GENOME-WIDE ANALYSIS OF EPIGENETICS AND ALTERNATIVE PROMOTERS IN CANCER CELLS

DISSERTATION

Presented in Partial Fulfillment of the Requirements for

the Degree Doctor of Philosophy in the Graduate

School of The Ohio State University

By

Jiejun Wu, M.D. & M.S.

*****

The Ohio State University

2007

Dissertation Committee:

Dr. Christoph Plass, Adviser

Dr. Tim H.-M. Huang, Adviser

Dr. Donald Harry Dean

Dr. Amanda Simcox

Dr. Huey-Jen Lin

Approved by:

_____

Adviser
Graduate Program in Molecular Genetics

ABSTRACT

Genome-wide approaches, such as ChIP-chip, have been widely applied to explore the patterns of epigenetic markers and the interactions between DNA and proteins. Compared to candidate gene studies, the application of epigenomic and genomic tools in these fields provides more comprehensive understanding of normal and abnormal events in cells, such as those biological changes promoting cancer development.

In the first part, I studied the relations between two well-known epigenetic markers, DNA methylation and histone modifications. Previous studies of individual genes have shown that in a self-enforcing way, dimethylation at histone 3 lysine 9 (dimethyl-H3K9) and DNA methylation cooperate to maintain a repressive mode of inactive genes. Less clear is whether this cooperation is generalized in mammalian genomes, such as the mouse genome. Here I use epigenomic tools to simultaneously interrogate chromatin modifications and DNA methylation in a mouse leukemia cell line, L1210. Histone modifications on H3K9 and DNA methylation in L1210 were profiled by both global CpG island array and custom mouse promoter array analysis. I used chromatin immunoprecipitation microarray (ChIP-chip) to examine acetyl-H3K9 and dimethyl-H3K9. I found that the relative level of acetyl-H3K9 at different chromatin positions has a wider

range of distribution than that of dimethyl-H3K9. I then used differential methylation hybridization (DMH) and restriction landmark genome scanning (RLGS) to analyze the DNA methylation status of the same targets investigated by ChIP-chip. The results of epigenomic profiling, which have been independently confirmed for individual loci, show an inverse relationship between DNA methylation and histone acetylation in regulating gene silencing. In contrast to the previous notion, dimethyl-H3K9 seems to be less distinct in specifying silencing for the genes tested. This study demonstrates in L1210 leukemia cells a diverse relationship exists between histone modifications and DNA methylation in the maintenance of gene silencing. Acetyl-H3K9 shows an inverse relationship between DNA methylation and histone acetylation in regulating gene silencing as expected. However, dimethyl-H3K9 seems to be less distinct in relation to promoter methylation. Meanwhile, a combination of epigenomic tools is of help in understanding the heterogeneity of epigenetic regulation, which may further our vision accumulated from single-gene studies.

In the second part, I profiled the multiple promoter usage in breast cancer cells. Various independent lines of evidence have suggested that a large fraction of human genes have multiple independently regulated promoters with distinct transcription start sites. Understanding which promoter is employed in which cellular condition is key to unraveling gene regulatory networks within the cell. To

this end, we have designed a custom microarray platform that tiles roughly 35,000 alternative putative promoters from nearly 7,000 genes in the human genome. To demonstrate the utility of this platform, I have analyzed the pattern of promoter usage in E2-treated and untreated MCF7 cells and show widespread usage of alternative promoters. Most intriguingly, I show that the downstream promoter in E2-sensitive multiple promoter genes tends to be very close to the 3'-terminus of the gene sequence, suggesting exotic mechanisms of expression regulation in these genes.

Taken together, I showed genome-wide diverse relations between DNA methylation and histone modifications in leukemia cell line L1210. I also characterized the alternative promoter usage in breast cell line MCF-7.

Dedicated to my family

ACKNOWLEDGMENTS

I would like to thank my advisors, Drs. Christoph Plass and Tim H.-M. Huang, for providing me the opportunity to work in such an exciting research field. I wish to thank them for sharing with me their knowledge, experience and expertise selflessly. The research projects put forth within this thesis could not have been accomplished without their mentorship and direction. They also helped me learn, think and work independently, which, I believe, will benefit me greatly for the rest of my life.

My graduate program committee, specifically Drs. Donald Dean and Amanda Simcox, provided encouraging and continued trust on my research. I would like to thank them for their support and insightful remarks.

I also wish to thank all the members of Dr. Plass and Dr. Huang's laboratories for their unselfish help and friendship, which made every day much more comfortable. Special thanks go to Yue-Zhong Wu, Chunhui Liu, Joseph Liu, Drs. Shu-Huei Wang and Li Yu, who helped me with some technical issues. I am grateful to Dr. Dustin Potter for his expertise in data analysis. Special thanks to Benjamin Rodriguez and Kristi Bennett for proofreading my manuscripts. I want to say thank-you to all other colleagues on campus, especially Dr. Ramana

Davuluri and Gregory Singer, with whom I completed the alternative promoter project.

I also would like to thank all my friends outside these groups, especially my previous advisor, Dr. Mark Muller, my friends Drs Rong Ju, Kui Liu, Yunfei Wang, and Feng Yang. Their friendship added another fun dimension to my graduate life. There are other friends I may not mention here, whose support, especially during the difficult times, made my life as smooth as possible.

Finally, I also thank my family for their continued support. They are the source of all my strength to face up to any difficulties. Every tiny bit of accomplishment that I have ever achieved belongs to them.

VITA

March, 1974……..Born, China

1998…….M. D., Hebei Medical University, China

2001……..M. S., The Capital University of Medical Sciences, China

2001-present…….Graduate Research Associate
The Ohio State University

PUBLICATIONS

Wu J, Smith LT, Plass C, Huang TH. ChIP-chip comes of age for genome-wide functional analysis. Cancer Res. 2006 Jul 15;66(14):6899-902.

Wu J, Wang SH, Potter D, Liu JC, Smith LT, Wu YZ, Huang TH, Plass C. Diverse histone modifications on histone 3 lysine 9 and their relation to DNA methylation in specifying gene silencing. BMC Genomics. 2007 May 24;8:131.

Singer GAC*, **Wu J***, Yan P, Plass C, Huang TH, Davuluri RV. Genome-wide analysis of alternative promoters of human genes using a custom promoter tiling array. Genome Res. Under review.

Liu Z, Xie Z, Liu S, Pavlovicz RE, **Wu J**, Plass C, Li PK, Li C, Huang TH, Marcucci G, Chan KK. DNA methylation modulation by a sesquiterpene lactone parthenolide. Cancer Res. Under review.

Liu Z*, **Wu J***, Xie Z, Liu S, Chen P, Plass C, Huang TH, Marcucci G, Chan KK. A Specific and Sensitive LC-MS/MS Method for Accurate Quantification of Gene Specific DNA Methylation (GSDM). In preparation.

Wu J, Huang TH. ChIP-chip meets nutrition: probing genome-wide dietary effects. Nutrition reviews. In preparation.

*These authors contributed equally to this work.

FIELDS OF STUDY

Major Field: Molecular Genetics

TABLE OF CONTENTS

Abstract……………………..……………………….……………………..……….....ii

Dedication……………………..……………………….……………………..……..........v

Acknowledgments…………………………………………………… ……..…………vi

Vita…..………………………………………………………………………………...viii

List of Tables………………………………..……………………………………......xiii

List of Figures……………………………..……………………..…………......xiv

Part I: DNA methylation and histone modifications in gene regulation

Chapters:

1 From epigenetics to epigenomics ……………..……………………………………1

1.1 DNA methylation…..……………………………..………………………4

1.2 Histone modifications...…....……………...…………..…...……………..…...12

1.2.1 Histone acetylation……………………………………………...14

1.2.2 Histone methylation……………………...……………………..…17

1.3 Crosstalks between DNA methylation and histone Modifications……19

1.4 Epigenetics and disease ………………....…………….............……21

1.5 Epigenomics ………………………………………………........……..22

2 ChIP-chip Comes of Age for Genome-wide Functional Analysis……24

2.1 Introduction……..……..……..24

2.2 The strategy and development of ChIP-chip ……25

2.3 ChIP-chip applications in genome-wide functional analysis ….....29

2.4 Other high-throughput methods beyond ChIP-chip ……...34

2.5 Perspective on future development of ChIP-chip ….....……...37

2.6 Conclusion on ChIP-chip……41

3 Diverse histone modifications on histone 3 lysine 9 and their relation to DNA methylation in specifying gene silencing ……....47

3.1 Introduction…….....……..…47

3.2 Methods…..…...…….....……...….....49

3.3 Results..... 54

3.4 Discussion.....59

4 Summary and perspectives……83

Part II: Alternative promoters and estrogen regulation

Chapters:

5 Transcriptional regulatory elements of gene regulation and estrogen receptors in breast cancer……..…….....85

5.1 Conventional and novel viewpoints on gene transcription…..……..85

5.1.1 The transcription machinery……86

5.1.2 *Cis* -acting transcription regulatory elements……...87

5.1.3 Novel perspectives on diversity of promoter regulation…….89

5.1.4 Analysis of regulatory elements……..……...91

5.2 The biological role of estrogen receptors in breast cancer……92

6 Genome-wide analysis of alternative promoter usage in estrogen-stimulated MCF7 cells using a custom promoter tiling array platform.....94

6.1 Introduction…….....……..…….94

6.2 Methods.....97

6.3 Results.....103

6.4 Discussion.....……..……..…….....108

7 Summary and perspectives……..127

References.....129

LIST OF TABLES

Table 3.1 The targets on selected custom mouse array and control genes unmethylated discovered by RLGS in mouse leukemia.....64

Table 3.2 The targets selected for COBRA and ChIP-PCR concentrations...66

Table 3.3 The primers used in COBRA, ChIP-PCR, promoter landscaping analysis and RT-PCR……..67

Table 6.1 Activity of promoters with various combinations of supporting evidence……..111

Table 6.2 Promoters validated by ChIP-PCR and the lines of evidence used to identify them……112

LIST OF FIGURES

Figure 2.1 A general ChIP-chip procedure.....43

Figure 2.2 Comparison of the procedures of ChIP-chip with other high throughput techniques for studying DNA-protein or RNA-protein interactions.....45

Figure 3.1 The distribution of histone modifications screened by ChIP-chip with mouse CpG island array…….....69

Figure 3.2 The distribution of histone modifications against DNA methylation identified by ChIP-chip and DMH with CpG island array……71

Figure 3.3 RLGS analysis……73

Figure 3.4 Histone modifications and DNA methylation status in mouse promoter regions……75

Figure 3.5 DNA methylation and histone modifications identified by COBRA and ChIP-PCR……77

Figure 3.6 The promoter landscape of histone modifications……79

Figure 3.7 The effect of TSA and AzadC on gene expression……...81

Figure 6.1 Alternative promoter selection……113

Figure 6.2 ChIP-chip ……..……115

Figure 6.3 ChIP-PCR……..…117

Figure 6.4 Promoter usage of NCOA7……119

Figure 6.5 Promoter usage of EIF3S9……..……..121

Figure 6.6 Multiple promoter usage in MCF-7 cells with E2 treatment…....123

Figure 6.7 Different patters of alternative promoters in MCF-7 with E2 treatment.....…125

PART I

DNA METHYLATION AND HISTONE MODIFICATIONS IN GENE REGULATION

CHAPTER 1

FROM EPIGENETICS TO EPIGENOMICS

Since 1995, the genomes of over 180 organisms have been sequenced[1]. Especially in 2003, the completion of the Human Genome project indicated one profound shift from the genomic era into the post-genomic era[2]. In the genomic era, the main focus was on clarifying the linear arrangement of the DNA sequence; in the post-genomic era, it is to understand how DNA sequences determine gene activities and the complex biology of organisms. It is now time to explore the astoundingly complicated network of regulation and functions of these genetic codes.

Three layers of regulatory networks are incorporated to regulate gene functions. The core layer is the DNA sequence, which contains not only protein

coding genes, but also regulatory units, such as promoters, enhancers, insulators, silencers and even some non-coding elements of unknown functions. The genetic information in all cells of one organism is the same and therefore its differential expression has to be based on more adjustable systems. The outer layer of regulatory network includes various signal transduction pathways and other metabolic processes. Protein factors are still in the center of these processes, which are both cause and consequence of gene regulation. The core layer contains relatively permanent information, which can be stably transferred from parents to the next generation, whereas main function of the outer layer is the flexibility to control the temporal and spatial expression of genomes. Between these two layers is the intermediate one, epigenetics, whose definition is still in evolution but at least includes DNA methylation and histone modifications. Different from the permanent DNA sequence and the temporary outer layer information, the epigenetic information contains signatures that are both stable to be inherited by cells and reversible without changes of DNA sequences.

The integration within and between these three layers adds more complexity to this regulatory network. It is known that environmental stimuli are relayed in a network to enter the nucleus. Promoters, enhancers and other regulatory elements cooperate to drive gene transcription. The interactions within the epigenetic network are well known. Intergenic regulatory elements require certain chromatin conformation and protein factors to perform their regulatory functions; epigenetic markers are sometimes led by transcription factors to specific location of the genome and chromatin modifications recruit protein

factors to regulate corresponding genes; Transcription factors recognize specific binding elements and access active chromatin domains to drive gene transcription. Here it deserves caution that temporary regulation signals may lead to epigenetic change that will be transferred to the next generation of cells. These interactions between different levels of gene regulation allow sophisticated control of biological activities.

Besides genetic information, the stability of epigenetic signatures may be critical for those important biological activities transferred or reserved such as those occurring in cancer stem cells or learning and memory processes. In addition, the stability of epigenetic markers also justifies the possibility of annotating their occurrence along DNA strands as the Human Epigenome Project aims to. However, we have to keep in mind that these epigenetic markers are much more changeable and heterogenous than the DNA sequence. Epigenetic markers are not the same in all cells of an organism. Therefore, efficient and reliable epigenomic tools such as ChIP-chip are necessary to provide a panoramic view of epigenetic regulation in different cells and under different conditions. At last, the inter-related regulatory networks make it indispensable that the combination of epigenomic tools with other high throughput tools of system biology. Since these regulatory networks are complicated and highly dynamic, not all changes inside these networks will definitely result in irreversible differences in biological features. Some gain or loss of DNA methylation in certain locations may not affect gene transcription.

Therefore, it is necessary to discriminate key steps resulting in disease from those redundant changes.

The word “epigenetic” was originally coined to categorize those phenomena that could not be explained by genetic principles[3]. Currently epigenetics can be defined as the study of these phenomena that can change gene expression and cellular phenotype without changes of genotype, the DNA sequences[4, 5]. Epigenetic topics may include paramutation, position effects variegation, gene imprinting, X inactivation, DNA methylation, histone modifications, chromatin modeling and non-coding RNAs. We will focus on the two most studied fields, DNA methylation and histone modifications.

1.1 DNA methylation

DNA methylation occurs at the carbon atom five positions of cytosine, which is the only covalent modification in eukaryotic genomes and has been found in every vertebrate examined. In eukaryotes, this modification take places primarily in the context of the dinucleotide “CpG”[6]. However, there have been reports of non-CpG methylation in human embryonic stem cells and in plants[7, 8]. In the human genome, methylated cytosines accounts for approximately 1% of the total DNA and 70-80% of all CpG dinucleotides[9, 10].

The amount of CpG dinucleotides is lower than other dinucleotide combinations in the human genome[11]. It is believed that this is caused by methylated cytosine, which is chemically unstable and prone to deamination, and in the end is converted to thymidine. However, there are regions of the genome,

called CpG islands, rich in clusters of CpG dinucleotides[12, 13]. CpG islands were originally characterized by an unmethylated CpG content greater than 50% and about five times more frequently than the expected CpG rate[14]. These unmethylated CpG islands are mostly located in the promoters and first exons of active genes, and are resistant to DNA methylation in part because of active transcription status. For example, binding of the Sp1 transcription factor seems to be critical for protection of CpG islands from methylation[15]. To exclude those CpG-rich sequences that are not associated with promoters, such as Alu repetitive elements, more strict criteria were introduced as DNA stretch of from 500bp to 4kb, with CpG content of more than 55% and the ratio of observed CpG over expected CpG in excess of 0.65[16].

DNA methylation is catalyzed by DNA methyltransferase (DNMT), which can be further categorized into three classes by their substrates:N4-methyladenine DNMT, N6-methyladenine DNMT and C5-methylcytosine DNMT. All of these three classes have been found in prokaryotes and only C5-methylcytosine DNMT is present in eukaryotes. In the methylation reaction mediated by these DNMTs, the target cytosine is everted from the DNA helix and inserted deep into the active site of DNMT and a cysteine SH group from the active site of DNMT initiates a nucleophilic attack at the C6 position of the target cytosine, The resulting intermediate complex acquires a methyl group from the methyl donor, S-adenosylmethionine (SAM) and produces methyl-cytosine[6, 17].

In mammalian cells, there are different members of DNMTs that share the same catalytic domain but possess different N-termini. DNMT1 is the largest

family member at 184 kDa[18]. The C-terminus of DNMT1 contains the catalytic domain as in other DNMTs and the large N-terminus contains the regulatory domains[19]. DNMT1 is thought to be a maintenance methyltransferase because it has a 10-40-fold preference for hemimethylated DNA and localizes to replication foci in mammalian cell nuclei[20, 21]. DNMT1 plays an essential role in life. DNMT1$^{-/-}$ mice showed a genome-wide hypomethylation, chromosome instability, abnormal imprinting and embryonic lethality[22-24]. Other isoforms of DNMT1 have been identified. The oocyte-specific DNMT1o and sperm specific DNMT1p may have tissue specific methylation function at some imprinted loci[25, 26]. DNMT2 contains all the conserved methyltransferase motifs but lacks the large amino-terminus[27]. It seems that DNMT2 can bind CpG sites[28]. However, the biological activity of DNMTs has not yet been determined. There were reports that DNMT2 may possess demethylation activity, which could not been confirmed in subsequent studies[29, 30]. Another family of enzymes, including DNMT3a and DNMT3b, do not exhibit preference for hemimethylated sites and demonstrate the *de novo* methyltransferase activity[31, 32]. DNMT3a and DNMT3b are intermediate in size in comparison with DNMT1 and DNMT2 and have smaller N-terminal regions. Both DNMT3a and DNMT3b are required for de novo methylation following the genome-wide demethylation in the mouse ES cells[33]. It has been shown that DNMT3a$^{-/-}$ mice were born normal but died by 4 weeks of age[33]. Mutation within the C-terminus of human DNMT3b is found to be associated with immunodeficiency, centromeric instability and facial anomalies (ICF) syndrome, which is characterized by abnormal hypomethylation

in centromeric satellite sequences and genome instability[34]. Another member of DNMT3s is DNMT3L, which does not contain several C-terminal catalytic motifs and probably dose not process catalytic activity[35, 36]. However, DNMT3L can help enhance the methylation activity of DNMT3a and DNMT3b[37]. For example, DNMT3L is required by DNMT3a to establish methylation in some imprinted genes in both male and female germ lines[35, 36]. DNMT3L$^{-/-}$ mice appear normal at birth but both sexes are sterile[38].

The process of demethylation remains elusive. There are passive and active demethylation pathways. During DNA replication, DNMT1 resides in the replication foci, through its interaction with the DNA polymerase clamps, and converts the hemimethylation sites to fully methylated sequences [39]. Passive demethylation is dependent on DNA replication and occurs when DNMT1 fails to maintain the DNA methylation pattern [40]. Active demethylation is processed independent of DNA replication, which has been observed in the genome-wide demethylation of zygotic paternal genomes after fertilization, and in the local demethylation of transfected DNA molecules [41, 42]. Although it seems clear that there is active DNA demethylation occurring in the genome, the enzymes responsible have been a mystery for long time. In one report, an enzyme isolated from nuclear extracts of chicken embryos, later purified as a glycosylase, promoted an active demethylation of DNA by base excision. In other reports, the DNA glycosylase did not show enough glycosidase activity and it is more likely to be a DNA repair enzyme [43, 44]. The ribozyme-like demethylase reported by Weiss et al has been placed in doubt when later reports indicated that the

purified preparations were not sensitive to RNAase [45, 46]. There are also a series of reports by Szyf's group showing that MBD2 had specific demethylase activity for mCpG DNA [29, 47, 48]. Unfortunately, this work could not be reproduced by other groups [49-52]. Recently, Gadd45a (growth arrest and DNA-damage-inducible protein 45 alpha), a nuclear protein involved in maintenance of genomic stability, DNA repair and suppression of cell growth, was reported to actively erase DNA methylation by promoting DNA repair and release epigenetic silencing[53].

In terms of genome structure, DNA methylation plays multiple roles. These roles include stabilizing the genome, facilitating mutation, and affecting genome compartmentalization. It has been demonstrated that DNA methylation might mask the homologous recombination between repetitive sequences. In fungal and mammalian cells, the recombination frequencies are reduced greatly when recombination hotspots are methylated [54, 55]. DNMT1-/- cells, or those treated with 5-Aza-2'-dC, in which methylation is removed or inhibited, show increased genomic rearrangements [56-58].

CpG sites are hotspots of genomic mutation, which has been demonstrated in the human germline [59]. Inactivating mutations can occur in tumor suppressor genes. For the p53 gene, about 25% of all mutations are at CpG sites in human cancers, and almost 50% of p53 mutations occur at methylation sites in colon cancer [60]. Methylation-dependent mutations are believed to arise from two mechanisms. One is the increased spontaneous deamination rate of 5-methylcytosine versus cytosine to thymine [61]. The other

is the repair deficiency of the deamination product of 5-methylcytosine versus that of cytosine [62].

DNA methylation is also believed to participate in chromatin compartmentalization. Clusters of inactive, methylated rDNA repeats are located in neuronal, perinucleolar heterochromatin [63]. Chromatin structure is highly related to its transcription status. Methylated CpG islands are present in the inactive, heterochromatic X chromosome, while unmethylated CpG islands are associated with an open structure that is deficient in H1, and is enriched in acetylated forms of histones H3 and H4 [64, 65].

In mammalian cells, the main role of DNA methylation is the repression of gene expression. Cytosine methylation in the regulatory elements of transposons, imprinted genes and genes on the inactive X chromosome are significant examples for DNA methylation repression. CpG methylation is believed to be protective for the mammalian genome, inhibiting the transcription of transposable elements. Most CpG dinucleotides are present not in CpG islands, but dispersed in CpG-poor genomic DNA, mainly in the parasitic elements, such as Alu elements and L1 elements. Alu's and L1 elements are scattered throughout the entire genome [66]. Because of the strong promoter capabilities, methylation of these CpG dinucleotides silences the transcription, and limits the spread of these parasitic elements, so that the threat to genomic stability is reduced. If activated, these elements are able to integrate and disrupt target genes [67, 68].

More attention is given to methylated CpG islands, which mostly reside in the promoter and the first exon of cellular genes. Methylation on CpG islands shows typical repressive function, and aberrant methylation of CpG appears to be involved in gene repression. Initial studies came from the correlation between gene activity and DNA hypomethylation. Tissue-specific genes are demethylated in expressing lineages, but are heavily methylated in cells that do not express them. Most housekeeping genes, however, are constitutively expressed, containing unmethylated CpG islands in all cell types [69, 70]. The exceptions are imprinted genes, or genes on the inactive X chromosome, such as H19 and HPRT, which are silenced by methylation of CpG islands [71]. The indirect evidence of repressive function of DNA methylation is the function of 5-azacytidine, which can demethylate and reactivate silenced genes[20, 72]. How this specific pattern of DNA methylation is established is not completely clear. It has been described that there are two genome-wide demethylation and remethylation waves in germ cells and preimplantation embryos, which are critical to methylation reprogramming of the genome. The control mechanisms are still unknown [73].

Several mechanisms have been described on how DNA methylation leads to subsequent transcriptional repression. Some transcription factors are unable to bind to their targets DNA when the DNA is methylated, as in the case of transcription factor AP-2[74, 75]. Alternatively, DNA methylation is believed to give rise to chromatin structural changes which may also mask the accessibility for transcription factor binding[71]. Another interesting mechanism comes from

the identification of the MBD family members, which are able to bind methylated DNA, and repress gene transcription[76]. The members of this family include MeCP2, MBD1, MBD2 and MBD3. MeCP2 is the first founding member of the MBD family described at the molecular level to have methyl-CpG-binding domain (MBD) and transcriptional repression domain (TRD) mediating the binding of methylated DNA and transcriptional repression[76, 77]. *In vivo* tethering assays, in which a Gal4-MBD fusion protein is expressed to recognize the Gal4-binding sites and 'tether' the target protein fragment, show a transcriptional repression function of MeCP2. MBD1 can also mediate transcriptional repression but has a preference for densely methylated sequences[78]. MBD2 and MBD3 are the only two members of MBD family with high sequence similarities. MBD3 is believed to be the component of the multiprotein complex Mi-2, which functions as a transcriptional repressor[79]. The function of MBD2 is rather ambiguous, though it was reported to be the component of MeCP1 complex with transcriptional repression function[30]. MBD2 has a MBD, a coiled-coil domain and a minimal repression domain near the MBD[29, 80]. There is no report on the function of the coiled-coil domain. The minimal repression domain mediates gene expression and interacts with the transcriptional repressor, Sin3A , which interacts directly with HDACs[81]. There is evidence from Mbd3(-/-) and Mbd2(-/-) mice that MBD3 and MBD2 play distinctive, but interacting, roles in mouse development[50]. MBD3 is necessary for embryonic development, while MBD2 knockout mice are viable, but with a maternal nurturing defect. The double

mutant of both MBD2 and MBD3 has a stronger phenotype than either single knockout.

1.2 Histone modifications

Another group of epigenetic markers include acetylation and methylation of histone tails [82]. Acetylation and mono-, bi- or trimethylation at specific amino acids of histones can impose permissive or repressive functions on gene transcription[82]. These histone modifications are catalyzed by various enzymes, which combine with other protein factors forming functional complexes to control the temporal and spatial modification of histones.

The eukaryotic chromosome is composed of protein and DNA, which form the protein-DNA complex called chromatin. Chromatin assumes various appearances in different stages of the cell cycle. Only in M phase can the individual chromosome be visible as the condensed form. During the other stages of the cell cycle, chromosomal DNA is highly dispersed and indistinguishable. So there are two different forms of chromatin defined: One is called heterochromatin, which is highly condensed and transcriptionally inactive; the other is called euchromatin, which is open and relaxed, and is accessible to transcription factors to initiate active transcription[83].

The basic unit of chromatin is called the nucleosome, which contains approximately 150 base pairs of DNA and a core histone octamer[84]. This histone octamer has two copies of each H2A, H2B, H3 and H4. The core histones are evolutionarily conserved, which suggests they may play important

roles in cell activities. While the globular C-terminal domains of these histones form the nucleosome scaffold wrapped with DNA, the N-terminal “tails” are flexible and protrude outward from the nucleosome. Histone modifications are imposed on the specific amino acids of these histone tails. H2A has four variants: H2AX, H2AZ, H2ABbd and MarcroH2A, which are believed to play important role in the histone-mediated regulation of chromatin metabolism[85]. H3 also has four variants: H3.1, H3.2, H3.3 and Cenp-A. Of these, H3.1 and H3.2 are involved in replication-dependent chromatin assembly and H3.3 is enriched at transcription active regions[86-88]. The nucleosome is linked by 30-40 bp of DNA and the linker histone H1. It is believed that these basic units can further undergo multiple packagings and eventually form a comprehensive chromosome, whose conformation switches between heterochromatin and euchromatin corresponding to the inactive and active transcription statuses[89].

The histone tail protrudes from the nucleosome and contains 20-35 residues[90], which are affected by a variety of post-translational modifications including acetylation, methylation, phosphorylation, ubiquitination and ADP-ribosylation [91]. These modifications may contribute to chromatin conformation and provide gene regulation information. The recent hypothesis of a “histone code” predicts that the information of distinct histone modifications can be read by other proteins to promote downstream events, either gene activation or repression. Currently, the most extensively studied modifications are histone acetylation and methylation, which will be described in detail.

1.2.1 Histone acetylation

Acetylation and deacetylation of the ε-amino groups of lysine residues in histone tails have long been linked to transcription activity. Histone acetylation was first discovered in 1964 and was proposed to regulate gene function, which was later confirmed by the finding that hyperacetylation was associated with transcription activation[92, 93]. Acetylation of lysine residues neutralizes their positive charge, thus decreasing the electrostatic interaction between DNA and histones. The reduced affinity of histone to DNA leads to alteration of histone-histone interactions between adjacent nucleosomes and the interactions between histones and other regulatory proteins[94, 95]. As result, these changes help establish a permissive chromatin environment for transcription. Some frequent targets of acetylation include lysines 9, 14, 18 and 23 in H3 and lysines 5, 8, 12 and 16 in H4[95]. Unexpectedly, acetylation at lysine 12 of H4 in yeast and *Drosophila* and some acetyltransferases, such as Sas3 and Hat1 in yeast were found contribute to gene silencing[96-99]. However, the original proposal that the neutralization of the positive charge of lysine residues and a resulting decrease in electrostatic interactions between DNA and histones is the major consequence of histone acetylation was challenged after discovery of histone acetyltransferase (HAT)[100]. An idea that has gained more popularity is that histone tails can be modified by histone modification enzymes and read by non-histone proteins in a residue-specific manner[101].

The first HAT identified was called Gcn5, which then was found to be associated with other transcription factors in a multisubunit complex that can

regulate Gcn5 specificity and recruitment to target promoters[95, 102, 103]. Since the identification of Gcn5, numerous other proteins possessing HAT activity have been found and frequently these HATs form protein complexes with other components. These HAT complexes can be divided into several families associated with different transcription adaptors or coactivators. The first family of HATs is called GNAT superfamily (Gcn 5-related *N*-acetyltransferase), which includes proteins involved in transcription initiation (Gcn5 and PCAF), elongation (Elp3), histone deposition and telomeric silencing (Hat1)[104]. The p300/CBP family is another HAT family, which contains members such as coactivators for multiple transcription factors[95]. The MYST family of HATs is named after its member MOA, Ybf2/Sas3, Sas2 and Tip60, which form complexes with other subunits and involved in transcriptional elongation, replication[105], dosage compensation of X chromosome[106], DNA-damage responses and the regulation of apoptosis[107]. Other HATs can be grouped as basal transcription factors, such as TAFII250, or as nuclear hormone-receptor cofactors, such as ACTR and SRC1. It was also reported that many HATs have non-histone proteins as substrates, such as p53 tumor suppressor and transcription factors TFIIE and TFIIF. The subunits or different forms of these HAT complexes have been demonstrated to be critical for regulating the specificity of acetyltransferase activity. One “two-step classification” hypothesis classified the lysines of histone tails into three cases and six groups [108], which were recognized by catalytic domains of different HATs[108, 109].

The acetylation of lysines in the N-terminal tails on histones is reversible and can be removed by a histone deacetylase (HDAC). HDACs catalyze histone deacetylase activity using a charge-delay mechanism and Zn^{2+}, a crucial cofactor for this reaction. Inhibitors such as Trichostatin A (TSA) function by displacing the zinc atom[110]. There are dozens of deacetylase enzymes in human, divided into three classes based on their sequence similarity with yeast reduced potassium dependency 3 (Rpd3), histone deacetylase 1 (Hda1) and silent information regulator 2 (Sir2): class I HDACs (HDAC1, 2, 3, and 8) are similar to the yeast Rpd3 and localized to the nucleus; class II HDACs (HDAC4, 5, 6, 7, 9, and 10) are homologous to the yeast Hda1 and are found in both the nucleus and cytoplasm; and class III HDACs including SIRT1–7, are similar to the yeast Sir2 and NAD-dependent enzymes, which have a structure distinct from the other two classes[111].

HDACs form various complexes to promote gene regulation. The most studied group of HDACs are mammalian class I, of which HDAC1/2 has been found in three complexes: the Sin3, NuRD, and CoREST complexes[112]. These complexes serve as corepressors for numerous transcription repressors. For example, neuron-restrictive repressor REST recruits both Sin3 and CoREST complexes to repress target transcription[113]. Unliganded thyroid hormone receptor recruits Sin3 and Mi2/NuRD for non-targeted deacetylation and the NCoR or SMRT complexes, which contain HDAC3, 4 and 5, for targeted deacetylation[114].

The regulatory functions of both HATs and HDACs are regulated at multiple levels. The first level is to regulate enzyme availability. Both HAT and HDAC are recruited by many transcription factors to regulate transcription. The ability of HATs and HDACs to be recruited can be changed to control their availability to a given signaling event. The second level is regulating enzyme activity by post-translational modifications or protein-protein interactions. The third level is to regulate the quantity of these enzymes in the cell, and has been shown during development[115].

The histone code of acetylation can be read by proteins that interact with histones. One group of proteins contains bromodomains which have been reported to interact with specific acetylated lysines of histones. The interaction between bromodomain and acetylated lysines is believed to stimulate nucleosome remodeling, further acetylation or recruitment of TFIID[116, 117]. In contrary, Sir3 and Tup1 are proposed to interact with hypoacetylated histone and repress gene transcription[118-120].

1.2.2 Histone methylation

Protein methylation is a covalent modification commonly occurring at carboxyl groups of glutamate, leucine, arginine, histidine and lysine residues[121]. Histones are methylated on arginine and lysine residues only. The focus here will be on histone methylation. Histone methylation has been shown to function in both transcriptional repression and activation[122]. Methylation of H3K9 H4K20 and H3-K27 are associated with transcriptional silencing. H3K4,

H3K36 and H3K79 are associated with gene activation. Increased complexity comes in part from the fact that methylation of lysine can occur in one of three forms: mono-, di-, or trimethyl. In mice, for example, pericentric heterochromatin is specifically enriched in trimethyl-H3K9 and H4K20[123, 124]. By contrast, mono- and dimethyl-H3K9 and H4K20 are found in euchromatin.

The first lysine specific histone methyltransferase (HMT) is SUV39h1[125]. Since its identification, numerous other HMTs have been found and characterized by their possession of the SET domain, which contributes the methyltransferase activity[126]. Based on the similarity between the human SET domains and their relationships to SET domains in yeast, human SET proteins can be divided into four families: SUV39 family, SET1 family, SET2 family and RIZ family. SUV39 family has the two most studied members, SUV39h1 and G9a[126]. The SUV39h1 displays an exquisite site-selectivity towards H3K9 methylation[125]. Methylation of H3K9 provides the binding sites of heterochromatic protein 1 (HP1), which forms complexes with SUV39h1 and generates a highly compact heterochromatic subdomains[127]. Other reports showed that SUV39h1 is also involved in transcriptional repression at euchromatic loci and the SUV39h1-HP1 complex is recruited by Rb protein to regulate cell cycle controlling genes such as cyclin E[128]. G9a is a ‘dual’ HMTase methylating both H3K9 and H3K27[129]. G9a is not localized to the repressive chromatin. However, loss of G9a in G9a-deficient embryonic stem cells abolished methylation mostly in euchromatic regions, which implies that G9a may be involved in gene-specific repression[130]. Recently the enzymes

involved in histone methylation have expanded from HMT to histone demethylases, which have two major families, LSD1 and JMJD2A, and are targeted to H3K4, H3K9 and H3K36[131]. It is likely that more protein factors will be found to play different roles in this highly regulated network of histone modifications.

1.3 The crosstalk between DNA methylation and histone modifications

DNA methylation and histone modifications play either repressive or activating roles in gene transcription [132] [82]. One of the critical questions about epigenetic regulation is how these epigenetic markers lead to transcription regulation. Currently, the mechanism of epigenetic regulation seems to link DNA methylation and histone modifications [82, 133].

Therefore, to understand the mechanism of epigenetic regulation, it is necessary to clarify the crosstalk between DNA methylation and histone modifications [82]. The first strong evidence of the connection between DNA methylation and histone deacetylation came from the comparison of bulk chromatin and chromatin associated with unmethylated CpG islands [65]. The chromatin along unmethylated DNA was enriched with hyperacetylated histone 3 and 4. The studies, based on in vitro transfected constructs containing the hsp70 promoter or tk gene, showed that unmethylated constructs were associated with hyperacetylation and remain active [71, 134]. Further support for the close association of DNA methylation and histone acetylation was the discovery of protein complex connections between DNA methylation and histone acetylation,

such as the co-immunoprecipitated methyl-CpG binding protein 2 (MeCP2) and histone deacetylase (HDAC) [135, 136]. It was then proposed that the repressive complexes, containing both MDB and HDAC, and also other repressive partners, direct DNA methylation for transcription repression [137].

The crosstalk between DNA methylation and histone methylation, especially methyl-H3K9, is more complicated. Studies in fungi, plants, and mammals all indicate that methyl-H3K9 may control DNA methylation in heterochromatin [138-140]. The association of DNA methyltransferase with heterochromatin protein 1 (HP1) supported the model that binding of HP1 with methyl-H3K9 recruits DNA methyltransferase to methylate DNA, establishing the heterochromatic state [141]. There is also evidence showing that DNA methylation impacts histone methylation [142-146], and that DNA methylation might exert a positive feedback on lysine methylation. To reconcile these two seemingly distinct mechanisms, a self-enforcing network of epigenetic regulation was proposed in which epigenetic order flows from histone methylation to DNA methylation and histone acetylation, ending again at histone methylation [82, 137]. The current self-enforcing model implies close correlation of histone modifications and DNA methylation. However, recent reports demonstrated that these epigenetic markers have a varying degree of autonomy, which will be discussed in the next chapter.

1.4 Epigenetics and disease

Abnormal methylation has been identified in cancer and other genetic diseases, including ICF, Rett and fragile X syndromes, further implicating the importance of DNA methylation on genomic functions[147]. The DNA methylation pattern undergoes various changes in carcinogenesis: both genome-wide hypermethylation and regional hypomethylation. There is an overall reduction in DNA methylation in cancer, which was first demonstrated by HPLC, though the degrees of demethylation is considerably diverse[148]. About 8%-10% of DNA methylation reductions were detected in benign and cancer polyps compared with the normal mucosa in colon tissue [149]. Hypomethylation is very often found in parasitic elements. As described above, the activation of these elements may lead to the disruption of other genes. Hypomethylation of L1 elements is often seen in chronic lymphocytic leukemia compared to normal mononuclear blood cells [150]. On the other hand, CpG islands with dense CpG content are the primary targets for hypermethylation in cancer. Compared with normal tissues, the CpG island in the promoter region of the calcitonin gene located on chromosome 11p was heavily methylated in human solid tumors [151]. Some potential tumor suppressor genes in the region of chromosome 11q are hypermethylated simultaneously [152]. Recently more and more non-mutated tumor suppressor genes have been identified as becoming silenced by promoter CpG island methylation [153]. More details have been discussed in one recent review[154].

A number of histone acetyltransferases or their protein partners have been found to be rearranged or mutated in cancer and other diseases[155]. For example, mutation of the CBP locus, which is one of the HATs, is the genetic basis for Rubinstein-Taybi syndrome and patients with this disease are predisposed to higher rates of cancer [156]. It has also been shown that mice lacking SUV39h1 and SUV39h2 will develop leukemia [157].and MLL, a protein of SET1 family, has also been linked to cancer[158].

Since many diseases such as cancer have a significant epigenetic etiology, epigenetics has gained more and more focus in translational studies or clinical therapy, which include using epigenetic markers for early detection screening of tumors, as markers of tumor prognosis, as the predictor of response to chemotherapy, and especially, developing a novel therapeutic option, called epigenetic therapy. In fact, many agents have been discovered to alter DNA methylation or histone modifications and some of them are being evaluated in clinical trials[159].

1.5 Epigenomics

Since DNA methylation and histone modifications are genome-wide epigenetic markers and proposed therapeutic interventions will affect the whole genome, it is necessary to introduce epigenomic tools to assess the gene targets affected genome-wide. The emerging Human Epigenome Project reflects this strategy [160]. Recently array-based studies profiled the histone lysine methylation states at distinct repeat classes [161]. ChIP-chip has been used to

map and link histone modifications to transcription status [162]. We also have established one strategy combining DMH and histone ChIP-chip to screen for tumor suppressors [163]. To clarify the remaining questions about the crosstalk between DNA methylation and histone modifications, we created our own custom array containing a group of selected targets identified previously in leukemia [164], which will be discussed in Chapter 3. In Chapter 2, we will focus on one of the revolutionary epigenomic tools, ChIP-chip.

CHAPTER 2

CHIP-CHIP COMES OF AGE FOR GENOME-WIDE FUNCTIONAL ANALYSIS

A short version of this chapter was published in Cancer Res. 2006 Jul 15;66(14):6899-902 by Wu J, Smith LT, Plass C and Huang TH.

2.1 Introduction

The completion of the Human Genome Project provides a road map for thorough interrogation of gene functions [164, 165]. This new endeavor is also made possible with further technological advances[165]. In addition to identifying novel transcription factor targets, current studies may shift our attention to genome-wide characterization of histone modifications and DNA methylation[160, 166]. These epigenetic processes are closely related to normal development and disease processes, including carcinogenesis[167]. The importance of this type of study is further echoed by a recent proposal for the Human Epigenome Project[160, 166]. This requires genome-wide technologies with high-throughput capability. One of these techniques is Chromatin ImmunoPrecipitation (ChIP), which can be broadly applied for the analysis of interactions among epigenetic markers, *cis*-acting elements and regulatory proteins[168, 169]. Hybridization of immunoprecipitated DNA on chip-arrays (ChIP-chip) has also been extensively

used as genome-wide tool to screen the binding position of protein factors[162, 170-173]. Because the use of ChIP-chip begins to gain popularity in the research community, the present review provides a timely overview of recent developments and applications.

2.2 The strategy and development of ChIP-chip

ChIP-chip combines ChIP and microarray techniques together (Figure 2.1). Briefly, cells or tissues are treated with formaldehyde so that interacting DNA and protein are crosslinked *in vivo*. DNA bound by protein is then sheared typically by sonication to ~0.2 to 2 kb. The DNA-protein complexes are then immunoprecipitated with specific antibodies against the protein. After immunoprecipitation, the crosslinking of the complexes is then reversed by incubation at 65^{o}C, and DNA elution. The eluted DNA is labeled with fluorescence dyes and hybridized to microarrays. In parallel, control DNA extracted from cells directly, without immunoprecipitation, or other reference DNA are used as controls. The hybridized slide is then scanned and analyzed. Comparison of immunoprecipitated DNA with reference control help to profile the binding position of specific proteins in the genome[170, 171, 174-176].

Compared to the amplification of specific target sequences from immunoprecipitated material (ChIP-PCR), ChIP-chip is a genome-wide “reverse-genetic” approach. To screen protein targets with ChIP-PCR, prior knowledge has to be applied to predict putative protein targets[168, 169]. ChIP-PCR is limited by the number of targets that can feasibly be checked compared with

microarray, which contains thousands of probes printed on glass slides. Another advantage of ChIP-chip is that it targets those genes directly bound and potentially regulated by the protein factor. With classical expression arrays, directly regulated genes and those changed secondarily cannot be distinguished easily. Therefore, once established, the high- throughput property of microarray combined with ChIP has made the ChIP-chip method widely used recently.

Ren and his colleagues pioneered the ChIP-chip technique by making a yeast intergenic DNA array[174]. Using this genomic array, they identified novel targets of the yeast transcription factors Gal4 and Ste12, and their expression levels were independently verified by expression microarray analysis[174]. A subsequent study by Iyer *et al.* profiled the binding sites of the cell-cycle transcription factor SBF and MBF in yeast[175]. ChIP-chip studies were further extended to bacterial, plant, and mammalian systems[177-181].

The extent of ChIP-chip application depends in part on the development of DNA microarray technology, especially the availability of arrayed slides for these organisms. In humans, one of the first ChIP-chip experiments adopted was the use of a CpG island array for screening novel E2F4 targets by our laboratory[179]. This array panel of CpG island clones was further developed to include 12,000 target sequences and has been deposited in the University of Toronto Microarray Facility for wider usages by the research community[182, 183]. Although the CpG island array can be applied without prior sequence knowledge to identify target genes, one limitation is that CpG islands may cover only 50-60% of whole human gene promoters and repeat sequences in this array

heavily interfere with the hybridization and later analysis[164, 165]. Another system is a selected promoter array, which contains 13,000 printed human fragments[184]. This array panel does not contain repeat sequences, but the distribution of sequences may be biased by only selecting known 5'-end regions of genes. Both of these arrays leave genomic gaps in the intergenic regions, which may be important for protein binding. To address this limitation, prototype tiling arrays, which cover human chromosomes 21 and 22, have been constructed to explore intergenic bindings of estrogen receptor alpha, chromosome-wide maps of histone modifications, and other ChIP-chip studies[185-190]. Recently, human genome-wide array with PCR amplicons printed has further been improved to cover more than 90% of human non-repetitive DNA sequences[191]. Different from PCR amplicon-based strategy, a new oligonucleotide array technique helped build another human array representing all human non-repetitive regions with resolution at 100-base pair[192]. A microarray containing the entire human genome sequence would be ideal to ChIP-chip, which is currently not available. ChIP-chip may be applied in other systems, like murine or viral, once such DNA arrays are available. Other arrays for specific research purposes have also been established and will be addressed later[182, 193, 194].

Critical for ChIP-chip experiments is the amount of starting material required for successful microarray hybridization. The number is highly variable depending on the quality of the antibody, binding frequency of protein to DNA, and other possible unknown factors. In one previous report, 1-10ng of ChIP DNA

was pulled down from total 30-60ug of *Drosophila* DNA[195]. With cDNA expression array, at least 2 μg of cDNA is required, which is difficult to obtain with ChIP-chip due to experiment limitations and expense. Up to 50 individual ChIP DNA samples have been pooled to get enough DNA for array hybridization[170]. An alternative approach to enrich the starting material is through PCR amplification. Ligation-mediated PCR, random primed PCR amplification, and other RNA polymerase-based liner amplification methods have been successfully applied to ChIP-chip[174, 175, 196]. However, PCR bias may be a concern, especially with mammalian systems, where large amounts of repeat sequences may skew data.

Another concern is background DNA that is pulled down by non-specific interactions of protein and DNA. In one typical ChIP-chip experiment based on pull-down by Suz12, more than 50% of the targets, with 3-fold enrichment compared with control group, were false-positives[182]. With selective human promoter array and appropriate threshold criteria, it was claimed that the false-positive rate could be less than 16%[184]. Attention should be paid to several key basics for ChIP-chip, such as antibody quality, immunoprecipitation handling, optimization of array hybridization conditions and data normalization and analysis. It is also necessary to establish appropriate controls. Generally, genomic DNA is used as an input control, and samples from no antibody or IgG groups are used as negative controls. Other control designs, such as transformed cell lines *versus* empty vector cell lines, wild type target *versus*

mutation target, and with *versus* without drug treatment, can all be considered as control options.

2.3 ChIP-chip applications in genome-wide functional analysis

The ChIP-chip studies discussed above indicate two strategies using ChIP-chip[174, 175]. One is to screen and identify binding targets of protein factor without prior knowledge. The other is to map protein's binding location to provide a genome-wide binding profile, which may require more accurate raw data for statistical analysis.

ChIP-chip has shown its power in probing both histone modification and transcription regulation in a high throughput and genome-wide way. Gene transcription can be regulated by covalent modification on the "tail" of chromatin histones. Systematic approaches are needed to perform global studies to provide a genome-wide perspective of histone regulation. ChIP-chip can be applied in mapping histone modifications. Two strategies are used in these studies. One is to detect the distribution of histone modifications using antibodies specially targeting these modifications[197]. The other is using ChIP-chip to locate, not the modified histones, but the enzymes that catalyze the histone modification reactions[198, 199]. In yeast, when mutations are introduced to a specific histone modifying enzyme, the changes of histone modifications can indicate targets[197]. This kind of study can also take advantage of combining ChIP-chip with other high throughput strategies, especially expression profiling, to establish the correlation of histone marks with transcription activity[197, 200-202]. Most

work is done in yeast and mammalian systems, though work with plant and *Drosophila* have also been reported[178, 194, 203].

ChIP-chip combined with yeast genomic microarray delineated the distributions of some modified histones, such as acetyl-H3 and H4, actyl-H3 (lysine 9 and 14), acetyl-H4 (lysine 5, 8, 12 and 16) and methyl-H3 (lysine 4), and the locations of some histone modifying enzymes, such as methyltransferase (Set1), deacetylase (Rpd3, Hda1, Hst1 and Hos), acetyltransferase (Gcn5 and Esa1), and other related factors (Rsc, Sth1, Ume1 and Ume6)[197-200, 204-206]. Some factors, such as acetyltransferases Gcn5 and Esa1, are generally distributed in the promoters of active genes. In contrast, histone deacetylases, like Rpd3, Hst1 and Hda1, are recruited to functionally distinct classes of genes. Mapped regions show various correlations with expression data from expression arrays[197]. For example, it is interesting to find that globally, methyl-H3K4 is more frequently distributed in coding regions but not promoter regions as previously assumed. This is one example how ChIP-chip provides a wider view than gene by gene studies[207, 208]. Another example is that telomere-proximal regions show significant histone hypoacetylation, which confirmed previous reports[197].

Three different arrays have been applied in studying human histone modifications with ChIP-chip. One is island array that showed strong correlation between CpG methylation and histone modifications[163]. With the help of these two features, five novel genes were identified and confirmed in the colon cancer cell line, SW-48, as being epigenetically silenced. The second array is the cDNA

array, which provided new information with respect to the distribution of histone methylation patterns in the coding regions of human genes[209]. Finally, the tiling array mapped H3 markers (di- and trimethyl-K4, acetyl-H3K9 and acetyl-H3K14) to nonrepetitive regions of human chromosome 21 and 22[187]. This indicates that modified histones are distributed in a “punctated” pattern with ~1-kb gaps between each other. Tri-methyl-H3K4 and acetyl-H3 are located at 5’ end of active genes, while di-methyl-H3K4 has not been significantly associated to the 5’ end of active genes. To compare histone modifications between human and mouse, the authors also mapped dimethyl-H3K4 to some loci of mouse genome with analogous tiling mouse arrays. The locations of histone methylation show strong conservation between human and mouse. It is a challenge to investigate the genome-wide status of histone modifications in mammalian genomes because of the larger genome size and much more complicated genome architectures, as well as a large amount of repeat sequences and highly fragmented gene structures. However, current findings indicate that ChIP-chip can provide a wider view of histone modifications in mammalian genomes than gene-by-gene studies.

ChIP-chip was also applied to identify regulatory binding elements. Identifying DNA binding elements and regulatory networks are major challenges to decipher in normal and diseased cells. Conventional methods, such as footprinting and gel shift assays can be laborious[210, 211]. With the information of sequencing projects provided, computational biologists created various algorithms to identify regulatory elements *in silico*. These algorithms are mainly

based on comparative “phylogenetic footprinting”, coregulation clustering, or both. Phylogenetic footprinting requires the interrogating sequences of the species with appropriate evolutionary distance so that the ratio of conserved functional DNA to conserved non-functional background is significant enough for statistical identification[212]. The drawbacks of phylogenetic footprinting are that part of regulatory elements will be missed and non-functional conserved noise may be mistaken as conserved regulatory elements[213]. Even the putative motif may not be functional. Expression profiling data have been applied to clustering analysis and then the identified coregulated groups can be used to search for overrepresented regulatory sequence elements and regulation modules[214-216] However, as stated above, expression profiles cannot distinguish direct regulation from regulation indirectly. It is significant that in a direct and *in vivo* way, the genome-wide binding data from ChIP-chip can be used for bioinformatic analysis to identify binding elements. Based on ChIP-chip binding data from yeast, some new algorithms have been devised to extract binding motifs[217, 218] One can further combine the data from ChIP-chip, expression profiling, comparative genomic and published literature to reduce background noise and to identify true interactions[219]. These strategies have been used to construct the overview scenario of motif architecture. It was found that DNA binding sites are distributed in promoters from a single motif to more complicated cooperating motifs. This motif architecture is dynamic and responds to environmental stimuli. One drawback of ChIP-chip regarding motif identification is that the immunoprecipitated DNA by specific antibody may not be bound with protein

directly because formaldehyde crosslinks not only interacting protein-DNA but also protein-protein. Two alternative ways to avoid the interference from indirect DNA protein binding are DNA immunoprecipitation with microarray detection (DIP-chip) or protein binding microarrays (PBMs)[220, 221]. DIP-chip processes protein/DNA interaction *in vitro* with purified protein of interest and genomic DNA without formaldehyde crosslinking[220]. PBMs is one modified DNA microarray strategy using the transcription factor itself instead of precipitated DNA as template to hybridize the probes on DNA arrays, and then visualize the bound proteins by fluorescence[221].

The information of binding elements identified by ChIP-chip leads to the discovery of modules that can be used to build a global view of regulatory network. For example in yeast, it was shown that cell cycle transcription activators which function during one stage of the cell cycle have the ability to regulate transcription activators functioning in the subsequent stage thus forming a connected regulatory network[222]. Various regulation loops are formed from simple positive or negative loops to more complicated ones involving multiple input or output signals[223]. They can be further assembled into a global network providing information about the regulation pathways. In human stem cells, ChIP-chip was applied to reconstructing the core regulation circuitry based on location analysis of three transcription factors, OCT4, SOX2, and NANOG[224]. Meanwhile, efforts have been made to establish more sophisticated algorithms to extract regulation modules from ChIP-chip data[225-227].

ChIP-chip has also been extended or modified for other purposes. Some custom arrays contain oligonucleotides or fragments covering several kilobases of promoter regions, and can be applied to check the binding position of regulator or histone modifications pertaining to a single gene of interest[182]. Another custom tiling array centered on heterochromatic knob on the long arm of chromosome 4 has been used to study how transposable elements determine heterochromatin and regulate genes epigenetically [194]. To study gene expression, DNA methylation, and histone acetylation in parallel, a progeny array panel, called expressed CpG island sequence tags, was designed from CpG island arrays. This array contains DNA fragments located in the promoter and first exon region of genes[193]. With antibody against 5-methylcytidine instead of protein factor, another modified technique defined DNA methylation profile by immunoprecipitation followed with microarray analysis[228].

2.4 Other high throughput methods beyond ChIP-chip

Since the ideal array covering all human chromosomes is not available, alternative options have begun to address this limitation (Figure 2.2a). One way is to digest genomic DNA with restriction enzymes creating blunt ends and to clone the DNA fragments precipitated by ChIP into a plasmid vector[229]. One drawback of the cloning strategy is a high ratio of false positive targets caused by nonspecific inserts. In addition, cloning is difficult for large scale scanning across all the human genome because of high labor and expenditure cost. Another possible way is to establish SELEX (systematic evolution of ligands by

exponential enrichment) genomic library for sequencing, in which random amplification and gel purification are applied so that the target fragments with a selected size range can be used for library construction (Figure 2.2a)[230]. This ChIP-SELEX has been applied to identify targets of steroid receptor coactivator-3 with moderate success, although similar nonspecific questions remain[231].

Additionally, combining ChIP with modified serial analysis of gene expression, which can be termed ChIP-SAGE (also SACO, GMAT or STAGE), can be used (Figure 2.2a)[232-234]. In ChIP-SAGE, DNA is first pulled down by ChIP assay. The immunoprecipitated DNA is then applied to SAGE instead of microarray hybridization to create DNA library with ChIP DNA inserted. Once sequenced, all the bound DNA fragments will be counted. With ChIP-SAGE, unknown human targets of E2F4 have been identified which have not been discovered by CpG island ChIP-chip[234]. ChIP-SAGE has also applied to identify CREB binding sites, which are often situated in intronic regions of target genes. It has been shown that a large part of CREB targets, without CpG islands, can be identified with this strategy[232]. Another application of ChIP-SAGE is mapping H3 and H4 acetylation in yeast, and diacetyl-H3K9 and K14 in human T cells[233, 235]. In human T cells, ChIP-SAGE showed histone acetylation in a distributed pattern across the whole genome, with active chromatin associated with acetylation islands[235]. Even though ChIP-SAGE has its own advantages over ChIP-chip, the requirement for extensive DNA sequencing makes it less convenient than ChIP-chip. In general, ChIP-SAGE may be used as a complementary choice to the ChIP-chip technique.

Immunoprecipitation-based ChIP-chip requires high quality antibodies, which may be unavailable, especially for newly emerging protein factors. One alternative technique, called targeted gene methylation (TAGM) or DNA adenine methyltransferase identification (DamID), utilizes the DNA methyltransferase to "mark" the positions of DNA-protein interactions (Figure 2.2b)[236, 237]. Dam, a DNA methyltransferase, is fused with the DNA binding factor of interest and the recombined protein is then expressed in host cells. The binding of fusion proteins with target DNA will lead the fused Dam to target and methylate DNA in the vicinity. By comparing this DNA with that from cells transfected only with Dam, the positions bound with Dam/fusion protein can be identified by DNA microarray hybridization. This technique has been applied successfully in the *Drosophila* system[237-241]. However, DamID has limitations over ChIP-chip. First, it requires more time to express the fusion protein than formaldehyde crosslinking of endogenous protein. Secondly, DamID cannot be applied for detecting post-translational changes, such as histone modifications. Thirdly, the resolution of DamID is less precise than that of ChIP-chip, even though both of the methods can produce similar binding maps[242].

ChIP-chip is applied to extrapolate the crosstalk between DNA and protein, which regulate the process of transcription (Figure 2.2a). Beside transcription-level regulation, RNA becomes modified, transported between subcellular compartments and translated into protein. Crosstalk between RNA and protein will regulate these processes. Because this is a necessary process in the cell, a strategy combining RNA immunoprecipitation (RNA-IP) with microarray

has been applied to understand these RNA-protein interactions (Figure 2.2c)[243, 244]. The success of ChIP-chip helped promote the integration of immunoprecipitation and microarray for investigating RNA-protein interactions. The immunoprecipitated RNA was first combined with mouse and yeast microarrays containing cDNA or open reading frames[243-245]. Then other arrays were used in RNA–IP microarrays, including human cDNA and whole-genome yeast arrays[246-250]. With these strategies, various RNA-binding protein factors, especially those mediating RNA localization and exporting have been studied. Subsets of transcripts linked to specific factors have been identified, and portray a complicated posttranscriptional regulatory network.

2.5 Perspective on future development of ChIP-chip

Generally, the ChIP-chip protocol can be completed in three steps: ChIP, post-ChIP DNA handling and microarray analysis. Some modifications can be made. Although formaldehyde crosslinking is currently used, other crosslinking methods may be applied. UV treatment has been used to study the genome-wide protein-DNA interaction, even though no comparison between UV and formaldehyde crosslinking has been reported[251]. However, formaldehyde crosslinks both DNA-protein and protein-protein interactions. In contrast, UV treatment crosslinks only protein-DNA without involving the interactions of protein-protein. This may be of help to selectively detect the direct interaction of protein and DNA without protein-protein interference[252]. Meanwhile, to make sure that there is a stable coupling between the protein complex and DNA, other

chemicals besides formaldehyde have been used to intensify the crosslink between proteins[253]. Other covalent trapping of protein-DNA complexes, such as the covalent binding of DNA methyltransferase and 5-aza-2'-deoxycytidine, may be modified in order to provide another marker for *in vivo* mapping[28, 254].

Unbiased amplification of ChIP DNA would be beneficial to data quality from ChIP-chip. Efforts to amplify genomic DNA homogenously include carefully designed amplification processes, adding chemical additives, and application of more efficient enzymes[196, 255-259]. For example, Phi29 DNA polymerase can amplify the whole genome DNA in a linear way without significant bias, and has been used in microarray studies[258, 259]. Recently, one commercial technique called chromatin Immunoprecipitation-guided ligation and selection (ChIP-GLAS), which allows for the comparison between ChIP DNA samples before PCR amplification, may bypass amplification bias and interference from repeat sequences (http://www.avivasysbio.com).

As we have discussed above, false positives may be frequent in ChIP-chip[182] [184]. In two reports with similar yeast arrays, one group found that 20% of Rpd3 targets have Ume6 binding sites, while the other study showed that only limited numbers of genes are targeted by both Rpd3 and Ume6[197, 199]. It cannot be ruled out that background noise caused these inconsistencies. For screening purposes, high false positive rates may not cause fatal concern, because specific targets are confirmed by ChIP-PCR. However, for profiling or mapping purposes, this can be a serious issue. Generally, the top list of enriched targets by ChIP-chip is used to confirm the data quality. This, however, cannot

guarantee that the final map reflects the real spectrum of target distributions. Considering that these targets are dynamic in DNA-protein binding, we should be cautious when interpreting ChIP-chip data. Optimizing the current ChIP-chip experiments, or further developing the technique, may make ChIP-chip mapping more convincing. Meanwhile, incorporating some stringent statistical parameters may be helpful to perform quality control for ChIP-chip. Recently, optimization of ChIP-chip decreased the false positive rate to less than 1% and false negative rate to 20-25%[224, 260].

Currently, data analysis used for ChIP-chip analysis is heavily adopted from expression microarray studies and as such similar assumptions are made. For instance, one issue of ChIP-chip analysis is how to normalize raw data. For expression array, the assumption is made that only portions of transcripts are changed in the cell tested compared with the control, so that the total signal of transcripts from the test sample and the control should be equal. ChIP-chip, however, is based on antibodies and the DNA pulled down, such as those against universal histone modifications, may be significantly changed, making the assumption of proportional pull-down invalid, as in normal expression microarray. Currently, attention is focused on creating algorithms to utilize the ChIP-chip data; to combine ChIP-chip with other platforms; and to finally establish some predictable models of transcription regulation. In addition, the basic work of design, application and analysis of ChIP-chip itself may be equally important to improve the reliability of these models.

One question is how to isolate functional parts of genomes for array making. Development of various genomic microarray platforms will promote the utility of ChIP-chip. The slides for the mouse CpG islands are now under development. Two recent studies may provide new clues to establish promoter arrays. In one study, yeast genome was fractioned into coding and noncoding regions by formaldehyde crosslinking and phenol-chloroform extraction. The promoters upstream heavily transcribed genes are efficiently enriched in the noncoding fraction[261]. In another study, ChIP with the pre-initiation complex, pulls down known and unknown putative promoters from the human cell line[191, 192]. Barring nonspecific background, these enriched promoter fragments may be applied to establish promoter libraries and arrays for further ChIP-chip study.

Combination of ChIP-chip with other high throughput methods may be beneficial. This means both combining multiple ChIP-chips and incorporating ChIP-chip with other high throughput methods. *In vivo*, yeast RNA polymerase III forms a protein complex with TFIIIC and TFIIIB, and is responsible for transcription initiation of small untranslated RNAs. With multiple ChIP-chips, the proteins, C160, Brf1 and Bdp1, t95, components of Pol III were colocalized with TFIIIC and TFIIIB at 94% of tRNA genes[262]. This combination can help validate the ChIP-chip data and also provide information how the Pol III transcription machinery corroborates to initiate gene transcription. Since the binding of protein factor with DNA targets does not mean it is functional, the incorporation of ChIP-chip with expression array is applied frequently to provide both binding and functional information of the protein factor on its target[226,

263, 264]. Another example is the combination of ChIP-chip with DMH, one method of detecting DNA methylation as mentioned above[163, 193, 259]. Recently, the yeast two hybrid has been applied to create the network of protein-protein interactions, and this proteome-map can be connected to the transcriptome[265, 266]. Theoretically, the information of protein-DNA crosstalks from ChIP-chip, combined with that from protein-protein crosstalks, can provide input signals to reconstruct the regulatory network. The expression array may indicate the output signals from this network. It is promising that along with the appropriate computational strategy, emerging high throughput methods from different platforms can work cooperatively to produce a clearer picture of the regulatory network, and disease-related changes.

2.6 Conclusion on ChIP-chip

The success of ChIP-chip is not surprising. New biological questions continue to drive the development of new techniques. DNA sequencing, expression microarray, proteomic 2D electrophoresis gel, and other system tools have helped understand the structure and amount of cell components. New genome-wide, high throughput tools, like ChIP-chip, are necessary to study the activities of key components, like epigenetic modifications and DNA-protein binding in cells. ChIP-chip has been frequently used in basic biological studies and may be modified further and expanded to other aspects, such as human diseases. Finally, the large amount of discoveries by ChIP-chip, and other high

throughput techniques, may be connected and organized with emerging bioinformatics to add to our knowledge on life and diseases.

Figure 2.1: A general ChIP-chip procedure. Formaldehyde is used to crosslink the DNA and interacting protein in cells. Cells are sonicated to shear DNA stands to a desired length. After sonication, immunoprecipitation is performed to pull down specific DNA protein complexes. In parallel, control DNA, which has not been immunoprecipitated by antibody, is used as the reference control. The DNA protein complex is then reversed at 65^{o}C and the DNA is eluted. Eluted DNA and control DNA are labeled with different fluorescent dyes and hybridized to a microarray. The hybridized slide is then scanned and analyzed to identify the enriched targets from immunoprecipitation.

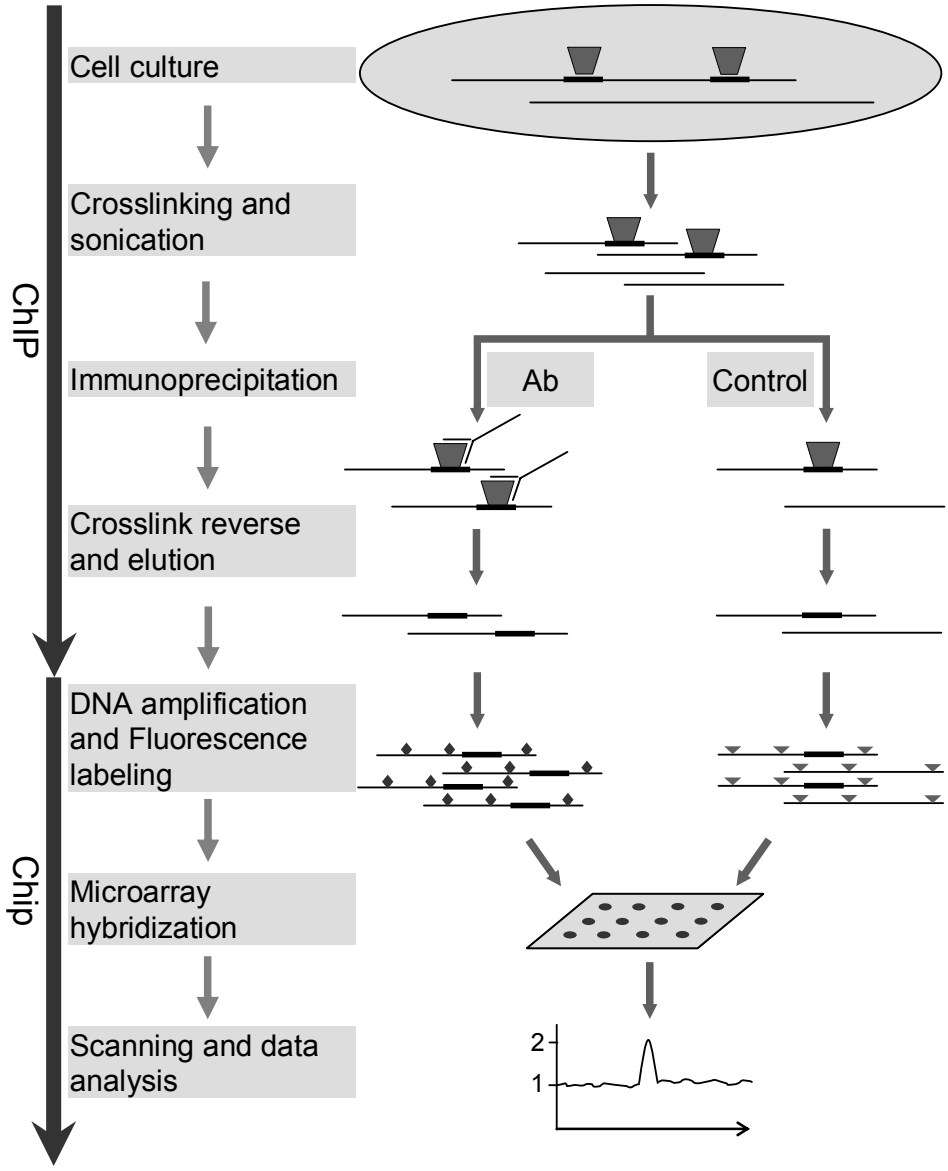
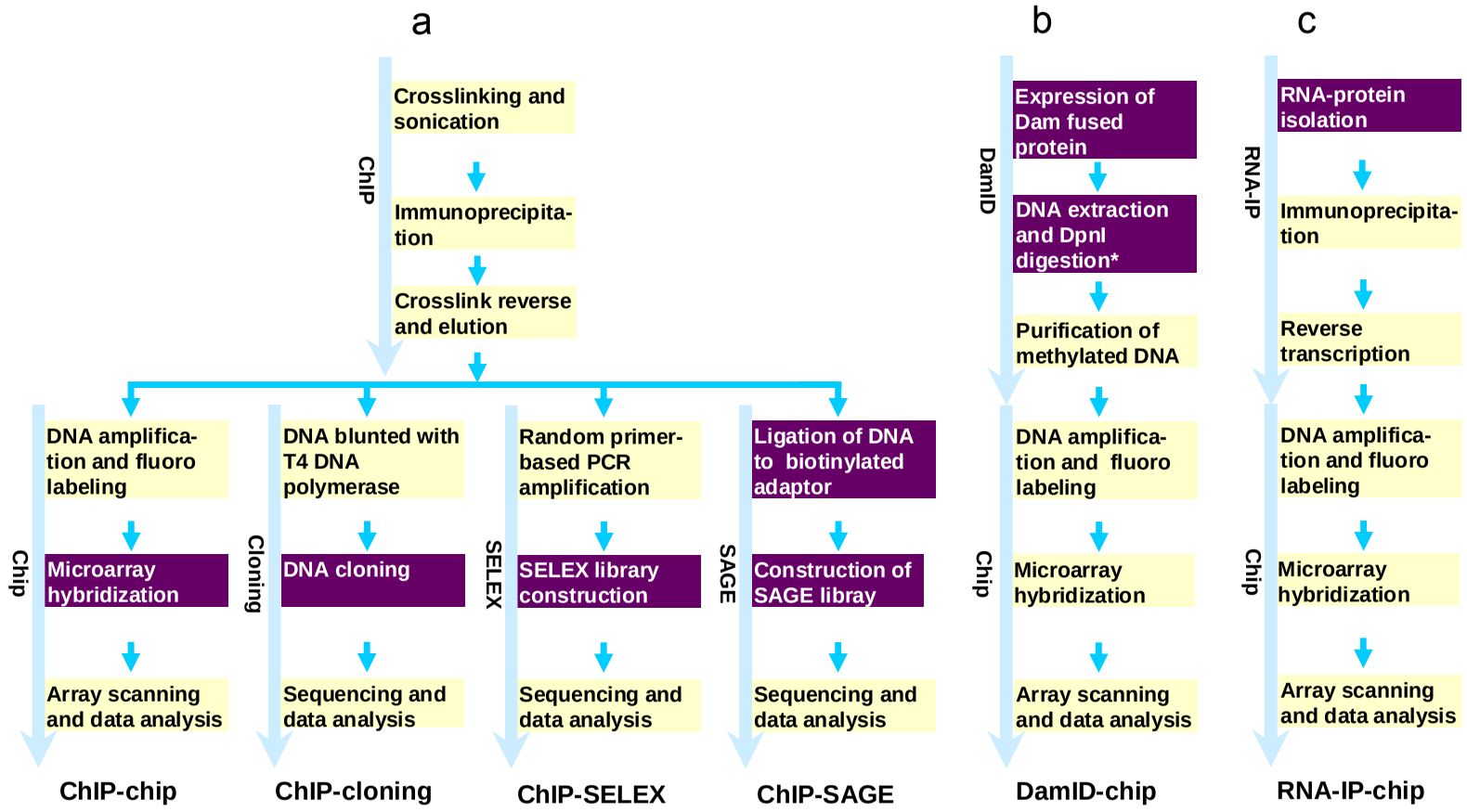

Cell culture
Crosslinking and sonication
Immunoprecipitation
Ab
Control
Crosslink reverse and elution
DNA amplification and Fluorescence labeling
Microarray hybridization
Scanning and data analysis
ChIP
Chip
2
1

Figure 2.2. Comparison of the procedures of ChIP-chip with other high throughput techniques for studying DNA-protein or RNA-protein interactions. The procedures of figure a-c are general principles. Key steps are highlighted with dark grey text box. DpnI (*star*) is a methylation sensitive restriction enzyme.


a
b
c
ChIP
Crosslinking and sonication
Immunoprecipita-tion
Crosslink reverse and elution
Chip
DNA amplifica-tion and fluoro labeling
Microarray hybridization
Array scanning and data analysis
ChIP-chip
Cloning
DNA blunted with T4 DNA polymerase
DNA cloning
Sequencing and data analysis
ChIP-cloning
SELEX
Random primer-based PCR amplification
SELEX library construction
Sequencing and data analysis
ChIP-SELEX
SAGE
Ligation of DNA to biotinylated adaptor
Construction of SAGE libray
Sequencing and data analysis
ChIP-SAGE
DamID
Expression of Dam fused protein
DNA extraction and DpnI digestion*
Purification of methylated DNA
Chip
DNA amplifica-tion and fluoro labeling
Microarray hybridization
Array scanning and data analysis
DamID-chip
RNA-IP
RNA-protein isolation
Immunoprecipita-tion
Reverse transcription
Chip
DNA amplifica-tion and fluoro labeling
Microarray hybridization
Array scanning and data analysis
RNA-IP-chip

CHAPTER 3

DIVERSE HISTONE MODIFICATIONS ON HISTONE 3 LYSINE 9 AND THEIR RELATION TO DNA METHYLATION IN SPECIFYING GENE SILENCING

3.1 Introduction

It is well known that DNA methylation plays a repressive role in gene transcription, both in heterochromatin and in repressed, protein-coding, euchromatin [132]. Recent work demonstrated that DNA methylation cooperates with histone modifications to perform this repressive function [133]. Acetylation and methylation on histone 3 lysine 9 (acetyl-H3K9 and methyl-H3K9, respectively) are two of the best studied modifications. Acetyl-H3K9 is known to be associated with active transcription, and methyl-H3K9 with repressed transcription [82]. To better understand the mechanisms of epigenetic regulation, it is necessary to clarify the crosstalk, including the distribution patterns, between these epigenetic markers [82]. Some reports showed the physical interaction between histone deacetylase and histone methyltransferase [125, 267]. Meanwhile, removal of acetylation has been shown to be a necessary step for histone methyltransferase activity [125, 267]. It is believed that histone acetylation and histone methylation act in concert to regulate gene transcription.

Studies in fungi and plant, and, to a lesser degree, in mammals indicate that methyl-H3K9 may control DNA methylation in heterochromatin [138-140]. Current knowledge also supports the idea that repressive complexes, containing both methyl binding domain (MBD) proteins and histone deacetylases (HDACs), in combination with other repressor proteins, direct DNA methylation and subsequently transcriptional repression [137]. Additional evidence shows that DNA methylation impacts histone methylation and that DNA methylation might exert a positive feedback on lysine methylation [142-146]. To reconcile these two seemingly distinct mechanisms, a self-enforcing network of epigenetic regulation has been proposed: histone methylation impacts DNA methylation and histone acetylation which in turn impacts histone methylation [82, 137].

The current self-enforcing model implies close correlation of histone modifications and DNA methylation, especially the crosstalk between methyl-H3K9 and DNA methylation [82]. However, recent reports demonstrated that these epigenetic markers have varying degrees of autonomy [142, 269-276]. For example in *Arabidopsis*, Trichostatin A (TSA), a histone deacetylase inhibitor, and 5'-aza-2-deoxycytidine (AzadC), a demethylating agent, do not always produce redundant outcomes. Most surprisingly, they may even demonstrate antagonistic effects as opposed to the expected synergistic effects [268]. In *Arabidopsis,* where DNA methylation is not crucial for survival, methyl-H3K9 marks heterochromatin independent of DNA methylation [142]. In mammals, no close association between methyl-H3K9 and DNA methylation was discovered for: imprinted gene loci on distal chromosome 7 [269, 270], the inactive X

chromosome in ICF and Rett syndrome cells [271], *FMR1* in fragile X syndrome [272], or *MGMT, LHR* in human cancers [273, 274].

Here we analyze profiles of DNA methylation and histone modifications in a mouse leukemia genome to better understand the relationship between these epigenetic events. Using genome-wide data, we demonstrate that histone acetylation and histone methylation show a distinct degree of autonomy with respect to promoter methylation.

3.2 Methods

Cell culture

Mouse leukemia cell line L1210 (American Type Culture Collection, Manassas, VA) was grown in Dulbecco's modified Eagle's medium (Cellgro, Herndon, VA) plus 10% FBS in plastic tissue culture plates in a humidified atmosphere containing 5% CO_2 at 37°C. The cells were grown to 90% confluency before being harvested.

Chromatin immunoprecipitation microarray (ChIP-chip)

Five millions of L1210 cells were crosslinked with 1% formaldehyde for 10 min, and then 0.125 M glycine was used to stop the crosslinking. Chromatin immunoprecipitation was performed by using ChIP assay kit (Upstate Biotechnology, Charlottesville, VA) as described previously [183]. The antibodies against acetyl-H3K9 (AcH3K9, 06-942) and dimethyl-H3K9 (diMeH3K9, ab-7312) were purchased from Upstate Biotechnology (Charlottesville, VA) and Abcam

(Cambridge, MA), respectively. Pooled DNA (up to 10) from multiple ChIPs and input DNA were labeled by Cy5 and Cy3 fluorescent dyes (Amersham, Buckinghamshire, UK) and then were cohybridized to the mouse 9.2k CpG island array (UHN microarray center, Ontario, Canada) or mouse custom array. Post-hybridization washes were performed as previously described [275]. The washed slides were scanned by a GenePix 4000A scanner (Axon, Union City, CA), and the acquired microarray images were analyzed with GenePix Pro 6.0 software (Axon, Union City, CA). Duplicate hybridizations were performed for each antibody and the quality of replicate chips was examined by scatter plot and Pearson's correlation analysis (from 0.77-0.82) [275]. After excluding the spots flagged for bad quality, normalized Cy5/Cy3 ratios of these loci were calculated by GenePix Pro 6.0 [183].

Differential methylation hybridization

Differential methylation hybridization (DMH) was performed essentially as described ([183, 275]). Briefly, 2 μg of genomic DNA were digested by *Mse*I to produce small fragments and then H-24/H-12 PCR linkers (5' – AGGCAACTGTGCTATCCGAGGGA

T-3' and 5'-TAATCCCTCG-GA-3') were ligated to the digested DNA fragments. The DNA samples were further digested with two methylation-sensitive endonucleases, *Hpa*II and *Bst*UI, and amplified by PCR using H-24 as a primer. After amplification, DNA from L1210 and DBA2 was labeled with Cy5 and Cy3

dye, individually. Hybridization and later analysis were performed as described above in ChIP-chip section.

Mouse custom array

PCR was performed to amplify the promoter regions (-700 bp to +300bp from the transcription start site) with mouse genomic DNA as template (see table 3.1). To ensure the reproducibility of each PCR and to prevent nonspecific amplification, PCR products (500-bp on average) were individually verified by 1.2% agarose gel electrophoresis. PCR products and the control repetitive DNA were then mixed with 50% dimethylsulfoxide and spotted in triplicate to GAPS II coated slides (Corning, Acton, MA) by Affymetrix/GMS 417 Arrayer (Affymetrix, Santa Clara, CA). Arrays were incubated in a desiccator overnight. Spotted DNA was rehydrated by holding slides over boiling water for 5 seconds and then placed on a hot plate for 2 seconds. UV (300mJ) cross-linking was used to immobilize spotted DNA. Slides were then stored in a desiccator at ambient temperature.

Restriction landmark genome scanning (RLGS)

High molecular weight DNA was extracted from L1210 cells and normal DBA2 mouse tissue. Subsequently, RLGS was performed as previously described [276]. Paired RLGS profiles, obtained from L1210 and DBA2, were overlaid and the difference between the two profiles was detected by visual inspection. Analysis was independently validated by at least one additional

investigator. All selected targets for ChIP-chip were analyzed by comparing the RLGS profiles from L1210 and DBA2.

Combined bisulfite restriction analysis (COBRA)

In vitro methylated DNA (representing 100% methylated DNA) and the DNA from a DBA2 mouse (representing 0% methylated DNA) were used as controls. Two micrograms of DNA from L1210 cells was treated with 3M sodium bisulfite overnight and then amplified by PCR. Primers were designed to amplify both methylated and unmethylated alleles of sodium bisulfite-treated DNA. PCR products were purified by the gel extraction kit (Qiagen, Valencia, CA) and then digested by *Bst*UI (CG↓CG) restriction enzyme (NEB, Ipswich, MA). The digested fragments were separated on an 8% polyacrylamide gel. The primers are listed in Table 3.3.

Chromatin immunoprecipitation- quantitative polymerase chain reaction

ChIP was conducted the same way as in ChIP-chip. DNA pool from ChIP and input control was first measured by spectrophotometer (NanoDrop, Wilmington, DE). Quantitative PCR with SYBR green-based detection (Applied Biosystems, Foster City, CA) was performed as described previously [277]. In brief, primers are designed according the promoter structure of selected genes (Figure 3.5A). Quantitative ChIP-PCR values were normalized against values from a standard curve (50 to 0.08 ng, R^2>0.99) constructed by input DNA with the same primer sets. The primers are listed in Table 3.3.

Trichostatin A (TSA) and 5-aza-2'-deoxycytidine (AzadC) treatment

Cells were split the day before treatment and then treated with TSA (Sigma, St. Louis, MO), AzadC (Sigma, St. Louis, MO) or the combination of the two drugs. 1, 2.5 and 5μM of AzadC in dimethylsulfoxide was applied to cells every 24 h for 1, 3 or 5 days. 300 nM TSA in demethylsulfoxide was used to treat cells for 24 h. For combination treatment, 1μM of AzadC daily for 1, 3 or 5 days was followed with 300 nM TSA for 24 h. Cells treated with medium containing dimethylsulfoxide served as a control.

Quantitative reverse transcription-polymerase chain reaction

Total RNA was extracted from drug treated and untreated cells. Two μg RNA was first treated with DNase I (Invitrogen, Carlsbad, CA) to remove potential DNA contamination and then was reverse transcribed with SuperScript II reverse transcriptase (Invitrogen, Carlsbad, CA). Quantitative RT-PCR was performed by using SYBR green (Applied Biosystems, Foster City, CA) as a marker for DNA amplification on a 7500 Real-Time PCR System apparatus (Applied Biosystems, Foster City, CA). The relative mRNA level of a given locus was calculated by relative quantitation of gene expression (Applied Biosystems, Foster City, CA) with GAPDH mRNA (based on amplification efficiency) as an internal control.

3.3 Results

Global profiling of acety-H3K9, dimethyl-H3K9, and DNA methylation in L1210 cells

We first performed chromatin ChIP-chip on the mouse leukemia cell line, L1210, with antibodies detecting either acetyl-H3K9 or dimethyl-H3K9. ChIP products were hybridized onto the mouse 9.2K CpG island microarray. Immunoprecipitated DNAs from acetyl- or dimethyl-H3K9 ChIPs were compared individually with total genomic DNA input. Increased hybridization signals indicated an enrichment of a specific histone modification for a given CpG island locus (red signals in Figure 3.1A and B). The scatter plot, with fold changes plotted against geometric mean of signal intensities, showed that the relative level of acetyl-H3K9 has a wider range of distribution than the intensity index seen for dimethyl-H3K9 (Figure 3.1C and D).

Next, DMH was performed using the mouse CpG island microarray. Because L1210 originated from the mouse strain DBA2, we used genomic DNA derived from this mouse strain as a control for assessing DNA methylation in L1210. The DMH assay was used to evaluate the methylation status of *Bst*UI and *Hpa*II sites, located within or nearby CpG islands. A 2-fold increased intensity was used as a cutoff for scoring positive loci for DNA methylation [275]. Those loci scoring positive (>2-fold) in DMH and/or ChIP-chip data were then used to compare the acety-H3K9 and dimethyl-H3K9 levels against their promoter methylation level (Figure 3.2). Overall, we observed a trend that high levels of acetyl-H3K9 (>2-fold) were preferentially present in unmethylated CpG islands

while acetylation levels less than 2-fold are correlated with hypermethylated loci in L1210 (Figure 3.2A). However, the distribution pattern of dimethyl-H3K9 against DNA methylation status was not as distinct as that of acetyl-H3K9 in this cell line (Figure 3.2B).

Subpanel profiling of acety-H3K9, dimethyl-H3K9, and DNA methylation in L1210 cells

To independently confirm these genome-wide findings, we focused the epigenetic analysis to a subset of promoter CpG islands. We first used the restriction landmark genome scanning (RLGS) technique to identify hypermethylated loci in L1210 compared to the control DBA2. Loss of CpG island sequences in RLGS indicates potential *Not*I hypermethylation present in this leukemia cell line (Figure 3.3) [278]. Of the 1300 sites screened, we identified a total of 435 (or 33%) RLGS fragment losses. The identified DNA methylation pattern of L1210 was similar to profiles obtained from the leukemia samples derived from a mouse model of NK/T acute lymphoblastic leukemia, with numerous commonly methylated sequences (data not shown) [164].

We then used a subset of 71 (including 54 hypermethylated loci and 17 unmethylated sequences) promoter CpG islands identified in RLGS to establish a custom mouse promoter array for ChIP-chip assays. The 5' fragments of these targets, including their transcription start site or *Not*I restriction site, were amplified by PCR and printed on glass slides (see Table 3.1 for location). Immunoprecipitated DNAs from L1210 using antibodies against acetyl- or

dimethyl-H3K9 were then used to hybridize this custom microarray panel. The results showed that the level of acetyl-H3K9 had a wider range of distribution than that of dimethyl-H3K9 (Figure 3.4A). Similar to the results obtained from the global microarray, the dimethyl-H3K9 level of these 71 loci was not as high as expected even though most of the targets were methylated in L1210. To clarify the relationship between histone modifications and DNA methylation, we plotted the distribution patterns of histone acetylation or methylation *versus* DNA methylation in these 71 loci (Figure 3.4B). Consistent with the previous reports [65], this interval plot analysis showed that acety-H3K9 was reversely correlated with DNA methylation ($p<0.01$). However, dimethyl-H3K9 and DNA were not found to be significantly correlated. Altogether, the results confirmed the aforementioned genome-wide findings that acetyl-H3K9 and dimethyl-H3K9 may have distinct autonomies with respect to DNA methylation in this subpanel of loci in L1210 cells.

Confirmation of acety-H3K9, dimethyl-H3K9, and DNA methylation profiles in individual CpG island loci

To further confirm the DNA methylation status and histone modifications in individual genes, combined bisulfite restriction analysis (COBRA) and ChIP-PCR were performed on 12 genes chosen from the subpanel list (Table 3.2 and Figure 3.5A). Two known genes, *p19ARF* and *ID4*, were used as unmethylated and methylated controls, respectively (Figure 3.5)[164, 279]. Of these selected genes, six were deemed active, while the rest were inactive, according to their

associated epigenetic marks (Table 3.2). Of the 12 targets examined, the methylation status of 10 genes was confirmed in L1210 cells (Figure 3.5B). Two genes (*BC011343* and *Dscaml1*), found to be methylated by RLGS (Table 3.1), were not confirmed by COBRA, which may be caused by different restriction enzymes used in these two methods. The COBRA data showed an "all or nothing" methylation status in the promoters of targets examined. In the tested sites of these promoters, there seems to be no partial methylation detected by COBRA, which facilitates our further analysis.

ChIP, followed by real-time PCR, was applied to evaluate both acetyl- and dimethyl-H3K9 enrichment in the promoter regions of these targets. The enrichment levels of these 12 loci were compared between ChIP-DNA and genomic DNA (Figure 3.5C). Eleven of the 12 loci showed the same trend of histone modifications as derived by ChIP-chip assays (Table 3.2 and Figure 3.5C). In this regard, the acetyl-H3K9 level was inversely correlated with the DNA methylation status. Not surprisingly, the dimethyl-H3K9 level showed no significant difference with respect to the DNA methylation status of these loci, suggesting that the acquisition of dimethyl-H3K9 is less dependent on DNA methylation in the protein-coding genes than acetyl-H3K9. Meanwhile, the overall enrichment of acetyl-H3K9 varied greatly, but dimethy-H3K9 varied to a lesser extent, which is consistent with our ChIP-chip data. The results of the two control genes, *p19ARF* and *ID4*, are consistent with our other data (Figure 3.5C).

The chromatin landscape of the promoter regions (6-kb) in two genes, *Ran* and *Zic3*, was analyzed in greater detailed by ChIP-PCR. Twelve sets of

primers were used in ChIP-PCR to cover their promoter regions from -3.5 to 2.5-kb away from the respective transcription start sites. The overall levels of acetyl-H3K9 were lower than those of dimethyl-H3K9 in the promoter region of the inactive *Zic3* gene, but significantly increased in the active *Ran* gene in L1210 cells (Figure 3.6). In contrast, dimethyl-H3K9 levels were only slightly higher than those of acetyl-H3K9 in most parts of the interrogating regions, except in some regions of the active *Ran* promoter. The highest level of dimethyl-H3K9 was found in transcribed regions but not in the promoter of repressed genes.

The effects of Trichostatin A (TSA) and 5-aza-2'-deoxycytidine (AzadC) on gene re-expression

Four genes were studied following treatment with TSA, a histone deacetylase inhibitor, and AzadC, a DNA methyltransferase inhibitor. The expression of four selected genes, *Tjp1*, *Zic3*, *Ran* and *Cog8*, were examined under different dosage of treatments for 1, 3 and 5 days. *Zic3* and *Tjp1* were associated with repressive epigenetic markers (Figure 3.5B and C), and their transcription in L1210 could not be detected with quantitative RT-PCR prior to drug treatment (Figure 3.7A and B). AzadC alone was able to de-repress *Tjp1*, but not *Zic3*. TSA alone did not reactivate *Tjp1* or *Zic3*. However, a combination of TSA and AzadC was able to activate the expression of *Zic3*, and showed a synergistic effect. For *Tjp1*, the addition of TSA, however, had no or little additional effect on transcription, and its re-expression was due to AzadC treatment. Even though histone acetylation and DNA methylation were closely

correlated for both *Tjp1* and *Zic3*, these two epigenetic markers may affect gene function to differing extents. These results are also consistent with previous reports that DNA methylation is a dominant repressive factor, and that TSA alone may not de-repress gene transcription if the gene is densely methylated [280]. Two genes, *Ran* and *Cog8*, were expressed in L1210 and not methylated in their promoters (Figure 3.5B and C). The expression of *Ran* was increased under low concentrations (1-5 μM) of AzadC at day 1. However, both *Ran* and *Cog8* showed reduced expression to varying degrees following prolonged treatments (5 days) of AzadC and/or TSA (Figure 3.7C and D).

3.4 Discussion

The relationship between histone modifications and DNA methylation in maintaining gene silencing has been studied at the chromosomal level [281]. The model shows that a cooperation between methyl-H3K9 and DNA methylation is found in heterochromatin regions and major satellite repeats [281]. In euchromatic regions or at the individual gene level, controversial results have been reported in regards to the distribution and function of histone methylation in mammals [269, 270, 272-274].

Using genome-wide profiling techniques in L1210 leukemia cells, our present study shows distinct levels of autonomy in histone modifications in relation to DNA methylation of multiple protein-coding genes. Specifically, we demonstrate an inverse relationship between DNA methylation and histone acetylation in regulating transcription of these genes in mouse leukemia cells.

However, methyl-H3K9 seems to be ambiguous in specifying silencing of some genes tested. Our findings might not fit into the model that histone methylation and DNA methylation are closely corollated in maintaining the repressive state of genes [82]. It should be noted that the establishment of this prior model is solely based on the observation of a few genes [143, 144, 146, 282, 283]. The present findings, however, are based on genome-wide profiling of these epigenetic components in multiple genes. Here we would like to propose an alternative model, in which histone methylation is distributed throughout the whole genome, including transcribed regions, and can be reversed by histone demethylase. However, DNA methylation is the final repressive lock, which can not be easily removed. In the “context” of stabilized chromatin, regions of histone acetylation “islands” are used to keep the active conformation at specific positions.

In addition to the above explanation, one additional suggestion is that the promoter region of protein-coding genes is not the prime target for this histone modification (i.e., methyl-H3K9). Recent discoveries have shown that the mouse promoter regions of *hemoglobin beta adult major chain* and *GATA-2* have lower levels of both di- and tri-methylation of H3K9 than those in major satellite repeats and transcribed regions [284]. It is possible that regulatory mechanisms of histone methylation of H3K9, especially its crosstalk with DNA methylation, are different depending on chromatin locations and other unknown factors.

Our data shows the diverse status of histone modifications in relation to DNA methylation in mouse leukemia cells, providing new clues to the understanding of epigenetic regulation in mouse genome. In this regard,

epigenetic components that specify active or inactive chromatin play different roles, but are cooperative, under different circumstances during mammalian development. Crosstalk between H3K9 methylation and DNA methylation are evolutionarily conserved from fungi to plants to mammals [82, 122]. While genetic studies have shown that while H3K9 methylation is completely responsible for establishing DNA methylation in heterochromatic regions of *Neurospora*, this correlation is only partially established in *Arabidopsis* [138, 139]. Meanwhile, some reports have shown that the distribution of histone methylation is dependent on DNA methylation in plants, but not in fungi [138, 285]. In mouse embryonic stem cells lacking Dnmt1, Dnmt3a or Dnmt3b, no trimethyl-H3K9 redistribution is observed [140]. In double-null mouse ES cells for Suv39h, a histone methyltransferase, DNA methylation profiles are only changed in pericentric satellite repeats, but not in other repeat sequences [140]. Another histone methyltransferase, G9a, specifically affects imprinted genes depending on the development of embryonic stages [283]. Other gene studies have also produced controversial results between the correlation of histone modifications and DNA methylation [143, 286, 287]. From these studies, it is obvious that epigenetic redundancy, resulting from complex interactions among different chromatin components, is implemented to safeguard the stability of repressed chromatin structure.

Possible “heterogeneity” of epigenetic regulation is also revealed by our TSA/AzadC treatment study. Repressive epigenetic marks may be different in the 4 genes analyzed, as the same drug treatment produced differential effects of

expression in these loci. Meanwhile, alternative pathways may exist for TSA or AzadC that affect their upstream regulators genes, which also regulate the expression of these genes.

Here we need to keep in mind that only dimethylation of H3K9 was examined in this study, and further investigation is essential to delineate the relation of mono- and trimethyl-H3K9 methylation to DNA methylation. It is known that in mouse, different types of methylation at H3K9 are distributed with various patterns in chromatin [82, 122]. For example, trimethyl-H3K9 is over abundant in heterochromatin, whereas mono- and dimethyl-H3K9 are predominantly in euchromatin. Since we were more interested in the epigenetic modifications in euchromatin, dimethyl-H3K9 was selected in this study. Both mono- and trimethyl-H3K9 methylation deserve further study so that “heterogeneity” of epigenetic regulation can be well understood. We selected a single mouse leukemia cell line, L1210, in this study and conclusions drawn in this system will need to be validated in other mammalian cells.

ChIP-chip, RLGS, and DMH are genome-wide techniques, which can be readily applied to epigenomic studies. CpG island arrays have been widely used in human epigenetic studies [288]. However, very little work has been done combining mouse CpG island arrays with other epigenomic tools. Current knowledge of crosstalk between histone modifications and DNA methylation comes mainly from a series of experimental strategies, including complex interaction, genetic studies and sequence characterization [137]. In addition, one main direction is to delineate specific epigenetic marks that implicate cellular

functions, such as cell-lineage determination and stem cell differentiation. These challenges require epigenomic tools, as those described in this study and two recent reviews [82, 122]. Studies have described the use of ChIP-chip to investigate the correlation between histone modifications and gene transcription from yeast to human [162]. Because of technical limitations, few epigenomic tools have been reported in the mouse. The implementation of CpG island and custom microarrays makes it possible to interrogate complex epigenetic networks in mammalian systems.

Spot on RLGS	GB Accession	Name	NotI site	Not I position[a]	CpG island	chip position[b]	mouse leukemia(8)
5F75	AF002701	Gfra2	5' end	-114	Yes	-543, 87	8[c]
3F70	AF105268	Gpc6	5' end	1062	Yes	-295, 254	8
3F56	U36760	Foxg1	5' end	662	Yes	1353, 1900	7
6D12	AK080865	mRNA	5' end	518	Yes	-494, 37	7
2H54	BE863773	EST	5' end	-294	Yes	-291, 162	7
4E24	AB041591	Slc16a11	5' end	482	Yes	-54, 417	7
5F70	AF288666	Sema6a	5' end	88	Yes	-316, 211	7
5F59	AB021964	Igsf4	5' end	806	Yes	-406, 122	7
2F37	BC063061	Sox3	5' end	1087	Yes	-250, 258	7
2G41	X75018	Id4	5' end	210	Yes	-529, 4	7
3F09	AF334801	Pcdh10	Body	2960	Yes	-486, 45	6
2G81	AK047294	Cstad	Body	10072	No	-215, 85	6
3C06	No EST	-	-	chr1:15,464,778	Yes	Chr1:15,464,320-15,464,777	6
3E75	AJ010949	Cacna2d3	5' end	-777	Yes	-310, -845	6
1F21	D70849	Zic3	5' end	676	Yes	195, 526	6
2G50	AY371925	Mamdc1	5' end	-1144	Yes	-499, 44	6
5G21	D14340	Tjp1	5' end	525	Yes	-290, 251	6
4F59	BC014822	Ephb3	5' end	-34	Yes	-257, 249	6
4G54	BC065168	Kbtbd9	5' end	571	Yes	26, 517	5
1F30	AK089717	mRNA	Body	39864	Yes	39447, 39940	5
4D49	D83206	Vmp	5' end	-47	Yes	-249, 247	5
5E02	AK041673	mRNA	5' end	451	Yes	-784, -282	5
5D36	AF191211	Nr4a3	5' end	684	Yes	-111, 429	5
3D03	AF263913	Fign	5' end	-361	Yes	-558, -97	5
2F57	BC058660	Tgfbli4	5' end	-59	Yes	10, 522	5
3F14	AK013491	mRNA	5' end	416	Yes	-357, 151	4
4B19	AK086839	EST	-		Yes		4
5C23	AK053913	Slit2	5' end	-1655	Yes	-586, -190	4
4E58	AK039431	mRNA	5' end	123	Yes	-558, -90	4
4D27	M95599	Hoxa2	5' end	401	No	-240, 349	4
4G96	BE987673	mRNA	5' end	-1	Yes	-610, -137	3
3G60	AK033940	4631416L12Rik	Body	1756	Yes	-404, 97	3
4D26	No EST	-	-	chr10:18,242,971	Yes	chr10:18,243,737-18,244,237	3
4G84	AK076525	mRNA	5' end	205	Yes	1995, 2504	3
2G34	AK004677	Klhl13	5' end	22	Yes	-387, 125	2
2G35	AK039308	mRNA	5' end	139	Yes	-472, 16	2
2G48	L24118	Tnfaip2	5' end	732	Yes	-252, 215	2
2F89	BE985330	EST	5' end	6	Yes	-318, 180	2
1G51	Y00051	Ncam1	5' end	451	Yes	-348, 127	2
3G119	AK080973	mRNA	5' end	-173	Yes	-355, 130	2
2F72	BC011343	mRNA	5' end	131	Yes	-648, -127	1
2G09	AK087498	mRNA	5' end	443	Yes	-288, 224	1
3C13	AK006521	mRNA	5' end	182	Yes	-426, 116	1
1E27	S75970	mRNA	5' end	-1098	Yes	-238, 236	1
5G61	AF487346	Dscaml1	5' end	-57	Yes	-297, 153	1
2D30	No EST	-	-	chr10:40,268,207	No	chr10:40,268,715-40,269,277	0
1F25	s75907	mTR2R1	5' end	-1098	Yes	-314, 150	0
2F39	AK044806	mRNA	Body	1642	Yes	-310, 189	0
6D20	AK052809	mRNA	5' end	53	Yes	329, 875	0
3E11	BC042784	mRNA	3' end	48262	Yes	48032, 48524	0
4G73	BC085500	Irx3	body	1653	Yes	-604, 7	0

6C25	AK078614	Ptprk	5' end	270	Yes	-462, 37	0
2C51	AK015334	mRNA	5' end	-208	Yes	-414, 82	0
1D26	bf456945	EST	5'end	4	yes	-317, 138	0
1F08	AK075900	Ran	5'end	-575	yes	-528, 12	0
2G93	AK081709	Ppra	5' end	-700	yes	-321, 160	0
3D04	AB041610	Cog8	5' end	21	yes	-896, -418	0
3D07	AF121344	Tspan5	5' end	86	yes	-314, 192	0
3d63	BC016883	Golga5	5' end	-35	yes	-317, 99	0
3E'54	BI689556	ETS	5' end	617	yes	-374, 162	0
3F61	BC030401	mRNA	5' end	236	yes	-175, 331	0
3G35	BC008101	Rasl11b	5' end	-277	yes	-507, 27	0
3G90	BC069933	Npepl1	5' end	226	yes	-754, -264	0
4B20	AK086839	mRNA	5' end	333	yes	-321, 169	0
4C14	AF179424	gata4	Body	4463	yes	-370, 126	0
4C18	AF459435	Slc6a8	5' end	781	yes	-30, 437	0
4D15	AK002749	Ndufa6	5' end	19	yes	-465, 84	0
4E'29	AK051764	mRNA	5' end	-1590	yes	-331, 165	0
4E'47	AF113751	Nup210	5' end	79	yes	-436, 109	0
5G22	AK037877	mRNA	5' end	64	no	-294, 218	0
5G90	AK009348	mRNA	5' end	-222	yes	-715, -206	0
5H105	BC010216	Rara	5' end	236	yes	-687, -162	0
5H33	No EST	-	-	chr4:119,672,344	yes	chr4:119,672,010-119,672,488	0
6G41	AY196089	Dot1l	5' end	227	yes	-268, 263	0
6G67	BB578608	ETS	5' end	246	yes	-402, 133	0
7E'09	AK122293	Dnajc6	5, end	272	yes	-407, 111	0

Table 3.1. The targets on selected custom mouse array and control genes unmethylated discovered by RLGS in mouse leukemia.

[a, b] The numbers in this column mean the distance from transcription start site.

[c] The numbers indicate methylated samples identified by RLGS from a total of 8 test samples.

RLGS ID	GB Accession	Name	CpG island	mouse leukemia(8)	RLGS[b]	A/in[c]	M/in[d]	Function
2F72	BC011343	mRNA	Yes	1[a]	-	↑	↓	Unknown
3D04	AB041610	Cog8	Yes	0	+	↑	↓	Component of oligomeric golgi complex 8.
5G61	AF487346	Dscaml1	Yes	1	-	↑	↓	Down syndrome cell adhesion molecule-like 1. cell adhesion molecule that can mediate cation- independent homophilic binding activity.
1F08	AK075900	Ran	Yes	0	+	↑	↓	Member RAS oncogene family. GTP-binding protein involved in nucleocytoplasmic transport.
6G41	AY196089	Dot1l	Yes	0	+	↑	↓	Histone H3 methyltransferase DOT1.
3D07	AF121344	Tspan5	Yes	0	+	↑	↓	Tetraspanin 5.
4C18	AF459435	Slc6a8	Yes	0	+	↓	↑	Solute carrier family 6 (neurotransmitter transporter, creatine), member 8. Required for the uptake of creatine.
2G50	AY371925	Mamdc1	Yes	6	-	↓	↑	MAM domain containing 1. May involved in cell-cell interactions
1F21	D70849	Zic3	Yes	6	-	↓	↑	Zinc finger protein of the cerebellum 3. Probably functions as a transcription factor in the earliest stages of the left-right (lr) body axis formation
5G21	D14340	Tjp1	Yes	6	-	↓	↑	Tight junction protein 1. Homologous to the product of the 'discs large-1' tumor suppressor gene of Drosophila. may be involved in tight junction.
4E24	AB041591	Slc16a11	Yes	7	-	↓	↑	Solute carrier family 16 (monocarboxylic acid transporters), member 11.
4G54	BC065168	Kbtbd9	Yes	5	-	↓	↑	Kelch repeat and BTB (POZ) domain containing 9.

Table 3.2. The targets selected for COBRA and ChIP-PCR.

[a] The numbers indicate methylated samples identified by RLGS from a total of 8 test samples.

[b] "-", Spot lost in L1210 identified by RLGS; "+", spot detected in L1210 by RLGS.

[c, d] A/in, M/in indicate the level of acetyl-H3K9 and dimethyl-H3K9

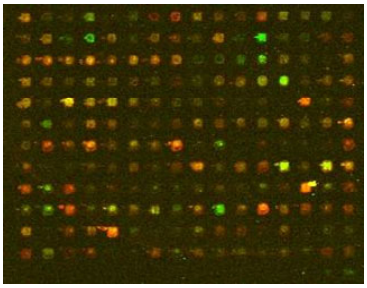
Name	5' Primer	3' Primer	Annealing temperature,ºC
COBRA			
p19ARF	ggtttttggttattgtgaggattta	aacctttcctacctaatccaaaatt	56
BC011343	tggggagtttgtttagtgttagagt	ctcccttaaaaaaacaacaccttc	55
Cog8	gtatttgtagttttgttggaag	taaaaaaaataaaaaataaaaacctcctaa	55
Ran	taattgttgtttaggagggagttag	ccaaatcctatatccacaaacac	55
Dot1l	ggttttgtgatttttataaagaggg	acttaaacctcaactccaacttctc	55
Tspan5	tgtttttatgaatttagaggatgtt	aaacctaactacccccctacc	55
ID4	atttgagatatttttttttaaaattt	aaccacccaaactataacc	50
Dscaml1	tttaggttggggttggagtagtat	ccacataccataaaaaaaccctatt	55
Slc6a8	atagttatttggggttgggt	taaacctaaaaaatctcatatatatc	55
Mamdc1	ttattatttgtaggagagggaggat	cctctacccaacttaaacatctctac	55
Zic3	taggttagttggtttttttattggtg	aaatttatcaatcctatccc	55
Tjp1	ggttgggtatgtttagtggtt	aaaaattcaaataaaacaaaaatcc	55
Slc16a11	ttatttttgtgttggttagttgttt	acctaaaaataaaaatcctaaatac	55
Kbtbd9	ggtattttaagagttttaatttgatt	aaacaaaaaaactaacttccctctc	55
ChIP-PCR			
p19ARF	ttatagatggactcggagcaagg	ccccctagcagtagctgcg	58
BC011343	gaggacggtccaggcttta	gacccggatgtacagggtaa	58
Cog8	ccccacctcacctcacatag	agcatcggagagaaccttca	58
Ran	ttgttgctccgccctctc	gactggagctggaaagatgg	58
Dot1l	tggtctctctcgctttcctc	accgacgcacgcacttac	58
Tspan5	aagccaccctctttttcagg	cgagagagacgagggaacac	58
ID4	ggcgatccaccttagtcgaag	tttgtgagcgacaatcggc	58
Dscaml1	tttccccttacatggcagac	cagtctcgatcctgctcctc	58
Slc6a8	aaagagcgctgaaaacggta	atctggcgtgtccaagtctc	58
Mamdc1	agagcagatccgcacacc	tctaaactcccgccagtctc	58
Zic3	cgcgctcttgagtagaggag	ggaggagaagggaggagaaa	58
Tjp1	cacaggagtttggggttctc	ttacttgctaggcggtttcc	58
Slc16a11	ctatccctaggcctggttcc	gtggggaaacacctgtgaat	58
Kbtbd9	aggagggttgatggataggg	aggaacgagaggaggtggac	58
Promoter lanscaping			
Ran-1	tagaggtgactcggtcgttaagagta	tggaactggaattacagaatagttgtg	58

Ran-2	taaacagatgcttaaaacaatgcactagt	ccatattctgtagtccatttgaaaacac	58
Ran-3	ctaccttcaatcaagttgaccacagat	acgcctgctctttcatatgtcttat	58
Ran-4	tggaaggcataatggtgagagtg	gggcagtaccctagactgaacaaa	58
Ran-5	atgcccgcttgagtgtattctc	tgtcctggcaaccagcct	58
Ran-6	ttgctggtcaattgctgctc	tccaggcgttcagcatcc	58
Ran-7	gtccgctgcgtctccg	cctcgacgctaccttccaga	58
Ran-8	ccgctcgtcttccatacca	ccttggatgtagtagccatcgc	58
Ran-9	tggtgtcgtttgatcatatgctg	tgacccgggagctttcc	58
Ran-10	taaagacgaatgaatatccttgtgatctt	accccggttataaacacccc	58
Ran-11	ctaactggcatagagatctggtacga	aatagattttgccttcactttcctgt	58
Ran-12	tttagaatgtcctggaatggagattat	agttgtagttacttttggcagaaatgtc	58
Zic3-1	atgtaaaccccagtaagccaaagt	tcagacaggaccactcgaacc	58
Zic3-2	ccacaacagaattcgaaatggtc	taggagtgcagttgttcagtgagg	58
Zic3-3	aggaatgctagtccccaactacc	cagatacccggataagcgagg	58
Zic3-4	gagggcagaaccggaaaag	ccatctctctgtagcaaacaacaact	58
Zic3-5	acacgaaaagcacagtcactgtct	gacacataccagaaacaagcatagatg	58
Zic3-6	ctgttctatcacgggacaaggg	gagcctactaacggtaattcggag	58
Zic3-7	gcgctgccaatcattgtgt	aatcactcactcctcgcacataaa	58
Zic3-8	cgcgctcttgagtagaggag	ggaggagaagggaggagaaa	58
Zic3-9	cgagcagtcttcacgctcc	gttgtccacgtgccctgtg	58
Zic3-10	aagaagagctgcgaccgga	ccaatagcagacgtggttgttc	58
Zic3-11	cgccgagacctttcagtacc	ggaacttagaactcggcaaaagc	58
Zic3-12	actaccaggcttagcaaaaccg	aaccaagctgggtaggacaatg	58
RT-PCR			
Zic3	tgcgacaagtcctacacaca	ctatagcgggtggagtggaa	58
Tjp1	gcagccaaagaaggcttaga	ggaggtcaaggaggaaaagg	58
Cog8	agttctgcactgccttcctg	gtccaaggtttccgtgctta	58
Ran	tgtgtggcaacaaagtggat	ctggcaagccagaggaaag	58
GAPDH	cggtgtgaacggatttggc	tttgatgttagtggggtctcgc	58

Table 3.3. The primers used in COBRA, ChIP-PCR, promoter landscaping analysis and RT-PCR.

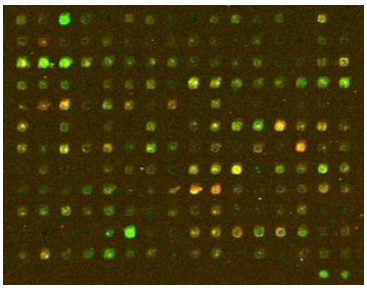
Figure 3.1. The distribution of histone modifications screened by ChIP-chip with mouse CpG island array. A and B. Representative results from the mouse CpG island arrays hybridized by ChIP DNA, labeled with Cy5, and input control, with Cy3. C and D. Scatter plot of histone modification level to density index. Histone modification level is indicated by the fold enrichment of ChIP DNA *vs* input DNA.

A



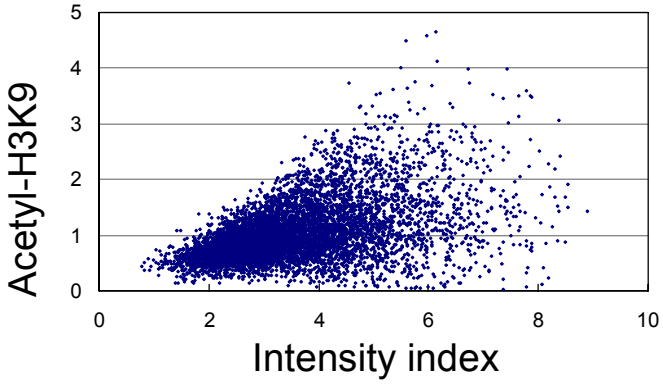
Acetyl-H3K9

B



Dimethyl-H3K9

C



D

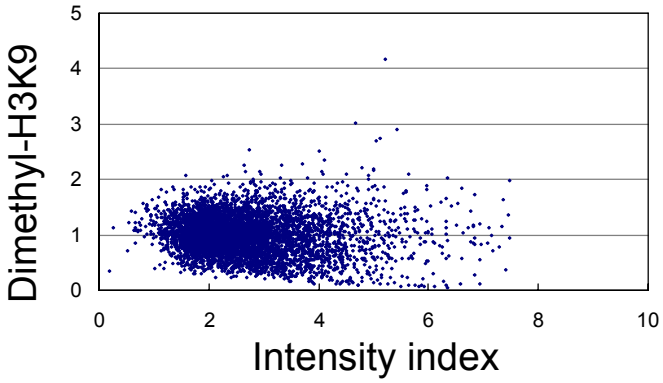


Figure 3.2. The distribution of histone modifications against DNA methylation identified by ChIP-chip and DMH with CpG island array. A and B. Positive targets (>2-fold change) from ChIP-chip (both acetyl-H3K9 and dimethyl-H3K9) and/or DMH are selected. The histone acetylation and methylation levels are plotted against DNA methylation status. Histone modification levels in L1210 cells were examined by comparison between ChIP DNA enriched with acetyl- and dimethyl-H3K9 antibody, and the input DNA. DNA methylation status was screened by DMH at *Bst*UI and *Hpa*II sites. Mean plots with error bounds were used to indicate the level of histone modifications.

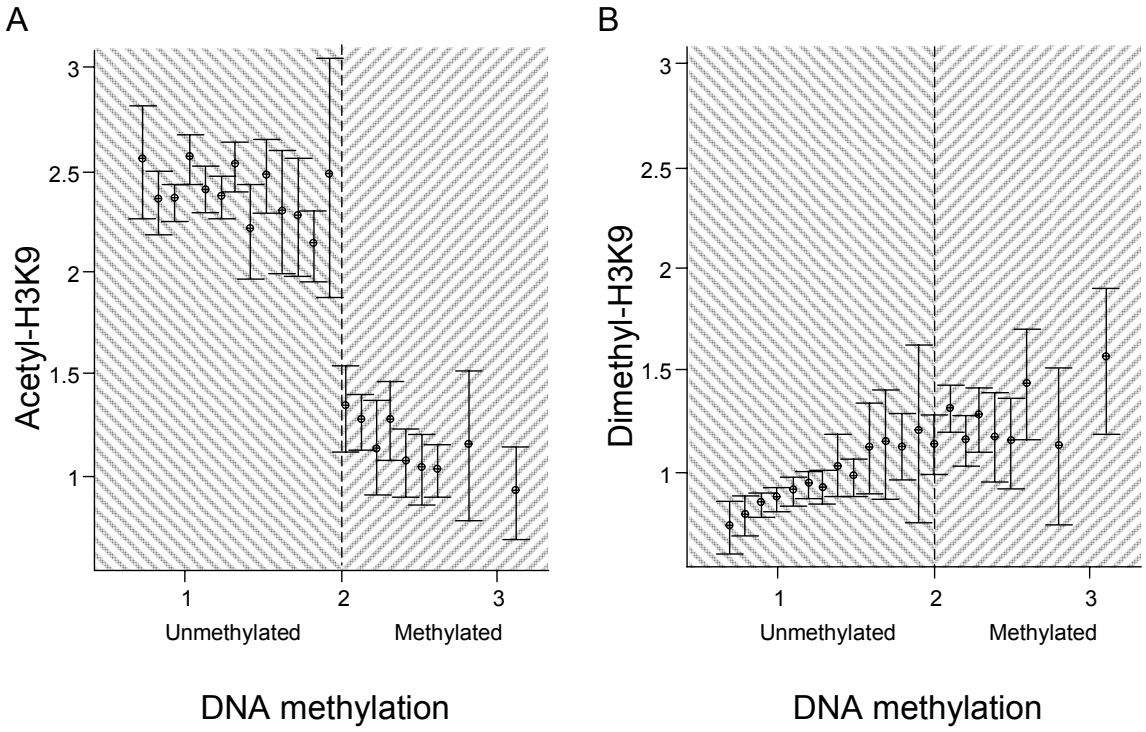

A
B
Acetyl-H3K9
Dimethyl-H3K9
3
2.5
2
1.5
1
1
2
3
Unmethylated
Methylated
DNA methylation
DNA methylation

Figure 3.3. RLGS analysis. A and B. RLGS sections from L1210 and DBA2. Fragment (arrow) was present in DBA2 mouse but lost in L1210, indicating that the gene was methylated in L1210 and resistant to methylation-sensitive *Not*I restriction enzyme digestion. C and D. Sections showing RLGS fragment (arrow) detected in both L1210 and DBA2 mouse, indicating that the gene was cut by *Not*I and not methylated in both L1210 and DBA2.

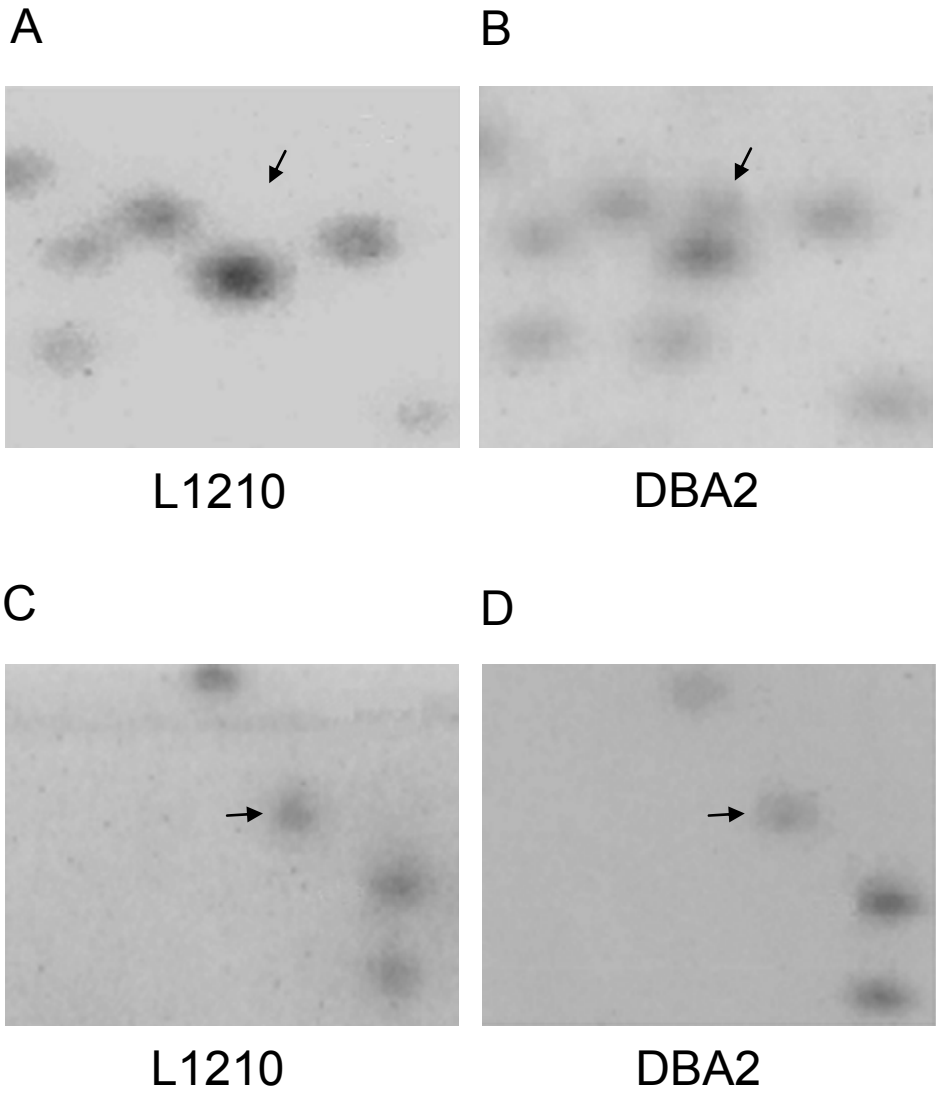

A
B
L1210
DBA2
C
D
L1210
DBA2

Figure 3.4. Histone modifications and DNA methylation status in mouse promoter regions. A. Histogram of histone acetylation and dimethylation identified by ChIP-chip with mouse custom promoter array. Brown and green bars indicate the frequency of acetyl-H3K9 and dimethyl-H3K9. Frequency curves are shown as follows: continuous line, acetyl-H3K9; dash line, dimethyl-H3K9. B. Interval plots of histone modifications versus DNA methylation. Histone modifications were plotted with two groups: unmethylated and methylated genes identified by RLGS. The interval bars indicate the distribution of histone modifications with horizontal lines at the endpoints of the 95% confidence interval for the mean and a symbol at the mean.

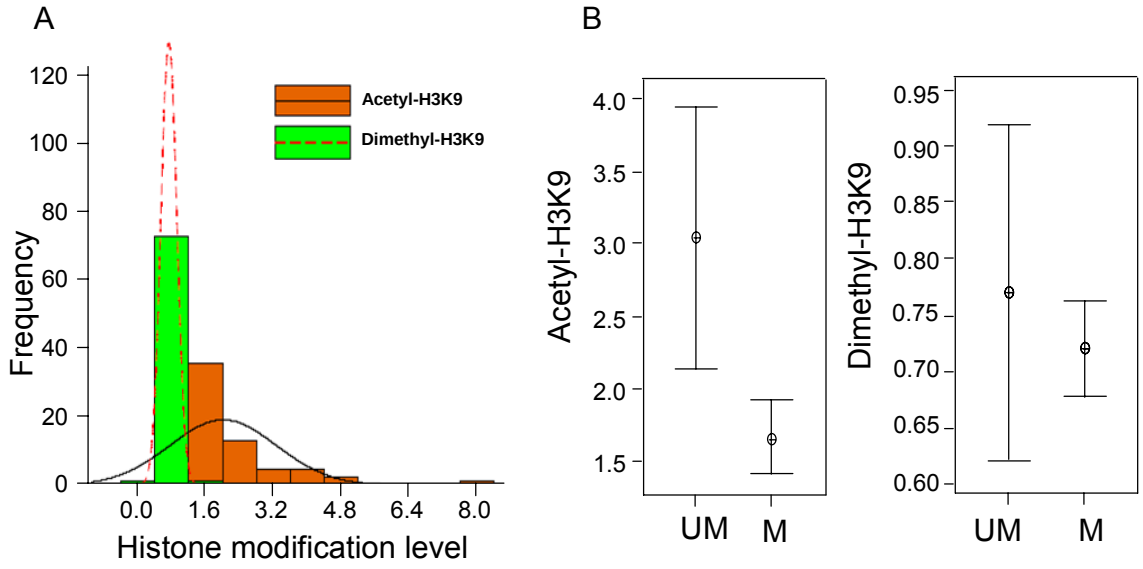

A
B
Acetyl-H3K9
Dimethyl-H3K9
Frequency
Histone modification level
0.0
1.6
3.2
4.8
6.4
8.0
UM
M

Figure 3.5. DNA methylation and histone modifications identified by COBRA and ChIP-PCR. A. The targets examined by COBRA and ChIP-PCR. *Not*I sites and regions for COBRA and ChIP-PCR are indicated. B. DNA methylation identified by COBRA. DNA from DBA2 mouse (U) and bisulfite treated DNA (M) were negative and positive controls, respectively. DNA from L1210 (L) was compared with controls. Arrows and stars indicate the methylated and unmethylated fragments, individually. C. Quantitative ChIP-PCR was conducted to examine the individual histone modifications of the same 12 selected genes including both methylated (right portion of graph) and unmethylated at CpG sites (left portion of graph). The histone modification level of each gene was compared with input genomic DNA from L1210 and indicated with bars: blank, acetyl-H3K9; black, dimethyl-H3K9. Each error bar represents the standard deviation calculated from triplicates.

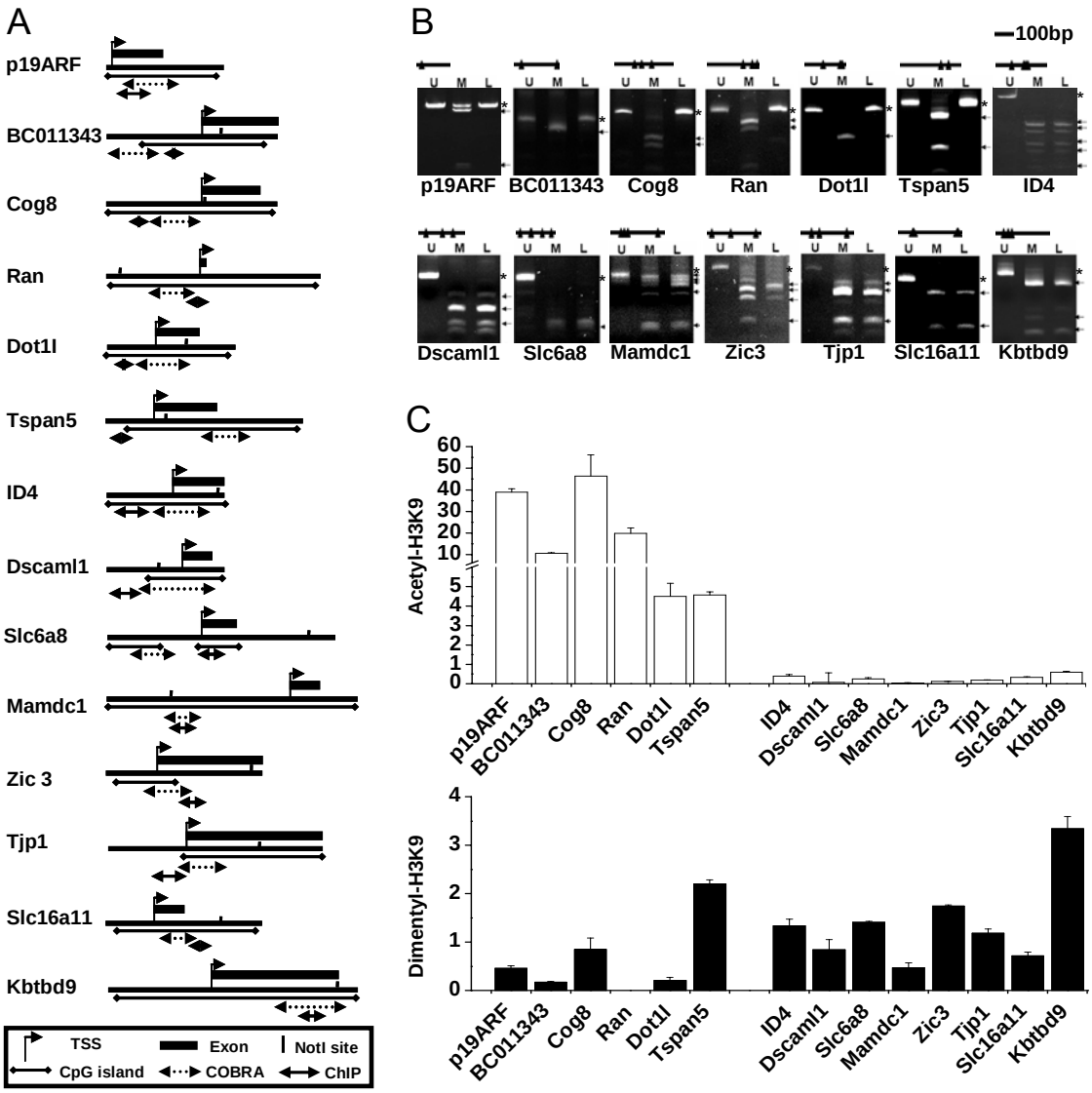
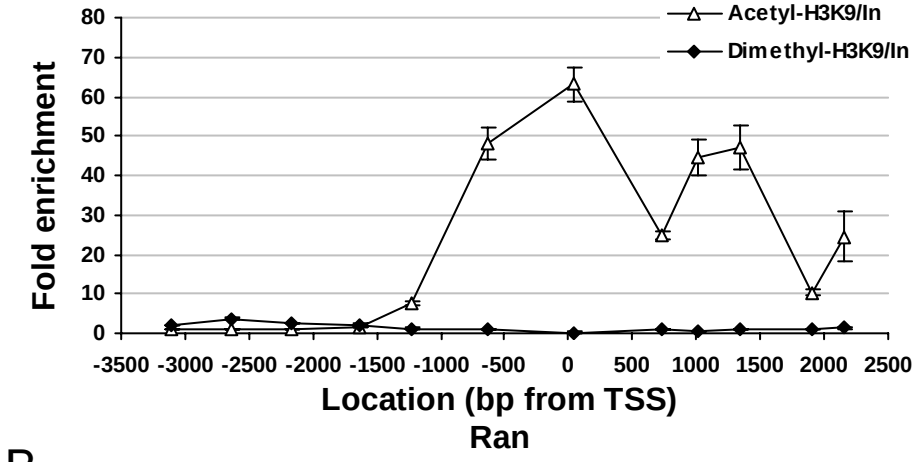

A
p19ARF
BC011343
Cog8
Ran
Dot1l
Tspan5
ID4
Dscaml1
Slc6a8
Mamdc1
Zic 3
Tjp1
Slc16a11
Kbtbd9
TSS
Exon
NotI site
CpG island
COBRA
ChIP
B
100bp
U M L
p19ARF
BC011343
Cog8
Ran
Dot1l
Tspan5
ID4
Dscaml1
Slc6a8
Mamdc1
Zic3
Tjp1
Slc16a11
Kbtbd9
C
Acetyl-H3K9
Dimentyl-H3K9

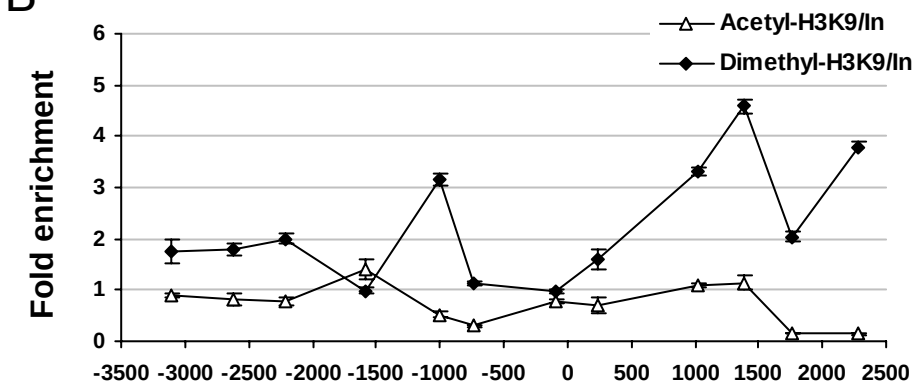
Figure 3.6. The promoter landscape of histone modifications. The fold enrichment by ChIP-chip with acetyl-H3K9 and dimethyl-H3K9 antibodies was screened by 12 sets of primers. The 12 PCR positions for each gene are indicated as the distances from transcription start site (0). Each error bar represents the standard deviation calculated from triplicates.

A


Acetyl-H3K9/In
Dimethyl-H3K9/In
Fold enrichment
80
70
60
50
40
30
20
10
0
-3500 -3000 -2500 -2000 -1500 -1000 -500 0 500 1000 1500 2000 2500
Location (bp from TSS)

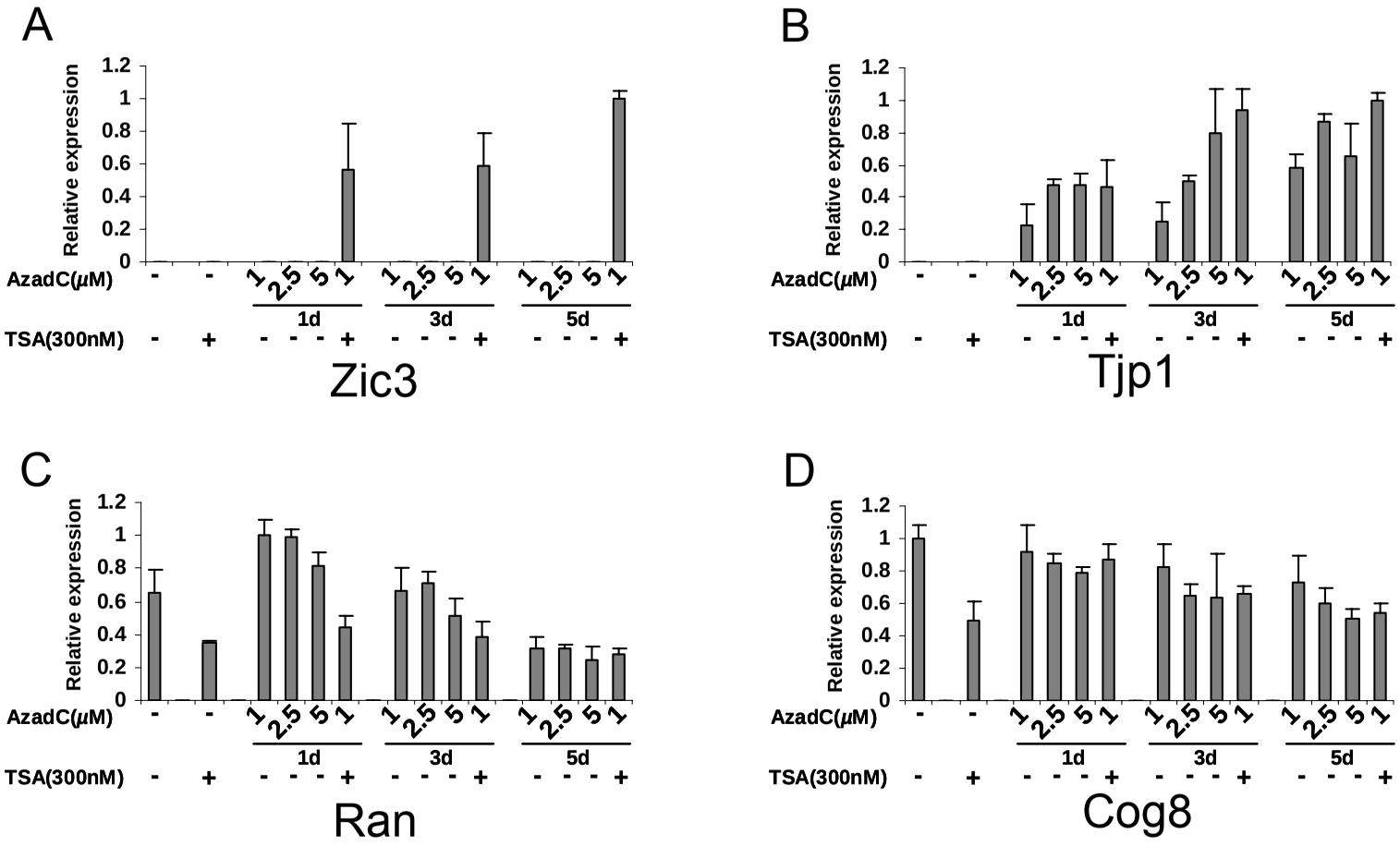
Ran

B


Acetyl-H3K9/In
Dimethyl-H3K9/In
Fold enrichment
6
5
4
3
2
1
0
-3500 -3000 -2500 -2000 -1500 -1000 -500 0 500 1000 1500 2000 2500
Location (bp from TSS)

Zic3

Figure 3.7. The effect of TSA and AzadC on gene expression. L1210 cells were treated without or with TSA (300 nM for 24 h), AzadC (1, 2.5 and 5μM for 1, 3 and 5 days), or the combination of TSA and AzadC (1μM of AzadC for 1, 3 and 5 days followed by 300 nM of TSA for 24 h). Expression of targets were identified by quantitative RT-PCR. **A** and **B.** *Zic3* and *Tjp1* are methylated and associated with hypoacetyl-H3K9. **C** and **D.** *Ran* and *Cog8* are unmethylated and associated with hyperacetyl-H3K9. Each error bar represents the standard deviation calculated from triplicates.


A
Relative expression
1.2
1
0.8
0.6
0.4
0.2
0
AzadC(μM)
- - 1 2.5 5 1 1 2.5 5 1 1 2.5 5 1
1d
3d
5d
TSA(300nM)
- + - - - + - - - + - - - +
Zic3
B
Relative expression
1.2
1
0.8
0.6
0.4
0.2
0
AzadC(μM)
- - 1 2.5 5 1 1 2.5 5 1 1 2.5 5 1
1d
3d
5d
TSA(300nM)
- + - - - + - - - + - - - +
Tjp1
C
Relative expression
1.2
1
0.8
0.6
0.4
0.2
0
AzadC(μM)
- - 1 2.5 5 1 1 2.5 5 1 1 2.5 5 1
1d
3d
5d
TSA(300nM)
- + - - - + - - - + - - - +
Ran
D
Relative expression
1.2
1
0.8
0.6
0.4
0.2
0
AzadC(μM)
- - 1 2.5 5 1 1 2.5 5 1 1 2.5 5 1
1d
3d
5d
TSA(300nM)
- + - - - + - - - + - - - +
Cog8

CHAPTER 4

SUMMARY AND PERSPECTIVES

In summary, we have performed integrative epigenomic studies and found a diverse relationship between histone modifications and DNA methylation for the maintenance of gene silencing. Acetyl-H3K9 appears to have an inverse relationship with promoter methylation in protein-coding genes. In contrast, methyl-H3K9 seems to be less distinctly related to promoter methylation. This work also demonstrates the importance of using genome-wide approaches to decipher complex epigenetic regulation in the cells.

Because of the heterogeneity of epigenetic regulation, we have to be cautious when we talk about epigenetic therapy and need to delineate the dynamic patterns of these epigenetic markers in various biological statuses of different cells. Besides DNA methylation and histone modifications at H3K9, attention should be directed to other epigenetic markers which may play a critical role in gene regulation [289]. In different cells, tissues or organisms, these epigenetic markers may function in unique ways. Even the same epigenetic pathway may even show altered regulatory function under different environmental stimuli in the same cell type. Therefore, application of epigenomic

tools in the Human Epigenome Project will be the next promising and necessary step to understand the mystery of life.

PART II

ALTERNATIVE PROMOTERS AND ESTROGEN REGULATION

CHAPTER 5

TRANSCRIPTIONAL REGULATORY ELEMENTS OF GENE REGULATION AND ESTROGEN RECEPTORS IN BREAST CANCER

5.1 Conventional and novel viewpoints on gene transcription

Besides the epigenetic modifications discussed above, other steps that are involved in regulation of gene transcription include transcription initiation, transcript processing, transport to the cytoplasm, translation of mRNA, mRNA stability, protein activity and stability[290, 291]. However, it is believed that most regulation occurs at the step of transcription initiation[292]. To adjust the level of transcription in a precise spatial and temporal expression pattern, complex *Cis*-acting DNA elements and numerous *trans*-acting factors interact and are involved in the regulation of transcription activity. These *Cis*-acting DNA regulatory elements include both core promoters and other regulatory elements,

such as enhancers, insulators and silencers. The *trans*-acting factors, traditionally being proteins, include both general transcription factors, such as RNA polymerase II, and other regulatory transcription factors. For the last several years, novel perspectives have also been introduced based on new evidences. For example, it is now clear that multiple promoters and alternative promoter usage are wide spread in human genome [293]. Another example is the long-range interaction between the regulatiory complex and the promoters of its targets, such as the model for ER-mediated transcription [188].

5.1.1 The transcription machinery

Bacteria have only one single RNA polymerase, while eukaryotes possess three: RNA polymerases I, II and III. Eukaryotic RNA Pol II is responsible for transcription of protein-coding genes. However, it cannot initiate transcription of eukaryotic genes without the cooperation of six other general transcription factors, TFIIA TFIIB, TFIID, TFIIE, TFIIF and TFIIH[294]. RNA polymerase II and other general transcription factors assemble on the core promoter in an ordered fashion forming the transcription preinitiation complex (PIC). The assembly of PIC will direct RNA polymerase II to the transcription start site and these general transcription factors are sufficient to drive only low levels of transcription, so-called basal transcription. Transcription activity is greatly stimulated by another groups of factors, called activators, which in general, are sequence-specific DNA-binding proteins[295]. In addition to DNA binding domains, these activators may contain dimerization, ligand-binding and/or activation domains. Activators recruit

or accelerate the assembly of general transcription factors. One group of best-studied examples is nuclear hormone receptors, including estrogen receptor α and β. The activity of an activator may be modulated by the third of factors called co-regulators, which do not bind specific DNA sequence but interact with DNA-bound activators to work together to positively or negatively regulate transcription[296].

5.1.2 *Cis* -acting transcription regulatory elements

Gene regulation at the DNA level is brought about by interplay between transcription factors and *cis*-acting transcriptional regulatory elements, such as promoters, enhancers, insulators and silencers.

The promoter is composed of the core promoter and the proximal promoter[295, 297]. A core promoter is the region that can extend ~35 bp upstream and/or downstream of the transcription start site (TSS) and serve as the docking site for the basal transcriptional machinery. It may include four identified promoter elements: TATA box, initiator element, downstream promoter element and TFIIB recognition element[297, 298]. The TATA box was the first core promoter element identified in eukaryotic protein-coding genes[299]. In humans, the TATA box is present 25-30 bp upstream of the TSS and is bound by the TATA-binding protein, a key component of the TFIID complex. Binding of TATA with TATA-binding protein initiates PIC formation[297]. A second element of core promoter, the initiator, contains a pyrimidine-rich sequence surrounding the transcription start site. The initiator binds with TFIID or other transcription

factor and functions similarly with TATA box, capable of directing transcription initiation either alone or in conjunction with TATA box or other promoter elements[300]. The downstream promoter element is located at -27 to -31 bp of the TSS and is required for TFIID binding[301]. It may increase the transcription activity by several fold independent of TATA box [302, 303]. The interaction between TFIIB and TFIIB-Recognition Element, a well-studied element, clearly enhances the assembly of a preinitiation complex and transcription initiation. The proximal promoter is defined as the region immediately upstream of the core promoter, and typically contains multiple binding sites for transcription activators[297].

Enhancers are DNA sequences that can activate transcription after binding of activator proteins[304]. The enhancer activity is independent of its own orientation on DNA and its position to the promoter, either upstream or downstream of TSS. It is believed that enhancer-promoter communications impart transcription regulation in a spatial- or temporal-specific manner during development and differentiation[305]. Various models are proposed to describe the communication between enhancer and promoter, such as the DNA looping model, which states that the interaction of an enhancer-bound activator with a protein factor of the general transcription machinery at the promoter is accompanied by looping of the intervening DNA [306]. Insulators are those sequences that can block genes from being affected by the neighboring sequences and they have two main functions: one is to interfere with communications between enhancer and promoter in a position-dependent

manner without interfering with the enhancer activity[307]; the other is to prevent the spread of repressive chromatin[308]. One example is *gypsy* in flies. Once placed between an enhancer and promoter, it will specifically block the enhancer communication with the promoter[309]. Another example is the imprinting control region located upstream of the H19 gene that regulate allele-specific transcription of H19 and another gene, Igf2[310]. Silencers are specific sequences that flank promoters. Their functions are similar with enhancers independent of orientation and position[311]. However, silencers recruit repressor instead of activator complexes to down-regulate gene transcription.

5.1.3 Novel perspectives on diversity of promoter regulation

Most core promoters, however, do not have all four elements. A statistical analysis of ~10,000 predicted human promoters showed that the known core promoter motifs may not be as universal as previously thought[312]. In humans, it is estimated that only 10-20% of promoters contain a functional TATA box, while 85% contain an initiator[313]. With development of new experimental technologies capable of genome-wide discovery and characterization of core promoters, it is clear now that most mammalian genes do not conform to the conventional model of promoter architecture and usage[314]. Higher rate of genome region, including so-called ‘transcriptional dark matter’, was transcribed. Overlapping transcripts were widely found. Therefore, many of these transcripts are believed to be antisense transcription involved in regulation of target transcription. Meanwhile there are widespread occurrence of multiple promoter

regions and alternative promoter usage[293]. Here we will focus only on the topic of alternative promoters.

Recently, CAGE tag analysis provided the first truly global insight into the sequence structure of core promoters and the dynamic expression associate with these promoters[315]. At last, two classes of promoters can be defined: sharp type promoters and broad type promoters[314]. Sharp type promoters have a relatively tight defined TSS, whereas broad type promoters have a wide distribution of many TSS in a 100 bp range. Sharp type promoters are the conventional promoters associated with TATA box and frequently with other transcription factor binding sites. The broad type promoters, however, have less TATA box and less consensus binding sites. Instead, the broad type promoters often have CpGs islands. It was also found that single type promoters were highly regulated and tissue-specific, and broad type promoters were found more with ubiquitously expressed genes. Moreover, different TSS utilization at broad type promoter are activated in different cell types[316]. In addition to multiple TSS usage within one promoter region, genome-wide analysis of alternative promoters of protein coding genes showed that in vertebrates, 20-52% of the genes have alternative promoters[192, 317]. The biological significance of such wide-spread alternative promoters is not clear even though single gene studies have shown the alternative promoter in these genes may receive differential signaling information to generate different protein isoforms[318].

5.1.4 Analysis of regulatory elements

Since alternative promoter usage was found to be a common phenomenon[319], complete characterization of promoter regions of these transcripts is critical to understand such transcription regulation. Both computational and experimental strategies have been applied to determine and analyze promoter sequences. Compared with prokaryotic promoters, the complexity of eukaryotic promoters makes it difficult to develop successful computational algorithm to identify and analyze them. Prokaryotic promoters have three consistent features: the TATA box, CA dinucleotide and the spacer between them, which have been used successfully as the signatures for promoter prediction[320]. However, similar features in eukaryotic promoters are less consistent as discussed above, so that eukaryotic promoters are difficult to be recognized reliably based on these features. Other efforts of promoter prediction, such as those based on transcription factor binding analysis and gene recognition, were also limited because of the diversity of the eukaryotic genome. The alignment of cDNA to genome sequences or analysis of evolutionally conserved regions have helped delineate the architecture and function of regulation sequences in promoter region, however, both have their limitations and can not provide direct evidence of promoter activities [315].

Recently, intensive efforts have been invested in establishing genome-wide profiling methods to identify regulation regions, including transcriptional starting sites and the upstream promoter sequences in human and mouse. Currently, three ways have been applied for this purpose and will be described in

the next chapter in more detail.

5.2 The biological role of estrogen receptors in breast cancer

The normal functions of estrogen are involved in growth and maintenance of female reproductive tissues and other non-reproductive tissues such as the cardiovascular tissue, bone and central nervous system. Estrogen action is mediated through binding to its cognate receptor, the estrogen receptor[321]. There are two forms of estrogen receptors, ERα and ERβ, of which ERα is the most well-studied[322-324]. The estrogen receptors are members of the nuclear receptor superfamily of transcription factors. There are two genomic pathways utilized by estrogen receptors to perform gene regulation. One is ligand-dependent pathway where estrogen receptor requires estrogen to form heterodimers or homodimers. Upon dimerization, estrogen receptor binds to estrogen response elements at the regulatory region of the target genes and recruit coactivators or corepressors to modulate gene transcription[325-328]. Another pathway is independent of estrogen and mediates transcription by interacting with other protein at AP1 and Sp1 sites[329]. In addition, estrogen can mediate rapid activation of several signaling pathways without producing RNA and protein, which is called nongenomic estrogen signaling[330-332]

Hormonal exposure is the best characterized risk factor for breast cancer. One of the earliest piece of evidences came from the observation of breast tumor regression after removal of ovaries, the major site of estrogen production in premenopausal women[333]. More evidence supports the association of estrogen exposure with the increased risk of breast cancer[334-338]. It was also

found that breast cells expressing ERα are sensitive to estrogen-mediated proliferation and regression upon estrogen ablation[339-341]. Since it is clear that ERα is important for breast cancer development, selective estrogen receptor modulators, such as tamoxifen, are used to block the interaction between estrogen and estrogen receptor and come to be the most effective therapeutic strategy for breast caner[342].

CHAPTER 6

GENOME-WIDE ANALYSIS OF ALTERNATIVE PROMOTER USAGE IN ESTROGEN-STIMULATED MCF7 CELLS USING A CUSTOM PROMOTER TILING ARRAY PLATFORM

Gregory Singer and Jiejun Wu contributed equally to this work.

6.1 Introduction

The regulation of human gene expression is known to be an extraordinarily complex process. Nevertheless, one could easily believe that the combinatorial interaction of multiple transcription factors within the gene promoter is sufficient to explain this complexity. However, genes with more than one promoter have been known for some time [343], and recent studies using independent lines of evidence have suggested that a large proportion of human genes have more than one independently regulated promoter [317, 344, 345]. As shown in Figure 6.1A, alternative promoters can take many different forms, producing a wide variety of transcripts and proteins from a single gene locus. Needless to say, these various promoters greatly increase the regulatory control that the cell has over the expression of the gene.

Alternative promoters are of particular interest because their aberrant expression has been linked to a number of diseases, particularly cancer. There are a number of experimentally-characterized well known genes that support the assertion of one gene-multiple promoters, for example TP53 [346], MYC [347], CYP19 [348], BRCA1 [349], P73 [350], MID1 [351], Cathespin B [352], SRC [353], kallikrien 6 [354] and TGF-β3 [355], to name a few. CYP19A1 is well-known example that has 5 known alternative promoters, many of which are separated by 10kb and therefore are regulated by completely non-overlapping promoters. Alternative first exons Ex-1.1, Ex-1.3/Ex-1.4, and Ex-1f splice with Ex-2 to encode the 5'UTRs of the aromatase P450 mRNA in the placenta, adipose tissue, and brain, respectively. Additionally, in gonads, the transcription starts just 39 bp upstream of translation initiation codon in exon-2. The use of alternative non-coding first exons in the CYP19 transcripts does not alter the protein structure, as the different 5'UTRs splice into a common second exon (exon-2) that contains translation initiation codon. Indeed, it is known that the various promoters are used in a tissue-dependent manner, but the promoter upstream of exon Ex-1.4 is aberrantly expressed in breast cancer tissue, aggravating the disease [348].

Many putative gene promoters have been identified through the mapping of ESTs to the genome (Acembly [356] or ECGene [357]), or through sequence conservation studies with other organisms [358] and *de novo* computational prediction (e.g., FirstEF [359], DragonGSF [360]). Databases such as MPromDb [361] and H-DBAS [362] provide information about well-curated promoters and

alternative spliced transcripts identified by aligning completely sequenced and precisely annotated full-length cDNAs [317]. Recently, intensive efforts have been invested in establishing genome-wide profiling methods to identify the regulation regions, including alternative transcription starting sites and the upstream promoter regions in human and mouse genomes. Currently, three ways were applied for this purpose. One is based on the decreased nucleosome occupancy and increased sensitivity to DNase of the active promoter regions. The two approaches, called DNase-chip and DNase-array, have been created to detect those transcribed promoters and transcripts [363] [364]. The second one is called the cap analysis gene expression (CAGE), combining full-length cDNA library with SAGE technology to screen those 5' parts of transcripts [365]. The third one is using ChIP-chip to profile the binding position of the RNA polymerase II preinitiation complex [192]. The data from these studies provide evidence of large-scale alternative splicing and wide-spread use of alternative promoters throughout the mammalian genomes. Most of these methods cannot predict the mRNA sequence produced from that promoter, and therefore constructing a traditional cDNA microarray to detect their expression is impossible. Moreover, two promoters may produce mRNA isoforms that are nearly indistinguishable, again making expression microarrays difficult to design. One alternative is to use ChIP-chip to detect the binding of RNA polymerase II to the genome. To that end, we have designed a custom microarray that contains probes complementary to sequences tiling the regions around known and putative

promoters in the human genome. The array contains 237,217 non-control probes that tile 34,486 promoters in 6,949 genes (see Figure 6.1B)

It is well known that estrogen receptor can induce both increases and decreases in gene expression, and that these events can then affect cell division and breast cancer progression [321, 366]. Characterization of the regulation of these estrogen-responsive target genes will be beneficial to understanding their effects in breast cancer development. Here, genome-wide analyses using ChIP-chip have been performed in order to recognize ER-sensitive promoters and to demonstrate the diversity mechanisms used by ER to regulate its targets genes. Interestingly, we find that within a single gene with more than one promoter, downstream promoters are much more likely to be affected by E2 treatment than upstream promoters.

6.2 Methods

Target identification

We considered three sources of evidence for identifying promoter targets for our microarray. The first was the 5'-end of genes as identified in the UCSC Known Gene track, which is largely based on the alignment of RefSeq mRNAs to the human genome [367]. A second line of evidence was the database of CAGE tags sequenced by the Riken group [368]. These tags capture ~20 bases at the 5' end of messenger RNAs, and have been mapped back to the human genome. We used the UCSC LiftOver tool to convert Riken's hg17 human genome

coordinates to the more recent hg18 genome. Our final line of evidence was *ab initio* promoter predictions generated by the FirstEF program [359].

Each line of evidence identifies a transcription start site (TSS). We considered TSSs separated by >500bp to be distinct promoters—a commonly used criterion. Although there are undoubtedly transcription factor binding sites that extend beyond this region, this distance is great enough for the core promoters of each TSS to be distinct [369], and we can therefore consider these TSSs to be independently regulated to a large extent. TSSs were clustered using a neighbor-joining algorithm [370] until all clusters were separated by at least 500 bases. The coordinates of these clusters were then extended 200 bases up- and downstream.

Probe selection

Each promoter region was aligned to the genome using BLAT [371] in order to discover regions that are not unique. Alignments that were longer than 55 bases (90% of the probe length) were masked, as were 60mers within the sequence that had >85% or <50% G+C. From the remaining unmasked regions of each promoter, probes were selected such that the average spacing would be roughly 100 bases, but that the spacing between two successive probes would be no more than 300 bases. In the end, the true average spacing is 80 bases.

Gene selection

Not all genes could be put on the array, so to prioritize we assigned each gene a score. Three points were awarded for each promoter supported by "known gene" evidence, two points for those supported by CAGE tag evidence [368], and one point for FirstEF [359] evidence. Genes were then ranked by their total score, and only the best-scoring genes were included on the array. In the end, the roughly 244,000 probes cover 34,486 promoter regions from 6,949 genes, with a median tiling coverage of 5 probes per promoter. The median number of promoters per gene on the array is 3, although the range is from 1 to over 30 (Figure 6.1B).

Cell culture

MCF-7 human breast cancer cells (American Type Culture Collection, Manassas, VA) were maintained in growth medium (MEM with 2 mM L-glutamine, 0.1 mM non-essential amino acids, 50 units/ml penicillin, 50 µg/ml streptomycin, 6 ng/ml insulin, and 10% FBS) as described by Fan et al [372]. Prior to all experiments, cells were cultured in hormone-free basal basal medium (phenol-red free MEM with 2 mM L-glutamine, 0.1 mM non-essential amino acids, 50 units/ml penicillin, 50 µg/ml streptomycin, and 3% charcoal-dextran stripped FBS) for three days.

Chromatin immunoprecipitation on microarray (ChIP-chip) assay

Five million MCF-7 cells with and without E2 treatment (10 nM, 3 h) were crosslinked with 1% formaldehyde for 10 min, at which point 0.125 M glycine was used to stop the crosslinking. Chromatin immunoprecipitation was performed using a ChIP assay kit (Upstate Biotechnology, Charlottesville, VA) as described [183]. The antibodies against Pol II were purchased from Santa Cruz Biotechnology (Santa Cruz, CA). Ligation-mediated PCR was applied to 20 ng of ChIP DNA and input DNA as described by Ren et al [174]. The primers used in ligation-mediated PCR were: oligo JW102, 5'-GCGGTGACCCGGGAGATCTGAA TTC-3' and JW103 5'-GAATTCAGATC-3'. Two µg of amplified ChIP DNA and input DNA were then labeled by Cy5 and Cy3 fluorescent dyes (Amersham, Buckinghamshire, UK) and were then cohybridized to the custom alternative promoter array. The slides were washed with three wash buffers (Buffer I. 6X SSPE + 0.005% sarcosine; Buffer 2, 0.06X SSPE; Buffer 3, anti-oxidant mixture in acetonitrile purchased from Agilent) in series at room temperature.

Chromatin immunoprecipitation- quantitative polymerase chain reaction

ChIP was conducted in the same manner as in the ChIP-chip experiments, described above. The pooled DNA from ChIP and input control were first measured by spectrophotometer (NanoDrop, Wilmington, DE). Quantitative PCR with SYBR green-based detection (Applied Biosystems, Foster City, CA) was performed as described previously. In brief, primers were designed using Primer Express software (Applied Biosystems, Foster City, CA) according

to the promoter structure of selected. Quantitative ChIP-PCR values were normalized against values from a standard curve (50 to 0.08 ng, R-squared > 0.99) constructed by input DNA with the same primer sets.

Quantitative reverse transcription-polymerase chain reaction

Total RNA was extracted from MCF-7 cells with or without E2 treatment. Two μg of RNA was first treated with DNase I (Invitrogen, Carlsbad, CA) to remove potential DNA contamination and then was reverse transcribed with SuperScript II reverse transcriptase (Invitrogen, Carlsbad, CA). Quantitative RT-PCR was performed by using SYBR green (Applied Biosystems, Foster City, CA) as a marker for DNA amplification on a 7500 Real-Time PCR System apparatus (Applied Biosystems, Foster City, CA). The relative mRNA level of a given locus was calculated by relative quantization of gene expression (Applied Biosystems, Foster City, CA) with glucose phosphate isomerase mRNA as an internal control.

Microarray analysis

The washed slides were scanned by a GenePix 4000A scanner (Axon, Union City, CA) and the acquired microarray images were analyzed using GenePix 6.0 software. Briefly, the user-selectable laser power settings for Cy5 (635 nm, red) and Cy3 (532 nm, green) were adjusted so that the overall Cy5 to Cy3 ratios were close to 1 and that the signal intensities spanned the entire spectrum with minimal signal saturation at the high intensity range. When these conditions were satisfied, the microarray was scanned and a grid file was loaded

to mark the general location of the scanned image. The GenePix 6.0 software performed a spot finding function and captured intensity-related information in a GPR file.

GPR files were passed through a custom-built quality control filter which flagged all probes that didn't meet all of the following criteria in both the green and red channels: (1) % > B + 2SD greater than 30; (2) median – background > 0; (3) signal-to-noise ratio greater than 1.5. These filtered results were then normalized using the default parameters (plus Lowess normalization) in Agilent's Chip Analytics software version 1.3. A post-normalization MA plot is shown in Figure 6.2B. We then used a modified version of the mixture model described by Khalili et al [373] to classify probes into one of two groups: bound or not bound. Figure 6.2C shows the fit of the gamma+normal mixture model to our data. One benefit of this type of analysis is that we are able to directly estimate our false positive rates based on a probe's probability of assignment to the "unbound" distribution.

Since each promoter contains many probes, for each promoter region we chose the probe with the best p-value for inclusion into the "bound" distribution and compared these across the various experimental treatments and replicates. We used the following criteria to classify promoters within individual experiments: strongly bound promoters had probes that were classified in the "bound" distribution with a p-value less than 0.05. Weakly bound promoters were those that did not significantly fall within the "unbound" distribution with a p-value of 0.05. Unbound promoters were those whose probes fell within the "unbound"

distribution with a p-value less than 0.05. As Figure 6.2D illustrates, by combining replicate experiments, we were able to classify each promoter into “highly on” (both replicates were strongly bound), “medium on” (one replicate was strongly bound, the other weakly bound), “low on” (both replicates are weakly bound), “weakly off” (replicates don't agree, so we fall back on the null hypothesis of no binding), or “strongly off” (both replicates show an unbound state).

6.3 Results

Alternative promoter array

Using evidence from three sources: UCSC Known Genes [367], FirstEF [359], and Riken CAGE tags [368], we find evidence for more than 185,000 transcription start sites separated by 500 bases or more in the human genome. We took a gene-centric approach to our microarray design, choosing genes that had two or more known or putative promoters. In the end, about 34,000 known or putative promoters were selected for our array, covering about 7,000 genes. The median number of promoters per gene is three (Figure 6.1B). 60mer probes were then designed to tile a region -200 to +200 surrounding each known and putative promoter. Because of limitations on probe design, not all regions could be effectively covered but on average the spacing is about 80 bases between probes.

Genome-wide profile of promoter usage

ChIP-chip with antibody against Pol II was performed in MCF-7 cell lines without and with E2 treatment for 3 hours, as described in the Methods. The Expectation Maximization (EM) algorithm of Khalili et al [373] was modified from the original Gamma-Normal-Gamma fit to a simple Gamma-Normal fit that appeared to be more appropriate for our data. Figure 6.2C shows that the algorithm clearly defines two distinct distributions, representing the unbound probes (in red) and the bound probes (in green). A nice feature of the algorithm is that probes can be assigned to each distribution with a certain probability, allowing us to increase or reduce the stringency of our assignments easily. We defined strong candidates for RNA Polymerase II activity as those probes that fell within the green distribution with a p-value of at most 0.05. However, we also defined a second, weaker condition: those probes that are not significantly part of the larger unbound (red) distribution at a p-value of 0.05. This latter group would encompass the “grey area” that lies between the two distributions. The “best” probe from each promoter was used to evaluate the activity of the promoter as a whole. Figure 6.2D shows the proportion of active promoters in MCF7 cells at different quality thresholds. At least 65% of the promoters (both putative and known) are inactive in this cell line, whereas ~17% of the promoters have strong evidence for being active.

We tested ten promoters that we predicted to be active with high confidence and eight promoters that were predicted to be inactive in MCF7 cells. ChIP-PCR experiments showed that these predictions were for the most part

accurate (Figure 6.3)—seven out of the ten positive targets microarray analysis were confirmed to be bound to RNA polymerase II. Similarly, all but one of the negative samples showed no evidence of RNA polymerase II binding. Although the binding of RNA polymerase II to the promoter region needn't correlate to gene expression because of posttranscriptional events, we find that a rough correspondence does exist. For example, two promoters in the gene NCOA7 were shown to bind to RNA polymerase II with a "low" level of confidence, although in the absence of E2 the upstream promoter was predicted to be "strongly off" (Figure 6.4A). These qualitative results were verified by qRT-PCR (Figure 6.4B). Compare these results to the gene EIF3S9, whose most upstream promoter was "highly on" in both treatments (Figure 6.5A). The qRT-PCR experiments show a correspondingly high level of expression of that gene isoform (Figure 6.5B).

Identification of novel promoters

As shown in Table 1, each promoter on the array is supported by different lines of evidence. The most common promoters are those that are supported by the ubiquitous CAGE tags throughout the genome. However, only 14% of the roughly 18,000 such promoters were found to be active at "high" or "medium" confidence levels. Of course, it is important to note that a negative result does not necessarily indicate an inaccurate promoter prediction; these promoters may be active in different cell types, or under different environmental conditions. Therefore, these numbers should be seen as a lower limit. By far, the greatest

success rate was found for CpG-related promoters that had all lines of evidence supporting them—UCSC Known Genes, CAGE tags, and FirstEF predictions—of which 68% were found active. Interestingly, the data indicate that CpG-related promoters that are supported by CAGE tags and FirstEF predictions also enjoy a high rate of success—far better than either CAGE tags or FirstEF predictions alone. 16% of non-CpG-related promoters in this category were found to be active, while an impressive 39% of CpG-related promoters supported by CAGE and FirstEF results were found to be active. In all, if we consider all promoters not supported by KnownGenes to be “novel”, then out of 20,879 promoters, 3,172 (15%) were active in at least one treatment. If we eliminate promoters supported only by CAGE tags, then 601 out of 1,977 promoters (30%) are found to be active. Of the ten genes selected for validation in Figure 6.3, eight fall into the novel category (i.e., no mRNA evidence) and six of these were confirmed (see Table 2). These surprising results indicate that large numbers of undiscovered, unannotated promoters exist within human genes. Notably, we have discovered 303 new and active promoters that lie more than 500 bases upstream of the currently-defined 5’ end of the gene, suggesting that a significant fraction of the current gene annotations may not be 5’-complete. One of these promoters was upstream of SOX12, and was verified to bind to RNA polymerase II (Figure 6.3).

Differential use of multiple promoters with estrogen stimulation

Our hypothesis is that treatment with E2 should affect the promoter activity of some genes in the genome. For this analysis, we defined “active” as promoters with “high”, “medium” or “low” confidence. For genes with a single active promoter, 2,697 were found to be active in both E2- and E2+ treatments (see Supplemental Table 1). 178 genes were inactivated by E2, while 77 were activated by E2. This bias is highly significant (p=2.5e-10), indicating that more promoters are inactivated by E2 than are activated, which supports the previous report about estrogen-mediated early-downregulated genes [374]. When we consider the two active promoter genes, we find 993 in which both promoters are active and not affected by E2 treatment (see Supplemental Table 2). More interesting are the cases where one promoter is affected by E2 treatment. For 25 genes, the upstream promoter is activated by E2 (Figure 6.6A; also see the gene NCOA7 in Figure 6.4), while in 61 genes the upstream promoter is inactivated by E2 (Figure 6.6B)—a more than 2:1 bias in favor of inactivation, just as we saw in the single active promoter gene case, and also significant (p=0.000175). Curiously, this same bias is not present when we examine the downstream promoters, where 62 are activated by E2 (Figure 6.6C) and 64 are inactivated by E2 (Figure 6.6D). An unexpected observation was that there are a total of 127 downstream promoters affected by E2 treatment, while only 87 upstream promoters are affected—a significant bias (p=0.00625). These intriguing patterns provide some insight into the regulatory control of genes and their isoforms by E2. To investigate this phenomenon further, we examined the locations of active

promoters within each gene. As shown in Figure 6.7A, for genes with a single active promoter that is insensitive to E2 treatment there is a strong tendency for that promoter to be located at the 5' end of the annotated gene—not a surprising finding. Similar trends are observed in genes with two active promoters that are insensitive to E2 treatment, where the upstream promoter is again located near the 5' end of the gene, while the location of the downstream promoter is uniformly distributed throughout the length of the gene (Figure 6.7B). However, a surprising change is observed if one of the promoters is E2-sensitive, where we then find that there is a very strong tendency for the downstream promoter to be close to the 3' end of the gene (Figure 6.7C). The mechanisms behind this bias remain a mystery, though we propose possible explanations in the Discussion.

6.4 Discussion

Although genome tiling arrays are increasingly becoming a viable alternative to focused microarrays, they remain significantly more expensive than focused microarrays, and their signal-to-noise ratio is far worse—both because of the larger numbers of inactive probes and because of the lack of probe design considerations. Another alternative mechanism for studying alternative promoters is the use of traditional expression arrays that have been designed to specifically interrogate particular gene isoforms. Unfortunately, in a large number of cases, mRNA isoforms are not known for putative promoters, and many isoforms that originate at different promoters differ only in the first exon---a small percentage of

the entire molecule, making it difficult to distinguish between the various isoforms.

Here, we have presented a novel 244k microarray that is capable of measuring alternative promoter usage in over 34,000 putative promoters from nearly 7,000 genes. This platform is suitable for indirect expression analyses using RNA polymerase II ChIP-chip as we have shown in this paper, but it is also suitable for methylation based studies using DMH or meDIP experiments (since more than 5000 of the putative promoters fall within CpG islands), or for ChIP-chip experiments using other proteins of interest, such as transcription factors or histone modification signatures. We have demonstrated clear evidence for alternative promoter activities within genes, including the verification of a number of promoters that were heretofore considered putative. These results suggest that a large fraction of the genes in the human genome possess undiscovered promoters and transcription start sites, which agrees with findings based on the mapping of ESTs to the genome [356, 357], and the mapping of 5' oligo cap cDNA libraries to the genome [317].

Most intriguingly, we discovered that there is a distinct bias for the downstream promoter in E2-sensitive two-promoter genes to be very close to the 3' end of the gene, whereas no such bias exists in E2-insensitive genes. These promoters are very unlikely to produce a functional transcript of any sort, and we therefore speculate that its purpose is merely to regulate the expression of the transcript initiated at the upstream promoter by "blocking" the progression of the RNA polymerase II complex. This "stalling" mechanism has been observed in

other contexts. For example, inhibiting DNA replication was recently found to cause RNA polymerase II to stall during the transcription of p21 [375]. Similarly, the cofactor of BRCA1 (COBRA1) is known to cause stalling of the RNA polymerase II complex proximal to the promoter [376, 377]. However, we can think of no reason for "blocking" promoters to have a bias towards the 3' end of the gene, since this blocking action could be realized at any point relative to the primary promoter. An alternative possibility is that promoters near the 3' end of the gene are driving expression of an interfering RNA, either antisense to the primary transcript or that is capable of inhibiting the formation and progression of the RNA polymerase II complex at the primary promoter [378]. Such noncoding, interfering RNAs are known to regulate expression of the dhfr gene in humans, for example, although in this case the interfering RNA is transcribed from a promoter that lies upstream of the primary promoter [379, 380]. Much more work will need to be performed in the future to identify the regulatory action that these 3'-UTR promoters have on their primary transcripts, if any.

Evidence	CpG-related	Number of promoters	Number of active promoters	(%)
KnownGene	Yes	28	2	7.10%
	No	792	84	10.60%
CAGE	Yes	616	38	6.20%
	No	18286	2533	13.90%
FirstEF	Yes	129	26	20.20%
	No	184	48	26.10%
KnownGene, CAGE	Yes	93	9	9.70%
	No	910	81	8.90%
KnownGene, FirstEF	Yes	63	25	39.70%
	No	45	10	22.20%
CAGE, FirstEF	Yes	1123	440	39.20%
	No	541	87	16.10%
KnownGene, CAGE, FirstEF	Yes	3202	2170	67.80%
	No	154	60	39.00%

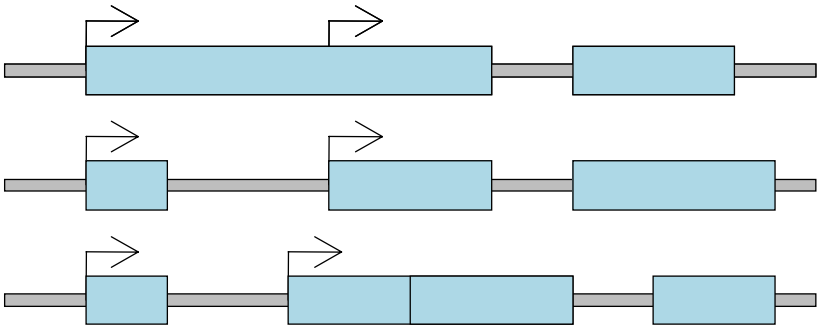
Table 6.1: Activity of promoters with various combinations of supporting evidence.

Gene Symbol	Genomic location (hg18)	Evidence
ACPT	chr19:55989804-55990068	CAGE tags
PPP2R2A	chr8:25959258-25959687	FirstEF
ZNF85	chr19:20897730-20898104	KnownGene
APEG1	chr2:220021650-220021924	FirstEF + CAGE tags
KCNK3	chr2:26804455-26804795	FirstEF + CAGE tags
PPFIA3	chr19:54322857-54323287	FirstEF + CAGE tags
SHRM	chr4:77829597-77830023	FirstEF + CAGE tags
SOX12	chr20:253659-254066	FirstEF + CAGE tags

Table 6.2: Promoters validated by ChIP-PCR and the lines of evidence used to identify them.

Figure 6.1: Alternative promoter selection. Alternative promoters can take on several forms (A): Two promoters on a single exon (top); alternative first exons (middle); a downstream promoter is located within the intron region of another isoform (bottom). The median number of promoters per gene on our microarray is three (B). There are a significant number of single-promoter genes on the array, but these are invariably share a bidirectional promoter with multi-promoter genes.

A



B

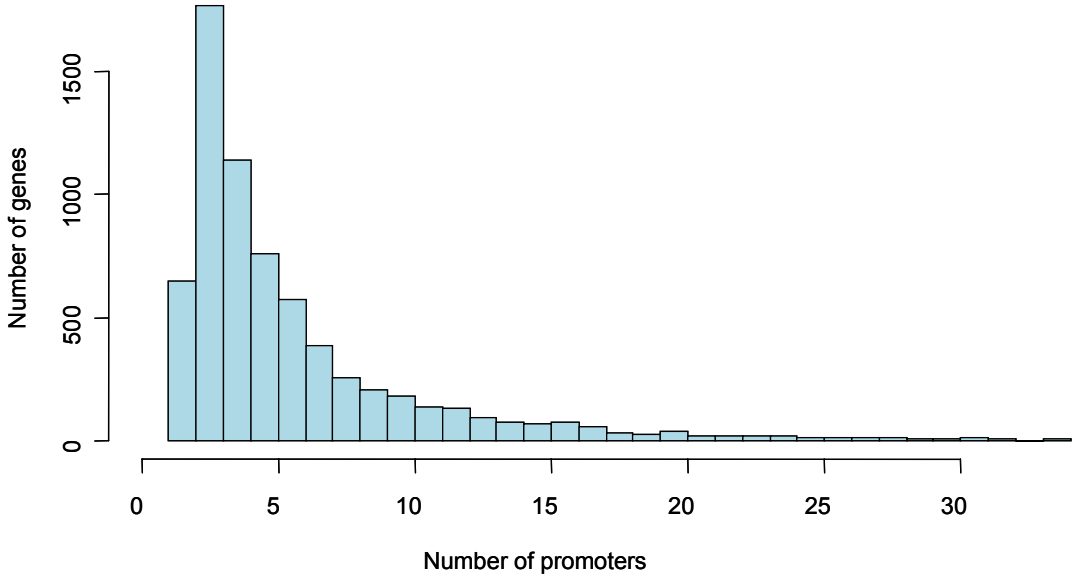

Number of genes
0
500
1000
1500
0
5
10
15
20
25
30
Number of promoters

Figure 6.2: ChIP-chip. (A). MA plot for a control experiment, after normalization (B; M = log2(Red/Green); A= log2(Red*Green)/2). Fit of the gamma+normal model to the log ratio of red versus green channels (C). The red portion of the histogram shows probes that belong to the unbound distribution with $p<0.05$. The green portion of the histogram are probes that belong to the bound distribution with $p<0.05$. The grey areas in between are ambiguous. Our model allows us to annotate promoters as being active or inactive at different confidence levels (D). “High on” indicates strong evidence for RNA Polymerase II binding in both replicates (the probes fall within the green portion of panel C); “Medium on” indicates strong evidence for RNA Polymerase II binding in one replicate, and weak evidence in the other (i.e., the probes fall outside of the red area in panel C). “Low on” indicates weak evidence in both replicates. “Low off” indicates inconsistency between the replicates, and finally “Strong off” indicates a high probability that no binding occurred (i.e., probes fall within the red portion of panel C).

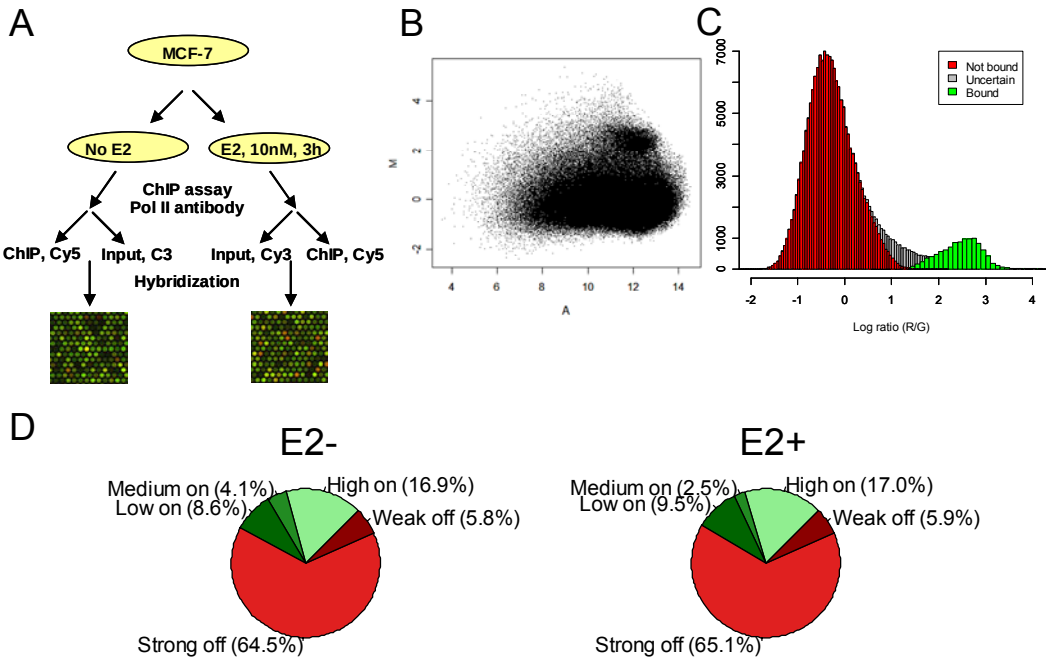

A
MCF-7
No E2
E2, 10nM, 3h
ChIP assay
Pol II antibody
ChIP, Cy5
Input, C3
Input, Cy3
ChIP, Cy5
Hybridization
B
M
A
C
Not bound
Uncertain
Bound
Log ratio (R/G)
D
E2-
Medium on (4.1%)
Low on (8.6%)
High on (16.9%)
Weak off (5.8%)
Strong off (64.5%)
E2+
Medium on (2.5%)
Low on (9.5%)
High on (17.0%)
Weak off (5.9%)
Strong off (65.1%)

Figure 6.3: ChIP-PCR. Seven out of ten promoters were confirmed to be active based on ChIP-PCR assays (green bars). Similarly, all but one of the promoters called as negative showed no evidence for RNA polymerase II binding (red bars). Error bars indicate standard errors from the mean, based on three replicates.

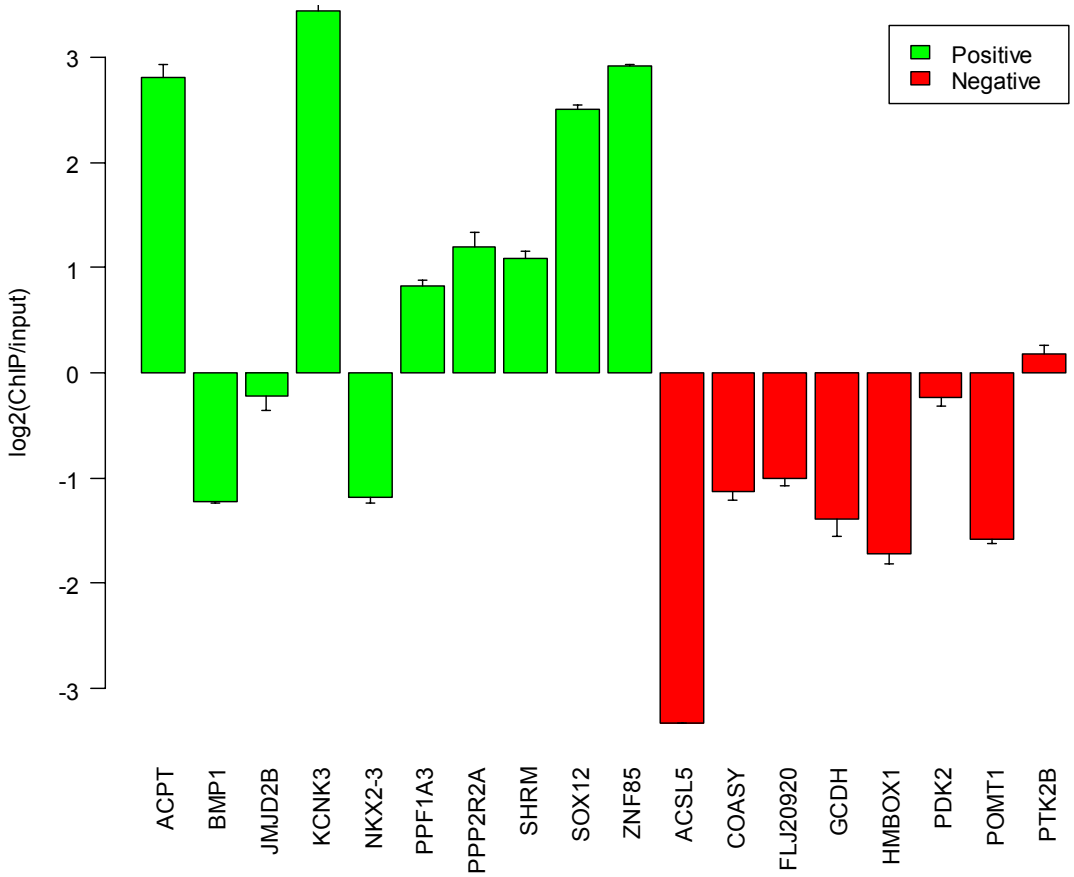

Positive
Negative
log2(ChIP/input)
3
2
1
0
-1
-2
-3
ACPT
BMP1
JMJD2B
KCNK3
NKX2-3
PPF1A3
PPP2R2A
SHRM
SOX12
ZNF85
ACSL5
COASY
FLJ20920
GCDH
HMBOX1
PDK2
POMT1
PTK2B

Figure 6.4: Promoter usage of NCOA7. Shown here are the first four exons of the gene NCOA7, spanning a region of roughly 32kb (A). Exon 3 is spliced out of the transcript initiated at Exon 1, but Exon 4 is common to both transcripts. The ChIP-chip microarray analysis indicated that the first promoter is inactive in the control experiment, but is activated with E2 treatment at a low level, a result that is verified by qRT-PCR results (B). The second promoter was predicted to be active at a low level with and without E2 treatment, which again was verified (C). Error bars indicate standard errors from the mean for three replicates.

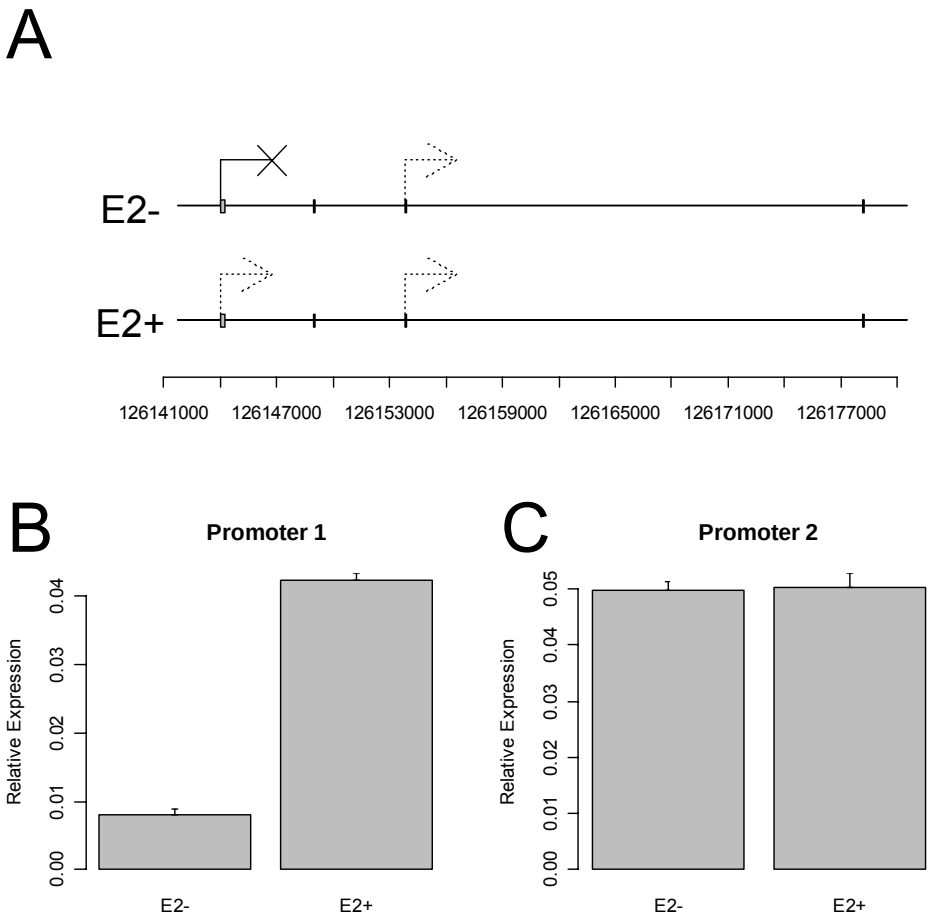
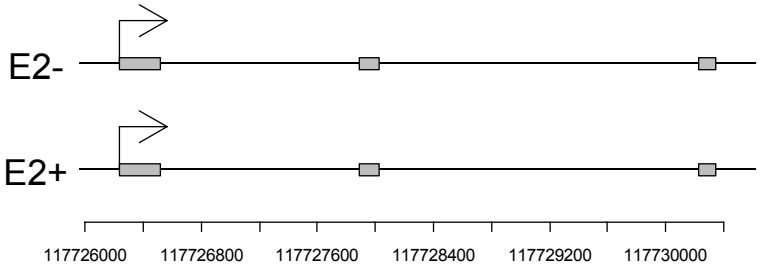

A
E2-
E2+
126141000
126147000
126153000
126159000
126165000
126171000
126177000
B
Promoter 1
Relative Expression
0.00
0.01
0.02
0.03
0.04
E2-
E2+
C
Promoter 2
Relative Expression
0.00
0.01
0.02
0.03
0.04
0.05
E2-
E2+

Figure 6.5: Promoter usage of EIF3S9. Shown here are the first three exons of the gene EIF3S9, spanning a region of approximately 3.7kb (A). This promoter was shown to be highly active in both treatments, which was verified by qRT-PCR.

A


E2-
E2+
117726000
117726800
117727600
117728400
117729200
117730000

B

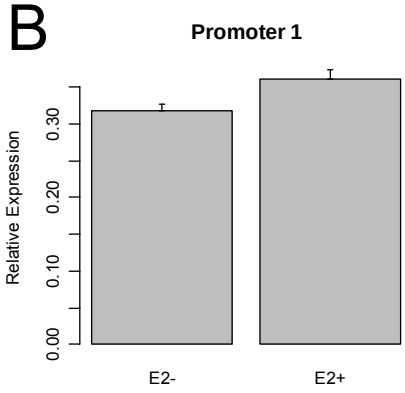

Promoter 1
Relative Expression
0.00
0.10
0.20
0.30
E2-
E2+

Figure 6.6: Multiple promoter usage in MCF-7 cells with E2 treatment. We found a total of 212 genes that had exactly two promoters that were active in one of these experiments. Of these, the upstream promoter was activated by E2 in 25 genes (A), and was inactivated by E2 in 61 cases (B). The downstream promoter was activated by E2 treatment in 62 cases (C), and inactivated by E2 in 64 cases (D).

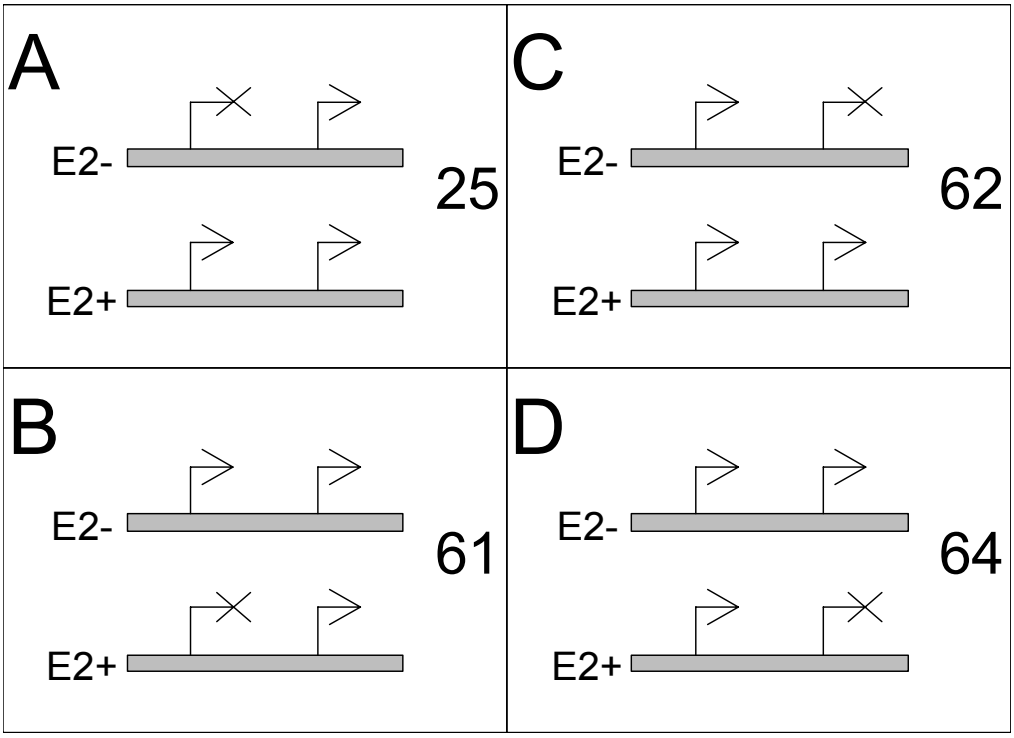
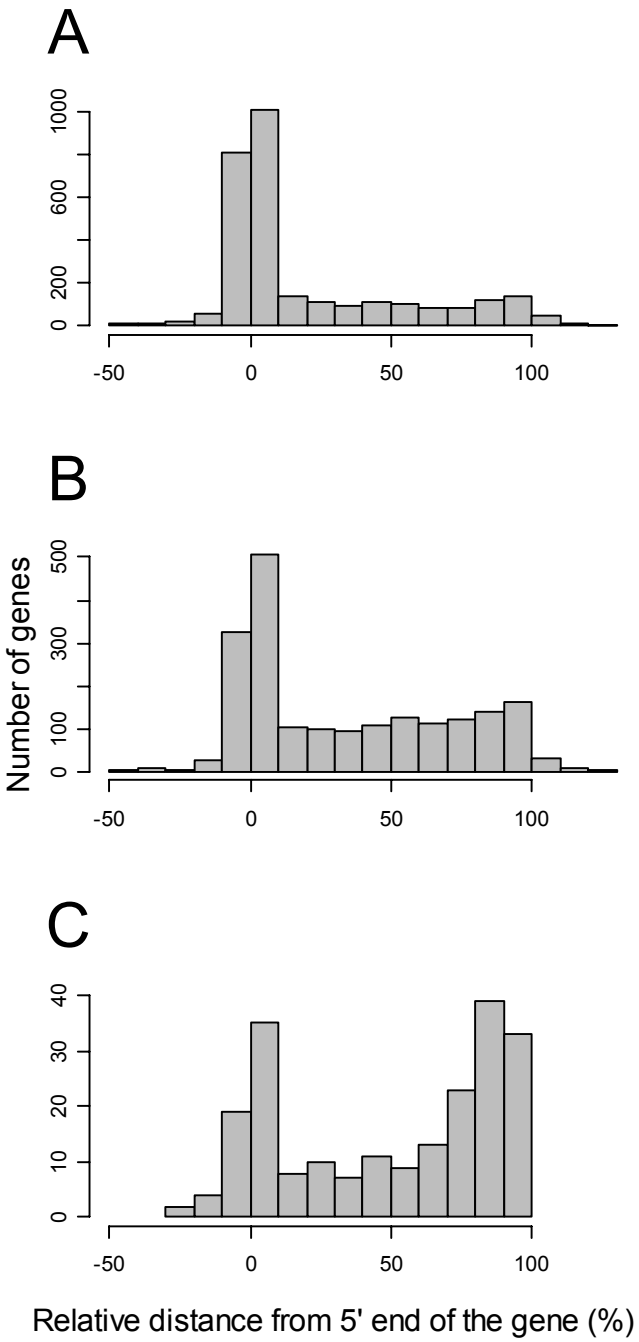

A
E2-
E2+
25
C
E2-
E2+
62
B
E2-
E2+
61
D
E2-
E2+
64

Figure 6.7: Different patters of alternative promoters in MCF-7 with E2 treatment. For genes with a single active promoter, there is a strong tendency for that promoter to be located at the 5’ end of the annotated gene (A). Similar trends are observed in the case of genes with two active promoters where neither is affected by E2 treatment. Here, one of the promoters is likely to be at the 5’ end of the gene, while the other promoter can occur anywhere else along the gene length with roughly equal probability (B). A different pattern is observed in genes with two active promoters where one is affected by E2 treatment (either activated or inactivated). In this case we can see that, as before, the upstream active promoter is likely to be located at the 5’ end of the gene, but the downstream promoter is strongly biased towards the 3’ end of the gene (C).


A
B
C
Number of genes
0
200
600
1000
0
100
300
500
0
10
20
30
40
-50
0
50
100
Relative distance from 5' end of the gene (%)

CHAPTER 7

SUMMARY AND PERSPECTIVES

We designed the alternative promoter array covering all the regions that may have multiple promoters based on current studies. With this alternative promoter array, we performed ChIP-chip to extract information of promoter usage in breast cancer cell lines under estrogen stimulation. We then showed that multiple promoters were used in MCF-7 cells and under E2 treatment, alternative promoters were activated differentially.

One current focus of functional genomics is to determine transcripts expressed in genomes and how these transcripts are regulated. In contrast to the original concept that one gene is regulated by its specific promoter in eukaryotes, alternative promoter usage was found to be a common phenomenon. Consequently, complete characterization of promoter regions of these transcripts is necessary to determine their functions and regulation. This study is a step toward addressing this challenge. No clear interpretation has been found for the results we produced. However, we can not rule out the possibility that the usage of alternative promoters in breast cancer cells is relative to temporal and spatial expression of various transcripts under specific conditions. Meanwhile the

findings in this study also raise more questions. For example, will alternative usage of promoters play a role in breast cancer development? Or how are alternative promoters in the same region controlled by upstream signals? Answers to these questions may be critical to understand the differential translations of genetic information in both normal and diseased cells.

REFERENCES

1. Roberts, L., et al., A history of the Human Genome Project. Science, 2001. 291(5507): p. 1195.

2. Collins, F.S., et al., A vision for the future of genomics research. Nature, 2003. 422(6934): p. 835-47.

3. Waddington, C.H., The epigenotype. Endeavour, 1942. 1: p. 18-20.

4. Goldberg, A.D., C.D. Allis, and E. Bernstein, Epigenetics: a landscape takes shape. Cell, 2007. 128(4): p. 635-8.

5. Shi, Y., Taking Epigenetics Center Stage. Cell, 2007. 128(4): p. 639-40.

6. Bestor, T.H., The DNA methyltransferases of mammals. Hum Mol Genet, 2000. 9(16): p. 2395-402.

7. Haines, T.R., D.I. Rodenhiser, and P.J. Ainsworth, Allele-specific non-CpG methylation of the Nf1 gene during early mouse development. Dev Biol, 2001. 240(2): p. 585-98.

8. Lukens, L.N. and S. Zhan, The plant genome's methylation status and response to stress: implications for plant improvement. Curr Opin Plant Biol, 2007. 10(3): p. 317-22.

9. Bird, A., DNA methylation patterns and epigenetic memory. Genes Dev, 2002. 16(1): p. 6-21.

10. Worm, J. and P. Guldberg, DNA methylation: an epigenetic pathway to cancer and a promising target for anticancer therapy. J Oral Pathol Med, 2002. 31(8): p. 443-9.

11. Bird, A.P., DNA methylation and the frequency of CpG in animal DNA. Nucleic Acids Res, 1980. 8(7): p. 1499-504.

12. Cooper, D.N. and M. Krawczak, Cytosine methylation and the fate of CpG dinucleotides in vertebrate genomes. Hum Genet, 1989. 83(2): p. 181-8.

13. Clark, S.J. and J. Melki, DNA methylation and gene silencing in cancer: which is the guilty party? Oncogene, 2002. 21(35): p. 5380-7.

14. Gardiner-Garden, M. and M. Frommer, CpG islands in vertebrate genomes. J Mol Biol, 1987. 196(2): p. 261-82.

15. Baylin, S.B. and J.G. Herman, DNA hypermethylation in tumorigenesis: epigenetics joins genetics. Trends Genet, 2000. 16(4): p. 168-74.

16. Osborne, L.R., et al., A 1.5 million-base pair inversion polymorphism in families with Williams-Beuren syndrome. Nat Genet, 2001. 29(3): p. 321-5.

17. Pradhan, S. and P.O. Esteve, Mammalian DNA (cytosine-5) methyltransferases and their expression. Clin Immunol, 2003. 109(1): p. 6-16.

18. Smith, S.S., et al., Mechanism of human methyl-directed DNA methyltransferase and the fidelity of cytosine methylation. Proc Natl Acad Sci U S A, 1992. 89(10): p. 4744-8.

19. Yen, R.W., et al., Isolation and characterization of the cDNA encoding human DNA methyltransferase. Nucleic Acids Res, 1992. 20(9): p. 2287-91.

20. Jones, P.A. and S.M. Taylor, Cellular differentiation, cytidine analogs and DNA methylation. Cell, 1980. 20(1): p. 85-93.

21. Pradhan, S., et al., Recombinant human DNA (cytosine-5) methyltransferase. I. Expression, purification, and comparison of de novo and maintenance methylation. J Biol Chem, 1999. 274(46): p. 33002-10.

22. Gaudet, F., et al., Induction of tumors in mice by genomic hypomethylation. Science, 2003. 300(5618): p. 489-92.

23. Li, E., C. Beard, and R. Jaenisch, Role for DNA methylation in genomic imprinting. Nature, 1993. 366(6453): p. 362-5.

24. 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.

25. Mertineit, C., et al., Sex-specific exons control DNA methyltransferase in mammalian germ cells. Development, 1998. 125(5): p. 889-97.

26. Howell, C.Y., et al., Genomic imprinting disrupted by a maternal effect mutation in the Dnmt1 gene. Cell, 2001. 104(6): p. 829-38.

27. Yoder, J.A. and T.H. Bestor, A candidate mammalian DNA methyltransferase related to pmt1p of fission yeast. Hum Mol Genet, 1998. 7(2): p. 279-84.

28. Liu, K., et al., Endogenous assays of DNA methyltransferases: Evidence for differential activities of DNMT1, DNMT2, and DNMT3 in mammalian cells in vivo. Mol Cell Biol, 2003. 23(8): p. 2709-19.

29. Bhattacharya, S.K., et al., A mammalian protein with specific demethylase activity for mCpG DNA. Nature, 1999. 397(6720): p. 579-83.

30. Ng, H.H., et al., MBD2 is a transcriptional repressor belonging to the MeCP1 histone deacetylase complex. Nat Genet, 1999. 23(1): p. 58-61.

31. Robertson, K.D., et al., The human DNA methyltransferases (DNMTs) 1, 3a and 3b: coordinate mRNA expression in normal tissues and overexpression in tumors. Nucleic Acids Res, 1999. 27(11): p. 2291-8.

32. Xie, S., et al., Cloning, expression and chromosome locations of the human DNMT3 gene family. Gene, 1999. 236(1): p. 87-95.

33. 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.

34. Xu, G.L., et al., Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene. Nature, 1999. 402(6758): p. 187-91.

35. Chedin, F., M.R. Lieber, and C.L. Hsieh, The DNA methyltransferase-like protein DNMT3L stimulates de novo methylation by Dnmt3a. Proc Natl Acad Sci U S A, 2002. 99(26): p. 16916-21.

36. Kaneda, M., et al., Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting. Nature, 2004. 429(6994): p. 900-3.

37. Suetake, I., et al., DNMT3L stimulates the DNA methylation activity of Dnmt3a and Dnmt3b through a direct interaction. J Biol Chem, 2004. 279(26): p. 27816-23.

38. Hata, K., et al., Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice. Development, 2002. 129(8): p. 1983-93.

39. Leonhardt, H., H.P. Rahn, and M.C. Cardoso, Intranuclear targeting of DNA replication factors 1. J.Cell Biochem.Suppl, 1998. 30-31: p. 243-249.

40. Laird, P.W., Oncogenic mechanisms mediated by DNA methylation 3. Mol.Med.Today, 1997. 3(5): p. 223-229.

41. Mayer, W., et al., Demethylation of the zygotic paternal genome 3. Nature, 2000. 403(6769): p. 501-502.

42. Paroush, Z., et al., Dynamics of demethylation and activation of the alpha-actin gene in myoblasts 1. Cell, 1990. 63(6): p. 1229-1237.

43. Jost, J.P., Nuclear extracts of chicken embryos promote an active demethylation of DNA by excision repair of 5-methyldeoxycytidine 1. Proc.Natl.Acad.Sci.U.S.A, 1993. 90(10): p. 4684-4688.

44. Jost, J.P., et al., The RNA moiety of chick embryo 5-methylcytosine- DNA glycosylase targets DNA demethylation 2. Nucleic Acids Res., 1997. 25(22): p. 4545-4550.

45. Weiss, A., et al., DNA demethylation in vitro: involvement of RNA 1. Cell, 1996. 86(5): p. 709-718.

46. Swisher, J.F., et al., Analysis of putative RNase sensitivity and protease insensitivity of demethylation activity in extracts from rat myoblasts 1. Nucleic Acids Res., 1998. 26(24): p. 5573-5580.

47. Cervoni, N., S. Bhattacharya, and M. Szyf, DNA demethylase is a processive enzyme 3. J.Biol.Chem., 1999. 274(13): p. 8363-8366.

48. Ramchandani, S., et al., DNA methylation is a reversible biological signal 1. Proc.Natl.Acad.Sci.U.S.A, 1999. 96(11): p. 6107-6112.

49. Wade, P.A., et al., Mi-2 complex couples DNA methylation to chromatin remodelling and histone deacetylation. Nat Genet, 1999. 23(1): p. 62-6.

50. Hendrich, B., et al., Closely related proteins MBD2 and MBD3 play distinctive but interacting roles in mouse development 2. Genes Dev., 2001. 15(6): p. 710-723.

51. Ng, H.H., et al., MBD2 is a transcriptional repressor belonging to the MeCP1 histone deacetylase complex 4. Nat.Genet., 1999. 23(1): p. 58-61.

52. Sansom, O.J., et al., Deficiency of Mbd2 suppresses intestinal tumorigenesis 1. Nat.Genet., 2003. 34(2): p. 145-147.

53. Barreto, G., et al., Gadd45a promotes epigenetic gene activation by repair-mediated DNA demethylation. Nature, 2007. 445(7128): p. 671-5.

54. Maloisel, L. and J.L. Rossignol, Suppression of crossing-over by DNA methylation in Ascobolus 1. Genes Dev., 1998. 12(9): p. 1381-1389.

55. Hsieh, C.L. and M.R. Lieber, CpG methylated minichromosomes become inaccessible for V(D)J recombination after undergoing replication 2. EMBO J., 1992. 11(1): p. 315-325.

56. Chen, R.Z., et al., DNA hypomethylation leads to elevated mutation rates 10. Nature, 1998. 395(6697): p. 89-93.

57. Miniou, P., et al., Abnormal methylation pattern in constitutive and facultative (X inactive chromosome) heterochromatin of ICF patients 1. Hum.Mol.Genet., 1994. 3(12): p. 2093-2102.

58. Ji, W., et al., DNA demethylation and pericentromeric rearrangements of chromosome 1 1. Mutat.Res., 1997. 379(1): p. 33-41.

59. Cooper, D.N. and H. Youssoufian, The CpG dinucleotide and human genetic disease 1. Hum.Genet., 1988. 78(2): p. 151-155.

60. Greenblatt, M.S., et al., Mutations in the p53 tumor suppressor gene: clues to cancer etiology and molecular pathogenesis 1. Cancer Res., 1994. 54(18): p. 4855-4878.

61. Coulondre, C., et al., Molecular basis of base substitution hotspots in Escherichia coli 2. Nature, 1978. 274(5673): p. 775-780.

62. Ciotta, C., et al., Increased somatic recombination in methylation tolerant human cells with defective DNA mismatch repair 3. J.Mol.Biol., 1998. 276(4): p. 705-719.

63. Akhmanova, A., et al., Characterisation of transcriptionally active and inactive chromatin domains in neurons 1. J.Cell Sci., 2000. 113 Pt 24: p. 4463-4474.

64. Antequera, F. and A. Bird, Number of CpG islands and genes in human and mouse 1. Proc.Natl.Acad.Sci.U.S.A, 1993. 90(24): p. 11995-11999.

65. Tazi, J. and A. Bird, Alternative chromatin structure at CpG islands. Cell, 1990. 60(6): p. 909-20.

66. Cooper, D.N. and M. Krawczak, Cytosine methylation and the fate of CpG dinucleotides in vertebrate genomes 1. Hum.Genet., 1989. 83(2): p. 181-188.

67. Montagna, M., et al., Identification of a 3 kb Alu-mediated BRCA1 gene rearrangement in two breast/ovarian cancer families 1. Oncogene, 1999. 18(28): p. 4160-4165.

68. Kazazian, H.H., Jr. and J.V. Moran, The impact of L1 retrotransposons on the human genome 4. Nat.Genet., 1998. 19(1): p. 19-24.

69. Bird, A.P., CpG-rich islands and the function of DNA methylation 1. Nature, 1986. 321(6067): p. 209-213.

70. Yeivin, A. and A. Razin, Gene methylation patterns and expression 2. EXS, 1993. 64: p. 523-568.

71. Kass, S.U., N. Landsberger, and A.P. Wolffe, DNA methylation directs a time-dependent repression of transcription initiation. Curr Biol, 1997. 7(3): p. 157-65.

72. Taylor, S.M., P.A. Constantinides, and P.A. Jones, 5-Azacytidine, DNA methylation, and differentiation 7. Curr.Top.Microbiol.Immunol., 1984. 108: p. 115-127.

73. Reik, W., W. Dean, and J. Walter, Epigenetic reprogramming in mammalian development 4. Science, 2001. 293(5532): p. 1089-1093.

74. Cervoni, N. and M. Szyf, Demethylase activity is directed by histone acetylation 1. J.Biol.Chem., 2001. 276(44): p. 40778-40787.

75. Comb, M. and H.M. Goodman, CpG methylation inhibits proenkephalin gene expression and binding of the transcription factor AP-2 1. Nucleic Acids Res., 1990. 18(13): p. 3975-3982.

76. Hendrich, B. and A. Bird, Identification and characterization of a family of mammalian methyl-CpG binding proteins 1. Mol.Cell Biol., 1998. 18(11): p. 6538-6547.

77. Lewis, J.D., et al., Purification, sequence, and cellular localization of a novel chromosomal protein that binds to methylated DNA 2. Cell, 1992. 69(6): p. 905-914.

78. Fujita, N., et al., Mechanism of transcriptional regulation by methyl-CpG binding protein MBD1 2. Mol.Cell Biol., 2000. 20(14): p. 5107-5118.

79. Wade, P.A., et al., A multiple subunit Mi-2 histone deacetylase from Xenopus laevis cofractionates with an associated Snf2 superfamily ATPase 2. Curr.Biol., 1998. 8(14): p. 843-846.

80. Boeke, J., et al., The minimal repression domain of MBD2b overlaps with the methyl-CpG-binding domain and binds directly to Sin3A 1. J.Biol.Chem., 2000. 275(45): p. 34963-34967.

81. Yu, F., J. Thiesen, and W.H. Stratling, Histone deacetylase-independent transcriptional repression by methyl-CpG-binding protein 2 2. Nucleic Acids Res., 2000. 28(10): p. 2201-2206.

82. Fuks, F., DNA methylation and histone modifications: teaming up to silence genes. Curr Opin Genet Dev, 2005. 15(5): p. 490-5.

83. Wolffe, A.P. and H. Kurumizaka, The nucleosome: a powerful regulator of transcription. Prog Nucleic Acid Res Mol Biol, 1998. 61: p. 379-422.

84. Kornberg, R.D., Chromatin structure: a repeating unit of histones and DNA. Science, 1974. 184(139): p. 868-71.

85. Ausio, J., Histone variants--the structure behind the function. Brief Funct Genomic Proteomic, 2006. 5(3): p. 228-43.

86. Tagami, H., et al., Histone H3.1 and H3.3 complexes mediate nucleosome assembly pathways dependent or independent of DNA synthesis. Cell, 2004. 116(1): p. 51-61.

87. Wu, F. and A.S. Lee, YY1 as a regulator of replication-dependent hamster histone H3.2 promoter and an interactive partner of AP-2. J Biol Chem, 2001. 276(1): p. 28-34.

88. Chow, C.M., et al., Variant histone H3.3 marks promoters of transcriptionally active genes during mammalian cell division. EMBO Rep, 2005. 6(4): p. 354-60.

89. Allan, J., et al., The structure of histone H1 and its location in chromatin. Nature, 1980. 288(5792): p. 675-9.

90. Fischle, W., Y. Wang, and C.D. Allis, Histone and chromatin cross-talk. Curr Opin Cell Biol, 2003. 15(2): p. 172-83.

91. Van Holde, K.E., Chromatin. 1988: New York, Springer.

92. 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.

93. Hebbes, T.R., A.W. Thorne, and C. Crane-Robinson, A direct link between core histone acetylation and transcriptionally active chromatin. Embo J, 1988. 7(5): p. 1395-402.

94. Grant, P.A. and S.L. Berger, Histone acetyltransferase complexes. Semin Cell Dev Biol, 1999. 10(2): p. 169-77.

95. Roth, S.Y., J.M. Denu, and C.D. Allis, Histone acetyltransferases. Annu Rev Biochem, 2001. 70: p. 81-120.

96. Turner, B.M., A.J. Birley, and J. Lavender, Histone H4 isoforms acetylated at specific lysine residues define individual chromosomes and chromatin domains in *Drosophila* polytene nuclei. Cell, 1992. 69(2): p. 375-84.

97. Braunstein, M., et al., Efficient transcriptional silencing in Saccharomyces cerevisiae requires a heterochromatin histone acetylation pattern. Mol Cell Biol, 1996. 16(8): p. 4349-56.

98. Reifsnyder, C., et al., Yeast SAS silencing genes and human genes associated with AML and HIV-1 Tat interactions are homologous with acetyltransferases. Nat Genet, 1996. 14(1): p. 42-9.

99. Kelly, T.J., et al., Type B histone acetyltransferase Hat1p participates in telomeric silencing. Mol Cell Biol, 2000. 20(19): p. 7051-8.

100. Brownell, J.E. and C.D. Allis, An activity gel assay detects a single, catalytically active histone acetyltransferase subunit in Tetrahymena macronuclei. Proc Natl Acad Sci U S A, 1995. 92(14): p. 6364-8.

101. Kimura, A., K. Matsubara, and M. Horikoshi, A decade of histone acetylation: marking eukaryotic chromosomes with specific codes. J Biochem (Tokyo), 2005. 138(6): p. 647-62.

102. Brownell, J.E., et al., Tetrahymena histone acetyltransferase A: a homolog to yeast Gcn5p linking histone acetylation to gene activation. Cell, 1996. 84(6): p. 843-51.

103. Wu, J. and M. Grunstein, 25 years after the nucleosome model: chromatin modifications. Trends Biochem Sci, 2000. 25(12): p. 619-23.

104. Grant, P.A., A tale of histone modifications. Genome Biol, 2001. 2(4): p. REVIEWS0003.

105. John, S., et al., The something about silencing protein, Sas3, is the catalytic subunit of NuA3, a yTAF(II)30-containing HAT complex that interacts with the Spt16 subunit of the yeast CP (Cdc68/Pob3)-FACT complex. Genes Dev, 2000. 14(10): p. 1196-208.

106. Vogelauer, M., et al., Global histone acetylation and deacetylation in yeast. Nature, 2000. 408(6811): p. 495-8.

107. Ikura, T., et al., Involvement of the TIP60 histone acetylase complex in DNA repair and apoptosis. Cell, 2000. 102(4): p. 463-73.

108. Dhalluin, C., et al., Structure and ligand of a histone acetyltransferase bromodomain. Nature, 1999. 399(6735): p. 491-6.

109. Jacobson, R.H., et al., Structure and function of a human TAFII250 double bromodomain module. Science, 2000. 288(5470): p. 1422-5.

110. McLaughlin, F. and N.B. La Thangue, Histone deacetylase inhibitors open new doors in cancer therapy. Biochem Pharmacol, 2004. 68(6): p. 1139-44.

111. Yang, X.J. and E. Seto, Collaborative spirit of histone deacetylases in regulating chromatin structure and gene expression. Curr Opin Genet Dev, 2003. 13(2): p. 143-53.

112. Grozinger, C.M. and S.L. Schreiber, Deacetylase enzymes: biological functions and the use of small-molecule inhibitors. Chem Biol, 2002. 9(1): p. 3-16.

113. Ballas, N., et al., Regulation of neuronal traits by a novel transcriptional complex. Neuron, 2001. 31(3): p. 353-65.

114. Li, J., et al., Specific targeting and constitutive association of histone deacetylase complexes during transcriptional repression. Genes Dev, 2002. 16(6): p. 687-92.

115. Legube, G. and D. Trouche, Regulating histone acetyltransferases and deacetylases. EMBO Rep, 2003. 4(10): p. 944-7.

116. Hassan, A.H., et al., Function and selectivity of bromodomains in anchoring chromatin-modifying complexes to promoter nucleosomes. Cell, 2002. 111(3): p. 369-79.

117. Agalioti, T., G. Chen, and D. Thanos, Deciphering the transcriptional histone acetylation code for a human gene. Cell, 2002. 111(3): p. 381-92.

118. Hecht, A., et al., Histone H3 and H4 N-termini interact with SIR3 and SIR4 proteins: a molecular model for the formation of heterochromatin in yeast. Cell, 1995. 80(4): p. 583-92.

119. Carmen, A.A., L. Milne, and M. Grunstein, Acetylation of the yeast histone H4 N terminus regulates its binding to heterochromatin protein SIR3. J Biol Chem, 2002. 277(7): p. 4778-81.

120. Edmondson, D.G., M.M. Smith, and S.Y. Roth, Repression domain of the yeast global repressor Tup1 interacts directly with histones H3 and H4. Genes Dev, 1996. 10(10): p. 1247-59.

121. Clarke, S., Protein methylation. Curr Opin Cell Biol, 1993. 5(6): p. 977-83.

122. Martin, C. and Y. Zhang, The diverse functions of histone lysine methylation. Nat Rev Mol Cell Biol, 2005. 6(11): p. 838-49.

123. Peters, A.H., et al., Partitioning and plasticity of repressive histone methylation states in mammalian chromatin. Mol Cell, 2003. 12(6): p. 1577-89.

124. Schotta, G., et al., A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin. Genes Dev, 2004. 18(11): p. 1251-62.

125. Rea, S., et al., Regulation of chromatin structure by site-specific histone H3 methyltransferases. Nature, 2000. 406(6796): p. 593-9.

126. Kouzarides, T., Histone methylation in transcriptional control. Curr Opin Genet Dev, 2002. 12(2): p. 198-209.

127. Lachner, M. and T. Jenuwein, The many faces of histone lysine methylation. Curr Opin Cell Biol, 2002. 14(3): p. 286-98.

128. Nielsen, S.J., et al., Rb targets histone H3 methylation and HP1 to promoters. Nature, 2001. 412(6846): p. 561-5.

129. Tachibana, M., et al., Set domain-containing protein, G9a, is a novel lysine-preferring mammalian histone methyltransferase with hyperactivity and specific selectivity to lysines 9 and 27 of histone H3. J Biol Chem, 2001. 276(27): p. 25309-17.

130. Tachibana, M., et al., G9a histone methyltransferase plays a dominant role in euchromatic histone H3 lysine 9 methylation and is essential for early embryogenesis. Genes Dev, 2002. 16(14): p. 1779-91.

131. Metzger, E. and R. Schule, The expanding world of histone lysine demethylases. Nat Struct Mol Biol, 2007. 14(4): p. 252-4.

132. 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.

133. Bird, A.P. and A.P. Wolffe, Methylation-induced repression--belts, braces, and chromatin. Cell, 1999. 99(5): p. 451-4.

134. Eden, S., et al., DNA methylation models histone acetylation. Nature, 1998. 394(6696): p. 842.

135. Nan, X., et al., Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. Nature, 1998. 393(6683): p. 386-9.

136. Jones, P.L., et al., Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. Nat Genet, 1998. 19(2): p. 187-91.

137. Richards, E.J. and S.C. Elgin, Epigenetic codes for heterochromatin formation and silencing: rounding up the usual suspects. Cell, 2002. 108(4): p. 489-500.

138. Tamaru, H. and E.U. Selker, A histone H3 methyltransferase controls DNA methylation in Neurospora crassa. Nature, 2001. 414(6861): p. 277-83.

139. Jackson, J.P., et al., Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase. Nature, 2002. 416(6880): p. 556-60.

140. Lehnertz, B., et al., Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin. Curr Biol, 2003. 13(14): p. 1192-200.

141. Fuks, F., et al., The DNA methyltransferases associate with HP1 and the SUV39H1 histone methyltransferase. Nucleic Acids Res, 2003. 31(9): p. 2305-12.

142. Soppe, W.J., et al., DNA methylation controls histone H3 lysine 9 methylation and heterochromatin assembly in Arabidopsis. Embo J, 2002. 21(23): p. 6549-59.

143. Fahrner, J.A., et al., Dependence of histone modifications and gene expression on DNA hypermethylation in cancer. Cancer Res, 2002. 62(24): p. 7213-8.

144. Espada, J., et al., Human DNA methyltransferase 1 is required for maintenance of the histone H3 modification pattern. J Biol Chem, 2004. 279(35): p. 37175-84.

145. Fuks, F., et al., The methyl-CpG-binding protein MeCP2 links DNA methylation to histone methylation. J Biol Chem, 2003. 278(6): p. 4035-40.

146. Padjen, K., S. Ratnam, and U. Storb, DNA methylation precedes chromatin modifications under the influence of the strain-specific modifier Ssm1. Mol Cell Biol, 2005. 25(11): p. 4782-91.

147. Robertson, K.D. and A.P. Wolffe, DNA methylation in health and disease 1. Nat.Rev.Genet., 2000. 1(1): p. 11-19.

148. Gama-Sosa, M.A., et al., The 5-methylcytosine content of DNA from human tumors 5. Nucleic Acids Res., 1983. 11(19): p. 6883-6894.

149. Feinberg, A.P., et al., Reduced genomic 5-methylcytosine content in human colonic neoplasia 2. Cancer Res., 1988. 48(5): p. 1159-1161.

150. Dante, R., et al., Methylation patterns of long interspersed repeated DNA and alphoid repetitive DNA from human cell lines and tumors 2. Anticancer Res., 1992. 12(2): p. 559-563.

151. Baylin, S.B., et al., DNA methylation patterns of the calcitonin gene in human lung cancers and lymphomas 1. Cancer Res., 1986. 46(6): p. 2917-2922.

152. de Bustros, A., et al., The short arm of chromosome 11 is a "hot spot" for hypermethylation in human neoplasia 1. Proc.Natl.Acad.Sci.U.S.A, 1988. 85(15): p. 5693-5697.

153. Plass, C., Cancer epigenomics. Hum Mol Genet, 2002. 11(20): p. 2479-88.

154. Robertson, K.D., DNA methylation and human disease. Nat Rev Genet, 2005. 6(8): p. 597-610.

155. Yang, X.J., The diverse superfamily of lysine acetyltransferases and their roles in leukemia and other diseases. Nucleic Acids Res, 2004. 32(3): p. 959-76.

156. Petrij, F., et al., Rubinstein-Taybi syndrome caused by mutations in the transcriptional co-activator CBP. Nature, 1995. 376(6538): p. 348-51.

157. Peters, A.H., et al., Loss of the Suv39h histone methyltransferases impairs mammalian heterochromatin and genome stability. Cell, 2001. 107(3): p. 323-37.

158. Caldas, C. and S. Aparicio, Cell memory and cancer--the story of the trithorax and Polycomb group genes. Cancer Metastasis Rev, 1999. 18(2): p. 313-29.

159. Esteller, M., Cancer epigenomics: DNA methylomes and histone-modification maps. Nat Rev Genet, 2007. 8(4): p. 286-98.

160. Jones, P.A. and R. Martienssen, A Blueprint for a Human Epigenome Project: The AACR Human Epigenome Workshop. Cancer Res, 2005. 65(24): p. 11241-11246.

161. Martens, J.H., et al., The profile of repeat-associated histone lysine methylation states in the mouse epigenome. Embo J, 2005. 24(4): p. 800-12.

162. Hanlon, S.E. and J.D. Lieb, Progress and challenges in profiling the dynamics of chromatin and transcription factor binding with DNA microarrays. Curr Opin Genet Dev, 2004. 14(6): p. 697-705.

163. Kondo, Y., et al., Chromatin immunoprecipitation microarrays for identification of genes silenced by histone H3 lysine 9 methylation. Proc Natl Acad Sci U S A, 2004. 101(19): p. 7398-403.

164. Yu, L., et al., Global assessment of promoter methylation in a mouse model of cancer identifies ID4 as a putative tumor-suppressor gene in human leukemia. Nat Genet, 2005. 37(3): p. 265-74.

165. Hood, L., A personal view of molecular technology and how it has changed biology. J Proteome Res, 2002. 1(5): p. 399-409.

166. Rakyan, V.K., et al., DNA methylation profiling of the human major histocompatibility complex: a pilot study for the human epigenome project. PLoS Biol, 2004. 2(12): p. e405.

167. Egger, G., et al., Epigenetics in human disease and prospects for epigenetic therapy. Nature, 2004. 429(6990): p. 457-63.

168. Orlando, V., Mapping chromosomal proteins in vivo by formaldehyde-crosslinked-chromatin immunoprecipitation. Trends Biochem Sci, 2000. 25(3): p. 99-104.

169. Kuo, M.H. and C.D. Allis, In vivo cross-linking and immunoprecipitation for studying dynamic Protein:DNA associations in a chromatin environment. Methods, 1999. 19(3): p. 425-33.

170. Oberley, M.J., et al., High-throughput screening of chromatin immunoprecipitates using CpG-island microarrays. Methods Enzymol, 2004. 376: p. 315-34.

171. Bernstein, B.E., et al., The use of chromatin immunoprecipitation assays in genome-wide analyses of histone modifications. Methods Enzymol, 2004. 376: p. 349-60.

172. Buck, M.J. and J.D. Lieb, ChIP-chip: considerations for the design, analysis, and application of genome-wide chromatin immunoprecipitation experiments. Genomics, 2004. 83(3): p. 349-60.

173. Weinmann, A.S., Novel ChIP-based strategies to uncover transcription factor target genes in the immune system. Nat Rev Immunol, 2004. 4(5): p. 381-6.

174. Ren, B., et al., Genome-wide location and function of DNA binding proteins. Science, 2000. 290(5500): p. 2306-9.

175. Iyer, V.R., et al., Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF. Nature, 2001. 409(6819): p. 533-8.

176. Bowtell, D.D. and J.F. Sambrook, DNA Microarrays: A Molecular Cloning Manual. 2002, new york: Cold Spring Harbor Laboratory Press. 453-508.

177. Laub, M.T., et al., Genes directly controlled by CtrA, a master regulator of the Caulobacter cell cycle. Proc Natl Acad Sci U S A, 2002. 99(7): p. 4632-7.

178. Chua, Y.L., et al., Microarray analysis of chromatin-immunoprecipitated DNA identifies specific regions of tobacco genes associated with acetylated histones. Plant J, 2004. 37(6): p. 789-800.

179. Weinmann, A.S., et al., Isolating human transcription factor targets by coupling chromatin immunoprecipitation and CpG island microarray analysis. Genes Dev, 2002. 16(2): p. 235-44.

180. Ren, B., et al., E2F integrates cell cycle progression with DNA repair, replication, and G(2)/M checkpoints. Genes Dev, 2002. 16(2): p. 245-56.

181. Horak, C.E., et al., GATA-1 binding sites mapped in the beta-globin locus by using mammalian chIp-chip analysis. Proc Natl Acad Sci U S A, 2002. 99(5): p. 2924-9.

182. Kirmizis, A., et al., Silencing of human polycomb target genes is associated with methylation of histone H3 Lys 27. Genes Dev, 2004. 18(13): p. 1592-605.

183. Leu, Y.W., et al., Loss of estrogen receptor signaling triggers epigenetic silencing of downstream targets in breast cancer. Cancer Res, 2004. 64(22): p. 8184-92.

184. Odom, D.T., et al., Control of pancreas and liver gene expression by HNF transcription factors. Science, 2004. 303(5662): p. 1378-81.

185. d'Adda di Fagagna, F., et al., A DNA damage checkpoint response in telomere-initiated senescence. Nature, 2003. 426(6963): p. 194-8.

186. Martone, R., et al., Distribution of NF-kappaB-binding sites across human chromosome 22. Proc Natl Acad Sci U S A, 2003. 100(21): p. 12247-52.

187. Bernstein, B.E., et al., Genomic maps and comparative analysis of histone modifications in human and mouse. Cell, 2005. 120(2): p. 169-81.

188. Carroll, J.S., et al., Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1. Cell, 2005. 122(1): p. 33-43.

189. Cawley, S., et al., Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs. Cell, 2004. 116(4): p. 499-509.

190. Euskirchen, G., et al., CREB binds to multiple loci on human chromosome 22. Mol Cell Biol, 2004. 24(9): p. 3804-14.

191. Kim, T.H., et al., Direct isolation and identification of promoters in the human genome. Genome Res, 2005. 15(6): p. 830-9.

192. Kim, T.H., et al., A high-resolution map of active promoters in the human genome. Nature, 2005. 436(7052): p. 876-80.

193. Shi, H., et al., Triple analysis of the cancer epigenome: an integrated microarray system for assessing gene expression, DNA methylation, and histone acetylation. Cancer Res, 2003. 63(9): p. 2164-71.

194. Lippman, Z., et al., Role of transposable elements in heterochromatin and epigenetic control. Nature, 2004. 430(6998): p. 471-6.

195. Orlando, V., H. Strutt, and R. Paro, Analysis of chromatin structure by in vivo formaldehyde cross-linking. Methods, 1997. 11(2): p. 205-14.

196. Liu, C.L., S.L. Schreiber, and B.E. Bernstein, Development and validation of a T7 based linear amplification for genomic DNA. BMC Genomics, 2003. 4(1): p. 19.

197. Bernstein, B.E., et al., Methylation of histone H3 Lys 4 in coding regions of active genes. Proc Natl Acad Sci U S A, 2002. 99(13): p. 8695-700.

198. Reid, J.L., et al., Coordinate regulation of yeast ribosomal protein genes is associated with targeted recruitment of Esa1 histone acetylase. Mol Cell, 2000. 6(6): p. 1297-307.

199. Kurdistani, S.K., et al., Genome-wide binding map of the histone deacetylase Rpd3 in yeast. Nat Genet, 2002. 31(3): p. 248-54.

200. Robyr, D., et al., Microarray deacetylation maps determine genome-wide functions for yeast histone deacetylases. Cell, 2002. 109(4): p. 437-46.

201. Kurdistani, S.K., S. Tavazoie, and M. Grunstein, Mapping global histone acetylation patterns to gene expression. Cell, 2004. 117(6): p. 721-33.

202. Chen, L. and H. Zhao, Gene expression analysis reveals that histone deacetylation sites may serve as partitions of chromatin gene expression domains. BMC Genomics, 2005. 6(1): p. 44.

203. Schubeler, D., et al., The histone modification pattern of active genes revealed through genome-wide chromatin analysis of a higher eukaryote. Genes Dev, 2004. 18(11): p. 1263-71.

204. Ng, H.H., et al., Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity. Mol Cell, 2003. 11(3): p. 709-19.

205. Robert, F., et al., Global position and recruitment of HATs and HDACs in the yeast genome. Mol Cell, 2004. 16(2): p. 199-209.

206. Ng, H.H., et al., Genome-wide location and regulated recruitment of the RSC nucleosome-remodeling complex. Genes Dev, 2002. 16(7): p. 806-19.

207. Strahl, B.D., et al., Methylation of histone H3 at lysine 4 is highly conserved and correlates with transcriptionally active nuclei in Tetrahymena. Proc Natl Acad Sci U S A, 1999. 96(26): p. 14967-72.

208. Boggs, B.A., et al., Differentially methylated forms of histone H3 show unique association patterns with inactive human X chromosomes. Nat Genet, 2002. 30(1): p. 73-6.

209. Miao, F. and R. Natarajan, Mapping global histone methylation patterns in the coding regions of human genes. Mol Cell Biol, 2005. 25(11): p. 4650-61.

210. Brenowitz, M., et al., Quantitative DNase footprint titration: a method for studying protein-DNA interactions. Methods Enzymol, 1986. 130: p. 132-81.

211. Garner, M.M. and A. Revzin, A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: application to components of the Escherichia coli lactose operon regulatory system. Nucleic Acids Res, 1981. 9(13): p. 3047-60.

212. Kellis, M., et al., Sequencing and comparison of yeast species to identify genes and regulatory elements. Nature, 2003. 423(6937): p. 241-54.

213. Pennacchio, L.A. and E.M. Rubin, Genomic strategies to identify mammalian regulatory sequences. Nat Rev Genet, 2001. 2(2): p. 100-9.

214. Stormo, G.D., DNA binding sites: representation and discovery. Bioinformatics, 2000. 16(1): p. 16-23.

215. Tavazoie, S., et al., Systematic determination of genetic network architecture. Nat Genet, 1999. 22(3): p. 281-5.

216. Segal, E., et al., Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. Nat Genet, 2003. 34(2): p. 166-76.

217. 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.

218. Smith, A.D., et al., Mining ChIP-chip data for transcription factor and cofactor binding sites. Bioinformatics, 2005. 21 Suppl 1: p. i403-i412.

219. Harbison, C.T., et al., Transcriptional regulatory code of a eukaryotic genome. Nature, 2004. 431(7004): p. 99-104.

220. Liu, X., et al., DIP-chip: rapid and accurate determination of DNA-binding specificity. Genome Res, 2005. 15(3): p. 421-7.

221. Mukherjee, S., et al., Rapid analysis of the DNA-binding specificities of transcription factors with DNA microarrays. Nat Genet, 2004. 36(12): p. 1331-9.

222. Simon, I., et al., Serial regulation of transcriptional regulators in the yeast cell cycle. Cell, 2001. 106(6): p. 697-708.

223. Lee, T.I., et al., Transcriptional regulatory networks in Saccharomyces cerevisiae. Science, 2002. 298(5594): p. 799-804.

224. Boyer, L.A., et al., Core transcriptional regulatory circuitry in human embryonic stem cells. Cell, 2005. 122(6): p. 947-56.

225. Bar-Joseph, Z., et al., Computational discovery of gene modules and regulatory networks. Nat Biotechnol, 2003. 21(11): p. 1337-42.

226. Xu, X., L. Wang, and D. Ding, Learning module networks from genome-wide location and expression data. FEBS Lett, 2004. 578(3): p. 297-304.

227. Zhang, Z. and M. Gerstein, Reconstructing genetic networks in yeast. Nat Biotechnol, 2003. 21(11): p. 1295-7.

228. Weber, M., et al., Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. Nat Genet, 2005. 37(8): p. 853-62.

229. Weinmann, A.S., et al., Use of chromatin immunoprecipitation to clone novel E2F target promoters. Mol Cell Biol, 2001. 21(20): p. 6820-32.

230. Singer, B.S., et al., Libraries for genomic SELEX. Nucleic Acids Res, 1997. 25(4): p. 781-6.

231. Labhart, P., et al., Identification of target genes in breast cancer cells directly regulated by the SRC-3/AIB1 coactivator. Proc Natl Acad Sci U S A, 2005. 102(5): p. 1339-44.

232. Impey, S., et al., Defining the CREB regulon: a genome-wide analysis of transcription factor regulatory regions. Cell, 2004. 119(7): p. 1041-54.

233. Roh, T.Y., et al., High-resolution genome-wide mapping of histone modifications. Nat Biotechnol, 2004. 22(8): p. 1013-6.

234. Kim, J., et al., Mapping DNA-protein interactions in large genomes by sequence tag analysis of genomic enrichment. Nat Methods, 2005. 2(1): p. 47-53.

235. Roh, T.Y., S. Cuddapah, and K. Zhao, Active chromatin domains are defined by acetylation islands revealed by genome-wide mapping. Genes Dev, 2005. 19(5): p. 542-52.

236. Carvin, C.D., et al., Targeted cytosine methylation for in vivo detection of protein-DNA interactions. Proc Natl Acad Sci U S A, 2003. 100(13): p. 7743-8.

237. van Steensel, B., J. Delrow, and S. Henikoff, Chromatin profiling using targeted DNA adenine methyltransferase. Nat Genet, 2001. 27(3): p. 304-8.

238. Orian, A., et al., Genomic binding by the *Drosophila* Myc, Max, Mad/Mnt transcription factor network. Genes Dev, 2003. 17(9): p. 1101-14.

239. Greil, F., et al., Distinct HP1 and Su(var)3-9 complexes bind to sets of developmentally coexpressed genes depending on chromosomal location. Genes Dev, 2003. 17(22): p. 2825-38.

240. Bianchi-Frias, D., et al., Hairy transcriptional repression targets and cofactor recruitment in *Drosophila*. PLoS Biol, 2004. 2(7): p. E178.

241. Sun, L.V., et al., Protein-DNA interaction mapping using genomic tiling path microarrays in *Drosophila*. Proc Natl Acad Sci U S A, 2003. 100(16): p. 9428-33.

242. Holland, S., et al., Comparison of Dam tagging and chromatin immunoprecipitation as tools for the identification of the binding sites for S. pombe CENP-C. Chromosome Res, 2005. 13(1): p. 73-83.

243. Tenenbaum, S.A., et al., Identifying mRNA subsets in messenger ribonucleoprotein complexes by using cDNA arrays. Proc Natl Acad Sci U S A, 2000. 97(26): p. 14085-90.

244. Takizawa, P.A., et al., Plasma membrane compartmentalization in yeast by messenger RNA transport and a septin diffusion barrier. Science, 2000. 290(5490): p. 341-4.

245. Shepard, K.A., et al., Widespread cytoplasmic mRNA transport in yeast: identification of 22 bud-localized transcripts using DNA microarray analysis. Proc Natl Acad Sci U S A, 2003. 100(20): p. 11429-34.

246. Penalva, L.O., et al., RNA-binding proteins to assess gene expression states of co-cultivated cells in response to tumor cells. Mol Cancer, 2004. 3: p. 24.

247. Hieronymus, H. and P.A. Silver, Genome-wide analysis of RNA-protein interactions illustrates specificity of the mRNA export machinery. Nat Genet, 2003. 33(2): p. 155-61.

248. Inada, M. and C. Guthrie, Identification of Lhp1p-associated RNAs by microarray analysis in Saccharomyces cerevisiae reveals association with coding and noncoding RNAs. Proc Natl Acad Sci U S A, 2004. 101(2): p. 434-9.

249. Gerber, A.P., D. Herschlag, and P.O. Brown, Extensive association of functionally and cytotopically related mRNAs with Puf family RNA-binding proteins in yeast. PLoS Biol, 2004. 2(3): p. E79.

250. Duttagupta, R., et al., Global analysis of Pub1p targets reveals a coordinate control of gene expression through modulation of binding and stability. Mol Cell Biol, 2005. 25(13): p. 5499-513.

251. Law, A., et al., Direct cloning of DNA that interacts in vivo with a specific protein: application to RNA polymerase II and sites of pausing in *Drosophila*. Nucleic Acids Res, 1998. 26(4): p. 919-24.

252. Zhang, L., et al., Detecting DNA-binding of proteins in vivo by UV-crosslinking and immunoprecipitation. Biochem Biophys Res Commun, 2004. 322(3): p. 705-11.

253. Kurdistani, S.K. and M. Grunstein, In vivo protein-protein and protein-DNA crosslinking for genomewide binding microarray. Methods, 2003. 31(1): p. 90-5.

254. Verdine, G.L. and D.P. Norman, Covalent trapping of protein-DNA complexes. Annu Rev Biochem, 2003. 72: p. 337-66.

255. Nagesha, H.S., M. Yu, and L.F. Wang, Application of linker-ligation-PCR for construction of phage display epitope libraries. J Virol Methods, 1996. 60(2): p. 147-54.

256. Zou, N., et al., Random priming PCR strategy to amplify and clone trace amounts of DNA. Biotechniques, 2003. 35(4): p. 758-60, 762-5.

257. Baskaran, N., et al., Uniform amplification of a mixture of deoxyribonucleic acids with varying GC content. Genome Res, 1996. 6(7): p. 633-8.

258. Dean, F.B., et al., Comprehensive human genome amplification using multiple displacement amplification. Proc Natl Acad Sci U S A, 2002. 99(8): p. 5261-6.

259. Buckley, P.G., et al., A full-coverage, high-resolution human chromosome 22 genomic microarray for clinical and research applications. Hum Mol Genet, 2002. 11(25): p. 3221-9.

260. Pokholok, D.K., et al., Genome-wide map of nucleosome acetylation and methylation in yeast. Cell, 2005. 122(4): p. 517-27.

261. Nagy, P.L., et al., Genomewide demarcation of RNA polymerase II transcription units revealed by physical fractionation of chromatin. Proc Natl Acad Sci U S A, 2003. 100(11): p. 6364-9.

262. Harismendy, O., et al., Genome-wide location of yeast RNA polymerase III transcription machinery. Embo J, 2003. 22(18): p. 4738-47.

263. Wyrick, J.J. and R.A. Young, Deciphering gene expression regulatory networks. Curr Opin Genet Dev, 2002. 12(2): p. 130-6.

264. van Steensel, B., Mapping of genetic and epigenetic regulatory networks using microarrays. Nat Genet, 2005. 37 Suppl: p. S18-24.

265. Rual, J.F., et al., Towards a proteome-scale map of the human protein-protein interaction network. Nature, 2005.

266. Ge, H., et al., Correlation between transcriptome and interactome mapping data from Saccharomyces cerevisiae. Nat Genet, 2001. 29(4): p. 482-6.

267. Vaute, O., et al., Functional and physical interaction between the histone methyl transferase Suv39H1 and histone deacetylases. Nucleic Acids Res, 2002. 30(2): p. 475-81.

268. Chang, S. and C.S. Pikaard, Transcript profiling in Arabidopsis reveals complex responses to global inhibition of DNA methylation and histone deacetylation. J Biol Chem, 2005. 280(1): p. 796-804.

269. Lewis, A., et al., Imprinting on distal chromosome 7 in the placenta involves repressive histone methylation independent of DNA methylation. Nat Genet, 2004. 36(12): p. 1291-5.

270. Umlauf, D., et al., Imprinting along the Kcnq1 domain on mouse chromosome 7 involves repressive histone methylation and recruitment of Polycomb group complexes. Nat Genet, 2004. 36(12): p. 1296-300.

271. Gartler, S.M., et al., Normal histone modifications on the inactive X chromosome in ICF and Rett syndrome cells: implications for methyl-CpG binding proteins. BMC Biol, 2004. 2(1): p. 21.

272. Pietrobono, R., et al., Molecular dissection of the events leading to inactivation of the FMR1 gene. Hum Mol Genet, 2005. 14(2): p. 267-77.

273. Zhao, W., et al., The essential role of histone H3 Lys9 di-methylation and MeCP2 binding in MGMT silencing with poor DNA methylation of the promoter CpG island. J Biochem (Tokyo), 2005. 137(3): p. 431-40.

274. Zhang, Y., N. Fatima, and M.L. Dufau, Coordinated changes in DNA methylation and histone modifications regulate silencing/derepression of luteinizing hormone receptor gene transcription. Mol Cell Biol, 2005. 25(18): p. 7929-39.

275. Yan, P.S., et al., Dissecting complex epigenetic alterations in breast cancer using CpG island microarrays. Cancer Res, 2001. 61(23): p. 8375-80.

276. Costello, J.F., C. Plass, and W.K. Cavenee, Restriction landmark genome scanning. Methods Mol Biol, 2002. 200: p. 53-70.

277. Cheng, A.S., et al., Combinatorial Analysis of Transcription Factor Partners Reveals Recruitment of c-MYC to Estrogen Receptor-alpha Responsive Promoters. Mol Cell, 2006. 21(3): p. 393-404.

278. Smiraglia, D.J., et al., Excessive CpG island hypermethylation in cancer cell lines versus primary human malignancies. Hum Mol Genet, 2001. 10(13): p. 1413-9.

279. Chim, C.S., et al., Frequent DAP kinase but not p14 or Apaf-1 hypermethylation in B-cell chronic lymphocytic leukemia. J Hum Genet, 2006. 51(9): p. 832-8.

280. Cameron, E.E., et al., Synergy of demethylation and histone deacetylase inhibition in the re-expression of genes silenced in cancer. Nat Genet, 1999. 21(1): p. 103-7.

281. Sims, R.J., 3rd, K. Nishioka, and D. Reinberg, Histone lysine methylation: a signature for chromatin function. Trends Genet, 2003. 19(11): p. 629-39.

282. Bachman, K.E., et al., Histone modifications and silencing prior to DNA methylation of a tumor suppressor gene. Cancer Cell, 2003. 3(1): p. 89-95.

283. Xin, Z., et al., Role of histone methyltransferase G9a in CpG methylation of the Prader-Willi syndrome imprinting center. J Biol Chem, 2003. 278(17): p. 14996-5000.

284. Vakoc, C.R., et al., Histone H3 lysine 9 methylation and HP1gamma are associated with transcription elongation through mammalian chromatin. Mol Cell, 2005. 19(3): p. 381-91.

285. Tariq, M., et al., Erasure of CpG methylation in Arabidopsis alters patterns of histone H3 methylation in heterochromatin. Proc Natl Acad Sci U S A, 2003. 100(15): p. 8823-7.

286. Nguyen, C.T., et al., Histone H3-lysine 9 methylation is associated with aberrant gene silencing in cancer cells and is rapidly reversed by 5-aza-2'-deoxycytidine. Cancer Res, 2002. 62(22): p. 6456-61.

287. Kondo, Y., L. Shen, and J.P. Issa, Critical role of histone methylation in tumor suppressor gene silencing in colorectal cancer. Mol Cell Biol, 2003. 23(1): p. 206-15.

288. Wu, J., et al., ChIP-chip comes of age for genome-wide functional analysis. Cancer Res, 2006. 66(14): p. 6899-902.

289. Sridhar, V.V., et al., Control of DNA methylation and heterochromatic silencing by histone H2B deubiquitination. Nature, 2007. 447(7145): p. 735-8.

290. Sauer, F. and R. Tjian, Mechanisms of transcriptional activation: differences and similarities between yeast, *Drosophila*, and man. Curr Opin Genet Dev, 1997. 7(2): p. 176-81.

291. Maniatis, T., S. Goodbourn, and J.A. Fischer, Regulation of inducible and tissue-specific gene expression. Science, 1987. 236(4806): p. 1237-45.

292. Mitchell, P.J. and R. Tjian, Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. Science, 1989. 245(4916): p. 371-8.

293. Kapranov, P., A.T. Willingham, and T.R. Gingeras, Genome-wide transcription and the implications for genomic organization. Nat Rev Genet, 2007. 8(6): p. 413-23.

294. Thomas, M.C. and C.M. Chiang, The general transcription machinery and general cofactors. Crit Rev Biochem Mol Biol, 2006. 41(3): p. 105-78.

295. Maston, G.A., S.K. Evans, and M.R. Green, Transcriptional Regulatory Elements in the Human Genome. Annu Rev Genomics Hum Genet, 2006. 7: p. 29-59.

296. Lemon, B. and R. Tjian, Orchestrated response: a symphony of transcription factors for gene control. Genes Dev, 2000. 14(20): p. 2551-69.

297. Smale, S.T. and J.T. Kadonaga, The RNA polymerase II core promoter. Annu Rev Biochem, 2003. 72: p. 449-79.

298. Wray, G.A., et al., The evolution of transcriptional regulation in eukaryotes. Mol Biol Evol, 2003. 20(9): p. 1377-419.

299. Goldberg, M., Sequence analysis of *Drosophila* histone genes. 1979, Stanford University.

300. Corden, J., et al., Promoter sequences of eukaryotic protein-coding genes. Science, 1980. 209(4463): p. 1406-14.

301. Butler, J.E. and J.T. Kadonaga, The RNA polymerase II core promoter: a key component in the regulation of gene expression. Genes Dev, 2002. 16(20): p. 2583-92.

302. Burke, T.W. and J.T. Kadonaga, *Drosophila* TFIID binds to a conserved downstream basal promoter element that is present in many TATA-box-deficient promoters. Genes Dev, 1996. 10(6): p. 711-24.

303. Burke, T.W. and J.T. Kadonaga, The downstream core promoter element, DPE, is conserved from *Drosophila* to humans and is recognized by TAFII60 of *Drosophila*. Genes Dev, 1997. 11(22): p. 3020-31.

304. Jack, J., et al., Expression of the cut locus in the *Drosophila* wing margin is required for cell type specification and is regulated by a distant enhancer. Development, 1991. 113(3): p. 735-47.

305. Banerji, J., L. Olson, and W. Schaffner, A lymphocyte-specific cellular enhancer is located downstream of the joining region in immunoglobulin heavy chain genes. Cell, 1983. 33(3): p. 729-40.

306. Vilar, J.M. and L. Saiz, DNA looping in gene regulation: from the assembly of macromolecular complexes to the control of transcriptional noise. Curr Opin Genet Dev, 2005. 15(2): p. 136-44.

307. Burgess-Beusse, B., et al., The insulation of genes from external enhancers and silencing chromatin. Proc Natl Acad Sci U S A, 2002. 99 Suppl 4: p. 16433-7.

308. Recillas-Targa, F., et al., Position-effect protection and enhancer blocking by the chicken beta-globin insulator are separable activities. Proc Natl Acad Sci U S A, 2002. 99(10): p. 6883-8.

309. Gerasimova, T.I. and V.G. Corces, Boundary and insulator elements in chromosomes. Curr Opin Genet Dev, 1996. 6(2): p. 185-92.

310. Banerjee, S., et al., Igf2/H19 imprinting control region (ICR): an insulator or a position-dependent silencer? ScientificWorldJournal, 2001. 1: p. 218-24.

311. Ogbourne, S. and T.M. Antalis, Transcriptional control and the role of silencers in transcriptional regulation in eukaryotes. Biochem J, 1998. 331 (Pt 1): p. 1-14.

312. Gershenzon, N.I. and I.P. Ioshikhes, Synergy of human Pol II core promoter elements revealed by statistical sequence analysis. Bioinformatics, 2005. 21(8): p. 1295-300.

313. Suzuki, Y., et al., Identification and characterization of the potential promoter regions of 1031 kinds of human genes. Genome Res, 2001. 11(5): p. 677-84.

314. Sandelin, A., et al., Mammalian RNA polymerase II core promoters: insights from genome-wide studies. Nat Rev Genet, 2007. 8(6): p. 424-36.

315. Shiraki, T., et al., Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage. Proc Natl Acad Sci U S A, 2003. 100(26): p. 15776-81.

316. Kawaji, H., et al., Dynamic usage of transcription start sites within core promoters. Genome Biol, 2006. 7(12): p. R118.

317. Kimura, K., et al., Diversification of transcriptional modulation: large-scale identification and characterization of putative alternative promoters of human genes. Genome Res, 2006. 16(1): p. 55-65.

318. Muller, M., et al., One, two, three--p53, p63, p73 and chemosensitivity. Drug Resist Updat, 2006. 9(6): p. 288-306.

319. Landry, J.R., D.L. Mager, and B.T. Wilhelm, Complex controls: the role of alternative promoters in mammalian genomes. Trends Genet, 2003. 19(11): p. 640-8.

320. Prestridge, D.S., Predicting Pol II promoter sequences using transcription factor binding sites. J Mol Biol, 1995. 249(5): p. 923-32.

321. Pearce, S.T. and V.C. Jordan, The biological role of estrogen receptors alpha and beta in cancer. Crit Rev Oncol Hematol, 2004. 50(1): p. 3-22.

322. Kuiper, G.G., et al., Cloning of a novel receptor expressed in rat prostate and ovary. Proc Natl Acad Sci U S A, 1996. 93(12): p. 5925-30.

323. Mosselman, S., J. Polman, and R. Dijkema, ER beta: identification and characterization of a novel human estrogen receptor. FEBS Lett, 1996. 392(1): p. 49-53.

324. Tremblay, G.B., et al., Cloning, chromosomal localization, and functional analysis of the murine estrogen receptor beta. Mol Endocrinol, 1997. 11(3): p. 353-65.

325. Evans, R.M., The steroid and thyroid hormone receptor superfamily. Science, 1988. 240(4854): p. 889-95.

326. Klein-Hitpass, L., et al., An estrogen-responsive element derived from the 5' flanking region of the Xenopus vitellogenin A2 gene functions in transfected human cells. Cell, 1986. 46(7): p. 1053-61.

327. Chen, J.D. and R.M. Evans, A transcriptional co-repressor that interacts with nuclear hormone receptors. Nature, 1995. 377(6548): p. 454-7.

328. Horlein, A.J., et al., Ligand-independent repression by the thyroid hormone receptor mediated by a nuclear receptor co-repressor. Nature, 1995. 377(6548): p. 397-404.

329. Kushner, P.J., et al., Estrogen receptor pathways to AP-1. J Steroid Biochem Mol Biol, 2000. 74(5): p. 311-7.

330. Filardo, E.J., et al., Estrogen-induced activation of Erk-1 and Erk-2 requires the G protein-coupled receptor homolog, GPR30, and occurs via trans-activation of the epidermal growth factor receptor through release of HB-EGF. Mol Endocrinol, 2000. 14(10): p. 1649-60.

331. Kahlert, S., et al., Estrogen receptor alpha rapidly activates the IGF-1 receptor pathway. J Biol Chem, 2000. 275(24): p. 18447-53.

332. Migliaccio, A., et al., Tyrosine kinase/p21ras/MAP-kinase pathway activation by estradiol-receptor complex in MCF-7 cells. Embo J, 1996. 15(6): p. 1292-300.

333. McGovern, E.L. and W.E. Perrin, Treatment of Inoperable Carcinoma with Newer Chemotherapeutic Agents: Report of Three Cases. J Am Osteopath Assoc, 1963. 63: p. 330-3.

334. Kelsey, J.L. and M.D. Gammon, The epidemiology of breast cancer. CA Cancer J Clin, 1991. 41(3): p. 146-65.

335. Hulka, B.S., E.T. Liu, and R.A. Lininger, Steroid hormones and risk of breast cancer. Cancer, 1994. 74(3 Suppl): p. 1111-24.

336. Russo, J., L.K. Tay, and I.H. Russo, Differentiation of the mammary gland and susceptibility to carcinogenesis. Breast Cancer Res Treat, 1982. 2(1): p. 5-73.

337. Colditz, G.A., et al., The use of estrogens and progestins and the risk of breast cancer in postmenopausal women. N Engl J Med, 1995. 332(24): p. 1589-93.

338. Magnusson, C., et al., Breast-cancer risk following long-term oestrogen- and oestrogen-progestin-replacement therapy. Int J Cancer, 1999. 81(3): p. 339-44.

339. Clarke, R.B., et al., Dissociation between steroid receptor expression and cell proliferation in the human breast. Cancer Res, 1997. 57(22): p. 4987-91.

340. Johnston, S.R., et al., Modulation of Bcl-2 and Ki-67 expression in oestrogen receptor-positive human breast cancer by tamoxifen. Eur J Cancer, 1994. 30A(11): p. 1663-9.

341. Jordan, V.C. and S. Koerner, Tamoxifen (ICI 46,474) and the human carcinoma 8S oestrogen receptor. Eur J Cancer, 1975. 11(3): p. 205-6.

342. Jordan, V.C., S. Gapstur, and M. Morrow, Selective estrogen receptor modulation and reduction in risk of breast cancer, osteoporosis, and coronary heart disease. J Natl Cancer Inst, 2001. 93(19): p. 1449-57.

343. Ayoubi, T.A. and W.J. Van De Ven, Regulation of gene expression by alternative promoters. Faseb J, 1996. 10(4): p. 453-60.

344. Bajic, V.B., et al., Mice and men: their promoter properties. PLoS Genet, 2006. 2(4): p. e54.

345. Cooper, S.J., et al., Comprehensive analysis of transcriptional promoter structure and function in 1% of the human genome. Genome Res, 2006. 16(1): p. 1-10.

346. Bourdon, J.C., et al., p53 isoforms can regulate p53 transcriptional activity. Genes Dev, 2005. 19(18): p. 2122-37.

347. Marcu, K.B., S.A. Bossone, and A.J. Patel, myc function and regulation. Annu Rev Biochem, 1992. 61: p. 809-60.

348. Agarwal, V.R., et al., Use of alternative promoters to express the aromatase cytochrome P450 (CYP19) gene in breast adipose tissues of cancer-free and breast cancer patients. J Clin Endocrinol Metab, 1996. 81(11): p. 3843-9.

349. Orban, T.I. and E. Olah, Emerging roles of BRCA1 alternative splicing. Mol Pathol, 2003. 56(4): p. 191-7.

350. Li, C.Y., J. Zhu, and J.Y. Wang, Ectopic expression of p73alpha, but not p73beta, suppresses myogenic differentiation. J Biol Chem, 2005. 280(3): p. 2159-64.

351. Landry, J.R., et al., The Opitz syndrome gene Mid1 is transcribed from a human endogenous retroviral promoter. Mol Biol Evol, 2002. 19(11): p. 1934-42.

352. Podgorski, I. and B.F. Sloane, Cathepsin B and its role(s) in cancer progression. Biochem Soc Symp, 2003(70): p. 263-76.

353. Dehm, S.M. and K. Bonham, Regulation of alternative SRC promoter usage in HepG2 hepatocellular carcinoma cells. Gene, 2004. 337: p. 141-50.

354. Pampalakis, G., et al., Cloning and characterization of novel isoforms of the human kallikrein 6 gene. Biochem Biophys Res Commun, 2004. 320(1): p. 54-61.

355. Archey, W.B., et al., Methylation of CpGs as a determinant of transcriptional activation at alternative promoters for transforming growth factor-beta3. Cancer Res, 1999. 59(10): p. 2292-6.

356. Thierry-Mieg, D., et al., AceView: Identification and functional annotation of cDNA-supported genes in higher organisms. Unpublished.

357. Kim, N., S. Shin, and S. Lee, ECgene: genome-based EST clustering and gene modeling for alternative splicing. Genome Res, 2005. 15(4): p. 566-76.

358. Siepel, A. and D. Haussler, Computational identification of evolutionarily conserved exons. RECOMB '04, 2004.

359. Davuluri, R.V., I. Grosse, and M.Q. Zhang, Computational identification of promoters and first exons in the human genome. Nat Genet, 2001. 29(4): p. 412-7.

360. Bajic, V.B. and S.H. Seah, Dragon gene start finder: an advanced system for finding approximate locations of the start of gene transcriptional units. Genome Res, 2003. 13(8): p. 1923-9.

361. Sun, H., et al., MPromDb: an integrated resource for annotation and visualization of mammalian gene promoters and ChIP-chip experimental data. Nucleic Acids Res, 2006. 34(Database issue): p. D98-103.

362. Takeda, J., et al., H-DBAS: alternative splicing database of completely sequenced and manually annotated full-length cDNAs based on H-Invitational. Nucleic Acids Res, 2007. 35(Database issue): p. D104-9.

363. Crawford, G.E., et al., DNase-chip: a high-resolution method to identify DNase I hypersensitive sites using tiled microarrays. Nat Methods, 2006. 3(7): p. 503-9.

364. Hillier, L.W., et al., Generation and annotation of the DNA sequences of human chromosomes 2 and 4. Nature, 2005. 434(7034): p. 724-31.

365. Kodzius, R., et al., CAGE: cap analysis of gene expression. Nat Methods, 2006. 3(3): p. 211-22.

366. Frasor, J., et al., Profiling of estrogen up- and down-regulated gene expression in human breast cancer cells: insights into gene networks and pathways underlying estrogenic control of proliferation and cell phenotype. Endocrinology, 2003. 144(10): p. 4562-74.

367. Hsu, F., et al., The UCSC Known Genes. Bioinformatics, 2006. 22(9): p. 1036-46.

368. Kawaji, H., et al., CAGE Basic/Analysis Databases: the CAGE resource for comprehensive promoter analysis. Nucleic Acids Res, 2006. 34(Database issue): p. D632-6.

369. Jin, V.X., et al., Genome-wide analysis of core promoter elements from conserved human and mouse orthologous pairs. BMC Bioinformatics, 2006. 7: p. 114.

370. Saitou, N. and M. Nei, The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol, 1987. 4(4): p. 406-25.

371. Kent, W.J., BLAT--the BLAST-like alignment tool. Genome Res, 2002. 12(4): p. 656-64.

372. Fan, M., et al., Diverse gene expression and DNA methylation profiles correlate with differential adaptation of breast cancer cells to the antiestrogens tamoxifen and fulvestrant. Cancer Res, 2006. 66(24): p. 11954-66.

373. Khalili, A., et al., Gamma-Normal-Gamma Mixture Model for Detecting Differentially Methylated Loci in Three Breast Cancer Cell Lines. Cancer Informatics, 2007. 2: p. 43-54.

374. Carroll, J.S., et al., Genome-wide analysis of estrogen receptor binding sites. Nat Genet, 2006. 38(11): p. 1289-97.

375. Mattia, M., et al., p53-Dependent p21 mRNA elongation is impaired when DNA replication is stalled. Mol Cell Biol, 2007. 27(4): p. 1309-20.

376. Aiyar, S.E., et al., Attenuation of estrogen receptor alpha-mediated transcription through estrogen-stimulated recruitment of a negative elongation factor. Genes Dev, 2004. 18(17): p. 2134-46.

377. Aiyar, S.E., et al., Regulation of clustered gene expression by cofactor of BRCA1 (COBRA1) in breast cancer cells. Oncogene, 2007. 26(18): p. 2543-53.

378. Martens, J.A., L. Laprade, and F. Winston, Intergenic transcription is required to repress the Saccharomyces cerevisiae SER3 gene. Nature, 2004. 429(6991): p. 571-4.

379. Martianov, I., et al., Repression of the human dihydrofolate reductase gene by a non-coding interfering transcript. Nature, 2007. 445(7128): p. 666-70.

380. Blume, S.W., et al., The 5'-untranslated RNA of the human dhfr minor transcript alters transcription pre-initiation complex assembly at the major (core) promoter. J Cell Biochem, 2003. 88(1): p. 165-80.