]. Therefore, we employed a machine learning approach to integrate ChIP-seq (chromatin immunoprecipitation with massively parallel DNA sequencing) data of histone modifications and whole genome bisulphite sequencing (WGBS) data (DNA methylation profiles) to identify the enhancer activity changes (*i.e* gain or loss of enhancers/promoters) in tumors in comparison to normal renal tissues. Then we investigated the correlation of these activity changes with gene expression changes of ccRCC. We also analysed the TFs that specifically bind these altered enhancers and their association with gene expression changes. In addition, we investigated the possible drivers of these epigenetic changes, specifically the Von Hippel-Lindau (VHL) tumor suppressor, because mutations of

the VHL, which is detected in ~75% of sporadic cases, are considered as the main drivers of ccRCC.

1.2 Epigenetic regulation of gene expression

DNA methylation and post-translational modifications of histone proteins are the two main types of epigenetic modifications[22,23]. In this section, I will briefly introduce these modifications and their role in regulating gene expression.

1.2.1 Histone modification

The nuclear DNA is packaged around octamers of histone proteins (consisting of two H3-H4 and two H2A-H2B dimers), in units referred to as nucleosomes, which are then compacted into a higher order complex structure called chromatin. This multi level chromatin organization occurs through a hierarchy of histone-dependent interactions (**Figure 1).** DNA and nucleosomes are often referred to as "beads on a string", where the string is the DNA and beads represent nucleosomes[3,4]. Chromatin structure is mainly divided into heterochromatin (condensed) and euchromatin (open chromatin or relaxed) based on the nucleosome positioning. Generally, regulatory factors (RFs) like TFs can access the euchromatin DNA; hence genes in euchromatin are transcriptionally active. On the other hand, heterochromatin state is associated with gene silencing[4].

Amino acids of the highly basic N-terminal tails of histone proteins (the amino-terminal ends of the histone protein chains) protrude from nucleosomes and are chemically modified by regulatory proteins and enzymes. These chemical modifications are often called histone tail modifications or simply histone modifications. The inter-nucleosomal interactions and recruitment of other chromatin remodelling factors (Chromatin Remodelling Complexes; CRCs) are affected by these histone tail modifications, thus regulating nucleosome movement and unwinding. These changes allow inaccessible regulatory elements on DNA become accessible to influence gene expression, or accessible regions become inaccessible, depending on the type of histone modification[24-26].

There is an ever-growing list of histone modifications which includes acetylation, methylation, phosphorylation, deamination, ADP ribosylation, ubiquitination and sumoylation. The combination of all these post-translational chromatin modifications is the major determinant of the chromatin structure and, thus, gene expression[27]. Amongst these

modifications, acetylation and methylation are the two best-characterized histone modifications.

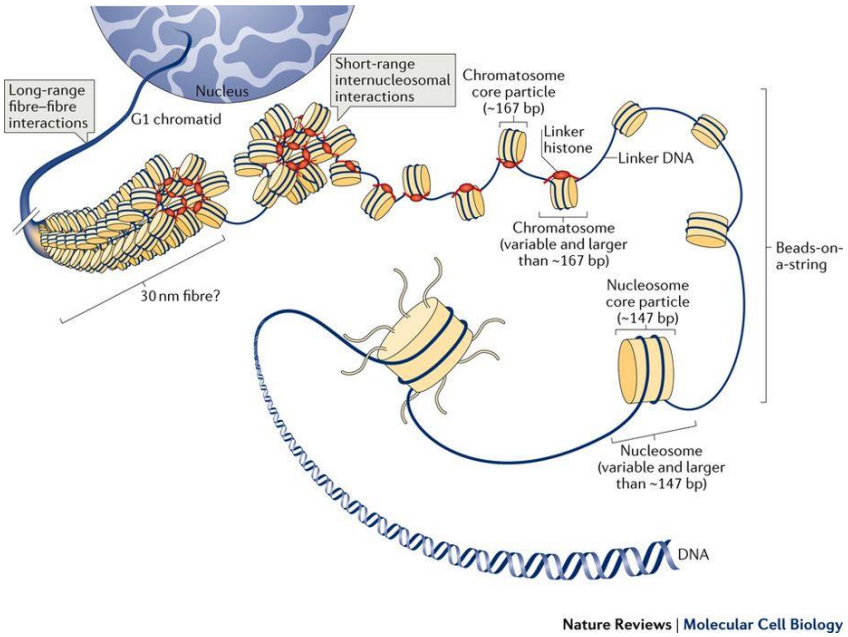


Figure 1: Multiple levels of chromatin folding: The nuclear DNA is packaged around nucleosome core particles consisting of histone proteins, and is compacted into a higher order complex structure called chromatin through a hierarchy of histone-dependent interactions. Adapted from Fyodorov, D. V., *et al*. *Nature Reviews Molecular Cell Biology* 19.3 (2018): 192. (https://doi.org/10.1038/nrm.2017.94) with permission from Springer Nature[28]

1.2.1.1 Histone acetylation

Histone acetylation is considered as a transcriptional activator (leading to high DNA accessibility) and its highly dynamic state is regulated by the opposing activity of two families of enzymes, histone acetyltransferases (HATs) and histone deacetylases (HDACs)[26,29,30] (**Figure 2**). The HATs catalyse the acetylation process. In contrast, HDACs reverse this process by removing the acetyl group, hence predominantly considered as transcription repressors[26].

1.2.1.2 Histone methylation

Histone methylation is governed by histone methyl transferases (HMTs) that catalyse the transfer of a methyl group to specific amino acids in histone tails, such as lysine and arginine in H3 histone[26]. Depending on the type, HMTs can either mono-, di- or tri-methylate (degree of methylation) a residue on the histone tail. For instance, lysine can be mono-, di- or tri-methylated[31,32]. Unlike acetylation, transcriptional activity of methylation depends on the

specific degree of methylation and the residue affected. For example, H3K4me3, H3K4me1 and H3K36me3 are associated with transcriptional activation while H3K9me3, and H3K27me3 are leading to transcriptional repression[33,34]. These methylation patterns are reversible; histone demethylases (HDMTs) catalyze the removal of methyl group from the affected residues [4,18].

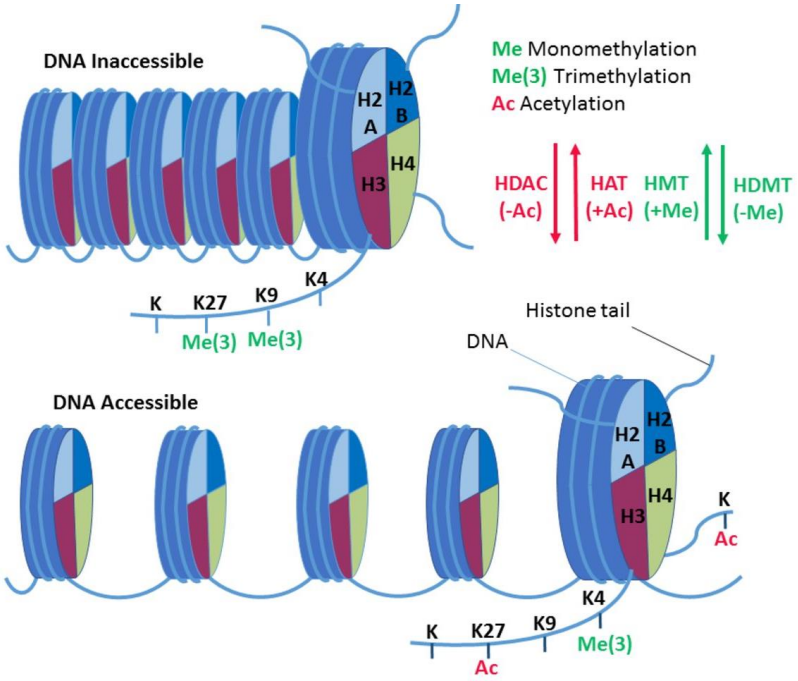


Figure 2: Histone modification: Histone tails can be modified by the addition and removal of acetyl and methyl groups on specific residues. These modifications are catalyzed by specific enzymes; HAT and HDAC regulate histone acetylation and deacetylation, respectively, while HMT and HDMT catalyse histone methylation and demethylation, respectively. Each type of histone modification has a specific effect on transcriptional activity. Repressive histone modifications facilitate the dense packaging of DNA and restrict the accessibility for RFs. Activating histone modifications loosen DNA packaging and result in transcriptional activation. Abbreviations: histone acetyltransferase (HAT); histone deacetylase (HDAC); histone demethylase (HDMT); histone methyltransferase (HMT). Adapted from Meddens C.A., *et al*., *Gut* 2019;68:928-941[35] (http://dx.doi.org/10.1136/gutjnl-2018-317516) under the Creative Commons Attribution Non Commercial License.

1.2.2 Histone code and the cross talk between different histone modifications

The histone code is a hypothesis describing the cross talk and combinatorial effect of different histone modifications that function as a molecular “code”, recognized and used by

other non-histone regulatory proteins to possibly regulate chromatin state and gene expression. These modifications include all the possible histone modifications, which are regulated by various histone modifying enzymes[36,37]. Deciphering how the histone code could be translated into biological response is essential to accurately predict the consequences of histone modifications. Depending on the set of histone modifications deposited at a specific region, distinct "readouts" of the epigenetic information will occur (e.g. active, inactive, or posied)[38]. In addition, different histone combinations are associated with different genomic features (*i.e:* promoters or enhancers), which means that a combination of histone modifications can be used to identify these functionally distinct genomic regions[39]. **Figure 3** illustrates an example of chromatin states and features encoded by different combinations of histone modifications.

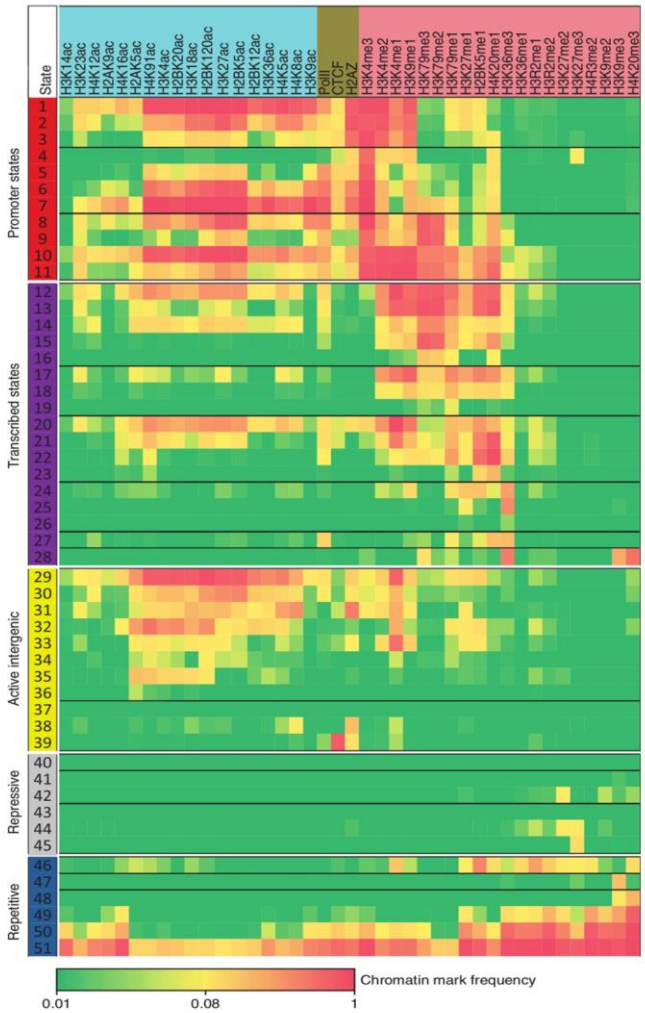


Figure 3: The histone code at different levels of genomic features. The combinations of histone modifications and their associations to various genomic states. (*i.e:* promoter states, transcribed states). Each genomic state is associated with a combination of histone marks.

The frequency which they occur showing the relative degree of association is indicated by a colour scale spanning values between 0 and 1; light blue: acetylation marks, pink: methylation marks, brown: CTCF/Pol2/H2AZ. Adapted from Ernst, J., & Kellis, M., *Nature biotechnology*, 28(8), 817. DOI: 10.1038/nbt.1662, with permission.

1.2.3 Chromatin remodelling

RFs (*e.g:* sequence-specific activators and repressors, mediator complexes, and general transcription factors) need to interact with DNA to regulate gene expression. Therefore, higher order compacted chromatin should be remodelled to a more loosely packed structure to allow regulatory factors to access DNA. In this regard, chromatin remodeling complexes **(**CRCs) play an ensemble of crucial roles associated with nucleosome dynamics, including the continuous shuffling of the nucleosome positions (nucleosome sliding), transforming the conformation of nucleosomal DNA, nucleosome assembly and disassembly, and facilitation of the RF interaction with DNA[40-43]. Not only activation of genes but also gene repression can be mediated by CRCs, e.g. by establishing proper density and spacing of nucleosomes leading to compacted higher order chromatin structures[44].

1.2.4 DNA methylation

While DNA methylation is not limited to 5-methylcytosine (5mC), the term DNA methylation is often used for this particular type of modification, in which the fifth carbon of the pyrimidine ring of cytosines becomes methylated. In most cells, 5mC modification largely takes place at sites in which the cytosine is followed by a guanine residue (CpG), which is called CpG methylation. Overall, 70-80% of all CpGs are methylated in mammals[45,46]. DNA methylation is generally associated with transcriptional silencing, particularly in repression of endogenous repeat elements such as transposons, as well as some tissue-specific genes during development and differentiation. Furthermore, it also plays a vital role in genomic imprinting and X-chromosome inactivation in female cells[47].

CpGs are predominantly found in endogenous repeat elements, promoter CpG islands, and intergenic regions[46,48,49]. The methylation status of many of these regions is associated with the expression of nearby genes. For example, CpG methylation in promoters and enhancers are negatively correlated with gene expression, since methylation of these regions often results in heterochromatin[46,50,51]. The correlation between CpG DNA methylation and gene expression was first shown in a study of CpG islands (CGIs)[45]. CGIs are regions with at least ~200 bp in length and over 50% GC composition[52,53]. CGIs are dispersed across the

genome, with ~45-50% of all human promoters associated with CGIs. The majority of CGIs are unmethylated throughout the development, hence leading to transcriptional activation[50,53,54].

DNA methylation at CpGs is carried out by a specific set of enzymes called DNA methyltransferases (DNMTs) including DNMT1, DNMT3A, DNMT3B[45]. DNMT3a and DNMT3b are *de novo* methyltransferases, DNMT1 predominantly maintains the established DNA methylation pattern during cell replication[55]. DNA methylation can be actively reversed through TET family enzymes, or passively reversed through DNA replication without *de novo* methylation[46].

1.2.5 Interplay of different epigenetic modifications, TFs and gene expression

Usually, a cross-talk or an interplay of different types between epigenetic modifications (shown in **Figure 4**) is required to regulate various nuclear processes including DNA repair and gene expression in a well coordinated manner. Histone modifications and DNA methylation are dependent on each other and their combinatorial effect may drive gene expression[56]. For instance, histone modifications such as H3K27me3 and H3K9me3 may influence DNA methylation, which in turn may serve as a template for the deposition of certain histone modifications after DNA replication[57]. This is exemplified by direct interactions between histone-modifying enzymes like HMTs and DNA methylation enzymes such as DNMTs[58]. However, the chronological sequence of these relationships is still unknown. HP1 (heterochromatin protein 1) also represents a connection between DNA methylation and histone modifications[59]. HP1 binds to both H3K9me3 and DNMT1 and influences DNA methylation patterns[56]. Of note, it regulates transcription of genes in both open chromatin and heterochromatin[60]. In addition, site-specific histone modifications usually attract CRCs toward a particular genomic region and/or modulate the efficacy of their enzymatic reactions. Likewise, CRCs can also influence and sometimes directly regulate histone modifications[7].

DNA methylation and repressive histone modifications usually mediate the transition from active to silent chromatin states, whereas DNA demethylation and active histone marks generally direct conversion of silent chromatin to the active state[61]. There seems to be many ways in which silent chromatin can repress transcription. Restriction of the binding of proteins like TFs and RNA Polymerase II is a common mechanism[62], since these proteins initiate gene transcription upon binding[63]. Another mechanism can be through recruitment of

specific repressor proteins – for example, binding of MeCP2 and other methyl-CpG binding domain proteins (MBDs) is strictly dependent on DNA methylation[64]. These proteins often function as transcriptional repressors[65]; for example, MBDs recruit repressor complexes to methylated DNA. These repressor complexes interact with other HDACs and CRCs, leading to the formation of a more compact and transcriptionally inactive chromatin[66]. Moreover, transcriptional repressors compete with TFs for binding to methylated CpG thereby repressing transcription[67].

On the other hand, TFs interact with a variety of proteins that methylate or demethylate DNA and modify histones[63]. A few of them are highlighted here: TFs can recruit and form complexes with DNMTs and thus modulate promoter methylation[68]. In addition, TFs function as recruiters of histone-modifying enzymes to nucleosomes. For instance, specific TFs repress or activate gene expression through recruitment of HDACs or HATs during histone acetylation[69]. The role of TFs in recruitment of chromatin modifying proteins provides a mechanistic explanation for region-specific epigenetic modifications: TFs usually recognize specific patterns of DNA sequence[70], and therefore each TF binds to a specific set of genomic regions. In fact, histone marks can be predicted accurately from TF-binding patterns[71], supporting the notion that epigenetic modifications are guided by the binding of TFs to DNA. Overall, the cross-talks between epigenetic modifications and TFs form a complex and dynamic network whose abnormalities can lead to many diseases, including cancer[7,72]

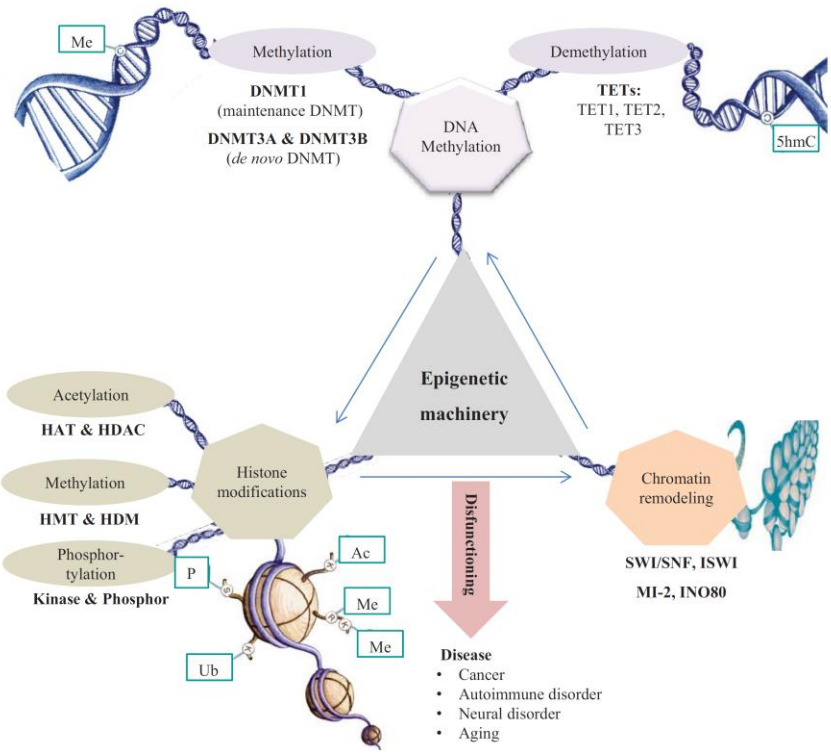


Figure 4: Interplay of different epigenetic modifications: i) DNA methylation, ii) histone proteins and iii) chromatin remodeling are three main specific epigenetic modifications, and regulation of gene expression depends on the well coordinated interactions between these modifications. Dysregulation of these interactions results in several diseases including cancer. DNMT: DNA Methyl Transferase, HAT: Histone Acetylation Transferase, HDAC: Histone Deacetylase, HMT: Histone Methyltransferase, HDM: Histone Demethyl Transferase. Adapted from Sahar, O. S. and Nussler. A. K., *Epigenetic Diagnosis & Therapy* 1.1 (2015): 5-13 (https://doi.org/10.2174/2214083201666150220233430)[72] with permission of Bentham Science Publishers Ltd.

1.3 Regulatory elements and their connection with epigenetic modifications and gene expression

Enhancers and promoters are the most well characterized types of DNA elements involved in transcriptional regulation[73]. In this section, I will briefly introduce these regulatory features of the genome, their role in mediating gene expression, and their interplay with each other.

1.3.1 Promoters

Promoter is the fundamental unit of regulation required for the transcription of a particular gene, and is located typically upstream of the TSS of a gene[74]. The transcription initiation involves the recruitment of transcriptional co-activators, general TFs, and RNA polymerase II to the promoter[8]. Generally, a promoter is composed of three main portions: core promoter, proximal promoter and distal promoter[75].

The core promoter encompasses the TSS and the minimal stretch of DNA sequence that directs the RNA polymerase II to the transcription initiation site. Mainly four DNA elements have been found to be associated with the function of core promoter: the TATA box, the TFIIB recognition element (BRE), the initiator (Inr), and the downstream promoter element (DPE). TATA box is found in some promoters and bears the binding site for TATA-binding protein (TBP), which is a subunit of the TFIID general TF complex. The BRE is located immediately upstream of the TATA box of some promoters and increases the binding affinity of TFIIB for the promoter. The Inr is located in the immediate region surrounding the TSS and is sufficient to direct accurate initiation in TATA-less promoters. DPE, which is a functional analogous of TATA-box, is located downstream of TSS and allows TFIID to bind cooperatively with Inr in the absence of TATA element in order to initiate transcription[76,77].

However, the operational range of a DPE is yet to be discovered. In addition, proximal promoter regions usually extend up to 1Kb upstream from TSS and contain many TF binding sites necessary to increase the binding affinity of RNA polymerase II to core promoter. Moreover, distal promoter is located beyond the proximal promoter and its function is largely unknown, but it shows comparatively weaker influence on gene transcription than the proximal promoter[75,77,78].

Active promoters are characterized by co-occupancy of H3K4me3 and H3K27ac histone marks[79]. In contrast, poised (bivalent or inactive) promoters, which are essential for germ cell identity, are delineated by concurrent presence of H3K4me3 and H3K27me3. Even though poised promoters are not associated with active gene expression, they are not marked by DNA methylation and can transit into the active form through temporal and special regulation.

1.3.2 Enhancers

Enhancer is a type of distal and cell type-specific regulatory element. Enhancers can regulate one or many genes in varying distances from the TSS, and at different orientations (enhancer can be located either upstream or downstream of TSS), or at different genomic locations (e.g. inter-genic or intra-genic regions)[80-82]. A central feature of enhancers is their ability to act as integrated TF binding platforms by providing clustered recognition sites for multiple TFs[83]. However, several factors influence the binding of TFs to enhances. For example, DNA hyper-methylation at enhancers mostly restricts TF binding, whereas hypo-methylation favours the binding of TFs[84]. The most distinctive feature of enhancers is the presence of specific histone modifications: Enhancers are commonly flanked by H3K4me1/2 histone modifications, and active enhancers are distinguished from poised enhancers by the presence of H3K27ac[85,86].

Currently, different models have been proposed to explain the mechanisms by which enhancers regulate gene expression (**Figure 5**). Amongst them, the looping model is the most widely accepted model to date, as it can explain many features of the enhancer-mediated gene expression[87,88]. This model proposes a direct physical contact between an enhancer and a target promoter through the creation of a DNA loop, following the binding of regulatory factors including TFs and co-activators[89,90]. In response to environmental or developmental signals, TFs bind enhancers, often followed by recruitment of transcriptional co-activators. These co-activators, for instance, CBP/p300 and MLL3/MLL4, have been shown to modulate

acetylation of histone H3K27 or monomethylate histone H3K4 on enhancers, respectively. Then, enhancer-promoter interaction is formed and stabilized by factors such as cohesin and the mediator complex through formation of a DNA loop. This allows interaction of enhancer-bound TFs and the co-activators with the basal transcription machinery on promoters and stimulation of transcription[80].

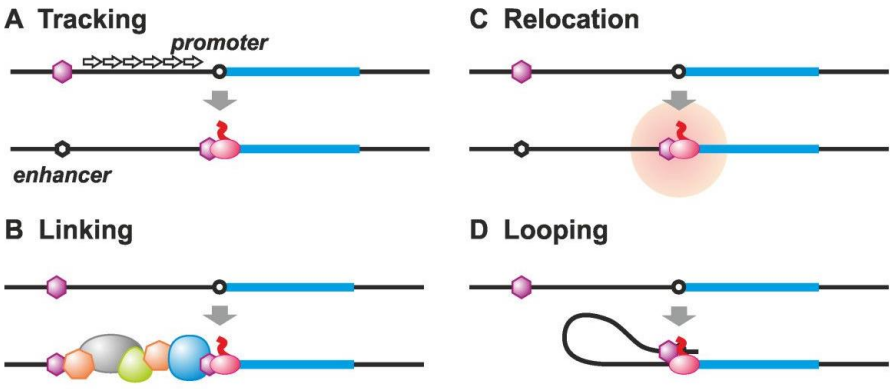


Figure 5: Existing models for the function of enhancers: The four existing models that explain how enhancers regulate gene transcription are depicted. A. The tracking model: a protein like transcription factor (TF) (purple hexagon) recruited onto the enhancer and tracks along the DNA strand towards the promoter. Then the TF associates with RNA polymerase (pink oval) when it reaches the promoter and stimulates transcription. B. The linking model: a TF is loaded on to the enhancers and drives protein polymerization towards the promoter. C. The relocation model: a gene relocates to sub compartments in the nucleus (pink halo) facilitating the interaction with correspondent promoter and enhancer and then initiate the transcription. D. The looping model: the enhancer is bound by TFs and recruit other regulatory factors. Then it forms a DNA loop and comes closer to the correspondent promoter and interact with basal transcription machinery in order to initiate the transcription. Adapted from Kolovos, P., *et al.*, *Epigenetics & chromatin* 5.1 (2012): 1 (https://doi.org/10.1186/1756-8935-5-1)[81] under Creative Commons Attribution License.

1.3.3 TFs and gene expression

TFs are one of the major classes of proteins that regulate the expression of genes followed by binding to the DNA. TFs usually compete with nucleosomes in order to access DNA. The binding affinity of TFs depends on several factors, including DNA methylation and histone modifications in or around binding regions[91]. TFs use their DNA binding domains to bind specific DNA sequences within gene promoters or enhancers and recruiting

regulatory co-factors using their effector domains – the effector domain of a TF can lead to activation or repression of expression depending on its molecular function and interacting partners (**Figure 6**). Trans-activating domains recruit transcription machinery, including RNA polymerase II, and other TFs or protein complexes such as CRCs to promote transcription, whereas repressive domains recruit chromatin silencing proteins[92,93,94]. TFs are often classified based on the three dimensional structure of their DNA-binding domains[95]. For example, zinc finger, helix-turn-helix (homeodomain), helix-loop-helix and leucine zipper are types of DNA-binding domains that are commonly found in TFs[95].

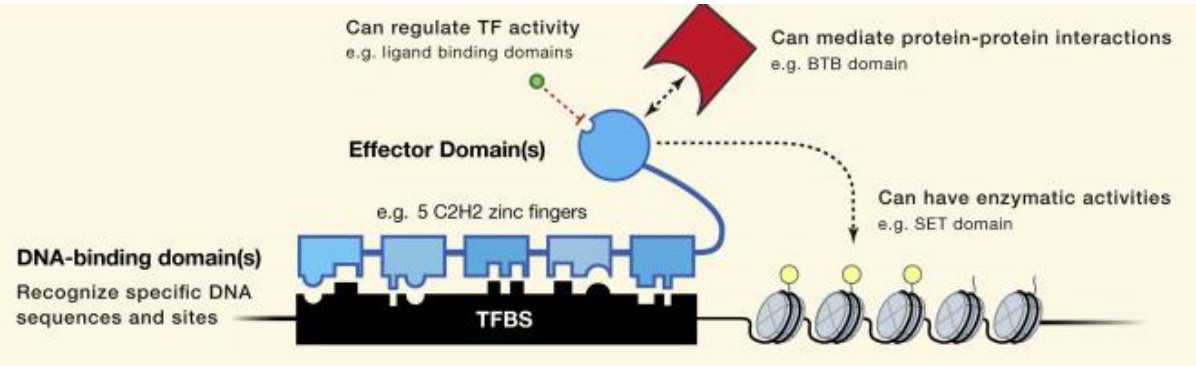


Figure 6: A Schematic of a TF and its domains. Adapted from Lambert, S. A., *et al.* (2018). *Cell* 172(4): 650-665 with permission (DOI: https://doi.org/10.1016/j.cell.2018.01.029) [70]

The majority of protein-coding genes show tissue- and signal-specific expression, mediated by a large set of DNA sequence-specific TFs. This sequence specificity is mediated by the different sequence and structural elements of the DNA binding domains of TFs[96]. Among TFs, transcription activators can bind to the enhancer regions and physically interact with the promoter regions through DNA looping. Then, they facilitate the recruitment of basal transcription machinery or pre-initiation complex, which is composed of RNA polymerase II and a large number of general TFs (*e.g.* TFIIB, TFIID and TFIIF), to the core promoter region. Of note, some TFs directly bind to the promoter region and recruit the basal transcriptional machinery and other proteins such as co-factors. The recruitment of basal transcriptional machinery is an essential step in transcription since RNA polymerase II cannot independently bind to the DNA. After the recruitment to the core promoter, RNA polymerase initiates the transcription of the gene in the presence of other general TFs[97,98]. On the other hand, repressor proteins can inhibit transcription by obstructing the assembly of transcription machinery on the promoter[98]. In addition, repressor TFs inhibit transcription through other

mechanisms for example, establishing a repressive epigenetic environment (*e.g.* KRAB proteins)[99].

Chromatin remodeling by eviction or sliding nucleosomes provides transcription machinery access to regulatory regions in promoters, which is often required for the activation of transcription [100]. The SWI/SNF family of CRCs can disassemble nucleosomes to promote gene expression, while the ISW1 family closely assembles nucleosomes leading to transcriptional repression[43]. TFs often facilitate the recruitment of these CRCs to the DNA. For example, SWI/SNF complexes are recruited to target genes through association with c-Myc and CCAAT/enhancer-binding protein β[101].

1.4 Epigenetic modifications in cancer

The epigenetic landscape created by the interplay of epigenetic modifications regulates the formation and maintenance of different cell types, preserving the cellular identity throughout developmental stages of an organism. Therefore, alterations of epigenetic landscape in cells can contribute to a spectrum of diseases, in particular to cancer[5,16,102], which manifest their effect through global or local dysregulation of gene expression profiles[5,14]. For example, epigenome abnormalities can result in silencing of a tumor suppressor gene (TSG) or activation of a proto-oncogene, independently from genetic alterations or in conjunction with them. These dysregulations may act as drivers in cancer initiation that confer growth advantage to cancer cells, and therefore are positively selected during cancer evolution[103]. However, it is important to note that we are still at the door step of understanding epigenome alterations in cancer, and more studies are required to explore hidden territories in cancer epigenetics compared to cancer genome alterations that have been studied for decades.

1.4.1 DNA methylation aberrations in cancer

Aberrant DNA methylation was the first type of epigenetic abnormality identified in cancer[5,104]. Global hypo-methylation (reduction of DNA methylation levels in tumor tissues relative to the corresponding normal tissue from which the tumors were derived) and hyper-methylation (increase of DNA methylation levels in tumor relatively to normal tissues) are two distinguished characteristics of the cancer epigenome.

1.4.1.1 DNA hypo-methylation in cancer

The reduction of methylation occurs at genome-wide level and specifically at various genomic locations including promoters, enhancers and gene bodies[105-107]. Some gene promoters transition to a hypomethylated state in cancer, which is associated with an increase in gene expression. The demethylated genes mainly accommodate functions that are essential for cancer progression and metastasis, including cell growth, cell adhesion and communication, cell signaling, cellular migration, and invasion of normal tissue[108]. For example, the promoter of the engulfment and cell motility 3 (*ELMO3*) gene undergoes demethylation followed by over-expression of the gene in lung cancer. ELMO3 is involved in cellular migration and it is suggested to be associated with metastatic spread of lung cancer cells[109].

Hypo-methylation of enhancers is also associated with elevated gene expression in many cancers. For instance, Xiong *et al.* showed that recurrent hypo-methylation of enhancer of C/EBPβ gene activates a self-reinforcing enhancer-target loop, therefore overexpressing this gene in hepatocellular carcinoma, which is associated with poor prognosis[110]. However, DNA hypo-methylation alone is insufficient for enhancer activation and requires other histone modifications such as H3K27ac and H3K4me1[86,111].

In addition, hypo-methylation within the transcribed regions or gene body is usually negatively correlated with gene expression[112], and can also lead to activation of alternative transcription start sites within the gene body and, therefore, aberrant transcripts[113].

Even though the exact mechanisms of loss of DNA methylation in cancer and its functional consequences are not yet fully understood, we have already started to dissect these mechanisms. One leading possibility of global hypo-methylation is that, malfunction of DNMTs and TETs due to mutations[114].

1.4.1.2 DNA hyper-methylation in cancer

Hyper-methylation of one or few specific regions are observed in some cancers. Among those, hyper-methylation of promoter CGIs is widely studied in many malignancies. Hyper-methylation of a gene-associated CGI was first reported in the calcitonin gene promoter in two malignancies in the mid-1980s: small cell lung carcinoma and lymphoma[115]. Since then, CGI hyper-methylation has been reported in almost every tumor type[116] and it is integrally associated with transcriptional silencing of TSGs[114]. TSGs that are repressed by promoter hyper-methylation are involved in crucial cellular processes such as DNA repair, cell cycle progression, cell adhesion, apoptosis and angiogenesis[5]. Hyper-methylation of TSG

promoters, such as *BRCA1*, *APC*, *MLH1* and *CDKN2A*, has been reported in multiple cancers including glioblastoma, pancreatic, breast, colorectal and ovarian cancers[8].

In most of the cancer types, hyper-methylation is observed in ~5%–10% of genes associated with CGIs[114]. However, in some specific cancers, including low-grade gliomas and certain colorectal cancers, extensive and frequent hyper-methylation of thousands of GCIs is a characteristic feature. This phenomenon is known as CpG island methylator phenotype (CIMP)[117,118]. Similar to hyper-methylation of promoter CGIs, some enhancer regions are hyper-methylated and thereby lead to gene silencing[113]. On the other hand, gene body methylation is usually associated with elevation of gene expression level[119].

1.4.2 Histone modification aberrations in cancer

1.4.2.1 Histone acetylation in cancer

Histone acetylation is mostly localized at enhancers, promoters, and the gene body[120], and is associated with active gene expression. As such, hyper-acetylation of histones on regions such as enhancers and promoters that are associated with proto-oncogenes may play a role in activating the expression of these genes[79,114]. Conversely, hypo-acetylation of histones, which often co-occurs with DNA methylation, may contribute to gene silencing[27,114]. Among many potential mechanisms for hyper- and hypo-acetylation of histones, genetic or epigenetic aberrations in the HAT and HDAC genes are mostly studied. These enzymes also acetylate/de-acetylate a broad range of non-histone proteins, including p53, Rb, and MYC, which often form complexes with multi-subunit chromatin-modifiers. Abnormal modifications of these non-histone proteins often lead to alterations in many cellular pathways that support tumorigenesis[27,121].

1.4.2.2 Histone Methylation in cancer

It has been reported that many HMTs are associated with cancer[27]. Among them, the MLL family of HMTs is affected by genetic abnormalities, either via loss of function, translocation or rearrangements in many forms of cancer[27,122]. For example, MLL1, which is specifically responsible for H3K4 methylation, is frequently translocated in myeloid and lymphoid leukemias[27,122]. Enhancer of zeste homolog 2 (EZH2), which is the catalytic component of the Polycomb repressive complex 2 (PRC2), is a HMT responsible for di- and trimethylation of H3K27. PRC2 involved in transcriptional repression by reducing the accessibility of both TFs and CRCs such as SWI/SNF to DNA[123]. EZH2 has been reported to

have both oncogenic and tumor suppressor functions in numerous cancers[124]. For example, in several cancers including breast, prostate, and bladder cancers, EZH2 is overexpressed and is considered as an oncogene, leading to H3K27me3 accumulation[125]. In contrast, loss-of-function mutations of EZH2 suggests a potential tumor-suppressor role in myeloid malignancies[126] and KRAS-driven lung adenocarcinoma[124].

On the other hand, numerous types of HDMTs are implicated in cancer. LSD1, a type of HDMT, is frequently overexpressed in many cancers including acute myloid leukemia[126], ER-negative breast cancer[127], and neuroblastoma[128], and therefore is considered a classic oncogene[126].

Additionally, global changes of histone methylation are also reported in several cancers. For example, lower levels of H3K9me3 in non-small cell lung cancer[129], H3K4me2 in adenocarcinoma[130], H3K4me1, H3K9me2, and H3K9me3 in prostrate cancer[131] were observed. These alterations are associated with poor survival and worst prognosis in these cancers[132].

1.4.3 Accumulations of epigenetic alterations at enhancers

Genome-wide profiling of histone modifications has helped with the discovery of novel enhancers. The loss and gain of enhancers, relative to normal cells, which can be determined by the presence or absence of histone marks in a locus-specific manner, is a prominent feature in several cancers including ccRCC[79], colorectal cancer[133] and acute myeloid leukemia[111]. This differential activity of enhancers results in altered cancer gene expression. For example, in colorectal cancer, enhancers that are active in cancer cells but not in normal cells are found near the up-regulated genes[133].

What drives these locus-specific gain and loss of enhancers remains an outstanding question. One possibility is that these regions are affected by somatic mutations that introduce transcription factor binding sites. In support of this hypothesis, gained enhancers in colorectal cancer are commonly occupied by a set of RFs including AP-1 and cohesin factors[134]. Another example is recruitment of MYB to the TAL1 enhancer in T-cell acute lymphoblastic leukaemia as a consequence of a somatic mutation at the enhancer[135,136]

In addition, many other mechanisms and models have been suggested, related to epigenetic modifications underlying the dysregulation of enhancer activity. One hypothesis suggests that genetic alterations in chromatin remodeling factors and/or co-activators may

play a role. In support of this, *ARID1A*, *EP300* (*P300*), and *MLL3/4* genes are frequently mutated in bladder cancer, hepatocellular carcinoma, non-hodgkin lymphoma, medulloblastoma, breast cancer and colon cancer[8]. For instance, mutations in *MLL3/4* are thought to destabilize the MLL3/4 protein, thereby impacting the methylation of histone proteins at enhancers[137]. Other proteins whose dysfunction may result in aberrant enhancer activity in cancer include HDACs, HDMTs, HMTs and HATs, DNA methylation/de-methylation enzymes such as DNMT3A and TET, and enhancer-promoter interaction stabilizers such as cohesion [138-141].

1.4.4 Identification of regulatory elements

Alterations of activity of regulatory elements is one of the critical epigenetic features of tumorigenesis[79,114,142]. Therefore, many studies focused on detecting these activity changes in cis-regulatory elements. Chen *et al*. performed one of the largest enhancer activation studies using TCGA RNA-seq data across different cancer types. The premise of this study was that enhancers often produce RNAs called enhancer RNAs (eRNAs), which can serve as a proxy for enhancer activity. This study provided a systematic view of enhancer activity in different cancers based on expression of enhancer RNAs[143]. However, one limitation of their study was that, not all activated enhancers act as transcriptional units to produce enhancer RNAs[144]. Amongst the other methods to detect activity of regulatory elements, combination of specific histone marks, for instance H3K4me1 and H3K27ac for active enhancers, is widely used[85,145]. The first indication that active regulatory elements may be distinguished by specific histone modifications came from analyses of the data from the pilot phase of the ENCODE project[145,146]. Currently, ChromHMM[147] and Segway[148] are two widely used method to predict chromatin states such as active promoters and active enhancers by integrating different histone marks. The former uses a multivariate Hidden Markov model and latter uses a dynamic Bayesian network model to integrate data[12]. Some studies integrate histone modification data with other data types such as regulatory protein binding sites[149] and chromatin accessibility[149] in order to improve the prediction accuracy of detecting activity changes of regulatory elements.

1.5 Epigenetic abnormalities in ccRCC

1.5.1 Loss of histone modifiers and chromatin remodeler genes

Although ~80% of ccRCCs exhibit loss-of-function mutations in the VHL gene, VHL knockout in mice is not sufficient for inducing ccRCC[150]. Therefore, other genes must also be required for the development of this malignancy. Chromosome 3p hosts three frequently mutated genes that encode epigenetic modifiers: *PBRM1*, *BAP1* and *SETD2*, along with *VHL*[151]. Interestingly, a deletion of ~50 Mb on chromosome 3p with one or more of those genes is detected in 90% of sporadic ccRCCs[16]. In addition, other histone modifier genes, including *KDM5C* and *KDM6A* are mutated in ccRCC. These observations suggest a key role for epigenomics aberrations in ccRCC tumorigeneses and/or progression. **Table 1** Summarizes the functions of these genes.

Table 1: Functions of frequently mutated epigenetic modifiers in ccRCC

Gene	Main Functions of the protein
PBRM1	Chromatin remodelling, regulating replication and transcription by interact with numerous TFs[152].
BAP1	Regulate the expression of polycomb target genes in ccRCC[153].
SETD2	Tri-methylate histone H3K36[154], repair DNA mismatches[155] and double-strand breaks[156], recruitment of DNMT3A to non-promoter regions[154], maintenance of constitutive and facultative heterochromatin[157]
KDM6A	Demethylate H3K27me3[158]
KDM5C	Demethylate H3K4me1-3[159,160]

Other than the frequent mutations in well-known histone modifier and chromatin remodeler genes, dysregulation of many other genes have been reported. Some of these are briefly described below. JMJD3, a HDMT involved in the demethylation of H3K27, is over-expressed in ccRCC resulting in decreased H3K27 methylation level in tumors relative to adjacent non-tumor tissues[161]. Mixed-lineage leukemia protein 2 (MLL2), an HMT which directs H3K4 tri-methylation, exhibits altered expression in ccRCC tumors[158]. KDM6A interacts with MLL2, which also associates with KDM5C[162]. Nevertheless, the role of MLL2 in ccRCC tumorigenesis is currently unknown. As another example, over-expression of EZH2 is associated with metastasis and worse clinical outcome in ccRCC [163].

1.5.2 Other histone modification aberrations in ccRCC

Several studies reported alterations of histone modification in RCC by profiling of histone marks. Mosashvilli and collogues[164] profiled several histone marks including H3K9Ac and H3K18Ac in 193 RCC patients (including 142 ccRCC cases) using immunohistochemistry in a tissue microarray and found that global deacetylation of histone H4 was positively correlated with pathological stage and nuclear grade. In addition, they found that acetylation of histone H3 is associated with systemic metastatic spread and RCC progression. Moreover, low H3K18ac levels were associated with tumor progression[164,165]. Another study which profiled histone methylation in same patients described that global H3K4me1-3 levels were inversely correlated with lymph node involvement and distant metastasis in RCC[165]. In addition, another study showed that lower H3K27me3 levels were observed in patients with distant metastasis[166]. They also reported that the lower level of global methylation levels of H3K27me1-3 is associated with advanced pathological stage, higher Fuhrman grade and vascular invasion in RCC tumors[166].

1.5.3 DNA methylation aberrations in ccRCC

Many genes have been found to be inactivated through tumor-specific promoter hyper-methylation in ccRCC[167]. Among these genes, some have been suggested to play a role in ccRCC tumorigenesis, including genes that are involved in regulating essential cancer-associated pathways such as metastasis (*e.g.* RAP1GAP and CYTIP), cell cycle (*e.g.* RASSF1, KILLIN, and BTG3), Wnt signalling pathway (*e.g.* SFRP5 and WIF-1), DNA mismatch repair (*e.g.* MSH2)[168], and Keap1/Nrf2 pathway (*e.g,* Keap1)[16]. In addition, Xing *et al*. reported a comprehensive list of genes that were silenced by promoter methylation in ccRCC tumors[34].

Specifically, the CIMP has been observed in 20% of ccRCC tumors[18,169,170]. Arai *et al*. have identified methylation in 17 genes including FAM150A, GRM6, ZNF540, ZFP42, ZNF154 that are hallmarks of CIMP in ccRCC[171]. CIMP-positive ccRCCs are relatively more aggressive and are associated with worse patient outcome. These tumors also exhibit increased activity of anaerobic glycolysis pathway, which provides energy to tumor cells[16,172].

In addition to the promoter hyper-methylation, Caroline *et al.* have shown that many kidney-specific intronic enhancers in RCC are targeted by aberrant methylation changes which impact TF binding, resulting in transcriptional dysregulation[173]. In ccRCC, the

overexpression of JAGGED1 (JAG1), a ligand in the NOTCH signaling pathway, is another example of how aberrant hypomethylation at enhancer region, co-existing with H3K4me1, is associated with stimulation of gene expression[174].

Furthermore, changes in 5hmC, regulated by TETs, are also associated with changes in gene expression. ccRCC tumors show global 5hmC reduction compared to adjacent, normal kidney tissue[175]. However, further studies are required to understand the mechanisms underlying for 5hmC reduction and their role in tumorigenesis.

1.5.4 Functions of VHL and its role in epigenetic modifications in ccRCC

VHL is a multi-functional protein, whose best known function relates to the regulation of the hypoxia signalling pathway by targeting hypoxia inducible factors (HIFs) for proteasomal degradation. Nevertheless, given the multi-functional nature of VHL, it has been noted that the consequence of VHL inactivation in ccRCC is much broader than the activation of HIFs[176].

1.5.4.1 Characteristics and general function of VHL

Alternative splicing of the VHL gene transcript results in two VHL proteins of different sizes (213 residues with ~30 kDa and 160 amino acids with 18-19 kDa), which have similar capacities for tumor suppression[177-179]. In most ccRCC tumors, both proteins are inactivated by genetic or epigenetic mechanisms; Most related genetic changes are caused due to loss of function mutations. [16,154]. VHL can act as a multi-functional protein – it acts as an adapter that recruits different effector proteins[180,181]. In ccRCC, VHL is best known as the substrate-binding subunit of a SCF-type E3 ubiquitin ligase complex[181]. VHL regulates the ubiquitylation of HIF subunits (HIF1 and 2) for proteasomal degradation, thereby regulating the HIF stability in an oxygen dependent manner.

1.5.4.2 HIF-VHL gene regulation

HIF TFs are heterodimers composed of two of five HIF protein subunits: HIF1α, HIF2α, HIF3α, HIF1β and HIF2β. HIF1α is constitutively expressed while HIF2α is mainly restricted to endothelial, lung, kidney and hepatic cells. Both HIF1α and HIF2α are stabilized at hypoxic conditions and thereby form a dimer with stable HIF1β. This protein-dimer binds with DNA elements called hypoxia-response elements (HREs), subsequently activating the transcription of hundreds of target genes (hypoxia responsive genes). HIFs mainly activate the transcription of HRGs as an acute or chronic adaptive response to hypoxic (low oxygen)

tension[16,180,182]. Furthermore, HIFs (specially HIF1α and HIF2α) regulate genes associated with many cellular pathways including proliferation and survival (*e.g: EGFR*, *TGFR-1*) and angiogenesis (*e.g. VEGF* and *PDGF*)[18,180,183].

HIF1α exists in hydroxylated and non-hydroxylated forms under normal and hypoxic conditions, respectively. VHL specifically recognizes the hydroxylated form of HIF1α and targets it for degradation. Because VHL function is lost in the majority of ccRCCs, HIF is constitutively active in these cancers independent of the normaxic/hypoxic condition of the cell, thereby driving the transcriptional activation of the hypoxia responsive genes[182,184,185].

1.5.4.3 VHL-mediated DNA methylation

Recently, our lab in collaboration with another research group at University of Toronto investigated the DNA methylation patterns in ccRCC cell lines in which VHL has been stably reconstituted, comparing the changes relative to VHL-deficient ccRCC cell lines, as well as the relationship between these changes and those seen in patient tumor samples vs. normal kidney tissues (using 450k methylation microarrays). The authors found that VHL reconstitution rescues a modest proportion of the methylation patterns observed in normal kidney cells. Interestingly, they also found that VHL-mediated methylation changes are hypoxia-independent[186]. However, the extent of these methylation changes, the specific affected areas, their functional consequences, and the mechanisms that drive them as a response to VHL are yet to be determined.

1.5.4.4 Epigenetic remodeling in VHL-deficient ccRCC

In a recent study, Yao and colleagues[79] investigated the consequences of VHL deficiency on chromatin alteration at *cis*-regulatory elements, specifically promoters and enhancers/super-enhancers. This is the most comprehensive study that reports multiple epigenome profiles in ccRCC with more than 10 samples using ChIP-seq data. They found substantial chromatin level changes in these cis-regulatory elements by profiling H3K4me3, H3K4me1 and H3K27ac histone marks in both tumor and normal samples. Interestingly, a set of gained and lost promoters and enhancers in ccRCC were identified in the study. They also investigated the effect of VHL on these alterations and found that they are mediated by both HIF-dependent as well as HIF-independent manner. The study found that the genes associated with enhancers are enriched in ccRCC specific processes such as HIF and pro-angiogenic pathways and metabolism. Moreover, they discovered a set of oncogenes that were affected by epigenetic modifications at super-enhancers, including *VEGFA* and *EPAS1*.

Intriguingly, they also found a master regulator of ccRCC, ZNF395, which is associated with a gained super enhancer that leads to overexpression of that gene – the authors reported that knock-down of ZNF395 results in decreased cell proliferation and viability both *in vitro* and *in vivo* in ccRCC[79,187].

1.6 Transcriptome changes in ccRCC and their drivers

Other than the studies of epigenetic modifications and associated gene expression changes, a number of studies reported the transcriptome changes in ccRCC. For example, Scelo and colleagues used 94 patient samples from four European countries and analysed transcriptomic changes in tumor relative to normal samples[188]. They highlighted significant alterations in several pathways including focal adhesion and phosphatidylinositide 3-kinase (PI3K) pathways mainly due to genetic modifications (i.e somatic mutations)[188]. Interestingly, several other studies reported that epigenetic modification is also responsible for a substantial fraction of these gene expression changes in ccRCC. For instance, a study by Bhagat *et al.* revealed that both genetic (*i.e:* copy number alteration) and epigenetic alterations (*i.e* DNA methylation changes) lead to NOTCH pathway activation in ccRCC[189]. Another study by Wozniak *et al.* reported that down-regulated genes are enriched in many pathways including metabolic and catabolic processes. Simultaneously, upregulated genes were associated with pathways such as immune and hypoxia responses. However, they proposed that only 7% of these gene expression alterations could be explained by epigenetic changes[190]. As I will discuss in the next chapters, this number is likely a gross under-estimation of the role of epigenetic alterations in mediating gene expression changes in ccRCC.

1.7 Mapping of epigenetic modifications and data generation

Over the last few years, epigenetics has become a major focus of scientific research, with a growing number of studies examining genome-wide aberrations of epigenetic modifications across diseases. Of the many assays used, chromatin immunoprecipitation and bisulphite treatments are the two most common techniques for assessing the genome-wide profiles of histone modifications and DNA methylation, respectively[191]. In combination with these assays, microarray (*e.g:* 450K microarrays) and high-throughput sequencing technologies have provided the capability to produce genome-wide high-resolution maps of DNA methylation and histone modifications in normal tissues and diseases such as cancer[192].

ChIP-seq and WGBS are widely used high-throughput sequencing methods that have are used towards comprehensive studies of epigenetic modifications.

i. ChIP-seq has been used not only to map genome-wide epigenetic modifications, specifically histone marks, but also TF binding maps. Chromatin immunoprecipitation enriches DNA fragments that are bound by specific protein or nucleosomes with specific histone marks. After crosslinking and pull-down of the protein of interest, DNA fragments are sequenced and analyzed to identify the protein binding sites[193].
ii. WGBS determines the genome-wide DNA methylation status of cytosines by treating the DNA with sodium bisulphite followed by high-throughput sequencing[193]. This method directly estimates the absolute methylation levels at single-CpG resolution. However, WGBS requires deep coverage of the entire genome,[194] and therefore comes with a significant experimental cost.

Many large-scale initiatives have been established for systematic mapping of epigenetic and related data in normal tissues and across diseases[12]. These include projects by the Alliance for the Human Epigenome and Disease (AHEAD) Task Force[195], the ENCyclopedia Of DNA Elements (ENCODE) Project Consortium[196,197], the Human Epigenome Project (HEP) Consortium[198], RoadMap Epigenome Project Consortium[21] and the International Human Epigenome Consortium (IHEC)[199]. The Cancer Genome Atlas (TCGA)[200] and International Cancer Genome Consortium[201] provide many data types including gene expression and epigenomic specifically for cancers. Moreover, Cistrome is a centralized database of histone modifications[202], which integrates TCGA gene expression data with public ChIP-seq data in cancer. Different types of datasets such as DNA microarrays, ChIP-seq and WGBS generated by these projects are freely available to the scientific community.

1.8 Applications of machine learning on epigenetic data analysis

Machine learning is a rapidly developing field based on pattern recognition[203]. It is an automated process of detecting patterns in large-scale datasets using computer-based statistical models, where a fitted model may then be used for classification of previously unseen patterns on new datasets[204]. Two major machine-learning approaches that have been used in biological data analysis include supervised and unsupervised learning. Supervised algorithms are used when there are labeled training data of two or more classes of interest and unsupervised algorithms are used when the samples are not labeled[205]. Algorithms that are widely used for supervised learning include but are not limited to support vector machines,

regression, artificial neural networks, decision trees, and random forests. Some of the most commonly used unsupervised algorithms include non-negative matrix factorization, principal component analysis, K-mean clustering, and hierarchical clustering among many others [203,205].

In supervised learning, the goal is to predict the (unobserved) label of an object given its (observed) properties[206]. For example, we might want to predict whether a particular genomic region represents an enhancer or not (unobserved label), based on the histone marks that are present in that region (observed properties). A supervised machine-learning algorithm can use a "training set", e.g. a set of genomic regions with known labels and histone marks to learn a model that connects the histone mark patterns to the labels, and then uses this model to predict the labels of other genomic regions that are potentially uncharacterized. The process of training a supervised machine-learning model usually consists of several steps such as data gathering and preparation (which depends on the problem that is being addressed), training[207] and hyper-parameter tuning of the machine-learning model[208], and validation of the model to ensure that its performance is acceptable and that the model is generalizable (i.e. it performs well even on cases that are not included in the training set, such as a held-out validation set or an independent test set) [207-209].

Machine learning is widely used in various biological domains including genomics, epigenomics and proteomics[210]. Examples include a large number of studies that have shown the power of different machine learning algorithms for the analysis of epigenetic data and predicting gene expression. For example, support vector machines have been used for genomic mapping of methylation patterns for all 22 human autosomes[211] and prediction of methylated CpGs[205]. Orozco *et al*., introduced a DNA methylation based random forest classifier to aid in the diagnosis of brain metastases[212]. Moreover, multiple linear regression and multivariate adaptive regression splines have been utilized to build predictive models of gene expression as a function of histone modifications[213].

1.9 Hypothesis

Abnormal epigenetic alterations in ccRCC affect *cis*-regulatory elements, and therefore contribute to abnormal gene expression in ccRCC.

1.10 Objectives

1. To characterize epigenome changes in ccRCC and investigate their role in mediating gene expression changes.
2. To identify potential drivers of these epigenetic changes and their connection to gene expression in ccRCC.

CHAPTER 2. MATERIALS AND METHODS

In order to understand the epigenetic changes in ccRCC and its connection to gene expression changes, we began by analyzing the genome-wide maps of three histone modifications (H3K27ac, H3K4me3 and H3K4me1 as well as DNA methylation (WGBS data) in the primary tumors of four ccRCC patients as well as matching normal tissues. We used a machine-learning approach to combine these maps and identify enhancers and proximal regulatory elements that are gained or lost in tumor compared to normal tissue. We also used machine-learning to connect gain or loss of enhancers/promoters to gene expression changes in ccRCC. We then investigated the TFs that bind these gained/lost elements and, therefore, may mediate the associated gene expression changes. Finally, we performed RNA-seq on ccRCC cell line models in two oxygen conditions (*i.e:* hypoxic and normoxic) that are deficient of VHL or express an ectopically reconstituted VHL, in order to understand the role of VHL in mediating these epigenetic and gene expression changes. The following sections describe the details of the methods we used to obtain and analyze these data.

2.1 Patient Information

The ccRCC patient samples that underwent epigenome profiling were provided by CAGEKID consortium. Epigenome and gene expression profiling experiments of these samples were performed by McGill Epigenome Mapping Centre (EMC) as described previously[214-216]. Refer to **Supplementary Table S1** for detailed patient information.

The data were downloaded from the EMC data portal. (https://genomequebec.mcgill.ca/nanuqMPS/project/ProjectPage/projectId/9015)

2.2 Histone ChIP-seq Analysis

ChIP-seq raw reads of eight patient samples (four pairs of normal and tumor tissues) were filtered by phred quality score (phred33>=30) and length (>=32), followed by adapter trimming using Trimmomatic[217] (version 0.22). Single end sequencing tags were mapped against the human reference genome (GRCh38) using bowtie2[218] (version 2.2.9) with default parameters. Reads with mapping quality (MAPQ score) >30 were chosen using Samtools[219] (version 1.3) for subsequent analysis. Significant broad peaks were called using MACS2[220] (version 2.1.1.20160309) with a FDR threshold of 0.05.

2.3 RNA-Seq differential gene expression Analysis

RNA-Seq raw reads that were obtained from RCC4 cell lines (2 replicates per each oxygen condition) or ccRCC patients (four sample pairs) were filtered by phred quality score (phred33>=30) and length (>=32), and adapters were removed by Trimmomatic[217] (version 0.22). Paired end sequencing tags were mapped against the human reference genome (GRCh38) using HISAT2[221] (2.0.4). Only reads with mapping quality >30 were used for subsequent analysis. HTSeq-count[222] (version 0.9.1) was used to count reads using "intersection-strict" and "reverse" parameters and genome annotation from Gencode[223] (GRCh38). The resulted raw counts were used for differential gene expression analysis using DESeq2[224].

In parallel, raw read counts of 29 matched tumor-normal paired samples of ccRCC patients with VHL mutations were obtained from the CAGEKID cohort[188] and differential gene expression analysis was performed using DESeq2[224] similar to what is described above.

2.4 Defining proximal regulatory elements (PREs)

Overlapping H3K4me3 and H3K27ac broad peaks were extracted from tumor and normal samples. Their width was extended by 100bp from both sides to reach a total width of 200bp. Only regions within ±10kb of TSS were retained. All the regions identified across samples were then pooled together and overlapping regions between normal and tumor samples were merged using Bedtools[225] merge (version 2.26.0). The length of each element was normalized to 1kb or 10kb for downstream analyses.

2.5 Defining enhancers

Same procedure of defining PREs was followed except that the overlapping H3K27ac and H3K4me1 broad peaks were obtained from all the samples, removing regions within ±10kb of any gene body.

2.6 Average DHS signals at the centre of enhancers

Kidney DHS data of 89-day old female fetus obtained from chromatin accessibility assays using DNase I hypersensitivity were downloaded as a wig file from the Gene Expression Omnibus (GEO)[226] repository (GEO Accession number GSM1027338[227]). The file was then converted to bigWig format using "wigtobigwig" utility from ENCODE

portal[228]. All enhancers from normal and tumor samples were centre aligned. We then calculated the average DHS signal value at different distances from the centre of enhancers using bwtool[229] aggregate (version 1.0). Random genomic coordinates with the same amount of enhancers were produced by Bedtools[225] random (version 2.26.0). Random regions and all enhancers from Yao *et al.*[79] were used to generate average DHS signals for comparison with our enhancers.

2.7 Overlapping enhancers from tumor and normal samples with GenoSTAN enhancers

GenoSTAN[230] enhancer data were downloaded from https://i12g-gagneurweb.in.tum.de/public/paper/GenoSTAN/. Bedtools[225] intersect (version 2.26.0) was used to calculate the percentages of overlapping coordinates of GenoSTAN enhancers[230] with our enhancers (pooled from both normal and tumor samples), along with two sets of random coordinates (as negative control) as well as enhancers from a previous study[79]. Overlapping percentages were calculated by incrementing the number of cell types in which GenoSTAN enhancers were identified.

2.8 WGBS data analysis

Raw reads were filtered based on their quality (phred33 >= 30) and length (n >= 50). Illumina adapters were then trimmed using Trimmomatic[217] (version 0.36). The trimmed reads were aligned per sequencing lane to the pre-indexed reference genome by Bismark[231] (version 0.18.2) with bowtie2[218] (version 2.3.1) in paired-end mode and with default parameters. BAM files from different lanes were merged using Picard (version 2.9.0). We then removed duplicated reads using the Bismark[231] function "deduplicate_bismark". Methylation calls were obtained using the Bismark[231] function "bismark_methylation_extractor". Finally, methylKit[232] was used to find differentially methylated CpGs using extracted methylation values (FDR <0.05; hyper-methylated: $\Delta\beta > 0$; hypo-methylated : $\Delta\beta < 0$).

2.9 Classification of gained and lost elements of ccRCC

Two separate supervised random forest machine learning models, generated by R package "randomForest"[233,234] (version 4.6-14), were used for the gain/loss classification of enhancers and promoters. The following predictor variables were included in each model

(query location corresponds to the genomic region that we want to classify as either gained or lost):

- The number of hypo- or hyper-methylated CpGs within 1kb or 10kb of each query location (four features).
- The number of total CpGs with 1kb or 10kb of each query location (two features; included in order to control for the effect of CpG number).
- The total number of regulatory elements that were detected across normal samples within 1kb or 10kb of each query location (two features).
- The total number of regulatory elements that were detected across tumor samples within 1kb or 10kb of each query location (two features).

As the **gold standard or target** for training, gained and lost events of regulatory elements identified in a previous study[79] were used. Model performance was determined by cross-validation (leaving one chromosome out in each round, training the model on the remaining chromosomes, and testing the performance on the left-out chromosome). Trained models were then used to classify gained and lost elements in ccRCC tumors across the whole dataset.

2.10 GSEA pre-ranked test

Log2fold change of gene expression in tumor vs. normal samples and their base mean expression values were used to calculate IHW[235] weights. All genes with IHW weight more than the 10% of maximum IHW weight in the matrix were extracted and ranked by their log2 fold change. This ranked gene list was used for the pre-ranked test using GSEA[236] software to identify enriched gene sets from a custom Gene Matrix Transposed (GMT) database file created using target genes of enhancer-associated regulatory factors.

2.11 450k methylation microarray data analysis

Raw 450MD files of patients (from 79 VHL mutated patients) and cell lines (RCC4 and 786-O; three sample pairs per each cell line) were obtained from a previous study by Robinson *et al.*[186], and the Minfi[237] Bioconductor R package (version 1.20.2) was used to analyse 450k methylation data. Raw methylation data was normalized using functional normalization (preprocessFunnorm function[238]) and all the loci with single nucleotide polymorphisms (SNPs) were dropped. Beta values were calculated using the formula M/(M+U+100); where M and U denote the methylated and unmethylated signals,

respectively, and 100 is a pseudo count added to prevent division by zero and stabilize the ratio for probes with low signal.

2.12 RCC4 cell culture and reagents

RCC4 ccRCC subclones stably expressing HA-VHL (RCC4-VHL; VHL+) or empty plasmid (RCC4-MOCK; VHL-) cells were kindly gifted by Prof. William G. Kaelen, Harvard University, USA in collaboration with with Prof. Janusz Rak, McGill University, Canada. They were maintained in DMEM (Gibco) supplemented with 1 μg/ml G418, 100 μg/ml penicillin/streptomycin and 10% (vol/vol) heat-inactivated FBS (Sigma), and incubated at 37 °C and 5% (vol/vol) CO_2 in a humidified incubator. For hypoxia treatment, cells were maintained at 1% O_2 for 3 days in a humidified hypoxia chamber at 37 °C.

2.13 RNA isolation and RNA-Seq of RCC4 cells

Total RNA was extracted from cultured cells in both oxygen conditions using miRNeasy kit (Qiagen). Briefly, cells were washed with PBS and lysed with 700μl QIAzol reagent for each well in a 6-well plate. Lysates were processed according to the supplier protocols for total RNA isolation. RNA Quantification was done using the Nanodrop spectrophotometer. The isolated RNA was used to generate rRNA-depleted first-strand cDNA libraries using TruSeq Stranded Total RNA-LT (Ribo-Zero Gold, Illumina). The libraries were then sequenced on an Illumina HiSeq 4000 PE100 platform, producing a minimum of 50 million paired-end 100-bp reads per sample.

2.14 The enrichment analysis of regulatory factor binding sites

Coordinates of the binding sites of 161 known regulatory factors (RFs) from ENCODE[196] (using Regulation "Txn Factor ChIP" track") were downloaded from the UCSC Genome Browser[239], and normalized to a total width of 400bp from the centre. Four background sets were used to examine the relative enrichment of each RF within 1kb of the lost or gained enhancers, as follows: i) the union of gained and lost enhancers, ii) the union of gained and lost enhancers that have at least one binding site for at least one RF iii) the set of all the enhancers irrespective of their gain or loss score, and iv) all enhancers with at least one binding site for at least one RF. Hypergeometric test was performed for each RF and the significance of the enrichment was calculated, followed by p-value correction for multiple hypothesis testing (FDR).

CHAPTER 3. RESULTS

3.1 Epigenome aberrations on regulatory elements can be identified by histone modifications

To investigate the epigenetic alterations on regulatory elements and their connection to dysregulation of gene expression in ccRCC tumors, we first generated histone chromatin immunoprecipitation sequencing (ChIP-seq) profiles (three histone marks associated with active transcription: H3K27ac, H3K4me3, and H3K4me1) in four primary tumor/normal matched pairs (refer to **Supplementary Table S1** for detailed information of patients). In total, we produced 684,392,482 uniquely mapped ChIP-seq reads (mapping quality >30; refer to **Supplementary Table S2** for ChIP-seq statistics).

Since histone modifications are best represented by broad peaks, we then obtained broad peaks from our ChIP-seq samples (using MACS2[220]). On average, 91.2% of H3K27ac, 98.9% of H3K4me1 and 93.3%of H3K4me3 broad peaks that we obtained from normal kidney tissues overlapped with respective peaks from adult kidney tissues in the Roadmap Epigenomics dataset[21], which was substantially higher than expected by chance (**Supplementary Figure S1A**). This observation suggests that overall the histone marks we identified are consistent with other datasets from similar tissues.

We further examined our ChIP-seq data for enrichment of histone marks at the upstream flanking region, gene body and downstream flanking region of genes in normal tissue samples in order to investigate the distribution of different histone modifications relative to gene structure and their relationship to gene expression. Based on RNA-Seq data that we generated from the same tumor/normal samples, we found that histone marks associated with active transcription (H3K27ac, H3K4me1 and H3K4me3) are enriched near the TSS of highly expressed genes and depleted in genes with low expression (**Supplementary Figure S1B**), providing further support that our ChIP-seq-based epigenetic measurements mirror gene expression.

To investigate whether epigenetic differences between tumor and normal tissues reflect gene expression differences, as suggested in previous studies[240,241], we first compared ChIP-seq signals of histone marks around transcription start sites (TSSs) between tumor and normal tissue samples to identify cancer-associated alterations in histone modification patterns. We then performed a differential gene expression analysis using RNA-Seq data of

same tumor and normal samples to identify up- and down-regulated genes (using DESeq2[224]) in tumors. A comparison between cancer-associated alterations in histone marks and gene expression patterns revealed that up-regulation in tumors accompanies gain of activating histone marks near TSS (**Supplementary Figure S1C**). Overall, these analyses confirm the expected relationships between (differential) gene expression and (differential) histone modification in ccRCC tumor and matching normal samples, paving the way for a more detailed analysis of the expression and epigenetic landscapes of ccRCC, as described in the next sections.

3.2 Active distal and proximal regulatory regions in ccRCC

To further understand the contribution of epigenome abnormalities to cancer-associated gene expression patterns in ccRCC, we focused our analysis on regulatory elements. Different categories of regulatory elements are characterized by co-occurrence of specific histone modifications. For example, active distal enhancers (*i.e.* enhancers that are away from TSS) are specifically delineated by co-occurrence of H3K4me1 and H3K27ac, while active promoters are associated with the simultaneous presence of H3K27ac and H3K4me3[79,85,242]. Therefore, we defined active distal and proximal regulatory elements (PREs) in ccRCC by integrating patterns of the three histone mark broad peaks from our ChIP-seq data as described below. In each normal or tumor sample, we identified regions within ±10kb distance from TSSs that showed overlapping H3K27ac and H3K4me3 peaks and labelled them as PREs. These PREs include core promoter elements that are usually located within hundreds of base pairs from TSS[243], as well as other regulatory elements in their vicinity that may contribute to gene expression. Similarly, active distal enhancers were defined as regions with co-occurring H3K27ac and H3K4me1 marks in areas outside of ±10kb vicinity of any gene (refer to **Figure 7A-C** and methods section for complete details of the procedure).

Overall, we found 24476 putative active enhancers, of which 2347 and 15571 were exclusively detected in normal and tumor tissues, respectively (normal-exclusive enhancers, for example, are defined as those that are not within 1kb of any enhancer found in tumor samples). Also, 6558 enhancers were common in both tissue types. Likewise, from a total of 20824 PREs identified in our analyses, 2506 and 3884 were specific to normal and tumor tissues, respectively, while 14434 were shared in both (**Figure 7D**).

Notably, the defined enhancers (either from normal or tumor sample) exhibited high enrichment in regions with kidney-specific open chromatin status (DNase I hypersensitive

site (DHS) signals[227]) as compared to randomly selected regions on the genome (**Figure 7E**). Moreover, a majority of the identified enhancers (72.7%) coincided with GenoSTAN enhancers[230] from 127 cell lines whereas percentage greater than expected by chance (**Figure 7F**). In addition, over a third of enhancers and promoters (37.7%) identified in our study had also been found in a recent paper[79] reporting on the epigenome landscapes of ccRCC (**Supplementary Figure S1D**). Lastly tumor-specific and normal-specific active regulatory elements identified in our study were significantly enriched for gained and lost elements, in ccRCC, respectively, as reported by Yao *et al.* (**Supplementary Figure S1E**). Overall, these observations confirmed that our defined enhancer and promoter elements are consistent with available data on gene regulatory elements of kidney tissue, but also contain a substantial number of new regulatory regions that have not been previously identified.

3.3 Differential DNA methylation correlates with differential activity of regulatory elements

To investigate DNA methylation patterns at the regulatory elements and their contribution to the activity of regulatory element, we generated WGBS profiles for three tumor/normal sample pairs (for which DNA samples were available) that were profiled for histone modification marks and one additional pair (refer to **Supplementary Table S1** for detailed information of patients). We obtained methylation data for 13,492,181 CpGs (see **Supplementary Table S3** for WGBS statistics), among which 855,694 and 133,383 loci were hypo- and hyper-methylated, in tumors compared to normal samples, respectively (FDR < 0.05).

By superimposing differentially methylated CpGs (hypo- or hyper-methylated in tumors) with coordinates of our regulatory elements (both enhancers and PREs) we identified 1816 and 95 enhancers (out of 24476) with a statistically significant excess of hypo- and hyper-methylated CpG dinucleotides, respectively, while the equivalent figures for PREs are 1071 and 212 (binomial test; IHW-adjusted FDR < 0.1). This analysis suggests that enhancers and PREs are hotspots for differential methylation, although our statistical power is limited by the number of CpGs that are present in each regulatory element.

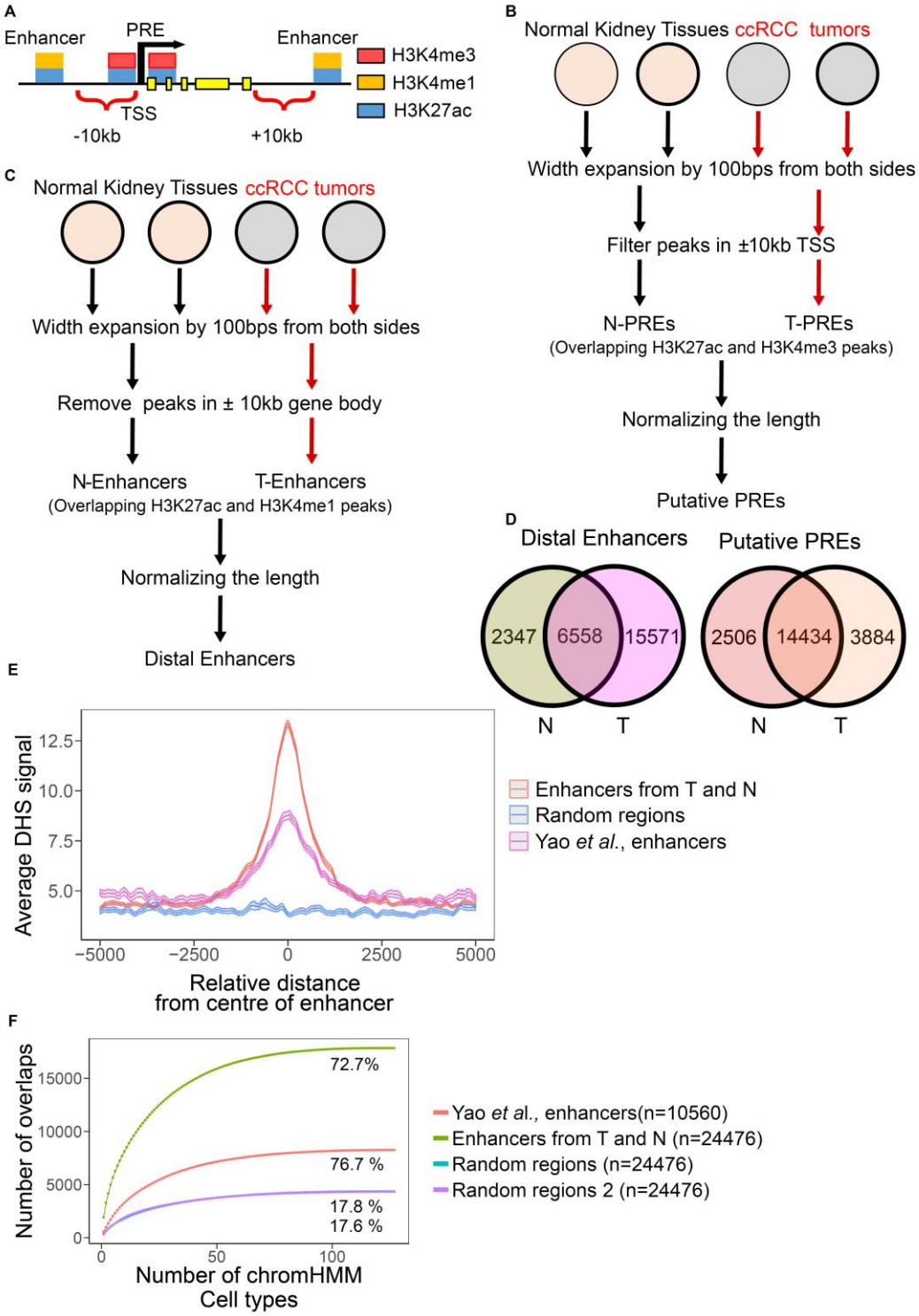


Figure 7: *Cis*-regulatory elements of the genome are characterized by histone modifications: A. Proximal Regulatory Elements (PREs) are defined by the co-presence of H3K4me3, H3K27ac, and proximity to TSS within 10 kb. Putative distal enhancers are defined by the co-occurrence of H3K4me1, H3K27ac, and mutual exclusivity with PREs; **B.** Overlapping H3K4me3 and H3K27ac broad peaks are obtained from both tumor and normal samples and their width is extended by 100bp from both sides (total width becomes 200bp).

Only regions within ±10kb of TSS are retained. All the regions are then pooled together and overlapping regions are merged. For downstream analyses, in order to identify overlapping features (such as DNA methylation), the length of each element is normalized to 1kb (core element) or 10kb; C. Defining enhancers follows the same procedure as **Figure 7B** by obtaining overlapping H3K4me1 and H3K24ac broad peaks from both tumor and normal samples, with the difference that regions within ±10kb of any gene body are removed; **D.** The Venn diagram shows the number of enhancers and PREs identified in tumor (T) and normal (N) tissues and their overlap. Overlapping elements are identified based on 1kb length extension (core elements); **E.** Enrichment of enhancers from this study, randomly selected regions and enhancers identified in ccRCC from a previous study (Yao *et al.,*[79]) in regions with kidney-specific open chromatin status (DNase I hypersensitive site (DHS) signals[227]) obtained from left kidney of 89 days old female fetus (GEO Accession number GSM1027338[227]); **F.** Number of overlaps of enhancers from T and N, two random genomic coordinates and enhancers from a pervious study (Yao *et. al.,*[79]) between GenoStan[230] enhancers (from 127 ENCODE and Roadmap Epigenomics cell types) with the accumulation of number of cell types. Percentage values show the number of overlaps in each comparison at 127 cell lines.

Next, we examined the relationship between activity status of regulatory elements, as defined by active histone marks, and their DNA methylation patterns (refer to **Supplementary Figure S2** for the relationship between number of CpGs and statistical power). Regulatory elements with significant excess of hypo-methylated CpGs were more frequently found in tumor samples, whereas regulatory elements with excess of hyper-methylated CpGs were more likely to be found in normal samples, suggesting that DNA hypo-methylation is in fact a strong indicator of tumor-specific regulatory elements (**Figure 8A**). To verify these findings, we examined the distribution of hypo- and hyper-methylated CpGs in a list of gained and lost regulatory elements in ccRCC, which has been identified in a recent study[79]. We observed that regulatory elements that are marked predominantly with hyper-methylated CpGs in our data are also enriched for lost regulatory elements reported by Yao *et al.* while those regulatory elements that are characterized by hypo-methylated CpGs in our study show enrichment in their gained elements (**Figure 8B**). Taken together, our findings corroborate previous data about epigenetic alterations in ccRCC, and suggest that abnormal DNA methylation patterns along with alterations of histone marks may be involved in dysregulation of active regulatory elements in ccRCC.

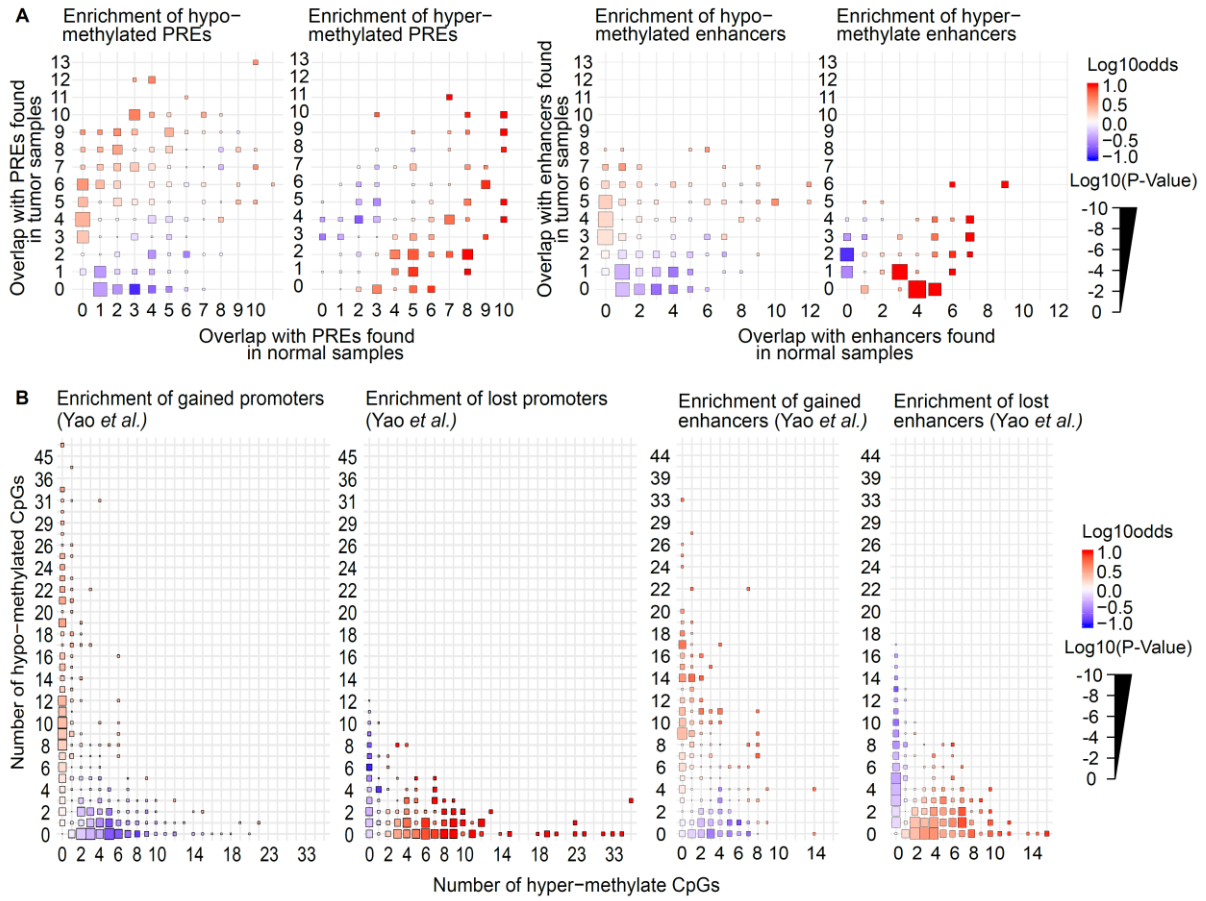


Figure 8: DNA methylation is a good indicator of gain and loss of regulatory elements: A. Distribution of significantly hypo- and hyper-methylated (binomial test; IHW-adjusted FDR < 0.1) regulatory elements with respect to the frequency of their observation in tumor and normal samples. The x- and y-axes correspond to the number of times an overlapping element is observed in any normal and tumor sample, respectively. The color of each box denotes the enrichment of the regulatory element type indicated on top of each panel (red: enrichment; blue: depletion). For example, the red box at x=0 and y=4 in the left graph indicates that hypo-methylated PREs are enriched among those that overlap zero elements in normal samples and four elements in tumor samples (relative to what would be expected by chance). The box size corresponds to the P-value of enrichment or depletion, and the color gradient corresponds to logarithm of odds (Fisher's exact test); **B.** Distribution of gained and lost regulatory elements, as defined by a previous study[79], with respect to their methylation status in our data. The x- and y-axes correspond to the number of hyper- and hypo-methylated CpGs that overlap each element, respectively. Annotations are similar to **panel (A)**.

3.4 Integrative analysis of histone modifications and DNA methylation uncovers the landscape of gain and loss of active regulatory elements in ccRCC

Since ChIP-seq and DNA methylation patterns from our data were well correlated with previously reported gained and lost active regulatory elements, we hypothesized that the employment of a supervised machine learning model (random forest) that integrates information across histone modifications and DNA methylation should provide more power for the identification of gain and loss events in ccRCC. We utilized our ChIP-seq and WGBS data from patient samples as predictor variables to train separate random forest models for enhancers and PREs in order to distinguish gain events from loss events (using gain and lost events identified in a previous study[79] as the gold standard for training; see Methods for details). Receiver operating characteristic (ROC) curve analyses revealed high sensitivity and specificity for our approach in identifying regulatory elements (leave-one-chromosome-out cross-validation), with area under the curve (AUC) values of 0.995 and 0.999 for enhancers and PREs (P-value < 2.59×10^{-73}), respectively **(Supplementary Figure S3A**). When we applied this random forest classifier to the entire set of regulatory elements, at a score cutoff of 0.95 the classifier identified 5345 gained enhancers, 2466 lost enhancers, 11503 gained PREs and 694 lost PREs in ccRCC tumors. Notably, we identified 4066 gained enhancers, 1747 lost enhancers, 7866 gained PREs and 575 lost PREs in ccRCC that have not been reported in previous epigenetic studies of ccRCC[79].

To investigate whether the random forest model has learned biologically meaningful rules for classification of regulatory elements, we examined the properties of regulatory elements that were identified by the classifier as gained, lost, or stable (*i.e.* not changed between tumor and normal). As **Figure 8A** shows, enhancers that have mostly hypo-methylated CpGs and few hyper-methylated CpGs are mostly labeled as "gained enhancers" by the classifier, as expected, and the opposite pattern is observed in lost enhancers. Comparably, regions that have balanced hypo- and hyper-methylated CpGs are mostly labeled as "stable enhancers". The same pattern is observed in PREs that are identified as gained, lost or stable by the PRE-classification model (**Supplementary Figure S3B**). Moreover, the regulatory elements status that is identified by the random forest models are well correlated with their sample of origin. For example, as **Figure 9B,** shows, the regions

that have more overlap with tumor sample enhancer elements are more likely to be labeled as gained enhancers, and vice versa (see **Supplementary Figure S3C** for PREs).

Would it be possible to identify gained or lost regulatory elements simply based on their sample of origin? Our analyses highlight cases where the sample of origin alone is not informative about gain or loss of active regulatory elements, but the model uses additional information from DNA methylation to make an informed decision. For example, in cases where an enhancer overlaps one element identified from a tumor sample and one element identified from a normal sample, the model classifies it as a gained enhancer when there are substantial number of hypo-methylated CpGs in the enhancer (**Supplementary Figure S3D**). Similarly, there is a case in which an enhancer overlaps multiple elements identified from tumor samples and no elements identified in normal samples (perhaps due to lack of statistical power or due to sample preparation issues), but the model still classifies it as a "lost enhancer" due to the presence of a large number of hyper-methylated CpGs (**Supplementary Figure S3E**). These observations suggest that our classifier utilizes information across histone marks and CpG methylation data to accurately predict a comprehensive and reliable set of gained and loss *cis*-regulatory elements.

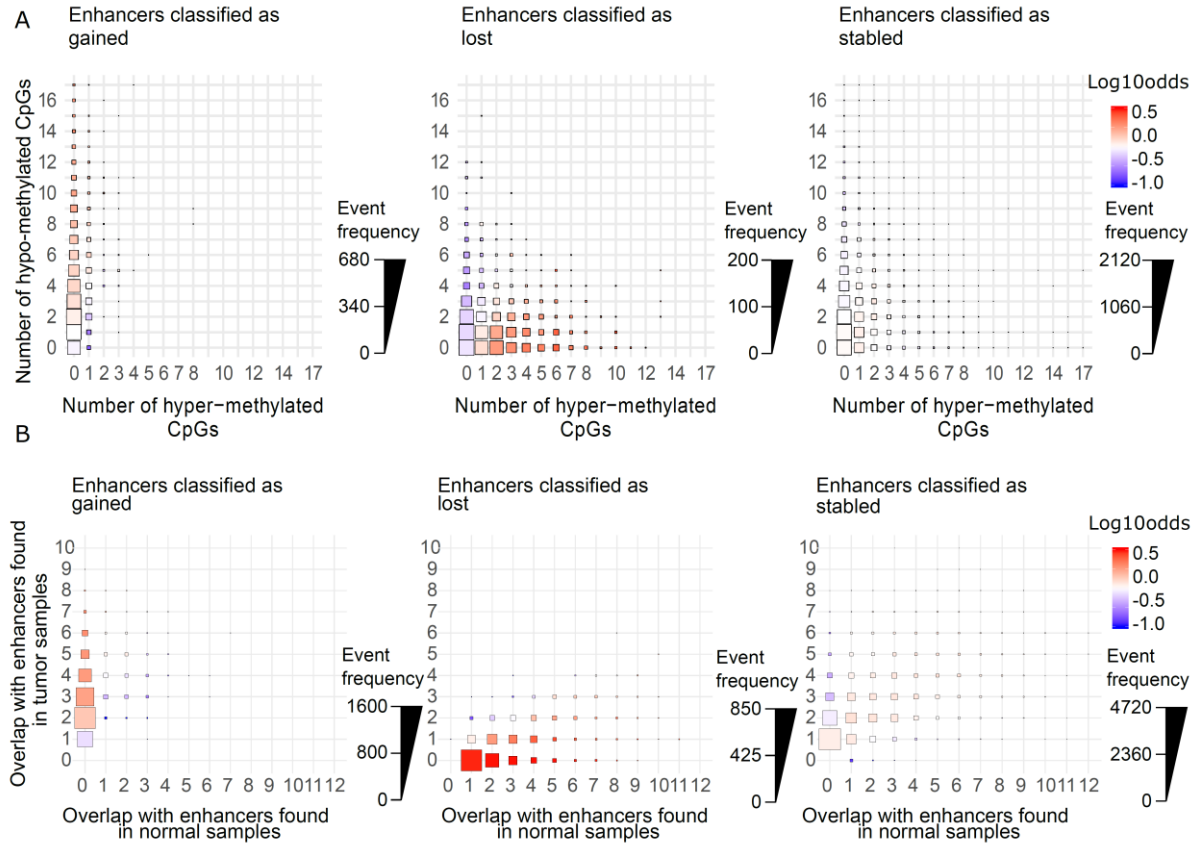


Figure 9: Investigating the inner working of machine learning classifier: **A.** Distribution of hypo- and hyper-methylated CpGs among enhancers that are classified as gained, lost, or stable by the random forest classifier. The size of each box denotes the number of enhancers (with the label that is indicated on top of the panel) that overlap the specified number of hypo- and hyper-methylated CpGs. The color gradient represents the logarithm of fold enrichment relative to what would be expected by chance (i.e. what would be expected from the null hypothesis that enhancers from all three classes have the same distribution of hypo- and hyper-methylated CpGs); **B.** Similar to **panel (A)**, but x- and y-axes correspond to the number of overlaps of each region with the enhancer elements identified from normal and tumor samples, respectively.

3.5 Gain and loss of active regulatory elements can predict the differential expression of associated genes

Changes in enhancer and promoter activity state (*i.e* gain or loss of active marks) alter the gene expression of nearby genes[244]. Therefore, we sought to investigate whether the status (gain or loss) of active enhancers and PREs that we identified in ccRCC is associated with ccRCC gene expression patterns. To identify ccRCC-associated gene expression patterns we preformed a differential gene expression analysis using RNA-seq data of 29 patient-matched primary ccRCC tumor and normal kidney sample pairs from the CAGEKID cohort[188]. Overall, we obtained 5685 and 5689 significantly up- and down-regulated protein coding genes, respectively (FDR < 0.05; refer to **Supplementary Figure S4A** for distribution of expression changes). We then assigned each enhancer and PRE to its closest protein-coding gene, and examined the relationship between status (gained or lost) of active regulatory elements in tumors and ccRCC-associated expression (up- or down-regulation) of their closet gene.

As shown in **Figure 10A,** up- and down-regulated genes in ccRCC are more likely to be associated with gained and lost active regulatory elements, respectively. This observation is valid for both enhancers and PREs, and points to the notion that alterations in the activity state of regulatory elements are associated with changes in expression of nearby genes, and may serve to predict their expression patterns.

To test this possibility, we employed a machine learning strategy to predict changes in the expression of genes from alterations in the activity status of their nearby regulatory

elements. To train the classifier, we used gain/loss scores, which we had calculated through integrating data from ChIP-seq and WGBS for each regulatory element, and the distances between the regulatory elements and their associated TSS as predictors. The target was the ccRCC-associated changes in expression of the gene in each given regulatory element-gene pairs (significant up-regulation, significant down-regulation, or no change, based on the results of differential gene expression). We used cross-validation to evaluate the performance of the classifier, following leave-one-chromosome-out strategy. ROC analyses indicates AUC values of 0.95 and 0.88 for predicting up-regulated genes based on enhancer and PRE activity, respectively, and AUC values of 0.966 and 0.9 (P-value < 2.59×10^{-73}) for predicting down-regulated genes based on enhancer and PRE activity, respectively (**Supplementary Figure S4B**), indicating high accuracy of our classifiers. These results suggest that a large fraction of gene expression changes in ccRCC can be explained by the gain or loss of active enhancers and/or PREs.

To examine whether our classifier has learned meaningful relationships between regulatory element activity and gene expression, we visualized the gain/loss scores of regulatory elements associated with genes that our model classifies as up-regulated, down-regulated, or no-change. Interestingly, genes that are labeled as up-regulated by our model show an enrichment of gained enhancers near their TSS. In contrast, we can observed an enrichment of lost enhancers near the TSS of genes that the classifier labels as down-regulated, and genes that are predicted to have no differential expression are more likely to be associated with stable enhancers, i.e. those that have neither a strong gain or loss score (**Figure 10B**). Overall, these results suggest that our machine-learning classifier has identified biological relevant and interpretable rules to identify up-regulated, down-regulated, or stable genes based on the activity and distance of nearby enhancers. Similar analyses on the PREs also revealed interpretable connections between PRE activity and differential gene expression (**Supplementary Figure S4C**).

3.6 Gained and lost enhancers harbor binding sites for specific regulatory factors

Enhancers harbor binding sites for TFs, and the interaction of these TFs with co-factors such as transcriptional co-activators, RNA polymerase, and other regulatory factors mediate the effect of enhancer on gene expression. Therefore, we examined the regulatory factors

(RFs) that are likely involved in causing and/or mediating the effect of gain or loss of enhancers on gene expression.

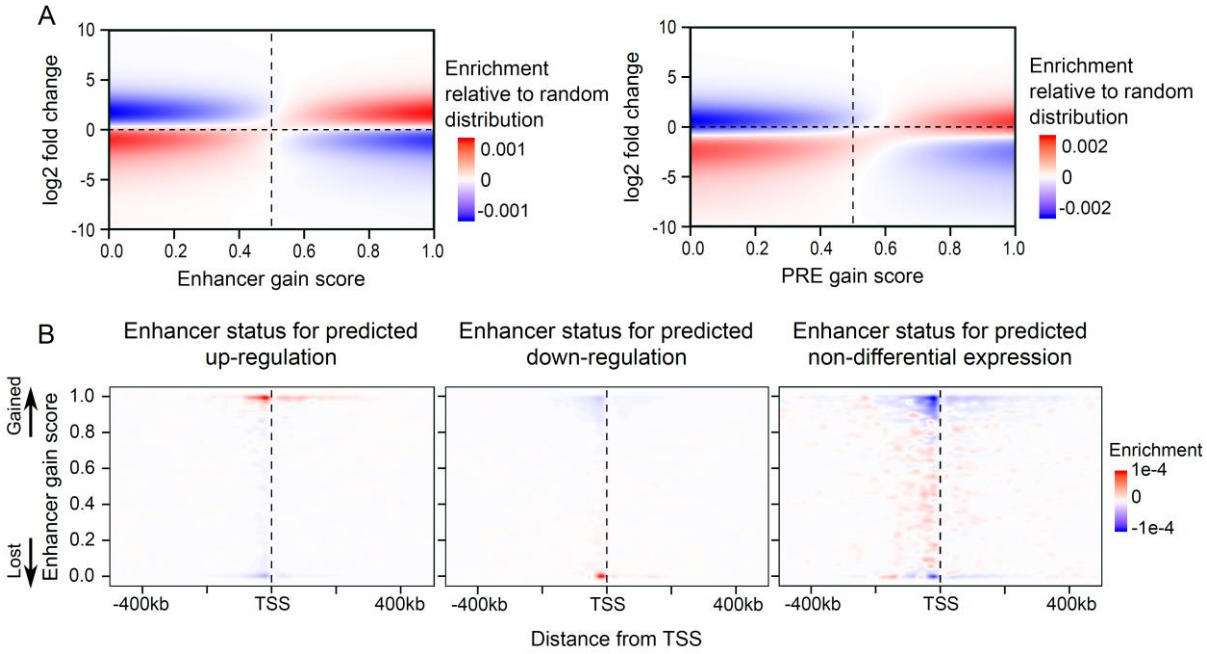


Figure 10: Relationship of gene expression changes and gain or loss of active regulatory elements: A. Visual representation of the relationship between differential gene expression genes and regulatory element activity in ccRCC. The x- and y- axis represent the regulatory element gain score and log2 fold-change of the expression of the associated gene (in tumor vs. normal), respectively. Note that a gain score of 0.5 depicts stable regulatory elements (gained regulatory elements have scores that range from 0.5 to 1 and lost regulatory elements have scores that range from 0.5 to 0). The color gradient shows the enrichment of genes at each region of the scatterplot relative to what would be expected if the x and y variables were independent of each other. Therefore, for example, the red color at the upper-right quartile of the left graph indicates that gained enhancers are more likely to be associated with up-regulated genes than would be expected by chance, and the blue color at the lower-right quartile means that gained enhancers are less likely to be associated with down-regulated genes; **B.** Visualization of the enhancer status for predicted up-, down-regulated and non-differentially expressed genes relative to the enhancer location from TSS. Enhancer gain score is shown on the y axis while distance from TSS is shown on the x axis. Red indicates enrichment and blue shows depletion. Therefore, for example, the red color at the top middle part of the left graph shows an enrichment of gained enhancers near the TSS of genes that are predicted to be up-regulated.

To perform this analysis, we obtained the coordinates of the binding sites of 161 known RFs from ENCODE[196] for 91 cell types, including human embryonic kidney cells (HEK293). We then examined the enrichment of these binding sites within 1kb region of the gained or lost enhancers. To ensure that the enrichment analysis is not confounded by unseen biases in RF binding site mapping or in identification of enhancers, we focused on RFs that are significantly enriched in our gained or lost enhancers relative to four different background sets; i) the union of gained and lost enhancers, ii) the union of gained and lost enhancers that have at least one binding site for at least one RF binding sites iii) the set of all the enhancers irrespective of their gain or loss score, and iv) all enhancers with at least one binding site for at least one RF (hypergeometric test; refer to method section for details of each test). Overall, we identified 32 RFs that were significant in all four enrichment analyses (FDR < 0.05; **Figure 11A**). Of these 32 RFs, 25 were enriched in gained enhancers and the remaining seven were enriched in lost enhancers. Consistent with these results, the target genes of all gain-associated RFs are significantly up-regulated in tumor relative to normal, and the targets of four out of seven loss-associated RFs are down-regulated (**Supplementary Figure S5**; GSEA pre-ranked test[236]; FDR < 0.05).

Next, we sought to investigate the biological and cellular processes that may be affected by target genes of these RFs. Using CPDB[245], we performed pathway enrichment analysis for up-regulated genes that are associated with at least one gained enhancer and that have a binding site for at least one gain-associated RF. Similarly, we performed this analysis for down-regulated targets of loss-associated RF that have a nearby lost enhancer. To test robustness of the results, we performed three enrichment analyses for each gene set using three background gene sets: i) background 1: default CPDB background, ii) background 2: all genes with significant differential expression between tumors and normal samples, iii) background 3: all genes associated with at least one enhancer. Then, we identified pathways that were significantly enriched in the examined gene set in all three analyses (FDR < 0.01; **Figure 11B** shows the significant pathways in all three tests). Interestingly, results revealed that HIF transcription factor network, which is frequently dysregulated in ccRCC[81], is a top up-regulated pathway associated with enhancer dysregulation. Among genes affected by HIF-enhancer reprogramming are *VEGFA, EDN1* and *SLC2A1*, which are well-established overexpressed genes in ccRCC, suggesting that enhancer activation plays a key role in mediating the effect of HIFs in inducing the expression of hypoxia-responsive genes. In addition, our results showed that pathways associated with immune system, including

interferon gamma signaling, cytokine signaling, and T-cell receptor signaling pathways, are also positively regulated by enhancer activation in ccRCC. Taken together, our results show that cooperation between enhancer remodeling and RFs whose binding sites coincide with enhancers may play a key role in ccRCC-associated gene expression patterns, and may underlie dysregulation of ccRCC pathways such as HIF signalling.

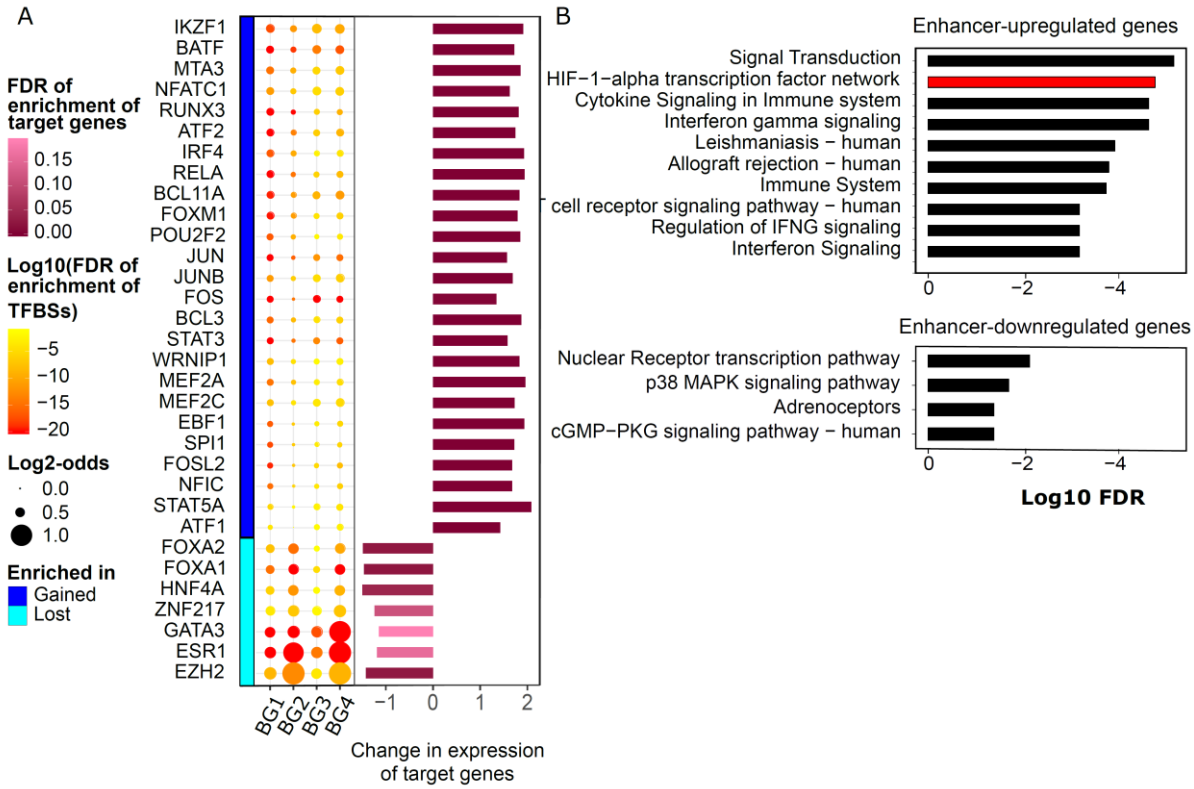


Figure 11: Regulatory factors whose binding sites are enriched in gained or lost enhancers: **A.** Enrichment of Regulatory factors (RFs) whose binding sites are enriched in gained/lost enhances and enrichment of their target genes in differentially expressed genes of tumor relative to normal samples. Only the RFs whose binding sites are significantly enriched in 1kb region of gained or lost enhancers relative to four different backgrounds are shown (background 1: all gained and lost enhancers, background 2: all gained and lost enhancers with at least one RF binding site, background 3: all enhancers, background 4: all enhancers with at least one RF binding site). Overall, 32 RFs are significant in all four enrichment analyses (hypergeometric test, FDR < 0.05). The circle size indicates the log2 odds of enrichment and the color of circles indicates the significance (log10 of FDR). The bar plot on the right shows the enrichment of the targets these RFs among up- or down-regulated genes (GSEA[236] pre-ranked test). The x-axis of the bar plot corresponds to the GSEA normalized enrichment score, and the color of each bar represents the statistical significance of the test

(FDR). BG; background; RFBS: RF binding site; **B.** Pathways that are enriched among up- and down-regulated targets of gain-associated or loss-associated RFs. Significantly enriched pathways are obtained from CPDB[245] by performing pathway enrichment analyses using up-regulated target genes assigned to gained enhancers and down-regulated target genes linked with lost enhancers as the test gene sets. Three background gene sets were used (background 1: default CPDB background, ii. background 2: all significant differentially expressed genes, iii. background 3: all the target genes associated with at least one enhancer), and only the significant pathways common in all the three analyses (FDR < 0.01) are shown in this figure (for the full list of significant pathways refer to **Supplementary Table S4**).

3.7 Altered enhancers and associated gene expression changes can be partially reversed by VHL

A recent study[186] has shown that introducing the wild type VHL to VHL-deficient ccRCC cell lines can trigger the reprogramming of DNA methylation patterns toward the normal kidney tissue state in a hypoxia-independent manner. This observation suggests a direct link between VHL deficiency and ccRCC epigenome patterns, which may also affect enhancer malfunction. Therefore, we sought to examine possible roles of VHL in enhancer reprogramming in ccRCC.

First, to investigate the effects of VHL on DNA methylation reprogramming, we used available data to analyze DNA methylation patterns in two ccRCC cancer cell-lines carrying VHL loss-of-function mutations (RCC4 and 786-O) and their derivatives in which wild-type VHL is stably re-expressed to restore the function of VHL. We obtained genome-wide DNA methylation profiles, which were generated using 450K DNA methylation microarray data (Infinium HumanMethylation450 BeadChip/ 450K MD) for both VHL reconstituted (VHL+) and VHL-deficient (VHL-) versions of the two cell lines, which were maintained in 21% O_2 (normoxic)[186]. The 450K arrays can probe the methylation state at 485,764 annotated cytosine positions across the genome[246]. We compared the genome-wide status of 5mC levels in VHL+ cells relative to VHL- cells in both RCC4 and 786-O cell lines by a differential methylation analysis. This analysis uncovered 109268 and 110191 significantly hyper-methylated ($\Delta\beta > 0$, FDR < 0.05) loci and 45310 and 81808 hypo-methylated ($\Delta\beta < 0$, FDR < 0.05) loci in RCC4 and 786-O, respectively (**Supplementary Figure S6A**). The results indicate that the extent of DNA hyper-methylation in VHL reconstituted cells, particularly in

the RCC4 cell line, is greater than the hypo-methylation, suggesting that reconstitution of VHL may promote DNA methylation in ccRCC cells.

To investigate the distribution of VHL-mediated DNA methylation changes across the genome, we analyzed the prevalence of the identified differentially methylated CpGs with respect to different genomic region annotations, focusing on the CpGs that behave consistently in both the RCC4 and 786-O cell lines in response to VHL reconstitution. We observed that the prevalence of hyper-methylated loci is higher than hypo-methylated loci in all genomic regions in both cell lines (**Supplementary Figure S6B**), suggesting that the effect of VHL on the DNA methylation pattern is global (genome-wide), and results predominantly in hyper-methylation in ccRCC cells.

We examined whether DNA methylation changes that are driven by VHL in ccRCC cell lines mirror the DNA methylation changes that are observed in patient tumors compared to normal tissues. We first looked at the overall correlation between VHL-driven DNA methylation changes in ccRCC cell lines and differential DNA methylation between tumor and normal. As shown in **Figure 12A**, there is an overall negative correlation, suggesting that loss of VHL may drive part of the DNA methylation differences between tumor and normal tissue. This negative correlation is most prominent when we focus on VHL-mediated methylation in RCC4 cells, suggesting that this cell line may more faithfully recapitulate the VHL-driven epigenetic changes in ccRCC. This can also be seen when we look at the significant differentially methylated CpGs in cell lines and their overlap with significant differentially methylated CpGs in ccRCC tumors (Fisher's Exact test; P-value < 2.2×10^{-16}; **Figure 12B**), which shows a particularly significant overlap between CpGs that are hyper-methylated after VHL reconstitution in RCC4 cells and those that are hypo-methylated in patient tumors compared to normal tissue. Therefore, we focus on the RCC4 cell line in the rest of our analyses.

To understand the relationship between these methylation changes with REs, we investigated the distribution of hyper- and hypo-methylated CpGs around TSS, as well as around DHS's that were obtained from kidney tissue. As shown in **Figure 12C**, there is a marked depletion of both hypo- and hyper-methylated CpGs at TSS's, both in RCC4 cells and in patient samples. In contrast, when we focus on DHS's that are far away from genes (which are likely distal regulatory elements such as enhancers), there is a sharp enrichment of CpGs that are hyper-methylated after VHL-reconstitution in RCC4 cells, and similarly a

sharp enrichment of hypo-methylated CpGs in tumors vs. normal tissues (**Figure 12D**). These observations suggest that in RCC4 cells VHL largely affects distal regulatory elements such as enhancers, in a manner that mirrors differential CpG methylation in tumors.

Interestingly, we found that the binding sites of certain regulatory factors are enriched near hypo- or hyper-methylated CpGs. Specifically, we performed a systematic analysis of 161 regulatory factors (using their binding sites from ENCODE[196]), in order to identify RFs whose binding sites are significantly enriched with hypo- or hyper-methylated CpGs, separately, relative to all differentially methylated CpGs in the methylation array. We performed this analysis separately for differentially methylated CpGs in RCC4 and 786-O (VHL+ vs. VHL-) as well as patient samples (tumor vs. normal). Overall, we identified 14 RFs that are associated with hypo-methylation after VHL reconstitution and hyper-methylated in tumors, compared to 13 RFs that are associated with hyper-methylation after VHL reconstitution and hypo-methylated in tumors (**Figure 12E**). Intriguingly, a considerable number of RFs that are enriched near tumor hypo-methylated CpGs (and VHL hyper-methylated CpGs) overlap those that we found to be enriched in tumor-gained enhancers (**Figure 11A**). Also, we found EZH2 to be associated with tumor hyper-methylated CpGs, VHL hypo-methylated CpGs, and tumor-gained enhancers. These results suggest that these RFs may be involved in mediating the effect of VHL-driven epigenetic changes in ccRCC.

Finally, we examined the extent to which VHL-driven epigenetic changes contribute to transcriptome remodelling in ccRCC. To do this, we performed RNA-sequencing on VHL+ and VHL- RCC4 cell lines. We examined each of these cell line versions in two oxygen conditions (hypoxia and normoxia) in order to understand hypoxia-dependent and hypoxia-independent effects of VHL on gene expression. After measuring the differential expression of each gene in VHL+ relative to VHL- cells in each growth condition, we focused on genes that are up-regulated in ccRCC tumors due to gain of enhancer activity (*i.e.* those identified in **section 3.6**). As **Figure 12F** shows, genes that are up-regulated in ccRCC due to enhancer gain are significantly enriched among genes that are down-regulated after VHL reconstitution in RCC4 cells. Interestingly, this pattern can be seen in both hypoxic and normoxic conditions, although the enrichment among down-regulated genes appears to be stronger in the normoxic condition, suggesting that there might be a mixture of hypoxia-dependent and hypoxia-independent effects. Despite this statistically significant enrichment, overall only 10% of the genes that are up-regulated in tumors due to gained enhancers are inhibited by

VHL, suggesting that enhancer-mediated up-regulation of genes can only be partially reversed by VHL reconstitution, and therefore a large majority of ccRCC epigenetic changes are likely caused by factors other than VHL.

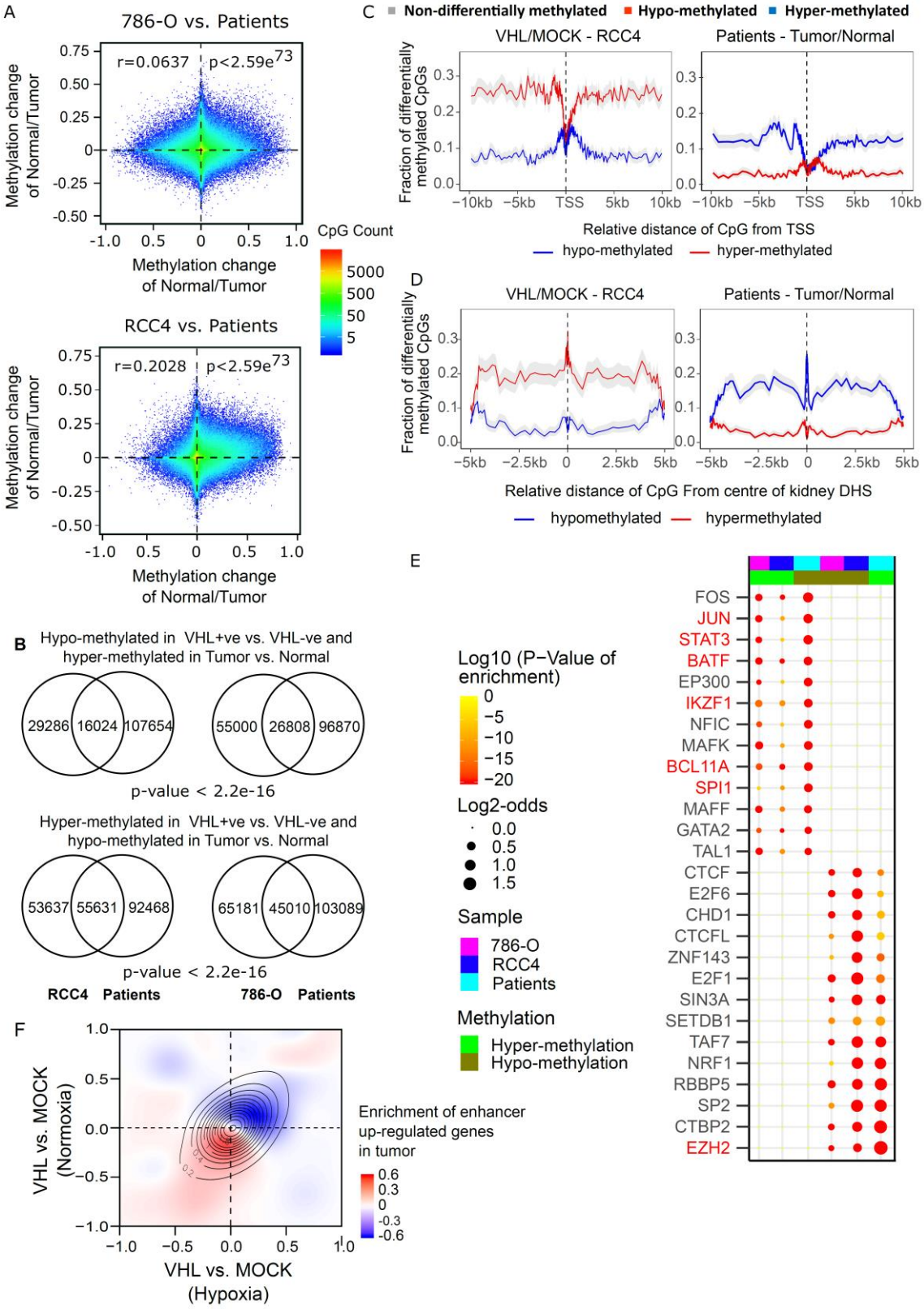


Figure 12: VHL reconstitution partially recovers the changes occurred in ccRCC: A. Correlation between differentially methylated CpGs in cell lines and patients. The scatter plot shows the density and distribution of Δβ values of all CpGs for 786-O (top) and RCC4 (bottom); **B.** Number of VHL specific, ccRCC specific and CpG loci consistently differentially methylated among cell lines and patients. Number of differentially methylated CpGs ($0.05 < FDR$; $|\Delta\beta| > 0$) only found in cell lines (RCC4 and 786-O) and patients CpGs differentially methylated in patients consistently methylated in reverse direction in cell lines are listed in Venn diagrams. P-value < 2.2e-16 (Fisher's exact test); **C.** Distribution of the fraction of differentially methylated loci around TSSs. Distribution of the ratio between number of hyper- hypo-methylated CpGs and all the CpGs at a position around TSSs for RCC4 and the same overlapping CpGs in patients (left to right). The CpGs and their respective distances from a TSS are grouped into bins and average number of CpGs and average distance per bin were calculated. Finally, the mean CpG ratio plotted against the mean distance separately for hyper- and hypo-methylated CpGs; **D.** Distribution of VHL-mediated differential CpG methylation with respect to DHSs (DHS data from fetus kidney tissue; DHS regions that are within 5kb of a gene body were removed to retain only distal REs). In each graph, the x-axis shows the distance relative to the center of DHS, and y-axis shows the fraction of CpGs at each given distance that are either significantly hyper-methylated ($\Delta\beta > 0$, $FDR < 0.05$, red) or hypo-methylated ($\Delta\beta < 0$, $FDR < 0.05$, blue). Analysis was done in RCC4 cell lines (VHL+ vs. VHL-) and patient samples (tumor vs. normal) separately, including only CpGs that were common in both datasets; **E.** Enrichment of regulatory factor binding sites (RFBSs) near differentially methylated CpGs in tumor/normal samples or VHL+/VHL- cell lines. The analysis was performed separately for the cell lines (RCC4 and 786-O) and patient samples. Only RFs with significant enrichment in all three analyses are included. The circle size indicates the log2-odds of enrichments, while the color gradient represents the logarithm of P-value (Fisher's exact test); **F.** The density plot shows the enrichment and depletion of genes that are up-regulated in ccRCC tumors ($FDR < 0.01$) due to enhancer gain (random forest classifier score > 0.95) with respects to differential gene expression in VHL+ vs. VHL- cells in two oxygen conditions. The contours represent the probability density function for all genes. Red indicates the enrichment and blue shows the depletion of enhancer-mediated up-regulated genes.

CHAPTER 4. DISCUSSION

In this study, we explored the landscape of epigenome aberrations in ccRCC, and investigated the relationship between alterations in gene expression and epigenetic fingerprints of *cis*-regulatory elements (regulatory elements), specifically distal enhancers and proximal regulatory elements (PREs). In addition, we examined the possible role of VHL, the major driver of the disease, in enhancer reprogramming and the associated transcriptome remodelling in ccRCC.

4.1 Discovery of gained and lost regulatory elements in ccRCC

Several studies showed that enrichment of specific combination of histone modifications, such that H3K4me1 and H3K4me3 occurring simultaneously with H3K27ac are associated with active enhancers and promoters, respectively[85,86,247]. In addition, the interplay between histone modifications and DNA methylation in human cancers has been reported in numerous studies.[248-250] Moreover, the reciprocal relationships between active regulatory elements and DNA methylation patterns have also been used to identify activity changes in regulatory elements in the genome[86]. Considering those, in this study we integrated information across histone modifications and DNA methylation (by utilizing two separate supervised machine learning models - random forest, for enhancers and PREs) as together, they provide more power for the identification of gain and loss events in ccRCC. We identified thousands of gained and lost regulatory elements and interestingly, among them, 71.2% of gain and lost events were novel and had not been previously reported in ccRCC, based on comparison to the only other study that we are aware of that has tried to comprehensively characterize regulatory elements and their epigenetic state in ccRCC[79]. Although, we did not perform experiments to identify optimum number of samples for better performance, we suggest that performance of active regulatory element identification is positively correlated with the number of samples. Capabilities of using this method in other tissues and cancer types merit further exploration.

4.2 Relationship between gain and loss of regulatory elements with gene expression changes in ccRCC

Several studies have shown that enhancers affect gene expression independently of their orientation (can be located either upstream or downstream of the gene TSS) and at

various distances from their target promoters[251,252]. The effective distance of enhancer-promoter interactions can be highly variable. For example, while most enhancers often act on the closest gene promoter, some enhancers can bypass neighbouring genes and regulate more distantly-located genes along the chromosome. Also, a single enhancer may regulate multiple genes[252,253]. Therefore, following a simple *ad hoc* procedure[254], we assigned each enhancer and PRE to its closest protein-coding gene and examined the relationship between cancer-associated changes in enhancer activity status (gained or lost) and differential expression (up- or down-regulation in tumors relative to normal tissue) of the closet gene. Using a supervised random-forest based machine learning strategy, we showed that the gain and loss of regulatory elements in ccRCC are predictive of the changes that happen in the expression patterns of nearby genes. Our results are consistent with previous studies which show enhancer activation positively correlated with gene up-regulation in several cancers including breast cancer[142] and adult T-cell leukemia[255].

4.3 Regulatory factors-enhancer complexes drive activation of ccRCC pathways

Enhancer activity depends on binding of regulatory factors, including transcription factors and co-activators to the enhancer region and their interactions with RNA polymerase. Out of the 161 regulatory factors examined in our study, binding sites of 25 and 7 were significantly enriched in gained and lost enhancers, respectively. Consistent with these results, the target genes of the RFs associated with gained enhancers were significantly up-regulated in tumors, whereas target genes of RFs associated with lost enhancers were down-regulated in tumors relative to normal samples, but only four of them reached statistical significance after multiple testing correction.

In our study, we also investigated the biological pathways and cellular processes that were affected by target genes of regulatory factors that we had identified. Notably, most of the target genes were involved in cellular processes that have previously been connected to ccRCC. For example, cellular responses to stress and immune system pathways were enriched in up-regulated genes associated with gained enhancers. Among up-regulated genes, we observed the HIF1α and HIF2α transcription factor network genes, which are associated with the hypoxia signalling pathway, the main driver pathway of ccRCC[256,16]. It is noteworthy that although the data set of regulatory factor binding sites that we analyzed did not include binding sites for HIF1α and 2α, both HIF1α and HIF2α transcription factor

networks were enriched in up-regulated target genes that were associated with TFs whose binding sites were enriched in gained enhancers in tumors. In other words, when we looked at the target genes of TFs that are enriched in gained enhancers (excluding HIFs), we still observed that these genes are significantly enriched for the HIF signalling pathway. This suggests that while HIFs directly regulate expression of several genes (*e.g.* VEGF) through binding to hypoxia responsive elements (HRE)[257], they may also contribute to the function of other TFs whose target genes are involved in cellular process that are affected by hypoxia signaling[258] (**Supplementary Table S5**). Abnormal activation of hypoxia signaling is involved in angiogenesis, glycolysis, cell proliferation, invasion and metastasis in ccRCC[19]. In addition, both innate and adaptive immune system related pathways, including interferon gamma signaling, T-cell receptor signaling pathway, Toll-like receptors cascades, and the AP1-transcription factor network were among those enriched in enhancer-associated up-regulated genes in ccRCC. Abnormal innate and adaptive immune responses are involved in oncogenesis by facilitating the selection of aggressive clones, and stimulating cancer cell proliferation and metastasis[259]. Further investigations on these specific pathways may prove to be of value for prognostic and diagnostic purposes in ccRCC. Overall, our study suggests that enhancer dysregulation plays a major role in over-expression of pathways that are integral to ccRCC biology. Interestingly, however, we did not observe any significant association between lost enhancers and genes that are enriched in main pathways that are inactivated in ccRCC. This may suggest that while upregulation of driver genes in ccRCC is mediated by specific TF-enhancer complexes, down-regulation of genes is caused by other mechanisms.

4.4 Changes of enhancer activity status affects expression of cancer-related genes in ccRCC

We also observed that expression of several well-established dysregulated genes in ccRCC is associated with abnormal changes in the activity of enhancers. For example, *MYC*, *VEGFA* and *EGFR* were significantly up-regulated in ccRCC tumors and are associated with gained enhancers. MYC is an oncogenic TF that is essential for promoting cell cycle progression, angiogenesis, cell growth and proliferation in ccRCC[260,261]. VEGFA, a hypoxia responsive gene, is responsible for inducing proliferation and migration of vascular endothelial cells and angiogenesis in ccRCC[262]. EGFR is a receptor tyrosine kinase that mediates numerous important aspects of cell biology that are related to ccRCC

tumorigenesis[263,264]. On the other hand, GATA3, a protein that is responsible for inhibiting adipocyte differentiation, is down regulated in ccRCC and is associated with a lost enhancer. This enhancer loss may at the GATA3 locus may lead to adipogenic trans-differentiation, which supports tumorigenesis and metastasis[265]. JAGGED1 (JAG1), a ligand in notch pathway, is commonly overexpressed and reported to be associated with loss of CpG methylation at H3K4me1-associated enhancer regions[174]. Interestingly *JAGGED1* was up regulated in our ccRCC tumors, potentially as a consequence of nearby gained enhancer. Although its pathological role in renal cell carcinoma is still unclear, patients with overexpression of *JAG1* have poor outcome[266]. These observations show that abnormal epigenetic patterns may explain aberrant expression of many genes that are connected to the biology or clinical outcome of ccRCC.

4.5 VHL-mediated DNA methylation reprogramming and its role in enhancer regulation in ccRCC

A recent study reported that von Hippel-Lindau (VHL), the most frequently mutated gene in ccRCC, is able to reprogram the DNA methylome in ccRCC[186]. Therefore, we studied the potential involvement of VHL-mediated DNA methylation changes in remodeling the regulatory landscape of ccRCC.

In our study, we observed a global DNA hyper-methylation after VHL re-constitution in two VHL-deficient ccRCC cell lines (RCC4 and 786-O), with a pattern that was significantly negatively correlated with DNA methylations levels in patient samples harbouring loss-of-function VHL mutations. This suggests that VHL deficiency may at least partially drive the DNA methylation differences between tumor and normal tissues. We note that a recently published study by Artemov *et al*.[267] reported opposite results to us by showing global hypo-methylation upon inactivating VHL by CRISPR/Cas9 in Caki-1 cell line. However, a few pitfalls of their study can be summarized as follows: Caki-1 cell line is a VHL-positive (expressing wild-type VHL) cell line, which had been isolated from a skin metastatic site[268], whereas we selected two appropriate models of ccRCC which represent the canonical malfunction of the driver pathway of ccRCC, which is the inactivation of VHL (primary ccRCC cell lines RCC4 and 786-O) and compared our results with actual primary tumor tissues (RCC4 and 786-O cell lines are VHL-deficient and resemble actual tumors). Although Caki-1 cells express wildtype VHL they may have other modifications downstream of VHL that still renders VHL non-functional, and therefore VHL CRISPR in Caki-1 may not

recapitulate the functional consequences of VHL deficiency in kidney epithelial cells (because VHL is already not doing what it is supposed to do due to downstream alterations). In addition, they have performed functional genomics experiments only on one cell lines (Caki-1) but our functional work (VHL-reconstitution) were based on two cell lines, supporting reproducible observations. Therefore, their results may also be confounded by cell line-specific patterns.

In addition, VHL+ RCC4 cells showed an elevated level of methylation in distal DHS regions of genes compared to that in proximal areas of TSS. These observations suggest that in RCC4 cells, VHL largely affects distal regulatory elements such as enhancers, in a manner that mirrors differential CpG methylation in tumors. Therefore, focusing on enhancers, we uncovered several regulatory factors associated with dysregulated enhancers that there affected by VHL-driven DNA methylation changes. Among them, BATF, JUN, SPI1 and STAT3 were enriched in gained enhancers, and EZH2 was associated with lost enhancers.

Considering the biological functions of these regulatory factors, BATF can form a heterodimer with JUN proteins to bind to AP1 transcription factor motifs, and is critical for T helper type 17 differentiation, growth and survival in anaplastic large cell lymphoma[269]. STAT3 is an oncogenic TF controlling inflammation, cell proliferation, survival, and differentiation in normal tissue as well as in ccRCC tumor growth[270]. EZH2 is a histone methyl transferase enzyme that catalyzes H3K27 tri-methylation, a suppressive histone mark. It is a subunit and one of the catalytic components of polycomb repressive complex 2 (PRC2), which contributes to polynucleosome compaction and leads to transcriptional repression by reducing the access of both TFs and chromatin remodelers such as SWI/SNF to DNA[123]. In addition, it directly controls the DNA methylation[271]. Target genes of EZH2 are involved in various biological functions including cell cycle, cell proliferation and cell differentiation[272]. In ccRCC, over-expression of EZH2 is associated with metastasis and worse clinical outcome[163]. Overall, our results suggest that these regulatory factors may play a crucial role in reprogramming major biological and cellular pathways associated with cancer, warranting future studies for functional validation of these candidate factors.

Finally, we examined the extent of ccRCC-associated transcriptional changes that can be explained by VHL-driven epigenetic changes in ccRCC. Our results showed that VHL might be responsible for a mixture of hypoxia-dependent and hypoxia-independent effects on gene up-regulation due to gain of enhancer activity. Overall, our results suggest that only

10% of enhancer-upregulated genes in ccRCC can be inhibited by VHL reconstitution. This suggests that a major fraction of epigenetic changes that drive enhancer-mediated gene upregulation in ccRCC is not mediated directly by VHL. On the other hand, given that our results uncover a global effect on DNA methylation by VHL, further studies are necessary to examine the function consequences of these alterations beyond enhancer regulation.

CHAPTER 5. CONCLUSIONS AND FUTURE DIRECTIONS

Here, we characterized the epigenome changes of *cis*-regulatory elements, specifically the enhancers and proximal regulatory elements in ccRCC, and investigated their role in gene expression changes in ccRCC. We identified thousands of differentially activated enhancers and proximal elements in ccRCC and found that most of them are associated with alterations in the expression of their target genes. We also found a set of regulatory factors whose binding sites are enriched in gained and lost *cis*-regulatory elements, suggesting that they may modulate the expression of the genes associated with these regulatory elements. Finally, we investigated the potential involvement of VHL in these epigenetic alterations. Our analysis revealed that only ~10% of gene expression changes in ccRCC can be reversed by VHL-driven epigenome alterations on enhancers, and that these VHL-driven changes are mixture of hypoxia-dependent and independent events. Moreover, we discovered several potential regulatory factors whose functions are regulated by VHL-mediated DNA methylation and that are associated with differentially activated enhancers in tumor relative to normal. These factors are possible candidates for functional validations for future studies. Overall, our study provides a better understanding of the molecular mechanisms that underlie the ccRCC initiation, progression and/or metastatic.

CHAPTER 6. REFERENCES

1. Russo, V.E., Martienssen, R.A. & Riggs, A.D. *Epigenetic mechanisms of gene regulation*, (Cold Spring Harbor Laboratory Press, 1996).
2. Bird, A. DNA methylation patterns and epigenetic memory. *Genes & Development* **16**, 6-21 (2002).
3. Handy, D.E., Castro, R. & Loscalzo, J. Epigenetic modifications: basic mechanisms and role in cardiovascular disease. *Circulation* **123**, 2145-2156 (2011).
4. Felsenfeld, G. & Groudine, M. Controlling the double helix. *Nature* **421**, 448 (2003).
5. Sharma, S., Kelly, T.K. & Jones, P.A. Epigenetics in cancer. *Carcinogenesis* **31**, 27-36 (2010).
6. Hanahan, D. & Weinberg, R.A. Hallmarks of cancer: the next generation. *Cell* **144**, 646-74 (2011).
7. Swygert, S.G. & Peterson, C.L. Chromatin dynamics: interplay between remodeling enzymes and histone modifications. *Biochimica et biophysica acta* **1839**, 728-736 (2014).
8. Zhou, S., Treloar, A.E. & Lupien, M. Emergence of the Noncoding Cancer Genome: A Target of Genetic and Epigenetic Alterations. *Cancer discovery* **6**, 1215-1229 (2016).
9. Kellis, M. *et al.* Defining functional DNA elements in the human genome. *Proc Natl Acad Sci U S A* **111**, 6131-8 (2014).
10. Graur, D. *et al.* On the immortality of television sets: "function" in the human genome according to the evolution-free gospel of ENCODE. *Genome Biol Evol* **5**, 578-90 (2013).
11. Wittkopp, P.J. & Kalay, G. Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. *Nat Rev Genet* **13**, 59-69 (2011).
12. Kagohara, L.T. *et al.* Epigenetic regulation of gene expression in cancer: techniques, resources and analysis. *Briefings in functional genomics* **17**, 49-63 (2017).
13. Patel, S.A. & Vanharanta, S. Epigenetic determinants of metastasis. *Molecular oncology* **11**, 79-96 (2017).
14. Chik, F., Szyf, M. & Rabbani, S.A. Role of Epigenetics in Cancer Initiation and Progression. in *Human Cell Transformation: Role of Stem Cells and the Microenvironment* (eds. Rhim, J.S. & Kremer, R.) 91-104 (Springer New York, New York, NY, 2012).

15. Lawrence, M.S. *et al.* Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* **505**, 495-501 (2014).
16. Mehdi, A. & Riazalhosseini, Y. Epigenome Aberrations: Emerging Driving Factors of the Clear Cell Renal Cell Carcinoma. *International Journal of Molecular Sciences* **18**(2017).
17. Ferlay, J. *et al.* Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. *International Journal of Cancer* **136**, E359-E386 (2015).
18. Morris, M.R. & Latif, F. The epigenetic landscape of renal cancer. *Nat Rev Nephrol* **13**, 47-60 (2017).
19. Siegel, R.L., Miller, K.D. & Jemal, A. Cancer statistics, 2018. *CA: A Cancer Journal for Clinicians* **68**, 7-30 (2018).
20. Becket, E. *et al.* Identification of DNA Methylation-Independent Epigenetic Events Underlying Clear Cell Renal Cell Carcinoma. *Cancer research* **76**, 1954-1964 (2016).
21. Roadmap Epigenomics, C. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317-330 (2015).
22. Sandoval, J. & Esteller, M. Cancer epigenomics: beyond genomics. *Current Opinion in Genetics & Development* **22**, 50-55 (2012).
23. Henikoff, S. & Smith, M.M. Histone variants and epigenetics. *Cold Spring Harbor perspectives in biology* **7**, a019364-a019364.
24. Campos, E.I. & Reinberg, D. Histones: Annotating Chromatin. *Annual Review of Genetics* **43**, 559-599 (2009).
25. Fedorova, E. & Zink, D. Nuclear architecture and gene regulation. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1783**, 2174-2184 (2008).
26. Bannister, A.J. & Kouzarides, T. Regulation of chromatin by histone modifications. *Cell research* **21**, 381-395 (2011).
27. Audia, J.E. & Campbell, R.M. Histone Modifications and Cancer. *Cold Spring Harbor perspectives in biology* **8**, a019521-a019521.
28. Fyodorov, D.V., Zhou, B.-R., Skoultchi, A.I. & Bai, Y. Emerging roles of linker histones in regulating chromatin structure and function. *Nature Reviews Molecular Cell Biology* **19**, 192 (2017).
29. Xhemalce, B., Dawson, M. A. and Bannister, A. J. Histone Modifications. in *Reviews in Cell Biology and Molecular Medicine* (2011).

30. Allfrey, V.G., Faulkner, R. & Mirsky, A.E. ACETYLATION AND METHYLATION OF HISTONES AND THEIR POSSIBLE ROLE IN THE REGULATION OF RNA SYNTHESIS. *Proceedings of the National Academy of Sciences of the United States of America* **51**, 786-794 (1964).
31. Kleer, C.G. *et al.* EZH2 is a marker of aggressive breast cancer and promotes neoplastic transformation of breast epithelial cells. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 11606-11611 (2003).
32. Sato, A., Asano, T., Ito, K., Sumitomo, M. & Asano, T. Suberoylanilide hydroxamic acid (SAHA) combined with bortezomib inhibits renal cancer growth by enhancing histone acetylation and protein ubiquitination synergistically. *BJU International* **109**, 1258-1268 (2012).
33. Ramakrishnan, S., Ellis, L. & Pili, R. Histone modifications: implications in renal cell carcinoma. *Epigenomics* **5**, 453-462 (2013).
34. Xing, T. & He, H. Epigenomics of clear cell renal cell carcinoma: mechanisms and potential use in molecular pathology. *Chinese journal of cancer research = Chung-kuo yen cheng yen chiu* **28**, 80-91 (2016).
35. Meddens, C.A., van der List, A.C.J., Nieuwenhuis, E.E.S. & Mokry, M. Non-coding DNA in IBD: from sequence variation in DNA regulatory elements to novel therapeutic potential. *Gut* **68**, 928 (2019).
36. Gray, S.G. Chapter 17 - The Potential of Epigenetic Compounds in Treating Diabetes. in *Epigenetics in Human Disease* (ed. Tollefsbol, T.O.) 331-367 (Academic Press, San Diego, 2012).
37. Lee, J.-S., Smith, E. & Shilatifard, A. The language of histone crosstalk. *Cell* **142**, 682-685 (2010).
38. Jenuwein, T. & Allis, C.D. Translating the Histone Code. *Science* **293**, 1074 (2001).
39. Ernst, J. & Kellis, M. Discovery and characterization of chromatin states for systematic annotation of the human genome. *Nature biotechnology* **28**, 817-825 (2010).
40. Becker, P.B. & Hörz, W. ATP-Dependent Nucleosome Remodeling. *Annual Review of Biochemistry* **71**, 247-273 (2002).
41. Narlikar, G.J., Fan, H.-Y. & Kingston, R.E. Cooperation between Complexes that Regulate Chromatin Structure and Transcription. *Cell* **108**, 475-487 (2002).

42. Rippe, K. *et al.* DNA sequence- and conformation-directed positioning of nucleosomes by chromatin-remodeling complexes. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 15635-15640 (2007).
43. Tyagi, M., Imam, N., Verma, K. & Patel, A.K. Chromatin remodelers: We are the drivers!! *Nucleus (Austin, Tex.)* **7**, 388-404 (2016).
44. Clapier, C.R., Iwasa, J., Cairns, B.R. & Peterson, C.L. Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes. *Nature Reviews Molecular Cell Biology* **18**, 407 (2017).
45. Christman, J.K. 5-Azacytidine and 5-aza-2′-deoxycytidine as inhibitors of DNA methylation: mechanistic studies and their implications for cancer therapy. *Oncogene* **21**, 5483 (2002).
46. Jang, H.S., Shin, W.J., Lee, J.E. & Do, J.T. CpG and Non-CpG Methylation in Epigenetic Gene Regulation and Brain Function. *Genes* **8**, 148 (2017).
47. Reik, W. Stability and flexibility of epigenetic gene regulation in mammalian development. *Nature* **447**, 425 (2007).
48. Law, J.A. & Jacobsen, S.E. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nature reviews. Genetics* **11**, 204-220 (2010).
49. Deaton, A.M. & Bird, A. CpG islands and the regulation of transcription. *Genes & development* **25**, 1010-1022 (2011).
50. Lister, R. *et al.* Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462**, 315-322 (2009).
51. Chodavarapu, R.K. *et al.* Relationship between nucleosome positioning and DNA methylation. *Nature* **466**, 388-392 (2010).
52. Ioshikhes, I.P. & Zhang, M.Q. Large-scale human promoter mapping using CpG islands. *Nature Genetics* **26**, 61 (2000).
53. Antequera, F. & Bird, A. Number of CpG islands and genes in human and mouse. *Proceedings of the National Academy of Sciences of the United States of America* **90**, 11995-11999 (1993).
54. Laurent, L. *et al.* Dynamic changes in the human methylome during differentiation. *Genome research* **20**, 320-331 (2010).
55. Feng, J. & Fan, G. Chapter 4 - The Role of DNA Methylation in the Central Nervous System and Neuropsychiatric Disorders. in *International Review of Neurobiology*, Vol. 89 67-84 (Academic Press, 2009).

56. Jin, B., Li, Y. & Robertson, K.D. DNA methylation: superior or subordinate in the epigenetic hierarchy? *Genes & cancer* **2**, 607-617 (2011).
57. Cedar, H. & Bergman, Y. Linking DNA methylation and histone modification: patterns and paradigms. *Nat Rev Genet* **10**, 295-304 (2009).
58. Ikegami, K., Ohgane, J., Tanaka, S., Yagi, S. & Shiota, K. Interplay between DNA methylation, histone modification and chromatin remodeling in stem cells and during development. *Int J Dev Biol* **53**, 203-14 (2009).
59. Hayashihara, K. *et al.* The middle region of an HP1-binding protein, HP1-BP74, associates with linker DNA at the entry/exit site of nucleosomal DNA. *J Biol Chem* **285**, 6498-507 (2010).
60. Cheutin, T. *et al.* Maintenance of stable heterochromatin domains by dynamic HP1 binding. *Science* **299**, 721-5 (2003).
61. Jaenisch, R. & Bird, A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nature Genetics* **33**, 245 (2003).
62. Watt, F. & Molloy, P.L. Cytosine methylation prevents binding to DNA of a HeLa cell transcription factor required for optimal expression of the adenovirus major late promoter. *Genes Dev* **2**, 1136-43 (1988).
63. Ahsendorf, T., Müller, F.-J., Topkar, V., Gunawardena, J. & Eils, R. Transcription factors, coregulators, and epigenetic marks are linearly correlated and highly redundant. *PLOS ONE* **12**, e0186324 (2017).
64. Billard, L.M., Magdinier, F., Lenoir, G.M., Frappart, L. & Dante, R. MeCP2 and MBD2 expression during normal and pathological growth of the human mammary gland. *Oncogene* **21**, 2704-12 (2002).
65. Boyes, J. & Bird, A. DNA methylation inhibits transcription indirectly via a methyl-CpG binding protein. *Cell* **64**, 1123-34 (1991).
66. Nan, X. *et al.* Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature* **393**, 386-389 (1998).
67. Huang, Q. *et al.* Mechanistic Insights Into the Interaction Between Transcription Factors and Epigenetic Modifications and the Contribution to the Development of Obesity. *Front Endocrinol (Lausanne)* **9**, 370 (2018).
68. Hervouet, E., Vallette, F.M. & Cartron, P.F. Dnmt1/Transcription factor interactions: an alternative mechanism of DNA methylation inheritance. *Genes Cancer* **1**, 434-43 (2010).

69. Xin, B. & Rohs, R. Relationship between histone modifications and transcription factor binding is protein family specific. *Genome Research* **28**, 321-333 (2018).
70. Lambert, S.A. *et al.* The Human Transcription Factors. *Cell* **172**, 650-665 (2018).
71. Benveniste, D., Sonntag, H.-J., Sanguinetti, G. & Sproul, D. Transcription factor binding predicts histone modifications in human cell lines. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 13367-13372 (2014).
72. Sahar Olsadat, S. & Andreas, K.N. DNA Methylation: A Possible Target for Current and Future Studies on Cancer? *Epigenetic Diagnosis & Therapy (Discontinued)* **1**, 5-13 (2015).
73. Andersson, R. Promoter or enhancer, what's the difference? Deconstruction of established distinctions and presentation of a unifying model. *Bioessays* **37**, 314-23 (2015).
74. Shandilya, J. & Roberts, S.G. The transcription cycle in eukaryotes: from productive initiation to RNA polymerase II recycling. *Biochim Biophys Acta* **1819**, 391-400 (2012).
75. Makalowski, W., Pande, A., Brosius, J., Raabe, C.A. & Makalowska, I. Transcriptional interference by small transcripts in proximal promoter regions. *Nucleic Acids Research* **46**, 1069-1088 (2018).
76. Kutach, A.K. & Kadonaga, J.T. The downstream promoter element DPE appears to be as widely used as the TATA box in Drosophila core promoters. *Molecular and cellular biology* **20**, 4754-4764 (2000).
77. Juven-Gershon, T., Hsu, J.-Y., Theisen, J.W. & Kadonaga, J.T. The RNA polymerase II core promoter - the gateway to transcription. *Current opinion in cell biology* **20**, 253-259 (2008).
78. Hasegawa, M. *et al.* Regulation of the Human FcεRI α-Chain Distal Promoter. *The Journal of Immunology* **170**, 3732-3738 (2003).
79. Yao, X. *et al.* VHL Deficiency Drives Enhancer Activation of Oncogenes in Clear Cell Renal Cell Carcinoma. *Cancer Discovery* **7**, 1284 (2017).
80. Herz, H.-M., Hu, D. & Shilatifard, A. Enhancer malfunction in cancer. *Molecular cell* **53**, 859-866 (2014).
81. Kolovos, P., Knoch, T.A., Grosveld, F.G., Cook, P.R. & Papantonis, A. Enhancers and silencers: an integrated and simple model for their function. *Epigenetics & chromatin* **5**, 1-1 (2012).

82. Bulger, M. & Groudine, M. Functional and mechanistic diversity of distal transcription enhancers. *Cell* **144**, 327-39 (2011).
83. Palstra, R.-J. & Grosveld, F. Transcription factor binding at enhancers: shaping a genomic regulatory landscape in flux. *Frontiers in genetics* **3**, 195-195 (2012).
84. Aran, D., Sabato, S. & Hellman, A. DNA methylation of distal regulatory sites characterizes dysregulation of cancer genes. *Genome Biol* **14**, R21 (2013).
85. Creyghton, M.P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proceedings of the National Academy of Sciences* **107**, 21931-21936 (2010).
86. Sharifi-Zarchi, A. *et al.* DNA methylation regulates discrimination of enhancers from promoters through a H3K4me1-H3K4me3 seesaw mechanism. *BMC genomics* **18**, 964-964 (2017).
87. Simonis, M., Kooren, J. & de Laat, W. An evaluation of 3C-based methods to capture DNA interactions. *Nat Methods* **4**, 895-901 (2007).
88. Cook, P.R. A model for all genomes: the role of transcription factories. *J Mol Biol* **395**, 1-10 (2010).
89. Ptashne, M. & Gann, A. Transcriptional activation by recruitment. *Nature* **386**, 569-77 (1997).
90. Dorsett, D. Distant liaisons: long-range enhancer-promoter interactions in Drosophila. *Curr Opin Genet Dev* **9**, 505-14 (1999).
91. Li, M. *et al.* Dynamic regulation of transcription factors by nucleosome remodeling. *eLife* **4**, e06249 (2015).
92. Transcription Factors. in *Reference Module in Biomedical Sciences* (Elsevier, 2014).
93. Alberini, C.M. Transcription factors in long-term memory and synaptic plasticity. *Physiol Rev* **89**, 121-45 (2009).
94. Frietze, S. & Farnham, P.J. Transcription factor effector domains. *Sub-cellular biochemistry* **52**, 261-277 (2011).
95. Yang, V.W. Eukaryotic Transcription Factors: Identification, Characterization and Functions. *The Journal of Nutrition* **128**, 2045-2051 (1998).
96. Phillips, T. Regulation of Transcription and Gene Expression in Eukaryotes. *Nature Education* **1(1):199**(2008).
97. Carlberg, C. & Molnár, F. The Basal Transcriptional Machinery. in *Mechanisms of Gene Regulation* (eds. Carlberg, C. & Molnár, F.) 37-54 (Springer Netherlands, Dordrecht, 2014).

98. Orphanides, G., Lagrange, T. & Reinberg, D. The general transcription factors of RNA polymerase II. *Genes Dev* **10**, 2657-83 (1996).
99. Ying, Y. *et al.* The Krüppel-associated box repressor domain induces reversible and irreversible regulation of endogenous mouse genes by mediating different chromatin states. *Nucleic acids research* **43**, 1549-1561 (2015).
100. Dechassa, M.L. *et al.* SWI/SNF has intrinsic nucleosome disassembly activity that is dependent on adjacent nucleosomes. *Molecular cell* **38**, 590-602 (2010).
101. Xu, Y.Z., Thuraisingam, T., Marino, R. & Radzioch, D. Recruitment of SWI/SNF Complex Is Required for Transcriptional Activation of the SLC11A1 Gene during Macrophage Differentiation of HL-60 Cells. *Journal of Biological Chemistry* **286**, 12839-12849 (2011).
102. Jones, P.A. & Baylin, S.B. The epigenomics of cancer. *Cell* **128**, 683-692 (2007).
103. Stratton, M.R., Campbell, P.J. & Futreal, P.A. The cancer genome. *Nature* **458**, 719-724 (2009).
104. Riggs, A.D. & Jones, P.A. 5-Methylcytosine, Gene Regulation, and Cancer. in *Advances in Cancer Research*, Vol. 40 (eds. Klein, G. & Weinhouse, S.) 1-30 (Academic Press, 1983).
105. Yang, L. *et al.* DNMT3A Loss Drives Enhancer Hypomethylation in FLT3-ITD-Associated Leukemias. *Cancer Cell* **29**, 922-934 (2016).
106. Mendizabal, I., Zeng, J., Keller, T.E. & Yi, S.V. Body-hypomethylated human genes harbor extensive intragenic transcriptional activity and are prone to cancer-associated dysregulation. *Nucleic acids research* **45**, 4390-4400 (2017).
107. Rodriguez, J. *et al.* Chromosomal Instability Correlates with Genome-wide DNA Demethylation in Human Primary Colorectal Cancers. *Cancer Research* **66**, 8462 (2006).
108. Stefanska, B. *et al.* Definition of the landscape of promoter DNA hypomethylation in liver cancer. *Cancer Res* **71**, 5891-903 (2011).
109. Soes, S. *et al.* Hypomethylation and increased expression of the putative oncogene ELMO3 are associated with lung cancer development and metastases formation. *Oncoscience* **1**, 367-74 (2014).
110. Xiong, L. *et al.* Aberrant enhancer hypomethylation contributes to hepatic carcinogenesis through global transcriptional reprogramming. *Nat Commun* **10**, 335 (2019).

111. Qu, Y. *et al.* Cancer-specific changes in DNA methylation reveal aberrant silencing and activation of enhancers in leukemia. *Blood* **129**, e13 (2017).
112. Rauch, T.A., Wu, X., Zhong, X., Riggs, A.D. & Pfeifer, G.P. A human B cell methylome at 100-base pair resolution. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 671-678 (2009).
113. Pfeifer, G.P. Defining Driver DNA Methylation Changes in Human Cancer. *International journal of molecular sciences* **19**, 1166 (2018).
114. Baylin, S.B. & Jones, P.A. Epigenetic Determinants of Cancer. *Cold Spring Harbor perspectives in biology* **8**, a019505.
115. Baylin, S.B. *et al.* DNA Methylation Patterns of the Calcitonin Gene in Human Lung Cancers and Lymphomas. *Cancer Research* **46**, 2917 (1986).
116. Esteller, M. CpG island hypermethylation and tumor suppressor genes: a booming present, a brighter future. *Oncogene* **21**, 5427 (2002).
117. Weisenberger, D.J. *et al.* CpG island methylator phenotype underlies sporadic microsatellite instability and is tightly associated with BRAF mutation in colorectal cancer. *Nature Genetics* **38**, 787 (2006).
118. Malta, T.M. *et al.* Glioma CpG island methylator phenotype (G-CIMP): biological and clinical implications. *Neuro-Oncology* **20**, 608-620 (2017).
119. Liang, G. & Weisenberger, D.J. DNA methylation aberrancies as a guide for surveillance and treatment of human cancers. *Epigenetics* **12**, 416-432 (2017).
120. Schneider, R. & Di Cerbo, V. Cancers with wrong HATs: the impact of acetylation. *Briefings in Functional Genomics* **12**, 231-243 (2013).
121. Pasqualucci, L. *et al.* Inactivating mutations of acetyltransferase genes in B-cell lymphoma. *Nature* **471**, 189-195 (2011).
122. Smith, E., Lin, C. & Shilatifard, A. The super elongation complex (SEC) and MLL in development and disease. *Genes & Development* **25**, 661-672 (2011).
123. Veneti, Z., Gkouskou, K.K. & Eliopoulos, A.G. Polycomb Repressor Complex 2 in Genomic Instability and Cancer. *International journal of molecular sciences* **18**, 1657 (2017).
124. Wang, Y. *et al.* Ezh2 Acts as a Tumor Suppressor in Kras-driven Lung Adenocarcinoma. *Int J Biol Sci* **13**, 652-659 (2017).
125. Bracken, A.P. *et al.* EZH2 is downstream of the pRB-E2F pathway, essential for proliferation and amplified in cancer. *Embo j* **22**, 5323-35 (2003).

126. Harris, W.J. *et al.* The histone demethylase KDM1A sustains the oncogenic potential of MLL-AF9 leukemia stem cells. *Cancer Cell* **21**, 473-87 (2012).
127. Lim, S. *et al.* Lysine-specific demethylase 1 (LSD1) is highly expressed in ER-negative breast cancers and a biomarker predicting aggressive biology. *Carcinogenesis* **31**, 512-20 (2010).
128. Schulte, J.H. *et al.* Lysine-specific demethylase 1 is strongly expressed in poorly differentiated neuroblastoma: implications for therapy. *Cancer Res* **69**, 2065-71 (2009).
129. Song, J.S., Kim, Y.S., Kim, D.K., Park, S.I. & Jang, S.J. Global histone modification pattern associated with recurrence and disease-free survival in non-small cell lung cancer patients. *Pathol Int* **62**, 182-90 (2012).
130. Seligson, D.B. *et al.* Global levels of histone modifications predict prognosis in different cancers. *Am J Pathol* **174**, 1619-28 (2009).
131. Ellinger, J. *et al.* Global levels of histone modifications predict prostate cancer recurrence. *Prostate* **70**, 61-9 (2010).
132. Chervona, Y. & Costa, M. Histone modifications and cancer: biomarkers of prognosis? *American journal of cancer research* **2**, 589-597 (2012).
133. Akhtar-Zaidi, B. *et al.* Epigenomic enhancer profiling defines a signature of colon cancer. *Science* **336**, 736-9 (2012).
134. Cohen, A.J. *et al.* Hotspots of aberrant enhancer activity punctuate the colorectal cancer epigenome. *Nature communications* **8**, 14400-14400 (2017).
135. Mansour, M.R. *et al.* Oncogene regulation. An oncogenic super-enhancer formed through somatic mutation of a noncoding intergenic element. *Science* **346**, 1373-7 (2014).
136. Deligezer, U., Akisik, E.E., Erten, N. & Dalay, N. Sequence-Specific Histone Methylation Is Detectable on Circulating Nucleosomes in Plasma. *Clinical Chemistry* **54**, 1125 (2008).
137. Hu, D. *et al.* The MLL3/MLL4 branches of the COMPASS family function as major histone H3K4 monomethylases at enhancers. *Mol Cell Biol* **33**, 4745-54 (2013).
138. Yang, L., Rau, R. & Goodell, M.A. DNMT3A in haematological malignancies. *Nat Rev Cancer* **15**, 152-65 (2015).
139. Huang, Y. & Rao, A. Connections between TET proteins and aberrant DNA modification in cancer. *Trends Genet* **30**, 464-74 (2014).

140. Solomon, D.A., Kim, J.S. & Waldman, T. Cohesin gene mutations in tumorigenesis: from discovery to clinical significance. *BMB Rep* **47**, 299-310 (2014).
141. Plass, C. *et al.* Mutations in regulators of the epigenome and their connections to global chromatin patterns in cancer. *Nat Rev Genet* **14**, 765-80 (2013).
142. Li, Q.-L. *et al.* The hyper-activation of transcriptional enhancers in breast cancer. *Clinical Epigenetics* **11**, 48 (2019).
143. Chen, H. *et al.* A Pan-Cancer Analysis of Enhancer Expression in Nearly 9000 Patient Samples. *Cell* **173**, 386-399.e12 (2018).
144. Ding, M. *et al.* Enhancer RNAs (eRNAs): New Insights into Gene Transcription and Disease Treatment. *Journal of Cancer* **9**, 2334-2340 (2018).
145. Heintzman, N.D. *et al.* Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet* **39**, 311-8 (2007).
146. Birney, E. *et al.* Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**, 799-816 (2007).
147. Ernst, J. & Kellis, M. ChromHMM: automating chromatin-state discovery and characterization. *Nat Methods* **9**, 215-6 (2012).
148. Hoffman, M.M. *et al.* Unsupervised pattern discovery in human chromatin structure through genomic segmentation. *Nat Methods* **9**, 473-6 (2012).
149. Fernández, M. & Miranda-Saavedra, D. Genome-wide enhancer prediction from epigenetic signatures using genetic algorithm-optimized support vector machines. *Nucleic acids research* **40**, e77-e77 (2012).
150. Kapitsinou, P.P. & Haase, V.H. The VHL tumor suppressor and HIF: insights from genetic studies in mice. *Cell Death Differ* **15**, 650-9 (2008).
151. Gossage, L. *et al.* Clinical and pathological impact of VHL, PBRM1, BAP1, SETD2, KDM6A, and JARID1c in clear cell renal cell carcinoma. *Genes Chromosomes Cancer* **53**, 38-51 (2014).
152. Reisman, D., Glaros, S. & Thompson, E.A. The SWI/SNF complex and cancer. *Oncogene* **28**, 1653-68 (2009).
153. Sato, Y. *et al.* Integrated molecular analysis of clear-cell renal cell carcinoma. *Nat Genet* **45**, 860-7 (2013).
154. Comprehensive molecular characterization of clear cell renal cell carcinoma. *Nature* **499**, 43-9 (2013).
155. Li, F. *et al.* The histone mark H3K36me3 regulates human DNA mismatch repair through its interaction with MutSalpha. *Cell* **153**, 590-600 (2013).

156. Carvalho, S. *et al.* SETD2 is required for DNA double-strand break repair and activation of the p53-mediated checkpoint. *Elife* **3**, e02482 (2014).
157. Chantalat, S. *et al.* Histone H3 trimethylation at lysine 36 is associated with constitutive and facultative heterochromatin. *Genome Res* **21**, 1426-37 (2011).
158. Dalgliesh, G.L. *et al.* Systematic sequencing of renal carcinoma reveals inactivation of histone modifying genes. *Nature* **463**, 360-363 (2010).
159. Guo, X. & Zhang, Q. The Emerging Role of Histone Demethylases in Renal Cell Carcinoma. *Journal of kidney cancer and VHL* **4**, 1-5 (2017).
160. Niu, X. *et al.* The von Hippel-Lindau tumor suppressor protein regulates gene expression and tumor growth through histone demethylase JARID1C. *Oncogene* **31**, 776-86 (2012).
161. Shen, Y. *et al.* Expression and significance of histone H3K27 demethylases in renal cell carcinoma. *BMC Cancer* **12**, 470 (2012).
162. Kluzek, K., Bluyssen, H.A. & Wesoly, J. The epigenetic landscape of clear-cell renal cell carcinoma. *Journal of kidney cancer and VHL* **2**, 90-104 (2015).
163. Xu, B. *et al.* Enhancer of Zeste Homolog 2 Expression Is Associated With Metastasis and Adverse Clinical Outcome in Clear Cell Renal Cell Carcinoma: A Comparative Study and Review of the Literature. *Archives of Pathology & Laboratory Medicine* **137**, 1326-1336 (2013).
164. Mosashvilli, D. *et al.* Global histone acetylation levels: prognostic relevance in patients with renal cell carcinoma. *Cancer Sci* **101**, 2664-9 (2010).
165. Ellinger, J. *et al.* Prognostic relevance of global histone H3 lysine 4 (H3K4) methylation in renal cell carcinoma. *Int J Cancer* **127**, 2360-6 (2010).
166. Rogenhofer, S. *et al.* Global histone H3 lysine 27 (H3K27) methylation levels and their prognostic relevance in renal cell carcinoma. *BJU Int* **109**, 459-65 (2012).
167. Ricketts, C.J., Hill, V.K. & Linehan, W.M. 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* **9**, e85621-e85621 (2014).
168. Pagliaro, L. & Shenoy, N. Sequential pathogenesis of metastatic VHL mutant clear cell renal cell carcinoma: putting it together with a translational perspective. *Annals of Oncology* **27**, 1685-1695 (2016).
169. The Cancer Genome Atlas Research, N. *et al.* Comprehensive molecular characterization of clear cell renal cell carcinoma. *Nature* **499**, 43 (2013).

170. Malouf, G. *et al.* Association of CpG island methylator phenotype with clear-cell renal cell carcinoma aggressiveness. *Journal of Clinical Oncology* **32**, 4574-4574 (2014).
171. Arai, E. *et al.* Single-CpG-resolution methylome analysis identifies clinicopathologically aggressive CpG island methylator phenotype clear cell renal cell carcinomas. *Carcinogenesis* **33**, 1487-93 (2012).
172. Arai, E. *et al.* Genome-wide DNA methylation profiles in both precancerous conditions and clear cell renal cell carcinomas are correlated with malignant potential and patient outcome. *Carcinogenesis* **30**, 214-21 (2009).
173. Hu, C.Y. *et al.* Kidney cancer is characterized by aberrant methylation of tissue-specific enhancers that are prognostic for overall survival. *Clin Cancer Res* **20**, 4349-60 (2014).
174. Bhagat, T.D. *et al.* Notch Pathway Is Activated via Genetic and Epigenetic Alterations and Is a Therapeutic Target in Clear Cell Renal Cancer. *J Biol Chem* **292**, 837-846 (2017).
175. Munari, E. *et al.* Global 5-Hydroxymethylcytosine Levels Are Profoundly Reduced in Multiple Genitourinary Malignancies. *PLoS One* **11**, e0146302 (2016).
176. Riazalhosseini, Y. & Lathrop, M. Precision medicine from the renal cancer genome. *Nat Rev Nephrol* **12**, 655-666 (2016).
177. Blankenship, C., Naglich, J.G., Whaley, J.M., Seizinger, B. & Kley, N. Alternate choice of initiation codon produces a biologically active product of the von Hippel Lindau gene with tumor suppressor activity. *Oncogene* **18**, 1529 (1999).
178. Schoenfeld, A., Davidowitz, E.J. & Burk, R.D. A second major native von Hippel-Lindau gene product, initiated from an internal translation start site, functions as a tumor suppressor. *Proc Natl Acad Sci U S A* **95**, 8817-22 (1998).
179. Minervini, G. *et al.* Isoform-specific interactions of the von Hippel-Lindau tumor suppressor protein. *Scientific Reports* **5**, 12605 (2015).
180. Frew, I.J. & Moch, H. A Clearer View of the Molecular Complexity of Clear Cell Renal Cell Carcinoma. *Annual Review of Pathology: Mechanisms of Disease* **10**, 263-289 (2015).
181. Hsu, T. Complex cellular functions of the von Hippel–Lindau tumor suppressor gene: insights from model organisms. *Oncogene* **31**, 2247 (2011).

182. Shenoy, N. & Pagliaro, L. Sequential pathogenesis of metastatic VHL mutant clear cell renal cell carcinoma: putting it together with a translational perspective. *Annals of Oncology* **27**, 1685-1695 (2016).
183. Lv, X. *et al.* The role of hypoxia-inducible factors in tumor angiogenesis and cell metabolism. *Genes & diseases* **4**, 19-24 (2016).
184. Cockman, M.E. *et al.* Hypoxia inducible factor-alpha binding and ubiquitylation by the von Hippel-Lindau tumor suppressor protein. *J Biol Chem* **275**, 25733-41 (2000).
185. Bader, H.L. & Hsu, T. Systemic VHL gene functions and the VHL disease. *FEBS letters* **586**, 1562-1569 (2012).
186. Robinson, C.M. *et al.* Consequences of VHL Loss on Global DNA Methylome. *Scientific Reports* **8**, 3313 (2018).
187. Ricketts, C.J. & Linehan, W.M. Insights into Epigenetic Remodeling in VHL-Deficient Clear Cell Renal Cell Carcinoma. *Cancer Discovery* **7**, 1221 (2017).
188. Scelo, G. *et al.* Variation in genomic landscape of clear cell renal cell carcinoma across Europe. *Nature Communications* **5**, 5135 (2014).
189. Bhagat, T.D. *et al.* Notch Pathway Is Activated via Genetic and Epigenetic Alterations and Is a Therapeutic Target in Clear Cell Renal Cancer. *Journal of Biological Chemistry* **292**, 837-846 (2017).
190. Wozniak, M.B. *et al.* Integrative Genome-Wide Gene Expression Profiling of Clear Cell Renal Cell Carcinoma in Czech Republic and in the United States. *PLOS ONE* **8**, e57886 (2013).
191. DeAngelis, J.T., Farrington, W.J. & Tollefsbol, T.O. An overview of epigenetic assays. *Molecular biotechnology* **38**, 179-183 (2008).
192. Martín-Subero, J.I. & Esteller, M. Profiling Epigenetic Alterations in Disease. in *Epigenetic Contributions in Autoimmune Disease* (ed. Ballestar, E.) 162-177 (Springer US, Boston, MA, 2011).
193. Park, P.J. ChIP-seq: advantages and challenges of a maturing technology. *Nature reviews. Genetics* **10**, 669-680 (2009).
194. Stevens, M. *et al.* Estimating absolute methylation levels at single-CpG resolution from methylation enrichment and restriction enzyme sequencing methods. *Genome research* **23**, 1541-1553 (2013).
195. Jones, P.A. & Martienssen, R. A blueprint for a Human Epigenome Project: the AACR Human Epigenome Workshop. *Cancer Res* **65**, 11241-6 (2005).

196. Davis, C.A. *et al.* The Encyclopedia of DNA elements (ENCODE): data portal update. *Nucleic Acids Res* **46**, D794-d801 (2018).
197. The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science* **306**, 636-40 (2004).
198. Rakyan, V.K. *et al.* DNA methylation profiling of the human major histocompatibility complex: a pilot study for the human epigenome project. *PLoS Biol* **2**, e405 (2004).
199. Bujold, D. *et al.* The International Human Epigenome Consortium Data Portal. *Cell Systems* **3**, 496-499.e2 (2016).
200. Cancer Genome Atlas Research, N. *et al.* The Cancer Genome Atlas Pan-Cancer analysis project. *Nature genetics* **45**, 1113-1120 (2013).
201. Zhang, J. *et al.* International Cancer Genome Consortium Data Portal--a one-stop shop for cancer genomics data. *Database : the journal of biological databases and curation* **2011**, bar026-bar026 (2011).
202. Liu, T. *et al.* Cistrome: an integrative platform for transcriptional regulation studies. *Genome Biol* **12**, R83 (2011).
203. Shijie, F., Yu, C., Cheng, L. & Fanwang, M. Machine Learning Methods in Precision Medicine Targeting Epigenetic Diseases. *Current Pharmaceutical Design* **24**, 3998-4006 (2018).
204. Heung, B. *et al.* An overview and comparison of machine-learning techniques for classification purposes in digital soil mapping. *Geoderma* **265**, 62-77 (2016).
205. Holder, L.B., Haque, M.M. & Skinner, M.K. Machine learning for epigenetics and future medical applications. *Epigenetics* **12**, 505-514 (2017).
206. Hastie, T., Tibshirani, R. & Friedman, J. Overview of Supervised Learning. in *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (eds. Hastie, T., Tibshirani, R. & Friedman, J.) 9-41 (Springer New York, New York, NY, 2009).
207. Chicco, D. Ten quick tips for machine learning in computational biology. *BioData mining* **10**, 35-35 (2017).
208. Arlot, S. & Celisse, A. A survey of cross-validation procedures for model selection. *Statist. Surv.* **4**, 40-79 (2010).
209. Refaeilzadeh, P., Tang, L. & Liu, H. Cross-Validation. in *Encyclopedia of Database Systems* (eds. Liu, L. & ÖZsu, M.T.) 532-538 (Springer US, Boston, MA, 2009).
210. Tarca, A.L., Carey, V.J., Chen, X.-w., Romero, R. & Drăghici, S. Machine Learning and Its Applications to Biology. *PLOS Computational Biology* **3**, e116 (2007).

211. Das, R. *et al.* Computational prediction of methylation status in human genomic sequences. *Proc Natl Acad Sci U S A* **103**, 10713-6 (2006).
212. Orozco, J.I.J. *et al.* Epigenetic profiling for the molecular classification of metastatic brain tumors. *Nature communications* **9**, 4627-4627 (2018).
213. Xu, X., Hoang, S., Mayo, M.W. & Bekiranov, S. Application of machine learning methods to histone methylation ChIP-Seq data reveals H4R3me2 globally represses gene expression. *BMC Bioinformatics* **11**, 396 (2010).
214. Histone Modifications (https://epigenomesportal.ca/edcc/doc/experiment_ChIP-Seq.html).
215. Transcriptome (https://epigenomesportal.ca/edcc/doc/experiment_RNA-Seq.html).
216. Methylome (https://epigenomesportal.ca/edcc/doc/experiment_WGB-Seq.html).
217. Bolger, A.M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics (Oxford, England)* **30**, 2114-2120 (2014).
218. Langmead, B. & Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nature methods* **9**, 357-359 (2012).
219. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)* **25**, 2078-2079 (2009).
220. Zhang, Y. *et al.* Model-based Analysis of ChIP-Seq (MACS). *Genome Biology* **9**, R137 (2008).
221. Kim, D., Langmead, B. & Salzberg, S.L. HISAT: a fast spliced aligner with low memory requirements. *Nature methods* **12**, 357-360 (2015).
222. Pyl, P.T., Anders, S. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166-169 (2014).
223. Harrow, J. *et al.* GENCODE: the reference human genome annotation for The ENCODE Project. *Genome research* **22**, 1760-1774 (2012).
224. Love, M.I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology* **15**, 550-550 (2014).
225. Quinlan, A.R. & Hall, I.M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842 (2010).
226. Edgar, R., Domrachev, M. & Lash, A.E. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* **30**, 207-10 (2002).
227. Bernstein, B.E. *et al.* The NIH Roadmap Epigenomics Mapping Consortium. *Nat Biotechnol* **28**, 1045-8 (2010).

228. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74 (2012).
229. Pohl, A. & Beato, M. bwtool: a tool for bigWig files. *Bioinformatics (Oxford, England)* **30**, 1618-1619 (2014).
230. Zacher, B. *et al.* Accurate Promoter and Enhancer Identification in 127 ENCODE and Roadmap Epigenomics Cell Types and Tissues by GenoSTAN. *PLOS ONE* **12**, e0169249 (2017).
231. Krueger, F. & Andrews, S.R. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics (Oxford, England)* **27**, 1571-1572 (2011).
232. Akalin, A. *et al.* methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome biology* **13**, R87-R87 (2012).
233. Breiman, L. Random Forests. *Machine Learning* **45**, 5-32 (2001).
234. Liaw, A. & Wiener, M. Classification and Regression by randomForest. *R News* **2**, 18--22 (2002).
235. Ignatiadis, N., Klaus, B., Zaugg, J.B. & Huber, W. Data-driven hypothesis weighting increases detection power in genome-scale multiple testing. *Nature methods* **13**, 577-580 (2016).
236. Subramanian, A. *et al.* Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences* **102**, 15545 (2005).
237. Aryee, M.J. *et al.* Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* **30**, 1363-1369 (2014).
238. Fortin, J.-P. *et al.* Functional normalization of 450k methylation array data improves replication in large cancer studies. *Genome biology* **15**, 503-503 (2014).
239. Kent, W.J. *et al.* The Human Genome Browser at UCSC. *Genome Research* **12**, 996-1006 (2002).
240. Bernhart, S.H. *et al.* Changes of bivalent chromatin coincide with increased expression of developmental genes in cancer. *Scientific reports* **6**, 37393-37393 (2016).
241. Kurdistani, S.K. Histone modifications as markers of cancer prognosis: a cellular view. *British journal of cancer* **97**, 1-5 (2007).
242. Heintzman, N.D. *et al.* Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nature Genetics* **39**, 311 (2007).

243. Maston, G.A., Evans, S.K. & Green, M.R. Transcriptional Regulatory Elements in the Human Genome. *Annual Review of Genomics and Human Genetics* **7**, 29-59 (2006).
244. Sur, I. & Taipale, J. The role of enhancers in cancer. *Nature Reviews Cancer* **16**, 483 (2016).
245. Kamburov, A., Wierling, C., Lehrach, H. & Herwig, R. ConsensusPathDB--a database for integrating human functional interaction networks. *Nucleic acids research* **37**, D623-D628 (2009).
246. Sandoval, J. *et al.* Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome. *Epigenetics* **6**, 692-702 (2011).
247. Heyn, H. *et al.* Epigenomic analysis detects aberrant super-enhancer DNA methylation in human cancer. *Genome biology* **17**, 11-11 (2016).
248. Kondo, Y. Epigenetic cross-talk between DNA methylation and histone modifications in human cancers. *Yonsei medical journal* **50**, 455-463 (2009).
249. Okitsu, C.Y. & Hsieh, C.-L. DNA methylation dictates histone H3K4 methylation. *Molecular and cellular biology* **27**, 2746-2757 (2007).
250. Nan, X. *et al.* Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature* **393**, 386 (1998).
251. Visel, A., Rubin, E.M. & Pennacchio, L.A. Genomic views of distant-acting enhancers. *Nature* **461**, 199-205 (2009).
252. Pennacchio, L.A., Bickmore, W., Dean, A., Nobrega, M.A. & Bejerano, G. Enhancers: five essential questions. *Nature Reviews Genetics* **14**, 288 (2013).
253. Mohrs, M. *et al.* Deletion of a coordinate regulator of type 2 cytokine expression in mice. *Nature Immunology* **2**, 842 (2001).
254. Chen, C.-H. *et al.* Determinants of transcription factor regulatory range. *bioRxiv*, 582270 (2019).
255. Wong, R.W.J. *et al.* Enhancer profiling identifies critical cancer genes and characterizes cell identity in adult T-cell leukemia. *Blood* **130**, 2326-2338 (2017).
256. Schödel, J. *et al.* Hypoxia, Hypoxia-inducible Transcription Factors, and Renal Cancer. *European urology* **69**, 646-657 (2016).
257. Tsuzuki, Y. *et al.* Vascular endothelial growth factor (VEGF) modulation by targeting hypoxia-inducible factor-1alpha--> hypoxia response element--> VEGF cascade differentially regulates vascular response and growth rate in tumors. *Cancer Res* **60**, 6248-52 (2000).

258. Semenza, G.L. A compendium of proteins that interact with HIF-1α. *Experimental cell research* **356**, 128-135 (2017).
259. Gonzalez, H., Hagerling, C. & Werb, Z. Roles of the immune system in cancer: from tumor initiation to metastatic progression. *Genes & development* **32**, 1267-1284 (2018).
260. Tang, S.W. *et al.* MYC pathway is activated in clear cell renal cell carcinoma and essential for proliferation of clear cell renal cell carcinoma cells. *Cancer Lett* **273**, 35-43 (2009).
261. Stefan, E. & Bister, K. MYC and RAF: Key Effectors in Cellular Signaling and Major Drivers in Human Cancer. *Curr Top Microbiol Immunol* **407**, 117-151 (2017).
262. Ma, X. *et al.* MicroRNA-185 inhibits cell proliferation and induces cell apoptosis by targeting VEGFA directly in von Hippel-Lindau-inactivated clear cell renal cell carcinoma. *Urol Oncol* **33**, 169.e1-11 (2015).
263. Matusan-Ilijas, K. *et al.* EGFR expression is linked to osteopontin and Nf-kappaB signaling in clear cell renal cell carcinoma. *Clin Transl Oncol* **15**, 65-71 (2013).
264. Cossu-Rocca, P. *et al.* EGFR kinase-dependent and kinase-independent roles in clear cell renal cell carcinoma. *Am J Cancer Res* **6**, 71-83 (2016).
265. Tun, H.W. *et al.* Pathway Signature and Cellular Differentiation in Clear Cell Renal Cell Carcinoma. *PLOS ONE* **5**, e10696 (2010).
266. Wu, K., Xu, L., Zhang, L., Lin, Z. & Hou, J. High Jagged1 expression predicts poor outcome in clear cell renal cell carcinoma. *Jpn J Clin Oncol* **41**, 411-6 (2011).
267. Artemov, A.V., Zhigalova, N., Zhenilo, S., Mazur, A.M. & Prokhortchouk, E.B. VHL inactivation without hypoxia is sufficient to achieve genome hypermethylation. *Scientific reports* **8**, 10667-10667 (2018).
268. Breuksch, I. *et al.* Integrin α5 triggers the metastatic potential in renal cell carcinoma. *Oncotarget* **8**, 107530-107542 (2017).
269. Schleussner, N. *et al.* The AP-1-BATF and -BATF3 module is essential for growth, survival and TH17/ILC3 skewing of anaplastic large cell lymphoma. *Leukemia* **32**, 1994-2007 (2018).
270. Urbschat, A. *et al.* Expression of the anti-inflammatory suppressor of cytokine signaling 3 (SOCS3) in human clear cell renal cell carcinoma. *Tumour Biol* **37**, 9649-56 (2016).
271. Vire, E. *et al.* The Polycomb group protein EZH2 directly controls DNA methylation. *Nature* **439**, 871-4 (2006).

272. Gan, L. *et al.* Epigenetic regulation of cancer progression by EZH2: from biological insights to therapeutic potential. *Biomarker research* **6**, 10-10 (2018).
273. Schodel, J. *et al.* High-resolution genome-wide mapping of HIF-binding sites by ChIP-seq. *Blood* **117**, e207-17 (2011).
274. Stow, L.R., Jacobs, M.E., Wingo, C.S. & Cain, B.D. Endothelin-1 gene regulation. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* **25**, 16-28 (2011).
275. Sowter, H.M., Raval, R.R., Moore, J.W., Ratcliffe, P.J. & Harris, A.L. Predominant role of hypoxia-inducible transcription factor (Hif)-1alpha versus Hif-2alpha in regulation of the transcriptional response to hypoxia. *Cancer Res* **63**, 6130-4 (2003).

APPENDICES

Supplementary Materials

Supplementary methods

Histone broad peaks comparison with reference peaks

Histone Chip-Seq broad peaks of unconsolidated adult kidney from road map epigenome[21] peaks were downloaded (link at the end). Coordinates were converted to GRCh38 from hg19. Bedtools[225] shuffle (version 2.26.0) was used to generate a set of randomly shuffled histone mark peaks of kidney tissues from road map epigenome project. Overlap percentage was calculated comparing H3K27ac, H3K4me1 and H3K4me3 broad peaks from both normal and tumor tissues (from four ccRCC patients) with corresponding histone broad peaks from kidney tissues and their randomly shuffled genomic regions. (https://egg2.wustl.edu/roadmap/data/byFileType/peaks/unconsolidated/broadPeak/)

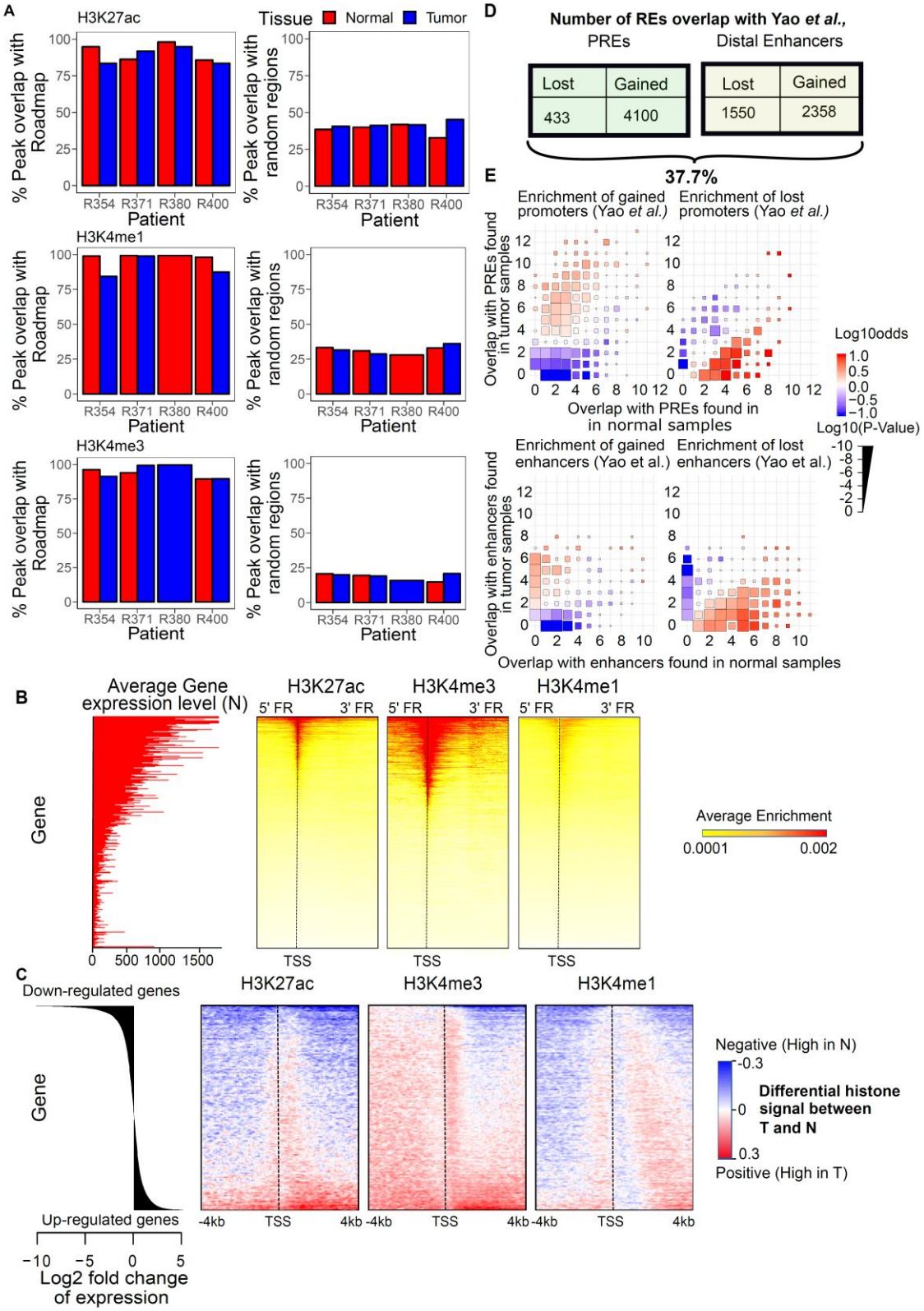
Visualizing the histone enrichment profiles surrounding TSS

Chip-Seq signal files generated by MACS2[220] of normal samples of patient LR354 were obtained (for H3K27ac, H3K4me3 and H3K4me1) and signal of each coordinate was normalized by total area under the curve of each sample. Normalized signal, files (bedgraph) were used for subsequent analysis. For one gene, gene body and similar width to gene body of surrounding flanking regions were considered for the visualization. ChIP-Seq signals of all three regions were separately averaged to 800bp each (one row in the heatmap). The same procedure was applied to all the genes and genes were ordered by the total average signal across rows of H3K27ac enrichment profile and this order was used as a reference for other histone marks. Average base mean expression values normalized to library size of each normal sample were used to compare with histone enrichment profiles.

Visualizing the differential ChIP-Seq signal for histone marks around TSS

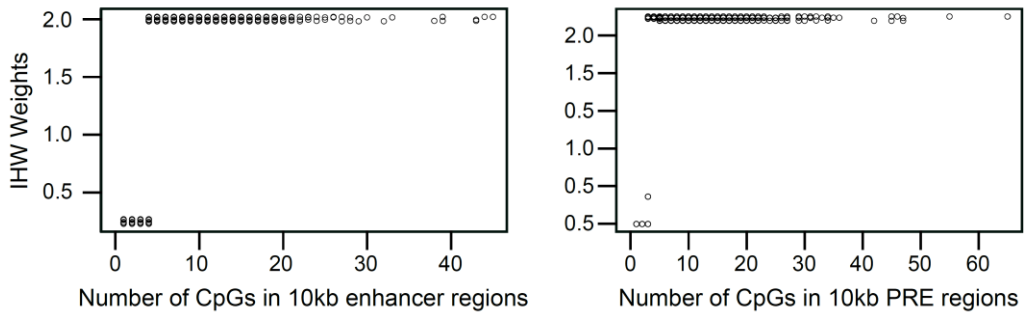
Normalized ChIP-Seq signal files of both tumor and normal samples were used and 4kb region of TSS of each gene was selected. Genes were ordered based on the log2 fold change of expression of four patients comparing tumor relative to normal samples. Signal difference of tumor relative to normal samples were then calculated and used to visualize the change of histone modification.

Supplementary figures

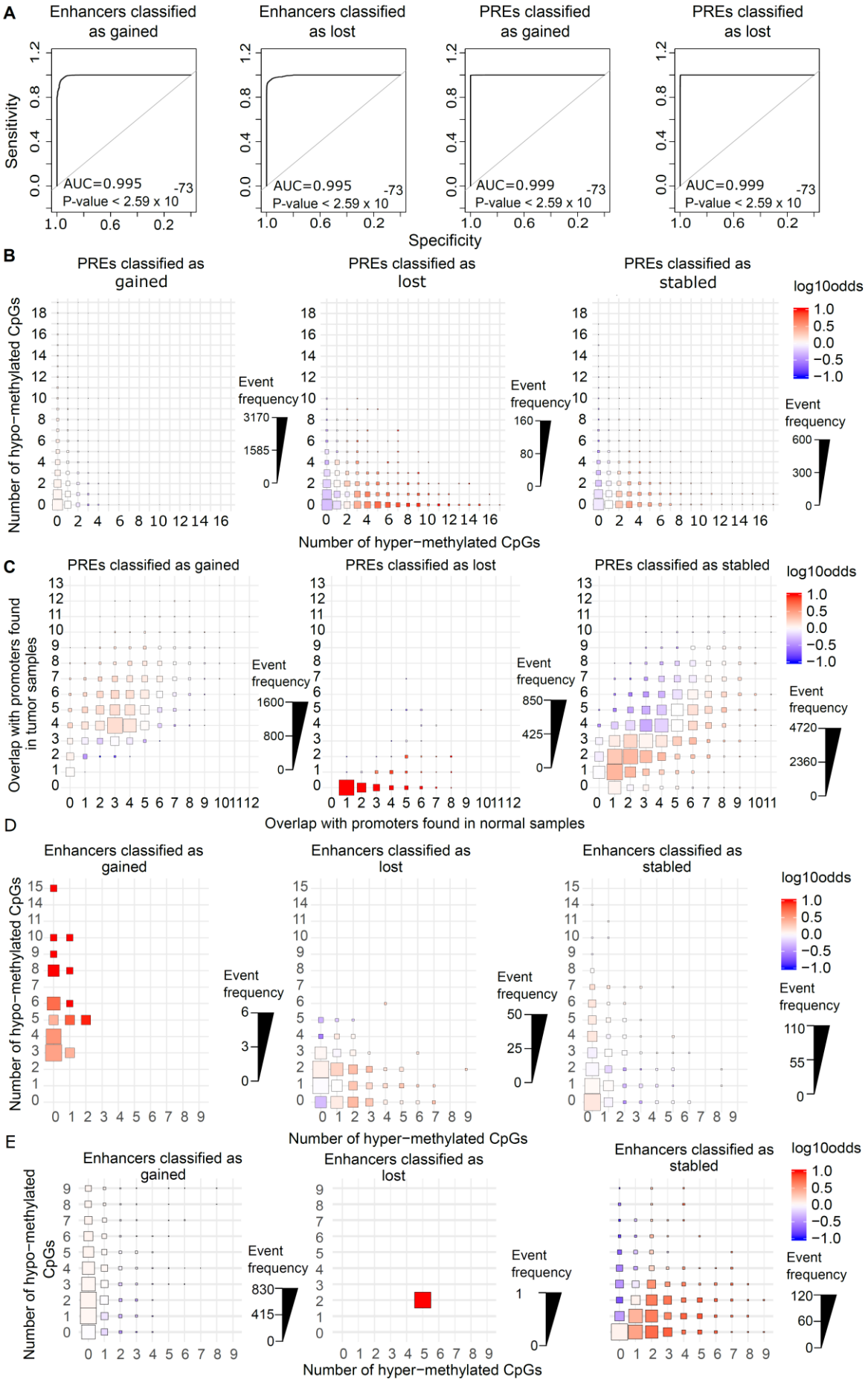


Supplementary Figure S1: A. Comparison of histone mark peaks identified in this study with peaks from Roadmap Epigenomics dataset: Percentage of overlapping of H3K27ac, H3K4me1 and H3K4me3 broad peaks from both normal and tumor tissues (from four ccRCC

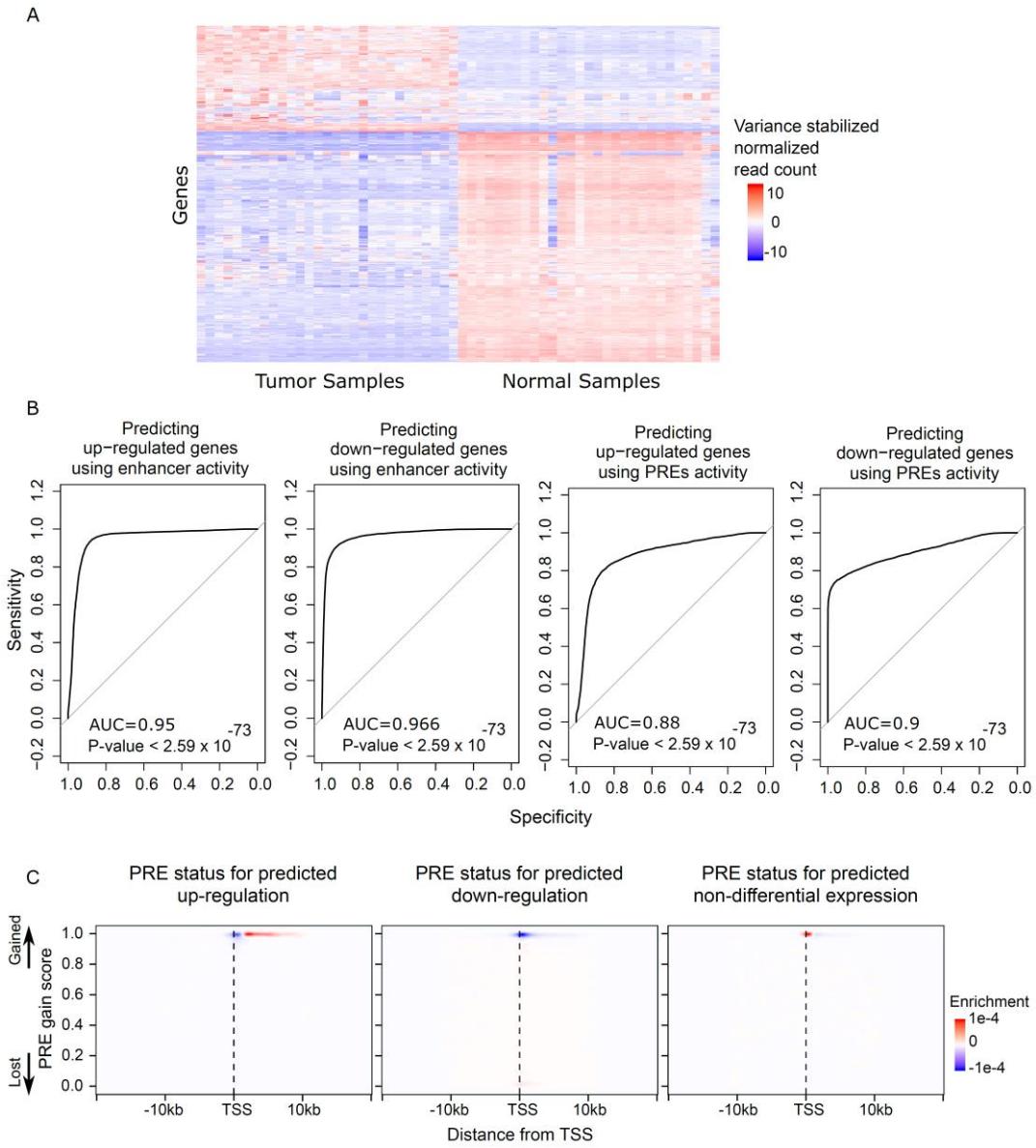
patients) with corresponding histone broad peaks from unconsolidated adult kidney tissues (left) or their genomic regions were randomly shuffled (right). Note the higher percentage of overlap with actual RoadMap Epigenomics peaks than random genomic regions. H3K4me3 from the tumor sample of patient LR380 and H3K4me1 from its normal sample were excluded from this study due to low data quality; **B.** Histone enrichment profiles across the upstream flanking region, gene body, and downstream flanking region of genes: Each row represents one gene, with genes sorted by their average expression (shown in the left graph). The color gradient represents the average signal for H3K27ac, H3K4me3 and H3K4me1, obtained from normal samples (red represents the highest signal, and yellow denotes the lowest signal). The upstream flanking region, downstream flanking region, and gene body are normalized to have the same length for visualization purposes. FR: flanking region; **C.** Differential ChIP-seq signal for histone marks around TSS and its relationshi to differential gene expression between tumor (T) and normal (N) tissue samples: Each row represents one gene, sorted by their log2 fold change in gene expression as shown in the left graph. The color gradient represents log-fold change in average histone mark signal between T and N (red: enrichment in T; blue: enrichment in N tissue). H3K27ac and H3K4me3 panels were generated using data from patient LR354; H3K4me1 was analyzed using data from RL400 (selected due to availability of high-quality data in both T and N tissues). **D.** Number of regulatory elements (REs) in this study overlap with gained and lost regulatory elements in ccRCC defined by Yao *et al.*[79]: Number of overlaps of all core elements (without considering the origin tissue) with a previous study (Yao *et al.*) that identified a set of regulatory elements in ccRCC are shown in the tables. Percentage value denotes total percentage of regulatory elements in our study found in Yao *et. al.*, study; **E.** Distribution of tumor- and normal-specific regulatory elements from a previous study[79] with respect to the frequency of their observation in our tumor and normal samples. The x- and y-axes correspond to the number of times an overlapping element is observed in any normal and tumor sample, respectively. Color and annotations are similar to **Figure 8A**.



Supplementary Figure S2: Relationship between number of CpGs and statistical power (IHW) in PREs (left) and enhancers (right): Two-sided exact binomial test was used to calculate the significance of methylation imbalance (between hyper- and hypo-methylated CpGs) on the 10kb region around enhancers and PREs. Then independence hypothesis weighing (IHW)[235] was used to calculate the adjusted P-value, optimizing for FDR < 0.1 using the level of significance (P-value) and the number of differentially methylated CpGs on the element (n).


A
Enhancers classified as gained
Enhancers classified as lost
PREs classified as gained
PREs classified as lost
Sensitivity
Specificity
AUC=0.995
P-value < 2.59 x 10 -73
AUC=0.995
P-value < 2.59 x 10 -73
AUC=0.999
P-value < 2.59 x 10 -73
AUC=0.999
P-value < 2.59 x 10 -73
B
PREs classified as gained
PREs classified as lost
PREs classified as stabled
Number of hypo−methylated CpGs
Number of hyper−methylated CpGs
Event frequency
log10odds
C
PREs classified as gained
PREs classified as lost
PREs classified as stabled
Overlap with promoters found in tumor samples
Overlap with promoters found in normal samples
Event frequency
log10odds
D
Enhancers classified as gained
Enhancers classified as lost
Enhancers classified as stabled
Number of hypo−methylated CpGs
Number of hyper−methylated CpGs
Event frequency
log10odds
E
Enhancers classified as gained
Enhancers classified as lost
Enhancers classified as stabled
Number of hypo−methylated CpGs
Number of hyper−methylated CpGs
Event frequency
log10odds

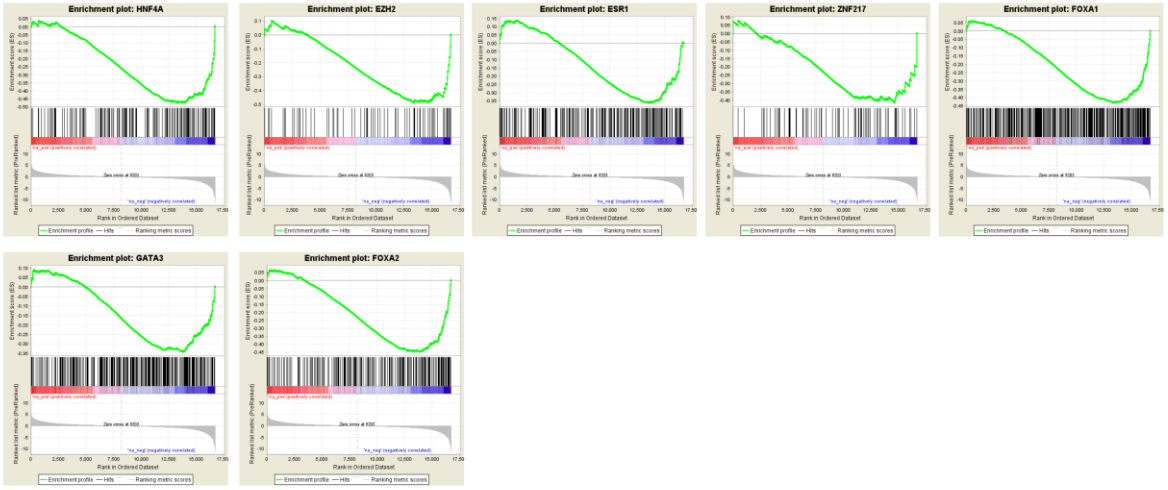
Supplementary Figure S3: A. Receiver operating characteristic (ROC) curves for identifying gained and lost enhancers and PREs. Cross-validation on the gold standard data set (*i.e.* gained/lost promoters and enhancers reported in a previous study[79]) was used to produce the ROC curves in order to evaluate the classification approach. In this cross-validation, all enhancers from one chromosome were held-out for testing, and the classifier was trained on the other chromosomes. Then, the held-out data were used for determining gain and loss events (leave-one-chromosome-out cross-validation). This ensures that local information will not be leaked across nearby elements during the training and testing phases; **B.** Distribution of hypo- and hyper-methylated CpGs among PREs that are classified as gained, lost, or stable by the random forest classifier. Annotations are similar to **Figure 3A**; **C.** Similar to **panel (B)**, but x- and y-axes correspond to the number of overlaps of each region with the PREs identified from normal and tumor samples, respectively; **D.** Similar to **Figure 3A** but, distribution only for where enhancer overlaps one element identified from a tumor sample and one element identified from a normal sample. **E.** Similar to **Figure 3A** but, distribution only for where enhancer overlaps more than one element identified from a tumor sample and no element identified from a normal sample



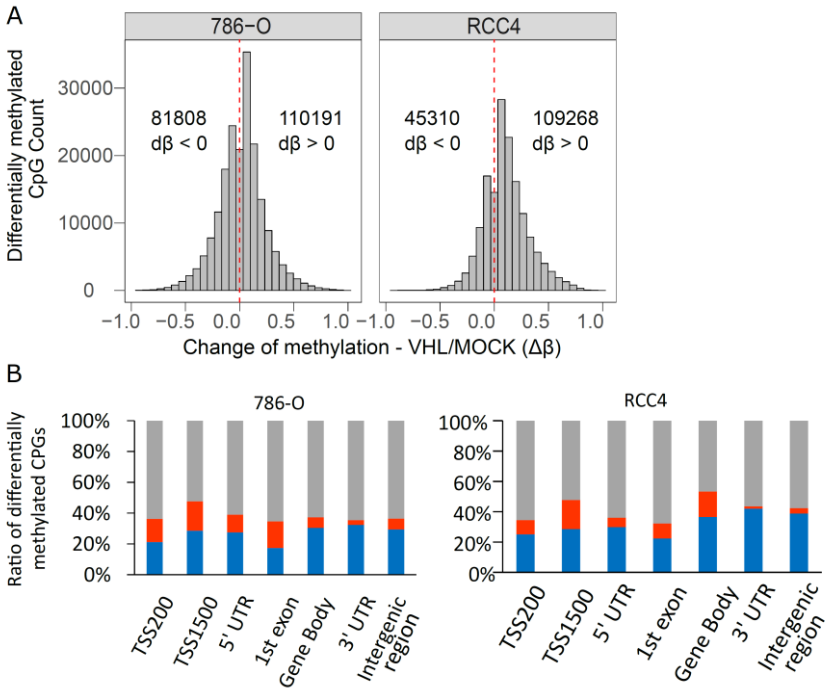
Supplementary Figure S4: A. Distribution of RNA-Seq gene expression values among tumor and normal samples. Color gradient shows the variance-stabilized values of read counts normalized to have an average of zero for each sample. Top 1000 genes with highest variance are shown in the figure; **B.** Receiver operating characteristic (ROC) curves for predicting the expression of genes associated with regulatory elements: Leave-one-out cross-validation was used to evaluate the performance of our classifier for predicting up- or down-regulated genes (FDR < 0.01) using regulatory elements activity; **C.** Visualization of the PRE status for predicted up-, down-regulated and non-differentially expressed genes relative to the PRE location from TSS. PRE gain score is shown on the y axis while distance from TSS is shown on the x axis. Red indicates enrichment and blue shows depletion. Therefore, as an example, the red color at the top middle part of the left graph shows an enrichment of gained PREs downstream and in the vicinity of TSS of genes that are predicted to be up-regulated.

Regulatory factors enriched in gained enhancers and GSEA results for their target genes

Regulatory factors enriched in lost enhancers and GSEA results for their target genes



Supplementary Figure S5: GSEA enrichment plots for target genes associated with regulatory factors. Target genes of each RF whose binding sites are significantly enriched in gained or lost enhancers were selected, followed by a GSEA pre-ranked test to examine whether the targets of the RF are enriched in up- or down-regulated genes (tumor vs. normal). The figure shows the enrichment plots generated by the GSEA software for all selected 32 RFs.



Supplementary Figure S6: A. Distribution of differentially methylated CpGs relative to change of methylation (Δβ). Number of significantly differentially methylated (VHL+ve vs. MOCK) CpGs (0.05 <FDR) relative to the change of methylation (Δβ) for 786-O and RCC4 cell lines. Overall, 110191 and 109268 loci are hyper-methylated, and 81808 and 45310 loci are hypo-methylated in 786-O and RCC4 respectively; **B**. The distribution of the ratio between hyper-, hypo-methylated, non-differentially methylated CpGs and total CpGs of different genomic regions. The distribution of the ratio of differentially methylated CpGs for RCC4 (VHL/MOCK) cell lines and patient (T/N) samples;

Supplementary tables

Supplementary Table S1: Clinical information of patients; M: male, F: female

Patient ID	Sex	Ethnicity	Country	Tumor Stage	Tumor Laterality	Age category	Tobacco usage	Profiled data type
LR354	M	European ancestry	UK	1	Right kidney	<55	Current smoker	ChIP-Seq, RNA-Seq, WGBS
LR371	M	European ancestry	UK	1	Left kidney	65-69	Ex-smoker	ChIP-Seq, RNA-Seq
LR380	M	European ancestry	UK	1	Left kidney	70+	Ex-smoker	ChIP-Seq, RNA-Seq, WGBS
LR398	F	European ancestry	UK	3	Right kidney	60-64	Ex-smoker	WGBS
LR400	F	European ancestry	UK	1	Left kidney	55-59	Never	ChIP-Seq, RNA-Seq, WGBS

Supplementary Table S2: ChIP-seq analysis statistics; T: tumor, N: Normal

Patient ID	Sample type	Histone modification	Raw read count	Overall alignment rate (%)	Tags after filtering in treatment	Tags after filtering in control
RL354	N	H3K27ac	31870072	98.00	25719060	15586321
RL354	T	H3K27ac	19049026	98.32	16800047	25431284
RL371	N	H3K27ac	29,868,488	98.39	25615646	25709754
RL371	T	H3K27ac	41470878	98.95	31103595	22728163
RL380	N	H3K27ac	31758758	99.05	27653117	11734774
RL380	T	H3K27ac	29962914	98.66	25526062	17604874
RL400	N	H3K27ac	21745067	98.29	17067870	43933173
RL400	T	H3K27ac	68034144	99.21	52067078	43234735
RL354	N	H3K4me3	29351725	97.10	21019122	17185401
RL354	T	H3K4me3	38076319	97.53	31156666	26512299
RL371	N	H3K4me3	31003950	98.41	23808649	27250881
RL371	T	H3K4me3	31584539	96.15	23143882	23502332
RL380	N	H3K4me3	35762186	98.72	27893559	18002713
RL380	T	H3K4me3	11771595	97.46	9652163	18070989
RL400	N	H3K4me3	60162306	96.00	24688722	39709982
RL400	T	H3K4me3	46787546	92.82	37872958	43234735
RL354	N	H3K4me1	63451450	95.19	51654811	15586321
RL354	T	H3K4me1	63144291	98.46	17312283	25431284
RL371	N	H3K4me1	66393517	96.07	56061156	25709754
RL371	T	H3K4me1	71001106	98.41	52115362	22728163
RL380	N	H3K4me1	35748586	95.37	28952502	11734774
RL380	T	H3K4me1	15649816	96.66	13110606	17604874
RL400	N	H3K4me1	67382824	98.62	48141488	38403486

RL400	T	H3K4me1	64914968	98.76	41127912	37916260

Supplementary Table S3: WGBS analysis statistics

Patient ID	Sample	Raw reads	Trimmed reads	Aligned reads	Mapping efficiency (%)
LR354	N_1	2,062,426,418	1,717,219,528	1,090,172,288	63.48
LR354	N_2	203,503,296	147,839,506	109,551,370	74.10
LR354	T_1	2,060,689,278	1,862,258,018	1,167,620,986	62.70
LR380	N_1	2,036,265,582	1,897,001,532	1,219,271,256	64.27
LR380	N_2	201,092,252	153,696,260	115,504,900	75.15
LR380	T_1	2,239,250,024	2,041,092,698	1305273884	63.95
LR380	T_2	235,831,200	167,402,550	122,748,676	73.32
LR398	N_1	1,936,284,554	1,809,086,682	1,227,681,344	67.86
LR398	T_1	2,130,245,932	1,944,291,960	1,291,919,996	66.44
LR400	N_1	776,882,358	748,832,184	653,434,186	87.26
LR400	T_1	799,885,498	772,774,862	675,394,396	87.39

Supplementary Table S4: Significantly enriched pathways of up-regulated target genes of gained enhancer-RF pairs

Pathway	FDR	Source
Interferon gamma signaling	0.001633	Reactome
Cytokine Signaling in Immune system	0.001892	Reactome
Immune System	0.003067	Reactome
T cell receptor signaling pathway – (human)	0.003708	KEGG
Leishmaniasis (human)	0.003708	KEGG
Adaptive Immune System	0.006734	Reactome
Interferon Signaling	0.006734	Reactome
Th1 and Th2 cell differentiation - (human)	0.006734	KEGG
HIF-1-alpha transcription factor network	0.00705	PID
Toll Like Receptor 4 (TLR4) Cascade	0.011596	Reactome
Regulation of IFNG signaling	0.012772	Reactome
Receptor-ligand binding initiates the second proteolytic cleavage of Notch receptor	0.014103	Reactome
Toll-Like Receptors Cascades	0.014103	Reactome
Direct p53 effectors	0.014103	PID
Allograft rejection - (human)	0.014103	KEGG
Signaling by NOTCH1 HD Domain Mutants in Cancer	0.015924	Reactome
Constitutive Signaling by NOTCH1 HD Domain Mutants	0.015924	Reactome
Toxoplasmosis - (human)	0.015924	KEGG
Signaling by NOTCH2	0.019254	Reactome
Signaling by PTK6	0.019271	Reactome
Signaling by Non-Receptor Tyrosine Kinases	0.019271	Reactome
Measles - (human)	0.019271	KEGG

Autoimmune thyroid disease - (human)	0.019271	KEGG
Alpha4 beta1 integrin signaling events	0.022168	PID
Tuberculosis - (human)	0.022168	KEGG
Signaling by NOTCH	0.02238	Reactome
Signal Transduction	0.02238	Reactome
MyD88:Mal cascade initiated on plasma membrane	0.02238	Reactome
Toll Like Receptor TLR1:TLR2 Cascade	0.02238	Reactome
Toll Like Receptor TLR6:TLR2 Cascade	0.02238	Reactome
Toll Like Receptor 2 (TLR2) Cascade	0.02238	Reactome
Signaling by NOTCH1 PEST Domain Mutants in Cancer	0.02238	Reactome
Signaling by NOTCH1 in Cancer	0.02238	Reactome
Constitutive Signaling by NOTCH1 PEST Domain Mutants	0.02238	Reactome
Signaling by NOTCH1 HD+PEST Domain Mutants in Cancer	0.02238	Reactome
Constitutive Signaling by NOTCH1 HD+PEST Domain Mutants	0.02238	Reactome
HIF-2-alpha transcription factor network	0.02238	PID
IL12-mediated signaling events	0.02238	PID
Cell adhesion molecules (CAMs) - Homo sapiens (human)	0.02238	KEGG

Supplementary Table S5: Up-regulated target genes associated with HIF1α and HIF2α transcription factor network and RFs associated with these target genes. 2nd and 3rd columns state whether these target genes are directly regulated by HIF1α and/or HIF2α[273-275]. HIF1α and HIF2α may contribute to the recruitment of other TFs whose target genes are involved in cellular process that are affected by hypoxia signaling.

Target gene	**HIF1 α**	**HIF2 α**	**RFs associated with target the gene**
ABCG2	no	no	NFIC
HMOX1	no	no	FOS, SPI1, JUN, FOSL2, STAT3
BHLHE40	no	no	NFATC1, MEF2C, JUNB, ATF1, IKZF1, STAT5A
ID2	no	no	ATF1, ATF2, BCL11A, BCL3, FOS, FOSL2, FOXM1, IKZF1, IRF4, JUN, JUNB, MEF2A, MEF2C, MTA3, NFATC1, NFIC, POU2F2, RELA, RUNX3, SPI1, STAT3, STAT5A, WRNIP1
PFKFB3	yes	yes	BATF, BCL11A, BCL3, EBF1, FOXM1, IKZF1, IRF4, MEF2A, NFIC, POU2F2, RELA, RUNX3, SPI1, STAT3, WRNIP1
CXCR4	no	no	JUN, NFIC, RELA, STAT3, IRF4, NFIC, JUNB, SPI1, JUN, ATF1, ATF2, BATF, BCL11A, EBF1, FOS, FOXM1, IKZF1, IRF4, MTA3, POU2F2, RELA, RUNX3, SPI1, STAT3, WRNIP1
JUN	no	no	ATF1, ATF2, BCL3, FOS, FOSL2, FOXM1, JUN, JUNB, MEF2A, MEF2C, NFIC, RELA, RUNX3, SPI1, STAT3, STAT5A
CREB1	no	no	SPI1
ETS1	no	no	FOS, FOSL2, JUN, RELA, RUNX3, SPI1, STAT3
VEGFA	no	yes	EBF1

SLC2A1	yes	no	FOSL2, NFIC
CITED2	yes	yes	ATF2, BATF, BCL11A, EBF1, FOS, JUN, RELA, RUNX3, SPI1, STAT
EDN1	yes	no	BCL3, FOS, FOSL2, JUN, RUNX3, STAT3
NDRG1	yes	yes	ATF1, FOS, JUN, MEF2A, RELA, RUNX3, STAT3
PLIN2	no	no	ATF1, ATF2, BATF, BCL3, EBF1, FOS, FOSL2, FOXM1, JUN, JUNB, MTA3, NFIC, RUNX3, STAT3, STAT5A
EPAS1	no	no	IKZF1, MEF2A, IKZF1, IKZF1, MEF2A, RELA, MEF2C, IRF4, FOXM1, MEF2A, IRF4, IKZF1, BCL3, EBF1, FOS, POU2F2, FOSL2, MTA3, POU2F2, POU2F2, MTA3, FOXM1, ATF2, BATF, BCL11A, BCL3, EBF1, FOS, FOSL2, FOXM1, IRF4, JUN, JUNB, MEF2A, MEF2C, MTA3, NFATC1, NFIC, POU2F2, RELA, RUNX3, SPI1, STAT3, STAT5A
FLT1	no	no	ATF1, BATF, EBF1, FOS, FOSL2, JUN, JUNB, MEF2A, RELA, RUNX3, SPI1, STAT3, STAT5A

Copyright clearance

SPRINGER NATURE LICENSE
TERMS AND CONDITIONS

May 31, 2019

This Agreement between Mr. Pubudu Nawarathna Mudiyanselage ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number	4598280655471
License date	May 29, 2019
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Nature Biotechnology
Licensed Content Title	Discovery and characterization of chromatin states for systematic annotation of the human genome
Licensed Content Author	Jason Ernst, Manolis Kellis
Licensed Content Date	Jul 25, 2010
Licensed Content Volume	28
Licensed Content Issue	8
Type of Use	Thesis/Dissertation
Requestor type	non-commercial (non-profit)
Format	print and electronic
Portion	figures/tables/illustrations
Number of figures/tables /illustrations	1
High-res required	no
Will you be translating?	no
Circulation/distribution	1,001 to 2,000
Author of this Springer Nature content	no
Title	Characterization of epigenetic changes and their connection to gene expression abnormalities in clear cell renal cell carcinoma
Institution name	n/a
Expected presentation date	Jul 2019
Portions	Figure 1A
Requestor Location	Mr. Pubudu Nawarathna Mudiyanselage 3600 ave du parc Apt 901 Montreal, QC H2X 3R2 Canada Attn: Mr. Pubudu Nawarathna Mudiyanselage
Total	**0.00 USD**

SPRINGER NATURE LICENSE TERMS AND CONDITIONS

May 31, 2019

This Agreement between Mr. Pubudu Nawarathna Mudiyanselage ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number	4585431225100
License date	May 10, 2019
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Nature Reviews Molecular Cell Biology
Licensed Content Title	Emerging roles of linker histones in regulating chromatin structure and function
Licensed Content Author	Dmitry V. Fyodorov, Bing-Rui Zhou, Arthur I. Skoultchi, Yawen Bai
Licensed Content Date	Oct 11, 2017
Licensed Content Volume	19
Licensed Content Issue	3
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	print and electronic
Portion	figures/tables/illustrations
Number of figures/tables /illustrations	1
High-res required	no
Will you be translating?	no
Circulation/distribution	1,001 to 2,000
Author of this Springer Nature content	no
Title	Characterization of epigenetic changes and their connection to gene expression abnormalities in clear cell renal cell carcinoma
Institution name	n/a
Expected presentation date	Jul 2019
Portions	Figure 1
Requestor Location	Mr. Pubudu Nawarathna Mudiyanselage 3600 Ave du parc Apt 901 Montreal, QC H2X 3R2 Canada Attn: Mr. Pubudu Nawarathna Mudiyanselage
Total	**0.00 USD**

Copyright Clearance Center

Confirmation Number: 11814452
Order Date: 05/12/2019

Customer Information

Customer: Pubudu Nawarathna Mudiyanselage
Account Number: 3001451250
Organization: Pubudu Nawarathna Mudiyanselage
Email: pubudu.nawarathna@mail.mcgill.ca
Phone: +94 716103126
Payment Method: Invoice

This is not an invoice

Order Details

Epigenetic Diagnosis & Therapy

Billing Status: **N/A**

Order detail ID: 71897004
ISSN: 2214-0832
Publication Type: Journal
Volume:
Issue:
Start page:
Publisher: Bentham Science Publishers Ltd.

Permission Status: ✔ **Granted**
Permission type: Republish or display content
Type of use: Thesis/Dissertation
Order License Id: 4586350040026

Requestor type	Not-for-profit entity
Format	Print, Electronic
Portion	chart/graph/table/figure
Number of charts/graphs/tables /figures	1
The requesting person/organization	Pubudu Nawarathna
Title or numeric reference of the portion(s)	Fig 1
Title of the article or chapter the portion is from	DNA Methylation: A Possible Target for Current and Future Studies on Cancer?
Editor of portion(s)	N/A
Author of portion(s)	Nussler, Andreas ; Sajadian, Sahar
Volume of serial or monograph	1
Issue, if republishing an article from a serial	1
Page range of portion	6
Publication date of portion	Apr 17, 2015
Rights for	Main product
Duration of use	Current edition and up to 5 years

Note: This item was invoiced separately through our **RightsLink service.** More info **$ 0.00**

This Agreement between Mr. Pubudu Nawarathna Mudiyanselage ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

License Number	4630920524787
License date	Jul 16, 2019
Licensed Content Publisher	Elsevier
Licensed Content Publication	Cell
Licensed Content Title	The Human Transcription Factors
Licensed Content Author	Samuel A. Lambert,Arttu Jolma,Laura F. Campitelli,Pratyush K. Das,Yimeng Yin,Mihai Albu,Xiaoting Chen,Jussi Taipale,Timothy R. Hughes,Matthew T. Weirauch
Licensed Content Date	Feb 8, 2018
Licensed Content Volume	172
Licensed Content Issue	4
Licensed Content Pages	16
Start Page	650
End Page	665
Type of Use	reuse in a thesis/dissertation
Intended publisher of new work	other
Portion	figures/tables/illustrations
Number of figures/tables /illustrations	1
Format	both print and electronic
Are you the author of this Elsevier article?	No
Will you be translating?	No
Original figure numbers	1A
Title of your thesis/dissertation	Characterization of epigenetic changes and their connection to gene expression abnormalities in clear cell renal cell carcinoma
Expected completion date	Jul 2019
Estimated size (number of pages)	140
Requestor Location	Mr. Pubudu Nawarathna Mudiyanselage 3600 avenue du parc Apt 901 Montreal, QC H2X3R2 Canada Attn: Mr. Pubudu Nawarathna Mudiyanselage
Publisher Tax ID	GB 494 6272 12
Total	0.00 CAD

Ethical approval for using patient samples

Patient samples were obtained from a previously published study by Scelo *et al*., 2014 through CAGEKID consortium. They recruited patients undergoing nephrectomy for suspected renal cancer during the period December 2008 to March 2011 at St James's University Hospital in Leeds, UK to the study after informed consent was obtained. Ethical approvals were obtained from the Leeds (East) Local Research Ethics Committee and the International Agency for Research on Cancer Ethics Committee. All sampling and clinical data collection was undertaken according to predefined standard operating procedures following guidelines from the International Cancer Genome Consortium.

Epigenome and gene expression profiling experiments of these samples were performed by McGill University Epigenome Mapping Centre (EMC). The data are available to download from EMC data portal.
(https://genomequebec.mcgill.ca/nanuqMPS/project/ProjectPage/projectId/9015)