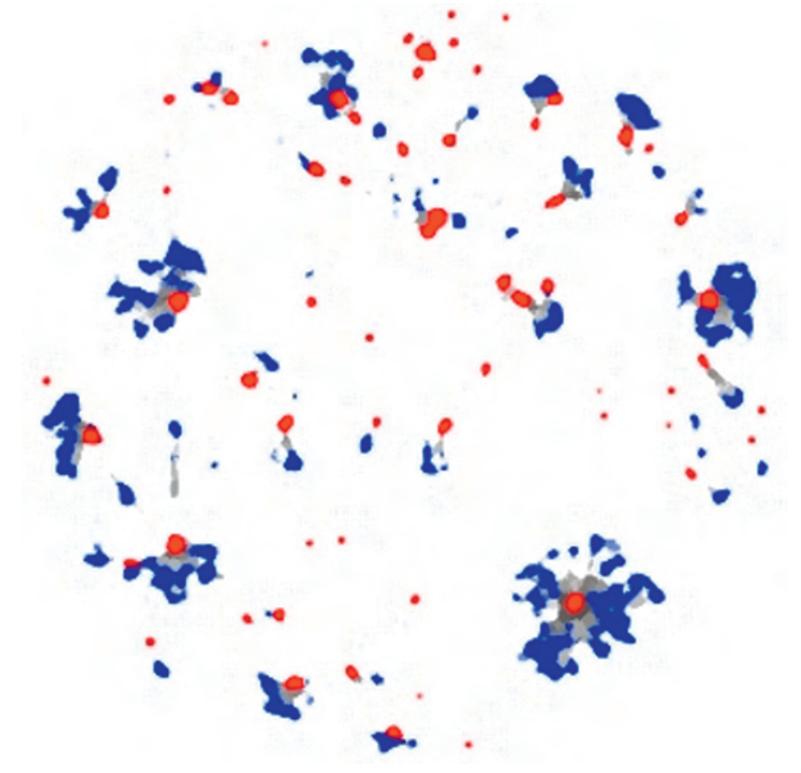


Thesis for doctoral degree (Ph.D.)  
2010

# NETWORKS IN EPIGENETICS



Thesis for doctoral degree (Ph.D.) 2010

NETWORKS IN EPIGENETICS

Kuljeet Sandhu



**Karolinska  
Institutet**

**200**  
1810 – 2010 *Years*

Kuljeet Sandhu



**Karolinska  
Institutet**

**200**  
1810 – 2010 *Years*

Department of Microbiology, Tumor and Cell biology  
Karolinska Institutet, Stockholm, Sweden

# NETWORKS IN EPIGENETICS

Kuljeet Sandhu



**Karolinska  
Institutet**

Stockholm 2010

All previously published papers were reproduced with permission from the publisher.

Published by Karolinska Institutet.

© Kuljeet Sandhu, 2010

ISBN 978-91-7409-876-1

Printed by



[www.reproprint.se](http://www.reproprint.se)

Gårdsvägen 4, 169 70 Solna

*Dedicated to our small-world*

*“All complex systems have their Achilles’ heel”*

**-LINKED, Albert-Laszlo Barabasi**

*“Single leader is a bad design”*

**-SYNC, Steven Strogatz**



## ABSTRACT

Networks are ubiquitous. It is only recently, we started exploring what purpose they serve in real world. Biological research supposedly has, and will be, benefited most from the network science because of inherent complexity embedded in its multilayered organization. While the metabolic networks (MNs), gene-regulatory networks (GRNs) and protein-interaction networks (PINs) are being studied extensively, networks in context of epigenetics are largely ignored. Moreover, chromatin interactions networks (CINs), i.e., physical cross-talk among chromatin loci within nucleus, are rather in its infant stage. The present thesis aims to unravel and characterize networks in epigenetics using experimental and systems approach.

We studied a CIN centered on one of the *Achilles' heels* in mammalian development, namely *H19* imprinting control region (*H19-ICR*). Malfunctioning of this small genomic locus has pleiotropic consequences. In particular, it breaks the barrier of parthenogenesis, predisposes the animal for cancer and hyper-sensitizes the growth factor *Igf2* post-receptor signaling. To explore CIN around *H19-ICR* (H-CIN), we devised a proximity-ligation based high resolution assay named as Circular Chromosome Conformation Capture, abbreviated as 4C. The method identifies all unknown interacting partners of a genomic locus of choice, *H19-ICR* in our case. Using 4C in combination with high throughput arrays, we uncovered wide-spread *cis*- and *trans*- interactions of *H19-ICR* in different mouse lineages. Incorporating an error in *H19-ICR*, that abolishes binding of CTCF protein, revealed allele-specific epigenetic regulation of H-CIN. Further analysis suggested that *H19-ICR* could influence the gene expression at distance. Dedicated experiments on embryonic stem cells and *in-vitro* derived embryoid bodies show extensive reprogramming of H-CIN, which appears to be regulated by dynamic movement of *H19-ICR* itself, followed up by upregulation of proximal genes during differentiation. Genes proximal to interaction sites show general traits of being developmentally regulated. Moreover, we see a significant over-representation of imprinted domains from 13 different chromosomes that further prompted us to uncover an imprinted CIN consistently in embryonic stem cells, embryoid bodies, somatic and germ-line lineages using quantitative *in-situ* experiments. Interestingly, CTCF binding sites within *H19-ICR* determines the physical proximity among imprinted loci and also transfect the epigenetic states, namely replication timing, of other imprinted loci during germ-line development. Comparative analysis on embryonic and germ-line stem cells indicates that the transvection of multiple imprinted loci by *H19-ICR* requires germ-line transmission. In brief, the study of H-CIN suggests that a single genomic locus could *trans*-regulate the epigenetic states of other genomic loci pleiotropically. Given the fact that *H19-ICR* is the ancient most ICR and replication asynchrony is possibly an early epigenetic feature of imprinted genes, our results might suggest that transvection through H-CIN might represent a possible evolutionary interplay in the establishment of imprinted clusters in placental mammals.

We further explored a possible link to CTCF/Cohesins, the popular partners in orchestrating higher order chromatin structures in *cis*. Analysis, however, negates a general link to these and suggests multiple mechanisms of CTCF mediated *trans*-interactions. As this being studied, in parallel we search for the conformational feature of CTCF protein that might facilitate its binding to diverse promoters and co-factors. We predict that CTCF contains structurally disordered regions alternating to zinc fingers and towards the open terminals suggesting a possibility of multi-conformation dynamics in CTCF structure that might contribute to its multiple interactions and eventually its diverse functions. This observation was an induction for the paper IV in the thesis wherein we report intrinsic structural disorder consistently in most chromatin modifiers. The physical feasibility ascribed by structural disorder might explain involvement of chromatin remodeling factors in diverse nuclear functions.

It is increasingly being realized that CINs often associate with co-regulons, which in turn may associate with PINs. Interestingly, imprinted genes, besides their CIN, also organize their gene regulatory network known as imprinted gene network (IGN). On similar lines, we attempt to analyze their PIN that might be an indirect consequence of CIN followed by IGN. Systems analyses of available human PIN data uncovered a highly central and tightly bound network module of imprinted gene-products and their interacting partners (IGPN) dedicated to imprinted gene-function. The robustness of human interactome is significantly compromised by this network and its malfunctioning makes the human interactome vulnerable to errors. We further show association of this network with several complex disorders. The study opens up novel systems perspectives in understanding imprinted gene-function in mammals.

Thus, the thesis deciphers novel networks implicated in epigenetics. In particular, we uncovered erroneous perturbations in imprinted interactomes, i.e., H-CIN and IGPN, which helps understanding their epigenetic and functional pleiotropy in mammalian imprintome.



## LIST OF PUBLICATIONS

The thesis is based on following articles:

- I. Zhao Z<sup>§</sup>, Tavoosidana G<sup>§</sup>, Sjölander M, Göndör A, Mariano P, Wang S, Kanduri C, Lezcano M, **Sandhu KS**, Singh U, Pant V, Tiwari V, Kurukuti S, Ohlsson R\*. Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nat Genet.* 2006 Nov;38(11):1341-7.
- II. **Sandhu KS**<sup>§</sup>, Shi C<sup>§</sup>, Sjölander M<sup>§</sup>, Zhao Z<sup>§</sup>, Gondor A, Liu L, Tiwari VK, Guibert S, Emilsson L, Imreh MP, Ohlsson R\*. Nonallelic transvection of multiple imprinted loci is organized by the *H19* imprinting control region during germline development. *Genes Dev.* 2009 Nov 15;23(22):2598-603.
- III. **Sandhu KS**, Zhao Z, Shi C, Kanduri C, Ohlsson R\*. Genome-wide chromatin interactions of mouse *H19* imprinting control locus negate a general link to CTCF. *Manuscript.*
- IV. **Sandhu KS**\*. Intrinsic disorder explains diverse nuclear roles of chromatin remodeling proteins. *J Mol Recognit.* 2009 Jan-Feb;22(1):1-8.
- V. **Sandhu KS**\*. Network of imprinted gene-products and partners ascribes error intolerance to human interactome. *Under review*

<sup>§</sup> Co-first author

\* Corresponding author

# CONTENTS

1. General context	1
2. Epigenetics in 3D	2
2.1. Epigenetics	2
2.2. Nonrandom chromatin organization	3
2.3. Chromatin Interactions Networks (CINs)	4
2.4. Factors mediating chromatin interactions	5
2.5. Role of crowding forces	5
2.6. Mosaic networks of genomic functions	7
3. Epigenetics and mammalian development	7
3.1. Genomic imprinting	7
3.2. Epigenetic reprogramming	9
3.3. Reprogramming of replication timing	10
4. Epigenetics and Protein Interaction Networks (PINs)	10
5. CINs vs PINs: spin together?	13
6. Summary of methods	14
6.1. Analysis of CINs	14
6.2. Analysis of PINs	15
6.3. Analysis of structural disorder	16
7. Results	17
7.1. Specific aims of the thesis	17
7.2. Summary of paper 1	17
7.3. Summary of paper 2	17
7.4. Summary of paper 3	19
7.5. Summary of paper 4	19
7.6. Summary of paper 5	20
8. Conclusions	21
9. Acknowledgements	22
10. References	23

## LIST OF ABBREVIATIONS

3C	Chromosome Conformation Capture
4C	Circular Chromosome Conformation Capture
BIND	Biomolecular Interaction Network Database
BWS	Beckwith Wiedemann Syndrome
ChIA	Chromatin Interactome Analysis
ChromDB	Chromatin DataBase
CIN	Chromatin Interaction Network
CT	Chromosomal Territory
CpG	Cytosine paired with Guanine
CREMOFAC	Chromatin REMdeling FACtor database
DIP	Database of Interacting Proteins
DMR	Differentially Methylated Region
DNA	Deoxy-ribo Nucleic Acid
ERE	Estrogen Response Element
ESC	Embryonic Stem cell
EB	Embryoid bodies
FRAP	Fluorescence Recovery After Photo-bleaching
FISH	Fluorescent <i>in-situ</i> hybridization
GRN	Gene Regulatory Network
H3K4	Histone 3 Lysine 4
H3K27	Histone 3 Lysine 27
HCG	High Cytosine-Guanine
H-CIN	<i>H19</i> -ICR CIN
HIN	Human Interactome Network
HPRD	Human Protein Reference Database
ICR	Imprinting Control Region
IGN	Imprinted Gene Network
IGP	Imprinted Gene-products and Partners
IGPN	IGP-Network
Kb	Kilobases
LCG	Low Cytosine-Guanine
LDR	Long Disordered Region
Mb	Megabases
NPC	Neural Precursor Cell
OMIM	Online Mendelian Inheritance in Man
PCR	Polymerase Chain Reaction
PET	Paired End Tag
PGC	Primordial Germ Cell
PIN	Protein Interaction Network
PPI	Protein-Protein Interaction
PRE	Polycomb Response Element
QQ	Quantile-Quantile
RNA	Ribo Nucleic Acid
SAM	S-adenosylmethionine
TCR	Transcription Coupled Repair
TRANSFAC	TRANScription FACtor database
WWW	World Wide Web
VIP	Very Important Person

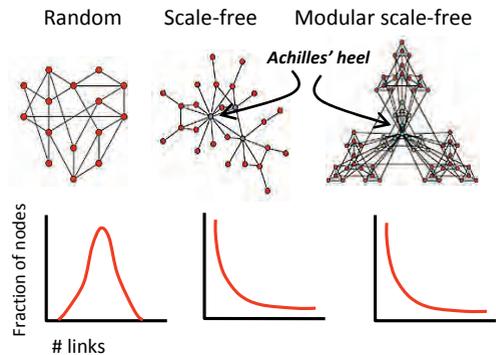
**Note:** Gene name abbreviations are skipped



## 1. General context

Components in the world around us are never independent, they have an intrinsic property of being connected. Connected to an extent that we regard the world as *small-world*. A prevailing view about real world networks for several decades has been its randomness. The surprising realization of real world networks being non-random came very late towards the end of last decade. A break-through discovery, that most real world networks from our society to world wide web, to organized crime, to ecosystems, to cell biology share the same basic skeleton of network architecture, has revolutionized our thinking of complex systems<sup>1</sup>. They belong to the same mathematical equation determining their nonrandom topology. The power-law equation  $p(k) \approx k^{-\lambda}$ , where  $p(k)$  is the proportion of nodes having sufficiently large  $k$  connections and  $\lambda$  is the a constant *degree exponent* whose value typically ranges  $2 < \lambda < 3$ , states that there would be very few nodes having large number of connections, while most other will have very few connections<sup>1</sup> (f.1). Though nature, in general, follows its favorite Gaussian distribution, it imposes power-law almost flawlessly when the system experiences self organization or a transition from chaos to order. Interestingly, power-law also acts in complex activity dynamics of real world, which is characterized by *bursts*, an inherent pattern of sudden intense activity spaced by lengthy intervals of inactivity<sup>2</sup>. Temporal pattern of our e-mail communication, earth-quacks, heart-beats, gene-expression are few examples of *burstiness*<sup>2</sup>. Thus, the power-law is increasingly being recognized as a generic organizing principle across several complex systems. Presence of this *trade mark* in most real world networks argued strongly against the random views of networks and eventually abandoned those. Such nonrandom network topology, also termed as scale-free topology, teaches us how complex systems acquire robustness<sup>1</sup>. Any random malfunctioning error in the scale free network is absorbed since most nodes do not connect sufficiently enough to rest of the network and thus fail to propagate the perturbation caused by the error. Moreover, given a sufficiently large scale-free network, a significant proportion of all nodes can be deleted without affecting its *small-world* property. At the same time, however, we have certain highly connected nodes or hubs that tie the whole network together. Selectively attacking these hubs would be fatal to the network. That is *Achilles' heel* (f.1). We have *Achilles' heel* in all complex systems<sup>1</sup>. Google on WWW, a manager or coordinator in an organization, P53 gene in cell, Arlanda airport in national/international flight network of Sweden etc are few examples of *Achilles' heel*. Therefore, the real world networks harbor robustness against random errors along with vulnerability to targeted attacks. Robustness explains why life persists even after several internal errors in the cell, and why food chains in our ecosystems sustain even after gradual loss of

several species. The vulnerability, on the other hand, explains why we see wide-spread black-out when certain node in our city's power-grid fails and why we get cancer by malfunctioning of P53 gene. However, single hub failures generally do not propagate after some extent, since there would be other hubs that would still keep the network alive. Additionally, the perturbation might also get trapped locally in organized modules of networks. Thus, the chance to have a multiple hub failure that could cause severe system-wide perturbation would approach zero given the power-law distribution of connections. That is, the robustness prevails. Nevertheless, *Achilles' heels* in the complex systems forces us to pay more attention to them. We learn how to control infectious disease outbreak by narrowing down to hubs, how to crack organized crime by targeting their key players, how to save the WWW from hackers by securing its hubs and how to cure a disease by keeping the cell's hubs healthy and functioning. A vast majority of biological scientists indeed focus on such *Achilles' heels* that were identified by their classical approach of introducing errors in the system and observing the phenotypic outcome.



**Fig. 1.** Origin of scale-free network model. Random network model of Erdős & Rényi, which prevailed for about half a century, has Gaussian distribution of node degrees. However, the real world networks are not random and possess a power-law (or scale-free) distribution of node-degrees as uncovered by Barabasi and group. A step further, Barabasi and group showed how a scale-free network could also be modular, as in the case of protein interaction and gene regulatory networks.

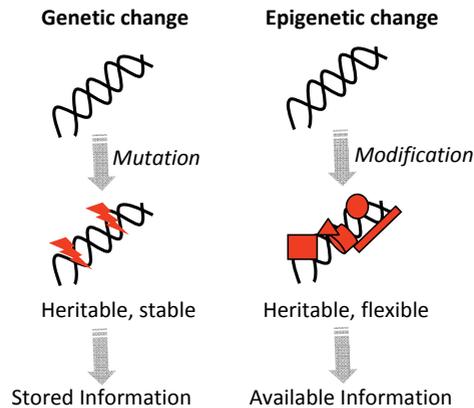
The present thesis is based on few such *Achilles' heels*, certainly not the only, in mammalian systems. Unlike traditional reductionist approach, the thesis identifies novel networks and their perturbations in context of epigenetics and mammalian development. In particular, we uncover an extensive chromatin interaction network (CIN) of a candidate *Achilles' heel* in mammalian development, namely *H19* imprinting control region or *H19-ICR*. We attempt to characterize functional and mechanical insights of these interactions. We further

extend the analysis to protein interaction network (PIN) and identified a development and metabolism related, highly central network-module that compromises with robustness of whole human PIN when malfunction. In yet another study we uncovered a common structural feature supposedly responsible for high degree and diversity of protein interactions, again in context of epigenetics. The text in the thesis makes some prior assumption about readers' basic understanding of molecular biology and gene regulation.

## 2. Epigenetics in 3D

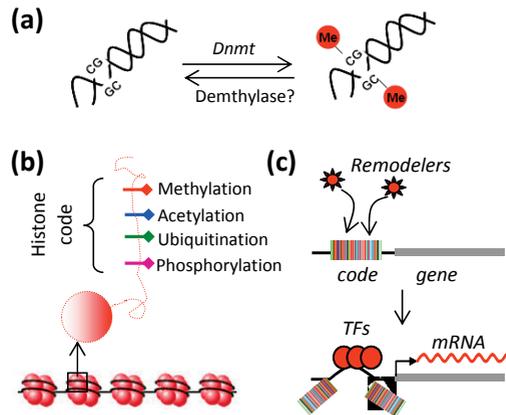
### 2.1. Epigenetics

After decades of work in genetics, research focus is now redirected towards epigenetics, i.e. the molecular changes, other than in underlying DNA sequence, that alter gene function. Like genetic changes (mutations), epigenetic changes are also heritable, though not necessarily. One of the earliest examples of heritable epigenetic change is the peloric variants of *Linaria* flowers, first described by Linnaeus. The gene responsible for this change inherits the silencing epigenetically<sup>3</sup>. In contrast to genetic changes, epigenetic changes are reversible and thus endow functional flexibility and diversity to the genome (f.2). Agouti gene determines coat color in mouse offspring depending on its epigenetic states in the mating parents<sup>4</sup>. Therefore, epigenetic changes determine the availability of genetic information stored in the DNA (f.2).



**Fig. 2.** Genetic vs. epigenetic change. As the term itself suggests ('epi' =on the top), the epigenetic change is placed on the top of DNA by diverse modifications (DNA methylation, histone modifications, non-coding RNAs etc) unlike genetic change that involves the change in DNA sequence itself (mutation). The epigenetic changes are reversible, thus ascribe flexibility to the genome and eventually determines the available content of the stored information in DNA.

There are several mechanisms of epigenetic modifications in eukaryotes. The most studied one is DNA methylation (f.3a), i. e. substitution of a hydrogen atom with methyl (-CH<sub>3</sub>) group at C5 position on cytosine base. This is the only known chemical modification of DNA molecule itself in eukaryotes. DNA methylation has been extensively studied at CpG residues present primarily at regulatory sequence elements like proximal promoters of the genes.



**Fig 3.** (a) DNA methylation at CpG residues. (b) Diverse post-translational modifications on histone tails (the histone code). (c) Histone code on gene promoter is recognized by chromatin remodeling complexes that subsequently remodels the nucleosomal array that enables the transcription factors to access DNA.

Based on CpG distribution, all promoters in mammalian genome(s) are categorized into two classes, those with low CpG density (LCG) and those with high CpG density (HCG) or CpG islands (CGI)<sup>5</sup>. HCG promoters, in general, associate with housekeeping genes, while LCGs are prominent in tissue specific genes<sup>5</sup>. Both these promoters have distinct modes of action. DNA methylation at HCGs usually results in chromatin condensation, thus, forming a repressive structure inaccessible to transcriptional machinery. LCGs, in contrast, are methylated at sparse CpG residues that restrict the binding of transcription factors to the DNA. Further, the eukaryotic DNA, in general, is wrapped around by octamers of histones (nucleosomal structures) to form a thick chromatin fiber (f.3b). This provides an additional epigenetic layer to the gene regulatory machinery by controlling DNA accessibility to the transcription factors. Diverse post-translational modifications (f.3b) are carried out on histone tails by histone modifying enzymes, which in turn can have distinct downstream effects on transcription and other nuclear functions<sup>6</sup>. Most of the studied modifications occur at the proximal promoter, gene body, enhancers and

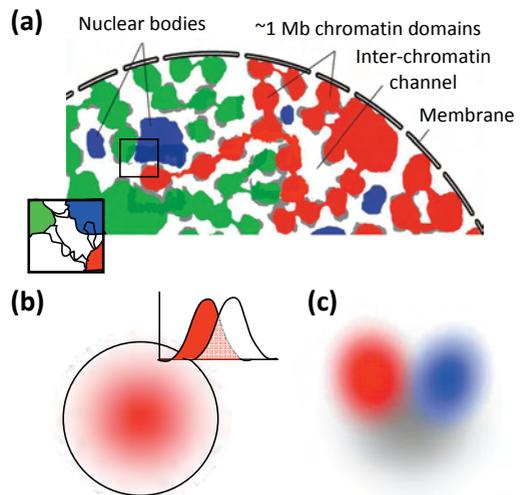
insulator elements<sup>7</sup>. The chromatin marks at enhancers, in particular, vary in different cell types and regulate cell-type specific transcriptional program<sup>7</sup>. In brief, the DNA methylation and diverse histone modifications together paint the genome in mosaic of active, inactive and poised states of chromatin, which in turn restrict the search space for regulatory proteins like chromatin remodeling factors. Chromatin remodeling factors make multimeric assemblies that recognize and reconfigure the chromatin states to enable the transcription factors access the cis-regulatory elements<sup>8</sup> (f.3c). Above these layers of regulation, is the layer of topological determinants. It is believed that higher order 3D orientation of chromatin fiber associates with function, however until recently there was lack of direct evidence of these phenomenon partly due to technical limitations. Subnuclear localization of genes, spatial crowding of nuclear factors, long-range communications among regulatory elements are the known topological determinants in nucleus that regulates expressivity of the genome and are discussed further.

## 2.2. Non-random chromatin organization

Eukaryotic nucleus is a complex, though highly (self-) organized organelle that contains several architectural and functional sub-compartments regulating essential biological processes like transcription, replication and DNA repair. Chromosomes, polymers of several gigadaltons, occupy specific territorial space known as chromosomal territories<sup>9</sup> (CTs) (f.4a). CTs conform into fractal organization (f.4a) that is driven by inherent properties of polymers harboring self-similarity and further stabilized by crowding (discussed later) forces asserted by chromatin interacting proteins. The radial organization of CTs is remarkably associated with gene density of the chromosome<sup>9, 10</sup>. Gene dense chromosomes, irrespective of length, are located towards interior of the nucleus, while gene poor chromosomes are arranged towards nuclear periphery<sup>9, 10</sup> (f.4b). Though, the non-random radial distribution of CTs is evident, precise positioning may not be seen all the cells in a population. This hints that there may be some probabilistic code for radial positional of CTs. In fact, probabilistic model based on gene density strongly suggests the existence of a probabilistic global positioning code, whereas deterministic models fail to show the same effect<sup>11</sup>. Moreover, territories themselves have preferred geometry determined by gene density and centromere/telomere orientations (f.4c). Again, the gene rich domains preferably locate to the periphery, while gene poor regions reside inside the territory<sup>12</sup>, however the claims are controversial<sup>10</sup>. Taken together, the chromosomal domains in mammalian nucleus seem to adapt gene density based organization besides their self organizing polymeric property.

Further, in the absence of transcriptional activity, chromatin loci are generally remained confined in small

intra-nuclear volumes for hours<sup>13</sup>. Although chromatin has random *Brownian* motion within a short range in its local environment, large scale random movements are nearly absent<sup>13</sup>. The mobility is further restricted for the chromatin loci located towards the nuclear periphery than the interior ones<sup>13</sup>. Attachment with nuclear membrane, accumulation of heterochromatin and spatial constraints are the proposed factors restricting the mobility of chromatin in peripheral regions<sup>13</sup>. Moreover, distinct nuclear structures like nuclear speckles, nucleolus etc (discussed later) can also restrict the large-scale random chromatin movements by trapping the chromatin domains.



**Fig. 4.** Nonrandom chromatin organization. **(a)** Cartoon presentation of chromosomal territories (CTs) in the nucleus. CTs conform into fractal organization of ~1Mb domains. Chromatin fibers from neighboring domains can loop out to intermingle or to access factors in nuclear bodies like nuclear-speckles. **(b)** Non-random radial distributions of gene-dense (red) and gene-poor (white) chromosomal regions. Distribution plot depicts distance of gene-rich and gene-poor regions from the center of the nucleus **(c)** Centro(grey)- and telo(red/blue)-mere dependent topology of chromosomal territory.

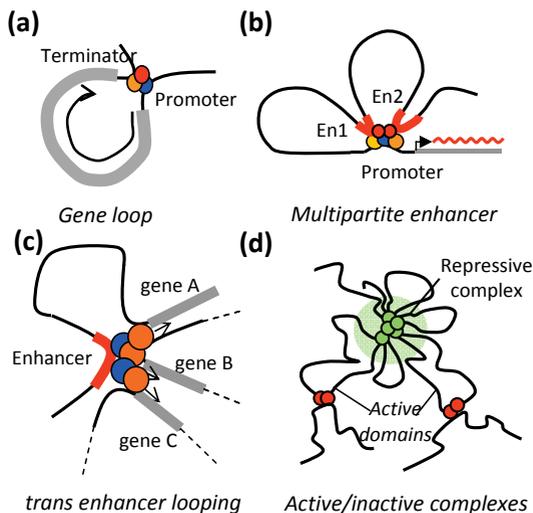
Spatial location of the gene in the nucleus often matters to its expression<sup>14</sup>. It has been observed that dragging a gene from nuclear interior to the periphery, by an inducible system, silences the gene<sup>15</sup>. *Lamin*, a nuclear envelop protein, strikingly associates with gene poor regions and form a repressive chromatin environment. Moreover, nuclear pore complexes have also been shown to associate with gene activity in collaboration with chromatin modifications<sup>16, 17</sup>.

Thus, the chromatin organization in the nucleus is far from being random and could associate with genomic functions.

### 2.3. Chromatin Interaction Networks (CINs)

Besides the global nonrandom arrangement of chromatin, there are numerous local regulatory topologies within CTs or in the inter-territorial space. To begin with, the genes themselves can make self-loops to bring promoter and terminator together to regulate its rapid transcription in a cyclic manner and to keep transcriptional memory over short periods of repression and re-induction<sup>18</sup> (f.5a). This has been shown for several genes in yeast and rRNA genes in mammals<sup>19-21</sup>. Remarkably, self-looping is also observed in relatively shorter genes (~1 Kb) raising concerns over feasibility of chromatin looping<sup>21</sup>. Theoretically, two chromatin loci need to be around 10kbp apart to form an optimum chromatin loop<sup>22</sup>. Since naked DNA is more flexible and can loop at 500bp distance<sup>22</sup>, nucleosome dynamics at transcribing locus could actually relax the chromatin fiber<sup>23</sup>. Interactions at short distances could also arise due to DNA condensation, supercoiling, higher affinity protein-protein interactions or additional force applied by strong DNA bending proteins. All these aspects together could reason for such experimental observations. Second, the enhancers located several kb away from the promoter has a mechanism to communicate with promoter. One of the characteristics of enhancer action is that it retains the memory of its initial contact with *RNA pol2* or transcription factors that determines its recurrent action on the transcriptional activity. Theoretical random jump model also confirms this observation<sup>24</sup>. Logically, this memory could be hypothesized in the form of a higher order topological structure formed by enhancer-promoter communication (f.5b). Such structures are, in turn, determined by the probability of site juxtaposition. The model of enhancer action by long-range looping, that could form structural memory for the recurrent transcription, has been long-hypothesized. However, until recently there was lack of direct evidence of such phenomenon due to technical limitations. With the emergence of 3C (Chromosome Conformation Capture)<sup>25</sup>, it is now feasible to screen such long range interactions. Using ChIP and 3C, it has been demonstrated that long range enhancer-promoter communications regulate the timing of transition between poised and active state of genes<sup>26</sup>. Elegant examples are worked out that show multipartite nature of enhancers (f.5b). *TNF* gene is shown to have two distal enhancers interacting with the proximal promoter in the activated T-cells<sup>27</sup>. Multiple estrogen response elements (EREs) are found to interact together at the promoter of *GREB1* gene<sup>28</sup>. Interestingly, Estrogen response has also been shown to induce *Actin/Myosin* directed inter-chromosomal long range interactions leading to activation of genes<sup>29</sup>. *GREB1* gene, previously demonstrated to harbor *cis*-interactions with several ERE elements, is also a part of this *trans*- interaction network. Thus, the same environmental signal induces several kind of *cis*- and *trans*- interactions, that makes the scenario highly

complicated. More recently, complete CIN of Estrogen response elements has been mapped by a naive *ChIA-PET* technique, which has uncovered a high proportion of complex *cis*-interactions as compared to duplex interaction strengthening bipartite nature of ERE enhancers<sup>30</sup>. *Trans* interactions, however, were rather few in their dataset, indicating that ERE elements primarily function in *cis*<sup>30</sup>. The authors clearly show that ERE-CIN associates with the expression of the estrogen responsive genes<sup>30</sup>.



**Fig. 5.** Distinct higher order topologies in gene regulation. (a) Gene-loops regulating recurrent transcription in cyclic manner. (b) Multiple enhancers looping onto single gene-promoter. (c) Long range *cis*- and *trans*-looping of single enhancer to multiple promoters. (d) Large repressive complex having active domains looped out.

The *trans*-communication among loci on different CTs can occur by distinct modes. Extensive intermingling has been suggested on the territory edge of neighboring CTs<sup>31</sup> (f.4a). Interestingly, particular locus can also loop out from its CT to a substantial distant to regulate interaction in *trans* (f.5c). Probably one of the most interesting examples worked out so far is *trans*-interaction of *H-enhancer* of olfactory receptor promoter looping to promoters of other olfactory receptors from distinct chromosomes<sup>32</sup> and ensuring the stochastic selection and expression of a single receptor at a time. As compared to Estrogen regulated interactions, the interactions here seem to have random and mutually exclusive choice of an interacting partner at a time to the bait. A recently uncovered CIN, including *cis/trans* interactions of hundreds of genes mediated by Klf1 transcription factor, is shown to associate with their transcriptional co-regulation<sup>33</sup>.

---

Moreover, a CIN of 13 subunits of Cytochrome-*c*-oxidase from 13 different chromosomes, assuring their co-expression<sup>34</sup>, further consolidates the existence of widespread CINs in context of transcriptional co-regulation.

The networks of long range *cis*- and *trans*- interactions can also form large repressive or poised complexes with a potential to be activated at certain time (f.5d). Polycomb Response Elements (PRE) collide with each other to form a CIN of repressive Bithorax complex<sup>35</sup>. Interestingly, the active domains loop out from this repressive complex upon activation<sup>35</sup>.

These evidences together represent a conserved mechanism for activation as well as repression of genes by higher order topological structures or CINs. In most of the cases studied so far, the regulation of such structures is clearly shown to be accompanied by the gene function. Thus, the 3-dimensional higher order chromatin structure is associated functional fate of the gene(s).

#### 2.4. Nuclear factors associated with CINs

The factors involved in chromatin interactions can be diverse. RNA polymerase, general transcription factors, activators, repressors themselves can derive such interactions. However, there are few specific factors that, supposedly, are specialized to orchestrate higher order structures in the nucleus. Among these CTCF, Cohesin and SATB1 are being considered as major architectural proteins in the nucleus<sup>36, 37</sup>. CTCF, a multifunctional protein and well known enhancer blocker and boundary element, is shown to regulate some of its, if not all, functions via long range interactions<sup>38</sup>. Complex chromatin conformation at *H19/Igf2* maternal locus<sup>39</sup>, inducible looping at *HLA-DQAI* (*MHC-II* genes) region<sup>40</sup> and homologous pairing during X-inactivation are shown to be mediated by CTCF<sup>41</sup>. Since CTCF is ubiquitously expressed, its DNA binding is regulated by mechanisms other than its differential expression, for example by binding to methylation free domains<sup>42, 43</sup>, diverse post-translational modifications<sup>44-47</sup>, transcription of non-coding RNA<sup>48</sup>, histone modifications<sup>49, 50</sup>, nucleolar localization<sup>51</sup>, chromatin remodeling<sup>52</sup>, interaction with other factors like *YY1*<sup>53</sup>, RNApol2<sup>54</sup>, Cohesin<sup>55</sup>, nuclear lamina<sup>56</sup> etc. Around 15,000 CTCF binding sites have been identified in mammalian genome that follows gene density quite strikingly<sup>57</sup> and since gene dense regions are shown to be physically interacting<sup>58</sup>, there is a possibility that CTCF plays, though not necessarily, a global role in long range interactions genome wide. Furthermore, cohesins are shown to associate with CTCF binding genome wide uncovering non-canonical role of cohesins in transcriptional regulation<sup>55, 59, 60</sup>. Moreover, cohesins mediates higher order chromatin organization by physically linking CTCF to chromatin<sup>36, 61, 62</sup>. However, it remained to be seen if cohesins participate in CTCF mediated CINs genome-wide. We explored some of these aspects in paper III in the thesis. SATB1, on the other

hand, has been demonstrated to form a cage like structures tethering chromatin loops around<sup>37</sup> (discussed later).

Non-coding RNAs are being considered as an integral component of chromatin structure<sup>63</sup> silencing several genes in *cis* and *trans*<sup>64-69</sup>. Some reports support their possible role in organizing higher order chromatin conformation<sup>66, 70-72</sup>, while others do not<sup>35, 73</sup>. Some of the unpublished data from our lab suggests the disruption of chromatin interaction complexes when treated the cross-linked chromatin with RNase. However, it remains an interesting question to be answered in next few years.

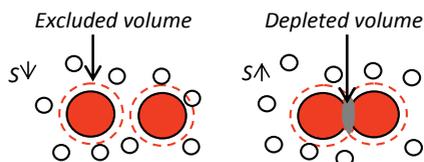
Moreover, the functional assemblies in the nucleus themselves could mediate interactions among distant chromatin loci. These are discussed in next section.

#### 2.5. Role of crowding forces

The regulatory factors primarily occupy the space in inter-chromatin compartments either within territories or inter-territorial space. These regulatory factors are generally crowded in discrete foci by a process known as molecular crowding. Theoretical studies uncover that the phenomenon of *depletion attraction* (f.6) can lower the free energy of the system, which would otherwise be higher because of lower entropy of molecules in highly condensed environment, by volume exclusion upon macro-molecular interaction and serves as main driving force behind molecular crowding in the nucleus<sup>74</sup>. The molecular crowding increases- 1.) the effective concentration (thermodynamic activity measured per unit of available volume) of macromolecules and 2.) the association constants of relatively slower macro-molecular reactions *in-vivo*, while dissociation constants remain unaffected<sup>75-78</sup>. One of the best studied examples of spatial crowding is nucleolus. There are 1-4 nucleoli in each mammalian nucleus. These are the hubs of several regulatory factors involved in ribosome biogenesis, mRNA metabolism, cell cycle regulation, signal recognition particles biogenesis, protein folding and primary miRNA processing<sup>79</sup>. The nucleolar crowding is dependent on RNApol1 transcription. Inhibiting RNApol1 leads to dis-integration of nucleoli<sup>80</sup>. Dynamics of nucleoli also changes during cell-cycle and there occurs an organized disassembly onset of M-phase<sup>81</sup>.

The expressive and repressive states of genome itself could maintain their respective states by a positive feed-back mechanism between fractal organization of chromatin and crowding forces asserted by chromatin interacting proteins<sup>82</sup>. Euchromatin shows higher fractal dimension than heterochromatin, thus ascribing wider search space for transcriptional machinery<sup>82</sup>. Heterochromatin, on the other hand, makes a compact foci due to relatively lower fractal dimension, which is further maintained by crowding of heterochromatin associated proteins<sup>82</sup>. Further, the pattern of heterochromatin foci might differ in distinct cell types partly due to relative expressivity of genome, distinct

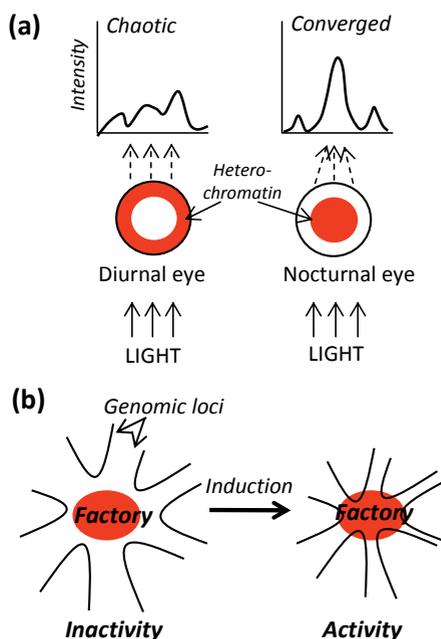
nuclear packing and thermodynamic forces. Interestingly, rod cells in nocturnal animals show peculiar pattern of heterochromatin, which could function as convex lens and focus the scattered light efficiently onto retina<sup>83</sup> (f.7a). This hints at evolutionary forces that could modulate nuclear architecture and ascribe an additional constructive dimension to it, i.e., moon-lighting of heterochromatin besides its regular function of repressing genes constitutively.



**Fig. 6.** Molecular crowding in nucleus. Entropically favorable forces like *depletion attraction* could bring spatial crowding of macromolecules in the nucleus. S stands for entropy.

Another appealing example of spatial crowding is the spatially concentrated foci of active RNAPol2, also termed as '*transcription factories*' (f.7b). These are considered as potential sites of ongoing transcription in the nucleus<sup>84</sup>. The idea logically emerged from the fact that number of nascent transcription sites in the mammalian nuclei is far less than the total RNAPol2 molecules. Therefore, ~75000 nascent transcripts are concentrated in ~2400 sites of ~80nm that are extranucleolar<sup>85</sup>. It is proposed that transcription factories are pre-determined structures where distal gene loci move for transcription and it is not a result of stochastic crowding of DNA bound RNAPol2, though it is to be seen whether the movement of genes is directional or diffusive<sup>84</sup>. Inhibition of transcriptional initiation by heat shock induces partial dissociation of distal gene loci from transcription factories, while blocking transcriptional elongation does not affect the gene localization to the transcription factory<sup>86</sup>. It seems that the chromatin loop organization is not strictly dependent on transcription factory and these structures could be formed prior to entering the transcription foci. This is also supported by another report where by blocking transcriptional initiation and elongation by  $\alpha$ -amanitin has no effect on long range interactions<sup>73</sup>. Distant interactions as well as active/inactive chromatin marks are maintained at  *$\beta$ -globin* locus after inhibiting RNAPol2<sup>73</sup>. It is suggested that chromatin marks in collaboration with other factors are responsible for long range interactions, if not active RNAPol2. Recently, using an artificial system, Xu et al showed that active genes with identical promoters co-localize in a same specialized factory<sup>87</sup>. Specialization of transcription factories hints towards some specificity primarily determined by gene regulatory structure and the factors binding to it. Molecular crowding is also reported

for several well known transcription factors. Surprisingly, the transcription factor foci do not overlap with transcription factories<sup>88</sup>. It is hypothesized that these sites represent incomplete regulatory complexes that are devoid of RNAPol2 or alternatively these could function as storage sites of specific factor, which could be fetched when needed. Such foci are seen to be positionally stable for shorter time periods (in minutes), but are comparatively mobile over longer periods. FRAP (Fluorescence recovery after photo-bleaching) analyses reveal that there exists dynamic movements of factors inside-out these foci strengthening the hypothesis that these sites could regulate instant accessibility of the factor<sup>88</sup>.



**Fig. 7. (a)** Heterochromatin pattern (red) in rod cells of diurnal and nocturnal animals. Inverse pattern in nocturnal eye facilitate the convergence of light as a convex lens. **(b)** The concept of '*molecular factories*' in the nucleus. Upon induction, several genomic loci can move to the factory to access the regulatory factors.

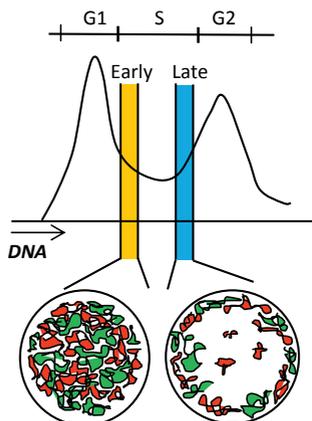
Moreover, such nuclear assemblies are also reported for the factors prominently involved in architecting higher order chromatin structures. SATB1 has been demonstrated to form a cage like structure in thymocyte nucleus tethering several chromatin domains to itself and modulating chromatin marks in a lineage specific manner<sup>37</sup>. Importantly, in breast cancer SATB1

reorganizes the chromatin structure and changes the expression of hundreds of genes that eventually leads to metastasis<sup>89</sup>. However, the contribution of crowding forces in such large scale reorganizations remains to be explored.

Thus, the crowding forces in the nucleus serve regulatory purpose, particularly to compartmentalize distinctly the transcriptional active/inactive domains and by regulating effective concentrations & association constants.

## 2.6. Mosaic networks of genomic functions

Long range interactions and nuclear sub-compartments are not limited to transcription, they also extend to other regulatory functions like splicing, replication and DNA repair. Nuclear speckles are the storage sites of splicing factors. Evidences hint that nuclear speckles are dynamic nuclear hubs for chromosomal interactions for transcriptional regulation<sup>29</sup>. Inhibiting the active transcription rapidly erases these nuclear factories<sup>90</sup>. Splicing, in general, occurs outside nuclear speckles, though introducing exogenous pre-mRNAs are shown to assemble at these sites<sup>91</sup>. Interestingly, deleting the crucial intron or exon that are known to recruit splicing factors, leads to de-localization of pre-mRNA from nuclear speckle sites suggesting that intron-containing active genes assemble at these sites for instant access to splicing machinery<sup>91, 92</sup>. Moreover, bulk of chimeric transcripts in EST databases and evidences of physical proximity among translocation prone loci in cancers hints at inducible CINs of gene fusions and *trans* splicing<sup>93-95</sup>.



**Fig. 8.** Segregation and sub-nuclear localization of transcription and replication domains during S-phase. Transcription foci are shown in red, while replication foci are in green.

Live cell imaging in yeast has revealed the existence of replication factories, the sites of bulk DNA synthesis<sup>96</sup>. It is argued that the replication factories are the

consequence of DNA replication itself, strengthening function to sub-nuclear structure paradigm. Further, there is an association between timing of replication and gene expression. Active genes generally replicate early in the S-phase, whereas inactive genes replicate late<sup>97</sup>. It is hypothesized that early replication can specify permissive chromatin states for transcription, while late replication could determine heterochromatinization of silent domains. However, the compartmentalization of replicating and transcribing domains is almost mutually exclusive and they change their nuclear occupancies during their progressive activation/inactivation in cell-cycle without overlapping with each other<sup>98</sup> (f.8). This suggests a highly dynamic mosaic compartmentalization of essential genomic functions in mammalian nuclei. Though this model is conflicting to a previous observation<sup>99</sup>. Moreover, it has been postulated that the timing of replication is also determined by spatial nuclear position of the domain<sup>100</sup>. The origins are marked for early or late replication timing in the late telophase or early G1 phase, a phase when chromosomes undergo global rearrangements after mitosis<sup>101</sup>. The foci of early replicating domains are randomly distributed across nucleus, while late replicating foci are generally located towards nuclear periphery in the early G1 phase, however this association is disrupted in late G1 or S-phase suggesting that the origins are probably licensed for early or late replication in the early G1 phase probably by orchestrating chromosomal loci.

DNA repair centers are inducible nuclear sub-organelles where damaged DNA molecules assemble and followed up by sequential recruitment of DNA repair proteins<sup>102</sup>. The fact is based on the observation that the number of DNA repair centers are far less than number of DNA damage sites in the nucleus. Sometimes the DNA repair sites co-localize with transcription/nuclear speckle sites, particularly in the case of Transcription Coupled DNA repair (TCR), wherein the stalled RNA polymerase on damaged gene induces the progressive recruitment of DNA repair factors. Similarly, replication machinery is also shown to be targeted to DNA repair foci<sup>103</sup>.

Since transcription, replication and chromatin structure are also shown to be the determinants of the mammalian genome architecture linearly<sup>104</sup>, it can be concluded safely that various nuclear functions together drive the organization of mammalian genome compositionally as well as spatially.

## 3. Epigenetics and mammalian development

### 3.1. Genomic Imprinting

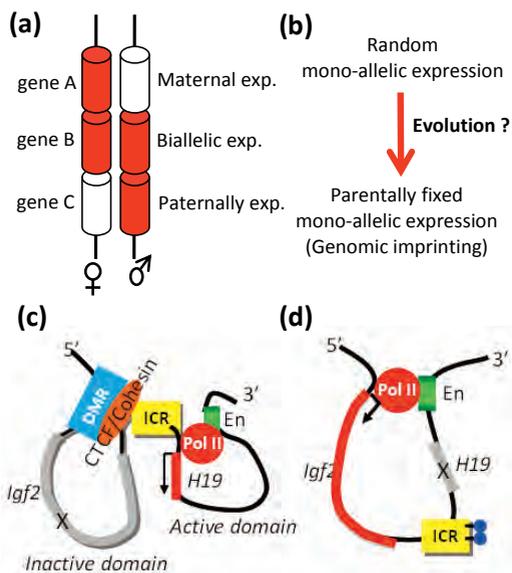
The parental genomes in diploid cells are, in general, epigenetically equivalent. However, wide-spread mono-allelic expression (f.9a) suggests partial epigenetic dimorphism in the genome<sup>105</sup>. Parentally fixed mono-allelic expression, i.e., when the mono-allelic expression

is inherited in parent of origin specific manner, is known as genomic imprinting. The genomic imprint is found in insects, flowering plants and placental mammals and widely considered to be originated by parental conflict. *Parental conflict* theory<sup>106</sup> states that the inequality in parental genomes due to imprinting evolved due to contrasting interest of each parental genome with respect to fitness of their genes. The paternal genome tends to maximize the growth resources for the offspring in order to outcompete other paternal genomes, while maternal genome tries to save resources for itself<sup>106</sup>. Accordingly many of the paternally expressed genes are growth factors and many of the maternally expressed genes are growth limiting. There are other theories of imprinting also, though not as widely accepted as *parental conflict* theory. Some authors suggest that the imprinting is the result of defense against foreign DNA like viral genome<sup>107</sup>. The theory is based on the fact that DNA methylation that is seen as defense mechanism against foreign DNA in prokaryotes as well as eukaryotes is prominently involved in imprinting too. Another theory namely *intra-locus sexual conflict*<sup>108</sup> is based on the assumption that males are more likely to transmit high male fitness, whereas female would transmit female fitness.

As a result, evolution will direct paternal genome to repress the maternally inherited genes and vice-versa. Genomic imprinting, here, is suggested as a mean to moderate the severe outcome of such sexual conflict. A more recent theory states that imprinting with maternal expression may be evolved when natural selection favors co-adaptation of maternal and offspring traits<sup>109</sup>. Ohlsson and others suggested a more straight forward hypothesis that imprinting could have evolved by gradual parental fixation of random mono-allelic expression of the genes that have lethal effect if silenced on both alleles<sup>110</sup> (f.9b).

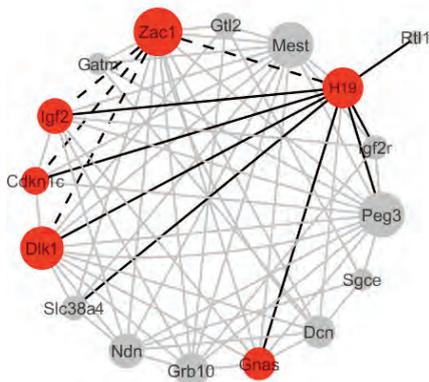
Most imprinted genes are located in clusters on chromosomes and often regulated by a single imprinting control region (ICR)<sup>111</sup>. Differential epigenetic modifications on ICR regulate allele-specific silencing of imprinted genes in the cluster<sup>111</sup>. In general, sequence features<sup>112, 113</sup>, differential methylation and allelic representation of H3K4 and H3K9 trimethylation marks<sup>114</sup> mark their identification. Moreover, few ICRs transcribe long noncoding RNAs that in collaboration with other chromatin marks set up cluster-wide imprinting program<sup>64-66, 111</sup>. Interestingly, long range interactions are also implicated in genomic imprinting suggesting it as widespread gene-regulation phenomenon. Complex loop formation mediated by CTCF/cohesin binding at *H19* imprinting control region (ICR) keeps maternal *Igf2* allele in a repressive domain by restricting access to enhancers, i.e. insulation<sup>39, 61, 115-117</sup>. The loop could serve as epigenetic memory since it is maintained through mitosis<sup>115</sup>. On paternal site, the methylation of ICR inhibits *CTCF* binding and *Igf2* accesses the enhancer<sup>39</sup>. More details are illustrated in figure-9c,d. ICR in this particular locus is of special interest as error in this region has pleiotropic consequences. It causes parthenogenesis, predisposes animal for cancer, hyper-sensitizes *Igf2* post-receptor signaling in mice<sup>118-120</sup>, abrogates several *cis* interactions in cancer cells<sup>121</sup>. Moreover, *H19*-ICR is the ancient most ICR known so far suggesting its functional criticality<sup>122</sup>.

Perturbations in regular genomic imprinting are strongly associated with cancer and several complex diseases like Angelman syndrome, Prader-Willi Syndrome, Beckwith-Wiedmann Syndrome. Dysregulation of regular imprinting pattern in germ-line development is one of the main causes for these syndromes<sup>123</sup>. Moreover, loss of *Igf2* imprinting is associated with several cancers<sup>119, 124</sup>. Epigenetic haploidy of imprinted genes is considered to basic cause of their susceptibility. However, this may not be the sole reasoning. Evidences suggest that imprinted genes function pleiotropically, i.e., error at one imprinted locus can affect several other imprinted loci as observed in the case of *Zac1* and *H19*<sup>125-127</sup>. Knocking down *Zac1*, a development regulatory protein and *H19* a precursor miRNA alters the expression of other imprinted genes<sup>125-127</sup>. Moreover, *H19* is proposed to be a tumor suppressor



**Fig. 9.** (a) Allele-specific expression. (b) A probable model for the evolution of genomic imprinting from random mono-allelic expression. (c, d) Higher order chromatin structure of *Igf2/H19* locus on maternal chromosome (c) and paternal chromosome (d). CTCF and Cohesins mediate chromatin looping and restrict *Igf2* in a repressive domain on maternal chromosome

gene<sup>128</sup>. Imprinted genes are also shown to be co-regulated during development<sup>125</sup>. Thus, an imprinted gene network (IGN) has been proposed (f.10), which might regulate whole imprinted gene-function in a pleiotropic manner. Interestingly, multiple imprinted loci show hypomethylation of maternally methylated ICRs in BW syndrome suggesting role of certain *trans* acting factors in maintaining imprinting in somatic cells at maternally methylated ICRs, which when malfunction could induce mosaic imprinting error.<sup>129</sup>



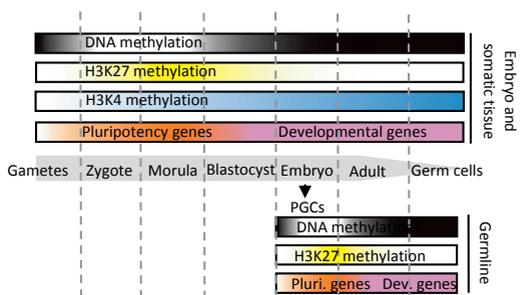
**Fig. 11.** Imprinted Gene Network (IGN). Size of the node is scaled to corresponding degree of interaction. Red nodes represent the network uncovered by knock-out /transgenic experiments. Black solid-lines represent the network of *H19* ncRNA, while dashed-lines represent *Zac1* network. Rest of the network (grey) is uncovered by correlation meta-analysis of microarray data.

### 3.2. Epigenetic reprogramming

Precursor cells, starting from a pluripotent state, gradually differentiate into diverse cell types during development. Flexible gene expression scheme in pluripotent cells is precisely reprogrammed into more defined and fixed cell-type specific scheme<sup>130</sup>. Such large scale reprogramming of gene expression is, in turn, carried out by extensive acquisition and erasure of various epigenetic marks<sup>130</sup> (f.12). Histone modifications provide a flexible mean for such reprogramming events, while DNA methylation ascribes comparatively more stable change. During differentiation pluripotency related genes get silenced by histone modifications and DNA methylation<sup>130</sup>. A majority of the development related genes harbor H3K4me3, an activation mark, as well as H3K27me3, a repressive mark in pluripotent cells ascribing a potential to be activated or repressed during differentiation to a particular cell type<sup>131</sup>. Such multivalent marks have also been suggested to enhance the affinity and specificity of chromatin remodeling assemblies to the modified histones<sup>132</sup>. This phenomenon is attributed to both the thermodynamic and kinetic reasons. Thermodynamically,

there is effective reduction in the entropy due to restriction in rotational and translational freedom of binding. Reaction kinetics explains the enhancements in reaction rates by increasing effective concentration, not by increasing rate constants<sup>132</sup>. Thus, multi-valency brings flexibility as well as thermodynamic ease in the system.

During germ-line development, the DNA methylation and repressive histone modifications (such as H3K9 methylation) are erased in PGCs (Primordial Germ Cells) and pluripotency related genes are re-expressed<sup>130</sup> (f.12). This erasure ensures equivalent epigenotype in germ-line cells of male and female embryos. This is the only stage in mammalian development when both the parental genomes have almost the same epigenotype. Later on in the gametogenesis the DNA gets methylated and development related genes are re-expressed<sup>130</sup>.



**Fig. 12.** Epigenetic reprogramming during mammalian development. Shown are the gradients of various epigenetic modifications in somatic and germ cells with respect to developmental stage.

The epigenetic reprogramming of imprinted genes differ from rest of the genome<sup>130</sup>. Parental epigenetic marks are re-established in gametes and retained persistent during the embryonic development, unlike non-imprinted genes that undergo another reprogramming event<sup>130</sup>. Epigenetic reprogramming at *Igf2/H19* locus has been studied in detail<sup>133</sup>. The *H19*-ICR and *Igf2* DMRs are methylated in sperms and unmethylated in oocytes<sup>133</sup>. While *H19*-ICR retains the germline imprints further after fertilization, paternal *Igf2* DMRs get demethylated soon after fertilization and *Igf2* shows bi-allelic expression in blastocyst stage<sup>133</sup>. At post implantation stage, imprints are re-established and *Igf2* again shows imprinted expression. In male germ-line, *H19*-ICR paternalizes, i.e acquires imprints on both alleles, during spermatogonia→spermatocyte transition, which parallels with *H19*-ICR's association with BORIS, a CTCF paralogue specifically expressed in spermatocytes (paper II). Moreover, differential timing of imprint acquisition is observed for paternal and maternal ICRs. While paternal allele is methylated slightly earlier, ~50% of maternal alleles remain hypomethylated until the completion of

meiosis<sup>134</sup>. However, recently it is shown that *H19*-ICR could act autonomously and its function is not dependent on germ-line methylation suggesting that germ-line methylation may not be a primary imprint<sup>135</sup>. Thus, methylation could serve as stabilizer of some other primary epigenetic mark say histone modifications or replication timing (?)

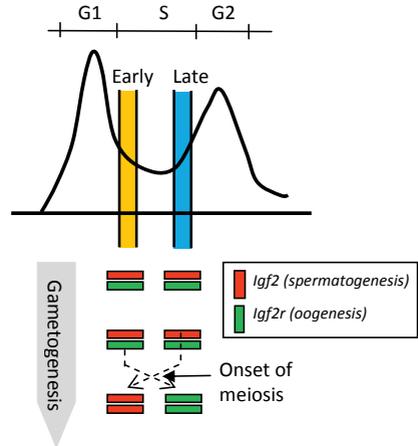
In yet another event, paternal x-chromosome in female 2-cell stage embryo undergoes global inactivation by repressive chromatin environment driven by noncoding RNA *Xist*<sup>130</sup>. However, later at blastocyst stage, x-inactivation is first reversed and then randomly re-inactivated<sup>130</sup>. This re-inactivation, though random, is preserved irreversible in the life of the cell. During germ-line development x-inactivation is also erased to generate epigenetically equivalent gametes<sup>130</sup>.

### 3.3. Reprogramming of replication timing

A genome wide study of replication timing in mouse reveals extensive alteration in replication timing patterns during ES cell differentiation<sup>136</sup>. The changes strikingly associate with gene density, transcription and sub-nuclear localization. Unique patterns of reprogramming suggest that replication timing is a distinct epigenetic signature that associates with other genetic and epigenetic marks<sup>136</sup>. More recently, Gondor and Ohlsson represented an interesting hypothesis that epigenetic reprogramming factors may determine the timing of replication and establishment/inheritance of epigenetic marks may require progression through S-phase<sup>137</sup>. Their model proposes the coordination of replication timing through unique spatial clustering of origins before replication, which in turn may explain how epigenetic reprogramming events propagate through large domains during lineage specification<sup>137</sup>. Though, this promising model has implications in understanding development, evolution and diseases, detailed mechanistic understanding of this epigenetic mark (?) remains to be solved.

Asynchronous replication is considered as hallmark of imprinted genes<sup>138</sup>. In most cases, the paternal allele replicates early, while maternal late regardless of their imprinting status<sup>138</sup>. Asynchronous replication and sub-nuclear localization of *Igf2/H19* imprinted domain is shown to be under the control of *H19*-ICR, though the report shows opposite pattern of replication timing of *Igf2/H19* parental alleles, a conflicting observation to most others<sup>139</sup>. Asynchronous replication timing of imprinted domains is established in gametes and maintained further on in the development<sup>140</sup>. This asynchrony is erased in the germ cells just before onset of meiosis by placing parental marks on the maternal/paternal copies in a gamete-specific manner. For example, late replicating allele of *Igf2* (paternally expressed) shifts to early replication just before meiosis during spermatogenesis. Similarly, early replicating allele of *Igf2r* (maternally expressed) shifts to late replication

during oogenesis<sup>140</sup> (f.13). It is proposed that *trans*-acting factors specifically activated in S-phase recognizes certain *cis*-elements at the time of replication and once the timing pattern is set, it can be inherited to daughter cells by regulated appearance of *trans*- factors<sup>140</sup>. Since, allele-specific methylation is a one of characteristics of imprinted genes, it can provide a mechanism to set up the asynchronous replication timing, however evidences do not favor this. Mono-allelic expression of some imprinted genes can retain in the absence of DNA methyltransferases<sup>141</sup>. Moreover, it has been shown that asynchronous replication timing of imprinted genes is independent of DNA methylation<sup>139, 142</sup>. Differential conformational environment of the parental alleles might contribute in setting up such epigenetic marks. However, reprogramming of CINs and their association with other epigenetic marks remains unaddressed. Paper II addresses the possible link of CINs in regulating reprogramming of replication timing during germ-line development.



**Fig. 13.** Replication reprogramming of imprinted genes in germ line development. The late replicating allele of *Igf2* replicates early just before the onset of meiosis during spermatogenesis, while early replicating allele of *Igf2r* locus shifts to late replication in oogenesis. Though, this reprogramming happens at different developmental stages in males (8 day old mice) and females (13,5 dpc embryo), it remains to be associated with on the onset of meiosis.

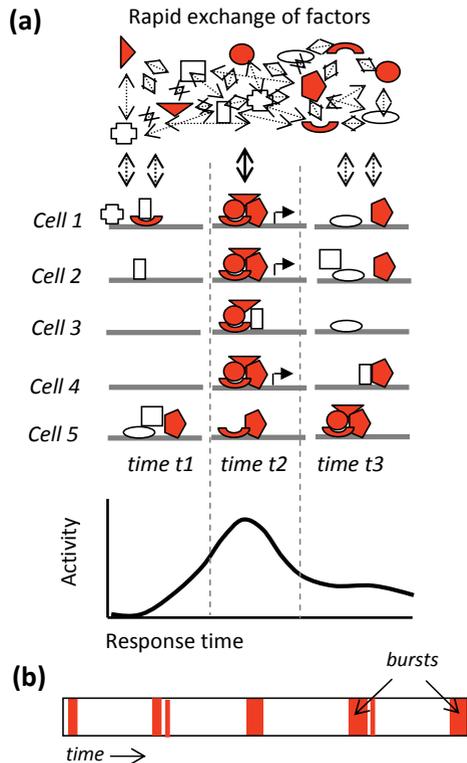
### 4. Epigenetics and Protein Interaction Networks (PINs)

Mammalian nucleus is densely crowded environment and macromolecules keep on striking each other as per their diffusion rates in natural physiological conditions. Their binding to relatively static structures like chromatin, however, limits their mobility<sup>143</sup>. Most nuclear proteins also exchange between chromatin-bound and free states quite rapidly resulting in a highly dynamic interplay among nuclear factors and chromatin<sup>144-146</sup>. Crowding forces increases the association constants of

macromolecular interactions that bring assembly of multi-protein complexes targeting essential genomic functions. However, how random collisions would determine the assembly of correct complex and how does complex recognize the correct target on chromatin? Live cell imaging together with systems biology has uncovered interesting insights into protein interaction dynamics and their interplay with the chromatin<sup>146</sup>. These studies support a combinatorial probabilistic model wherein macromolecules undergo a rapid association-dissociation dynamics and finally converge to a correct, active complex<sup>146</sup>. The affinity for the correct complex formation and target recognition is determined by low dissociation rate, while specificity is brought by *kinetic proofreading*<sup>147, 148</sup>, a phenomenon wherein the several high energy intermediates are formed by ATP-driven mechanism and the complexes with low dissociation rates are selected thermodynamically. Such probabilistic model suggests that a large proportion of proteins would reside in incomplete and inactive complexes and often interact with chromatin non-specifically with relatively low residence time (~fraction of a second). That, indeed, reasons why only 1% of total RNA polymerases specifically bind to promoters and engage in full length transcription, while ~15% remain transiently bound and ~85% are in free diffusive state<sup>146</sup>. Moreover, the model also suggests that assembly of correct complex would, in fact, vary cell to cell and resulting in transcriptional heterogeneity in a cell population, which is minimum at certain time point<sup>149</sup> (f.14a). The stochasticity in gene expression could further lead to transcriptional *bursts* (f. 14b), i.e. long period of inactivity intervened by short pulses or *bursts* of activity<sup>150, 151</sup>.

One of the main constituents of macromolecular assemblies that bind to chromatin are the chromatin modifying factors. These factors either modify or reconfigure nucleosomes to modulate DNA accessibility for transcription<sup>152-154</sup>. Besides their own biochemically stable complexes, these factors transiently interact with transcriptional activators and repressors<sup>8, 155-158</sup>. Chromatin modifying factors can be sub-grouped in two categories- ATP-dependent<sup>154</sup> chromatin remodelers that use ATPase activity to reconfigure nucleosomes and ATP-independent<sup>159</sup> histone modifiers which include Histone Deacetylase (HDAC), Histone Acetyl Transferase (HAT), Histone Methyl Transferase (HMT), Poly-ADP Ribose Polymerase (PARP), Histone Demethylase, Histone kinase, Histone ubiquitinase, Histone deubiquitinase, and Histone Proline isomerase<sup>159</sup>. Based on structural similarities, ATP dependent remodelers are further grouped into several sub-families like SWI/SNF, ISWI, CHD, INO80, and Rad<sup>160</sup>. Several functional domains such as ATPase, Chromo, Bromo, ARID, SWIRM, SANT and PHD are specific to these proteins<sup>160</sup>. Their indispensability, as revealed by mouse knock-out experiments<sup>161</sup>, suggests lack of redundancy in these

factors questioning then why do a cell needs several ATP-consuming chromatin remodelers just to open up the chromatin. The answer lies in accumulated literature that suggest several additional nuclear roles of chromatin remodeling complexes including splicing<sup>162</sup>, cohesin loading & chromosome segregation<sup>163-167</sup>, higher order chromatin structure<sup>168</sup>, telomere regulation<sup>169, 170</sup>, DNA replication<sup>171-173</sup>, check point control and DNA repair<sup>174-180</sup>.



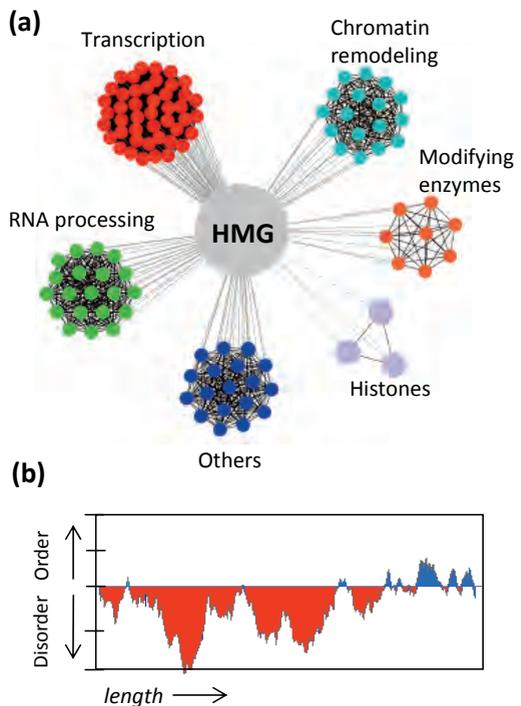
**Fig. 14. (a)** Combinatorial probabilistic model of macromolecular assembly and transcriptional heterogeneity. Red colored factors are the components of correct transcriptional assembly, which shows heterogeneous binding pattern across a cell population. However, as the response time progresses, number of cells having correct assembly increases and further decreases as the assembly dissociates. **(b)** Transcriptional *bursts* in a single cell.

Although advent in high through put technologies does allow mapping of chromatin-protein interactions at genome scale, interaction of large macro-molecular assemblies with chromatin is yet to be explored in detail. A proteomics approach focused on chromatin need to be accelerated in higher eukaryotes. Nevertheless, protein-protein interaction data obtained basically from yeast two hybrid screens and co-immunoprecipitations suggests

diverse interactions of chromatin remodeling proteins in yeast<sup>155</sup>. Yeast PIN has revealed their interactions with other factors related to transcription, DNA repair, DNA recombination, metabolism, cell cycle, ribosome biogenesis, RNA processing, sporulation, transport, cell organization, stress response and protein degradation classes suggesting a high degree of diversity<sup>155</sup>. Further, chromatin remodeling enzymes show programmed dynamics during development<sup>181</sup>. In particular, combinatorial assembly of chromatin remodelers is observed during lineage specification<sup>181</sup>. Differential composition of these assemblies determines the pluripotency or the lineage commitment of cells during differentiation process<sup>181</sup>. Combinatorial assemblies of chromatin remodeling complexes is, in fact, being proposed as a mean to acquire spatiotemporal diversity in gene expression in brain<sup>181</sup>.

What makes chromatin remodelers so versatile? What feature brings their diverse macromolecular interactions and combinatorial assemblies? These questioned are addressed in paper IV in the thesis. The induction for the hypothesis in paper IV comes from an observation that the wide range of interactions of HMG (f.15a), a group of chromatin interacting proteins, is attributed to its structural feature<sup>182, 183</sup>. HMG proteins are architectural transcription factors involved in assembling specific macromolecular complexes onto chromatin. It bends the DNA and that in turn facilitates the binding of other factors<sup>184</sup>. High centrality/hub-ness of HMG is primarily attributed to its intrinsic structural disorder (f.15b) accompanied by post-translational modifications<sup>185</sup>. We introduce intrinsic disorder in the following section.

The protein structure–function paradigm, which states that protein must fold to acquire its function, is based on the assumption that specific molecular recognition and binding in the cell requires precise geometrical orientation that is achieved by rigid three-dimensional structure of the protein. A rather late realization that many proteins, 1/3 of human proteome, remain completely or partially unfolded/disordered in their native functional state has prompted transformation of “protein structure–function” paradigm to “protein trinity” paradigm, which states “The native proteins can exist in any of the three thermodynamic states—ordered, molten globule, and random coil”<sup>186</sup>. Any one of these states, not necessarily the ordered one, can be the native functional state of the protein<sup>186</sup>. Intrinsically disordered proteins are discovered more recently, partly because the biased biochemical and biophysical methods selectively work for globular proteins. Disordered regions do not crystallize into fixed structure and diffract x-rays chaotically. Thus these regions are either excluded from the studies or are shown as missing coordinates, a trait which earlier limited their identification and later used to characterize the same. Disordered regions in protein renders larger search radius for initial molecular recognition<sup>187, 188</sup> and their transition



**Fig. 15.** HMG-PIN and structural disorder. **(a)** Shown are the diverse protein interactions of HMG protein. Proteins belonging to same functional class are interlinked and clustered. **(b)** Intrinsic structural disorder in HMG protein as predicted by Charge-hydrophobicity method.

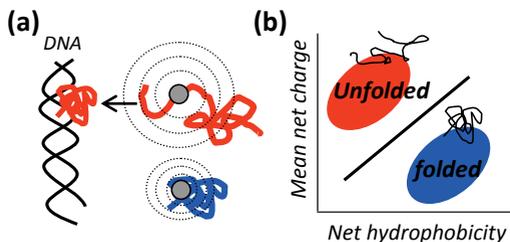
to rigid conformations upon binding with interacting partners decreases the conformational entropy to form a compact, but reversible macromolecular assembly<sup>187</sup> (f.16a). Reversible transitions between disordered and ordered conformations might regulate biological function of the protein either by acting as conformational switch<sup>189</sup> or by stabilizing against protein degradation<sup>190</sup>. Intrinsic disorder also allows multiple interactions by conforming into several flexible conformers and is a property of *date* hubs<sup>191</sup>, characterized by transient interaction with respect to time and space, in PPI as well as transcription regulatory network<sup>192</sup>. HMG hub serves as an excellent example demonstrating how intrinsic disorder could contribute to higher diversity in molecular interactions. Moreover, another prominent hub TP53 that has >300 interacting partners also harbors long disordered regions<sup>193</sup>.

Intrinsically disordered regions are compositionally distinct from ordered regions. Disordered regions are enriched in hydrophilic and charged amino acids. Interestingly, a simple rule based on net charge and

hydropathy could distinguish disordered proteins from ordered quite effectively<sup>194, 195</sup>. Natively disordered proteins cluster in the high net charge-low hydrophobicity half of the plot while other way side the ordered proteins<sup>194, 195</sup> (f.16b). Following is the equation that demarcates best the ordered and disordered proteins.

$$\langle R \rangle = 2.785 \langle H \rangle - 1.151$$

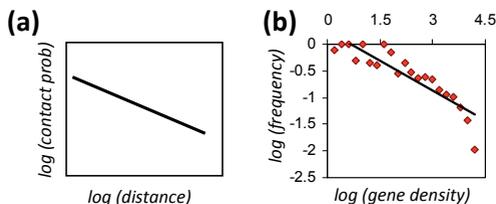
Where,  $\langle R \rangle$  and  $\langle H \rangle$  are the mean net charge and mean net hydrophobicity of the protein. Thus, based on sequence characteristics, several supervised and unsupervised algorithms have been developed for the prediction of local and global structural disorder in proteins<sup>196-202</sup>.



**Fig. 16.** Intrinsic structural disorder. (a) Unfolded conformations renders greater capture radius for macromolecular interactions. Their coupled folding and binding ascribes reversibility in interaction. (b) Charge-hydrophobicity plot (Uversky's plot) discriminates folded and unfolded proteins quite effectively.

## 5. CINs vs PINs : spin together?

Like PINs, CINs also follow power-law, though in slightly different context. The frequency of interactions at closer distance across chromosomal length is quite high and decreases rapidly as the distance between interacting loci increases, thus conforming into fractal organization following a power-law distribution<sup>203</sup> (f.17a). Moreover, chromatin interactions are shown to follow gene density, which itself follows power-law distribution (?) as shown in the following figure-17b. Even, if the interactions do not follow gene-density (paper III), they cluster at preferred locations intervened by large non-interacting regions on chromosomes and thus follow a power law (?)



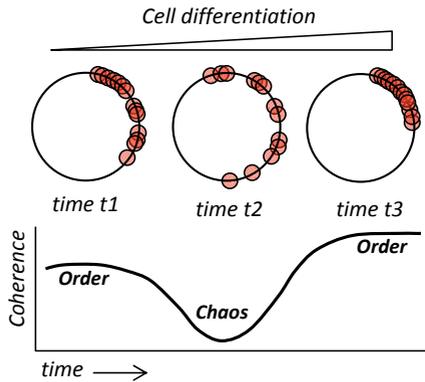
**Fig. 17.** Power law distribution in CINs. (a) log-log plot of contact probability and among genomic loci in *cis* and corresponding genomic distance (b) log-log plot of gene density and its frequency as measured in *mm6* genome assembly of *Mus*.

Another common and most fundamental aspect about CINs and PINs is that physical interactions, besides their random collisions, are accompanied by certain biological function. However, a bigger question is whether CINs and PINs spin together, i. e., could CINs determine their downstream PINs. It is increasingly being uncovered that co-regulated genes show physical proximity among their genomic loci in the nuclear space, most probably to access the transcription machinery specific to these genes<sup>33, 34</sup>. Co-regulated genes, in turn, could associate with physical interactions among their protein-products. Several reports claim nonrandom frequent protein interactions among co-expressed genes in prokaryotes and yeast<sup>204, 205</sup>. In higher eukaryotes, however the correlation is relatively weaker<sup>206</sup>. This could attribute to following reasons. First, there might be higher noise, than anticipated earlier, in gene-expression data which fails to uncover weakly co-expressed genes, second could be the partial lack of correlation between gene and protein expression levels and third might be the conditional protein-protein interaction under certain environment which is regulated by post-translational mechanisms. Nevertheless, a comprehensive analysis on multiple eukaryotic species has revealed that protein-products of genes that are consistently co-expressed in multiple species, indeed, interact frequently<sup>206</sup>. A probable reason to this is to control protein-protein interaction by balancing their concentrations right through transcription level.

In an intriguing report, Rajapakse et al<sup>207</sup> demonstrated that chromosomal arrangements indeed associates with co-expression of corresponding genes during lineage-differentiation. The association is measured in terms of relative entropy. Both the chromosomal association and co-regulated gene-expression show higher entropy (chaos) upon lineage commitment of precursor cells and eventually ends up in highly ordered state upon lineage specification. The association of chromosomal topological arrangement and co-regulated gene-expression was further modeled using Kuramoto's coupled oscillator model, wherein globally coupled oscillators exhibit transition from incoherence to coherence as their coupling strength passes certain threshold. Higher the association between chromosomes and co-regulated genes is, greater the coherence or order and *vice-versa* (f.18). This is probably the first report that generalize the observation of co-regulons' association with chromosomal topologies at genomic scale.

Therefore, by extrapolation, it is possible that stereo-clustering of genes in the nucleus might indirectly regulate protein-protein interactions *via* regulating co-regulons. Since imprinted genes conform CIN (paper 2) and GRN networks as discussed earlier, we attempt to address if these networks eventually regulate imprinted gene-function *via* remotely determining their co-operation at protein level. We drifted away further and pruned our analysis in broader terms of imprinted gene-function in

mammals. Concepts of network science, as discussed in *Methods*, are implemented to uncover systems properties of human imprintome.



**Fig. 18.** Coupled oscillator model of chromosomal association with co-regulated gene-expression. The model estimates how each chromosome (oscillator, red circle) associates (oscillates) with other chromosomes as a function of its share to co-regulated gene-expression, *i.e.* measures the degree of coherence, during cell differentiation. Dispersed circles represent incoherent state where chromosomes are in different phases, while clustered circles depict coherent state where all chromosomes are at closer phases.

## 6. Summary of Methods

The thesis, *in toto*, comprises of several experimental and computational methods that sometimes differ significantly from one paper to the other. To simplify the diversity and comprehensiveness of methods used, we brief them as following:

### 6.1. Analysis of CINs

#### 6.1.1. Experimental system

Most of the experimental work in the thesis, by the author, is performed on mouse cell-lines, namely R1 embryonic stem cells, their derived embryoid bodies (14 days, LIF withdrawal method) and MEF cells from SD7 X BL/6 mice. Some of the work is done on somatic and male germ-line cells on mice crosses generated by Pant et al<sup>208</sup>. these crosses harbor substitution mutation in three out of total four CTCF-binding sites in *H19-ICR* in parent of origin specific manner.

The male germ-line cells, namely spermatogonia, spermatocyte and round spermatids are distinguished based on heterochromatin pattern and nuclear morphology under confocal microscope. In spermatogonia, heterochromatin blobs appear scattered randomly in the nuclear space, while spermatocytes these blobs are enlarged and generally arranged towards nuclear periphery, also the size of nucleus itself is slightly larger

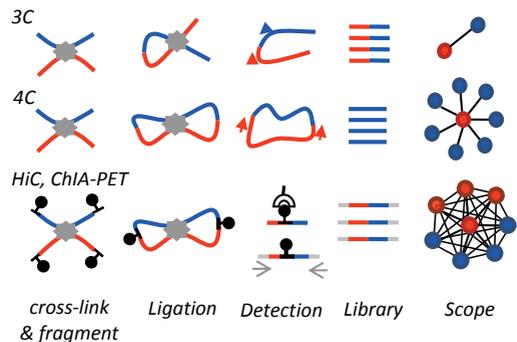
than spermatogonia. The spermatid nuclei are small and have a single big blob of heterochromatin in the center.

#### 6.1.2. Circular Chromosome Conformation Capture (4C)

4C assay<sup>209, 210</sup> is a logical extension of 3C technique<sup>25</sup>. By incorporating circularization and subsequent inverse PCR steps, we can identify all unknown interacting partners of a known genomic bait in an unbiased manner<sup>209</sup>. Detailed strategy is shown in figure-19. In brief, we -

- ✓ Cross-link the chromatin with 1% formaldehyde  
*Low % of formaldehyde is used to avoid over-crosslinking of chromatin that in turn could affect digestion efficiency.*
- ✓ Digest with a 4-cutter restriction enzyme  
*4-cutter, instead of 6-cutter, restriction enzyme ensures the higher resolution of our 4C assay*
- ✓ Ligate the open ends in a diluted condition.  
*Diluted condition prefers intra-molecular over inter-molecular ligation*
- ✓ De-crosslink the chromatin at 60° C.  
*Heat dehydrates formaldehyde to it's gaseous form.*
- ✓ Inverse, nested PCR of circular 4C product  
*Inverse PCR from known bait region amplifies all unknown interacting partners ligated to the bait. Nested primers assure the specificity of PCR.*

Technological advances now allow screening of all-to-all interactions (Hi-C<sup>203</sup> & ChIA-PET<sup>30</sup>). A comparison of 3C, 4C and HiC is sketched in figure-19.



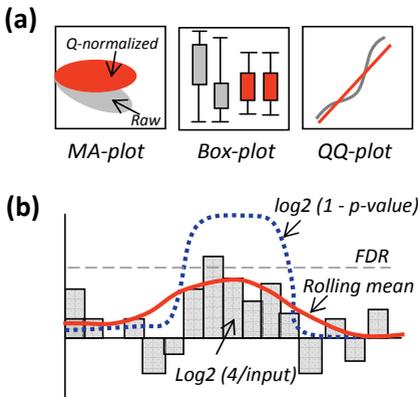
**Fig. 19.** Exploring CINs. Shown are the proximity-ligation based methods to identify chromatin interactions. The methods, basically, differ in their detection methods. 3C identifies one-to-one interaction by selecting primers from two known regions of interest. 4C identifies one-to-all interactions by strategically placing primers on the known region of interest. HiC or ChIA-PET identifies all-to-all interactions by incorporating tagged linkers to the open ends and further capturing the same on the ligated junctions in sheared DNA. Pulled down DNA fragments are subjected to pair end tag (PET) sequencing.

### 6.1.3. Tiling array analysis and data-mining

Amplified 4C product was further hybridized to Nimblegen tiling arrays. The data (4c & input genomic DNA) was quantile normalized across channels and samples. MA-plots, box-plots and quantile-quantile plots were used for quality control (f.20a). Signals were identified by rolling mean method<sup>211</sup> (f.20b). Following  $p$ -value<sup>211</sup> was assigned to each sliding window:

$$P = 1 - \text{ERF}\left(\frac{\bar{\omega}}{\sigma/\sqrt{n}}\right)$$

Where  $\bar{\omega}$  is average  $\log_2(4c/\text{input})$  of sliding window,  $\sigma$  is standard deviation of background distribution (normal distribution of background was confirmed using QQ plot) and  $n$  is the number of probes in the sliding window<sup>211</sup>. Signal map and R-package were used for data visualization. Genomic data from UCSC and published genome-wide studies were used to annotate our 4C library. All computational programming and statistical testing were done on PERL and R platforms.



**Fig. 20.** 4C-array analysis. (a) Tiling array data from Nimblegen is quantile-normalized across channels and samples and quality assured by MA-, box- and QQ-plots. (b) Normalized  $\log_2(4c/\text{input})$  signals were processed using rolling mean method. P-value to each rolling window is assigned and further corrected using Bonferroni method.

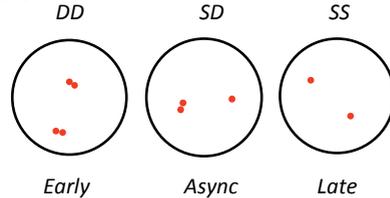
### 6.1.4. In situ analysis of CINs

Interactions identified by 4C were validated and quantified by extensive 3D DNA FISH experiments. FISH also allows us to study dynamics of interactions, to certain extent, with respect to space/time and chromosomal territory. In particular, we applied triple color DNA FISH to study the cross-talk among imprinted domains from 7 different chromosomes.

### 6.1.5. Replication timing analysis

Due to limited availability of germ-line cells, we studied the replication timing pattern by DNA FISH method. S-

phase nuclei are demarcated by staining PCNA. Single dot or singlet of stained DNA of interest suggests unrepliated allele, while doublet signifies replicated allele. Thus, distinct patterns of allelic replication, i.e. singlet-singlet (SS), singlet-doublet (SD) and doublet-doublet (DD), suggest higher probability of alleles replicating late, replicating asynchronously and replicating early in S-phase respectively (f.21). This method marks the end-point of replication and therefore is of higher resolution as compared to BrdU pulse labeling methods.



**Fig. 21.** Analysis of replication timing by DNA FISH. Doublet-doublet signal suggests that both alleles are already replicated, i.e., higher probability to be replicated early in the S-phase. Singlet-doublet signal represents asynchronous replication where one allele is already replicated and other is not. Singlet-singlet signal represents unrepliated alleles, i.e. higher probability of replicating late in the S-phase.

## 6.2. Analysis of PINs

Human PPI data was obtained from Cytoscape<sup>212</sup> datasets. The data is composite repository of protein or domain interaction data from DIP<sup>213, 214</sup>, BIND<sup>196</sup>, IntAct<sup>215</sup>, HPRD<sup>216</sup> databases and several published articles<sup>217-219</sup>. It includes 61,226 interactions in total. Network was constructed and analyzed using data-structures and algorithms available in *igraph* library of R-package. Network visualization was done on Cytoscape<sup>212</sup> platform. *Spring-embedded*<sup>212</sup> lay-out algorithms that push highly connected nodes towards interior and weakly connected nodes towards periphery were used.

### 6.2.1. Network connectivity

Connectivity is defined as ration of number of observed connections to expected connections in a network<sup>220</sup>.

$$\text{conn} = \frac{2E}{V(V-1)}$$

Where  $E = \#$  observed edges,  $V = \#$  of vertices. The quantity  $[V(V-1)]/2$  is total possible number of edges in a graph. The value of  $\text{conn}$  ranges from 0 to 1.

### 6.2.2. Transitivity

Transitivity or clustering coefficient measures the probability that the neighboring vertices of a given vertex are connected. The local transitivity of a vertex  $v$  is ration

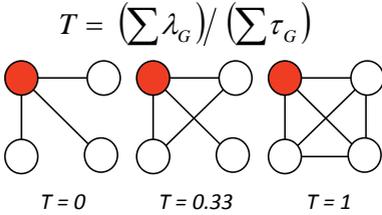
of triangles ( $\lambda_G$ ) to total triplets ( $\tau_G$ ) centered on the vertex in graph ( $G$ )<sup>220</sup> (f.22).

$$T(v) = \lambda_G(v) / \tau_G(v)$$

$T(v)$  is averaged over all vertices ( $n$ ) to get mean local transitivity of a network<sup>220</sup>.

$$\bar{T} = \frac{1}{n} \sum_{i=1}^n T_i$$

Global transitivity is the ration of total number of triangles to number of triplets in a network<sup>220</sup>.



**Fig. 22.** Transitivity of a vertex (red) in three different scenarios. Case-1 does not have a triangle centered on the vertex of interest, thus shows  $0/3=0$  transitivity. Case-2 and -3 have 1 and 3 triangles centered on the vertex corresponding to  $1/3=0.33$  and  $3/3=1$  transitivity respectively.

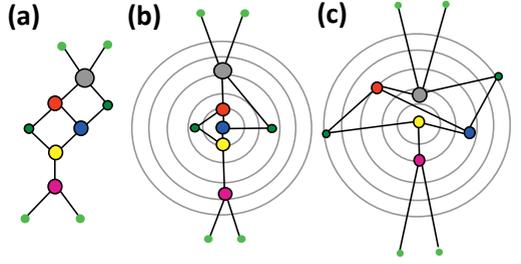
### 6.2.3. Centrality

Centrality, as the name itself suggests, measures how central a vertex is in the network. The simplest centrality measure is degree, i.e. number of connections, of a vertex. However, degree may not always reveal the true global centrality of a vertex. Certain vertices interact among themselves exclusively and thus forms highly intra-connected network modules that are less connected to outside the network, for example, network module of ribosomal proteins. Thus, we used other centrality measures which consider the connectivity of the vertex to rest of the network, namely *betweenness*<sup>220</sup> and *closeness*<sup>220</sup>. *Betweenness* measures number of all shortest paths passing through a vertex, while *closeness* measures the (inverse of) average length of shortest path to all other vertices in the network<sup>220</sup> (f.23). Unweighted breadth-first search algorithm, built in *iGraph*, is used to find the shortest paths. By definition, *betweenness centrality* ( $C_B$ ) of a vertex  $v$  in graph  $G = (V, E)$  is given by,

$$C_B(v) = \sum_{s \neq v \neq t \in V} \sigma_{st}(v)$$

Where  $\sigma_{st}(v)$  is the shortest path between  $s$  and  $t$  vertices passing through vertex  $v$ . Similarly *Closeness centrality* ( $C_C$ ) is given as

$$C_C(v) = (V-1) / \left( \sum_{v \neq i} \sigma_{vi} \right)$$



**Fig. 23.** Measures of vertex centrality. (a) A toy network of  $V=11, E=12$ . The size of the node corresponds to its degree. (b) Closeness centrality reveals blue vertex as the most central followed by red and yellow. (c) Betweenness centrality reveals yellow vertex as the most central followed by grey and pink vertices. Note that red and blue vertices are not closer to yellow anymore. The closeness of vertices does not necessarily mean that they should also fall inbetween the shortest paths of the network. Thus, the application of these measures also associates to the context of the question. However, for large networks where order of vertex-centrality in a sub-network is irrelevant, the choice of method may not matter, as in our case, to a greater extent.

Where  $V$  is the total number of vertices,  $\sigma_{vi}$  is the shortest path between vertex  $v$  and  $i$

## 6.3. Analysis of Intrinsic Disorder

### 6.3.1. Data

Protein sequences of chromatin remodeling factors were downloaded from CREMOFAC<sup>221</sup> database, which includes 720 sequences from 49 different organisms. Redundant sequences were filtered using *CD-Hit* algorithm taking 40% identity criteria, which led to 147 non-redundant chromatin remodeling protein sequences. Sequences for other chromatin proteins were taken from ChromDB<sup>222</sup> database and again filtered using *CD-Hit*<sup>223</sup>.

### 6.3.2. Prediction of intrinsic disorder

As mentioned earlier, intrinsically disordered regions can be predicted by sequence properties alone. An index ( $I_F$ ) of protein folding potential can be derived from following equation<sup>201</sup>.

$$I_F = 2.785 \langle H \rangle - |\langle R \rangle| - 1.151$$

Where,  $\langle R \rangle$  and  $\langle H \rangle$  are the mean net charge and mean net hydrophobicity of the protein. *FoldIndex* algorithm uses this equation in a sliding window strategy. Another supervised algorithm, *IUPred*<sup>224</sup>, which uses amino acids' strength of making stabilized contacts in a given window, is also used for consensus results. It estimates side chain pair-wise interaction energies in a window around a given residue and discriminate disordered regions from ordered as these would show distinctive range of interaction energies<sup>224</sup>.

---

### 6.3.3. Sequence complexity

Low sequence complexity often associates with unfolded conformations<sup>225</sup>. We calculated sequence complexity using Shannon's entropy<sup>225</sup>.

$$S = - \sum_{i=1}^N \left( \frac{N_i}{L} \right) \log_2 \left( \frac{N_i}{L} \right)$$

Where,  $N_i$  = # amino acid  $i$  &  $L$  = length of the sequence.

## 7. Results

### 7.1. Specific aims of the thesis

- ✓ To identify chromatin interaction networks (CINs) in the nucleus. The lack of appropriate techniques to explore such structures genome wide motivated us to invent novel technology.
- ✓ To explore the epigenetic regulation of CINs.
- ✓ To study the role of CINs in essential genomic functions like transcription.
- ✓ To study the fate and role of CINs in epigenetic reprogramming during development.
- ✓ To gain mechanistic insight of CINs, in particular, to study the factors (CTCF/cohesins) involved in these interactions.
- ✓ To identify protein conformational feature that endow diverse interactions to CTCF and, in general, to other chromatin-interacting proteins
- ✓ To study the PINs, if any, of CINs.

### 7.2. Summary of Paper I

#### *A high resolution assay to explore chromatin interactions*

We present a novel assay "Circular Chromosome Conformation Capture (4C)" to identify genome wide interacting partners of a region of choice or bait in an unbiased manner. The assay is briefed in the *Methods* section. We identified wide range of chromatin interactions of *H19-ICR* locus (H-CIN) confirming the efficiency of our assay. The 4C assay was further validated by 3C and DNA-FISH experiments. We specifically argue against random collisions in our 4C data by incorporating several technical controls in the assay. The non-randomness is further reflected in genetic and epigenetic features of interacting sequences. The wide range of *trans* interactions of *H19-ICR* appears to be unique as most studies elsewhere primarily reported interactions in *cis*<sup>30, 58</sup> (*unpublished results in collaboration*). This probably hints at hub-ness of *H19-ICR* which might be attributed to hyper dynamics (being towards terminal?) of this locus.

Our 4C assay, when compared to other versions<sup>226</sup>, is of higher resolution which is determined by the choice of restriction enzyme and the design of array. By using 4-

cutter restriction enzyme (paper I, II & III) and performing a genome-wide tiling array at 52 bp resolution (paper II & III), we certainly cover interactions at higher resolution. High resolution of 4C further allows us map the interacting sequence which can further be subjected to sequence analysis as in the paper III.

#### *Epigenetic control of interactions*

We show that the long range *trans*-interactions identified by 4C are dependent on maternally inherited CTCF binding sites in *H19-ICR*. Mutation in CTCF binding sites on maternal *H19-ICR* leads to significant loss of frequent *trans*-interactions. This observation was validated by 3C and DNA FISH analysis of randomly taken interactors from 4C library. Moreover, we observe over-representation of imprinted genes in the 4C library, further strengthening the epigenetic nature of these interactions. These epigenetic features also argue strongly against random collisions.

#### *Role of H-CIN in transcriptional regulation*

To gain a functional insight of H-CIN, we narrowed down to an imprinted gene *Impact* located upstream to a metabolic gene *Osbpl1a* that interacts with *H19-ICR* in *trans*. We further checked the interaction between DMR of *Impact* gene with *H19-ICR* by 3C. Interestingly, *Impact* DMR interacts with maternal *H19-ICR* and mutation in maternal ICR enhances *Osbpl1a* expression by ~2.5 fold, while represses *Impact* by ~2 fold. This suggests that *H19-ICR* can influence transcription at distance mimicking as an *insulator* (?) in *trans*.

#### *H-CIN and ES cell differentiation*

To further explore the regulation of H-CIN, we studied their frequencies in undifferentiated embryonic stem cells and *in-vitro* derived embryoid bodies. We see substantial change in the pattern of interactions upon differentiation, particularly, the high frequency interactions are more prominent in embryoid bodies as compared to ES cells. Low frequency interactions did not show any significant representation in either ES or embryoid bodies. The data, though limited, hints at reprogramming of interactions during ES cell differentiation. This prompted us to further explore the H-CIN at genome scale (paper II & III).

### 7.3. Summary of Paper II

#### *Reprogramming of H-CIN during ES cell differentiation*

The H-CIN data in paper I limits to the interacting sequences identified by 4C cloning followed by sequencing. We further used tiling array approach to identify an unbiased genome-wide H-CIN. Significantly enriched regions identified in the genome wide 4C screen were further selected to design dedicated tiling array for 4C analysis of mouse embryonic stem (ES) cells and derived embryoid bodies (EB). Reproducible 4C data

from ES and EB samples show striking re-orientation of interactions during differentiation. *H19-ICR* experiences highly complexed intra-chromosomal environment in ES cells, while gets loosen in EBs. Concomitant to this, 3D DNA FISH reveals that *H19-ICR* loops out of its chromosomal territory significantly more often in EBs than in ESCs. This hints that a particular region in the genome might regulate its *cis*- and *trans*- interactions by its own dynamic movement.

### Imprinted CIN

In adherence to our earlier report, we found significant over-representation of imprinted loci in H-CIN. The imprinted genes, regardless of their expression status, were consistently over-represented in ESCs, EBs and neonatal liver 4C datasets. To further uncover the underlying network, we perform extensive triple color DNA FISH experiments in ES cells. Quantitative analysis of all-to-all interactions revealed a wide-spread interaction network of imprinted loci from 7 distinct chromosomes. The imprinted loci primarily interact in a pair-wise manner ('date' interactions), whereas simultaneous interactions ('party' interactions) are comparatively rare. This could attribute to 1) highly dynamic nature of interactions and 2) cost asserted by physical constraints for having simultaneous interactions. Furthermore, we observe differential site preference for *cis*- and *trans*-interactions. The *cis*- interactions, in general, happens inside the chromosomal territory, while *trans*- interactions happens on the edge or outside of territory suggesting that the movement of bait (*H19-ICR*) can, in fact, regulate the mutually exclusive repertoire of *cis*- and *trans*-interactions.

### H-CIN and Epigenetic reprogramming in germ-line

Consistent over-representation of imprinted genes in different cell types may suggest stable inheritance of parental epigenetic mark during post-zygotic development. To elucidate the epigenetic fate of imprinted CIN, we extended our analysis to male germ-line development. We studied interactions together with an epigenetic feature replication timing, which is known to undergo reprogramming at imprinted loci at onset of meiosis during gametogenesis. We recapture the imprinted CIN in spermatogonial cells, which is subsequently lost in spermatocytes and round spermatids and hence parallels epigenetic reprogramming. Spermatogonia inheriting mutations in *CTCF* binding sites in maternal *H19-ICR*, fail to show H-CIN suggesting *H19-ICR* as an organizer of imprinted CIN in germ-line. Further, asynchronous replication timing of all imprinted domains tested is significantly lost when maternal copy of *H19-ICR* inherits mutation in *CTCF* binding sites. This strongly suggests that *H19-ICR* can place epigenetic mark in *trans* and delay the epigenetic reprogramming in male germ-line development until the onset of meiosis.

Moreover, the late replicating allele shifts to early replication as suggested by analysis on somatic lineage. Therefore, we speculate that the H-CIN is maternal in male germ-line. By extrapolation, it might be paternal in female germ-line.

Though the maternal mutation in *H19-ICR* abolishes the proximity among imprinted loci in germ cells, we were puzzled to see that proximity is retained in somatic cells. We argue that this might relate to following possibilities- 1.) The overall proximity among loci may not change, however the close interactions, which cannot be confirmed by DNA FISH, are lost. 2.) Regions in the *Igf2/H19* domain, other than *H19-ICR*, might retain interactions. 3) Interactions in somatic lineages might have some other function, for example, these might regulate co-expression of imprinted genes that has been reported elsewhere<sup>125</sup>. This also prompted us to pursue the study presented in paper V.

Sequence analysis of interacting regions and their immediate neighbors hints at strand bias data not shown in the paper, f.24) near interacting sequences. Strand bias has earlier been shown as a genomic feature of replication origins in mammalian genomes<sup>104</sup>. Thus, probable collisions of replication origins in our data cannot be denied. *H19-ICR*, in fact, has been shown to harbor replication origins in nearby regions (data from a collaborator, not shown). Thus, it can be speculated that *CTCF* sites on *H19-ICR*, along with other epigenetic regulators, might delay the firing of replication origin on maternal allele of imprinted genes and thus keeping alleles replicating asynchronously.

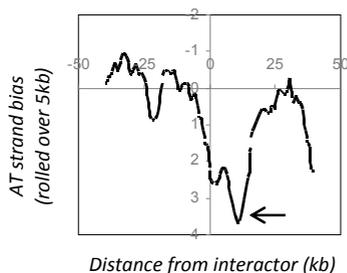
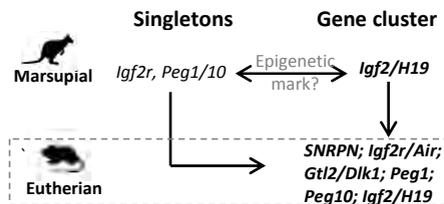


Fig. 24. Strand bias near interacting regions of *H19-ICR*

### *H19-ICR*: an ancient epigenetic organizer?

*Igf2/H19* locus is the only clustered imprinted gene locus in marsupials. Rest of the known imprinted genes are either not imprinted in marsupials or imprinted as singleton. Our data might suggest a possible mechanism involved in the evolution of the imprinted gene clusters from marsupials to eutherians (f.25). Other regions in the genome might have acquired epigenetic mark for imprinting (replication timing?) initially by interacting to *H19-ICR* (?) (f.25) and gained their own ICR later in

the evolution. This may also strengthen the proposal that replication timing could be an ancient mean to establish imprinting followed up by more stable marks like DNA methylation to maintain it.



**Fig. 25.** A hypothetical model, partly supported by our data, for evolution of imprinted gene clusters in eutherians.

Since most of our analysis is centered on *H19*-ICR, it can be argued that our analysis is biased and regulation of imprinted CIN and non-allelic transvection may also be contributed by other imprinted loci. We do not rule out this possibility. However, some of our analysis argues against this. We cross-checked the property of transvection when promoter of *Kenq1ot1*, another imprinted gene which codes for noncoding RNA and regulate domain-wide imprinting, is deleted. We did not see any change in replication timing of *Igf2/H19* domain while *vice-versa* is true (data not shown), suggesting that *H19*-ICR might act as a hub of transvection in mammalian imprintome.

In brief, the study uncovers how a single genomic locus could regulate epigenetic states at distance in a pleiotropic manner.

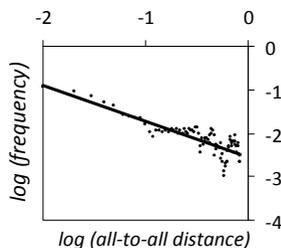
#### 7.4. Summary of paper III

##### Genome wide CIN

Using 4C in combination with tiling array, we uncovered genome-wide interactions of *H19*-ICR in the pooled sample of neonatal liver, neonatal brain, embryonic stem cell and embryoid bodies. The genome-wide data uncovers wide-spread *cis*- and *trans*- interactions of *H19*-ICR. We see preferred clustering of interactions on spatially distinct regions on most chromosomes arguing against the random collisions. The interacting regions strongly associate with transcriptional units, not necessarily gene density, suggesting functional nature of these interactions.

Non random clustering of interactions prompted us to explore if power-law could emerge from our data. We calculated all-to-all distances among interacting sequences and plotted their distribution. The log-log plot (f.26) indeed suggests that the data follows a power law (chromatin interaction *bursts*?). Several factors like fractal organization of chromatin, dynamics of individual

interacting loci, genomic/epigenomic structures could contribute to this. Though any further interpretation of power-law with our limited CIN data would just be a weak speculation, role of trade-mark features of power-law (robustness, *Achilles' heel*) in CIN cannot be ruled out.



**Fig. 26.** Power-law (?) in H-CIN. Shown is the log-log plot of genomic distances (all-to-all) and their frequency in H-CIN.

##### Developmentally regulated H-CIN

H-CIN shows significant association with gene-expression during ES cell differentiation, bivalent chromatin marks and higher tissue specificity of proximal genes. Though ES interactors do not associate with bivalent chromatin marks, EB interactors show strong association with bivalent marks of ES cells. This could suggest that in ES cells H-CIN has heterogeneous population of interactions, while the whole H-CIN organizes itself towards more of developmentally regulated genes. As *Igf2* gene gets upregulated during ES cell differentiation and the whole *Igf2/H19* domain also loops out from its chromosomal territory, it is possible that *Igf2/H19* domain moves from an transcriptionally poised compartment to a comparatively more active compartment. This is also supported by time-series expression data of ES cell differentiation. Genes proximal to interactors show higher expression as the cells differentiate.

Although the gene ontology analysis did not reveal any over-representation of particular functional class, the above observations does hint at development related function of H-CIN.

##### Lack of CTCF/Cohesin's association with H-CIN

CTCF and Cohesins association is increasingly being proposed as global regulator of higher order chromatin conformation. Our analysis, however, contradicts this. We did not observe any over-representation of CTCF binding sites in our H-CIN library. Further, siRNA mediated silencing of *Rad21*, a cohesin subunit, does not alter the physical proximity among interactors. Enrichment of various other factor binding sites, as revealed by mapping TRANSFAC matrices, suggests a possibility of diverse mechanisms by which CTCF mediates chromatin interactions. Recently CTCF-binding sites are categorized

into low, medium and high occupancy sites, which interestingly differ in their epigenetic and functional fate<sup>227</sup>. As the consensus sites are being explored extensively, the protein aspects of CTCF, which could endow multiple interactions to it, remain ignored. Next paper in the thesis addresses this issue to some extent.

## 7.5. Summary of paper IV

### *Structural disorder in chromatin remodeling proteins*

Diverse nuclear functions of chromatin remodeling factors prompted us to explore the structural feature ascribing the multiple interactions. Structural disorder, earlier, has been shown to mediate transient interactions in PPI and transcriptional networks. Using computational methods, we consistently observed greater structural disorder in chromatin remodeling factors. The observation might explain the physical feasibility of diverse interactions of chromatin remodeling proteins with chromatin, DNA, RNA, transcription, replication, splicing and DNA repair factors.

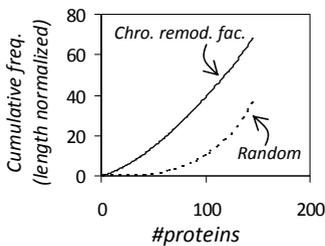


Fig. 27. Structural disorder in chromatin remodeling factors

Since chromatin remodeling factors are, in general, large multi-domain proteins, it can be argued that larger protein length, which in turn could accommodate disordered regions, is the unique feature here instead of structural disorder. We rule out this possibility by normalizing the data to the protein length. Chromatin remodeling factors show consistently higher structural disorder than random datasets (f.27, data not shown in the paper).

### *Low Sequence Complexity in observed LDRs*

Visual inspection of chromatin remodeling and polycomb/trithorax group of proteins suggested presence of tandem repeats of hydrophilic and charged amino acids. We consolidate the observation by calculating sequence complexity and repeat frequencies. Presence of repeats might hint at evolutionary origin of these long disordered regions (LDRs) *via* repeat expansion.

### *Structural disorder in other Chromatin proteins*

A further analysis on wide range of chromatin binding proteins revealed significant intrinsic disorder in most

chromatin proteins. This may either suggest a necessity of disordered conformations for chromatin interaction or a greater diversity in interacting partners of chromatin proteins than anticipated so far. As LDRs are known to be implicated in coupled folding and binding, which brings reversible interactions (as discussed earlier), we further explored Protein Data Bank for structural evidence claiming the same. Though we could not find the instances of coupled folding and binding partly due to limited availability of disordered structures, with the help of few examples we show that disordered regions are indeed directly involved in interaction with naked or modified DNA and with other chromatin related proteins.

### *Intrinsic disorder in CTCF*

One of the cases that we highlight in the article is CTCF protein. CTCF is known to recognize diverse promoters by different combinations of its 11 zinc fingers. CTCF also interacts with number of other proteins and undergo several post-translational modifications. We show that CTCF harbors disordered conformations on its terminals and between the Zn fingers, which in turn renders substantial flexibility to make diverse conformations and interact with diverse promoters and co-factors. Moreover, we find several protein/DNA interaction domains on free disordered tails of CTCF. In particular, we uncover disordered PEST (Pro-Glu-Ser-Thr) motifs, which are widely known to have multiple functional roles like regulating protein degradation (both ubiquitin and calpain mediated), protein-protein interactions and protein phosphorylation<sup>228</sup>. Thus, the study uncovers previously undiscovered features of CTCF protein that could endow multiple mechanisms of its function.

## 7.6. Summary of paper V

### *Network of Imprinted gene-products & partners (IGPN)*

Massive amount of publicly available data allow us to perform systems analysis to test certain hypotheses. We pulled down PPI and gene-expression data of imprinted genes and subjected to extensive theoretical analysis. The analyses uncovered novel insights into imprinted gene-function. Implementing several topological descriptors clearly suggested existence of a network module of imprinted gene-products and their interacting partners, dedicated to imprinted gene functions like cellular biosynthesis, development, metabolism, cell cycle and transcriptional regulation.

As the data suggest that imprinted gene-products are significantly closer to each other as compared to other proteins in HIN, we propose a hypothesis that “*CIN regulating IGPN regulating IGPN*” (f.28), however it is not thoroughly supported yet and hence we do not claim this as such in the paper.

### *High Centrality and dating in IGPN*

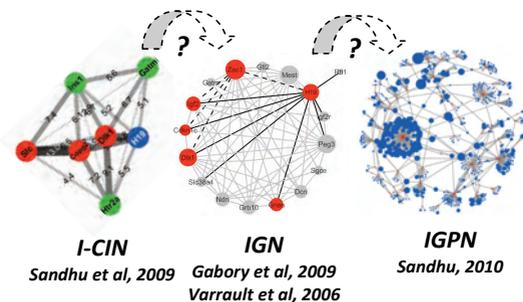


Fig. 28. Imprinted interactomes

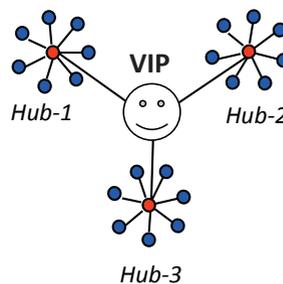


Fig. 29. VIP club phenomenon

Centrality descriptors suggest how important a vertex is in a network. We find that IGPN vertices are highly central to human interactome network (HIN). Imprinted gene-products interact with highly central hubs in HIN which makes IGPN an *Achilles' heel* in HIN. Its malfunction leads to expansion of HIN topology, i.e. longer routes to traverse. In particular, *date* interactions among partners belonging to distinct functional classes in IGPN appears to be more central than the *party* interactions restricted to the same function suggesting that cross-transitivity is a critical component in IGPN as reported elsewhere also<sup>229</sup>. The *date* interactions, which are more central in IGPN, could indeed propagate the perturbation wider than the *party* interactions.

#### Association with complex diseases

Imprinted genes often associate with complex disorders either genetically or epigenetically. Since imprinted genes are part of an extensively connected network module, IGPN, it is possible that significant proportion of this network interplays in disease development. We find significant representation of IGPN genes in OMIM's Morbid entries suggesting their genetic association with diseases. This clearly relates topological centrality of IGPN to their functional indispensability. We further explore the gene expression data for complex disorders and find ~88% of total IGPN genes perturbed in disease phenotypes. Moreover, genes dysregulated in multiple diseases show greater centrality in IGPN network, hinting at their possible topological role in propagating perturbation in the network

#### Imprinted gene-products as VIP club

VIP club phenomenon<sup>230</sup> values certain vertices in a network that do not interact with many partners but with most influential partners (f.29). VIP club escape the cost of having extensive interactions and could manipulate the hubs in the network without being exposed themselves to the mass abruptly. Imprinted gene-products might be analogues to VIP club as we see they interact to top hubs and bring them together to constitute a central network module in human interactome.

## 8. Conclusions

We decipher and annotate novel networks in context of epigenetics. In particular, we establish a novel high resolution, high throughput assay, 4C, to identify one-to-all chromatin interactome(s) in genome. The assay further enabled us uncovering an imprinted chromatin interactome centered on *H19-ICR*, which associates with long range non-allelic transvection of epigenetic states across most of imprintome. In particular, CTCF sites on *H19-ICR* determine the physical proximity among most imprinted loci and transduce their epigenetic states during germline development. Thus, the probable *Achilles' heel* in mammalian development, i.e. *H19-ICR*, proves its identity of being so. We propose that *H19-ICR* could act as a hub (?) of transvection in mammalian imprintome.

We show that physical proximity among imprinted or non-imprinted loci *in trans* may not be strictly mediated by CTCF-Cohesin combination, in contrast to their emerging role in linking relatively short range *cis* complexes. We propose multiple mechanisms of CTCF mediated interactions, which could partially be attributed to structural disorder in CTCF protein.

As the network properties of imprinted genes are increasingly being realized, we show that this trait could indeed be extended to their protein interaction network. Their greater "*small world-ness*" strongly supports existence of their network-module. We show that imprinted gene-products tie together the highly central hubs (*Achilles' heels*) of human interactome, which could impose higher vulnerability to errors in human interactome. We propose that imprinted gene-products could act as *VIP club*, which themselves escape the cost of being hub; instead they connect to most central hubs and probably modulate them.

Our study opens up novel networks perspective in epigenetics, which could endow pleiotropy to the system. Extensive implementation of such holistic strategy in future could help uncovering novel insights in epigenetics which would otherwise go unnoticed in traditional reductionist approach.

## 9. Acknowledgements



I sincerely thank to all who associate with me during my stay in Sweden. In particular, I appreciate the **extraordinary generosity of Swedish people** that I enjoyed throughout my stay.

I also express my gratitude to our open access knowledge resources, i.e. the public databases and search engines, especially WWW hub **Google** and the biomedical literature hub **PubMed**. My salute to open access/source platforms of **PERL** and **R** ! I could code almost everything I needed in the thesis using the two.

I am grateful to **Karolinska Institute** for considering me in their graduate program. Financial support from **Vetenskapsrådet**, **Cancerfonden**, **Barncancerfonden**, and **HEROIC** are duly acknowledged.



Thank you **Rolf** for exposing me to the world of Epigenetics. Your able guidance is much appreciated. I enjoyed the company of my lab-mates, they are the best I ever met. Special thanks to **Chengxi** and **Mikael** not only for our collaborative work, but for guiding me through experiments that eventually decided my thesis. Thanks **Nori** for your expert guidance and valuable suggestions every now and then. I loved talking to you! **Samer**, your managerial skills are much appreciated, thanks for enriching the lab with useful resources. Thanks **Anita** for exposing me to experimental technicalities of 4C. Wish **Carole** and **Sylvain** a very happy life in Montpellier, many thanks pals for the nice time you shared with us. I would miss our Friday outings with lab-mates, esp. our favorite Thai restaurant! Thanking you **Marta** for making our life easier with the cells and lab purchases. Thanks **Helena/Rosemarie** for technical/administrative assistance and your warmth discussions. Thanks a ton **Chandra** for your kind favor in tough goings. Ni Hao Ma **Xinqi**? I wish we could play more of ping-pong, Hej **Olle**, I love your sense of humor! Hi **Merlin**, happy doing Chinese Mafia ☺

I heartedly thank my best friends **Mythily** and **Bjorn** for your honest support in difficult times. Hey **B**, we will have collaboration in future for sure ☺. Hi **Vidya**, if you happened to see my thesis, I acknowledge your lessons on computer programming. Hi **GP**, days spent with you will remain unforgettable. I appreciate your curiosity over everything and your broader sense of science. I acknowledge my mentor **Dr. Raghuram** with whom I started my scientific career, exposure in your lab was a major boost to my scientific enthusiasm. Thank you very much for your kind support throughout my career. I acknowledge my academic stay at **IBAB**, **IBM** and **IGIB**. Exposure in these labs helped me acquiring computational skills.

With all my respect and love, I thank my **family** for their ever-lasting support and encouragement. Without their co-operation, my PhD would remain a dream. I love my family to the deepest of my heart! I thank all my **Indian friends** in Uppsala and Stockholm. The name-list is lengthy, perhaps not beyond **A-Z** (**Adwair-Zidaan** ☺). I am lazy enough to name all and moreover I will feel bad if I miss one. Instead, I convey a **HUGE THANKS** to all of you. I wish I could enjoy Indian parties with you. Corridor friends (again you are too many!), you all leave a sweet memory of late night cooking, kitchen jokes, strange recipes and corridor parties. I will miss you!

## 10. References

1. Barabási A-L. *Linked : how everything is connected to everything else and what it means for business, science, and everyday life*. New York: Plume, 2003.
2. Barabási A-L. *Bursts : the hidden pattern behind everything we do*. New York, N.Y.: Dutton.
3. Cubas P, Vincent C, Coen E. An epigenetic mutation responsible for natural variation in floral symmetry. *Nature* 1999; 401:157-61.
4. Morgan HD, Sutherland HG, Martin DI, Whitelaw E. Epigenetic inheritance at the agouti locus in the mouse. *Nat Genet* 1999; 23:314-8.
5. Saxonov S, Berg P, Brutlag DL. A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proc Natl Acad Sci U S A* 2006; 103:1412-7.
6. Peterson CL, Lanier MA. Histones and histone modifications. *Curr Biol* 2004; 14:R546-51.
7. Heintzman ND, Hon GC, Hawkins RD, Kheradpour P, Stark A, Harp LF, Ye Z, Lee LK, Stuart RK, Ching CW, Ching KA, Antosiewicz-Bourget JE, Liu H, Zhang X, Green RD, Lobanenkov VV, Stewart R, Thomson JA, Crawford GE, Kellis M, Ren B. Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* 2009; 459:108-12.
8. Varga-Weisz P. ATP-dependent chromatin remodeling factors: nucleosome shufflers with many missions. *Oncogene* 2001; 20:3076-85.
9. Cremer T, Kreth G, Koester H, Fink RH, Heintzmann R, Cremer M, Solovei I, Zink D, Cremer C. Chromosome territories, interchromatin domain compartment, and nuclear matrix: an integrated view of the functional nuclear architecture. *Crit Rev Eukaryot Gene Expr* 2000; 10:179-212.
10. Kupper K, Kolb A, Biener D, Dittrich S, von Hase J, Thormeyer T, Fiegler H, Carter NP, Speicher MR, Cremer T, Cremer M. Radial chromatin positioning is shaped by local gene density, not by gene expression. *Chromosoma* 2007; 116:285-306.
11. Kreth G, Finsterle J, von Hase J, Cremer M, Cremer C. Radial arrangement of chromosome territories in human cell nuclei: a computer model approach based on gene density indicates a probabilistic global positioning code. *Biophys J* 2004; 86:2803-12.
12. Clemson CM, Hall LL, Byron M, McNeil J, Lawrence JB. The X chromosome is organized into a gene-rich outer rim and an internal core containing silenced nongenic sequences. *Proc Natl Acad Sci U S A* 2006; 103:7688-93.
13. Mearini G, Fackelmayr FO. Local chromatin mobility is independent of transcriptional activity. *Cell Cycle* 2006; 5:1989-95.
14. Verschure PJ, van Der Kraan I, Manders EM, van Driel R. Spatial relationship between transcription sites and chromosome territories. *J Cell Biol* 1999; 147:13-24.
15. Reddy KL, Zullo JM, Bertolino E, Singh H. Transcriptional repression mediated by repositioning of genes to the nuclear lamina. *Nature* 2008; 452:243-7.
16. Brickner DG, Cajigas I, Fondufe-Mittendorf Y, Ahmed S, Lee PC, Widom J, Brickner JH. H2A.Z-mediated localization of genes at the nuclear periphery confers epigenetic memory of previous transcriptional state. *PLoS Biol* 2007; 5:e81.
17. Brown CR, Kennedy CJ, Delmar VA, Forbes DJ, Silver PA. Global histone acetylation induces functional genomic reorganization at mammalian nuclear pore complexes. *Genes Dev* 2008; 22:627-39.
18. Tan-Wong SM, Wijayatilake HD, Proudfoot NJ. Gene loops function to maintain transcriptional memory through interaction with the nuclear pore complex. *Genes Dev* 2009; 23:2610-24.
19. O'Sullivan JM, Tan-Wong SM, Morillon A, Lee B, Coles J, Mellor J, Proudfoot NJ. Gene loops juxtapose promoters and terminators in yeast. *Nat Genet* 2004; 36:1014-8.
20. Nemeth A, Guibert S, Tiwari VK, Ohlsson R, Langst G. Epigenetic regulation of TTF-I-mediated promoter-terminator interactions of rRNA genes. *EMBO J* 2008; 27:1255-65.
21. Singh BN, Hampsey M. A transcription-independent role for TFIIIB in gene looping. *Mol Cell* 2007; 27:806-16.
22. Rippe K. Making contacts on a nucleic acid polymer. *Trends Biochem Sci* 2001; 26:733-40.
23. Li Q, Barkess G, Qian H. Chromatin looping and the probability of transcription. *Trends Genet* 2006; 22:197-202.
24. Murugan R. Action at distance along the DNA: a random jump model on enhancer action. *J Theor Biol* 2007; 248:696-705.
25. Dekker J, Rippe K, Dekker M, Kleckner N. Capturing chromosome conformation. *Science* 2002; 295:1306-11.
26. Vernimmen D, De Gobbi M, Sloane-Stanley JA, Wood WG, Higgs DR. Long-range chromosomal interactions regulate the timing of the transition between poised and active gene expression. *EMBO J* 2007; 26:2041-51.
27. Tsytsykova AV, Rajsbaum R, Falvo JV, Ligeiro F, Neely SR, Goldfeld AE. Activation-dependent intrachromosomal interactions formed by the TNF gene promoter and two distal enhancers. *Proc Natl Acad Sci U S A* 2007; 104:16850-5.
28. Deschenes J, Bourdeau V, White JH, Mader S. Regulation of GREB1 transcription by estrogen receptor alpha through a multipartite enhancer spread over 20 kb of upstream flanking sequences. *J Biol Chem* 2007; 282:17335-9.
29. Hu Q, Kwon YS, Nunez E, Cardamone MD, Hutt KR, Ohgi KA, Garcia-Bassets I, Rose DW, Glass CK, Rosenfeld MG, Fu XD. Enhancing nuclear receptor-induced transcription requires nuclear motor and LSD1-dependent gene networking in interchromatin granules. *Proc Natl Acad Sci U S A* 2008; 105:19199-204.
30. Fullwood MJ, Liu MH, Pan YF, Liu J, Xu H, Mohamed YB, Orlov YL, Velkov S, Ho A, Mei PH, Chew EG, Huang PY, Welboren WJ, Han Y, Ooi HS, Ariyaratne PN, Vega VB, Luo Y, Tan PY, Choy PY, Wansa KD, Zhao B, Lim KS, Leow SC, Yow JS, Joseph R, Li H, Desai KV, Thomsen JS, Lee YK, Karuturi RK, Herve T, Bourque G, Stunnenberg HG, Ruan X, Cacheux-Rataboul V, Sung WK, Liu ET, Wei CL, Cheung E, Ruan Y. An oestrogen-receptor-alpha-bound human chromatin interactome. *Nature* 2009; 462:58-64.
31. Branco MR, Pombo A. Intermingling of chromosome territories in interphase suggests role in translocations and transcription-dependent associations. *PLoS Biol* 2006; 4:e133.
32. Lomvardas S, Barnea G, Pisapia DJ, Mendelsohn M, Kirkland J, Axel R. Interchromosomal interactions and olfactory receptor choice. *Cell* 2006; 126:403-13.
33. Schoenfelder S, Sexton T, Chakalova L, Cope NF, Horton A, Andrews S, Kurukuti S, Mitchell JA, Umlauf D, Dimitrova DS, Eskiw CH, Luo Y, Wei CL, Ruan Y, Bieker JJ, Fraser P. Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells. *Nat Genet*; 42:53-61.
34. Dhar SS, Ongwijitwat S, Wong-Riley MT. Chromosome conformation capture of all 13 genomic Loci in the transcriptional regulation of the multisubunit biogenic cytochrome C oxidase in neurons. *J Biol Chem* 2009; 284:18644-50.
35. Lanzaolo C, Roure V, Dekker J, Bantignies F, Orlando V. Polycomb response elements mediate the formation of chromosome higher-order structures in the bithorax complex. *Nat Cell Biol* 2007; 9:1167-74.
36. Gause M, Schaaf CA, Dorsett D. Cohesin and CTCF: cooperating to control chromosome conformation? *Bioessays* 2008; 30:715-8.
37. Cai S, Han HJ, Kohwi-Shigematsu T. Tissue-specific nuclear architecture and gene expression regulated by SATB1. *Nat Genet* 2003; 34:42-51.
38. Ohlsson R, Lobanenkov V, Klenova E. Does CTCF mediate between nuclear organization and gene expression? *Bioessays*; 32:37-50.
39. Kurukuti S, Tiwari VK, Tavoosidana G, Pugacheva E, Murrell A, Zhao Z, Lobanenkov V, Reik W, Ohlsson R. CTCF binding at the H19 imprinting control region mediates maternally inherited higher-order chromatin conformation to restrict enhancer access to Igf2. *Proc Natl Acad Sci U S A* 2006; 103:10684-9.
40. Majumder P, Gomez JA, Chadwick BP, Boss JM. The insulator factor CTCF controls MHC class II gene expression and is required for the formation of long-distance chromatin interactions. *J Exp Med* 2008; 205:785-98.
41. Xu N, Donohoe ME, Silva SS, Lee JT. Evidence that homologous X-chromosome pairing requires transcription and Ctfc protein. *Nat Genet* 2007; 39:1390-6.
42. Schoenherr CJ, Levorse JM, Tilghman SM. CTCF maintains differential methylation at the Igf2/H19 locus. *Nat Genet* 2003; 33:66-9.
43. Mukhopadhyay R, Yu W, Whitehead J, Xu J, Lezcano M, Pack S, Kanduri C, Kanduri M, Ginjaia V, Vostrov A, Quitschke W, Chermukhin I, Klenova E, Lobanenkov V, Ohlsson R. The binding sites for the chromatin insulator protein CTCF map to DNA methylation-free domains genome-wide. *Genome Res* 2004; 14:1594-602.
44. Klenova EM, Chermukhin IV, El-Kady A, Lee RE, Pugacheva EM, Loukinov DI, Goodwin GH, Delgado D, Filippova GN, Leon J, Morse HC, 3rd, Neiman PE, Lobanenkov VV. Functional phosphorylation sites in the C-terminal region of the multivalent multifunctional transcriptional factor CTCF. *Mol Cell Biol* 2001; 21:2221-34.

45. Yu W, Ginjala V, Pant V, Chernukhin I, Whitehead J, Docquier F, Farrar D, Tavoosidana G, Mukhopadhyay R, Kanduri C, Oshimura M, Feinberg AP, Lobanenkov V, Klenova E, Ohlsson R. Poly(ADP-ribosylation) regulates CTCF-dependent chromatin insulation. *Nat Genet* 2004; 36:1105-10.
46. El-Kady A, Klenova E. Regulation of the transcription factor, CTCF, by phosphorylation with protein kinase CK2. *FEBS Lett* 2005; 579:1424-34.
47. MacPherson MJ, Beatty LG, Zhou W, Du M, Sadowski PD. The CTCF insulator protein is posttranslationally modified by SUMO. *Mol Cell Biol* 2009; 29:1714-25.
48. Ong CT, Corces VG. Modulation of CTCF insulator function by transcription of a noncoding RNA. *Dev Cell* 2008; 15:489-90.
49. Splinter E, Heath H, Kooren J, Palstra RJ, Klous P, Grosveld F, Galjart N, de Laat W. CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus. *Genes Dev* 2006; 20:2349-54.
50. Lutz M, Burke LJ, Barreto G, Goeman F, Greb H, Arnold R, Schultheiss H, Brehm A, Kouzarides T, Lobanenkov V, Renkawitz R. Transcriptional repression by the insulator protein CTCF involves histone deacetylases. *Nucleic Acids Res* 2000; 28:1707-13.
51. Torrano V, Navascues J, Docquier F, Zhang R, Burke LJ, Chernukhin I, Farrar D, Leon J, Berciano MT, Renkawitz R, Klenova E, Lafarga M, Delgado MD. Targeting of CTCF to the nucleolus inhibits nucleolar transcription through a poly(ADP-ribosylation)-dependent mechanism. *J Cell Sci* 2006; 119:1746-59.
52. Ishihara K, Oshimura M, Nakao M. CTCF-dependent chromatin insulator is linked to epigenetic remodeling. *Mol Cell* 2006; 23:733-42.
53. Donohoe ME, Zhang LF, Xu N, Shi Y, Lee JT. Identification of a Ctf cofactor, Yy1, for the X chromosome binary switch. *Mol Cell* 2007; 25:43-56.
54. Chernukhin I, Shamsuddin S, Kang SY, Bergstrom R, Kwon YW, Yu W, Whitehead J, Mukhopadhyay R, Docquier F, Farrar D, Morrison I, Vigneron M, Wu SY, Chiang CM, Loukinov D, Lobanenkov V, Ohlsson R, Klenova E. CTCF interacts with and recruits the largest subunit of RNA polymerase II to CTCF target sites genome-wide. *Mol Cell Biol* 2007; 27:1631-48.
55. Rubio ED, Reiss DJ, Welsh PL, Disteche CM, Filippova GN, Baliga NS, Aebersold R, Ranish JA, Krumm A. CTCF physically links cohesin to chromatin. *Proc Natl Acad Sci U S A* 2008; 105:8309-14.
56. Guelen L, Pagie L, Brasset E, Meuleman V, Faza MB, Talhout W, Eussen BH, de Klein A, Wessels L, de Laat W, van Steensel B. Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature* 2008; 453:948-51.
57. Kim TH, Abdullaev ZK, Smith AD, Ching KA, Loukinov DI, Green RD, Zhang MQ, Lobanenkov VV, Ren B. Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. *Cell* 2007; 128:1231-45.
58. Simonis M, Klous P, Splinter E, Moshkin Y, Willemsen R, de Wit E, van Steensel B, de Laat W. Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nat Genet* 2006; 38:1348-54.
59. Parelho V, Hadjur S, Spivakov M, Leleu M, Sauer S, Gregson HC, Jarmuz A, Canzonetta C, Webster Z, Nesterova T, Cobb BS, Yokomori K, Dillon N, Aragon L, Fisher AG, Merkschlagler M. Cohesins functionally associate with CTCF on mammalian chromosome arms. *Cell* 2008; 132:422-33.
60. Wendt KS, Yoshida K, Itoh T, Bando M, Koch B, Schirghuber E, Tsutsumi S, Nagae G, Ishihara K, Mishiro T, Yahata K, Imamoto F, Aburatani H, Nakao M, Imamoto N, Maeshima K, Shirahige K, Peters JM. Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature* 2008; 451:796-801.
61. Nativio R, Wendt KS, Ito Y, Huddleston JE, Uribe-Lewis S, Woodfine K, Krueger C, Reik W, Peters JM, Murrell A. Cohesin is required for higher-order chromatin conformation at the imprinted IGF2-H19 locus. *PLoS Genet* 2009; 5:e1000739.
62. Hadjur S, Williams LM, Ryan NK, Cobb BS, Sexton T, Fraser P, Fisher AG, Merkschlagler M. Cohesins form chromosomal cis-interactions at the developmentally regulated IFNG locus. *Nature* 2009; 460:410-3.
63. Rodriguez-Campos A, Azorin F. RNA is an integral component of chromatin that contributes to its structural organization. *PLoS One* 2007; 2:e1182.
64. Nagano T, Mitchell JA, Sanz LA, Pauler FM, Ferguson-Smith AC, Feil R, Fraser P. The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. *Science* 2008; 322:1717-20.
65. Pandey RR, Mondal T, Mohammad F, Enroth S, Redrup L, Komorowski J, Nagano T, Mancini-Dinardo D, Kanduri C. Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. *Mol Cell* 2008; 32:232-46.
66. Terranova R, Yokobayashi S, Stadler MB, Otte AP, van Lohuizen M, Orkin SH, Peters AH. Polycomb group proteins Ezh2 and Rnf2 direct genomic contraction and imprinted repression in early mouse embryos. *Dev Cell* 2008; 15:668-79.
67. Khalil AM, Guttman M, Huarte M, Garber M, Raj A, Rivea Morales D, Thomas K, Presser A, Bernstein BE, van Oudenaarden A, Regev A, Lander ES, Rinn JL. Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc Natl Acad Sci U S A* 2009; 106:11667-72.
68. Li JT, Zhang Y, Kong L, Liu QR, Wei L. Trans-natural antisense transcripts including noncoding RNAs in 10 species: implications for expression regulation. *Nucleic Acids Res* 2008; 36:4833-44.
69. Mariner PD, Walters RD, Espinoza CA, Drullinger LF, Wagner SD, Kugel JF, Goodrich JA. Human Alu RNA is a modular transacting repressor of mRNA transcription during heat shock. *Mol Cell* 2008; 29:499-509.
70. Clemson CM, Hutchinson JS, Sara SA, Ensminger AW, Fox AH, Chess A, Lawrence JB. An architectural role for a nuclear noncoding RNA: NEAT1 RNA is essential for the structure of paraspeckles. *Mol Cell* 2009; 33:717-26.
71. Vitali P, Royo H, Marty V, Bortolin-Cavaillle ML, Cavaillle J. Long nuclear-retained non-coding RNAs and allele-specific higher-order chromatin organization at imprinted snoRNA gene arrays. *J Cell Sci* 123:70-83.
72. Sasaki YT, Ideue T, Sano M, Mituyama T, Hirose T. MENP5/10/beta noncoding RNAs are essential for structural integrity of nuclear paraspeckles. *Proc Natl Acad Sci U S A* 2009; 106:2525-30.
73. Palstra RJ, Simonis M, Klous P, Brasset E, Eijkelkamp B, de Laat W. Maintenance of long-range DNA interactions after inhibition of ongoing RNA polymerase II transcription. *PLoS ONE* 2008; 3:e1661.
74. Marenduzzo D, Finan K, Cook PR. The depletion attraction: an underappreciated force driving cellular organization. *J Cell Biol* 2006; 175:681-6.
75. Minton AP. Macromolecular crowding. *Curr Biol* 2006; 16:R269-71.
76. Minton AP. The influence of macromolecular crowding and macromolecular confinement on biochemical reactions in physiological media. *J Biol Chem* 2001; 276:10577-80.
77. Minton AP. Implications of macromolecular crowding for protein assembly. *Curr Opin Struct Biol* 2000; 10:34-9.
78. Minton AP. Macromolecular crowding and molecular recognition. *J Mol Recognit* 1993; 6:211-4.
79. Andersen JS, Lam YW, Leung AK, Ong SE, Lyon CE, Lamond AI, Mann M. Nucleolar proteome dynamics. *Nature* 2005; 433:77-83.
80. Olson MO, Dundr M, Szebeni A. The nucleolus: an old factory with unexpected capabilities. *Trends Cell Biol* 2000; 10:189-96.
81. Sirri V, Hernandez-Verdun D, Roussel P. Cyclin-dependent kinases govern formation and maintenance of the nucleolus. *J Cell Biol* 2002; 156:969-81.
82. Bancaud A, Huet S, Daigle N, Mozziconacci J, Beaudouin J, Ellenberg J. Molecular crowding affects diffusion and binding of nuclear proteins in heterochromatin and reveals the fractal organization of chromatin. *EMBO J* 2009; 28:3785-98.
83. Solovei I, Kreysing M, Lanctot C, Kosem S, Peichl L, Cremer T, Guck J, Joffe B. Nuclear architecture of rod photoreceptor cells adapts to vision in mammalian evolution. *Cell* 2009; 137:356-68.
84. Osborne CS, Chakalova L, Brown KE, Carter D, Horton A, Debrand E, Goyenechea B, Mitchell JA, Lopes S, Reik W, Fraser P. Active genes dynamically colocalize to shared sites of ongoing transcription. *Nat Genet* 2004; 36:1065-71.
85. Jackson DA, Iborra FJ, Manders EM, Cook PR. Numbers and organization of RNA polymerases, nascent transcripts, and transcription units in HeLa nuclei. *Mol Biol Cell* 1998; 9:1523-36.
86. Mitchell JA, Fraser P. Transcription factories are nuclear subcompartments that remain in the absence of transcription. *Genes Dev* 2008; 22:20-5.
87. Xu M, Cook PR. Similar active genes cluster in specialized transcription factories. *J Cell Biol* 2008; 181:615-23.
88. Grande MA, van der Kraan I, de Jong L, van Driel R. Nuclear distribution of transcription factors in relation to sites of transcription and RNA polymerase II. *J Cell Sci* 1997; 110 (Pt 15):1781-91.

89. Han HJ, Russo J, Kohwi Y, Kohwi-Shigematsu T. SATB1 reprogrammes gene expression to promote breast tumour growth and metastasis. *Nature* 2008; 452:187-93.
90. Mintz PJ, Spector DL. Compartmentalization of RNA processing factors within nuclear speckles. *J Struct Biol* 2000; 129:241-51.
91. Melcak I, Melcakova S, Kopsky V, Vecerova J, Raska I. Presplicingosomal assembly on microinjected precursor mRNA takes place in nuclear speckles. *Mol Biol Cell* 2001; 12:393-406.
92. Huang S, Spector DL. Intron-dependent recruitment of pre-mRNA splicing factors to sites of transcription. *J Cell Biol* 1996; 133:719-32.
93. Lin C, Yang L, Tanasa B, Hutt K, Ju BG, Ohgi K, Zhang J, Rose DW, Fu XD, Glass CK, Rosenfeld MG. Nuclear receptor-induced chromosomal proximity and DNA breaks underlie specific translocations in cancer. *Cell* 2009; 139:1069-83.
94. Mani RS, Tomlins SA, Callahan K, Ghosh A, Nyati MK, Varambally S, Palanisamy N, Chinnaiyan AM. Induced chromosomal proximity and gene fusions in prostate cancer. *Science* 2009; 326:1230.
95. Unneberg P, Claverie JM. Tentative mapping of transcription-induced interchromosomal interaction using chimeric EST and mRNA data. *PLoS One* 2007; 2:e254.
96. Kitamura E, Blow JJ, Tanaka TU. Live-cell imaging reveals replication of individual replicons in eukaryotic replication factories. *Cell* 2006; 125:1297-308.
97. Gilbert DM. Replication timing and metazoan evolution. *Nat Genet* 2002; 32:336-7.
98. Wei X, Samarabandu J, Devdhar RS, Siegel AJ, Acharya R, Berezney R. Segregation of transcription and replication sites into higher order domains. *Science* 1998; 281:1502-6.
99. Hassan AB, Errington RJ, White NS, Jackson DA, Cook PR. Replication and transcription sites are colocalized in human cells. *J Cell Sci* 1994; 107 ( Pt 2):425-34.
100. Panning MM, Gilbert DM. Spatio-temporal organization of DNA replication in murine embryonic stem, primary, and immortalized cells. *J Cell Biochem* 2005; 95:74-82.
101. Dimitrova DS, Prokhorova TA, Blow JJ, Todorov IT, Gilbert DM. Mammalian nuclei become licensed for DNA replication during late telophase. *J Cell Sci* 2002; 115:51-9.
102. Lisby M, Rothstein R, Mortensen UH. Rad52 forms DNA repair and recombination centers during S phase. *Proc Natl Acad Sci U S A* 2001; 98:8276-82.
103. Solomon DA, Cardoso MC, Knudsen ES. Dynamic targeting of the replication machinery to sites of DNA damage. *J Cell Biol* 2004; 166:455-63.
104. Huvet M, Nicolas S, Touchon M, Audit B, d'Aubenton-Carafa Y, Arneodo A, Thernes C. Human gene organization driven by the coordination of replication and transcription. *Genome Res* 2007; 17:1278-85.
105. Gimelbrant A, Hutchinson JN, Thompson BR, Chess A. Widespread monoallelic expression on human autosomes. *Science* 2007; 318:1136-40.
106. Moore T, Haig D. Genomic imprinting in mammalian development: a parental tug-of-war. *Trends Genet* 1991; 7:45-9.
107. Barlow DP. Methylation and imprinting: from host defense to gene regulation? *Science* 1993; 260:309-10.
108. Day T, Bonduriansky R. Intralocus sexual conflict can drive the evolution of genomic imprinting. *Genetics* 2004; 167:1537-46.
109. Gross L. A new theory for the evolution of genomic imprinting. *PLoS Biol* 2006; 4:e421.
110. Ohlsson R, Paldi A, Graves JA. Did genomic imprinting and X chromosome inactivation arise from stochastic expression? *Trends Genet* 2001; 17:136-41.
111. Reik W, Walter J. Genomic imprinting: parental influence on the genome. *Nat Rev Genet* 2001; 2:21-32.
112. Luedt PP, Dietrich FS, Weidman JR, Bosko JM, Jirtle RL, Hartemink AJ. Computational and experimental identification of novel human imprinted genes. *Genome Res* 2007; 17:1723-30.
113. Kim J. Multiple YY1 and CTCF binding sites in imprinting control regions. *Epigenetics* 2008; 3:115-8.
114. Dindot SV, Person R, Strivens M, Garcia R, Beaudet AL. Epigenetic profiling at mouse imprinted gene clusters reveals novel epigenetic and genetic features at differentially methylated regions. *Genome Res* 2009; 19:1374-83.
115. Burke LJ, Zhang R, Bartkuhn M, Tiwari VK, Tavosoidana G, Kurukuti S, Weth C, Leers J, Galjart N, Ohlsson R, Renkawitz R. CTCF binding and higher order chromatin structure of the H19 locus are maintained in mitotic chromatin. *EMBO J* 2005; 24:3291-300.
116. Han L, Lee DH, Szabo PE. CTCF is the master organizer of domain-wide allele-specific chromatin at the H19/Igf2 imprinted region. *Mol Cell Biol* 2008; 28:1124-35.
117. Kanduri C, Pant V, Loukinov D, Pugacheva E, Qi CF, Wolffe A, Ohlsson R, Lobanekov VV. Functional association of CTCF with the insulator upstream of the H19 gene is parent of origin-specific and methylation-sensitive. *Curr Biol* 2000; 10:853-6.
118. Kono T, Obata Y, Wu Q, Niwa K, Ono Y, Yamamoto Y, Park ES, Seo JS, Ogawa H. Birth of parthenogenetic mice that can develop to adulthood. *Nature* 2004; 428:860-4.
119. Sakatani T, Kaneda A, Iacobuzio-Donahue CA, Carter MG, de Boom Witzel S, Okano H, Ko MS, Ohlsson R, Longo DL, Feinberg AP. Loss of imprinting of Igf2 alters intestinal maturation and tumorigenesis in mice. *Science* 2005; 307:1976-8.
120. Kaneda A, Wang CJ, Cheong R, Timp W, Onyango P, Wen B, Iacobuzio-Donahue CA, Ohlsson R, Andraos R, Pearson MA, Sharov AA, Longo DL, Ko MS, Levchenko A, Feinberg AP. Enhanced sensitivity to IGF-II signaling links loss of imprinting of IGF2 to increased cell proliferation and tumor risk. *Proc Natl Acad Sci U S A* 2007; 104:20926-31.
121. Vu TH, Nguyen AH, Hoffman AR. Loss of IGF2 imprinting is associated with abrogation of long-range intrachromosomal interactions in human cancer cells. *Hum Mol Genet*; 19:901-19.
122. Smits G, Mungall AJ, Griffiths-Jones S, Smith P, Beury D, Matthews L, Rogers J, Pask AJ, Shaw G, VandeBerg JL, McCarrey JR, Renfree MB, Reik W, Dunham I. Conservation of the H19 noncoding RNA and H19-IGF2 imprinting mechanism in therians. *Nat Genet* 2008; 40:971-6.
123. Falls JG, Pulford DJ, Wylie AA, Jirtle RL. Genomic imprinting: implications for human disease. *Am J Pathol* 1999; 154:635-47.
124. Ohlsson R. Loss of IGF2 imprinting: mechanisms and consequences. *Novartis Found Symp* 2004; 262:108-21; discussion 21-4, 265-8.
125. Varrault A, Gueydan C, Delalbre A, Bellmann A, Houssami S, Aknin C, Severac D, Chotard L, Kahli M, Le Digarcher A, Pavlidis P, Journot L. Zac1 regulates an imprinted gene network critically involved in the control of embryonic growth. *Dev Cell* 2006; 11:711-22.
126. Gabory A, Ripoche MA, Le Digarcher A, Watrin F, Ziyayat A, Forne T, James H, Ainscough JF, Surani MA, Journot L, Dandolo L. H19 acts as a trans regulator of the imprinted gene network controlling growth in mice. *Development* 2009; 136:3413-21.
127. Jager S, Schonherr N, Spengler S, Ranke MB, Wollmann HA, Binder G, Eggermann T. LOT1 (ZAC1/PLAGL1) as member of an imprinted gene network does not harbor Silver-Russell specific variants. *J Pediatr Endocrinol Metab* 2009; 22:555-9.
128. Yoshimizu T, Miroglio A, Ripoche MA, Gabory A, Vernucci M, Riccio A, Colnot S, Godard C, Terris B, Jammes H, Dandolo L. The H19 locus acts in vivo as a tumor suppressor. *Proc Natl Acad Sci U S A* 2008; 105:12417-22.
129. Blik J, Verde G, Callaway J, Maas SM, De Crescenzo A, Sparago F, Cerrato F, Russo S, Ferraiuolo S, Rinaldi MM, Fischetto R, Lalatta F, Giordano L, Ferrari P, Cubellis MV, Larizza L, Temple IK, Mannens MM, Mackay DJ, Riccio A. Hypomethylation at multiple maternally methylated imprinted regions including PLAGL1 and GNAS loci in Beckwith-Wiedemann syndrome. *Eur J Hum Genet* 2009; 17:611-9.
130. Reik W. Stability and flexibility of epigenetic gene regulation in mammalian development. *Nature* 2007; 447:425-32.
131. Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, Fry B, Meissner A, Wernig M, Plath K, Jaenisch R, Wagschal A, Feil R, Schreiber SL, Lander ES. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* 2006; 125:315-26.
132. Ruthenburg AJ, Li H, Patel DJ, Allis CD. Multivalent engagement of chromatin modifications by linked binding modules. *Nat Rev Mol Cell Biol* 2007; 8:983-94.
133. Lopes S, Lewis A, Hajkova P, Dean W, Oswald J, Forne T, Murrell A, Constancia M, Bartolomei M, Walter J, Reik W. Epigenetic modifications in an imprinting cluster are controlled by a hierarchy of DMRs suggesting long-range chromatin interactions. *Hum Mol Genet* 2003; 12:295-305.
134. Davis TL, Trasler JM, Moss SB, Yang GJ, Bartolomei MS. Acquisition of the H19 methylation imprint occurs differentially on the parental alleles during spermatogenesis. *Genomics* 1999; 58:18-28.
135. Gebert C, Kunkel D, Grinberg A, Pfeifer K. H19ICR methylation requires an imprinted environment only in the male germ line. *Mol Cell Biol* 2009.

136. Hiratani I, Ryba T, Itoh M, Yokochi T, Schwaiger M, Chang CW, Lyou Y, Townes TM, Schubeler D, Gilbert DM. Global reorganization of replication domains during embryonic stem cell differentiation. *PLoS Biol* 2008; 6:e245.
137. Gondor A, Ohlsson R. Replication timing and epigenetic reprogramming of gene expression: a two-way relationship? *Nat Rev Genet* 2009; 10:269-76.
138. Kitsberg D, Selig S, Brandeis M, Simon I, Keshet I, Driscoll DJ, Nicholls RD, Cedar H. Allele-specific replication timing of imprinted gene regions. *Nature* 1993; 364:459-63.
139. Gribnau J, Hochedlinger K, Hata K, Li E, Jaenisch R. Asynchronous replication timing of imprinted loci is independent of DNA methylation, but consistent with differential subnuclear localization. *Genes Dev* 2003; 17:759-73.
140. Simon I, Tenzen T, Reubinoff BE, Hillman D, McCarrey JR, Cedar H. Asynchronous replication of imprinted genes is established in the gametes and maintained during development. *Nature* 1999; 401:929-32.
141. Caspary T, Cleary MA, Baker CC, Guan XJ, Tilghman SM. Multiple mechanisms regulate imprinting of the mouse distal chromosome 7 gene cluster. *Mol Cell Biol* 1998; 18:3466-74.
142. Cerrato F, Dean W, Davies K, Kagotani K, Mitsuya K, Okumura K, Riccio A, Reik W. Paternal imprints can be established on the maternal *Igf2-H19* locus without altering replication timing of DNA. *Hum Mol Genet* 2003; 12:3123-32.
143. Phair RD, Misteli T. High mobility of proteins in the mammalian cell nucleus. *Nature* 2000; 404:604-9.
144. Houtsmuller AB, Vermeulen W. Macromolecular dynamics in living cell nuclei revealed by fluorescence redistribution after photobleaching. *Histochem Cell Biol* 2001; 115:13-21.
145. Gorski SA, Dunder M, Misteli T. The road much traveled: trafficking in the cell nucleus. *Curr Opin Cell Biol* 2006; 18:284-90.
146. Dinant C, Luijsterburg MS, Hofer T, von Bornstaedt G, Vermeulen W, Houtsmuller AB, van Driel R. Assembly of multiprotein complexes that control genome function. *J Cell Biol* 2009; 185:21-6.
147. Blossey R, Schiessel H. Kinetic proofreading of gene activation by chromatin remodeling. *HFSP J* 2008; 2:167-70.
148. Hopfield JJ. Kinetic proofreading: a new mechanism for reducing errors in biosynthetic processes requiring high specificity. *Proc Natl Acad Sci U S A* 1974; 71:4135-9.
149. Voss TC, Schiltz RL, Sung MH, Johnson TA, John S, Hager GL. Combinatorial probabilistic chromatin interactions produce transcriptional heterogeneity. *J Cell Sci* 2009; 122:345-56.
150. Golding I, Cox EC. Eukaryotic transcription: what does it mean for a gene to be 'on'? *Curr Biol* 2006; 16:R371-3.
151. Kaern M, Elston TC, Blake WJ, Collins JJ. Stochasticity in gene expression: from theories to phenotypes. *Nat Rev Genet* 2005; 6:451-64.
152. Kouzarides T. Chromatin modifications and their function. *Cell* 2007; 128:693-705.
153. Lusser A, Kadonaga JT. Chromatin remodeling by ATP-dependent molecular machines. *Bioessays* 2003; 25:1192-200.
154. Gangaraju VK, Bartholomew B. Mechanisms of ATP dependent chromatin remodeling. *Mutat Res* 2007; 618:3-17. Epub 2007 Jan 21.
155. Cairns BR. Chromatin remodeling complexes: strength in diversity, precision through specialization. *Curr Opin Genet Dev* 2005; 15:185-90.
156. van Vugt JJ, Ranes M, Campsteijn C, Logie C. The ins and outs of ATP-dependent chromatin remodeling in budding yeast: biophysical and proteomic perspectives. *Biochim Biophys Acta* 2007; 1769:153-71. Epub 2007 Feb 9.
157. Neely KE, Hassan AH, Brown CE, Howe L, Workman JL. Transcription activator interactions with multiple SWI/SNF subunits. *Mol Cell Biol* 2002; 22:1615-25.
158. Cavellani E, Asp P, Percipalle P, Farrants AK. The WSTF-SNF2h chromatin remodeling complex interacts with several nuclear proteins in transcription. *J Biol Chem* 2006; 281:16264-71. Epub 2006 Apr 9.
159. Kouzarides T. Chromatin modifications and their function. *Cell* 2007; 128:693-705.
160. Mohrmann L, Verrijzer CP. Composition and functional specificity of SWI2/SNF2 class chromatin remodeling complexes. *Biochim Biophys Acta* 2005; 1681:59-73.
161. Ko M, Sohn DH, Chung H, Seong RH. Chromatin remodeling, development and disease. *Mutat Res* 2008; 647:59-67.
162. Batsche E, Yaniv M, Muchardt C. The human SWI/SNF subunit Brm is a regulator of alternative splicing. *Nat Struct Mol Biol* 2006; 13:22-9. Epub 2005 Dec 11.
163. Sillibourne JE, Delaval B, Redick S, Sinah M, Doxsey SJ. Chromatin Remodeling Proteins Interact with Pericentrin to Regulate Centrosome Integrity. *Mol Biol Cell* 2007; 11:11.
164. Lejeune E, Bortfeld M, White SA, Pidoux AL, Ekwall K, Allshire RC, Ladurner AG. The Chromatin-Remodeling Factor FACT Contributes to Centromeric Heterochromatin Independently of RNAi. *Curr Biol* 2007; 17:219-24. Epub 2007 Jul 5.
165. Huang J, Laurent BC. A Role for the RSC chromatin remodeler in regulating cohesion of sister chromatid arms. *Cell Cycle* 2004; 3:973-5. Epub 2004 Aug 8.
166. Huang J, Hsu JM, Laurent BC. The RSC nucleosome-remodeling complex is required for Cohesin's association with chromosome arms. *Mol Cell* 2004; 13:739-50.
167. Hsu JM, Huang J, Meluh PB, Laurent BC. The yeast RSC chromatin-remodeling complex is required for kinetochore function in chromosome segregation. *Mol Cell Biol* 2003; 23:3202-15.
168. Corona DF, Siriaco G, Armstrong JA, Snarskaya N, McClymont SA, Scott MP, Tamkun JW. ISWI regulates higher-order chromatin structure and histone H1 assembly in vivo. *PLoS Biol* 2007; 5:e232.
169. Pandita TK, Hunt CR, Sharma GG, Yang Q. Regulation of telomere movement by telomere chromatin structure. *Cell Mol Life Sci* 2007; 64:131-8.
170. Yu EY, Steinberg-Neifach O, Dandjinou AT, Kang F, Morrison AJ, Shen X, Lue NF. Regulation of telomere structure and functions by subunits of the INO80 chromatin remodeling complex. *Mol Cell Biol* 2007; 11:11.
171. Collins N, Poot RA, Kukimoto I, Garcia-Jimenez C, Dellaire G, Varga-Weisz PD. An ACF1-ISWI chromatin-remodeling complex is required for DNA replication through heterochromatin. *Nat Genet* 2002; 32:627-32. Epub 2002 Nov 18.
172. Bell SP, Stillman B. ATP-dependent recognition of eukaryotic origins of DNA replication by a multiprotein complex. *Nature* 1992; 357:128-34.
173. Falbo KB, Shen X. Chromatin remodeling in DNA replication. *J Cell Biochem* 2006; 97:684-9.
174. Osley MA, Tsukuda T, Nickoloff JA. ATP-dependent chromatin remodeling factors and DNA damage repair. *Mutat Res* 2007; 618:65-80.
175. Gaillard H, Fitzgerald DJ, Smith CL, Peterson CL, Richmond TJ, Thoma F. Chromatin remodeling activities act on UV-damaged nucleosomes and modulate DNA damage accessibility to photolyase. *J Biol Chem* 2003; 278:17655-63. Epub 2003 Mar 11.
176. Chai B, Huang J, Cairns BR, Laurent BC. Distinct roles for the RSC and Swi/Snf ATP-dependent chromatin remodelers in DNA double-strand break repair. *Genes Dev* 2005; 19:1656-61.
177. Yu Y, Waters R. Histone acetylation, chromatin remodelling and nucleotide excision repair: hint from the study on MFA2 in *Saccharomyces cerevisiae*. *Cell Cycle* 2005; 4:1043-5. Epub 2005 Aug 21.
178. Morrison AJ, Highland J, Krogan NJ, Arbel-Eden A, Greenblatt JF, Haber JE, Shen X. INO80 and gamma-H2AX interaction links ATP-dependent chromatin remodeling to DNA damage repair. *Cell* 2004; 119:767-75.
179. Gong F, Fahy D, Smerdon MJ. Rad4-Rad23 interaction with SWI/SNF links ATP-dependent chromatin remodeling with nucleotide excision repair. *Nat Struct Mol Biol* 2006; 13:902-7. Epub 2006 Oct 1.
180. Durr H, Hopfner KP. Structure-function analysis of SWI2/SNF2 enzymes. *Methods Enzymol* 2006; 409:375-88.
181. Ho L, Crabtree GR. Chromatin remodelling during development. *Nature*; 463:474-84.
182. Sgarra R, Zammiti S, Lo Sardo A, Maurizio E, Arnoldo L, Pegoraro S, Giaccotti V, Manfioletti G. HMG A molecular network: From transcriptional regulation to chromatin remodeling. *Biochim Biophys Acta*; 1799:37-47.
183. Rattner BP, Yusufzai T, Kadonaga JT. HMG N proteins act in opposition to ATP-dependent chromatin remodeling factors to restrict nucleosome mobility. *Mol Cell* 2009; 34:620-6.
184. Stros M. HMG B proteins: Interactions with DNA and chromatin. *Biochim Biophys Acta*; 1799:101-13.
185. Dunker AK, Cortese MS, Romero P, Iakoucheva LM, Uversky VN. Flexible nets. The roles of intrinsic disorder in protein interaction networks. *Febs J* 2005; 272:5129-48.
186. Dunker AK, Obradovic Z. The protein trinity--linking function and disorder. *Nat Biotechnol* 2001; 19:805-6.
187. Sugase K, Dyson HJ, Wright PE. Mechanism of coupled folding and binding of an intrinsically disordered protein. *Nature* 2007; 447:1021-5. Epub 2007 May 23.

188. Levy Y, Onuchic JN, Wolynes PG. Fly-casting in protein-DNA binding: frustration between protein folding and electrostatics facilitates target recognition. *J Am Chem Soc* 2007; 129:738-9.
189. Sandhu KS, Dash D. Dynamic alpha-helices: conformations that do not conform. *Proteins* 2007; 68:109-22.
190. Tompa P, Prilusky J, Silman I, Sussman JL. Structural disorder serves as a weak signal for intracellular protein degradation. *Proteins* 2008; 71:903-9.
191. Singh GP, Ganapathi M, Dash D. Role of intrinsic disorder in transient interactions of hub proteins. *Proteins* 2007; 66:761-5.
192. Singh GP, Dash D. Intrinsic disorder in yeast transcriptional regulatory network. *Proteins* 2007; 68:602-5.
193. Oldfield CJ, Meng J, Yang JY, Yang MQ, Uversky VN, Dunker AK. Flexible nets: disorder and induced fit in the associations of p53 and 14-3-3 with their partners. *BMC Genomics* 2008; 9:S1.
194. Uversky VN. Natively unfolded proteins: a point where biology waits for physics. *Protein Sci* 2002; 11:739-.
195. Uversky VN. What does it mean to be natively unfolded? *Eur J Biochem* 2002; 269:2-12.
196. Ward JJ, McGuffin LJ, Bryson K, Buxton BF, Jones DT. The DISOPRED server for the prediction of protein disorder. *Bioinformatics* 2004; 20:2138-9.
197. Shimizu K, Hirose S, Noguchi T. POODLE-S: web application for predicting protein disorder by using physicochemical features and reduced amino acid set of a position-specific scoring matrix. *Bioinformatics* 2007; 23:2337-8.
198. Dosztanyi Z, Csizmek V, Tompa P, Simon I. IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics* 2005; 21:3433-4.
199. Lindner R, Russell RB, Neduva V, Gibson TJ. GlobPlot: Exploring protein sequences for globularity and disorder. *Nucleic Acids Res* 2003; 31:3701-8.
200. Su CT, Chen CY, Hsu CM. iPDA: integrated protein disorder analyzer. *Nucleic Acids Res* 2007; 35:W465-72.
201. Prilusky J, Felder CE, Zeev-Ben-Mordehai T, Rydberg EH, Man O, Beckmann JS, Silman I, Sussman JL. FoldIndex: a simple tool to predict whether a given protein sequence is intrinsically unfolded. *Bioinformatics* 2005; 21:3435-8.
202. Xue B, Dunbrack RL, Williams RW, Dunker AK, Uversky VN. PONDR-FIT: A meta-predictor of intrinsically disordered amino acids. *Biochim Biophys Acta*; 1804:996-1010.
203. Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragozy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, Sandstrom R, Bernstein B, Bender MA, Groudine M, Gnirke A, Stamatoyannopoulos J, Mirny LA, Lander ES, Dekker J. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 2009; 326:289-93.
204. Ge H, Liu Z, Church GM, Vidal M. Correlation between transcriptome and interactome mapping data from *Saccharomyces cerevisiae*. *Nat Genet* 2001; 29:482-6.
205. Grigoriev A. A relationship between gene expression and protein interactions on the proteome scale: analysis of the bacteriophage T7 and the yeast *Saccharomyces cerevisiae*. *Nucleic Acids Res* 2001; 29:3513-9.
206. Bhardwaj N, Lu H. Correlation between gene expression profiles and protein-protein interactions within and across genomes. *Bioinformatics* 2005; 21:2730-8.
207. Rajapakse I, Perlman MD, Scalzo D, Kooperberg C, Groudine M, Kosak ST. The emergence of lineage-specific chromosomal topologies from coordinate gene regulation. *Proc Natl Acad Sci U S A* 2009; 106:6679-84.
208. Pant V, Mariano P, Kanduri C, Mattsson A, Lobanenko V, Heuchel R, Ohlsson R. The nucleotides responsible for the direct physical contact between the chromatin insulator protein CTCF and the H19 imprinting control region manifest parent of origin-specific long-distance insulation and methylation-free domains. *Genes Dev* 2003; 17:586-90.
209. Zhao Z, Tavosoidana G, Sjolinder M, Gondor A, Mariano P, Wang S, Kanduri C, Lezcano M, Sandhu KS, Singh U, Pant V, Tiwari V, Kurukuti S, Ohlsson R. Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nat Genet* 2006; 38:1341-7.
210. Gondor A, Rougier C, Ohlsson R. High-resolution circular chromosome conformation capture assay. *Nat Protoc* 2008; 3:303-13.
211. Buck MJ, Nobel AB, Liebig JD. ChIPOTle: a user-friendly tool for the analysis of ChIP-chip data. *Nucleic Acid Biol* 2005; 6:R97.
212. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 2003; 13:2498-504.
213. Xenarios I, Rice DW, Salwinski L, Baron MK, Marcotte EM, Eisenberg D. DIP: the database of interacting proteins. *Nucleic Acids Res* 2000; 28:289-91.
214. Bader GD, Donaldson I, Wolting C, Ouellette BF, Pawson T, Hogue CW. BIND--The Biomolecular Interaction Network Database. *Nucleic Acids Res* 2001; 29:242-5.
215. Hermjakob H, Montecchi-Palazzi L, Lewington C, Mudali S, Kerrien S, Orchard S, Vingron M, Roechert B, Roepstorff P, Valencia A, Margalit H, Armstrong J, Bairoch A, Cesareni G, Sherman D, Apweiler R. IntAct: an open source molecular interaction database. *Nucleic Acids Res* 2004; 32:D452-5.
216. Mishra GR, Suresh M, Kumaran K, Kannabiran N, Suresh S, Bala P, Shivakumar K, Anuradha N, Reddy R, Raghavan TM, Menon S, Hanumanthu G, Gupta M, Upendran S, Gupta S, Mahesh M, Jacob B, Mathew P, Chatterjee P, Arun KS, Sharma S, Chandrika KN, Deshpande N, Palvankar K, Raghavath R, Krishnakanth R, Karathia H, Rekha B, Nayak R, VishnuPriya G, Kumar HG, Nagini M, Kumar GS, Jose R, Deepthi P, Mohan SS, Gandhi TK, Harsha HC, Deshpande KS, Sarker M, Prasad TS, Pandey A. Human protein reference database--2006 update. *Nucleic Acids Res* 2006; 34:D411-4.
217. Rual JF, Venkatesan K, Hao T, Hirozane-Kishikawa T, Dricot A, Li N, Berriz GF, Gibbons FD, Dreze M, Ayivi-Guedehoussou N, Klitgord N, Simon C, Boxem M, Milstein S, Rosenberg J, Goldberg DS, Zhang LV, Wong SL, Franklin G, Li S, Albaladejo JS, Lim J, Fraughton C, Lammas E, Cevik S, Bex C, Lamesch P, Sikorski RS, Vandenhoute J, Zoghbi HY, Smolyar A, Bosak S, Sequerra R, Doucette-Stamm L, Cusick ME, Hill DE, Roth FP, Vidal M. Towards a proteome-scale map of the human protein-protein interaction network. *Nature* 2005; 437:1173-8.
218. Stelzl U, Worm U, Lalowski M, Haenig C, Brembeck FH, Goehler H, Stroedicke M, Zenkner M, Schoenherr A, Koeppen S, Timm J, Mintzlaff S, Abraham C, Bock N, Kietzmann S, Goedde A, Toksoz E, Droegge A, Krobitsch S, Korn B, Birchmeier W, Lehrach H, Wanker EE. A human protein-protein interaction network: a resource for annotating the proteome. *Cell* 2005; 122:957-68.
219. Ramani AK, Bunescu RC, Mooney RJ, Marcotte EM. Consolidating the set of known human protein-protein interactions in preparation for large-scale mapping of the human interactome. *Genome Biol* 2005; 6:R40.
220. Kolaczyk ED, SpringerLink (Online service). Statistical analysis of network data methods and models. Springer series in statistics. New York ; London: Springer, 2009:xiii, 386 p.
221. Shipra A, Chetan K, Rao MR. CREMOFAC--a database of chromatin remodeling factors. *Bioinformatics* 2006; 22:2940-4.
222. Gendler K, Paulsen T, Napoli C. ChromDB: the chromatin database. *Nucleic Acids Res* 2008; 36:D298-302.
223. Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 2006; 22:1658-9.
224. Dosztanyi Z, Csizmek V, Tompa P, Simon I. IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics* 2005; 21:3433-4. Epub 2005 Jun 14.
225. Romero P, Obradovic Z, Li X, Garner EC, Brown CJ, Dunker AK. Sequence complexity of disordered protein. *Proteins* 2001; 42:38-48.
226. Ohlsson R, Gondor A. The 4C technique: the 'Rosetta stone' for genome biology in 3D? *Curr Opin Cell Biol* 2007; 19:321-5.
227. Essien K, Vigneau S, Apreleva S, Singh LN, Bartolomei MS, Hannenhalli S. CTCF binding site classes exhibit distinct evolutionary, genomic, epigenomic and transcriptomic features. *Genome Biol* 2009; 10:R131.
228. Sandhu KS, Dash D. Conformational flexibility may explain multiple cellular roles of PEST motifs. *Proteins* 2006; 63:727-32.
229. Han JD, Bertin N, Hao T, Goldberg DS, Berriz GF, Zhang LV, Dupuy D, Walhout AJ, Cusick ME, Roth FP, Vidal M. Evidence for dynamically organized modularity in the yeast protein-protein interaction network. *Nature* 2004; 430:88-93.
230. Naoki Masuda NK. VIP-club phenomenon: Emergence of elites and masterminds in social networks. *Social Networks* 2006; 28:297-309.

**NB: Figures in the thesis are sketched based on information cited elsewhere in the text.**

