Comprehensive Summaries of Uppsala Dissertations
from the Faculty of Science and Technology 972

Chromatin Insulators and CTCF: Architects of Epigenetic States during Development.

BY

RITUPARNA MUKHOPADHYAY

ACTA UNIVERSITATIS UPSALIENSIS
UPPSALA 2004

Dissertation presented at Uppsala University to be publicly examined in Lindahlsalen, EBC, Uppsala, Tuesday, May 18, 2004 at 13:30 for the degree of Doctor of Philosophy. The examination will be conducted in English.

Abstract

Mukhopadhyay, R. 2004. Chromatin Insulators and CTCF: Architects of Epigenetic States during Development.. Acta Universitatis Upsaliensis. *Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology* 972. 39 pp. Uppsala. ISBN 91-554-5952-8

A controlled and efficient coordination of gene expression is the key for normal development of an organism. In mammals, a subset of autosomal genes is expressed monoallelically depending on the sex of the transmitting parent, a phenomenon known as genomic imprinting.

The imprinted state of the *H19* and *Igf2* genes is controlled by a short stretch of sequences upstream of *H19* known as the imprinting control region (ICR). This region is differentially methylated and is responsible for the repression of the maternally inherited *Igf2* allele. It harbors hypersensitive sites on the unmethylated maternal allele and functions as an insulator that binds a chromatin insulator protein CTCF. Hence the *H19* ICR, which plays an important role in maintaining the imprinting status of *H19* and *Igf2*, was shown to lose the insulator property upon CpG methylation.

Another ICR in the *Kcnq1* locus regulates long-range repression of *p57Kip2* and *Kcnq1* on the paternal allele, and is located on the neighboring subdomain of the imprinted gene cluster containing *H19* and *Igf2*, on the distal end of mouse chromosome 7. Similarly to the *H19* ICR, the *Kcnq1* ICR appears to possess a unidirectional and methylation-sensitive chromatin insulator property in two different somatic cell types. Hence, methylation dependent insulator activity emerges as a common feature of imprinting control regions.

The protein CTCF is required for the interpretation and propagation of the differentially methylated status of the *H19* ICR. Work in this thesis shows that this feature applies genomewide. The mapping of CTCF target sites demonstrated not only a strong link between CTCF, formation of insulator complexes and maintaining methylation-free domains, but also a network of target sites that are involved in pivotal functions. The pattern of CTCF *in vivo* occupancy varies in a lineage-specific manner, although a small group of target sites show constitutive binding.

In conclusion, the work of this thesis shows that epigenetic marks play an important role in regulating the insulator property. The studies also confirm the importance of CTCF in maintaining methylation-free domains and its role in insulator function. Our study unravels a new range of target sites for CTCF involved in divergent functions and their developmental control.

Keywords: DNA methylation, CTCF, Chromatin insulators, Microarray

Rituparna Mukhopadhyay, Department of Evolutionary Biology, Department of Animal Development and Genetics, Norbyv. 18A, Uppsala University, SE-75236 Uppsala, Sweden

ISSN 1104-232X
ISBN 91-554-5952-8
urn:nbn:se:uu:diva-4241 (http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-4241)

To

My

Parents

List of Papers

I. Holmgren, C., Kanduri, C., Dell, G., Ward, A., **Mukhopadhyay, R**., Kanduri, M., Lobanenkov, V., and Ohlsson, R.: **CpG methylation regulates the *Igf2/H19* insulator.** *Curr Biol,* 11, 1128-30. (2001).

II. Kanduri, C., Fitzpatrick, G., **Mukhopadhyay, R.**, Kanduri, M., Lobanenkov, V., Higgins, M., and Ohlsson, R.: **A differentially methylated imprinting control region within the *Kcnq1* locus harbors a methylation-sensitive chromatin insulator.** *JBC*, 277, 18106-18110. (2002).

III. **Mukhopadhyay, R**., Yu, W., Whitehead, J., Xu, J., Lezcano, M., Pack, S., Kanduri, C., Kanduri, M., Ginjala, V., Vostrov, A., Quitschke, W., Chernukhin, I., Klenova, E., Lobanenkov, V., and Ohlsson, R.: **The binding sites for the chromatin insulator protein CTCF map to DNA methylation-free domains genomewide.** (*Submitted*).

IV. **Mukhopadhyay, R.,** Whitehead, J., Lezcano, M., Yu, W., Mattsson, A., and Ohlsson, R.: **Lineage-specific patterns of occupancy for target sites of the chromatin insulator protein CTCF.** (*Manuscript*).

CONTENTS

ABBREVIATIONS

Inr	Initiator
Scs/scs'	Special chromatin structures
CTCF	CCCTC binding factor
CpG	Cytosine paired with guanine dinucleotides
Dnmt	DNA methyltransferases
MeCP2	Methyl CpG binding protein
MBD	Methyl binding domain
PGC	Primordial germ cell
LTR	Long terminal repeats
LINE	Long interspersed nuclear elements
SINE	Short interspersed nuclear elements
HDAC	Histone deacetyltransferase
HAT	Histone acetyl transferase
HMT	Histone methyl transferase
HP1	Heterochromatic protein1
PWS	Prader-willi syndrome
AS	Angelman syndrome
Igf2	Insulin like growth factor 2
BWS	Beckwith-wiedemann syndrome
DMR	Differentially methylated region
ICR	Imprinting control region
ChIP	Chromatin immunoprecipitation
Kb	Kilo base
Bp	Base pair
MNase	Micrococcal nucleases
DNase	Deoxy ribo nucleases

INTRODUCTION

One of the beautiful happenings is the creation of life on earth followed by evolution to form multicellular organisms. Being a biologist, one of the fascinating facts is that the development of a multicellular organism begins with a single cell – the fertilized egg or zygote, which divides mitotically to give rise to all different cell types constituting the adult organism. Through the process of differentiation the zygote gives rise to hundreds of different cell types. All these different types of cells harbor the same genetic material but the morphological and functional differences between them is established by extremely complicated levels of gene regulation during different stages of development. Gene expression in different cell-lineages is influenced by a multitude of factors and is controlled at different levels. Regulation of gene expression not only involves genetic elements like promoters and enhancers, but also epigenetic factors like DNA methylation and chromatin organization play essential roles. Whether a particular gene is expressed or not and in which type of cell lineage, depends therefore on different epigenetic states. The work in this thesis involves investigating the epigenetic aspect of gene regulation at two different loci with a further extension to genome wide studies.

1. Transcriptional control of gene expression

The spatiotemporal pattern of gene expression is mainly controlled at the level of transcription during the process of cellular differentiation and morphogenesis. Transcriptional regulation of eukaryotic genes is a multistep process which involves the assembling of multi-protein complexes on gene regulatory regions. These regulatory regions primarily contain promoters, enhancers, silencers and insulators.

1.1. Promoters

Promoters are essential for gene expression and are located close to the transcriptional initiation site, usually within 50-100 bp. The core promoter consists of an A/T-rich sequence 25-30 bp upstream of the transcription start site with consensus TATAAA, called the TATA box. In addition, Inr (for Initiator) promoters have been found which are devoid of TATA box sequences

and transcription initiation is dependent on the consensus sequence YY-CAYYYYY (Y=T/C). Many TATA-containing genes also contains sequences having homology to Inr (87). Studies from Drosophila suggest that there could be two other classes of promoters, one in which TATA boxes are paired with Inr elements (88) and the other category in which Inr elements are combined with downstream promoter elements (DPE), which are located 30 bp downstream of the start site. However, many promoters, including a number of promoters within CpG islands, lack any of the above mentioned core elements (86). In the absence of activating *cis*-elements, the basal transcription machinery assembled on core promoters initiates very low level of transcription. Physiological level of gene expression is achieved by activating regulatory transcription factors which in turn interact with enhancers or upstream promoter elements.

1.2. Enhancers and silencers

In eukaryotic organisms, the spatiotemporal expression of genes is determined by regulatory sequences called enhancers and silencers. Enhancers bind factors which facilitates the initiation of transcription at the promoter. The enhancer-mediated activation operates in both a position and an orientation-independent manner. Enhancers have been described in both higher eukaryotes and bacteria where they function over a long distance but surprisingly, in case of yeast enhancer-like elements (UASs) work over only a short distance (68).

Enhancer-promoter communication over a large distance requires the use of special facilitating mechanisms. Several models have been proposed to explain enhancer action over a distance. In all of these models it is assumed that activators work through a protein-protein interaction along with a component of the transcription machinery. The activator can either interact with a stable target such that a protein already exists bound to the activated promoter, or the target is created through enhancer-induced recruitment of a protein to the promoter. The first model is the `looping model´ which suggests that DNA is looped so that activator factors bound to the enhancer could interact with the promoter region of the gene (8).

A second model is the `tracking model´ which suggests that the enhancer-bound complex actively scans the entire length of DNA in search of the promoter (37). The third model is the `facilitated tracking model´ which suggests that enhancer bound factors loop or track out several times until they reach the promoter (7). Finally, the fourth `linking model´ suggests that a chain of protein complexes extends along the chromatin fiber to the promoter, where the gene activation is mediated. (10).

Silencers are another type of regulatory element involved in negatively controlling the transcriptional activity of genes. Silencer elements are not well characterized, in contrast to enhancers, but they share features similar to

that of enhancers: They target to specific promoters and are able to function in an orientation-independent manner. Silencers are suggested to repress transcription by creating a repressive type of chromatin that spreads and silences gene expression, as exemplified by the lysozyme silencer (4). In yeast the silencing of the telomeric regions is caused by the formation of a multi-protein complex which recruits silent information regulator (SIR) proteins, leading to formation of repressed chromatin (25).

1.3. Insulators and boundary elements

Eukaryotic genomes are organized into domains containing individual genes and gene clusters, which facilitates distinct patterns of expression both during development and differentiation of cells. Regulatory elements like enhancers are present genomewide to control target genes in *cis* over considerable distances. Insulators or boundary elements restrict the expression patterns and avoid unwanted activation from *cis* regulatory elements, thereby maintaining proper expression domains.

Chromatin boundaries or insulators depict a class of regulatory elements that posses a common ability to protect genes from inappropriate signals emanating from their surrounding environment (102). Insulator elements have been described in many organisms including *Drosophila* and vertebrates and are perceived to function in two principally different manners.

The first category acts as **enhancer blockers** in which the insulator blocks the action of an enhancer on a promoter. Enhancer blocking can only occur if the insulator is situated between the enhancer and the promoter, and not if it is placed in some other position. Insulator elements were first discovered in *Drosophila* and were found to contain *DNase I*-hypersensitive sites located near boundaries of the heat-shock gene locus (*hsp70)*. These elements were called special chromatin structures (*scs/scs´*) and were shown to protect a transgene against chromosomal position effects when they flanked the gene. Similarly the *gypsy* transposable element can block the effect of an enhancer when located between the enhancer and a promoter (26, 73). The proteins Zw-5 and BEAF-32 binds to *scs* and *scs´* elements and facilitate insulator function, whereas Su (Hw), a zinc finger protein, binds to the *gypsy* insulator and mediates its function. The first vertebrate insulator element discovered was located at the 5' end of the chicken β-globin locus. A 250 bp DNA fragment mapped at this region harbors the insulator activity and it recruits the vertebrate protein CTCF, leading to formation of the insulator complex. Similarly, the insulator sequence upstream of the *H19* gene maintains proper expression of the neighboring genes in a parent-of-origin specific manner in both humans and mice (48). By extrapolation, the mammalian genome should be littered with strategically positioned insulator elements. This perception was borne out with our studies in which we have found more than 200 potential insulator sequences in the mouse genome.

Many of these sequences reside in the intergenic regions, suggesting their role in regulating the neighboring genes (paper III). One of the sequences, for example, is located on chromosome 10 between an upstream gene which codes for 60S acidic ribosomal protein and a downstream gene coding for transactivator interacting with Cbp/P300, emerges as having potential insulator function (Clone 6, paper III).

The second category includes **barriers**, which prevent the spreading of nearby condensed chromatin that might otherwise lead to silencing of the expression of neighboring genes. Some insulators are able to act as both enhancer-blockers and barriers whereas others, like in yeast, serve primarily as barriers. Yeast insulator elements have been found at the telomeres and at the mating-type loci, and they appear to separate active from silenced chromatin. However, the vertebrate insulator element at the chicken β-globin locus also acts as a barrier. It was found that within the genome, this element marks a boundary between the open, *DNase I*-sensitive and acetylated chromatin of the β-globin locus and the more condensed chromatin located outside the locus. Later studies showed that a second boundary or insulator element, also marked by hypersensitivity to *DNase I*, is present at the 3' end of the chicken β-globin gene. It was found to have both ability to protect against position effects as well as enhancer-blocking function (6, 12). We found in our studies that the chromatin insulator protein CTCF bound sequences belonging to heterochromatic region, thereby stressing the possibility that those sequences could act as barriers and maintain expression domains within the heterochromatic regions (paper III).

The mechanism behind the function of insulators and how they regulate gene expression is still not very clear. However, many mechanisms have been proposed. One of them is by looping, where the insulator elements loop to form a closed domain, and disrupts the enhancer-promoter communication. Another mechanism proposed was promoter decoy in which the insulator imitates the promoter by interacting with some or all protein components of the transcription complex, and hence fools the enhancer into interacting with the insulator instead of the promoter. Building up of higher order chromatin conformation at the insulator sequence site could also physically block the communication between enhancer and promoter. The bulky insulator complex formed could also loop out and interact with another insulator complex, thus preventing interaction between promoter and enhancer which in this scenario will be situated in different loops.

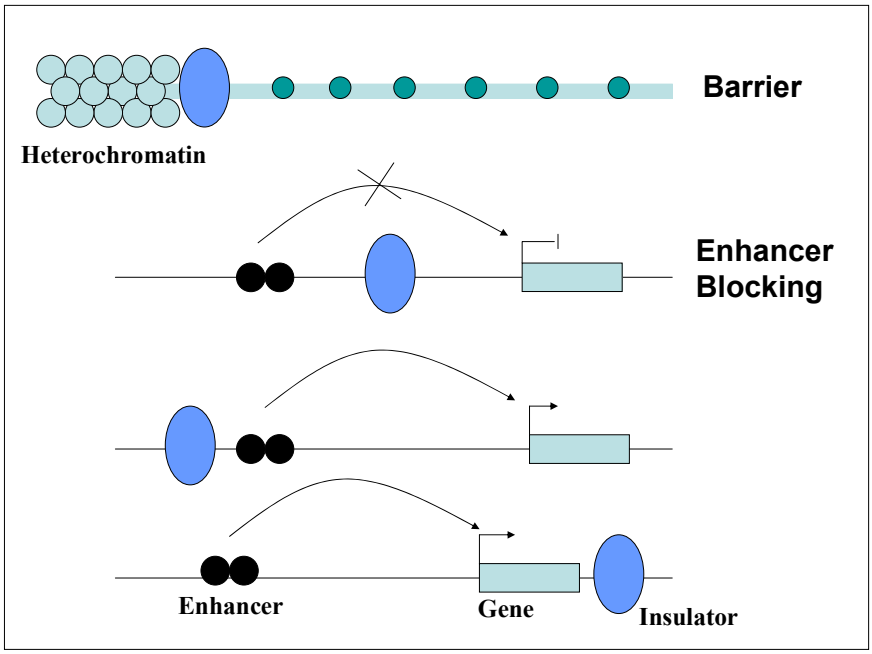

Figure 1. Two different kinds of insulator functions.

2. Epigenetic control of gene expression

Epigenetics can be defined as a heritable change in gene expression that are not accompanied by changes in DNA sequence (70). Epigenetic processes are essential for development and differentiation and such mechanisms are important for protecting cellular functions from being hijacked by viral genomes for their own benefit. It is a well documented fact that different levels of gene expression in different lineages depend on mechanisms that affect the packaging of genetic material without altering the DNA sequences. Such modifications can both enhance and repress transcription which most likely involves changing the chromatin structure as well.

2.1. DNA methylation

In mammalian genomes, methylation of cytosine residues at CpG dinucleotides is one of the major epigenetic modifications and has a profound effect on gene expression. 70-80% of the 5-positioned carbon of cytosine in 5'-CpG-3' dinucleotide is modified by a methyl group. Unmethylated CpG dinucleotides are mainly clustered in the CpG-rich sequences known as CpG islands. Genes involved in cancer show hypermethylation of CpG islands.

DNA methylation is catalyzed by three different DNA methyltransferases, all of which contain a highly conserved C-terminal catalytic domain and a variable N-terminal extension. DNA methyltransferase 1 (Dnmt1) is the maintenance enzyme that methylates hemi-methylated CpG dinucleotides in the nascent strand of DNA and restores the methylation pattern after DNA replication. It is ubiquitously expressed and has a C-terminal catalytic domain consisting of 500 residues and an 1100 residue N-terminal extension that regulates the substrate specificity and targeting of the methyltransferase to different nuclear and cellular sites. Dnmt1 knockout mice show genome wide demethylation and developmental arrest at E8.5 stage (60). In mature oocytes, an oocyte –specific form Dnmt1o, is reported to be present at high levels and is maintained until early zygotic stages. Gene targeting experiments in mice suggest that it plays an important role in maintaining methylation marks for maternally imprinted genes (40).

DNA methyltransferases 3a and 3b are *de novo* methyltransferases which are required for the initiation of *de novo* methylation and for establishment of new DNA methylation patterns during development. In addition to the catalytic domain, Dnmt 3a and 3b contain a proline-tryptophan-tryptophan-proline (PWWP) domain and a plant homeodomain (PHD) -like Zn finger domain. Dnmt3a knockout mice die at around 4 weeks of age; conversely Dnmt3b knockout mice show demethylation of minor satellite DNA, and embryonic lethality at E14.5- E18.5. Double knockout mice show failure to initiate *de novo* methylation after implantation and developmental arrest at E8.5. Another member of the Dnmt 3 family, Dnmt 3L, was identified by screening genome data bases. It was found to be highly expressed in undifferentiated ES cells and the expression was lost upon differentiation. Reports suggest that it is also expressed during gametogenesis (9). Studies have shown that it co-operates with the Dnmt 3 family to carry out *de novo* methylation of maternally imprinted genes during oogenesis and early mouse development (35, 67).

Dnmt 2 is another class of methyltransferases. Although it has the same sequence motifs that are conserved among all methyltransferases, its biological function is yet unknown. Homologues of Dnmt 2 have been identified in fission yeast, flies, amphibians and mammals.

DNA binding activities of methylation-sensitive transcriptional factors like E2F, ETS, and AP2 are hindered by DNA methylation, but methyl-CpG binding proteins are required along with this to abolish the binding of methylation-insensitive transcriptional factors like Sp1 and YY1. MeCP2, MBD1, MBD2, MBD3, MBD4 are the five members of the family and are characterized by having a conserved methylated DNA-binding domain (MBD). MeCP2 was found to be located on the X chromosome and was the first true member of the family of proteins that recognize methylated CpG. The gene is mutated in Rett syndrome patients. Knockout mice show complex neurological defects, increased anxiety-related behavior and seizures.

MBD1, which is the largest member of the family, MBD2 and MBD3 are found to be involved in methylation mediated transcriptional repression, whereas MBD4 has been associated with DNA-repair enzymes, and has a DNA glycosylase activity (3, 69).

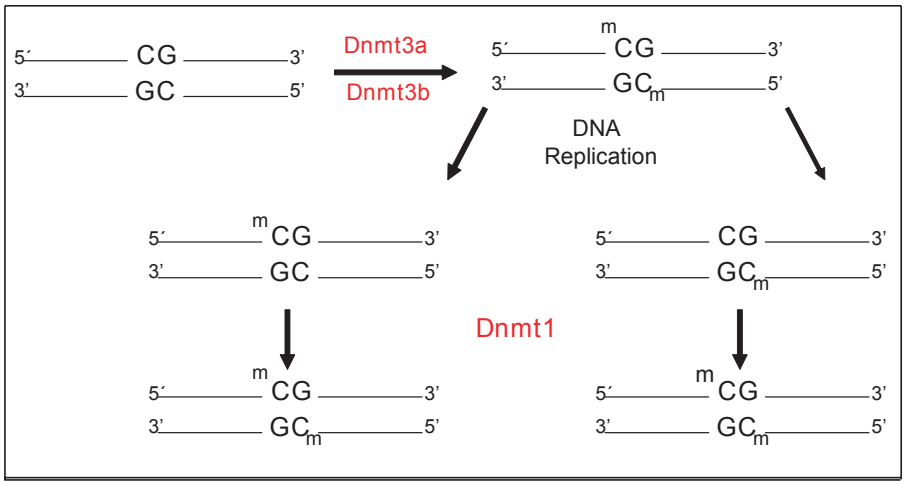

Figure 2. DNA methylation.

It has been observed that global demethylation and methylation events occur in primordial germ cells (PGCs). At around embryonic days 11.5-12.5, the PGCs undergo demethylation at imprinted loci which erases the existing parental imprinting marks. The imprinting marks are re-established during gametogenesis process: maternal-specific genomic imprints are re-established during growth and maturation of oocytes whereas paternal specific marks are established during the differentiation of the spermatocytes from the leptotene to pachytene stages of meiosis (35). Imprinting marks are re-established through *de novo* methylation by Dnmt3a and 3b. During spermatogenesis the methylation pattern is re-established along with histone hypoacetylation (36, 46). After fertilization, both maternal and paternal genomes again undergo rapid reprogramming. Both maternal and paternal chromosomes undergo demethylation in the zygote and by the blastocyst stage most of the methylation marks that were inherited from the gametes are erased, however the methylation marks on imprinted genes are protected, thus preserving the parental imprints. After implantation, embryonic DNA-methylation patterns are re-established through lineage-specific *de novo* methylation. Studies have shown that the establishment of embryonic methylation pattern requires both *de novo* and maintenance methyltransferase activities (41, 61, 72, 82).

2.2. DNA methylation in cancer

DNA methylation plays a crucial role in cancer development as it activates or represses certain cancer-related genes. Hypermethylation of CpG islands and promoter regions often leads to silencing of tumor suppressor genes. In a variety of primary tumors, the cell cycle related inhibitors are hypermethylated, thus assisting the cancer cells to escape senescence and continue proliferating. Studies have shown that in the case of breast and uterine tumors, aberrant methylation of the estrogen and progesterone receptor renders the cancer cells unresponsive to the steroid hormones. Along with the hypermethylation of the CpG islands, the genome of the cancer cell undergoes global hypomethylation. Hypomethylation can reactivate intragenomic parasitic DNA and repeats. The most obvious potential role for DNA hypomethylation would be in the activation of oncogenes. Finally, hypomethylation can affect imprinted genes. It has been proposed that disruption of the epigenetic machinery and their patterns is a major hallmark of human cancer (17, 18, 49, 90).

3. Heterochromatin and repeat elements

In eukaryotes, genomes are packaged into a dynamic, hierarchical chromatin structure which regulates critical cellular processes, such as transcription, DNA replication and repair. The fundamental repeating unit of chromatin is the nucleosome, which consists of 146 base pairs of DNA wrapped around an octamer of the core histone proteins H2A, H2B, H3, and H4. Linker histones of the H1 class are thought to bind at the exit and entry points of DNA as it winds around the nucleosome, establishing a higher level of organization. The carboxyl-terminal of the core histones binds to DNA whereas the amino terminal tails are freely available for different modifications which are discussed later in this section (14, 23).

To facilitate balanced regulation of gene expression and chromosome behavior, DNA is organized into structurally distinct domains. Two types of chromosomal domains are euchromatic domains, which are defined as more accessible and transcriptionally active portions of the genome and heterochromatic domains, which are condensed and transcriptionally silent. Heterochromatin was first recognized in late 1920s as heavily stained bodies in interphase nuclei by E. Heitz (31). Constitutive heterochromatin is commonly found at centromeric, peri-centromeric and telomeric regions and is mostly composed of long stretches of satellite repeats. They remain condensed and are packaged in permanently inactive form during interphase. Facultative heterochromatin refers to the euchromatic region which is condensed and packaged into a compact heterochromatin-like structure in a de-

velopmentally regulated manner as in the inactive X chromosome in female mammalian cells (16, 27).

3.1. Repeats

Evidence suggests that repetitive DNA elements play an important role in targeting heterochromatin formation in higher eukaryotes. Heterochromatic regions consist predominantly of repetitive DNA, including satellite sequences and transposable elements, which comprise a major portion of any complex genome. These regions don't harbor many genes hence they are typically gene-poor. Unique or single copy sequences are those which do not share homology with any other sequences in the same genome; on the other hand sequences that share homology with one or more other genomic regions are considered to be repeated or multi-copy. All organisms harbor substantial amounts of repetitive DNA, which in higher eukaryotes typically, makes up the bulk of entire genome. Repeats could be broadly classified into two categories: those which are clustered into tandem arrays and repeats that are dispersed throughout the genome (16).

Tandemly repeated DNA: The characteristic feature of the eukaryotic genome is the presence of tandemly repeated DNA which is also called satellite DNA. They constitute a major bulk at centromere and many other places in eukaryotic chromosomes. The repetitive DNA-rich constitution of the centromere could be one of the possible reason that it is replicated last in the chromosome (13). In addition, there are two other varieties of satellite DNA, namely minisatellites and microsatellites. Minisatellites form clusters up to 20 kb with repeat units up to 25 bp in length. A well documented example of a minisatellite is telomeric DNA, which in case of humans consists of hundreds of copies of the motif 5'-TTAGGG-3'. Microsatellite clusters are shorter, usually less than 150 bp and the repeat unit is 13 bp or less in length.

Interspersed repeats: Interspersed repeats are formed when a copy of the repeat unit appears elsewhere in the genome distant from the location of original sequence. This could occur very efficiently through transposition and interspersed repeats are known to harbor transpositional activity. RNA transposons or retroelements are characteristic features of eukaryotic genomes but have not been reported in prokaryotes. The retroelements, which have long terminal repeats at either end and play a role in the transposition process are described as **LTR elements.** These include **retroviruses**, **endogenous retroviruses (ERVs)** and **retrotransposons**. Retroviruses are viruses which have RNA as the genetic material and they infect many types of vertebrate. After infecting the cell, the RNA is converted to DNA with the help of reverse transcriptase and this DNA integrates into host genome. New viruses are formed by copying the integrated DNA into RNA and packaging them into the viral coats. Endogenous retroviruses are those retroviral genomes which have been incorporated into vertebrate chromosomes and have

lost their capacity to form viruses, although some of them could remain active for shorter periods (76, 77). Retrotransposons harbor sequences similar to ERVs but are characteristic of non-vertebrate eukaryotic genomes like plants, fungi etc. There are other retroelements which do not have LTRs and are called **retroposons**. These include the **LINE** and **SINE** families. LINEs or L1 repeats are long interspersed nuclear elements which contain a gene that resembles reverse-transcriptase and may probably be involved in the retrotransposition process. The mouse L1 DNA family is considered very old and homologous repetitive families have been found in a wide variety of organisms including protists and plants (66). The SINE family is referred to as short interspersed nuclear elements that are devoid of a reverse transcriptase gene but probably they borrow reverse transcriptase enzymes from other retroelements and can use it to transpose. Alu repeats in the human genome are a well suited example of the SINE family. In mouse, B1 and B2 with relatively short repeat units of 140 bp and 190 bp in length respectively, belong to SINE type.

3.2. Factors in heterochromatin assembly

Cellular mechanisms have evolved to sense repeated elements and neutralize them by different chromatin modifications. Histone modifications and DNA methylation direct the chromosomal silencing of DNA repeats in plants, animals and fungi. Studies performed on organisms ranging from yeast to mammals suggest strongly that histones and their posttranslational modifications play a pivotal role in the assembly of heterochromatin. Histone modifications include acetylation, methylation and ubiquitination of lysine (K) residues, phosphorylation of serine (S) and threonine (T) residues, methylation of arginine (R) residues, and ADP ribosylation at glutamate (E) residues. As mentioned earlier, the core histone proteins consist of a carboxyl-terminal domain, making up the nucleosome scaffold, and free amino-terminal, which has the potential sites for different modifications (29). These modifications could affect chromosome function through two distinct mechanisms. First, the electrostatic charge of the histone is altered in most of these modifications which in turn changes its structural properties or its binding to DNA. Second, new binding surfaces for protein recognizing modules are created which recruit complexes to their proper site of action. The bromodomains, which can recognize acetylated lysine, and the chromodomains, which can recognize methylated lysine, are examples of such recognition modules. Struhl and Allis proposed the `histone code´ hypothesis, according to which specific combinatorial sets of histone modification signals can dictate the recruitment of particular transacting factors to accomplish specific functions (89).

Histone acetylation is one of the better understood modifications. Hyperacetylated histones are mostly associated with activated genomic regions,

whereas deacetylation mainly results in repression and silencing (32, 33, 96). Two types of factors that maintain the acetylation status of histones are the HDACs (histone deacetylases) and the HATs (histone acetyl transferases) (53). Recent studies have shown that acetylation and deacetylation reactions occur continuously, generating a steady-state level of global or bulk histone acetylation (98). Another important modification is histone methylation, which appears to have multiple effects on chromatin function. Methylation of lysine residues on H3 (K4, K9, K27, K36 and K79) and H4 (K20 and K79) and of arginine residues on H3 (R2, R17 and R26) and H4 (R3) have been reported. Methylation of H3-K9 is associated with repression and silencing in many species, whereas methylation of H3-K4 is associated with active or permissive chromatin regions. Recent studies using antibodies that can distinguish between di- and tri-methylated H3-K4, revealed that trimethylation is specifically associated with active transcription whereas dimethylated K4 exists in both active and repressed genes (42, 105). Many different histone methyltransferases (HMTs) have been attributed for this modification, which are usually characterized by the SET domain protein motif. Several families of histone H3-K9 methyltransferases have been identified, such as Suv39h1, Suv39h2, G9a in mammalian cells, which indicates that different histone methyltransferses are involved in targeting different regions of the genome, thereby regulating the chromatin structure and gene expression. G9a methylates histone H3-K9 in the euchromatic region whereas Suv39h methylates specifically H3 in the pericentromeric heterochromatic region (60).

In addition, heterochromatic protein 1 (HP1) is found to be present in constitutive heterochromatic regions of many eukaryotes and thought to play important role in regulating heterochromatin-mediated silencing and chromosome structure. HP1 belong to a class of multifunctional chromatin-associated proteins, which was originally identified in *Drosophila melanogaster*, and predominantly localizes to pericentric heterochromatin. This protein is highly conserved and several organisms have more than one HP1 family member. In mammals there are three isoforms of HP1, namely HP1α, HP1β and HP1γ (27). In human and mouse cells, HP1α is found in centromeric regions, HP1β is distributed widely on the chromosome, and HP1γ localizes mostly to euchromatic regions. Similar kinds of distribution pattern of the three HP1 proteins have been reported in *Drosophila*. HP1 has three functional domains; an N-terminal chromodomain (CD) which provides a potential site for protein-protein interactions and has been reported to bind to histone H3 methylated at K9, a central hinge domain (HD) and a C-terminal chromoshadow domain (CSD) which is responsible for dimerization and interaction of HP1s with other chromosomal proteins (27, 85, 93). Recent biochemical and genetic experiments have led to a model of heterochromatin assembly based on histone modification patterns. HP1 binds specifically to H3-mK9 through its chromodomain and interacts with the spe-

cific HMT required for this modification through its chromoshadow domain (84).

4. Genomic imprinting

Genomic imprinting leads to expression of genes in parent-of-origin specific manner, which indicates that either of the parental alleles is inactivated during the course of development. The first evidence that came in support of imprinting was from the nuclear transfer experiments which showed that both maternal and paternal contributions are important for normal development of the embryo (79, 91). Further studies with conceptuses displaying uniparental disomy, where both chromosomes are inherited from single parent, revealed abnormal patterns of gene expression, leading to either under or over-expression of genes. The first imprinted gene was identified in 1991 and to date more than 70 imprinted genes have been identified (http://www.mgu.har.mrc.ac.uk/imprinting/imprin-ref.html) and many more predicted (http://fantom2.gsc.riken.jp/EICODB/).

Three important characteristics have been assigned to the imprinted genes, the first one being monoallelic expression, which indicates that the expression of a gene depends on the sex of the transmitting parent. The second one indicates that imprinted genes mostly reside within clusters of genes which are located at the evolutionary conserved domains. The third characteristic is that imprinted genes are generally found to be associated with a differentially methylated region in the parental alleles.

Disruption of imprinting is associated with many growth and development related disorders. Prader-Willi syndrome (PWS) and Angelman syndrome (AS) were the first definitive examples of imprinting mutations in humans and they involve the same chromosomal location of 15q11-13. PWS is associated with severe hypotonic hyperphagia leading to obesity, facial dysmorphology, hypogonadism and mental retardation, and AS is a neurobehavioral disorder that is characterized by mental retardation, seizures and ataxia. PWS is caused by loss of function of imprinted genes like *SNRPN*, *ZNF127* and *IPW* on the paternal allele of region 15q11-13 whereas AS results from lack of a functional maternal copy of the *UBE3A* at same location (38). Beckwith-Wiedemann syndrome (BWS) involves pre- and post-natal overgrowth, organomegaly, hemihyperplasia, omphalocele, ear lobe and renal abnormalities. It has been found to be associated with a diversity of genetic and epigenetic alterations on chromosome 11p15 in human. Hypermethylation of *H19* and biallelic expression of *Igf2* have been detected in one type of BWS whereas other types of BWS has been found to be associated with methylation and transcription errors at KvDMR and *KCNQ1OT1* (101).

Haig and Westoby proposed the parent-offspring conflict theory to account for the evolution of imprinting in mammals (103). According to this theory, genetic conflict exists between parents over the allocation of maternal resources to the embryo. The mother being the sole source of nutrients to the developing embryo tries to balance her resources by allocating among all her offspring. On the other hand, the father is only interested in ensuring the fitness of his offspring by extracting maximal maternal resources. Hence all the growth promoting genes would be paternally expressed whereas growth inhibiting genes would be maternally expressed. Maternal repression of the *Igf2* gene, which is a growth promoting gene, is consistent with this hypothesis and the paternal repression of the *Igf2r* gene, which is involved in transporting excess of *Igf2* to lysosomes for destruction, is maternally expressed. Imprinting is expected in viviparous species where there is contribution of maternal resources to the embryo and the potential for polyandry, whereas such kind of pressure for maternal resources is lacking in case of oviparous animals. Imprinting has been observed in marsupials and eutherian mammals but is lacking in monotremes and birds (44, 74, 103).

4.1. The imprinted gene cluster

One of the most intensively studied imprinted gene clusters resides within an 800 kb region on the distal end of mouse chromosome 7 and the orthologous region on human chromosome 11p15.5. It contains at least nine imprinted genes, including the maternally expressed *Imptl, Ipl*, *Cdkn1c* ($p57^{kip2}$), *Kcnq1 (Kvlqt1)*, *Ascl2 (Mash2)* and *H19* genes and the paternally expressed *Lit1*, *Igf2* and *Ins2* genes. The imprinted cluster is currently believed to be divided into two sub-domains: one of them contains the *H19*, *Igf2* and *Ins2* genes and the second sub-domain contains the *Ascl2 (Mash2)*, *Tssc6*, *Tssc4*, *Kcnq1 (Kvlqt1)*, *Kcnq10t1 (Lit1)*, *Cdkn1c* ($p57^{kip2}$), *Tssc3, Cd81* and *Nap2* genes (1, 47, 78).

4.2. *H19/Igf2* locus

The *H19/Igf2* locus is the best studied example among the imprinted genes. *H19* encodes an untranslated RNA (92) whereas *Igf2* encodes a fetal growth factor. These two genes are situated 90 kb apart and share a common set of endodermal and mesodermal enhancers 3' of *H19*. Deletion of endodermal enhancers results in loss of expression of *H19* or *Igf2* following maternal or paternal inheritance (59). A differentially methylated region (DMR) which is also referred to as the imprinting control region (ICR) has been found approximately 2 kb upstream of the *H19* promoter. Deletion of the ICR in knockout experiments shows reactivation of *H19* and loss of *Igf2* expression upon paternal transmission and a reciprocal effect upon maternal transmission, thus leading to loss of imprinting for both *Igf2* and *H19* (94, 95). This

region is unmethylated on the maternal allele whereas in case of paternal allele it is heavily methylated and methylation spreads up to the *H19* promoter, leading to silencing of this gene (45). Studies from our lab and two other labs have shown that on the maternal allele this region is unmethylated, harbors nuclease hypersensitive sites, and is important for silencing of *Igf2* (5, 34, 48). It has been shown that the CCCTC binding factor (CTCF), a zinc finger protein, binds to several sites on the unmethylated ICR (5, 49, 71), which is essential for enhancer blocking function and sets up the insulator complex. Studies from our lab showed that mutation of CTCF target sites abolishes the binding of CTCF and hence disrupts the insulator activity (49). First experimental evidence that the insulator function is methylation-sensitive was provided by our work based on transfection experiments with an episomal vector (paper I).

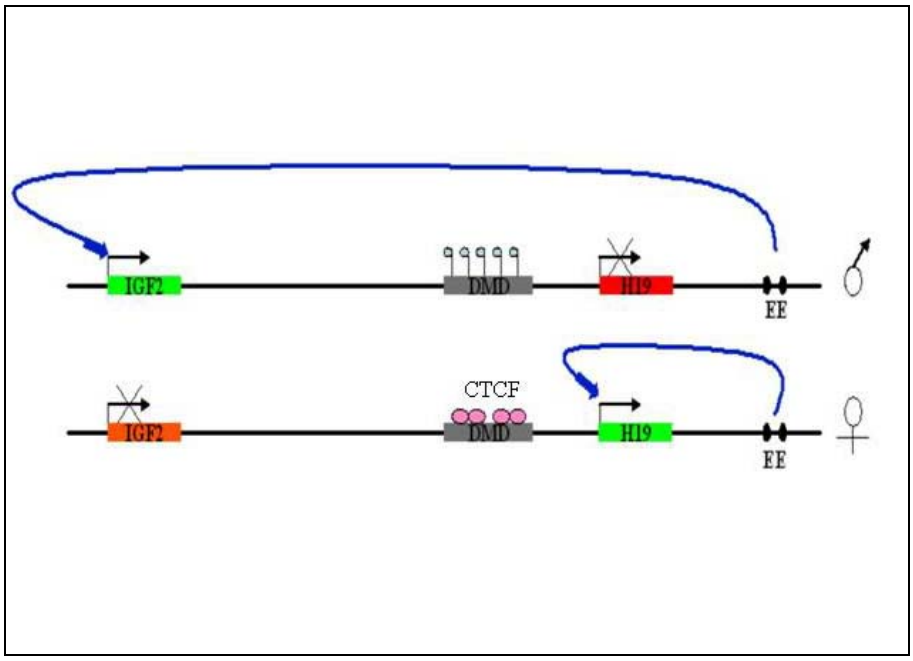

Figure 3. The *H19/IGF2* locus. The DMD or ICR region is methylated on the paternal allele and unmethylated on the maternal allele. *IGF2* is expressed from the paternal allele and *H19* from the maternal allele.

Another independent observation showed that insulator activity was dependent on its distance from *H19* promoter (28) suggesting that its function is distance-dependent. Recently it has been shown by generating a knock-in mouse model that the CTCF-*H19* complex is vital for monoallelic expression of *Igf2* and plays a crucial role in maintaining a methylation-free domain at the ICR (48, 75, 83).

4.3. *Kcnq1* locus

Kcnq1, previously named *Kvlqt1*, was first identified from the linkage studies while looking for the cause of cardiac arrhythmia. The *Kcnq1* gene is responsible for long QT syndrome (LQTS) and also found to be associated with deafness (81). It encodes a protein which is co-assembled with the product of the *Kcne1 (MinK)* gene in the heart to form a cardiac-delayed rectifier-like K+ current. During the early stages of development *Kcnq1* is expressed from the maternal allele, but the paternal allele also becomes active in late juveniles and adults, leading to complete biallelism. It has been shown to be expressed monoallelically in fetal lung and kidney but biallelically in the case of fetal heart (30).

An imprinting control region (ICR), referred as KvDMR, was detected within intron 10 of the *Kcnq1* gene. This region is methylated on the maternal allele, but remains unmethylated on the paternal allele and found to be associated with a large non-coding RNA that is transcribed in an opposite orientation to *Kcnq1* in both humans and mice (65). KvDMR methylation is associated with active transcription of *Kcnq1*, *Cdkn1c* and other neighboring genes in the case of the maternal allele, while in the paternal allele it is unmethylated and leads to active transcription of antisense transcript *Lit1*, whereas all other neighboring genes are silenced. Knockout experiments in which KvDMR is deleted show reactivation of genes both distal and proximal to KvDMR during paternal transmission, whereas no effect was seen upon maternal transmission (24). As mentioned earlier, BWS is associated with disruption of imprinting at the *Kcnq1* locus. Hence, it could be explained that in BWS there is loss of methylation at the maternal allele of KvDMR which leads to its activation with the subsequent silencing of maternally expressed genes like *Cdkn1c*. In our studies, by using transfection experiments, we have found that KvDMR behaves as insulator in two different cell-lines, namely Hep3B and Jurkat. Further, we found that this insulator function was methylation-sensitive. Apart from the *H19/Igf2* locus this was second evidence which showed that insulator activity of the ICR in an imprinted domain acts in a methylation dependent manner (paper II).

4.4. The CCCTC-binding factor, CTCF

The transcriptional processes are mediated by multifunctional proteins, many of which bind to the nucleic acid (55). The versatile zinc finger proteins, which are multifunctional in nature bind to DNA, and also mediate protein-protein interactions which are also linked to human development and diseases including cancer. Some examples of such types of proteins are CTCF, ZNF74, YY1, WT1, GL1 and many other proteins (56).

The transcriptional regulator CTCF contains a central 11 zinc finger (ZF) DNA binding domain that is flanked by 267 amino acids on the N-terminal

side and 150 amino acids on the C-terminal side (11, 100). It was discovered for its ability to bind to an unusually long 50 bp GC-rich sequence containing the core repeat sequence CCCTC within the chicken *c-myc* promoter (63). This 727 amino acid CCCTC- binding factor (CTCF) has extremely divergent binding sites. The molecular mass of full-length CTCF is 82 kDa, but due to the presence of sequences in the N- and C- terminus, it migrates at approximately 130 kDa on SDS-PAGE gels (52). CTCF appears to be present in multiple forms as has been observed in chicken, having molecular masses of 55, 70, 73, 80, and 97 kDa (51). CTCF is an ubiquitously expressed (63), and exceptionally highly conserved multifunctional protein that binds through combinatorial use of its 11 ZFs to target sites with remarkable sequence variation (19).

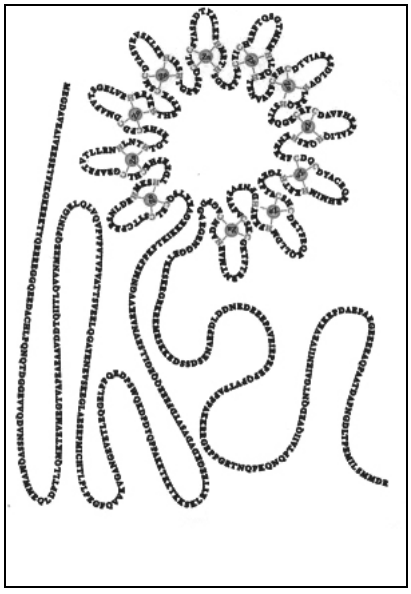

Figure 4. The multifunctional CTCF protein.

The chicken CTCF gene contains 8 exons; exon1 and exons 3 to 7 are relatively smaller in size, each being less than 500 bp, whereas exons 2 and 8 are relatively large, and 797 and 1569 bp respectively. Exon 2 harbors the translational start site and encodes most of the amino-terminal domain upstream of the first zinc-finger. The 11 ZFs are encoded by exons 3 to 6, with several individual fingers shared between two neighboring exons. The AT-rich DNA binding domain positioned downstream of the 11^{th} zinc finger are encoded by exons 6 and 7 (50). The organization of human and mouse CTCF are essentially identical. The mammalian CTCF gene, as compared to the chicken CTCF, contains five evolutionarily new introns that have high density of Alu and LINE family repeats(71).

Many important binding sites for CTCF are known to date. In addition, we have identified more than 200 CTCF target sites which include intronic, exonic and intergenic regions in the mouse genome. Interestingly we have found that CTCF may be associated with different functions as the target sites map to loci which include genes involved in metabolism, neurogenesis, apoptosis and signaling pathways (paper III). A report has shown that CTCF binds to microsatellite repeats (2). In our studies, we found that CTCF is associated with different types of repeats like LTRs, LINEs and SINEs. Another interesting observation was that CTCF co-localizes with HP1 in the nuclear compartment, supported by detection of many sequences which probably belongs to heterochromatic regions of mouse genome as they remain unidentified in the mouse databases (paper III).

CTCF, a putative tumor suppressor

Human CTCF is encoded by a single-copy gene localized at chromosome locus 16q22.1, a region frequently deleted in several different human malignancies (57). This region is found to be frequently deleted in both sporadic breast tumors (15) and prostate cancers (58), which suggests that CTCF may be a common candidate for a tumor suppressor gene at this region (20). Further, rearrangements of the CTCF exons encoding the zinc finger DNA binding domain was displayed in a fresh breast tumor sample (20). Subsequent studies have identified four mutations, all of which were missense mutations in the zinc finger DNA binding domain of CTCF. These mutations caused a selective loss of DNA-binding ability in the mutant CTCF (21). It has been observed that many of the CTCF-regulated genes are frequently deregulated in human tumors. The hypermethylation at the DM1 locus in congenital myotonic dystrophy (DM) abrogated the CTCF binding, leading to disruption of CTCF-mediated insulator activity (22). Further, we have also found CTCF target sites to be associated with cancer-related loci as exemplified by one of the target sequence that resides in the intronic region of a gene which encodes for a putative prostate cancer suppressor located on chromosome 8 (clone 294, paper III).

Repression and activation by CTCF

CTCF has been found to be involved in transcriptional repression at the chicken lysozyme silencer and *c-myc* genes. The transcriptional repression is established by CTCF directly interacting with promoter elements and upstream silencer elements of the genes. It is suggested that CTCF recruits histone deacetylases and hence mediates silencing (64). CTCF harbors strong transcriptional silencing domains, one of which is located in the zinc finger region. It has been reported that the placement and sequence of the CTCF binding sites at chicken and human *c-myc* genes is quite divergent (19). CTCF has also been implicated in transcriptional activation of the amyloid β-protein precursor (APP) which is cleaved to produce the amyloid β-protein (99).

The chromatin insulator protein CTCF

As mentioned earlier insulators are DNA elements that prevent the activation of promoters by an inappropriate enhancer and/or stop the spread of condensed chromatin. CTCF sites are present in all of the vertebrate enhancer-blocking elements examined so far, notably the chicken *β-globin* insulator. As mentioned in the beginning we identified around 200 sequences in the mouse genome which harbor insulator activity through the binding of CTCF (paper III). CTCF plays an important role in the regulation of the *H19* and *Igf2* gene expression and the insulator property is methylation-sensitive. As mentioned earlier, our studies showed that the insulator function is disrupted when the ICR in the *H19/Igf2* locus is methylated (paper I). Further on the genomewide level when we cross referenced our CTCF target site library with MBD2, which binds to clusters of CpGs, we found that CTCF maintains methylation-free domains. It could be possible that this property of CTCF to set up the insulator function which is methylation sensitive could be universal (paper III). Although involvement of CTCF in insulator function is well established, the underlying mechanism remains unclear. Many models have been proposed to explain the insulator function. Recent studies from the Felsenfeld lab show that CTCF forms a complex with nucleophosmin and localizes at the nucleolar surface, suggesting that CTCF tethers its target sites to create separate loop domain structures, thereby abrogating the enhancer-promoter communication (104). Another interesting observation from our lab is that CTCF binds both constitutively and in a lineage-specific manner to its target sites. Lineage-specific occupancy could depend on the epigenetic marks or availability of different protein partners in different tissues, as CTCF is also localized to non-CTCF target site through protein-protein interactions (paper IV).

5. An overview of the microarray approach

The interference of transcription factor binding by the chromatin organization to the availability of their binding sites *in vivo*, is poorly understood (97). Such curiosity compelled researchers to develop methods which would enable a closer look into this kind of complicated regulation. The approaches finally generated the development of genome-wide analysis of *in vivo* transcription factor binding by using the "chromatin immunoprecipitation (ChIP) on a chip" technique. DNA prepared from chromatin immunoprecipitation against a transcription factor of interest is analyzed by hybridizing to a microarray containing DNA sequences from concerned organisms. The first effort was made in year 2000 by Richard. A. Young's group, in which they used this method to monitor the binding of transcription activators in yeast. They prepared ChIP DNA for the transcriptional activator Gal4, followed by amplification using ligation mediated PCR, and hybridized to a DNA mi-

croarray containing all yeast intergenic sequences. It was found that of the 1286 Gal4 "consensus" binding sites in the *S. cerevisiae* genome, only 10 were bound by Gal4 *in vivo* (80). This was followed by two subsequent studies from Patrick O. Brown's group, where they determined the distribution of repressor-activator protein 1 (Rap1), heterodimeric transcription factors SBF (Swi4p/Swi6p) and MBF (Swi6p/Mbp1p). It was reported that Rap1 recognizes both coding and intergenic DNA sequences, but that binding of Rap1 to the genome was highly specific to intergenic regions with the potential to act as promoters (43, 62). This approach was also used in mapping GATA-1 binding throughout the β-globin locus (39). As confidence was gained in this approach, researchers used this method to determine target sites for different factors involved in transcription. In 2002, Michael Grunstein described the genome-wide distribution of histone deacetylase (54). Hence, this method is already established to determine the *in vivo* occupancy of factors, which provides vital information as to the complicated network involved in transcriptional regulation.

We used the "ChIP on chip" approach to study the *in vivo* occupancy of the chromatin insulator protein CTCF. We took a different approach, by using the ChIP DNA itself to make the microarray. We generated a CTCF target site library using an antibody against the C-terminal domain of CTCF and purifying the chromatin immunoprecipitated DNA. This DNA was ligated to linkers, PCR amplified and cloned into pGEM vector to generate the plasmid library. As a result of this, we obtained clones which were either intergenic or situated within intron and exon of a gene. Random primers were used to PCR amplify the ChIP DNA to generate probes for hybridization. By using this approach, we were able to map target sites involved in multiple cellular functions including metabolism, neurogenesis, and apoptosis and signaling, and were able to cross-reference them with DNA methylation states in order to study epigenetic control of the CTCF-based network throughout the genome (paper III). Later studies were done to determine the developmental regulation of CTCF target site occupancy, (paper IV) using the same approach.

AIMS OF THE PRESENT STUDY

The preliminary work of this thesis was aimed at investigating the epigenetic aspects in controlling imprinting at the *H19*/*Igf2* locus. Studies were conducted on a short stretch of sequence upstream of the *H19* gene known as the imprinting control region (ICR). Subsequent studies were carried out for another imprinted region, namely *Kcnq1,* where the imprinting is epigenetically regulated by the *Kcnq1* ICR. Further expanding the study to the genomewide level unravels the importance of CTCF, a multifunctional protein, in setting up the insulator complex and maintaining methylation-free domains. Finally our study depicts the lineage-specific variation in CTCF binding which is developmentally regulated. The following questions arose during the course of our investigations:

- Is the insulator property of the *H19* ICR methylation-sensitive?
- Is the insulator property of the *Kcnq1* ICR similar to that of the *H19* ICR?
- What are the other CTCF target sites in the mouse genome?
- Is CTCF involved in setting up insulator complexes genomewide?
- Does CTCF maintain methylation-free domains throughout the genome?
- Does CTCF occupancy differ in a lineage-specific manner?

RESULTS AND DISCUSSIONS

Paper I

The *H19* ICR in the *H19*/*Igf2* locus was shown to play an important role in setting up the insulator complex (48). The ICR is methylated on the paternal allele and unmethylated on the maternal allele. On the maternal unmethylated allele a protein called CTCF binds and forms the insulator complex (5). Point mutations of CTCF target sites within the *H19* ICR abrogates its insulator activity (49). Since the *H19* ICR is differentially methylated we wanted to find out whether the insulator activity was methylation-sensitive.

The *H19* ICR insulator activity is methylation-sensitive

In order to address this issue we *in vitro* methylated and mock-methylated the *H19* ICR, followed by ligation into the pREP vector between the *H19* reporter gene and SV40 enhancer (as shown in Figure 5).

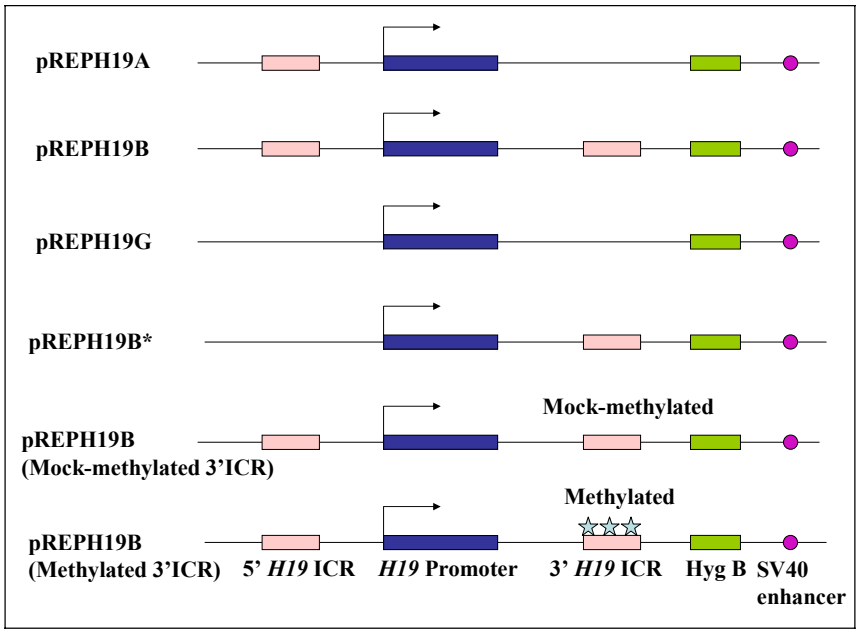

Figure 5. Schematic map of pREP constructs.

The ligated mix was purified and transfected into the Hep3B cell line followed by selection of single cell clones. Genotyping was done to select the correct construct for both methylated and mock-methylated clones using southern hybridization. Methylation status was confirmed by digestion with the methylation sensitive *Hha I* enzyme followed by southern analysis. Total RNA was extracted from the cells and was subjected to RNase protection assay. The assay showed that the expression of *H19* was lowered to the basal level in case of the mock-methylated clone, but the methylated clones showed higher levels of *H19* expression, comparable to the pREP construct with no insulator activity (Figure 3, paper I).

Chromatin conformation of *H19* ICR

Nuclei from both methylated and mock-methylated clones were prepared and subjected to *MNase* and *DNase* treatment. The mock-methylated clones showed a digestion pattern comparable to that of mouse fetal liver. Methylated clones however did not show any hypersensitive sites, suggesting that protein-DNA interactions are disrupted when the ICR is methylated.

Our studies showed that the insulator function is methylation-sensitive and the formation of hypersensitive sites is also methylation-dependent, further suggesting that the binding of CTCF, as this is a part of insulator complex, is also methylation-sensitive. These observations assist in explaining the reason for biallelic expression of *Igf2* in some human cancers.

Paper II

Our previous studies have shown that the imprinting control region at the *H19/IGF2* locus is involved in the formation of a chromatin insulator complex in a methylation-dependent manner, thereby controlling the monoallelic, parent-of-origin dependent expression of the*Igf2* and *H19* genes. In order to examine if this was a characteristic feature for imprinting control regions, we extended our studies of the *Kcnq1* ICR. *Kcnq1* is located on another subdomain of the imprinted gene cluster, residing at the distal end of mouse chromosome 7 and orthologous region of human chromosome 11p15.5, which is associated with Beckwith-Wiedemann syndrome (BWS). The *Kcnq1* ICR acts as a regulatory region in this domain and has been found to reside within intron 10 of *Kcnq1* gene. The region is methylated on the maternal allele, whereas on paternal allele it is unmethylated and associated with the transcription of a non-coding RNA in the opposite orientation to the *Kcnq1* gene. We investigated the characteristics of the *Kcnq1* ICR using episomal based assays.

The *Kcnq1* ICR acts as chromatin insulator in two different cell lines

We exploited two different insulator assay systems to test the possibility whether the *Kcnq1* ICR possesses a chromatin insulator property. The first system was the integrated chromatin insulator assay, whereby the fragment

to be tested was incorporated between promoter vδ1 and enhancer Eδ4 in the E-p-neo-scs vector (as shown in Fig 6). The *scs* (special chromatin structure) is a known insulator element which is active in mammalian cells. It was included in the vector in order to restrict the activation of vδ1 promoter by adjacent enhancer element. This was followed by transfecting the linearized construct into the human T cell leukemia Jurkat cell line and allowing it to grow to obtain drug-resistant colonies. The number of neo-resistant colonies obtained in p-neo-scs, which is a construct without enhancer, is taken as baseline for comparison. The construct with insulator activity should give a similar number of colonies as in p-neo-scs, which gives the basal activity of promoter in absence of enhancer. The results obtained by comparing neo-resistant colonies formed by different constructs show that the *Kcnq1* ICR acts as a unidirectional chromatin insulator (Figure 1b, paper II).

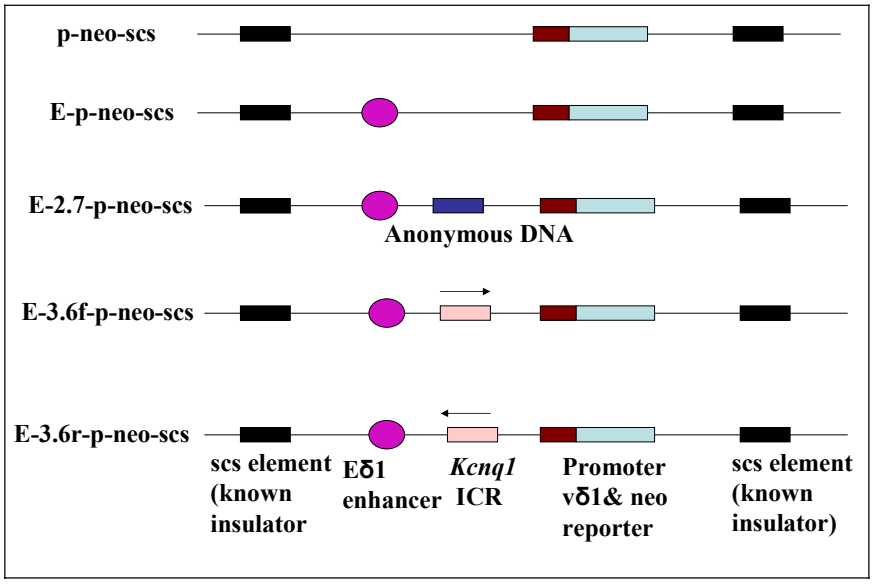

Figure 6. Schematic map of neo-scs constructs.

The second system was the episomal based insulator assay, in which the *Kcnq1* ICR was inserted between the SV40 enhancer and the mouse *H19* reporter gene, followed by transfecting the constructs into the Hep3B cell line. Total RNA was extracted and an RNase protection assay was performed. The assay showed that the expression of the *H19* reporter gene was lowered when the *Kcnq1* ICR was placed between SV40 enhancer and *H19* reporter gene and was comparable to the expression of *H19* reporter gene

with the pREP*H19*B construct (with *H19* ICR).The expression of the *H19* reporter gene remain unaffected for other constructs hence confirming that *Kcnq1* acts as unidirectional insulator (Figure 1c, paper II).

The insulator activity of the *Kcnq1* ICR is methylation dependent

We generated *in vitro* methylated and mock-methylated *Kcnq1* ICR fragments and ligated them to the pREP episomal vector (Figure 7). The ligation mix was transfected into the Hep3B cell line, followed by the selection of single cell clones. The inserts of those clones were analyzed for correct episomal constructs using southern hybridization. The methylation status of the ICR was tested by *Hha I* digestion followed by southern analysis. Total RNA was extracted and upon performing an RNase protection assay we found that unmethylated or mock-methylated clones harbor insulator activity, whereas methylated clones show loss of insulator function to various degrees (Figure 3, paper II).

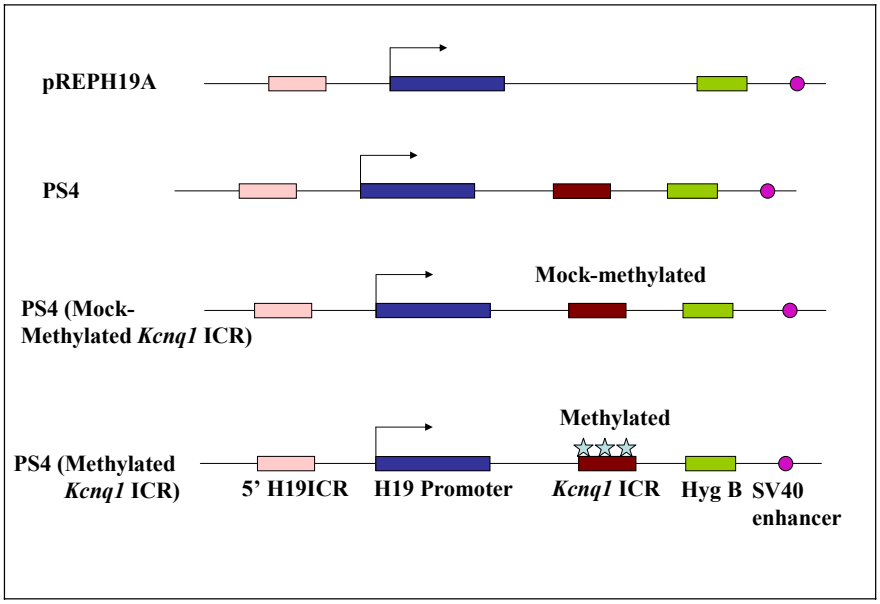

Figure 7. Schematic map of mock-methylated and methylated constructs.

Chromatin conformation of the *Kcnq1* ICR

Three *DNase I* hypersensitive sites were mapped within the *Kcnq1* ICR in both mouse fetal and adult liver. These sites were recapitulated in the mock-methylated clones, whereas methylated clones did not show any hypersensi-

tive sites. This result indicated that the DNA-protein interactions are abrogated by methylation.

Overall our studies show that the *Kcnq1* ICR harbors insulator activity in a methylation dependent manner, and that chromatin organization differs in methylated and unmethylated regions. It also suggests that methylation-dependent chromatin insulator activity could be a common feature of imprinting control regions.

Paper III

Previous *in vitro* studies (5, 34, 49) and the studies from paper I clearly indicated that the chromatin insulator protein CTCF plays an important role in setting up the insulator complex in methylation-dependent manner. Recent studies from a mouse knock-in model have further indicated the importance of CTCF in maintaining a methylation-free *H19* imprinting control region when inherited maternally. The region is subjected to *de novo* methylation by losing CTCF binding (75, 83). These emerging facts made us curious to find out whether occupancy of CTCF plays a universal role in maintaining methylation-free domains. CTCF binds to a variety of target sequences through combinatorial use of its 11 zinc fingers, which makes the genomewide mapping of CTCF target sites difficult to predict. However this hurdle was overcome by generating a CTCF target site library from chromatin immunoprecipitated DNA (ChIP) using a C-terminal CTCF antibody in mouse fetal liver. Several interesting observations were documented, which are discussed below.

CTCF target site occupancy in mouse fetal liver

The ChIP DNA obtained from the mouse fetal liver was ligated to T7 and T3 linkers, followed by PCR amplification and size selection and they were cloned into pGEM vector. The DNA was prepared from these clones and inserts PCR amplified using the T7 and SP6 primers and spotted on poly-lysine coated slides to generate CTCF target site library microarrays. Probes for hybridization were prepared from ChIP DNA using an N-terminal CTCF antibody or MBD2 antibody, and DNA obtained from *in vitro* binding assays, in which complexes formed between library sequences and recombinant CTCF is immunopurified.

Comparing the hybridizations from the ChIP DNA using the N-terminal CTCF antibody and *in vitro* binding DNA, and visualizing these outcomes by generating scatter plots, led to several important conclusions. It was found that there is moderate agreement between the *in vivo* and *in vitro* binding, which may indicate the possibility that chromatin organization restricts the availability of target sites *in vivo*. Secondly, there were sequences which showed strong *in vivo* binding but no *in vitro* binding, which suggests that protein-protein interactions could localize CTCF to non-CTCF target sites. It

was also found that the binding of CTCF shows a continuous distribution pattern, indicating that majority of the sequences interact with CTCF. However, while accounting for such type of interactions it should be accepted that there exists a gradient of affinity between CTCF and target site sequences. Finally, many repeat sequences were found to interact with CTCF like LTRs, LINEs, SINEs and microsattelite repeats. It has been shown that CTCF binds to microsatellite repeats and that these microsattelite sequences play a role in regulating specific gene expression (2).

Mapping of CTCF target sites to the loci with pivotal functions

Following sequencing and bioinformatic analysis of the CTCF target site library clones, it was found that two thirds of the sequences in the CTCF target site library could not be identified in the mouse genome databases. However southern blot and PCR analysis confirmed that those targets are present in mouse genome and ruled out the possibility of contamination. Hence we propose that those sequences belong to the heterochromatic regions in the mouse genome. Analyzing the distribution of CTCF and heterochromatic protein 1 (HP1) in the nucleus, it was found that CTCF colocalizes with HP1, which supports the idea that there is a possible link between CTCF and heterochromatin. 55 identified sequences mapped to loci with pivotal functions in the mouse genome, which included target sites in the introns, exons and intergenic regions. Further information about these target sites was obtained from the Gene Ontology database, showing that target sites mapped to loci which are involved in multiple cellular functions such as signaling and apoptotic pathways, neurogenesis, metabolism and clusters of olfactory or pheromone receptor genes. In addition to this two target sites mapped to imprinted domains (*Grb10* and *Snrpn*) and four others were identified in the EICO library of candidate imprinted genes.

Screening for insulator function using a novel assay

To examine the insulator function, two target site sequences were cloned into the episomal vector (see below Figure 8), and by performing an RNase protection assay it was shown that they were able to block the *H19* reporter gene expression to various degrees (Figure 4a, paper III). However to account for the insulator function for all the target site sequences we developed a novel approach in which sequences were inserted between the SV40 enhancer and a toxin-A reporter gene with *H19* promoter in the episomal based toxin construct (see below Figure 8), and were subjected to hygromycin selection. Only those sequences which could successfully prevent the enhancer-promoter communication can survive the selection. The surviving clones were used to prepare DNA and were then hybridized to the microarray containing clones from CTCF target site library. The *H19* ICR was ligated at the same position to act as a positive control. A few of the clone sequences from the target site library were also individually subjected to selection and they were found to block the enhancer-promoter communication to various degrees, while two other clones with no insulator function failed to survive

(Figure 4b, paper III). It was observed from the scatter plots that insulator strength shows a correlation with *in vitro* binding affinity of CTCF with a continuous distribution of the stringency (Figure 4e, paper III), suggesting that insulator function is correlated to the *in vitro* binding strength.

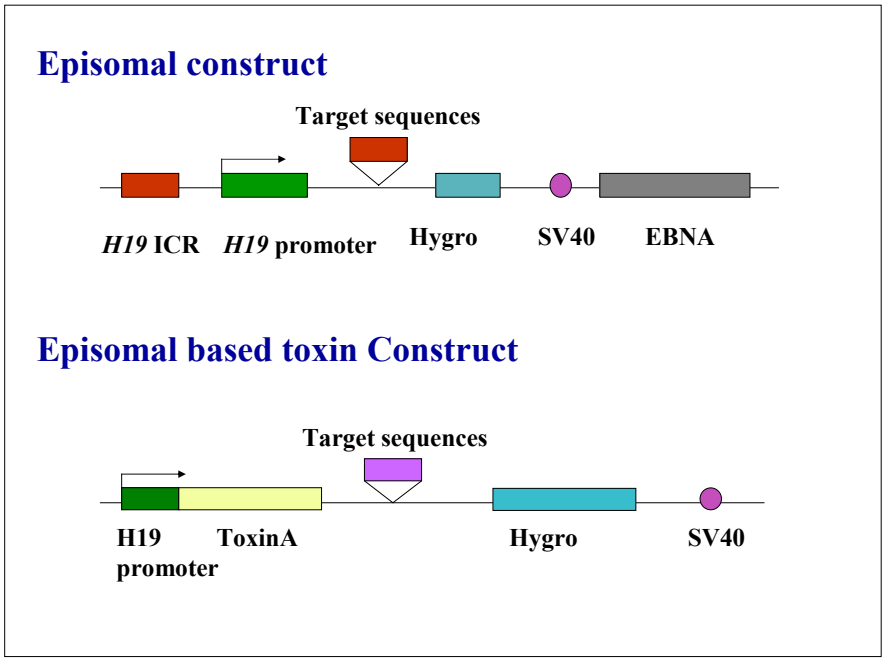

Figure 8. Schematic map of pREP episomal and toxin construct.

CTCF target sites map to methylation-free domains

The CTCF target site library was cross-referenced with DNA methylation marks by hybridizing the probes generated from immunopurified sequences using antibodies against 5-methylcytidine and a methyl-binding protein (MBD2) which interacts with clusters of methylated CpGs. Scatter plots were generated to compare the CTCF binding with methylation status and it was observed that CTCF maintains methylation free domains as visualized from MBD2 binding (Figure 5c, paper III). In case of the binding of 5-methylcytidine, which binds to single CpGs, there exists a small level of background which reflects the flexibility of methylation status at single CpG sites.

Overall our results have unraveled a CTCF-based network in many important pathways such as growth, neurogenesis and metabolism. In addition there appears to be a correlation between CTCF and epigenetic states throughout the genome. An important observation reported here involves the insulator function, which is probably dictated by chromatin modifications.

The insulator complex appears to function in a continuous or analog mode, suggesting role of stochastic events in the regulation of gene expression. Since many of the target sites remain unidentified, we suggest that they belong to heterochromatic regions based on the co-localization with HP1; it appears that CTCF target sites are shared between euchromatic and heterochromatic domains. Hence it could be possible that CTCF maintains expression domains within the heterochromatic region.

Paper IV

As discussed in the beginning of this thesis, a single cell gives rise to different cell types although they consist of the same genetic material. This emphasizes the fact that the expression patterns of certain genes are developmentally regulated. Epigenetic marks can regulate these expression patterns to a great extent, which would also involve dictating the binding pattern for certain factors in a tissue-specific manner. Since CTCF is a multi-functional constitutively expressed protein and our previous studies have shown that CTCF is an important component of insulator complex and also maintains methylation free domains throughout the genome (paper III), we wanted to study its binding pattern in different tissues.

Lineage-specific occupancy of CTCF target sites

In order to study the different lineage-specific binding pattern, we prepared immunoprecipitated DNA using the C-terminal specific CTCF antibody from mouse fetal liver, brain, thymus, skin, undifferentiated and differentiated embryonic stem cells. The ChIP DNA was hybridized to the microarrays generated from the mouse CTCF target site library.

The binding patterns for different tissues were visualized using scatter plots. It was found that most CTCF target sites interact with CTCF in mouse fetal liver but the occupancy of CTCF target sites is greatly reduced in the case of brain, thymus and skin (figure 1b-e, paper IV). On the other hand, almost all of the target sites are occupied in the case of undifferentiated stem cells, but again a reduced level of occupancy was observed for differentiated cells (figure 1f-g, paper IV). Since stem cells are pluripotent in nature the cell fate remains undecided until they undergo differentiation. It could be possible that it has an entirely different chromatin conformation; hence all the target sites are available. The availability of target sites can also be explained in terms of availability of protein partners for CTCF. In our previous study we have documented that CTCF can be localized to a non-CTCF target site by protein-protein interactions. Hence if all different protein partners for CTCF are present then it shows the possibility that all the binding sites will be occupied by CTCF. Another fact that emerged from our observations is that target sites could be divided into two categories, one which shows a constitutive binding pattern and the other showing lineage-specific binding

(facultative). The lineage-specific occupancy of CTCF target sites could be due to different methylation status and varied chromatin conformations in different cell types.

Another interesting observation was that the *H19* ICR shows constitutive binding for CTCF, whereas *KvDMR*, *IGF2r* and *Snrpn* show lineage-specific binding. Undifferentiated stem cells show strong binding for all three except *Snrpn*. In liver all four genes show binding but *H19* is the strongest of all. Other hypothetical imprinted genes also show lineage-specific binding patterns. Although there is a link between CTCF bindings and imprinting, it shows that CTCF binding could be one of the motifs in maintaining and propagating the imprinted states as in the case of the *H19*/*Igf2* locus. Hence it could be concluded that only a small fraction of imprinting control regions have constitutive binding whereas large number of them show lineage-specific features.

Lineage-specific occupancy is not dictated by CTCF expression

The *in vitro* and *in vivo* binding of CTCF was compared for ES cells and differentiated ES cells. This comparison revealed that undifferentiated ES cells undergo both epigenetic modification and modification of protein-protein interactions for non-CTCF target site sequences during the process of differentiation. Since the occupancy for CTCF target sites was varying enormously between different tissue types, there was a possibility that different levels of CTCF expression could be one of the reasons for such variation. Western blot analysis suggests that lineage-specific CTCF target site occupancy for low-affinity target sites is possibly related to the levels of CTCF protein. This is supported by the observation in skin, where a very small fraction of CTCF target sites are occupied *in vivo,* which in turn is comparable to the presence of a lower level of the CTCF protein. However, abundance of protein is not correlated with the *in vivo* occupancy in the case of brain.

Hence we conclude that occupancy of CTCF target sites varies in a lineage-specific manner and is developmentally regulated. Different epigenetic marks could possibly affect the occupancy in a tissue-specific manner. The availability of different protein partners and formation of higher order chromatin conformation can also be crucial players dictating the target site occupancy in different tissues.

CONCLUDING REMARKS

Despite the fact that all cells harbor the same genetic material, they display different morphological and functional differences which are maintained by complicated gene regulation involving genetic and epigenetic factors. The work in this thesis indicates the complex epigenetic aspect of gene regulation, with a special emphasis on two different imprinting control regions on mouse chromosome 7. The thesis presents data which proves that insulator activity, which is important in maintaining spatiotemporal patterns of gene expression, is methylation-sensitive and is a characteristic feature of imprinting control regions.

Subsequently, studies on the multifunctional protein CTCF, which is involved in setting up the insulator complex, indicates that it maintains methylation-free domains genomewide. The CTCF target sites are associated with loci with many important pivotal functions. Further, its involvement in the insulator complex throughout the genome indicates that it plays crucial role in maintaining expression domains throughout the mouse genome.

Finally, the lineage-specific occupancy of CTCF unravels the fact that the CTCF target site occupancy is developmentally regulated, which are controlled by different epigenetic marks and other protein factors present during the development.

Thus the studies indicate the complexity of mechanisms involved in regulating gene expression and show that epigenetic marks are involved to a great extent in maintaining proper expression domains. This could be an important implication for developmental abnormalities and many human diseases because most of them involve disruption or alteration of the epigenetic state. This alteration further abrogates the insulator function, leading to improper expression of genes.

ACKNOWLEDGEMENTS

This work was carried out at the department of Development and Genetics, Uppsala University. I wish to express my gratitude to following people:

Rolf Ohlsson, head of the department and supervisor of this thesis work for introducing me to the field of epigenetics, enthusiastic scientific discussions and for giving me the freedom to explore the exciting area of microarrays. Thanks for your generous help and support.

My colleagues, **Chandrasekhar Kanduri**, **Claes Holmgren**, **Meena Kanduri**, **Joanne Whitehead** (great help with the linguistics), **Piero Mariano**, **Magda Lezcano**, **Noopur Thakur**, **Srinivasulu Kurukuti**, **Vijay Tiwari**, **Anita Gönder**, **Rosita Bergström**, **Gholmreza Tovossi**, **Zhihu Zhao**, **Pandey** and **Rolf Ericsson** (generous assistance with computers and ugly blot scanning) for their assistance in different ways.

Special thanks to **Vinod** for being excellent friend and for never-ending arguments! **Wenqiang** for being nice friend and great project partner, all the scientific & non-scientific discussions and affectionate nature. **Li** & **Yue Yue** for all the fun.

Mouse genetics group members, **Reinald Fundele**, **Wei Shi**, **Umashankar Singh**, **Farhan Rizvi**, **Tong Sun** for their help and scientific discussions.

The wonderful ladies, **Helena Malmikumpu** for untiring help and affection, **Anita Mattson** for enormous help with animal facility, **Rose-marie Andersson** for efficiently taking care of all financial details, **Ambar Khodabandah** for the help with sequencing

The present and previous members, **Carina**, **Carl**, **Åke**, **Graham**, **Jan**, **Gary**, **Erik**, **Lai & Lisa**, **Vasu**, **Jungwang**, **Liang**, **Marie**, **Arwen**, **Yo nino**, **Gun-Britt**, **Lennart**, **Svante**, **Olivier** and **Oyvind** for their help.

Stefan Gunnarsson, **Gary Wife**, **Stefan Ås** for help with computers and microscopes.

Rudbeck laboratory members, **Anders Isaksson**, **Jacob**, **Niclas** for their help with the microarray printing facility. **Jan Dumanski**, **Carl**, **Kiran** and **Patrick** for their help with arrays.

Members from other departments, **Pikul**, **Young A** and other members from **Compys** and **Ekotox** for their help.

My previous supervisors and mentors, **Amulya Panda**, **Pramood Upadhyay**, **Devapriya Choudhury** for their guidance and encouragement in pursuing a scientific career.

My Best friends, **Gargi**, **Parul** and **Nannu,** for always being there for me.

My Stockholm-friends, **Rajeev & Parvati**, **Raju**, **Oscar**, **Shankha**, **Vipin & Nilima**, **Amarjit**, **Reetakshi**, **Vikas** for all the great parties and memorable times.

The **Indian community of Uppsala** for their help in different ways.

My PUBT janta and NII bunch, **Maityda & Munia**, **Jhala & bhabhi**, **Anand**, **Manoj**, **Ashu**, **Jagan**, **Rajeev**, **Mandy & Tresa**, **Shweta**, **Deepak**, **Srikanth**, **Jona & Surendra**, **Swapna**, **Pritha**, **Parimalda & Bhagyalaxmidi**, **Shantanuda & Arpita**, **Munmun**, **Shamimda**, **Sudeep & Sangita**, **Malini**, **Harpreet**,and **Sourabh** for making my world so beautiful.

My parents, for being the best, words are not enough to express my feelings for you. **My little brother Arijit**, (my loving PD) for endless jokes, love and your ability to cheer me always. **Sandeepan** for being friendly and understanding. Thanks for pleasant time and company.

All **my relatives** for the support and love.

All others whose names I might have forgot.

All mice, whose sacrifice made this work possible.

This work was financially supported by the Swedish Science Research Council, the juvenile Diabetes Research Foundation International, the Swedish Cancer Research Foundation, the Swedish Pediatric Cancer Foundation, the Wallenberg and Lundberg Foundations, the Swedish Royal Academy of Science, the Von Hofsten Foundation, the Lars Hieta Foundation and the Helge Ax:son Johnson Foundation.

REFERENCES

1. **Ainscough, J. F., R. M. John, and M. A. Surani.** 1998. Mechanism of imprinting on mouse distal chromosome 7. Genet Res **72:**237-45.
2. **Arnold, R., W. Maueler, G. Bassili, M. Lutz, L. Burke, T. J. Epplen, and R. Renkawitz.** 2000. The insulator protein CTCF represses transcription on binding to the (gt)(22)(ga)(15) microsatellite in intron 2 of the HLA-DRB1(*)0401 gene. Gene **253:**209-14.
3. **Ballestar, E., and A. P. Wolffe.** 2001. Methyl-CpG-binding proteins. Targeting specific gene repression. Eur J Biochem **268:**1-6.
4. **Baniahmad, A., C. Steiner, A. C. Kohne, and R. Renkawitz.** 1990. Modular structure of a chicken lysozyme silencer: involvement of an unusual thyroid hormone receptor binding site. Cell **61:**505-14.
5. **Bell, A. C., and G. Felsenfeld.** 2000. Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. Nature **405:**482-5.
6. **Bell, A. C., A. G. West, and G. Felsenfeld.** 2001. Insulators and boundaries: versatile regulatory elements in the eukaryotic. Science **291:**447-50.
7. **Blackwood, E. M., and J. T. Kadonaga.** 1998. Going the distance: a current view of enhancer action. Science **281:**61-3.
8. **Bondarenko, V. A., Y. V. Liu, Y. I. Jiang, and V. M. Studitsky.** 2003. Communication over a large distance: enhancers and insulators. Biochem Cell Biol **81:**241-51.
9. **Bourc'his, D., G. L. Xu, C. S. Lin, B. Bollman, and T. H. Bestor.** 2001. Dnmt3L and the establishment of maternal genomic imprints. Science **294:**2536-9.
10. **Bulger, M., and M. Groudine.** 2002. TRAPping enhancer function. Nat Genet **32:**555-6.
11. **Burcin, M., R. Arnold, M. Lutz, B. Kaiser, D. Runge, F. Lottspeich, G. N. Filippova, V. V. Lobanenkov, and R. Renkawitz.** 1997. Negative protein 1, which is required for function of the chicken lysozyme gene silencer in conjunction with hormone receptors, is identical to the multivalent zinc finger repressor CTCF. Mol Cell Biol **17:**1281-8.
12. **Burgess-Beusse, B., C. Farrell, M. Gaszner, M. Litt, V. Mutskov, F. Recillas-Targa, M. Simpson, A. West, and G. Felsenfeld.** 2002. The insulation of genes from external enhancers and silencing chromatin. Proc Natl Acad Sci U S A **99 Suppl 4:**16433-7.

13. **Csink, A. K., and S. Henikoff.** 1998. Something from nothing: the evolution and utility of satellite repeats. Trends Genet **14:**200-4.
14. **Dillon, N., and R. Festenstein.** 2002. Unravelling heterochromatin: competition between positive and negative factors regulates accessibility. Trends Genet **18:**252-8.
15. **Driouch, K., F. Dorion-Bonnet, M. Briffod, M. H. Champeme, M. Longy, and R. Lidereau.** 1997. Loss of heterozygosity on chromosome arm 16q in breast cancer metastases. Genes Chromosomes Cancer **19:**185-91.
16. **Elgin, S. C., and S. I. Grewal.** 2003. Heterochromatin: silence is golden. Curr Biol **13:**R895-8.
17. **Esteller, M., M. F. Fraga, M. F. Paz, E. Campo, D. Colomer, F. J. Novo, M. J. Calasanz, O. Galm, M. Guo, J. Benitez, and J. G. Herman.** 2002. Cancer epigenetics and methylation. Science **297:**1807-8; discussion 1807-8.
18. **Esteller, M., and J. G. Herman.** 2002. Cancer as an epigenetic disease: DNA methylation and chromatin alterations in human tumours. J Pathol **196:**1-7.
19. **Filippova, G. N., S. Fagerlie, E. M. Klenova, C. Myers, Y. Dehner, G. Goodwin, P. E. Neiman, S. J. Collins, and V. V. Lobanenkov.** 1996. An exceptionally conserved transcriptional repressor, CTCF, employs different combinations of zinc fingers to bind diverged promoter sequences of avian and mammalian c-myc oncogenes. Mol Cell Biol **16:**2802-13.
20. **Filippova, G. N., A. Lindblom, L. J. Meincke, E. M. Klenova, P. E. Neiman, S. J. Collins, N. A. Doggett, and V. V. Lobanenkov.** 1998. A widely expressed transcription factor with multiple DNA sequence specificity, CTCF, is localized at chromosome segment 16q22.1 within one of the smallest regions of overlap for common deletions in breast and prostate cancers. Genes Chromosomes Cancer **22:**26-36.
21. **Filippova, G. N., C. F. Qi, J. E. Ulmer, J. M. Moore, M. D. Ward, Y. J. Hu, D. I. Loukinov, E. M. Pugacheva, E. M. Klenova, P. E. Grundy, A. P. Feinberg, A. M. Cleton-Jansen, E. W. Moerland, C. J. Cornelisse, H. Suzuki, A. Komiya, A. Lindblom, F. Dorion-Bonnet, P. E. Neiman, H. C. Morse, 3rd, S. J. Collins, and V. V. Lobanenkov.** 2002. Tumor-associated zinc finger mutations in the CTCF transcription factor selectively alter tts DNA-binding specificity. Cancer Res **62:**48-52.
22. **Filippova, G. N., C. P. Thienes, B. H. Penn, D. H. Cho, Y. J. Hu, J. M. Moore, T. R. Klesert, V. V. Lobanenkov, and S. J. Tapscott.** 2001. CTCF-binding sites flank CTG/CAG repeats and form a methylation-sensitive insulator at the DM1 locus. Nat Genet **28:**335-43.
23. **Fischle, W., Y. Wang, and C. D. Allis.** 2003. Histone and chromatin cross-talk. Curr Opin Cell Biol **15:**172-83.

24. **Fitzpatrick, G. V., P. D. Soloway, and M. J. Higgins.** 2002. Regional loss of imprinting and growth deficiency in mice with a targeted deletion of KvDMR1. Nat Genet **32:**426-31.
25. **Gasser, S. M., and M. M. Cockell.** 2001. The molecular biology of the SIR proteins. Gene **279:**1-16.
26. **Gerasimova, T. I., and V. G. Corces.** 2001. Chromatin insulators and boundaries: effects on transcription and nuclear organization. Annu Rev Genet **35:**193-208.
27. **Gilbert, N., S. Boyle, H. Sutherland, J. de Las Heras, J. Allan, T. Jenuwein, and W. A. Bickmore.** 2003. Formation of facultative heterochromatin in the absence of HP1. Embo J **22:**5540-50.
28. **Ginjala, V., C. Holmgren, E. Ulleras, C. Kanduri, V. Pant, V. Lobanenkov, G. Franklin, and R. Ohlsson.** 2002. Multiple cis elements within the Igf2/H19 insulator domain organize a distance-dependent silencer. A cautionary note. J Biol Chem **277:**5707-10.
29. **Goll, M. G., and T. H. Bestor.** 2002. Histone modification and replacement in chromatin activation. Genes Dev **16:**1739-42.
30. **Gould, T. D., and K. Pfeifer.** 1998. Imprinting of mouse Kvlqt1 is developmentally regulated. Hum Mol Genet **7:**483-7.
31. **Grewal, S. I., and D. Moazed.** 2003. Heterochromatin and epigenetic control of gene expression. Science **301:**798-802.
32. **Grunstein, M.** 1997. Histone acetylation in chromatin structure and transcription. Nature **389:**349-52.
33. **Grunstein, M.** 1997. Molecular model for telomeric heterochromatin in yeast. Curr Opin Cell Biol **9:**383-7.
34. **Hark, A. T., C. J. Schoenherr, D. J. Katz, R. S. Ingram, J. M. Levorse, and S. M. Tilghman.** 2000. CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. Nature **405:**486-9.
35. **Hata, K., M. Okano, H. Lei, and E. Li.** 2002. Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice. Development **129:**1983-93.
36. **Hazzouri, M., C. Pivot-Pajot, A. K. Faure, Y. Usson, R. Pelletier, B. Sele, S. Khochbin, and S. Rousseaux.** 2000. Regulated hyperacetylation of core histones during mouse spermatogenesis: involvement of histone deacetylases. Eur J Cell Biol **79:**950-60.
37. **Herendeen, D. R., G. A. Kassavetis, and E. P. Geiduschek.** 1992. A transcriptional enhancer whose function imposes a requirement that proteins track along DNA. Science **256:**1298-303.
38. **Hitchins, M. P., and G. E. Moore.** 2002. Genomic imprinting in fetal growth and development. Expert Rev Mol Med **2002:**1-19.
39. **Horak, C. E., M. C. Mahajan, N. M. Luscombe, M. Gerstein, S. M. Weissman, and M. Snyder.** 2002. GATA-1 binding sites mapped in the beta-globin locus by using mammalian chIp-chip analysis. Proc Natl Acad Sci U S A **99:**2924-9.
40. **Howell, C. Y., T. H. Bestor, F. Ding, K. E. Latham, C. Mertineit, J. M. Trasler, and J. R. Chaillet.** 2001. Genomic imprinting dis-

rupted by a maternal effect mutation in the Dnmt1 gene. Cell **104:**829-38.

41. **Howlett, S. K., and W. Reik.** 1991. Methylation levels of maternal and paternal genomes during preimplantation development. Development **113:**119-27.
42. **Iizuka, M., and M. M. Smith.** 2003. Functional consequences of histone modifications. Curr Opin Genet Dev **13:**154-60.
43. **Iyer, V. R., C. E. Horak, C. S. Scafe, D. Botstein, M. Snyder, and P. O. Brown.** 2001. Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF. Nature **409:**533-8.
44. **John, R. M., and M. A. Surani.** 2000. Genomic imprinting, mammalian evolution, and the mystery of egg-laying mammals. Cell **101:**585-8.
45. **Kaffer, C. R., M. Srivastava, K. Y. Park, E. Ives, S. Hsieh, J. Batlle, A. Grinberg, S. P. Huang, and K. Pfeifer.** 2000. A transcriptional insulator at the imprinted H19/Igf2 locus. Genes Dev **14:**1908-19.
46. **Kafri, T., M. Ariel, M. Brandeis, R. Shemer, L. Urven, J. McCarrey, H. Cedar, and A. Razin.** 1992. Developmental pattern of gene-specific DNA methylation in the mouse embryo and germ line. Genes Dev **6:**705-14.
47. **Kanduri, C., G. Fitzpatrick, R. Mukhopadhyay, M. Kanduri, V. Lobanenkov, M. Higgins, and R. Ohlsson.** 2002. A differentially methylated imprinting control region within the Kcnq1 locus harbors a methylation-sensitive chromatin insulator. J Biol Chem **277:**18106-10.
48. **Kanduri, C., C. Holmgren, M. Pilartz, G. Franklin, M. Kanduri, L. Liu, V. Ginjala, E. Ulleras, R. Mattsson, and R. Ohlsson.** 2000. The 5' flank of mouse H19 in an unusual chromatin conformation unidirectionally blocks enhancer-promoter communication. Curr Biol **10:**449-57.
49. **Kanduri, C., V. Pant, D. Loukinov, E. Pugacheva, C. F. Qi, A. Wolffe, R. Ohlsson, and V. V. Lobanenkov.** 2000. Functional association of CTCF with the insulator upstream of the H19 gene is parent of origin-specific and methylation-sensitive. Curr Biol **10:**853-6.
50. **Klenova, E. M., S. Fagerlie, G. N. Filippova, L. Kretzner, G. H. Goodwin, G. Loring, P. E. Neiman, and V. V. Lobanenkov.** 1998. Characterization of the chicken CTCF genomic locus, and initial study of the cell cycle-regulated promoter of the gene. J Biol Chem **273:**26571-9.
51. **Klenova, E. M., R. H. Nicolas, H. F. Paterson, A. F. Carne, C. M. Heath, G. H. Goodwin, P. E. Neiman, and V. V. Lobanenkov.** 1993. CTCF, a conserved nuclear factor required for optimal transcriptional activity of the chicken c-myc gene, is an 11-Zn-finger protein differentially expressed in multiple forms. Mol Cell Biol **13:**7612-24.

52. **Klenova, E. M., R. H. Nicolas, S. U, A. F. Carne, R. E. Lee, V. V. Lobanenkov, and G. H. Goodwin.** 1997. Molecular weight abnormalities of the CTCF transcription factor: CTCF migrates aberrantly in SDS-PAGE and the size of the expressed protein is affected by the UTRs and sequences within the coding region of the CTCF gene. Nucleic Acids Res **25:**466-74.
53. **Kouzarides, T.** 1999. Histone acetylases and deacetylases in cell proliferation. Curr Opin Genet Dev **9:**40-8.
54. **Kurdistani, S. K., D. Robyr, S. Tavazoie, and M. Grunstein.** 2002. Genome-wide binding map of the histone deacetylase Rpd3 in yeast. Nat Genet **31:**248-54.
55. **Ladomery, M.** 1997. Multifunctional proteins suggest connections between transcriptional and post-transcriptional processes. Bioessays **19:**903-9.
56. **Ladomery, M., and G. Dellaire.** 2002. Multifunctional zinc finger proteins in development and disease. Ann Hum Genet **66:**331-42.
57. **Lasko, D., W. Cavenee, and M. Nordenskjold.** 1991. Loss of constitutional heterozygosity in human cancer. Annu Rev Genet **25:**281-314.
58. **Latil, A., O. Cussenot, G. Fournier, K. Driouch, and R. Lidereau.** 1997. Loss of heterozygosity at chromosome 16q in prostate adenocarcinoma: identification of three independent regions. Cancer Res **57:**1058-62.
59. **Leighton, P. A., J. R. Saam, R. S. Ingram, C. L. Stewart, and S. M. Tilghman.** 1995. An enhancer deletion affects both H19 and Igf2 expression. Genes Dev **9:**2079-89.
60. **Li, E.** 2002. Chromatin modification and epigenetic reprogramming in mammalian development. Nat Rev Genet **3:**662-73.
61. **Li, E., T. H. Bestor, and R. Jaenisch.** 1992. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. Cell **69:**915-26.
62. **Lieb, J. D., X. Liu, D. Botstein, and P. O. Brown.** 2001. Promoter-specific binding of Rap1 revealed by genome-wide maps of protein-DNA association. Nat Genet **28:**327-34.
63. **Lobanenkov, V. V., R. H. Nicolas, V. V. Adler, H. Paterson, E. M. Klenova, A. V. Polotskaja, and G. H. Goodwin.** 1990. A novel sequence-specific DNA binding protein which interacts with three regularly spaced direct repeats of the CCCTC-motif in the 5'-flanking sequence of the chicken c-myc gene. Oncogene **5:**1743-53.
64. **Lutz, M., L. J. Burke, G. Barreto, F. Goeman, H. Greb, R. Arnold, H. Schultheiss, A. Brehm, T. Kouzarides, V. Lobanenkov, and R. Renkawitz.** 2000. Transcriptional repression by the insulator protein CTCF involves histone deacetylases. Nucleic Acids Res **28:**1707-13.
65. **Mancini-DiNardo, D., S. J. Steele, R. S. Ingram, and S. M. Tilghman.** 2003. A differentially methylated region within the gene

Kcnq1 functions as an imprinted promoter and silencer. Hum Mol Genet **12:**283-94.

66. **Martin, S. L., D. Branciforte, D. Keller, and D. L. Bain.** 2003. Trimeric structure for an essential protein in L1 retrotransposition. Proc Natl Acad Sci U S A **100:**13815-20.
67. **Meehan, R. R.** 2003. DNA methylation in animal development. Semin Cell Dev Biol **14:**53-65.
68. **Mongelard, F., and V. G. Corces.** 2001. Two insulators are not better than one. Nat Struct Biol **8:**192-4.
69. **Nakao, M.** 2001. Epigenetics: interaction of DNA methylation and chromatin. Gene **278:**25-31.
70. **Nephew, K. P., and T. H. Huang.** 2003. Epigenetic gene silencing in cancer initiation and progression. Cancer Lett **190:**125-33.
71. **Ohlsson, R., R. Renkawitz, and V. Lobanenkov.** 2001. CTCF is a uniquely versatile transcription regulator linked to epigenetics and disease. Trends Genet **17:**520-7.
72. **Okano, M., D. W. Bell, D. A. Haber, and E. Li.** 1999. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell **99:**247-57.
73. **Oki, M., and R. T. Kamakaka.** 2002. Blockers and barriers to transcription: competing activities? Curr Opin Cell Biol **14:**299-304.
74. **O'Neill, M. J., R. S. Ingram, P. B. Vrana, and S. M. Tilghman.** 2000. Allelic expression of IGF2 in marsupials and birds. Dev Genes Evol **210:**18-20.
75. **Pant, V., P. Mariano, C. Kanduri, A. Mattsson, V. Lobanenkov, R. Heuchel, and R. Ohlsson.** 2003. The nucleotides responsible for the direct physical contact between the chromatin insulator protein CTCF and the H19 imprinting control region manifest parent of origin-specific long-distance insulation and methylation-free domains. Genes Dev **17:**586-90.
76. **Patience, C., Y. Takeuchi, and R. A. Weiss.** 1997. Infection of human cells by an endogenous retrovirus of pigs. Nat Med **3:**282-6.
77. **Patience, C., D. A. Wilkinson, and R. A. Weiss.** 1997. Our retroviral heritage. Trends Genet **13:**116-20.
78. **Paulsen, M., K. R. Davies, L. M. Bowden, A. J. Villar, O. Franck, M. Fuermann, W. L. Dean, T. F. Moore, N. Rodrigues, K. E. Davies, R. J. Hu, A. P. Feinberg, E. R. Maher, W. Reik, and J. Walter.** 1998. Syntenic organization of the mouse distal chromosome 7 imprinting cluster and the Beckwith-Wiedemann syndrome region in chromosome 11p15.5. Hum Mol Genet **7:**1149-59.
79. **Pfeifer, K.** 2000. Mechanisms of genomic imprinting. Am J Hum Genet **67:**777-87.
80. **Ren, B., F. Robert, J. J. Wyrick, O. Aparicio, E. G. Jennings, I. Simon, J. Zeitlinger, J. Schreiber, N. Hannett, E. Kanin, T. L. Volkert, C. J. Wilson, S. P. Bell, and R. A. Young.** 2000. Ge-

nome-wide location and function of DNA binding proteins. Science **290:**2306-9.

81. **Robbins, J.** 2001. KCNQ potassium channels: physiology, pathophysiology, and pharmacology. Pharmacol Ther **90:**1-19.
82. **Santos, F., B. Hendrich, W. Reik, and W. Dean.** 2002. Dynamic reprogramming of DNA methylation in the early mouse embryo. Dev Biol **241:**172-82.
83. **Schoenherr, C. J., J. M. Levorse, and S. M. Tilghman.** 2003. CTCF maintains differential methylation at the Igf2/H19 locus. Nat Genet **33:**66-9.
84. **Shankaranarayana, G. D., M. R. Motamedi, D. Moazed, and S. I. Grewal.** 2003. Sir2 regulates histone H3 lysine 9 methylation and heterochromatin assembly in fission yeast. Curr Biol **13:**1240-6.
85. **Singh, P. B., and S. D. Georgatos.** 2002. HP1: facts, open questions, and speculation. J Struct Biol **140:**10-6.
86. **Smale, S. T.** 2001. Core promoters: active contributors to combinatorial gene regulation. Genes Dev **15:**2503-8.
87. **Smale, S. T., and D. Baltimore.** 1989. The "initiator" as a transcription control element. Cell **57:**103-13.
88. **Smale, S. T., A. Jain, J. Kaufmann, K. H. Emami, K. Lo, and I. P. Garraway.** 1998. The initiator element: a paradigm for core promoter heterogeneity within metazoan protein-coding genes. Cold Spring Harb Symp Quant Biol **63:**21-31.
89. **Strahl, B. D., and C. D. Allis.** 2000. The language of covalent histone modifications. Nature **403:**41-5.
90. **Strathdee, G., and R. Brown.** 2002. Aberrant DNA methylation in cancer: potential clinical interventions. Expert Rev Mol Med **2002:**1-17.
91. **Surani, M. A., S. C. Barton, and M. L. Norris.** 1984. Development of reconstituted mouse eggs suggests imprinting of the genome during gametogenesis. Nature **308:**548-50.
92. **Svensson, K., R. Mattsson, T. C. James, P. Wentzel, M. Pilartz, J. MacLaughlin, S. J. Miller, T. Olsson, U. J. Eriksson, and R. Ohlsson.** 1998. The paternal allele of the H19 gene is progressively silenced during early mouse development: the acetylation status of histones may be involved in the generation of variegated expression patterns. Development **125:**61-9.
93. **Thiru, A., D. Nietlispach, H. R. Mott, M. Okuwaki, D. Lyon, P. R. Nielsen, M. Hirshberg, A. Verreault, N. V. Murzina, and E. D. Laue.** 2004. Structural basis of HP1/PXVXL motif peptide interactions and HP1 localisation to heterochromatin. Embo J **23:**489-99.
94. **Thorvaldsen, J. L., K. L. Duran, and M. S. Bartolomei.** 1998. Deletion of the H19 differentially methylated domain results in loss of imprinted expression of H19 and Igf2. Genes Dev **12:**3693-702.
95. **Thorvaldsen, J. L., M. R. Mann, O. Nwoko, K. L. Duran, and M. S. Bartolomei.** 2002. Analysis of sequence upstream of the endoge-

nous H19 gene reveals elements both essential and dispensable for imprinting. Mol Cell Biol **22:**2450-62.

96. **Turner, B. M.** 2000. Histone acetylation and an epigenetic code. Bioessays **22:**836-45.
97. **Urnov, F. D.** 2002. A feel for the template: zinc finger protein transcription factors and chromatin. Biochem Cell Biol **80:**321-33.
98. **Vogelauer, M., J. Wu, N. Suka, and M. Grunstein.** 2000. Global histone acetylation and deacetylation in yeast. Nature **408:**495-8.
99. **Vostrov, A. A., and W. W. Quitschke.** 1997. The zinc finger protein CTCF binds to the APBbeta domain of the amyloid beta-protein precursor promoter. Evidence for a role in transcriptional activation. J Biol Chem **272:**33353-9.
100. **Vostrov, A. A., M. J. Taheny, and W. W. Quitschke.** 2002. A region to the N-terminal side of the CTCF zinc finger domain is essential for activating transcription from the amyloid precursor protein promoter. J Biol Chem **277:**1619-27.
101. **Weksberg, R., J. Nishikawa, O. Caluseriu, Y. L. Fei, C. Shuman, C. Wei, L. Steele, J. Cameron, A. Smith, I. Ambus, M. Li, P. N. Ray, P. Sadowski, and J. Squire.** 2001. Tumor development in the Beckwith-Wiedemann syndrome is associated with a variety of constitutional molecular 11p15 alterations including imprinting defects of KCNQ1OT1. Hum Mol Genet **10:**2989-3000.
102. **West, A. G., M. Gaszner, and G. Felsenfeld.** 2002. Insulators: many functions, many mechanisms. Genes Dev **16:**271-88.
103. **Wilkins, J. F., and D. Haig.** 2003. What good is genomic imprinting: the function of parent-specific gene expression. Nat Rev Genet **4:**359-68.
104. **Yusufzai, T. M., H. Tagami, Y. Nakatani, and G. Felsenfeld.** 2004. CTCF tethers an insulator to subnuclear sites, suggesting shared insulator mechanisms across species. Mol Cell **13:**291-8.
105. **Zhang, Y., and D. Reinberg.** 2001. Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails. Genes Dev **15:**2343-60.

Acta Universitatis Upsaliensis

Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology

Editor: The Dean of the Faculty of Science and Technology

A doctoral dissertation from the Faculty of Science and Technology, Uppsala University, is usually a summary of a number of papers. A few copies of the complete dissertation are kept at major Swedish research libraries, while the summary alone is distributed internationally through the series *Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology*. (Prior to October, 1993, the series was published under the title "Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science".)

Distribution:
Uppsala University Library
Box 510, SE-751 20 Uppsala, Sweden
www.uu.se, acta@ub.uu.se

ISSN 1104-232X
ISBN 91-554-5952-8