The Control of the Epigenome

MAGDA LEZCANO
Abstract

The genetic information required for the existence of a living cell of any kind is encoded in the sequence information scripted in the double helix DNA. A modern trend in biology struggles to come to grip with the amazing fact that there are so many different cell types in our body and that they are directed from the same genomic blueprint. It is clear, that the key to this feature is provided by epigenetic information that dictates how, where and when genes should be expressed. Epigenetic states “dress up” the genome by packaging it in chromatin conformations that differentially regulate accessibility for key nuclear factors and in coordination with differential localizations within the nucleus will dictate the ultimate task, expression.

In the imprinted Igf2/H19 domain, this feature is determined by the interaction between the chromatin insulator protein CTCF and the unmethylated H19 imprinting control region. Here I show that CTCF interacts with many sites genome-wide and that these sites are generally protected from DNA methylation, suggesting that CTCF function has been recruited to manifest novel imprinted states during mammalian development.

This thesis also describes the discovery of an epigenetically regulated network of intra and interchromosomal complexes, identified by the invented 4C method. Importantly, the disruption of CTCF binding sites at the H19 imprinting control region not only disconnects this network, but also leads to significant changes in expression patterns in the interacting partners.

Interestingly, CTCF plays an important role in the regulation of the replication timing not only of the Igf2 gene, but also of all other sequences binding this factor potentially by a cell cycle-specific relocation of CTCF-DNA complexes to subnuclear compartments.

Finally, I show that epigenetic marks signifying active or inactive states can be gained and lost, respectively, upon exposure to stress. As many genes belonging to the apoptotic pathway are upregulated we propose that stress-induced epigenetic lesions represent a surveillance system marking the affected cells for death to the benefit of the individual. This important observation opens our minds to the view of new intrinsic mechanisms that the cell has in order to maintain proper gene expression, and in the case of misleads there are several check points that direct the cell to towards important survival decisions.

Keywords: Epigenetics, chromosome interactions, replication timing, histone modifications, epigenetic surveillance

Magda Lezcano, Department of Physiology and Developmental Biology, Animal Development and Genetics, Norbyvägen 18 A, Uppsala University, SE-75236 Uppsala, Sweden

© Magda Lezcano 2006

ISSN 1651-6214
ISBN 91-554-6680-X
urn:nbn:se:uu:diva-7190 (http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-7190)
To my sweet mother

Dedicado a mi dulce madre
List of Papers

This thesis is based on the following papers, which will be referred to in the text by their Roman numerals.


Reprints were made with permission of the publishers.
## Contents

List of Papers ................................................................. 5  
Abbreviations ................................................................. 8  
Introduction ........................................................................ 9  
  Epigenetics ................................................................. 10  
  Gene Expression and Regulation ................................. 11  
  Methylation ................................................................. 14  
  Histones ....................................................................... 17  
  Chromatin ................................................................. 21  
  The CTCF Protein ..................................................... 23  
  Replication Timing ..................................................... 25  
  Genome Organization ............................................... 28  
  Long Range Interactions ............................................ 29  
Aims of the Present Studies ............................................. 34  
  Paper I: ................................................................. 34  
  Paper II: ................................................................. 34  
  Paper III: ................................................................. 35  
  Paper IV ................................................................. 35  
Results and Discussion ............................................... 36  
  Paper I ................................................................. 36  
  Paper II ................................................................. 38  
  Paper III ................................................................. 40  
  Paper IV ................................................................. 42  
Conclusions ........................................................................ 44  
Summary in Swedish .................................................... 46  
  Epigenomets reglering ................................................. 46  
Acknowledgements ..................................................... 48  
References ......................................................................... 52
## Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>ATP</td>
<td>Adenosine 5'-triphosphate</td>
</tr>
<tr>
<td>PEV</td>
<td>Position-effect variegation</td>
</tr>
<tr>
<td>HAT</td>
<td>Histone acetyltransferase</td>
</tr>
<tr>
<td>LCR</td>
<td>Locus control region</td>
</tr>
<tr>
<td>TDP</td>
<td>Timing decision point</td>
</tr>
<tr>
<td>ZF</td>
<td>Zinc finger</td>
</tr>
<tr>
<td>ORF</td>
<td>Open reading frame</td>
</tr>
<tr>
<td>APP</td>
<td>Amyloid β-protein precursor</td>
</tr>
<tr>
<td>kb</td>
<td>Kilobases</td>
</tr>
<tr>
<td>ACH</td>
<td>Active chromatin hubs</td>
</tr>
<tr>
<td>CT</td>
<td>Chromosome territories</td>
</tr>
<tr>
<td>IC</td>
<td>Interchromatin compartment</td>
</tr>
<tr>
<td>3C</td>
<td>Chromosome conformation capture</td>
</tr>
<tr>
<td>MAR</td>
<td>Matrix attachment region</td>
</tr>
<tr>
<td>DMRs</td>
<td>Differentially methylated regions</td>
</tr>
<tr>
<td>BP</td>
<td>Base pairs</td>
</tr>
<tr>
<td>GSTP1</td>
<td>Pi class glutathione S-transferase gene</td>
</tr>
<tr>
<td>NTD</td>
<td>amino-terminal “tail” domain</td>
</tr>
<tr>
<td>ICR</td>
<td>Imprinting control region</td>
</tr>
<tr>
<td>DSBs</td>
<td>Double-strand breaks</td>
</tr>
<tr>
<td>ChIP</td>
<td>Chromatin immunoprecipitated DNA</td>
</tr>
<tr>
<td>FISH</td>
<td>Fluorescence in situ hybridization</td>
</tr>
<tr>
<td>LTRs</td>
<td>Long terminal repeats</td>
</tr>
<tr>
<td>4C</td>
<td>Circular Chromosome Conformation Capture</td>
</tr>
<tr>
<td>BrdU</td>
<td>5-bromo-2′-deoxyuridine</td>
</tr>
<tr>
<td>PI</td>
<td>Propidium iodide</td>
</tr>
<tr>
<td>qPCR</td>
<td>quantitative real time PCR</td>
</tr>
</tbody>
</table>
Introduction

The wonderful Mother Nature have created life in such a complex and still so perfect way that to understand its mechanisms or to get a brief idea of its magnificent works it has and will require centuries and millions of efforts.

In biology, we have tried to defined and organized organisms according to their differences and similarities. But still each one of us is different from the rest. Even within an individual, cells from various organs perform different tasks in order to maintain a living multi-cellular organism. It’s incredible that despite having such diversity, the genetic information written in the deoxyribonucleic acid (DNA) is the same among all cells within an individual. So, how do we achieve such a difference? The main issue relies on the fact that not all the genes are expressed at the same time in all cells. There is a differential expression of genes in different cells and at different times and this is the basis for the diversity among cells with a variety of specialized functions within an individual.

After fertilization the fertilized egg or zygote will divide and these new cells have the potential to differentiate into any type of cell, meaning that they are totipotent. Then, the differentiation procedure begins and cells start to loose the possibility to differentiate into any cell types. There is a process of specification of the three embryonic germ layers: ectoderm (give rise to skin and neural lineages), mesoderm (generates blood, bone, muscle, cartilage, and fat), and endoderm (contributes to the formation of tissues of the respiratory and digestive tracts). This specification is irreversibly maintained through later embryonic stages and throughout adulthood. The cells become restricted to their own new cell lineage, they also become mortal and they will not be able to divide for unlimited times.

The possibility of performing animal cloning let us open our eyes and realize that under certain circumstances a cell can be reprogrammed. The versatility of a totipotent cell could be regained and the proper mechanisms that stabilize a certain gene expression could be modified. This has a very important relevance since understanding how to dismantle the repressive marks will solve many dilemmas of modern and ancient genetics. With this in mind, it’s also important to understand how the genetic information is acquired, under which circumstances and mechanisms is maintained the commitment of a differentiated cell. At last, how the system takes care of erroneous regulation of gene expression, and how certain cells overcome the strict control of abnormal patterns and thus lead to a cancerogenic type.
Epigenetics

The term was first used by Conrad H. Waddington (1905–1975), as the junction of *epigenesis* and *genetics* resulted in *epigenetics*, the study of development, emphasizing its fundamental dependence on genetics and its interest in processes. *Genotype + Epigenotype = Phenotype*. It has been more commonly used as; the study of heritable changes in gene expression that occurs without a change in the sequence of the DNA. Waddington speculated that every gene affects several different processes, and that genes work together to form gene networks. During development, these networks themselves were thought to undergo changes in competence. Waddington described wing development in *Drosophila melanogaster* as affected by more than 40 genes, creating gene networks (Waddington C.H., 1939). In addition, he stated that there is no simple one-to-one relation between a gene and a phenotypic character; as such a relation only exists between the phenotype and the genotype as a whole. Epigenetics represented for quite a while a collecting terminology for many genetic observations, which indicated a lack of uniformity and reciprocity of the F1 stating a non-Mendelian segregation or unpredictable instability of an apparently mutant allele.

Paramutation represents one of the classic epigenetic phenomena. A typical example is that in certain hybrids carrying a combination of two different, specific alleles, one allele appears as if it is contaminated by the other (i.e., causes a phenotype, which may or may not be entirely similar to that of the other allele, but seemingly dependent on the other allele). One of the classic examples of paramutation was described by Brink (1956), for the maize R locus. He combined the R-r allele, conferring dark purple seed color, with the R-stippled allele, conferring purple stippled seed phenotype. A proportion of F1 hybrids displayed slight purple color and, upon self-fertilization, the reduced coloration trait co-segregated with an apparently modified, paramutated form of the R-r allele, which was designated R-r0. Thus, it appeared that R stippled modified in *trans* the R-r allele.

Another exception of Mendel’s rule is the phenomenon of imprinting (Surani M.A., 1984), which means that the parental contributions to the offspring are unequal. In other words, the paternal and maternal genomes may have different effects (imprinting) on the developing offspring, by modification of expression of one of the alleles, depending on the direction of the cross. Genomic imprinting is a good model to study differential expression, since the two alleles carries the same DNA sequence but are differentially regulated at a single cell level.

A distinguishing characteristic of imprinted genes is their association with regions of differential methylation, this will be discussed in detail below. In addition, it is indicated that not only DNA methylation, and methylation of lysine residues in histone tails plays major roles in gene silencing, but *per se*
the chromatin conformation of the region. This will be analyzed further in detail on the coming chapters.

It is commonly assumed that epigenetic marks are “reset” in the germ line with each generation, as is the case with imprinted loci, but there are exceptions that bring the rule into question. Epimutations in tumors, and at mutant or translocated genes, are somatic events (Jones P.A., 2002). Several familial autosomal dominant cancer syndromes have been found from inactivating mutations in one of the alleles of a tumor suppressor gene and tumors develop from cells that have lost the remaining (functional) allele by mutation or epimutation as a somatic event (Garber J.E., 2005).

Gene Expression and Regulation

Gene expression is under strict control. Proteins must be deployed at correct time, location and intensity, since abnormal gene expression can be deleterious for the organism. Gene expression is a complex, multi-step process with many checkpoints to be modulated.

The “Openness” of the chromatin is crucial for transcription rate of genes (Figure 1). When chromatin is “open”, the DNA is more accessible to transcription factors and polymerases, and thus the genes are actively transcribed. Methylation of CpG-islands of genes or gene regulatory sequences (promoters, enhancers) lowers the transcription rate and together with the chromatin conformation are heritable ways of controlling the transcription rate without changing the sequence of DNA. Pre-transcriptional control is access by binding regulatory proteins (transcription factors) to cis-regulatory sequences of genes.

![Figure 1. Chromatin structure regulates transcriptional activity.](image)

Figure 1. Chromatin structure regulates transcriptional activity. Nucleosomes consist of DNA (black line) wrapped around histone octomers (purple). Post-translational modification of histone tails by methylation (Me), phosphorylation (P) or acetylation (Ac) can alter the higher-order nucleosome structure. Nucleosome structure can be regulated by ATP-dependent chromatin remodelers (yellow...
cylinders), and the opposing actions of histone acetyltransferases (HATs) and histone deacetylases (HDACs). Methyl-binding proteins, such as the methyl-CpG-binding protein (MECP2), target methylated DNA (yellow) and recruit HDACs. A. DNA methylation and histone deacetylation induce a closed-chromatin configuration and transcriptional repression. B. Histone acetylation and demethylation of DNA relaxes chromatin, and allows transcriptional activation. (From Johnstone R.W. 2002) (Permission granted from The Nature Publishing Group).

The concepts outlined above have been a join of many efforts from the last several decades. In an attempt to give credits to those outstanding minds, we will recapitulate key studies that have contributed to build up the current models that describe how genes are expressed.

The field of transcription analysis was launched with the discovery of RNA polymerase just 50 years ago in 1957, when Vernon Ingram linked a defect in the gene that caused sickle-cell anemia with a single amino-acid change in the hemoglobin protein. The question was how the information contained in the DNA was transferred into protein information. As more scientific knowledge was gathered, the concepts became more clear.

A general model for bacterial gene regulation was proposed by Jacob and Monod in 1961. They predicted the existence of repressors encoded by non-structural genes. Repressors control the transcription of single or groups of genes. They were regulated by metabolites or other environmental factors, which could promote or antagonize their function. Repressors acted through operators that functioned in cis with the regulated genes; however, the precise nature of the repressors and how they interacted with operators was still unclear.

The nucleosome hypothesis represents one of the great paradigm shifts in the understanding of eukaryotic gene expression. Its key concept is that eukaryotic DNA is tightly packaged around a core of structural proteins called histones, to generate a nucleosome array that is fundamental for controlling gene expression (Kornberg, 1974).

As already mentioned, generally speaking, eukaryotic cells do not discard DNA as they differentiate. Cellular differentiation therefore has to be explained as a consequence of differential gene expression. The first models were described in 1975 by Arthur Riggs, and by Robin Holliday and John Pugh, who discussed DNA methylases in bacteria. Their models, proposed that the properties of these enzymes, in particular their preference for hemi-methylated substrates made them ideally suited to establish stable differentiated expression states in the absence of genetic mutation. They further proposed that the sequence-specific binding of these enzymes would have a gene-regulatory role.

Adrian Bird and colleagues subsequently provided further evidence linking DNA methylation and gene expression. In 1985, they characterized the small fraction of the mouse genome that is frequently cleaved by a
methylation-sensitive restriction enzyme. These sequences, which would come to be known as CpG islands, are CpG-rich fragments with low or undetectable levels of methylation. The available literature indicated that CpG islands were typically located near the 5′ ends of genes, and the authors predicted correctly that genes might be associated with "methylation-free" zones near sequences of regulatory significance.

But in a higher context it is important to take into account the influence of chromatin structure. So, the studies of Han and Grunstein (1988) opened up the view when they showed that nucleosome loss, through histone depletion, resulted in the increased transcription of numerous genes in yeast. This provided the first in vivo evidence that nucleosomes can repress gene activity, and it has become so important that Segal and collaborators (2006) attempted the mammoth task of predicting the positions of all the nucleosomes in the genome, based solely on computer models of DNA sequence characteristics.

The Grunstein group lay the first cornerstone by showing that histone tails had specific functions. Studies on histone H4 revealed that its tail is dispensable for yeast growth (Kayne P.S., 1988), but is required to repress the activation of specific mating loci through the binding of a repressor protein. The same group showed that the H4 tail also had a gene-activating function and that histone modifications, namely acetylation at specific residues were required for gene transcription. But, of course, the post-translational modification of histone tails and their binding to regulatory proteins is a fundamental theme in transcriptional regulation.

Now, chromatin configuration could also be modified in the cell and important studies by Craig Peterson (1994), Robert Kingston and Michael Green (1994) described the yeast and human Swi/Snf complexes. These 10-subunit complexes were able to stimulate binding of the Gal4 transcription factor to nucleosomal DNA in an ATP-dependent manner, a functional test of what is now known as chromatin remodeling. The Swi/Snf complexes directly interacted with nucleosomal DNA and altered the pattern of nuclease cleavage, changing in this way, the DNA topology. It was proposed that the Swi/Snf complex would interact with a transcriptional activator to localize it to the region targeted for disruption, and would then contacts the DNA to disrupt interactions with the histones and increase accessibility for transcription factors.

For a long time, heterochromatin was considered to be the 'dark matter' of the genome: highly condensed, late replicating and associated with lack of gene expression. Various lines of evidence indicated that heterochromatin could lead to transcriptional silencing, but the mechanism was unknown. Studies of position-effect variegation (PEV) in fruit flies, and of the centromeric and mating-type regions in fission yeast, have proved to be a key to light on the mysteries of heterochromatin structure and function.
The Grewal (1997) group used the mating-type region of fission yeast to study the factors that favour heterochromatin formation in \textit{cis}. Little was known about them, except that repeat elements tended to be heterochromatic. This was an important issue, because as was known from PEV, the heterochromatic state could spread to inactivate nearby genes. The authors found that \textit{cenH}, a centromere-homologous repeat that normally lies at the silent mating locus, is sufficient to bring about heterochromatin formation at ectopic sites. This ability is associated with H3K9 methylation and recruitment of \textit{Swi6} (a yeast HP1 counterpart) and, importantly, requires the RNAi machinery.

In 1996, it was shown that histone acetylases and deacetylases were, in fact, well-known transcriptional regulators. These studies provided the first clear connection between histone acetylation and transcriptional regulation; leading to a better understanding of how histone modifications control chromatin structure and gene expression. In first, Allis and colleagues cloned a histone acetyltransferase (HAT) from \textit{Tetrahymena thermophila}. The cloned HAT turned out to have a high level of sequence relatedness to the yeast \textit{Gcn5} (Kuo M. H., 1996), a transcriptional activator, with almost 80% sequence identity in some domains. Consistent with the sequence homology, the authors showed that recombinant \textit{Gcn5} had HAT activity \textit{in vitro}. In the second study, Schreiber and colleagues purified a mammalian histone deacetylase. This approach, combined with the microsequencing of purified proteins, yielded that one of the proteins had 60% sequence identity to the yeast \textit{Rpd3} (Taunton J., 1996), a characterized transcriptional repressor. The two antagonistic enzymatic activities had the opposite functional effects: activation or repression on transcription.

Genetic screens in fruit flies and fission yeast had shown that the fruitfly suppressor of variegation 3-9 (\textit{Su(var)3-9}) and fission yeast \textit{clr4} were important for establishing and propagating heterochromatin a higher-order chromatin structure that is repressive for transcription. Work on the mammalian homologues had extended this link, but a mechanistic understanding of how these proteins controlled heterochromatin formation was lacking. In 2000, Jenuwein and colleagues showed that the mammalian homologue of \textit{Su(var)3-9} was a histone lysine methyltransferase that selectively methylated histone H3 at Lysine 9 (Melcher M., 2000).

**Methylation**

DNA methylation in vertebrates typically occurs at CpG sites (that is, where a cytosine is directly followed by a guanine in the DNA sequence); this process of methylation results in the conversion of the cytosine to 5-methylcytosine (Richards E. J., 1997). The formation of Me-CpG is catalyzed by an enzyme known as DNA methyltransferase. A variety of
DNA methyltransferases exist to modify the DNA in a variety of patterns. Some DNA methyltransferases act primarily in conjunction with replication to perpetuate methylation patterns from "mother" strands to newly synthesized "daughter" strands. Other DNA methyltransferases can add methyl groups to DNA strands that have no pre-existing methylation. CpG sites are uncommon in vertebrate genomes but are often found at higher density near vertebrate gene promoters where they are collectively referred to as CpG islands. The methylation state of these CpG islands can have a major impact on gene activity/expression as previously mention.

Using a knockout mouse model, Li and collaborators observed the importance of methylation and of the DNA methyltransferase enzymes in the developing embryo. In these experiments, all the resulting embryos died at the morula stage (Okano M., 1999). This is explained by the fact that in early development (fertilization to 8-cell stage), the eukaryotic genome is demethylated (Figure 2). From the 8-cell stage to the morula, de novo methylation of the genome occurs, modifying and adding epigenetic information to the genome, and thus the knockout model will have great difficulty in setting the correct epigenetic information. By blastula stage, the methylation is thought to be complete and no genome wide changes occur later on.

**Figure 2. DNA methylation reprogramming cycle.** The cycles of DNA methylation reprogramming in the mouse can be divided into two distinct time periods: one associated with gametogenesis, and the other with preimplantation development. By the primitive streak stage of embryonic development (E 7.5), the primordial germ cells (PGCs). These cells are initially highly methylated but rapidly lose this methylation around ED 8 at the time they begin to migrate toward their final destination in the genital ridge between E 10.5 and E 11.5. Here PGCs continue to proliferate and undergo differentiation into distinct male and female germlines. By ED 12.5, most sequences in both genders have become maximally demethylated. This loss of methylation is associated with the erasure of germline imprints (black line). Thereafter, both male (blue) and female (red) gametes are arrested, the female in meiotic prophase and the male in mitosis. In males, de novo methylation of the genome, including imprinted loci, takes place prenatally, while in the female germline, de novo methylation occurs postnatally in growing oocytes. The
Reacquisition of DNA methylation requires the activity of a de novo MTase that, together with Dnmt3L, restores the imprinted information content to both germlines. Fertilization signals the second round of methylation reprogramming when the maturing paternal pronucleus is demethylated in a replication independent or active manner. Thereafter, the zygote and embryo are demethylated by a replication dependent or passive process owing to the general exclusion of the maintenance methylase Dnmt1o from the nucleus. Despite genomewide demethylation, imprinted methylation is maintained during this period. De novo methylation roughly coincides with the differentiation of the first two lineages of the blastocyst stage. These lineages are asymmetrically modified epigenetically with the inner cell mass being hypermethylated in comparison to the trophectoderm, reflecting the DNA methylation status of their somatic and placental derivatives, respectively. Imprinted genes are depicted with a black line throughout. Maternal and paternal alleles of nonimprinted genes and genomic sequences are depicted in red and blue, respectively (From Dean W., 2005).

It is important to mention that cytosine methylation have most negative effects on gene regulation via the involvement of other proteins that binds specifically to the DNA when it is methylated. Indeed, a number of methylcytosine binding proteins have been identified and several are found in close association with one or more histone deacetylases. These findings suggest that methylation brings local histone deacetylation, which could facilitate methylation of one or more deacetylated lysines on the histones and subsequent recruitment of repressor proteins that prevent transcription factors from gaining access to the gene (Ng H., 1999).

Numerous examples from plants and animals indicate that methylation induced gene silencing can result in a lasting condition, giving rise to epimutations; this is refer to a change in the epigenetic state of a given locus, which, in turn, may alter expression from the associated gene(s). The change occurs in somatic tissue and is propagated to daughter cells. Like mutations, epimutations involve stable alterations; but unlike mutations, these alterations are not based on changes in the DNA sequence.

Findings have lead to the hypothesis that promoter methylation may be one of the mechanisms of tumor-suppressor gene inactivation in cancer. This prediction was confirmed by the discovery in a subset of sporadic tumors of erroneous promoter methylation involving tumor-suppressor genes such as Rb, VHL and P16. Studies with the ER gene in colonic tumors led to the discovery that in some instances, this aberrant methylation actually begins in the normal-appearing cells of the affected tissues.

In general, genes whose promoters are affected by CpG island methylation are not expressed in tumor cells, and its expression can be restored by treatment with the methylation inhibitor 5-deoxy-azacytidine. In the case of tumor-suppressor genes, promoter methylation and coding sequence mutations are mutually exclusive events, suggesting in fact that methylation gives the same selective advantage to cells as mutational
inactivation, and is a *bona fide* alternate mechanism of gene inactivation in cancer.

Neoplasia can be characterized by "methylation imbalance" in which genome-wide hypomethylation is accompanied by localized hypermethylation and an increase in the expression of DNA methyltransferase (s) (Baylin J., 1998). The overall methylation state in a cell might also be a precipitating factor in carcinogenesis as evidence suggests that genome-wide hypomethylation can lead to chromosome instability and increased mutation rates (Chen Y., 1998). The methylation state of some genes can be used as a biomarker for tumourigenesis. For instance, hypermethylation of the pi-class glutathione S-transferase gene (GSTP1) appears to be a promising diagnostic indicator of prostate cancer (Nakayama J., 2004).

In cancer, the dynamics of genetic and epigenetic gene silencing are very different. Somatic genetic mutation leads to a block in the production of functional proteins from the mutant allele. If a selective advantage is conferred to the cell, the cell expands clonally to give rise to a tumor in which all cells lack the capacity to produce certain proteins. In contrast, epigenetically mediated gene silencing occurs gradually. It begins with a subtle decrease in transcription, given by a decrease in protection of the CpG island from the spread of flanking heterochromatin (Jones PA, 2002).

Histones

The enormous length of the eukaryotic genome requires that it is packaged into a stable structure that can be replicated and propagated during mitosis and is sufficiently malleable and modifiable to enable access to genetic information. Such structure is acquired by the formation of chromatin, which contains DNA and its associated proteins.

The subunit of nucleosomal arrays and chromatin fibers as mentioned before is the nucleosome core: a nucleoprotein complex consisting of 147 base pairs of DNA wrapped around an octamer of core histones proteins H2A, H2B, H3 and H4 (Richmond TJ, 1984). It is interesting that between core histones proteins H2A, H2B, H3 and H4. It is the tail domain of histone H4 the key and responsible for the 30 nm fiber formation by making contact with the H2A and H2B residues on the surface of the adjacent nucleosomes (Dorigo B, 2004; Dorigo B, 2003).

However, in some instances, specialized histone variants are found in place of the canonical histones, enabling the encoding of epigenetic information through defined or “specialized” nucleosomes arrays.

The histone “code” hypothesis has explained the complexity of the pattern of histone modifications (Strahl B.D., 2000; Turner B.M., 2000; Jenuwein T., 2001; Nakayama J., 2001). This hypothesis states that a
specific histone modification, or combination of modifications, can affect distinct downstream cellular events by altering the structure of chromatin and/or generating a binding platform for effectors proteins, which specifically recognized the modification(s) and initiate events that lead to gene transcription or silencing.

High-resolution crystallographic studies have shown that the majority of the histone octamer is found within the superhelical DNA gyre (White C.L., 2001). However, between 14 and 38 amino terminal residues of each core histone extend outside the disk-shape complex as N and C terminal 'tail' domains of histones and various studies have shown that these tails are substrates for a collection of posttranslational modifications, including acetylation and methylation of lysines (K) and arginines (R), phosphorylation of serines (S) and threonines (T), ubiquitylation and sumoylation of lysines, as well as ribosylation takes place. Adding further is the fact that some of the histone residues can accept one to three methyl groups and this is exemplified in the case of lysine and arginine residues, which exists in mono-, di- or trimethylated forms. Nevertheless, the majority of these post-translational marks occur on the amino-terminal and carboxy-terminal histone tail domains, but recent studies have documented that post-translation modifications in the central domains of the histones play a functional role in various biological functions. It is the eight amino-terminal tail domain that directs the formation of chromatin higher-order structures (Gordon F, 2005; Dorigo B, 2004; Dorigo B, 2003). However, the primary mechanism by which tail modifications act appears to be through altering the ability of non-histone proteins to interact with chromatin (Martin, 2005; Jenuwein, 2001).

Unlike acetylation, which generally correlates with transcriptional activation, histone lysine methylation can signal either activation or repression depending on the particular lysine residue that is methylated (Martin C., 2005).

Studies have shown that site-specific combinations of histone modifications correlate well with particular biological functions (Table 1). For instance; the combination of H4 K8 acetylation, H3 K14 acetylation, and H3 S10 phosphorylation is often associated with transcription. Conversely, tri-methylation of H3K9 and the lack of H3 and H4 acetylation correlate with transcriptional repression in higher eukaryotes. Methylation of K9, in turn, recruits the binding of repressor proteins, such as HP1 (Heterochromatin Protein 1) that helps in establishing highly compacted and transcriptionally inactive regions of chromatin (Nakayama M., 2001). Particular patterns of histone modifications also correlate with global chromatin dynamics, as diacetylation of histone H4 at K4 and K12 is associated with histone deposition at S phase, and phosphorylation of histone H2A (at S1 and T119) and H3 (at T3, S10 and S28) appear to be hallmarks of condensed mitotic chromatin.
In between nucleosomes there is the linker DNA that is composed of short unstructured amino-terminal “tail” domains (NTD), a central globular winged-helix domain and a long (~100 amino acid) highly basic, unstructured C-terminal domain (Wolffe AP, 1997). Binding of linker histones to nucleosomes and the linker DNA between adjacent nucleosomes affects as well the chromatin structure. Linker histones protect an additional 20 bp of DNA from nuclease digestion due to formation of a stem-loop structure in the linker DNA upon H1 binding (Bednar J, 1998). It also stabilizes the extensively folded 30 nm state, without fundamentally altering folding pathways specified by the core histone NTDs (Carruthers LM, 2000).

**Table 1.** Most common histone modifications with their possible biological functions. (From Peterson C.L., 2004)

<table>
<thead>
<tr>
<th>Modification</th>
<th>Histone</th>
<th>Site</th>
<th>Enzyme</th>
<th>Possible function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acetylation</td>
<td>H2A</td>
<td>K4 (S. cerevisiae)</td>
<td>Esa1</td>
<td>Transcriptional activation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>K5 (mammals)</td>
<td>Tip60, p300/CBP</td>
<td>Transcriptional activation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>K7 (S. cerevisiae)</td>
<td>Har1, Esa1</td>
<td>?</td>
</tr>
<tr>
<td></td>
<td>H2B</td>
<td>K5</td>
<td>ATF2</td>
<td>Transcriptional activation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>K11 (S. cerevisiae)</td>
<td>Gcn5</td>
<td>Transcriptional activation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>K12 (mammals)</td>
<td>p300/CBP, ATF2</td>
<td>Transcriptional activation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>K16 (S. cerevisiae)</td>
<td>Gcn5, Esa1</td>
<td>Transcriptional activation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>K15 (mammals)</td>
<td>p300/CBP, ATF2</td>
<td>Transcriptional activation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>K20</td>
<td>p300</td>
<td>Transcriptional activation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>K4</td>
<td>Esa1, Hpa2</td>
<td>Transcriptional activation</td>
</tr>
<tr>
<td>H3</td>
<td>K9</td>
<td>?</td>
<td>Gcn5, SRC-1</td>
<td>Histone deposition, Transcriptional activation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>K14</td>
<td>Gcn5, PCAF, Esa1, Tip60, SRC-1, Elp3, Hps2, hTFIHC90, TAF1, Sas2, Sas3, p300</td>
<td>Transcriptional activation, DNA repair, Transcriptional activation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>K18</td>
<td>Gcn5 (SAGA/STAGA complex), p300, CBP</td>
<td>Transcriptional activation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>K23</td>
<td>Gcn5 (SAGA/STAGA complex)</td>
<td>Transcriptional activation</td>
</tr>
<tr>
<td>Methylation</td>
<td>H3</td>
<td>K4</td>
<td>Set1 (yeast)</td>
<td>Set9 (vertebrates)</td>
</tr>
<tr>
<td>-------------</td>
<td>----</td>
<td>----</td>
<td>--------------</td>
<td>------------------</td>
</tr>
<tr>
<td>H4</td>
<td>K4</td>
<td>Set1 (yeast)</td>
<td>Set9 (vertebrates)</td>
<td>MLL, Trx</td>
</tr>
<tr>
<td>K9</td>
<td>Suv39h, Cle4</td>
<td>G9a</td>
<td>SETDB1</td>
<td>Dim-5, Kryptonite</td>
</tr>
<tr>
<td>K27</td>
<td>Ezh2</td>
<td>Transcriptional silencing</td>
<td>X inactivation (Me****)</td>
<td></td>
</tr>
<tr>
<td>K36</td>
<td>Set2</td>
<td>Transcriptional silencing</td>
<td>Transcriptional repression?</td>
<td></td>
</tr>
<tr>
<td>K79</td>
<td>Dot1p</td>
<td>Euchromatin</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R17</td>
<td>CARM1</td>
<td>Transcriptional activation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>K27</td>
<td>Ezh2</td>
<td>Transcriptional silencing</td>
<td>X inactivation (Me****)</td>
<td></td>
</tr>
<tr>
<td>K36</td>
<td>Set2</td>
<td>Transcriptional silencing</td>
<td>Transcriptional repression?</td>
<td></td>
</tr>
<tr>
<td>K79</td>
<td>Dot1p</td>
<td>Euchromatin</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: * Me1: mono-methylation, ** Me2: di-methylation, *** Me3: tri-methylation.
Chromatin

DNA is never found as a naked molecule in animal or plant cell. DNA is always found in association with proteins, most of them histones. Since the DNA must fit into the cell nucleus which is a few micrometres in diameter; so it is compacted in length of approximately 10,000 to 50,000 –fold since it range in length from 10 million to 100 billion base pairs (bp). This DNA-protein complex is the template for a number of essential cell processes including transcription and replication. As such, these processes depend on the precise structure of the chromatin fibers in order to take place. The basic structural unit of chromatin is the nucleosome (Kornberg R., 1974). Nucleosomes comprise of around 147 base pairs of DNA wrapped in a left-handed superhelix 1.7 times around a core histone octamer (2 molecules each of histones H2A, H2B, H3 and H4) (Figure 3). This 11 nm fiber is often referred to as “beads on a string”. Histone H1 or a related "linker" histone binds to the 40-70 bp of linker DNA that separates adjacent core particles and helps compact the beads-on-a-string into fibers ~30 nm in diameter (Finch J.T., 1976).

![Figure 3. Chromatin Structure](image)

Figure 3. Chromatin Structure. It is appreciable the long stretch of chromatin that have been folded to fit into the nucleus. It is also shown the distribution of nucleosomes, linker chromatin regions and the histone composition of the nucleosomes (Kindly provided by Michael Shogren-Knaak).

When cells divide, the chromatin is seen as distinct chromosomes, duplicating, with an equal partition of each set of chromosomes then traveling to each of the new daughter cells. When the new chromosomes reach the new cells, they begin to unravel into long thin extended 10 nm microfibrils called euchromatin or condensed coiled masses called heterochromatin.
The features that distinguish heterochromatin from euchromatin are described in Table 2 and involve both the binding of heterochromatin proteins such as HP1 and the post-transcriptional modifications of histones tails as they protrude from the nucleosomes.

Table 2. Differences between Euchromatin and Heterochromatin.

<table>
<thead>
<tr>
<th>Euchromatin</th>
<th>Heterochromatin</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lightly staining</td>
<td>Darkly staining</td>
</tr>
<tr>
<td>Decondensed, open</td>
<td>Condensed, closed</td>
</tr>
<tr>
<td>DNase sensitive</td>
<td>DNase less sensitive</td>
</tr>
<tr>
<td>Transcriptionally poised or active</td>
<td>Silent except for $\beta$ heterochromatin</td>
</tr>
<tr>
<td>Loosely positioned nucleosomes</td>
<td>Ordered, regularly spaced nucleosomes</td>
</tr>
<tr>
<td>Acetylated H3 and H4 histones</td>
<td>Hypoacetylated histones</td>
</tr>
<tr>
<td>H3K4me</td>
<td>H3K9me, HP1 (Swi6), HP2, Chp1 proteins</td>
</tr>
<tr>
<td>Early replicating</td>
<td>Late replicating</td>
</tr>
</tbody>
</table>

A particular chromatin state, once established, is then somatically maintained as a stable heritable epigenetic state. Establishment and maintenance involve methyl-CpG-binding proteins, ATP-dependent chromatin remodeling complexes, and histone-modification enzymes that acetylate, phosphorylate, methylate, and ubiquitinate core histone tails. Lysines at the amino-terminal ends of the core histones are the predominant sites of known regulatory modifications as already mention. Active genes are preferentially associated with highly acetylated histones whereas inactive genes are associated with hypoacetylated histones. Histone acetylation and deacetylation are thought to exert their regulatory effects on gene expression by altering the accessibility of the nucleosomal DNA to DNA-binding transcription activators, other chromatin modifying enzymes or multi-subunit chromatin remodeling complexes capable of displacing nucleosomes (Fuks F., 2000; Guschin D., 2000).

Constitutive or permanent heterochromatin can be classified in two types. $\alpha$ heterochromatin is characterized by late replicating, G band positive blocks of tandem repetitive sequences that are usually devoid of genes. $\beta$ heterochromatin was first described by Heitz (1934) in Drosophila polytene chromosomes as a diffusely banded region between the centric $\alpha$ heterochromatin and euchromatin. $\beta$ heterochromatin is thought as a junk yard for old retrotransposons.

Chromatin states tend to spread to nearby areas and so, insulators and barriers prevent the spreading of chromatin states across heterochromatin/euchromatin transitions (Gerasimova T.I., 2000; Bell A.C., 2001; Farrell C.M., 2002). It has been shown that there are initiation sites of heterochromatin as well as termination sites; this last one is also referred to
as heterochromatin insulators (Locke J., 1988). In mammals, such barriers can be facultative, for example, depending on the methylation status of CpG dinucleotides as exemplified in CTCF protein binding sites. This mechanism is used in the regulation of the imprinted gene activity and it will be discussed below.

The CTCF Protein

This protein was first described as being involved in the transcriptional regulation of the chicken c-myc gene, it was named CTCF (for ‘CCCTC-binding factor’) (Lobanenkov V.V., 1990). It was also found to bind a number of different sequences in the human, mouse and avian myc promoters; it binds to an unusually long 50 bp GC-rich sequence containing the core repeat sequence CCCTC. It negatively regulates myc in both mammals and birds. It is ubiquitously expressed. It contains a central 11 zinc finger (ZF) DNA binding domain that is flanked by 167 amino acids on the N-terminal side and 150 amino acids on the C-terminal side (Burcin M., 1997). Compared with the chicken CTCF gene, the mammalian counterparts have five evolutionarily new introns that have a high density of different classes of Alu- and LINE-family repeats. These differences do not change the CTCF open reading frame (ORF), indicating its conservation for the estimated 300 million years of evolution since birds and humans separated; also its identity rises to 100% for the region containing the 11 ZFs (Filippova G.N., 1996). The transcription repression is established by the interaction of CTCF with promoter elements and upstream silencer elements of the particular gene. It is suggested that CTCF recruits histone deacylases and hence mediates silencing (Lutz M., 2000).

But its functions have been shown to be quite diverse. To mention some of the most relevant it could be mention its implication in transcriptional activation of the amyloid β-protein precursor (APP) (Vostrov A.A., 1997). It is mainly recognized as an insulator, which refers to either a DNA sequence which serves to delimit an expression domain, or to the functional protein factor which binds to this sequence. Insulators are referred to as boundary elements, and serve to block the spread of the heterochromatin along the chromosome. Similarly, insulators can prevent the inappropriate action of transcriptional enhancers or silencers, bound cofactors, or covalent modification of the DNA on a neighboring promoter (Bell A.C., 2001). Not simply passive barriers, insulators are active participants in eukaryotic gene regulation. Insulators have been also found to be related with chromosome architecture, as observed from studies of the facet-strawberry allele of Notch (Vazquez J., 2000).

In this way, insulators serve to provide physical boundaries between active and silent chromatin domains. The insulator activity of CTCF was
first described in the chicken β-globin locus (Bell A.C., 1999) and then later it was also shown in the mouse and human (Farrell C.M., 2002), the chicken α-globin locus (Valadez-Graham V., 2004), the mouse T receptor locus (Magdinier F., 2004) and the mouse Pax6 locus (Li T., 2005). It is also implicated in maintaining a methylation free domain surrounding the BRCA1 promoter, where loss could lead to breast tumors (Butcher D.T., 2004) since, methylation of the DNA binding site for the insulator protein can block binding and insulator activity completely.

A more direct way of nullifying the action of an insulator is to prevent binding of the protein responsible for its activity. It has been shown that the imprinting mechanism at the Igf2/H19 locus uses this strategy (Bell A.C., 2000). Imprinting results in expression of H19 only from the maternally transmitted allele and of Igf2 only from the paternal allele. It is also observed that methylation of the imprinting control region (ICR) of this locus could somehow abolish insulator activity by inhibiting the binding of CTCF, which plays an important role in the insulator activity (Bell A.C., 2000; Hark A.T., 2000; Kanduri C., 2000; Holmgren C., 2001), allowing the downstream enhancer to activate Igf2 expression on the paternal allele. However, the Igf2 gene itself also contains differentially methylated regions (DMRs), with DMR1 being a methylation-sensitive silencer (Constancia M., 2000; Eden S., 2001) and DMR2 being a methylation sensitive activator (Murrell A., 2001). Additional enhancer sequences located 5’ of the ICR (Drewell R.A., 2002; Charalambous M., 2004) suggest that a simple insulator function of the ICR is not sufficient to explain all aspects of maternal silencing of Igf2.

To explain the mechanisms of this silencing, Reik’s group have shown the formation of a tight loop between the maternal H19 DMR and the Igf2 DMR1 were Igf2 is confined in an inactive domain away from the enhancers. In the case of the paternal allele, the H19 DMR interacts with the Igf2 DMR2, allowing Igf2 expression and not of H19 (Murrel A., 2004). But, this model has more elements, as shown that CTCF plays a role promoting the interaction between matrix attachment region (MAR)3 and DMR1, creating a tight loop around the maternal Igf2 locus that could be responsible for the silencing in this allele (Kurukuti S., 2006). Even more CTCF has also been associated with long long-range chromatin looping and local histone modifications in the β-globin locus (Splinter E., 2006). Also recently Ling (2006) showed interactions between Igf2-H19 ICR on the maternal chromosome were CTCF binds and Wsb1/Nf1 located in chromosome 11 and they showed that decrease in protein level of CTCF or mutation of the binding sites will disrupt this association. Implying that CTCF contribute to the formation of a high order chromatin structure that organized chromosome territories (discuss further in the papers of this thesis).

To widen up CTCF spectrum, it has been shown that the function of poly(ADP-ribosyl)ation (PARlation) that was traditionally associated with DNA repair and genotoxic stress, is also important in enhancer/promoter
complexes and so related to transcription regulation since, CTCF itself was found PARlated at the N-terminal domain when analyzed at the Igf2/H19 locus (Yu W., 2004).

Moreover, a new and novel form of epigenetic memory during cell division is shown by CTCF binding to mitotic chromosomes. It has been shown that, the chromatin loop organized by the CTCF-bound, differentially methylated region at the Igf2/H19 locus can be detected in mitosis. In contrast, the enhancer/promoter loop of the same locus is lost in mitosis (Burke L.J., 2005).

Replication Timing

Since the 1960s there has been evidence showing that during S phase in the cell cycle there is coordination between transcription and timing of DNA replication. The supporting evidence as well as new studies that create a certain discrepancy in this model will be discussed as follow.

Initially, S-phase dependent gene activation and epigenetic marks on chromatin provided means to discriminate transcriptionally active and repressive states. The inheritance of these repressive chromatin states during cell cycle can be achieved through a replication-coupled nucleosome assembly pathway and it is correlated with self-reinforced mechanisms concerning histone modification, DNA methylation, and some heterochromatin proteins. The first studies showed that genetically inactive heterochromatin, contained in Giemsa dark chromosome bands or ‘G bands’, replicates late during S phase, whereas most transcription takes place in Giemsa light or ‘R bands’, which replicate early during S phase (Stambrook PJ., 1970). Later on, studies at single genes showed that nearly all transcriptionally active genes replicate early in S phase, and more than half of the developmentally regulated genes replicate late when they are not expressed (Goldman MA., 1984; Hatton KS., 1988). Using the locus control region (LCR) of the human β-globin gene in mouse transgenes, Cedar’s group was able to show that the LCR contains dominant cis-acting elements that can regulate the time at which mammalian replication origins fire (Simon I., 2001).

Later, from the outcome of several studies, there has been a development of models describing the replication timing and transcription and these models are two reciprocal but not mutually exclusive working models. In the first model, transcriptional potential is established by synthesizing DNA at times when specific proteins are available for assembly into chromatin. For example, early replicating DNA would have a competitive advantage for binding limiting concentrations of transcriptional activators (Gilbert DM., 1986), whereas proteins that facilitate the assembly of heterochromatin would be available only during late S phase (Rountree MR., 2000). An
alternative model proposes that the structure of heterochromatin delays the initiation of replication, perhaps by restricting access of essential replication proteins to chromatin (Heun P., 2001).

An interesting model to study replication timing is allelic asynchrony. In monoallelically expressed genes the replication timing is asynchronous such as in imprinted genes (Simon I., 1999), genes encoding olfactory receptors were the decision of paternal replication is in random manner (Chess A., 1994) and the female X chromosome (Wutz A., 2000). This asynchrony is established early in development; the Cedar’s group could observe the asynchrony of imprinting genes after fertilization. But in germ cells this pattern is erased before meiosis and reset later in gametogenesis (Simon I., 1999).

Several studies have also suggested that replication timing is re-established in each cell cycle and by modifications of chromatin that take place as sequences are re-positioned after mitosis. In mammalian cells, the events that establish a replication program can be taken apart by incubating nuclei from synchronized cells in extracts from replication-competent Xenopus eggs (Dimitrova D.S., 1999; Li F., 2001). These studies have defined a discrete window of time in early G1 phase during which the replication-timing program is established, known as the ‘timing decision point’ (TDP). These experiments suggest a model in which chromatin regulators disperse during mitosis and are then re-concentrated into subnuclear compartments by a clustering of related chromosomal domains. Chromatin coming into contact with these compartments would be modified in trans by the locally high concentrations of these regulators, setting thresholds for replication timing. The determining event is probably not histone modifications, as patterns of histone acetylation and methylation are largely maintained during mitosis, but it could be the association of other chromatin regulators such as chromodomain proteins.

Discrepancies between the general association of transcription and replication timing in the S-phase were first elucidated by the studies in exponentially growing CHO-K1 Hamster cells were the newly synthesized DNA was labeled with 5-mercuri-dCTP and [3H]dTPP and so studied in single-copy gene level that can later be compared with their expression levels. This showed that there was a general correlation between transcriptional activity and replication in the first half of S phase, but examination of specific genes revealed a number of exceptions. Approximately 25% of total poly(A) RNA was transcribed from the late-replicating DNA (Taljanidisz J., 1989).

In genome wide studies in budding yeast, the relationship between replication timing and transcription was not totally corroborated (Raghuraman M.K., 2001). Raghuraman and colleagues identified the temporal program of replication timing, its origins and the time and frequency. But unexpectedly no global correlation between transcription and
replication timing was observed. Also, looking at genes of embryonic origins in *Drosophila* the connection is ambivalent (Schubeler D., 2002). It was shown that 30% of the early replicating genes were inactive and the same percent of late replicating genes were active. One possible explanation could be the large size of the replication timing domains that will include several different genes and possibly include genes with different transcription competence. In lymphocytes studies, it was also clear the association of leukocyte-specific genes that regardless of their expression they were always replicating early (Azuara V., 2003).

Analyzing the human genome, several studies demonstrated that early replicating regions contain relatively more active genes and, also appeared to include GC rich regions (White E.J., 2004; Woodfine K., 2004; Jeon Y., 2005).

This dynamic interplay between transcription and replication timing could be explained by the interplay of developmental gene expression and the housekeeping genes expressed in most tissues. It seems that the dynamics of these two groups of genes could have different regulation in terms of replication timing. In one sense, the control of developmental stages should have a more restricted way and in a certain manner the control of housekeeping genes should be straight forward. This explanation comes from the studies in the β-globin gene (Cimborna D.M., 2000; Simon I., 2001) gene and the immunoglobulin gene loci (Ermakova O.V., 1999; Zhou J., 2002). Also, the Fisher and Merkenschlager’s group has analyzed the replication timing of two developmental different cells, mouse embryonic stem cells and *in vitro* differentiated neural progenitors cells, showing an interesting mark of stem cells having most of the so called “stem cell markers” replicating early and upon differentiation they change to late replication. In contrast some of the neural specific genes studied were always replicating late in differentiated cells (Perry P., 2004). It is very important to note that they also found a change in chromatin accessibility asses by histone deacetylation as the cells differentiate and this in consequence could be related or a consequence of the change to late replication timing in the case of the *Rex1* locus.

Replication timing could also be important in the sense of considering long range interactions such as looping systems (described further). Regarding distal regulatory elements and how they could contact the gene promoter over a long distance, it could be explained by suggesting that DNA polymerases are immobilized by attaching to the large architecture where they reel their templates and extrude the newly synthesized DNA stand (Cook P.R., 1999). Now, by a replication dependent model there will be an opportunity for physically tethering the distantly separated elements together.

Finally, abnormal allelic asynchrony has become a useful clinical marker for predicting malignant cancers, and defects in replication timing have been

**Genome Organization**

The nuclear volume contains morphologically distinct higher-order chromatin domains, such as condensed heterochromatin, and numerous membranes-less proteinaceous sub-compartments, including the nucleolus and multitude of small nuclear bodies. A nuclear compartment is defined as a macroscopic region within the nucleus that is morphologically and/or functionally distinct from its surrounding (Dundr M., 2001; Matera A.G., 1999). The physically distinct nature of each compartment not only contributes to spatially partition the nucleus but also creates distinct functional subdomains within the nucleus. The absence of membranes provokes the question of how nuclear compartments are formed and maintained. An alternative possibility is that nuclear bodies are largely the results of the sum of many, likely transient and non-specific, interactions among its resident proteins.

Chromosomes themselves are non-randomly arranged within the nuclear space and occupy preferential positions relative to the center of the nucleus and relative to each other. The fact that both negative regulation, as well as positive regulation can take place simultaneously in close proximity suggest that the involved control mechanisms act locally, possibly at the single gene level.

The central question into understand whether nonrandom gene positioning is the cause or consequence of gene transcription. The association of a gene locus with a particular nuclear landmark may be strictly a reflection of the gene’s functional status. It could be that the positioning of the gene may precede changes in gene activity and might thus be a prerequisite for proper function and even serve a regulatory role. The recruitment of a certain loci to a specific location enhancing its translation will only be optimized by the association with a transcription domain.

The accessibility of eukaryotic DNA is dependent upon the hierarchical level of chromatin organization. New evidence has open the thoughts of long range interactions and regulations and the possible physical contact of regulatory elements either intra or interchromosomally. The description of active chromatin hubs is becoming very interesting. The position of activate transcription as well as exclusion in the case of needing to repress transcription in heterochromatic regions fits perfectly with the need to control different elements. Organized factories or territories of active transcription controlled by time and space in the nucleus and in its inheritance regulations in terms of epigenetic marks is giving the explanations of many unsolved biological questions of the past.
In order to stimulate the activity of a promoter in a large distance, there are cis-acting elements such as enhancers. They could be located in different position in respect with the promoter where they act.

Several models explain the communication between enhancer and promoters. The looping model explains the formation of a loop of DNA to be able to overcome large distances between distant elements. The tracking or scanning model, describes how factors are localized at the enhancer regions and from there they roll in a read out manner until they find the promoter region. The linking model is a combination of the above models and it explains an establishment of modified chromatin domains between the enhancer and the promoter by facilitator proteins which generates a progressive chain of higher order complexes along the chromatin fiber and so establish a communication without direct contact. But perhaps, it could be that there is no unique model, but instead a combinatorial repertoire that will fit the aspects of differential gene expression.

**Long Range Interactions**

Gene expression is controlled by promoter sequences located upstream of the transcriptional start site of the gene. There are also so called cis-regulatory elements that can be found around or within the gene itself. These regulatory elements are a few hundred base pairs and are rich in binding sites for transcription factors; they are also hypersensitive to digestion by DNase I, giving a reflection of local rearrangement of nucleosomes and possibly a local distortion in DNA topology (Leach K.M., 2001). Transcriptional enhancers function at a distance from their target genes to increase the use of promoters and have similar organizational properties discussed previously.

Many developmental important genes are control by complex transcriptional enhancers called locus control region (LCR), characterized also by DNase I hypersensitive sites. Using mouse models, it was shown that an important difference between LCRs and enhancers is the ability to have integration-position-independent and copy number dependent expression pattern (Grosveld F, 1987). With this, it’s assumed that LCRs provide a dominant chromatin opening characteristic. Unlike insulators, LCRs do not need to flank a gene. For example, the combination of four hypersensitive sites comprising the β-globin LCR with a β-globin gene is sufficient to counteract repressive signals coming from either side of any integration site in the genome (Li Q., 1999). It also has a strong enhancer activity and the ability to establish domains of histone modifications characterized by acetylation of histone H3 and H4 and methylation of H3K4 (Hebbes T.R., 1994; Litt M.D., 2001). LCRs are also implicated in establishment of replication timing and DNA methylation patterns (Simon I., 2001).
In most cases enhancers or LCRs are located tens of kilobases (kb) away from their target gene, but it has been also found that there could be also distances of up to a megabase (Lettice LA, 2003). Having this in account, an erroneous activation of the nearby genes could be possible, giving the possibility of a promiscuous activation. In other scenario, one has to keep in mind that transcription silencing could be achieved by the transcription repression of the condense chromatin state and this condensation is actually self-propagating, leading to the possibility of accidentally repressing genes that must be active.

In this context, a possible way of controlling a accidentally miss activation by distant enhancers or a nearby chromatin repression could by organizing genes in chromatin domains that are maintained independently of their surroundings by the establishment of chromatin boundaries. Specialized DNA sequences termed insulators (discussed earlier) can establish these boundaries (West A.G., 2002).

The interaction of the LRC with its target gene is a long range interaction. This interactions could be explained by the formation of active chromatin hubs (ACH) that brings DNA elements together, even if they are some kb apart from each other (see Figure 4); and this is established by the affinity of enhancer-promoter interactions giving an autonomy of adjacent genes rather than the presence of insulating elements (Dillon N., 2000). Interesting information come from the studies in the β-globin locus, were the β-major and β-minor loci interact with a hypersensitive site located 40-60 kb away; but do not interact with the non expressing embryonic β-globin. Similar results were obtained by separate groups using different technologies, 3C (chromosome conformation capture) and RNA-TRAP (Tolhuis B., 2002; Carter D., 2002).

**Figure 4. Representation of the Active Chromatin Hub.** Diagram of a gene cluster embedded in heterochromatin (green), with active genes and cognate cis-regulatory sequences in red and inactive genes in blue. The cluster of regulatory sequences and active genes forms an active chromatin hub (ACH) (kindly provided by de Laat W).
Entry of new regulatory elements may further stabilize or destabilize existing interactions and alter the expression levels of genes present in the ACH. The stable formation of an ACH is the critical event buffering against position effects in transgenic experiments. The observation that multiple combinations of cis-regulatory elements can confer position-independent expression on transgenes suggests that these combinations share the capacity to form a stable ACH anywhere in the genome.

In this context it is still needed to understand the organization of structures that coordinates with certain areas or spatial locations in the nucleus and given like this a certain gene activity. Genes that are silenced and associated with heterochromatin states are usually found in the nuclear periphery, the contrary of active genes that are found more towards the interior of the nucleus. Having this in mind, we can define chromosome territories (CT) as a spatial localization in the nucleus and interchromatin compartment (IC) as a compartmentalized structure that contains macromolecular complexes that are required for all biological functions like replication, transcription, splicing and repair. These areas can also form “channels” along which RNA is transported out of the nucleus (Kramer J., 1994). An ordered nuclear structure and the concept of specific chromosomes positions in interphase had started to become more tangible by the use of fluorescence in situ hybridization (FISH) (Manuelidis L., 1985). Later on, in neuronal cells, large chromatin movements have been noted during differentiation (Manuelidis L., 1990; De Boni U., 1994) or in pathological situations (Borden J., 1988). Large-scale chromatin movements were also observed during the interphase in Drosophila cells (Csink A.K., 1998). In human, chromosome 18 and 19 have a similar DNA content (85 and 67 Mb, respectively), the gene-poor chromosome 18 territories are typically found at the nuclear periphery and interesting to note that this position is establish early in the cell cycle, whereas the gene-rich chromosome 19, have the tendency to adopt a more internal position (Croft J.A., 1999). It was also shown by Bickmore’s laboratory, that these positions of chromosome 18 and 19 are changed if transcription levels are modified.

Regulatory and coding sequences of active genes can interact with the transcription machinery only when they are positioned at the surface of chromatin domains that line the IC, or on chromatin loops that extend into the IC, giving the topology of transcriptionally active genes (Figure 5 e,f). It is also known that genes located over 40 megabases apart in mouse chromosome 7 can interact in cis in creating in this way transcription factories and it is thought that chromatin loops from nearby chromosomes could also occupy the same factory (Osborne C.S., 2004). Also it has been shown the intra-chromosomal interaction between the promoters of Il4, Il5 and Il13 and the T-helper-cell 2 cytokine (T₄2) LCR, these interactions are spread over a 120 kb region. Importantly, there are also interchromosomal interactions between IFN-γ in chromosome 10 and the regulatory regions in
**Figure 5. Model of functional nuclear architecture.** Structural features that support the chromosome-territory–interchromatin-compartment (CT–IC) model are shown. These features are drawn roughly to scale on an optical section taken from the nucleus of a living HeLa cell. **A.** A giant chromatin loop with several active genes (red) expands from the CT surface into the IC space. **B.** CTs contain separate arm domains for the short (p) and long chromosome arms (q), and a centromeric domain (asterisks). Inset: topological model of gene regulation. Top, actively transcribed genes (white) are located on a chromatin loop that is remote from centromeric heterochromatin. Bottom, recruitment of the same genes (black) to the centromeric heterochromatin leads to their silencing. **C.** CTs have variable chromatin density (dark brown, high density; light yellow, low density). Loose chromatin expands into the IC, whereas the densest chromatin is remote from the IC. **D.** CT showing early-replicating chromatin domains (green) and mid-to-late-replicating chromatin domains (red). Each domain comprises ~1 Mb. Gene-poor chromatin (red), is preferentially located at the nuclear periphery and in close contact with the nuclear lamina (yellow), as well as with in-foldings of the lamina and around the nucleolus (nu). Gene-rich chromatin (green) is located between the gene-poor compartments. **E.** Higher-order chromatin structures built up from a hierarchy of chromatin fibers. Inset: this topological view of gene regulation.
indicates that active genes (white dots) are at the surface of convoluted chromatin fibers. Silenced genes (black dots) may be located towards the interior of the chromatin structure. F. The CT–IC model predicts that the IC (green) contains complexes (orange dots) and larger non-chromatin domains (aggregations of orange dots) for transcription, splicing, DNA replication and repair. G. CT with ~1-Mb chromatin domains (red) and IC (green) expanding between these domains. Inset: the topological relationships between the IC, and active and inactive genes. The finest branches of the IC end between ~100-kb chromatin domains. Top: active genes (white dots) are located at the surface of these domains, whereas silenced genes (black dots) are located in the interior. Bottom: alternatively, closed ~100-kb chromatin domains with silenced genes are transformed into an open configuration before transcriptional activation. Taken from (Cremer T., 2001) (permission granted from Nature Publishing Group).

An important observation is the possibility that genes that share IC in a given time could also be prompt to have higher potential of chromosomal translocations since it may be an exchange formation requiring that two (or more) genomic loci come together spatially as they do in a shared IC (Hlatky L., 2002). Holley and colleagues (2002), developed a model for evaluating radiation-induced chromosomal exchanges by explicitly taking into account interphase (G(0)/G(1)) chromosome structure, nuclear organization of chromosomes, the production of double-strand breaks (DSBs), and the subsequent rejoining in a faithful or unfaithful manner. Recently it has also been visualized this possibility with experiments using ultrathin cryosections and electron microscopy of different human cell lines (Branco M., 2006).

But biology is not that simple, and there are observations that the RNA polymerase II enzyme is present in the whole nucleoplasmic space offering a homogenous distribution of local concentration points (Jackson D.A., 1998); this tends toward a view in which transcription is not spatially restricted. Also, polymerases and other giant macromolecular complexes of the cell have a predicted tendency to associate in nonspecific entropy driven macrostructures (Marenduzzo D., 2006), possibly explaining the relative immobility of transcription factories relative to translocating DNA molecules. Thus giving more opportunities for an open discussion and more work for the scientific community trying to elucidate and come up with the real truth of how the system works.
Aims of the Present Studies

Paper I:

As more information gathers, CTCF is a clearly important protein in the cell. It’s a ubiquitously expressed protein and very conserved in chordates. Given so, it is an important task to understand further roles and key regulations of this protein. It has a wide variety of target sequences, due to the versatility of its 11 ZF, which makes the genome wide mapping of CTCF binding sites quite difficult even through the immense tools of bioinformatics.

To approach the task of lighting more the secrets of CTCF, it was decided to create a CTCF target site library from chromatin immunoprecipitated DNA (ChIP) using a specific C-terminal CTCF antibody in mouse fetal liver cells. In combination, this library was used to spot microarray slides as a method that could be used for further screening. The occupancy information coming from the library will allow to identity more possible roles as well as to identify more chromatin insulator regions where CTCF acts as a boundary and also explore more about its methylation sensitive binding properties.

Paper II:

Important information has been showing the versatility of chromosome territories when it comes to gene expression. Moreover, chromosome interactions are being proposed as an intrinsic method for ensuring the proper activation of genes depending on several factors given by a cell specificity and developmental order. Different methods have been used to gather this information as described earlier, but the biggest inconvenience is the previous assumption of the interaction which leads to a lack in discovering possible new key partners.

To elucidate more interactions and thus gather more information that ultimately will lead to the formulation of new models, it was decided to approach these interactions as the fact that sequences must be in close physical proximity in order to interact, allowing us to interlace them by high chances of ligating them together in formaldehyde cross-linked material.
Paper III:
Several lines of evidence suggest a relation between replication timing and gene expression, although new evidence seems to have some discrepancy with the original model. The possible relation is being focused more on the actual epigenetic status of the gene and with this its chromatin conformation. In studies from several groups on the replication timing status of the imprinted $H19/Igf2$ locus this apparent controversy is being exemplified. Based on the fact that it is an imprinting locus and thus has a differential allele regulation and also different epigenetic status, this locus is quite tempting to use as a perfect model for a better understanding of these issues. Besides, this locus also has relation with the protein CTCF, as mentioned earlier and in paper I, and has various roles in the dynamics of other imprinted genes and on many epigenetic regulatory systems.

In order to clarify this issue, we set up the task to analyze by different methods the replication status of the two parental alleles in the $H19/Igf2$ locus in wild type mice and in mice where the imprinting domain is disrupted. Also, since the differentiation procedure gives rise to genome wide modifications not only in organization of the epigenetic status but obviously leading to changes in expression; we were also proposed to analyzed in mouse embryonic stem cells the CTCF occupancy and its relation with replication timing, opening views for its functionality.

Paper IV
As any normal leaving specimen, any given cell could encounter several stress events at any point. The intrinsic response to these events is given by the ability of the cell to overcome and rectify a possible mis-regulation in gene expression. In order to give starting points that will help to clarify and generate models of cell response, we generated particular stress situations that were expected to generate cell responses given by changes in epigenetic marks and as expected also changes in the transcription repertoire.
Results and Discussion

Paper I

In order to create a library of DNA sequences binding to CTCF, we performed ChIP assays on mouse fetal liver cells using a specific C-terminal CTCF antibody. The ChIP DNA obtained was ligated to T7 and T3 linkers and PCR amplified. These products were cloned into pGEM vector that was used to perform a second amplification with T7 and SP6 primers in order to use this material for sequencing and elimination of duplicates. Since extensive amplification was needed and possible bias as a consequence, it was performed bandshift assays and multiplex PCR from randomly selected clones. After these, 266 clones were selected and spotted on poly-L-lysine coated glass slides creating in this way a CTCF target site library microarrays that was used for further studies.

After the analysis of the sequenced material, it was clear that the majority of the clones were not identified in the existing mouse genome libraries. The explanation could be that most of the genome libraries are not covering well the heterochromatic regions, opening the remark that CTCF highly binds these regions and increasing the correlation with late replication as it will be studied in paper III. To clarify this possibility, the association with a well recognized heterochromatin protein HP1β was analyzed by immunofluorescent studies in mouse fibroblast cells, showing significant co-localization.

In order to study further these set of sequences, it was found that 55 sequences were uniquely mapped within the genome, from these, 21 were intronic, 2 exonic, one covered an intron/exon boundary, 2 were in known imprinted domains (Grb10 and Snrpn) and the remainder were intergenic. To further characterize these target sites, a gene ontology database was used, showing that 9 clones mapped to loci implicated in cancer, 4 to loci involve in the ubiquitin pathway, 5 to G protein signaling pathways, 3 to the Wnt signaling pathway, 3 to apoptotic pathways, 6 to neurogenesis and 4 to olfactory or pheromone receptor genes, to mention the most relevant. It is relevant to note that no already known CTCF target sites were identified in the library, indicating that only a fraction of the binding sites were identify by the library and that the real number of CTCF binding sites is quite numerous. It is also important to note that among the sequences of the library several repeat sequences were identified, like long terminal repeats.
(LTRs), LINEs, SINEs, CpG islands and microsatellite repeats; these repeat sequences as already shown, play important roles in regulating specific gene expression.

Next, to validate these data, it was important to compare the in vitro binding of the library with immunopurified in vitro formed complexes between recombinant CTCF and the library; these complexes were labeled with Cy3 and the entire library was labeled with Cy5. These labeled samples were hybridized to the spotted microarrays slides described above. Following normalization procedures, it was clear that there is a wide variation in binding affinity, including a subset of clones that showed essentially no binding.

The identification of the in vivo binding that will confirm the binding probability as well as the specificity of the initial library was performed by hybridizing to the microarray slides the ChIP DNA from mouse fetal liver using this time an N-terminal CTCF antibody. By generating scatter plots comparing the in vitro binding and the in vivo binding, one could see some agreement confirming the binding of CTCF to the sequences of the library and some disagreement especially among the low signal of the weaker binding sequences. This could show that the binding of CTCF to a particular sequence in vivo is regulated by other factors, like accessibility for instance and these factors are not being exposed in the in vitro assay. It was also observed that there are highly binding sequences in the in vivo assay that are not in the in vitro assay, demonstrating a possible protein-protein interaction of CTCF with an unknown array of other proteins.

Since CTCF has been shown to act as an insulator in the H19/Igf2 locus it was important to assess the insulator abilities of the sequences from the library. For that reason, the pooled library was inserted between the SV40 enhancer and a toxin-A reporter gene with the H19 promoter in an episomal based toxin construct (Figure 6). Only sequences which could successfully prevent the enhancer-promoter communication can survive the selection. After transfection into JEG-3 cells, DNA was prepared from the surviving clones and hybridized to the microarray library and compare with the in vitro binding assay. This showed a great correlation proving the insulator activity of the library. Some individual clones were also use in the same assay, showing that they block the enhancer-promoter communication to various degrees, but two clones showed no insulator function at all since they failed to survive.

Figure 6. Schematic map of toxin construct.
At last, since CTCF binds in methylation free sites, it was important to test the library in this context. In this case, the microarray library was hybridized with labeled immunoprecipitated sequences using antibodies against 5-methylcytosine and the methyl-binding protein MBD2 which interacts with cluster of methylated CpGs. Scatter plots were generated and it was shown that the vast majority of CTCF binding sites are unmethylated. The few methylated binding sites could be imprinted genes in which one of the alleles could be methylated or that methylation is at the flanking sequences and so giving a normal background.

Paper II

To identify new key chromosomal interactions between the H19/Igf2 locus and other regions in the whole genome; it was important to achieve a method that will show physical close proximity of the different unknown sequences with the known H19/Igf2 locus. This method was called 4C (Circular Chromosome Conformation Capture). In order to observe these interactions, crosslinked DNA-protein complexes were subject to lengthy ligation procedures in which circular DNA structures were generated. In this scenario if a certain sequence had come in close position to the H19/Igf2 locus due to protein DNA interactions, this proximity will be revealed as the two interacting sequences will be ligated together and then, detected after reverse crosslinking and PCR amplification using nested primers that are located in the known region of the circular structure, the H19/Igf2. By cloning and sequencing of these products one could then identify and so map to the corresponding genome sequence.

Even before analyzing the sequences in detail, it is important the observation of a great variety of sizes in the amplified material showing a great variety of interacting partners. The analysis of neonatal mouse liver in this assay revealed 114 unique sequences. From which in 18 instances 2 or more different sequences from different chromosomes were found in the same clone, suggesting the promiscuous interaction between several chromosomes. It was also observed the occurrence of 16 clones in 2 or more independently established samples. Due to the fact that the H19/Igf2 locus is located in chromosome 7, it was also very notorious the over representation of sequences from the same chromosome indicating favoring interactions in cis; but in general sequences from all chromosomes were also obtained. It was also obtained sequences from 10 known imprinted domains and 11 from candidate imprinted domains. In general, the interactions mapped within intergenic and even more to intronic regions. Interestingly there a distinct mark showing that sequences proximal to H19 ICR showed no allelic preference, whereas sequences distal to the ICR were specific for the maternal allele.
In the H19/Igf2 locus it has been also described quite vastly the binding of CTCF in the ICR which helps to maintain the imprinted domain. It is also shown that CTCF is strictly related with interactions between different parts of the H19/Igf2 locus (Kurukuti S., 2006). So, disturbing the binding of CTCF in this locus will presumably disrupt some CTCF dependent interactions. For this purpose, it was used a knock in mouse model in which the 3 of the CTCF binding sequences were disturbed and so no CTCF is shown to bind in this region, depending the parental identity within the cross being used (Pant V., 2003). The allele carrying the disturbed binding is the 142* crossed with SD7; this mouse harbor the distal chromosome 7 imprinting region of the Mus spretus origin in a Mus musculus background. By performing 4C on 142* X SD7 and SD7 X 142* mouse liver cells it was shown that the majority of interactions (73%) were coming when the CTCF binding sites were not mutated, implying a major role of CTCF in the organization of these interactions.

An independent procedure known as the 3C assay (Dekker J., 2002) confirmed the interaction between identified sequences and the H19/Igf2 locus. In this assay it was also possible to discriminate the parental allele incurring in the interaction, and this allowed confirmation that the allele responsible for the interaction was the one with the normal CTCF binding sites and not in the case of the mutated allele.

In an attempt to actually visualize the interaction in vivo in mouse neonatal liver cells, 3D DNA FISH was performed using a region in the H19/Igf2 locus and sequences identified from the 4C library the Osbp1/a in chromosome 18 and Abcg2 in chromosome 6. The observed sequences were indeed interacting with each other. As controls, non interacting sequence regions in chromosomes 9 and 15 were not observed in close proximity.

To confirm the functional implication of these interactions, the expression levels of Osbp1/a and the 5’ imprinted gene Impact was assessed by comparing mRNA levels on both mouse crosses (142* X SD7 and SD7 X 142*). Results showed that there is a physical interaction between the maternal H19 ICR and Osbp1/a / Impact loci that influence the transcriptional activation of Impact and the repression of Osbp1/a.

Since the previous observation confirmed that chromosome interactions are related with gene expression; it was also important to compare two developmental different samples that are expected have different expression repertoires. In this case, mouse embryonic stem cells and in vitro differentiated stem cells were compared by 4C. As expected, many interactions were present in one or the other samples and some interactions were maintained. This showed that interactions in stem cells were lost or gained as the differentiation procedure took place. It was statistically different whether maternal specific or biallelic interactions were taking place in the two set of samples. Pronounced maternal interaction took place in the differentiated sample.
Paper III

As has been discussed earlier, the cells have several mechanisms to regulate the expression of genes, all this mechanism interact together. As it has been shown in paper I and also in paper II, CTCF plays an important role in setting key regulatory mechanisms for the imprinting of the \textit{H19/Igf2} locus. In consequence, it was important now to study the key regulation that CTCF imposes in the replication timing of the two parental alleles of this locus. For that reason, mouse fetal liver cells from the previously described crosses 142* X SD7 and SD7 X 142* were labeled with 5-bromo-2’-deoxyuridine (BrdU) in order to identify newly replicated sequences. Next, cell cycle fractionation of the S-phase was performed in accordance with PI (propidium iodide) staining (see figure 7). Following this, cells were lysed and the DNA was sonicated in order to be able to immunopurify the nascent DNA that was labeled with BrdU. This purified material was subject to quantitative real time PCR (qPCR). In this case, since we are able to identify parental allele, it was shown in agreement with others (Kawame H., 1995; Greally J.M., 1998; Simon I., 1999; Gribnau J., 2003); that the paternal allele replicates early and the maternal allele replicates late in the \textit{H19/Igf2} locus.

Figure 7. Scheme of the cell cycle fractionation

The next question was to see how is the replication timing varied between the mouse 142* X SD7 and SD7 X 142* crosses, since as described early the 142* allele carries a mutation that inhibits CTCF binding and thus there is a \textit{de novo} CpG methylation that consequently leads to biallelic expression of \textit{Igf2} (Pant V., 2004). For this case, the method describe above was performed on neonatal liver cells and followed by qPCR designed to be able to discriminate between parental alleles due to a polymorphism in the mouse strains. In this case, it was shown that due to the lack of binding of CTCF, the maternal mutated allele was shifted to early replication timing. Interestingly, this applies to two different cells types with different expression levels of \textit{H19} and \textit{Igf2} the brain and liver. This observation leads...
to an interesting thought that have also been discussed previously (Raghuraman M.K., 2001) that there is no straight correlation between expression and replication timing. There must be more parameters controlling these decisions.

Since the striking results above showed that CTCF has a role to play in the settlement of late replication in the mutated maternal inherited \( H19/Igf2 \) locus; it was also interesting to see this from a more genome wide position. In this case, and since we had characterized the CTCF library described in paper I, we analyzed the samples mentioned above from both mouse crosses (142* X SD7 and SD7 X 142*) and these samples were subjected to hybridization onto the microarray slides. The observation was also clear: the majority of the CTCF binding sites were associated with a late replication repertoire. Individual clones were also selected and analyzed by qPCR to confirm the data.

The next important observation was carried out by screening for changes in CTCF occupancy during embryonic stem cell differentiation. In this case, undifferentiated and \textit{in vitro} differentiated cells were used for ChIP analysis followed by hybridization to microarray slides and selection of individual clones for confirmation purposes. Results showed that there is a bulk of sequences that have the same degree of interaction in both populations of cells; this could be interpreted as the group of interactions required in all developmental stages as well as some constitutive interactions. In a subpopulation of sequences, there was easily observed a primary interaction in undifferentiated cells, showing probably interactions that are lost upon differentiation and which even could be interactions needed as a stem cell characteristic. From these differential interactions some clones were picked and subject to further analysis by S-phase fractionation and BrdU immunoprecipitation as previously described for the mouse neonatal liver cells. This analysis also revealed a high link between CTCF and late replication in both cells populations.

In addition, the control of replication can be regulated by various factors including factors located in the LCR that could be situated many kb distal from the replication origins. In order to gather a better understanding of the results so far described, it was important to observe replication origin by analyzing the samples so far examined (neonatal liver cells as well as undifferentiated and differentiated stem cells) in a high resolution (104 bp) microarray. It was interesting to observe that there was not apparent difference in BrdU incorporation in a region that includes the \( H19 \) gene and the \( H19 \) ICR between early and late replication in mouse fetal liver and undifferentiated and differentiated stem cells. The outcome observation is that there is a very define region between the \( H19 \) gene and the \( H19 \) ICR that is constantly replicated (both early and late S-phase).

As discussed earlier, CTCF is a very versatile protein that enables interactions with very distinct genomic domains. To explore this aspect in
terms of replication timing; the interaction between the 5′-end of the Igf2 locus and the 3′-end of the H19 locus was analyzed by using the conventional 3C method (outlined in paper II). This analysis showed that the 5′ and the 3′ regions are in close proximity, and by analyzing the interaction in an allele perspective, this interaction was shown to be almost exclusive to the paternally inherited domain. On the lack of CTCF binding, these interactions in the paternal allele becomes more exaggerated. This result indirectly led to infer the importance of CTCF in the interaction of several domains within the H19/Igf2 region that leads to a differential treatment of the replication origins in co-ordination with the imprinting repertoire of this locus.

The next important observation was in relation with the nuclear architecture of the cell. For this we used 3D FISH in order to ask for the association of the Igf2/H19 locus and the nucleoli cells from the two mouse crosses we have been using. The final observation was that the Igf2/H19 locus is highly associated with the nucleoli when the binding of CTCF was not disrupted in the maternal allele. With this observation, we confirm the importance of CTCF binding on specific sequences that are required to be localized in a late replication context.

Paper IV

It is important to understand the different mechanisms that a cell may use in order to overcome stress situations that will diminish its viability; and also to identify important features of those mechanisms that the cell and an organism develop in order to continue after deleterious regulations. In this scenario we have decided to artificially stress human 293T cells by exposing them to a stress imposed by a permeabilization procedure that also includes a cold shock. The first findings were the observed change in two well characterized epigenetic histone modifications, H3K9me3 that is associated with repressed chromosomal states and H3K4me3 associated with open chromosomal states. Using immunohistochemical staining it is observed a decrease staining pattern in H3K9me3 and a more nuclear periphery localization of the signal, leading us to speculate a re-deposition of repressed sequences in comparison with the untreated cells. The opposite effect is observed at H3K4me3 staining, were the pattern is increased as expected from this opposing mark.

Next, it was important to understand in a better way the possible cause of these changes, so we replace NTPs in the energy pump mixture from the protocol used before and used instead dNTPs since they are not energy donors. Again we could see big changes in the H3K9me3 (being decreased) and H3K4me3 (being increased). So we could then make the conclusion that
the responsible for most of the changes was that the cells were exposed according to the protocol to a cold shock.

Since we have been observing changes in epigenetic marks that are associated with levels of transcription, it was then necessary to estimate expression changes on these cells. By affymetrix expression analysis combined with gene ontology associations, we could identify several key pathways that are affected by the treatment. Of importance is the observation of two fold increase expression in the heat and cold shock genes. Suggesting that the cells were indeed stress by the cold shock performed at the beginning of the treatment. We also observed gene changes in cancer, epigenetic, metabolic, apoptosis, biosynthesis, DNA replication, cell cycle to name a few. In general, this information implies that the epigenome could be affected by stress signals and that the cell may response in several different ways to overcome these issues. The degree of the response will be dictated by an intrinsic epigenetic surveillance mechanism that will activate for instance apoptosis pathways upon relevant genome wide lesions.
Conclusions

The work presented here in this thesis is part of the already enormous efforts by this group as well as by others, where work is focused on creating a better understanding of gene regulation, and the importance of epigenetics in setting up the landmarks that will be recognized by the transcription machinery.

Accumulating evidence show the role of insulators as important factors in maintaining spatiotemporal marks that will maintain a certain boundary status corresponding to the need of marking or delimiting certain chromatin configuration, depending of the transcriptional repertoire of the locus they are acting on. This role of insulators was studied in the case of the protein CTCF acting as part of an insulator complex in the H19/Igf2 locus and in many other positions in the whole genome as illustrated in paper I.

Subsequently, another new emerging important factor that coordinates and perchance facilitates the expression of possible related genes in terms of cellular transcription is illustrated by the formation of chromosomal interactions. In paper II we showed how versatile is the H19/Igf2 locus in facilitating these interactions with a great variety of different loci in virtually all the chromosomes in the mouse genome. It will be of great relevance to visualize if all these interactions are present in one single cell, giving a remarkable possibility that big areas of the nucleous are just specialized in the transcription process or in regulatory functions, but for obvious limitations in material and the sensitivity of the method, we have our results in terms of pooled populations. In any case, the interactions are there to verify the concept of a big network and cooperation of the whole genome, and to discard the idea of the transcription machinery traveling here and there around the nucleus.

Later on, as a continuation of the concepts mentioned above, we have shown in paper III another clear link in the regulation of replication timing by the association of CTCF, which, as mentioned in paper I, interacts in different parts across the whole genome. This DNA binding protein could have the role of bringing together certain regulated gene loci to a particular position prompting late replication. Then the obvious question could be in regards to different repertoires within the genome. In this remark, it is important as well to mention that, as also shown in paper I, in order for CTCF to bind to a certain locus, it must have a certain chromatin
configuration such as methylation free, and this is obviously regulated by other factors that will act on a cell type specific and developmental context.

The last contribution of this work is the basic observation of changes in epigenomic marks after stress situations. An epigenetic surveillance mechanism that will be guarding the cells fitness and by any given stress situation several mechanisms will be induce in order to help the repair mechanisms or if the impact is un-reparable the cell will be driven toward cell death pathways.
Summary in Swedish

Epigenomets reglering

Underbara Moder Natur har skapat liv som är mycket komplext men ändå så perfekt och för att förstå dess mekanismer eller för att få en idé om hur storartat det fungerar krävs oerhört mycket arbete.

Den genetiska informationen som krävs för alltyper av levande celler kodas i sekvensinformationen som lagras i dubbelhelix, DNA. En trend inom biologin är att försöka förstå det fantastiska i att det är så många olika celltyper i människans kropp, vilka krävs för att bygga upp de olika organ som utgör den multicellulära organsim som människan är. Dessa olika celler härstammar och instrueras från samma skiss. Nyckeln till denna egenskap tillhandahålls genom vad man kallar i vetenskapen epigenetisk information, vilken dikterar hur, när och var gener ska aktiveras. Epigenetik fastställer generens aktivitet genom att packa det i speciella konformationer av DNA och proteiner, vilka kallas kromatinstrukturen. Dess konformationer reglerar möjligheten för åtkomst av nyckelfaktorer som läser informationen som ligger lagrat i DNA. Nyckelfaktorena i koordination med olika fysiska lokalisationer i cellkärnan bestämmer den ultimata uppgiften, expression. Det är viktigt att vi inte ser kärnan i cellen som en platt struktur, utan som en tredimensionell organiserad struktur där olika delar har sina speciella funktioner och att DNA inte är en rigid struktur utan är en formbar och dynamisk struktur.


Den här avhandlingen beskriver även upptäckten att epigenetiskt reglerade nätverk av intra- och interkromosomala komplex, identifierade genom den nya uppfunna 4C-metoden. Viktigt är att störningen av CTCFs
bindningsställen i H19-kontrollregionen inte bara avbryter förbindelsen i nätverket utan leder också till signifikanta förändringar i uttrycksnivåer i interagerande komponenter. Det är intressant att notera att CTCF spelar en viktig roll vid regleringen av tiden för replikering, inte bara för Igf2-genen utan även för alla andra sekvenser som binder till denna faktor, möjligtvis genom en cellcykel-specifik förflyttning av CTCF-DNA komplext till subnukleära delar vilket leder till olika replikationsmönster.

Slutligen, jag har visat att epigenetisk beteckningar som betecknar aktiva och inaktiva transkriptionstillstånd kan genereras eller försvinna beroende på stresssituationer som vi har introducerat. Genom att många gener som är involverade i apoptos, vilka är relaterade till celldöd, är uppreglerade föreslår vi att stressinducerade epigenetiska skador utgör ett övervakningssystem och som markerar celler som ska dö, vilket är fördelaktigt för organsimen. Denna viktiga observation öppnar för möjligheterna att nya inneboende mekanismer som cellen har för att upprätthålla korrekt genaktivitet och i de fall då det går fel finns det flera kontrollstationer som leder cellen mot viktiga överlevnadsbeslut.
Acknowledgements

The work presented here was a result of great interlaces efforts among all the people previous and current members at the Department of Animal Developmental and Genetics, in Uppsala University.

The first is my supervisor Professor Rolf Ohlsson, to whom I should express my gratitude for showing me the marvelous of the epigenetic world. In addition, for allowing me to see different important ways to approach to difficult and challenging experiments.

I should also give my thanks to Dr. Chandrasekhar Kanduri, for all the very important guidance in relation with crucial experiments and for all the very constructing comments that have of course help me to grow in laboratory experience.

To my formal supervisor Dr. Marcela Salazar for introducing me to the marvelous of questioning, reasoning, analyzing and in general all what is about to work in science.

My great appreciation, to Professor Lars Ährlund-Richter for bringing me here to Sweden and for showing me the wonderful world of embryonic stem cells.

My thanks also, to Professor Reinald Fundele and his group for the interesting comments.

Probably I will miss to mention some also important people that in one way or the other have built me up. However, any way thanks to whom I may miss writing here. Thanks to all the formal members of ZUB that I have been in touch with, many thanks to Joanne Whitehead for all the teaching at the beginning and for all the relevant comments work related since she left. To Vinod Pant also for important things I have learned from you. To Rituparna Mukhopadhyay as well for key things that lead me to important learning. My biggest thanks to Rolf Ericsson for all the help and guidance when I started to settle here in Sweden. To Sreenivasulu Kurukuti for all your wise knowledge and advice and for those important techniques you showed me. To Irina Kholodnyuk for so many discussions not only scientific but also life experience; all those advices have really made me think for some time and probably have shaped me. Thanks to the current members of ZUB. The first of all, Carina Östman, you have simply no idea of all the things you have though me, starting from skiing, which I think
have become one of my biggest sports activities, ice-skating that have lead me to discover even more, the beautiful nature of Sweden and so I could just keep on going. Thanks for showing me your courage, your strength in life, your incredible motivation and energy that just seems simply never ending. We all know we will have Carina for many many years and the important is that it will be with a gorgeous smile all the time. Thanks to Helena Malmikumpu you also have taught me many things, one specially be comprehensive. Many thanks to Claes Holmgren, you also taught me some things in the beginning and then it has been fun to have you around in the office. Thanks to you Tiger (Zhihu Zhao) for all your cooperation when I needed. Taras Shevchuk you also showed me many important aspects of lab work. Maybe I will not have enough time and space to give my special thanks to Rosita Bergström for all those scientific and specially non scientific conversations we have had, and all those advices and comments you gave me; and even more important for those horse back riding moments here in Sweden and in Estonia. Anita Göndör, you also have share with me good scientific and non-scientific long conversations, thanks also for hearing me on those crucial moments. To you Sha Wang for your nice collaborations, for nice advices in the lab and for those nice dinners we had together were you introduced me to the wonderful of China. To you, Piero Mariano, thanks also for those smoking pauses especially in winter. To Radha Raman Pandey, thanks for all the help and collaboration in the lab, but honestly thanks for all those moments you made me laugh when we were skiing and skating together. To Gholamreza Tavoosidana, thanks also for your collaboration and for all the talks, we had recently. Carole Rougier for your great help in the last part of this thesis and for teaching me the importance of wearing kneepads. To Mikael Sjölinde for your help here in the writing of the thesis as well as for helping me with some lab work and correcting me my Swedish homework's. This last also to Rebeca, you are one of the sweetest persons I have ever meet; you have such a wonderful heart. Thanks to Sylvan Guibert for your French help. Faizan for your collaboration. To Gunnar and Xuping for teaching me when I was teaching you. Rose Marie for taking care of the important things, the money. Jan Grawe my biggest and sincere thanks; I do not think I could have made it without your help, thanks for your expert help with the confocal and the facs. In this regards also thanks to Stefan for teaching me and helping me in the confocal at BSA.

Now it comes the people that actually have help me the most and the most important have been there and will be there even after this dissertation. My personal Best Friends. Thanks to you Andrea, my fore sweetest friend in the whole world. You have been there unconditioned even with such distance and with you taught schedule, you have showed me how to be strong and always with a smile in your face, even in the cruelest moments of our lives. Gracias mi chibchombiana preferida. To you Andres, it is unconceivable all
the moments we have share and the most important is that we have been there for each other especially in the bad ones. Thanks for your sweetness and loving care friendship. Thanks to you Rafael, for holding our hands, to us Andres and me. My dear Lilian you also have been there unconditional for me in good and bads, and even more not just only by word but actually helping me with all of you. Thanks to you Oscar, it was such a surprise that life made us to meet again here in Sweden, but thanks for all your hospitality and friendship here in Sweden and there in our precious, land. My dearest Marley the joy of speaking with you is indescribable, you know, thanks for all your friendship for all your care, all the moments we have share, all the warm feelings especially in those bad moments. Thanks for being there unconditionally giving me all your support and advice. Thanks to you Cinthya for introducing me even though it was for short but to the nations here in Uppsala. Ruben, thanks for teaching me how to cook, incredible but true. Juanra, incredible all your help in some experiments. Florecita, gracias por los bonitos momentos cuando estabas aqui en suecia. To you Sol, how a wonderful friend you are, thanks for your hospitality here in Stockholm and later there in Paris, for all our long conversations thanks for showing me to see different ways in life, gracias mil por tu amistad. To Lina for your great support here in Sweden.

To the Jägargatan 20 gang, and here I quote form you Kyriakos “is probably one of the most known addresses for international students in Stockholm” and then I quote for you Katta and the Hotel California Song, You can check out any time you like, but you can never leave. Yes as I suppose we will never forget those moments. Thanks Xiang Hua, Ia Ladestam, Xidan Li, Tara Malekshahi, Rafaele Attia, Anja Schue, Varinia Soledad, Karina Iliescu, Alexander Bondar, Hanna Karlsson, Asim, Astrid, Mahbub, Fabio, Evi, Alain.

To Elena Kashuba for those great advices and those important things you have done for me.

To Urban Forsberg and Karin Almgren, for allowing me to share those nice moments with your sweet and loving family. Thanks Urban for those brainstorming conversations, thanks for Nice and Åre and my biggest and most admirations to you Karin, one of the most outstanding examples in my life, your care with your gorgeous family and such a perfect example of wonderful success in your work. As I have said to Urban and to your son, you are simply a superhero.

Thanks to you, my sweet heart for all your comprehension in those long hours I spend in the lab, for your patience and for your love.

To my wonderful country, even though we have so hard times, some day, some how things will get better. We should persist and work hard because we have a gorgeous country. In addition, even if I sound pathetic, I should follow the example of Shakira and say, “This is pa´ mi gente”.
The last but actually THE MOST IMPORTANT, my mother. To whom I dedicate all I ever do in my life. Because it is from you Ma, that I have the courage and strength, the support and the shelter I will ever need. You have given your life and dedication to me and I have no words and no means to express my love for you, the only way is to say, Mami te quiero.
References


enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. Hum Mol Genet. 12: 1725-35.


replication-timing analysis of human chromosome 22 at high resolution and different developmental states." Proc Natl Acad Sci USA 101: 17771-17776.


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”.)