The Epigenetics of Gene Transcription and Higher Order Chromatin Conformation

VIJAY KUMAR TIWARI
Dissertation presented at Uppsala University to be publicly examined in Lindahlsalen, Evolutionary Biology Centre, Norbyvägen 18, Uppsala, Friday, February 10, 2006 at 13:00 for the degree of Doctor of Philosophy. The examination will be conducted in English.

Abstract

It is becoming increasingly clear that long-range control of gene expression is mediated through direct physical interactions between genes and regulatory elements, either intra- or interchromosomally. In addition to transcriptional initiation, formation of active chromatin hubs seem to be crucial for increased transcriptional efficiency as well as insulation from neighbouring heterochromatin environment. Regulatory factors apparently have an important role in organization of such functional modules in a development and differentiated-dependent fashion.

The relevance of trans-acting factors in the ‘choice’ process of X-Chromosome Inactivation (XCI) was highlighted by our observations where CTCF was shown to occupy a homologous position on the active mouse and human Xist/Mist promoters and its binding affinity was altered in familial cases of opposite skewed X-inactivation patterns.

The paradigm of genomic imprinting, i.e. the Igf2-H19 locus, manifests its imprinted states through the H19 Imprinting Control Region (ICR). The repression of the maternal Igf2 allele depends on the insulator properties of the H19 ICR when this interacts with CTCF.

The studies here detected a novel kind of CTCF-dependent tightly closed pocket- like higher order structure exclusively on maternal allele which was found to be essential for imprinted Igf2 expression as well as maintenance of precise epigenetic marks at various Differentially Methylated Regions (DMRs) across this locus.

Despite the highly condensed state of the mitotic chromosome, the insulator protein CTCF was found to constitutively occupy its known target sites. Furthermore, pivotal CTCF-dependent long-range regulatory loops within Igf2-H19 locus were found to survive mitotic compaction and such mechanisms might serve as a novel kind of epigenetic memory to minimize transcriptional chaos and to reset proper expression domains in the daughter cells as soon as cells exit mitosis.

Our observations also suggest that the epigenetic reprogramming of H19 ICR during spermatogenesis is initiated by a CTCF-dependent recruitment of chromatin remodeling factor Lsh to the H19 ICR followed by completion of the imprint acquisition process by a replacement of CTCF with its closely related parologue termed BORIS.

Overall, this thesis unravels the novel roles for CTCF as an architectural factor in the organization of higher order chromatin conformations and transcriptional regulation.

Keywords: DNA-protein interactions, Nuclear organization, Transcriptional regulation, Higher order chromatin conformations, Epigenetic reprogramming

Vijay Kumar Tiwari, Department of Physiology and Developmental Biology, Animal Development and Genetics, Norbyvägen 18 A, Uppsala University, SE-75236 Uppsala, Sweden

© Vijay Kumar Tiwari 2006

ISSN 1651-6214
ISBN 91-554-6444-0
urn:nbn:se:uu:diva-6302 (http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-6302)
To My Parents
List of papers:

This thesis is based on following papers, which will be refereed to by their Roman numerals:


II Kurukuti S #, Tiwari VK #, Tavosidana G #, Pugacheva E, Murrell A, Zhao Z, Lobanenkov V, Reik W and Ohlsson R: CTCF binding at the H19 ICR imparts maternally-inherited higher order chromatin conformations and regional epigenotypes to restrict enhancer access to Igf2 Under revision in PNAS (# = shared first authors)


Other relevant reading:

Paper I and III were reproduced by the kind permission of the respective publishers-Oxford University Press and Nature Publishing Group.
CONTENTS:

ABBREVIATIONS 8

INTRODUCTION 9

1. Conventional models of gene regulation and novel perspectives 9
   1.1. Linear view of transcriptional regulation 11
   1.1.1. Promoter-Enhancer communication 11
   1.1.2. Mechanisms of transcriptional induction by enhancers 14
   1.1.3. Insulator action 14
   1.1.4. Regulating insulator activity 16
   1.2. Long-range chromatin interactions: regulation in clusters 16
       1.2.1. The active chromatin hub 18
   1.3. Interchromosomal interactions: gene regulation in three dimensions 22
   1.4. Nuclear organization and multilayered gene regulatory mechanisms 24
       1.4.1. Transcription and replication factories 24
       1.4.2. Chromosome positioning and nuclear localization 26
       1.4.3. Chromosome Territory 28
       1.4.4. Chromatin fluidity 30
       1.4.5. Compartmentalization of transcription in the nucleus 33

2. The molecular determinants of DNA looping 34

3. Epigenetic control of gene expression 36

4. Maintenance of epigenetic states 38

5. Genomic imprinting 40

6. X-chromosome inactivation 42

7. Epigenetic reprogramming in development 45
   7.1. Reprogramming in early embryos 45
   7.2. Reprogramming in germ cells 46
   7.3. CTCF, BORIS, Lsh and epigenetic reprogramming during male gametogenesis 48

8. Long-range chromatin interactions and disease 49

AIMS OF THE PRESENT STUDY 51

RESULTS AND DISCUSSION 52

CONCLUDING REMARKS 64

ACKNOWLEDGEMENTS 66

REFERENCES 68
**ABBREVIATIONS:**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>3C</td>
<td>Chromosome Conformation Capture</td>
</tr>
<tr>
<td>ACH</td>
<td>Active Chromatin Hub</td>
</tr>
<tr>
<td>BORIS</td>
<td>Brother of the regulator of imprinted sites</td>
</tr>
<tr>
<td>ChIP</td>
<td>Chromatin immunoprecipitation</td>
</tr>
<tr>
<td>CTCF</td>
<td>CCCTC binding factor</td>
</tr>
<tr>
<td>CT</td>
<td>Chromosome Territory</td>
</tr>
<tr>
<td>DMD</td>
<td>Differentially methylated domain</td>
</tr>
<tr>
<td>Dnmt/DNMT</td>
<td>DNA methyltransferase</td>
</tr>
<tr>
<td>DMR</td>
<td>Differentially methylated region</td>
</tr>
<tr>
<td>HAT</td>
<td>Histone acetyl transferase</td>
</tr>
<tr>
<td>Hbb</td>
<td>Beba-globin gene</td>
</tr>
<tr>
<td>HDAC</td>
<td>Histone deacetylase</td>
</tr>
<tr>
<td>HP1</td>
<td>Heterochromatin Protein 1</td>
</tr>
<tr>
<td>HS</td>
<td>Hypersensitive site</td>
</tr>
<tr>
<td>IC</td>
<td>Interchromatin Compartment</td>
</tr>
<tr>
<td>ICD</td>
<td>Interchromosome Domain</td>
</tr>
<tr>
<td>ICM</td>
<td>Inner Cell Mass</td>
</tr>
<tr>
<td>ICR</td>
<td>Imprinting control region</td>
</tr>
<tr>
<td>Igf2/IGF2</td>
<td>Insulin-like growth factor 2</td>
</tr>
<tr>
<td>LCR</td>
<td>Locus Control Region</td>
</tr>
<tr>
<td>LOI</td>
<td>Loss of Imprinting</td>
</tr>
<tr>
<td>Lsh</td>
<td>Lymphoid-specific Helicase</td>
</tr>
<tr>
<td>MARs</td>
<td>Matrix Attachment Regions</td>
</tr>
<tr>
<td>PARP</td>
<td>Poly (ADPribosyl) polymerase</td>
</tr>
<tr>
<td>PGC</td>
<td>Primordial Germ Cells</td>
</tr>
<tr>
<td>RNA-TRAP</td>
<td>RNA-Tagging and Recovery of Associated Proteins</td>
</tr>
<tr>
<td>TE</td>
<td>Trophoectoderm</td>
</tr>
<tr>
<td>XCI</td>
<td>X-Chromosome Inactivation</td>
</tr>
</tbody>
</table>
INTRODUCTION:
Cells transmit information from one generation to another via two distinct routes, genetic and epigenetic. While the genetic inheritance is based on the DNA code, epigenetic information comprises modifications occurring directly on DNA or on the chromatin, which does not involve changes in DNA sequence. A multicellular organism starts its life from single fertilized egg which differentiates into many different cell types. The later process depends solely on the precise regulation of differential gene expression. The regulation of gene expression can take place at different stages, from initiation of transcription to post-translational mechanisms. The major kind of gene regulation differs for different genes, for examples, for one gene mRNA stability might play a role in deciding its levels while for others, rate of transcriptional initiation as such may be crucial. Recently DNA looping and higher order chromatin organization has emerged as key regulators of gene transcription.

1. Conventional models of gene regulation and novel perspectives:
For the last few years, our understanding of gene regulation has moved from an initial notion of one-dimensional array of regulatory elements next to each other on the same thread of DNA as the gene to an appreciation that genes are associated with group of proteins, forming multimolecular complexes that are arranged in structures generically known as chromatin. Subsequent findings about long-range interactions, existence of several regulatory sub-loops within the same locus, formation of active chromatin hubs and creation of functional expression domains introduced another dimension onto the scene. With the recent evidences, it is has become apparent that long-range control of gene expression is mediated through direct physical interactions between genes and regulatory elements, either intra- or interchromosomally which themselves are regulated during the different stages of development and differentiation. Beside transcriptional initiation, formation of active chromatin hubs seem to be crucial for increased transcriptional efficiency as well as insulation from neighbouring heterochromatic environment. Regulatory factors apparently have an important role in organization of such functional modules in a development and differentiated-dependent fashion. A sophisticated picture of gene regulation emerges in our mind when we combine the available information about the multiple ways in which gene transcription can be regulated with the nuclear organization in general, for example, significance of chromosome territories, replication and transcription factories, radial and relative positioning of chromosomes, the inheritance of mitotic chromosome positioning and nuclear compartmentalization e.g., existence of functionally dedicated nuclear bodies (like Cajal bodies, IGCs, PML bodies etc). This reinforces us to think that precise gene regulation is outcome of interplay between multiple forces and all of them might act together, in various combinations or in a step-by-step manner to ensure proper reading of the genetic and epigenetic information in terms of gene expression at correct timing. This all envisages that nucleus itself has evolved different strategies for precise regulation of gene transcription during the process of differentiation and development. The picture below depicts the gradual evolution in the concepts of regulation of gene transcription which are discussed in considerable details subsequently (Fig 1).
Figure 1. Novel concepts in transcriptional regulation

(A) The simplest model for regulation suggests that the cis-acting enhancers act by simple looping over to reach promoters and induce transcription in one-to-one action manner. P stands for promoter and E for Enhancer.

(B) Active Chromatin Hubs: Multiple cis-regulatory elements join together in proper spatial orientation to create a functionally dedicated compartment which ensures proper gene expression in a development and differentiation-specific manner.

(C) Interchromosomal interactions: the genes on different chromosomes that are coregulated within a common cell lineage are brought together in close physical proximity in the nucleus and this trans-association is decisive for their expression patterns. Thus, chromatin hub-like structures could be composed of partners from different chromosomes as well.

(D) Multilayered gene regulatory mechanisms: multiple regulatory players operate in various combinations to set up proper gene expression profile during development and differentiation. (See text for details)

1.1. Linear view of transcriptional regulation:

1.1.1 Promoter-Enhancer communication:
Enhancers are distally located cis-acting elements that stimulate the activity of promoters over large distances [1]. Enhancers can be located upstream or downstream of the target gene and in some cases even within the gene, usually in an intronic region and this activity is orientation independent. One of the first beautiful demonstrations of enhancer action was done by Schaffner and coworkers who demonstrated in vitro that enhancers can stimulate transcription in trans, by coupling an enhancer to a promoter containing plasmid via a biotin-streptavidin bridge [2]. Similar activation was observed when enhancer and promoter containing plasmids were injected as intertwined catenates into frog oocytes [3].

Various models have been proposed to explain how enhancers can activate promoters from considerably large distances up to several kb [Fig 1(A), Fig 2]. The
“looping model” states that transcriptional activators bind to the enhancer, followed by a looping of the DNA to get in touch with the promoter to interact with the transcription machinery [Fig 2(ii)]. In the other model, known as the “tracking model” (also called “scanning” or “entry inside” model), factors localize at the enhancer and utilise it as a starting point [Fig 2(iii)]. These factors then track along the DNA until they find a promoter region. Besides these, two more models have been proposed to explain enhancer-mediated gene activation. The “facilitated tracking” or “looping tracking” model is a combination of the first two models, suggesting that the enhancer-bound factors loops out (or tracks) several times until it reaches the promoter. The “linking model” suggests the establishment of modified chromatin domains between the enhancer and the promoter by facilitator proteins which generate a progressive chain of higher order complexes along the chromatin fibre and thereby establishing communication with the promoter without enhancer communicating directly with promoter [Fig 2(iv)]. Most likely several enhancer-mediated activation mechanisms co-exist, and both independently and in combination, they activate genes. Despite the presence of these models, it has been difficult to fully explain gene activation mechanisms based on any of these models alone. Evidences exit to indicate that different genes might use different mechanisms to stimulate gene transcription.

Figure 2. Different models for enhancer action

Recently the looping model of LCR (locus control region)-promoter interactions was confirmed within the β-globin locus in vivo by the method called 3C (Chromosome Conformation Capture) which calculates the physical proximity of two given DNA segments based on the frequency of formaldehyde induced cross-linking between them[4]. The results show that the whole LCR is in close proximity with the transcriptionally active promoters, but not to the spacer DNA, which loops out. A clear link to direct interaction between LCR and transcribed genes was obtained using a modified RNA FISH method called RNA TRAP [5]. The data suggest that a hypersensitive site of LCR is positioned over
a distance less than 100–300 nm from transcribing regions of the *Hbb-b1* and *Hbb-b2* globin genes.

Another study strongly supported tracking along the DNA [7]. In this study chromatin immunopurification was used to show that that factors associate with the enhancer, subsequently leave it and track along the DNA to reach the promoter *in vivo*.

An elegant study carried out on HNF-4α promoter-enhancer further supported the facilitated-tracking or looping-tracking models [6]. The researchers studied the order of recruitment of factors to the HNF-4α regulatory regions upon the initial activation of the gene during enterocyte differentiation. Initially, an independent assembly of regulatory complexes at the proximal promoter and the upstream enhancer regions was observed which was subsequently followed by the tracking of the entire DNA-protein complex formed on the enhancer along the intervening DNA until it encountered the proximal promoter. This movement correlated with a unidirectional spreading of histone hyperacetylation.

The properties of *Drosophila* CHIP protein provided support for linking (spreading–looping) model. Interestingly, CHIP cannot bind to DNA directly, but can interact with numerous transcription factors and facilitate their action over a distance in vivo [8, 9]. It was proposed that CHIP is recruited by an activator protein bound at an enhancer; which then works as a protein "bridge" between the activator bound at the enhancer and other proteins having multiple weak/non-binding sites between the enhancer and promoter [10, 11]. As a result, a wave of small protein-stabilized chromatin loops is initiated at the enhancer and moves towards the promoter.

**DNA supercoiling and enhancer-promoter communication:**

Extensive studies in prokaryotes led to the proposal of two mechanisms to explain how DNA supercoiling can facilitate enhancer-promoter communication over large distances: “Slithering” and “Branch Collision” [12, 13] (Fig 3). The ‘slithering model’ suggests that intertwined DNA helixes can slide relative to each other on super coiled DNA at a high rate and this sliding greatly increases the probability of enhancer-promoter getting in close proximity. Alternatively, according to ‘branch collision model’, frequent collisions between the branches formed on super coiled DNA could facilitate communication between the enhancer and promoter localized on different branches of the same DNA molecule. The figure 3 depicts these models where enhancer and promoter are indicated by black and white boxes.

Based on computer simulations of Brownian dynamics of super coiled DNA, it has been proposed that slithering is the predominant mechanism facilitating enhancer-promoter communication over distances up to 10 kb [14]. This is yet to be tested experimentally. Does DNA supercoiling facilitate enhancer–promoter communication in eukaryotes? The bulk of a eukaryotic genome does not contain unconstrained DNA supercoiling [15] and there is no enzyme (a gyrase) to generate it. However, there are several ways including nucleosome displacement or remodeling, transcript elongation [16, 17, 18, and 19] to generate negative supercoiling in eukaryotes. In support of this, transcriptionally active regions have been shown to contain considerable levels of unconstrained DNA supercoiling [20]. Moreover, DNA supercoiling is needed for the action of eukaryotic transcriptional and recombination enhancers over a large distance in *vitro* [21, 22].
1.1.2. Mechanisms of Transcriptional induction by enhancers:
An enhancer recruits activators and/or RNA Pol II complexes onto it and when it physically communicates with promoters, delivers those proteins to promoter elements initiating transcription [24]. The enhancer action might facilitate recruitment of promoter binding proteins indirectly. Alternatively, the enhancer-promoter interaction might result in a conformational change and/or activity of a protein already bound at the promoter. Independently or cooperatively to above mentioned speculations, this interaction between the protein occupied enhancers and promoters may result in recruitment of other proteins such as general transcription factors, histone acetyl transferases, chromatin remodelers, and RNA Pol II which subsequently start the real play for transcription induction [25]. An enhancer might also lead to relocation of a promoter away from repressive environment or it itself can organize an active compartment (for example, active chromatin hubs, discussed later), where the promoter is recruited for activation. Recent evidences indicate that enhancers might organize some kind of active compartment dedicated to Pol II transcription where the promoters are recruited in a regulated fashion for activation [26]. A crude possibility would be that enhancers acquire their full activating function/potential after they have encountered the correct competent promoter (ascertained by specific protein-protein and DNA protein interactions). The full potential of a promoter would mean its epigenetic state in terms of chromatin modifications, the protein profile around it etc. Nevertheless, the final outcome of promoter-enhancer interaction is transcriptional induction above basal levels through overcoming rate-liming steps that can be different at different promoters. There are data suggesting that the level of transcriptional activation by enhancers is determined by the frequency of enhancer-promoter communication, the stability of this interaction and in case the enhancer is shared by multiple promoters, their competence and strength of interaction. [27, 28, 29, 30, 31].

1.1.3. Insulator action
Insulators are specific DNA sequences that after binding of the corresponding proteins have two primary functional activities. (1) Position-dependent enhancer blocking: insulators block enhancer action only when placed between an enhancer and a promoter but not upstream or downstream of an enhancer–promoter pair [32, 33, and 34]. Enhancers can work over a large distance from target promoters; insulators can also block enhancer activity being placed over various distances from the promoter [35]. (2) The ability to form chromatin boundaries and therefore confer position-independent transcription to transgenes.
stably integrated in the genome [36, 37, 38, and 39]. A locus flanked by insulators functions as an independent unit of gene expression being isolated from both positive and negative effects of the surrounding chromatin as a result of the boundary and enhancer-blocking activity of insulators. The observation that two insulators flanking an enhancer or promoter are required to achieve full repression of stable transgenes suggest that the insulators might functionally interact forming a chromatin loop [40]. Such looping can also potentially interfere with other regulatory interactions across the locus. In vivo experiments supported this notion that insulators may modulate enhancer-promoter interaction by interacting with each other and facilitating the formation of chromatin loops [41, 42]. Blanton J et al provided support to the looped domain model of “boundary” action where they are thought to delimit structural domains by interacting with each other or with some other nuclear structure. The authors showed that Drosophila scs and scs’ boundary proteins, ZW5 and BEAF, respectively, interact with each other bringing the former two elements in close proximity [43]. On the other hand, a single insulator can block enhancer action on linear DNA in some cases suggesting that insulator-induced loop formation is not essential for the repression and that insulators could block a tracking component of enhancer action [44]. Finally, some insulators can functionally interact with promoters [45], suggesting that ‘promoter decoy’ kind of mechanisms might exist in some cases. All these observations are consistent with the looping–scanning mechanism of enhancer action.

For several years we had fun with the definition of insulators in connection with enhancer-blocking and heterochromatin barrier activities but the recent observations suggest that sometimes these activities could be separable. A good example is β-globin locus where enhancer-blocking and barrier activity has been attributed to two different proteins namely, CTCF and USF respectively whose binding site lies within the 5’ HS4 insulator element [46]. Here USF mediates recruitment of enzymes that acetylate histones and methylates H3K4 in the immediate vicinity and confers barrier activity to these elements. This was the first evidence that a vertebrate insulator element was shown to function at the level of histone code to resist the encroachment of transcriptionally repressive chromosomal condensation process. Drosophila SF1 insulator element has both activities separable as well [47]. There are other examples where CTCF has only enhancer-blocking activities and no barrier activity [48, 49]. If we talk about separable activities of enhancer-blocking and barrier function, there might be following ways in which these insulator elements execute these functions:

A. Enhancer-blocking activity:

Based on the literature available so far, there could be different possible ways in which insulators could function as enhancer-blockers. The ultimate goal of enhancer blocking activity is to somehow interfere with the message of transcription activation coming from an enhancer to reach the promoter. It can physically block the enhancer-promoter communication by interfering with the loop formation and/or tracking of factors required for transcriptional activation. Alternatively, active insulators can set up new independent loops to separate promoter and enhancer in two domains which are inaccessible to each other [50]. The other model of insulator action proposed is physical tethering to subnuclear structures such as nuclear periphery (in case of the gypsy element from Drosophila) [51], nucleolar periphery (in case of the chicken HS4 element) [52]. The new loops that might be formed in this way would dissect the promoter-enhancer communication possibly due to torsional constraints or steric interference etc. The other possibility is that insulator can recruit some proteins which actively act to disrupt the enhancer-promoter connection. This could also happen by recruitment of some chromatin
modifying proteins leading to formation of chromatin non-permissive for flexibility and competence required for enhancer to activate promoter [53].

B. Barrier activity:
Insulators were first discovered as elements needed to block the spread of heterochromatin into active chromatin and thus acting as borders. The studies from yeast to vertebrates have indicated several ways in which these elements might enforce their barrier activity. It could similarly happen by physical tethering to nuclear substructures such as nucleophosmin or nuclear membrane as discussed before which would physically obstruct the spread of heterochromatin due to physical constraints. They could also create nucleosome displacement and gap thereby altering the histones that the heterochromatin factors need to act upon [54]. Alternatively, insulators can compete with heterochromatin propagation by directly binding to histones (nucleosome masking). Here, a transcription factor CTF-1 interacts directly with histone H3 and this interaction confers barrier activity [55]. The histone-binding domain of the protein was shown to act as barrier against the propagation of a SIR-containing histone-hypoacetylated, silent chromatin structure from telomere. The insulator can recruit chromatin modifying enzymes and lead to antagonistic modifications as to the approaching heterochromatin marks e.g. USF is responsible for recruiting a number of histone modifying enzymes that are responsible for modification profile of the HS4 element and removal of USF binding site or knock-down of USF leads to loss of HS4 barrier activity [46].

1.1.4. Regulating insulator activity:
Recent evidences suggested various ways in which insulator activity could be regulated:
1. Direct methylation of an insulator element leads to loss of its activity in some cases (e.g. CTCF binding at H19 ICR is methylation-sensitive) [56, 57].
2. The activity of insulator binding proteins could be sensitive to post-translation modifications (e.g. CTCF is shown to interact with PARP and gets poly(ADPribosyl)ated and this is shown to regulate its chromatin insulation activity [58].
3. Certain co-factors might modulate the activity of insulator protein [e.g. the activity of CTCF binding elements can be modulated by neighbouring binding sites for thyroid hormone receptor (TR)] [59]. The addition of thyroid hormone led to abrogation of the enhancer-blocking activity without losing CTCF from chromatin. In addition to this, a recent study by Gilson and colleagues discovered that a POZ-zinc transcription factor Kaiso interacts with CTCF and this binding reduces the enhancer-blocking activity of CTCF in an insulation assay [60].
4. Pairing between insulator elements may be a common mechanism of mediating “insulator bypass” of those insulator elements which function by tethering to subnuclear foci (for instance, two copies of the gypsy insulator positioned in tandem neutralize one another’s enhancer blocking activity possibly by pairing to each other by means of proteins like GAGA and mod (mdg4) [51, 61].

1.2. Long-range chromatin interactions-regulation in clusters:
The connection between higher order chromatin architecture and transcription regulation has become clear recently with impressive evidences coming from studies of β-globin locus. In this case direct physical interactions and looping have been demonstrated, where a remote enhancer interacts with the globin gene during transcription alongwith certain other elements in the locus. This was also one of the first clear demonstrations that clustering of regulatory elements is key to creating and maintaining active domains and regulating transcription [62, 63]. Point to be noted is that out of nine distal regulatory elements that
seems to cluster around active Hbb genes, only one is known classical ‘enhancer’ and other sequence elements involved have insulator like properties, intragenic promoter activity or unknown functions. They coalesce into what has been called an ‘active chromatin hub’ (discussed later in detail).

Long-range looping interactions have also been demonstrated in imprinted and cytokine gene clusters, and between chromatin boundary elements. In case of Igf2 - H19 imprinted locus, although Banerjee and Smallwood several years ago proposed how higher order looping could be involved in gene regulation here, it did not receive much attention because of lack of strong experimental evidences [64]. Recently Murrell et al used a GAL4 knock-in approach as well as the chromosome conformation capture (3C) technique to show that the differentially methylated regions (DMRs) in the imprinted genes Igf2 and H19 interact in a parent-of-origin specific manner in mice. These interactions partition maternal and paternal chromatin into distinct loops. This generates a simple epigenetic switch for Igf2 through which it moves between an active and a silent chromatin domain and these domains are accessible and inaccessible to downstream enhancers respectively. These results, for the first time demonstrated the occurrence of long-range chromatin interactions at an imprinted locus. There are indications of looping and interactions at human IGF2 -H19 locus as well showing that chromatin higher order folding-mediated control of imprinting regulation might be well conserved [65].

![Figure 4. A chromatin-loop model for Igf2/H19 imprinting proposed by Murrell et al. The epigenetic status of all the DMRs (methylated, filled lollipops; unmethylated, open lollipops) is shown on both maternal and paternal alleles. On maternal allele, the loop resulting from H19 ICR- Igf2 DMR1 interaction places the Igf2 promoters in the silent compartment (hypothetical) whereas the switching of the same loop to H19 ICR-DMR2 on paternal allele allows enhancer access to the Igf2 promoters driving its expression.](image-url)
Flavell RA and colleagues discovered the formation of a pre-ACH (active chromatin hub, discussed later in detail) kind of structure in Th2 cytokine locus consisting of elements in the promoter region in non-expressing cell types and naïve T-cells prior to transcriptional activation of the cytokine genes [66]. The LCR became part of this structure as the cells differentiate into different cell types which express cytokine genes. The transcription factors GATA 3 and STAT 6 were shown to be essential for the establishment and/or maintenance of these interactions.

Rett syndrome (RTT) is an X-linked dominant neurodevelopmental disorder affecting almost exclusively girls. Although mutations in methyl-CpG-binding protein (MeCP2) are strongly correlated with RTT, gene expression patterns are not significantly altered in MeCP2-deficient cells. A recent study by Horike et al identified MeCP2-mediated histone modification and formation of a higher order chromatin loop structure specifically associated with silent chromatin at the Dlx5–Dlx6 locus in normal cells, and its absence thereof in RTT patients where, instead Dlx5-Dlx6 interacted with far distant sequences, forming active chromatin associated loop[67]. The altered expression of Dlx5 observed here through loss of silent chromatin loop formation provides a molecular mechanism underlying RTT. In addition, this study provided a novel role for MeCP2 in higher order chromatin organization and imprinting. These observations also linked Rett syndrome (RTT) with genomic imprinting by identifying a maternally expressed gene, DLX5, as a target for MeCP2. LOI for DLX5 was found in there of four LCLs (Lymphoblastoid cell lines) isolated from individuals with RTT. The same was true in mouse model for RTT.

1.2.1. The active chromatin hub (ACH):

The thoughts about how genomes manage to set up independent expression profiles of individual genes originally came from electron miCRoscopy studies of lampbrush chromosomes some 50 years back showing that genome were structurally organized in large loops of varying sizes [68]. This led to the speculation that such structures represent functional domains of specific gene expression. The expression domains containing a gene or gene cluster with all its cis regulatory elements needs to be independent of influence of other such nearby domains. Insulators or boundary elements were proposed to be the ‘genomic guardians’ setting up independent expression modules and preventing transcriptional chaos.

But above all, there are numerous tissue-specific genes that do not occupy a physically separate chromosomal domain (or ‘functional expression module’). The human α-globin gene cluster is a classical example of a locus showing autonomous expression without occupying a physically isolated entity chromosome16. It lies immediately adjacent to several ubiquitously expressed genes and its major cis-regulatory element resides in an intron of one of these housekeeping genes, 40 kb away from the α-globin genes (Figure 5 below) [69]. There are several other examples of similar occurrence, some of which are indicated below in the figure. They include chicken lysozyme locus, mouse β-globin cluster, human β-globin cluster and human growth hormone cluster.

Several years ago, it was shown that that multiple combinations of cis-regulatory elements like LCRs (Locus Control Regions) can confer position-independent and copy-number-dependent expression on a linked transgene in transgenic mice [70]. Thus, there is no strict requirement for elements defined as insulators to be at the borders of position independent expression modules.

Using the Chromosome Conformation Capture (3C) technique, Tolhuis et al found that the HS of LCR, located 40-60 kb away interacts specifically with the active (expressing) distal β-globin genes (β-major and β-minor) in fetal liver with intervening
Figure 5. Overlapping gene expression modules. Genomic organizations of overlapping gene expression modules are schematically presented. Note that the scale in kilobases is indicated below each genomic fragment. The arrows represent hypersensitive sites (figure adapted from Vyas P et al. 1992).

DNA containing the inactive genes [embryonic β-globin genes (εγ and βh1)], positioned in between the LCR and adult genes, looping out. Similar results, obtained by completely different methodology, were also reported by others [62]. This group developed RNA TRAP, which involves targeting of horseradish peroxidase (HRP)-labeled probes to nascent RNA transcripts, followed by quantitation of HRP-catalyzed biotin deposition on chromatin nearby. Tolhuis et al also found that spatial clustering with the active genes is not confined to the cis-regulatory elements of the LCR but it also includes two sets of HS at either side of the locus, 130 kb apart from each other. This includes 3'HS1 and two additional 5'HS (called 5'HS-60.7 and 5'HS-62.5) with previously unknown function and might have a regulatory role. Importantly, the transcriptionally silent OR (Olfactory Receptor) genes located in between the 50HS and the LCR, like the inactive embryonic globin genes, do not participate in clustering, but loop out. They also found that β-globin locus adopts a linear confirmation in the nucleus of brain cells where all genes are inactive [4].

Similar analysis was done by Palstra et al in primitive erythroid cells in blood from embryos at 10.5 d post-coitum (d.p.c.), which predominantly express embryonic globin genes εγ and βh1. The results indicated that this clustering or core Chromatin hub (CH) [where active gene(s) are not part of it)] seems to be conserved between primitive and definitive erythroid cells, two developmentally different types of cells that express a different subset of globin genes. This core CH includes the two hypersensitive sites at −60 kb, all hypersensitive sites of the LCR and 3' HS1. There is a developmental switch and contacts between individual globin genes and the core CH of mouse β-globin locus is established which corresponded well with the switching in expression patterns. Thus, here a pre-set CH seem to contacted by different globin genes during development and differentiation forming Active Chromatin Hub (ACH) depending on the cell type. This structural change correlates well with the developmental switch in
expression of the genes[26]. They next analyzed the conformation of the human β-globin locus at different stages of development. The mouse and human β-globin gene loci have a high degree of nucleotide sequence conservation, particularly at regulatory regions. The results obtained with transgenic human β-globin gene locus showed formation of a core ‘chromatin hub’ (CH), similar to mouse, consisting of 3’HS and HS at LCR, which was unchanged in primitive and erythroid cells. The structural changes observed here in CH were related to positioning of genes relative to core CH, depending on which gene(s) join core CH, forming ‘active chromatin hub’ (ACH) and this correlated very well with the transcriptional activity of those genes. Overall, these results indicate that the overall spatial organization of the β-globin gene cluster is conserved from mice to humans.

Based on the above findings and further experiments on in vitro differentiation of 1/11 erythroid progenitor cells led researchers draw few conclusive models. The β-globin locus is in transcriptionally poised state in progenitor cells adopting ‘core CH’ configuration as described. On induction of differentiation, contact between gene(s) and the LCRs are established forming ‘active chromatin hub’ (ACH) [see Fig 2 (B), Fig 6] and the β-globin genes are expressed. Overall they strongly suggest that regulatory elements surrounding the β-globin genes in cis create an erythroid- suggest specific developmentally stable nuclear compartment dedicated to RNA polymerase II transcription. RNA pol II density on active β-globin genes was found to be much higher than on most active genes and such chromatin hub kind of compartment formed by chromatin-associated regulatory elements might primarily act to increase the transcription efficiency simply because that would result in high local concentration of DNA binding sites for transcription factors and subsequently recruits them to the active centre [71, 72]. These factors might include chromatin modifying enzymes such as histone acetylases (HATs), Chromatin remodeling machinery, general transcription activators, tissue specific transcription factors etc and all of them together at an elevated concentration boost gene expression much above basal levels. This would create a competent environment for transcription within ACH irrespective of surrounding inactive or heterochromatic features. The ACH need not occupy a fixed position in nucleus; instead it could be a dynamic fluid entity. It could be found outside or inside chromosome territory or in interchromatin domain (ICD) compartment.

Figure 6. Progenitor cells contain a core CH formed due to spatial clustering of cis-regulatory elements without genes being part of it. As cells differentiate, different genes join the cluster for activation depending on cell type, forming ‘active chromatin hub’ (ACH) (figure adapted from Palstra RJ et. al.2003).

As cells differentiate into two different cell types, due to cell-type specific differences, the ACH composition and organization may differ, driving differential gene
expression and thus same LCR or any other regulatory element might boost expression of different genes in different cell types. The same might hold true during development where at different stages, the same locus might form different kinds of active chromatin hubs composed of different constituents.

An ACH is at least composed of two components - an enhancer and a promoter. The interaction productivity between the two will decide the transcriptional outcome. Depending on the nature of neighbouring chromatin, some other elements would have evolved to stabilize promoter-enhancer complex or to help blocking the encroachment of heterochromatin, if there is any. Perhaps, that is why we see additional components being part of the ACH besides promoter and enhancer. Promoters and other regulatory elements may or may not compete for interaction with the ACH, depending on the nature of the element and related factors [e.g. OR promoters do not compete with the b-globin ACH, whereas an SV40 promoter does [73].

How could the spatial clustering of regulatory elements in 3D space we described here prevent encroachment of heterochromatin mediated silencing from nearby chromatin? The high concentration of DNA binding site at ACH would result in high local recruitment of factors like HATs, chromatin remodelling machinery etc. This gain support from observation that all the elements being part of \( \beta \)-globin ACH have high histone acetylation levels [74, 75]. Others have proposed that local hyperacetylation of histones could interfere with linear spreading of heterochromatin modification [76]. The high levels of transcriptionally permissive modifications at ACH could actively protect against repressive neighbourhoods [77]. An alternative mechanism comes from 'nucleosome gap model', which suggests that disruption of a continuous nucleosome array can cause a barrier for the spreading of silent chromatin [78]. This model was also based on the observation that multiple types of \textit{cis}-regulatory elements can form a barrier, and being hypersensitive was argued to be the common denominator crucial for this activity [78]. An alternative possibility would be that the LCR or any other \textit{cis}-regulatory component of ACH occupies specific nuclear positioning by tethering to subnuclear structures like nuclear lamin in case of insulators and this in turn provides a physical autonomy from nearby chromatin profile. This proposal still waits for experimental evidence.

The way in which overlapping genes set up independent expression profiles could also be explained by ACH formation. Since the active genes are part of ACH while inactive ones are not, the transcriptional outcome seems to be dependent on promoter competition for ACH. The experimental evidences also suggest that actual structure of ACH (nature of elements, their orientation etc), distances between promoters and regulatory elements, promoter affinity [79] and gene order are crucial in determining outcome of this competition and in deciding the functional autonomy of adjacent genes. The presence of fetal \( \beta \)-globin genes in between the LCR and the adult globin genes precludes inappropriate activation of the latter in embryonic erythroid cells and this indicates the importance of gene order in this competition. Replacing the fetal \( \beta \)-globin promoter by a spectrin promoter causes premature expression of the distal adult \( \beta \)-globin gene (loss of competition) indicating promoter affinity and specificity are also crucial determinant here [79].

In last few years, results from several labs have given hints on how do regulatory elements come together in the nuclear space to form a 'functional expression module' or an 'ACH'. Three models have been proposed: random collision, tracking and linking. The random collision model proposes that flexibility of the chromatin fibre causes transcription factors bound to distant regulatory sequences to randomly interact with each other and to the surrounding chromatin; affinity between constituents in DNA-protein and protein-protein interactions involved will determine the duration of these interactions and
thereby the loop persistence. Both the tracking and linking model have been discussed before during enhancer-promoter communication discussion. Grosveld and colleagues further investigated the role of most actives sites of the LCR and the promoter of β-globin in the maintenance of the ACH [80]. The β-globin gene promoter deletion in the context of full LCR doesn’t affect the conformation of ACH with the β-globin gene still remaining in proximity of it indicating the transcriptional activity of β-globin genes is not required to organize ACH. Additional deletions of HS2 and HS3 showed that HS3, but not HS2, in combination of β-globin gene promoter is crucial for the maintenance of ACH at the definitive stage. The authors conclude that multiple interactions between regulatory regions (here LCR) and gene(s) (here β-globin gene) are required to stabilize active chromatin hub (ACH) and to maintain its spatial configuration in vivo.

They compared the creation of ACH to that of folding of an enzyme to create an active site “pocket”. Initial efforts using random collision, tracking or linking etc for folding of the locus might be non-specific in nature, but this will provide basic structure required for proper folding or establishing correct spatial organization (active site) of LCR and allow interactions with the genes (which in the analogy is the substrate). The complex formed by hypersensitive sites, such as seen by Palstra et al. (2003) in precursor cells, was referred to as “chromatin hub” (CH), whereas the complex involving hypersensitive sites and active genes (i.e., the flanking sites, the LCR, and the genes) is named the “active chromatin hub” (ACH). When different sets of genes are present in a locus, they may use the same or different flanking sites to initiate the “folding” process, followed by the interaction of specific sequences (like the globin LCR HS1–4) to create a compartment, the ACH, to transcribe the different genes in different tissues [81]. Recently, Drissen et al showed that EKLF, an erythroid specific transcription factor which is necessary for adult β-globin gene transcription, is required for progression from the chromatin hub (CH) present in erythroid precursors to a fully active ACH [82]. Dekker and colleagues recently showed that in mouse, haematopoietic transcription factor GATA-1 and its cofactor FOG-1 are required for the physical interaction between the β-globin major gene promoter and β-globin LCR [83]. GATA-1 binds β-major globin gene promoter in a tissue-specific manner independently of the LCR (loop formation by GATA-1 occurs also in the absence of new protein synthesis) and induces loop formation with concomitant induction of β-globin gene transcription.

1.3 Interchromosomal interactions – gene regulation in three dimensions:

The phenomenon of transvection is well understood in Drosophila but such trans-regulation is very poorly known in mammals [84]. Recently Spilianakis et al. revealed a novel mechanism of gene regulation where they produced the first clear evidence of inter-chromosomal communication in mammals showing that genes on different chromosomes that are destined to be expressed within a common cell lineage are brought in close physical proximity in the nucleus [85].

To explain how genes that are far from each other in the genome, either in cis (on the same chromosome) or trans (on different chromosome), can be coordinated to be expressed together, or to preclude the expression of one another, required a leap into another dimension [86]. As well known, the three-dimensional positioning of a gene within the nucleus might influence gene expression pattern e.g. genes present in an area containing repressive or heterochromatin factors are silent whereas the same in an areas containing activating euchromatin proteins are transcriptionally on. How do genes find their appropriate location in the nucleus of a cell, and how are genes that must be expressed herded into active neighbourhoods? The authors discovered that in the naive TH cell, the two gene complexes - TH1 cytokine gene locus (containing the cytokine gene Ifny) and TGF2
cytokine gene locus (containing the cytokine genes Il4, Il5 and Il13) are close together in a region of the nucleus (confirmed by two independent methods of different sensitivity and precision, namely 3C and FISH). On the basis of findings that the T_h2 and Ifnγ loci have a very low percentage localization with heterochromatin regions (HP1 staining) and other subsequent expression analysis, they propose that co-localization of the T_h2 and the Ifnγ loci poises or prepares these loci for their rapid expression upon stimulation rather than holding them in a repressive environment. Upon receiving a specific stimulus, the gene to be activated (for example Ifnγ after a TH1 stimulus) is allowed to begin expression, whereas its counterpart that is to remain silent (in this case the TH2 genes) is moved, presumably to a more repressed region of the nucleus [Fig 1(C)].

Spilianakis et al. provide some evidence that following stimulation, the inter-chromosomal interactions in the naive cell are replaced by intra-chromosomal ones between regulatory elements within the same gene region. Upon differentiation of naive T cells to effector cells the interactions change: in T_h1 cells the interchromosomal association between the Ifnγ gene and the T_h2 cytokine locus is significantly reduced and is apparently replaced by another T_h1-specific interaction between the Ifnγ gene and Ifnγ CNS2, a tissue-specific transcriptional enhancer. Thus the interchromosomal interaction somehow seems to make the Ifnγ gene promoter competent for transcription and therefore soon after the stimulus, the newly established intrachromosomal contacts between the Ifnγ gene and Ifnγ CNS2 could drive expression of Ifnγ gene quickly. Moreover, the specific elements involved in inter-chromosomal interactions does seem to regulate to some extent the expression of the genes involved, as deletion of a regulatory element (RHS7) from the TH2 LCR on chromosome 11 disturbs the interchromosomal associations, and this is coupled with delayed expression kinetics of Ifnγ from chromosome 10. Apparently, a very interesting finding of this study was the observations that interchromosomal interactions are found specifically in naive T cells which have the potential to transcribe both Ifnγ and TH2 cytokine genes upon activation. On differentiation, they form cell types expressing either Ifnγ or T_h2 cytokine genes. They lose the potential to express both concomitant with the loss of interchromosomal contact between the loci and replaced by new intrachromosomal interactions along with the restricted expression of either set of genes.

This strengthens the possibility that dynamic intra-and interchromosomal interactions between certain genetic loci might regulate their transcriptional activation or silencing. In addition, LCRs can help to regulate the expression of co-ordinately regulated genes by regulating interchromosomal interactions. It will be interesting to explore such interchromosomal interactions in other co-ordinately regulated genes to conclude whether it is a common feature in the nucleus of higher organisms. Genome has significant clustering of coregulated genes and that might help such interchromosomal interactions to operate easily. Versteeg and coworkers showed that genes that are highly expressed in a large number of different human cell types are strongly clustered on a number of chromosomes [87]. Some of such clusters might correspond to chromatin that forms giant loops bulging out of chromosome territories. A clear illustration of a large cluster of highly transcribed genes that forms a loop structure has been described by [88]. If similar gene clusters occur on another chromosome and are coregulated, they might come together due to similar cis-regulatory element(s) they have, which might lead to accumulation of high concentration of relevant transcription factors.

Observation made with electron microscopy state that although most often the chromosomes are observed as individual units with distinct borders, decondensed chromatin from neighbouring territories can be seen in contact in limited regions [89]. This might be the result of interchromosomal communication between chromosomally separated loci, or the genes from them happen to share the same protein.
factory (for example, a heterochromatin body like HP1 or an euchromatic body, a nuclear body like Cajal body, transcription factory etc). Alternatively, they might be seen in contact by default of the ongoing independent processes. The interchromosomal interactions have implications to explain preferential chromosome neighbourhoods and perhaps thereby preferred translocations.

1.4 Nuclear organization and multilayered gene regulatory mechanisms:
It has become clear that it is important not to look gene-specific transcription as an individual unit per se; rather multiple players organize this successful orchestra. These players include nuclear localization of a gene within different nuclear subcompartments and with respect to constitutive heterochromatin, its interaction with structural proteins like nuclear lamina, specific ‘radial’ and ‘relative’ chromosome positioning, positioning of the gene with respect to chromosome territories, existence of transcription and replication factories, presence of functionally dedicated nuclear bodies like IGCs (Interchromatin granule clusters), Cajal bodies, nucleolus, PML-NBs (PML nuclear bodies) etc, nuclear matrix and various regulatory intra- and interchromosomal interactions. Most of these players play a role one or the other time during the transcription life cycle of every gene. Few of these have been discussed in previous sections; some more are discussed below.

1.4.1. Transcription and replication factories:
Direct visualization of newly synthesized DNA and pulse labelling studies revealed that DNA synthesis is carried out at specialized replication sites called “replication factories or foci” each of which is shared by several active replicons and these appear in recognizable patterns that persist for hours during S phase in mammalian cells [90, 91, and 92]. Throughout S phase, these factories were found to assemble dynamically. Those replicons which are found in close vicinity to those foci or factories are temporarily recruited to these subcompartments and finally the newly synthesized DNA regresses back to chromatin rich regions, away from the replication site [92, 93]. Within nucleus, there seems to be compartmentalization into subdomains permissive and repressive for replication and perhaps this property is rendered by their constituent factors. Gribnau J et al found that in the Igf2 -H19 locus, early-replicating allele is localized more toward the interior of the nucleus relative to the late-replicating locus [94]. The human β-globin locus replicates late in S phase and associates with pericentric heterochromatin in cells that do not express β-globin. In contrast, replication timing of the locus is shifted in erythrocytes to an earlier point in S phase and the locus localizes away from pericentric heterochromatin [95]. In yeast the late replicating origins were found to be enriched in the nuclear periphery, whereas early-replicating origins were found to be distributed randomly [96]. However, this enrichment of late-replicating origins in the nuclear periphery is lost as cells progress into S phase.

It has also been shown that pol II transcription occurs in discrete foci known as “transcription factories” dispersed throughout the nucleoplasm of mammalian cells, each comprising several Pol II complexes mostly with hyperphosphorylated form of pol II (See figure 7). Labelling of nascent RNA in nuclei produces multiple foci when visualized with light microscopy. However, foci represent collections of nascent RNA around multiple polymerases which constitute those transcription factories. The observations that a gene is only temporarily associated with a polymerase envisages that active Pol II transcription is most likely to constitute a dynamic element of subnuclear organization [97, 98, 99]. This is also supported by observations by Osborne et al where they showed that most active genes are not transcribed continuously; rather they undergo transcription “on-off” cycles [100].
As compared to the localization of RNA pol II, the distribution of GTFs (General transcription factors) is somewhat more variable. Surprisingly, except for TFIIH, little correlation was observed between the localization of GTFs and the localization of RNA pol II LS and transcription sites [101]. In addition to this broad distribution, many transcription-associated proteins also localize to other nuclear domains, such as nuclear speckles, the OPT domain, Cajal bodies, promyelocytic leukemia (PML) nuclear bodies, and heat shock factor 1 (HSF1) granules and is discussed later. The transcription factors studied to date show rapid exchange rates at transcription sites, with residency times ranging from seconds to minutes, supporting a “hit-and-run” model [102] but whether it is factors running and hitting the DNA elements or vice-versa, is not clear.

Keeping in view the existence of immobilized polymerases and transcription factories, a simple model for the transcription cycle for a gene would involve looping out of gene promoters to attach to these factories and bind polymerase complexes, followed by transcription and release of nascent transcript into the interchromosomal compartment where factors are present for RNA processing and transport. The transcription termination is followed by dissociation of these loops away from polymerase factories. There is a possibility that transcription termination factors act by causing destabilization of this interaction or repulsion from these polymerase factories by one or other mechanisms and this needs further experimental verification. Point to be noted is that once a gene promoter has met these factories, the rounds of transcription might vary from gene to gene, depending on several factors like its epigenetic potential, stability and productivity of this interaction, strength of termination etc. But sooner or later, it has to regress back into its chromosome territory and come out again to look for nearby pol II factories to reinitiate transcription. The factors which could determine the rate of transcription reinitiation might depend on factors which decide the proximity and cooperative stability of multiple protein–DNA (e.g., TATA binding proteins) and protein–protein (e.g., TAFs) interactions between the holoenzyme and the promoter. If the gene is part of a big expression domain, multiple cis-regulatory elements (e.g., enhancers and other hypersensitive sites) may cooperate to activate transcription, establishing a dynamic equilibrium that positions the DNA–protein initiation complex closer to the transcription site. An LCR or enhancer in some way might help increasing the probability of attachment of the chromatin loop containing the gene(s) to any of these key transcription cycle interactions. Here, certain locus-specific transcription factors might play a very crucial role as architectural proteins during the precise folding and stability of such chromatin higher order interactions.

Interestingly, few observations made few years ago showed that the number of RNA polymerase II transcription factories in a cell are countable and seem to be limiting in vivo whereas the number of active transcription units in the cell are several thousand at a time [103]. This finding provokes the possibility that several active genes could occupy the same factory in a given unit of time. Indeed, these speculations hold true as a recent report by Osborne et al showed that genes separated by upto 40 Kb on chromosome 7 frequently co-localize to the same transcription factory in a transcription-dependent manner [100] (Fig 7). Their observation that genes fluctuate between transcriptionally ‘on-off’ states suggests that the transcriptionally ‘on’ state apparently correlates with factory occupancy and the ‘off’ state with relocation away from these factories. This reinforces our thoughts about regulation of gene transcription. This observation also envisage that there is possibility of transcriptional regulation by controlling chromatin mobility as well which in turn might decide the frequency and strength of dynamic association of these loci to the transcription factories.
1.4.2. Chromosome positioning and nuclear localization:

Newly developed three dimensional imaging methods are providing new perspectives on nuclear architecture. The positions of the interphase chromosomes show distinctly heritable patterns through cell division [104]. More recent studies, again employing FISH in fixed cells, have focused on two modes of chromosome positions. The first is radial location within the nucleus measured from the centre to the nuclear envelope (called ‘radial positioning’). Most of the evidences so far led to consensus that gene rich chromosomes are preferentially found in the nuclear interior, while gene poor chromosomes mostly localize closer to the nuclear envelope [105, 106, 107]. Radial organization was found to be evolutionarily conserved in primates [108] and the radial arrangement of early and late replicating chromatin domains could be shown to persist over two successive cell generations [104]. The second mode of chromosome positions that has been documented is neighbourhood relationships between two specific chromosomes (called ‘relative positioning’). Here certain chromosomes seem to prefer specific chromosomes as their neighbours more frequently than expected from a random arrangement of chromosomes [109]. This is especially important for the question of homolog pairing and the proximity of chromosomes frequently undergoing translocations [110, 111]. Most interestingly, the physical proximity of chromosomes undergoing translocations in mouse lymphoma cells was found to be conserved in the closely related normal splenocytes [109]. This showed that physical proximities may promote chromosomal rearrangements. Observations from few labs also correlated replication timing and nuclear localization as discussed in previous section as well [112, 113].

The specific nuclear repositioning of genes has been correlated with transcriptional activation and silencing as well. In one study, researchers compared the positions of two closely linked homeobox (Hox) genes — one active, the other inactive — relative to their chromosome territory. The active gene repositioned up to 1 μm away from the inactive gene and its chromosome territory showing that gene expression correlates with chromatin decondensation and increased mobility [114]. The second study, by Zink et al., investigated the nuclear organization of the cystic fibrosis transmembrane conductance
regulator gene (CFTR) and compared it with nearby flanking genes [115]. They showed that CFTR adopted a more interior position when expressed, although the inactive flanking genes remained constitutively attached to heterochromatin at the nuclear periphery. In this case, although the alteration in transcription status led to repositioning of genes, driving this gene into a different environment was not enough to affect its transcription which suggests that transcription activity affected the nuclear positioning and not vice-versa. These are beautiful examples indicating that the nuclear localization of a gene could be individually controlled in a transcription-dependent manner.

Studies from Gerlich et al. concluded that chromosome positions are faithfully transmitted during mitosis and no global rearrangements occurred in interphase [116]. Interactions with the nuclear envelope after mitosis have been proposed to as a possible mechanism for establishing preferred positions [117]. Also the length of pericentric heterochromatin has been correlated to delay sister chromatid separation due to proportionate cohesion binding [118]. Future studies are awaited to uncover novel mechanisms and explanations behind such faithful transmission of radial positioning. What could be the biological function of behind maintaining such chromosome positioning patterns? Gerlich et al speculate that inheritance of chromosome order could function as an epigenetic mechanism to ensure proper gene expression patterns. It is well known that nucleus is functionally divided into zones of active and silent nature comprised of different kinds of factors [119, 120, 121], and such patterns might help regulating activation or silencing events with higher efficiency [122]. The concentration of various general and specific transcription factors, euchromatin and heterochromatin factors as well and RNA pol II (transcription factories) varies in the nucleus and could affect chromatin structure including those of promoters and enhancers and thereby the probability of transcription. Therefore, if a chromosome is positioned in the daughter cell nucleus similar to its mother cell, this will place the genes into similar subnuclear environment; and in that way it might facilitate maintaining the global gene expression pattern in mammalian cells. An interesting observation in this study was that the accuracy of transmission was not found to be 100% and there was some intermixing which might be an indication that although to a large extent the positioning patterns are maintained, the daughter cells will resemble mother cells but they could most likely change chromatin structure within context to allow variation depending on need. Moreover, since related cell types have been shown to have similar patterns of nuclear organization, it is probable that these are established during differentiation or play a role in the process of differentiation [123] and in disease [124, 125, 126, and 110]

It was found that gene activity is probably inherently compartmentalized along the chromosome in humans [87]. This study generated a ‘Human Transcriptome Map’ which revealed that genome is non-randomly organized into regions of high and low gene expression. The regions of high gene expression are separated by large regions of low activity. This may be a strategy in regulation of proper gene expression. This kind of organization is analogous to the finding that lineage-restricted or coregulated genes are being clustered in the genome of the model organisms like mouse, fly and others. There are increasing evidences that genome is non-randomly organized; co-expressed genes have a tendency to be adjacent or grouped along the genome [127]. This might facilitate sharing of regulatory elements like enhancers and insulators, creation of a hub of similar cis-acting sequences which increase the efficiency and specificity of regulation and better operation of development and differentiation-controlled long-range regulatory intrachromosomal interactions.

Several studies have given beautiful examples which relate positioning of a gene with respect to chromosome territory and different nuclear subcompartment with its
transcriptional activity. In a recent study, Anne et al find a distinctive nuclear distribution patterns for two best-studied genes NANOG and OCT-4 involved in pluripotency in human ES cells [128]. Chromosome 12p, a region of the human genome that contains clustered pluripotency genes including NANOG, occupies a more central nuclear position in ES cells than in differentiated cells. For chromosome 6p harbouring Oct-4, they detect a relocalisation of the OCT4 locus, to a position outside its chromosome territory although no overall change in nuclear chromosome position was observed. Therefore it seems that dynamic reorganization of chromatin at higher order level occurs concomitant with differentiation process and the level and method could vary from gene to gene. Understanding such changes globally would enhance our understanding of stem cell biology [128].

A beautiful observation by Brown and colleagues demonstrated the role of long-range organization in differentiated gene expression where they studied the position of several differentiation-regulated genes in immature or mature B cells and found that inactive genes associate with heterochromatic domains, yet this association is only seen in cycling cells. In quiescent cells, association of repressed loci with pericentric heterochromatin is lost, being re-established only after mitogenic activation. Spatial reorganization was also reported for genes repressed during T-lymphocyte activation (Rag-1 and TdT) [129]. In summary, they reported two examples (Rag-1 and TdT) of dynamic repositioning of genes in the nuclei of lymphocytes that occurs in response to mitogenic stimulation or following differentiation induced gene-silencing. This raised an interesting possibility that locus relocation might have a role in maintaining heritable-gene expression. In murine-developing B-cells, the lymphocyte specific transcription factor Ikaros colocalizes with pericentric heterochromatin (PCH) through direct DNA binding [130]. These were one of the first studies to link higher order chromatin conformation directly with gene transcription. When the T-cell specific genes and developmentally regulated B-cell genes are inactive, they associate with heterochromatic Ikaros clusters. Ikaros seems to occupy its binding sites in gene’s regulatory element and then recruit the gene to PCH through Ikaros-binding sites found in constitutive heterochromatin (CH) [131]. Thus, a protein-mediated relocation to a heterochromatic region could lead to transcriptional silencing of genes in a development and differentiation controlled fashion.

1.4.3. Chromosome Territory (CT):
A chromosome territory is defined as the distinct physical form a chromosome acquires in the interphase nucleus. Various evidences have lead to the proposal of two kinds of models: interchromosome domain model (ICD) and interchromatin compartment model (IC), both of which are compared below in the figure 8 [132]. The ICD model envisages that the active genes (green circles) are preferentially localized at the periphery of CTs which would make genes easily accessible to transcription machinery (e.g. transcription factories) and splicing apparatus found in ICD (dashed area). This facilitates quick transport of processed transcripts through nuclear pores for export as well. On the other hand, the silent genes (red squares) or intergenic loci will stay inside of these territories rendering them inaccessible to factors present in ICD. Thus precise gene expression, processing and transport depends on selective positioning of genes with respect to surfaces of chromosome territories. There are several supporting studies for ICD model. For example, an elegant study showed that the ANT2 gene (located at Xq24–25 and subject to X inactivation) resides at the periphery of the active X territory, but on the interior of the inactive X territory (where it is silenced) [133]. By contrast, ANT3 (located at the pseudoautosomal region, Xp22.3) escapes inactivation and is peripheral in both X territories. The point to remember is that whole inactive X is preferentially located towards nuclear periphery. There are reports that some
transcriptionally active gene-dense loci can apparently stretch large distances ‘outside’ of the chromosome territory (possibly into the ICD) [88].

![Figure 8](image.png)

*Figure 8.* The diagram compares the IC and ICD models (figure adapted from Williams RR, 2003).

In contrast to these previous observations there is now a large body of evidence that several basal machineries regulating processes such as transcription, replication and repair can readily gain access to the chromosome interior. The data from Mahy et al. showed that both ubiquitous and tissue restricted genes occur within the interior of chromosome territories instead of being confined to the periphery of the territories [134]. Several labs have shown with both confocal light microscopy and electron microscopy studies that chromosome territories are rather open structures consisting of distinct condensed chromatin domains and a considerable amount of interchromatin space (dashed area) [135, 136, 137, 138] i.e., the space between the compact chromatin domains, which is largely devoid of chromatin [139]. The IC model was proposed following observations that the chromatin-devoid space (dashed area), and regulatory factors, extended through the interior of the territory. Active genes would be found on the surface of the territory or on the surface of the condensed chromatin subdomains (CC), whereas silent genes or intergenic loci could be found within the condensed chromatin domains. The non-chromatin or interchromatin space extends like continuous channels both inside and between CTs. The new transcript synthesis taking place at the interphase (perichromatin compartment) between condensed chromatin and the interchromatin compartment most probably leads to deposition of nascent RNA products directly into the interchromatin compartment from where the processing factors will take care of subsequent steps. It has been observed that certain active loci can also reside ‘extraterritorially’ (ET) on large (megabase) chromatin loops extending from the territory surface. There are few examples available to support the IC model. A study in wheat nuclei supports IC model which revealed that sites of transcription initiation, visualized by BrUTP incorporation were present in the interior of chromosome territories [140]. In addition, immunolabelling of nascent BrdU-labelled RNA together with chromosome territory painting in human fibroblasts revealed that RNA transcripts are found throughout the interior of chromosome territories.
territories, although notably, in regions of decondensed chromatin and not in regions of condensed chromatin [135]. Transcription sites and also Polycomb group silenced loci occur concentrated at the surface of compact chromatin domains [137, 138, 135]. Such actively regulated sites occurring outside of compact chromatin might represent fine chromatin loops that extend into the interchromatin space emanating away from compact chromatin domains. In addition, work of Mahty et al. showed that a ubiquitously expressed gene often colocalized with unlabeled or less intensely labeled areas of the chromosome territory, whereas the linked intergenic locus is positioned frequently within intensely labelled relatively condensed subdomains of the chromosome territory [134].

1.4.4. Chromatin fluidity:
Live cell studies have confirmed that no large-scale movements of chromatin take place but rather their motion is restricted within confined volumes with chromatin domains being allowed to have local rearrangements only [141]. In order to monitor the movements of individual genetic loci on different human chromosomes, Bickmore and co-workers randomly integrated a lac operator array into the human genome in order to visualize chromatin organization and remodeling in real time and selected clones that contained a single integration site of the array (128-mer array) at different chromosomal positions. This was the first study of chromatin motion of different loci individually which differed in gene density, replication timing, intranuclear positioning and association with nuclear components. In general, this study indicated a variable degree of constraint on chromatin movement on different genomic regions; chromatin associated with the nucleolus or nuclear periphery was more restricted in its movement than chromatin associated with other nucleoplasmic regions, indicating that these structures may act as anchoring sites and might confer rigidity to chromatin motion [142].

The restricted motion might also be caused by special protein complexes or other molecular tethers like Matrix Attachment Regions (MARs). Chromatin movement studies have been consistent with a model of Brownian motion that is constrained by interactions with immobile nuclear structures. However, when observations to study chromatin movements are made for a long period of time, apparently it seems that genomic loci can make much larger, long-range movements. As proposed by Peter Fraser and colleagues recently, there is strong possibility that the short-range movements we see in short-time studies result from large scale movements which take long time [143]. Overall, chromatin domain movements are allowed to extent far beyond their “chromosome territory” (CT).

In the light of ICD and IC models discussed previously, even these small scale movements may be sufficient to localize a gene or locus to a relevant nuclear body, repressive or active compartment, to some protein factory or to establish inter- and intrachromosomal interactions. It is noted that the territory as a whole, viewed by live-cell painting (fluorescent nucleotide incorporation at S phase and subsequent passage) does not move to any great extent during interphase [139]. Interestingly, all these observations suggest a novel level of gene regulation by controlling the activity of the molecular tethers which keep constraint on chromatin motion. Again speculating, these tethers may allow differential chromatin mobility which would affect various mobility-based processes such as going out of chromosome territory for gene regulatory purposes such as to reach transcription factories, different protein-based bodies etc and to establish chromosomal contacts in cis and in trans. Currently, the molecular nature of tethers is not known, but candidates include components of the nuclear envelope and lamina, nuclear pore complexes, heterochromatin bodies like HP1 and Polycomb, nuclear matrix and nucleolus (nucleophosmin), already assembled protein factories like Pol II complexes, chromosome-chromosome interactions etc [Fig 1(D)]. Each one of these have been shown to associate...
with genomic loci in a function-based manner. Nuclear matrix or scaffold is a speculated
three dimensional filamentous protein network supposed to give framework to maintain the
overall size and shape of the nucleus in interphase. It is supposed to act as a structural
attachment site for the DNA loops; the attached bases being called Matrix or Scaffold
Attachment Regions (MARs / SARs). The existence of a nuclear matrix has recently been
disputed with the suggestion that proteins aggregate together artificially under the
conditions used for matrix preparation. However, there is much evidence that argues for its
existence [144]. These all tethers are probably temporary in nature. The permanent
attachments are thought to provide structural integrity to the chromosome and include the
structural components of the chromatin itself like SMC (Structural Maintenance of
Chromosome) proteins [145].

There are several indications that the nuclear envelope (NE) might have an
important role to play in gene regulatory mechanisms [97]. Besides the interactions
necessary for nuclear assembly, the nuclear envelope-chromatin association is correlated
with gene repression. On one hand, the nuclear lamina (e.g., lamins, LBR, or LEM
proteins) impounds folded chromatin from transcription, whereas on the other, the nuclear
envelope would play a more active role by stabilizing repressive complexes. Alternatively,
lamins might directly interfere with transcription.

The complex interactions of lamins and lamin binding proteins (e.g. LBR)
with DNA and with chromatin-associated proteins (e.g. histones, HP1, HA95 and BAF) at
the nuclear periphery suggest roles for these proteins in higher order chromatin
organization by providing specific chromatin-docking site at the NE and by structurally
organizing chromatin fibers in the three-dimensional nuclear space [146] As discussed
before, this interaction could be very important in organizing higher order loop domains
involved in various steps of gene regulation including chromatin insulator function,
transcriptional regulation, boundary activity etc. As discussed previously, the nuclear
envelope is almost exclusively associated with heterochromatin events; heterochromatic
entities like telomere, centromere, repetitive DNA and inactive genes are preferentially
positioned near the nuclear periphery. Insertion of an enhancer to induce ectopic expression
of a gene in some cases is accompanied with relocation away from periphery. In the
interphase cells of female mammals, the inactive (heterochromatinized) X chromosome
also sits near the periphery whereas the active X in the same cell occupies a more interior
position [147, 148]. Also, highly silenced human chromosome 18 occupies more peripheral
position in the nucleus as compared to highly active chromosome 19 [106]. These
evidences clearly show a direct correlation with transcriptional repression.

Direct evidences for nuclear lamina involvement in chromatin organization came from
studies of lamina deficient cells. Chromatin attachment to nuclear lamins was abnormal in
human emerin-deficient cells, in mouse lacking lamin A/C, in Drosophila cells lacking
lamin Dm., and in C. elegans cells lacking Ce-lamin [105,149, 150, 151, 152].

Interestingly, the cells expressing various mutant lamins show defect in
replication process indicating that Lamin proteins directly or indirectly may be involved in
regulating DNA replication process [153, 154, and 155]. The Drosophila gypsy element (a
retrotransposon and insulator) associates with the nuclear envelope through binding to
Su(hw) protein, but it becomes detached from the envelope and looses insulator function if
Su(hw) is mutated. Insulator elements, such as gypsy, are regulatory sequences that are
thought to organize independent genomic regions of transcriptional activity (or repression)
and thus point towards a functional link between a tether site and transcription control
[156]. There are evidences for positive role for nuclear lamin as well. Expression of a
dominant -negative form of lamin mutant lacking the head domain disrupts lamin A
filaments and inhibits specifically RNA polymerase II activity in both mammalian cells and
transcriptionally active embryonic nuclei from *Xenopus laevis* (not Pol I and Pol III) and involves redistribution of TATA binding protein [157]. Another study demonstrated that Pol II transcription is organized by internal (nucleoplasmic) lamins and does not involve peripheral lamins [158].

In addition, components of nuclear lamina may directly influence transcription by interacting with transcription factors and/or chromatin remodeling complexes. For example, an integral protein of the inner membrane (Ring Finger Binding protein, RFBP) is directly shown to interact with RUSH proteins, the SWI/SNF remodeling factors [159]. Lamin proteins have also been shown to interact with E2F transcriptional complexes which regulate G1-S phase transition [160]. In addition, lamins A/C associate directly with the hyperphosphorylated, active form of retinoblastoma protein (pRb) [161].

<table>
<thead>
<tr>
<th>Nuclear lamin protein</th>
<th>Chromatin binding partners</th>
<th>Major binding sites(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lamin</td>
<td>5′DNA, MAPs/NARs</td>
<td>lamin N-terminal domain</td>
</tr>
<tr>
<td></td>
<td>telomeric and centromeric DNA</td>
<td></td>
</tr>
<tr>
<td></td>
<td>microphase chromatin</td>
<td></td>
</tr>
<tr>
<td></td>
<td>active chromosomes</td>
<td></td>
</tr>
<tr>
<td></td>
<td>core histones H2A and H2B</td>
<td>lamin C-terminal domain</td>
</tr>
<tr>
<td>LBR</td>
<td>chromatin</td>
<td>N-terminal nucleoplasmic domain</td>
</tr>
<tr>
<td></td>
<td>5′DNA</td>
<td>residues 1–33</td>
</tr>
<tr>
<td></td>
<td>RFI</td>
<td>residues 97–174</td>
</tr>
<tr>
<td></td>
<td>H4A95</td>
<td>ND</td>
</tr>
<tr>
<td>LAP2α</td>
<td>DNA</td>
<td>LEM-like domain (residues 1–50)</td>
</tr>
<tr>
<td></td>
<td>chromatin</td>
<td>residues 1–85</td>
</tr>
<tr>
<td></td>
<td>RFI</td>
<td>C-terminal LAP2α-specific domain</td>
</tr>
<tr>
<td></td>
<td>RFI A95</td>
<td>residues 226–443</td>
</tr>
<tr>
<td>LAP2β</td>
<td>DNA</td>
<td>LEM-like domain (residues 1–50)</td>
</tr>
<tr>
<td></td>
<td>chromatin</td>
<td>residues 1–85</td>
</tr>
<tr>
<td></td>
<td>RFI A95</td>
<td>LEM domain (residues 111–132)</td>
</tr>
<tr>
<td>Sin3αα</td>
<td>RFI</td>
<td>LEM domain (residues 1–50)</td>
</tr>
<tr>
<td>Sin3αβ</td>
<td>RFI</td>
<td>LEM domain (residues 1–50)</td>
</tr>
<tr>
<td>MAN1</td>
<td>RFI</td>
<td>LEM domain (residues 1–50)</td>
</tr>
<tr>
<td>RFBP</td>
<td>RUSH</td>
<td>RFBP conformationally flexible loop</td>
</tr>
<tr>
<td>5A</td>
<td>latephase chromatin</td>
<td>Binding requires four domains:</td>
</tr>
<tr>
<td></td>
<td>polytene chromosomes</td>
<td>residues 1–117, 270–306, 397–472</td>
</tr>
<tr>
<td></td>
<td>active chromosomes</td>
<td>and 506–696</td>
</tr>
<tr>
<td></td>
<td>histone H2B</td>
<td></td>
</tr>
</tbody>
</table>

**Table 1.** Various nuclear lamin proteins, their chromatin interactions and major docking sites (adapted from Mattout-Drubetzki A and Gruenbaum Y, 2003)

Although strong connections exist between the nuclear periphery and silencing, a genetic screen in *S. cerevisiae* to identify chromatin boundary activities (BAs) has identified several transportins that mediate their epigenetic function by specific physical tethering of the *cis*-acting boundary elements to the Nuclear Protein complexes (NPC) [36]). A summary of most of the known interactions between nuclear lamin proteins and chromatin is shown in the table [162].
The supporting role for LBR in chromatin organization came from studies of an autosomal dominant disorder called Pegler-Huet anomaly (PHA) by mutations in the LBR gene showing characteristic features like aberrant chromatin organization and nuclear lobulation in blood granulocytes [163]. There are evidences that LBR-HP1 binding to chromatin is sensitive to histone modifications indicating that histone code might influence gene expression by controlling chromatin-nuclear periphery interactions [164, 165,166]. There are other examples to relate nuclear lamina structure and function. One interesting example is mutations in the A-type lamin gene (LMNA, with major splicing variants A and C) resulting in heterogeneous group of genetic disorders, collectively referred to as “laminopathies” [167]. The wide variety of dysfunction and symptoms seen here reflect a genomewide relation with these lamin proteins.

Overall, based on evidences so far, nuclear lamina is clearly involved directly or indirectly in many nuclear activities, including DNA replication, transcription, nuclear and chromatin organization, cell cycle regulation, development and differentiation, and apoptosis.

1.4.5. Compartmentalization of transcription in nucleus:
Transcription is highly compartmentalized in mammalian nuclei. Besides the existence of transcription factories and few transcription factors as specific bodies, there are some function-based foci (includes nuclear speckles, Cajal bodies, and PML bodies, nucleolus etc) which can arise through the self-assembly of essential components without an underlying template or scaffold. A very good example is formation of nucleolus where rRNA genes are transcribed by pol I complexes. Observations that IGCs (Intercromatin Granule Clusters) contain transcribed genes, RNA and splicing factors like SR proteins indicates that it is most probably involved in mRNA processing [168]. Cajal bodies are likely a site of SnoRNP and SnRNP processing [169]. It also contains guide RNAs and might be involved in RNA editing. Thus IGCs and Cajal bodies present evidence that machinery involved in the processing of transcripts is organized within the nucleus. Although not very clear, PML-NBs have been given regulatory role in transcription [170], given the fact that they directly interact with factors like CBP and pRB [171] and transcriptionally robust loci are associated with PML-NBs [172]. They also contain proteasomes and may have a role in degrading transcriptional regulators in response to external stimuli. It is also possible that they play a role in transcription possibly as a reservoir of concentrated transcription factors. The all nuclear bodies seem to follow the common principle of “form follows function” i.e. they loose their structure on perturbation of their function/relation proteins [127].

Thus, the transcription is functionally well-compartmentalized in the nucleus. The active transcription centers contain all the components required to establish the transcription complex (transcription factories and other regulatory factors), perform the transcription cycle, generate the nascent transcript, and process this into mature mRNA. Chromatin domains and loops are generated by any contact between chromatin and the active center. Furthermore, the stability of any attachments, and hence levels of expression, will depend on extremely complex interactions that reflect the affinity of individual proteins for DNA and the local concentrations of these proteins.
2. The molecular determinants of DNA looping:

Our discussion so far makes us realize that most of the events involved in gene regulation seem to require looping of chromatin as an essential component. It was first discovered in the ara operon of Escherichia coli [173] although that time it was already suspected to be present in eukaryotic enhancers. Besides transcriptional regulation, apparently it is involved in other basic cellular processes such as replication, recombination, and chromatin condensation during mitosis, chromosomal positioning, chromosomal tethering etc.

Most of these looping events are carried out by sequence specific DNA binding proteins. Stability of loops will depend highly on the specificity and strength of protein-protein interactions mediating their formation. The flexibility of molecular tethers holding chromatin, DNA supercoiling and nucleosome positioning can also facilitate the process. The loop formation could take place in several alternative loop geometries depending on DNA-protein and protein-protein interactions and in that case binding of an architectural protein might help stabilize only a specific loop conformation. Together with superhelicity of DNA, concerted action of specific and non-specific DNA binding proteins to form DNA loops must also contribute to the above process (Fig 9) (modified from [174].

There could be following ways in which the DNA loops can be formed and stabilized.

(a) Two similar or different proteins occupying cognate sites on DNA interact with each other directly e.g. between DNA elements bound to the c-Myb and CCAAT/enhancer binding protein (C/EBP)b proteins in the transactivation of the myeloid genes [175].

(b) A bi-dentate protein binds to two distant DNA sites. Bi-dentate proteins have two identical or non-identical DNA-binding sites. e.g. Ikaros proteins seems to have binding sites both in pericentric heterochromatin (PCH) and in the regulatory elements of genes it targets to PCH as discussed previously. Another example is the LacI tetramer in prokaryotes which is actually a dimer of dimers. Each dimer binds to one operator site and thus a lac repressor tetramer can bind two independent pieces of operator and bring them together [176].

(c) An adaptor protein bridges two proteins bound at spatially separated sites on DNA, thus closing the loop. Apparently cAMP-responsive-element-binding protein (CREB)-binding protein (CBP) acts like adaptor protein in the b-interferon enhancesome.

(d) In above mentioned [(a) to (c)], chromatin higher order folding might be facilitated by a DNA sequence-specific architectural protein (discussed later in detail).

(e) In case of inter- and intrachromosomal interactions between individual loops and the targeting of chromatin to particular sites of repression and activation, besides the above factors operating within loops, protein-protein interactions between two interacting loops (in first case) or between the targeting site and the looped DNA (in later case) is crucial for determining stability and final outcome of such interactions.

Thus a variety of factors could contribute to whether a particular loop will be formed (figure 10) [177]. The most important seems to be proteins occupying the distant interacting sites. The other players include DNA supercoiling (in general, the effective concentration of one point near another is increased by supercoiling), the distance between the interacting sites (As the length of DNA between sites increases, the energetic requirement for both bending and twisting diminishes, but the local concentration of sites also decreases and the entropic cost increases) the flexibility and deformability of intervening DNA (depends on architectural protein occupancy, molecular tethers, base composition and sequence of the interacting sites), the orientation of the sites (If two sites are disoriented around the DNA, twisting of the DNA between the sites may be required to orient the sites properly for looping; on the other hand, if the sites are already properly oriented, torsional stiffness of the DNA holds the sites in the correct orientation and increases the ease of loop formation) and the three-dimensional freeness of the interacting
sites, histone code, other regulatory factor(s) occupancy and yet some unknown contributing factors[177] [178].

Formation of loops along a dsDNA via collisions of sites brought together by conformational diffusion is an important mechanism behind “higher-order” organization of dsDNA in vivo as is well supported by observations that genomic loci exhibit Brownian diffusion. As discusses before, the stability of the loops might be increased by the binding of an architectural protein at a specific position of the looped DNA. The architectural protein introduces a bend with a characteristic stereo specificity or it might induce conformational changes in chromatin (see figure 9, the dark round oval bodies bound here and there) which in turn might facilitate subsequent steps e.g. in EBV, a DNA-bending architectural protein HMGB1 is shown to promote enhanceosome assembly by a viral activator on a target gene [179]. There are several examples of these proteins in prokaryotes [174]. In eukaryotes, such architectural protein is started to be discovered like SATB1, Batch1, LEF-1, and DEK etc. Lymphoid Enhancer Factor-1 (LEF-1) has been proposed to act as architectural transcriptional factor that activates transcription in association with other proteins [180]. The ubiquitous proto-oncogene DEK is found associated with chromatin during the entire cell cycle. It is shown to change the topology of DNA in chromatin and protein-free DNA by introducing positive supercoils. DEK also stimulates intermolecular catenation of circular DNA molecules in the presence of topoisomerase II [181]. The properties shown by DEK makes it a good candidate as architectural factor in eukaryotes. HMG1 and 2 (HMG box proteins) are related architectural proteins which bend DNA and bind preferentially to distorted DNA structures. They might be recruited to specific sites by DNA-protein or protein–protein interactions where they mainly act as architectural facilitators in the assembly of nucleoprotein complexes e.g. in affecting recombination and initiation of transcription [182]. An interesting study done by Terumi Kohwi-Shigemitsu and colleagues showed that in thymocyte nuclei, SATB1, a cell-type specific proteins, forms cage-like patterns throughout the nucleus circumscribing heterochromatin. This provides attachment sites for
specialized genomic regions including Myc, and leads to formation of a tissue-specific nuclear architecture and is well correlated with proper regulation of distant genes [183]. Using SATB1 null thymocytes, it was also demonstrated that this tethering of the selected sequences onto its network helps maintaining proper histone modification patterns over a region of roughly 10 Kb covering genes regulated by SATB1 and those genes were found disturbed in these null thymocytes. Bach1 is a BTB/POZ transcription factor, and is known to mediate the communication between cis elements located far away from each other by introducing the formation of organized DNA loops [184]. SP1 is also shown to mediate DNA loop formation [185]. There are several proteins known which induce a kink or bend in the DNA sequence they bind and this is high time to explore the known proteins about their role as architectural proteins to facilitate loop formation and/or stabilize them. CTCF, the vertebrate insulator protein has been already implicated to harbour such activity [186, 187]. It is quite possible that out of plethora of transcription factors present in eukaryotes, some possess dual properties i.e. besides their specific properties, they also possess architectural properties such as ability to induce a kink or bend the DNA, cause conformational changes and thereby facilitate the chromatin looping processes required for various gene regulatory mechanisms.

3. Epigenetic control of gene expression:

The genetic material of all the cells is same and therefore differential gene expression and its spatio-temporal regulation might depend on the mechanism that read the information of the same genetic material differently without altering the DNA sequence. That is where epigenetic modifications decide the final activation or repression of gene expression, mostly by affecting chromatin structure. In order for these epigenetic modifications to be functional on a long term basis, they need both to remain stable to constitute a cell-memory through cell divisions, and to be erased when needed, e.g. during gametogenesis. Several different epigenetic modifications have been described and two of the important ones are discussed here, namely histone modifications and DNA methylation.

Histones undergo to a variety of post-translational modifications, principally in their conformationally flexible N-terminal tails. The specific residues here undergo a wide variety of modifications which include phosphorylation, ubiquitination, acetylation and mono-, di- or tri-methylation. The number of possible permutations of the various modifications is extremely large and has been proposed to constitute a ‘histone code’, which epigenetically confers different degrees of transcriptional competence on different regions of chromatin [188]. The post-translational modification of histones show a high level of diversity and complexity and the interplay between them is not understood to a large extent.

The best understood epigenetic modification of DNA in mammals is methylation of cytosine at position C5 in CpG dinucleotides. The mammalian DNA methylation machinery has two constituents, the DNA methyltransferases (DNMTs), which establish and maintain DNA methylation patterns, and the methyl-CpG binding proteins (MBDs), which ‘read’ these methylation marks [189, 190]. There is also clear evidence that some kind of DNA demethylase has a role in regulating DNA methylation patterns during embryonic development, although the activity responsible is yet to be identified [191]. DNA methylation is crucially involved in regulating many cellular processes including embryonic development, transcription, chromatin structure, X chromosome inactivation, genomic imprinting and chromosome stability. In normal cells, DNA methylation occurs predominantly in repetitive genomic regions, including satellite DNA and parasitic elements (such as long interspersed transposable elements (LINES), short interspersed transposable elements (SINES) and endogenous retroviruses). Approximately 60-90% of
all CpG sequences in the genome are methylated, while unmethylated CpG dinucleotides are mainly clustered in the CpG rich sequence termed CpG islands which are mostly associated with promoter regions of active genes. CpG dinucleotides are under-represented in the mammalian genome due to the conversion of methylated cytosine to thymidine via spontaneous deamination, followed by stabilisation of the genome during replication in the germline. CpG islands, particularly those associated with (housekeeping mainly) promoters, are generally unmethylated, although an increasing number of exceptions are being identified. Specific cis-acting sequences seem to protect them from global wave of de novo methylation occurring during reprogramming event at the time of implantation. DNA methylation usually represses transcription directly, by inhibiting the binding of specific transcription factors. For example, the abnormal methylation that occurs on the fragile X mental retardation 1 (FMR1) promoter prevents binding of nuclear respiratory factor (NRF1; also known as α-PAL) and inhibits transcription of this gene, causing fragile X syndrome [192]. It could also cause transcriptional suppression indirectly, by recruiting methyl-CpG-binding proteins and their associated repressive chromatin remodelling activities such as histone deacetylases (HDACs) and histone methyl transferase (HMTs) that alters the local chromatin environment [193]. Methylation can also affect nucleosomal positioning [194]. It could also lead to a general decrease in DNase I sensitivity and alteration in higher order structure that make these regions less accessible to the transcription machinery [195]. The exceptions to the suppression role are also known e.g. an intergenic region (DMR2) in mouse Igf2 augments transcription on methylation [196].

Little is known about how DNA methylation is targeted to specific regions; however, this probably involves interactions between DNMTs and one or more chromatin-associated proteins. Early models for the control of DNA methylation proposed two-steps: First the methylation pattern is established by de novo DNMTs active on unmethylated DNA (de novo methylation) and then this pattern is maintained through replication by a semiconservative mechanism by copying the methylation residues from parental strands onto the newly synthesized nascent DNA by maintenance DNMTs specific for the hemi-methylated DNA resulting from replication (maintenance methylation).

<table>
<thead>
<tr>
<th>Protein</th>
<th>Function</th>
<th>Mutant Phenotype</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Methyltransferases</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dnmt1</td>
<td>Maintenance of methylation</td>
<td>Embryonic lethal, loss of imprinting and X-linked gene expression, ES cells viable</td>
</tr>
<tr>
<td>Dnmt1o</td>
<td>Oocyte-specific isoform</td>
<td>No phenotype</td>
</tr>
<tr>
<td>Dnmt2</td>
<td>Non-CpG methylation in Drosophila</td>
<td></td>
</tr>
<tr>
<td>Dnmt3a, Dnmt3b</td>
<td>De novo methyltransferases, establishment of methylation</td>
<td>Embryonic lethal, ICF syndrome</td>
</tr>
<tr>
<td>Dnmt3L</td>
<td>No catalytic activity, colocalizes with Dnmt3a and Dnmt3b</td>
<td>Abnormal maternal imprinting</td>
</tr>
<tr>
<td><strong>Methyl binding proteins</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MeCP2</td>
<td>Methyl binding proteins, recruit HDACs</td>
<td>RTT</td>
</tr>
<tr>
<td>MBD1</td>
<td>Methyl binding proteins, recruit HDACs</td>
<td></td>
</tr>
<tr>
<td>MBD2</td>
<td>Methyl binding proteins, recruit HDACs</td>
<td>Behaviour abnormalities</td>
</tr>
<tr>
<td>MBD3</td>
<td>Methyl binding proteins, recruit HDACs</td>
<td>Lethal</td>
</tr>
<tr>
<td>MBD4</td>
<td>Repair enzyme</td>
<td>Increased mutation frequency</td>
</tr>
<tr>
<td><strong>Histone-modifying proteins</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HDAC1</td>
<td>Histone deacetylase</td>
<td>Embryonic lethal</td>
</tr>
<tr>
<td>Suvar39</td>
<td>Lys9 methylation in histone H3</td>
<td>Embryonic lethal, chromosomal instability, increased tumor risk</td>
</tr>
</tbody>
</table>

Table 2. Key epigenetic regulatory proteins, their function and phenotype (adapted from Jaenisch R and Bird A, 2003).
The phenotypes of some selected mouse mutants of epigenetic regulatory factors are listed in table 2 [197].

There are several indications that DNA methylation and histone methylation are intimately linked and may control each other, but which one comes first becomes complicated due to reports from both ways. There are indications that environmental factors like diet can influence epigenetic states, which directly links environment to the regulation of gene expression [198]. Transitions between silent and transcriptionally competent chromatin states seem to be highly dynamic and depend on a balance between the factors that promote a transcriptionally active state such as HATs and those that sustain a silent state and, such as HDACs. A shift in the balance between an active and a silent chromatin conformation could be simply achieved by regulating those factors, resulting in transcriptionally active or inactive state.

There is increasing body of evidence that genes are regulated by several layers of repression mechanisms e.g. DNA methylation, histone modifications, sequence-dependent repression (involves protein factors that recognize and bind to local cis-acting sequences and bring about transcription repression), relocation to different nuclear subcompartment (viz. lymphoid-specific zinc-finger protein, IKAROS dependent relocation of a gene region from its euchromatic position within the nucleus to specialized regions of heterochromatin), DNA replication (strong correlation between late-replication timing and gene repression) and moreover recently explored inter-and intrachromosomal interactions. These processes could act independently, guide each other or can work in coordination to ensure proper expression patterns. X-chromosome inactivation in female mammals represents a good example where several gene repression mechanisms operate together, for example it involves a non-coding transcript (Xist RNA), late-replication timing, DNA methylation, histone modifications, nuclear localization etc all of which are used to maintain heterochromatin state of the inactive-X. Once inactive the choice of which X will be inactive, the pattern is maintained clonally in the daughter cells and need not be regenerated with each cell division during development [199, 200].

4. Maintenance of epigenetic states:
The various cell types in a multicellular organism are genotypically identical and yet phenotypically different. This is due to differences in the epigenetic patterns of genes and various genomic regions that exist between the different cell groups and are executed for differential gene expression. The crucial task is the stable maintenance of these marks over several cycles of mitotic cell divisions for faithful transmission of this information.

The replication fork provides an opportune situation for both creating and propagating the epigenetic state of chromatin. Maintenance methylases (Dnmt1) recognize the newly formed hemi-methylated strand after replication and add methyl groups to that similar to the parental ones. In that way, the methylation patterns once established by de novo methylases (Dnmt3a & Dnmt3b) are faithfully inherited from mother to the daughter cells. While forming the two daughter strands from a parental strand, the histones present in the parental strand are randomly distributed to the daughter ones; accompanied by new histone deposition as well. A good example to support the correlation between replication timing and chromatin status is the experiment done long back where replication-competent reporter plasmids were transfected into Rat1 cells at different times during S phase. The exogenous DNA was assembled into transcriptionally active, hyperacetylated chromatin if injected in early S phase, while it was assembled into transcriptionally inert, hypoacetylated chromatin when injected late. Once the DNA acquires a particular replication profile, the transcriptional state was maintained through cell divisions [201].
The correlation of replication timing with the type of chromatin assembled can be explained by the fact that a histone deacetylase, HDAC2, and a chromatin remodeling complex, WICH, specifically localize to late-replicating foci [202]. Recent evidences from our lab has revealed that the vertebrate insulator protein CTCF binding sites strongly correlate with late-replication timing of the associated loci (Bergstrom R et al unpublished observations). Thus, these proteins besides executing insulator activity, might have yet unexplored role in establishing epigenetic states throughout mammalian genome by having an impact on DNA replication. There could be similar other proteins in the genome and they possibly execute their function by placing the target sequences to a nuclear environment permissive for early/late-replication.

There may be several mechanisms to stably transmit the pre-existing acetylation patterns to newly assembled chromatin following replication [203]. One attractive observation invokes the possibility that HATs and HDACs may remain in vicinity and associate with newly assembled chromatin preceding replication [204, 205]. A second proposal is that these factors HATs and HDACs may be targeted to methylated regions so that the acetylation patterns could be propagated alongwith methylation during replication. These gains supported from observations that Dnmt1 interacts with histone deacetylase [206, 207, and 208]. The recruitment of MeCP2 to methylated CpG dinucleotides represents a major mechanism by which DNA methylation can repress transcription. MeCP2 silences gene expression partly by recruiting histone deacetylase (HDAC) activity, resulting in chromatin remodeling [209]. MeCP2 also associates with HTMase in vivo and that this interaction delivers Lys (9) of histone H3 to the recruited site [210]. The same might hold true during replication and MeCP2 can guide establishment of histone acetylation and methylation patterns to the newly synthesized DNA in DNA methylation guided fashion. The maintenance of histone methylation patterns during the replication process is explained by an excellent proposal that HMTs are recruited to chromatin by existing methylated residues, as is suggested to be the case with H3K9me3 in heterochromatin [211]. This kind of positive feedback loop could enforce modification of newly deposited histones. In some cases where DNA methylation is guiding force for histone methylation patterns, it just needs the former signatures already established [212]. In the case of methylation patterns of active genes, the continuation of transcription might also be sufficient to transmit these modifications to newly deposited histones.

The spatial organization of DNA synthesis contributes in a manner similar to temporal factors: The “replication foci” observed might be due to coordinated regulation of similarly replicating loci where they join together to share a common set of chromatin modifying factors [213]. If different types of chromatin come close together to form a subcompartment of the nucleus, it might result in a high local concentration of chromatin modifying factors. Although long-term inactivation does not always correlate with late replication, it does correlate with relocation to a subnuclear domain in which heterochromatin factors such as HP1 concentrate. HP1 binding usually persists during replication and this might allow quick transfer of this information to the daughter strands to get packaged into heterochromatin. During replication, HP1 is sown to interact with chromatin assembly factor p150CAF-1 and this delivers HP1 molecules to heterochromatic sites, where they are subsequently retained by further interactions with methylated H3-K9 and RNA [93]. Also as has been discussed previously, the maintenance of chromosome positions through cell division possibly helps to re-establish epigenetic state and transcription patterns in daughter cells similar to mother cell, by placing genes and regulatory elements in similar kind of microenvironment in the nucleus [116].

The genes of the Polycomb group (PcG) and trithorax group (trxG) are part of a widely conserved cell memory system which prevents changes in cell identity by
maintaining transcription patterns which are set in the early stages of embryonic life, throughout development, and in adulthood (Reviewed [214]). This multiprotein assembly is targeted to specific regions of the genome where it effectively freezes the embryonic expression status of a gene, be it active or inactive, and propagates that state stably throughout development. PcG and trxG control, respectively, repressed and active transcriptional states of several loci in the genome, including developmentally and cell cycle-regulated genes. There is strong evidence that the memory function encoded by these two groups of genes is achieved through regulation of higher order chromatin structures. PcG gene products form large multimeric protein complexes in Drosophila, mouse and human. PcG mediated gene silencing can be directed by DNA elements in cis, defined as PcG response elements (PRE). Trimethyl K9 and dimethyl K27 on histone H3 appear to be specifically enriched at PcG target sites but there could be additional marks as well. On the other hand, several trxG members act at elements defined as TRE (that overlap with PRE) via chromatin remodeling and induction of histone modifications that increase chromatin accessibility to transcription factors. Thus, chromatin structure appears to contain the molecular imprint underlying cell memory and epigenetic inheritance here. Are there any similarities in memory propagation by PcG/trxG and DNA methylation in mammals? Yes there are, first they both can suppress (and activate) transcription in a heritable manner. In addition, PcG/trxG lock in the expression status in which they have had no part in setting up which is similar to DNA methylation in few cases (e.g. CpG-island methylation on the X chromosome) Also, both of these systems are activated primarily during a discrete window of time during early phases of development. It is possible that even within a species; different tissues employ PcG/trxG and DNA methylation interchangeably. This is supported by observations that X chromosome inactivation in extraembryonic tissues of the mouse requires polycomb group protein Eed and DNA methylation in embryo proper (discussed later). Knock-out of the eed gene results in reactivation of the inactive X in extraembryonic tissue, but in somatic cell types, inactive X chromosome keep on maintaining its imprinted status [215]. In contrast, Dnmt1 mutations reactivate the inactive X of the embryo proper, but not the extraembryonic inactive X [216]. The finding that certain CpG islands on the inactive X chromosome are methylated in somatic cells but not in extraembryonic tissues [217] fits with the view that methylation replaces Pc-G in somatic tissues [218].

5. Genomic imprinting:
Genomic imprinting affects several mammalian genes. Here, the epigenetic mechanisms mark the two parental alleles during gametogenesis, resulting in allele-specific expression depending on the parent-of origin in the developing embryo and adult. Imprint acquisition at a locus requires that the two alleles be differentially marked in oogenesis and spermatogenesis, and these marks escape reprogramming after fertilization, and that they are reliably transmitted propagated during development. The epigenetic modifications responsible for establishing and maintaining imprinted gene expression are thought to involve DNA methylation, chromatin modifications, and asynchronous replication timing. Recent findings about occurrence of long-range chromatin interactions add a new dimension to the proper imprinting regulation [50]. Most imprinted genes are found clustered in large chromosomal domains which indicate that they might be co-regulated. There exist specific cis control elements, imprinting control regions (ICRs), which are required for regional control of imprinting, imprinted gene expression, and asynchronous replication timing. These imprinting control regions (ICRs) are regulated by epigenetic modifications. ICRs are up to several kilobases in length, and are rich in CpG dinucleotides (many correspond to CpG islands). Another hallmark of ICRs is that they have DNA
methylation on one of the two parental alleles in parent-of-origin specific manner. These
marks are resistant to global demethylation during the early cleavages of the
preimplantation embryo. Following fertilisation, allelic methylation marks are maintained
throughout development and they mediate imprinted gene expression [219, 220].

To date, DNA methylation is the only known epigenetic mark to be required
for maintenance of imprinted gene expression (except a recent study by Lewis et al 2004)
[221]. Table 2 highlights the importance of few regulatory factors in the epigenetic
regulation. Using a conditional targeting approach, we know for sure that Dnmt3a (but not
Dnmt3b) is the methyltransferase that adds methyl groups to ICRs, for some in the female
and for others in the male germ line [222]. The question that puzzles us is how these are
specifically targeted to those regions. Alternatively, if the acquired methylation imprints
are result of the de novo methylation during germline development, how the same region is
protected from it in one of the two germlines. One possibility is that the binding of non-
histone proteins to individual ICRs could prevent acquisition of methylation in one of the
two germ lines. This seems to hold true for Igf2 -H19 locus where CTCF binds to the
unmethylated maternal allele on ICR in somatic cells. Recently two independent studies,
including one from our own lab, where point mutations were introduced at H19 ICR to
abolish binding of CTCF and when these mutations were passaged through female germ
line led to aberrant hypermethylation of maternal allele [223, 224]. The postulate that
CTCF could protect H19 ICR from DNA methylation in the female germ line was
confirmed by a recent study in which CTCF expression was downregulated during
oogenesis using a transgenic RNA-interference (RNAi) approach [225]. What remains
unclear is whether in male germ line there exist factor(s) directing active DNA methylation
at ICR or this methylation state is acquired as a default due to the absence of factors like
CTCF which could protect it from genomewide methylation occurring during
reprogramming events.

Igf2r ICR and Kcnq1 ICR are two independent imprinted domains which
seem to be essential for silencing along the paternal chromosome. Both of them seem to
involve non-coding RNA mediated repression of nearby genes on paternal chromosome
which might bring about other epigenetic modifications needed for silencing. The allelic
repression along the imprinted Igf2 r and Kcnq1 domains has marked similarities with X-
chromosome inactivation. For instance, there is experimental evidence for the involvement
of Eed (PcG proteins) and non-coding RNA in both imprinted domains and the same is true
for inactive X [226,215, 227]. As coregulated genes are clustered in imprinted domains,
likewise on human X-chromosome, genes subjected to XCI induced silencing and those
that escape it are blocked into large domains-which suggest that regional control elements
determining the transcriptional status in both cases and in addition, in both cases it might
involve higher order chromatin organization [228].

In a recent paper by Lewis et al, authors show that imprinted paternal
repression at certain ‘distal’ genes in the Kcnq1 imprinting cluster in extra-embryonic
placenta tissue is independent of Dnmt1 function [229]. This is the first example for
regulation of imprinting by other means independent of DNA methylation. The special case
being here extra-embryonic development, where H3K9me2 and/or PRC2 protein actions
are required for imprinted silencing independent of DNA methylation.

Recently Paul D. Soloway and colleagues have shown that Imprinted expression of Rasgrf1
is regulated by methylation-sensitive CTCF binding on maternal DMD situated 30 kb 5’ of
the promoter where it acts as an enhancer blocker. CTCF binding to the paternal allele is
prevented by repeat-mediated methylation, allowing expression of Rasgrf1 [230].

Mouse Igf2 has three DMRs: maternally methylated, placenta-specific
DMR0 located at exon U1 and paternally methylated DMR1 and DMR2 located upstream
of promoter 1 and in exon 6, respectively (Fig 4). Deletion of the maternal H19 DMR results in loss of imprinting of Igf2 with biallelic expression in most tissues studied [231], and deletion of the maternal Igf2 DMR1 results in biallelic expression of Igf2 in mesodermal tissues [232]. Deletion of Igf2 DMR2 doesn’t affect imprinting but reduces transcriptional activation of Igf2 [196]. Igf2 DMR2 is one of the rare examples known so far where methylation induces transcription of the related gene. Thus, DMR1 has a (methylation-sensitive) silencer function and DMR2 has an (methylation-sensitive) activator function. There are observations suggesting that some other marks present in ICRs, independent of DNA methylation, lead to the localization of the two parental imprinted alleles to different subnuclear compartments with different replication characteristics and thus result in asynchronous replication [94].

Methylation studies have shown that on the maternal chromosome, the H19 ICR protects Igf2 DMR1 and DMR2 from methylation, and DMR1 protects the DMR2 from methylation in a hierarchical manner [233]. The observations that the methylated status of all DMRs on maternal alleles is not due to linear spreading of epigenetic state suggested long-range chromatin interactions and raised possibility that the DMRs interact directly such that the H19 ICR protects the Igf2 DMRs from becoming methylated in somatic tissues. Recently Murrell et al showed that differential interaction of H19 ICR with Igf2 DMR1 (maternal allele) and Igf2 DMR2 (paternal allele) partitions the whole locus into active and repressive domains and decides Igf2 expression by physically regulating enhancer access across the locus [50].

With the evidences available so far, between human and mouse, there seems to be a nearly perfect conservation of imprinting with one notable exception of Igf2 r, which is imprinted in mice but not in humans [234], [235]. There are various disorders known which result from deregulation of imprinting. A comprehensive list of putatively imprinted mammalian ‘genes’ was derived from the published literature and is currently summarized in the ‘Catalogue of imprinted genes and parent-of-origin effects’ (www.otago.ac.nz/IGC) and the MRC Mammalian Genetics Unit mouse imprinting data website (http://www.mgu.har.mrc.ac.uk/research/imprinting).

6. X-chromosome inactivation:
X-chromosome inactivation (XCI) operates in mammals to achieve dosage equivalence for most X-linked genes between the two X-chromosomes in female and the single X-chromosome in males. This represents a remarkable example of epigenetic inheritance because once an X is chosen to be inactive during early embryogenesis; the state is maintained through several generations in somatic cells maintaining the repressed status of thousands of genes without any error.

A crucial component to the inactivation procedure is X-Inactivation Centre (XIC). The Xist gene is part of it and codes for more than 17 Kb untranslated RNA which spreads in cis to cover the future X to be inactivated. A study demonstrated that “counting” (where only a single X will remain active per diploid autosome set, with inactivation of supernumerary X chromosomes) is assured precisely by a 20 Kb bipartite domain within XIC region 3’ of Xist [236]. The “choice” function whereby one of two X chromosomes is selected for inactivation (or to stay active?) seem to be influenced by multiple elements spread across the Xic region like Tsix, Xist, Xce etc and the targeted deletions in these regions have shown to lead to non-random X-inactivation [248]. The choice of which X to be inactivated is random (except extraembryonic tissues). In marsupials and in extraembryonic tissues in mouse, an unknown parental imprint leads to preferential inactivation of the paternal X chromosome and it likely affects one of the many elements involved in X chromosome choice. In mouse, an X controlling element (Xce) maps 3’ of
\textit{Xist} and can skew X-inactivation patterns from the normal 50:50 ratio \cite{237}. Another important element is \textit{Tsix} gene which has transcription start site 12 kb downstream of \textit{Xist} and it codes for few non-coding RNAs antisense to \textit{Xist}. Mouse knock-out studies have shown that \textit{Tsix} negatively regulated the expression of \textit{Xist} by a mechanism requiring overlapping transcription which subsequently stops \textit{Xist} from inactivating that X-chromosome \cite{238,239}. Another component of choice is \textit{Xite}, which lies upstream of \textit{Tsix} and it positively regulates \textit{Tsix} expression \cite{240}. \textit{Xite} harbours developmentally regulated hypersensitive sites and transcription start sites that give rise to low-level intergenic transcripts.

The current model illustrates that autosomal \textit{trans}-acting factors interact with X-linked “counting” element and block inactivation of single X in diploid cells. The identification of binding sites for factors like CTCF in \textit{Tsix} promoter region suggests its role in count and choice possibly through \textit{Tsix}/\textit{Xist} transcriptional regulation \cite{241}. \textit{Trans}-acting factors are also involved in interpreting the \textit{Xist} signal and establishing the heterochromatic state e.g. the Polycomb group proteins Eed/Enx (H3K27 methylation), Polycomb proteins Ring1A/B (mediating histone H2A ubiquitylation) \cite{242}. In addition, localized \textit{Xist} is essential for the accumulation of the histone variants macroH2A1 and 2 on the inactive X chromosome to form the macrochromatin body (MCB) \cite{243,244}. Because MCB formation is a relatively late event, appearing several days after the first appearance of \textit{Xist} expression and the first signs of inactivation, it is important for maintaining rather than initiating heterochromatinization \cite{245}. Interestingly, a study identified that chromosomal association and silencing abilities of \textit{Xist} RNA are functionally separable\cite{246}. BRCA1, the breast and ovarian tumor suppressor, also localizes to the inactive X and is especially interesting because it is the first protein found to be necessary for proper \textit{Xist} localization and stabilization of the inactive state \cite{247}.

Upon differentiation of ES cells in vitro, one of the first things that occur is the upregulation of \textit{Xist} RNA which starts coating the future inactive-X in cis which subsequently starts recruiting factors required for silencing involving histone modifications typical for heterochromatin. Besides few differences in opinion, there is general consensus about order of events occurring during developmental acquisition of X inactivation in mouse is shown in figure 11. The kinetics of X-inactivation during preimplantation development could be briefly summarized here (Reviewed in \cite{248}). Onset of \textit{Xist} expression takes place on paternal X-chromosome (Xp) at 2-cell stage followed by coating around at 4-cell stage which triggers initial silencing of Xp. There seems to be a resistance imprint on maternal X (Xm) that prevents the maternally inherited X to express \textit{Xist} and to get it inactivated at this stage. Subsequently, there is (transient) recruitment of Eed/Enx proteins, macroH2A association and H3K27 (due to Eed/Enx1 HMTase activity) or H3K9 methylation (HMTase not clear) of the X-chromosome. By the time blastocyst stage is reached, essentially all cells contain an inactive Xp associated with the PcG proteins Eed and Enx1. In the extraembryonic lineages (trophectoderm and primitive endoderm), the Xp is maintained inactive and supposedly gets frozen further by a shift in replication time (to early replication in this case). However in the ICM, the Xp in all cells gets reactivated, sheds off \textit{Xist} coating and PcG proteins and the histone methylation marks are gradually reversed. Thus there is a stage when the cells that will subsequently contribute to the epiblast (embryo proper) contain two active X-chromosomes before there is random inactivation of either Xm or Xp. The random inactivation involves similar steps as in the beginning of cleavage post-fertilization such as \textit{Xist} RNA coating, Eed association, histone methylation, late replication, macro H2A association and DNA methylation in a stepwise manner which ultimately leads to silenced state of one of the X-chromosomes that is
faithfully maintained through cell divisions in soma. In the female germ line, which is set aside subsequently, the inactive X becomes reactivated just prior to meiosis.

Figure 11. The Kinetics of X-chromosome inactivation during development (see text for details) (figure adapted from Heard E. 2004).

Meiotic sex inactivation (MSCI) is a process where there is transient inactivation of the single X chromosome that is observed during the pachytene stage of meiosis in males forming the densely stained sex chromatin body or X-Y body. There seems to be some differences in the mechanism and the role of MSCI and somatic X inactivation. Also there seems to be some mechanistic differences between mouse and human somatic X-chromosome inactivation as well. Although most of the genes on inactive X are silent, as much as quarter of genes on the same chromosome escape this repression and are transcribed. It would be of great interest to study the mechanism behind such escape. *Xist* inactivation and late replication happens several days before de novo methylation of CpG islands and therefore DNA methylation does not seem to be involved in establishing X-inactivation. However, it is involved in perpetuating the inactive state, which works in cooperation with e.g. continuous expression of *Xist* and deacetylation of histones etc [249]. Interestingly as discussed before, the maintenance of random X-inactivation in the embryo depends on DNA methylation, whereas maintenance of imprinted X-inactivation in extraembryonic tissues is dependent on Eed function and *Xist* expression but independent of DNA methylation [215]. To conclude, PcG proteins and associated chromatin modifying enzymes may be mediators of critical transition from the reversibly inactive, *Xist*-dependent phase of inactivation, to the irreversible, *Xist*-independent phase defined by Wutz and Jaenisch [250].
7. Epigenetic reprogramming in development:
Mammalian development is characterised by two phases of DNA methylation reprogramming that occurs initially during germ cell development and then during preimplantation as discussed here in detail (figure 12):

7.1. Reprogramming in early embryos:
The maternal genome is arrested at metaphase II with its genome packaged with histones. The paternal genome delivered by mature sperm is densely packed with protamines and relatively less histones. Soon after fertilization, maternal genome completes meiosis and the paternal genome undergoes massive chromatin remodeling which includes replacement of protamines with hyperacetylated histones and subsequent further modifications of histones. This is followed by genomewide demethylation event in the paternal genome which is completed before DNA replication begins in the male pronucleus (active demethylation) [251]. Interestingly there are regions in the paternal genome which escape this demethylation wave, for example, heterochromatin in and around centromeres IAP transposons and paternally methylated imprinted genes (e.g. H19, Ras Grf1 but not Igf2) and this might have to do with maintaining chromosome stability, suppression of IAP transposition events and the maintenance of parental imprinting, respectively [252,253,254]. Both the mechanism and function of paternal genome demethylation as well as how the demethylase activity is targeted to or excluded from specific sequences is not known. This paternal demethylation might provide an opportunity to reprogramme the paternal germline imprints by the maternally produced oocyte cytoplasm [255]. This gets support from observations that this zygotic demethylation of paternal genome is observed only in species with imprinting and the majority of germline imprints are found to be maternal. The other hypothesis is that it might be essential for proper transcriptional activation of paternal genome in embryo [256].

As cleavage progresses, there is further loss of DNA methylation which depends on DNA replication owing to the fact that Dnmt1 protein inherited from oocyte (Dnmt1o) is excluded from the nucleus during first three cleavage divisions [257]. This thus leads to ‘passive demethylation’ of both the genomes as there is no maintenance methylase around. Interestingly, imprinted genes retain their germline imprinting which provokes search for candidate DNMT(s) responsible and their targeting to proper sequences where DNA methylation is maintained during these stages of development. Dnmt1o enters into the nucleus only at 8-cell stage and it is speculated to be involved in maintaining methylation at imprinted loci [258]. Thus, the embryo’s genome is passively demethylated during early cell cycles before blastula. De novo methylation precedes after this demethylation event and roughly coincides with the differentiation of the first two lineages of the blastocyst stage: the inner cell mass (ICM) (embryo proper) and the trophectoderm (extraembryonic lineages e.g. placenta). There is global epigenetic asymmetry in the two lineages with the placenta relatively undermethylated for different sequences examined [259]. Also, placenta seems to be relatively independent of DNA methylation for maintenance of X-chromosome inactivation and imprinting at later stages (discussed in previous section). These differences are apparent around blastocyst stage where active and passive demethylation operating together have already resulted in low methylation levels in TE. In contrast, ICM gets extensive de novo methylated. The reason behind this de novo methylation differences could be attributed to the observation that the de novo methylase Dnmt3b is absent in TE but is detected in ICM in the blastocyst stage (expression beginning at E4.5) [260]. This DNA methylation differences are conserved among mouse, bovine, sheep and rabbit and thus seems to be important for the function or allocation of these lineages. The important
question that remains is how the unmethylated allele of imprinted genes is protected from de novo methylation carried out by Dnmt3a and Dnmt3b around the time of implantation. There are also differences in the histone modification patterns among the two lineages e.g. higher levels of repressive histone methylation patterns are detected in ICM than in TE. There are evidences indicating that this epigenetic asymmetry established in the blastocyst is indeed important for development. Perhaps there is need of special epigenetic controls in the ICM which form embryo proper. Although modifications are being re-established in much of the genome is the ICM, reprogramming of marks already established may occur to restricted parts of the genome. As discussed previously, there is reprogramming of inactive X-chromosome and one of the X is randomly inactivated.

At fertilization, maternal genome is already abundant in DNA methylation as well as chromatin modifications associated both with active and repressive nature. The progressive histone modification of the paternal genome goes on presumably leading to a chromatin state equivalent to that of maternal genome. This does not exclude paternal genome from acquiring unique epigenetic marks early on developments which are important for imprinting and X-chromosome inactivation [261][262]. The purpose behind this Postzygotic epigenetic reprogramming (demethylation and remethylation) could be removal of acquired epigenetic modifications (as for germ-line reprogramming) especially those acquired during gametogenesis. This could also be involved in first lineage decisions during mammalian development since methylation is intimately linked with differentiation [263][264].

7.2. Reprogramming in germ cells:

Primordial germ cells (PGCs) are derivatives of epiblast cells and they first arise from primitive streak at E7.5 (embryonic day 7.5) from where they begin to migrate on around E8.5 to the genital ridge reaching there by E11.5. Early PGCs are thought to harbour all the epigenetic marks the epiblast cells have including imprinted gene expression, random X-inactivation and DNA methylation patterns. Around the time when PGCs arrive the genital ridge, many of the mentioned epigenetic marks are erased. The methylation in imprinted genes and single copy genes is lost with the major demethylation event occurring between E11.5 and E12.5 [265]. This demethylation is completed by E13 to E14. The PGCs have entered gonads by this stage and most, if not all, DMRs in imprinted genes become demethylated due to this demethylation wave going on since couple of days. The imprinted genes express biallelically as well [266]. These demethylation events are a process of ‘active demethylation’ as Dnmt1 is still around in the nucleus. The similarities and differences between active demethylation here and one taking place in the zygote should be explored. There seem to be some interesting differences like, paternally methylated DMRs e.g. H19 ICR is demethylated in PGCs but not in zygotes [265]. But even this extensive erasure of epigenetic marks leaves some imprints since later on in spermatogenesis when H19 ICR becomes methylated again, the originally methylated alleles become de novo methylated at an earlier stage than the originally unmethylated ones [267]. Perhaps other marks like histone modifications could guide the process.

Once the genomes of male and female PGCs are demethylated, the cells enter mitotic (male) until about birth when mitosis of spermatogonial cells is resumed) and meiotic arrest in prophase of meiosis I (female) (this arrest is released on puberty). This arrest has been implicated to be of significance during development for example, it could be advantageous not to replicate hypomethylated genomes (e.g. demethylated centromeres are condensed and functionally altered) [268] and also demethylation leads to higher frequency of chromosome structural abnormalities [269].
Several days later, remethylation takes place in male germ line (bit earlier than female) at the prospermatogonia stage (E15 to E16 and onwards). This remethylation is followed by re-entry of these cells into mitosis and subsequent meiosis to produce mature sperms. The limited evidences available so far indicate that the remethylation of imprinted genes and the single copy sequences occurs at the same time. Remethylation in female germ line takes place but later than male after birth during the growth of oocytes. The enzymes that lead to this de novo methylation are still unknown. There could be various roles of reprogramming in the germ cells, the most important being resetting of imprints. The other purpose could be removal of acquired epigenetic modifications, which can be influenced by individual genetic and environmental factors [255] [258] [270] [271].

![Figure 12](image)

**Figure 12.** Epigenetic reprogramming taking place in the germline (A) and preimplantation embryos (B) (see text for details). Dark grey is used for maternal genome and light grey is for paternal genome in figure A and up to (\(\text{H}\)) mark in figure B and after the mark these colours represent EX and EM respectively. EM and EX stand for embryonic and extraembryonic lineages respectively. Upper dashed line in (B) represents methylated imprinted genes and some repeat sequences. The lower dashed line in the same figure shows unmethylated imprinted genes (figure adapted from Reik, Dean et al. 2001).

The epigenetic reprogramming has been studied at *Igf2-H19* locus to a relatively large extent. The three DMRs in murine *Igf2-H19* domain (*Igf2 DMR1, Igf2 DMR2 and H19 ICR*) are fully methylated in sperm and unmethylated in oocytes [233]. *H19 ICR* seem to harbour germline methylation imprint since the methylated status of paternal allele and unmethylated status of maternal allele is maintained in the embryo after fertilization. In contrast, both DMRs in *Igf2* loose paternal methylation soon after fertilization (by Morula and blastocyst stage, all the paternal methylation is erased consistent with observation for most germline imprints). *Igf2* is biallelically
expressed at blastocyst stage. How H19 ICR is protected from demethylation is not understood. Paternal methylation imprints become established at post implantation stage (Igf2 shows monoallelic expression again). Interestingly, in largely mesodermal tissues such as kidney, an increase in DMR1 methylation takes place before the increase in DMR2 methylation temporally. As discussed elsewhere in this thesis, this locus exhibits control of epigenetic modifications by a hierarchy of DMRs which led to proposal of long-range interactions. Here it was shown that maternal H19 ICR protects Igf2 DMR1 and Igf2 DMR2 and Igf2 DMR1 protects Igf2 DMR2 from de novo methylation during postzygotic embryonic development. We don’t know what protects these regions from de novo methylation during oogenesis. The factors which lead to methylation of all three DMRs during spermatogenesis are also not known.

7.3. CTCF, BORIS, Lsh and epigenetic reprogramming during male gametogenesis:

BORIS (Brother Of the Regulator of Imprinted Sites) is a protein named after its discovery where it was found to share similarity with CTCF in all the zinc fingers [272]. Homology analyses revealed that BORIS is a parologue of CTCF, having been generated by a duplication event (descended with divergence from a common ancestral gene followed by transposition event). These two proteins only differ at their C and N-termini and this in association with findings that BORIS and CTCF can compete for similar sites; we speculate that the same target site might elicit different responses depending on whether it is binding CTCF or BORIS because the protein-partners may be different due to differences at their ends. This could also deregulate those genes which are involved in cell cycle and growth normally regulated by CTCF. Interestingly, BORIS is only expressed in adult testis and its aberrant activation is frequently being noticed in several human cancers and this places this in the unique family of ‘Cancer-Testis’ genes (CT genes) which are normally expressed only in testis but abnormally activated in different malignancies [272]. But BORIS is a unique CT gene, since unlike others it has a somatic counterpart with the same DNA binding domain but with the properties of a tumour suppressor.

Interestingly, when Loukinov et al compared the staining pattern for BORIS, CTCF and 5-meC using specific antibodies, they found that the two proteins BORIS and CTCF were expressed in a mutually exclusive manner during male germine development [272]. The BORIS-CTCF switching is proposed to be intimately linked with the initiation of de novo methylation post-erasure [273]. Our recent observations say that BORIS is expressed at high level both in spermatocytes and spermatids at low level in spermatogonia, suggesting that BORIS is activated at spermatogonia stage [Paper (IV)]. On the other hand, CTCF is expressed throughout spermatogenesis, although the signal is highest in spermatids. Our additional observations led us propose that the epigenetic reprogramming of H19 ICR during spermatogenesis is initiated by a CTCF-dependent recruitment of chromatin remodeling factor Lsh to the H19 ICR followed by completion of the imprint acquisition process by a replacement of CTCF with BORIS [Paper (IV)].

Two Recent studies, by Hong JA and Vatolin et al., indicate that BORIS is activated in a variety of human cancers, and that competition between BORIS and CTCF may contribute to epigenetic perturbations in these malignancies. Overall, the results indicated that CTCF binding to promoter of a specific class of gene encoding cancer-testis (CT) antigens such as NY-ESO-1 was associated with silencing with the gene whereas switching from CTCF to BORIS occupancy at the same site coincided with derepression of the same gene in lung cancer cells [274]. Similar kind of results were obtained in another independent study [275].

Presumably, BORIS might interfere with CTCF-dependent transcriptional regulation in cancer cases when overexpressed. The observation that CTCF and BORIS
staining almost doesn’t overlap in male germline cells [Paper IV] indicates that aberrant BORIS upregulation in cancer cells and in epigenetic reprogramming events during spermatogenesis may lead to relocation of CTCF-bound sequence to new microenvironment where they are differently regulated. This change in nuclear positioning might also lead to changes in replication timing, relocation to a transcriptionally repressive or active environment causing direct disturbances. CTCF to BORIS switch might also cause transcriptional deregulation for pivotal genes e.g. those associated with cycle related control, cell growth and thus facilitate cancer progression events. This gets support from observations that CTCF binds the proximal exonic region of hTERT (the catalytic subunit for telomerase activity) and inhibits its transcription and hence, aberrant expression of BORIS might lead to derepression of the gene, thereby promoting cell survival in cancer cases [283].

Lymphoid-specific helicase (Lsh) belongs to a family of SNF2 chromatin-remodeling proteins, and is very crucial for normal embryonic development and its targeted deletion is lethal [276]. Loss of Lsh resulted in global loss of CpG methylation including repetitive elements and single copy genes. The loss of CpG methylation was most substantial around pericentric heterochromatin sequences. Lsh -/- cells were shown to have abnormal patterns of histone methylation [277]. Later on, observations suggested that Lsh is recruited by intact pericentric heterochromatin structure which then assists in maintaining intact heterochromatin structure by establishing CpG methylation patterns [277]. A role for Lsh cooperation with DNA methylation to mediate epigenetic control of imprinting was demonstrated by findings that it associates with the promoter of Cdkn1c and the deletion of Lsh resulted in biallelic expression of the gene concomitant with the substantial reduction of CpG methylation at the 5’ DMR of the Cdkn1c gene [278]. Recently, researchers presented evidences that Lsh is involved in the process of de novo methylation at retroviral genes, episomal targets and endogenous sequences and that Lsh interacts with Dnmt3a and Dnmt3b in ES cells but not with Dnmt1. [279]. This indicates that Lsh, being a member of the SNF2 chromatin remodeling family might function to facilitate access to some chromatin templates and thus promote targeting of Dnmt3a or 3b to specific genomic targets or , it may facilitate the stable association of these methyltransferases with their specific targets and promote the methylation efficiency of nucleosomal templates or condensed chromatin. Our unpublished observations indicate that Lsh is recruited by CTCF to unmethylated H19 ICR allele during male germline development suggests that it might be linked with initial events in the epigenetic reprogramming process that lead to conversion of maternal to paternal epigenotype [Paper (IV)].

9. Long-range chromatin interactions and disease:
The extensive involvement of long-range chromatin movement and interactions raise the possibility that various transcription deregulation disorders might have its roots in these higher order chromosome conformations. We have already exemplified Rett Syndrome in detail which is the first human disorder where loss of looping has been directly implicated [67]. The variety of symptoms seen in the heterogeneous group of genetic disorders called ‘laminopathies’ indicate that it is directly/indirectly due to alteration in lamin-regulated gene expression in a genomewide fashion [167]. As already said, the preferential spatial organization of translocation-prone loci might have implications for the genesis of cancers (preferred chromosomal neighbourhood, interchromosomal interactions etc). Interchromosomal interactions observed also implicate that mutation of a gene could affect transcription of another gene(s) found in trans and this could have implication in the disorders where we don’t see a direct correlation between the mutated gene and the phenotype observed [280]. In support of above, there are unpublished observations from
our lab where we find that mutations of CTCF target sites at a particular locus can affect epigenetic state and transcriptional potential of another locus in trans which adds novel perspectives to our understanding about various disorders involving aberrant gene regulation (Tiwari VK et al unpublished observations). Nuclear compartmentalization might be advantageous to expose or restrain a specific gene or gene locus from an elevated or decreased concentration of a specific factor. In this context the formation of aggregates or inclusion bodies in nuclei of patients with polyglutamine diseases may be an essential process to maintain cell survival under stressed disease condition [281]. The various imprinting disorders could be due to loss or disturbance of looping interactions found at loci such as Igf2-H19 (Murrell et al 2004) due to diverse reasons including loss of insulator function, altered epigenetic patterns of various cis elements involved directly or indirectly in establishing and/or maintaining long-range interactions. The alterations in differentiation regulated inter- and intrachromosomal interactions might pave a role for development related disorders and could have links to cancer development as well since there are some incidences where the stem cells have been intimately related with cancer development [282]. This is high time to develop techniques to compare genomewide looping interactions in normal vs. diseased state which might give us new insights into the mechanisms of disease progression as well as new directions to develop potential diagnostic markers.
AIMS OF THE PRESENT STUDY

This thesis highlights novel roles for CTCF in epigenetic regulation of gene transcription and higher order chromosome conformation.

X chromosome inactivation (XCI) operates in mammals to achieve dosage equivalence for most X-linked genes between the two X chromosomes in females and the single X chromosome in male. Our study here was aimed to find a correlation between the known familial cases of skewed X-inactivation patterns with the binding affinity of CTCF in the mouse/human \textit{Xist/XIST} promoter region.

Genomic imprinting affects several mammalian genes and leads to allele-specific expression depending on the parent of origin of the allele. The major studies in this work were aimed at studying complete higher order chromatin organization of \textit{Igf2}\,-\textit{H19} locus, the paradigm of genomic imprinting, in interphase cells with a subsequent analysis in mitotic cells. The investigations aimed to find a group of long-range interactions and the factors which are pivotal for imprinting regulation at this locus.

The epigenetic reprogramming taking place during germline development following erasure of epigenetic marks on parental alleles is a crucial process for precise resetting of imprints and thereby proper development. The final study aimed to find out the role of CTCF to BORIS switch observed during male germline development in epigenetic reprogramming events.

The following questions were the main concerns during the course of investigation:

- Is it possible that the differential affinity of CTCF to the same target site interpreted differently on a functional basis?

- Does activation of \textit{Igf2} promoters by downstream enhancers need a physical contact between them?

- What is the mechanism by which imprinting control element \textit{H19} ICR physically insulates \textit{Igf2} promoters from downstream enhancers exclusively on maternal allele?

- What roles do CTCF target sites play in the higher order chromatin organization as well as coordinating regional epigenetic marks at \textit{Igf2}\,-\textit{H19} locus?

- How the epigenetic modifications at \textit{Igf2}\,-\textit{H19} locus are co-ordinately regulated in a hierarchical manner? Do long-range regulatory loops contribute to the stable maintenance of the epigenetic state here?

- What happens to the long-range regulatory loops during mitosis when the chromosomes are highly condensed?

- Does CTCF to BORIS switch during spermatogenesis have any role to play in the maternal to paternal epigenotype reprogramming?
RESULTS AND DISCUSSIONS

PAPER I

The phenomenon of dosage compensation operates in female mammals resulting in one X chromosome getting inactivated randomly in soma, parent-of-origin dependent in extraembryonic tissues. The inactivation process involves coating by a noncoding RNA Xist in cis which in turn seems to be negatively regulated by another noncoding RNA Tsix (expressed on active X in mouse). There are several other regulatory elements like Xite and other non-coding transcripts at 5' end of Xist and the interplay between them is not very clear. But overall, all of these known elements and perhaps additional candidates on both sides on Xist play together to set a stage for a balance between sense and antisense transcription prior to Xist transcription and subsequent X inactivation. Skewing in X-inactivation pattern could result from affecting any of the known regulatory elements including the Xist promoter activity.

Interestingly we found two human cases of skewed X-inactivation with great similarity at the molecular level. In one case, families carrying a germline mutation C (-43) G in Xist promoter region showed preferential inactivation of the X chromosome bearing it [285]. Another was a case of ring X-chromosome of a 3.5 year old girl, who carried C (-43) A germline mutation at the same position and the X-chromosome bearing it was associated with escape from X inactivation [285]. This was an interesting observation where mutation at the same position was resulting in opposite choices for X-inactivation and there could be two major possibilities. First, the mutations might have created a novel binding site where recruitment of new factor leads to different fates of X-inactivation choice. Alternatively, the two kinds of mutations are affecting the affinity of same transcription factor in opposite way for Xist promoter. Using various experimental approaches, we attempt to solve the mystery and try to explain the causal mechanism behind such skewed X-inactivation patterns observed.

CTCF specifically recognizes mouse/human XIST/Xist promoters in vitro:

We carried out a systematic search for CTCF binding on human and mouse by EMSA using several long overlapping fragments for XIST/Xist promoter and recombinant CTCF protein. Sequence-specific binding of CTCF with one fragment was observed for both human and mouse Xist promoters, for the region which contains the XIST/Xist transcriptional start site. Hendrich et al have already shown that this region is nearly 78% homologous in several species including human, mouse, rabbit, and horse. This in turn potentiates the possible functional role of this region [286].

CTCF occupies active mouse/human XIST/Xist promoters in vivo:

Chromatin Immunoprecipitation (ChIP) experiments were carried out using CTCF antibody to validate the above in vitro results in human and mouse in vivo. Experiments with primary cultures of human mammary epithelial cells (HMEC) and normal human dermal fibroblasts (NHDF) derived from male and female revealed that CTCF pulls down Xist promoter only in female human cells. Instead, e-MYC Oncogene 5'-insulator site (N-site) was pulled down in both. The control antibody did not give any signals. Since, females have one inactive and one active X chromosome, but males have only one active X which lead us to speculate that CTCF may be binding only to the active Xist promoter (inactive X chromosome) but due to the absence of markers we could not validate this possibility completely.
To get a deeper insight into the study of CTCF-active *Xist* promoter (inactive X) association, we switched our studies to mouse placentas. Placenta is an extra-embryonic tissue that exhibits non-random X-inactivation and only the X-chromosome coming from father gets inactivated. We used dispersed formaldehyde –crosslinked cells from male and female placentas derived from interspecific *M. musculus* (C57BL6J) and *M. spretus* crosses for ChIP, where we knew polymorphism in *Xist* promoter between two strains to distinguish between alleles. Anti-CTCF antibody pulled down *Xist* promoter only from female placentas whereas *H19* ICR was equally precipitated by CTCF in both male and female placentas. To directly verify that CTCF might be associated with active *Xist* promoter i.e. inactive X-chromosome only, we sequenced this ChIP DNA sample. This approach was based on an A/C polymorphism at -88 upstream of the transcriptional start site of the *Xist* gene in *M. musculus* (C57BL/6J) compared with *M. spretus*. Interestingly, the sequencing showed that the CTCF antibody pulled down only the paternally inherited allele (from *M. spretus*), confirming our predictions that CTCF occupies *Xist* promoter of the inactive X chromosome in female mouse placenta and, by extrapolation, also to female humans.

The identification of this CTCF binding site alongwith the earlier mapped target sites for it in the chromatin insulator region at the 3' of *Xist* suggested a dual function of CTCF in the X-inactivation process and might involve higher order chromatin conformations [241].

**Germline mutations in CTCF target site at Xist promoter in families with skewed X inactivation and CTCF occupancy:**

The occurrence of two different kinds of point mutations at the same site upstream of *Xist* transcription start site with opposite skewing patterns of X-inactivation choice and the fact that it occurs in a highly conserved fragment harbouring CTCF binding site lead us speculate that these mutations might be affecting CTCF binding affinity. To verify our speculations, we carried out EMSA using in vitro translated CTCF and nuclear extracts individually with human promoter fragment from -108 to +31 containing either wild type sequence or C(-43)G or C(-43)A mutations. Interestingly, C(-43)G mutations (preferential inactivation) dramatically enhanced the affinity of the probe as compared to the wild type whereas C(-43)A mutations (escape from X-inactivation) abolished CTCF binding. Competitive EMSA carried out using labelled probes from wild type and two mutation bearing fragments with cold competitors derived from known high affinity CTCF target site at DMD4 at mouse *H19* ICR revealed that an excess of cold competitor was required to compete out CTCF binding with probe bearing C (-43)G mutation as compared to the wild type probe and C (-43)A mutation probe. More competitive EMSA were done with probe swapping and using other target sites as competitors and all of them ended up with conclusion that C(-43)G mutations led to increase in CTCF binding to the promoter region and C(-43)A mutations was associated with reduction of CTCF binding. Further experiments revealed that there were differences in base-contacting points in CTCF-DNA complexes formed with the wild type and mutated *Xist* promoters. Interestingly CTCF was found to utilize different combinations of ZFs (zinc fingers) for interaction with the wild type and C (-43) G mutant *Xist* minimal promoter.

**Enhanced affinity of CTCF to C (-43) G mutant Xist minimal promoter is paralleled by gain of enhancer-blocking activity:**

CTCF binding to its target site parallels with enhancer-blocking activity in an affinity-dependent manner (Mukhopadhyay et al 2004). We wanted to test the same here whether the increased affinity of CTCF protein to the C(-43)G mutant promoter leads to
any increment of enhancer-blocking activity. To test this, we cloned various fragments like mouse and human wild type XIST/Xist promoter, mutant promoters (both mutants), dG contact mutant forms of promoter (eliminates CTCF binding), known HS4 insulator in between the β-globin LCR (enhancer) and the neomycin reporter gene. We counted the number of colonies after transfection and neomycin selection. While the wild type and dG mutant insert plasmids showed negligible enhancer-blocking activity, C(-43)G mutation lead to increase in this activity as compared to the controls but the effect was not as strong as the HS4 insulator for some reasons in the latter. This gives strength to the possible function role of CTCF binding here.

We speculate that the insulator protects from upstream repressive cis elements in the affected patients with C (-43) G mutation [286]. Unpublished data from our collaborators say that CTCF is a component of RNA Pol II and has extensive overlap with binding site with the same genomewide (Klenova E and colleagues). This led us think that the CTCF target site discovered here might facilitate the recruitment of the pre-initiation RNA polymerase complex to enhance XIST–transcription. Increased affinity at C (-43) G mutated promoter will result in more stable or frequent recruitment of the same, thereby enhanced Xist transcription and preferential inactivation of the X-chromosome harbouring it.

The data here indicate that the mutations occurring here must be interfering with the random choice of XCI in affected patients. The choice mechanism in somatic cells might be governed by the regulating the affinity of CTCF with the wild type XIST/Xist promoters. There are various possible candidates for it like Tsix transcript itself here which could affect CTCF-XIST/Xist promoter interactions prior to the onset of the inactivation process as its transcription process extends beyond the Xist promoter (Lee JT et al 1999). Since we know that CTCF has a pivotal role in mediating higher order chromatin organizations from paper II, we could extrapolate the same here and postulate that XCI requires CTCF-dependent organization of higher order chromatin conformation that relies on the stability of CTCF-XIST/Xist promoter complex during early stages of XCI before the choice is made.

Paper II:

The imprinted regulation of H19 and Igf2 expression involves binding of the vertebrate insulator protein, CTCF, to the maternally hypomethylated differentially methylated region, called H19 ICR. Igf2 region also has DMRs, DMR1 (methylation-sensitive repressor), DMR2 (methylation-sensitive activator). All the three DMRs (Igf2 DMR1 and DMR2, H19 ICR) are unmethylated maternally and methylated paternally. This methylation status is not due to linear spreading of methylation and it was shown that that H19 ICR protects Igf2 DMR1 and Igf2 DMR2 against de novo methylation in the soma in preimplantation conceptuses and embryonic stem cells (Lopes et al 2003). The proposal that this might involve higher order chromatin conformation got strength by recent observation by Murrell et al where they showed that H19 ICR physically interacts with Igf2 DMR1 and Igf2 DMR2 on maternal and paternal alleles respectively which was thought to partition Igf2 - H19 domain into active and silent compartment (Murrell 2004). This would allow enhancer access to Igf2 only on maternal allele. Although the model was warmly welcomed, it did not suffice for all the answers we needed. For example, what proteins play role in such looping organization, are the proteins already known to bind to various regions in this domain e.g. GCF2 (Igf2 DMR1), CTCF (H19 ICR) etc have a role to play here? Also, we did not know how exactly H19 ICR controls methylation of Igf2 DMR1 and Igf2
DMR2. Do various MARs in this locus have any role in imprinted gene expression, for example, in cooperation with DMRs in organizing proper loops? Moreover, we needed answers for questions like the mechanism by which enhancers activate monoallelic transcription at this locus and how CTCF interferes with its action. Does the higher order organization involve any other regulatory regions of the locus besides \textit{Igf2} DMRs and \textit{H19} ICR and if any, what role do they have to play? To find answers to these questions, we carried out a systematic 3C analysis across the entire \textit{Igf2} -\textit{H19} domain, and we analyzed the role of CTCF target sites at ICR on higher order chromatin conformations and epigenetic state of the locus using CTCF-target sites knock-in mice. The findings are discussed briefly here:

**Parent-of-origin dependent interaction of downstream enhancers across \textit{Igf2} -\textit{H19} locus:**

We used Chromosome Conformation Capture (3C) technique to scan for physical interaction of endodermal enhancers along the entire \textit{Igf2} -\textit{H19} locus (figure 1)[287]. The principle of the technique is very simple; briefly, cells are treated with formaldehyde to crosslink proteins to other proteins nearby and DNA. This crosslinked complex is then subjected to cleavage by a restriction enzyme, followed by ligation at low DNA concentration. Under such conditions, ligations between crosslinked DNA fragments, which are intramolecular, is strongly favoured over ligations between random fragments, which are intermolecular. After ligation, the crosslinks are reversed and ligation products are detected and quantified by polymerase chain reaction (PCR). The crosslinking frequency of two specific restriction fragments, as measured by the amount of corresponding ligation product, is proportional to the frequency with which these two genomic sites interact (Dekker et al., 2002). Thus, 3C analysis provides information about the spatial organization of chromosomal regions in vivo (figure below).

**Figure 1. Chromosome Conformation Capture (3C) methodology**

We chose EcoRI for this enhancer scan study since its recognition sites fall in most of the reasons of interest. Using en4 (enhancer 4) as bait, we analyzed its crosslinking frequency with the different Eco RI fragments of the locus in the crosses C57 X SD7, 142* X SD7 and SD7 X 142* mice. 142* mice are CTCF target sites mutant at \textit{H19} ICR created...
by point mutation at three of the four CTCF target sites and does not bind CTCF [223]. The neutralization of crosslinking frequency and corrections for allelic bias in the PCR amplification step was important to make conclusions in this assay. The PCR products were restriction digested with restriction enzyme whose recognition sequence is specifically present in SD7 en4 fragment to distinguish the allelic-origin of these interactions. Neonatal liver cells from wild type (C57XSD7) and paternal inheritance of the mutation (SD7X142*) gave identical results. The en4 region was in close physical proximity on both parental alleles from L23 mrp gene upto the H19 ICR. Further analysis showed that the maternal Igf2 promoters do not have any access to the enhancers consistent with observations that maternal Igf2 is silent. The data doesn’t support the previous proposals that H19 ICR blocks enhancer action beyond it completely since we found that there was a substantial interaction between en4 and IGS1 and to a lesser extent with HSS and IGS2. Further S’ of these regions; however, there was no enhancer access to the maternal allele. This also shows that suppression of the paternal H19 gene doesn’t require its exclusion from proximity with enhancers and might be determined by its epigenetic status (for example, methylated promoter resulting from methylation spread from H19 ICR on paternal allele) and some other factors. Interestingly, the same study done with maternal inheritance of the mutation (142* X SD7) revealed that the whole of the Igf2 region, including the promoters, was now accessible to the en4 enhancers clearly demonstrating that the CTCF target sites have a direct role in preventing the access of downstream enhancers to the Igf2 promoters on maternal allele.

These results revision our thinking about insulator action where we see that enhancer-blocking is not an absolute phenomenon since we observe minimal interactions of enhancer with upstream of ICR as well. Although the enhancer action here is supposed to be blocked at H19 ICR, we did not see an absolute hindrance; rather it is diminished slowly as a gradient beyond ICR. The enhancer-blocking, being a biological system, can never be an absolute phenomenon and can never be all-or-none. The hypersensitive sites (found on both parental alleles) in the intervening region between Igf2 and H19 genes which are shown to have enhancer functions in both imprinted and non-imprinted tissues might have role to play in this leakiness.

The H19 ICR is in close physical proximity with Igf2 cis-regulatory elements in parent of origin-specific manner:

Since H19 ICR has been implicated in several studies to be crucially important for imprinting regulation, we thought next to analyze how this region regulates enhancer access along the locus by the same methodology as above but this time using H19 ICR as common platform, instead of en 4. A Fau l polymorphic restriction site (present only in C57) 61 bp downstream of the EcoR1 restriction site within the ICR was utilized to discriminate between parental alleles. The results were interesting: the ICR was found to be in close physical proximity to Igf2 DMR1, supporting previous observations by Murrell et al and MAR3 exclusively on maternal chromosome whereas these fragments were excluded from interaction on paternal chromosome, instead it was in close proximity to P1/DMR2 fragments [50]. This promoted the idea that on the maternal allele, the Igf2 promoters are located in a tightly closed loop generated due to physical proximities of H19 ICR, Igf2 DMR1 and MAR3 and this loop renders the Igf2 promoters non-communicable with downstream enhancers exclusively on maternal allele. The paternal chromosome conformation where ICR-P1/DMR2 complex is formed lacks this closed pocket and allows activation of Igf2 promoters by enhancers. Interestingly, when the CTCF target site mutations are inherited maternally, we observed the loss of ICR-DMR1 and ICR-MAR3 complex, indicating that the closed pocket structure is lost. This corresponds well with the
biallelic Igf2 transcription observed in these mice [223] along with our data in previous section that enhancers now get access to the maternal Igf2 as well. Thus, CTCF target sites are very crucial for this tight loop formation exclusive on maternal allele which is suppressive for Igf2 transcription.

These results emphasize reforms in the simple looping model proposed by Murrell et al as we find here that there are multiple long-range interactions occurring at this locus and it is difficult to underestimate the importance of any of them. It is possible that the downstream enhancers may be metastable part of a transcription factory or an active chromatin hub (ACH) on both alleles and that is where the different parts of the locus are differentially recruited. The recruitment in turn might be the result of the interplay between epigenetic status of various regulatory regions, stability and specificity of various interdependent and independent loops formed throughout the locus. Interestingly, there is possibility that multiple interactions occurring throughout the locus are required to stabilize the active chromatin hub [80]. A deeper study using various mutant/deletion mice for various regions of this locus is needed to signify the clear importance of these several interactions across the locus.

CTCF binds Igf2 DMR1 on maternal allele and relies on long-range Interactions with CTCF-occupied H19 ICR:

To investigate deeply about the H19 ICR-DMR1 interaction, we chose another enzyme site Hind III for 3C analysis. Choosing a new restriction enzyme will do two jobs here: first this site falls more precisely in the ICR and DMR1 region and second, this will make sure that the results we get are not affected by fragment size/enzyme sites. The results were comparable showing that ICR interacts with DMR1 only on maternal allele and this physical proximity is lost on maternal inheritance of CTCF target site mutations. If this interaction occurs between CTCF occupied ICR and DMR1, CTCF could be part of the complex where these two remote sequences are tethered. In order to investigate this, we combined the 3C analysis with chromatin immunopurification (ChIP), a method termed ChIP-loop assay [67]. Here formaldehyde cross-linked chromatin from livers of wild type mice was restriction digested with Hind III enzyme, as for the 3C step in the beginning, followed by immunoprecipitation using a CTCF antibody. Later, this DNA-protein complex was ligated under high dilution conditions, reverse crosslinked, and amplified using specific primers as for the usual 3C assay. As expected, we found that CTCF indeed is part of the complex that holds ICR and DMR1 together. The PCR product was specifically seen in CTCF antibody ChIPed samples but was absent in controls including the control serum.

Previous observations from our lab that different CTCF-DNA complexes have the ability to interact with each other allured us to think that ICR-DMR1 complex observed here might involve CTCF occupancy on both partners of the loop [186]. Gel-shift assays using various overlapping fragments of Igf2 DMR1 revealed that two overlapping fragments were interacting with CTCF. This CTCF positive fragment harbours five of the seven CpGs in DMR1 we have analyzed by bisulfite assay later on. Since CTCF is known to have methylation-sensitive binding properties, we assayed for it here and found that only CpG number 5 interacts with CTCF in a methylation-sensitive manner [56] [57]. To recapitulate the same results in vivo, we carried out ChIP assay in liver samples. The results showed that indeed CTCF occupies DMR1, exclusively on maternal allele. Interestingly, this CTCF occupancy on DMR1 was lost when the mutated H19 ICR was inherited maternally. So far in our knowledge, this is the first time where it is shown that CTCF binding at a particular genomic region depends on remote cis acting sequences and requires long-range interactions to maintain them.
**Igf2 DMR1-CTCF-H19 ICR complex regulates the epigenetic state of Igf2 DMR1 and Igf2 DMR2:**

The observation that (methylation-sensitive) CTCF binding at Igf2 DMR1 depends on long-range interactions with CTCF-occupied H19 ICR in vivo, prompted us to test whether the loss of CTCF binding at ICR and thereby this loop has resulted in any aberrant methylation patterns at DMR1. Bisulfite analysis was carried on liver DNA extracted from mice carrying the paternal (control) and maternal inheritance of the mutant H19 ICR allele. The control inheritance showed that the maternal DMR1 was comparatively less methylated than the paternal one in consistent with previous reports [223]. Instead, on maternal inheritance of the mutation; there was overall increase in methylation of maternal DMR1 allele. A close analysis of the CpG number 5, where CTCF occupancy was found to be methylation-sensitive, showed that there is a striking change at this CpG while comparing results from mice with maternal and paternal inheritance of the mutation. In the latter case, the paternal allele was highly methylated at this particular CpG whereas the maternal allele was hypomethylated. But the maternal derivation of mutant H19 ICR resulted in hypermethylation of both alleles at this site. This clearly demonstrates that loss of CTCF occupancy at CpG number 5 of Igf DMR1 led to loss of protection of this CpG against de novo methylation. This analysis also emphasizes that certain CpGs in an island could be functionally more crucial than other ones around. Additionally, there is also possibility that different CpGs might have different roles to play in a genomic segment. Since Igf2 DMR1 acts as a silencer for Igf2 transcription on the maternal allele where it is undermethylated and it is inactivated by hypermethylation on paternal allele. This might work together with differential enhancer access to ensure monoallelic Igf2 expression exclusively from paternal allele. On maternal inheritance of the mutant H19 ICR allele, paternal DMR1 is getting hypermethylated and this might lead to loss of its silencer activity there, in turn permitting Igf2 transcription in the situation when enhancers have biallelic access to the Igf2 promoters.

As discussed before, Igf2 DMR2 is shown to be a methylation-dependent activator and it augments Igf2 transcription on paternal allele where it is hypermethylated. Since H19 ICR has also been proposed to protect this region from de novo methylation on maternal allele, we carried of bisulfite analysis of this region in the mutant mice. Interestingly, we found that the maternal inheritance of the mutation resulted in hypermethylation of maternal Igf2 DMR2 very similar to the paternal allele whereas the control (paternal inheritance of mutation) showed methylation levels comparable to the wild type mice. This hypermethylation observed here may be coordinating with biallelic enhancer access and hypermethylation of maternal Igf2 DMR1 observed here to result in biallelic Igf2 transcription observed in these mice because now this might help boost Igf2 transcription on maternal allele as well, being methylation-dependent activator. Thus H19 ICR also prevents de novo methylation of the maternally inherited DMR2 allele and again it relies directly on CTCF occupancy at H19 ICR and thereby on ICR- DMR1 loop existing on maternal allele.

Thus overall, Altered methylation in DMR1 and DMR2 could additionally contribute to expression of the maternal Igf2 allele, besides the biallelic enhancer action observed here on the maternal inheritance of the CTCF target site mutant H19 ICR allele. Also it is interesting to realize that CTCF binding sites could regulate the epigenetic potential of regions situated several Kb in cis.

These results are first to explain imprinted gene expression on the basis of detailed higher order chromatin analysis of whole locus. Especially, the result gives a deep insight into the mechanisms by which ICRs/insulator proteins function by regulating three dimensional spatial organization of a particular locus. If it is a general property of
insulators could be tested by recently identified number of CTCF-dependent chromatin insulators throughout the genome [187].

Paper III:

The onset of mitosis is characterized by chromosome condensation and global silencing of all three polymerase driven transcription. Long back, Gazit et al made the observation that hypersensitive sites are preserved for some genes in mitosis as compared to interphase chromatin [288]. The precise location of these sites could vary for some genes showing regulatory differences between mitotic vs interphase stage [289]. The condensed structure of mitotic chromosomes was found to exclude many transcription factors e.g. RNA polymerases [290] and to retain some e.g. Poly (ADP-ribose) polymerase [291]. Although these correlations were made, there was no functional explanation for the exclusion or inclusion of these factors. The constitutively bound factors were thought to be either important for condensed chromosome structure or markers for accurate propagation of gene expression throughout the cell cycle. In a recent study by Chen D et al, using FRAP (fluorescence recovery after photobleaching) the authors demonstrated that RNA polymerase I transcription factors had very high exchange kinetics with rDNA genes in the condensed NORs (Nucleolar Organizer Regions) [292]. This shows that condensed mitotic chromatin might allow selective accessibility for certain proteins.

Our findings from paper II revealed that imprinting regulation at \( Igf2 - H19 \) locus requires CTCF-dependent long-range chromosomal interactions among various regions on maternal allele, but we didn’t know the fate of the regulatory loops in the locus when the chromosomes are condensed rod-like entities during mitosis. This was an important question to ask because the genome is divided into distinct functional or expression domains and each one of them needs to be insulated from each other so as not to allow transcriptional chaos. This also applies for other CTCF-dependent and independent insulators/boundaries throughout the genome. This prompted us to investigate in detail about the CTCF binding on mitotic chromatin and therein specific targets as well as the higher order structures regulated by CTCF in mitosis.

**CTCF is associated with mitotic chromosomes and it requires C-terminal zinc fingers:**

HeLa cells were fixed at different stages of mitosis and with ethanol/acetic acid and probed with anti-CTCF antibody followed by FITC-conjugated secondary antibody. DNA was stained with propidium iodide stain. Confocal imaging found that CTCF was constitutively bound to mitotic chromosomes in all stages of mitosis from prophase to telophase. It also stained centrosomes and midbody in addition to the chromosomes here. A western blotting done on biochemically purified mitotic chromosome fraction gave strong signal for CTCF confirming that CTCF is part of mitotic chromatin. To verify these findings in more living conditions, confocal microscopy was done on mitotic chromosomes in cell lines expressing GFP fused at C or N termini of CTCF which confirmed that CTCF is an integral part of mitotic chromosomes. To confirm whether this CTCF binding is the result of direct DNA binding through its zinc fingers, transient transfections were carried out using various GFP-deletion constructs. The observations revealed the mitotic chromosome targeting region maps to the C-terminal half within the zinc finger domain (DNA binding property of CTCF), implicating this mitotic chromosome adherence is the result of direct interaction with chromatin.
CTCF constitutively occupies known target sites in mitosis both in human and mouse:

The observations made do not explain that the occupancy we see here is due to CTCF binding to its specific target sequences or something else. To solve this issue, we selected few already known targets of CTCF in human viz $\beta$-globin insulator (Tanimoto et al, 2003), the $H19$ locus (Bell and Felsenfeld, 2000), the DM-1 locus (Filippova et al, 2001) and the myc-N site (Lutz et al, 2003) and mouse viz. amyloid $\beta$-protein precursor (APP) promoter (Unpublished), the $H19$ ICR at the mouse Igf2-$H19$ locus and two additional CTCF target sites Igf2 R (unpublished) and #396 [187]. Chromatin immunoprecipitation (ChIP) for CTCF carried out on asynchronous and mitotic chromatin prepared from HeLa cells (human) and NIH3T3 cells (mouse) revealed that the target sites are constitutively occupied by CTCF in chromatin from interphase as well as mitosis.

Higher order chromatin conformation of Igf2-$H19$ locus in mitotic cells:

Given the possibility that the mitotic chromosomes are highly accessible to some transcription factors, it is possible that some of these factors like CTCF might be functioning to maintain the individual expression domains during mitosis as well. This raised the issue whether that the long-range interactions observed at Igf2-$H19$ locus [50], Paper II) survive mitotic compaction or they are re-established after every round of cell division. In order to find the answers, we performed 3C as described in paper I for some of the key interactions occurring at the locus on chromatin prepared from asynchronous (interphase) and mitotic cells. Analysis of a short-range intrachromosomal interaction within Calreticulin gene which is used to normalize crosslinking frequency in 3C revealed that it is maintained in mitotic chromosomes. To make sure that there are no random ligation products getting enriched here rather the products we see are products of specific intra-molecular ligation events, we took one primer from calreticulin and one from Igf2 DMR1 for 3C PCR which subsequently resulted in no signal although the band was present in positive control for PCRs. We thought that the strong compaction might result in non-specific chromatin contacts and that is what is detected with calreticulin. To make sure, we analysed a long-range interaction in Erc3 gene, separated by approximately 14 Kb, being ubiquitously transcribed, and gives a positive signal in interphase cells for the same [26]. Interestingly, this interaction was lost in mitotic cells consistent with shut down of Erc3 transcription in mitosis. This ensures the specificity of signals we get in 3C even in highly condensed mitotic chromatin. Thereafter, we tested the CTCF-mediated long-range interactions between Igf2 DMR1 and $H19$ ICR which are separated by more than 80 kb in cis. Strikingly, in several samples analyzed, we consistently found that this interaction in maintained in mitotic cells as compared to interphase cells. This was tested using two different primer combinations from both ends and both gave identical results. This entails us to think that CTCF and key regulatory long-range regulatory loops dependent on it might serve as a novel kind of epigenetic memory to minimize transcriptional chaos a and to reset proper expression domains in the daughter cells as soon as cells exit mitosis (figure 2). The loops which survive mitosis might be very crucial in deciding higher order organization of a particular locus. They could function as “guide loops” which, once cells exit mitosis, function directly or indirectly in setting up other loops in the locus.

We then thought to analyze another significant loop i.e. downstream enhancers interacting with Igf2 promoters which are separated by around 100 Kb. The 3C signal was seen in asynchronous cells but was absent in mitotic chromatin. Perhaps the enhancer interaction with Igf2 promoters is transcription driven and therefore the factory or chromatin hub organized by enhancers might involve Pol II as an essential component
which being excluded from mitotic chromosomes here fails to organize such higher order domains.

Figure 2. CTCF-dependent long-range interactions constitute epigenetic memory through mitosis

Paper IV:

As we have discussed, *H19* ICR is differentially methylated on the two alleles in a parent-of-origin specific manner, with the maternal allele unmethylated and paternal methylated. This difference is acquired during germline development, where the allele acquires methylation during male gametogenesis whereas the same is rendered methylation-free during female gametogenesis. The switching from female to male epigenotype i.e. acquisition of methylation mark at the *H19* ICR is completed by the time of first meiotic division in adult mouse and human testis [267] [293]. There is some delay in the reprogramming (de novo methylation) of maternal allele here as compared to the paternal allele during male gametogenesis which might be due to its originally methylation-protected status.

The results from our lab and collaborators have shown that the methylation status of *H19* ICR is dependent on CTCF target sites therein. The extensive reprogramming (acquisition of de novo methylation) during male germline development might be allowed on maternal allele only once the insulator function here is compromised to allow the machinery to act upon. Our search for such candidates ended up with discovery of a CTCF parologue, expressed during male gametogenesis only [272]. The gene was named BORIS (Brother Of Regulator of Imprinted Status), owing to the fact that it has same 11 Zinc fingers with more than 95% similarity at the amino acid level, the only difference between two being at the C and N-terminus. Lsh, an SNF2 family member, is an important
regulatory of DNA as well as histone methylation. It is widely expressed during development but its expression gets restricted to the thymus and testis in the adult mouse. H19 upstream regions is considerably hypomethylated in Lsh-/− mice [276]. Here, we tried to explore the roles of Lsh and BORIS during epigenetic reprogramming of the maternal H19 allele of ICR during male gametogenesis.

**CTCF and BORIS expression during male germline development:**

It was very important to determine the precise timing of BORIS activation in the adult mouse testis to precisely link BORIS with the reprogramming process of H19 ICR. Using high affinity antibody against a BORIS epitope that is common among several BORIS variants, we show that BORIS is expressed at high level both in spermatocytes and spermatids. The low level of expression seen in spermatogonia led us speculate that BORIS is activated at this stage. Similar analysis with CTCF showed that CTCF is expressed throughout spermatogenesis, although the signal is highest in spermatids. Surprisingly, the immunofluorescence confocal analysis of the squash preparations of staged testis cells showed that there is very little or no overlap in CTCF and BORIS distribution in spermatocytes or round spermatids. The BORIS seems to be confined to the inside of the nuclear membrane in these cells which in turn raises the possibility that BORIS activation might be linked with relocation of some CTCF target sites during reprogramming events to other nuclear sites, mostly towards periphery.

**BORIS occupies H19 ICR in a methylation-sensitive manner and requires intact CTCF target sites:**

The similarity in DNA binding zinc fingers between CTCF and BORIS provides a possibility of same target site preferences. A gel-shift analysis using wild type H19 ICR and CTCF #3 mutated H19 ICR fragment revealed that BORIS occupies H19 ICR and the binding is lost when the CTCF target sites are mutated. CTCF is known to bind H19 ICR in a methylation-sensitive manner [56] [57]. We next assessed whether BORIS possess the same property. Gel-shift assay carried out with an in vitro methylated H19 ICR fragment revealed that BORIS, like CTCF binds H19 ICR in a methylation-sensitive manner. Chromatin Immunoprecipitation (ChIP) carried out in adult testis showed that BORIS antibody pulls down only maternal allele of H19 ICR confirming the allele-specific BORIS occupancy in vivo.

**BORIS associated with H19 ICR during early stages of epigenetic reprogramming:**

The above results suggested that BORIS interacts with H19 ICR during the imprint acquisition process. To conclude precisely whether the BORIS-H19 ICR complex formation takes place before or after the imprint acquisition process, we carried out bisulfite analysis of CTCF #3 with the CTCF and BORIS ChIP material. The CTCF was shown to occupy exclusively unmethylated sequences. On the other hand, we found a mixture of methylated and non-methylated CpGs in the H19 ICR molecules interacting with BORIS. Although it is quite possible that somatic cells within the testis contributes to the CTCF ChIP-bisulfite data, the distinct expression pattern of BORIS during male germline development suggests that BORIS is associated with H19 ICR during early stages of epigenetic reprogramming events.

**Lsh binds unmethylated H19 ICR allele and this interaction relies on intact CTCF target sites:**

Immunofluorescence analysis showed that Lsh and CTCF are actually coexpressed only in spermatogonia. To investigate whether Lsh forms a complex with CTCF, we
carried out co-immunoprecipitation assays of nuclear extracts from adult testis. Lsh complexes were purified using two different affinity purified antibodies against Lsh and subsequently analyzed for presence of CTCF and BORIS by Western blot. CTCF was strongly in the immunoprecipitate but not BORIS, indicating that Lsh forms a complex with CTCF.

To directly verify whether Lsh directly binds H19 ICR, we carried out ChIP assays from adult testes of mice carrying maternal and paternal inheritance of CTCF target site mutations. The results showed that Lsh pulls down H19 ICR only when the wild type allele is inherited maternally; suggesting that Lsh binds only the unmethylated allele and this binding requires the contact nucleotides for CTCF.

Inheritance of mutated H19 ICR allele doesn’t affect overall methylation acquisition process:

We next wanted to analyze the effect of CTCF target site mutations at H19 ICR on the imprint acquisition process. Towards this end, we purified sperm DNA and restricted by the methylation-sensitive enzyme Hha I followed by Southern blot analysis. All Hha I sites were found methylated, showing that the inheritance of mutant H19 ICR allele did not have any serious effect on the methylation acquisition process. To cover individual CpGs not included in the southern analysis here, we carried out our bisulfite analysis of the region around CTCF target #3. The results supported the southern results to a large extent suggesting that the methylation acquisition process is not significantly affected due to the absence of CTCF, Lsh or BORIS binding at H19 ICR.

Collectively, these data show that the CTCF target sites in H19 ICR create resistance for epigenetic reprogramming process and that specific mechanisms have evolved in the testis to overcome this obstacle. Our observation that CTCF appears to recruit Lsh to the maternal H19 ICR presumably in spermatogonia is potentially interesting in this aspect since Lsh is a chromatin remodeling factor essential for genomewide de novo methylation. It is shown previously that changing positions of the nucleosome positioning sequences (NPSs) at H19 ICR results in loss of both CTCF target site occupancy and insulator function [29]. In light of those observations, we postulate that the positioned nucleosomes within the H19 ICR might create an obstacle to the reprogramming events and speculate that CTCF recruits Lsh to the H19 ICR to begin nucleosome sliding. Further, there is CTCF to BORIS switch in occupancy at H19 ICR in spermatogonia committed to spermatocyte differentiation due to activation of BORIS at this stage. The observation that BORIS doesn’t bind methylated H19 ICR and the CpGs in H19 ICR sequences pulled down by the BORIS antibody in formaldehyde cross-linked adult testis are largely unmethylated, this event most likely takes place during early stages of the imprint acquisition process. The maternal inheritance of CTCF target sites at H19 ICR and thereby absence of CTCF, Lsh and BORIS occupancy doesn’t hamper the imprint acquisition process which suggests that the CTCF to BORIS switch is unlikely to involve active recruitment of any de novo methylation machinery. Instead, the switch might act to release the observed resistance against methylation provided by CTCF which in turn might be facilitated by nuclear relocalization of the target sequence to new microenvironment since CTCF and BORIS distribution in male germline cells almost doesn’t overlap.
CONCLUDING REMARKS

This thesis highlights the relation between epigenetics, gene regulation and higher order chromatin conformations. The evidences so far suggest that long-range control of gene expression is mediated through direct physical interactions between genes and regulatory elements, either cis or in trans. Besides transcriptional initiation, formation of active chromatin hubs seem to be crucial for increased transcriptional efficiency as well as protection from neighbouring heterochromatic regions. Transcription factors such as CTCF apparently have an important role in organization of such functional modules in a development and differentiated-dependent fashion. The observations here impinge upon the fact that organizing loop domains might be a very crucial mechanism behind imprinting regulation and long-range chromatin interactions are key to cellular processes. The discussion also illustrates here that each gene’s transcription cycle might be operated by several different kinds of regulatory mechanisms, each one of them acting in a regulated manner at different time points.

The relevance of trans-acting factors in the ‘choice’ process of XCI was strengthened by our observations where mouse and human were shown to bind CTCF at a homologous position preferentially on active XIST/Xist promoters. The binding was conserved in human, mouse, rabbit, horse and rabbit supporting the functional relevance of this occupancy. Interestingly we found that CTCF binding affinity was enhanced on the point mutant Xist promoter where the X-chromosome harbouring it was getting preferentially inactivated. This increased affinity was correlated with increased insulator activity of the fragment. Studies here with two familial cases of opposite skewed X-inactivation patterns added new dimensions to our understanding of X-inactivation process that the choice process could be affected by simply regulating the binding affinity of a regulatory factor to its target site and such process might operate during distinct stages of development, for example in our case, before the ‘choice’ of the X-chromosome to be inactivated is made.

Using 3C assay, we first time demonstrated convincingly the role of CTCF target sites at H19 ICR in organizing chromatin conformations at Igf2-H19 locus, which in turn is very crucial for imprinting regulation. We also showed how the epigenetic status of different DMRs at this locus is regulated by CTCF contact nucleotides at H19 ICR in a hierarchical manner. Overall, we find distinct set of physical interactions with H19 ICR which set up different kinds of loop conformations on paternal and maternal allele, allowing Igf2 to express or be silent respectively.

Another study started with beautiful demonstration that CTCF constitutively occupies its target sites on mitotic chromosomes both in human and mouse cells and despite the highly condensed status of mitotic chromosome, it was not getting excluded. Subsequently, 3C methodology was used to investigate few long-range chromatin interactions in mitosis. Interestingly, we found that H19 ICR-Igf2 DMR1 interaction survives mitotic compaction whereas transcription-dependent interaction Igf2 promoter-downstream enhancer was lost. This provokes the possibility that certain crucial interactions might survive M-phase and serve as an epigenetic memory to set up proper functional domains and expression profile in daughter cells as soon as they exit mitosis.

The last study showed here that Lsh, a chromatin remodeling factor that is essential for de novo methylation of the entire genome, is recruited to the maternal allele of H19 ICR during male germline development, probably in a complex with the chromatin insulator protein CTCF. Moreover, BORIS, a CTCF paralog with exclusive expression in testis, replaces CTCF at the maternal H19 ICR allele during early stages of the imprint acquisition process. This led us propose that the imprint acquisition process involves
initiation of epigenetic reprogramming by a CTCF-dependent recruitment of Lsh to the $H^{19}$ ICR followed by completion of the imprint acquisition process by a switch in CTCF to BORIS occupancy.

Overall, the thesis unravels the novel roles for CTCF as a pivotal factor in regulation of chromatin higher order structure formation and transcriptional regulation. We discovered novel role for CTCF as a decisive factor for X-chromosome inactivation process where choice could be modulated just by regulating its affinity for the target site. We, for the first time, convincingly demonstrate how CTCF target sites at $H^{19}$ ICR set up proper imprinted expression patterns at whole locus by regulating various long-range interactions and coordinating regional epigenetic marks. These studies also discovered that CTCF occupancy at its target sites and dependent crucial regulatory loops survive mitosis, maintaining insulated domains and constitute a novel kind of epigenetic memory to daughter cells. Our observations suggest that the epigenetic reprogramming of $H^{19}$ ICR during male gametogenesis is initiated by a CTCF-dependent recruitment of chromatin remodeling factor Lsh to the $H^{19}$ ICR followed by completion of the imprint acquisition process by a replacement of CTCF with BORIS.
ACKNOWLEDGEMENTS

The present thesis is the result of work carried out at Department of Animal Development and Genetics, EBC, Uppsala University. Efforts from lot of people have helped me directly or indirectly to reach this stage of my life and I wish to express my gratitude towards them.

First and foremost I would like to express my sincere thanks to Prof. Rolf Ohlsson, head of the department and my PhD supervisor, who provided me with the opportunity to work under his guidance. Besides always being there to discuss good science, he shaped my personality and made me appreciate the realities of life. I am thankful to him for training me to fight difficulties and challenges, be it scientific or personal!

I am thankful to Dr Chandrasekhar Kanduri, my co-supervisor and my inspiration to come to Uppsala whose kind personality combined with everlasting enthusiasm to do good science always inspires us. I express my gratitude to him for all he did for me.

My colleagues Reza (luxury moments), Tiger, Piero (the Friday ‘pub hero’), Claes, Rosita (multifaceted help), Magda (great help in initial settlements), Anita, Taras (fun talk), Irina, Sha, Olga and Mikael who helped me in one or the other way during my lab life and also added a smile to my face whenever possible. I wish a life full of happiness to all of them and pray that they all become bright stars in the sky of science.

I thank Mouse genetics group members: Reinald Fundele, Yang, Jana, Umashankar and others for being part of the scientific community of the department.

The excellent lab manager Helena Malmikumpu, Financial in-charge and administrative support Rose-Marie Löfberg, and the best animal facility manager and great support Anita Mattson, the sequencing lady Ambar Khodabandah for all the help provided.

Stefan Ås, a great person, for his friendship and round the clock assistance with computers. Stefan Gunnarsson, Gary wife for anytime needed help with microscopes.

I also thank all the past members of the department: Meena Kanduri, Farhan Rizvi, Sreenivasulu Kurukuti, Joanne Whitehead (ceaseless help), Vinod Pant (great discussions and guidance), Rtituparna Mukhopadhyay, Wenxiang Yu, Rolf Ericsson, Carina Östman, Carl-erik Cantell and Graham Budd for their help and guiding me in many things, especially during my early days in the lab.

My corridor gang and best company: Jitendra (Cycle Jackson), Dharam (Sardar), Nimesh (Mallu), Puneet (Halwai ji Lucknow wale), Pandey (Sota Siyaar) and Faizaan (Moulawi sahib), Rayees & Sadique (The Hyderabadi mystery). It will be hard to forget moments we shared together and the peaks of fun we had. I have no better words than just saying-‘you meant everything to me in Uppsala’. You warden will miss you all a lot………

I convey special thanks to Jitendra and Rayees for their huge efforts during the final preparation of this thesis.

My special friends from Sweden, Thomas and Vendela, who added to my good memories about this country. Their care and love for me is unforgettable and they own a very special place in my heart.
The Indian community at Uppsala, who never lets anyone, feel home-sick.

My special gratitude to all the mice whose sacrifice made this work possible.

My previous supervisor, Prof Paturu Kondaiah, Indian Institute of Science, Bangalore, India, who taught me that being a good scientist and a good person are not mutually exclusive. His everlasting support always gives me strength.

I am deeply indebted to the place and people where I did my Masters degree, Banaras Hindu University, Varanasi, India, especially Prof S C Lakhotia, Prof Rajiv Raman, Prof Mercy J Raman, Prof M M Chaturvedi and other members of MHG and Zoology department for recognizing my potential and always encouraging me to take giant steps into the field of Science. I am what they made me!

My friends from India, especially Rajesh Yadav, for his best possible support in ups and downs of my life; Yogesh Mishra for his brotherly affection; Moushami Mallik for her support and care for me.

My relatives back home, whose continuous support and trust in my capabilities was the driving force under difficult situations.

My sister Renu, her daughter Gargi, brother-in-law Raghuvar, my late brother’s wife Anita and my nephew Akash (Manu) for their love, affection and care for me.

My late brother Vipin, who used to teach me “Sky is not the limit”. His memories always keep my eyes wet.

My parents, for being THE BEST, for their unconditional love, care, support, guidance, encouragement and a wonderful sense of humor. I now understand what Yudhistira exactly meant in Mahabharata by saying-“Mother is bigger than the Earth and Father is higher than the Sky”. I dedicate all my achievements in life so far to them. I am nothing without them.

And all others whose names I might have missed.

The work was financially supported by the Swedish Science Research Council, the Swedish Cancer Research Foundation the Swedish Pediatric Cancer Foundation, the Lundberg Foundation, the von Hofsten Foundation, the Wallenberg Foundation, the Juvenile Diabetes Research Foundation International and Stiftelsen Wenner-Gren’ska Samfondet.
REFERENCES:


Dimitrova DS, Gilbert DM. 1999. The spatial position and replication timing of chromosomal domains are both established in early G1 phase. Mol Cell 4:983-93.


Yan Q, Cho E, Lockett S, Muegge K. 2003. Association of Lsh, a regulator of DNA


Acta Universitatis Upsaliensis

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology 138

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 Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology. (Prior to January, 2005, the series was published under the title “Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology”.)

Distribution: publications.uu.se
urn:nbn:se:uu:diva-6302