University of California, Merced

A Systems Biology Approach to Epigenetic Gene Regulation

by

Stephen P. Wilson

A dissertation submitted in partial fulfillment for
the degree of Doctor of Philosophy
in the
Program for Quantitative and Systems Biology
School of Natural Sciences

Summer 2019

Committee in Charge:
Professor Fabian V. Filipp, Advisor
Professor Michael D. Cleary, Chair
Professor Ramendra Saha
Professor Zhong Wang

The Dissertation of Stephen Patrick Wilson is approved, and is acceptable in quality and form for publication on microfilm and electronically:

_____

Professor Fabian V. Filipp, Advisor

_____

Professor Michael D. Cleary, Chair

_____

Professor Ramendra Saha

_____

Professor Zhong Wang

University of California, Merced
2019

"Far better is it to dare mighty things, to win glorious triumphs, even though checkered by failure... than to rank with those poor spirits who neither enjoy nor suffer much, because they live in a gray twilight that knows not victory nor defeat."

Theodore Roosevelt

University of California, Merced

Abstract

Program for Quantitative and Systems Biology
School of Natural Sciences

Doctor of Philosophy

by Stephen P. Wilson

The ability to control when, and how much of the genetic code is being expressed is the underlying principle behind gene regulation. Control of gene production is able to influence a cell's phenotype by determining which structural components of the cell's observable traits (shape, growth, and behavior) are made. In multicellular organism's different cell types are able to arise from the same genetic code due to a difference in the patterns of genes being expressed. Essentially anywhere in the process of gene expression from transcription, RNA processing, translation, and post-translational modifications of the protein is subject to regulation. As transcription is the first step in the process of gene expression, it is the first level of regulation for influencing the cell phenotype. The actions of transcription factors, histone modifiers, and other proteins work together to influence RNA polymerase's ability to complete the process of transcription. The actions of transcription factors are able to influence transcription by controlling the ability of RNA polymerase to be recruited to the start of a protein coding region and histone modifiers can rearrange the histones of the chromatin causing entire regions of a chromosome to become exposed or sequestered. These transcriptional regulators are able to work in a combinatorial fashion with one another to either activate and/or repress wide repertoires of transcriptional targets. Mapping out a network of interactions between these transcriptional regulators in gene expression programs allows researchers to understand how each protein is able to influence the phenotype of the cell, and how mutations to any of these transcriptional regulators are able to drive the cell into a diseased state. In the case of cancer, changes in the mechanisms of gene regulation brought on by mutations to these transcriptional regulators may drive the cell's hyper proliferative state. With the creation of next generation sequencing researchers are now better able to define where regulation is taking place in the genome, and how much it is able to influence gene expression. This gives researchers the ability to build these gene regulatory networks and evaluate their impact on gene expression. The subsequent chapters of this dissertation are a reflection of my published work investigating the contribution of oncogenic processes to gene regulatory networks in cancer through the study of hyperactivating somatic mutation of a histone modifier, changes in transcription factor response element specificity, epigenetic regulation of transcription factor signaling, and a transcription factor coactivation network.

Acknowledgements

If I have seen further, it is by standing on the shoulders of Giants. It is my hope to thank everyone either in writing or in person for the support, dedication, and guidance into forging myself into who I am today.

I am thankful to Dr. Fabian Filipp who took me on first as an undergraduate intern into his lab, and then into his lab as a graduate student. I would be remiss to mention his mentorship into getting me into graduate school and achieving everything thus far. I would like to thank each member of my committee past and present: Dr. Michael Cleary, Dr. Ramendra Saha, Dr. Zhong Wang, and Dr. Emilia Huerta-Sanchez for their support during my foray in the graduate program. Your mentorship and honesty have made me a better scientist.

To the collaborators I had the pleasure of working with on the projects described in this work I thank you for your hard work, dedication, perseverance, and patience while I worked on my contributions. I would like to explicitly thank them: Dr. Tiffen, Dr. Gallagher, and Dr. Hersey of the Melanoma Immunology and Oncology group at the University of Sydney; Dr. Qi, and Dr. Fan of the Biochemistry department at the University of Maryland; and Dr. Sahgal of the Center for Molecular Oncology at the Queen Mary University in London.

To my fellow graduate students of the Systems biology and Cancer metabolism group: I have learned a great deal by working with you. The fellowship we developed in eating lunch together and including each other in our lives outside of lab are among the highlights of my time here at UC Merced.

Next, I would also like to thank my fellow graduate students of the RadioBio podcast group for helping to refine my science communication skills and putting up with me during the course of starting the club. Truly I have worked with the greatest minds of my generation. I look forward listening to future episodes.

Next, I would like to thank the efforts of Dr. Chadwick Burgdorf, Dr. Li Miao, and the support staff of the Counseling and Psychological services department here at the University of California, Merced. It is during our darkest moments that we must focus to see the light. I could not have achieved this without your council.

Finally to my family: thank you for your encouragement while pursuing this degree. It would not have been possible without all of you. Tracy, Laura, Aubrey, Brian 2.0, and Carol -I have tried my best to live up to your examples. Mom and Dad -I am truly grateful to everything you have done for me. Your hard work, love, support, guidance, and patience are what made this achievement possible.

Table of Contents

ABSTRACT V

ACKNOWLEDGEMENTS VI

CHAPTER ONE: INTRODUCTION 1

1.1 OVERVIEW OF GENE REGULATION CONTRIBUTING TO CANCER BIOLOGY 1
1.1.1 Oncogenes 1
1.1.2 Tumor Suppressor 3
1.2 EPIGENETICS 3
1.2.1 DNA methylation 3
1.2.2 Histone Modifications 4
1.2.2.1 Covalent histone modifications 4
1.2.2.1.1 Histone Acetylation 5
1.2.2.1.2 Histone Methylation 7
1.2.2.1.3 Histone Phosphorylation 9
1.2.2.1.4 Histone ubiquitination 10
1.2.2.1.5 Histone Sumoylation 10
1.2.2.1.6 Histone ADP-ribosylation 11
1.2.2.1.7 Histone proline isomerization 11
1.2.2.2 ATP dependent chromatin remodeling Complexes 12
1.3 TRANSCRIPTION FACTORS IN GENE REGULATION AND CANCER 12
1.4 IN SUMMARY 13
1.5 REFERENCES 14

CHAPTER 2: METHODOLOGY 32

2.1 STUDYING PHENOTYPIC SHIFTS CAUSED BY TRANSCRIPTIONAL REGULATORS IN CANCER 32
2.2. OVERVIEW OF APPROACH TO PRIMARY RESEARCH 32
2.2.1 Using cancer cell lines as a model system 32
2.2.2 The Cancer Genome Atlas Program 32
2.3 GENOMICS TECHNIQUES 34
2.3.1 ChIP-seq in order to identify genomic locations 34
2.3.2 Bisulfite DNA methylation Sequencing for identification of DNA methylation events 35
2.3.3 Genome alignment 36
2.3.4 Peak Calling 36
2.3.5 Motif Analysis 36
2.3.6 Gene set enrichment analysis 37
2.4 TRANSCRIPTOMICS 37
2.4.1 Microarray 37
2.4.2 RNA-seq 38

2.5 References 39

CHAPTER THREE: SOMATIC COPY NUMBER AMPLIFICATION AND HYPERACTIVATING SOMATIC MUTATIONS OF EZH2 CORRELATE WITH DNA METHYLATION AND DRIVE EPIGENETIC SILENCING OF GENES INVOLVED IN TUMOR SUPPRESSION AND IMMUNE RESPONSES IN MELANOMA. 48

Methods 47
Results 47
Discussion 52
Conclusion 55
References 55

CHAPTER FOUR: REFINEMENT OF THE ANDROGEN RESPONSE ELEMENT BASED ON CHIP-SEQ IN ANDROGEN-INSENSITIVE AND ANDROGEN RESPONSIVE PROSTATE CANCER CELL LINES. 61

Methods 60
Results 61
Discussion 69
Conclusion 70
References 70

CHAPTER FIVE: THE HISTONE DEMETHYLASE KDM3A REGULATES THE TRANSCRIPTIONAL PROGRAM OF THE ANDROGEN RECEPTOR IN PROSTATE CANCER CELLS. 76

Results 75
Discussion 81
Methods 84
References 87

CHAPTER SIX: A NETWORK OF EPIGENOMIC AND TRANSCRIPTIONAL COOPERATION ENCOMPASSING AN EPIGENOMIC MASTER REGULATOR IN CANCER. 93

Results 92
Discussion 95
Conclusion 98
Methods 98
References 99

CHAPTER SEVEN: CONCLUSIONS 104

7.1 Summary of contributions 104
7.2 Somatic Copy Number Amplification and Hyperactivating Somatic Mutations of EZH2 Correlate with DNA Methylation and Drive Epigenetic Silencing of Genes Involved in Tumor Suppression and Immune Responses in Melanoma 104
7.2.1 Summary of results 104

7.2.2 Applying cancer subtyping into the development of precision medicine 105
7.2.3 Future perspectives on SKCM cohort studies 106
7.3 REFINEMENT OF THE ANDROGEN RESPONSE ELEMENT BASED ON CHIP-SEQ IN ANDROGEN-INSENSITIVE AND ANDROGEN RESPONSIVE PROSTATE CANCER CELL LINES. 107
7.3.1 Summary of results 107
7.3.2 Applying response element refinement strategies into motif databases 109
7.3.3 Future perspectives on response element alterations in cancer 110
7.4 THE HISTONE DEMETHYLASE KDM3A REGULATES THE TRANSCRIPTIONAL PROGRAM OF THE ANDROGEN RECEPTOR IN PROSTATE CANCER CELLS. 111
7.4.1 Summary of results 111
7.4.2 Applying multiomic strategies to the regulation of transcription 115
7.4.3 Future Perspectives on multiomic analysis of KDM3A 116
7.5 A NETWORK OF EPIGENOMIC AND TRANSCRIPTIONAL COOPERATION ENCOMPASSING AN EPIGENOMIC MASTER REGULATOR IN CANCER. 116
7.5.1 Summary of results 116
7.5.2 Applying interactomics for studying biological pathways 117
7.5.3 Future Perspectives on the gene regulatory network of KDM3A 119
7.6 CONCLUSIONS 119
7.7 REFERENCES 104

Chapter One: Introduction

1.1 Overview of gene regulation contributing to Cancer biology

The World Health Organization defines cancer as an abnormal cell growth that can arise from any normal tissue within the human body. All cells that have undergone the transformation into cancer gain eight hallmark characteristics of cancer: Cell growth without the proper stimulation from their environment, independence from anti-growth signals from their environment, developing mechanisms for ignoring programmed cell death, the ability to undergo limitless cell division, gaining the ability to promote the grow of new blood vessels, the ability to invade tissue in other locations in the human body, changing the rules for how energy is utilized within the cell, and avoiding detection from the immune system[1, 2]. The progression of these traits from a healthy cell to a cancerous one occurs through the accumulation of multiple alterations that are either genetic or epigenetic in nature and can cause the physiological circuitry of the cell to go haywire. Genetically speaking these alterations called mutations can create the activation and inactivation of two classes of genes found in cancer biology: oncogenes and tumor suppressors. Alterations in the epigenetic regulation of genes within a cell are achieved without making changes to the DNA sequence itself [3]. The current state of the field for cancer biology is building an understanding of the overlap and intersection of both genetic and epigenetic modifications for the development of cancer. It is becoming well accepted that the transformation of a cell's phenotype to a cancerous one is achieved both through changes in genetic and epigenetic regulation.

1.1.1 Oncogenes

An oncogene is defined as a gene that assists a cell with growth but then is altered through changes in its coding sequence or is expressed at high levels within the cell [4]. Oncogenes give us insight into the molecular mechanism for how a cell is able to acquire its hallmark characteristics during the formation of cancer [4]. While oncogenes were initially discovered as a product generated by tumor forming viruses it is believed that most human cancers are not caused by viruses but may arise due to exposure of cancer-causing agents [5]. Conversion of a gene into an oncogene can occur through three basic methods of activation:

- a mutation within the coding region of the gene or its regulatory region that may cause an increase in the protein activity or a loss in its ability to be regulated.
- An increase in the amount of the protein due to deregulation of its expression, mRNA stability, or a duplication of the gene within the cell.
- A chromosomal translocation that either places the gene in a region with higher expression or a fusion with a 2nd gene, which increases oncogenic activity [6].

Most cancer treatments available today target oncogenic proteins [6]. It is the goal of oncogenomics, a branch of genomics, to understand the structure, function, and mapping of cancer associated genes through the use of bioinformatics [7]. It is essential to better

understand the mechanisms of oncogene activation, as well as the changes in the gene's function during oncogenesis.

1.1.2 Tumor Suppressor

A tumor suppressor is defined as a gene that acts as a deterrent to the progression of cancer. Often a tumor suppressor is mutated resulting in the loss of its ability to function or is changed to be expressed minimally in the cell. In terms of a cell becoming more cancerous it is believed that the loss of tumor suppressor genes is more damaging than the activation of oncogenes [8]. The canonical functions of tumor suppressors are described as repressors of the cell cycle, the response mechanism to DNA damage, initiators of programmed cell death, keeping cells adhered to their extracellular matrix, and repairing DNA damage [9].

1.2 Epigenetics

There are two classifications of genes that are important in the development of cancer: the activation of oncogenes and the deactivation of tumor suppressors. A classical way at looking at how oncogenes and tumor suppressors become activated or deactivated is that their genetic code is changed to affect their function and gene expression. Current understanding is that there may also be some epigenetic changes that may interact with oncogenes and tumor suppressors in facilitating a change of their behavior in the cell [10, 11]. Conrad Waddington was the first to define epigenetics as heritable transition in a cell's phenotype that were autonomous of changes in the sequence of DNA [3]. A brief observation into the linguistics of epigenetics shows that the prefix epi suggests a factor to cause cells to express themselves differently in addition to traditional genetic basis. Adrian Bird later describes epigenetics as "the structural adaptation of chromosomal regions so as to register, signal or perpetuate altered activity states"[12]. Examples of epigenetic mechanisms that alter gene expression without changing the underlying DNA sequence are DNA methylation and histone modification.

1.2.1 DNA methylation

Methylation of Cytosine residues in CpG islands is associated with regulation of transcription, chromatin structure, X chromosome inactivation, genomic imprinting, and chromosome stability[13]. Normally in somatic cells DNA methylation patterns are passed down to daughter cells exactly as they are found in the parental cell before cell division [14]. In cancer, changes of the methylation state of CpG islands are often changed during malignant transformation of the cell where normally unmethylated CpG islands become startlingly methylated resulting in a decrease of expression [13, 15]. It is estimated that cancer cell genomes contain anywhere between 20-50% less DNA methylation marks than in normal cells [16-19]. These methylated regions are not limited to promoter and enhancer regions but the gene body too [15]. Epigenetic regulation of tumor suppressors and oncogenes through DNA methylation find that in cancer cells CpG

islands found within tumor suppressor gene promoters are hypermethylated, while conversely the CpG islands found within oncogene promoter regions are hypomethylated [16]. This epigenetic regulation favors transcriptional activation of oncogenes and transcriptional repression of tumor suppressor genes.

DNA methylation is carried out by DNA methyltransferases (DNMTs) with three prominent ones found in cancer: DNMT1, DNMT3a, and DNMT3b. DNMT1 sees hemimethylated DNA created during DNA replication and methylated CpG marks on the newly synthesized DNA strand so that it matches with the parental strand [20]. While DNMT3a and DNMT3b are able to carry out a similar methylation function for newly synthesized DNA strands their main function is to establish new methylation marks during embryogenesis [21]. Methyl-binding proteins seek DNA methylation and can assist in recruiting histone modifying enzymes to alter nearby chromatin [22].

1.2.2 Histone Modifications

Within the nucleus DNA is wrapped around proteins called histones forming a structure known as nucleosomes and depending on how tight the DNA is wrapped around the histones gene expression is affected. Histone modifications were first proposed to influence regulation of transcription back in 1964 [23]. Since then it's been confirmed that these histone marks influence transcription by changing between open and closed states: histones wrapped around DNA in an open state grants access to the gene body to be expressed, while on the contrary a compact state blocks access to the transcriptional machinery that facilitate gene expression. Remodeling the chromatin can change these states. Chromatin remodeling changes the packaging of the nucleosomes allowing regions of transcriptional regulation to be exposed or hidden. Chromatin remodeling is typically carried out by covalent histone modifications or by ATP dependent chromatin remodeling complexes.

1.2.2.1 Covalent histone modifications

Protein complexes that catalyze the addition or removal of specific chemical groups to the histone carry out these covalent histone modifications. Depending on the chemical modifications added to the histone the affinity between DNA and the histone can change to facilitate a tightly wrapped nucleosome or a loosely wrapped nucleosome. Researchers have pieced together which covalent modifications are associated with the open and closed states of nucleosomes through what is called the histone code hypothesis [24]. Similar to how changes in the genetic code may shift a cell towards a cancerous phenotype, a shift in the histone code is able to push a cell into a neoplastic one. Changing covalent modifications on the histones surrounding the gene body of oncogenes and tumor suppressors would alter the accessibility of these genes and therefore influence their gene expression levels. Global studies of histone marks in healthy and cancerous cells shows shifts in the levels of some histone marks [25-29]. The histone marks typically found on histones are acetylation, methylation, phosphorylation,

ubiquitination, sumoylation, ADP-ribosylation, deimination, and proline isomerization [22, 30-37]. The alterations of these covalent histone marks in gene regulation are discussed in the subsequent sections below.

1.2.2.1.1 Histone Acetylation

Histones are able to be acetylated due to a transfer of an acetyl group from acetyl-coA [38]. The levels of histone acetylation correlate with the availability of acetyl-CoA: high amounts of histone acetylation marks when acetyl-CoA is plentiful and low amounts of histone acetylation when it is scarce [39, 40]. Within the histone code the process of adding acetyl groups to lysine residues found on histone tails neutralizes the positive charge of lysine resulting in a weakened electrostatic interaction between the histone and the negatively charged DNA. Conversely if an acetyl group is removed from a lysine residue its overall positive charge is restored allowing for a stronger electrostatic interaction between the histone and the DNA. These marks are maintained by two definitive groups of enzymes called Histone acetyltransferases (HATs) and histone deacetylases (HDACs).

The acetylated open state of chromatin results in transcription factors and other chromatin modifiers being able to be recruited and is experimentally associated with actively transcribed genes [41-43]. As a result, researchers began to build an interaction map of HAT proteins with the transcription factors they interact with [44]. As a result, a study in the coactivation preferences of HAT protein complex p300/CBP with transcription factors identified hundreds of these interactions raising the issue of how essential is HATs to the building of transcriptional machinery [45]. This led to the theory that HAT cooperation with transcription factors may assist HAT in targeting nearby histones for acetylation to further aid in the activation of transcription [46]. Besides adding acetylating nearby histones HAT may play an additional role of regulating transcription by adding acetyl groups to the transcription factors with which they interact. Acetylation of key lysine residues in the primary structure of transcription factors STAT3, c-Myb, STAT3, and E2F1 increased the transactivating ability of these transcription factors to recognize their binding sites in DNA [47, 48]. However, it was also observed that HAT acetylation of FOX01 resulted in a loss of its overall positive charge and the transcription factor's ability to recognize its DNA binding sites [49]. Either way protein acetylation may affect signal transduction of transcription factors.

HDACs restore the positive charge of the lysine residue and are associated with inducing transcriptional inactivation. There are 18 likely human HDAC proteins in humans that utilize zinc or NAD+ mechanisms to remove acetyl groups from lysine substrates [50]. It is important to take into consideration that HDACs are also able to control the acetylation status of histone tail residues and are also able to add acetyl groups to cytoplasmic proteins and transcription factors [51]. Further repression of gene expression can be achieved through HDAC interaction with DNA methyltransferases [52, 53]. Also similar to HATs researchers began summarizing non-histone proteins targeted by HDACs and found several transcription factors in that list [54]. HDACs may interact

directly with the transcription factor or through a coregulator protein may interact as an intermediary to influence transcription [55-59]. Additionally, HDACs may use transcription factors to help identify nearby histones for deacetylation to further facilitate transcriptional repression [55, 56].

In order to understand the role histone acetylation played in the development of cancer researchers began comparing the overall levels of histone acetylation between healthy and cancerous cells. What they found in some models was that the levels of acetylation decreased as the tumor growth progresses resulting in a decrease in the total amount of acetylation found on the histone protein tail [27, 60]. Further investigation found increased acetylation specifically H2AK5, H3K9, H3K56, H4K5, and H4K8 across several cancer types [61-63]. Conversely decreased or loss of acetylation was found at H3K9, H3K14, H4K12, and H4K16 was found in others [63, 64]. As HATs and HDACs are specific towards the particular histone tail residues they modify, it is important to focus on the changes made to this precise class of proteins to evaluate their role in changing the epigenome in the development of cancer.

Across cancer cell types different HAT members are altered in some way resulting in their activity being strongly correlated to the development of cancer [65-70]. Different HAT proteins are found mutated, truncated, or completely deleted across lymphomas and various squamous cell carcinomas [69-77]. However, their specific role in cancer may be dependent on the genomic alteration and on the protein. For example, HAT CBP/300 is often found deleted or mutated resulting in a loss of function where it is believed to be a tumor suppressor in some cancer types [69, 72, 78]. Conversely p300 is over expressed in other cancers where its regulation of fatty acid synthase, and lipid metabolism supports cancer growth [55, 79-81]. Regardless HATs are found to interact with various transcription factors in regulating key biological processes relevant to cancer: DNA repair, cell growth, senescence, differentiation, and apoptosis [82].

As histone deacetylation typically results in a closed chromatin state HDACS in cancer are typically expected to be recruited for facilitating abnormal gene silencing [83]. While somatic mutations of HDACS are not prominent in cancer, their expression appears to be increased [83]. HDACs appear to use several molecular mechanisms in promoting the progression of cancer. First and foremost is the dysregulation of the cell cycle by facilitating the removal of acetyl marks from the promoters of p21, p27, and p57 to negatively regulate their expression [84-86]. Besides repressing the expression of cell cycle inhibitors HDACs are found to control transition of the cell through the cell cycle at G1, and the G2/M phase [87, 88]. Investigation into this effect found that knockdown of HDAC5 resulted in a down-regulation of proteins that normally progress the cell cycle: cyclin D and cyclin dependent kinases 2,4, and 6 [89]. It is possible that a similar form of regulation exists between HDAC5 and cyclin D/ CDK2/4/6 when compared to the transcriptional switch of HDAC10 deacetylating the histones near the promoter of an inhibitor of cyclin A2 [87]. Another way HDACs are able to promote the progression of cancer is through the mediation of DNA damage pathways in the genome. One of the mechanisms HDACs are able to achieve this is by facilitating non homologous end

joining by removing acetyl marks from H3K56 and H4K16 residues [90]. Loss of HDAC1/2 expression resulted in a decrease in the function of the ATM-mediated DNA damage response-signaling pathway [91]. However, it should be noted that HDACs with their ability to catalyze the removal of acetyl marks on non-histone proteins HDACs are able to interact acetylated oncogenes to restore their function [92].

1.2.2.1.2 Histone Methylation

The arginine and lysine residues of the histone proteins are able to be methylated and can alter chromatin structure to both and open and closed states contributing either to activating or repressing transcription [93]. Histone methylation differs from histone acetylation due to the fact that it doesn't change the charge of the histone. Histone methylation is carried out by histone methyltransferases (HMTs) and were first discovered in the 1960s [23, 94]. In order to methylate the histone HMTs require the metabolite S-adenosylmethionine (SAM) to donate a methyl group to the histone [95]. Methionine is a required precursor to SAM that is produced through the methionine salvage cycle or the folate cycle [95]. If there is a reduction in the availability of methionine through defects in its metabolic pathways directly affects SAM levels resulting in a decrease in histone methylation [96-98]. For decades histone methylation was regarded as a permanent post-translational modification until histone demethylases (HDMs) were uncovered in 2004 and showed the methylation status of histones could be rewritten [99]. HDM requires the TCA cycle intermediate alpha-ketoglutarate as a substrate for the removal of methyl marks on histones [100]. Alpha-ketoglutarate is enzymatically produced from isocitrate by Isocitrate dehydrogenase proteins and through conversion of glutamine [101-104].

Of all the methylation patterns of histone proteins studied, the methylation of the lysine residues of the histone tail is the most clearly defined. While there are many lysine residues on a histone tail that can be methylated, the ones examined the most are H3K4, H3K9, H3K27, H3K36, H3K79, and H4K20. Unlike the process of histone acetylation the process of adding methylation marks to these histone residues may either weaken or strengthen the electrostatic interaction between the histone and DNA depending on which residue is methylated. For example histone methylation of residues H3K4, H3K36, and H3K79 result in weakening the interaction between the histone and DNA allowing accessibility of the transcriptional machinery [105]. In contrast the methylation of H3K9, H3K27, and H4K20 result in increased interaction between the histone and DNA resulting in repression of transcription [105]. Adding or removing these methyl residues on the histone allows this aspect of the histone code to play a role in determining gene expression, maintaining the genome, maturation, differentiation, and the cell cycle [106]. That is why recurrent mutations of lysine methyltransferases (KMTs) and lysine demethylases (KDMs) are prevalent across several cancer types [107-110]. Although loss of function mutations in KMTs and KDMs are more frequently reported than gain of function mutations in cancer may suggest that this class of epigenetic regulators function

as a type of tumor suppressor; this may not be the case [111]. Overexpression of these proteins in cancer appear to be able to stimulate cell proliferation [112, 113].

KMTs are also able to target non-histone proteins as well resulting in an addition of up to three methyl groups to exposed lysine residues [114-120]. KMTs are able to regulate transcription factors through a variety of mechanisms. The first being that methylation of transcription factors trigger degradation [121, 122]. Secondly lysine methylation was determined to play a role in cellular localization when newly methylated transcription factors migrated into the nucleus [123, 124]. Methylation of transcription factors has the ability to either inhibit or promote their DNA binding [125-129]. KDM4A assembles into a complex with the ER in ER-positive breast cancer and with the AR in prostate cancer functioning as a transcriptional co-activator in both examples [130].

Arginine methylation is carried out by protein arginine N-methyltransferases (PRMTs) and are grouped based on the type of protein the target arginine is found in. Types one and two PRMT are able to add methyl groups exclusively to arginines found on histone proteins, and PRMTs types three and four are able to add them only to arginine residues of non-histone proteins [131]. Similar to the methylation of lysine residues on the histone, arginine methylation may either weaken or strengthen the electrostatic interaction between the histone and DNA depending on which residue is methylated. Methylation of H3R2, H3R8, H3R17, and H4R3 residues result in weakening the interaction between the histone and DNA allowing accessibility of the transcriptional machinery [132-134]. In contrast methylation of H3R2, and H4R3 result in an increase of transcriptional repression due to increased interaction between the histone and DNA [134-136]. Often lysine and arginine methylation states oppose each other in regulating gene expression. Transcription by blocking H3K4 methylation and inversely H3K4me3 prevents methylation of H3R2 [137, 138]. Arginine demethylation was first mentioned in 2007 [139]. It is not as well understood as lysine demethylation. Currently PAD4 and JMJD6 are the only two histone arginine demethylases known [139]. Some histone lysine demethylases containing the JmjC domain are also able to carry out arginine demethylation [140, 141]. An example of this is lysine demethylase JMJD1B which is able to remove methyl marks from H3K9 and H4R3 [142]. Besides methylation arginine residues are also susceptible to diminution by peptidyl arginine deiminases converting the residue to citrulline [143]. The process of reverting citrulline back to arginine isn't well studied [144]. Deimination marks are able to repress expression of target genes by blocking the activating methylation marks that would normally be added to H3R2, H3R8, H3R17, and H3R26 [30]. Conversely deaminations of arginine residues are avoided if these residues are demethylated [30].

At one point more than 100 different non-histone proteins were reported to contain methylated arginine in their primary structure [145]. H4R17 is linked to gene activation and coactivation with the nuclear hormone receptor [146]. H4R17 methylation occurs simultaneously with recruitment of ESR1 [147]. Histone arginine methylation of p53 silences its transcriptional function [148]. It can also target arginine residues of tumor suppressors directly [149]. Up regulation of PRMT expression and dysregulation

are observed in several cancer types [150-156]. H4R3 and H3R8 methylation represses tumor suppressor gene ST7 and CDKN2A [55, 157-159].

1.2.2.1.3 Histone Phosphorylation

The process of adding a negatively charged phosphate group to serine, threonine, and tyrosine residues on histone proteins results in an overall negative charge of the histone and is repelled from interacting with the overall negative charge of DNA resulting in a change in the overall structure of the local chromatin environment [160]. The phosphorylation status of the histone protein is controlled by the actions of two classes of proteins: kinases that add phosphoryl groups to the histone and phosphatases that remove them. While normally kinases and phosphatases are thought of as participating exclusively in the signaling transduction process in the cytoplasm, but they can also be observed carrying out the phosphorylation of histones in the nucleus [160-162].

Most studies into the effects of phosphorylation of histone protein tail residues have looked at how the addition of a phosphoryl group affects other covalent modifications on nearby residues. For example increased phosphorylation of H3S10, H3T11, and H3S28 resulted in an increase of acetylation at H3K9 and H3K14 residues resulting in chromatin shifting towards an open state [163-168]. This type of correlation of histone marks working towards transcriptional activation also occurs with the phosphorylation of H3T11 and H3T6 in response to androgen receptor dependent gene activation showed removal of repressive methylation marks on H3K9 by a KDM, and with phosphorylation of H3T6 preventing the removal of the transcriptionally activating methylation marks on H3K4 by a KDM [169, 170]. A possible explanation of this phenomenon is that one residue is able to influence its neighboring residues repeatability of covalent modifications. It could also be that cooperation between two histone residues working in combination are able to influence transcription such as the phosphorylation of H3S28 simply interacts with H3K27 acetylation in stimulating transcriptional activation [171, 172]. While it may seem that histone phosphorylation primarily favors an open chromatin state, the phosphorylation of H3S10 is also associated with condensation of chromatin during mitosis and meiosis [173-176].

Besides influencing gene regulation histone phosphorylation is able to oversee cellular responses to DNA damage. Phosphorylation of S139 of H2AX is one of the first steps of recognizing a double stranded break in DNA [177-179]. The Phosphorylated H2AX subunits are found several kilobases on either side of the break and serves as a binding site for the protein MCD1 that recruits DNA repair proteins [4, 55, 179-182]. This mechanism of identifying DNA damage is used in nonhomologous end joining, homologous end joining, and DNA base excision repair [177, 183, 184]. Phosphatase removal of the phosphoryl mark from S139 of H2AX is essential for the cell cycle to progress, as phosphorylation of this mark is also induced upon death receptor activation and can result in apoptosis [185, 186].

1.2.2.1.4 Histone ubiquitination

Ubiquitin is a 76 amino acid polypeptide containing seven lysine residues that was first observed being ligated to histone protein H2A's K199 residue back in 1975 [187, 188]. The process of adding ubiquitin to lysine residues involves the coordination of three enzymes: ubiquitin-activating enzymes (E1s), ubiquitin-conjugating enzymes (E2s), and ubiquitin ligases (E3s) [189-191]. Removing ubiquitin is achieved through a class of isopeptidases called deubiquitinases [192]. Ubiquitin is able to influence chromatin structure by creating a steric hindrance between itself and other histone residues affecting whether or not the transcriptional machinery has access to DNA [193]. Histones marked with ubiquitin are also able to serve as a protein binding site for transcriptional regulators [194, 195].

Ubiquitination of histone subunits unlike other proteins are not marked for degradation but for transcriptional regulation [196]. While all histone proteins (H1, H2A, H2B, H3, and H4) are observed to be capable of being ubiquitinated, the biological function of some (H1, H3, and H4) are not fully understood [196-201]. Of the remaining histone proteins H2A appears to be the most abundantly ubiquitinated with 10% of all histones having the ubiquitin mark compared to H2B having only 1% on all nucleosomes [202-205].

Not only are does it appear that there are differences in the abundances of H2A and H2B ubiquitination, but also in the function the ubiquitin mark has on each protein. For example ubiquitination of H2A are enriched in the promoter of polycomb target genes and serves as the recruitment site for polycomb repressive complex 1 mediated transcriptional silencing [206, 207]. In contrast H2B ubiquitination stimulates transcriptional activation by disrupting the structure of the chromatin [208-211]. The mechanism of H2B ubiquitination is also used for controlling the processes of histone disassembly and reassembly in chromatin remodeling [212-216].

Addition of ubiquitin to non-histone proteins induces their degradation through proteasomes, endoplasmic reticulum mediated pathways, and the lysosomes [217, 218]. This allows the ubiquitin mark to regulate gene expression through ubiquitin mediated degradation of transcriptional regulators. An example of this is the ability of ubiquitination being able to degrade HATs, HDACS, HMTs, and HDMTs [219-226]. Ubiquitin plays a role in regulating transcription through proteasome destruction of transcription factors or proteasome independent mechanisms [227]. Ubiquitination of p53 activates its proteasomal degradation [228].

1.2.2.1.5 Histone Sumoylation

Sumoylation is when the small ubiquitin-like modifier (SUMO) protein is covalently attached to its target. Unlike ubiquitination sumoylation blocks proteasomal degradation when bound to a lysine residue [229, 230]. Sumoylation happens mainly on the histone H4 subunit but may be found on all the core histone proteins [231]. Histone sumoylation is associated with transcriptional repression by facilitating a closed chromatin state [32, 231, 232].

Most of the studies on the effects of sumoylation to chromatin have to deal with what happens when SUMOs are covalently added to HDACS, HDM, HMT, and transcription factors [32, 233-243]. The addition of the SUMO group appears to influence target recognition for these proteins. For example when the sumoylation of HDAC1 and HDAC4 was blocked their histone deacetylation activity was reduced [244, 245]. Similarly when KDM6B is sumoylated it was reported to occupy its target genes more often [233]. Contrary to epigenetic factors sumoylation of transcription factors lose the ability to activate transcription due to the SUMO group altering their protein stability [246-249]. Sumoylation of transcription factors can serve as a signal to recruit additional chromatin regulatory proteins [250-252]. For example sumoylated ELK1 is able to recruit HDAC2 [253].

1.2.2.1.6 Histone ADP-ribosylation

ADP-ribosylation is the transfer of an ADP-ribose group from nicotinamide adenine dinucleotide (NAD+) to an amino acid side chains [254]. This is catalyzed through the action of Poly-ADP-ribose-polymerase (PARP) proteins and is reversible through the action of poly-ADP-ribose-glycohydrolase (PARG) proteins [255]. The process of ADP-ribosylation may result in a mono-ADP-ribosylation (MAR) chain or a poly-ADP-ribosylation (PAR) chain [256]. MARylation is the ADP-ribosylation mark commonly found in the cytoplasm, and typically outnumbers the PARylation mark that is commonly found in the nucleus [257, 258]. In the absence of NAD+, PARP proteins contribute to the closed state of chromatin; however with the addition of NAD+ to PARP bound nucleosomes, the chromatin is able to lead to the open state of the chromatin [259, 260]. The ADP-ribosylated state of histones is reversible by degradation of the PAR marks through PARG [55, 259, 260]. It is theorized that ADP-ribosylation may affect chromatin structure by competing with other histone modifications for the same amino acid or by creating steric hindrance [31]. The H4 histone protein is preferentially ADP-ribosylated when it is hyperacetylated [261]. ADP-ribosylation may co-occur with histone phosphorylation, but it appears that the size of the ADP-ribose group may block kinase accessibility to its target site [262, 263].

1.2.2.1.7 Histone proline isomerization

Proline isomerization is when a peptide-propyl isomerase (PPIase) catalyzes a proline residue to shift between its cis and trans isomers [264]. Due to H3P38 being close to H3K36 the isomerization at P38 is able to influence H3K36's methylation [33, 265]. The H3P38 is in the cis position the histone tail is pulled closer to the DNA resulting in a steric hindrance that crowds the tail preventing histone methyltransferases from recognizing it. When H3P38 is in the trans position H3K36 is able to be methylated [266].

1.2.2.2 ATP dependent chromatin remodeling Complexes

ATP dependent chromatin remodeling complexes utilize energy derived from hydrolysis of ATP for moving, ejecting, or restructuring nucleosomes. ATP dependent chromatin remodelers oversee almost every single chromosomal mechanism and their deregulation leads to cancer [267, 268]. All remodeler ATPases can be broken down into four groups: imitation switch (ISWI), chromodomain helicase DNA binding (CHD), switch/sucrose non-fermentable (SWI/SNF) and INO80 [269-271]. These four groups are able to influence chromatin organization through assisting in the assembly of nucleosomes, controlling accessibility of chromatin, and editing nucleosomes. After DNA replication members of the ISWI and CHD help form the nucleosome octamers and spreading them apart at established lengths [272-276]. Members of the SWI/SNF family of remodelers are able to make chromatin more accessible allowing transcription activators or repressors to recognize their DNA binding sites [277]. Members of the INO80 remodeling group are able to remove and replace a histone and replace it with either a canonical or a variant histone protein [278-280]. Chromatin remodelers are able to target histones through the modifications added to the proteins [281]. Targeting may be able to function through the unified actions of histone modifications, transcriptional activators, and transcriptional repressors [282].

1.3 Transcription factors in gene regulation and cancer

Transcription factors are a nomenclature for types of protein that regulate the pace of genetic information being transcribed from DNA to messenger RNA through binding to a specific DNA sequence [283-287]. The DNA binding domain of a transcription factor is their distinguishing feature as its structure defines its binding site recognition pattern and serves as the mechanism by which transcription factors are grouped [288-291]. Once bound to their DNA sequence, transcription factors may influence the initiation of transcription by either recruiting RNA polymerase (activators) or by blocking the recruitment of RNA polymerase (repressors) [292-296].

Transcription factors are able to regulate through control of their synthesis, nuclear localization, accessibility of the DNA binding site, and the availability of other cofactors. Transcription factors are able to regulate their own transcription by binding to the DNA of their own gene and down regulate itself through a negative feedback loop [297]. While proteins are transcribed in the nucleus, they are translated in the cell's cytoplasm. In order for a transcription factor to return back to the nucleus they must receive a nuclear localization signal, which may mean binding with a ligand, becoming phosphorylated, or interacting with other proteins to form a complex [298-300].

Due to transcription factors being able to regulate gene expression affecting a variety of cellular processes including but not limited to differentiation, development, intracellular signaling, the cell cycle, and metabolism [301-303]. Many human diseases are attributed to mutations in transcription factors affecting their ability to regulate these

cellular processes [304]. Transcription factors in the context of cancer are able to play a role as a tumor suppressor or an oncogene [305].

1.4 In summary

During the course of this introductory chapter the reader is introduced to the idea that gene expression can be regulated through the actions of transcription factors binding to their recognition sites in DNA serving either as activators or repressors of transcription. While these binding sequences appear plentiful throughout the genome only a small fraction are functional [306, 307]. This behavior of selective binding to recognition sites may be due to epigenetic marks controlling chromatin accessibility [308-310]. Conversely sometimes transcription factors themselves are able to recruit chromatin modifying proteins to alter chromatin structure [311]. It is speculated that epigenetic modifications are not randomly distributed across the genome and that they are able to target specific regions through the use of transcription factors, noncoding RNAs, and histone remodelers [312]. Whichever the scenario investigation into the functional impact of coordinated action between histone modifying enzymes and transcription factors may lend to its target specificity or at the very least create accessibility for DNA binding of specific transcription factors [313-317]. Within the context of cancer; mutations in the functional domains of either transcription factors or the histone modifying proteins may deregulate the transcriptional program it oversees. The subsequent chapters of this dissertation explore the details of this phenomenon.

1.5 References

1. Hanahan, D. and R.A. Weinberg, The hallmarks of cancer. Cell, 2000. 100(1): p. 57-70.
2. Hanahan, D. and R.A. Weinberg, Hallmarks of cancer: the next generation. Cell, 2011. 144(5): p. 646-74.
3. Dupont, C., D.R. Armant, and C.A. Brenner, Epigenetics: definition, mechanisms and clinical perspective. Semin Reprod Med, 2009. 27(5): p. 351-7.
4. Becker, W.M., The world of the cell. 7th ed. 2009, San Francisco: Pearson/Benjamin Cummings. xxviii, 791, 89 p.
5. Parkin, D.M., The global health burden of infection-associated cancers in the year 2002. Int J Cancer, 2006. 118(12): p. 3030-44.
6. Yokota, J., Tumor progression and metastasis. Carcinogenesis, 2000. 21(3): p. 497-503.
7. Strausberg, R.L., et al., Oncogenomics and the development of new cancer therapies. Nature, 2004. 429(6990): p. 469-74.
8. Weinberg, R.A., The biology of cancer. Second edition. ed. 2014, New York: Garland Science, Taylor & Francis Group. xx, 876, A 6, G 30, I 28 pages.
9. Sherr, C.J., Principles of tumor suppression. Cell, 2004. 116(2): p. 235-46.
10. Novak, K., Epigenetics changes in cancer cells. MedGenMed, 2004. 6(4): p. 17.
11. Banno, K., et al., Epimutation and cancer: a new carcinogenic mechanism of Lynch syndrome (Review). Int J Oncol, 2012. 41(3): p. 793-7.
12. Bird, A., Perceptions of epigenetics. Nature, 2007. 447(7143): p. 396-8.
13. Robertson, K.D., DNA methylation and human disease. Nat Rev Genet, 2005. 6(8): p. 597-610.
14. Bird, A., DNA methylation patterns and epigenetic memory. Genes Dev, 2002. 16(1): p. 6-21.
15. Baylin, S.B. and P.A. Jones, A decade of exploring the cancer epigenome - biological and translational implications. Nat Rev Cancer, 2011. 11(10): p. 726-34.
16. Esteller, M., Aberrant DNA methylation as a cancer-inducing mechanism. Annu Rev Pharmacol Toxicol, 2005. 45: p. 629-56.
17. Egger, G., et al., Epigenetics in human disease and prospects for epigenetic therapy. Nature, 2004. 429(6990): p. 457-63.
18. Herman, J.G. and S.B. Baylin, Gene silencing in cancer in association with promoter hypermethylation. N Engl J Med, 2003. 349(21): p. 2042-54.
19. Feinberg, A.P. and B. Tycko, The history of cancer epigenetics. Nat Rev Cancer, 2004. 4(2): p. 143-53.
20. Li, E., T.H. Bestor, and R. Jaenisch, Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. Cell, 1992. 69(6): p. 915-26.
21. Okano, M., et al., DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell, 1999. 99(3): p. 247-57.

22. Klose, R.J. and A.P. Bird, Genomic DNA methylation: the mark and its mediators. Trends Biochem Sci, 2006. 31(2): p. 89-97.
23. Allfrey, V.G., R. Faulkner, and A.E. Mirsky, Acetylation and Methylation of Histones and Their Possible Role in the Regulation of Rna Synthesis. Proc Natl Acad Sci U S A, 1964. 51: p. 786-94.
24. Jeffries, M.A., Epigenetic editing: How cutting-edge targeted epigenetic modification might provide novel avenues for autoimmune disease therapy. Clin Immunol, 2018. 196: p. 49-58.
25. Aprelikova, O., et al., The epigenetic modifier JMJD6 is amplified in mammary tumors and cooperates with c-Myc to enhance cellular transformation, tumor progression, and metastasis. Clin Epigenetics, 2016. 8: p. 38.
26. Dang, W., et al., Histone H4 lysine 16 acetylation regulates cellular lifespan. Nature, 2009. 459(7248): p. 802-7.
27. Fraga, M.F., et al., Loss of acetylation at Lys16 and trimethylation at Lys20 of histone H4 is a common hallmark of human cancer. Nat Genet, 2005. 37(4): p. 391-400.
28. Richon, V.M., et al., Histone deacetylase inhibitor selectively induces p21WAF1 expression and gene-associated histone acetylation. Proc Natl Acad Sci U S A, 2000. 97(18): p. 10014-9.
29. Vire, E., et al., The Polycomb group protein EZH2 directly controls DNA methylation. Nature, 2006. 439(7078): p. 871-4.
30. Cuthbert, G.L., et al., Histone deimination antagonizes arginine methylation. Cell, 2004. 118(5): p. 545-53.
31. Hassa, P.O., et al., Nuclear ADP-ribosylation reactions in mammalian cells: where are we today and where are we going? Microbiol Mol Biol Rev, 2006. 70(3): p. 789-829.
32. Nathan, D., et al., Histone sumoylation is a negative regulator in Saccharomyces cerevisiae and shows dynamic interplay with positive-acting histone modifications. Genes Dev, 2006. 20(8): p. 966-76.
33. Nelson, C.J., H. Santos-Rosa, and T. Kouzarides, Proline isomerization of histone H3 regulates lysine methylation and gene expression. Cell, 2006. 126(5): p. 905-16.
34. Nowak, S.J. and V.G. Corces, Phosphorylation of histone H3: a balancing act between chromosome condensation and transcriptional activation. Trends Genet, 2004. 20(4): p. 214-20.
35. Shilatifard, A., Chromatin modifications by methylation and ubiquitination: implications in the regulation of gene expression. Annu Rev Biochem, 2006. 75: p. 243-69.
36. Sterner, D.E. and S.L. Berger, Acetylation of histones and transcription-related factors. Microbiol Mol Biol Rev, 2000. 64(2): p. 435-59.
37. Zhang, Y. and D. Reinberg, Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails. Genes Dev, 2001. 15(18): p. 2343-60.

38. Pietrocola, F., et al., Acetyl coenzyme A: a central metabolite and second messenger. Cell Metab, 2015. 21(6): p. 805-21.
39. Cai, L., et al., Acetyl-CoA induces cell growth and proliferation by promoting the acetylation of histones at growth genes. Mol Cell, 2011. 42(4): p. 426-37.
40. Shi, L. and B.P. Tu, Acetyl-CoA and the regulation of metabolism: mechanisms and consequences. Curr Opin Cell Biol, 2015. 33: p. 125-31.
41. Choudhary, C., et al., Lysine acetylation targets protein complexes and co-regulates major cellular functions. Science, 2009. 325(5942): p. 834-40.
42. Heintzman, N.D., et al., Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat Genet, 2007. 39(3): p. 311-8.
43. Wang, Z., et al., Combinatorial patterns of histone acetylations and methylations in the human genome. Nat Genet, 2008. 40(7): p. 897-903.
44. Trisciuoglio, D., M. Di Martile, and D. Del Bufalo, Emerging Role of Histone Acetyltransferase in Stem Cells and Cancer. Stem Cells Int, 2018. 2018: p. 8908751.
45. Dearnaley, D.P., M.G. Ormerod, and J.P. Sloane, Micrometastases in breast cancer: long-term follow-up of the first patient cohort. Eur J Cancer, 1991. 27(3): p. 236-9.
46. Chan, H.M. and N.B. La Thangue, p300/CBP proteins: HATs for transcriptional bridges and scaffolds. J Cell Sci, 2001. 114(Pt 13): p. 2363-73.
47. Sano, Y. and S. Ishii, Increased affinity of c-Myb for CREB-binding protein (CBP) after CBP-induced acetylation. J Biol Chem, 2001. 276(5): p. 3674-82.
48. Wang, R., P. Cherukuri, and J. Luo, Activation of Stat3 sequence-specific DNA binding and transcription by p300/CREB-binding protein-mediated acetylation. J Biol Chem, 2005. 280(12): p. 11528-34.
49. Matsuzaki, H., et al., Acetylation of Foxo1 alters its DNA-binding ability and sensitivity to phosphorylation. Proc Natl Acad Sci U S A, 2005. 102(32): p. 11278-83.
50. Li, Y. and E. Seto, HDACs and HDAC Inhibitors in Cancer Development and Therapy. Cold Spring Harb Perspect Med, 2016. 6(10).
51. Yang, X.J. and E. Seto, The Rpd3/Hda1 family of lysine deacetylases: from bacteria and yeast to mice and men. Nat Rev Mol Cell Biol, 2008. 9(3): p. 206-18.
52. Dobosy, J.R. and E.U. Selker, Emerging connections between DNA methylation and histone acetylation. Cell Mol Life Sci, 2001. 58(5-6): p. 721-7.
53. Fuks, F., et al., Dnmt3a binds deacetylases and is recruited by a sequence-specific repressor to silence transcription. EMBO J, 2001. 20(10): p. 2536-44.
54. Witt, O., et al., HDAC family: What are the cancer relevant targets? Cancer Lett, 2009. 277(1): p. 8-21.
55. Aghdassi, A., et al., Recruitment of histone deacetylases HDAC1 and HDAC2 by the transcriptional repressor ZEB1 downregulates E-cadherin expression in pancreatic cancer. Gut, 2012. 61(3): p. 439-48.
56. Brehm, A., et al., Retinoblastoma protein recruits histone deacetylase to repress transcription. Nature, 1998. 391(6667): p. 597-601.

57. Chauchereau, A., et al., HDAC4 mediates transcriptional repression by the acute promyelocytic leukaemia-associated protein PLZF. Oncogene, 2004. 23(54): p. 8777-84.
58. Harms, K.L. and X. Chen, Histone deacetylase 2 modulates p53 transcriptional activities through regulation of p53-DNA binding activity. Cancer Res, 2007. 67(7): p. 3145-52.
59. Watamoto, K., et al., Altered interaction of HDAC5 with GATA-1 during MEL cell differentiation. Oncogene, 2003. 22(57): p. 9176-84.
60. Seligson, D.B., et al., Global histone modification patterns predict risk of prostate cancer recurrence. Nature, 2005. 435(7046): p. 1262-6.
61. Barlesi, F., et al., Global histone modifications predict prognosis of resected non-small-cell lung cancer. J Clin Oncol, 2007. 25(28): p. 4358-64.
62. Das, C., et al., CBP/p300-mediated acetylation of histone H3 on lysine 56. Nature, 2009. 459(7243): p. 113-7.
63. Van Den Broeck, A., et al., Loss of histone H4K20 trimethylation occurs in preneoplasia and influences prognosis of non-small cell lung cancer. Clin Cancer Res, 2008. 14(22): p. 7237-45.
64. Elsheikh, S.E., et al., Global histone modifications in breast cancer correlate with tumor phenotypes, prognostic factors, and patient outcome. Cancer Res, 2009. 69(9): p. 3802-9.
65. Avvakumov, N. and J. Cote, The MYST family of histone acetyltransferases and their intimate links to cancer. Oncogene, 2007. 26(37): p. 5395-407.
66. Di Martile, M., D. Del Bufalo, and D. Trisciuoglio, The multifaceted role of lysine acetylation in cancer: prognostic biomarker and therapeutic target. Oncotarget, 2016. 7(34): p. 55789-55810.
67. Iyer, A., D.P. Fairlie, and L. Brown, Lysine acetylation in obesity, diabetes and metabolic disease. Immunol Cell Biol, 2012. 90(1): p. 39-46.
68. Iyer, N.G., H. Ozdag, and C. Caldas, p300/CBP and cancer. Oncogene, 2004. 23(24): p. 4225-31.
69. Kishimoto, M., et al., Mutations and deletions of the CBP gene in human lung cancer. Clin Cancer Res, 2005. 11(2 Pt 1): p. 512-9.
70. Pasqualucci, L., et al., Inactivating mutations of acetyltransferase genes in B-cell lymphoma. Nature, 2011. 471(7337): p. 189-95.
71. Chen, M.K., et al., Overexpression of p300 correlates with poor prognosis in patients with cutaneous squamous cell carcinoma. Br J Dermatol, 2015. 172(1): p. 111-9.
72. Gayther, S.A., et al., Mutations truncating the EP300 acetylase in human cancers. Nat Genet, 2000. 24(3): p. 300-3.
73. Li, M., et al., High expression of transcriptional coactivator p300 correlates with aggressive features and poor prognosis of hepatocellular carcinoma. J Transl Med, 2011. 9: p. 5.
74. Li, Q., et al., PCAF inhibits hepatocellular carcinoma metastasis by inhibition of epithelial-mesenchymal transition by targeting Gli-1. Cancer Lett, 2016. 375(1): p. 190-198.

75. Sun, X.J., et al., The Role of Histone Acetyltransferases in Normal and Malignant Hematopoiesis. Front Oncol, 2015. 5: p. 108.
76. Xiao, X.S., et al., High Expression of p300 in Human Breast Cancer Correlates with Tumor Recurrence and Predicts Adverse Prognosis. Chin J Cancer Res, 2011. 23(3): p. 201-7.
77. Yokomizo, C., et al., High expression of p300 in HCC predicts shortened overall survival in association with enhanced epithelial mesenchymal transition of HCC cells. Cancer Lett, 2011. 310(2): p. 140-7.
78. Bryan, E.J., et al., Mutation analysis of EP300 in colon, breast and ovarian carcinomas. Int J Cancer, 2002. 102(2): p. 137-41.
79. Debes, J.D., et al., p300 in prostate cancer proliferation and progression. Cancer Res, 2003. 63(22): p. 7638-40.
80. Gang, X., et al., P300 acetyltransferase regulates fatty acid synthase expression, lipid metabolism and prostate cancer growth. Oncotarget, 2016. 7(12): p. 15135-49.
81. Zhong, J., et al., p300 acetyltransferase regulates androgen receptor degradation and PTEN-deficient prostate tumorigenesis. Cancer Res, 2014. 74(6): p. 1870-1880.
82. Kalkhoven, E., CBP and p300: HATs for different occasions. Biochem Pharmacol, 2004. 68(6): p. 1145-55.
83. Johnstone, R.W. and J.D. Licht, Histone deacetylase inhibitors in cancer therapy: is transcription the primary target? Cancer Cell, 2003. 4(1): p. 13-8.
84. Lagger, G., et al., Essential function of histone deacetylase 1 in proliferation control and CDK inhibitor repression. EMBO J, 2002. 21(11): p. 2672-81.
85. Yamaguchi, T., et al., Histone deacetylases 1 and 2 act in concert to promote the G1-to-S progression. Genes Dev, 2010. 24(5): p. 455-69.
86. Zupkovitz, G., et al., The cyclin-dependent kinase inhibitor p21 is a crucial target for histone deacetylase 1 as a regulator of cellular proliferation. Mol Cell Biol, 2010. 30(5): p. 1171-81.
87. Liberzon, A., et al., The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst, 2015. 1(6): p. 417-425.
88. Senese, S., et al., Role for histone deacetylase 1 in human tumor cell proliferation. Mol Cell Biol, 2007. 27(13): p. 4784-95.
89. Fan, J., et al., Down-regulation of HDAC5 inhibits growth of human hepatocellular carcinoma by induction of apoptosis and cell cycle arrest. Tumour Biol, 2014. 35(11): p. 11523-32.
90. Miller, K.M., et al., Human HDAC1 and HDAC2 function in the DNA-damage response to promote DNA nonhomologous end-joining. Nat Struct Mol Biol, 2010. 17(9): p. 1144-51.
91. Thurn, K.T., et al., Histone deacetylase regulation of ATM-mediated DNA damage signaling. Mol Cancer Ther, 2013. 12(10): p. 2078-87.
92. Bereshchenko, O.R., W. Gu, and R. Dalla-Favera, Acetylation inactivates the transcriptional repressor BCL6. Nat Genet, 2002. 32(4): p. 606-13.

93. Mosammaparast, N. and Y. Shi, Reversal of histone methylation: biochemical and molecular mechanisms of histone demethylases. Annu Rev Biochem, 2010. 79: p. 155-79.
94. Gremion, G., et al., [Physical performance and sedation: comparative study of the effects of a benzodiazepine (temazepam) and of a non-benzodiazepine hypnotic (zolpidem)]. Schweiz Z Sportmed, 1992. 40(3): p. 113-8.
95. Locasale, J.W., Serine, glycine and one-carbon units: cancer metabolism in full circle. Nat Rev Cancer, 2013. 13(8): p. 572-83.
96. Mentch, S.J., et al., Histone Methylation Dynamics and Gene Regulation Occur through the Sensing of One-Carbon Metabolism. Cell Metab, 2015. 22(5): p. 861-73.
97. Sadhu, M.J., et al., Nutritional control of epigenetic processes in yeast and human cells. Genetics, 2013. 195(3): p. 831-44.
98. Shiraki, N., et al., Methionine metabolism regulates maintenance and differentiation of human pluripotent stem cells. Cell Metab, 2014. 19(5): p. 780-94.
99. Shi, Y., et al., Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. Cell, 2004. 119(7): p. 941-53.
100. Kooistra, S.M. and K. Helin, Molecular mechanisms and potential functions of histone demethylases. Nat Rev Mol Cell Biol, 2012. 13(5): p. 297-311.
101. Altman, B.J., Z.E. Stine, and C.V. Dang, From Krebs to clinic: glutamine metabolism to cancer therapy. Nat Rev Cancer, 2016. 16(10): p. 619-34.
102. Chowdhury, R., et al., The oncometabolite 2-hydroxyglutarate inhibits histone lysine demethylases. EMBO Rep, 2011. 12(5): p. 463-9.
103. Fan, J., et al., Glutamine-driven oxidative phosphorylation is a major ATP source in transformed mammalian cells in both normoxia and hypoxia. Mol Syst Biol, 2013. 9: p. 712.
104. Xu, W., et al., Oncometabolite 2-hydroxyglutarate is a competitive inhibitor of alpha-ketoglutarate-dependent dioxygenases. Cancer Cell, 2011. 19(1): p. 17-30.
105. Barski, A., et al., High-resolution profiling of histone methylations in the human genome. Cell, 2007. 129(4): p. 823-37.
106. Sawan, C. and Z. Herceg, Histone modifications and cancer. Adv Genet, 2010. 70: p. 57-85.
107. Gonzalez-Perez, A., A. Jene-Sanz, and N. Lopez-Bigas, The mutational landscape of chromatin regulatory factors across 4,623 tumor samples. Genome Biol, 2013. 14(9): p. r106.
108. Kandoth, C., et al., Mutational landscape and significance across 12 major cancer types. Nature, 2013. 502(7471): p. 333-339.
109. Lawrence, M.S., et al., Discovery and saturation analysis of cancer genes across 21 tumour types. Nature, 2014. 505(7484): p. 495-501.
110. Zack, T.I., et al., Pan-cancer patterns of somatic copy number alteration. Nat Genet, 2013. 45(10): p. 1134-40.
111. McGrath, J. and P. Trojer, Targeting histone lysine methylation in cancer. Pharmacol Ther, 2015. 150: p. 1-22.

112. Hoffmann, I., et al., The role of histone demethylases in cancer therapy. Mol Oncol, 2012. 6(6): p. 683-703.
113. Rotili, D. and A. Mai, Targeting Histone Demethylases: A New Avenue for the Fight against Cancer. Genes Cancer, 2011. 2(6): p. 663-79.
114. Biggar, K.K., Z. Wang, and S.S. Li, SnapShot: Lysine Methylation beyond Histones. Mol Cell, 2017. 68(5): p. 1016-1016 e1.
115. Clarke, S.G., Protein methylation at the surface and buried deep: thinking outside the histone box. Trends Biochem Sci, 2013. 38(5): p. 243-52.
116. Dillon, S.C., et al., The SET-domain protein superfamily: protein lysine methyltransferases. Genome Biol, 2005. 6(8): p. 227.
117. Lanouette, S., et al., The functional diversity of protein lysine methylation. Mol Syst Biol, 2014. 10: p. 724.
118. Moore, K.E. and O. Gozani, An unexpected journey: lysine methylation across the proteome. Biochim Biophys Acta, 2014. 1839(12): p. 1395-403.
119. Petrossian, T.C. and S.G. Clarke, Uncovering the human methyltransferasome. Mol Cell Proteomics, 2011. 10(1): p. M110 000976.
120. Wang, Z.A. and W.R. Liu, Proteins with Site-Specific Lysine Methylation. Chemistry, 2017. 23(49): p. 11732-11737.
121. Lee, J.M., et al., RORalpha attenuates Wnt/beta-catenin signaling by PKCalpha-dependent phosphorylation in colon cancer. Mol Cell, 2010. 37(2): p. 183-95.
122. Yang, X.D., et al., Negative regulation of NF-kappaB action by Set9-mediated lysine methylation of the RelA subunit. EMBO J, 2009. 28(8): p. 1055-66.
123. Chuikov, S., et al., Regulation of p53 activity through lysine methylation. Nature, 2004. 432(7015): p. 353-60.
124. Subramanian, K., et al., Regulation of estrogen receptor alpha by the SET7 lysine methyltransferase. Mol Cell, 2008. 30(3): p. 336-47.
125. Ea, C.K. and D. Baltimore, Regulation of NF-kappaB activity through lysine monomethylation of p65. Proc Natl Acad Sci U S A, 2009. 106(45): p. 18972-7.
126. Gaughan, L., et al., Regulation of the androgen receptor by SET9-mediated methylation. Nucleic Acids Res, 2011. 39(4): p. 1266-79.
127. Huang, J., et al., Repression of p53 activity by Smyd2-mediated methylation. Nature, 2006. 444(7119): p. 629-32.
128. Liu, X., et al., Repression of hypoxia-inducible factor alpha signaling by Set7-mediated methylation. Nucleic Acids Res, 2015. 43(10): p. 5081-98.
129. Lu, T., et al., Regulation of NF-kappaB by NSD1/FBXL11-dependent reversible lysine methylation of p65. Proc Natl Acad Sci U S A, 2010. 107(1): p. 46-51.
130. Berry, W.L., et al., Oncogenic features of the JMJD2A histone demethylase in breast cancer. Int J Oncol, 2012. 41(5): p. 1701-6.
131. Gary, J.D. and S. Clarke, RNA and protein interactions modulated by protein arginine methylation. Prog Nucleic Acid Res Mol Biol, 1998. 61: p. 65-131.
132. Bedford, M.T., Arginine methylation at a glance. J Cell Sci, 2007. 120(Pt 24): p. 4243-6.

133. Lee, J. and M.T. Bedford, PABP1 identified as an arginine methyltransferase substrate using high-density protein arrays. EMBO Rep, 2002. 3(3): p. 268-73.
134. Wang, H., et al., Methylation of histone H4 at arginine 3 facilitating transcriptional activation by nuclear hormone receptor. Science, 2001. 293(5531): p. 853-7.
135. Gonsalvez, G.B., et al., Two distinct arginine methyltransferases are required for biogenesis of Sm-class ribonucleoproteins. J Cell Biol, 2007. 178(5): p. 733-40.
136. Nishioka, K. and D. Reinberg, Methods and tips for the purification of human histone methyltransferases. Methods, 2003. 31(1): p. 49-58.
137. Guccione, E., et al., Methylation of histone H3R2 by PRMT6 and H3K4 by an MLL complex are mutually exclusive. Nature, 2007. 449(7164): p. 933-7.
138. Kirmizis, A., et al., Arginine methylation at histone H3R2 controls deposition of H3K4 trimethylation. Nature, 2007. 449(7164): p. 928-32.
139. Chang, B., et al., JMJD6 is a histone arginine demethylase. Science, 2007. 318(5849): p. 444-7.
140. Markolovic, S., S.E. Wilkins, and C.J. Schofield, Protein Hydroxylation Catalyzed by 2-Oxoglutarate-dependent Oxygenases. J Biol Chem, 2015. 290(34): p. 20712-22.
141. Walport, L.J., et al., Arginine demethylation is catalysed by a subset of JmjC histone lysine demethylases. Nat Commun, 2016. 7: p. 11974.
142. Li, S., et al., JMJD1B Demethylates H4R3me2s and H3K9me2 to Facilitate Gene Expression for Development of Hematopoietic Stem and Progenitor Cells. Cell Rep, 2018. 23(2): p. 389-403.
143. Thompson, P.R. and W. Fast, Histone citrullination by protein arginine deiminase: is arginine methylation a green light or a roadblock? ACS Chem Biol, 2006. 1(7): p. 433-41.
144. Wang, Y., et al., Human PAD4 regulates histone arginine methylation levels via demethylimination. Science, 2004. 306(5694): p. 279-83.
145. Wei, H., et al., Protein arginine methylation of non-histone proteins and its role in diseases. Cell Cycle, 2014. 13(1): p. 32-41.
146. Bauer, U.M., et al., Methylation at arginine 17 of histone H3 is linked to gene activation. EMBO Rep, 2002. 3(1): p. 39-44.
147. Frietze, S., et al., CARM1 regulates estrogen-stimulated breast cancer growth through up-regulation of E2F1. Cancer Res, 2008. 68(1): p. 301-6.
148. Jansson, M., et al., Arginine methylation regulates the p53 response. Nat Cell Biol, 2008. 10(12): p. 1431-9.
149. Powers, M.A., et al., Protein arginine methyltransferase 5 accelerates tumor growth by arginine methylation of the tumor suppressor programmed cell death 4. Cancer Res, 2011. 71(16): p. 5579-87.
150. Hong, H., et al., Aberrant expression of CARM1, a transcriptional coactivator of androgen receptor, in the development of prostate carcinoma and androgen-independent status. Cancer, 2004. 101(1): p. 83-9.

151. Mathioudakis, N. and R. Salvatori, Adult-onset growth hormone deficiency: causes, complications and treatment options. Curr Opin Endocrinol Diabetes Obes, 2008. 15(4): p. 352-8.
152. Papadokostopoulou, A., et al., Colon cancer and protein arginine methyltransferase 1 gene expression. Anticancer Res, 2009. 29(4): p. 1361-6.
153. Takahashi, Y., et al., Aberrant expression of tumor suppressors CADM1 and 4.1B in invasive lesions of primary breast cancer. Breast Cancer, 2012. 19(3): p. 242-52.
154. Wang, L., S. Pal, and S. Sif, Protein arginine methyltransferase 5 suppresses the transcription of the RB family of tumor suppressors in leukemia and lymphoma cells. Mol Cell Biol, 2008. 28(20): p. 6262-77.
155. Yoshimatsu, M., et al., Dysregulation of PRMT1 and PRMT6, Type I arginine methyltransferases, is involved in various types of human cancers. Int J Cancer, 2011. 128(3): p. 562-73.
156. Zhong, J., et al., Identification and expression analysis of a novel transcript of the human PRMT2 gene resulted from alternative polyadenylation in breast cancer. Gene, 2011. 487(1): p. 1-9.
157. Chung, J., et al., Protein arginine methyltransferase 5 (PRMT5) inhibition induces lymphoma cell death through reactivation of the retinoblastoma tumor suppressor pathway and polycomb repressor complex 2 (PRC2) silencing. J Biol Chem, 2013. 288(49): p. 35534-47.
158. Pal, S., et al., Low levels of miR-92b/96 induce PRMT5 translation and H3R8/H4R3 methylation in mantle cell lymphoma. EMBO J, 2007. 26(15): p. 3558-69.
159. Pal, S., et al., Human SWI/SNF-associated PRMT5 methylates histone H3 arginine 8 and negatively regulates expression of ST7 and NM23 tumor suppressor genes. Mol Cell Biol, 2004. 24(21): p. 9630-45.
160. Baek, S.H., When signaling kinases meet histones and histone modifiers in the nucleus. Mol Cell, 2011. 42(3): p. 274-84.
161. Bungard, D., et al., Signaling kinase AMPK activates stress-promoted transcription via histone H2B phosphorylation. Science, 2010. 329(5996): p. 1201-5.
162. Dawson, M.A., et al., JAK2 phosphorylates histone H3Y41 and excludes HP1alpha from chromatin. Nature, 2009. 461(7265): p. 819-22.
163. Chadee, D.N., et al., Increased Ser-10 phosphorylation of histone H3 in mitogen-stimulated and oncogene-transformed mouse fibroblasts. J Biol Chem, 1999. 274(35): p. 24914-20.
164. Cheung, P., et al., Synergistic coupling of histone H3 phosphorylation and acetylation in response to epidermal growth factor stimulation. Mol Cell, 2000. 5(6): p. 905-15.
165. Clayton, A.L., et al., Phosphoacetylation of histone H3 on c-fos- and c-jun-associated nucleosomes upon gene activation. EMBO J, 2000. 19(14): p. 3714-26.
166. Clements, A., et al., Structural basis for histone and phosphohistone binding by the GCN5 histone acetyltransferase. Mol Cell, 2003. 12(2): p. 461-73.

167. Manning, G.S., A field-dissociation relation for polyelectrolytes with an application to field-induced conformational changes of polynucleotides. Biophys Chem, 1977. 7(3): p. 189-92.
168. Zhong, S., et al., Phosphorylation at serine 28 and acetylation at lysine 9 of histone H3 induced by trichostatin A. Oncogene, 2003. 22(34): p. 5291-7.
169. Metzger, E., et al., Phosphorylation of histone H3T6 by PKCbeta(I) controls demethylation at histone H3K4. Nature, 2010. 464(7289): p. 792-6.
170. Metzger, E., et al., Phosphorylation of histone H3 at threonine 11 establishes a novel chromatin mark for transcriptional regulation. Nat Cell Biol, 2008. 10(1): p. 53-60.
171. Gehani, S.S., et al., Polycomb group protein displacement and gene activation through MSK-dependent H3K27me3S28 phosphorylation. Mol Cell, 2010. 39(6): p. 886-900.
172. Lau, P.N. and P. Cheung, Histone code pathway involving H3 S28 phosphorylation and K27 acetylation activates transcription and antagonizes polycomb silencing. Proc Natl Acad Sci U S A, 2011. 108(7): p. 2801-6.
173. de la Barre, A.E., et al., Core histone N-termini play an essential role in mitotic chromosome condensation. EMBO J, 2000. 19(3): p. 379-91.
174. Sauve, D.M., et al., Phosphorylation-induced rearrangement of the histone H3 NH2-terminal domain during mitotic chromosome condensation. J Cell Biol, 1999. 145(2): p. 225-35.
175. Wei, Y., et al., Phosphorylation of histone H3 at serine 10 is correlated with chromosome condensation during mitosis and meiosis in Tetrahymena. Proc Natl Acad Sci U S A, 1998. 95(13): p. 7480-4.
176. Wei, Y., et al., Phosphorylation of histone H3 is required for proper chromosome condensation and segregation. Cell, 1999. 97(1): p. 99-109.
177. Downs, J.A., et al., Binding of chromatin-modifying activities to phosphorylated histone H2A at DNA damage sites. Mol Cell, 2004. 16(6): p. 979-90.
178. Morrison, A.J., et al., INO80 and gamma-H2AX interaction links ATP-dependent chromatin remodeling to DNA damage repair. Cell, 2004. 119(6): p. 767-75.
179. Shroff, R., et al., Distribution and dynamics of chromatin modification induced by a defined DNA double-strand break. Curr Biol, 2004. 14(19): p. 1703-11.
180. Rogakou, E.P., et al., Megabase chromatin domains involved in DNA double-strand breaks in vivo. J Cell Biol, 1999. 146(5): p. 905-16.
181. Rogakou, E.P., et al., DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139. J Biol Chem, 1998. 273(10): p. 5858-68.
182. Stewart, G.S., et al., MDC1 is a mediator of the mammalian DNA damage checkpoint. Nature, 2003. 421(6926): p. 961-6.
183. Fernandez-Capetillo, O., et al., H2AX: the histone guardian of the genome. DNA Repair (Amst), 2004. 3(8-9): p. 959-67.
184. Redon, C., et al., Yeast histone 2A serine 129 is essential for the efficient repair of checkpoint-blind DNA damage. EMBO Rep, 2003. 4(7): p. 678-84.

185. Macurek, L., et al., Wip1 phosphatase is associated with chromatin and dephosphorylates gammaH2AX to promote checkpoint inhibition. Oncogene, 2010. 29(15): p. 2281-91.
186. Solier, S. and Y. Pommier, The apoptotic ring: a novel entity with phosphorylated histones H2AX and H2B and activated DNA damage response kinases. Cell Cycle, 2009. 8(12): p. 1853-9.
187. Goldknopf, I.L., et al., Isolation and characterization of protein A24, a "histone-like" non-histone chromosomal protein. J Biol Chem, 1975. 250(18): p. 7182-7.
188. Nickel, B.E. and J.R. Davie, Structure of polyubiquitinated histone H2A. Biochemistry, 1989. 28(3): p. 964-8.
189. Blumenfeld, N., et al., Purification and characterization of a novel species of ubiquitin-carrier protein, E2, that is involved in degradation of non-"N-end rule" protein substrates. J Biol Chem, 1994. 269(13): p. 9574-81.
190. Ciechanover, A., et al., ATP-dependent conjugation of reticulocyte proteins with the polypeptide required for protein degradation. Proc Natl Acad Sci U S A, 1980. 77(3): p. 1365-8.
191. Villamil, M.A., Q. Liang, and Z. Zhuang, The WD40-repeat protein-containing deubiquitinase complex: catalysis, regulation, and potential for therapeutic intervention. Cell Biochem Biophys, 2013. 67(1): p. 111-26.
192. Uhlen, M., et al., Proteomics. Tissue-based map of the human proteome. Science, 2015. 347(6220): p. 1260419.
193. Johnsen, S.A., The enigmatic role of H2Bub1 in cancer. FEBS Lett, 2012. 586(11): p. 1592-601.
194. Hammond-Martel, I., H. Yu, and B. Affar el, Roles of ubiquitin signaling in transcription regulation. Cell Signal, 2012. 24(2): p. 410-21.
195. Hofmann, K., Ubiquitin-binding domains and their role in the DNA damage response. DNA Repair (Amst), 2009. 8(4): p. 544-56.
196. Thompson, L.L., et al., Regulation of chromatin structure via histone post-translational modification and the link to carcinogenesis. Cancer Metastasis Rev, 2013. 32(3-4): p. 363-76.
197. Jeusset, L.M. and K.J. McManus, Developing Targeted Therapies That Exploit Aberrant Histone Ubiquitination in Cancer. Cells, 2019. 8(2).
198. Jones, J.M., et al., The RAG1 V(D)J recombinase/ubiquitin ligase promotes ubiquitylation of acetylated, phosphorylated histone 3.3. Immunol Lett, 2011. 136(2): p. 156-62.
199. McClurg, U.L. and C.N. Robson, Deubiquitinating enzymes as oncotargets. Oncotarget, 2015. 6(12): p. 9657-68.
200. Thorslund, T., et al., Histone H1 couples initiation and amplification of ubiquitin signaling after DNA damage. Nature, 2015. 527(7578): p. 389-93.
201. Wang, H., et al., Histone H3 and H4 ubiquitylation by the CUL4-DDB-ROC1 ubiquitin ligase facilitates cellular response to DNA damage. Mol Cell, 2006. 22(3): p. 383-94.

202. Kalb, R., et al., BRCA1 is a histone-H2A-specific ubiquitin ligase. Cell Rep, 2014. 8(4): p. 999-1005.
203. West, M.H. and W.M. Bonner, Histone 2B can be modified by the attachment of ubiquitin. Nucleic Acids Res, 1980. 8(20): p. 4671-80.
204. Wu, L., et al., The RING finger protein MSL2 in the MOF complex is an E3 ubiquitin ligase for H2B K34 and is involved in crosstalk with H3 K4 and K79 methylation. Mol Cell, 2011. 43(1): p. 132-44.
205. Zhu, P., et al., A histone H2A deubiquitinase complex coordinating histone acetylation and H1 dissociation in transcriptional regulation. Mol Cell, 2007. 27(4): p. 609-21.
206. Nakagawa, T., et al., Deubiquitylation of histone H2A activates transcriptional initiation via trans-histone cross-talk with H3K4 di- and trimethylation. Genes Dev, 2008. 22(1): p. 37-49.
207. Osley, M.A., Regulation of histone H2A and H2B ubiquitylation. Brief Funct Genomic Proteomic, 2006. 5(3): p. 179-89.
208. Belotserkovskaya, R., et al., FACT facilitates transcription-dependent nucleosome alteration. Science, 2003. 301(5636): p. 1090-3.
209. Davie, J.R. and L.C. Murphy, Level of ubiquitinated histone H2B in chromatin is coupled to ongoing transcription. Biochemistry, 1990. 29(20): p. 4752-7.
210. Davie, J.R. and L.C. Murphy, Inhibition of transcription selectively reduces the level of ubiquitinated histone H2B in chromatin. Biochem Biophys Res Commun, 1994. 203(1): p. 344-50.
211. Li, B., M. Carey, and J.L. Workman, The role of chromatin during transcription. Cell, 2007. 128(4): p. 707-19.
212. Kari, V., et al., The H2B ubiquitin ligase RNF40 cooperates with SUPT16H to induce dynamic changes in chromatin structure during DNA double-strand break repair. Cell Cycle, 2011. 10(20): p. 3495-504.
213. Minsky, N., et al., Monoubiquitinated H2B is associated with the transcribed region of highly expressed genes in human cells. Nat Cell Biol, 2008. 10(4): p. 483-8.
214. Pavri, R., et al., Histone H2B monoubiquitination functions cooperatively with FACT to regulate elongation by RNA polymerase II. Cell, 2006. 125(4): p. 703-17.
215. Prenzel, T., et al., Estrogen-dependent gene transcription in human breast cancer cells relies upon proteasome-dependent monoubiquitination of histone H2B. Cancer Res, 2011. 71(17): p. 5739-53.
216. Zhu, B., et al., Monoubiquitination of human histone H2B: the factors involved and their roles in HOX gene regulation. Mol Cell, 2005. 20(4): p. 601-11.
217. Kulathu, Y. and D. Komander, Atypical ubiquitylation - the unexplored world of polyubiquitin beyond Lys48 and Lys63 linkages. Nat Rev Mol Cell Biol, 2012. 13(8): p. 508-23.
218. Wakabayashi, M., et al., Some biological aspects on the preputial gland of the female rat. Kobe J Med Sci, 1966. 12(2): p. 71-89.

219. Hirata, Y., et al., TRIM48 Promotes ASK1 Activation and Cell Death through Ubiquitination-Dependent Degradation of the ASK1-Negative Regulator PRMT1. Cell Rep, 2017. 21(9): p. 2447-2457.
220. Kwon, D.H., et al., MDM2 E3 ligase-mediated ubiquitination and degradation of HDAC1 in vascular calcification. Nat Commun, 2016. 7: p. 10492.
221. Liu, Y., et al., BRMS1 suppresses lung cancer metastases through an E3 ligase function on histone acetyltransferase p300. Cancer Res, 2013. 73(4): p. 1308-17.
222. Poizat, C., et al., Proteasome-mediated degradation of the coactivator p300 impairs cardiac transcription. Mol Cell Biol, 2000. 20(23): p. 8643-54.
223. Shima, Y., et al., PML activates transcription by protecting HIPK2 and p300 from SCFFbx3-mediated degradation. Mol Cell Biol, 2008. 28(23): p. 7126-38.
224. Van Rechem, C., et al., The SKP1-Cul1-F-box and leucine-rich repeat protein 4 (SCF-FbxL4) ubiquitin ligase regulates lysine demethylase 4A (KDM4A)/Jumonji domain-containing 2A (JMJD2A) protein. J Biol Chem, 2011. 286(35): p. 30462-70.
225. Wiper-Bergeron, N., et al., Stimulation of preadipocyte differentiation by steroid through targeting of an HDAC1 complex. EMBO J, 2003. 22(9): p. 2135-45.
226. Yu, Y.L., et al., Smurf2-mediated degradation of EZH2 enhances neuron differentiation and improves functional recovery after ischaemic stroke. EMBO Mol Med, 2013. 5(4): p. 531-47.
227. Conaway, R.C., C.S. Brower, and J.W. Conaway, Emerging roles of ubiquitin in transcription regulation. Science, 2002. 296(5571): p. 1254-8.
228. Brooks, C.L. and W. Gu, Ubiquitination, phosphorylation and acetylation: the molecular basis for p53 regulation. Curr Opin Cell Biol, 2003. 15(2): p. 164-71.
229. Desterro, J.M., M.S. Rodriguez, and R.T. Hay, SUMO-1 modification of IkappaBalpha inhibits NF-kappaB activation. Mol Cell, 1998. 2(2): p. 233-9.
230. Klenk, C., et al., SUMO-1 controls the protein stability and the biological function of phosducin. J Biol Chem, 2006. 281(13): p. 8357-64.
231. Shiio, Y. and R.N. Eisenman, Histone sumoylation is associated with transcriptional repression. Proc Natl Acad Sci U S A, 2003. 100(23): p. 13225-30.
232. Neyret-Kahn, H., et al., Sumoylation at chromatin governs coordinated repression of a transcriptional program essential for cell growth and proliferation. Genome Res, 2013. 23(10): p. 1563-79.
233. Bueno, M.T. and S. Richard, SUMOylation negatively modulates target gene occupancy of the KDM5B, a histone lysine demethylase. Epigenetics, 2013. 8(11): p. 1162-75.
234. de la Vega, L., et al., A redox-regulated SUMO/acetylation switch of HIPK2 controls the survival threshold to oxidative stress. Mol Cell, 2012. 46(4): p. 472-83.
235. Dhall, A., et al., Sumoylated human histone H4 prevents chromatin compaction by inhibiting long-range internucleosomal interactions. J Biol Chem, 2014. 289(49): p. 33827-37.
236. Girdwood, D., et al., P300 transcriptional repression is mediated by SUMO modification. Mol Cell, 2003. 11(4): p. 1043-54.

237. Huang, C., et al., SUMOylated ORC2 Recruits a Histone Demethylase to Regulate Centromeric Histone Modification and Genomic Stability. Cell Rep, 2016. 15(1): p. 147-157.
238. Lee, B. and M.T. Muller, SUMOylation enhances DNA methyltransferase 1 activity. Biochem J, 2009. 421(3): p. 449-61.
239. Riising, E.M., et al., The polycomb repressive complex 2 is a potential target of SUMO modifications. PLoS One, 2008. 3(7): p. e2704.
240. Spektor, T.M., et al., The UBC9 E2 SUMO conjugating enzyme binds the PR-Set7 histone methyltransferase to facilitate target gene repression. PLoS One, 2011. 6(7): p. e22785.
241. Wagner, T., et al., Sumoylation of HDAC2 promotes NF-kappaB-dependent gene expression. Oncotarget, 2015. 6(9): p. 7123-35.
242. Yang, S.H., et al., Dynamic interplay of the SUMO and ERK pathways in regulating Elk-1 transcriptional activity. Mol Cell, 2003. 12(1): p. 63-74.
243. Zheng, J., et al., SUMO-1 Promotes Ishikawa Cell Proliferation and Apoptosis in Endometrial Cancer by Increasing Sumoylation of Histone H4. Int J Gynecol Cancer, 2015. 25(8): p. 1364-8.
244. David, G., M.A. Neptune, and R.A. DePinho, SUMO-1 modification of histone deacetylase 1 (HDAC1) modulates its biological activities. J Biol Chem, 2002. 277(26): p. 23658-63.
245. Kirsh, O., et al., The SUMO E3 ligase RanBP2 promotes modification of the HDAC4 deacetylase. EMBO J, 2002. 21(11): p. 2682-91.
246. Kim, Y.H., C.Y. Choi, and Y. Kim, Covalent modification of the homeodomain-interacting protein kinase 2 (HIPK2) by the ubiquitin-like protein SUMO-1. Proc Natl Acad Sci U S A, 1999. 96(22): p. 12350-5.
247. Ross, S., et al., SUMO-1 modification represses Sp3 transcriptional activation and modulates its subnuclear localization. Mol Cell, 2002. 10(4): p. 831-42.
248. Sahin, U., H. de The, and V. Lallemand-Breitenbach, PML nuclear bodies: assembly and oxidative stress-sensitive sumoylation. Nucleus, 2014. 5(6): p. 499-507.
249. Tatemichi, Y., et al., Nucleus accumbens associated 1 is recruited within the promyelocytic leukemia nuclear body through SUMO modification. Cancer Sci, 2015. 106(7): p. 848-56.
250. Hannich, J.T., et al., Defining the SUMO-modified proteome by multiple approaches in Saccharomyces cerevisiae. J Biol Chem, 2005. 280(6): p. 4102-10.
251. Song, J., et al., Identification of a SUMO-binding motif that recognizes SUMO-modified proteins. Proc Natl Acad Sci U S A, 2004. 101(40): p. 14373-8.
252. Song, J., et al., Small ubiquitin-like modifier (SUMO) recognition of a SUMO binding motif: a reversal of the bound orientation. J Biol Chem, 2005. 280(48): p. 40122-9.
253. Yang, S.H. and A.D. Sharrocks, SUMO promotes HDAC-mediated transcriptional repression. Mol Cell, 2004. 13(4): p. 611-7.

254. Hottiger, M.O., et al., Toward a unified nomenclature for mammalian ADP-ribosyltransferases. Trends Biochem Sci, 2010. 35(4): p. 208-19.
255. Aravind, L., et al., The natural history of ADP-ribosyltransferases and the ADP-ribosylation system. Curr Top Microbiol Immunol, 2015. 384: p. 3-32.
256. Alvarez-Gonzalez, R. and H. Mendoza-Alvarez, Dissection of ADP-ribose polymer synthesis into individual steps of initiation, elongation, and branching. Biochimie, 1995. 77(6): p. 403-7.
257. Hilz, H., ADP-ribosylation of proteins--a multifunctional process. Hoppe Seylers Z Physiol Chem, 1981. 362(11): p. 1415-25.
258. Wielckens, K., et al., Protein-bound polymeric and monomeric ADP-ribose residues in hepatic tissues. Comparative analyses using a new procedure for the quantification of poly (ADP-ribose). Eur J Biochem, 1981. 117(1): p. 69-74.
259. Kim, M.Y., et al., NAD+-dependent modulation of chromatin structure and transcription by nucleosome binding properties of PARP-1. Cell, 2004. 119(6): p. 803-14.
260. Poirier, G.G., et al., Poly (ADP-ribosyl) ation of polynucleosomes causes relaxation of chromatin structure. Proc Natl Acad Sci U S A, 1982. 79(11): p. 3423-7.
261. Malik, N. and M. Smulson, A relationship between nuclear poly (adenosine diphosphate ribosylation) and acetylation posttranslational modifications. 1. Nucleosome studies. Biochemistry, 1984. 23(16): p. 3721-5.
262. Tanigawa, Y., et al., Mono (ADP-ribosyl) ation of hen liver nuclear proteins suppresses phosphorylation. Biochem Biophys Res Commun, 1983. 113(1): p. 135-41.
263. Wong, M., et al., Relationship between histone H1 poly (adenosine diphosphate ribosylation) and histone H1 phosphorylation using anti-poly (adenosine diphosphate ribose) antibody. Biochemistry, 1983. 22(10): p. 2384-9.
264. Follis, A.V., et al., Pin1-Induced Proline Isomerization in Cytosolic p53 Mediates BAX Activation and Apoptosis. Mol Cell, 2015. 59(4): p. 677-84.
265. Sadakierska-Chudy, A. and M. Filip, A comprehensive view of the epigenetic landscape. Part II: Histone post-translational modification, nucleosome level, and chromatin regulation by ncRNAs. Neurotox Res, 2015. 27(2): p. 172-97.
266. Youdell, M.L., et al., Roles for Ctk1 and Spt6 in regulating the different methylation states of histone H3 lysine 36. Mol Cell Biol, 2008. 28(16): p. 4915-26.
267. Lai, A.Y. and P.A. Wade, Cancer biology and NuRD: a multifaceted chromatin remodeling complex. Nat Rev Cancer, 2011. 11(8): p. 588-96.
268. Wilson, B.G. and C.W. Roberts, SWI/SNF nucleosome remodelers and cancer. Nat Rev Cancer, 2011. 11(7): p. 481-92.
269. Bartholomew, B., Regulating the chromatin landscape: structural and mechanistic perspectives. Annu Rev Biochem, 2014. 83: p. 671-96.
270. Clapier, C.R. and B.R. Cairns, The biology of chromatin remodeling complexes. Annu Rev Biochem, 2009. 78: p. 273-304.
271. Narlikar, G.J., R. Sundaramoorthy, and T. Owen-Hughes, Mechanisms and functions of ATP-dependent chromatin-remodeling enzymes. Cell, 2013. 154(3): p. 490-503.

272. Corona, D.F., et al., ISWI is an ATP-dependent nucleosome remodeling factor. Mol Cell, 1999. 3(2): p. 239-45.
273. Fei, J., et al., The prenucleosome, a stable conformational isomer of the nucleosome. Genes Dev, 2015. 29(24): p. 2563-75.
274. Ito, T., et al., ACF, an ISWI-containing and ATP-utilizing chromatin assembly and remodeling factor. Cell, 1997. 90(1): p. 145-55.
275. Lusser, A., D.L. Urwin, and J.T. Kadonaga, Distinct activities of CHD1 and ACF in ATP-dependent chromatin assembly. Nat Struct Mol Biol, 2005. 12(2): p. 160-6.
276. Torigoe, S.E., et al., Identification of a rapidly formed nonnucleosomal histone-DNA intermediate that is converted into chromatin by ACF. Mol Cell, 2011. 43(4): p. 638-48.
277. Boeger, H., et al., Removal of promoter nucleosomes by disassembly rather than sliding in vivo. Mol Cell, 2004. 14(5): p. 667-73.
278. Goldberg, A.D., et al., Distinct factors control histone variant H3.3 localization at specific genomic regions. Cell, 2010. 140(5): p. 678-91.
279. Mizuguchi, G., et al., ATP-driven exchange of histone H2AZ variant catalyzed by SWR1 chromatin remodeling complex. Science, 2004. 303(5656): p. 343-8.
280. Ruhl, D.D., et al., Purification of a human SRCAP complex that remodels chromatin by incorporating the histone variant H2A.Z into nucleosomes. Biochemistry, 2006. 45(17): p. 5671-7.
281. Suganuma, T. and J.L. Workman, Signals and combinatorial functions of histone modifications. Annu Rev Biochem, 2011. 80: p. 473-99.
282. Chatterjee, N., et al., Histone H3 tail acetylation modulates ATP-dependent remodeling through multiple mechanisms. Nucleic Acids Res, 2011. 39(19): p. 8378-91.
283. Beato, M. and K. Eisfeld, Transcription factor access to chromatin. Nucleic Acids Res, 1997. 25(18): p. 3559-63.
284. Eckersley-Maslin, M.A., C. Alda-Catalinas, and W. Reik, Dynamics of the epigenetic landscape during the maternal-to-zygotic transition. Nat Rev Mol Cell Biol, 2018. 19(7): p. 436-450.
285. Karin, M., Too many transcription factors: positive and negative interactions. New Biol, 1990. 2(2): p. 126-31.
286. Latchman, D.S., Transcription factors: an overview. Int J Biochem Cell Biol, 1997. 29(12): p. 1305-12.
287. Sikder, D. and T. Kodadek, Genomic studies of transcription factor-DNA interactions. Curr Opin Chem Biol, 2005. 9(1): p. 38-45.
288. Jin, J., et al., PlantTFDB 3.0: a portal for the functional and evolutionary study of plant transcription factors. Nucleic Acids Res, 2014. 42(Database issue): p. D1182-7.
289. Matys, V., et al., TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. Nucleic Acids Res, 2006. 34(Database issue): p. D108-10.
290. Mitchell, P.J. and R. Tjian, Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. Science, 1989. 245(4916): p. 371-8.

291. Ptashne, M. and A. Gann, Transcriptional activation by recruitment. Nature, 1997. 386(6625): p. 569-77.
292. Lee, T.I. and R.A. Young, Transcription of eukaryotic protein-coding genes. Annu Rev Genet, 2000. 34: p. 77-137.
293. Nikolov, D.B. and S.K. Burley, RNA polymerase II transcription initiation: a structural view. Proc Natl Acad Sci U S A, 1997. 94(1): p. 15-22.
294. Roeder, R.G., The role of general initiation factors in transcription by RNA polymerase II. Trends Biochem Sci, 1996. 21(9): p. 327-35.
295. Shlyueva, D., G. Stampfel, and A. Stark, Transcriptional enhancers: from properties to genome-wide predictions. Nat Rev Genet, 2014. 15(4): p. 272-86.
296. Yip, K.Y., et al., Classification of human genomic regions based on experimentally determined binding sites of more than 100 transcription-related factors. Genome Biol, 2012. 13(9): p. R48.
297. Pan, G., et al., A negative feedback loop of transcription factors that controls stem cell pluripotency and self-renewal. FASEB J, 2006. 20(10): p. 1730-2.
298. Bohmann, D., Transcription factor phosphorylation: a link between signal transduction and the regulation of gene expression. Cancer Cells, 1990. 2(11): p. 337-44.
299. Weigel, N.L. and N.L. Moore, Steroid receptor phosphorylation: a key modulator of multiple receptor functions. Mol Endocrinol, 2007. 21(10): p. 2311-9.
300. Whiteside, S.T. and S. Goodbourn, Signal transduction and nuclear targeting: regulation of transcription factor activity by subcellular localisation. J Cell Sci, 1993. 104 (Pt 4): p. 949-55.
301. Brown, M.S. and J.L. Goldstein, The SREBP pathway: regulation of cholesterol metabolism by proteolysis of a membrane-bound transcription factor. Cell, 1997. 89(3): p. 331-40.
302. Dias, S., et al., Effector Regulatory T Cell Differentiation and Immune Homeostasis Depend on the Transcription Factor Myb. Immunity, 2017. 46(1): p. 78-91.
303. Kang, J. and N. Malhotra, Transcription factor networks directing the development, function, and evolution of innate lymphoid effectors. Annu Rev Immunol, 2015. 33: p. 505-38.
304. Semenza, G.L., Transcription factors and human disease. Oxford monographs on medical genetics. 1998, New York: Oxford University Press. xiii, 368 p.
305. Libermann, T.A. and L.F. Zerbini, Targeting transcription factors for cancer gene therapy. Curr Gene Ther, 2006. 6(1): p. 17-33.
306. Landolin, J.M., et al., Sequence features that drive human promoter function and tissue specificity. Genome Res, 2010. 20(7): p. 890-8.
307. Spitz, F. and E.E. Furlong, Transcription factors: from enhancer binding to developmental control. Nat Rev Genet, 2012. 13(9): p. 613-26.
308. Dror, I., R. Rohs, and Y. Mandel-Gutfreund, How motif environment influences transcription factor search dynamics: Finding a needle in a haystack. Bioessays, 2016. 38(7): p. 605-12.

309. Inukai, S., K.H. Kock, and M.L. Bulyk, Transcription factor-DNA binding: beyond binding site motifs. Curr Opin Genet Dev, 2017. 43: p. 110-119.
310. Slattery, M., et al., Absence of a simple code: how transcription factors read the genome. Trends Biochem Sci, 2014. 39(9): p. 381-99.
311. Zhang, S., et al., Profiling the transcription factor regulatory networks of human cell types. Nucleic Acids Res, 2014. 42(20): p. 12380-7.
312. Bernstein, B.E., A. Meissner, and E.S. Lander, The mammalian epigenome. Cell, 2007. 128(4): p. 669-81.
313. Ge, Z., et al., Chromatin remodeling: recruitment of histone demethylase RBP2 by Mad1 for transcriptional repression of a Myc target gene, telomerase reverse transcriptase. FASEB J, 2010. 24(2): p. 579-86.
314. Kleine-Kohlbrecher, D., et al., A functional link between the histone demethylase PHF8 and the transcription factor ZNF711 in X-linked mental retardation. Mol Cell, 2010. 38(2): p. 165-78.
315. Outchkourov, N.S., et al., Balancing of histone H3K4 methylation states by the Kdm5c/SMCX histone demethylase modulates promoter and enhancer function. Cell Rep, 2013. 3(4): p. 1071-9.
316. Tahiliani, M., et al., The histone H3K4 demethylase SMCX links REST target genes to X-linked mental retardation. Nature, 2007. 447(7144): p. 601-5.
317. Visakorpi, T., et al., In vivo amplification of the androgen receptor gene and progression of human prostate cancer. Nat Genet, 1995. 9(4): p. 401-6.

Chapter 2: Methodology

2.1 Studying phenotypic shifts caused by transcriptional regulators in cancer

The phenotype of a cell depends on how genes are being expressed in that cell. Shifts in the phenotype of the cell are due to changes in gene expression, and often are caused by changes in the mechanisms that regulate the gene. The first level of regulation for gene expression occurs during the process of transcription where the actions of transcription factors, histone modifiers, and other proteins are able to influence RNA polymerase's ability to access the gene. These transcriptional regulators are able to work in a combinatorial fashion with one another to either activate and/or repress wide repertoires of transcriptional targets. Within the context of cancer; mutations in the functional domains of either transcription factors or the histone modifying proteins may deregulate the transcriptional program it oversees.

Problems that I would like to address in my research are how do the mutations found in these transcriptional regulators affect their function, where in the genome do, they act, how is gene expression affected, and how do these regulators interact with one another. Mutations within transcription factors and histone modifiers may alter their activity level, their ability to be regulated, or how their protein behaves in the cell. Combining genomic with transcriptomic studies will provide insight into where these transcriptional regulators localize genome wide, and their impact on gene expression. Lastly, building a network of co-localizing factors could describe the combinatorial regulation of several factors in gene regulation and may serve as a mechanism for target recognition.

2.2. Overview of approach to primary research

After the completion of the Human genome project many researchers believe that mapping the human epigenome is the next step [1, 2]. However an organism's epigenome varies on a number of factors including age, tissue type, and environmental triggers. Additionally disease aberrations of normal epigenomic events may occur. Yet mapping different epigenomic events across all these conditions will yield insight into epigenetic mechanisms in processes of aging and disease. This philosophy of mapping epigenomic events can also be applied to transcription factors allowing researchers to develop gene regulatory circuits in regard to gene expression. In order to model this, researchers are able to use cancer cell lines and patient data from the Cancer Genome Atlas.

2.2.1 Using cancer cell lines as a model system

In order to generate consistent sampling and to make the results reproducible the use of cell lines is desired [3]. The use of cancer cell lines in scientific research began 60 years ago and has become an integral part of life science research with thousands of cell lines being established from a variety of donor tissues. The use of immortalized cell lines over primary cells is advantageous due to their low cost, ability to be grown easily, and

ability to bypass ethical concerns with animal or human tissues [3]. While these cell lines originate form a tissue type; they have undergone several mutations to become immortal. This may generate some concern for how much the cell line is able to simulate the conditions found within patient's tumors. Comparing breast cancer tumors with breast cancer cell lines showed a similarity in gene expression patterns between the two [4, 5]. Quantitatively speaking there was a 72% agreement between breast cancer cell line and breast cancer tumor gene expression; however there appeared to be more copy number alterations found in the cell lines [6]. These results suggest that cell lines are able to mirror many aspects of cancer biology, however there are some differences that don't match completely with tumors. While cancer cell lines are not able to completely mirror the behavior of tumors; they are advantageous to researchers as they are able to grow a clonal population of cells for consistent sampling and reproducible results [3]. This pattern is found across several tissue types including ovarian, colorectal, melanoma, and lung cancer [7-10].

2.2.2 The Cancer Genome Atlas Program

The Cancer Genome Atlas (TCGA) is a cancer genomics program initiated by the National Cancer Institute and the National Human Genome Research Institute in 2006 to catalogue mutations found in cancer using genome sequencing and bioinformatics. This program has since grown to 2.5 petabytes of genomic, epigenetic, transcriptomic, and proteomic data for over 20,000 cancer samples across 33 different cancer types [11]. The goal of this project is to generate inclusive genomic, transcriptomic, epigenetic, and proteomic datasets with the sole purpose of creating large patient cohorts for each cancer type. This goal is advantageous to cancer biologists as the process of collecting and sequencing high quality data on this large scale would not be achievable by an individual research group. Likewise it is the hope that in making the datasets of the TCGA available to the public that the research community will be able to mine the data for better understanding of the molecular mechanisms behind cancer biology [12]. This consortium is a reliable source of datasets due to the TCGA having a strict set of criteria for including samples into its study. They collect the primary tumor from a patient that hasn't received treatment yet and a sample of normal tissue to use as a control. The Biospecimen Core Resource at the National Institute of Health ensures consistent assessment of its pathology, and then the DNA and RNA from the collected tissue is extracted. These samples then undergo a quality control process before they are sequenced and uploaded for genomic analysis. Access to this dataset allows researchers to discover abnormalities of DNA sequences, gene expression, and protein structure [13-19]. The TCGA has also been developing an account of different forms of epigenetic regulation found altered in cancer patients such as miRNAs, ncRNAs, and methylation [20-25]. The consortium is able to identify various forms of DNA alterations found in cancer such as substitutions, indels, gene fusions, and copy number alterations [13, 15, 26-28]. Altogether it is the goal of researchers to identify the distinct mutational characteristics of that cancer type in patients and place their altered roles in the context of

functional cellular pathways [29-31]. The TCGA can be used for finding mutations, predicting genes that may have some significance in prognosis, and regulatory network development [32, 33]. Tying the TCGA into research with cancer cell lines allows researchers to compare how well their cell line models correlate with what is seen in patients.

2.3 Genomics Techniques

2.3.1 ChIP-seq in order to identify genomic locations

A genome wide methodology is needed in order to determine the genomic locations of where regulatory events take place. The Chromatin immunoprecipitation sequencing (ChIP-seq) protocol is able to identify enriched binding sites for immunoprecipitated molecules [34]. This technique was first developed in 2007 to identify DNA sequences bound by a transcription factor [35]. Since then the applications of ChIP-seq have expanded to profiling histone modifications, histone variants, and nucleosome positioning. ChIP-seq is a common method in the field of genomics for understanding transcriptional regulation in disease, tissue specific epigenetic regulation, and chromatin organization [36-45]. The process of sample preparation for ChIP-seq can be summarized as the DNA and its associated protein of interest are cross linked, sheared into fragments, isolated following treatment with a protein specific antibody, and then lastly sequenced [46-48]. Prior to the invention of affordable next generation sequencing, the immunoprecipitated DNA regions were required to hybridize with DNA probes on an array in order to determine their genomic locations [49]. This proved problematic for experimental analysis in two ways: If there wasn't a corresponding DNA probe on the array it wasn't counted, and a mismatch between the ChIP DNA and the array probe would result in false identification of ChIP events. These problems in this experimental design are nullified in a ChIP-seq experiment due to the process of sequencing the DNA fragments means there is no dependence on the presence of the DNA probe or that proper hybridization will take place. The process of DNA hybridization also limited the resolution of the identified DNA fragments to only 30-100 base pairs, while the advent of sequencing following ChIP increased the resolution of the bound DNA to a single base pair [50]. With all that being said ChIP-seq has its limitations due to the cost of sequencing, and the quality of its data is dependent on the antibody used in the experiment. Ideally an antibody proven to be sensitive and specific towards its target will result in clean data while an antibody that binds non-specifically contributes to a noisy experiment [51]. Another limitation for this type of experiment is that in order for a ChIP-seq experiment to produce noise-free, consistent, and reproducible results an input of around 50 million cells is suggested [52, 53]. As DNA is being isolated from each cell within the sample, the binding patterns of the antibody act in an additive manner in creating the experiment signal. Lastly during the sample preparation and amplification stage of ChIP-seq there is potential for generating a bias towards DNA fragments with a

higher GC nucleotide content. Luckily there are bioinformatic packages that are able to perform quality control tests to report any biases that were generated during sequencing that can thereby be corrected. The raw data and all its processed forms are made available to other scientists through the Gene expression omnibus with the purpose of sharing data between laboratories to ensure the results are reproducible [54]. Information to be shared includes cell line, experimental protocol, whether the experiment uses biological or technical replicates, catalog/lot number for antibody, peak calling algorithm used, summary for the number of reads, and a link to control track.

2.3.2 Bisulfite DNA methylation Sequencing for identification of DNA methylation events

DNA methylation is an epigenetic mark associated with repression of transcriptional activity that involves the addition of a methylation group onto the carbon-5 position of cytosine residues of a CpG dinucleotide. Researchers are able to determine regions of DNA methylation by treating DNA with the chemical bisulfite, which is able to convert cytosine residues to uracil, but methylated cytosine residues are left alone. Following treatment with bisulfite the samples can be sequenced allowing researchers to generate a single nucleotide resolution of the methylation status of individual cytosine residues. In order for unmethylated cytosines to be converted into uracils the DNA needs to be single stranded, otherwise this would result in an incomplete conversion of cytosine molecules and will result in false positives for DNA methylation [55]. Some methods utilize Embedding DNA in agarose to help increase the conversion rate by physically keeping the DNA strands separate [56]. Other limitations to bisulfite sequencing are that the conditions associated for a complete conversion can often lead to the degradation of the incubated DNA [57]. Techniques such as cycling incubation temperature and monitoring the pH may minimize DNA degradation or allow the DNA to amplify during PCR [55, 58].

The DNA modification of cytosine into 5-hydroxymethylcytosine complicates bisulfite sequencing due its treatment with bisulfite results in its conversion to cytosine-5-methylsulfonate [59, 60]. This complicates bisulfite sequencing as it is unable to distinguish between 5-methylcytosine and 5-hydroxymethylcytosine [61]. In order to distinguish between these two marks additional steps were added to the bisulfite sequencing methodology. In order to obtain only 5-methylcytosine marks the oxidative bisulfite sequencing approach was developed where 5-hydroxymethylcytosine is chemically oxidized to the 5-formylcytosine state, which is able to be converted into uracil following bisulfite treatment [62]. In order to distinguish 5-hydroxymethylcytosine marks a modified method called Tet-assisted bisulfite sequencing (TAB-seq) is able to protect 5-hydroxymethylcytosine marks while reducing 5-methylcytosine to thymine [63].

2.3.3 Genome alignment

A daunting aspect of a ChIP-seq experiment is the mapping of billions of short DNA fragments to their respective position in the genome. The field of ChIP-seq analysis has progressively refined sorting algorithms that are able to balance the accuracy, speed, and flexibility necessary for tackling this task [64, 65]. The primary goal of these alignment tools is to map these sequences correctly to their genomic loci while allowing some flexibility in matching due to errors in sequencing, single nucleotide polymorphisms, indels, or differences between the genome of interest and the reference genome [64, 65]. While allowing reads that are mapped to multiple locations would increase the sensitivity of peak detection, it also increases the number of false positives [66]. Uniquely mapped reads contain enough information for most ChIP-seq analysis [67].

2.3.4 Peak Calling

In order to determine genomic regions of significant ChIP enrichment the computational method "peak calling" is used. These enriched regions are where the protein interacts with the DNA [68]. Peaks may have different qualities depending on the type of ChIP-seq experiment performed. ChIP-seq experiments investigating transcription factor binding have very defined sharp 100-200 base pair peaks compared to ChIP-seq experiments exploring histone marks that can cover a wider region [69]. A number of computational and statistical methods have been developed for distinguishing between signal and noise during the process of peak calling [70-73]. Differential peak calling identifies significant differences between two ChIP signals; most notably the experimental and control ChIP-seq experiments. A ChIP-seq control sample is usually naked DNA that was obtained or an immunoprecipitation with a nonspecific antibody. During the peak calling process, a threshold for the significance of ChIP enrichment compared to the control sample is set. Following the peak calling experiment significant peaks across ChIP samples can be used to detect significant differences between experimental conditions resulting in a list of genomic target regions that can be used to discover transcription factor motifs, and gene ontology regulated by that ChIP'ed factor [74-77].

2.3.5 Motif Analysis

It is possible to identify transcription factor binding sites within ChIP-seq peak regions. Transcription factors are able to recognize 6 to 20 base pair patterns in DNA called motifs [78]. It is possible to find these motifs by submitting sequences from peak regions identified during peak calling into motif finding algorithms where potential motifs are returned with their statistical significance [79-82]. These algorithms use position frequency matrices (PFM) to describe the preference of nucleotides at each position in the queried transcription factor binding site as a way to determine if a sequence can be statistically considered as a binding site [83, 84]. Public databases such

as JASPAR, TRANSFAC, and HOCOMOCO containing the binding preferences for hundreds of transcription factors [78, 85-89]. The degenerate specificity of transcription factors can result in predicted binding sites numbering in the millions, however when combined with ChIP-seq data it can range somewhere in the thousands [90, 91]. Ideally this methodology is used confirm true DNA binding events of transcription factors, but can also be used to associate histone modifying marks with transcription factors as the ChIP-seq signal of Histone modifications have a strong correlation with transcription factor motifs allowing researchers to speculate on the cooperation between transcription factors and histone marks [92].

2.3.6 Gene set enrichment analysis

Genome wide methodologies generate signal for a large number of genes making it difficult to determine the experimental impact [93]. Gene set enrichment analysis (GSEA) is used to identify over-representation of a class of genes or proteins in a large set of genes. This approach is useful for interpreting genome wide profiles [94]. Researchers are able to generate gene sets from their experiments but need to verify whether that gene set is involved in a biological process. This is achieved by comparing the input genes to reference sets and performing a statistical test to see if there is enrichment. A GSEA calculates an enrichment score for genes that are over-represented, the statistical significance of the enrichment score, and its bearing on a false discovery rate [94, 95]. The Molecular Signatures Database (MSigDB) has over 10,000 curated gene sets ranging from genomic location, canonical pathways, transcription factor targets, coexpression, oncogenesis, and immunological response [93].

2.4 Transcriptomics

While ChIP-seq can help identify genomic loci of histone marks and transcription factor binding sites, this technique alone doesn't explain how these genomic events affect gene expression. In order to achieve this, an experiment in studying the transcriptome is needed. The word transcriptome describes all of the RNA molecules found in a cell and is a reflection of which genes are being expressed at a given time. The field of transcriptomics investigates gene expression by examining the transcript levels of genes within a given population. Researchers are becoming more accustomed to using RNA sequencing (RNAseq) for studying the transcriptome of the cell, although an older technique called DNA microarrays can still be used. By combining either of these transcriptomic experiments to a ChIP-seq experiment a researcher is able to assess how genomic events impact gene expression [52, 69, 72, 96, 97].

2.4.1 Microarray

A microarray is a grid like format of microscopic probes attached to a membrane [98, 99]. These probes are a specific DNA sequence of a gene that recognizes and binds to fluorescently labeled complementary DNA (cDNA) made from mRNA extracted from

an organism [100]. Researchers are then able to measure gene expression for thousands of genes simultaneously by quantifying the amount of fluorescence given off at each probe [98]. Applications of microarray include profiling gene expression [101], comparative genomic hybridization [102, 103], and identifying single nucleotide polymorphisms [104]. A majority of microarray studies strive to compare gene expression between diseased cells and healthy cells [98]. Microarrays remain an extensively used approach in transcriptomic studies due to the effectiveness of their cost, and ease of analysis [105]. Microarrays are used to measure differential gene expression caused by knockout of transcription factors, and epigenetic interactions [106-109].

A two-channel microarray experiment compares gene expression across two conditions by having the probes treated with cDNA prepared from two samples that are labeled with two different fluorescent dyes [110]. This methodology examines the ratio of relative intensities for each fluorophore to identify differentially expressed genes between two conditions [111]. Two channel microarrays typically carry control probes to assist with normalization of the fluorescence measurements [112, 113]. This methodology generates a relative difference measurement between the two samples [100, 114].

Sources of variability in a microarray experiment are related to differences in arrays, dye labeling, efficiency in reverse transcription, and hybridization [115, 116]. These problems are generated during production of the arrays and probes but are typically resolved through careful quality control measures issues by the manufacturer. The best way to overcome any issue of variability is to incorporate replicates, perform dye-swapping experiments, and pooling samples to minimize biological variation [113, 116, 117].

Microarray results can be validated through RT-PCR, Northern blotting, Western blotting, and multiple microarray platforms [115, 116, 118]. Since a single microarray experiment measures gene expression levels for thousands of genes, it is not cost effective to verify every single gene. A list of key genes is verified depending on the scope and purpose of the experiment [99, 114]. A microarray platform may have excellent reproducibility without producing accurate or consistent measurements with other platforms due to the fact that a given probe is able to bind to the same number of labeled transcripts in repeated measurements of the same sample. Oligonucleotide microarrays produced by Affymetrix, Agilent and Codelink are able to provide correlating data sets suggesting high reproducibility [119, 120].

2.4.2 RNA-seq

Microarrays are slowly being replaced with RNAseq as it does not depend on the presence of probes for detecting gene expression and can identify transcript specific expression. RNAseq enables the entire transcriptome to be sequenced for quantifying gene expression [121, 122]. RNAseq is able to detect new transcripts, and splice junctions [123]. Although splice junctions assume that a number of reads span exon-exon junctions [124].

Studies comparing RNA-seq and microarray experiments report that the RNA-seq method was better for detecting lowly abundant transcripts, and isoforms [123, 125-129]. RNAseq is also the best method for providing the fewest false positives, highest true positives, and the best modeling of variation within the datasets [127, 130-133]. The RNAseq method allows for a higher number of differentially expressed genes compared to microarrays because sequencing is independent of probe availability and expression intensity [122, 124].

RNAseq, unlike microarrays, does not rely on transcript specific probes and can detect new transcripts, gene fusions, single nucleotide variants, and indels [122, 134]. RNAseq is not limited, like microarrays, in measuring gene expression due to prevalent background noise confounding the measurement of lowly expressed genes, and the signal gets saturated for highly expressed genes [122, 123, 134]. RNAseq has a higher specificity and sensitivity for detecting gene expression, especially lowly expressed genes [135-137]. The national institute of health (NIH) in recent years has begun allocating a majority of its grants in favor of new RNA-seq projects compared to microarray projects.

Expression profiling of samples is cheaper using microarrays instead of RNAseq. It makes sense to use microarray technology instead of RNA-seq if there are a large number of samples and cost is critical, for comparing with other microarray datasets from the same platform, or if there is an efficient in-house microarray workflow from sample collection to data analysis.

2.5 References

1. Bradbury, J., Human epigenome project--up and running. PLoS Biol, 2003. 1(3): p. E82.
2. Esteller, M., The necessity of a human epigenome project. Carcinogenesis, 2006. 27(6): p. 1121-5.
3. Kaur, G. and J.M. Dufour, Cell lines: Valuable tools or useless artifacts. Spermatogenesis, 2012. 2(1): p. 1-5.
4. Ross, D.T. and C.M. Perou, A comparison of gene expression signatures from breast tumors and breast tissue derived cell lines. Dis Markers, 2001. 17(2): p. 99-109.
5. Ross, D.T., et al., Systematic variation in gene expression patterns in human cancer cell lines. Nat Genet, 2000. 24(3): p. 227-35.
6. Neve, R.M., et al., A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. Cancer Cell, 2006. 10(6): p. 515-27.
7. Domcke, S., et al., Evaluating cell lines as tumour models by comparison of genomic profiles. Nat Commun, 2013. 4: p. 2126.
8. Lin, W.M., et al., Modeling genomic diversity and tumor dependency in malignant melanoma. Cancer Res, 2008. 68(3): p. 664-73.
9. Mouradov, D., et al., Colorectal cancer cell lines are representative models of the main molecular subtypes of primary cancer. Cancer Res, 2014. 74(12): p. 3238-47.

10. Sos, M.L., et al., Predicting drug susceptibility of non-small cell lung cancers based on genetic lesions. J Clin Invest, 2009. 119(6): p. 1727-40.
11. Hoadley, K.A., et al., Cell-of-Origin Patterns Dominate the Molecular Classification of 10,000 Tumors from 33 Types of Cancer. Cell, 2018. 173(2): p. 291-304 e6.
12. Cancer Genome Atlas Research, N., Comprehensive genomic characterization defines human glioblastoma genes and core pathways. Nature, 2008. 455(7216): p. 1061-8.
13. Bailey, M.H., et al., Comprehensive Characterization of Cancer Driver Genes and Mutations. Cell, 2018. 174(4): p. 1034-1035.
14. Fumagalli, D., et al., Principles Governing A-to-I RNA Editing in the Breast Cancer Transcriptome. Cell Rep, 2015. 13(2): p. 277-89.
15. Gao, Q., et al., Driver Fusions and Their Implications in the Development and Treatment of Human Cancers. Cell Rep, 2018. 23(1): p. 227-238 e3.
16. Han, L., et al., The Genomic Landscape and Clinical Relevance of A-to-I RNA Editing in Human Cancers. Cancer Cell, 2015. 28(4): p. 515-528.
17. Knijnenburg, T.A., et al., Genomic and Molecular Landscape of DNA Damage Repair Deficiency across The Cancer Genome Atlas. Cell Rep, 2018. 23(1): p. 239-254 e6.
18. Paz-Yaacov, N., et al., Elevated RNA Editing Activity Is a Major Contributor to Transcriptomic Diversity in Tumors. Cell Rep, 2015. 13(2): p. 267-76.
19. Shen, H., et al., Integrated Molecular Characterization of Testicular Germ Cell Tumors. Cell Rep, 2018. 23(11): p. 3392-3406.
20. Chiappinelli, K.B., et al., Inhibiting DNA Methylation Causes an Interferon Response in Cancer via dsRNA Including Endogenous Retroviruses. Cell, 2017. 169(2): p. 361.
21. Chiu, H.S., et al., Pan-Cancer Analysis of lncRNA Regulation Supports Their Targeting of Cancer Genes in Each Tumor Context. Cell Rep, 2018. 23(1): p. 297-312 e12.
22. Kim, H., et al., Integrative genome analysis reveals an oncomir/oncogene cluster regulating glioblastoma survivorship. Proc Natl Acad Sci U S A, 2010. 107(5): p. 2183-8.
23. Sumazin, P., et al., An extensive microRNA-mediated network of RNA-RNA interactions regulates established oncogenic pathways in glioblastoma. Cell, 2011. 147(2): p. 370-81.
24. Wang, Z., et al., lncRNA Epigenetic Landscape Analysis Identifies EPIC1 as an Oncogenic lncRNA that Interacts with MYC and Promotes Cell-Cycle Progression in Cancer. Cancer Cell, 2018. 33(4): p. 706-720 e9.
25. Yang, D., et al., Integrated analyses identify a master microRNA regulatory network for the mesenchymal subtype in serous ovarian cancer. Cancer Cell, 2013. 23(2): p. 186-99.

26. Bolton, K.L., et al., Association between BRCA1 and BRCA2 mutations and survival in women with invasive epithelial ovarian cancer. JAMA, 2012. 307(4): p. 382-90.
27. Davoli, T., et al., Tumor aneuploidy correlates with markers of immune evasion and with reduced response to immunotherapy. Science, 2017. 355(6322).
28. Zack, T.I., et al., Pan-cancer patterns of somatic copy number alteration. Nat Genet, 2013. 45(10): p. 1134-40.
29. Alexandrov, L.B., et al., Signatures of mutational processes in human cancer. Nature, 2013. 500(7463): p. 415-21.
30. Sanchez-Vega, F., et al., Oncogenic Signaling Pathways in The Cancer Genome Atlas. Cell, 2018. 173(2): p. 321-337 e10.
31. Stegh, A.H., et al., Glioma oncoprotein Bcl2L12 inhibits the p53 tumor suppressor. Genes Dev, 2010. 24(19): p. 2194-204.
32. Ellrott, K., et al., Scalable Open Science Approach for Mutation Calling of Tumor Exomes Using Multiple Genomic Pipelines. Cell Syst, 2018. 6(3): p. 271-281 e7.
33. Masica, D.L. and R. Karchin, Correlation of somatic mutation and expression identifies genes important in human glioblastoma progression and survival. Cancer Res, 2011. 71(13): p. 4550-61.
34. Schmidt, D., et al., Genome-scale validation of deep-sequencing libraries. PLoS One, 2008. 3(11): p. e3713.
35. Robertson, G., et al., Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. Nat Methods, 2007. 4(8): p. 651-7.
36. Deardorff, M.A., et al., HDAC8 mutations in Cornelia de Lange syndrome affect the cohesin acetylation cycle. Nature, 2012. 489(7415): p. 313-7.
37. Ernst, J., et al., Mapping and analysis of chromatin state dynamics in nine human cell types. Nature, 2011. 473(7345): p. 43-9.
38. Hansen, P., et al., Saturation analysis of ChIP-seq data for reproducible identification of binding peaks. Genome Res, 2015. 25(9): p. 1391-400.
39. Izumi, K., et al., Germline gain-of-function mutations in AFF4 cause a developmental syndrome functionally linking the super elongation complex and cohesin. Nat Genet, 2015. 47(4): p. 338-44.
40. Jeppsson, K., et al., The chromosomal association of the Smc5/6 complex depends on cohesion and predicts the level of sister chromatid entanglement. PLoS Genet, 2014. 10(10): p. e1004680.
41. Mikkelsen, T.S., et al., Comparative epigenomic analysis of murine and human adipogenesis. Cell, 2010. 143(1): p. 156-69.
42. Schaub, M.A., et al., Linking disease associations with regulatory information in the human genome. Genome Res, 2012. 22(9): p. 1748-59.
43. Shang, W.H., et al., Chromosome engineering allows the efficient isolation of vertebrate neocentromeres. Dev Cell, 2013. 24(6): p. 635-48.

44. Sutani, T., et al., Condensin targets and reduces unwound DNA structures associated with transcription in mitotic chromosome condensation. Nat Commun, 2015. 6: p. 7815.
45. Zuin, J., et al., A cohesin-independent role for NIPBL at promoters provides insights in CdLS. PLoS Genet, 2014. 10(2): p. e1004153.
46. Barski, A., et al., High-resolution profiling of histone methylations in the human genome. Cell, 2007. 129(4): p. 823-37.
47. Johnson, D.S., et al., Genome-wide mapping of in vivo protein-DNA interactions. Science, 2007. 316(5830): p. 1497-502.
48. Ren, B., et al., Genome-wide location and function of DNA binding proteins. Science, 2000. 290(5500): p. 2306-9.
49. Kim, T.H., et al., A high-resolution map of active promoters in the human genome. Nature, 2005. 436(7052): p. 876-80.
50. Rhee, H.S. and B.F. Pugh, Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution. Cell, 2011. 147(6): p. 1408-19.
51. Meyer, C.A. and X.S. Liu, Identifying and mitigating bias in next-generation sequencing methods for chromatin biology. Nat Rev Genet, 2014. 15(11): p. 709-21.
52. Furey, T.S., ChIP-seq and beyond: new and improved methodologies to detect and characterize protein-DNA interactions. Nat Rev Genet, 2012. 13(12): p. 840-52.
53. Schmidt, D., et al., ChIP-seq: using high-throughput sequencing to discover protein-DNA interactions. Methods, 2009. 48(3): p. 240-8.
54. Clough, E. and T. Barrett, The Gene Expression Omnibus Database. Methods Mol Biol, 2016. 1418: p. 93-110.
55. Fraga, M.F. and M. Esteller, DNA methylation: a profile of methods and applications. Biotechniques, 2002. 33(3): p. 632, 634, 636-49.
56. Olek, A., J. Oswald, and J. Walter, A modified and improved method for bisulphite based cytosine methylation analysis. Nucleic Acids Res, 1996. 24(24): p. 5064-6.
57. Grunau, C., S.J. Clark, and A. Rosenthal, Bisulfite genomic sequencing: systematic investigation of critical experimental parameters. Nucleic Acids Res, 2001. 29(13): p. E65-5.
58. Ehrich, M., et al., A new method for accurate assessment of DNA quality after bisulfite treatment. Nucleic Acids Res, 2007. 35(5): p. e29.
59. Kriaucionis, S. and N. Heintz, The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. Science, 2009. 324(5929): p. 929-30.
60. Tahiliani, M., et al., Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science, 2009. 324(5929): p. 930-5.
61. Huang, Y., et al., The behaviour of 5-hydroxymethylcytosine in bisulfite sequencing. PLoS One, 2010. 5(1): p. e8888.
62. Booth, M.J., et al., Quantitative sequencing of 5-methylcytosine and 5-hydroxymethylcytosine at single-base resolution. Science, 2012. 336(6083): p. 934-7.

63. Yu, M., et al., Tet-assisted bisulfite sequencing of 5-hydroxymethylcytosine. Nat Protoc, 2012. 7(12): p. 2159-70.
64. Langmead, B., et al., Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol, 2009. 10(3): p. R25.
65. Li, H. and R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics, 2009. 25(14): p. 1754-60.
66. Chung, D., et al., Discovering transcription factor binding sites in highly repetitive regions of genomes with multi-read analysis of ChIP-Seq data. PLoS Comput Biol, 2011. 7(7): p. e1002111.
67. Day, D.S., et al., Estimating enrichment of repetitive elements from high-throughput sequence data. Genome Biol, 2010. 11(6): p. R69.
68. Valouev, A., et al., Genome-wide analysis of transcription factor binding sites based on ChIP-Seq data. Nat Methods, 2008. 5(9): p. 829-34.
69. Pepke, S., B. Wold, and A. Mortazavi, Computation for ChIP-seq and RNA-seq studies. Nat Methods, 2009. 6(11 Suppl): p. S22-32.
70. Laajala, T.D., et al., A practical comparison of methods for detecting transcription factor binding sites in ChIP-seq experiments. BMC Genomics, 2009. 10: p. 618.
71. Liu, E.T., S. Pott, and M. Huss, Q&A: ChIP-seq technologies and the study of gene regulation. BMC Biol, 2010. 8: p. 56.
72. Park, P.J., ChIP-seq: advantages and challenges of a maturing technology. Nat Rev Genet, 2009. 10(10): p. 669-80.
73. Wilbanks, E.G. and M.T. Facciotti, Evaluation of algorithm performance in ChIP-seq peak detection. PLoS One, 2010. 5(7): p. e11471.
74. Machanick, P. and T.L. Bailey, MEME-ChIP: motif analysis of large DNA datasets. Bioinformatics, 2011. 27(12): p. 1696-7.
75. McLean, C.Y., et al., GREAT improves functional interpretation of cis-regulatory regions. Nat Biotechnol, 2010. 28(5): p. 495-501.
76. Thomas-Chollier, M., et al., RSAT peak-motifs: motif analysis in full-size ChIP-seq datasets. Nucleic Acids Res, 2012. 40(4): p. e31.
77. Welch, R.P., et al., ChIP-Enrich: gene set enrichment testing for ChIP-seq data. Nucleic Acids Res, 2014. 42(13): p. e105.
78. Matys, V., et al., TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. Nucleic Acids Res, 2006. 34(Database issue): p. D108-10.
79. Bailey, T.L., et al., MEME: discovering and analyzing DNA and protein sequence motifs. Nucleic Acids Res, 2006. 34(Web Server issue): p. W369-73.
80. Gordon, D.B., et al., TAMO: a flexible, object-oriented framework for analyzing transcriptional regulation using DNA-sequence motifs. Bioinformatics, 2005. 21(14): p. 3164-5.
81. Liu, X.S., D.L. Brutlag, and J.S. Liu, An algorithm for finding protein-DNA binding sites with applications to chromatin-immunoprecipitation microarray experiments. Nat Biotechnol, 2002. 20(8): p. 835-9.

82. Pavesi, G., et al., Weeder Web: discovery of transcription factor binding sites in a set of sequences from co-regulated genes. Nucleic Acids Res, 2004. 32(Web Server issue): p. W199-203.
83. Stormo, G.D., DNA binding sites: representation and discovery. Bioinformatics, 2000. 16(1): p. 16-23.
84. Wu, J. and J. Xie, Hidden Markov model and its applications in motif findings. Methods Mol Biol, 2010. 620: p. 405-16.
85. Bryne, J.C., et al., JASPAR, the open access database of transcription factor-binding profiles: new content and tools in the 2008 update. Nucleic Acids Res, 2008. 36(Database issue): p. D102-6.
86. Hume, M.A., et al., UniPROBE, update 2015: new tools and content for the online database of protein-binding microarray data on protein-DNA interactions. Nucleic Acids Res, 2015. 43(Database issue): p. D117-22.
87. Kulakovskiy, I.V., et al., HOCOMOCO: towards a complete collection of transcription factor binding models for human and mouse via large-scale ChIP-Seq analysis. Nucleic Acids Res, 2018. 46(D1): p. D252-D259.
88. Mathelier, A., et al., JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles. Nucleic Acids Res, 2016. 44(D1): p. D110-5.
89. Wingender, E., et al., TRANSFAC: a database on transcription factors and their DNA binding sites. Nucleic Acids Res, 1996. 24(1): p. 238-41.
90. Consortium, E.P., A user's guide to the encyclopedia of DNA elements (ENCODE). PLoS Biol, 2011. 9(4): p. e1001046.
91. Wasserman, W.W. and A. Sandelin, Applied bioinformatics for the identification of regulatory elements. Nat Rev Genet, 2004. 5(4): p. 276-87.
92. Kumar, V., et al., Uniform, optimal signal processing of mapped deep-sequencing data. Nat Biotechnol, 2013. 31(7): p. 615-22.
93. Liberzon, A., et al., The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst, 2015. 1(6): p. 417-425.
94. Subramanian, A., et al., Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A, 2005. 102(43): p. 15545-50.
95. Mootha, V.K., et al., PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nat Genet, 2003. 34(3): p. 267-73.
96. Angelini, C. and V. Costa, Understanding gene regulatory mechanisms by integrating ChIP-seq and RNA-seq data: statistical solutions to biological problems. Front Cell Dev Biol, 2014. 2: p. 51.
97. Zhou, V.W., A. Goren, and B.E. Bernstein, Charting histone modifications and the functional organization of mammalian genomes. Nat Rev Genet, 2011. 12(1): p. 7-18.
98. Butte, A., The use and analysis of microarray data. Nat Rev Drug Discov, 2002. 1(12): p. 951-60.

99. Conway, T. and G.K. Schoolnik, Microarray expression profiling: capturing a genome-wide portrait of the transcriptome. Mol Microbiol, 2003. 47(4): p. 879-89.
100. Li, X., et al., DNA microarrays: their use and misuse. Microcirculation, 2002. 9(1): p. 13-22.
101. Adomas, A., et al., Comparative analysis of transcript abundance in Pinus sylvestris after challenge with a saprotrophic, pathogenic or mutualistic fungus. Tree Physiol, 2008. 28(6): p. 885-97.
102. Moran, G., et al., Comparative genomics using Candida albicans DNA microarrays reveals absence and divergence of virulence-associated genes in Candida dubliniensis. Microbiology, 2004. 150(Pt 10): p. 3363-82.
103. Pollack, J.R., et al., Genome-wide analysis of DNA copy-number changes using cDNA microarrays. Nat Genet, 1999. 23(1): p. 41-6.
104. Hacia, J.G., et al., Determination of ancestral alleles for human single-nucleotide polymorphisms using high-density oligonucleotide arrays. Nat Genet, 1999. 22(2): p. 164-7.
105. Rai, M.F., et al., Advantages of RNA-seq compared to RNA microarrays for transcriptome profiling of anterior cruciate ligament tears. J Orthop Res, 2018. 36(1): p. 484-497.
106. Chernov, A.V., et al., Microarray-based transcriptional and epigenetic profiling of matrix metalloproteinases, collagens, and related genes in cancer. J Biol Chem, 2010. 285(25): p. 19647-59.
107. Colyer, H.A., R.N. Armstrong, and K.I. Mills, Microarray for epigenetic changes: gene expression arrays. Methods Mol Biol, 2012. 863: p. 319-28.
108. Essaghir, A., et al., Transcription factor regulation can be accurately predicted from the presence of target gene signatures in microarray gene expression data. Nucleic Acids Res, 2010. 38(11): p. e120.
109. Gorte, M., et al., Microarray-based identification of transcription factor target genes. Methods Mol Biol, 2011. 754: p. 119-41.
110. Shalon, D., S.J. Smith, and P.O. Brown, A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization. Genome Res, 1996. 6(7): p. 639-45.
111. Tang, T., et al., Expression ratio evaluation in two-colour microarray experiments is significantly improved by correcting image misalignment. Bioinformatics, 2007. 23(20): p. 2686-91.
112. Brown, P.O. and D. Botstein, Exploring the new world of the genome with DNA microarrays. Nat Genet, 1999. 21(1 Suppl): p. 33-7.
113. Xiang, Z., et al., Microarray expression profiling: analysis and applications. Curr Opin Drug Discov Devel, 2003. 6(3): p. 384-95.
114. Nakanishi, T., T. Oka, and T. Akagi, Recent advances in DNA microarrays. Acta Med Okayama, 2001. 55(6): p. 319-28.
115. Quackenbush, J., Computational analysis of microarray data. Nat Rev Genet, 2001. 2(6): p. 418-27.

116. Simon, R., M.D. Radmacher, and K. Dobbin, Design of studies using DNA microarrays. Genet Epidemiol, 2002. 23(1): p. 21-36.
117. Slonim, D.K., From patterns to pathways: gene expression data analysis comes of age. Nat Genet, 2002. 32 Suppl: p. 502-8.
118. Dopazo, J., et al., Methods and approaches in the analysis of gene expression data. J Immunol Methods, 2001. 250(1-2): p. 93-112.
119. Bakay, M., et al., Sources of variability and effect of experimental approach on expression profiling data interpretation. BMC Bioinformatics, 2002. 3: p. 4.
120. Bammler, T., et al., Standardizing global gene expression analysis between laboratories and across platforms. Nat Methods, 2005. 2(5): p. 351-6.
121. Mortazavi, A., et al., Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods, 2008. 5(7): p. 621-8.
122. Wang, Z., M. Gerstein, and M. Snyder, RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet, 2009. 10(1): p. 57-63.
123. Zhao, S., et al., Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells. PLoS One, 2014. 9(1): p. e78644.
124. Clark, T.A., C.W. Sugnet, and M. Ares, Jr., Genomewide analysis of mRNA processing in yeast using splicing-specific microarrays. Science, 2002. 296(5569): p. 907-10.
125. Bottomly, D., et al., Evaluating gene expression in C57BL/6J and DBA/2J mouse striatum using RNA-Seq and microarrays. PLoS One, 2011. 6(3): p. e17820.
126. Fu, X., et al., Estimating accuracy of RNA-Seq and microarrays with proteomics. BMC Genomics, 2009. 10: p. 161.
127. Marioni, J.C., et al., RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. Genome Res, 2008. 18(9): p. 1509-17.
128. Sirbu, A., et al., RNA-Seq vs dual- and single-channel microarray data: sensitivity analysis for differential expression and clustering. PLoS One, 2012. 7(12): p. e50986.
129. Zhang, W., et al., Effector CD4+ T cell expression signatures and immune-mediated disease associated genes. PLoS One, 2012. 7(6): p. e38510.
130. Conesa, A., et al., A survey of best practices for RNA-seq data analysis. Genome Biol, 2016. 17: p. 13.
131. Consortium, S.M.-I., A comprehensive assessment of RNA-seq accuracy, reproducibility and information content by the Sequencing Quality Control Consortium. Nat Biotechnol, 2014. 32(9): p. 903-14.
132. Law, C.W., et al., voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biol, 2014. 15(2): p. R29.
133. McCarthy, D.J., Y. Chen, and G.K. Smyth, Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. Nucleic Acids Res, 2012. 40(10): p. 4288-97.
134. Wilhelm, B.T. and J.R. Landry, RNA-Seq-quantitative measurement of expression through massively parallel RNA-sequencing. Methods, 2009. 48(3): p. 249-57.

135. Li, J., et al., Comparison of microarray and RNA-Seq analysis of mRNA expression in dermal mesenchymal stem cells. Biotechnol Lett, 2016. 38(1): p. 33-41.
136. Liu, Y., et al., RNA-Seq identifies novel myocardial gene expression signatures of heart failure. Genomics, 2015. 105(2): p. 83-9.
137. Wang, C., et al., The concordance between RNA-seq and microarray data depends on chemical treatment and transcript abundance. Nat Biotechnol, 2014. 32(9): p. 926-32.

Chapter Three: Somatic Copy Number Amplification and Hyperactivating Somatic Mutations of EZH2 Correlate with DNA Methylation and Drive Epigenetic Silencing of Genes Involved in Tumor Suppression and Immune Responses in Melanoma.

Utilizing the Cancer Genome Atlas (TCGA) skin cutaneous melanoma (SKCM) dataset a subset of 421 patients were observed to have the epigenetic regulator EZH2 of the Polycomb Repressive 2 complex to display oncogenic behavior through the mechanisms of either hyper activating mutations increasing its methyltransferase function, the EZH2 gene loci being amplified, and increased transcription. Patients containing these activating mutational events or increased EZH2 expression resulted in adverse patient survival compared to patients without these mutational events and normal EZH2 expression. As EZH2 is able to repress gene expression by trimethylating the histone 3 lysine 27 (H3K27) residue of nearby histones, and by recruiting DNA methyltransferases to methylate CpG islands. This paper attempts to identify transcriptional targets impacted by EZH2's oncogenic activation by examining the RNAseq and methylseq datasets from the identified patient cohort, and examining which genes became derepressed in cultured cells with the activating Y641 EZH2 mutation upon treatment with GSK126. The candidate genes identified are part of the tumor suppression, cell differentiation, cell cycle inhibition and repression of metastases as well as antigen processing and presentation pathways. These results highlight a systems biology methodology for identifying the oncogenic function of an epigenetic regulator using patient data from a large-scale consortium. The epigenetic repression of genes within the pathways identified within this study displays an epigenetic mechanism for a cancer cell to achieve hallmarks of cancer such as evading immune system detection, avoiding apoptosis, and avoiding senescence. Within this study I contributed computationally to the quantification of which EZH2 transcript was the most prominent in the cell, overlapping gene sets, and the gene set enrichment analysis of the cellular pathways impacted by EZH2 activity (figures 1,2,5).

NEOPLASIA
www.neoplasia.com
Volume 18 Number 2 February 2016 pp. 121–132 121

Somatic Copy Number Amplification and Hyperactivating Somatic Mutations of EZH2 Correlate With DNA Methylation and Drive Epigenetic Silencing of Genes Involved in Tumor Suppression and Immune Responses in Melanoma

Jessamy Tiffen*,1, Stephen Wilson†,1, Stuart J. Gallagher*, Peter Hersey* and Fabian V. Filipp†

*Melanoma Immunology and Oncology Group, The Centenary Institute, University of Sydney, Camperdown, NSW, Australia; †Systems Biology and Cancer Metabolism, Program for Quantitative Systems Biology, University of California Merced, Merced, CA 95343, USA

Abstract

The epigenetic modifier EZH2 is in the center of a repressive complex controlling differentiation of normal cells. In cancer EZH2 has been implicated in silencing tumor suppressor genes. Its role in melanoma as well as target genes affected by EZH2 are poorly understood. In view of this we have used an integrated systems biology approach to analyze 471 cases of skin cutaneous melanoma (SKCM) in The Cancer Genome Atlas (TCGA) for mutations and amplifications of EZH2. Identified changes in target genes were validated by interrogation of microarray data from melanoma cells treated with the EZH2 inhibitor GSK126. We found that EZH2 activation by mutations, gene amplification and increased transcription occurred in about 20% of the cohort. These alterations were associated with significant hypermethylation of DNA and significant downregulation of 11% of transcripts in patient RNASeq data. GSK126 treatment of melanoma lines containing EZH2 activation reversed such transcriptional repression in 98 candidate target genes. Gene enrichment analysis revealed genes associated with tumor suppression, cell differentiation, cell cycle inhibition and repression of metastases as well as antigen processing and presentation pathways. The identified changes in EZH2 were associated with an adverse prognosis in the TCGA dataset. These results suggest that inhibiting of EZH2 is a promising therapeutic avenue for a substantial fraction of melanoma patients.

Neoplasia (2016) 18, 121–132

Introduction

During cancer progression a tissue-specific dedifferentiation towards an immortal state takes place [1], a change that requires concerted alterations at the genomic, epigenomic, and transcriptional level [2]. The polycomb repressive complex (PRC) 2 is instrumental for chromatin remodeling and recruitment of proteins required for epigenetic modifications [1,3]. Crucial to PRC2 activity, the histone methyltransferase enhancer of zeste homolog 2 (EZH2) [GenBank:2146] tri-methylates lysine 27 of histone 3 (H3K27me3), leading to chromatin condensation and transcriptional repression. EZH2 can also direct DNA methylation via recruitment of DNA methyltransferases (DNMTs), thus linking histone methylation to DNA methylation [3]. The cellular networks targeted by EZH2 are essential in early development but downregulated in normal adult tissues.

In many types of cancers including lymphomas and leukemia, EZH2 is postulated to exert its oncogenic effects via aberrant histone and DNA methylation, causing silencing of tumor suppressor genes [4–9]. Recent studies have identified reversible H3K27me3 levels in response to aberrant EZH2 activity in melanoma suggesting suitability for pharmacological targeting [10–14]. In particular our recent studies have shown that small molecule inhibitors of EZH2 could induce cell cycle arrest and apoptosis of melanoma cells harboring somatic mutations of EZH2 [14].

Address all correspondence to: Fabian V. Filipp, PhD, MSc, 2500 North Lake Road, Merced, CA 95343.
E-mail: filipp@ucmerced.edu
[1] Equally contributing authors.
Received 8 November 2015; Revised 21 December 2015; Accepted 4 January 2016

1476-5586
http://dx.doi.org/10.1016/j.neo.2016.01.003

In this study, we capitalize on the druggability of EZH2 and shed light on its role as an epigenetic regulator. We apply a comprehensive systems biology approach to the skin cutaneous melanoma (SKCM) dataset of 471 patients and in total to 12366 Pan-cancer specimens of 32 tissues of The Cancer Genome Atlas (TCGA). We connect somatic mutations and somatic copy number alterations (SCNAs) of EZH2 to epigenetic and transcriptional control of its target genes. Methylation status and transcriptional activity of target genes is combined with the transcriptional response of cellular melanoma models of activating EZH2 mutations to treatment with an EZH2 inhibitor. The rationale behind combining transcriptional data from inhibitor studies is to reveal or confirm genes repressed by EZH2 activation.

Methods

We utilized files from 471 SNP arrays, 120 whole-genome, 339 whole-exome, and 440 clinical datasets with normal reference samples from 471 TCGA SKCM patients. In addition, we selected 458 patients of the SKCM cohort with complete methylome and transcriptome data. Genomic regions of TCGA SKCM data set aligned to HG19 were determined using the tool genomic identification of significant targets in cancer 2.0.21 at confidence level of 0.99 and cutoff q-value of 0.01. Somatic mutation and somatic copy number alterations were assessed for 32 different cancer tissues covering a total cohort size of 9833 and 6506 TCGA patients for somatic copy number alteration data and whole exome sequencing data, respectively (Supplementary Table 1). The study was carried out as part of IRB of the University of California Merced approved study dbGap ID 5094 "Somatic mutations in melanoma" and conducted in accordance with the Helsinki Declaration of 1975. The results shown are based upon next generation sequencing data generated by the TCGA Research Network http://cancergenome.nih.gov. Restricted access clinical, RNASeq, and whole-exome sequences were obtained from the TCGA genome data access center and the data portal.

Illumina HiSeq 2000 V2 RNA Sequencing by expectation-maximization normalized Log2 data was filtered for differential expression in patients with activating EZH2 mutations in two-tailed Z-tests and p-values below 0.05 in 458 and 12633 patients in TCGA SKCM and Pan-cancer, respectively. Pearson's correlation coefficient was calculated for paired differential methylation and RNASeq data classified according to moderate negative correlation ($-0.2 \geq \rho > -0.4$) or strong negative correlation ($-0.4 \geq \rho$) and associated with methylation dependent transcriptional silencing. Pairwise average-linkage in combination with Pearson's correlation was used as distance measure for both, column (patients) and row (genes or markers) hierarchical clustering. Methylation data was thresholded for differential methylation in patients with activating EZH2 mutations in two-tailed Z-tests and p-values below 0.05. Differentially regulated methylation markers were mapped to HG19 and only kept if a gene association was detected at least twice. Statistical hypotheses testing, in detail Fisher's exact method, was used to determine significant enrichment of somatic mutation in given patient cohorts. Microarray analysis was performed on two melanoma cell lines, IGR1 and MM386, each conducted as duplicates 48 hours after treatment with DMSO as control or 7.5 μM GSK126 (IC 50 value between 5-8 μM) [PubChem CID:68210102]. Transcriptomic data were normalized using the normal-exponential deconvolution method, corrected for multiple samples in two-tailed T-tests, and adjusted p-values below 0.05. Pathway enrichment was determined using the web-based gene set analysis toolkit at p-values below 0.05 and mapped onto the Kyoto Encyclopedia of Genes and Genomes (KEGG).

Somatic mutations of selected PRC genes were called by multi-step filtering after cohort selection, mapping of human genome and patient specific somatic references, assessment of recurrence, evolutionary conversation, basal mutation rate based on frequency of mutations of introns vs exons, and structural analysis [10]. TCGA patients showed recurring mutations TCGA-BF-A1PV-01 EZH2(Y641N), TCGA-D9-A1JW-06 EZH2(Y641F), and TCGA-EE-A3AF-06 EZH2(Y641N). EZH2 Y641 mutant melanoma cell lines C001 and MM386 were a generous gift from Dr. Chris Schmidt and Dr. Nick Hayward, QIMR, Brisbane, Australia. IGR1 cells were gifted from Dr. David Adams of the Welcome Trust Sanger Institute, Cambridge, UK. EZH2 wild type status Y641 of MELJD, ME1007, and KMJR138 cell lines was confirmed by sequenom or sanger sequencing, while IGR1 had in-frame point mutations EZH2(Y641N), MM386(Y641H) and C001(Y641S). All cells lines were authenticated by short tandem repeat validation. Cells were cultured in Dulbecco's modified Eagle medium (DMEM) supplemented with 10% fetal calf serum (AusGeneX, Brisbane, Queensland, Australia) and Pen/Strep (Sigma, St. Louis, MO, USA). Human dermal fibroblasts (HDF) and human epithelial melanocytes (HEM) were used as untransformed controls (ATCC, Manassas, VA, USA). HEM were cultured in M254 containing HGMS and all cells were maintained at 37°C in 5% CO_2.

Cell pellets were lysed with radioimmunoprecipitation (RIPA) buffer and subjected to western blot analysis. Total protein was determined using a BCA assay (Bio-Rad, Hercules, CA, USA). Labeled bands were detected by Immune-Star horseradish peroxidase chemiluminescence kit (Bio-Rad, Hercules, CA, USA) and images were captured by the Fujifilm LAS-4000 image system. Antibodies used were as follows: EZH2 (5246, Cell Signaling, Danvers, MA, USA), Beta Actin (AC-74, Sigma, St. Louis, MO, USA), p21 (SC-397, Santa Cruz Biotechnology, Dallas, TX, USA). GSK126 [PubChem CID:68210102] was purchased from Medchemexpress (New Jersey, NJ, USA) and dissolved in DMSO that was used as the vehicle control in all experiments. Cells were seeded in 6 well plates and treated the following day with DMSO, 2.5 μM, 5 μM or 7.5 μM GSK126. After 48 hours RNA was extracted from cells using the RNeasyPlus mini prep kit (Qiagen, Hilden, Germany), quantified using a Nanodrop (Thermo Scientific, Waltham, MA, USA) and 1 μg was reverse transcribed using SuperScriptIII (Life Technologies, Carlsbad, CA, USA). cDNA was amplified using the AB7900 real-time quantitative PCR (RT-qPCR) system (Life Technologies, Carlsbad, CA, USA) using Universal PCR Mastermix and Taqman probes (Life Technologies, Carlsbad, CA, USA) specific for CDKN1A [GeneBank:1026] (Hs00355782_m1) and normalized to levels of 18 s (Hs99999901_s1).

Results

Enrichment of Activating Somatic Mutations of EZH2 in Melanoma

EZH2 has somatic mutations in several regions of the molecule. The proportion of mutations that affect the different domains of EZH2 is similar between the melanoma cohort and the Pan-cancer average (DMNT binding domain 42. 9%/39.3%, CDYL binding domain 21.4%/16.7%, CXC domain 14.3%/13.1%, SET domain 21.4%/30.9% in SKCM and TCGA Pan-cancer respectively)

(Figure 1*A*, Supplementary Table 1). Alternative splice events in exons 3, 4, and 8 give rise to at least 5 mRNA isoforms (Figure 1*B*). In total, 100 coding mutations including 22 silent mutations were detected in EZH2: TCGA SKCM showed a mutation frequency of 4.1% in 339 exomes and TCGA Pan-cancer 1.3% in 6506 exomes. However, a somatic mutation in the center of the active site of the SET domain at residue Y641 was recurring. It was detected in 7 cases of all reported whole exome sequencing melanoma studies and 5 times across the TCGA Pan-cancer cohort. There were 2 cases in 278 metastatic tumors of TCGA SKCM; 1 case in 61 primary tumors of TCGA SKCM (Figure 1*C*); 2 cases in 121 metastatic Broad tumors; 2 cases in 91 metastatic Yale tumors [10,13,15]. The replacement of the JAK2 phosphorylation target residue Y641 alters EZH2 conformation, increasing its tri-methylation activity and stabilizing the protein [16,17]. Therefore it is expected that EZH2 Y641 mutants in melanoma cells exhibit increased H3K27me3 levels. High levels of activating mutations affecting the SET domain were also evident in other cancers such as Lymphoid Neoplasm Diffuse Large B-cell Lymphoma (DLBC) including 2 Y641 mutations, uterine corpus endometrial carcinoma (UCEC), lung cancer (LUAD), colorectal cancer (COAD), stomach cancer (STAD), and acute myeloid leukemia (LAML). Isoform-specific RNASeq analysis resolved splice variants of EZH2 (Figure 2*A*).

Major EZH2 Cancer Transcript NM_001203247 Lacks Extension of Exon 8, Hence Down-Stream Somatic Mutation Target Location is at Y641

Next, we investigated the TCGA RNASeq dataset to identify which isoforms of EZH2 were predominantly transcribed. Across the TCGA PAN-cancer panel, the predominantly transcribed isoform is mRNA transcript variant 3 [GeneBank:NM_001203247], 746 amino acids [GeneBank:NP_001190176], corresponding Uniprot entry [Swiss-Prot:Q15910-1], lacking a short segment at the end of exon 8 coding for 5 amino acids (Figure 2*A*). This is in contrast to the longest and first described transcript variant 1 [GeneBank:NM_004456], 751 amino acids [GeneBank:NP_004447], corresponding Uniprot isoform 2 [GeneBank:Q15910-2], [18]. While consistently being the major transcript across TCGA Pan-cancer as well as SKCM, Melanoma patients show 1.46-fold higher expression of the major EZH2 transcript compared to the TCGA Pan-cancer median in Z-tests with p-values below 1.0e-10 (Figure 2*A*). Patients with Y641 mutations share the common transcript [GeneBank:NM_001203247] and no switching of the 5' splice donor site of exon 8 was detected. Other alternative minor transcripts detected in cancer lacked mRNA segments in exon 3, 4, and 14 result in shorted isoforms, [GeneBank:NM_152998], 707 amino acids and [GeneBank:NM_001203248], 737 amino acids,

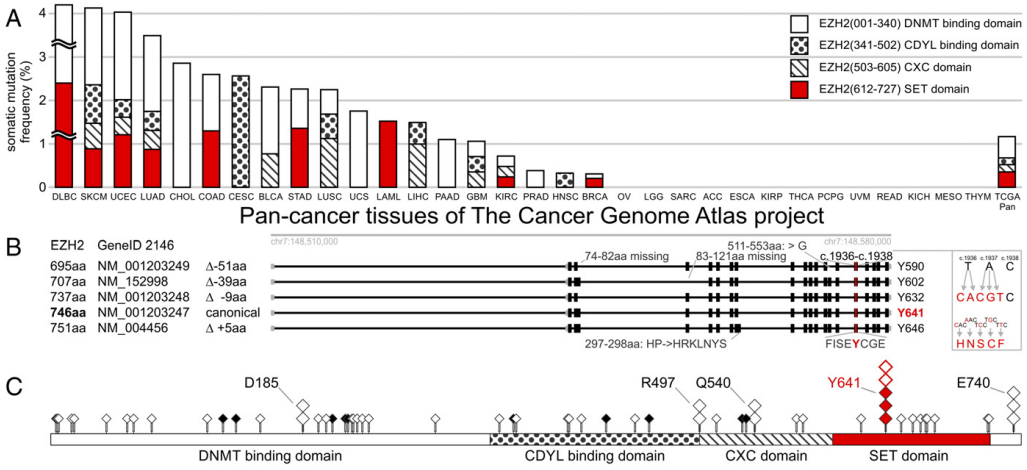


Figure 1. Distribution of somatic mutations in EZH2 in skin cutaneous melanoma (SKCM) and across Pan-cancer tissues of The Cancer Genome Atlas (TCGA) project. (A) EZH2 has a high somatic mutation frequency in melanoma of Pan-cancer tissues within TCGA. The frequency of non-silent somatic mutations (number of observed somatic mutations divided by cohort size) in different human cancer tissues within TCGA is sorted from most to least frequent. The fraction of affected protein domains of EZH2 is shaded white for DNMT binding domain, dotted for CDYL binding domain, striped for CXC domain, and red for SET domain. Lymphoid Neoplasm Diffuse Large B-cell Lymphoma (DLBC) has a mutation frequency of 14.6% exceeding the chosen y-axis range. Tildes indicate truncation of the DLBC data, while maintaining the relative domain distribution of observed DLBC mutations in DNMT and SET domains. TCGA Pan-cancer average across 6506 specimens is displayed at the very right of panel. (B) Isoform-specific transcripts of EZH2 affect numbering of somatic mutations in the C-terminal SET domain. Total amino acid count, NCBI Gene ID, amino acid difference to isoform NM_001203247, exon usage, and numbering of tyrosine in the SET-domain are provided. Somatic mutation observed in cDNA positions c.1936-1938 are boxed. (C) Somatic mutations of EZH2 in TCGA melanoma and Pan-cancer cohorts are plotted onto its amino acid sequence as filled and empty diamonds, respectively. Amino acid residues with recurring mutations across TCGA Pan-cancer tissues are shown as stacks of multiple diamonds: Y641 in red has three recurrences in melanoma, and two in lymphoma; D185H, R497Q, Q540P/*, K660E/R, and E740K/fs, each have recurrences in TCGA Pan-cancer.

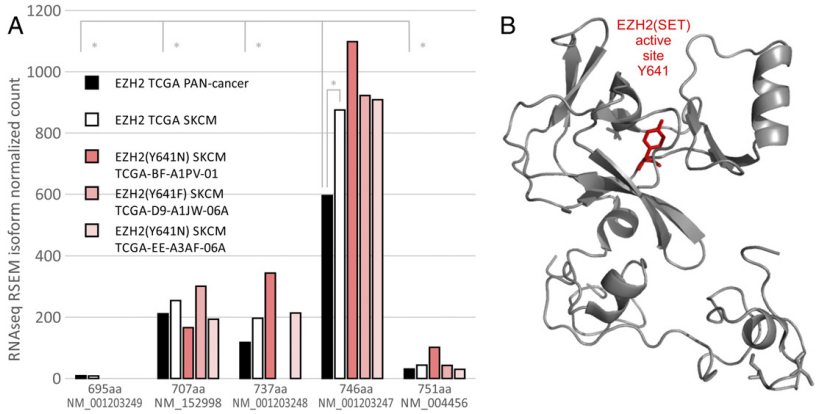


Figure 2. Isoform-specific RNASeq data across PAN-cancer panel assigns recurring somatic mutation of EZH2 in cancer to Y641 in center of SET domain. (A) Distribution of RNASeq data show isoform-specific counts for PAN-cancer and TCGA SKCM patients across 12633 and 458 specimens, respectively (somatic genotype and TCGA patient ID are indicated). Transcript NM_001203247 is the predominant isoform of wildtype and somatically mutated EZH2. We recommend mRNA transcript NM_001203247 and its corresponding 746 amino acid-long protein as canonical sequence framework. Asterisk indicated significant difference by Z-test with p-value below 1.0e-10 between isoforms or between TCGA Pan-cancer and SKCM cohorts. (B) The residue with the most frequent recurring somatic mutations of the polycomb repressive complex in SKCM (as well as in TCGA) is residue Y641 located in the active site of the SET domain of EZH2 plotted on ribbon structure of 4mi0.pdb.

which did not affect the SET-domain. Given the consistent and predominant transcriptional pattern of EZH2, we recommend cataloging of isoform [GeneBank:NM_001203247] as canonical transcript (Figures 1*B* and 2*A*). Numbering according to this transcript identifies the recurring somatic mutation in the SET-domain as Y641 (Figure 2*B*).

Significant Copy Number Gain of EZH2 in Melanoma

In addition to activating somatic mutations, TCGA melanoma patients showed strong somatic copy number alterations of EZH2 (Figure 3*A*). The EZH2 locus is over-expressed at the level of somatic copy number alterations (5.0% significant SCNA gain with a q-value below 0.01), which is higher than the TCGA Pan-cancer average of 1.8% (Figure 3*A*, Supplementary Table 2). The EZH2 locus at band 7q36.1 chr7:148,504,464-148,581,441 is in close proximity (8Mbp) to BRAF [GeneBank:673] and CDK5 [GeneBank:1020]. These and other proto-oncogenes share a hotspot at the end of the q-arm of chromosome 7 which is implicated in a significant amplification event with q-values below 10e-05 at a frequency of more than 50% for WGS (120 patients) and SNP-arrays (471 patients) (Figure 3*B*). Transcriptional levels of EZH2 are also increased in a subset of melanoma patients and 14.2% of patients showed differentially upregulated RNASeq levels. As expected, SCNA amplification of EZH2 correlates with mRNA upregulation and patients with SNCA gains showed increased EZH2 expression at the RNASeq level (Figure 3*C*). In comparison to Pan-cancer levels, the normalized RNASeq reads of EZH2 of SKCM patients are elevated by 46.7% (Figure 2*A*). Further, the pattern of SCNAs in melanoma is tightly linked to its mutational signature and tumor samples of patients with EZH2 and BRAF mutations are significantly enriched in somatic copy number amplifications with p-values below 0.05 and q-values below 10e-05 in comparison to BRAF wild type status (Figure 3*C*). EZH2 activation correlates with decreased patient survival (Figure 4). Kaplan–Meier curves of patients with EZH2 activating mutation, EZH2 amplification, or high normalized RNASeq counts show reduced patient survival of 1.8 years in comparison to curves of patients with EZH2 wild-type status (Figure 4*A*). RNASeq expression levels of EZH2 correlate with reduced overall median survival of melanoma patients of 1.8 years (Figure 4*B*).

Multi-Omics Study of Cancer Patient Data Correlates Epigenomic Silencing, DNA Hypermethylation, and Transcriptional Derepression by EZH2

We sought to identify EZH2 target genes in melanoma by combining three separate data sets. We combined gene transcription and DNA methylation data from the SKCM cohort in the TCGA with our own gene expression data based on response to small molecule inhibition of EZH2 in melanoma cell lines with activating mutations (Figure 5).

In SKCM, about 1/5th of the patient cohort is affected by somatic mutations, somatic copy number amplification, or transcriptional upregulation of EZH2. These tumors were considered likely to have high EZH2 activity and were compared to all other SKCM tumors. Comparing these cohorts, we overlaid gene sets of the TCGA SKCM transcriptome and methylome data looking for genes with both decreased expression and increased DNA methylation (Figure 5, *A–B*). In addition to histone tri-methylation, which compacts chromatin structure and represses transcription, EZH2 recruits DNA methyltransferases to methylate DNA and thereby enhances transcriptional repression. We therefore considered that EZH2 target genes would show decreased transcription and increased DNA methylation levels in the patient tumor data. To further increase confidence that these were EZH2 targets, we combined these results with expression array results from melanoma cells treated with the EZH2 inhibitor GSK126 [PubChem CID:68210102] (Figure 5, *A–C*).

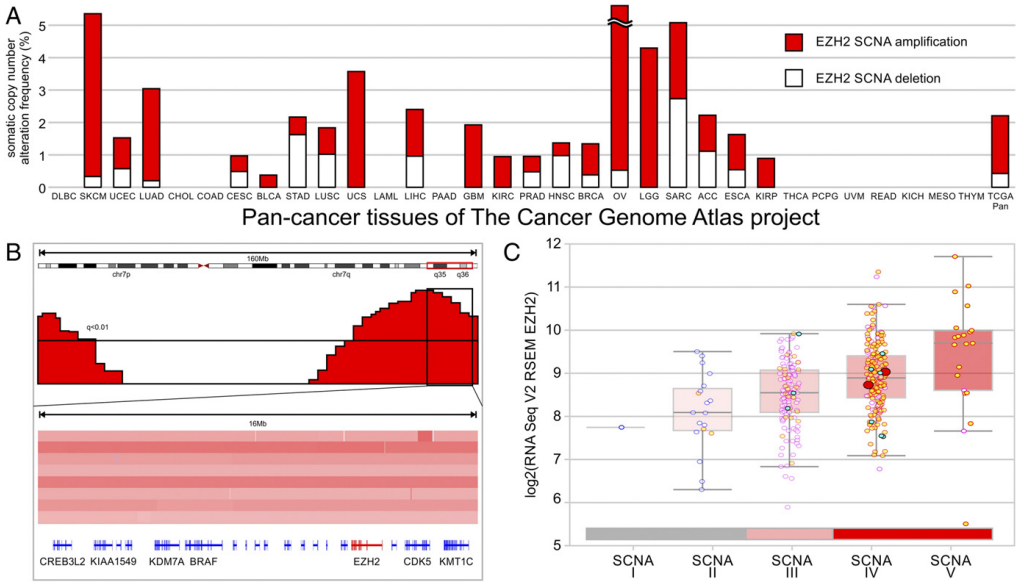


Figure 3. Somatic copy number amplification of EZH2 results in hyperactivation of methyltransfer activity of EZH2. (A) EZH2 is predominantly amplified at the somatic copy number level across a comprehensive panel of TCGA Pan-cancer patients. Frequency of EZH2 somatic copy number alterations (SCNAs) is shown in red for amplifications and white for deletions. Ovarian serous cystadenocarcinoma (OV) has a SCNAs frequency of 11.4% exceeding the chosen y-axis range indicated by tilde. TCGA Pan-cancer average across 9833 specimens is displayed at the very right of panel. (B) Amplification of chromosome band 7q includes EZH2 and other oncogenes. Segmented SNP array data of 8 representative patients is plotted relative to genomic coordinates. Bands 7q35 and 7q36 are framed on chromosome 7 and exonic regions of EZH2 are highlighted in red. Significance threshold of q = 0.01 is indicated as black line. (C) Somatic copy number alterations of EZH2 are plotted against the mRNA expression level of EZH2. Status of somatic copy number alterations are classified as deep deletions (SCNA I), shallow deletions (SCNA II), diploid (SCNA III), gain (SCNA IV), or amplification (SCNA V). Somatic copy number amplification is marked with red bar below data points. Deletions are circled in blue, amplifications in red. Patients with EZH2 mutations are highlighted as filled cyan circles and those with the Y641 mutation are shown as large red circles. Patients with BRAF mutations are marked with filled yellow circles.

As expected, patients with high EZH2 activity had differentially expressed genes that were mostly decreased (55.8% down) (Figures 5*B*, 6). Specifically, there were 2052 of 18088 transcripts differentially down regulated. Heatmaps of RNASeq data of top 50 genes with somatic EZH2 alterations visualize the transcriptional response (Figure 6). Also as expected, these patients showed mostly increased gene hypermethylation (74.7% up)—consisting of 82639 (61703 up, mapped to 12125 genes) of 485577 CG islands (Figure 5*B*).

Following treatment of melanoma cells with EZH2 inhibitor GSK126, H3K27me3 levels are expected to be decreased leading to increased transcriptional levels of EZH2 target genes. We performed protein immunoblotting to analyze the impact of EZH2 on H3K27me3 levels in melanoma cell lines. Common to all tested melanoma cell lines was presence of H3K27me3 that could be reduced by the EZH2 inhibitor GSK126 (Figure 7*A*). GSK126 is a competitive inhibitor of EZH2's methyltransferase activity that does not degrade EZH2 protein. Cell lines with active site mutations in the SET domain at Y641 had elevated H3K27me3 levels compared to wild-type cells. Upon treatment with GSK126, all tested cell lines had strongly reduced H3K27me3 levels (Figure 7*A*). In our experiments 539 transcripts showed significantly deregulated gene expression following GSK126 treatment with log2-fold changes above 2 and p-values below 0.05 (Figure 5*B*). 74.6% (402) of the deregulated genes were increased. Candidate EZH2 target genes were taken to be those both repressed in the SKCM cohort with EZH2 amplification or activation (threshold p-value below 0.05) and increase of transcription in melanoma cells following EZH2 inhibition by GSK126 (p-value below 0.05). The overlap of TCGA and drug response transcriptomes resulted in 98 responsive genes, of which 65 genes also showed DNA hypermethylation in TCGA tumors with high EZH2 activity (Figure 5*B* and *C*, Supplementary Table 3). These 65 candidate EZH2 target genes in melanoma included CDKN1A [GeneBank:1026] (Figures 7, *B*–*C* and 8), NDRG1 [GeneBank:10397], NFKB2 [GeneBank:4791], NFKBIA [GeneBank:4792], EPAS1 [GeneBank:2034], JUN [GeneBank:3725], JUND [GeneBank:3727], FOS [GeneBank:2353], FOSB [GeneBank:2354], IRF9 [GeneBank:10379], ITGA3 [GeneBank:3675], PLCG2 [GeneBank:5336], BCL6 [GeneBank:604], BOK [GeneBank:666], CD74 [GeneBank:972], HLA-A [GeneBank:3105], HLA-B [GeneBank:3106], HLA-DPA1 [GeneBank:3113], and HLA-F [GeneBank:3134] (Figure 8, Supplementary Table 4).

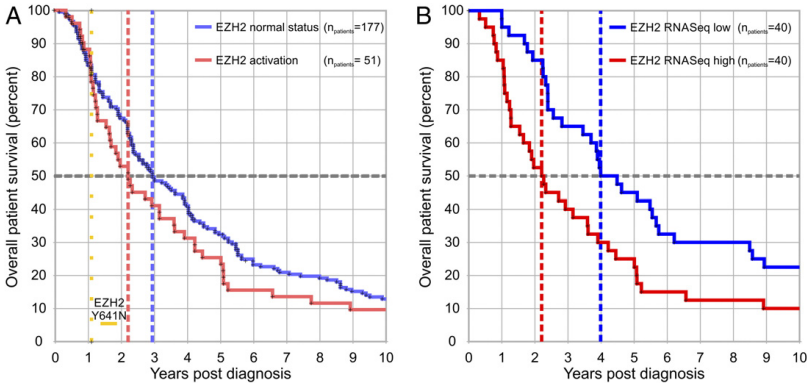


Figure 4. EZH2 is activated in melanoma and correlates with adverse patient survival. (A) Kaplan–Meier curves show shorter survival of TCGA SKCM patients with activated EZH2 (red) compared to patients with EZH2 wild-type status (blue). Median overall survival is indicated as dashed line (2.2 years post diagnosis for EZH2 activation defined as EZH2(Y641) mutation, SCNA amplification, or RNASeq upregulation; 3.0 years post diagnosis for EZH2 wild-type status; log-rank test p-value below 0.05). Survival record of patient with EZH2(Y641N) mutation is highlighted as dotted yellow line. RNASeq expression levels of EZH2 correlate with reduced overall median survival of melanoma patients. Patients with high EZH2 normalized RNASeq counts (red, top 20% of transcripts, median overall survival 2.2 years post diagnosis) show shorter survival in comparison to low EZH2 transcript levels (blue, bottom 20% of transcripts, median overall survival 4.0 years post diagnosis, log-rank test p-value below 0.05).

Multi-Omics Study Shows Enrichment of Transcriptional Silencing of Important Tumor Suppressors, Differentiation Factors, and Immune Response Genes

In order to assess any potential functional impact of high EZH2 activity in melanoma, we conducted a gene enrichment analysis of the 65 candidate EZH2 target genes derived from the above analysis. A KEGG pathway analysis showed that these candidate genes were significantly enriched in pathways deregulated in cancer including CDKN1A, NFKB2, NFKBIA, EPAS1, JUN, FOS, ITGA3, and PLCG2 (KEGG ID:05200; p-value = 6.33e-08; Supplementary Table 5). In addition, genes involved in antigen processing and presentation were significantly enriched (KEGG ID: 04612; p-value = 1.02e-05; Supplementary Table 5). We conducted a motif search of corresponding promotor and transcription start sites to identify transcriptional cooperation of PRC2. The identified target genes share a purine-rich motif of GGA(G/A)(G/A) at their promotors typical for recognition by E2F-related factors. We further assessed the DNA methylation pattern of this test group of silenced target genes. Unsupervised hierarchical clustering of their DNA methylation pattern recapitulated EZH2 status (Figure 9). The effect on DNA methylation was particularly pronounced in patients with activation mutations of EZH2 (Figure 10). Specifically we analyzed all CpG markers with significant elevation of methylation in patients with enhanced somatic EZH2 activity. The relative locations of methylation revealed accumulation of methylation sites at the transcription start site of candidate target genes. Besides chromatin histone tri-methylation by somatically activated EZH2 (Figure 7, Supplementary Table 4), such DNA hypermethylation pattern contributes to dampening the transcriptional response of affected target genes (Figures 6, 9, and 10).

Activating Somatic Mutations of EZH2 Y641 Contribute to Melanomagenesis by Tumor Suppressor Gene Silencing

Of the identified genes, we selected CDKN1A for characterization by RT-qPCR and western blot analysis (Figure 7). EZH2 may exhibit its oncogenic potential through CDKN1A by transcriptionally downregulating its gene product, the tumor suppressor and cell cycle regulator p21 (Figure 7*B*) [17,19]. While EZH2 is absent or lowly expressed in human epithelial melanocytes (HEM) or human dermal fibroblasts (HDF), it is highly expressed in melanoma cells MELJD, ME1007, KMJR138 (Figure 7*B*). We were able to identify three cellular melanoma models, IGR1, C001, and MM386, which recapitulate the Y641 point mutations. All three melanoma cell lines showed evidence of expression of EZH2 at the protein level (Figure 7*A*). The gene product of CDKN1A, p21, is absent in melanocytes but present in melanoma cells. In stark contrast, melanoma cell lines with activating EZH2 mutations show absence of p21 (Figure 7*B*). Using a small molecule EZH2 inhibitor, we monitored the transcriptional response of CDKN1A by RT-qPCR. MM386 and IGR-1 cells showed a dose dependent upregulation of CDKN1A by RT-qPCR (Figure 7*C*), following treatment with GSK126, which is a competitive inhibitor of EZH2s methyltransferase activity [20]. Taken together, the transcriptional reactivation of CDKN1A following EZH2 inhibition by GSK126 shows reversal of the transcriptional silencing observed with activating EZH2 mutations.

Interaction of EZH2 With DNA Remodelers

Lastly, we sought to investigate if EZH2 alterations co-occur with changes to other DNA remodelers. Overexpressed in many types of cancers, EZH2 is postulated to exert its oncogenic effects via aberrant methylation, causing silencing of tumor suppressor genes. It tri-methylates histones, ultimately causing gene repression by chromatin condensation thus blocking access of gene promotors to transcription initiation machinery and facilitates DNA methylation. The group of PRC genes, EZH1 [GeneBank:2145], EZH2 [GeneBank:2146], EED [GeneBank:8726], SUZ12 [GeneBank:23512], RBBP4 [GeneBank:5928], RBBP7 [GeneBank:5931], as well as PRC-associated DNMTs show high enrichment of somatic mutations affecting 10.5% and 15.5%, respectively of 343 patients

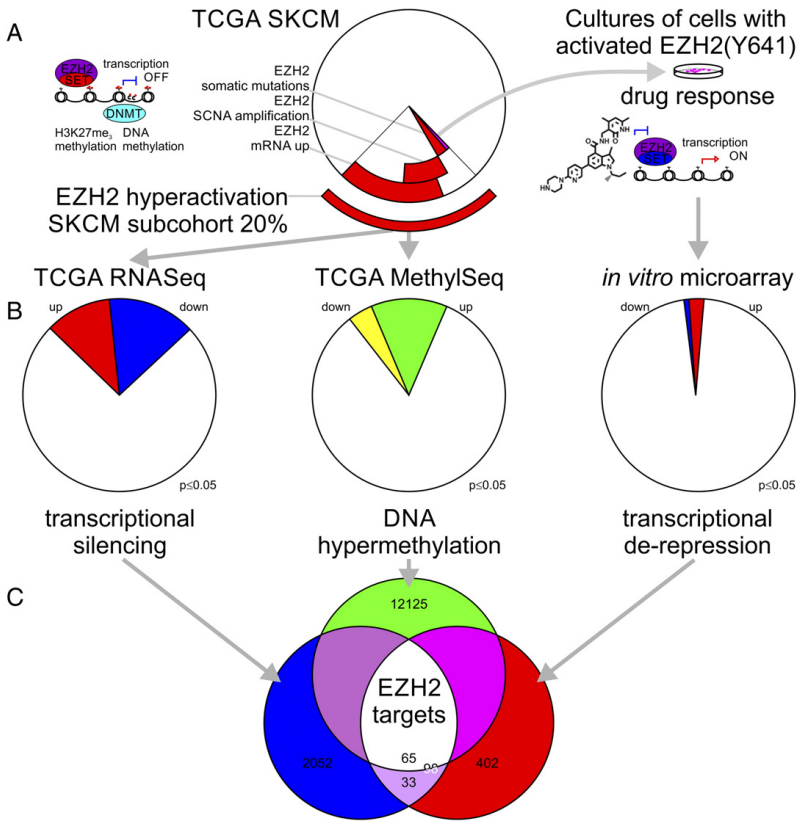


Figure 5. Multiomics analysis of EZH2 activation across different platform combined with responsiveness to drug treatment in melanoma. (A) The histone methyltransferase EZH2 transcriptionally silences target genes by tri-methylation of histone 3 K27. Cooperation with DNA methyltransferases helps manifest the transcriptional inactivation. 88 patients with EZH2 somatic mutations, somatic copy number amplification, and/or somatic mRNA upregulation were combined into a subhort comprising about 20% of all patients. Melanoma cells of TCGA patients with activating somatic mutations of EZH2(Y641) were cultured and subjected to a drug response assay using the EZH2 inhibitor GSK126. (B) Three genomics datasets, TCGA RNASeq, TCGA MethylSeq, and microarray of the drug response of cell lines, were filtered by p-value threshold below 0.05 and sorted by significantly up or down regulation. (C) EZH2 target genes show overlap of transcriptional silencing, DNA methylation, and responsiveness to drug exposure.

with whole exome sequencing data. The mutational signature of EZH2 (5.04% of SKCM patients) (Figure 1*A*) is mutually exclusive to non-synonymous somatic mutations in DNMT3A [GeneBank:1788] (occurring in 2.88% of patients) or DNMT3B [GeneBank:1789] (4.68% of patients) with an odds-ratio of 3.8, but is not strictly mutually exclusive to DNMT1 [GeneBank:1786] mutations (6.47% of patients) (Figure 10). EZH2-D142V in the DNMT binding domain and DNMT1-I531R co-occur as well as EZH2-C535W and DNMT1-R1466C. Further somatic mutations of PRC-associated DNMTs occur in methyltransfer domains, replication foci domains, and bromo-adjacent homology domains displaying an EZH2-distinct route of epigenetic changes in cancer. The mutual exclusivity (reflected in all cancers including blood-derived cancer with high EZH2 alterations LAML, acute myeloid leukemia, and DBLC, Lymphoid Neoplasm Diffuse Large B-cell Lymphoma) points towards a potential mechanism of somatic mutations in PRC2 driving cancer through EZH2 or alternatively DNMT3A/B (Figures 9, 10). Identification of genes whose expression is specifically modulated by EZH2, DNMT1 vs DNMT3A/B, and combination of EZH2 and DNMT inhibitors are the next logical steps in this line of research. Taken together, amplified methylation patterns as observed in EZH2-activated patients are part of the transcriptional response of affected target genes and regulate gene clusters critical for oncogenic signaling.

Discussion

This analysis on the large SKCM dataset from TCGA appears to have revealed several hitherto unappreciated aspects of EZH2 biology in melanoma. Although the analysis verified that activating mutations of EZH2 are relatively uncommon [15] when added to copy number

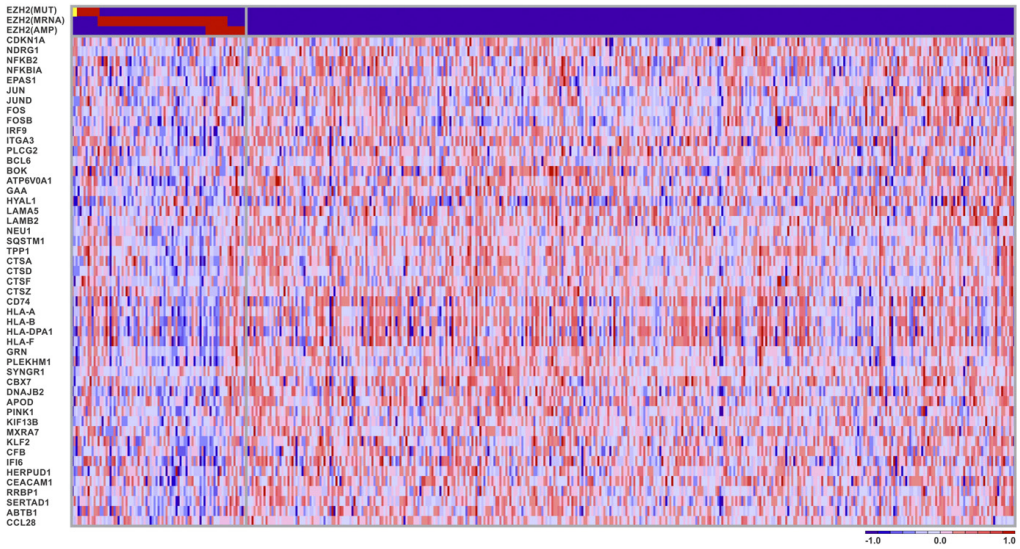


Figure 6. Activation of EZH2 and pattern of transcriptional silencing in melanoma. 88 of 470 cases in the skin cutaneous melanoma dataset show somatic mutations, somatic copy number amplification, or upregulation of transcription of EZH2. Top 50 genes are shown for patients with EZH2 alterations vs patients with EZH2 wild-type status. Transcripts show significant deregulation with p-values below 1.0e-04. RNASeq transcription levels are shown as log2 normalized heat map from low values in blue to high values in red according to color scheme legend. Somatic EZH2 status is indicated in red in the first three rows. Activating somatic mutations of Y641 in EZH2 are highlighted in yellow in the first row.

gains and amplification of gene expression the number of patients with dysregulation of EZH2 approaches 20% of patients (88 of 471 patients). Pathological hyperactivity of EZH2 can arise from somatic amplification persisting at the transcriptional level or from somatic mutations but both events eventually result in a disruption of the epigenetic homeostasis. By looking into isoform-specific RNASeq data of more than 12,000 patients we were able to map the major cancer transcript [GeneBank:NM_001203247] of EZH2, which is an essential step to unambiguously identify non-synonymous somatic mutations. Irrespectively of lengths of N-terminal domains, somatic mutations across different cancers hit a unique phosphorylation site in the center of the SET domain resulting in hyperactivation and increased tri-methylation activity. Somatic Y641 mutations correlate with hypermethylation and lead to prominent gene suppression in the TCGA dataset. Patients with somatic copy number amplifications of EZH2 showed a similar functional impact: gene suppression was most pronounced in cases where SCNA events translated to significant upregulation of EZH2 transcripts. Importantly, the SCNA event amplifying EZH2 impacts a broad genomic environment including CREB3L2 [GeneBank:64764], KIAA1549 [GeneBank:57670], KDM7A [GeneBank:80853], BRAF [GeneBank:673], CDK5 [GeneBank:1020], and KMT2C [GeneBank:58508]. This suggests that a broader genomic context of chromosome 7 is functionally important for the tumor instead of a focal event amplifying an isolated proto-oncogene in melanoma. Our analysis identified a coincidence of BRAF activation with EZH2 amplification providing a mechanistic link for previous studies, which showed an association between BRAF mutations and DNA hypermethylation [21].

Abnormal DNA methylation of CpG markers is a well-known epigenetic feature of cancer [22]. Melanoma exhibits global hypomethylation within the bulk genome and local hypermethylation at specific tumor suppressor genes [23,24]. Gene-specific DNA hypermethylation might serve as classifier for melanoma, as studies indicate that multilocus DNA methylation signature genes may differentiate melanomas from nevi [25]. The characterized and confirmed candidate target genes susceptible to reversible EZH2 histone modification upon drug treatment agreed with the transcriptional signature of TCGA SKCM patients with EZH2 hyperactivity. The majority of the 98 candidate target genes also showed high levels of DNA methylation. This is consistent with suggested concerted action of histone modifiers and DNA methyltransferases at the epigenomic level resulting in transcriptional repression [2]. Methylation data in TCGA showed that DNA methylation appeared to be the common cause of gene suppression associated with EZH2. This may not be that surprising given that DNA methyltransferases DNMT1, DNMT3A, and DNMT3B are closely associated with polycomb repressive complex 2 and are required for epigenetic modifications [1,3]. The mechanism of suppression of genes not associated with DNA methylation may represent transcriptional repression resulting from EZH2 mediated histone methylation but was not further explored in this analysis.

Recent studies in murine melanoma models and human melanoma cultures have supported a role for EZH2 in the proliferation and metastasis of melanoma and have linked these properties to expression of possible suppressor genes identified in TCGA [10–14]. Gene expression studies carried out on melanoma lines with known

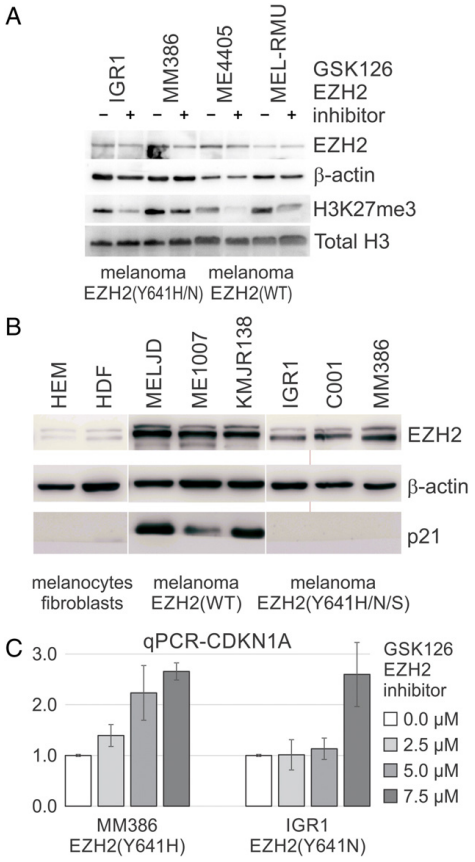


Figure 7. Histone methylation of H3K27 is reversible and responsive to EZH2 inhibition in melanoma cells. (A) Western blot analysis of EZH2 and H3K27me3 levels in melanoma cell lines with WT EZH2 status (right) and melanoma cell lines with activating Y641 mutations (left) in absence or presence of 48 hour treatment with 7.5 μM GSK126, a small molecule inhibitor of EZH2 activity. Hypermethylation of H3K27 of melanoma cells is reversible and responsive to EZH2 inhibitor treatment. (B) Activating mutations of EZH2 are associated with silenced CDKN1A expression in melanoma cell lines. Western blot analysis of EZH2 and CDKN1A expression in normal melanocytes and fibroblasts (left), melanoma cell lines with WT EZH2 status (middle), melanoma cell lines with activating Y641 mutations (right). Blots are from separate membranes developed at the same time. (C) Expression of CDKN1A following 48 h of treatment with increasing doses of EZH2 inhibitor GSK126. RT-qPCR values are relative to vehicle treated control (DMSO) normalized to 1.

activating mutations of EZH2 before and after treatment with the GSK126 inhibitor were a valuable resource to identify genes linked to EZH2 activation in the TCGA data set and allowed for detailed study of their epigenetic DNA makeup in SKCM patients. The test group of silenced candidate target genes recapitulated EZH2 status in unsupervised hierarchical clustering of their DNA methylation pattern and revealed accumulation of methylation sites at the transcription start site of candidate target genes.

The tissue-specific effector genes of the histone modifier EZH2 vary greatly. *In vitro* studies of cancer cell lines recapitulating activating Y641 mutations show a more than three-fold increase of tri-methylation of H3K27me3, a histone mark connected with gene silencing [11–14]. Activated EZH2 is associated with decreased transcription of tumor suppressor genes as well as antigen presentation in melanoma patients. Gene enrichment analysis of the genes combined from the GSK126 inhibitor studies with that of genes repressed in the TCGA dataset identified genes with well-known tumor suppressor activity, regulation of cell division, invasion and metastases. JUN, JUND, FOS, and FOSB implicated in our study are part of the AP-1 transcription factor complex that induces differentiation. EZH2 knockout studies in mice showed that in early development, EZH2 in the skin prevents premature AP-1 binding and maintains the epidermal progenitors until the precise moment differentiation is appropriate, sparked by a decline in EZH2 [26,27]. It has been suggested that aberrant EZH2 activity in cancer maintains the cells in a stem like state, therefore EZH2 inhibitors may represent a strategy to induce differentiation and repress tumor growth. Transcription of putative targets CDKN1A, BCL6, BOK, NDRG1, and NFKB is significantly reduced in SKCM patients with EZH2 activation. Although involved in different signaling axes—often with feed-back loops and dual impact in tumor promotion and suppression—these important regulators have in common that they can promote proliferative events, if suppressed [28].

We were able to confirm transcriptional reactivation of the tumor suppressor CDKN1A in EZH2 Y641 mutated melanoma cells, using an EZH2 inhibitor. CDKN1A encodes the cyclin dependent kinase inhibitor p21, which activates multiple tumor suppressor pathways including cell cycle arrest, differentiation and cellular senescence. These findings are consistent with a previous report that demonstrated EZH2 depletion in melanoma lead to reactivation of p21 and inhibited the growth of xenografts in mice [17,19]. Patients with activating EZH2 mutations, show hypermethylation and transcriptional silencing of the tumor suppressor CDKN1A resulting in absence of p21, the gene product of CDKN1A. Inhibition of EZH2 by small molecule drugs reactivates CDKN1A and illustrates how epigenetic control by EZH2 can regulate the plasticity of melanoma.

An unexpected finding was the repression of many genes associated with immune responses in patients with EZH2 dysregulation. However, this is consistent with increasing appreciation that activated EZH2 is associated with decreased transcription of genes involved in antigen presentation in melanoma patients. Recent data suggests that specific oncogenic signals can mediate cancer immune evasion and resistance to immunotherapies [29] and specifically EZH2 can dampen the anti-tumor immune response via repression of MHC-II genes [30,31]. In support of this we identified downregulation of CD74 [GeneBank:972], an important chaperone that associates with MHC-II to regulate antigen presentation for immune responses, in addition to chemokines/chemokine receptors, CCL28 [GeneBank:56477], CCL3L1 [GeneBank:6349] and CCR7 [GeneBank:1236], known to activate T cells and B cells [29]. Modulation of immune potentiation via epigenetic signals may point toward new candidate targets for melanoma treatment. A group of human leukocyte antigen (HLA) genes stood out as enriched cluster in transcriptomic and epigenomic analysis. HLA genes controlled by epigenetic pattern may contribute to observed regulation and diversity of neoantigens in melanoma [32].

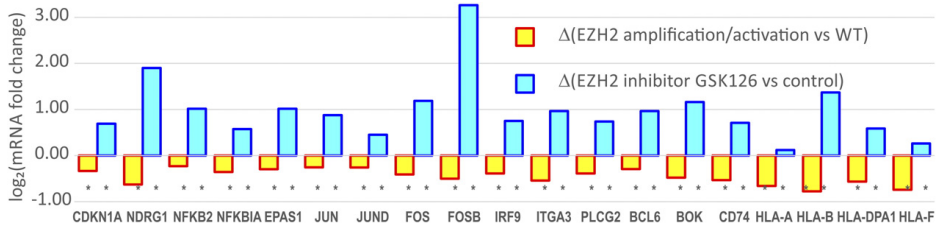


Figure 8. Inhibition of EZH2 with GSK126 reverses transcriptional silencing caused by somatic amplification or activating mutations of EZH2 in melanoma. Transcriptional suppression is observed in RNASeq data of SKCM patients with EZH2 amplification or activation (red frame, negative log2 fold changes). Presented genes showed increased levels of DNA methylation. Transcriptional suppression is reversed in microarray data of melanoma cell lines with activating Y641 mutations following treatment with EZH2 inhibitor (blue frame, positive log2 fold changes). Asterisk symbol below transcriptional data indicates significant deregulation of log2 fold changes with adjusted p-values below 0.05 threshold.

Conclusion

In summary these studies indicate that dysregulation of EZH2 is a relatively common occurrence in patients with melanoma and that this has negative implications for patient survival. Somatic events across different omics levels were associated with repression of genes involved in suppression of tumors as well as immune responses against cancers. The isoform-specific RNASeq analysis unambiguously identifies the major transcript of EZH2 and the copy number analysis shows functional amplification events of EZH2 across melanoma patients. The data also suggests that the previously described association of BRAF V600 mutations with methylation of DNA may be linked to overexpression of closely associated regions on chromosome 7. Repression of the EZH2 target genes appears to be largely due to both tri-methylation of H3K27 as well as methylation of DNA and this provides a basis for investigating combinations of both EZH2 and DNA methyltransferase inhibitors in patients with over expressed EZH2. Taken together, the PRC2 displays enhanced activity in melanoma. Hyperactivation of EZH2 by somatic copy number amplification, activating somatic mutations, or transcriptional upregulation correlates with DNA methylation and epigenetic silencing of genes involved in tumor suppression and immune responses in melanoma. Further studies are needed to determine whether inhibitors of EZH2 may also have a role immunotherapy with checkpoint inhibitors. Sophisticated chromatin immunoprecipitation experiments will closely decipher interaction of EZH2 with its target genes, chromatin modifiers and DNA methyltransferases. Taken together, EZH2 and its associated chromatin remodeling machinery represent a promising opportunity for therapeutic intervention in melanoma.

Competing Interest

The authors declare that there is no competing interest as part of the submission process of this manuscript.

Authors' Contributions

J.C.T., S.W. contributed equally to the work. J.C.T., S.W., S.J.G., P.H., F.V.F. designed the study, wrote the text, conducted experiments, performed the data analysis, and reviewed the final manuscript.

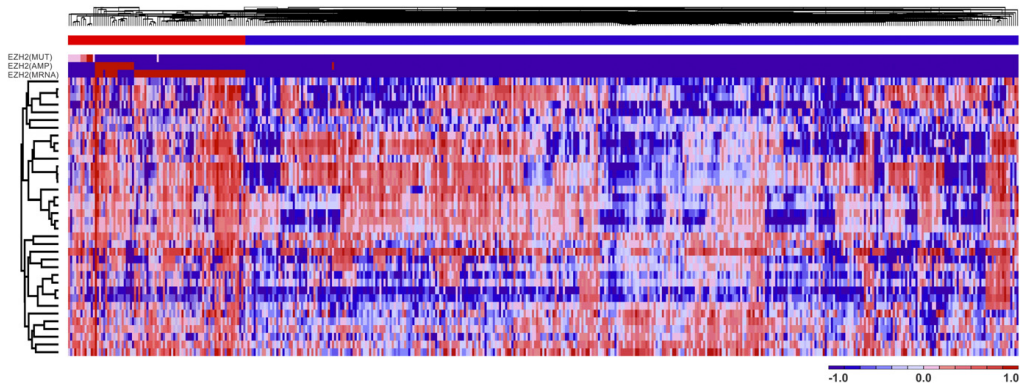


Figure 9. CG methylation markers recapitulate EZH2 activation patterns. Unsupervised hierarchical clustering of methylation markers of 461 patients aligns with EZH2 status. Presence of EZH2 activation is indicated as red bars on top of plot. EZH2 mutations, amplifications or mRNA upregulation are labelled in the first three rows.

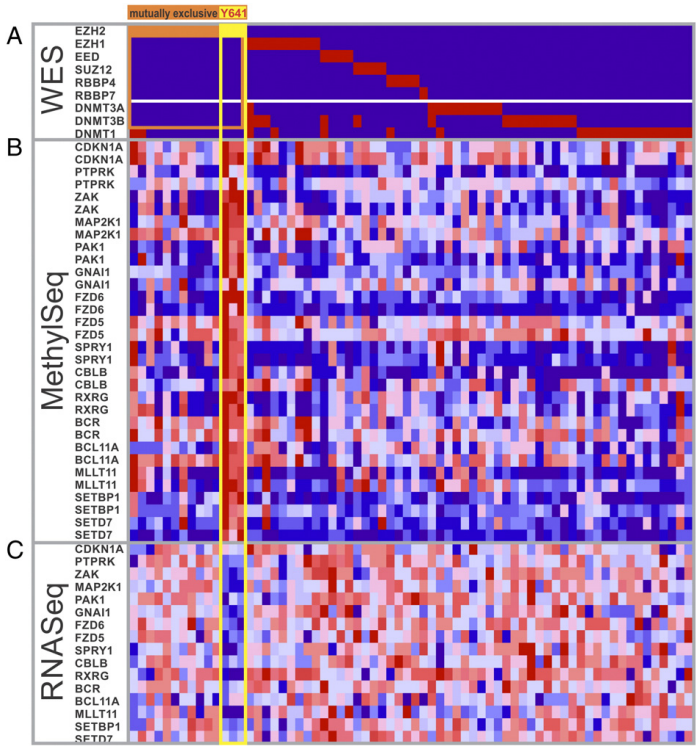


Figure 10. Somatic mutation of Y641 in EZH2 is associated with increased methylation of CpG markers and decreased transcription of tumor suppressors and epigenetic remodelers in melanoma patients. (A) Somatic mutations of EZH2 occur in a mutually exclusive setting in members of the polycomb repressive complex 2, or in *de novo* DNA methyltransferases. The activating somatic mutations of Y641 in EZH2 are marked in yellow in the first row. Transcriptomic and epigenomic data from patients with Y641F or Y641N (2x) are boxed in yellow. Other mutations of EZH2 are marked in orange in the first row. Mutually exclusive setting of EZH2 with EZH1, EED, SUZ12, RBBP4, RBBP7, or *de novo* DNA methyltransferases DNMT3A or DNMT3B is boxed in orange. (B) Representative CpG markers of candidate target genes show enhanced methylation in patients with mutated Y641 status. (C) RNASeq transcription levels of same candidate target genes are reduced.

Acknowledgements

We are thankful to all members of the TCGA Research Network for biospecimen collection, and data acquisition. F.V.F. is grateful for the support of grants CA154887 and CA176114 from the National Institutes of Health, National Cancer Institute. This work is supported by NHMRC program grant 633004, the Melanoma Institute of Australia, Cancer Institute NSW.

Appendix A. Supplementary data

The supplementary information contains tables on somatic mutations and whole-exome sequencing patient cohort (Supplementary Table 1), somatic copy number aberrations (Supplementary Table 2), transcriptomic data of identified target genes (Supplementary Table 3), methylome data of identified target genes (Supplementary Table 4), and pathway enrichment scores (Supplementary Table 5). Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.neo.2016.01.003.

References

[1] Mills AA (2010). Throwing the cancer switch: reciprocal roles of polycomb and trithorax proteins. *Nat Rev Cancer* **10**, 669–682.

[2] Baylin SB and Jones PA (2011). A decade of exploring the cancer epigenome - biological and translational implications. *Nat Rev Cancer* **11**, 726–734.

[3] Vire E, Brenner C, Deplus R, Blanchon L, Fraga M, Didelot C, Morey L, Van Eynde A, Bernard D, and Vanderwinden JM, et al (2006). The Polycomb group protein EZH2 directly controls DNA methylation. *Nature* **439**, 871–874.

[4] Ernst T, Chase AJ, Score J, Hidalgo-Curtis CE, Bryant C, Jones AV, Waghorn K, Zoi K, Ross FM, and Reiter A, et al (2010). Inactivating mutations of the histone methyltransferase gene EZH2 in myeloid disorders. *Nat Genet* **42**, 722–726.

[5] Ntziachristos P, Tsirigos A, Van Vlierberghe P, Nedjic J, Trimarchi T, Flaherty MS, Ferres-Marco D, da Ros V, Tang Z, and Siegle J, et al (2012). Genetic inactivation of the polycomb repressive complex 2 in T cell acute lymphoblastic leukemia. *Nat Med* **18**, 298–301.

[6] Morin RD, Johnson NA, Severson TM, Mungall AJ, An J, Goya R, Paul JE, Boyle M, Woolcock BW, and Kuchenbauer F, et al (2010). Somatic mutations altering EZH2 (Tyr641) in follicular and diffuse large B-cell lymphomas of germinal-center origin. *Nat Genet* **42**, 181–185.

[7] De Carvalho DD, Binato R, Pereira WO, Leroy JM, Colassanti MD, Proto-Siqueira R, Bueno-Da-Silva AE, Zago MA, Zanichelli MA, and Abdelhay E, et al (2011). BCR-ABL-mediated upregulation of PRAME is responsible for knocking down TRAIL in CML patients. *Oncogene* **30**, 223–233.

[8] Sneeringer CJ, Scott MP, Kuntz KW, Knutson SK, Pollock RM, Richon VM, and Copeland RA (2010). Coordinated activities of wild-type plus mutant EZH2 drive tumor-associated hypertrimethylation of lysine 27 on histone H3 (H3K27) in human B-cell lymphomas. *Proc Natl Acad Sci U S A* **107**, 20980–20985.

[9] McCabe MT, Graves AP, Ganji G, Diaz E, Halsey WS, Jiang Y, Smitheman KN, Ott HM, Pappalardi MB, and Allen KE, et al (2012). Mutation of A677 in histone methyltransferase EZH2 in human B-cell lymphoma promotes hypertrimethylation of histone H3 on lysine 27 (H3K27). *Proc Natl Acad Sci U S A* **109**, 2989–2994.

[10] Guan J, Gupta R, and Filipp FV (2015). Cancer systems biology of TCGA SKCM: Efficient detection of genomic drivers in melanoma. *Sci Rep* **5**, 7857–7866.

[11] Zingg D, Debbache J, Schaefer SM, Tuncer E, Frommel SC, Cheng P, Arenas-Ramirez N, Haeusel J, Zhang Y, and Bonalli M, et al (2015). The epigenetic modifier EZH2 controls melanoma growth and metastasis through silencing of distinct tumour suppressors. *Nat Commun* **6**, 6051–6067.

[12] Barsotti AM, Ryskin M, Zhong W, Zhang WG, Giannakou A, Loreth C, Diesl V, Follettie M, Golas J, and Lee M, et al (2015). Epigenetic reprogramming by tumor-derived EZH2 gain-of-function mutations promotes aggressive 3D cell morphologies and enhances melanoma tumor growth. *Oncotarget* **6**, 2928–2938.

[13] Tiffen J, Gallagher SJ, and Hersey P (2015). EZH2: an emerging role in melanoma biology and strategies for targeted therapy. *Pigment Cell Melanoma Res* **28**(1), 21–30.

[14] Tiffen JC, Gunatilake D, Gallagher SJ, Gowrishankar K, Heinemann A, Cullinane C, Dutton-Regester K, Pupo GM, Strbenac D, and Yang JY, et al (2015). Targeting activating mutations of EZH2 leads to potent cell growth inhibition in human melanoma by derepression of tumor suppressor genes. *Oncotarget* **6**(29), 27023–27036.

[15] Hodis E, Watson IR, Kryukov GV, Arold ST, Imielinski M, Theurillat JP, Nickerson E, Auclair D, Li L, and Place C, et al (2012). A landscape of driver mutations in melanoma. *Cell* **150**, 251–263.

[16] Antonysamy S, Condon B, Druzina Z, Bonanno JB, Gheyi T, Zhang F, MacEwan I, Zhang A, Ashok S, and Rodgers L, et al (2013). Structural context of disease-associated mutations and putative mechanism of autoinhibition revealed by X-ray crystallographic analysis of the EZH2-SET domain. *PLoS One* **8**(12), e84147.

[17] Sahasrabuddhe AA, Chen X, Chung F, Velusamy T, Lim MS, and Elenitoba-Johnson KS (2015). Oncogenic Y641 mutations in EZH2 prevent Jak2/beta-TrCP-mediated degradation. *Oncogene* **34**, 445–454.

[18] Hobert O, Jallal B, and Ullrich A (1996). Interaction of Vav with ENX-1, a putative transcriptional regulator of homeobox gene expression. *Mol Cell Biol* **16**, 3066–3073.

[19] Fan T, Jiang S, Chung N, Alikhan A, Ni C, Lee CC, and Hornyak TJ (2011). EZH2-dependent suppression of a cellular senescence phenotype in melanoma cells by inhibition of p21/CDKN1A expression. *Mol Cancer Res* **9**, 418–429.

[20] McCabe MT, Ott HM, Ganji G, Korenchuk S, Thompson C, Van Aller GS, Liu Y, Graves AP, Della Pietra III A, and Diaz E, et al (2012). EZH2 inhibition as a therapeutic strategy for lymphoma with EZH2-activating mutations. *Nature* **492**, 108–112.

[21] Hou P, Liu D, Dong J, and Xing M (2012). The BRAF(V600E) causes widespread alterations in gene methylation in the genome of melanoma cells. *Cell Cycle* **11**, 286–295.

[22] Rodriguez-Paredes M and Esteller M (2011). Cancer epigenetics reaches mainstream oncology. *Nat Med* **17**, 330–339.

[23] Hoon DS, Spugnardi M, Kuo C, Huang SK, Morton DL, and Taback B (2004). Profiling epigenetic inactivation of tumor suppressor genes in tumors and plasma from cutaneous melanoma patients. *Oncogene* **23**, 4014–4022.

[24] Liu S, Ren S, Howell P, Fodstad O, and Riker AI (2008). Identification of novel epigenetically modified genes in human melanoma via promoter methylation gene profiling. *Pigment Cell Melanoma Res* **21**, 545–558.

[25] Conway K, Edmiston SN, Khondker ZS, Groben PA, Zhou X, Chu H, Kuan PF, Hao H, Carson C, and Berwick M, et al (2011). DNA-methylation profiling distinguishes malignant melanomas from benign nevi. *Pigment Cell Melanoma Res* **24**, 352–360.

[26] Eferl R and Wagner EF (2003). AP-1: a double-edged sword in tumorigenesis. *Nat Rev Cancer* **3**, 859–868.

[27] Ezhkova E, Pasolli HA, Parker JS, Stokes N, Su IH, Hannon G, Tarakhovsky A, and Fuchs E (2009). Ezh2 orchestrates gene expression for the stepwise differentiation of tissue-specific stem cells. *Cell* **136**, 1122–1135.

[28] Lee ST, Li Z, Wu Z, Aau M, Guan P, Karuturi RK, Liou YC, and Yu Q (2011). Context-specific regulation of NF-kappaB target gene expression by EZH2 in breast cancers. *Mol Cell* **43**, 798–810.

[29] Spranger S, Bao R, and Gajewski TF (2015). Melanoma-intrinsic beta-catenin signalling prevents anti-tumour immunity. *Nature* **523**(7559), 231–235.

[30] Abou El Hassan M, Huang K, Eswara MB, Zhao M, Song L, Yu T, Liu Y, Liu JC, McCurdy S, and Ma A, et al (2015). Cancer Cells Hijack PRC2 to Modify Multiple Cytokine Pathways. *PLoS One* **10**, e0126466.

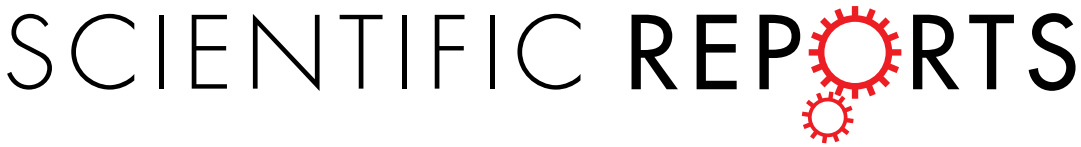
[31] Holling TM, Bergevoet MW, Wilson L, Van Eggermond MC, Schooten E, Steenbergen RD, Snijders PJ, Jager MJ, and Van den Elsen PJ (2007). A role for EZH2 in silencing of IFN-gamma inducible MHC2TA transcription in uveal melanoma. *J Immunol* **179**, 5317–5325.

[32] Carreno BM, Magrini V, Becker-Hapak M, Kaabinejadian S, Hundal J, Petti AA, Ly A, Lie WR, Hildebrand WH, and Mardis ER, et al (2015). Cancer immunotherapy. A dendritic cell vaccine increases the breadth and diversity of melanoma neoantigen-specific T cells. *Science* **348**, 803–808.

Chapter Four: Refinement of the androgen response element based on ChIP-Seq in androgen-insensitive and androgen responsive prostate cancer cell lines.

Transcription factor binding sites are sequence specific features present in promoters or enhancers of genes targets. Typically when performing a Chromatin immunoprecipitation sequencing (ChIPSeq) experiment on a transcription factor, a motif-based analysis is applied to identify its sequence specific binding sites to help verify its authenticity. This study applies model base search of the androgen receptor (AR) response element in androgen-insensitive and androgen-responsive prostate cancer cell lines. The 120,000 identified binding sites were classified based on their degeneracy, transcriptional involvement, oncogenic AR splice-variants, somatic copy number amplifications, and steroid treatment. This study refines the AR motif, and describes a workflow for optimizing response element identification in genomic sequences. Within this paper I provided the computational analysis of the data found in figures 1-3,5-7.

www.nature.com/scientificreports



OPEN

Received: 17 March 2016
Accepted: 03 August 2016
Published: 14 September 2016

Refinement of the androgen response element based on ChIP-Seq in androgen-insensitive and androgen-responsive prostate cancer cell lines

Stephen Wilson[1], Jianfei Qi[2] & Fabian V. Filipp[1]

Sequence motifs are short, recurring patterns in DNA that can mediate sequence-specific binding for proteins such as transcription factors or DNA modifying enzymes. The androgen response element (ARE) is a palindromic, dihexameric motif present in promoters or enhancers of genes targeted by the androgen receptor (AR). Using chromatin immunoprecipitation sequencing (ChIP-Seq) we refined AR-binding and AREs at a genome-scale in androgen-insensitive and androgen-responsive prostate cancer cell lines. Model-based searches identified more than 120,000 ChIP-Seq motifs allowing for expansion and refinement of the ARE. We classified AREs according to their degeneracy and their transcriptional involvement. Additionally, we quantified ARE utilization in response to somatic copy number amplifications, AR splice-variants, and steroid treatment. Although imperfect AREs make up 99.9% of the motifs, the degree of degeneracy correlates negatively with validated transcriptional outcome. Weaker AREs, particularly ARE half sites, benefit from neighboring motifs or cooperating transcription factors in regulating gene expression. Taken together, ARE full sites generate a reliable transcriptional outcome in AR positive cells, despite their low genome-wide abundance. In contrast, the transcriptional influence of ARE half sites can be modulated by cooperating factors.

Prostate cancer is the most common malignancy in American men[1,2]. Although localized prostate cancer is curable by surgery, the primary treatment for metastatic prostate cancer remains androgen deprivation therapy (ADT). However, the success of ADT is not guaranteed. Despite an initial response, advanced prostate cancer develops almost invariably resistance to ADT and progresses to a lethal disease stage called castration-resistant prostate cancer (CRPC). Thus, understanding the mechanisms underlying the development of CRPC is of critical importance for basic and clinical research.

Reactivation of the androgen receptor (AR, GeneBank: 367) under low androgen condition is believed to drive the development of CRPC[3,4]. AR belongs to the nuclear receptor superfamily and plays an important role in the physiology of normal prostate gland and progression of prostate adenocarcinoma (PRAD). The AR gene is located on the X chromosome at Xq11-12 and contains eight exons encoding a 919 amino acid-long protein. The AR protein consists of an N-terminal transactivation domain (NTD), a central DNA binding domain (DBD), a hinge region, and a C-terminal ligand-binding domain (LBD)[5]. Unliganded AR is sequestered in the cytoplasm by a chaperone complex. Upon ligand binding to the LBD, AR changes its conformation, dissociates from the chaperone complex, dimerizes, and translocates into the nucleus. Once translocated into the nucleus the AR dimer binds to the androgen response elements (AREs) present in promoter or enhancer elements of its target genes, and recruits co-activators or co-repressors to regulate gene expression.

[1]Systems Biology and Cancer Metabolism, Program for Quantitative Systems Biology, University of California Merced, 2500 North Lake Road, Merced, CA 95343, USA. [2]Marlene and Stewart Greenebaum Cancer Center, Department of Biochemistry and Molecular Biology, University of Maryland School of Medicine, 655 West Baltimore Street, Baltimore MD 21201, USA. Correspondence and requests for materials should be addressed to F.V.F. (email: filipp@ucmerced.edu)

Reactivation of the AR in CRPC can be due to changes of the AR or its steroid ligands[6–12]. Known alterations of the AR include somatic gene amplification and/or over-expression that increase AR mediated response to low androgen levels, AR mutations that change ligand specificity to allow for activation by other steroids, and generation of AR splice variants (AR-Vs) that lack the LBD and are constitutively active even in the absence of androgen. CRPC cells can also synthesize androgens themselves by conversion of testosterone derivatives or *de novo* by biosynthesis from cholesterol. Such intra-tumoral androgen synthesis permits maintenance of certain intracellular androgen levels and results in reconquered mitogenic AR activity[13–15].

Studies of AR chromatin binding by chromatin immunoprecipitation (ChIP) approaches like ChIP-on-chip or ChIP in combination with next generation sequencing (ChIP-Seq) have shed light onto the mechanisms of global regulation of AR activities in prostate cancer cell lines or tissues. Most AR chromatin binding studies were performed in the LNCaP cell line or its sublines[3,4,6,7,9–12], which are highly sensitive and responsive to androgen stimulation. LNCaP or its sublines express a full-length AR with point mutation of T877A at the LBD[16]. In contrast, there are limited studies on global AR binding in the CWR22Rv1 cell line[4,8], another AR-positive but androgen-insensitive PRAD cell line. The CWR22Rv1 prostate cancer cell line expresses AR full-length (AR(FL)) with a duplicated DBD in exon 3[17–19] and an AR splice variant, AR(V), lacking a LBD, thus becoming constitutively active[4,8,11,20]. In contrast to the LNCaP cell line where the AR depends on androgen activation, the CWR22Rv1 cell line shows constitutively active AR with limited changes in expression of AR target genes in the presence or absence of androgens[19,21].

AREs are well studied but poorly defined and have been shown to contain two hexamers with a three base-pair spacer with an inverted repeat in the second hexamer[22]. The sequence elements similar to this canonical ARE have been identified in some ChIP-Seq studies, whereas half AREs or tandem repeats of two hexamers were also found in other ChIP-Seq or ChIP-on-chip studies. In the past, studies revealed binding motifs adjacent to the AR binding sites but belonging to other transcription factor families such as the forkhead box A1 protein (FOXA1, GeneBank: 3169). Cooperative interactions facilitate chromatin binding of the AR and contribute to a promiscuous behavior of AREs[23–25]. AREs and adjacent transcription binding motifs have been well described in LNCaP cells but remain to be defined in CWR22Rv1 cells. Therefore, the purpose of our AR ChIP-Seq study is to further characterize the ARE and identify cooperation with adjacent transcription binding motifs in androgen-responsive and androgen-insensitive prostate cancer cell lines.

Methods

Cell culture. CWR22Rv1 is a human prostate carcinoma epithelial cell line derived from a xenograft that was serially propagated in mice after castration-induced regression and relapse of the parental, androgen-dependent CWR22 xenograft[26,27] (CRL-2505, American Type Culture Collection, Manassas, VA). The CWR22Rv1 prostate cancer cell line was kindly provided by Dr. James Jacobberger (Case Western Reserve University, Cleveland, OH), and are maintained in RPMI 1640 medium supplemented with 10% FBS and antibiotics. Cells are regularly tested to ensure that they are mycoplasma-free. All experimental protocols were approved by the Institutional Review Board at the University of California Merced. The study was carried out as part of IRB UCM13-0025 of the University of California Merced and as part of dbGap ID 5094 on somatic mutations in cancer and conducted in accordance with the Helsinki Declaration of 1975.

Knockdown of AR with shRNA. Lentiviral vectors encoding AR shRNA were purchased from Open Biosystems, and packaged in 293T cells by the calcium phosphate transfection. The supernatant containing lentiviral particles were collected 48 hours after transfection. CWR22Rv1 cells were transduced with the supernatant of lentiviral particles in the presence of polybrene (8 μg/ml) for 24 hours before replacement with the fresh growth media. Cells were analyzed at 72 hours post-transduction. The knockdown efficiency was confirmed by quantitative real time polymerase chain reaction (qRT-PCR) and Western-blot analysis (Supplementary Figure 1).

qRT-PCR analysis. Total RNA from prostate cancer cells was extracted using a mammalian RNA mini preparation kit (Sigma, GenElute, RTN10, Darmstadt, Germany) and then digested with deoxyribonuclease I (Sigma, AMPD1, Darmstadt, Germany). Complementary DNA (cDNA) was synthesized using random hexamers. Triple replicate samples were subjected to SYBR green (SYBR green master mix, Qiagen SABiosciences) qRT-PCR analysis in an Eco system (Illumina, San Diego). Gene expression profiles were analyzed using the ΔΔCT method. RT QPCR threshold cycle (CT) values were normalize the housekeeping gene cyclophilin A (PPIA, peptidylprolyl isomerase A, GeneBank: 5478). The following primers served for qRT-PCR analysis of human gene transcripts: PPIA: 5′-GACCCAACACAAATGGTTC-3′; 5′-AGTCAGCAATGGTGATCTTC-3′; AR: 5′-CTCCGCTGACCTTAAAGACATC-3′; 5′-TGCCCCCTAAGTAATTGTCCTT-3′.

Western-blot analysis. Whole cell lysates were harvested using radio-immunoprecipitation assay (RIPA) buffer composed of 50 mM trisaminomethane hydrochloride (Tris-HCl) pH 7.5, 150 mM sodium chloride (NaCl), 1% Triton X-100, 0.1% sodium dodecyl sulfate (SDS), 0.1% sodium deoxycholate, 1.0 mM EDTA, 1.0 mM sodium orthovanadate, and 1x protease inhibitor cocktail. Lysates were subjected to sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and proteins transferred to a nitrocellulose membrane (GE Healthcare Life Sciences). The membrane was probed with an AR antibody (Sigma EMD Millipore, PG21, 06-680, Darmstadt, Germany) or actin antibody (Sigma EMD Millipore, A2066, Darmstadt, Germany) followed by a secondary antibody conjugated to fluorescent dye, and blots were imaged using the odyssey detecting system (LI-COR Biotechnology).

Chromatin immunoprecipitation. Cells were crosslinked using 1% formaldehyde for 10 min at 298 K. Formaldehyde was diluted to a final concentration of 125 mM by adding 5 M glycine. Nuclear extracts were collected and sonicated to obtain 300 bp chromatin fragments using the Covaris S2 ultrasonicator (Covaris, Woburn, MA).

100 μg of chromatin was incubated with 5.0 μg of AR antibodies (Sigma EMD Millipore, PG21, 06–680, Darmstadt, Germany) overnight at 277 K followed by incubation with 30 μl of protein A/G beads for 4 hours. After four washes, crosslinking was reversed, and chromatin was digested with ribonuclease A (RNaseA) followed by proteinase K. The DNA was purified using spin columns. The size of the DNA was confirmed by a bioanalyzer (Agilent Biotechnologies, Savage, MD).

Next generation sequencing and ChIP-seq data analysis. The purified DNA library was sequenced using an Illumina HighSeq2000 at the Sanford-Burnham Medical Research Institute, National Genome Library Core Facility (Lake Nona, FL). This study included next generation sequencing reads of ChIP-Seq experiments of the androgen-independent CWR22Rv1 cell line as well as of the androgen-dependent LNCaP cell line. For the CWR22Rv1 cell line[27] we acquired total AR binding by ChIP-Seq. In addition, we compared the data to AR splice variant-specific isoforms AR(FL) and AR(V) as described[28]. For the LNCaP cell line[29] we assessed conditions of testosterone and ethanol treatment as described[30]. Sequenced regions were aligned to the reference human genome 19 using the Bowtie alignment program that utilizes an extended Burrows-Wheeler indexing for an ultrafast memory efficient alignment[31]. Peak calling utilized a model-based analysis of ChIP-Seq (MACS) algorithm[32,33]. The overlap analysis, plot of genomic location, sequence extraction, motif identification, and peak filtering were performed using ChIPseek: a web-based analysis for ChIP data[34]. ChIPseek also employs scripts from BEDtools[35] using a genome binning algorithm used by the UCSC genome browser to sort genomic regions into groups along the length of chromosome[36]. Data visualization was carried out using the integrative genomics viewer[37]. The tool genomic identification of significant targets in cancer (GISTIC), 2.0.21[38], was used to identify genomic regions that are significantly gained or lost across a set of paired normal and tumors samples of 492 specimen on Agilent SNP 6.0 gene expression microarrays G4502A_07_01. Arm-level amplification of HighSeq2000 data was estimated and compared to near diploid averages[39]. Events whose length was greater or less than 50% of the chromosome arm on which they resided were called arm-level or focal events, respectively. Segmented level 3 tumor copy number data relative to normal samples was used as input for GISTIC 2.0.21 and aligned to HG19. For significant loci and genes a cutoff q-value of 0.05 was applied.

Motif analysis based on position site specific matrix models. Computational response element searching algorithms are able to estimate a sequence's likelihood in belonging to the response element of the query transcription factor using position site specific matrices where each position in the query transcription factor model gives each of the four letters in the DNA alphabet a score based on the probability of that nucleotide being found at that position (Supplementary Table S1)[40]. ChIP-Seq derived ARE motif logos are deposited in transcription factor databases. Experimental transcription factor matrices based on 42956 and 79065 ChIP-Seq AR-binding events for ARE full and half sites, respectively, are referenced under accession M08907 and M08908 in TRANSFAC 2016.3, accession AR in Jaspar, and accession ANDR_HUMAN in HOCOMOCO. Summation into a logs-odd score is converted into a p-value assuming a zero-order background model, and all response elements less than the threshold are reported[41]. Motif discovery, motif enrichment, and motif scanning used the multiple expectation maximization for motif elicitation (MEME) and discriminative regular expression motif elicitation (DREME) suite software toolkits from a set of user supplied unaligned sequences for ChIP-Seq regions[42]. *De novo* motif analysis programs MEME and DREME identifies similar reoccurring DNA sequences, and allows easy submission genomic sequence databases to find similarity to previously studied DNA binding protein motifs[43,44]. After a motif of interest is discovered the genomic sequences of the ChIP sequenced data is scanned using the MEME suite software find individual motif occurrences (FIMO)[41] for individual motif occurrences using a position specific matrix to compute a log-likelihood ratio score for each submitted sequence. The position specific matrix is used further to analyze the sequenced data for motif enrichment for identifying potential co-activators within the data[45]. Transcription factor complexes were inferred from ChIP-Seq data using spaced motif analysis (SPAMO)[46].

Microarray analysis. CWR22v1 cells were transduced with lentiviral pLKO.1 control vector or AR shRNA for 72 h. Total RNA was isolated from cells, and 500 ng was used for synthesis of biotin-labeled cRNA using an RNA amplification kit (Ambion, Thermo Fisher Scientific, Waltham, MA). Biotinylated cRNA was labeled by incubation with streptavidin-Cy3 to generate a probe for hybridization with the GeneChip Human Transcriptome Array 2.0 (Affymetrix Inc, Santa Clara, CA). Four samples from two experimental groups ($n = 2$ per group) were hybridized to the chip to obtain raw gene expression data, which was processed to obtain raw data in the form of expression intensities. Raw data was then exported for further processing and analysis using R statistical software version 2.15 in combination with the BioConductor package[47]. The raw signal intensities were background corrected by using array-specific measures of background intensity based on negative control probes, prior to transformation and normalization using the variance stabilization (VSN)[48]. The dataset was then filtered to remove probes not detected (detection score < 0.95) in any sample. Differential expression between experimental groups was assessed by generating relevant contrasts corresponding to the two-group comparison and was evaluated using the linear models for microarray analysis (LIMMA) package[47]. Raw p-values were corrected for multiple testing using the false discovery rate controlling procedure of Benjamini and Hochberg, and adjusted p-values below 0.05 were considered significant[49]. Significant probe lists were then annotated using the relevant annotation file (HumanHT-12_V4_0_R1_15002873_B).

Results

Identification of ARE motifs based on ChIP-Seq pattern, models, or database knowledge. ChIP-Seq data represents an enrichment of loci related to the binding of the protein of interest selected by the immunoprecipitation. Our ChIP-Seq dataset contained 35073 broad peaks including 4731261 wiggle signals

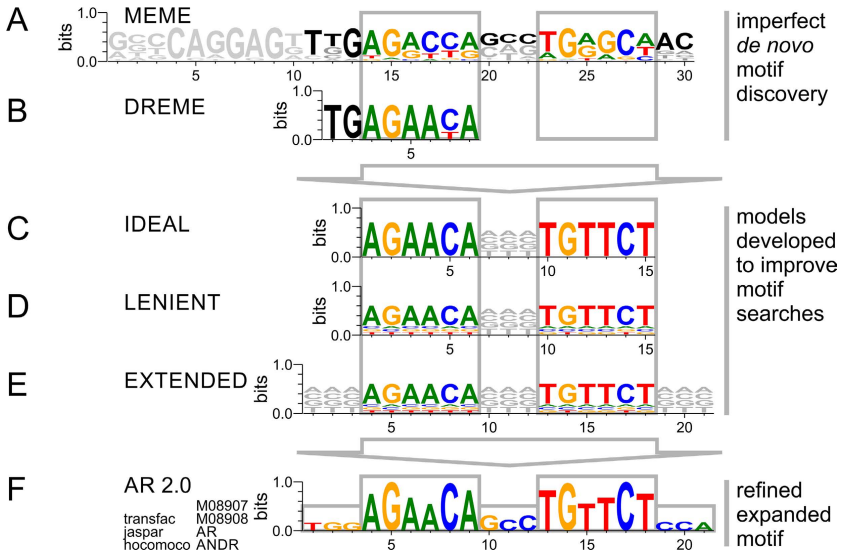


Figure 1. Identification of transcriptional motifs of the androgen receptor based on ChIP-Seq pattern, database knowledge, and position site specific matrix models. (**A**) Motif discovery based on fixed-length patterns or (**B**) short sequence pattern in conjunction with motif databases result in sparse, imperfect motifs. Model-based searches starting from (**C**) ideal model, (**D**) lenient model, or (**E**) extended model provides exhaustive description of motif space in ChIP-Seq experiment. (**F**) Using the identified AREs within our experiment a refined extended canonical ARE model is proposed and deposited in transcription factor databases under accession numbers M08907 (ARE full site), M08908 (ARE half site), AR, and ANDR.

detected by MACS and significantly enriched above the genomic control. 3017 (8.6%) of the detected peaks contained AREs, with many peaks showing multiple motif incidences. Furthermore, we detected DNA motifs resembling ARE sequences in about every 40th wiggle signal. Without putting any knowledge into the search for ARE motifs, we first attempted to conduct a pattern search independent of existing databases. A *de novo* motif discovery search showed an ungapped 30mer logo on the ChIP-Seq data using the MEME tool[50] (Fig. 1A). The 30mer contained not only a palindromic ARE-full site but also information on adjacent bases in the proximity of the ARE. Motif searches relied on existing database entries matching the ChIP-Seq data to a perfect but shortened ARE half site using the DREME tool[44] (Fig. 1B). Next, we attempted to identify ARE locations within our ChIP-Seq data using the FIMO tool[41] with a position site specific matrix as it is defined in motif databases (Supplementary Table S1). This proved problematic as the ARE identified by currently available database models (e.g. Jaspar model MA0007.2)[51] did not resemble the palindromic ARE sequence described in the literature and were shifted in frame[22,52–54]. In order to identify and describe AREs within our genome-wide ChIP-Seq dataset we implemented position site specific matrix (PSSM) models. The initial scan was based on a strict ideal 15mer motif search pattern of two hexamers with a 3mer spacer AGAACANNNTGTTCT (Fig. 1C). However, transcription factor binding is more lenient in its pattern recognition. A lenient 15mer motif search pattern not only dramatically improved detection of ARE occurrences but also allowed for classification of AREs with respect to deviation from a perfect motif (Fig. 1D). The lenient model is distinct from the ideal model such that it allowed for limited mismatch at individual base positions with the core hexamers of full site AREs as frequently observed in binding experiments. In an attempt to elucidate the pattern of the regions surrounding the ARE the lenient motif search pattern was expanded (Fig. 1E). In addition, a half site specific PSSM comprised of an isolated hexamer with flanking sequences captured all ARE motifs, which were lacking a complementary palindromic half site hexamer. Finally we refined and expanded the ARE motif discovered based on our experimental ChIP-Seq data referencing 42956 and 79065 events for ARE full and half site, respectively (Fig. 1F) (Supplementary Table S1).

Identification and base-specific classification of AREs into 5 tiers utilizing ChIP-Seq data. Nucleotide preference and deviation within a response element is critical in determining both the selectivity of transcription factors that bind to that sequence and the necessity for cooperating factors to assist with AR binding. The sequences detected by the ideal, lenient, half site, and extended PSSM model searches were quantified and sorted into tiers reflecting how much the sequence motifs deviated from an ideal ARE sequence (Supplementary Table S2). The ideal 15mer model of AGAACANNNTGTTCT had 71 AREs. The matches fell into tier 1, perfect motifs, with p-values below 8.34E-08. The lenient model identified the same matches in tier 1 and added less defined ARE into additional tiers. There were 71 AREs in tier 1 (perfect), 1583 matches in tier 2 (1bp off perfect),

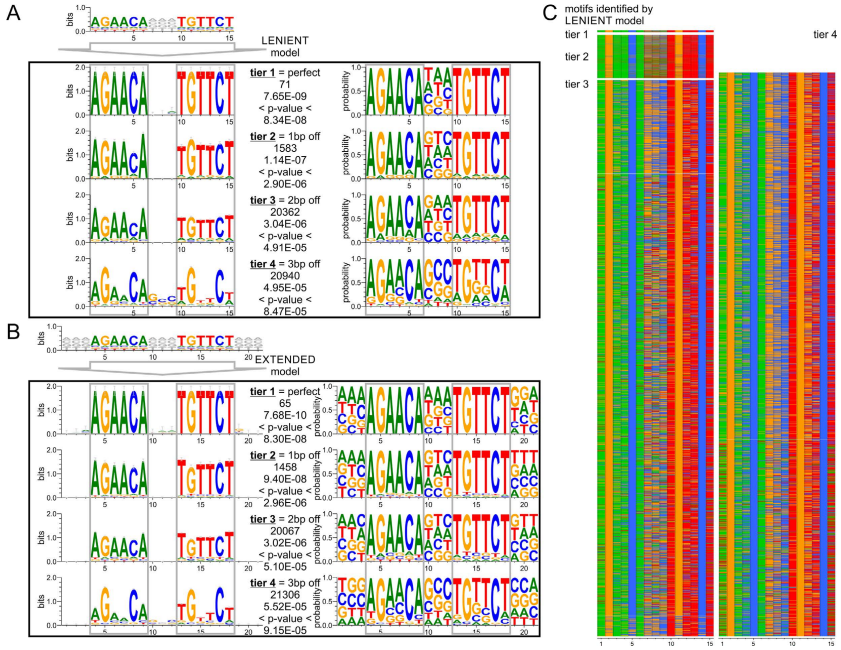


Figure 2. Identification and base-specific classification of AREs utilizing ChIP-Seq data. (**A**) ARE motifs discovered by ChIP-Seq data in combination with a lenient position site specific matrix model reflect the ideal 15mer motif of two hexamers with a 3mer spacer in its top hits. Additional tiers show up to three base pairs deviation off the ideal motif. (**B**) A search with an extended position site specific matrix model allows for insight into the environment of the core sequence motif. ARE logos on the left show the frequencies of the nucleotides scaled to the measure of conservation at each position while ARE logos on the right show the probability of a nucleotide being at that particular region. (**C**) Heatmap of discovered motifs by the lenient model shows conservation of G and C within each hexamer.

20362 in tier 3 (2bp off perfect), and 20940 in tier 4 (3bp off perfect) with p-values below 4.94E-05 (Fig. 2A) (Supplementary Table S1 and Supplementary Table S2). The motifs showed conservation of G and C in positions two and five respectively in the ARE hexamers. In addition, there was increased GC content in the spacer region of the response element. The extended model largely agreed with the lenient model. In tier 4, there were 21306 motifs detected with p-values below 5.52E-05 (Fig. 2B). Regions neighboring the core hexamers as well as the spacer region had increased GC content. The conserved G and C of the hexamer in the least defined AREs of tier 4 were comparable to tier 4 of the lenient model. The overall decreased detection by the extended model in tiers 1–3 in comparison to the lenient model is attributed to the nucleotides of truncated motifs at the border of the ChIP-Seq peak as well as masked repeats in the genome which were blocked out of the search. Heatmaps of all detected motifs by the lenient model highlighted the observed GC nucleotide preferences in positions 2,5,7,8,9,11, and 14 (numbering of 15mer) in contrast to varying nucleotide content in the AT dominated positions (Fig. 2C). While tier 4 of the ARE full site requires agreement in at least 9 bases over a length of 15 bases including the 3mer spacer region, there are many examples reported where the entire half of the ARE full site is degenerate[55]. Lastly, to comprehensively describe the genome-wide coverage of AREs, tier 5, focuses on ARE half sites. Tier 5 ARE half sites show a perfect hexamer, while not complying with the requirements of tiers 1 through 4. By comparing genomic coordinates for ARE half sites and full sites, we made sure to create association with the lowest tier possible of any ARE in question. Model-based motif searches offer the possibility to further expand the degeneracy, for example by allowing imperfect ARE half sites. However, increasing numbers of degenerate motifs create limited genomic enrichment detecting almost every gene. The 5 tiers listed provided genome-wide coverage while recognizing functional relevant content. Model-based motif searches offer the possibility to further expand the degeneracy, for example by allowing imperfect ARE half sites. However, genome-wide searches with additional degeneracy did not generate other motifs than already detected by lower tiers. In total, we detected 42956 ARE full sites (tiers 1–4) and 79065 ARE half sites (tier 5).

Genomic annotation and transcriptional regulation of ARE sites. Next, we sought to compare the functional content of different ARE tiers, in particular ARE half vs full sites. In agreement with the ChIP-Seq data, the majority of AREs falls into intergenic and intronic regions (Fig. 3A). While perfect AREs account only

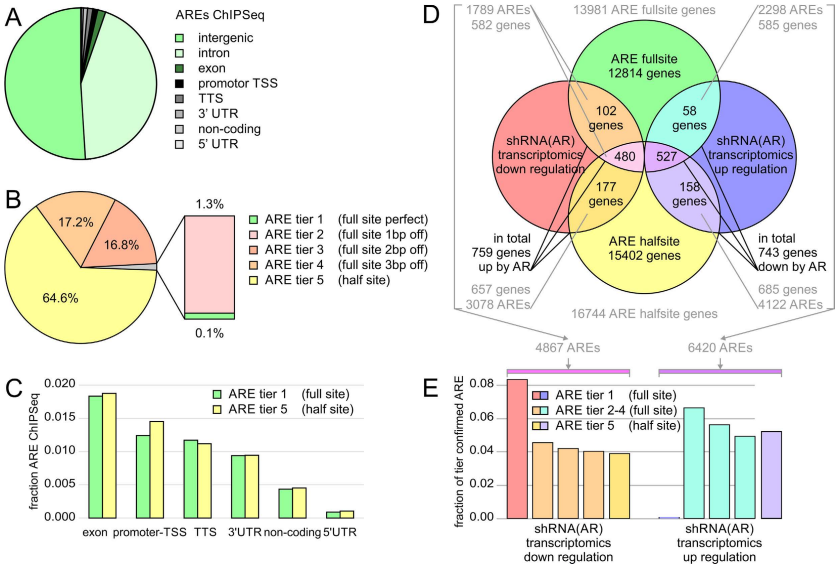


Figure 3. Genomic annotation and enrichment of AREs. (**A**) Identified AREs in ChIP-Seq experiment were annotated by genomic elements. (**B**) Classification of AREs into five tiers. There were 71 AREs in tier 1 (perfect palindromic ARE), 1583 matches in tier 2 (1 bp off perfect), 20593 in tier 3 (2 bp off perfect), 21031 in tier 4 (3 bp off perfect) and 79065 in tier 5 (half site) with p-values below 4.94E-05 in the search for motif matches. (**C**) Genomic location of ARE full and half sites. Intergenic and intronic locations are not shown. (**D**) Overlay of gene mapping of AREs identified by ChIP-Seq and transcriptomics experiments. Using this data we defined the group of 759 genes as positively regulated by AR activity (down in cell with shRNA knockdown of AR), and 743 genes as negatively regulated by AR activity. (**E**) AREs mapped by ChIP-Seq experiments were confirmed by transcriptomics experiment and matched to genes. AREs confirmed by transcriptomics experiment showed higher hit rate in better defined tiers of AREs. Hit rate plotted as fraction of tier confirmed by significant down or up regulation in stable shRNA knockdown experiment with p-values below 0.05. Different tiers of AREs suggest different effect on transcriptional outcome as well as necessity of modulation of weaker motifs by coordinating factors. Imperfect AREs make up majority of genome-wide recognition motifs.

for a fraction of the AR recognition sites, tier 3 and 4 comprise 96% of the detected full site AREs (Fig. 3B). These higher tiered AREs approximate to about twice as many isolated half sites (tier 5) for every full site (tier 1–4). ARE-harboring regions within the ChIP-Seq data were annotated to the human genome, which allowed these regions to be categorized based on their relative location and distance to the nearest gene loci. Promoter or transcription start sites (TSS, by default defined as minus 1000 bp to plus 100 bp from the start of the precursor mRNA-coding gene locus) and transcription termination sites (TTS, by default defined as minus 100 bp to plus 1000 bp from the end of the mRNA) genomic annotation are defined as being within ± 5000 bp window of the ends of the gene-coding body. Intergenic regions were defined as the remaining regions outside the gene body of TSS and TTS. The intergenic regions (50.9%) accounted for the majority of the peaks in agreement with previous ChIP-Seq experiments using AR antibodies (Fig. 3A)[56]. 18693 (43.5%) of the AREs were annotated as intronic regions, 788 (1.8%) as exonic regions, 533 (1.2%) as TSS regions, 503 (1.2%) as TTS regions, 404 (0.9%) as 3′ UTR regions, 186 (0.4%) as non-coding regions, and 38 (0.1%) as 5′ UTR regions (Fig. 3C). ARE half sites show a significantly higher fraction of TSS annotation of (1.4%), in particular bi-directional promoters, than ARE full sites (1.2%) with a p-value below 10E-04, suggesting beneficial genomic proximity for weaker half sites. However, when annotated to gene bodies there appeared to be the same proportion of full and half site containing genes, despite the larger number of ARE half sites (Fig. 3D) (Supplementary Table S3). Each tier reflected this annotation and no significant overrepresentation of functional elements were detected in individual tiers, despite ARE half sites showed more locations associated with TSS than ARE full sites. The functional role of more than 120,000 AREs detected by our searches were tested against the transcriptional response of AR knockdown in a microarray experiment. The large number of AREs indicate that there are occurrences where multiple AREs are associated with a single gene. When compared to the transcriptomic data, in total 759 genes were transcriptionally down regulated and 743 genes were up regulated upon shRNA AR knockdown. When looking at the fraction of confirmed AREs per tier there is a higher hit rate in ARE tiers closer to the canonical sequence (Fig. 3E). Despite their lower abundance in the genome compared to ARE half sites, ARE full sites are able to generate a reliable transcriptional outcome in AR positive cells.

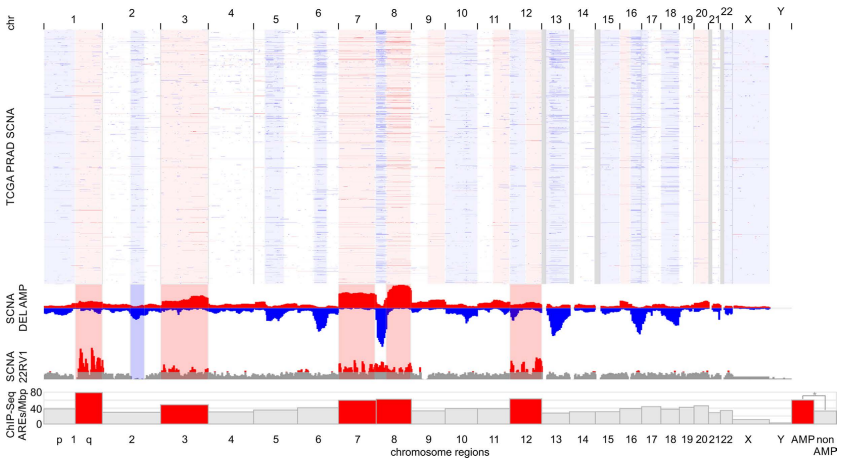


Figure 4. Copy number amplifications of the CWR22Rv1 cell line overlap with significant somatic copy number events in TCGA prostate adenocarcinoma patients and correlate with AR-ARE binding events by ChIP-Seq. Somatic copy number alteration (SCNA) profiles of 492 prostate adenocarcinoma (PRAD) patients in The Cancer Genome Atlas (TCGA) show broad arm-level events. Amplifications (AMP) are indicated in red; deletions (DEL) are indicated in blue. Low coverage next generation sequencing data of the CWR22Rv1 cell line reveals strong copy number amplification of chromosome arms 1q, 3p, 3q, 7p, 7q, 8p, 8q, 12p, and 12q. Utilization of androgen response elements by the androgen receptor is significantly elevated in amplified regions. Bar graph shows detected ARE-binding events by AR ChIP-Seq per megabase pair (Mbp).

Impact of somatic copy number alterations on genome-wide ARE utilization. Genotypic variation can modulate transcription factor binding, chromatin structure, and gene expression. In cancer progression, somatic copy number alterations (SCNAs) play critical roles by activating oncogenes and inactivating tumor suppressors[39]. In order to evaluate the significance of ARE utilization in the context of SCNAs, we determined recurring regions of copy number changes in 492 prostate adenocarcinoma (PRAD) patients in The Cancer Genome Atlas (TCGA) using the algorithm GISTIC. SCNA profiles of PRAD patients showed broad amplified arm-level events at 1q, 3p, 3q, 7p, 7q, 8p, 8q, 9q, 11q, 12q, 16p, 20p, and 20q and deleted regions at 1p, 6q, 8p, 10p, 10q, 12p, 13q, 15q, 16q, 17p, 18p, 18q, 22q, Xp, Xq, accompanied by focal events at 1p22, 1p31, 2q22, 3p13, 4q28, 5q11, 8p21, 11p11, 11q22, 12q24, 17q21, 19q13 (SCNA frequency more than 0.05; q-value less than 0.1) (Supplementary Table S4). Next, we determined SCNAs in the CWR22Rv1 cell line. Chromosome arms 1q, 3p, 3q, 7p, 7q, 8p, 8q, 12p, and 12q showed strong amplifications overlapping with SCNA regions identified in the TCGA PRAD cohort (Fig. 4). In order to assess if SCNAs modulate ARE binding, we determined the number of ChIP-Seq-detected AREs corrected for chromosome length. Regions with somatic copy number amplifications had significantly enhanced ARE utilization of 60.4 AREs/Mbp in contrast to euploidic genome regions of 32.3 AREs/Mbp with a p-value of 8.9e10-6 (Fig. 4). The amplified regions contained 5012 genes. In our ChIP-Seq and transcriptomic experiments 214 amplified genes classified as positive AR targets (down-regulation with shRNA knockdown) and showed exclusive enrichment of discrete pathways (p-value below 10e-03 and q-value below 0.1). Amplified and AR-responsive genes squalene epoxidase (SQLE, GeneBank: 6713), hydroxysteroid (17-beta) dehydrogenase 7 (HSD17B7, GeneBank: 51478), phosphomevalonate kinase (PMVK, GeneBank: 10654), lipoprotein lipase (LPL, GeneBank: 4023), v-myc avian myelocytomatosis viral oncogene homolog (MYC, GeneBank: 4609), NK3 homeobox 1 (NKX3-1, GeneBank: 4824), ELK4, ETS-domain protein (SRF accessory protein 1) (ELK4, GeneBank: 2005), PTK2B protein tyrosine kinase 2 beta (PTK2B, GeneBank: 2185), and zinc finger and BTB domain containing 10 (ZBTB10, GeneBank: 65986) are pathway members of the androgen response, steroid biosynthesis, and cholesterol homeostasis. In addition, amplified and AR-regulated genes showed enrichment in MTORC1 signaling, DNA replication, cell cycle, MYC targets, mismatch repair, homologous recombination, nucleotide excision repair, epigenetic regulators, and pathways in cancer. The detected ARE recognition by the androgen receptor displays a potential mechanism how SCNAs get translated to a functional, oncogenic level in prostate cancer.

Comparison of ARE utilization in androgen-insensitive and androgen-responsive prostate cancer cell lines. Androgen-insensitive cell lines, such as CWR22Rv1, express AR(FL) and AR splice variants[28]. In contrast, androgen-responsive prostate cancer models expressing exclusively AR(FL), such as LNCaP, offer insights into steroid-dependent gene regulation[30]. We applied the established model-based ARE annotation to condition-specific CWR22Rv1 and LNCaP ChIP-Seq samples and assessed ARE utilization dependent on AR splice isoforms as well as 5α-dihydrotestosterone treatment (Fig. 5). ARE binding by ChIP-Seq was assessed in

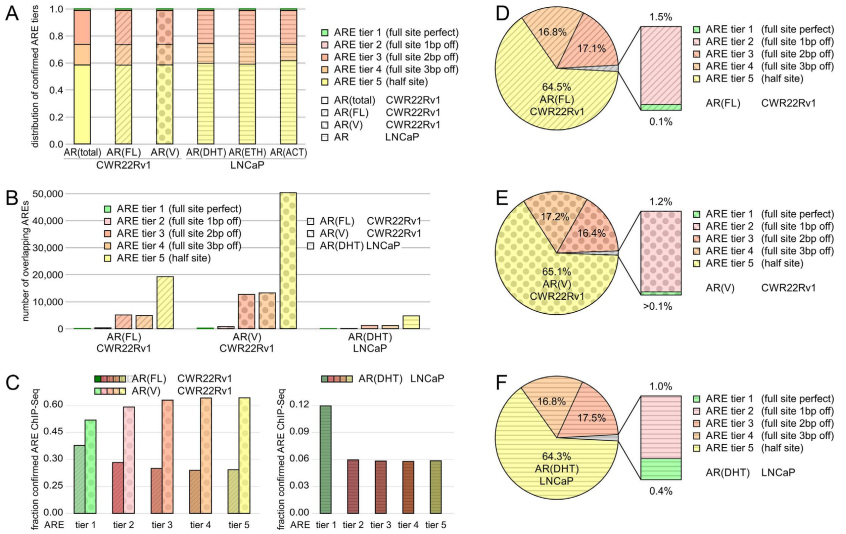


Figure 5. Utilization of androgen response elements in androgen-dependent LNCaP and androgen-independent CWR22Rv1 cellular models of prostate cancer. ARE binding by ChIP-Seq was assessed in the CWR22Rv1 cell line (total binding; binding by full-length androgen receptor, AR(FL); binding by variant androgen receptor, AR(V)) and in the LNCaP cell line (5α-dihydrotestosterone treatment, AR(DHT); ethanol treatment, AR(EtOH); functionally active steroid-bound AR corrected for EtOH background, AR(ACT)). (**A**) Comparison of AREs detected and confirmed by ChIP-Seq in AR(FL), AR(V), and AR(DHT) specimen. (**B**) Distribution of confirmed AREs by ChIP-Seq in CWR22Rv1 and LNCaP cells. (**C**) Fraction of confirmed AREs by ChIP-Seq in CWR22Rv1 and LNCaP cells. Pie charts visualizing distribution of ARE tiers detected and confirmed by ChIP-Seq in D) AR(FL) in CWR22Rv1 cells, (**E**) AR(V) in CWR22Rv1 cells, and (**F**) AR(DHT) in LNCaP cells. Tier 1 is highlighted in green, tier 2 in magenta, tier 3 in red, tier 4 in orange, tier 5 in yellow. ChIP-Seq analysis for AR(FL) is shaded with tilted lines, AR(V) with dots, and AR(DHT) with horizontal lines.

the CWR22Rv1 cell line for total AR binding, AR(total), for binding to full-length androgen receptor, AR(FL), for binding by variant androgen receptor, AR(V), in the LNCaP cell line for 5α-dihydrotestosterone treatment, AR(DHT), for ethanol treatment, AR(EtOH), and for functionally active steroid-bound AR corrected for EtOH background AR(ACT). All quantified conditions showed a similar distribution of ARE tiers 1–5 of about 0.001, 0.017, 0.244, 0.150, and 0.588, respectively (Fig. 5A). The data validates that utilization of perfect ARE tier 1 is a rare event, independent of AR splice-variants or steroid condition. Cross-validation of ChIP-Seq experiments of AR(total) vs AR(FL) confirmed 30,022 AREs assigned to AR(FL)-binding in the CWR22Rv1 cell line (Fig. 5B). AR(V) isoforms with 78,350 ARE ChIP-Seq events bind to DNA autonomous of full-length androgen receptor in the absence of androgen and modulate a unique set of genes that is not regulated by full-length androgen receptor (Fig. 5B). In contrast, AR(DHT) in the LNCaP cell line displayed a set of 7,361 AREs common to the CWR22Rv1 cell line (Fig. 5B). Next, we quantified the fraction of ARE tiers confirmed in overlapping ChIP-Seq experiments. AR(FL) showed an incremental reduced fraction with higher, less specific ARE tiers in the CWR22Rv1 cell line, while AR(V) isoforms had a stronger overlap with more degenerate motifs (Fig. 5C). The AR(DHT) LNCaP condition showed a trend similar to AR(FL) in the CWR22Rv1 cell line (Fig. 5C). The distribution of ARE tiers with AR splice-variants or with steroid treatment in the two tested PRAD cell lines showed most variation in the perfect AREs of tier 1. Notably, the frequency of perfect AREs correlates with AR specificity and increases from AR(V) isoforms to AR(FL) in the CWR22Rv1 cell line and quadruples in the AR(DHT) LNCaP condition (Fig. 5D–F). All evaluated conditions showed high agreement utilization of ARE tiers 3–5, half sites and imperfect full sites (Fig. 5D–F).

Network of transcriptional cooperation of the androgen receptor. We next sought to identify potential transcription factors that would cooperate with the AR to regulate gene expression. Using the Jaspar motif database, we grouped significant transcription factor logos within a window of ±160 bp from the ARE with a p-value of less than 0.05. Top hits included forkhead box (FOX), Krüppel-like factors (KLF), basic helix-loop-helix (BHLH), sterol regulatory element binding factor (SREBF), and v-myc avian myelocytomatosis viral oncogene homolog (MYC) families of transcription factors. Interestingly, several members of cooperating transcription factor families showed amplifications at the copy number level in the CWR22RV1 cell line as well as in TCGA PRAD patients. Detected somatic amplifications of transcription factors were maintained at

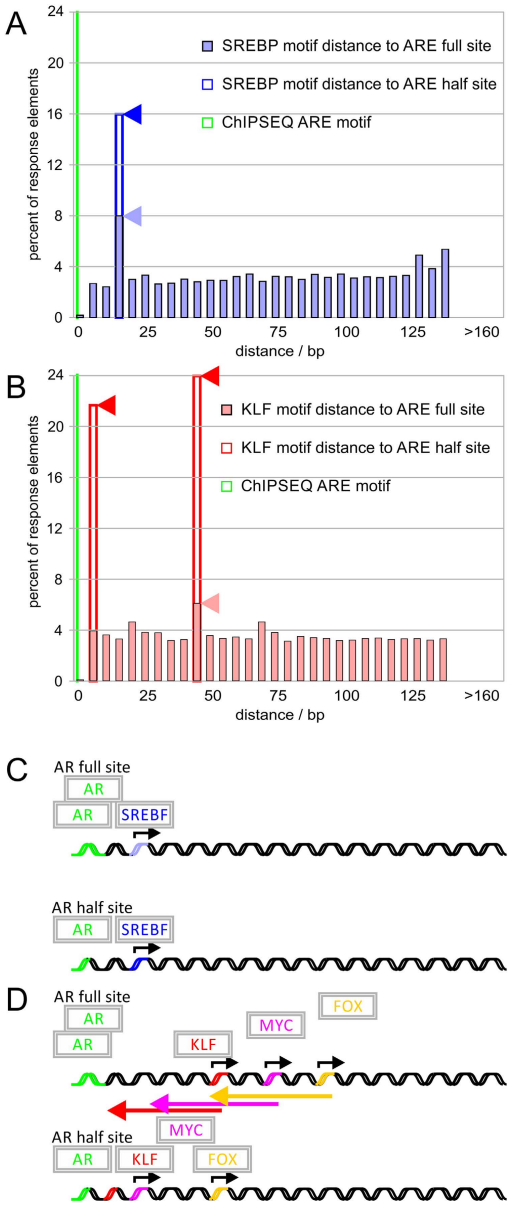


Figure 6. Enhancement of androgen response elements by cooperation of the androgen receptor with other transcription factors. Transcription factor complexes were inferred from ChIP-Seq data using spaced motif analysis in reference to detected AR binding sites (green). The affinity of weaker ARE transcription factor sites can be enhanced by cooperation with other transcription factors. (**A**) SREBF transcription factor family (blue) shows an increase in the percent of response elements found 15 bp from the ARE half site. (**B**) KLF transcription factor motif family (red) shows an increased fraction bound in AR ChIP-Seq signals. In addition, peaks of KLF motifs in 5 bp and 45 bp distance to ARE half sites shows increased proximity compared to association with ARE full sites. The weaker ARE half-site shows strong cooperation with (**C**) SREBF and closer motif distance with (**D**) KLF, MYC, and FOX, transcription factor families with the AR.

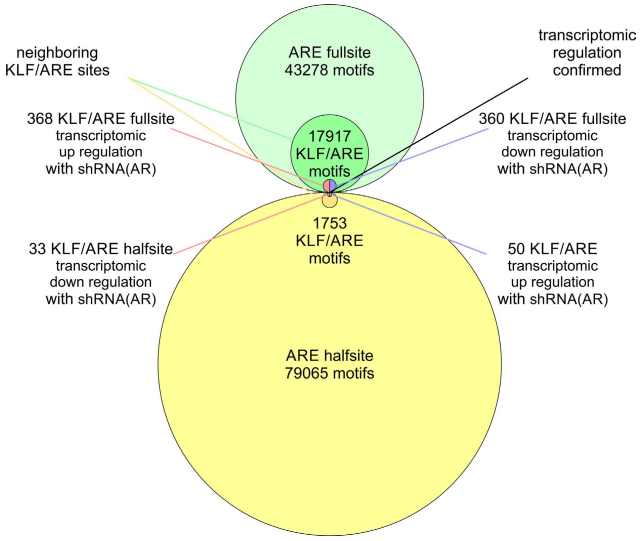


Figure 7. Synergy of ARE and KLF motifs in androgen receptor-mediated transcriptional responses. Quantification of KLF sites with respect to detected ARE full sites and half sites. 17917 full sites coincide with KLF sites. 25 fell into tier 1, 611 into tier 2, 8067 into tier 3, and 9214 into tier 4. Stronger cooperation of ARE full sites with KLF motifs result in larger transcriptional response (368 and 360 genes up and down in cells with shRNA knockdown of AR, respectively) in contrast to ARE half sites (33 and 50 genes up and down, respectively).

the transcriptional level. Somatic copy number amplifications on chromosome 3 included Krüppel-like family member Krüppel-like factor 15 (KLF15, GeneBank: 28999). Similarly, amplified regions on chromosome 7 included transcription factors v-myc avian myelocytomatosis viral oncogene homolog (MYC, GeneBank:4609) and forkhead box K1 (FOXK1, GeneBank: 28999). We then organized recorded distances of detected, enriched, and/or amplified transcription factor families to the ARE into histograms with 5bp bins and analyzed spacing of the transcription factor motifs. We noticed an increase of detected transcription factor motifs at 15 bp for KLF, and 45 bp for SREBF-related transcription factors between full and half site ARE (Fig. 6A,B). This suggests that for the AR to recognize any weaker half site response element, cooperation of other transcription factors might be required (Fig. 6C,D). Distance of KLF, MYC, or FOX motifs to ARE half sites was reduced in comparison with distance to ARE full sites (Fig. 6D). Taken together, the transcription factor network analysis (top hits in motif enrichment with p-values below 0.05) suggests that KLF, MYC, FOX, and SREBF families of transcription factors have the ability to utilize motifs in the cistrome of AREs and to cooperate with the AR.

Quantifying the transcriptional response between full and half site ARE with KLF. Next, we characterized transcription factor cooperation of the AR with the Krüppel-like family. We quantified the number of KLF sites with respect to the detected ARE tiers to evaluate effect of transcriptional cooperation with respect to motif degeneracy (Supplementary Table S5 and Supplementary Table S6). We found that 17917 full sites coincide with KLF sites. 25 fell into tier 1, 611 into tier 2, 8067 into tier 3, and 9214 into tier 4. When comparing between detected full and half sites there was a stronger cooperation of the ARE full sites with KLF motifs resulting in larger transcriptional response with 360 and 368 genes up and down regulated in contrast to ARE half sites, which had 50 and 33 genes up and down regulated (Fig. 7). Despite a larger number of weaker ARE half sites found in the proximity of KLF motifs, stronger AREs next to KLF motifs resulted in a larger transcriptional response. Genes associated with pathways in cancer as well as TP53 signaling were enriched in genes with KLF and ARE full site motifs with p-values below 0.05 (Supplementary Table S5). For KLF and ARE half site motifs pathways of extracellular matrix-receptor interaction and focal adhesion with p-values below 0.001 were found. Consequently, the data suggests that KLF may modulate the binding of AR with both weaker and stronger AREs, which control genes with distinct function.

Functional enrichment of androgen receptor binding sites. Gene set enrichment analysis of identified putative AR target genes revealed several functional clusters. In addition to the identified 759 and 743 genes confirmed by ChIP-Seq binding as well as transcriptional activity, we tested all ARE full site and half site tiers for pathway enrichment (Supplementary Table S5). Gene sets corresponding to full site AREs with activating gene expression (582 genes as positively regulated by AR activity with ARE full or half sites; 102 with exclusively full sites) revealed 41 including 21 exclusive pathways significantly enriched with p-values below 0.05.

For half site AREs (585 genes as positively regulated by AR activity with ARE full or half sites; 53 with exclusively full sites) 39 including 12 exclusive pathways were found with p-values below 0.05. Pathways in both sets included DNA replication, cell cycle control, and metabolic pathways. Of particular interest were pathways that were exclusively assigned to ARE full sites or half sites. The set of AR target genes with ARE full sites focused on pyrimidine metabolism, terpenoid backbone biosynthesis, one-carbon metabolism, glycine, serine and threonine metabolism with p-values below 0.001. The metabolic program—framed by genes with full site AREs—supports proliferative functions required for cellular maintenance. In contrast, the set of AR target genes with ARE half sites included steroid biosynthesis, terpenoid backbone biosynthesis, peroxisome, pentose phosphate pathway, glycerolipid metabolism, and mitogen activated protein kinase signaling pathway with p-values below 0.001. An enrichment of genes containing ARE half sites involved in lipid and steroid biosynthesis could point to the gender-, development-, and tissue-specific control in prostatic differentiation. Therefore, the gene set enrichment analysis suggests that AR targets genes controlled by ARE full sites and/or ARE half sites have common proliferative functions but also distinct biological functions involved in lipid metabolism.

Discussion

We refined AR-binding and AREs in AR-positive but androgen-insensitive CWR22Rv1 prostate cancer cells using ChIP-Seq and motif-guided genome-wide analysis. The independence of androgen ligand for AR activity makes the CWR22Rv1 cell line a favorable model for genome-scale characterization of AR binding events in prostate cancer. Therefore, one can accomplish a comprehensive, unprecedented picture of the ARE by using ChIP-Seq analysis of AR-specific immunoprecipitation. We classified AREs according to their degeneracy and their transcriptional involvement. We quantified ARE utilization in response to somatic copy number amplifications, AR splice-variants, and steroid treatment. Our AR ChIP-Seq mapping shows that a majority of AREs are imperfect AREs with several base pairs deviating from the canonical palindromic 15 mer ARE full site. Our results fit in with previous assessments that AR DNA binding is possible despite base pair deviation from the canonical ARE full site sequence[23]. Although imperfect AREs make up 99.9% of the motifs, the degree of degeneracy correlates negatively with validated transcriptional outcome. Weaker AREs, particularly ARE half sites, benefit from neighboring motifs or cooperating transcription factors in regulating gene expression. In addition, ARE half sites showed enrichment of mitogen activated protein kinase signaling required for stimulation of proliferation. Therefore, the ability to regulate weaker AREs carries weight for prostatic development and oncogenic control[23].

Somatic copy number alterations of TCGA prostate cancer patients correlated with amplifications of chromosome arms 1q, 3p, 3q, 7p, 7q, 8p, 8q, 12p, and 12q observed in the cellular CWR22Rv1 prostate cancer model. Aneuploidy of the CWR22Rv1 model was recognized early on during the genetic characterization of the metastatic cell line[27]. Most SCNA studies in prostate cancer have focused on AR gene amplification found in the majority (87%) of CRPC tumors[57–59]. The number of recognized ARE sites increases with the level of AR expression[60]. A study in LNCaP cells described AR recruitment to amplified chromosomal regions found in metastatic PRAD implicating AR co-amplification with SCNAs[61]. Similarly, our SCNA data showed a higher density of ARE binding events in amplified regions suggesting that ARE recognition and amplification might play a role in carrying SCNAs to a functional level in prostate cancer.

We applied genome-wide model-based motif searches to different prostate cancer models and investigated ARE occurrence dependent on AR-splice variants and steroid regulation. Derived from an androgen-dependent CWR22 mouse xenograft that relapsed during androgen deprivation, the CWR22Rv1 prostate cancer cell line is androgen-insensitive, expresses different AR isoforms[4,8,9,27,62], and is expected to display enhanced ARE recognition. ChIP-Seq data on the CWR22Rv1 cell line revealed 122021 AREs (32.0% ARE full sites; 68.0% ARE half sites). 88.5% of the detected AREs overlapped with ChIP-Seq data of published experiments after processing using the same bioinformatics workflow[28]. AR(V) isoforms recognize 2.6 times more AREs than AR(FL) in the isoform-specific ARE characterization of the CWR22Rv1 cell line. In contrast, the LNCaP cell line serves as a model for primary prostate tumors that are responsive to ADT therapy[29]. Using genome-wide model-based motif searches, the LNCaP cell line under testosterone-treatment displayed 6.0% overlap with the CWR22Rv1 cell line resembling the fraction of AREs recognized under androgen stimulation[30]. The fraction of AREs activated under androgen stimulation in LNCaP is small compared to the total number of AREs recognized in the CWR22Rv1 model. Dihydrotestosterone-activated AR response resembles a narrow, well-defined, physiological gene-expression program required for prostatic function[30,63]. In contrast, the altered and enhanced spectrum of AREs assigned to shortened, non-specific CWR22Rv1 AR(V) isoforms mediates an oncogenic gene expression program that is able to circumvent androgen deprivation, support continued proliferation, and drive CRPC[8,20,28].

AREs can be modulated by cooperation with other transcription factors, which can compensate for missing canonical contacts of imperfect AREs and nevertheless result in successful gene expression events[23,55]. The cistrome of investigated AREs actively participates in the AR controlled gene expression within CRPC[64,65]. A model of susceptibility to cooperation of weak AREs with neighboring transcription factors has already been confirmed for different transcriptional networks with the AR[55,23]. In our motif searches, the validated AR-cooperating transcription factor FOXA1 displayed motif enrichment, differential up-regulation, and high tumor expression, serving as test and validation data point. Overexpression of FOXA1 promotes cell cycle progression, CRPC signature, and has been described as playing a crucial role in assisting AR site recognition for weaker binding sites by creating excessive open chromatin sites[66–68]. The transcriptional program required for prostate-specific gene expression of steroid biosynthesis enzymes is distinctly enriched in ARE half sites and consistent with modular and tightly regulated tissue-specific control[69,70]. In PRAD, cooperation of SREBF and AR signaling has been linked to progression of prostate cancer cells[71]. SREBF targeted gene expression drops drastically after androgen deprivation therapy, but their transcription program re-emerges upon reactivation of the AR transcriptional program manifesting the hypothesis of transcriptional cooperation[72]. In addition, SREBF has been shown to be

recruited with the AR to target gene promoters[73]. In the Krüppel-like transcription factor family, motif enrichment and somatic amplification of chromosome 3 suggest KLF15 as candidate transcription factor to support the AR. While the genomic data provide strong evidence for cooperation, experimental validation will be required to solidify this finding. Other cellular models have linked KLF15 overexpression with enhanced recruitment of nuclear receptors[74,75]. Previous analyses of the KLF15 promoter detected AREs suggesting hormonal control of its transcription[76]. Development- and cell-specific expression of KLF isoforms may modulate AR signaling, where AREs and KLF motifs fall in close proximity. Taken together, promiscuous recognition of androgen receptor cognate sites guided by transcriptional cooperation may allow for dynamic, tissue-specific regulation.

A predominant signature of imperfect AREs is the GC content of the 3 bp spacer as well as the flanking regions of extended motifs. The GC content is richest in higher, less perfect tiers of AREs. CpG dinucleotides within the genome enable the cell to control ARE availability by DNA 5′-cytosine methylation[77,78]. Actively transcribed genomic regions, in particular promoter or transcription factor cognate sites, tend to have less DNA methylation[79–81]. Therefore, weaker AREs are subject to control of gene expression by the dynamic equilibrium of histone and DNA methyltransferases and demethylases[82]. Conformational change of the protein structure plays an important role in the ambiguity of ARE recognition. The AR is able to recognize weaker sites by first binding to a high affinity AGAACA site followed by a strong conformation change in the protein and possibly in the ARE[83,84]. While analyzing the different tiers of the ARE in our data we found that the 1–3 nucleotides would be off in only one of the hexamers recognized by the homodimer DBD. Additionally, as we progressed to the higher tiers the G and C in positions 2 and 5 in the AR hexamer remained conserved. Our work suggests that ARE full sites are able to bind to and recognize weaker sites through binding of a stronger half site within the full site and that the weaker site is more dependent upon the G and C residues for complete homodimer binding. Further, the GC content is also important for the spectrum of cooperating transcription factors with the AR. The KLF family recognizing GC/GT boxes has been implicated in regulation of oncogenic expression signatures in LNCaP and PC3 prostate cancer cell lines[85,86]. Transcription factors cooperating with the AR form an important regulatory hierarchy governing androgen-dependent gene expression in normal as well as malignant prostate tissue and offer potential new opportunities for therapeutic intervention.

Conclusion

Our data refined the recognition of ARE sequences within CWR22Rv1 and LNCaP prostate cancer cell lines. We expanded the nucleotide specificity of the ARE, identified potential modes of regulation these response elements are subject to, and outlined a protocol for identifying coordinating transcription factors to assist with weaker site recognition. While a major disadvantage remains in possible false-negatives being identified in computationally predicted sites, future experimental verification will have to determine if these response elements play a role in gene regulation. Future ChIP-Seq studies could look into prominent histone modifications accompanying detected response element sites within prostate cancer and facilitate insight how chromatin alteration affects AR gene targeting. Importantly, we identified significant differences in the genomic landscape of ARE full and half sites. Despite the fact that ARE half sites outnumber ARE full sites by 2-fold, stronger ARE were more frequently confirmed at the transcriptional level than weaker AREs. Nevertheless, weaker AREs are affected by AR expression or regulation, and may have strong functional impact by multiplicity and/or genomic proximity. ARE impact depends on somatic alterations, motif-receptor-binding specificity, tissue-specific lineage, cooperating factors (e.g. FOX, SREBF, MYC, KLF), distance to neighboring motifs, AR-splice variants, and steroid regulation.

References

1. Siegel, R., Ma, J., Zou, Z. & Jemal, A. Cancer statistics, 2014. *CA Cancer J Clin.* **64,** 9–29, doi: 10.3322/caac.21208 (2014).
2. Zhang, X. *et al.* Amplification and protein expression of androgen receptor gene in prostate cancer cells: Fluorescence hybridization analysis. *Oncol Lett.* **9,** 2617–2622, doi: 10.3892/ol.2015.3114 (2015).
3. Poelaert, F. *et al.* The role of androgen receptor expression in the curative treatment of prostate cancer with radiotherapy: a pilot study. *Biomed Res Int.* **2015,** 812815, doi: 10.1155/2015/812815 (2015).
4. Cao, B. *et al.* Androgen receptor splice variants activating the full-length receptor in mediating resistance to androgen-directed therapy. *Oncotarget* **5,** 1646–1656, doi: 10.18632/oncotarget.1802 (2014).
5. Bain, D. L., Heneghan, A. F., Connaghan-Jones, K. D. & Miura, M. T. Nuclear receptor structure: implications for function. *Annu Rev Physiol.* **69,** 201–220, doi: 10.1146/annurev.physiol.69.031905.160308 (2007).
6. Attard, G., Cooper, C. S. & de Bono, J. S. Steroid hormone receptors in prostate cancer: a hard habit to break? *Cancer Cell* **16,** 458–462, doi: 10.1016/j.ccr.2009.11.006 (2009).
7. Chen, C. D. *et al.* Molecular determinants of resistance to antiandrogen therapy. *Nat Med* **10,** 33–39, doi: 10.1038/nm972 (2004).
8. Dehm, S. M., Schmidt, L. J., Heemers, H. V., Vessella, R. L. & Tindall, D. J. Splicing of a novel androgen receptor exon generates a constitutively active androgen receptor that mediates prostate cancer therapy resistance. *Cancer Res.* **68,** 5469–5477, doi: 10.1158/0008-5472.CAN-08-0594 (2008).
9. Sun, S. *et al.* Castration resistance in human prostate cancer is conferred by a frequently occurring androgen receptor splice variant. *J Clin Invest* **120,** 2715–2730, doi: 10.1172/JCI41824 (2010).
10. Zhang, X. *et al.* Androgen receptor variants occur frequently in castration resistant prostate cancer metastases. *Plos One* **6,** e27970, doi: 10.1371/journal.pone.0027970 (2011).
11. Li, Y. *et al.* AR intragenic deletions linked to androgen receptor splice variant expression and activity in models of prostate cancer progression. *Oncogene* **31,** 4759–4767, doi: 10.1038/onc.2011.637 (2012).
12. Hu, R. *et al.* Ligand-independent androgen receptor variants derived from splicing of cryptic exons signify hormone-refractory prostate cancer. *Cancer Res.* **69,** 16–22, doi: 10.1158/0008-5472.CAN-08-2764 (2009).
13. Montgomery, R. B. *et al.* Maintenance of intratumoral androgens in metastatic prostate cancer: a mechanism for castration-resistant tumor growth. *Cancer Res.* **68,** 4447–4454, doi: 10.1158/0008-5472.CAN-08-0249 (2008).
14. Cai, C. & Balk, S. P. Intratumoral androgen biosynthesis in prostate cancer pathogenesis and response to therapy. *Endocr Relat Cancer* **18,** R175–R182, doi: 10.1530/ERC-10-0339 (2011).
15. Mostaghel, E. A. Steroid hormone synthetic pathways in prostate cancer. *Transl Androl Urol.* **2,** 212–227, doi: 10.3978/j.issn.2223-4683.2013.09.16 (2013).

16. Bohl, C. E., Wu, Z., Miller, D. D., Bell, C. E. & Dalton, J. T. Crystal structure of the T877A human androgen receptor ligand-binding domain complexed to cyproterone acetate provides insight for ligand-induced conformational changes and structure-based drug design. *J Biol Chem.* **282,** 13648–13655, doi: 10.1074/jbc.M611711200 (2007).
17. Marcias, G. *et al.* Identification of novel truncated androgen receptor (AR) mutants including unreported pre-mRNA splicing variants in the 22Rv1 hormone-refractory prostate cancer (PCa) cell line. *Hum Mutat.* **31,** 74–80, doi: 10.1002/humu.21138 (2010).
18. Li, Y. *et al.* Intragenic rearrangement and altered RNA splicing of the androgen receptor in a cell-based model of prostate cancer progression. *Cancer Res.* **71,** 2108–2117, doi: 10.1158/0008-5472.CAN-10-1998 (2011).
19. Tepper, C. G. *et al.* Characterization of a novel androgen receptor mutation in a relapsed CWR22 prostate cancer xenograft and cell line. *Cancer Res.* **62,** 6606–6614 (2002).
20. Yamamoto, Y. *et al.* Generation 2.5 antisense oligonucleotides targeting the androgen receptor and its splice variants suppress enzalutamide-resistant prostate cancer cell growth. *Clin Cancer Res* **21,** 1675–1687, doi: 10.1158/1078-0432.CCR-14-1108 (2015).
21. Li, Y. *et al.* Androgen receptor splice variants mediate enzalutamide resistance in castration-resistant prostate cancer cell lines. *Cancer Res.* **73,** 483–489, doi: 10.1158/0008-5472.CAN-12-3630 (2013).
22. Roche, P. J., Hoare, S. A. & Parker, M. G. A consensus DNA-binding site for the androgen receptor. *Mol Endocrinol.* **6,** 2229–2235, doi: 10.1210/mend.6.12.1491700 (1992).
23. Wang, Q. *et al.* A hierarchical network of transcription factors governs androgen receptor-dependent prostate cancer growth. *Mol Cell* **27,** 380–392, doi: 10.1016/j.molcel.2007.05.041 (2007).
24. Sahu, B. *et al.* Dual role of FoxA1 in androgen receptor binding to chromatin, androgen signalling and prostate cancer.*EMBO J* **30,** 3962–3976, doi: 10.1038/emboj.2011.328 (2011).
25. He, B. *et al.* GATA2 facilitates steroid receptor coactivator recruitment to the androgen receptor complex. *Proc Natl Acad Sci USA* **111,** 18261–18266, doi: 10.1073/pnas.1421415111 (2014).
26. Pretlow, T. G. *et al.* Xenografts of primary human prostatic carcinoma. *J Natl Cancer Inst* **85,** 394–398 (1993).
27. Sramkoski, R. M. *et al.* A new human prostate carcinoma cell line, 22Rv1. *In vitro Cell Dev Biol Anim* **35,** 403–409, doi: 10.1007/s11626-999-0115-4 (1999).
28. Lu, J. *et al.* The cistrome and gene signature of androgen receptor splice variants in castration resistant prostate cancer cells. *J Urol.* **193,** 690–698, doi: 10.1016/j.juro.2014.08.043 (2015).
29. Horoszewicz, J. S. *et al.* LNCaP model of human prostatic carcinoma. *Cancer Res.* **43,** 1809–1818 (1983).
30. Tan, P. Y. *et al.* Integration of regulatory networks by NKX3-1 promotes androgen-dependent prostate cancer survival. *Mol Cell Biol.* **32,** 399–414, doi: 10.1128/MCB.05958-11 (2012).
31. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10,** R25, doi: 10.1186/gb-2009-10-3-r25 (2009).
32. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9,** R137, doi: 10.1186/gb-2008-9-9-r137 (2008).
33. Feng, J., Liu, T. & Zhang, Y. Using MACS to identify peaks from ChIP-Seq data. *Curr Protoc Bioinformatics* Chapter 2, Unit 2 14, doi: 10.1002/0471250953.bi0214s34 (2011).
34. Chen, T. W. *et al.* ChIPseek, a web-based analysis tool for ChIP data. *BMC Genomics* **15,** 539, doi: 10.1186/1471-2164-15-539 (2014).
35. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26,** 841–842, doi: 10.1093/bioinformatics/btq033 (2010).
36. Kent, W. J. *et al.* The human genome browser at UCSC. *Genome Res.* **12,** 996–1006, doi: 10.1101/gr.229102. Article published online before print in May 2002 (2002).
37. Robinson, J. T. *et al.* Integrative genomics viewer. *Nat Biotechnol.* **29,** 24–26, doi: 10.1038/nbt.1754 (2011).
38. Mermel, C. H. *et al.* GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol.* **12,** R41, doi: 10.1186/gb-2011-12-4-r41 (2011).
39. Zack, T. I. *et al.* Pan-cancer patterns of somatic copy number alteration. *Nat Genet* **45,** 1134–1140, doi: 10.1038/ng.2760 (2013).
40. Bailey, T. L. & Gribskov, M. Combining evidence using p-values: application to sequence homology searches. *Bioinformatics* **14,** 48–54, doi: btb036 (1998).
41. Grant, C. E., Bailey, T. L. & Noble, W. S. FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27,** 1017–1018, doi: 10.1093/bioinformatics/btr064 (2011).
42. Bailey, T. L., Johnson, J., Grant, C. E. & Noble, W. S. The MEME Suite. *Nucleic Acids Res* **43,** W39–W49, doi: 10.1093/nar/gkv416 (2015).
43. Bailey, T. L. & Elkan, C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol.* **2,** 28–36 (1994).
44. Bailey, T. L. DREME: motif discovery in transcription factor ChIP-seq data. *Bioinformatics* **27,** 1653–1659, doi: 10.1093/bioinformatics/btr261 (2011).
45. McLeay, R. C. & Bailey, T. L. Motif Enrichment Analysis: a unified framework and an evaluation on ChIP data. *BMC Bioinformatics* **11,** 165, doi: 10.1186/1471-2105-11-165 (2010).
46. Whitington, T., Frith, M. C., Johnson, J. & Bailey, T. L. Inferring transcription factor complexes from ChIP-seq data. *Nucleic Acids Res.* **39,** e98, doi: 10.1093/nar/gkr341 (2011).
47. Smyth, G. K., Michaud, J. & Scott, H. S. Use of within-array replicate spots for assessing differential expression in microarray experiments. *Bioinformatics* **21,** 2067–2075, doi: 10.1093/bioinformatics/bti270 (2005).
48. Huber, W., von Heydebreck, A., Sultmann, H., Poustka, A. & Vingron, M. Variance stabilization applied to microarray data calibration and to the quantification of differential expression. *Bioinformatics* **18** Suppl 1, S96–104 (2002).
49. Benjamini, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B* (*Methodological*) **57,** 289–300 (1995).
50. Ma, W., Noble, W. S. & Bailey, T. L. Motif-based analysis of large nucleotide data sets using MEME-ChIP. *Nat Protoc* **9,** 1428–1450, doi: 10.1038/nprot.2014.083 (2014).
51. Mathelier, A. *et al.* JASPAR 2014: an extensively expanded and updated open-access database of transcription factor binding profiles. *Nucleic Acids Res.* **42,** D142–D147, doi: 10.1093/nar/gkt997 (2014).
52. Chen, H. *et al.* Genome-wide analysis of androgen receptor binding and gene regulation in two CWR22-derived prostate cancer cell lines. *Endocr Relat Cancer* **17,** 857–873, doi: 10.1677/ERC-10-0081 (2010).
53. Kaspar, F., Klocker, H., Denninger, A. & Cato, A. C. A mutant androgen receptor from patients with Reifenstein syndrome: identification of the function of a conserved alanine residue in the D box of steroid receptors. *Mol Cell Biol.* **13,** 7850–7858 (1993).
54. Verrijdt, G., Haelens, A. & Claessens, F. Selective DNA recognition by the androgen receptor as a mechanism for hormone-specific regulation of gene expression. *Mol Genet Metab* **78,** 175–185, doi: S1096719203000039 (2003).
55. Massie, C. E. *et al.* New androgen receptor genomic targets show an interaction with the ETS1 transcription factor. *EMBO Rep.* **8,** 871–878, doi: 10.1038/sj.embor.7401046 (2007).
56. Zhang, Y., Huang, Z., Zhu, Z., Liu, J. & Zheng, X. Network analysis of ChIP-Seq data reveals key genes in prostate cancer. *Eur J Med Res.* **19,** 47, doi: 10.1186/s40001-014-0047-7 (2014).
57. Visakorpi, T. *et al. In vivo* amplification of the androgen receptor gene and progression of human prostate cancer. *Nat Genet* **9,** 401–406, doi: 10.1038/ng0495-401 (1995).
58. Bubendorf, L. *et al.* Survey of gene amplifications during prostate cancer progression by high-throughout fluorescence *in situ* hybridization on tissue microarrays. *Cancer Res.* **59,** 803–806 (1999).

59. Qu, X. *et al.* A three-marker FISH panel detects more genetic aberrations of AR, PTEN and TMPRSS2/ERG in castration-resistant or metastatic prostate cancers than in primary prostate tumors. *Plos One* **8,** e74671, doi: 10.1371/journal.pone.0074671 (2013).
60. Urbanucci, A. *et al.* Overexpression of androgen receptor enhances the binding of the receptor to the chromatin in prostate cancer. *Oncogene* **31,** 2153–2163, doi: 10.1038/onc.2011.401 (2012).
61. Labbe, D. P. *et al.* Prostate cancer genetic-susceptibility locus on chromosome 20q13 is amplified and coupled to androgen receptor-regulation in metastatic tumors. *Mol Cancer Res.* **12,** 184–189, doi: 10.1158/1541-7786.MCR-13-0477 (2014).
62. Guo, Z. *et al.* A novel androgen receptor splice variant is up-regulated during prostate cancer progression and promotes androgen depletion-resistant growth. *Cancer Res.* **69,** 2305–2313, doi: 10.1158/0008-5472.CAN-08-3795 (2009).
63. Hu, R. *et al.* Distinct transcriptional programs mediated by the ligand-dependent full-length androgen receptor and its splice variants in castration-resistant prostate cancer. *Cancer Res.* **72,** 3457–3462, doi: 10.1158/0008-5472.CAN-11-3892 (2012).
64. Starick, S. R. *et al.* ChIP-exo signal associated with DNA-binding motifs provides insight into the genomic binding of the glucocorticoid receptor and cooperating transcription factors. *Genome Res.* **25,** 825–835, doi: 10.1101/gr.185157.114 (2015).
65. Wang, D. *et al.* Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA. *Nature* **474,** 390–394, doi: 10.1038/nature10006 (2011).
66. Jin, H. J., Zhao, J. C., Wu, L., Kim, J. & Yu, J. Cooperativity and equilibrium with FOXA1 define the androgen receptor transcriptional program. *Nat Commun.* **5,** 3972, doi: 10.1038/ncomms4972 (2014).
67. Sahu, B. *et al.* FoxA1 specifies unique androgen and glucocorticoid receptor binding events in prostate cancer cells. *Cancer Res.* **73,** 1570–1580, doi: 10.1158/0008-5472.CAN-12-2350 (2013).
68. Robinson, J. L. *et al.* Elevated levels of FOXA1 facilitate androgen receptor chromatin binding resulting in a CRPC-like phenotype. *Oncogene* **33,** 5666–5674, doi: 10.1038/onc.2013.508 (2014).
69. Swinnen, J. V., Ulrix, W., Heyns, W. & Verhoeven, G. Coordinate regulation of lipogenic gene expression by androgens: evidence for a cascade mechanism involving sterol regulatory element binding proteins. *Proc Natl Acad Sci USA* **94,** 12975–12980 (1997).
70. Heemers, H. *et al.* Androgens stimulate lipogenic gene expression in prostate cancer cells by activation of the sterol regulatory element-binding protein cleavage activating protein/sterol regulatory element-binding protein pathway. *Mol Endocrinol.* **15,** 1817–1828, doi: 10.1210/mend.15.10.0703 (2001).
71. Huang, W. C., Li, X., Liu, J., Lin, J. & Chung, L. W. Activation of androgen receptor, lipogenesis, and oxidative stress converged by SREBP-1 is responsible for regulating growth and progression of prostate cancer cells. *Mol Cancer Res.* **10,** 133–142, doi: 10.1158/1541-7786.MCR-11-0206 (2012).
72. Myers, R. B., Oelschlager, D. K., Weiss, H. L., Frost, A. R. & Grizzle, W. E. Fatty acid synthase: an early molecular marker of progression of prostatic adenocarcinoma to androgen independence. *J Urol.* **165,** 1027–1032, doi: S0022-5347(05)66596-2 (2001).
73. Suh, J. H. *et al.* Sterol regulatory element-binding protein-1c represses the transactivation of androgen receptor and androgen-dependent growth of prostatic cells. *Mol Cancer Res.* **6,** 314–324, doi: 10.1158/1541-7786.MCR-07-0354 (2008).
74. Hewitt, S. C. *et al.* Novel DNA motif binding activity observed *in vivo* with an estrogen receptor alpha mutant mouse. *Mol Endocrinol.* **28,** 899–911, doi: 10.1210/me.2014-1051 (2014).
75. Sasse, S. K. *et al.* The glucocorticoid receptor and KLF15 regulate gene expression dynamics and integrate signals through feed-forward circuitry. *Mol Cell Biol.* **33,** 2104–2115, doi: 10.1128/MCB.01474-12 (2013).
76. Asada, M. *et al.* DNA binding-dependent glucocorticoid receptor activity promotes adipogenesis via Kruppel-like factor 15 gene expression. *Lab Invest* **91,** 203–215, doi: 10.1038/labinvest.2010.170 (2011).
77. Kratz, C., Davignon, A., Chartrand, C. & Stanley, P. Simple d-transposition of the great arteries. Results of early balloon septotomy followed by two-stage surgical correction. *J Thorac Cardiovasc Surg.* **73,** 707–711 (1977).
78. Hu, S. *et al.* DNA methylation presents distinct binding sites for human transcription factors. *Elife* **2,** e00726, doi: 10.7554/eLife.00726 (2013).
79. Stadler, M. B. *et al.* DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature* **480,** 490–495, doi: 10.1038/nature10716 (2011).
80. Ziller, M. J. *et al.* Charting a dynamic DNA methylation landscape of the human genome. *Nature* **500,** 477–481, doi: 10.1038/nature12433 (2013).
81. Schubeler, D. Function and information content of DNA methylation. *Nature* **517,** 321–326, doi: 10.1038/nature14192 (2015).
82. Palacios, D., Summerbell, D., Rigby, P. W. & Boyes, J. Interplay between DNA methylation and transcription factor availability: implications for developmental activation of the mouse Myogenin gene. *Mol Cell Biol.* **30,** 3805–3815, doi: 10.1128/MCB.00050-10 (2010).
83. Schoenmakers, E. *et al.* Differential DNA binding by the androgen and glucocorticoid receptors involves the second Zn-finger and a C-terminal extension of the DNA-binding domains. *Biochem J.* **341** (Pt 3), 515–521 (1999).
84. Helsen, C. *et al.* Structural basis for nuclear hormone receptor DNA binding. *Mol Cell Endocrinol* **348,** 411–417, doi: 10.1016/j.mce.2011.07.025 (2012).
85. Lee, M. Y. *et al.* KLF5 enhances SREBP-1 action in androgen-dependent induction of fatty acid synthase in prostate cancer cells. *Biochem J* **417,** 313–322, doi: 10.1042/BJ20080762 (2009).
86. Kajita, Y. *et al.* The transcription factor Sp3 regulates the expression of a metastasis-related marker of sarcoma, actin filament-associated protein 1-like 1 (AFAP1L1). *PLoS One* **8,** e49709, doi: 10.1371/journal.pone.0049709 (2013).

Acknowledgements
F.V.F. is grateful for the support of grants CA154887 and CA176114 from the National Institutes of Health, National Cancer Institute. J.Q. is supported by grant CA154888 from the National Institutes of Health, National Cancer Institute. We are thankful to all members of the TCGA Research Network for biospecimen collection, and data acquisition. This work is supported by UC CRCC CRN-17-427258, GCR and HSRI program grants.

Author Contributions
F.V.F. and J.Q. designed the study and directed next generation sequencing. S.W. and F.V.F. conducted the ChIP-Seq data analysis. S.W., J.Q. and F.V.F. performed data interpretation, wrote the text, and reviewed the final manuscript.

Additional Information
Supplementary information accompanies this paper at http://www.nature.com/srep

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Wilson, S. *et al.* Refinement of the androgen response element based on ChIP-Seq in androgen-insensitive and androgen-responsive prostate cancer cell lines. *Sci. Rep.* **6**, 32611; doi: 10.1038/srep32611 (2016).

Chapter Five: The histone demethylase KDM3A regulates the transcriptional program of the androgen receptor in prostate cancer cells.

The lysine demethylase 3A (KDM3A) is a positive influence on the transcription of genes involved in the biological processes of spermatogenesis, metabolism, stem cell activity, and tumor progression by facilitating nucleosomes to exist in an open state by demethylating histone 3 lysine 9 residues (H3K9). This study matched transcriptomic and ChIPSeq profiles of KDM3A activity in prostate cancer cells. Combining histone KDM3A specific demethylation events with gene expression changes identified the transcriptional activation of gene targets involved in the cellular process of promoting androgen signaling. Matching ChIPSeq experiment of KDM3A in combination with ChIPSeq of the androgen receptor resulted in a gain of H3K9 methylation marks around androgen receptor binding sites of selected transcriptional targets in androgen signaling including positive regulation of KRT19, NKX3-1, KLK3, NDRG1, MAF, CREB3L4, MYC, INPP4B, PTK2B, MAPK1, MAP2K1, IGF1, E2F1, HSP90AA1, HIF1A, and ACSL3. This study also identified KDM3A's epigenetic regulation of gene targets involved in the hypoxia, glycolysis, and lipid metabolism processes of the cell. The approach of this study highlights how ChIPSeq and transcriptomic experiments can be applied to identify gene regulation of oncogenic pathways that contribute to cancer progression. Within this study I contributed to the computational analysis of the data supplied for figures 1-2, and 4.

www.impactjournals.com/oncotarget/ Oncotarget, 2017, Vol. 8, (No. 18), pp: 30328-30343

Research Paper

The histone demethylase *KDM3A* regulates the transcriptional program of the androgen receptor in prostate cancer cells

Stephen Wilson[1], Lingling Fan[2], Natasha Sahgal[3], Jianfei Qi[2], Fabian V. Filipp[1]

[1]Systems Biology and Cancer Metabolism, Program for Quantitative Systems Biology, University of California Merced, Merced, CA, USA

[2]Department of Biochemistry and Molecular Biology, Marlene and Stewart Greenebaum Comprehensive Cancer Center, University of Maryland School of Medicine, Baltimore, MD, USA

[3]Centre for Molecular Oncology, Barts Cancer Institute, Queen Mary University of London, London, United Kingdom

Correspondence to: *Fabian V. Filipp,* ***email:*** *filipp@ucmerced.edu*

Keywords: *cancer systems biology, epigenomics, ChIP-Seq, oncogene, prostate cancer*

Received: *July 18, 2016* ***Accepted:*** *September 09, 2016* ***Published:*** *March 03, 2017*

ABSTRACT

The lysine demethylase 3A (*KDM3A*, JMJD1A or JHDM2A) controls transcriptional networks in a variety of biological processes such as spermatogenesis, metabolism, stem cell activity, and tumor progression. We matched transcriptomic and ChIP-Seq profiles to decipher a genome-wide regulatory network of epigenetic control by KDM3A in prostate cancer cells. ChIP-Seq experiments monitoring histone 3 lysine 9 (H3K9) methylation marks show global histone demethylation effects of KDM3A. Combined assessment of histone demethylation events and gene expression changes presented major transcriptional activation suggesting that distinct oncogenic regulators may synergize with the epigenetic patterns by KDM3A. Pathway enrichment analysis of cells with *KDM3A* knockdown prioritized androgen signaling indicating that *KDM3A* plays a key role in regulating androgen receptor activity. Matched ChIP-Seq and knockdown experiments of *KDM3A* in combination with ChIP-Seq of the androgen receptor resulted in a gain of H3K9 methylation marks around androgen receptor binding sites of selected transcriptional targets in androgen signaling including positive regulation of *KRT19*, *NKX3-1*, *KLK3*, *NDRG1*, *MAF*, *CREB3L4*, *MYC*, *INPP4B*, *PTK2B*, *MAPK1*, *MAP2K1*, *IGF1*, *E2F1*, *HSP90AA1*, *HIF1A*, and *ACSL3*. The cancer systems biology analysis of KDM3A-dependent genes identifies an epigenetic and transcriptional network in androgen response, hypoxia, glycolysis, and lipid metabolism. Genome-wide ChIP-Seq data highlights specific gene targets and the ability of epigenetic master regulators to control oncogenic pathways and cancer progression.

INTRODUCTION

Methylation of histone lysine residues is a significant component of epigenetics and is associated with control of gene expression [1]. Specifically, methylation of lysine 9 of histone H3 (H3K9) has been recognized as hallmark of transcriptionally suppressed genes [2]. *KDM3A* (lysine demethylase 3A; Gene ID: 55818; also referred to as JMJD1A or JHDM2A) is crucial for gene regulation in a variety of biological activities such as spermatogenesis, metabolism, stem cell activity and tumor progression by demethylating mono- or di-methylated H3K9 [3–5]. Although the KDM3A protein regulates a wide array of target genes in tissue- and development-specific settings, chromatin modifiers often lack intrinsic DNA sequence specificity. Therefore, how KDM3A is targeted to specific genes is an area of current research interest and important for understanding epigenetic dysregulation in human disease.

KDM3A activity is deregulated in several cancers [3, 6–8]. In prostate adenocarcinoma (PRAD), KDM3A

functions as a transcriptional coactivator for the androgen receptor (*AR*; Gene ID: 367) [3, 9]. The ability to cooperate with the AR highlights a potential role of KDM3A as coactivator and driving force for sex-specific tissue development as well as for prostate cancer initiation and progression. In PRAD, androgen-dependent signaling plays a key role in the oncogenesis of prostate epithelial cells and the aggressiveness of the malignancy [10, 11]. The AR transcription factor belongs to the nuclear receptor superfamily and contains a C-terminal ligand-binding domain. Upon ligand binding, the AR undergoes a conformational change and dissociates from a cytosolic chaperone protein complex. Its ligand-bound conformation allows the AR to dimerize and to translocate into the nucleus [12]. Once in the nucleus, the activated AR dimer binds to androgen response elements (AREs) present in the promoter or enhancer of AR-regulated target genes and recruits co-activators or co-repressors to regulate gene expression [13]. In addition to the AR, KDM3A has been found to regulate expression and/or activity of several transcription factors such as *PPARG*, *KLF2*, *ESR1*, and *HOXA1* [14–17].

In order to further elucidate the impact of *KDM3A* on the epigenome, we performed chromatin immunoprecipitation in combination with next generation sequencing (ChIP-Seq) of its binding and demethylation activity. We quantified changes of H3K9me1 or H3K9me2 marks, the two substrates of KDM3A, and mapped AR-binding in the CWR22Rv1 prostate cancer cell line in combination with knockdown of *KDM3A*. Alteration of H3K9 methylation marks mapped to genomic locations coinciding with AR binding pinpoints target genes and oncogenic pathways cooperatively regulated by KDM3A and AR.

RESULTS

Genomic annotation and transcriptional regulation of KDM3A specific demethylase activity

Knockdown of *KDM3A* in CWR22Rv1 cells showed minor effects on global levels of H3K9me1 or H3K9me2 by Western blot analysis (Figure 1A), suggesting that KDM3A demethylates a small pool of methylated histone marks and regulates a specific set of gene targets. In order to establish the genome-wide impact of the epigenetic regulator KDM3A, we conducted a matched ChIP-Seq experiment using antibodies specific for histone marks H3K9me1 and H3K9me2 in combination with small hairpin RNA

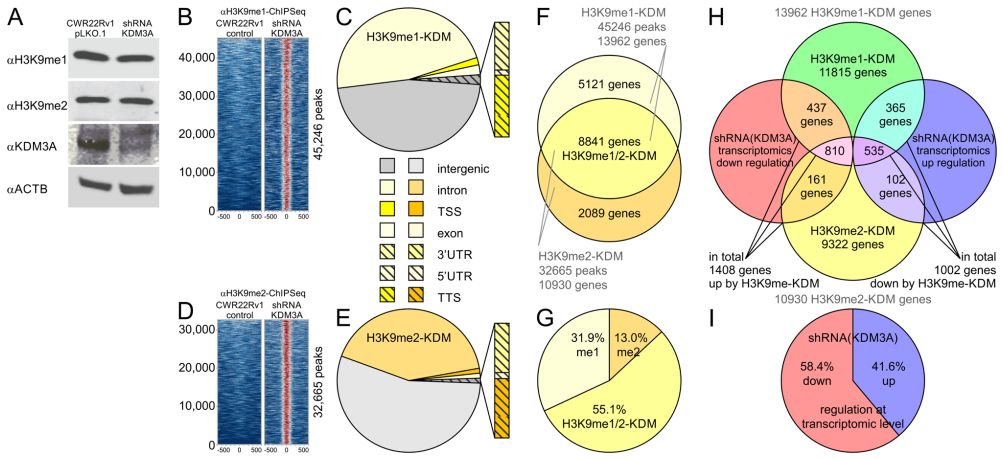


Figure 1: ChIP-Seq experiments with matched knockdown of *KDM3A* show gain of histone 3 lysine 9 methylation and transcriptional deactivation. A. Western blot of prostate cancer line CWR22Rv1 with antibodies against H3K9me1, H3K9me2, KDM3A (lysine demethylase 3A; Gene ID: 55818; also referred to as JMJD1A or JHDM2A), and beta-actin. ChIP-Seq experiments using **B-C.** H3K9me1-antibody and **D-E.** H3K9me2-antibody show specific gain of signal following small hairpin RNA (shRNA) knockdown of the histone methyltransferase *KDM3A* in the prostate cancer line CWR22Rv1. Genomic location of C) H3K9me1 and E) H3K9me2 sites identified by ChIP-Seq. **F.** H3K9me1 and H3K9me2 ChIP-Seq signals are overlapped and annotated. **G.** The majority of histone demethylase events due to KDM3A activity is detected by both, H3K9me1 and H3K9me2, ChIP-Seq antibodies. **H.** Overlay of gene mapping of histone methylation events identified by ChIP-Seq and transcriptomics experiments. Using this data we defined the group of 1408 genes as positively regulated by KDM3A activity (down in the prostate cancer line CWR22Rv1 with shRNA knockdown of *KDM3A*), and 1002 genes as negatively regulated by KDM3A activity. **I.** Transcriptomic impact of *KDM3A* knockdown shows 58.4% of gene activation (down in the prostate cancer line CWR22Rv1 with shRNA knockdown of *KDM3A*), and 41.6% of gene silencing.

(shRNA) knockdown of *KDM3A* in the CWR22Rv1 cell line (Supplementary Table 1). Histone lysine demethylation (KDM) events mediated by KDM3A were defined by gain of methylation ChIP-Seq signals following knockdown of *KDM3A* (Figure 1B–1E). Alterations of H3K9me1 and H3K9me2 histone marks upon *KDM3A* knockdown were evaluated in comparison to reference genomic DNA input or control non-coding shRNA samples. Overall, the peak counts of both H3K9me1 and H3K9me2 ChIP-Seq experiments showed a gain of signal (32244 to 34162 and 23353 to 46599, respectively). Since H3K9 methylation is a mark associated with the highly condensed heterochromatin state, we characterized the specific genomic regions associated with both H3K9 histone methylation marks. KDM events were functionally annotated by mapping bound regions to the human genome and by classifying them according to the nearest gene locus and relative position within coding regions. Promoter or transcription start sites (TSS) and transcription termination sites (TTS) genomic annotations are defined as being within ± 1000 bp of the gene-coding body. Intergenic regions were defined as the remaining regions outside the gene body. In the H3K9me1 ChIP-Seq experiment the intergenic regions were the most frequently found region with 21112 peaks (46.5%) followed by 21495 (47.7%) as intronic regions, 822 (1.8%) as exonic regions, 606 (1.3%) as promoter-TSS regions, 549 (1.2%) as TTS regions, 424 (0.9%) as 3'UTR regions, and 46 (0.1%) as 5' untranscribed (UTR) regions (Figure 1C). Similarly, the H3K9me2 ChIP-Seq had the intergenic region as its most frequent region with 18195 (55.9%) followed by 13167 (55.9%) as intronic regions, 373 (1.1%) as exonic regions, 355 (1.1%) as promoter-TSS regions, 246 (0.8%) as TTS regions, 204 (0.6%) as 3'UTR regions, and 25 (0.1%) as 5' UTR regions (Figure 1E). Taken together, ChIP-Seq profiles monitoring histone 3 lysine 9 methylation marks following *KDM3A* knockdown revealed selective histone demethylation effects of this epigenetic modifier.

Following genomic annotation, we were curious if there was a specific gene expression program underlying demethylation of H3K9me1 and H3K9me2. Half of the detected genes, 8841 (55.1%), contain both H3K9me1/2 marks (Figure 1F–1G). While the ChIP-Seq data shows that H3K9me1 has more annotated genes compared to H3K9me2 (5121 and 2089 respectively), both histone marks showed an equal fraction of genes being transcriptionally responsive to *KDM3A* knockdown according to the transcriptomic dataset. Overall, from the transcriptomic experiments, 1408 (58.4%; using a significance cutoff with adjusted p-values below 0.05) genes are reported as differentially down-regulated upon shRNA knockdown of *KDM3A* while 1002 genes are reported as up-regulated by *KDM3A* knockdown (Figure 1H–1I). The combination of ChIP-Seq histone demethylation events and transcriptomic assessment showed major transcriptional activation by KDM3A, suggesting that KDM3A may synergize with distinct transcriptional regulators for epigenetic control of gene expression.

Identification of an epigenetic and transcriptional network in androgen receptor signaling regulated by KDM3A

Following characterization of histone H3K9me1/2 marks we determined enrichment of transcriptional motifs associated with these histone marks controlled by KDM3A in prostate cancer cells. The goal of this analysis is to identify potential transcription factors that cooperate with KDM3A to regulate gene expression. Using the Jaspar motif database, we conducted an unbiased search for significant enrichment of transcription factor families (analysis of motif enrichment search with p-values below 0.05). Top hits included the androgen receptor, sterol regulatory element binding factor (SREBF), hypoxia inducible factor (HIF), activator protein 1 (AP1) complex of JUN/FOS, Krüppel-like factors (KLF), v-myc avian myelocytomatosis viral oncogene homolog (MYC), and forkhead box (FOX) families of transcription factors with significant enrichments and p-values below 1.0E-04 (Table 1). In addition, we enhanced simple ChIP-Seq-based searches with position-specific matrices to determine which transcription factor motifs were enriched compared to shuffled background sequences. The enrichment analysis showed the androgen response element (ARE) with 2915 incidences as one of the most frequent motifs detected (Table 1). Next, we analyzed significantly altered expression levels upon *KDM3A* shRNA knockdown (Figure 2A). At the transcriptional level, the androgen response gene set was the most enriched with a p-value and false discovery rate q-value each below 1.0E-20 (Figure 2B). Next, we sought inferred transcriptional regulators by comparing transcriptional targets to datasets that outline targets of transcription factors through the use of Ingenuity Pathway Analysis. The transcription factors AR, HIF, MYC, and AP1 complex were significantly enriched with p-values below 1.0E-07. Lastly, we merged ChIP-Seq profiles of H3K9Me1/2 and KDM3A transcriptional data focusing on 1408 annotated genes (overlap of H3K9ME1/2 ChIP-Seq with transcriptomic data that were down-regulated upon shRNA *KDM3A* knockdown) (Figure 1I, Supplementary Table 2). The data contained the highest enrichment ratio (26.7%) in a significantly enriched set of 27 genes in androgen signaling with p-values below 1.0E-17 and q-values below 1.0E-15 (Supplementary Table 3). In detail, putative KDM3A-regulated genes included pathways involved in androgen response, androgen receptor signaling, androgen biosynthesis, prostate cancer, pathways in cancer, cholesterol homeostasis, bile acid metabolism, aldosterone-regulated reabsorption, and progesterone regulation, hinting at the possibility of hormone nuclear steroid receptor involvement (enrichment with p-values below 2.62E-02 and q-values below 9.35E-02 correcting for multiple hypotheses testing) (Table 2). Interestingly for the concept of cooperative control, the pathways of

Table 1: Occurrence and enrichment by Fisher's exact test reveals enrichment of transcription factor motifs in KDM3A ChIP-Seq data

transcription factor	motif occurrences	adjusted p-values of motif enrichment H3K9me1 ChIP-Seq	adjusted p-values of motif enrichment H3K9me2 ChIP-Seq
TP53	3536	4.79E-10	5.58E-14
AR	2915	7.14E-07	2.00E-05
SREBF	1969	7.98E-50	8.24E-14
HIF	1964	1.88E-179	3.06E-142
FOS	1504	2.85E-256	2.60E-123
KLF	1244	4.06E-110	5.47E-81
MYC	767	1.68E-75	4.12E-63
FOX	584	0.00E-00	4.50E-276

Adjusted p-values accounting for multiple hypotheses testing using the false discovery rate controlling procedure of Benjamini and Hochberg of transcriptional motifs below 0.05 in H3K9me1/2-KDM3A ChIP-Seq data were considered significant.

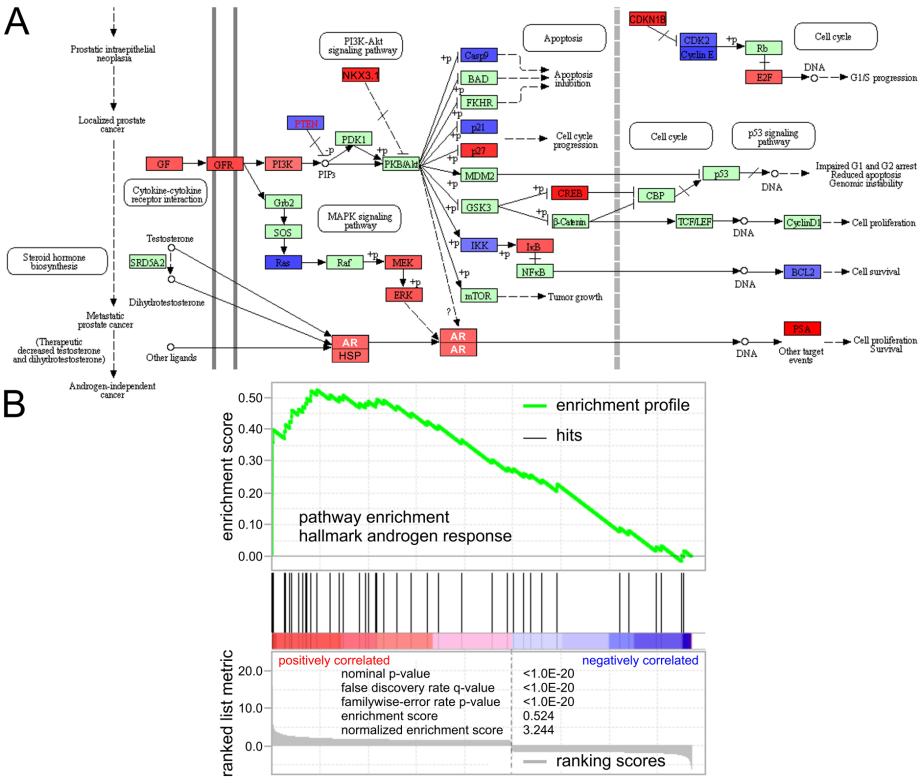


Figure 2: Knockdown of *KDM3A* results in epigenetic control and transcriptional activation of the androgen response. **A.** Map of transcriptional regulation of KDM3A on androgen signaling. Red indicates positive response of *KDM3A* on gene (down with shRNA knockdown of *KDM3A*); blue indicates negative response of *KDM3A* on gene (up with shRNA knockdown of *KDM3A*). **B.** Gene set enrichment analysis of ranked transcriptomic data upon shRNA knockdown of *KDM3A* indicated significant enrichment of hallmark gene set of androgen response with p-value and false discovery rate q-value below 1.0E-20.

Table 2: Enrichment of androgen-related signaling pathways in H3K9me1/2-KDM3A ChIP-Seq data

pathway	network	K	k	k/K	p-value	q-value
androgen response	hallmarks	101	27	0.267	2.77E-18	1.63E-16
pathways in cancer	kegg	328	36	0.110	4.49E-11	8.15E-10
androgen-mediated signaling	commons	130	21	0.162	1.58E-09	1.35E-07
regulation of androgen receptor activity	commons	108	17	0.157	8.25E-08	5.28E-06
cholesterol homeostasis	hallmarks	74	12	0.162	2.36E-06	1.69E-05
bile acid metabolism	hallmarks	112	14	0.125	8.35E-06	4.87E-05
aldosterone regulated sodium reabsorption	kegg	42	8	0.191	3.43E-05	1.42E-04
bladder cancer	kegg	42	8	0.191	3.43E-05	1.42E-04
coregulation of androgen receptor activity	commons	61	11	0.180	4.04E-06	1.97E-04
prostate cancer	kegg	89	11	0.124	8.43E-05	2.93E-04
colorectal cancer	kegg	62	9	0.145	1.05E-04	3.54E-04
androgen receptor signaling pathway	wiki	91	14	0.154	1.50E-06	8.35E-04
nongenotropic androgen signaling	commons	26	5	0.192	1.40E-03	1.58E-02
validated nuclear hormone receptor network	commons	65	7	0.108	5.20E-03	3.45E-02
progesterone-mediated signaling	kegg	86	8	0.093	7.10E-03	3.79E-02
bile acid secretion	kegg	71	7	0.099	8.40E-03	3.86E-02
bile acid biosynthesis	kegg	16	3	0.188	1.41E-02	4.68E-02
bile acid metabolism	commons	27	4	0.148	1.09E-02	5.33E-02
SREBF in cholesterol and lipid homeostasis	wiki	16	3	0.188	1.41E-02	7.96E-02
androgen biosynthesis	commons	8	2	0.250	2.62E-02	8.25E-02
cholesterol biosynthesis	wiki	18	3	0.167	1.96E-02	9.35E-02

The analysis is based on a query set of 1408 genes positively regulated by KDM3A with coincidence of demethylation and transcriptomic regulation (compare Figure 1). P-values below 0.05 from the hypergeometric distribution for K, the number of elements in the pathway of interest, k, the number of elements in the intersection of the input gene set, N=45956, all known human genes, and n=1408, the number of elements in the input gene set and false discovery rate q-values below 0.10 for multiple hypotheses testing according to the controlling procedure of Benjamini and Hochberg were considered significant.

regulation and coregulation of androgen receptor activity were also enriched with p-values and q-values below 4.04E-06 and 1.97E-04, respectively. *SLC26A2*, *FKBP5*, *KRT19*, *SORD*, *HOMER2*, *NDRG1*, *TPD52*, *INPP4B*, *PTPN21*, *ZMIZ1*, *PMEPA1*, *PPAP2A*, *TSC22D1*, *ACSL3*, *KLK3*, *NKX3-1*, *ELL2*, *MAP7*, *PTK2B*, *SMS*, *SPDEF*, *ABCC4*, *KLK2*, *MAF*, *TARP*, *AZGP1*, and *TMPRSS2* were key regulators of prostate cancer and AR signaling based on the KDM3A ChIP-Seq data (Figure 2A–2B). Transcriptional control of key players of cancer and AR signaling by KDM3A such as *KLK3*, *NKX3-1*, *MYC* were validated by chromatin immunoprecipitation coupled with quantitative real time polymerase chain reaction (ChIP-qRT-PCR) (Figure 3A–3B). Taken together, complementary analyses identified strong transcriptional networks including AR, MYC, FOX, KLF, AP1, and SREBF transcription factors that may be regulated by KDM3A. Androgen signaling was consistently identified

by all of these different enrichment approaches, suggesting a key role for KDM3A in regulating AR activity.

Matched KDM and AR ChIP-Seq experiments reveal coincidence of demethylase binding, demethylation and AR binding events

Knockdown of *KDM3A* in CWR22Rv1 cells resulted in loss of KDM3A ChIP-Seq binding accompanied by specific, matched gain of histone lysine 9 demethylation (Figure 4A–4C). Knockdown of *KDM3A* had little effect on the protein level of AR [3, 18]. We examined the alteration of AR binding by ChIP-Seq with an AR antibody following *KDM3A* knockdown and quantified the overlap of AR ChIP-Seq events with KDM3A binding and changes in epigenetic H3K9me1/2 marks (Figure 4A–4D). The activity-based ChIP-Seq array matched with knockdown of *KDM3A* resulted in 37525 peaks associated with KDM3A binding, 45246 and 32665 H3K9 mono- and di-demethylation (H3K9me1/2-KDM) events, respectively, and 34614 peaks for KDM3A-matched AR binding. Such an experimental design allows one to distinguish between histone demethylase binding (Figure 4E), epigenetic activity (Figure 4F), and coactivator binding events (Figure 4G). Gain of H3K9me1/2 was coupled to specific changes in AR binding in the *KDM3A* knockdown experiments (Figure 1A–1E). Overall 37.0% of the AR ChIP-Seq peaks with altered H3K9me1/2 signal were suppressed upon knockdown of *KDM3A*, while the remaining fraction was not affected. The genome-wide ChIP-Seq analysis is consistent with the biochemical data, demonstrating that KDM3A in effect recruited AR to target genes [3, 18]. KDM3A ChIP-Seq and H3K9me1/2-KDM ChIP-Seq in combination with matched knockdown of *KDM3A* produced an epigenetic network that overlaid with the AR ChIP-Seq data (Figure 4H–4J). In the case of matched and merged datasets of AR ChIP-Seq in combination with *KDM3A* knockdown, we identified in total 77911 H3K9me peaks (Figure 4F) and directly overlaid them with 34614 AR peaks containing 121700 ARE motifs (Figure 4I). Importantly, using such matched ChIP-Seq analyses, a set of 1912 genes was identified that showed an overlap of demethylation and AR binding events (KDM3A/AR ChIP, 2381 peaks, 1912 genes) (Figure 4I, Supplementary Table 1). Epigenetic events identified by ChIP-Seq were overlaid with transcriptomic data, defining a set of 421 genes that

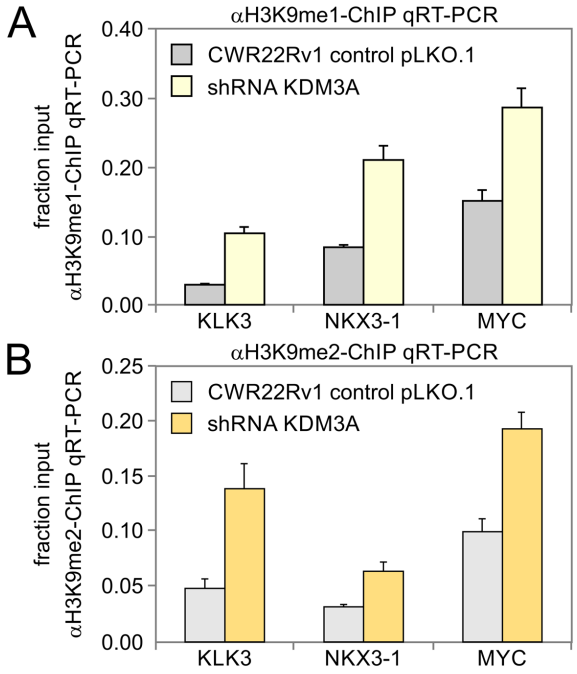


Figure 3: Detection of histone demethylase events by ChIP-qRT-PCR due to KDM3A activity in target genes involved in the androgen response. CWR22Rv1 prostate cancer cells transfected with pLKO.1 control or *KDM3A* shRNA were subjected to the chromatin immunoprecipitation coupled with quantitative real time polymerase chain reaction (ChIP-qRT-PCR) assay using immunoprecipitation with **A.** H3K9me1-antibody and **B.** H3K9me2-antibody. The precipitated chromatin fragments were analyzed by qRT-PCR using oligonucleotides for identified androgen response element regions of *KLK3*, *NKX3-1* or *MYC*.

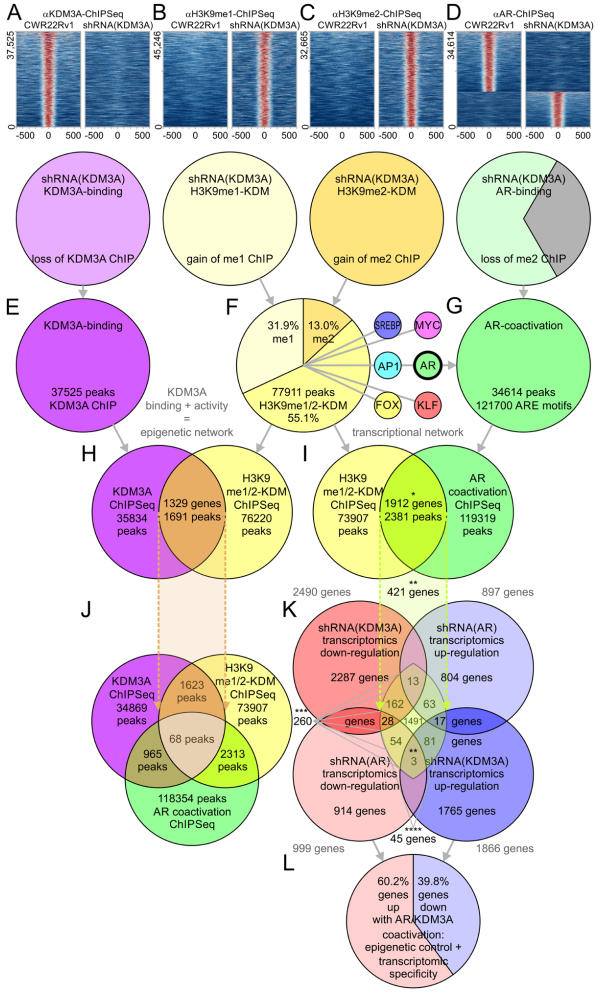


Figure 4: Matched ChIP-Seq and knockdown experiments of *KDM3A* in combination with ChIP of the androgen receptor show synergy of KDM3A and the androgen receptor. ChIP-Seq experiments in combination with *KDM3A* knockdown results in **A.** specific loss of KDM3A ChIP-Seq signal, **B.** specific gain of H3K9me1 ChIP-Seq signal for 97.1% of the observed histone marks, **C.** specific gain of H3K9me2 ChIP-Seq signal for 95.6% of the observed histone marks, and **D.** specific loss of ChIP-Seq signal for the androgen receptor (AR) with less than 2.3% retained binding. Peak calling utilizing a model-based analysis of ChIP-Seq algorithm results in **E.** 37525 peaks for KDM3A binding, **F.** 77911 peaks for H3K9 demethylation events (H3K9me1/2-KDM), and **G.** 34614 peaks for matched AR binding including 121700 androgen response elements (AREs). Motif analysis of KDM3A ChIP-Seq signals identifies pattern of transcription factor families including the AR motivating analysis of epigenetic and transcriptional cooperation between KDM3A and AR. **H.** KDM3A and H3K9 methylation ChIP-Seq signals are overlapped and annotated. **I.** H3K9 methylation and AR ChIP-Seq signals are overlapped and annotated. 1912 genes (gene set marked with *) showed coincidence of demethylation and AR binding events. **J.** Overlap of KDM3A and H3K9 methylation ChIP shows strong epigenetic network of KDM3A binding and activity. Overlap of KDM3A, H3K9 methylation, and matched AR ChIP shows participation of KDM3A in transcriptional activation. **K.** 421 genes (gene set marked with **) showed transcriptomic change upon *KDM3A*/*AR* knockdown in addition to coactivation detected by KDM/AR ChIP-Seq. 260 genes (gene set marked with ***) were positively regulated by KDM3A or AR activity (down in the prostate cancer line CWR22Rv1 with shRNA knockdown of *KDM3A*) and identified by KDM/AR ChIP-Seq. 45 genes genes (gene set marked with ****) showed overlap of transcriptomic response upon *KDM3A* and *AR* knockdown as well as coactivation detected by KDM/AR ChIP-Seq. **L.** Transcriptomic impact of *KDM3A* knockout shows 60.2% of gene activation (down in the prostate cancer line CWR22Rv1 with shRNA knockdown of *KDM3A*), and 39.8% of gene silencing.

had epigenetic marks (H3K9me1/2 ChIP), AR binding (AR ChIP), and a transcriptomic effect (differential expression in either *KDM3A* or *AR* knockdown experiments) (Figure 4K, Supplementary Table 4). Similar to the initial gene set based exclusively on H3K9me1/2 marks (57.2%) (Figure 1), 60.2% of genes scored as activated upon KDM3A/AR coactivation while 39.8% were silenced. Merged ChIP-Seq data of KDM3A/AR coactivation with transcriptomic data of *KDM3A* knockdown defined a set of 260 genes (whereas activation is interpreted as down-regulation upon shRNA knockdown) (Figure 4K–4L, Supplementary Table 4).

KDM3A and AR coactivation results in oncogenic pathway activation of AR signaling

H3K9 demethylation is known for stimulating gene expression [3, 5, 19, 20]. We used gene set enrichment analysis to identify signaling networks or functional clusters of genes controlled by both KDM3A and AR. At the gene level we studied pathway enrichment for the following three sets: A) 1912 genes defined by overlapping H3K9 demethylation/AR ChIP (Figure 4I, Supplementary Table 1-2); B) 421 genes defined by overlapping H3K9 demethylation/AR ChIP with differential expression upon *KDM3A*/*AR* knockdown (Figure 4K, Supplementary Table 2); and C) 260 genes that are down-regulated upon *KDM3A* or *AR* knockdown and having overlapping H3K9 demethylation/AR ChIP (Figure 4K, Supplementary Table 2). The gene sets were designed in a hierarchical fashion such that the parental set A of 1912 genes includes subset B of 421 genes, and B includes subset C of 260 genes (Supplementary Table 2). Such a hierarchical structure of gene sets allows one to monitor and identify requirements and conservation of a functional outcome. The gene set enrichment analysis revealed oncogenic activation of androgen signaling and metabolic pathways with p-values below 1.40E-02 and q-values below 4.44E-02 in hypoxic response, glycolysis, and lipogenic metabolism (Table 3, Supplementary Table 5). Enrichments of androgen response, metabolic pathways, hypoxia, aldosterone-regulated sodium reabsorption, glycolytic and glycerophospholipid metabolism pathways were conserved and enriched from the overlapping H3K9 demethylation/AR ChIP (1912 genes) to the genes displaying down-regulation upon *KDM3A*/*AR* knockdown and overlapping H3K9 demethylation/AR ChIP signal (260 genes). For example, androgen response and metabolic pathways showed incrementally higher enrichment with better defined input gene sets (significant enrichment of genes in androgen response with p-value of 7.51E-07 for 1912 gene set and decreased p-value of 5.23E-09 for 260 gene set). Similarly, pathways of glycolytic and glycerophospholipid metabolism showed consistently higher enrichment with the lowest p-values in the set of 260 genes, indicating that KDM3A demethylation targets these pathways and causes an up-regulation of gene expression. Of all genes detected by both H3K9 demethylation and AR ChIP-Seq experiments as well as *KDM3A* and *AR* differential expression following knockdown (core set of 45 genes; 28 genes up-regulated; 17 genes down-regulated), the AR response was the only significantly enriched pathway with a p-value of 5.29E-08. 7 genes, *NDRG1*, *PTK2B*, *ACSL3*, *KRT19*, *INPP4B*, *NKX3-1*, and *MAF*, with direct implication in AR signaling were present in all gene sets of the hierarchical enrichment analyses, thereby subject to KDM3A control by H3K9 demethylation, AR binding, and differential expression following knockdown (Supplementary Table 1-3, 5). In the androgen response pathway KDM3A and AR had strong ChIP-Seq activity (Table 4). The majority of epigenetic H3K9 demethylation events resulted in up-regulation of target gene activity and occurred at the TSSs of target genes. In addition, multiple introns, exons, and TTS of target genes were also implicated in KDM3A control. A network upstream regulator analysis implicated the AR to coordinate with epigenetic and transcriptional responses observed upon knockdown of *KDM3A* (Supplementary Table 6). At the transcriptional level, KDM3A may affect AR signaling directly by interacting with regulatory regions of *HSP90AA1* and *AR* genes. In addition, downstream effects of AR signaling were observed at *KRT19*, *NKX3-1*, *KLK3*, *TMPRSS2*, *PMEPA1*, *NDRG1*, *MAF*, *CREB3L4*, *MYC*, *INPP4B*, *PTK2B*, *MAPK1*, *MAP2K1*, *IGF1*, *E2F1*, *HIF1A*, *TARP*, *FKBP5*, *SPDEF*, *SMS*, *PPAP2A*, *SEPP1*, *UAP1*, *SORD*, *AZGP1*, *BCL2L1*, *ACSL3*, *CHUK*, and *CDKN1A* (Table 4).

DISCUSSION

Epigenetic regulators like KDM3A specifically demethylate histone marks H3K9me1 and H3K9me2, thereby playing a central role in the histone code. In cancer, demethylation and decondensation of chromatin can lead to dysregulated gene expression and transcriptional activation of gene targets [21]. Previous cell biological studies have suggested that KDM3A may stimulate transcription mediated by nuclear receptors and/or that KDM3A may be involved in activation of forkhead proteins during cell differentiation [17, 22]. While histone modifiers are not limited to specific DNA cognate sites, transcriptional specificity can be accomplished by cooperation with transcription factors recognizing distinct DNA motifs [13].

The chosen bioinformatics approach of quantifying changes of matched epigenetic remodeling in combination with transcription factor ChIP-Seq and transcriptomic analysis following knockdown of an epigenetic master regulator offers insight from two angles: it can identify in an unbiased way genome-wide cooperation of epigenetic remodeling with other members of the transcriptional machinery, and it can elucidate details of the interaction. In prostate cancer, steroid ligand dependency or independency can influence the prognostic outcome. The CWR22Rv1 cell line offers the benefit

Table 3: Pathway enrichment of transcriptional coactivation and matched ChIP-Seq experiments of H3K9me1/2 marks and the androgen receptor

pathway	network	1912 gene set (*)		421 gene set (**)		260 gene set (***)	
		p-value	q-value	p-value	q-value	p-value	q-value
androgen response	hallmarks	7.51E-07	1.27E-05	3.13E-07	2.46E-05	5.23E-09	1.23E-06
metabolic pathways	kegg	2.58E-06	1.55E-02	6.51E-09	6.00E-05	3.78E-07	1.80E-03
hypoxia	hallmarks	1.14E-11	8.99E-10	4.14E-10	9.77E-08	1.84E-06	2.17E-04
TNFα signaling via NFκB	hallmarks	1.96E-06	2.72E-05	1.34E-05	3.96E-04	1.28E-04	7.17E-03
estrogen response early	hallmarks	6.77E-09	2.66E-07	8.17E-05	1.48E-03	1.28E-04	7.17E-03
protein secretion	hallmarks	3.83E-07	8.22E-06	2.02E-07	2.38E-05	1.94E-04	7.63E-03
galactose metabolism	kegg	1.23E-03	5.79E-03	1.54E-03	1.65E-02	3.86E-04	1.30E-02
leukocyte transendothelial migration	kegg	1.04E-04	6.79E-04	6.66E-04	8.27E-03	5.04E-04	1.35E-02
adipocytokine signaling pathway	kegg	8.26E-06	8.49E-05	3.27E-04	5.15E-03	5.16E-04	1.35E-02
IL2 stat5 signaling	hallmarks	5.11E-07	9.27E-06	1.34E-05	3.96E-04	8.71E-04	1.47E-02
glycolysis	hallmarks	2.50E-04	1.52E-03	1.34E-05	3.96E-04	8.71E-04	1.47E-02
TP53 pathway	hallmarks	8.10E-05	5.62E-04	4.45E-04	5.84E-03	8.71E-04	1.47E-02
glycerophospholipid metabolism	kegg	7.61E-04	3.82E-03	6.49E-05	1.39E-03	8.73E-04	1.47E-02
aldosterone signaling	kegg	3.33E-04	1.92E-03	3.45E-05	8.15E-04	1.60E-03	2.51E-02
GNRH signaling pathway	kegg	5.01E-03	1.85E-02	2.10E-03	1.82E-02	2.38E-03	3.51E-02
endocytosis	kegg	2.74E-04	1.62E-03	1.30E-03	1.46E-02	3.50E-03	4.34E-02
glycolysis gluconeogenesis	kegg	1.40E-02	3.62E-02	2.26E-03	1.84E-02	4.85E-03	4.44E-02
interferon gamma response	hallmarks	2.09E-12	4.93E-10	1.34E-05	3.96E-04	5.09E-03	4.44E-02
estrogen response late	hallmarks	8.10E-05	5.62E-04	4.45E-04	5.84E-03	5.09E-03	4.44E-02
inflammatory response	hallmarks	2.00E-03	8.12E-03	4.45E-04	5.84E-03	5.09E-03	4.44E-02
heme metabolism	hallmarks	8.10E-05	5.62E-04	2.15E-03	1.82E-02	5.09E-03	4.44E-02

The analysis is focused on coactivation by KDM3A/AR ChIP and transcriptomic change upon knockdown. Input gene sets were defined by coactivation by KDM/AR ChIP-Seq (1912 genes; marked with *), coactivation by KDM/AR ChIP-Seq and transcriptomic change upon *KDM3A*/*AR* knockdown (421 genes; marked with **), oncogenic gene up-regulation by KDM3A/AR and KDM/AR ChIP coactivation (260 genes; marked with **) (compare Figure 4). Comprehensive lists of gene sets and enrichment ratios are deposited in the Supplementary Information. To determine significance of pathway enrichment thresholds of 0.05 and 0.10 were used for p-values and q-values, respectively according to multiple hypotheses testing using the controlling procedure of Benjamini and Hochberg.

of being able to probe a static, aggressive end point of the disease, while it has limitations due to lacking the dynamic ligand-dependent aspect of AR signaling. Since the CWR22Rv1 cell line expressed permanently activated AR without the ligand binding domain, it provides a stable model for studying the dynamic response to epigenetic regulation by *KDM3A* knockdown in combination with ChIP-Seq analysis. *KDM3A* knockdown abolished tumor formation in an orthotopic prostate tumor model using CWR22Rv1 cells [18]. Interestingly, *KDM3A* knockdown in other prostate cancer cell lines, including the androgen dependent LNCaP cells, blocked cell proliferation [18].

Androgen signaling is subject to multilevel control. In addition to agonist and antagonist ligand chaperones, intracellular localization and interactions with other transcription factors or histone modifiers can influence the transcriptional outcome. Multiple epigenetic regulators have been described to interact with the AR. KDM4C co-localizes with the androgen receptor in prostate carcinomas, and knockdown of *KDM4C* inhibits transcriptional activation and tumor cell proliferation [23]. KDM3A is involved in spermatogenesis by regulating expression of target genes such as *PRM1* and *TMP1*, which are required for packaging and condensation of

Table 4: Epigenetic profiles and transcription factor motifs in H3K9me1/2-KDM ChIP-Seq data of the androgen response pathway

Target gene	KDM TSS	KDM Exon	KDM Intron	KDM TTS	KDM3A regulation	AR regulation
KRT19	+	+	+	+	-4.5999	-0.7642
NKX3-1	+	-	-	+	-2.8073	-1.0951
KLK3	+	+	+	-	-1.3230	-0.5316
NDRG1	-	+	+	+	-1.2254	-
MAF	+	+	-	-	-1.1679	-
CREB3L4	+	+	+	+	-1.1671	-
MYC	+	+	+	+	-1.0253	-0.5261
INPP4B	+	+	+	+	-1.0252	-0.6637
PTK2B	-	+	+	-	-0.7523	-0.4538
MAPK1	+	-	+	+	-0.7441	-
MAP2K1	+	+	+	+	-0.7212	-
IGF1	+	+	+	+	-0.7062	-0.5278
E2F1	+	+	-	-	-0.6708	-
HSP90AA1	+	-	+	-	-0.6499	-
HIF1A	+	+	+	+	-0.6010	-
ACSL3	-	+	+	+	-0.5263	-0.8003
CHUK	+	-	+	-	0.4407	-
CDKN1A	+	-	+	-	0.7347	-

Transcriptomic validation of gene expression change of target gene following knockdown of *KDM3A* or *AR* is listed as LOG2 (fold change).

sperm chromatin [3]. Furthermore, its involvement in obesity resistance through regulation of metabolic genes such as *PPARA* and *UCP1* highlight a transcriptional network focused on lipid modifiers.

In this epigenetic and transcriptomic study we aimed to outline specific pathways of KDM3A demethylase action and enriched transcriptional networks under its control. Previous studies have outlined KDM3A expression levels in prostate cancer phenotypes [24], but KDM3A-regulated target pathways by ChIP-Seq analysis were unknown [24–26]. Motif-guided searches for cooperating transcription factors can link transcriptional programs with genome-wide histone modifications. Motifs of SREBP, HIF, AP1, KLF, MYC, and FOX families are enriched in the H3K9me1/2-KDM ChIP-Seq data and were described to play a role in prostate cancer progression [3, 9, 27–31]. However, of the enriched transcription factors characterized so far, only AR and MYC have strong biochemical links to KDM3A [3, 18]. Solely based on KDM3A regulation of ChIP-Seq and transcriptomic data, a gene set of 1408 genes revealed androgen-related signaling as top hit. Within the androgen response, transcriptionally validated genes, *NKX3-1* and *KLK3*, have been shown to have dynamically regulated histone modification states [32].

In a second step, matched ChIP-Seq studies of KDM3A and the AR allowed us to focus on a distinct transcriptional network. Aside from the androgen response, cellular metabolism is highly enriched in the executioner program of KDM3A and the AR. In particular, lipogenic and hypoxic metabolism stand out for their regulation in the combined KDM3A/AR ChIP-Seq and transcriptomic data. Previously established AR targets involved in cellular metabolism include *ACSL3*, which is involved in converting free long chain fatty acids into fatty acyl-CoA esters [33, 34]. Genome-wide ChIP-Seq profiles show that KDM3A regulates the expression of *FKBP5*, *CHUK*, *HSP90AA1*, and *VHL*. These genes were previously implicated as being regulated by the AR and their proteomic function is involved in AR folding, transactivation, and translocation in the nucleus [35–38]. *HSP90AA1* is described to be under the control of KDM3A [39–41]. Prominent shared targets of KDM3A and AR include gene targets associated with hypoxic metabolism. Transcriptional regulation in response to hypoxia is regulated by the actions of HIF1A and controls glycolytic

metabolism [42]. Previously, *KDM3A* gene expression was identified as one of the genes under control of HIF1A [43]. In addition, there is increased transcriptional activity of AR within castration-resistant prostate cancer by hypoxia [44]. A similar overlapping network of KDM3A demethylation, nuclear hormone signaling, and hypoxia is described in estrogen independent breast cancer models [16]. KDM3A's ability to regulate metabolic gene expression by controlling AR binding site availability in hypoxic cells may be the molecular action KDM3A utilizes to stimulate tumor progression [45]. The additional enriched pathways with shared KDM3A and AR regulation appear to the hypoxic cell response. Cytokine and metabolic signaling is able to induce expression of HIF1A and promote its activation in an oncogenic fashion [46–51].

The functional impact of coordinated action between a lysine demethylase and transcription factors may lend to its target specificity, or at the very least, create accessibility for DNA binding [52–56]. KDM3A controlled H3K9me1/2 ChIP-Seq data shows a strong enrichment of AR binding sites within the CWR22Rv1 castration-resistant prostate cancer cell line with continued expression of genes involved in the androgen response. However, about a third of the AR ChIP-Seq peaks were suppressed in response to increased H3K9me1/2 signal in the KDM3A experiments. It remains to be determined which other coactivators and corepressors take charge of AR binding sites in genomic locations where KDM3A has no effect on AR binding or transcriptional response. Notably, selected KDM3A- or AR-dependent genes show no clear association of modulated H3K9me1/2 marks within 50,000 bp around the gene body, suggesting that KDM3A can regulate gene expression either by accessing distant enhancers or by physical interaction with the transcriptional machinery independent of H3K9 demethylation. For example, KDM3A was identified to erase lysine 372 monomethylation of TP53, a protein methylation site crucial for the stability and pro-apoptotic function of chromatin-bound tumor suppressor [57]. Significantly, next to the AR, SREBF, and HIF, the conducted transcription factor motif analysis confirmed a strong presence of TP53 target sequences controlled by KDM3A. Extensive future ChIP-Seq studies of KDM3A as well as other candidate transcription factors associated with KDM3A will be necessary to further characterize the full spectrum of epigenetic and transcriptional control of the master regulator KDM3A.

In conclusion, the ChIP-Seq study refined the genomic sites of KDM3A-mediated H3K9me1/2 histone demethylation within the CWR22Rv1 prostate cancer cell line. The cancer systems biology analysis expanded underlying transcription factor motifs associated with oncogenic KDM3A demethylation, suggesting an underlying transcriptional network that directs transcriptional activation. Future experimental verification of epigenetic hotspots is needed to determine when detected response elements are functional in gene regulation. The matched transcriptomics and epigenomics approach identified an overlap between androgen receptor ChIP-Seq and KDM3A-regulated H3K9me1/2 ChIP-Seq. The comprehensive genome-wide mapping of matched ChIP-Seq profiles highlighted mechanistic details of how an epigenetic master regulator can exhibit control over selected transcriptional programs, such as metabolic pathways and hypoxia response in cancer.

MATERIALS AND METHODS

Tissue culture of prostate cancer cell lines

CWR22Rv1 is a human prostate carcinoma epithelial cell line derived from a xenograft that was serially propagated in mice after castration-induced regression and relapse of the parental, androgen-dependent CWR22 xenograft [58, 59] (CRL-2505, American Type Culture Collection, Manassas, VA). The CWR22Rv1 prostate cancer cell line was kindly provided by Dr. James Jacobberger (Case Western Reserve University, Cleveland, OH), and are maintained in RPMI 1640 medium supplemented with 10% fetal bovine serum and antibiotics. Cell cultures are regularly tested to ensure absence of mycoplasma. The CWR22Rv1 prostate cancer cell line expresses AR full-length with a duplicated DNA binding domain in exon 3 and AR splice variants, for example AR-v7, lacking a ligand binding domain. Thus, the CWR22Rv1 cell line displays constitutively active AR even in the absence of androgen [59]. In the ChIP-Seq study an AR antibody (PG21, 06-680, Sigma EMD Millipore, Darmstadt, Germany) was used recognizing both forms of the AR. All experimental protocols were approved by the Institutional Review Board at the University of California Merced. The study was carried out as part of IRB UCM13-0025 of the University of California Merced and as part of dbGap ID 5094 on somatic mutations in cancer and conducted in accordance with the Helsinki Declaration of 1975.

Knockdown of KDM3A with shRNA

Lentiviral vectors encoding *KDM3A* small hairpin RNA (shRNA), *AR* shRNA or lentiviral pLKO.1 control shRNA were purchased from Open Biosystems (GE Healthcare Dharmacon, Lafayette, CO), and packaged in human embryonic kidney 293T cells (CRL-3216, American Type Culture Collection, Manassas, VA) by calcium phosphate transfection. The supernatant containing lentiviral particles were collected 48 hours after transfection. CWR22Rv1 cells were transduced with the supernatant of lentiviral particles in the presence of polybrene (8 μg/ml) for 24 hours before replacement with the fresh growth media. Cells were analyzed at 48 hours post-transduction. The knockdown efficiency was confirmed by quantitative real time polymerase chain reaction (qRT-PCR) and Western-blot analysis.

qRT-PCR analysis

Total RNA from prostate cancer cells was extracted using a mammalian RNA mini preparation kit (RTN10-1KT, GenElute, Sigma EMD Millipore, Darmstadt, Germany) and then digested with deoxyribonuclease I (AMPD1-1KT, Sigma EMD Millipore, Darmstadt, Germany). Complementary DNA (cDNA) was synthesized using random hexamers. Triple replicate samples were subjected to SYBR green (SYBR green master mix, Qiagen SABiosciences) qRT-PCR analysis in an Eco system (Illumina, San Diego). Gene expression profiles were analyzed using the ΔΔCT method. qRT-PCR threshold cycle (CT) values were normalized using the housekeeping gene cyclophilin A (*PPIA*; peptidylprolyl isomerase A; Gene ID: 5478). The following primers served for ChIP-qRT-PCR validation of ChIP-Seq signal of H3K9me1/2-KDM around the AREs of AR target genes: *KLK3* (kallikrein related peptidase 3; Gene ID: 354; also known as PSA, prostate specific antigen): 5'-TGGGACAACTTGCAAACCTG-3'; 5'-CCAGAGTA GGTCTGTTTTCAATCCA-3'; *NKX3.1* (NK3 homeobox 1; Gene ID: 4824): 5'-GGTTCTGCTGTTACGTTTG-3'; 5'-CTTGCTTGCTCAGTGGAC-3'; *MYC* (v-myc avian myelocytomatosis viral oncogene homolog; Gene ID: 4609): 5'-CCAGCGAATTATTCAGAA-3'; 5'-AATTACCATTGACTTCCTC-3'.

Western-blot analysis

Whole cell lysates were harvested using radio-immunoprecipitation assay (RIPA) buffer composed of 50 mM trisaminomethane hydrochloride (Tris-HCl) pH7.5, 150 mM sodium chloride (NaCl), 1% Triton X-100, 0.1% sodium dodecyl sulfate (SDS), 0.1% sodium deoxycholate, 1.0 mM EDTA, 1.0 mM sodium orthovanadate, and 1x protease inhibitor cocktail. Lysates were subjected to sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and proteins transferred to a nitrocellulose membrane (GE Healthcare Life Sciences, Pittsburgh, PA). The membrane was probed with ChIP-grade H3K9me1 (ab9045, Abcam, Cambridge, MA), H3K9me2 (07-441, Sigma EMD Millipore, Darmstadt, Germany), KDM3A (A301-539A, Bethyl Laboratories, Montgomery, TX), or ACTB (A5441, Sigma EMD Millipore, Darmstadt, Germany) antibodies followed by a secondary antibody conjugated to fluorescent dye, and blots were imaged using the odyssey detecting system (LI-COR Biotechnology, Bad Homburg, Germany).

Chromatin immunoprecipitation

Cells were crosslinked using 1% formaldehyde for 10 min at 298 K. Formaldehyde was removed and cells were incubated with 125 mM glycine for 5 min at 298 K. Nuclear extracts were collected and sonicated to obtain 300 bp chromatin fragments using the Covaris S2 ultrasonicator (Covaris, Woburn, MA). 100 μg of chromatin was incubated with 5 μg of AR (PG21, 06-680, Sigma EMD Millipore, Darmstadt, Germany), 5 μg of KDM3A (A301-538A, Bethyl Laboratories, Montgomery, TX), 2 μg of H3K9me1 (ab9045, Abcam, Cambridge, MA), or 2 μg of H3K9me2 (07-441, Sigma EMD Millipore, Darmstadt, Germany) antibodies overnight at 277 K followed by incubation with 30 μl of protein A/G beads for 4 hours. After four washes, crosslinking was reversed. Chromatin was digested with ribonuclease A followed by proteinase K. Then the DNA was purified using spin columns. The size of the DNA was confirmed by a bioanalyzer (Agilent Biotechnologies, Savage, MD).

Next generation sequencing and ChIP-Seq analysis

The purified DNA library was sequenced using an Illumina HighSeq2000 at the Sanford-Burnham Medical Research Institute at Lake Nona, National Genome Library Core Facility. Sequenced regions were aligned to the reference human genome 19 using the Bowtie alignment program that utilizes an extended Burrows-Wheeler indexing for an ultrafast memory efficient alignment [60]. Peak calling utilized a model-based analysis of ChIP-Seq (MACS) algorithm [61, 62]. The overlap analysis, plot of genomic location, sequence extraction, motif identification, and peak filtering were performed using ChIPseek: a web-based analysis for ChIP data [63]. ChIPseek also employs scripts from BEDtools [64] using a genome binning algorithm used by the UCSC genome browser to sort genomic regions into groups along the length of chromosome [65]. Data visualization was carried out using the integrative genomics viewer [66].

Motif analysis based on position site-specific matrix models

Computational response element searching algorithms are able to estimate a sequence's likelihood in belonging to the response element of the query transcription factor using position site-specific matrix (PSSM) models where each position in the query transcription factor model gives each of the four letters in the DNA alphabet a score based on the probability of that nucleotide being found at that position [67]. Summation into a logs-odd score is converted into a p-value assuming a zero-order background model, and all response elements less than the threshold are reported [68]. Motif discovery, motif enrichment, and motif scanning used the multiple expectation maximization for motif elicitation (MEME) and discriminative regular expression motif elicitation (DREME) suite software toolkits from a set of user supplied unaligned sequences for ChIP-Seq regions [69]. *De novo* motif analysis programs MEME and DREME identify frequently detected DNA sequences patterns and similarity matches of recurring ChIP-Seq sequences with DNA motifs

of deposited studies in genomic sequence databases [68, 70]. After a motif of interest is discovered the genomic sequences of the ChIP sequenced data is scanned using the MEME suite software find individual motif occurrences (FIMO) [68] for individual motif occurrences using PSSMs to compute a log-likelihood ratio score for each submitted sequence. Sequence-specific matrix models are further used to analyze the next generation sequencing data for motif enrichment and potential coactivators [13, 71].

Microarray analysis

In order to quantify the transcriptomic effect of *KDM3A or AR* knockdown, a microarray profiling analysis was conducted on CWR22Rv1 knockdown cells. CWR22v1 cells were transduced with lentiviral pLKO.1 control shRNA vector, *AR*, or *KDM3A* shRNA for 48 h. Total RNA was isolated from cells, and 500 ng was used for synthesis of biotin-labeled cRNA using an RNA amplification kit (Ambion, Thermo Fisher Scientific, Waltham, MA). Biotinylated cRNA was labeled by incubation with streptavidin-Cy3 to generate a probe for hybridization with the GeneChip Human Transcriptome Array 2.0 (Affymetrix Inc, Santa Clara, CA) or the genome-wide transcriptome Human HT-12 V4.0 (Illumina Inc, San Diego, CA). Four samples from two experimental groups (n=2 per group) were hybridized to the chip to obtain raw gene expression data, which was processed to obtain raw data in the form of expression intensities. Raw data was then exported for further processing and analysis using R statistical software version 3.1 in combination with the BioConductor oligo, affy and genefilter packages [72]. The raw signal intensities were background corrected, LOG2 transformed, and quantile normalized to generate robust multi-array average (RMA) normalized intensities [73]. Quality control analyses did not reveal any outlier samples. Differential expression between experimental groups was assessed by generating relevant contrasts corresponding to the two-group comparison and was evaluated using the Linear Models for Microarray Analysis (LIMMA) package [72, 74]. Raw p-values were corrected for multiple hypotheses testing using the false discovery rate controlling procedure of Benjamini and Hochberg, and adjusted p-values below 0.05 were considered significant [75]. Genes with significantly altered expression levels with adjusted p-values below 0.05 following *KDM3A* knockdown were selected and analyzed through the use of Ingenuity Pathway Analysis (IPA, Qiagen, Redwood City, CA). Pathway enrichment of differentially expressed genes was determined by gene set enrichment analysis (GSEA) by pairing each gene with its expression value and ranking genes in descending order (Broad Institute, Cambridge, MA) [76, 77]. As we are testing multiple gene sets at the same time, the p-values need to be controlled for false positives. The false discovery rate estimation for the pathway enrichment is as summarized in q-values with a threshold of 0.10 controlling the error rate and correcting for multiple hypotheses testing according to Benjamini and Hochberg [75]. Acquired data of transcriptome profiling microarray analysis of CWR22Rv1 cells with *AR* knockdown using GeneChip Human Transcriptome Array 2.0, platform GPL16686 (Affymetrix Inc, Santa Clara, CA), is deposited in the NCBI GEO database under accession number GSE86547. Acquired data of CWR22Rv1 cells with *KDM3A* knockdown using hybridization with genome-wide transcriptome Human HT-12 V4.0, platform GPL10558 (Illumina Inc, San Diego, CA), is deposited under accession number GSE70498.

Availability of supporting data

The Supplementary Information contains tables on genome-wide mapping, annotation, and overlap of H3K9me1/me2 demethylation ChIP-Seq and AR ChIP-Seq (Supplementary Table 1), gene sets based on ChIP-Seq and transcriptomic data (Supplementary Table 2), pathway enrichment analysis based on H3K9me1/me2 demethylation ChIP-Seq gene set (Supplementary Table 3), transcriptomic response upon shRNA(*KDM3A*) and shRNA(*AR*) knockdown (Supplementary Table 4), hierarchical gene set enrichment analysis of identified KDM3A target genes (Supplementary Table 5), and an upstream regulator analysis based on ingenuity pathway analysis (Supplementary Table 6).

Author contributions

F.V.F. and J.Q. designed the study and directed next generation sequencing. S.W. and F.V.F. conducted the ChIP-Seq data analysis. L.F. and N.S. conducted transcriptomic validation of ChIP-Seq targets. S.W., J.Q., and F.V.F. performed data interpretation, wrote the text, and reviewed the final manuscript.

ACKNOWLEDGMENTS

F.V.F. is grateful for the support of grants CA154887 and CA176114 from the National Institutes of Health, National Cancer Institute. J.Q. is supported by grants CA154888 and CA207118 from the National Institutes of Health, National Cancer Institute. This work is supported by UC CRCC CRN-17-427258, GCR, and HSRI program grants.

CONFLICTS OF INTEREST

The authors declare that there is no competing interest as part of the submission process of this manuscript.

REFERENCES

1. Gonda TJ, Ramsay RG. Directly targeting transcriptional dysregulation in cancer. Nat Rev Cancer. 2015; 15:686–94.
2. Kooistra SM, Helin K. Molecular mechanisms and potential functions of histone demethylases. Nat Rev Mol Cell Biol. 2012; 13:297–311.
3. Yamane K, Toumazou C, Tsukada Y, Erdjument-Bromage H, Tempst P, Wong J, Zhang Y. JHDM2A, a JmjC-containing H3K9 demethylase, facilitates transcription activation by androgen receptor. Cell. 2006; 125:483–95.
4. Okada Y, Scott G, Ray MK, Mishina Y, Zhang Y. Histone demethylase JHDM2A is critical for Tnp1 and Prm1 transcription and spermatogenesis. Nature. 2007; 450:119–23.
5. Kuroki S, Matoba S, Akiyoshi M, Matsumura Y, Miyachi H, Mise N, Abe K, Ogura A, Wilhelm D, Koopman P, Nozaki M, Kanai Y, Shinkai Y, et al. Epigenetic regulation of mouse sex determination by the histone demethylase Jmjd1a. Science. 2013; 341:1106–09.
6. Krieg AJ, Rankin EB, Chan D, Razorenova O, Fernandez S, Giaccia AJ. Regulation of the histone demethylase JMJD1A by hypoxia-inducible factor 1 alpha enhances hypoxic gene expression and tumor growth. Mol Cell Biol. 2010; 30:344–53.
7. Uemura M, Yamamoto H, Takemasa I, Mimori K, Hemmi H, Mizushima T, Ikeda M, Sekimoto M, Matsuura N, Doki Y, Mori M. Jumonji domain containing 1A is a novel prognostic marker for colorectal cancer: *in vivo* identification from hypoxic tumor cells. Clin Cancer Res. 2010; 16:4636–46.
8. Yamada D, Kobayashi S, Yamamoto H, Tomimaru Y, Noda T, Uemura M, Wada H, Marubashi S, Eguchi H, Tanemura M, Doki Y, Mori M, Nagano H. Role of the hypoxia-related gene, JMJD1A, in hepatocellular carcinoma: clinical impact on recurrence after hepatic resection. Ann Surg Oncol. 2012; 19:S355-64.
9. Cleys ER, Halleran JL, Enriquez VA, da Silveira JC, West RC, Winger QA, Anthony RV, Bruemmer JE, Clay CM, Bouma GJ. Androgen receptor and histone lysine demethylases in ovine placenta. PLoS One. 2015; 10:e0117472.
10. Cao B, Qi Y, Zhang G, Xu D, Zhan Y, Alvarez X, Guo Z, Fu X, Plymate SR, Sartor O, Zhang H, Dong Y. Androgen receptor splice variants activating the full-length receptor in mediating resistance to androgen-directed therapy. Oncotarget. 2014; 5:1646–56. doi: 10.18632/oncotarget.1802.
11. Attard G, Cooper CS, de Bono JS. Steroid hormone receptors in prostate cancer: a hard habit to break? Cancer Cell. 2009; 16:458–62.
12. Bain DL, Heneghan AF, Connaghan-Jones KD, Miura MT. Nuclear receptor structure: implications for function. Annu Rev Physiol. 2007; 69:201–20.
13. Wilson S, Qi J, Filipp FV. Refinement of the androgen response element based on ChIP-Seq in androgen-insensitive and androgen-responsive prostate cancer cell lines. Sci Rep. 2016; 6:32611.
14. Jiang Y, Wang S, Zhao Y, Lin C, Zhong F, Jin L, He F, Wang H. Histone H3K9 demethylase JMJD1A modulates hepatic stellate cells activation and liver fibrosis by epigenetically regulating peroxisome proliferator-activated receptor γ. FASEB J. 2015; 29:1830–41.
15. Ohguchi H, Hideshima T, Bhasin MK, Gorgun GT, Santo L, Cea M, Samur MK, Mimura N, Suzuki R, Tai YT, Carrasco RD, Raje N, Richardson PG, et al. The KDM3A-KLF2-IRF4 axis maintains myeloma cell survival. Nat Commun. 2016; 7:10258.
16. Wade MA, Jones D, Wilson L, Stockley J, Coffey K, Robson CN, Gaughan L. The histone demethylase enzyme KDM3A is a key estrogen receptor regulator in breast cancer. Nucleic Acids Res. 2015; 43:196–207.
17. Mahajan K, Lawrence HR, Lawrence NJ, Mahajan NP. ACK1 tyrosine kinase interacts with histone demethylase KDM3A to regulate the mammary tumor oncogene HOXA1. J Biol Chem. 2014; 289:28179–91.
18. Fan L, Peng G, Sahgal N, Fazli L, Gleave M, Zhang Y, Hussain A, Qi J. Regulation of c-Myc expression by the histone demethylase JMJD1A is essential for prostate cancer cell growth and survival. Oncogene. 2016; 35:2441–52.
19. Goda S, Isagawa T, Chikaoka Y, Kawamura T, Aburatani H. Control of histone H3 lysine 9 (H3K9) methylation state via cooperative two-step demethylation by Jumonji domain containing 1A (JMJD1A) homodimer. J Biol Chem. 2013; 288:36948–56.
20. Tateishi K, Okada Y, Kallin EM, Zhang Y. Role of Jhdm2a in regulating metabolic gene expression and obesity resistance. Nature. 2009; 458:757–61.
21. Filipp FV. Crosstalk between epigenetics and metabolism—Yin and Yang of histone demethylases and methyltransferases in cancer. Brief Funct Genomics. 2017 Mar 24. doi: 10.1093/bfgp/elx001. [Epub ahead of print]
22. Herzog M, Josseaux E, Dedeurwaerder S, Calonne E, Volkmar M, Fuks F. The histone demethylase Kdm3a is essential to progression through differentiation. Nucleic Acids Res. 2012; 40:7219–32.
23. Wissmann M, Yin N, Müller JM, Greschik H, Fodor BD, Jenuwein T, Vogler C, Schneider R, Günther T, Buettner R, Metzger E, Schüle R. Cooperative demethylation by JMJD2C and LSD1 promotes androgen receptor-dependent gene expression. Nat Cell Biol. 2007; 9:347–53.
24. Suikki HE, Kujala PM, Tammela TL, van Weerden WM, Vessella RL, Visakorpi T. Genetic alterations and changes in expression of histone demethylases in prostate cancer. Prostate. 2010; 70:889–98.
25. Björkman M, Östling P, Härmä V, Virtanen J, Mpindi JP, Rantala J, Mirtti T, Vesterinen T, Lundin M, Sankila A, Rannikko A, Kaivanto E, Kohonen P, et al. Systematic knockdown of epigenetic enzymes identifies a novel histone

demethylase PHF8 overexpressed in prostate cancer with an impact on cell proliferation, migration and invasion. Oncogene. 2012; 31:3444–56.

26. Sowalsky AG, Xia Z, Wang L, Zhao H, Chen S, Bubley GJ, Balk SP, Li W. Whole transcriptome sequencing reveals extensive unspliced mRNA in metastatic castration-resistant prostate cancer. Mol Cancer Res. 2015; 13:98–106.
27. Huang WC, Li X, Liu J, Lin J, Chung LW. Activation of androgen receptor, lipogenesis, and oxidative stress converged by SREBP-1 is responsible for regulating growth and progression of prostate cancer cells. Mol Cancer Res. 2012; 10:133–42.
28. Shankar E, Song K, Corum SL, Bane KL, Wang H, Kao HY, Danielpour D. A Signaling Network Controlling Androgenic Repression of c-Fos Protein in Prostate Adenocarcinoma Cells. J Biol Chem. 2016; 291:5512–26.
29. Oktem G, Bilir A, Uslu R, Inan SV, Demiray SB, Atmaca H, Ayla S, Sercan O, Uysal A. Expression profiling of stem cell signaling alters with spheroid formation in CD133(high)/CD44(high) prostate cancer stem cells. Oncol Lett. 2014; 7:2103–09.
30. Sahu B, Laakso M, Ovaska K, Mirtti T, Lundin J, Rannikko A, Sankila A, Turunen JP, Lundin M, Konsti J, Vesterinen T, Nordling S, Kallioniemi O, et al. Dual role of FoxA1 in androgen receptor binding to chromatin, androgen signalling and prostate cancer. EMBO J. 2011; 30:3962–76.
31. Gurel B, Iwata T, Koh CM, Jenkins RB, Lan F, Van Dang C, Hicks JL, Morgan J, Cornish TC, Sutcliffe S, Isaacs WB, Luo J, De Marzo AM. Nuclear MYC protein overexpression is an early alteration in human prostate carcinogenesis. Mod Pathol. 2008; 21:1156–67.
32. Urbanucci A, Marttila S, Jänne OA, Visakorpi T. Androgen receptor overexpression alters binding dynamics of the receptor to chromatin and chromatin structure. Prostate. 2012; 72:1223–32.
33. Marques RB, Dits NF, Erkens-Schulze S, van Ijcken WF, van Weerden WM, Jenster G. Modulation of androgen receptor signaling in hormonal therapy-resistant prostate cancer cell lines. PLoS One. 2011; 6:e23144.
34. Qiao S, Tuohimaa P. The role of long-chain fatty-acid-CoA ligase 3 in vitamin D3 and androgen control of prostate cancer LNCaP cell growth. Biochem Biophys Res Commun. 2004; 319:358–68.
35. Shaw GL, Whitaker H, Corcoran M, Dunning MJ, Luxton H, Kay J, Massie CE, Miller JL, Lamb AD, Ross-Adams H, Russell R, Nelson AW, Eldridge MD, et al. The Early Effects of Rapid Androgen Deprivation on Human Prostate Cancer. Eur Urol. 2016; 70:214–18.
36. Echeverria PC, Picard D. Molecular chaperones, essential partners of steroid hormone receptors for activity and mobility. Biochim Biophys Acta. 2010; 1803:641-9.
37. Georget V, Térouanne B, Nicolas JC, Sultan C. Mechanism of antiandrogen action: key role of hsp90 in conformational change and transcriptional activity of the androgen receptor. Biochemistry. 2002; 41:11824–31.
38. Chen S, Chen K, Zhang Q, Cheng H, Zhou R. Regulation of the transcriptional activation of the androgen receptor by the UXT-binding protein VHL. Biochem J. 2013; 456:55–66.
39. Cheng MB, Zhang Y, Cao CY, Zhang WL, Zhang Y, Shen YF. Specific phosphorylation of histone demethylase KDM3A determines target gene expression in response to heat shock. PLoS Biol. 2014; 12:e1002026.
40. Kasioulis I, Syred HM, Tate P, Finch A, Shaw J, Seawright A, Fuszard M, Botting CH, Shirran S, Adams IR, Jackson IJ, van Heyningen V, Yeyati PL. Kdm3a lysine demethylase is an Hsp90 client required for cytoskeletal rearrangements during spermatogenesis. Mol Biol Cell. 2014; 25:1216–33.
41. Jain G, Voogdt C, Tobias A, Spindler KD, Möller P, Cronauer MV, Marienfeld RB. IκB kinases modulate the activity of the androgen receptor in prostate carcinoma cell lines. Neoplasia. 2012; 14:178–89.
42. Rankin EB, Giaccia AJ. Hypoxic control of metastasis. Science. 2016; 352:175–80.
43. Pollard PJ, Loenarz C, Mole DR, McDonough MA, Gleadle JM, Schofield CJ, Ratcliffe PJ. Regulation of Jumonji-domain-containing histone demethylases by hypoxia-inducible factor (HIF)-1alpha. Biochem J. 2008; 416:387–94.
44. Lunardi P, Beauval JB, Roumiguié M, Soulié M, Cuvillier O, Malavaud B. [Mechanisms of castration resistance: intratumoral hypoxia stimulates the androgen receptor expression]. [Article in French]. Prog Urol. 2016; 26:159–67.
45. Mimura I, Tanaka T, Wada Y, Kodama T, Nangaku M. Pathophysiological response to hypoxia - from the molecular mechanisms of malady to drug discovery: epigenetic regulation of the hypoxic response via hypoxia-inducible factor and histone modifying enzymes. J Pharmacol Sci. 2011; 115:453–58.
46. Filipp FV. Cancer metabolism meets systems biology: pyruvate kinase isoform PKM2 is a metabolic master regulator. J Carcinog. 2013; 12:14.
47. Filipp FV, Scott DA, Ronai ZA, Osterman AL, Smith JW. Reverse TCA cycle flux through isocitrate dehydrogenases 1 and 2 is required for lipogenesis in hypoxic melanoma cells. Pigment Cell Melanoma Res. 2012; 25:375–83.
48. Filipp FV, Ratnikov B, De Ingeniis J, Smith JW, Osterman AL, Scott DA. Glutamine-fueled mitochondrial metabolism is decoupled from glycolysis in melanoma. Pigment Cell Melanoma Res. 2012; 25:732–39.
49. Scharte M, Han X, Bertges DJ, Fink MP, Delude RL. Cytokines induce HIF-1 DNA binding and the expression of HIF-1-dependent genes in cultured rat enterocytes. Am J Physiol Gastrointest Liver Physiol. 2003; 284:G373–84.
50. Jung YJ, Isaacs JS, Lee S, Trepel J, Neckers L. IL-1beta-mediated up-regulation of HIF-1alpha via an NFkappaB/COX-2 pathway identifies HIF-1 as a critical link

between inflammation and oncogenesis. FASEB J. 2003; 17:2115–17.

51. Sharma V, Dixit D, Koul N, Mehta VS, Sen E. Ras regulates interleukin-1β-induced HIF-1α transcriptional activity in glioblastoma. J Mol Med (Berl). 2011; 89:123–36.
52. Visakorpi T, Hyytinen E, Koivisto P, Tanner M, Keinänen R, Palmberg C, Palotie A, Tammela T, Isola J, Kallioniemi OP. *In vivo* amplification of the androgen receptor gene and progression of human prostate cancer. Nat Genet. 1995; 9:401–06.
53. Tahiliani M, Mei P, Fang R, Leonor T, Rutenberg M, Shimizu F, Li J, Rao A, Shi Y. The histone H3K4 demethylase SMCX links REST target genes to X-linked mental retardation. Nature. 2007; 447:601–05.
54. Ge Z, Li W, Wang N, Liu C, Zhu Q, Björkholm M, Gruber A, Xu D. Chromatin remodeling: recruitment of histone demethylase RBP2 by Mad1 for transcriptional repression of a Myc target gene, telomerase reverse transcriptase. FASEB J. 2010; 24:579–86.
55. Outchkourov NS, Muiño JM, Kaufmann K, van Ijcken WF, Groot Koerkamp MJ, van Leenen D, de Graaf P, Holstege FC, Grosveld FG, Timmers HT. Balancing of histone H3K4 methylation states by the Kdm5c/SMCX histone demethylase modulates promoter and enhancer function. Cell Reports. 2013; 3:1071–79.
56. Kleine-Kohlbrecher D, Christensen J, Vandamme J, Abarrategui I, Bak M, Tommerup N, Shi X, Gozani O, Rappsilber J, Salcini AE, Helin K. A functional link between the histone demethylase PHF8 and the transcription factor ZNF711 in X-linked mental retardation. Mol Cell. 2010; 38:165–78.
57. Ramadoss S, Guo G, Wang CY. Lysine demethylase KDM3A regulates breast cancer cell invasion and apoptosis by targeting histone and the non-histone protein p53. Oncogene. 2017; 36:47-59.
58. Pretlow TG, Wolman SR, Micale MA, Pelley RJ, Kursh ED, Resnick MI, Bodner DR, Jacobberger JW, Delmoro CM, Giaconia JM, Pretlow TP. Xenografts of primary human prostatic carcinoma. J Natl Cancer Inst. 1993; 85:394–98.
59. Sramkoski RM, Pretlow TG 2nd, Giaconia JM, Pretlow TP, Schwartz S, Sy MS, Marengo SR, Rhim JS, Zhang D, Jacobberger JW. A new human prostate carcinoma cell line, 22Rv1. *In Vitro* Cell Dev Biol Anim. 1999; 35:403–09.
60. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009; 10:R25.
61. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, Liu XS. Model-based analysis of ChIP-Seq (MACS). Genome Biol. 2008; 9:R137.
62. Feng J, Liu T, Zhang Y. Using MACS to identify peaks from ChIP-Seq data. Curr Protoc Bioinformatics. 2011; Chapter 2:Unit 2.14.
63. Chen TW, Li HP, Lee CC, Gan RC, Huang PJ, Wu TH, Lee CY, Chang YF, Tang P. ChIPseek, a web-based analysis tool for ChIP data. BMC Genomics. 2014; 15:539.
64. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010; 26:841–42.
65. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. The human genome browser at UCSC. Genome Res. 2002; 12:996–1006.
66. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. Integrative genomics viewer. Nat Biotechnol. 2011; 29:24–26.
67. Bailey TL, Gribskov M. Combining evidence using p-values: application to sequence homology searches. Bioinformatics. 1998; 14:48–54.
68. Grant CE, Bailey TL, Noble WS. FIMO: scanning for occurrences of a given motif. Bioinformatics. 2011; 27:1017–18.
69. Bailey TL, Johnson J, Grant CE, Noble WS. The MEME Suite. Nucleic Acids Res. 2015; 43:W39-49.
70. Bailey TL, Elkan C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proc Int Conf Intell Syst Mol Biol. 1994; 2:28–36.
71. McLeay RC, Bailey TL. Motif Enrichment Analysis: a unified framework and an evaluation on ChIP data. BMC Bioinformatics. 2010; 11:165.
72. Smyth GK, Michaud J, Scott HS. Use of within-array replicate spots for assessing differential expression in microarray experiments. Bioinformatics. 2005; 21:2067–75.
73. Carvalho BS, Irizarry RA. A framework for oligonucleotide microarray preprocessing. Bioinformatics. 2010; 26:2363–67.
74. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015; 43:e47.
75. Benjamini Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J R Stat Soc B. 1995; 57:289–300.
76. Mootha VK, Lindgren CM, Eriksson KF, Subramanian A, Sihag S, Lehar J, Puigserver P, Carlsson E, Ridderstråle M, Laurila E, Houstis N, Daly MJ, Patterson N, et al. PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nat Genet. 2003; 34:267–73.
77. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci USA. 2005; 102:15545–50.

Chapter Six: A network of epigenomic and transcriptional cooperation encompassing an epigenomic master regulator in cancer.

This study proposes a workflow for identifying epigenetic and transcription factor cooperation in regulating a transcriptional network in cancer. By combining complementary -omics experiments a close teamwork of transcriptional and epigenomic machinery was discovered using gene promoter annotation with transcriptional motif enrichment. The ability for epigenetic factors to team up with specific transcription factors to regulate mitogenic and metabolic gene networks classifies these genes as master regulators when it comes to determining a cell's fate. Within the context of cancer oncogenic activation of these epigenetic factors are associated with the loss of proliferative control by disrupting previously established patterns of histone modifications. The therapeutic potential in understanding a network of cooperation between epigenomic and transcriptional factors allows for further understanding of mechanisms of chromatin regulation and may assist in the design of targeted therapeutics. Within this paper I contributed to the computational analysis of the data used in figures 1-5.

npj | Systems Biology and Applications

www.nature.com/npjsba

ARTICLE OPEN

A network of epigenomic and transcriptional cooperation encompassing an epigenomic master regulator in cancer

Stephen Wilson[1] and Fabian Volker Filipp[1]

Coordinated experiments focused on transcriptional responses and chromatin states are well-equipped to capture different epigenomic and transcriptomic levels governing the circuitry of a regulatory network. We propose a workflow for the genome-wide identification of epigenomic and transcriptional cooperation to elucidate transcriptional networks in cancer. Gene promoter annotation in combination with network analysis and sequence-resolution of enriched transcriptional motifs in epigenomic data reveals transcription factor families that act synergistically with epigenomic master regulators. By investigating complementary omics levels, a close teamwork of the transcriptional and epigenomic machinery was discovered. The discovered network is tightly connected and surrounds the histone lysine demethylase KDM3A, basic helix-loop-helix factors MYC, HIF1A, and SREBF1, as well as differentiation factors AP1, MYOD1, SP1, MEIS1, ZEB1, and ELK1. In such a cooperative network, one component opens the chromatin, another one recognizes gene-specific DNA motifs, others scaffold between histones, cofactors, and the transcriptional complex. In cancer, due to the ability to team up with transcription factors, epigenetic factors concert mitogenic and metabolic gene networks, claiming the role of a cancer master regulators or epioncogenes. Significantly, specific histone modification patterns are commonly associated with open or closed chromatin states, and are linked to distinct biological outcomes by transcriptional activation or repression. Disruption of patterns of histone modifications is associated with the loss of proliferative control and cancer. There is tremendous therapeutic potential in understanding and targeting histone modification pathways. Thus, investigating cooperation of chromatin remodelers and the transcriptional machinery is not only important for elucidating fundamental mechanisms of chromatin regulation, but also necessary for the design of targeted therapeutics.

npj Systems Biology and Applications (2018)4:24; doi:10.1038/s41540-018-0061-4

INTRODUCTION

Beyond genomic alterations, aberrant epigenomes contribute to many cancers, as demonstrated by widespread changes to DNA methylation patterns, redistribution of histone marks, and disruption of chromatin structure.[1] Altered epigenomes and transcriptomes are closely intertwined and share non-genomic mechanisms of dysregulation in cancer, and are therefore not just a passive by-product of cancer.[2] Epigenomic modifiers have the ability to affect the behavior of an entire network of cancer genes and can take on oncogenic roles themselves.[3] Furthermore, epigenetic factors cooperate and team up with transcription factors to control specific gene target networks.[4,5] In such works and in the following text, a *cis*-regulatory, synergistic molecular event between epigenetic and transcription factors is referred to as transcriptional cooperation (Fig. 1).

The combination of both transcriptomic and epigenomic profiling offers insight into different levels of gene regulation, transcription factor binding motifs, DNA and chromatin modifications, and how each component is coupled to a functional output. Chromatin remodelers and transcription factors are in close communication via recognition of post-translational histone modifications.[6] Thereby, they have the ability to harmonize and synchronize a dynamic exchange of chromatin between open, transcriptionally active conformations, and compacted, silenced ones.[3] Coordinated experiments interrogating transcriptional responses and chromatin binding via chromatin immunoprecipitation with next generation sequencing (ChIP-Seq) are well-equipped to capture different epigenomic and transcriptomic levels governing the circuitry of a regulatory network.[5]

Regulatory networks in biology are intrinsically hierarchical and governed by interactions and chemical modifications.[7,8] The regulome describes the interplay between genes and their products and defines the control network of cellular factors determining the functional outcome of a genomic element. The reconstruction of regulatory gene networks is stated as one of the main objectives of systems biology.[9,10] However, an accurate description of the regulome is a difficult task due to the dynamic nature of epigenetic, transcriptional, and signaling networks. Systems biology has the ability to integrate genome-wide omics data recorded by ChIP-Seq, assay for transposase-accessible chromatin using sequencing (ATAC-Seq), whole genome bisulfite sequencing (WGBS-Seq), and RNA sequencing (RNA-Seq) technology to identify gene targets of a regulatory event.[11] The integrated analysis of such data—on the one hand based on gene networks, on the other hand based on sequence features of high-resolution sequencing data—captures cooperation among regulators. Effective experimental design and data analysis of complementary epigenomic and transcriptomic platforms are required to decipher such epigenomic and transcriptional cooperation, which has a profound impact in development and disease.

[1]Systems Biology and Cancer Metabolism, Program for Quantitative Systems Biology, University of California Merced, 2500 North Lake Road, Merced, CA 95343, USA
Correspondence: Fabian Volker. Filipp (filipp@ucmerced.edu)

Received: 22 January 2018 Revised: 29 April 2018 Accepted: 7 May 2018
Published online: 1 July 2018

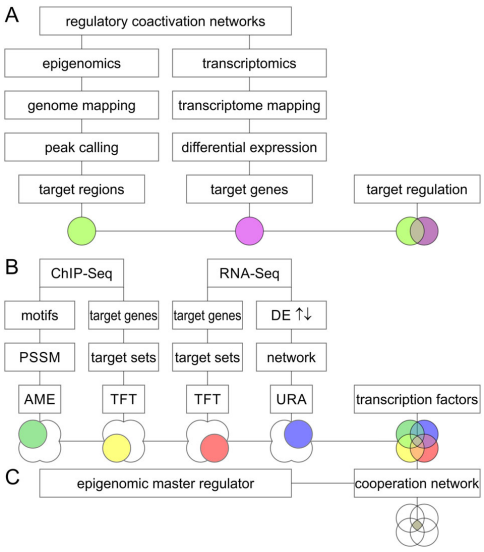


Fig. 1 Universal workflow for computational elucidation of regulatory cooperation networks. **a** By conjoining epigenomics and transcriptomics data, it is possible to define an effector network comprised of target genes affected by epigenomic regulation. The epigenomic effector network is regulated by chromatin binding or chromatin modification events resulting in gene expression changes. **b** Concerted analysis of chromatin immunoprecipitation ChIP-Seq and RNA-Seq data (or similar data) enables identification of epigenomic and transcriptomic master regulators and transcription factor networks. **c** By elucidating transcription factors associated with an epigenomic event or regulator, it is possible to identify a well-defined epigenomic-transcriptomic cooperation network supported by complementary multi-omics data. A color scheme denoting, both data types and systems biology analyses, is maintained throughout the entire document. Each color represents a specific analysis technique executed on either genome-wide epigenome and transcriptome profiles. The genome-wide intersection of epigenomic target regions (light green) with differentially expressed (DE) transcripts (purple) results in the effector network of regulated target genes. The intersection of analysis of motif enrichment (AME) and transcription factor target (TFT), and upstream regulator analysis (URA) approaches provides insights into cooperative networks of transcription factors associated with epigenomic regulators. Importantly, such genome-wide information can be accessed at the sequence or gene level providing different level of depth and resolution. ChIP-Seq data from AME (green) is enhanced by position site-specific matrix (PSSM) models of transcription factor motifs. TFT analysis can be performed on gene target sets derived from ChIP-Seq (yellow) or transcriptomics data (red). Furthermore, gene expression data contains valuable directional information indicated by arrows next to the gene expression data utilized by URA (blue), which incorporates hierarchical systems biology networks. The core analysis of the workflow includes multi-omics data integration between chromatin binding and differential gene expression events

We took advantage of published, information-rich transcriptomic and epigenomic data to study regulatory networks surrounding histone lysine demethylation. The presence or absence of methylation on histone lysine residues correlates with altered gene expression and is an integral part of the epigenetics code.[12] In particular histone 3 lysine 9 methylation (H3K9) is regarded as an epigenetic mark associated with suppressed gene activity.[13] The H3K9 lysine demethylase 3 A (KDM3A, also referred to as JMJD1A, Gene ID: 55818, HGNC ID: 20815) demethylates mono-methylated or di-methylated histone marks, thereby activating gene regulation within spermatogenesis, metabolism, stem cell activity and tumor progression.[14–16] Genome-wide ChIP-Seq data of KDM3A identified specific gene targets and transcriptional networks in androgen response, hypoxia, glycolysis, and lipid metabolism, emphasizing the importance of cooperation with transcription factors.[5] However, among epigenetic profiling experiments, a common observation is that enrichment studies provide significance for multiple transcription factors and not just one single, prioritized hit. This underscores the concept of transcriptional cooperation among epigenetic players but also emphasizes the need to design a reliable workflow that includes cross-validation with complementary, multi-omics platforms and analysis techniques.

RESULTS

Deciphering the regulatory landscape of an epigenomic and transcriptomic network

The regulatory landscape of an epigenomic player includes histone modifications, non-enzymatic chromatin interactions, cooperation with transcription factors, transcriptional modulation of gene target networks, and eventually stimulation of specific effector pathways. With the help of hierarchical experimental design, the complementary power of epigenomic and transcriptomic data can be leveraged, and thus used to address distinct levels of the regulome.[4] However, different genome-scale data platforms and analysis techniques result in the detection of significant, yet only occasionally overlapping, insight into regulatory networks. To address this, the intersection of multi-omics data levels is useful to augment and validate epigenetic regulation of transcriptional programs. At the same time, there are unique, platform-specific insights, which need to be analyzed accordingly.

Prioritization of cooperating transcription factors by integration of complementary data

The goal of our network analysis is to detect epigenomic and transcriptomic cooperation. Specifically, it is of interest to identify and prioritize transcription factors that are closely associated with an epigenomic factor by integrating complementary data (Fig. 1). Analysis of motif enrichment (AME) determines significant enrichment of transcription factor motifs among promotors or given sequences. Transcription factor target (TFT) analysis defines which transcription factor governs a set of target genes. Upstream regulator analysis (URA) integrates TFT networks with reconstructions of systems biology maps. Each platform provides a measure of significance for each detected transcription factor feature and corrects for multiple hypothesis testing using unbiased genome-wide data.[17] Importantly, computations of significance of enrichment may be performed at the level of either the gene or the sequence. Furthermore, different members of transcription factor families have the ability to recognize the same sequence motif. Therefore, dedicated searches may account for such ambiguity and overcome potential gene-specific database biases of individual transcription factors (Fig. 2). In a subsequent step, results of complementary platforms are intersected and compared. Taken together, combinations of high-throughput sequencing data deliver coordinates of epigenomic modification, enrichment of transcription factor motifs, transcriptional output, and networks of transcription factor targets (Figs. 1, 2).

Epigenomic switch of histone demethylation makes chromatin accessible and activates gene expression

Our approach is showcased by elucidating the epigenomic switch of KDM3A emphasizing its role as a master regulator. In order to

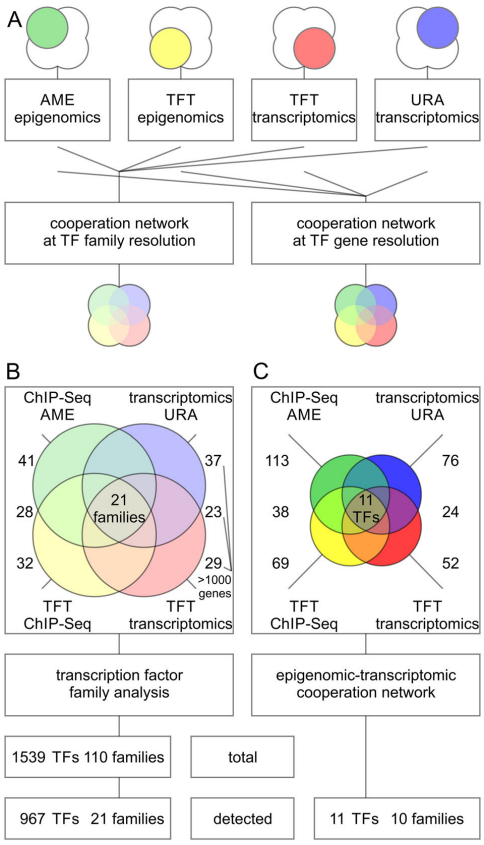


Fig. 2 Transcription factor enrichment associated with activity of an epigenetic modifier is assessed by complementary multi-omics platforms and resolved at the family and gene level. **a** Transcription factor (TF) enrichment associated with epigenomic activity is quantified using complementary omics platforms. ChIP-Seq data provides sequence-based insights on motifs (green) and genomic coordinates (yellow) of epigenomic activity. Transcriptomics data provides functional insight into regulated gene networks (red) and the direction of response (blue). The data was analyzed using analysis of motif enrichment (AME), transcription factor target analysis (TFT), or up-stream regulator analysis (URA). **b** On the one hand, the analysis is carried out at the level of transcription factor families based on position-specific matrix-assisted searches of structural motifs of transcription factor site recognition. This approach is sequence-based and considers the possibility that multiple, homologous members of transcription factor families have the ability to recognize the same transcription factor site. **c** On the other hand, the analysis is conducted at the transcript level with gene-specific insight into transcription factors and their expression levels. This approach takes advantage of regulatory networks, which are assigned to specific isoforms of homologous members of transcription factor families, and includes direction of regulation. The later approach yields a set of transcriptional coactivators that is about two orders of magnitude smaller and more specific than the transcription factor family-based approach

better understand the impact of KDM3A on transcriptional networks, coordinated ChIP-Seq and transcriptomic data of KDM3A binding and demethylation activity in combination with knockdown of KDM3A was utilized.[5] Such a combined array of matching epigenomics and transcriptomics experiments has the ability to focus on the cooperative forces of epigenetic regulation as well as its transcriptional consequences. A ChIP-Seq experiment offers direct insight into chromatin binding events and chemical modifications of histones. By overlaying genomic binding events with tracks of epigenomic marks, such as histone acetylation or methylation, associated with open or closed states of chromatin, a functional epigenomic landscape arises. Such ChIP-Seq profiles in combination with transcriptomics and functional genomics allow interrogation of the genome-wide impact of knockdown of a specific epigenomic regulator. Via genome-wide annotation and integration of sequencing reads, it becomes apparent that corresponding profiles of binding and histone modifications are reversed upon loss of function, mirroring the enzymatic function of the epigenetic modifier. Cooperative epigenomic and transcription factor binding coincides with promoter sites on meta gene coordinates enriched for histone lysine demethylation—overall indicators of transcriptionally activating epigenetic remodeling.

Regulation of transcriptional networks by H3K9 chromatin demethylation

Coordinated ChIP-Seq and transcriptomic data classified genome-wide interactions of the chromatin demethylase KDM3A using antibodies specific for KDM3A, and its histone marks H3K9me$_1$ and H3K9me$_2$, conjointly with shRNA knockdown of KDM3A in the CRL-2505 cell line. Transcriptomic impact of 4326 differentially expressed genes upon KDM3A knockdown showed 2460 genes as positively regulated by KDM3A activity (down in the CRL-2505 prostate cancer line with shRNA knockdown of KDM3A), and 1866 genes as negatively regulated by KDM3A activity. Using this data we defined the set of 56.9% differentially expressed genes as positively regulated by KDM3A activity (down in the CRL-2505 prostate cancer line with shRNA knockdown of KDM3A), and 43.1% of differentially expressed genes as negatively regulated by KDM3A activity. KDM3A binding locations were defined by a loss of ChIP-Seq signal following knockdown of KDM3A. Concurrently, H3K9me$_{1/2}$ histone marks following KDM3A knockdown are recognized as target regions of KDM3A histone lysine demethylation mediated by KDM3A. Changes in these ChIP-Seq marks upon KDM3A knockdown were contrasted against reference genomic DNA input or control non-coding shRNA samples. Quantification of the activity-based ChIP-Seq array matched with knockdown of KDM3A resulted in 37525 peaks associated with KDM3A binding, 45246 and 32665 H3K9 mono- and di-demethylation (H3K9me$_{1/2}$-KDM) events, respectively. Overall, the peak counts of both histone marks showed a gain of signal upon knockdown of KDM3A reflecting the demethylase activity. By integrating continuous ChIP-Seq signals, an average profile of a meta-gene can be generated and functional coordinates analyzed for regulatory control. In such a meta-gene profile, promotor regions are located within 1000 bp upstream of the gene-coding body, with the transcription start site (TSS) as the start of the gene-coding body at the zero position, and intergenic regions as the remaining regions outside of the gene body. KDM3A localized to the response element-rich promoter regions and demethylated H3K9me$_{1/2}$ histone marks in the proximity of the TSS. Taken together, the meta-gene analysis classified areas important for transcriptional regulation and defined genomic sequence coordinates relevant for cooperation with transcription factors.

Table 1. Detection of transcriptional cooperation by multi-omics integration of complementary data and analysis techniques

Symbol	Motif	TFClass	TF family	ChIP-Seq AME pval	ChIP-Seq TFT pval	Transcriptomics URA pval	Transcriptomics TFT pval
JUN	TGAGTCA	1.1.1	Jun-related factors	2.85E−16	1.55E−18	2.46E−03	2.40E−17
CEBPB	ATTGC GCAAT	1.1.8	C/EBP-related	4.04E−40	5.58E−09	0.00E+00	6.38E−09
MYOD	CAGGTG	1.2.2	MyoD / ASC-related factors	1.35E−38	6.97E−18	4.28E−02	6.36E−19
HIF1A	CACGC	1.2.5	PAS domain factors	1.88E−179	3.10E−02	0.00E+00	4.54E−17
SREBF1	CACATG	1.2.6	bHLH-ZIP factors	1.14E−46	1.76E−08	3.81E−02	1.64E−12
MYC	CACATG	1.2.6	bHLH-ZIP factors	9.41E−67	6.24E−21	1.97E−02	6.46E−31
AR	AGAACANNNTCTTGT	2.1.1	Steroid hormone receptors NR3 factors	1.02E−03	2.71E−02	3.30E−02	1.15E−02
SP1	CACCC	2.3.1	Three-zinc finger Krüppel-related factors	4.04E−190	2.40E−63	5.19E−03	9.68E−87
MEIS1	TGACA	3.1.4	TALE-type homeo domain factors	1.90E−04	6.26E−15	7.61E−03	7.37E−12
ZEB1	CACCTG	3.1.8	HD-ZF factors	1.02E−111	4.37E−18	2.88E−02	6.24E−19
ELK1	GGAAG	3.5.2	Ets-related factors	1.97E−03	1.43E−13	2.84E−02	1.78E−24

ChIP-Seq or transcriptomics data provide adjusted *p* values (pval) using analysis of motif enrichment (AME), transcription factor target analysis (TFT), or upstream regulator analysis (URA)

Accounting for motif similarity and structural homology of transcription factor families

The array of ChIP-Seq data was subjected to AME and TFT analysis, while the list of differentially expressed genes served as input for URA and TFT analysis (Fig. 2a). Each transcription factor hit was reported with its HGNC identification number, hierarchical classification of human transcription factors (TFClass) family barcode, motif logo, and significance corrected for multiple hypothesis testing using an adjusted *p* value cut-off of 0.05 (Table 1, Supplementary Information). Each individual analysis yielded between 29 and 41 significantly enriched transcription factor families, each corresponding to more than 1000 (1083 and up to 1292, respectively) associated genes (Fig. 2b). In comparison to the entire realm of 1539 existing transcription factors, such wide-ranging data tables provide little benefit, despite the impressive *p* values produced by analysis tools at first glance. For example hypoxia inducible factor 1 alpha subunit (HIF1A, Gene ID: 4609, HGNC ID: 4910, TFClass: 1.2.5) of the PAS domain factors (TFClass 1.2.5) is detected with an adjusted *p* value below 1.0E-100 by AME in the ChIP-Seq data. Other members of the same family like the aryl hydrocarbon receptor nuclear translocator (ARNT, Gene ID: 405, HGNC ID: 700, TFClass: 1.2.5) show similar significance, since the detection is based on the same sequence logo, highlighting the lack of ability to differentiate between structurally homologous transcription factors. Therefore, we intersected all four sets of AME ChIP-Seq, TFT ChIP-Seq, URA transcriptomics, and TFT transcriptomics, and narrowed down 21 transcription factor families supported by all datasets (Fig. 2a). Despite a considerable improvement of 21 projected families among 110 existing transcription factor families, the final set maps back to 967 transcription factors. In part, such lack of specificity is due to the large family of more than 3 adjacent zinc finger factors (TFClass: 2.3.3), whose motif was detected by the analysis but contains 487 members, accounting for almost a third of all transcription factors (Fig. 3a). Systems biology networks and enrichment studies provide insight into directionality of the response and draw attention to different sized effector networks (Fig. 3b, c). Only few nodes of the transcription factor target network were hyperconnected and showed promoter association with multiple transcription factors in epigenomics and transcriptomics datasets (Figs. 4, 5). Such a high degree of network connectivity speaks to a synergistic effect, where selected master regulators cooperate and act in sync, resulting in robust transcriptional output.

Multi-omics integration of complementary data yields well-refined target network

In order to improve the detected output, gene-specific systems biology networks were employed. In particular, URA and TFT analysis fueled by transcriptomic data provide useful insight. Databases of gene sets rely in part on experimental data of gene-specific knockdowns to characterize the impact of a transcription factor on a target network. Furthermore, a consistent directional response amplifies the significance of a detected hit. Therefore, such directional, gene-specific networks have the ability to overcome ambiguities. For example, members of the JUN-related factors (TFClass: 1.1.1) are detected but show different signs of regulation depending on the factor of interest, the response of its target genes, and the change of expression of the factor itself. After intersection of all four datasets, 11 transcription factors belonging to 10 transcription factor families were determined (Fig. 2c). This set of master regulators is supported by complementary omics platforms and different analysis techniques representing a high-confidence cooperation network of the epigenomic master regulator KDM3A (Fig. 5). The cooperation network includes cancer associated factors Jun proto-oncogene, AP-1 transcription factor subunit (JUN, Gene ID: 3725, HGNC ID: 6204, TFClass: 1.1.1), CCAAT/enhancer binding protein beta (CEBPB, Gene ID: 1051, HGNC ID: 1834, TFClass: 1.1.8), myogenic differentiation 1 (MYOD1, Gene ID: 3091, HGNC ID: 7611, TFClass: 1.2.2), HIF1A, sterol regulatory element binding transcription factor 1 (SREBF1, Gene ID: 379, HGNC ID: 11289, TFClass: 1.2.6), MYC proto-oncogene, bHLH transcription factor (MYC, Gene ID: 4609, HGNC ID: 7553, TFClass: 1.2.6), androgen receptor (AR, Gene ID: 367, HGNC ID: 644, TFClass: 2.1.1), Sp1 transcription factor (SP1, Gene ID: 6667, HGNC ID: 11205, TFClass: 2.3.1), Meis homeobox 1 (MEIS1, Gene ID: 4211, HGNC ID: 7000, TFClass: 3.1.4), zinc finger E-box binding homeobox 1 (ZEB1, Gene ID: 6935, HGNC ID: 11642, TFClass: 3.1.8), and ELK1, ETS transcription factor (ELK1, Gene ID: 2002, HGNC ID: 3321, TFClass: 3.5.2) representing less than 0.8% of all possible transcription factors (Figs. 4d, 5). Thereby it represents a hyperconnnected

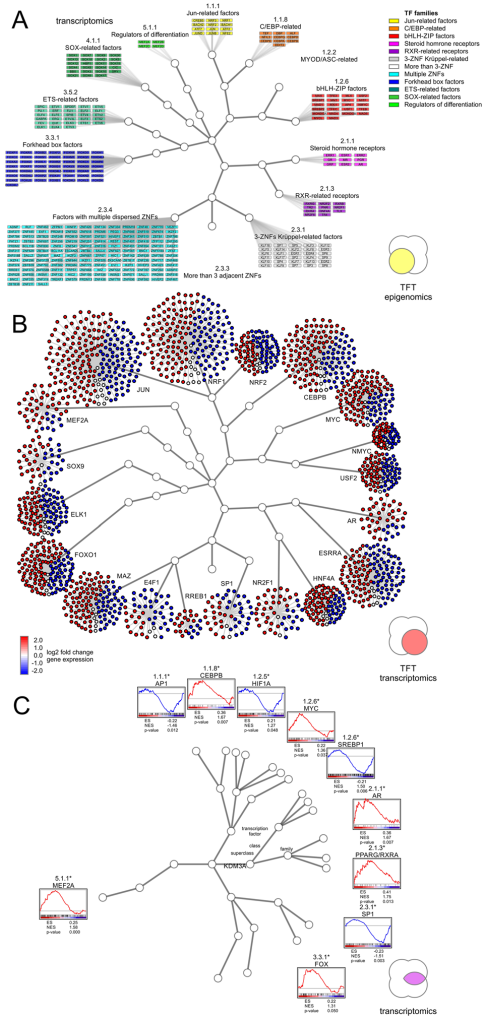


Fig. 3 Visualization of epigenomic and transcriptional cooperation illustrates redundancy and complexity of a target network. **a** Hierarchical trees of human transcription factors correspond to transcription factor superclass, class, and family from inward out. Transcription factor motifs often get recognized by multiple members of the same transcription factor family due to structural homology of DNA binding domains. The transcription factor target analysis (TFT) is carried out on sequence-specific epigenomics data. **b** Size of the transcriptional effector network and the direction of response are key parameters when evaluating target genes in epigenomic cooperation. The TFT analysis is based on differentially expressed transcripts with transcriptomic up and down response in red and blue, respectively. **c** Transcription factor target networks provide insight into enrichment and direction of the response. Identified cooperating transcription factors show agreement between complementary data of expression levels, direction of regulation, target sets, and hierarchical linkage

network of networks surrounding an epigenomic cooperation event. The identified factors can further be surveyed at the level of basal expression or regulation in patient-derived tumor specimens in TCGA underlining elevated expression in tumor progression. The analysis validates previously reported associations and contacts implicated in chromatin remodeling and discovered newly identified cooperative interactions (Table 2). For the specific role of KDM3A in cancer, epigenomic and transcriptional cooperation with transcription factors is key. KDM3A cooperates with mitogenic basic helix-loop-helix factors including MYC, HIF1A, and SREBF1, and derives a lipogenic program from association with nuclear receptors like AR. Ultimately, by overlaying the motif-specific and genomic data produced through matched experiments, epigenomic events can be correlated with the transcriptomic effect of histone remodelers and transcription factors.

DISCUSSION

ChIP-Seq based approaches provide sequence resolution but detection of enriched transcription factor motifs is ambiguous and is most appropriately accomplished at the transcription factor family level to account for and include homologous factors. In contrast, transcriptomics studies provide directionality of regulation—transcriptional activation or repression upon epigenomic activity—an important aspect lacking in coordinate-based ATAC-Seq or ChIP-Seq experiments. Integration of different sequence, gene, or network-based approaches prioritizes high-fidelity cooperation partners in epigenomic regulation. Therefore, any combination of complementary data from sequence, gene, or network-based approaches is identified as desirable input for reliable regulatory systems biology analyses.

For the oncogenic nature and target specificity of an epigenomic master regulator, epigenomic and transcriptional cooperation with transcription factors is key. KDM3A is able to support initiation of transcription by its ability to specifically remove mono-methylation and di-methylation marks from the H3K9 residue leading to chromatin de-condensation.[14] Transcriptionally silenced genes contain methylation marks on the H3K9 subunit[18] and arrays of ChIP-Seq experiments monitoring H3K9 methylation marks revealed global histone demethylation effects of KDM3A. Combined assessment of histone demethylation events and gene expression changes indicated major transcriptional activation, suggesting that distinct oncogenic regulators, in particular transcription factors, may synergize with the epigenetic patterns controlled by KDM3A. Furthermore, the epigenetic factor was shown to cooperate with the androgen receptor to control prostate tissue-specific gene target networks introducing the concept of the epioncogene and transcriptional cooperation.[19]

While KDM3A is able to control chromatin accessibility, the mechanism by which it targets specific genes is of current interest and may influence understanding of epigenetic dysregulation in human disease. While several cancers exhibit deregulated KDM3A activity, in prostate adenocarcinoma it functions as a transcriptional coregulator with the androgen receptor.[5,14,20,21] Such cooperative coactivation of the androgen receptor with KDM3A features a role for KDM3A as an active force in commencing oncogenesis in prostate epithelial cells. KDM3A is known to control the transcription and function of oncogenic transcription factors.[22,23] However, an expanded study outlining the effects of perturbed KDM3A H3K9 demethylation upon human transcription factor response element recognition in cancer has so far been missing.

The cooperation network includes previously validated interactions of MYC, HIF1A, and AR in cancer but also highlights close association of KDM3A with transcriptional networks of factors

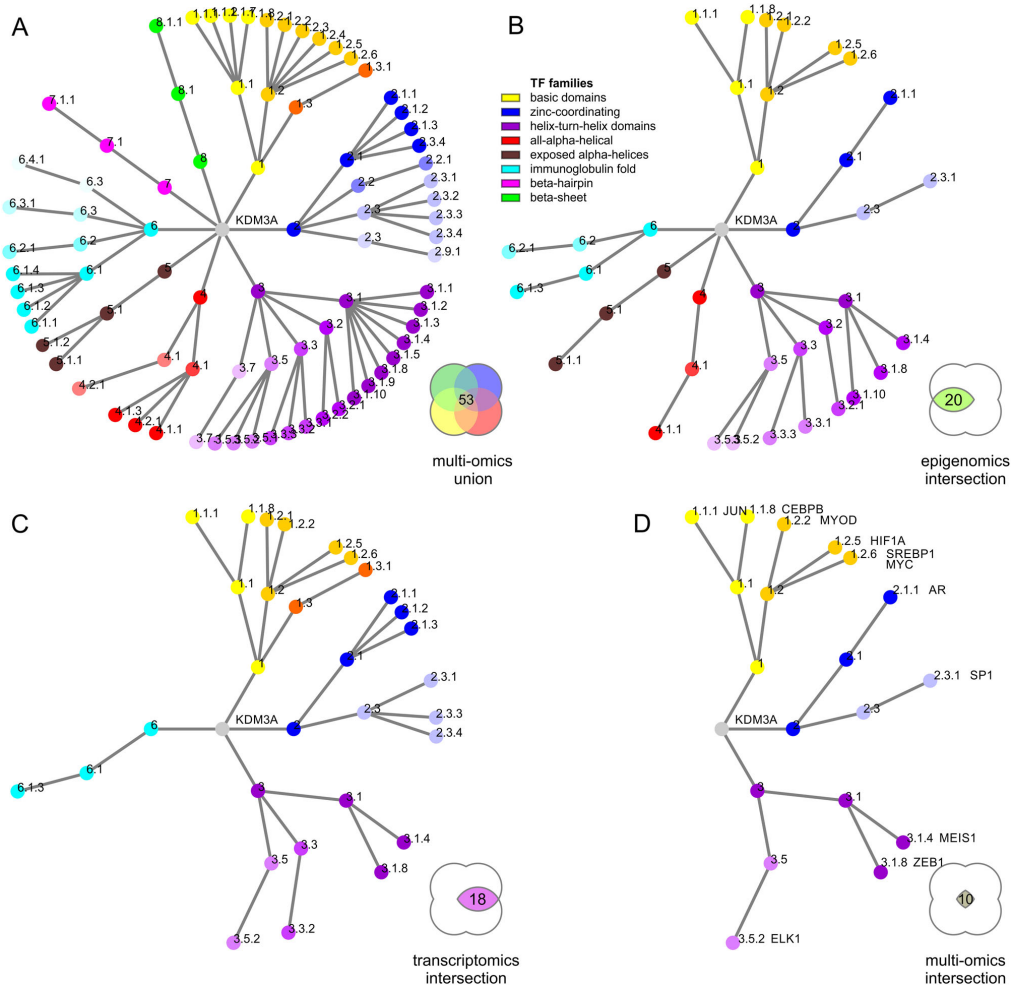


Fig. 4 Network of transcriptional cooperation of an epigenomic master regulator visualized by transcription factor family trees. **a** A network of transcription factors is detected using complementary epigenomic and transcriptomic data. The realm of potentially relevant transcription factor families is large yet many data points are not mirrored or validated by different platforms. **b**, **c** Confidence and mutual data support increases by integrating analysis of motif enrichment (AME), transcription factor target analysis (TFT), or up-stream regulator analysis (URA). For each data source B chromatin modifications or **c** differential expression of target genes detected transcription factors supported by at least two complementary techniques, AME and TFT for epigenomics data, and URA and TFT for transcriptomics. **d** High-confidence target network of transcription factors validated by four different omics platforms integrating ChIP-Seq based motifs and transcriptional networks. Legends of colored areas in Venn diagrams illustrate intersections of complementary datasets and analysis platforms with numbers of identified transcription factor families, respectively. The numbers next to individual nodes of the hierarchical family tree indicate transcription factor superclass, class, and family from inward out

rather studied in development and tissue differentiation like JUN, CEBPB, MYOD, SREBF1, SP1, MEIS1, ZEB1, or ELK1 (Table 1). By taking advantage of motif-specific target networks, KDM3A has the ability to induce glycolytic genes in urothelial bladder carcinoma.[24,25] Epigenomic regulation of SREBF1 activity has been reported to stimulate lipogenesis, and SREBP1 regulates lipid accumulation and cell cycle progression in androgen independent prostate cancer cell lines.[26–28] KDM3A regulates the transcriptional program of the AR, serves epigenomic master regulator by epigenomic and transcriptional cooperation of prostate adenocarcinoma.[5] Despite some factors including the forkhead box (FOX) factors (TFClass: 3.3.1) family were frequently detected at the transcription factor family levels, lack of consistent overlap of epigenomic and transcriptomic data eventually excluded prominent cancer drivers like forkhead box A1 (FOXA1, Gene ID: 3169, HGNC ID: 5021, TFClass: 3.3.1), forkhead box M1 (FOXM1, Gene ID: 2305, HGNC ID: 3818, TFClass: 3.3.1), forkhead box O1 (FOXO1, Gene ID: 2308, HGNC ID: 3819, TFClass: 3.3.1), or forkhead box O3 (FOXO3, Gene ID: 2309, HGNC ID: 3821, TFClass: 3.3.1). Despite FOX factors are known to cooperate with nuclear

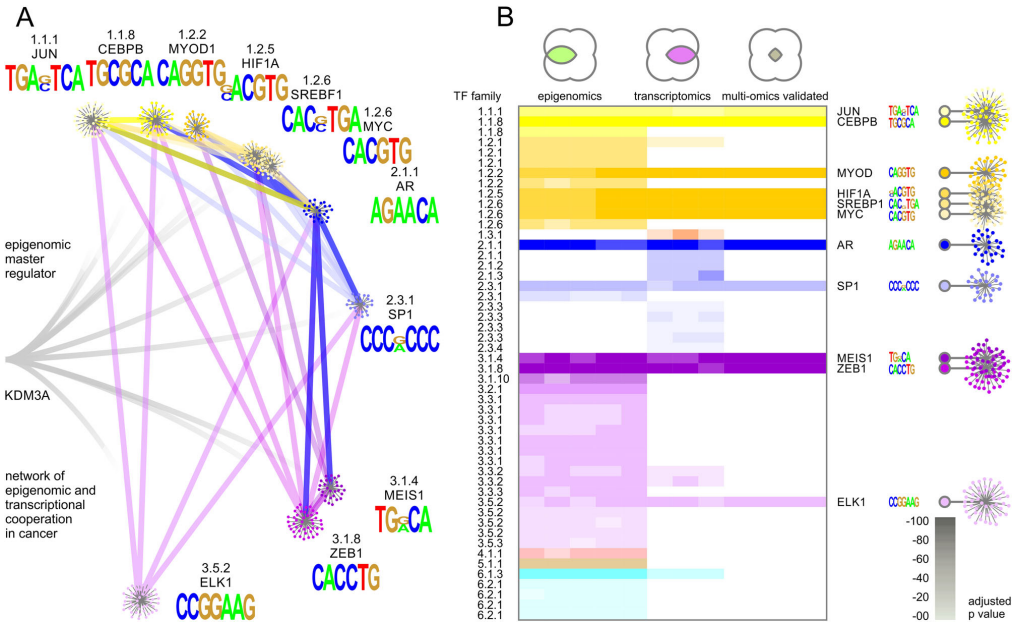


Fig. 5 The network of networks convoluted in epigenomic and transcriptional cooperation by an epigenomic master regulator displays specificity yet redundancy and hyperconnectivity. **a** A highly specific effector network is accomplished by epigenomic and transcriptional cooperation. Key features of the network are hyperconnectivity of utilized nodes and targeting of effector genes by multiple cooperating transcription factors. **b** Each identified node of the transcription factor network comprising epigenomic cooperation is supported by complementary epigenomic and transcriptomic data. Each cooperating transcription factor supports the epigenomic factor—in this case KDM3A—by contributing motif-specific recognition, by directing chromatin accessibility, and by controlling transcriptional coactivation

hormone receptors (TFClass: 2.1.1), it is possible that the putative association with KDM3A steams from the fact that the closely cooperating AR frequently has FOX motifs nearby. Systematic, genome-wide surveys have elucidated that FOX motifs are adjacent androgen response elements (AREs),[29,30] thereby facilitating cooperation at the level of transcription factors and promotion of prostate cancer progression.

Hundreds of transcription factors are significantly associated with each individual data analysis platform or high-throughput sequencing technology. Big data challenges can be overcome by systems biology analysis and integration of multi-omics data. Motif similarity is visualized by transcription factor family trees classifying superclass, class, and family of transcription factors (from inward to outward, Fig. 4) based on the characteristics of their DNA-binding domains. Single epigenomic or transcriptomic datasets examined by different analysis tools result in improved resolution but leave ambiguities. The footprints of cooperating transcription factors are found in cognate sequence motifs specific to their DNA binding domains. Such sequence motifs are more pronounced in events of cooperating epigenomic activity. Detected motif enrichment highlights the modularity, versatility, and efficacy of epigenomic cooperation, providing target specificity at genome-wide reach. The number of detected events in genome-wide epigenomic binding studies provides statistical power for sequence motif discovery and gene target enrichment. As a consequence, high-resolution epigenomic studies often arrive at multiple plausible solutions, though each suggested interaction or association may carry statistical significance. By intersecting complementary data platforms and analysis techniques, high-fidelity gene target networks involved in epigenomic and transcriptomic cooperation can be identified.

Within the regulome, epigenetic master regulators position themselves at the top of cellular hierarchies and control distinct phenotypic programs via reversible chemical modifications of chromatin, histone or nucleotide marks, without altering the core DNA sequence. Epigenetic oncogenes or tumor suppressors can arise when epigenetic master regulators are somatically activated or lost, and contribute to cancer initiation and progression.[31] In cancer, such epigenetic master regulators are found at the top of regulatory hierarchies, particularly in pathways related to cellular proliferation, survival, fate, and differentiation. For the manifestation of a genomic or non-genomic aberration of an epigenetic master regulator, it is a necessity that its own activity is affected by somatic mutation, copy number alteration, expression levels, protein cofactors, or methylation status. Epigenetic master regulators often accomplish target specificity of their phenotypic program by cooperation with members of the transcriptional machinery and therefore may depend on tissue-specific expression of such auxiliary factors. In cancer, an epigenetic master regulator populates an extreme state and is either permanently switched on or off. An epigenetic master regulator will become a cancer driver, if it is not functionally neutral but rather contributes to tumorigenesis or disease progression due to its hyperactive or deactivated state. Genomic profiling of cancer patients has the ability to identify coincidence or mutual exclusivity of somatic alterations of epigenomic and transcription factors. Extreme states of epigenetic master regulators by somatic loss or gain of function in cancer may emphasize preexisting cooperative interactions with transcription factors, which may be subtle and difficult to detect under normal circumstances. A defined challenge in the field of epigenetic master regulators is to identify cancer-specific vulnerabilities in gene targets and biological pathways that are

Table 2. Epigenomic and transcriptional cooperation events in cancer. Original findings and reported cooperation events of epigenomic regulators with transcription factor are enumerated

Symbol	TFClass	Cooperation event	Ref.
JUN	1.1.1	KDM3A assists in recruiting JUN to AP1 binding sites in regulating expression of CD44, MMP7, and PDGFRB in liver adenocarcinoma tumor formation	51
		KDM4A promotes a positive feedback loop by facilitating the binding of the AP1 complex to the promoters JUN and FOSL1 in squamous cell carcinoma cells	52
CEBPB	1.1.8	Novel event	
		KDM4B serves as a cofactor for CEBPB in preadipocytes and is recruited to the promoters of CEBPB regulated cell cycle genes	53
MYOD	1.2.2	Novel event	
		KDM4B regulates the expression of MYOD and physically interacts with MYOD thereby controlling myogenic differentiation	54
		KDM4C decreases MYOD degradation and increase MYOD transcriptional activity to facilitate skeletal muscle differentiation	55
HIF1A	1.2.5	KDM3A expression is stimulated by HIF1A binding to a response element in the promoter region of KDM3A	24
		KDM3A is regulated by HIF1A stimulating tumor formation in renal cell carcinoma	22
		KDM3A cooperates with HIF1A to induce glycolytic genes in urothelial bladder carcinoma	25
SREBF1	1.2.6	Novel event	
		KDM1A regulates SREBF1 binding to the FASN promotor stimulating lipogenesis	26
MYC	1.2.6	KDM3A stimulates MYC expression and attenuates its ubiquitin-dependent degradation by binding to a E3 ubiquitin ligase	56
		KDM3A regulates transcription of MYC and PAX3 by directly binding to their promotors and regulates their $H3K9me_2$ level in breast adenocarcinoma	57
		KDM4B binds the MYC/MAX motif and regulates expression of MYC signaling in neuroblastoma	58
		N-MYC physically interacts and recruits KDM4B. Additionally KDM4B is able to regulate the expression of MYC signaling in neuroblastoma	58
AR	2.1.1	KDM3A regulates the transcriptional program of the AR, serves epigenomic master regulator by epigenomic and transcriptional cooperation of prostate adenocarcinoma	5
		KDM3A facilitates transcriptional activation by hormone-dependent recruitment of the AR to target genes in prostate adenocarcinoma	14
		KDM1A and KDM4D bind to the AR and localize to ARE half sites in the promoter region of VEGFA in placental development	21
		KDM4A binds the AR and supporting urothelial bladder carcinoma initiation and progression	59
		KDM4B enhances AR transcriptional activity by demethylation and inhibits ubiquitination of the AR	60
SP1	2.3.1	Novel event	
		KDM4A silences SP1 by chromatin demethylation in breast adenocarcinoma	61
MEIS1	3.1.4	Novel event	
ZEB1	3.1.8	Novel event	
ELK1	3.5.2	Novel event	

frequently and consistently perturbed under the control of an epigenetic driver.

CONCLUSION

In conclusion, the identification of transcriptional cooperation and regulatory hierarchies highlights the importance of epigenetic regulators in mitogenic control and their potential as therapeutic targets. Epigenetic regulators such as KDM1A, KDM3A, KDM5A, KDM6A, KDM7A, EZH2, DOT1L, and others have been shown to be critical in oncogenesis and cancer resistance.[3,32,33] The discovery of the specific role of KDM3A in the interplay between a tissue-specific steroid receptor transcription factor and metabolic signaling provides a foundation for rational design of combination approaches where metabolic, epigenetic, and hormone-deprivation therapies may synergize. Our integrated multi-platform analysis reveals a complex molecular landscape of epigenomic and transcriptomic cooperation in cancer, providing avenues for precision medicine.[34] A close teamwork of the transcriptional and epigenomic machinery was discovered, in which one component opens the chromatin, another recognizes gene-specific DNA motifs, and others scaffold between histones, cofactors, and the transcriptional complex. This highlights a close connection between the epigenomic and transcriptomic machinery, albeit much of the underlying principles remain to be discovered. In conclusion, transcriptomics in combination with epigenomic profiling and measurement of chromatin accessibility enable global detection of epigenetic modifications and characterization of transcriptional and epigenetic footprints. Chromatin remodelers and transcription factors are in close communication via recognition of post-translational histone modifications and coordinate the dynamic exchange of chromatin between open, transcriptionally active conformations and compacted, silenced ones. In cancer, due to the ability to team up with transcription factors, epigenetic factors concert mitogenic and metabolic gene networks, claiming the role of a cancer master regulators or epioncogenes. Exploration into the cooperative roles of epigenetic histone modifiers and transcription factor families in gene regulatory networks contributes to our understanding of how a seemingly promiscuous epigenomic program is converted into a specific transcriptional response assisting in oncogenesis.

METHODS

Experimental design

Optimal experimental design mirrors different layers of the regulatory network and organizes sequencing assays in an array of coordinated experiments.[4,5,35] Coordinated epigenomic and transcriptomic profiles are well-equipped to capture different regulatory levels governing the circuitry of a cooperative network. The workflow introduced in our approach is

open and applicable to different epigenome and transcriptome profiles including ChIP-Seq, ATAC-Seq, RNA-Seq, and microarray experiments. Since the analysis primarily relies on differential expression, microarray or RNA-Seq data are equally applicable. Additionally, data on alternative transcripts may allow resolving regulation of splice-isoforms. For coordinated multi-omics analysis, it is important to integrate compatible datasets and look for matching conditions, each associated with presence or absence of a defined epigenomic event or factor. Genomic editing offers tools to conduct target-specific loss or gain of function studies such that an array of coordinated experiments can be assembled. Complementary epigenomic and transcriptomic data—on the one hand in form of significantly enriched genomic regions (epigenomic target regions), on the other hand as differentially expressed transcripts (target genes)—serves as input for four different analysis platforms (Figs. 1, 2).

Investigation of KDM3A as epigenetic switch in human cancer

KDM3A coordinates transcriptional activation by H3K9 demethylation thereby enabling chromatin accessibility and replacement of components of nucleosome-stalled polymerase complexes by tissue-specific transcription factors. The workflow is exemplified by reprocessing published data on KDM3A activity recorded with matching knockdown conditions in a human prostate carcinoma epithelial cellular model deposited in NCBI GEO entries GSE109748 and GSE70498[29,36,37] (CRL-2505, American Type Culture Collection, Manassas, VA). The utilized cellular model is a variant derived from a xenograft and simulates castration-induced regression and relapse typical of human prostate carcinoma epithelial cells independent of dihydroxytestosterone stimulation. Furthermore, KDM3A is activated by somatic copy number amplification in lung, prostate, uterine, bladder, testicular germ cell, ovarian, cervical, breast, sarcoma, melanoma, and other cancers making its cooperation network an important target of broad interest in oncology.

Data processing

Illumina HiSeq 2000 (Illumina, San Diego, CA) fastq files were aligned to the reference human genome 19 using the Bowtie software package.[38] Peak-calling utilized a model-based analysis of ChIP-Seq (MACS) algorithm.[39,40] Significant ChIP-Seq genomic locations relative to nearby gene bodies were annotated by ChIPSeek.[41] ChIP-Seq peak regions were sorted and filtered by BEDtools.[42] Average ChIP enrichment profiles over specific genomic features were calculated using the *cis*-regulatory element annotation system tool.[43,44] ChIPSeq binding profiles were visualized in the integrative genomics viewer (IGV).[45] Utilized conditions include ChIP-Seq profiles of antibodies specific for chromatin marks H3K9me$_1$, H3K9me$_2$, and the epigenomic modifier KDM3A in combination with small hairpin RNA (shRNA) knockdown of KDM3A matched with coordinated transcriptomic data of control and KDM3A knockdown cells using human transcriptome platform GPL10558 (HT-12 V4.0, Illumina, San Diego, CA).[5] The epigenomic and transcriptomics datasets contained 77911 features and 4356 differentially expressed transcripts upon KDM3A knockdown, respectively, with p values and q values below 0.05 adjusted for multiple hypothesis testing.

Network analysis and transcription factor target enrichment

Human transcription factors were annotated according to their Human Genome Organization (HUGO) Gene Nomenclature Committee (HGNC) identification number using the using the multi-symbol checker tool. Discovered transcription factors were classified by shared DNA binding domains according to the hierarchical classification of human transcription factors (TFClass) database.[46] Transcription factor binding and promoter sites were annotated utilizing transcription factor databases.[30,47] For transcription factor binding site searches we built manual or utilized deposited position site-specific matrices or sequence logos of curated, non-redundant transcription factor databases. Statistically significant enrichment of these transcription factor motifs was determined using find individual motif occurrences (FIMO) and motif enrichment tools of the motif-based sequence analysis toolkit (MEME) suite.[48] Upstream regulators were determined by ingenuity pathway analysis (IPA, Qiagen, Redwood City, CA) based on differentially expressed genes with an adjusted p value below 0.05. Significant enrichment of target gene networks with consistent transcription factor motifs was calculated for all target genes with annotated transcription factor motifs in the 3′ promoter region of their transcription start sites.[49]

Data availability

Data is deposited in NCBI GEO entries GSE109748 and GSE70498. Identified transcription factors and statistics are assembled in the Supplementary Information. All data supporting the findings of this study is openly available within the paper and the Supplementary Information deposited at the *npj Systems Biology and Applications* website. A preprint version of this manuscript is made available to the scientific community on the preprint server *bioRxiv* 309484.[50]

Supplementary Table 1–6 are compiled as Supplementary Information. Supplementary Table 1: Master regulators among epigenomic and transcriptomic cooperation network. Supplementary Table 2: Detection and hierarchical classification of human transcription factors. Supplementary Table 3: ChIP-Seq analysis of motif enrichment (AME). Supplementary Table 4: ChIP-Seq transcription factor target (TFT) analysis. Supplementary Table 5: Transcriptomics upstream regulator analysis (URA). Supplementary Table 6: Transcriptomics transcription factor target (TFT) analysis.

ACKNOWLEDGEMENTS

Practicing science is experiencing the creativity of life. Undertaking systems biology is bringing it to life. This work on cancer systems biology is generously supported by the National Institutes of Health and the National Science Foundation. Emilia Mili Filipp—fierce and fearless friend, protector and companion on all adventures of Franzi Violet Filipp and Leland Volker Filipp. F.V.F. is grateful for the support of grant CA154887 from the National Institutes of Health, National Cancer Institute and grant CRN-17-427258 by the University of California, Office of the President, Cancer Research Coordinating Committee, the Goethe Institute, Washington, DC, USA, and the Federal Foreign Office, Berlin, Germany.

AUTHOR CONTRIBUTIONS

S.W. and F.V.F. conducted the data analysis. F.V.F. designed the network study, directed the systems biology analysis, performed data interpretation, prepared the illustrations, and wrote the text. S.W. and F.V.F. reviewed the final manuscript.

ADDITIONAL INFORMATION

Supplementary information accompanies the paper on the *npj Systems Biology and Applications* website (https://doi.org/10.1038/s41540-018-0061-4).

Competing interests: The authors declare no competing interests.

Ethical approval: All experimental protocols were approved by the Institutional Review Boards at the University of California Merced. The study was carried out as part of IRB UCM13-0025 of the University of California Merced and as part of dbGap ID 5094 for study accession phs000178.v9.p8 on somatic mutations in cancer and conducted in accordance with the Helsinki Declaration of 1975.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

REFERENCES

1. Dawson, M. A. The cancer epigenome: concepts, challenges, and therapeutic opportunities. *Science* **355**, 1147–1152 (2017).
2. Leung, J. K. & Sadar, M. D. Non-genomic actions of the androgen receptor in prostate cancer. *Front. Endocrinol.* **8**, 2 (2017).
3. Filipp, F. V. Crosstalk between epigenetics and metabolism-Yin and Yang of histone demethylases and methyltransferases in cancer. *Brief. Funct. Genom.* **16**, 320–325 (2017).
4. Filipp, F. V. Epioncogenes in cancer—identification of epigenomic and transcriptomic cooperation networks by multi-omics integration of ChIP-Seq and RNA-Seq data. *Syst. Biol. Methods Mol. Biol.* **1800**, 101–121 (2019).
5. Wilson, S., Fan, L., Sahgal, N., Qi, J. & Filipp, F. V. The histone demethylase KDM3A regulates the transcriptional program of the androgen receptor in prostate cancer cells. *Oncotarget* **8**, 30328–30343 (2017).
6. Ahsendorf, T., Muller, 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).
7. Cheng, C. et al. An approach for determining and measuring network hierarchy applied to comparing the phosphorylome and the regulome. *Genome Biol.* **16**, 63 (2015).
8. Ay, A., Gong, D. & Kahveci, T. Hierarchical decomposition of dynamically evolving regulatory networks. *BMC Bioinform.* **16**, 161 (2015).

9. Filipp, F. V. Cancer metabolism meets systems biology: pyruvate kinase isoform PKM2 is a metabolic master regulator. *J. Carcinog.* **12**, 14 (2013).
10. Filipp, F. V. A gateway between omics data and systems biology. *J. Metab. Syst. Biol.* **1**, 1 (2013).
11. Gonda, T. J. & Ramsay, R. G. Directly targeting transcriptional dysregulation in cancer. *Nat. Rev. Cancer* **15**, 686–694 (2015).
12. Zentner, G. E. & Henikoff, S. Regulation of nucleosome dynamics by histone modifications. *Nat. Struct. Mol. Biol.* **20**, 259–266 (2013).
13. Kooistra, S. M. & Helin, K. Molecular mechanisms and potential functions of histone demethylases. *Nat. Rev. Mol. Cell Biol.* **13**, 297–311 (2012).
14. Yamane, K. et al. JHDM2A, a JmjC-containing H3K9 demethylase, facilitates transcription activation by androgen receptor. *Cell* **125**, 483–495 (2006).
15. Okada, Y., Scott, G., Ray, M. K., Mishina, Y. & Zhang, Y. Histone demethylase JHDM2A is critical for Tnp1 and Prm1 transcription and spermatogenesis. *Nature* **450**, 119–123 (2007).
16. Kuroki, S. et al. Epigenetic regulation of mouse sex determination by the histone demethylase Jmjd1a. *Science* **341**, 1106–1109 (2013).
17. Storey, J. D. & Tibshirani, R. Statistical significance for genomewide studies. *Proc. Natl Acad. Sci. USA* **100**, 9440–9445 (2003).
18. Barth, T. K. & Imhof, A. Fast signals and slow marks: the dynamics of histone modifications. *Trends Biochem. Sci.* **35**, 618–626 (2010).
19. Qi, J. & Filipp, F. V. An epigenetic master regulator teams up to become an epioncogene. *Oncotarget* **8**, 29538–29539 (2017).
20. Yamada, D. et al. Role of the hypoxia-related gene, JMJD1A, in hepatocellular carcinoma: clinical impact on recurrence after hepatic resection. *Ann. Surg. Oncol.* **19**, S355–364 (2012).
21. Cleys, E. R. et al. Androgen receptor and histone lysine demethylases in ovine placenta. *PLoS One* **10**, e0117472 (2015).
22. Krieg, A. J. et al. Regulation of the histone demethylase JMJD1A by hypoxia-inducible factor 1 alpha enhances hypoxic gene expression and tumor growth. *Mol. Cell. Biol.* **30**, 344–353 (2010).
23. Ohguchi, H. et al. The KDM3A-KLF2-IRF4 axis maintains myeloma cell survival. *Nat. Commun.* **7**, 10258 (2016).
24. Wellmann, S. et al. Hypoxia upregulates the histone demethylase JMJD1A via HIF-1. *Biochem. Biophys. Res. Commun.* **372**, 892–897 (2008).
25. Wan, W. et al. Histone demethylase JMJD1A promotes urinary bladder cancer progression by enhancing glycolysis through coactivation of hypoxia inducible factor 1alpha. *Oncogene* **36**, 3868–3877 (2017).
26. Abdulla, A. et al. Regulation of lipogenic gene expression by lysine-specific histone demethylase-1 (LSD1). *J. Biol. Chem.* **289**, 29937–29947 (2014).
27. Nambiar, D. K., Deep, G., Singh, R. P., Agarwal, C. & Agarwal, R. Silibinin inhibits aberrant lipid metabolism, proliferation and emergence of androgen-independence in prostate cancer cells via primarily targeting the sterol response element binding protein 1. *Oncotarget* **5**, 10017–10033 (2014).
28. Huang, W. C., Li, X., Liu, J., Lin, J. & Chung, L. W. Activation of androgen receptor, lipogenesis, and oxidative stress converged by SREBP-1 is responsible for regulating growth and progression of prostate cancer cells. *Mol. Cancer Res.* **10**, 133–142 (2012). 1541-7786.
29. Wilson, S., Qi, J. & Filipp, F. V. Refinement of the androgen response element based on ChIP-Seq in androgen-insensitive and androgen-responsive prostate cancer cell lines. *Sci. Rep.* **6**, 32611 (2016).
30. Kulakovskiy, I. V. et al. HOCOMOCO: towards a complete collection of transcription factor binding models for human and mouse via large-scale ChIP-Seq analysis. *Nucleic Acids Res.* **46**, D252–D259 (2018).
31. Tiffen, J., Wilson, S., Gallagher, S. J., Hersey, P. & Filipp, F. V. Somatic copy number amplification and hyperactivating somatic mutations of EZH2 correlate with DNA methylation and drive epigenetic silencing of genes involved in tumor suppression and immune responses in melanoma. *Neoplasia* **18**, 121–132 (2016).
32. Filipp, F. V. How cancer can become therapy-resistant—epigenetics might play a role. *Sci. Am.* **318**, 1, https://www.scientificamerican.com/article/how-cancer-can-become-therapy-resistant/ (2018).
33. Zecena, H. et al. Systems biology analysis of mitogen activated protein kinase inhibitor resistance in malignant melanoma. *BMC Syst. Biol.* **12**, 33 (2018).
34. Filipp, F. V. Precision medicine driven by cancer systems biology. *Cancer Metastas. Rev.* **36**, 91–108 (2017).
35. Rajewsky, N., Jurga, S. & Barciszewski, J. Transcriptional and epigenomic cooperation—systems biology: RNA technologies. *Syst. Biol. Methods Mol. Biol.* **1800**, 101–121 (2019).
36. Pretlow, T. G. et al. Xenografts of primary human prostatic carcinoma. *J. Natl Cancer Inst.* **85**, 394–398 (1993).
37. Sramkoski, R. M. et al. A new human prostate carcinoma cell line, 22Rv1. *In Vitro Cell. Dev. Biol. Anim.* **35**, 403–409 (1999).
38. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
39. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
40. Feng, J., Liu, T., & Zhang, Y. Using MACS to identify peaks from ChIP-Seq data. *Curr. Protoc. Bioinform* **34**, 1–14 (2011).
41. Chen, T. W. et al. ChIPseek, a web-based analysis tool for ChIP data. *BMC Genom.* **15**, 539 (2014).
42. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
43. Shin, H., Liu, T., Manrai, A. K. & Liu, X. S. CEAS: cis-regulatory element annotation system. *Bioinformatics* **25**, 2605–2606 (2009).
44. Ji, X., Li, W., Song, J., Wei, L. & Liu, X. S. CEAS: cis-regulatory element annotation system. *Nucleic Acids Res.* **34**, W551–554 (2006).
45. Thorvaldsdottir, H., Robinson, J. T. & Mesirov, J. P. Integrative genomics viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinform.* **14**, 178–192 (2013).
46. Wingender, E., Schoeps, T., Haubrock, M., Krull, M. & Donitz, J. TFClass: expanding the classification of human transcription factors to their mammalian orthologs. *Nucleic Acids Res.* **46**, D343–D347 (2018).
47. Khan, A. et al. JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic Acids Res.* **46**, D260–D266 (2018).
48. Bailey, T. L., Johnson, J., Grant, C. E. & Noble, W. S. The MEME suite. *Nucleic Acids Res.* **43**, W39–49 (2015).
49. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
50. Wilson, S. & Filipp, F. V. A network of epigenomic and transcriptional cooperation encompassing an epigenomic master regulator in cancer. *bioRxiv*, https://doi.org/10.1101/309484 (2018).
51. Nakatsuka, T. et al. Impact of histone demethylase KDM3A-dependent AP-1 transactivity on hepatotumorigenesis induced by PI3K activation. *Oncogene* **36**, 6262–6271 (2017).
52. Ding, X. et al. Epigenetic activation of AP1 promotes squamous cell carcinoma metastasis. *Sci. Signal.* **6**, 21–13 (2013). ra28S20-15.
53. Guo, L. et al. Histone demethylase Kdm4b functions as a co-factor of C/EBPbeta to promote mitotic clonal expansion during differentiation of 3T3-L1 pre-adipocytes. *Cell Death Differ.* **19**, 1917–1927 (2012).
54. Choi, J. H., Song, Y. J. & Lee, H. The histone demethylase KDM4B interacts with MyoD to regulate myogenic differentiation in C2C12 myoblast cells. *Biochem. Biophys. Res. Commun.* **456**, 872–878 (2015).
55. Jung, E. S. et al. Jmjd2C increases MyoD transcriptional activity through inhibiting G9a-dependent MyoD degradation. *Biochim. Biophys. Acta* **1849**, 1081–1094 (2015).
56. Fan, L. et al. Regulation of c-Myc expression by the histone demethylase JMJD1A is essential for prostate cancer cell growth and survival. *Oncogene* **35**, 2441–2452 (2016).
57. Zhao, Q. Y. et al. Global histone modification profiling reveals the epigenomic dynamics during malignant transformation in a four-stage breast cancer model. *Clin. Epigenetics* **8**, 34 (2016).
58. Yang, J. et al. The role of histone demethylase KDM4B in Myc signaling in neuroblastoma. *J. Natl Cancer Inst.* **107** (2015).
59. Kauffman, E. C. et al. Role of androgen receptor and associated lysine-demethylase coregulators, LSD1 and JMJD2A, in localized and advanced human bladder cancer. *Mol. Carcinog.* **50**, 931–944 (2011).
60. Coffey, K. et al. The lysine demethylase, KDM4B, is a key molecule in androgen receptor signalling and turnover. *Nucleic Acids Res.* **41**, 4433–4446 (2013).
61. Li, L. et al. JMJD2A-dependent silencing of Sp1 in advanced breast cancer promotes metastasis by downregulation of DIRAS3. *Breast Cancer Res. Treat.* **147**, 487–500 (2014).

Chapter Seven: Conclusions

7.1 Summary of contributions

The dynamics between epigenetic regulators and transcription factors in regulating gene expression and what alterations that may occur to these factors to cause the development of cancer is an important aspect for the field of gene regulation. The first study of this dissertation, the somatic copy number amplification and hyper-activation of the epigenetic factor EZH2 was determined to correspond to epigenetic silencing of genes involved in tumor suppression and the immune response through DNA methylation. The second study explored binding sites of the transcription factor known as the androgen receptor. The third study of histone demethylase KDM3A was found to regulate the transcriptional program of the transcription factor of the androgen receptor. The last study describes a gene regulatory network between the epigenomic master regulator KDM3A and the transcription factors with which it cooperates. The conclusions made from these studies, how they can be applied, and the future directions of study are discussed in the forthcoming section.

7.2 Somatic Copy Number Amplification and Hyperactivating Somatic Mutations of EZH2 Correlate with DNA Methylation and Drive Epigenetic Silencing of Genes Involved in Tumor Suppression and Immune Responses in Melanoma

7.2.1 Summary of results

It was determined prior to this study that the epigenetic modifier EZH2 was involved in silencing tumor suppressor genes across several cancer types, however the genes EZH2 targeted and their role in contributing to melanoma wasn't completely understood. In summary, the information obtained from the 471 TCGA SKCM patients identified that EZH2 was hyper-activated in melanoma through mechanisms of mutations within the gene, amplification of the gene body, and increased transcription for around 20% of these patients suggesting an unfavorable prognosis. The single nucleotide polymorphism DNA arrays from TCGA identified a recurrent mutation occurring at EZH2's SET domain at Y641. Other researchers have noted that the EZH2 Y641 residue is a target of JAK2 phosphorylation, and that resulting missense mutations of this residue results in stabilizing the protein and increasing its histone methylation activity [1, 2]. Therefore it is expected that due to these hyperactivating mutations of EZH2 in the catalytic SET domain of EZH2 in SKCM patients have increased histone methylation and therefore increased gene repression. In addition to mutations affecting EZH2 function, TCGA SKCM datasets identified that the EZH2 loci was one of several genes prone to amplification of the chromosomal band 7q36.1. This molecular event is important due to both having a mutation that results in the hyper-activation of gene activity and having a duplicated gene locus that encourages oncogenic events. Additionally RNAseq analysis of this cohort of patients showed an increase in the transcription of EZH2. Altogether

patients with EZH2 activating mutations, amplification, or high RNA-seq counts showed reduced survival compared to patients with wild-type EZH2.

In addition, adding trimethylation marks to histones EZH2 enables the recruitment of DNA methyltransferases to methylate DNA for further transcriptional repression. This information was used to identify EZH2 target genes from TCGA SKCM patient data by examining the intersection of genes with increased DNA methylation with genes with decreased transcription levels. These candidate genes were further validated as targets of EZH2 regulation by measuring changes in gene expression following the treatment of melanoma cell lines with activated EZH2 with the EZH2 inhibitor GSK126. Following treatment with the inhibitor GSK126 98 genes that were transcriptionally repressed regained gene expression and were associated with tumor suppression, cell differentiation, cell cycle inhibition, repression of metastases, and antigen processing and presentation pathways. EZH2 mediated repression of the transcription factors JUN, JUND, FOS, and FOSB of the AP-1 complex were found. Previous researchers found that in early development for mice EZH2 blocks AP-1 binding in the skin, thus maintaining epidermal progenitors until there is a decline in EZH2 expression resulting in differentiation [3, 4]. This describes a small regulatory circuit between an epigenetic modifier and transcription factors in determining a cell's phenotype. EZH2 treated with GSK126 also rescued the expression of cyclin dependent kinase inhibitor p21 (CDKN1A), an activator of multiple tumor suppressor pathways including cell cycle arrest, differentiation and cellular senescence. This is consistent with previous reports of EZH2 inhibition rescuing CDKN1A expression [2, 5]. Lastly EZH2 was able to repress genes associated with immune responses in patients. Oncogenic processes in cancer cells can assist with their evasion of the immune system making the cancer resistant to immunotherapies [6]. EZH2 achieves this by repressing the expression of major histocompatibility complex genes [7, 8]. Mining of TCGA SKCM datasets for EZH2 responsive genes identified EZH2 as being able to repress CD74, a chaperone protein that associates with MHC for regulating antigen presentation in addition to chemokine receptors [6].

7.2.2 Applying cancer subtyping into the development of precision medicine

TCGA has generated multiple -omic data types for nearly all patient cases: mutations by whole exome sequencing, gene expression by RNA sequencing, microRNA expression by small RNA sequencing, copy number by single nucleotide polymorphism (SNP) arrays, DNA methylation by Illumina arrays, protein expression by reverse phase protein arrays, and noncoding mutations and structural variants by whole genome sequencing [9]. Since its inception TCGA has greatly benefited the research community as a resource to better understand the molecular basis of cancer by identifying significant DNA-level alterations and defining molecular subtypes of cancer. With the large amount of cancer cases involved in TCGA, the significance of identified genes with alterations gain more statistical power [10]. Furthermore categorizing tumors into associated cohorts based on their molecular profiles within TCGA datasets allows for subtypes within the

cancer to be identified without making any major assumptions [11, 12]. Often different mutagenic events happen during the progression of carcinogenesis allowing for different combinations of genetic mutations to occur across patient samples for the same cancer type. Identifying these different combinations of mutagenic events allows researchers to define mutational signatures of a cohort and can provide clues into the specific mechanisms behind the characteristics of that cancer ideally by identifying the altered associated pathways [13]. Chapter Three details how patients within the SKCM dataset were sorted into a subtype based on screenings for activating mutations, copy number gains, and amplification of gene expression. Using this information one-fifth of patients diagnosed with SKCM were identified as having their cancer driven by EZH2 hyper-activation. Defining these subtypes in cancer represents the recognition that cancer may be heterogeneous and may be placed in the context of clinical trials for the development of precision medicine. This philosophy of precision medicine argues that healthcare should be personalized with medical decisions, treatments, practices, or products specific to individual patients. By understanding the molecular forces that drive the cancer phenotype for that specific patient, a personalized treatment plan may yield a better prognostic outcome. This study was able to identify candidate EZH2 target genes from TCGA SKCM datasets and evaluate the impact of GSK126 treatment for rescuing gene expression that was epigenetically silenced by EZH2. This study serves as an example of how scientists are able to identify oncogenes that drive the cancer phenotype, which fraction of the population is affected, and uncovering the cellular effects of reversing the oncogenic signal.

7.2.3 Future perspectives on SKCM cohort studies

In the future new analytical approaches or analysis parameters may be developed to refine the current understanding of the SKCM TCGA cohort where additional subtypes may be found. Currently TCGA data does not focus on patient treatment information in regard to the level of response to targeted therapy. In the future a study could be done to carry out the molecular profiles of patients from clinical trials which could help researchers define molecular signatures and associated pathways within SKCM that are responsive and non-responsive to the treatment. Based on the results of this publication, further research avenues investigating the role of EZH2 in melanoma could look into the effect of EZH2 mediated histone methylation's effect on suppressing gene expression. Another study could look into whether the inhibitors of EZH2 are able to play a role in immunotherapy.

7.3 Refinement of the androgen response element based on ChIP-Seq in androgen-insensitive and androgen responsive prostate cancer cell lines.

7.3.1 Summary of results

The specificity of target response element recognition for the transcription factor AR was improved through the use of ChIP-seq in the androgen-insensitive CWR22Rv1 and androgen responsive LNCaP prostate cancer cell lines. Initial attempts to identify the ARE binding site pattern without the use of existing motif databases yielded an imperfect full ARE site and a perfect half site motif pattern in the CWR22Rv1 ChIP-seq data. The following attempt to identify ARE locations using the position site specific matrix of the AR (MA0007.2) yielded genomic sequences that did not resemble the palindromic ARE sequence described in the literature resulting in the need to design a new position site specific matrix model for ARE site identification.

The combined effort of all the position site specific matrix models designed were able to identify 42956 ARE full and 79065 ARE half site events in CWR22Rv1 cells. These response elements were sorted into tiers and quantified: 71 AREs in tier 1 (perfect), 1583 matches in tier 2 (1bp off perfect), 20362 in tier 3 (2bp off perfect), and 20940 in tier 4 (3bp off perfect) with p-values below 4.94E-05. The motifs showed conservation of G and C in positions two and five respectively in the ARE hexamers. In addition, there was increased GC content in the spacer region of the response element. Lastly, to comprehensively describe the genome wide coverage of AREs, tier 5, focuses on ARE half sites. Tier 5 ARE half sites show a perfect hexamer, while not complying with the requirements of tiers 1 through 4. The functional role of the 120,000 AREs detected by the position site specific matrix model searches were tested against the transcriptional response of AR knockdown in a microarray experiment. In total 759 genes were transcriptionally down regulated, and 743 genes were upregulated upon shRNA AR knockdown. When looking at the fraction of confirmed AREs per tier there is a higher hit rate in ARE tiers closer to the canonical sequence. While 99% of AREs were imperfect, the transcriptional outcome correlated negatively with the degree of degeneracy.

The same model-based ARE annotation was also applied to condition-specific CWR22Rv1 and LNCaP ChIP-Seq samples and assessed ARE utilization dependent on AR splice isoforms as well as 5α-dihydrotestosterone (DHT) treatment. ARE binding by ChIP-Seq was assessed in the CWR22Rv1 cell line for total AR binding, AR(total), for binding to full-length androgen receptor, AR(FL), for binding by variant androgen receptor, AR(V), in the LNCaP cell line for 5α-dihydrotestosterone treatment, AR(DHT), for ethanol treatment, AR(EtOH), and for functionally active steroid-bound AR corrected for EtOH background AR(ACT). Data generated from these experimental conditions showed that the tier 1 "perfect" ARE is a rare event, independent of AR splice-variants or steroid condition. Cross-validation of ChIP-Seq experiments of AR (total) vs AR(FL) confirmed 30,022 AREs assigned to AR(FL)-binding in the CWR22Rv1 cell line (Fig. 5B). AR(V) isoforms with 78,350 ARE ChIP-Seq events bind to DNA autonomous of full-length androgen receptor in the absence of androgen and modulate a unique set of

genes that are not regulated by the full-length androgen receptor. In contrast, AR(DHT) in the LNCaP cell line displayed a set of 7,361 AREs common to the CWR22Rv1 cell line. Next, the quantified fraction of ARE tiers confirmed by overlapping ChIP-Seq experiments. AR(FL) showed an incremental reduced fraction with higher, less specific ARE tiers in the CWR22Rv1 cell line, while AR(V) isoforms had a stronger overlap with more degenerate motifs. The AR(DHT) LNCaP condition showed a trend similar to AR(FL) in the CWR22Rv1 cell line. Notably, the frequency of perfect AREs correlates with AR specificity and increases from AR(V) isoforms to AR(FL) in the CWR22Rv1 cell line and quadruples in the AR(DHT) LNCaP condition. All evaluated conditions showed high agreement utilization of ARE tiers 3–5, half sites and imperfect full sites.

Using the Jasper database, additional transcription factor binding sites were found within a window of ±160 bp from the ARE with a p-value of less than 0.05 were found. Top hits included forkhead box (FOX), Krüppel-like factors (KLF), basic helix-loop-helix (BHLH), sterol regulatory element binding factor (SREBF), and v-Myc avian myelocytomatosis viral oncogene homolog (MYC) families of transcription factors. An increase of detected transcription factor motifs at 15 bp was noticed for KLF, and 45 bp for SREBF-related transcription factors between full and half site ARE. This suggests that for the AR to recognize any weaker half site response element, cooperation of other transcription factors might be required. In order to determine this, the transcriptional response between full and half site ARE with KLF sites were quantified. 17917 ARE full sites coincide with KLF sites. 25 fell into tier 1, 611 into tier 2, 8067 into tier 3, and 9214 into tier 4. When comparing between detected full and half sites there was a stronger cooperation of the ARE full sites with KLF motifs resulting in larger transcriptional response with 360 and 368 genes up and down regulated in contrast to ARE half sites, which had 50 and 33 genes up and down regulated. Despite a larger number of weaker ARE half sites found in the proximity of KLF motifs, stronger AREs next to KLF motifs resulted in a larger transcriptional response.

Next the significance of ARE utilization in the context of somatic copy alterations were determined from prostate adenocarcinoma (PRAD) patients in The Cancer Genome Atlas (TCGA) and in the CWR22Rv1 cell line. Chromosome arms 1q, 3p, 3q, 7p, 7q, 8p, 8q, 12p, and 12q in the CWR22Rv1 cell line showed strong amplifications overlapping with somatic copy number alterations regions identified in TCGA PRAD cohort. Regions with somatic copy number amplifications had significantly enhanced ARE utilization of 60.4 AREs/Mbp in contrast to euploidic genome regions of 32.3 AREs/Mbp with a p-value of 8.9e10-6. The amplified regions contained 5012 genes. From the CWR22Rv1 ChIP-Seq and transcriptomic experiments 214 amplified genes were classified as positive AR targets (down-regulation with shRNA knockdown) and are pathway members of the androgen response, steroid biosynthesis, and cholesterol homeostasis. In addition, amplified and AR-regulated genes showed enrichment in MTORC1 signaling, DNA replication, cell cycle, MYC targets, mismatch repair, homologous recombination, nucleotide excision repair, epigenetic regulators, and pathways in cancer.

The identified 759 and 743 ARE genes confirmed by ChIP-Seq binding as well as transcriptional activity were tested for pathway enrichment by tiers. Gene sets corresponding to full site AREs with activating gene expression (582 genes as positively regulated by AR activity with ARE full or half sites; 102 with exclusively full sites) revealed 41 including 21 exclusive pathways significantly enriched with p-values below 0.05. For half site AREs (585 genes as positively regulated by AR activity with ARE full or half sites; 53 with exclusively full sites) 39 including 12 exclusive pathways were found with p-values below 0.05. Pathways in both sets included DNA replication, cell cycle control, and metabolic pathways. Of particular interest were pathways that were exclusively assigned to ARE full sites or half sites. The set of AR target genes with ARE full sites focused on pyrimidine metabolism, terpenoid backbone biosynthesis, one-carbon metabolism, glycine, serine and threonine metabolism with p-values below 0.001. The metabolic program—framed by genes with full site AREs—supports proliferative functions required for cellular maintenance. In contrast, the set of AR target genes with ARE half sites included steroid biosynthesis, terpenoid backbone biosynthesis, peroxisome, pentose phosphate pathway, glycerolipid metabolism, and mitogen activated protein kinase signaling pathway with p-values below 0.001. An enrichment of genes containing ARE half sites involved in lipid and steroid biosynthesis could point to the gender-, development-, and tissue specific control in prostatic differentiation. Therefore, the gene set enrichment analysis suggests that AR targets genes controlled by ARE full sites and/or ARE half sites have common proliferative functions, but also distinct biological functions involved in lipid metabolism.

7.3.2 Applying response element refinement strategies into motif databases

Transcription factors are a class of proteins conserved across all living organisms that are able to control gene expression by binding to DNA in a sequence-specific manner [14-16]. With the development of next generation sequencing techniques the ability to study transcription factor specificity has changed drastically in the last decade for both in vitro and in vivo studies allowing researchers to gain a broad overview of their binding and sequence specificity [17]. With these advances in sequencing technology researchers have been able to build and refine collections of transcription factor binding sites in databases such as TRANSFAC, JASPAR, UniPROBE, HOCOMOCO, CIS-BP, and Swiss Regulon [18-23]. Even though a great amount of effort has been placed in describing transcription factor response elements in these databases, some of the resulting response element models do not accurately represent the selectivity of transcription factor binding. This may be due to inadequate selectivity of transcription factor response element sequences into their model's design. Recognizing transcription factor response elements in DNA sequences requires a method for identifying canonical sequences within a level of leniency and while all sites identified cannot be functional in regulating gene expression, a transcriptomic experiment measuring the change of expression for a knocked-out gene will identify active response elements in gene regulation.

Furthermore experiments understanding the site-specific binding of transcription factors between cell types can lead to better understanding on cell specific transcriptional programs. Within the context of cancer, the mechanisms by which transcription factors function may differ greatly than in normal tissue. Identifying differences in transcription factor binding events between cancer phenotypes may identify changes in the gene targets of the transcription factor's gene expression program thus changing the cell's proliferative potential.

In cancer, mutagenic events are not limited only to the coding regions of genes, but also in the genomic regions that regulate them. Within the context of this research paper, mutations within the response elements of transcription factors are able to either increase or decrease transcription factor-DNA site recognition [24-33]. This means that degenerate response elements of transcription factors may acquire point mutations that transition them to a lower tier classification. The impact of these mutagenic response element events could result in a change in the gene regulation and expression for both proximal and distally located genes [25, 34-36]. Additionally, some nucleotides in the response element may be more important than the others for achieving DNA binding. Maintaining these residues may be required to retain its ability to act as an active binding site.

Combining transcriptomic experiments to response element analysis allows researchers to understand if its DNA binding results in gene activation or repression and allows for a model of gene regulatory networks to be established for that specific transcription factor. Additionally the genomic regions surrounding the response elements can search for secondary transcription factors that may be complacent in promoting a disease driven phenotype. This information may expand the regulatory network of genes due to the fact that transcription factors may co-localize or compete for DNA binding [37]. As transcription factors are able to operate in complex networks through thousands of bindings sites, crosstalk between these factors can also be applied to these databases [38].

7.3.3 Future perspectives on response element alterations in cancer

During this course of study, a prominent feature of the higher tiered AREs was the increase in the GC content of the response elements three base pair spacer and flanking regions of the response element. Additionally this study identified that there were classes of AREs unique to androgen-insensitive and androgen responsive prostate cancer cell lines. It is possible that due to the high GC content of these higher tiered AREs that the methylation of the CpG dinucleotides are able affect the shape of the response element and prevent AR from recognizing the site. Consequently some studies have revealed that some transcription factors preferentially bind to methylated DNA [39-41]. Investigation into the DNA methylation status of AREs in the androgen-insensitive and androgen responsive prostate cancer cell lines would reveal if this were the mechanism that makes the AREs tissue specific. Alternatively, these androgen-insensitive prostate cancer cell lines specific response elements may be prone to

mutational events that may transition the response elements to a lower tier where they may have increased affinity for the AR.

An additional direction that could be explored is that some transcription factors upon binding to their response elements have the ability to recruit epigenetic factors such as chromatin remodelers, chromatin modifiers, and DNA methyltransferases [42, 43]. Future ChIP-seq studies could look into prominent histone modifications that may accompany detected response element sites. Examining transcriptional coactivation of the AR with epigenetic factors prominent in prostate cancer may provide insight into the chromatin landscape that may accompany AR mediated gene regulation.

Lastly while transcription factors contain DNA binding domains, sometimes transcription factors may be able to bind to DNA indirectly by interacting with another transcription factor [44]. Performing a ChIP analysis for proteins associated with the AR may identify secondary transcription factors and cofactors it interacts with in prostate cancer.

7.4 The histone demethylase KDM3A regulates the transcriptional program of the androgen receptor in prostate cancer cells.

7.4.1 Summary of results

The H3K9 methylation mark is recognized as a hallmark of transcriptionally repressed gene. Previously the histone demethylase KDM3A was able to regulate the expression of genes involved in a variety of biological processes by demethylating mono- or dimethylated H3K9. KDM3A is deregulated in several cancers, and in prostate adenocarcinoma it cooperates with the AR to facilitate gene expression. In order to understand KDM3A's impact on transcriptional regulation, a ChIP-seq experiment, identified KDM3A binding and demethylation events was matched with a microarray experiment in the CWR22Rv1 prostate cancer cell line.

In order to establish the genome wide impact of the epigenetic regulator KDM3A, a matched ChIP-seq experiment using antibodies specific for histone marks H3K9me1 and H3K9me2 in combination with small hairpin RNA (shRNA) knockdown of KDM3A in the CWR22Rv1 cell line was conducted. Histone lysine demethylation (KDM) events mediated by KDM3A were defined by gain of methylation ChIP-Seq signals following knockdown of KDM3A. Overall, the peak counts of both H3K9me1 and H3K9me2 ChIP-Seq experiments showed a gain of signal (32244 to 34162 and 23353 to 46599, respectively). KDM events were functionally annotated by mapping bound regions to the human genome and by classifying them according to the nearest gene locus and relative position within coding regions. In the H3K9me1 ChIP-Seq experiment the intergenic regions were the most frequently found region with 21112 peaks (46.5%) followed by 21495 (47.7%) as intronic regions, 822 (1.8%) as exonic regions, 606 (1.3%) as promoter-TSS regions, 549 (1.2%) as TTS regions, 424 (0.9%) as 3′UTR regions, and 46 (0.1%) as 5′ untranscribed (UTR) regions. Similarly, the H3K9me2 ChIP-Seq had the

intergenic region as its most frequent region with 18195 (55.9%) followed by 13167 (55.9%) as intronic regions, 373 (1.1%) as exonic regions, 355 (1.1%) as promoter-TSS regions, 246 (0.8%) as TTS regions, 204 (0.6%) as 3′UTR regions, and 25 (0.1%) as 5′ UTR regions. These ChIP-Seq profiles following KDM3A knockdown revealed selective histone demethylation effects of this epigenetic modifier.

Following genomic annotation, the detected genes were checked to see if there was a specific gene expression program underlying the demethylation events of H3K9me1 and H3K9me2. Half of the detected genes, 8841 (55.1%), contain both H3K9me1/2 marks. While the ChIP-Seq data shows that H3K9me1 has more annotated genes compared to H3K9me2 (5121 and 2089 respectively), both histone marks showed an equal fraction of genes being transcriptionally responsive to KDM3A knockdown according to the transcriptomic dataset. Overall, from the transcriptomic experiments, 1408 (58.4%; using a significance cutoff with adjusted p-values below 0.05) genes are reported as differentially down regulated upon shRNA knockdown of KDM3A while 1002 genes are reported as up regulated by KDM3A knockdown. The combination of ChIP-Seq histone demethylation events and transcriptomic assessment showed major transcriptional activation by KDM3A, suggesting that KDM3A may synergize with distinct transcriptional regulators for epigenetic control of gene expression.

Following characterization of the H3K9me1 and H3K9me2 marks an analysis into the enrichment of transcriptional motifs associated with the histone marks controlled by KDM3A the CWR22Rv1 prostate cancer cell line was performed. The goal of this analysis is to identify potential transcription factors that cooperate with KDM3A to regulate gene expression. Using the Jaspar motif database, we conducted an unbiased search for significant enrichment of transcription factor families (analysis of motif enrichment search with p-values below 0.05). Top hits included the androgen receptor, sterol regulatory element binding factor (SREBF), hypoxia inducible factor (HIF), activator protein 1 (AP1) complex of JUN/FOS, Krüppel-like factors (KLF), v-Myc avian myelocytomatosis viral oncogene homolog (MYC), and forkhead box (FOX) families of transcription factors with significant enrichments and p-values below 1.0E-04. In addition, enhanced simple ChIP-Seq based searches with position-specific matrices to determine which transcription factor motifs were enriched compared to shuffled background sequences. The enrichment analysis showed the androgen response element (ARE) with 2915 incidences as one of the most frequent motifs detected. At the transcriptional level, the androgen response gene set was the most enriched with a p-value and false discovery rate q-value each below 1.0E-20. Transcriptional regulators were inferred by comparing transcriptional targets to datasets that outline targets of transcription factors through the use of Ingenuity Pathway Analysis. The transcription factors AR, HIF, MYC, and AP1 complex were significantly enriched with p-values below 1.0E-07. Lastly, by merging the ChIP-Seq profiles of H3K9Me1/2 and KDM3A transcriptional data focusing on 1408 annotated genes (overlap of H3K9ME1/2 ChIP-Seq with transcriptomic data that were down regulated upon shRNA KDM3A knockdown) contained the highest enrichment ratio (26.7%) in a significantly enriched set of 27 genes

in androgen signaling with p-values below 1.0E-17 and q-values below 1.0E-15. In detail, putative KDM3A-regulated genes included pathways involved in androgen response, androgen receptor signaling, androgen biosynthesis, prostate cancer, pathways in cancer, cholesterol homeostasis, bile acid metabolism, aldosterone-regulated reabsorption, and progesterone regulation, hinting at the possibility of hormone nuclear steroid receptor involvement. Interestingly for the concept of cooperative control, the pathways of regulation and coregulation of androgen receptor activity were also enriched with p-values and q-values below 4.04E-06 and 1.97E-04, respectively. SLC26A2, FKBP5, KRT19, SORD, HOMER2, NDRG1, TPD52, INPP4B, PTPN21, ZMIZ1, PMEPA1, PPAP2A, TSC22D1, ACSL3, KLK3, NKX3-1, ELL2, MAP7, PTK2B, SMS, SPDEF, ABCC4, KLK2, MAF, TARP, AZGP1, and TMPRSS2 were key regulators of prostate cancer and AR signaling based on the KDM3A ChIP-Seq data. Transcriptional control of key players of cancer and AR signaling by KDM3A such as KLK3, NKX3-1, MYC were validated by chromatin immunoprecipitation coupled with quantitative real time polymerase chain reaction (ChIP-qRT-PCR). Taken together, complementary analyses identified strong transcriptional networks including AR, MYC, FOX, KLF, AP1, and SREBF transcription factors that may be regulated by KDM3A. Androgen signaling was consistently identified by all of these different enrichment approaches, suggesting a key role for KDM3A in regulating AR activity.

Knockdown of KDM3A in CWR22Rv1 cells resulted in loss of KDM3A ChIP-Seq binding accompanied by specific, matched gain of histone lysine 9 demethylation. Knockdown of KDM3A had little effect on the protein level of AR. Investigation into the alteration of AR binding by ChIP-Seq with an AR antibody following KDM3A knockdown and quantifying the overlap of AR ChIP-Seq events with KDM3A binding and changes in epigenetic H3K9me1 and H3K9me2 marks resulted in 37525 peaks associated with KDM3A binding, 45246 and 32665 H3K9 mono- and di-demethylation (H3K9me1 and H3K9me2 -KDM) events, respectively, and 34614 peaks for KDM3A-matched AR binding. Overall 37.0% of the AR ChIP-Seq peaks with altered H3K9me1 and H3K9me2 signal were suppressed upon knockdown of KDM3A, while the remaining fraction was not affected. KDM3A ChIP-Seq and H3K9me1/2-KDM ChIP-Seq in combination with matched knockdown of KDM3A produced an epigenetic network that overlaid with the AR ChIP-Seq data. In the case of matched and merged datasets of AR ChIP-Seq in combination with KDM3A knockdown, a total of 77911 H3K9me peaks were identified and directly overlaid them with 34614 AR peaks containing 121700 ARE motifs. Importantly, using such matched ChIP-Seq analyses, a set of 1912 genes were identified that showed an overlap of demethylation and AR binding events. Epigenetic events identified by ChIP-Seq were overlaid with transcriptomic data, defining a set of 421 genes that had epigenetic marks (H3K9me1 and H3K9me2 ChIP), AR binding (AR ChIP), and a transcriptomic effect (differential expression in either KDM3A or AR knockdown experiments). Similar to the initial gene set based exclusively on H3K9me1 and H3K9me2 marks (57.2%), 60.2% of genes scored as activated upon KDM3A/AR coactivation while 39.8% were silenced. Merged ChIP-Seq data of KDM3A/AR

coactivation with transcriptomic data of KDM3A knockdown defined a set of 260 genes (whereas activation is interpreted as down-regulation upon shRNA knockdown). Gene set enrichment analysis was used to identify signaling networks or functional clusters of genes controlled by both KDM3A and AR. Pathway enrichment for the following three gene sets were performed: A) 1912 genes defined by overlapping H3K9 demethylation/AR ChIP; B) 421 genes defined by overlapping H3K9 demethylation/AR ChIP with differential expression upon KDM3A/AR knockdown; and C) 260 genes that are down regulated upon KDM3A or AR knockdown and having overlapping H3K9 demethylation/AR ChIP. The gene sets were designed in a hierarchical fashion such that the parental set A of 1912 genes includes subset B of 421 genes, and B includes subset C of 260 genes. The gene set enrichment analysis revealed oncogenic activation of androgen signaling and metabolic pathways with p-values below 1.40E-02 and q-values below 4.44E-02 in hypoxic response, glycolysis, and lipogenic metabolism. Enrichments of androgen response, metabolic pathways, hypoxia, aldosterone-regulated sodium reabsorption, glycolytic and glycerophospholipid metabolism pathways were conserved and enriched from the overlapping H3K9 demethylation/AR ChIP (1912 genes) to the genes displaying down-regulation upon KDM3A/AR knockdown and overlapping H3K9 demethylation/AR ChIP signal (260 genes). For example, androgen response and metabolic pathways showed incrementally higher enrichment with better defined input gene sets (significant enrichment of genes in androgen response with p-value of 7.51E-07 for 1912 gene set and decreased p-value of 5.23E-09 for 260 gene set). Similarly, pathways of glycolytic and glycerophospholipid metabolism showed consistently higher enrichment with the lowest p-values in the set of 260 genes, indicating that KDM3A demethylation targets these pathways and causes an up regulation of gene expression. Of all genes detected by H3K9 demethylation and AR ChIP-Seq experiments as well as KDM3A and AR differential expression following knockdown (core set of 45 genes; 28 genes up regulated; 17 genes down regulated), the AR response was the only significantly enriched pathway with a p-value of 5.29E-08. 7 genes, NDRG1, PTK2B, ACSL3, KRT19, INPP4B, NKX3-1, and MAF, with direct implication in AR signaling were present in all gene sets of the hierarchical enrichment analyses, thereby subject to KDM3A control by H3K9 demethylation, AR binding, and differential expression following knockdown. In the androgen response pathway KDM3A and AR had strong ChIP-Seq activity. The majority of epigenetic H3K9 demethylation events resulted in up regulation of target gene activity and occurred at the TSSs of target genes. In addition, multiple introns, exons, and TTS of target genes were also implicated in KDM3A control. A network upstream regulator analysis implicated the AR to coordinate with epigenetic and transcriptional responses observed upon knockdown of KDM3A. At the transcriptional level, KDM3A may affect AR signaling directly by interacting with regulatory regions of HSP90AA1 and AR genes. In addition, downstream effects of AR signaling were observed at KRT19, NKX3-1, KLK3, TMPRSS2, PMEPA1, NDRG1, MAF, CREB3L4, MYC, INPP4B, PTK2B, MAPK1, MAP2K1, IGF1, E2F1, HIF1A,

TARP, FKBP5, SPDEF, SMS, PPAP2A, SEPP1, UAP1, SORD, AZGP1, BCL2L1, ACSL3, CHUK, and CDKN1A.

7.4.2 Applying multiomic strategies to the regulation of transcription

Applying the biological analysis approach of "multiomics" where datasets from experiments studying different aspects of the same model organism are able to characterize a biological process through any sort of combination of experiments that are able to describe an organism's genome, proteome, transcriptome, epigenome, and metabolite [45, 46]. The development of whole -omic technologies allows for researchers to study multiple properties of gene expression for multiple genes quantitatively and simultaneously [47-51]. As a result different -omic data types may correlate with one another and may generate a regulatory hypotheses between the experiments. For example, within this study epigenomic histone marks were matched with identified transcription factor binding sites following knockout of the epigenetic master regulator allowed researchers to identify genome wide cooperation of epigenetic remodeling with other members of the transcriptional machinery. By combining different levels of experimental data researchers are able to search through complex biological systems to find relevant biomarkers with more confidence in their findings. For example the epigenetic modifier KDM3A was found to regulate a core set of genes within the androgen response pathway through the utilization of multiple ChIP-seq and a microarray analysis.

By using multiomic data sets researchers are able to model whole properties of cancer cells using a holistic approach to understand how different parts of a biological system are able to interact with one another to influence the function and behavior of that system [52]. For example, the simultaneous profiling of the epigenome and transcriptome of a cancer cell grants researchers a unique chance to directly measure the effect of differences in histone modifications has upon gene transcription [53].

An advantage of a multiomic approach is that researchers are able to focus on the causes and effects of a biological network better by utilizing quantitative procedures simultaneously and by integrating their results to model the system being studied [54]. In the advent of high throughput experimental techniques being developed researchers must be prepared to ask themselves during the course of progressing their research project if there is another way to integrate another theory, analysis, or model in order to acquire new quantitative data to better describe and validate their biological system [55]. Hopefully, as more researchers begin to adopt the approach of multiomic data set analysis, new methodologies and bioinformatic analyses will be developed as will the collection of multiomic data sets being generated allowing for better insight into biological systems.

7.4.3 Future Perspectives on multiomic analysis of KDM3A

In the future multiomic methodology will be commonplace for cancer biology researchers to use in understanding the relationship between DNA mutations with epigenetic modifications and their ability to regulate gene expression. Over time multiomic analysis will begin to shift from an analysis of a bulk population of cells that generate an average signal to a single cell population that allows researchers to have better resolution of the results of their experiment [47, 48, 56]. The development of single cell multiomic methodologies will generate new applications such as depicting cellular diversity, tracing cellular lineage, improved cell type detection, and understanding regulatory mechanisms between -omic levels [53].

Although the KDM3A protein regulates a wide array of target genes in tissue- and development-specific settings, chromatin modifiers often lack intrinsic DNA sequence specificity. Therefore, how KDM3A is targeted to specific genes is an area of current research interest and important for understanding epigenetic dysregulation in human disease. The functional impact of coordinated action between a lysine demethylase and transcription factors may lend to its target specificity, or at the very least, create accessibility for DNA binding. Future experimental verification of epigenetic hotspots is needed to determine when detected response elements are functional in gene regulation.

7.5 A network of epigenomic and transcriptional cooperation encompassing an epigenomic master regulator in cancer.

7.5.1 Summary of results

In epigenetic profiling experiments an enrichment of several transcription factors can be identified. Cooperation between these two classes of proteins are able to regulate gene expression by having one component open the chromatin, another one recognizing gene-specific DNA motifs, others scaffolding between histones, cofactors, and the transcriptional complex. Building a network of interactions between these proteins in regulating gene expression is a main goal of systems biology [7, 57, 58]. This publication showcases the impact of knocking down the histone lysine demethylase KDM3A on a transcriptional network using coordinated ChIP-Seq and transcriptomic data of KDM3A binding and demethylation activity. Upon KDM3A knockdown a total of 2460 genes were identified as positively regulated by KDM3A activity, and 1866 genes as being negatively regulated by KDM3A activity. Genomic KDM3A binding locations were defined by a loss of ChIP-Seq signal following knockdown of KDM3A. Additionally H3K9me1 and H3K9me2 histone marks following KDM3A knockdown were recognized as histone lysine demethylation targets of KDM3A. The ChIP-Seq experiment yielded 37525 peaks associated with KDM3A binding, 45246 and 32665 H3K9 mono- and di-demethylation (H3K9me1 and H3K9me2 -KDM) events. These matching epigenomics and transcriptomics experiments describe the genome wide impact of knocking down KDM3A as well as its resulting transcriptional outcome.

Candidate transcription factors were identified from ChIP-Seq data using an analysis of motif enrichment (AME) and Transcription factor target (TFT) analysis, while candidate transcription factors were identified from differentially expressed genes using an upstream regulator analysis (URA) and Transcription factor target (TFT) analysis techniques. Each analysis identified between 29 and 41 significant transcription factor families, each corresponding up to both 1083 and 1292 associated genes. Despite the impressive p-values generated by some of these analysis tools, specific transcription factor identification cannot be established. An example of this is the hypoxia inducible factor 1 alpha subunit (HIF1A) of the PAS domain factor family is detected with an adjusted p-value below 1.0E-100 by AME in the ChIP-Seq data similarly to other members of the same family like the aryl hydrocarbon receptor nuclear translocator (ARNT) due to a similarity of their recognition sequence. Therefore by intersecting the datasets of AME ChIP-Seq, TFT ChIP-Seq, URA transcriptomics, and TFT transcriptomics an inclusive list of 21 transcription factor families supported by all four datasets are generated. From these 21 identified transcription factor families a total of 967 transcription factors were implicated. Almost half of these transcription factors listed are due to the large family of the more than 3 adjacent zinc finger factors family that contains more than 487 transcription factor members. After building this network the transcriptional impact of KDM3A knock down onto the gene targets of transcription factors were explored, and the different size of intersection of transcriptional targets between KDM3A and the transcription factor could be explored. There were only a few genes of the network that were hyper connected with multiple transcription factors in epigenomics and transcriptomics datasets.

Review of transcriptomic data by URA and TFT analyses utilize gene set databases derived from previous experimental data of gene-specific knockdowns describes the effect of a transcription factor on its gene targets. These gene-specific analyses allowed for ambiguities in the transcription factor family network resulting in 11 transcription factors belonging to 10 transcription factor families being identified: AP-1 transcription factor subunit JUN, CCAAT/enhancer binding protein beta (CEBPB), myogenic differentiation 1 (MYOD1), HIF1A, sterol regulatory element binding transcription factor 1 (SREBF1), MYC, AR, Sp1, MEIS homeobox 1 (MEIS1), zinc finger E-box binding homeobox 1 (ZEB1), and ELK1. These transcription factors were previously reported to be associated and implicated in chromatin remodeling. By overlaying the motif-specific and genomic data produced through matching experiments, epigenomic events can be correlated with the transcriptomic effect of histone remodelers and transcription factors.

7.5.2 Applying interactomics for studying biological pathways

In the histone code, specific histone modifications can make the difference for chromatin to exist in open or closed states resulting in either transcriptional activation or repression of biological processes. By investigating interrelated epigenomic and transcriptomic datasets a tight partnership between transcriptional and epigenomic

machinery in regulating key biological processes can be made known. By defining a network of transcription factors that cooperate with the histone modifier KDM3A a small chunk of the interactome was described. The interactome is a facet of molecular biology where a whole set of molecular interactions that take place within a cell is described. Although typically used to describe the physical interactions between molecules such as a protein-protein interaction, the interactome can be applied to describing the indirect interactions among genes [59]. Of particular interest this study served as an example of using systems biology to describe a gene regulatory network built by the intersection of transcription factors, histone modifiers, and their target genes. The assembly of gene networks serve to help researchers better understand the behavior of pathways in the cell and is a step forward to fully mapping out the complexities of gene interactions at a cellular level [60].

As a general rule of thumb, the size of an organism's interactome correlates with the complexity of the biological organism [61]. It is impossible to completely describe a gene regulatory network as limitations of experimental methods will never fully capture every single interaction that occurs; additionally, some modules within a network can also be connected to specialized subnetworks [62]. For example the complete interactome of a cell contains biochemical networks, gene regulatory networks, protein interaction networks, and signaling networks. It is however the goal of interactomic studies to recognize and build these gene regulatory networks to the best of a researcher's ability to identify the cellular processes being regulated, identify potential biomarkers for diseases, and understand the consequences drug targeting may have on the network [63, 64]. Generating these gene regulatory networks to help design and understand the effect of pharmaceutical drugs upon a target network can lead to an evolution in researcher's understanding into the mode of action the drug has on an organism [65, 66].

An additional application of studying the interactome is studying the changes in the arrangement of a network due to alterations made in the proteins or how protein expression [67]. Looking at a gene regulatory network gives researchers a perspective of how a disease can be manifested from the network by the mutagenic events that alter the landscape of the interactome [68]. When building a gene regulatory network it is possible to identify nodes that are able to regulate themselves directly or indirectly in a feedback loop. Loss of control of any of these nodes due to mutagenic events may disrupt a feedback loop and can be responsible for shifting a cell into a diseased phenotype. By mapping out the complexities behind gene regulation, metabolic reactions, and protein-protein interactions researchers hope to gain insight into the regulatory mechanisms behind diseases. This methodology can be used to compare how metabolic proteins and their respective metabolites are altered from their normal metabolic network to a new diseased network [69]. A broader application of this methodology would be comparing the interactomes of deregulated gene regulatory networks across a variety of cancers to better classify and analyze the genetic relationships between them [69]. By mapping out gene regulatory networks from present -omic data sets allows researchers to infer new hypotheses that can then be tested by new experiments [70]. Thus investigating gene

regulatory networks of the cooperative interactions of chromatin remodelers and the transcriptional machinery will allow researchers to better understand mechanisms of determining cell identity, regulating chromatin, and designing drug targets.

7.5.3 Future Perspectives on the gene regulatory network of KDM3A

The ability to build the gene regulatory network of KDM3A and its cooperating transcription factors was successful in part due to the efforts of curated databases of transcriptional targets of various transcription factors derived from the datasets of other researchers. Expanding existing datasets describing gene targets of transcription factors and creating another database outlining gene target of epigenetic factors may serve to help describe the gene regulatory landscape of their model organism. Development of a database listing gene targets of epigenetic factors would allow researchers to develop an inverse gene regulatory network described in this paper: a network of epigenetic factors cooperating with a transcription factor across the genome.

Additional future directions could investigate the transcription factors identified in the KDM3A coactivation network further by surveying the basal expression, and somatic loss or gain of function of these factors across cancer types using TCGA. This could also prove beneficial for identifying their contributing oncogenic potential and biological relevance in cancer patients followed up by a study to describe the draggability of KDM3A or its transcription factors with various small molecule inhibitors. Lastly, the ability for KDM3A to localize to its genomic targets through transcription factors may be evaluated through loss of a transcription factor expression and its impact on the cell.

7.6 Conclusions

Cumulatively these results highlight how the different levels of gene regulation are investigated: from the classification of a cancer's cohort based on the hyper activation of an epigenetic regulator, altered specificity of transcription factor site recognition, the ability of an epigenetic factor to regulate the signaling pathway of a transcription factor, and the development of a gene regulatory network between an epigenetic regulator and the transcription factors with which it cooperates. Further development of genomic and transcriptomic techniques combined with the implementation of other -omic techniques will allow biologists to fully characterize the interactome of cellular states and processes and describe how their alterations lead to the development of diseased states for cells. Integration of patient data to cell line models for multiomic analysis will allow researchers to describe the molecular interactions that occur in the cell, how they drive cancer progression, and hopefully develop precise treatment options for patients.

7.7 References

1. Antonysamy, S., et al., Structural context of disease-associated mutations and putative mechanism of autoinhibition revealed by X-ray crystallographic analysis of the EZH2-SET domain. PLoS One, 2013. 8(12): p. e84147.
2. Sahasrabuddhe, A.A., et al., Oncogenic Y641 mutations in EZH2 prevent Jak2/beta-TrCP-mediated degradation. Oncogene, 2015. 34(4): p. 445-54.
3. Eferl, R. and E.F. Wagner, AP-1: a double-edged sword in tumorigenesis. Nat Rev Cancer, 2003. 3(11): p. 859-68.
4. Ezhkova, E., et al., Ezh2 orchestrates gene expression for the stepwise differentiation of tissue-specific stem cells. Cell, 2009. 136(6): p. 1122-35.
5. Fan, T., et al., EZH2-dependent suppression of a cellular senescence phenotype in melanoma cells by inhibition of p21/CDKN1A expression. Mol Cancer Res, 2011. 9(4): p. 418-29.
6. Spranger, S., R. Bao, and T.F. Gajewski, Melanoma-intrinsic beta-catenin signalling prevents anti-tumour immunity. Nature, 2015. 523(7559): p. 231-5.
7. Abou El Hassan, M., et al., Cancer Cells Hijack PRC2 to Modify Multiple Cytokine Pathways. PLoS One, 2015. 10(6): p. e0126466.
8. Holling, T.M., et al., A role for EZH2 in silencing of IFN-gamma inducible MHC2TA transcription in uveal melanoma. J Immunol, 2007. 179(8): p. 5317-25.
9. Creighton, C.J., The clinical applications of The Cancer Genome Atlas project for bladder cancer. Expert Rev Anticancer Ther, 2018. 18(10): p. 973-980.
10. Lawrence, M.S., et al., Discovery and saturation analysis of cancer genes across 21 tumour types. Nature, 2014. 505(7484): p. 495-501.
11. Eisen, M.B., et al., Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci U S A, 1998. 95(25): p. 14863-8.
12. Perou, C.M., et al., Molecular portraits of human breast tumours. Nature, 2000. 406(6797): p. 747-52.
13. Alexandrov, L.B., et al., Signatures of mutational processes in human cancer. Nature, 2013. 500(7463): p. 415-21.
14. Lambert, S.A., et al., The Human Transcription Factors. Cell, 2018. 172(4): p. 650-665.
15. Paillard, G. and R. Lavery, Analyzing protein-DNA recognition mechanisms. Structure, 2004. 12(1): p. 113-22.
16. Rohs, R., et al., Origins of specificity in protein-DNA recognition. Annu Rev Biochem, 2010. 79: p. 233-69.
17. Koboldt, D.C., et al., The next-generation sequencing revolution and its impact on genomics. Cell, 2013. 155(1): p. 27-38.
18. Hume, M.A., et al., UniPROBE, update 2015: new tools and content for the online database of protein-binding microarray data on protein-DNA interactions. Nucleic Acids Res, 2015. 43(Database issue): p. D117-22.

19. Khan, A., et al., JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. Nucleic Acids Res, 2018. 46(D1): p. D260-D266.
20. Kulakovskiy, I.V., et al., HOCOMOCO: a comprehensive collection of human transcription factor binding sites models. Nucleic Acids Res, 2013. 41(Database issue): p. D195-202.
21. Matys, V., et al., TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. Nucleic Acids Res, 2006. 34(Database issue): p. D108-10.
22. Pachkov, M., et al., SwissRegulon, a database of genome-wide annotations of regulatory sites: recent updates. Nucleic Acids Res, 2013. 41(Database issue): p. D214-20.
23. Weirauch, M.T., et al., Determination and inference of eukaryotic transcription factor sequence specificity. Cell, 2014. 158(6): p. 1431-1443.
24. Al Zadjali, S., et al., The beta-globin promoter -71 C>T mutation is a beta+ thalassemic allele. Eur J Haematol, 2011. 87(5): p. 457-60.
25. Claussnitzer, M., et al., Leveraging cross-species transcription factor binding site patterns: from diabetes risk loci to disease mechanisms. Cell, 2014. 156(1-2): p. 343-58.
26. Consortium, E.P., An integrated encyclopedia of DNA elements in the human genome. Nature, 2012. 489(7414): p. 57-74.
27. De Gobbi, M., et al., A regulatory SNP causes a human genetic disease by creating a new transcriptional promoter. Science, 2006. 312(5777): p. 1215-7.
28. Jeong, Y., et al., Regulation of a remote Shh forebrain enhancer by the Six3 homeoprotein. Nat Genet, 2008. 40(11): p. 1348-53.
29. Kulzer, J.R., et al., A common functional regulatory variant at a type 2 diabetes locus upregulates ARAP1 expression in the pancreatic beta cell. Am J Hum Genet, 2014. 94(2): p. 186-97.
30. Martin, D.I., S.F. Tsai, and S.H. Orkin, Increased gamma-globin expression in a nondeletion HPFH mediated by an erythroid-specific DNA-binding factor. Nature, 1989. 338(6214): p. 435-8.
31. Matsuda, M., N. Sakamoto, and Y. Fukumaki, Delta-thalassemia caused by disruption of the site for an erythroid-specific transcription factor, GATA-1, in the delta-globin gene promoter. Blood, 1992. 80(5): p. 1347-51.
32. Musunuru, K., et al., From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus. Nature, 2010. 466(7307): p. 714-9.
33. Weinhold, N., et al., Genome-wide analysis of noncoding regulatory mutations in cancer. Nat Genet, 2014. 46(11): p. 1160-5.
34. Claussnitzer, M., et al., FTO Obesity Variant Circuitry and Adipocyte Browning in Humans. N Engl J Med, 2015. 373(10): p. 895-907.
35. Mattei, R. and E.A. Carlini, Mazindol: anorectic and behavioral effects in female rats. Arch Int Pharmacodyn Ther, 1995. 330(3): p. 279-87.
36. Smemo, S., et al., Obesity-associated variants within FTO form long-range functional connections with IRX3. Nature, 2014. 507(7492): p. 371-5.

37. Wang, J., et al., Sequence features and chromatin structure around the genomic regions bound by 119 human transcription factors. Genome Res, 2012. 22(9): p. 1798-812.
38. Siersbaek, R., et al., Transcription factor cooperativity in early adipogenic hotspots and super-enhancers. Cell Rep, 2014. 7(5): p. 1443-1455.
39. Yin, Y., et al., Impact of cytosine methylation on DNA binding specificities of human transcription factors. Science, 2017. 356(6337).
40. Zhu, H., G. Wang, and J. Qian, Transcription factors as readers and effectors of DNA methylation. Nat Rev Genet, 2016. 17(9): p. 551-65.
41. Zuo, Z., et al., Measuring quantitative effects of methylation on transcription factor-DNA binding affinity. Sci Adv, 2017. 3(11): p. eaao1799.
42. Spitz, F. and E.E. Furlong, Transcription factors: from enhancer binding to developmental control. Nat Rev Genet, 2012. 13(9): p. 613-26.
43. Voss, T.C. and G.L. Hager, Dynamic regulation of transcriptional states by chromatin and transcription factors. Nat Rev Genet, 2014. 15(2): p. 69-81.
44. Burda, P., P. Laslo, and T. Stopka, The role of PU.1 and GATA-1 transcription factors during normal and leukemogenic hematopoiesis. Leukemia, 2010. 24(7): p. 1249-57.
45. Bersanelli, M., et al., Methods for the integration of multi-omics data: mathematical aspects. BMC Bioinformatics, 2016. 17 Suppl 2: p. 15.
46. Bock, C., M. Farlik, and N.C. Sheffield, Multi-Omics of Single Cells: Strategies and Applications. Trends Biotechnol, 2016. 34(8): p. 605-608.
47. Gawad, C., W. Koh, and S.R. Quake, Single-cell genome sequencing: current state of the science. Nat Rev Genet, 2016. 17(3): p. 175-88.
48. Huang, L., et al., Single-Cell Whole-Genome Amplification and Sequencing: Methodology and Applications. Annu Rev Genomics Hum Genet, 2015. 16: p. 79-102.
49. Tang, F., et al., mRNA-Seq whole-transcriptome analysis of a single cell. Nat Methods, 2009. 6(5): p. 377-82.
50. Wang, Y. and N.E. Navin, Advances and applications of single-cell sequencing technologies. Mol Cell, 2015. 58(4): p. 598-609.
51. Zong, C., et al., Genome-wide detection of single-nucleotide and copy-number variations of a single human cell. Science, 2012. 338(6114): p. 1622-6.
52. Tavassoly, I., J. Goldfarb, and R. Iyengar, Systems biology primer: the basic methods and approaches. Essays Biochem, 2018. 62(4): p. 487-500.
53. Hu, Y., et al., Single Cell Multi-Omics Technology: Methodology and Application. Front Cell Dev Biol, 2018. 6: p. 28.
54. Sauer, U., M. Heinemann, and N. Zamboni, Genetics. Getting closer to the whole picture. Science, 2007. 316(5824): p. 550-1.
55. Baitaluk, M., System biology of gene regulation. Methods Mol Biol, 2009. 569: p. 55-87.
56. Navin, N., et al., Tumour evolution inferred by single-cell sequencing. Nature, 2011. 472(7341): p. 90-4.

57. Filipp, F.V., Cancer metabolism meets systems biology: Pyruvate kinase isoform PKM2 is a metabolic master regulator. J Carcinog, 2013. 12: p. 14.
58. Filipp, F.V., A Gateway between Omics Data and Systems Biology. J Metabolomics Syst Biol, 2013. 1(1): p. 1.
59. Alonso-Lopez, D., et al., APID interactomes: providing proteome-based interactomes with controlled quality for multiple species and derived networks. Nucleic Acids Res, 2016. 44(W1): p. W529-35.
60. Azpeitia, E., et al., The combination of the functionalities of feedback circuits is determinant for the attractors' number and size in pathway-like Boolean networks. Sci Rep, 2017. 7: p. 42023.
61. Stumpf, M.P., et al., Estimating the size of the human interactome. Proc Natl Acad Sci U S A, 2008. 105(19): p. 6959-64.
62. Gao, L., P.G. Sun, and J. Song, Clustering algorithms for detecting functional modules in protein interaction networks. J Bioinform Comput Biol, 2009. 7(1): p. 217-42.
63. Barabasi, A.L., N. Gulbahce, and J. Loscalzo, Network medicine: a network-based approach to human disease. Nat Rev Genet, 2011. 12(1): p. 56-68.
64. Kiemer, L. and G. Cesareni, Comparative interactomics: comparing apples and pears? Trends Biotechnol, 2007. 25(10): p. 448-54.
65. Hopkins, A.L., Network pharmacology: the next paradigm in drug discovery. Nat Chem Biol, 2008. 4(11): p. 682-90.
66. Yildirim, M.A., et al., Drug-target network. Nat Biotechnol, 2007. 25(10): p. 1119-26.
67. Barabasi, A.L. and Z.N. Oltvai, Network biology: understanding the cell's functional organization. Nat Rev Genet, 2004. 5(2): p. 101-13.
68. Goh, K.I. and I.G. Choi, Exploring the human diseasome: the human disease network. Brief Funct Genomics, 2012. 11(6): p. 533-42.
69. Braun, P., E. Rietman, and M. Vidal, Networking metabolites and diseases. Proc Natl Acad Sci U S A, 2008. 105(29): p. 9849-50.
70. Bruggeman, F.J. and H.V. Westerhoff, The nature of systems biology. Trends Microbiol, 2007. 15(1): p. 45-50.