Principles underlying the organization and function of yeast genome

Arashdeep Singh

A thesis submitted for the partial fulfillment of

the degree of Doctor of Philosophy

Department of Biological Sciences

Indian Institute of Science Education and Research Mohali

Knowledge city, Sector 81, SAS Nagar, Manauli PO, Mohali 140306, Punjab, India.

August 2019

Declaration

The work presented in this thesis has been carried out by me under the guidance of Dr. Kuljeet Singh Sandhu at the Indian Institute of Science Education and Research Mohali. This work has not been submitted in part or in full for a degree, a diploma, or a fellowship to any other university or institute. Whenever contributions of others are involved, every effort is made to indicate this clearly, with due acknowledgement of collaborative research and discussions. This thesis is a bona fide record of original work done by me and all sources listed within have been detailed in the bibliography.

Arashdeep Singh

In my capacity as the supervisor of the candidate's thesis work, I certify that the above statements by the candidate are true to the best of my knowledge.

Dr. Kuljeet Singh Sandhu

Acknowledgements

I would like to express my sincere thanks to my PI, Dr. Kuljeet Singh Sandhu, who provided me an opportunity to perform my graduate research in his research group. His scientific ingenuity and novel scientific ideas will always inspire me for rest of my career in science. In addition to being a great researcher, he is also an excellent teacher and teaches the difficult concepts with ease. Apart from training me in genomics, bioinformatics, and computational biology, he also helped me to improve my presentation skills. Thank you sir for guiding and shaping my Ph.D research. Your vast knowledge and deep insights are indeed a treasure for anyone who is interested in understanding the regulation and evolution of genomes. I learned some of the most precious lessons from him which, I think, would not have been possible for me to learn otherwise.

I am also grateful to my thesis committee members; Dr. Rajesh Ramachandran and Dr. Shashi Bhushan Pandit, who reviewed my work time to time and provided me motivation and encouragement whenever I approached them.

I would also express my deepest gratitude to Dr. Shravan K. Mishra who guided me through the experimental results presented in this thesis. His critical thinking, scientific excellence, and humble nature will always motivate me to be a good scientist. Thank you sir for being kind, patient, and always discussing my project in depth whenever I approached for help. Discussions with you were always very productive and made the experiments more interesting for me. I surely have learned a lot from you.

I would also like to thank Dr. Lolitika Mandal, with whom I completed my thesis during BS-MS. It was indeed a great opportunity for me to attend your courses and work under your guidance. My sincere thanks is also due to Dr. Sudip Mandal for his critical inputs into my BS-MS thesis as well as for many sessions of interesting chats and discussions. A big thanks to Dr. Somdatta Sinha for being a great teacher as well as a mentor, and giving useful career advices.

My most sincere thanks also go to Dr. N. Sathyamurthy, the founding and former Director of IISER-Mohali for providing me an opportunity to do Ph.D. at IISER-Mohali. Thank you sir for nurturing one of the greatest places to do science in India during its budding days. I also thank Dr. Arvind, the current Director of IISER-Mohali, for his continuous support to assist the smooth passage of research at IISER-Mohali. I would also thank the Department of Biological Sciences (DBS) at IISER for organizing so many exciting sessions of the poster and oral presentations. It was indeed a great platform to exchange ideas and learn from our colleagues. I dually acknowledge the financial support from IISER-Mohali for funding my Ph.D. research and providing me travel grant to cover my registration fees for presenting a part of my work in one

of the scientific conferences at EMBL Germany. I also acknowledge the much useful financial support from DST-India and EMBL for participation in an international conferences.

It is very important to have positive working environment during the steep uphill journey of Ph.D. A big thank to my labmates; Meenakshi, Ken, Yachna, Harpreet, Monika, Mohan, and Jui, and many BS-MS students, who came to lab for projects, for maintaining a positive environment in the lab. Thanks to Meenakshi, a very good programmer, for organizing so many parties and making the beautiful photo collages of those times. I would like to thank Ken who is an extremely good computational biologist and discussions with him always helped me to better understand the fundamental concepts of the NGS analysis. Many thanks to Yachna, who has very good and deep understanding of biological systems and is always very humble and cheerful in dealing with everyone. It was indeed a great learning experience to discuss random scientific ideas with you. Also a big thank to Harpreet who often knew and told me about some of the coolest shortcuts and hacks for coding in R and Perl. Thank you Mohan, Monika (the curious one!) and Jui for being cheerful and happy people around.

My sincere thanks to Preeti, Ashutosh, Rivi, Priya, Paras, Deeksha, Nidhi, and Nitika for being awesome neighbors of our lab. I will always cherish the good time I had with you people. My special thanks to Preeti for her expert help in troubleshooting the Linux related issues at times.

I would also like to thank the people from Dr. Shravan's lab, who welcomed me wholeheartedly in their research group. My special thanks to Poulami, who is very cheerful person with great positive energy. Thank you for teaching me the experimental work and being always cool and light-minded. I would also thank the other members of this lab; Prashant, Karan, Rakesh, Sumanjeet, Kiran, Anupa, Poonam, and Praver, who were great colleagues and were always helpful. I will particularly remember lunch and dinner outings with Rakesh, Karan, Sumanjeet, and Praver. It was truly refreshing to be with you guys.

I would also thank my friends (outside the lab!) whose company immensely helped in keeping up with the personal and professional struggles of my Ph.D life at IISER Mohali. I would like to thank Kanika for being a good friend. Also, my sincere thanks go to Anshul, Sandeep, Ankit, Rajneesh, Swagatam, who are all lovely people and everyone was a unique piece (☺)! I am also thankful to my BS-MS friends; Dilraj, Nilmani, Keshav, Jithin, Sumit, Ankit, Vinay, Shashank, Nitish, Karishma, Manish, Manmeet, and Nikhil who provided me a great company throughout my stay at IISER-Mohali during BS-MS and later remained connected during Ph.D. My special thanks go to Jithin Paul, who, I believe, is a great life coach in addition to a brilliant mathematician in aspiration.

Lastly, my family provided me endless support and it's their love and blessings that helped me immensely to sail through my tough journey of Ph.D.

Synopsis

Despite the significant accumulation of genome-wide studies, the underlying principles governing organization and regulation of eukaryotic genome remains poorly understood. In this thesis, we addressed the following questions. (i) What constrains the three-dimensional organization of genome? (ii) What explains the genome-wide disconnect between binding and effect of chromatin regulators? We observed that the long-range convergence of genomic loci strongly associated with the spatial assortment of co-replicating and co-evolving domains through interactions among replication origins. Short-range intra-chromosomal clustering was constrained by the minimization of transcriptional noise of interacting genes, an observation that was evolutionarily conserved from prokaryotes to higher eukaryotes. In the second part of the thesis, we established that most chromatin regulators do not bind to their target promoters to regulate transcription, but to surveil the genome against genetic and epigenetic errors, presenting a simple explanation to the global discrepancy between binding and effect of chromatin regulators. The observations through this thesis deflated the popular claims that the co-expression of interacting genes is the major functional constraint shaping 3D genome architecture of yeast and that the chromatin regulators occupy promoter sites primarily for transcription regulation of genes.

Background

Part I: Understanding the 3D genome organization and regulation is one of the fundamental objectives of genome biology. The discovery of proximity ligation-based assay called 'chromosome conformation capture' (3C) (Dekker et al. 2002) and its derivatives including HiC (Dostie et al. 2006; Fullwood, Liu, Pan, Liu, Han, et al. 2009; Zhao et al. 2006) accelerated the growth of 3D genomics at an unprecedented rate in the past decade (de Wit and de Laat 2012). At the time this thesis work was proposed, relatively fewer studies had employed HiC to comprehensively analyze the 3D organization of eukaryotic genomes to understand the basic principles of chromatin folding (Dixon et al. 2012; Z. Duan et al. 2010; Feng et al. 2014; Le et al. 2013; Guoliang Li et al. 2012; Lieberman-Aiden et al. 2009; Mizuguchi et al. 2014; Rao et al. 2014). These studies provided the very first maps of long-range chromatin interactions in model experimental systems. Several studies attempted to explore the possible functional constraints which shaped the landscape of long-range intra and inter-chromosomal interactions among genes and popularly concluded co-expression and co-functionality as major underlying constraints (Chepelev et al. 2012; Diament, Pinter, and Tuller 2014; Homouz and Kudlicki 2013; Li et al. 2012; Sandhu 2012; Sandhu et al. 2012; Thévenin et al. 2014). However, these studies were limited by the fewer functional attributes tested and assumed genome-wide average scenario as control, which, as we argue through our observations, was not appropriate and unbiased approach.

Objective I: To delineate the functional variables best associated with inter- and intra-chromosomal interactions in the yeast genome and improvise on observations.

Part II: To avail the naked DNA to transcription factors, the chromatin undergoes chromatin remodeling through a dedicated repertoire of chromatin factors (Cairns 2009; Fry and Peterson 2001). However, deletion of a chromatin factor often does not alter the transcription of its bound target genes, a phenomenon known as global discrepancy or disconnect between binding and effect of chromatin regulatory factors (Lenstra and Holstege 2012). Functional redundancy due to the presence of paralogues genes can explain the differences between genome-wide locations and knock out studies of transcription factors (Gitter et al. 2009; W.-S. Wu and Lai 2015). Some studies have suggested condition-specific transcriptional effect of transcription factor binding to targets (Spivakov 2014). These observations are limited to transcription factors and have not been scrutinized for the chromatin factors. Despite their wide-spread implications in diverse genomic functions like transcription (B. Li, Carey, and Workman 2007), replication (MacAlpine and Almouzni 2013; Vincent, Kwong, and Tsukiyama 2008), and DNA repair (Luijsterburg and Van Attikum 2011; Morrison and Shen 2005; Price and D'Andrea 2013b), the possibility that the functions other than transcription might underlie observed disconnect between binding and effect of chromatin factors has not been explored. Instead, the community unsubstantially assumes that functional buffering through paralogues compensates the knockout effect of chromatin factors.

Objective II: To test the following hypotheses in order to explain the disconnect between chromatin factor binding and effect:

i) Functional buffering of chromatin factors through paralogues
ii) Condition-specific transcriptional roles of chromatin factors
iii) Non-transcriptional roles of chromatin factors

Principal findings of part I

Preferred domains of cis and trans interactions in yeast genome. Comparison of the genome-wide landscape of *cis* (intra-chromosomal) and *trans* (inter-chromosomal) interactions showed that yeast genome has preferred domains of *cis* or *trans* interacting regions. Interestingly, most of the genomic domains showing enrichment for *trans* or *cis* interactions were mutually exclusive. This suggested that the *cis* and *trans* chromosomal interactions might have evolved under different functional constraints in yeast genome, compelling us to perform separate analysis of *cis* and *trans* chromosomal interactions.

Distinct constraints for cis and trans interactions. Through the correlation analysis and PLSR analysis, we found that *trans* interaction frequency correlated strongly with the co-replication of the spatially interacting genes. On the other hand, *cis* interaction frequency strongly correlated

with co-fitness of the interacting genes. The distinction in the observed functional associations of *trans* and *cis* interactions aligned well with the observation of preferred domains of *trans* and *cis* interactions in yeast genome, reinforcing the idea of different functional constraints shaping the organization of *trans* and *cis* interactions. To further scrutinize the association of *trans* interaction frequency with co-replication of interacting loci, we compared the strength of co-localization among origins of replication against the random null model generated through bootstrapping. Interestingly, we found that early origins of replication showed non-random spatial clustering via *trans* interactions and late origins of replication, on the other hand, exhibited non-random spatial clustering via *cis* interactions, but not the vice versa. Consistently, we found that replication timing of the early replicating domains showed a stronger association with *trans* interaction frequency but not with *cis.* Similarly, replication timing of late replicating domains in yeast genome showed a stronger association with *cis* interaction frequency but not with *trans*. We found similar associations between co-varying mutation rates of early and late replication domains with *trans* and *cis* interaction frequency. Taken together, these observations highlight a model of 3D genome in yeast, where early and late replication factories are spatially segregated from each other via *trans* and *cis* chromosomal interactions.

Gene essentiality and minimization of expression noise constrained cis interactions. PLSR analysis showed that *cis* interaction frequency was maximally correlated with co-fitness of interacting genes. Additionally, we also observed that genes having greater fitness defect tend to have greater average *cis* interaction frequency. Furthermore, we observed that genes with extreme fitness defect such as lethality (i.e. the essential genes) were engaged in *cis* interactions of higher frequency as compared to the random null model. Upon closer inspection, we found that the essential gene clusters in yeast genome were engaged in short-range *cis* interactions of higher average interaction frequency as compared to non-essential genes. Based on these observations, we proposed that engagement of essential genes in short range *cis* interactions of higher average frequencies might represent a spatially restrained state of chromatin, which might be useful in minimization of their gene expression noise. Indeed, we observed that chromatin mobility data in yeast negatively correlated with *cis* interaction frequency and positively correlated with gene expression noise, supporting our proposal that short-range *cis* interactions of higher average frequency indeed might represent the spatially restrained state of chromatin associated with lower expression noise. The minimization of gene expression noise by high-frequency *cis* interaction was observed even in the genome organization of E.Coli and mESCs in addition to yeast, implying that the gene expression noise might have served as an important evolutionary selection force for the organization of 3D genome.

Principal findings of part II

Quantitative measurement of global disconnect. To develop a quantitative measure for the extent of disconnect for each factor, we defined effect-to-binding-ratio (EBR), which signifies the proportion of its binding targets which are transcriptionally affected upon deletion of the same factor. We observed that more than 70% of the chromatin factors have EBR values less than 10%, and maximum EBR of ~36% was observed for Tup1. These observations highlighted the widespread prevalence of disconnect between binding and effect of chromatin factors.

Presence of paralogues genes or condition-specific roles of chromatin factors did not explain the disconnect. We tested the presumption that the presence of paralogues genes of chromatin factors might account for the discrepancy between binding and effect by imparting functional buffering against their genetic deletions. We found that EBR of chromatin factors did not negatively correlate with the number of paralogues genes, indicating that chromatin factors having more number of paralogues genes did not show greater functional buffering as compared to factors having lesser number of paralogues genes. Additionally, the binding targets of chromatin factors that were co-occupied by their paralogues genes did not show greater tendency to be buffered against chromatin factor deletion as compared to the targets which were not co-occupied with paralogues genes. Further, EBR of chromatin factors paralogues scale negatively with the mean and variance of the average expression of their binding targets across multiple environmental conditions. These results showed that condition specificity of chromatin factors do not explain the discrepancy between their genome-wide location and effect.

Cell-cycle phase-specificity and functional distinction of affected and unaffected binding targets. Through the analysis of cell cycle expression pattern, we found that chromatin factors with or without effect did not exhibit cell cycle phase-specificity. However, we found that binding targets which were transcriptionally affected upon the deletion of chromatin factors were enriched for G2/M phase related metabolic genes. On the contrary, the unaffected binding targets tend to be G1/S phase specific showing the functional enrichment for chromatin organization, cell division, organelle fission etc. The observed functional distinction between affected and unaffected targets prompted us to explore whether the chromatin factors also had distinct functional preferences.

Functional distinction of chromatin factors with low and high EBR. We inferred the functions of chromatin factors through the gene ontology analysis of their protein interaction partners. We found that ontology terms related to genome integrity such as DNA repair, DNA damage response, DNA replication, and DNA metabolism form a separate cluster, while EBR and transcription clustered separately. The chromatin factors associated with genome integrity related function showed significantly lesser EBR as compared to transcription-associated chromatin

factors. These results suggested that chromatin factors lacking effect on transcription of their binding targets might implicate in genome surveillance. Furthermore, unaffected binding target genes had a significantly greater accumulation of DNA repair-related chromatin factors as compared to affected binding targets. Altogether, these observations suggested that chromatin factor binding at the unaffected binding targets might be implicated in mitigation of genetic and epigenetic errors.

Genetic and epigenetic errors at affected and unaffected binding targets. To test if the unaffected binding targets are buffered against genetic and epigenetic errors, we measured the accumulation of single nucleotide variants (SNVs), DNA double-strand breaks, *cis* variability and cryptic transcription at affected and unaffected binding targets. Interestingly, we found the unaffected target genes of chromatin factors had significantly fewer genetic and epigenetic errors as compared to transcriptionally affected target genes, possibly owing to the greater deposition of DNA repair-related factors at unaffected targets. To this end, we used the yeast genetics experiments to score the occurrence of mutagenesis in deletion strains of two chromatin factors; the Dot1 and Swc5. For this purpose, we inserted the Ura3 gene downstream of the binding locations of Dot1 and Swc5 in their corresponding deletion strains and wild-type backgrounds. We found that compared to the WT strains, Δdot1 and Δswc5 exhibited greater occurrence of spontaneous mutations in URA3 gene as assayed by their growth on FOA plates. Further, we found that the chromatin factors having lower EBRs had a greater number of negative genetic interactions and exhibited greater fitness defect in the data from chemical genomics screens. Taken together, these results indicated that chromatin factors lacking apparent effects on transcription might adopt essential roles for survival under different genetic or environmental conditions, not necessarily by regulating transcription, but through maintenance of genome integrity.

Conclusions

We performed comprehensive analyses to understand the principles underlying the 3D organization and function of the yeast genome. We have shown that: i) The organization of *cis* and *trans* chromosomal interactions is shaped by distinct functional constraints. While *trans* interactions strongly associated with the coordination of early replication, the *cis* interactions associated with the coordination of late replication, gene essentiality, and minimization of expression noise. ii) Widespread lack of transcriptional effects upon chromatin factor deletion cannot be accounted for by the presence of paralogues genes or the condition-specific roles of chromatin factors. Instead, we presented the evidence that the genome-wide binding of chromatin factors was important for genome surveillance, not necessarily the transcription.

This thesis is based on the following two manuscripts.

- **Singh A**, Bagadia M, Sandhu KS. Spatially coordinated replication and minimization of expression noise constrain three-dimensional organization of yeast genome. DNA Research. 2016; 23(2) : 155- 169.*doi*:10.1093/*dnares*/*dsw*005.
- **Singh A**, Choudhuri P, Keerthivasan C. Raanin, Mishra SK, Sandhu KS. A study of genome-wide disconnect between binding and effect of chromatin factors in yeast *(Manuscript)*.

Other publications.

- Bagadia M, **Singh A**, Singh Sandhu K. Three Dimensional Organization of Genome Might Have Guided the Dynamics of Gene Order Evolution in Eukaryotes. Genome Biology and Evolution. 2016; 8(3) : 946-954.*doi*: 10.1093/*gbe*/*evw*050.

List of tables

Table 1 Chromatin modifications and their functions. 12
Table 2 The chromatin remodeling complexes and their subunits..... 16
Table 3 A summary of 3C-based techniques 48
Table 4 Proximity ligation independent techniques to study 3D genome. 58
Table 5 Datasets for multivariate analysis 75
Table 6 Other datasets for the analysis of chromatin interactions..... 76
Table 7 Datasets for the analysis of discrepancy between binding and effect of chromatin factors in yeast. 77
Table 8 List of primers used for chromosomal tagging..... 90
Table 9 List of yeast strains used in this work. 91

List of abbreviations

ARS……..Autonomously replicating sequences

CT……..... Chromosomal territory

CUTs……....Cryptic unstable transcripts

DSBs…… Double strand breaks

EBR……..Effect-to-binding ratio

FOA……. 5-Fluoroorotic *acid*

GO……....Gene ontology

LAD……. Lamina associated domains

PCA……. Principal component analysis

PLSR……Partial least square regression

PPI……Protein-protein interactions

SGD……. *Saccharomyces* genome database

SNVs……....Single nucleotide variations

TAD…….... Topologically associated domains

TSS…….. Transcription start site

List of figures

Figure 1.1 Methylation modification of cytosine in DNA. 3
Figure 1.2 DNA methylation and imprinting. 4
Figure 1.3 'Beads on string' model of interphase chromatin. 6
Figure 1.4 The solenoid and zigzag model of 30nm chromatin fibre. 7
Figure 1.5 Some commonly studied covalent histone modifications. 9
Figure 1.6 Protein domains of ATP dependent chromatin remodeling enzymes. 18
Figure 1.7 Transcription initiation at repressed gene. 19
Figure 1.8 Chromatin remodeling at the site of transcription initiation. 20
Figure 1.9 Summary of chromatin remodeling during transcription elongation. 22
Figure 1.10 Distribution of GC content and gene-density across mouse chromosome 1. 26
Figure 1.11 Classification of the repeat elements in the genome. 30
Figure 1.12 Random diffusion vs. territorial model of chromosomal organization. 35
Figure 1.13 CT-ICD & CT-IC models of interphase chromosome organization. 37
Figure 1.14 Organization of heterochromatin in yeast and mammals. 41
Figure 1.15 Dynamic gene positioning with respect to nuclear periphery. 43
Figure 1.16 A schematic representation of 3C assay. 44
Figure 1.17 A schematic representation of HiC. 46
Figure 1.18 Summary of steps involved in commonly used C-based techniques. 47
Figure 1.19 Chromatin looping involved at H19/Igf2 locus. 50
Figure 1.20 Chromatin interaction pattern uncovered by 4C. 51
Figure 1.21 Contact map of HiC data. 53
Figure 1.22 Representation of systematic biases in HiC data. 55
Figure 1.23 Summary of chromatin and genome organization in 3D. 59
Figure 1.24 Interactions among genes anchored at transcription factory. 61
Figure 1.25 Possible modes of gene clustering proposed by different studies. 63
Figure 1.26 Disconnect between binding and effect of the chromatin modifiers. 70
Figure 2.1 Comparison of raw and normalized HiC data. 79
Figure 2.2 A schematic of workflow for the analysis of chromatin interaction data. 86
Figure 2.3 Schematic of URA3-based reporter and FOA assay 89
Figure 3.1 cis vs. trans interactions in yeast genome 95
Figure 3.2 The number of cis and trans interactions as a function of genomic coordinates. 96
Figure 3.3 Preferred domains of cis and trans interactions in yeast chromosome 4. 96
Figure 3.4 Correlation analysis of trans and cis interactions. 97
Figure 3.5 Variable map of PLSR analysis of the trans and cis interactions. 98
Figure 3.6 Robustness of the functional associates of trans and cis frequency. 99
Figure 3.7 Spatial clustering of replication origins. 101
Figure 3.8 Spatial segregation of early and late replication. 102
Figure 3.9 Covariance of mutation rate via cis and trans interactions 103
Figure 3.10 Association between nucleosome occupancy and chromatin interactions. 104
Figure 3.11 Nucleosome occupancy of cis-rich and trans-rich genomic domains. 104
Figure 3.12 Nucleosome occupancy on and around early and late origins of replication. 105
Figure 3.13 Spatial interactions among essential genes 106
Figure 3.14 Patterns of cis interactions of essential genes clusters. 107
Figure 3.15 cis interactions and gene expression noise. 108
Figure 3.16 Overlap between HindIII and EcoRI libraries of HiC data. 109
Figure 3.17 Chromatin mobility, cis interaction frequency and gene expression noise. 110

Figure 3.18 Validation of association between chromatin mobility and cis interaction frequency. 111
Figure 3.19 Peaks of higher average cis interaction frequency and the gene expression noise. 112
Figure 3.20 cis interactions and over expression toxicity. 112
Figure 3.21 cis interaction profile of essential genes in E.coli. 113
Figure 3.22 Expression noise and cis interaction frequency in E.coli and mESCs. 114
Figure 3.23 RNA-polII tethered interactions of high and low noise genes. 114
Figure 3.24 Quantifying the global disconnect between binding and effect of chromatin factors. 116
Figure 3.25 The distribution of Pearson's correlation coefficients between binding and effect of chromatin factors. 117
Figure 3.26 Association between disconnect and functional buffering. 118
Figure 3.27 Stress responsiveness and the disconnect. 119
Figure 3.28 Stress responsiveness of affected and unaffected targets. 119
Figure 3.29 Chromatin modifications of binding targets during yeast stress response 120
Figure 3.30 Enrichment of cell-cycle regulated genes among binding targets. 121
Figure 3.31 Analysis of cell cycle specificity of chromatin factors and their targets. 122
Figure 3.32 Gene ontology analysis of the binding targets. 123
Figure 3.33 EBR of chromatin factors from Venters et.al's classification 124
Figure 3.34 The GO analysis of protein interaction partners of chromatin factors. 125
Figure 3.35 Distinction among factors deposited at positively-affected, negatively-affected and unaffected binding targets 126
Figure 3.36 Analysis of the genetic errors at the binding targets. 127
Figure 3.37 Analysis of cryptic transcription and gene essentiality at the binding targets. 128
Figure 3.38 Fitness defects of chromatin factor deletions. 129
Figure 3.39 Selection of targets for FOA assay. 130
Figure 3.40 Results of FOA assay for Dot1 and Swc5. 131
Figure 4.1 A model for functional constraints underlying the 3D organization of yeast genome. 140
Figure 4.2 A model explaining the genome-wide disconnect of chromatin factors. 141

Table of Contents

Chapter 1 Introduction 1
1.1 Epigenetics and DNA methylation 3
1.2 Chromatin and its organization 5
1.2.1 Discovery of histones and 'beads on string' model of chromatin organization 5
1.2.2 The 30nm chromatin fibre and the solenoid model 6
1.2.3 Chromonema and higher levels of chromatin folding 7
Major highlights of the chromatin organization 8
1.3 Post-translational chromatin modifications 8
1.3.1 Histone acetylation 9
1.3.2 Histone methylation 10
1.3.3 Histone phosphorylation 11
1.3.4 Combinatorial complexity of chromatin modifications and the histone code 14
1.4 ATP dependent chromatin remodeling enzymes 15
1.4.1 SWI/SNF family remodelers 15
1.4.2 ISWI family remodelers 16
1.4.3 CHD family remodelers. 17
1.4.4 INO80/SWR1 family remodelers 17
1.5 Chromatin remodeling 18
1.5.1 Chromatin remodeling during transcription 19
1.5.1 Chromatin remodeling in DNA repair 22
1.5.2 Chromatin remodeling in DNA replication 24
1.6 Organization of genome in 1D 25
1.6.1 Non-random distribution of GC content across the genome 25
1.6.2 Gene dense and gene-poor regions 26
1.6.3 Repeat elements and their distribution in the genome 27
1.6.4 Distribution of CpG islands and DNA methylation 31
1.6.5 Gene clusters 32
1.7 Genome organization in 3D 34
1.7.1 Chromosome territories 34
1.7.2 Gene positioning within chromosomal territories 36
1.7.3 Organization of heterochromatin 38
1.7.4 The positioning of chromosome territories w.r.t nuclear lamina 41
1.7.5 Gene repositioning to accommodate transcriptional changes 42
1.8 Development and evolution of 3C based techniques 44
1.8.1 Circular Chromosome Conformation Capture (4C) 45

1.8.2 Chromosome Conformation Capture Carbon Copy (5C) 45
1.8.3 HiC or the '*all vs. all*' method 45
1.8.4 Chromatin Interaction Analysis using Paired-End Tag sequencing (ChIA-PET). 46
1.8.5 Other variants of HiC 47
1.9 Impact of 'C' based methods on the understanding of 3D genome 48
1.9.1 3C experiments and chromatin looping 48
1.9.2 The 4C & the discovery of intra- and inter-chromosomal interaction hubs 50
1.9.3 HiC and the 'topologically associated domains' (TADs) 52
1.9.4 Systematic biases in HiC data 55
1.10 Comparison between 3C-methods and microscopy 55
1.10.1 Discrepancy. 56
1.10.2 Reconciliation 56
1.11 Drifting away from proximity ligation. 57
1.11.1 Genome Architecture Mapping (GAM). 57
1.11.2 Split-Pool Recognition of Interactions by TAG Extension (SPRITE) 57
1.12 Introduction to the project I 60
1.12.1 Spatial co-localization of genes 60
1.12.1.1 Transcription 60
1.12.1.2 Replication 61
1.12.1.3 Other possible modes of gene clustering 62
1.12.2 Principles of large-scale organization of 3D genome are not fully understood 62
1.12.3 The integrative analysis in 3D genomics 63
1.12.4 The 3D organization of yeast genome 64
1.12.5 Constraints underlying 3D genome organization of yeast genome 65
1.13 Objectives of the project I 65
1.14 Introduction to project II 66
1.14.1 Investigation of histone code hypothesis 66
1.14.1.1 Genome-wide location of chromatin modifications 66
1.14.1.2 Genome-wide location of chromatin factors 67
1.14.1.3 Genome-wide transcriptional perturbation upon deletion of chromatin modifiers 68
1.14.2 The disconnect between binding and effect of chromatin factors 69
1.14.3 Determinants of genome-wide disconnect are not fully understood 70
1.15 Objectives of project II 71
Chapter 2 Materials and methods 73
2.1 Datasets 75
2.2 Data processing and statistical methods 78

2.3 Experimental methods 87
Chapter 3 Results 93
3.1 Results of the project I 95
3.1.1 Genome-wide landscape of *cis* and *trans* interactions in the yeast genome 95
3.1.2 Correlation analysis of *cis* and *trans* interactions 97
3.1.3 Spatial co-localization of origins of replication 100
3.1.4 Spatial segregation of early and late replication factories 101
3.1.5 Association between mutation rates and the chromatin interactions 102
3.1.6 Association between nucleosome occupancy and chromatin interactions 103
3.1.7 3D interactions among essential genes 106
3.1.8 Gene expression noise and *cis* interaction frequency 108
3.1.9 *cis* interaction frequency and toxic genes 112
3.1.10 Evolutionary conservation of noise-minimization *via* chromatin interactions 112
3.1.11 The possible mechanism of noise minimization through spatial interactions 114
3.2 Results of project II 116
3.2.1 Quantification of the disconnect between chromatin factors binding and effect 116
3.2.2 The analysis of functional buffering 117
3.2.3 Analysis of condition-specific roles of chromatin factors 118
3.2.4 Functional analysis of affected and unaffected binding targets 121
3.2.5 Functional distinction of chromatin factors with high and low EBR 123
3.2.6 The distinction among chromatin factors deposited at affected and unaffected targets 125
3.2.7 Analysis of genetic errors at binding targets. 126
3.2.8 Analysis of cryptic transcription (epigenetic errors) at binding targets 127
3.2.9 Fitness defects of chromatin factor deletions 129
3.2.10 Experiment to test the role of chromatin factors in genome integrity 130
Chapter 4 Discussion and conclusions 133
4.1 Functional and evolutionary constraints for 3D genome organization of the yeast 135
4.2 Analysis of genome-wide disconnect of chromatin factors in yeast 137
4.3 Conclusions 140
Appendix A 142
Appendix B 150
Appendix C 151
References 153

Chapter 1 Introduction

Literature review

1.1 Epigenetics and DNA methylation

The term 'epigenetics' was coined by Waddington in 1942 and was originally used to denote unknown phenomena which resulted in the development of different tissues and organs from the embryo despite the identical genetic information (Allis and Jenuwein 2016; Felsenfeld 2014; Waddington 1942). Decades after the rediscovery of Mendel's laws of inheritance (see for review Lenay 2000) the discovery of position effect variegation in *Drosophila* (Muller and Altenburg 1930), the transposable elements in maize (Mcclintock 1956), and imprinting in mouse (Barton, Surani, and Norris 1984; McGrath and Solter 1984; Surani, Barton, and Norris 1984) pointed towards the non-Mendelian form of inheritance. The discovery of X-chromosome inactivation in mouse (Lyon 1961; Ohno, Kaplan, and Kinosita 1959) further showed that the identical genetic material can be regulated by the additional information, possibly layered on the top of DNA sequence (reviewed in Allis and Jenuwein 2016; Felsenfeld 2014; Goldberg, Allis, and Bernstein 2007). The modern-day usage of the term 'epigenetics' refers to the modifications of the genome which can change the gene expression without affecting the underlying DNA sequence (Gayon 2016; Tollefsbol 2011). The research in past few decades have shown that epigenetic phenomena often involved the methylation of specific DNA sequences or the attachment of chemical modifications and remodeling of chromatin structure through the action of chromatin modifying enzymes (Allis and Jenuwein 2016; Felsenfeld 2014; Goldberg, Allis, and Bernstein 2007; Handel, Ebers, and Ramagopalan 2010; Handy, Castro, and Loscalzo 2011; Holliday 2006).

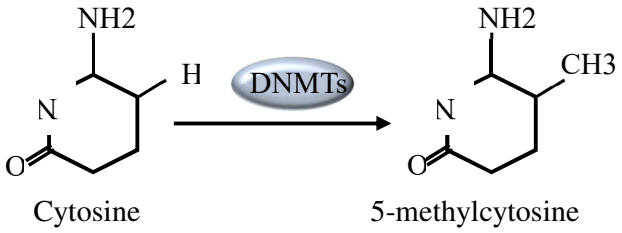

Figure 1.1 Methylation modification of cytosine in DNA.

Cytosine in the CpG residues can be methylated at its 5th position by the action of DNA

The methylation was discovered at the cytosine base of CpG residues in DNA in 1948 (Hotchkiss 1948) (figure 1.1). The initial proposal of involvement of DNA methylation in epigenetic regulation was made in 1975 by Holliday and Pugh who linked the random X-chromosome inactivation with the different levels of methylation in its maternal and paternal copies (R Holliday and Pugh 1975; also reviewed in Allis and Jenuwein 2016; Robin Holliday 2006). Soon DNA methylation was shown to be involved in transcriptional repression (Aharon Razin and

Riggs 1980; also reviewed in Allis and Jenuwein 2016) and was implicated the regulation of the genomic imprinting and developmental reprogramming (Brannan I and Bartolomei 1999; Elhamamsy 2017; C. Li et al. 2018; Lim and Maher 2010; Messerschmidt, Knowles, and Solter 2014; Paulsen and Ferguson-Smith 2001; Smith and Meissner 2013; Tycko 1997). Perhaps the best studied example of genomic imprinting involving DNA methylation is *Igf2* and *H19* genes (figure 1.2) which are expressed from paternal and maternal copies, respectively (Bell and Felsenfeld 2000) and were shown to be regulated *via* differential methylation of imprinting control region located ~2kb upstream of *H19* locus (Bell and Felsenfeld 2000; Hark et al. 2000; Kanduri et al. 2000; Lenay 2000).

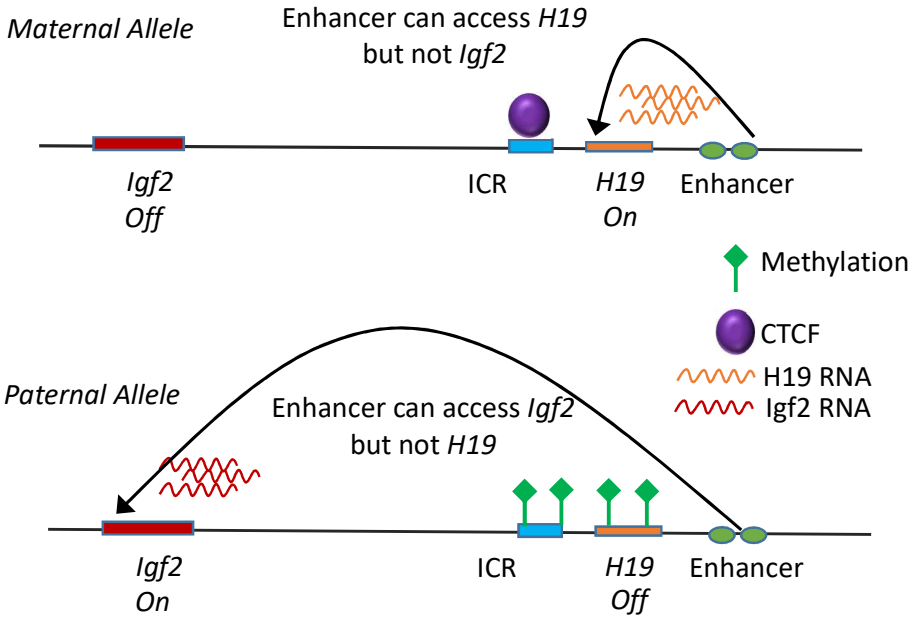

Figure 1.2 DNA methylation and imprinting.

Paternal allele (bottom) is methylated at ICR and *H19*, which prevents CTCF binding and block enhancer access of *H19* resulting in its repression. Maternal allele has unmethylated ICR, which can bind to CTCF and block the access of enhancer to *Igf2* resulting in its repression.

In addition to imprinting, the epigenetic regulation of gene expression is also linked with development, disease, and ageing (Ghavifekr Fakhr et al. 2013; Huh et al. 2013; Jin, Li, and Robertson 2011; Jones and Takai 2001; Kulis and Esteller 2010; Lim and Maher 2010; Saldanha and Watanabe 2015; Unnikrishnan et al. 2019).

The *de novo* DNA methylation in mammalian genomes is carried out by DNMT3a and DNMT3b DNA methyltransferases (DNMTs). Once the methylation patterns are established, they are stably maintained through the mitotic divisions by the activity of maintenance enzyme called DNMT1

(for review: Lee, Hore, and Reik 2014). The methylation patterns of one generation are rarely transmitted to the next generation because of their erasure in primordial germ cells (PGCs) and early embryo (Cowley and Oakey 2012; H. J. Lee, Hore, and Reik 2014; Seisenberger et al. 2012). But certain genomic regions are shown to be resistant to the erasure of methylation marks and can carry the epigenetic memory of one generation to the next generation (Daxinger and Whitelaw 2012). Though rare, a few examples of such escaper regions have been identified which include the retro-transposons such as intracisternal A particles (IAP) and certain CpG islands which are termed as 'variably erased' CpG islands (Cowley and Oakey 2012; H. J. Lee, Hore, and Reik 2014; Seisenberger et al. 2012).

Besides DNA methylation, the epigenetic regulation of gene expression has been also attributed to the covalent modification of histones (Ellis, Atadja, and Johnstone 2009; M. Lawrence, Daujat, and Schneider 2016; Quina, Buschbeck, and Di Croce 2006; Vaissière, Sawan, and Herceg 2008) and lately to the 3D organization chromatin inside the nucleus (D. Jost and Vaillant 2018; Kagohara et al. 2018; Mishra and Hawkins 2017). Since the major focus of this thesis is to understand the role of chromatin remodelers and 3D genome organization in the regulation of yeast genome, the detailed literature pertaining to these two fields is reviewed in the following sections review.

1.2 Chromatin and its organization

Historically, the organization of chromatin was studied using microscopy-based methods which were complemented by the biochemical and molecular biology-based assays towards the latter half of the 20th century. Some of the key insights obtained from this research are briefed below.

1.2.1 Discovery of histones and 'beads on string' model of chromatin organization

One of the first steps towards the understanding of the chromatin organization was taken in 1884 when Albrecht Kossel discovered a special class of positively charged proteins of alkaline nature and presented evidence for their DNA binding abilities (mentioned in Doenecke and Karlson 1984), naming them histones. Histones were later discovered to associate with DNA molecules to form chromatin fibre of 11nm (Chaffey 2003). The evidence about the physical nature of chromatin fibre came from the observations of repetitive patterns in the X-ray diffraction studies (Spencer et al. 1962) which raised the speculations of the subunit/repetitive structure of interphase chromatin. With the use of electron microscopy (EM) techniques, it was found that repetitive pattern in chromatin resembled the 'beads-on-string' arrangement in the interphase nucleus of rat thymus and chicken erythrocytes (Olins and Olins 1974). The contemporary biochemical assays of interphase chromatin involving the chemical crosslinking-based studies of histones supported the repetitive structure of chromatin (Kornberg 1974) and established that the histones assembled in octameric configuration to associate with DNA (Kornberg 1974; Thomas and Kornberg 1975).

The histone octamers were assembled from the association of two copies of each of the H2A, H2B, H3, and H4 and had a diameter of 125Å (E Morton Bradbury 1989). The positively charged surface of histone octamers provided an ideal substrate for wrapping off the negatively charged DNA, and the resulting structures which constitutred the 'beads-on-string' pattern (fugure 1.3) were termed as nucleosomes (E Morton Bradbury 1989). DNA formed nearly 1.65 turns around the nucleosome encompassing ~147 nucleotides and formed a linker region of around 20 nucleotides between two consecutive nucleosomes (E Morton Bradbury 1989) (figure 1.3). Histone-1, which was called as linker histone, bound and locked the DNA at the start and the exit of DNA coiling around the nucleosome, thereby tightly fixing the DNA onto the nucleosome (Bednar et al. 1998).

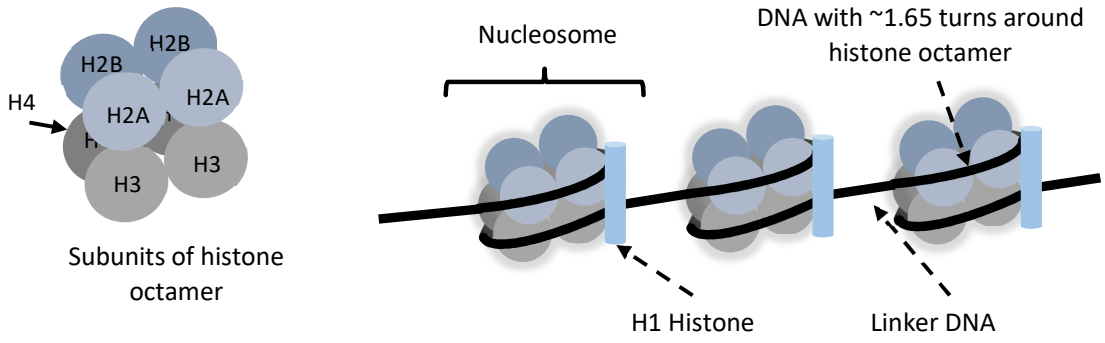

Figure 1.3 'Beads on string' model of interphase chromatin.
On the left is shown the subunits of histone octamer, which is assembled by two copies of each of H2A, H2B, H3, H4. On the right is representation of the organization of DNA around histone octamer to form nucleosomes, which is equivalent to the 11nm chromatin fibre.

1.2.2 The 30nm chromatin fibre and the solenoid model

To understand the higher order folding of 11nm chromatin fibre, several studies investigated the *in vitro* preparations of chromatin from interphase nucleus using electron microscopy (Finch and Klug 1976; Woodcock, Frado, and Rattner 1984; Worcel, Strogatz, and Riley 1981; also reviewed in Robinson, Fairall, Huynh, & Rhodes, 2006). These studies proposed that 11nm chromatin fibre was further coiled in a helical manner and resulted in a 30nm chromatin fibre. The 30nm chromatin fibre was proposed to be assembled through 'solenoid' (the two-start helix) (Robinson et al. 2006) or 'zigzag' (the one-start helix) model (Schalch et al. 2005). Both models implied the supercoiling of 11nm chromatin fibre in a helical manner but differed with respect to the arrangement of linker DNA. In the 'solenoid' model, linker DNA bent to assume the tangential position along the surface of the solenoid axis which made the consecutive nucleosomes as neighbours. In the 'zigzag' model, the linker DNA was straight (figure 1.4) and alternative

nucleosomes were neighbours in the folded configuration (described in Robinson et al., 2006; Tremethick, 2007; Woodcock et al., 1984).

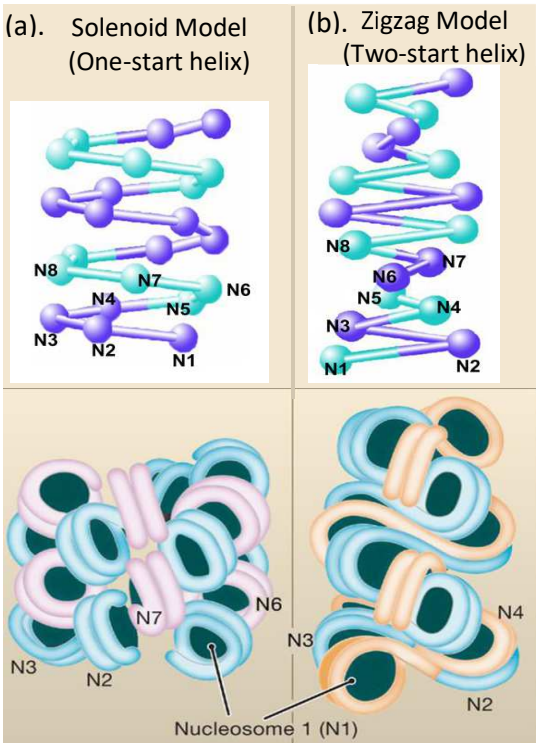

Figure 1.4 The solenoid and zigzag model of 30nm chromatin fibre.

The solenoid model is one-start helix where N^{th} nucleosome comes in physical proximity of $N+5^{th}$ nucleosome. Blue and magenta colors indicate the two alternative gyres of the helix in the lower panel.

b) The zigzag model is a two-start helix and Nth nucleosome is the neighbor to $N+1^{st}$ nucleosome. Blue and orange colors indicate the alternative nucleosome in the lower panel.

The picture is modified from Guohong Li and Reinberg (2011) with permission.

1.2.3 Chromonema and higher levels of chromatin folding

To understand chromatin structure beyond 30nm chromatin fibre, the ultra-structure of metaphase chromosome was subjected to extensive investigations and scrutiny. It was concluded that metaphase chromosome consisted of either a chromatin fibre of 200-300nm diameter which folded in helical manner and coexisted with radial loops to achieve the metaphase topology (Rattner and Lin 1985), or consisted of hierarchically folded fibres of multiple diameters involving the formation of 120, 240, 400-500, and 800-1,000Å chromatin fibres (A. S. Belmont, Sedat, and Agard 1987; Kireeva et al. 2004). These findings proposed the existence of 'chromonema', the superstructures made up of ~100-130nm chromatin fibre (Andrew S. Belmont

& Bruce, 1994; Kireeva et al., 2004; Sedat & Manuelidis, 1977; mentioned in Ou et al., 2017). However, the technological advancements in the state of art facilities, particularly the cryo-electron microscopy (cryo-EM) and adaptation of EM to *in vivo* systems led to the observations that did not support the existence of 30nm chromatin fibre under *in vivo* conditions. The earliest observation dated back to 1986, when no evidence was found for the existence of higher-order folding of chromatin and 11nm chromatin fibre was found to be the major form of chromatin constituting the metaphase chromosome in CHO and HeLa cell lines (Bancaud et al. 2009). Similar conclusions were drawn by other studies which showed that interphase chromatin behaved as loose thread consistent with 11nm chromatin fibre (Eltsov et al. 2008; Salmon et al. 2016). More recently, electron microscopy tomography (EMT) was coupled with better chromatin staining methods (Ou et al., 2017), which concluded that chromatin fibre predominantly existed in 5-24nm diameter range throughout the cell cycle and can be packed at different concentration densities depending on its activity and cell cycle phases (Ou et al., 2017). These observations have changed the overall understating and interpretation of chromatin folding.

Major highlights of the chromatin organization

(i) Linear DNA associates with histone octamer to form a higher order chromatin fibre of 11nm in diameter, which resembles the 'beads on string model'.

(ii) 11m chromatin fibre folds hierarchically to form 30nm 'solenoid' or 'zigzag' model of chromatin fibre, which can fold further into 100-300nm 'chromonema'.

(iii) The existence of higher order superstructures such as 'solenoids' and 'chromonema' is somewhat disputed based on observations made through of cryo-EM based studies.

1.3 Post-translational chromatin modifications

In 1950, even before the discovery of nucleosomes, Stedman and Stedman had proposed that the presence of histones in the DNA could result in the general repression of transcription (Stedman and Stedman 1950). Further research indicated the role of histones in the regulation of gene expression was not as straightforward as a general repressor and histones were found to be post-translationally modified to alter the DNA accessibility (Kouzarides 2007; Lawrence, Daujat, and Schneider 2016; Zhang and Reinberg 2001). The first report of post-translational modifications of histones was presented by Allfrey in the 1960s, who discovered the acetylation and methylation of histones and showed that these modifications have an influence on the RNA synthesis by the genome (Allfrey, Faulkner, and Mirsky 1964). Over the following years, many different types of histone modifications have been discovered (figure 1.5). Ever since their discovery, the scientists have undertaken tremendous efforts to understand the mechanism and detailed functional

characterization of these chemical modifications (reviewed in Bannister and Kouzarides 2011; Kouzarides 2007; M. Lawrence, Daujat, and Schneider 2016). Some of the common chromatin modifications are reviewed in the following subsections.

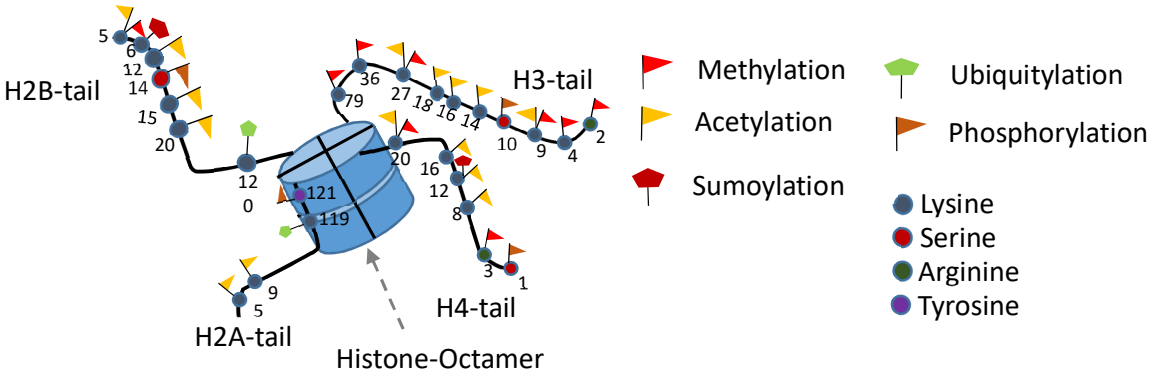

Figure 1.5 Some commonly studied covalent histone modifications.

N-terminal tails of histones bear various covalent modifications (colored flags).The colored dots on histone tails represent amino acid residues and the attached numbers indicate the position of the residues. Detailed citations for these modifications are given in table 1

1.3.1 **Histone acetylation**

Histone acetylation is one of the most well-characterized post-translational histone modifications. Even though the initial proposal for the involvement of histone acetylation in the regulation of RNA synthesis was made in 1964 (Allfrey, Faulkner, and Mirsky 1964), it took more than two decades to establish a clear role of histone acetylation in transcriptional regulation. It was speculated that histone acetylation involved the specific lysine residues of N-terminal tail of histones which imparted the negative charge to the histones, thereby weakening the interactions between DNA and nucleosomes (Allfrey, Faulkner, and Mirsky 1964; Brownell et al. 1996). Several studies provided indirect evidence for the indispensability of histone acetylation in gene activation. For example, Grunstein and colleagues showed that N-terminals of H4 histones were involved in the specific repression of silent mating loci in yeast (Kayne et al. 1988). Another study from the same group indicated that the N-terminal tails of H4 histones were required for the transcriptional activation of GAL1 and PHO5 promoters and ascribed this effect to the acetylation sites present on the N terminals of H4 (Durrin et al. 1991). A similar role of acetylable residues in the N-terminal tail of H3 was also discovered (Thompson, Ling, and Grunstein 1994).

The process of histone acetylation is carried out by a special class of enzymes called histone acetyltransferases (HATs), which catalyzes the transfer of acetyl group from acetyl coenzyme A to the specific lysine residues of the N terminal of histone H3/H4 (Loidl 1994). In 1996, Brownell et al. discovered a histone acetyltransferase (HAT) in Tetrahymena which had a strong homology

with the yeast transcriptional co-activator; Gcn5p (Brownell et al. 1996). The authors of this work also established the HAT activity of the Gcn5p and provided evidence for the direct link between histone acetylation and transcriptional activation (Brownell et al. 1996). Another study provided the evidence for enhanced DNA binding activity of transcription factors in the presence of acetylated H4 histones and consolidated the ideas of histone acetylation, chromatin accessibility and gene activation (Vettese-Dadey et al. 1996). Further research showed that histone acetylation was the main chromatin modification associated with transcriptional activation (Chuang et al. 2004; Clayton, Hazzalin, and Mahadevan 2006; Eberharter and Becker 2002; D. Y. Lee et al. 1993; Verdone et al. 2006). Several different types of HATs, including many of the known transcriptional co-activators such as Gcn5p, PCAF, TAF250, etc. have been discovered till date (reviewed in K. K. Lee and Workman 2007; Sterner and Berger 2000).

Histone acetylation is a reversible modification and the reverse reaction is catalyzed by histone deacetylases (HDACs). Many of the common transcriptional repressors such as Sin3, SMRT, HDAC2, Sir2 etc. possess HDAC-activity (Delcuve, Khan, and Davie 2012). For a detailed description of the different classes of HATs and HDACs, interested readers may refer to the following reviews (Carrozza et al. 2003; Kouzarides 2007; K. K. Lee and Workman 2007; Sterner and Berger 2000; Delcuve, Khan, and Davie 2012; Marmorstein 2001).

1.3.2 **Histone methylation**

Another well studied and characterized histone modification is the histone methylation. The very first discovery of methylation of histone dates back to 1964 (Kim and Paik 1965; Murray 1964) and its role in gene expression regulation was proposed during the same time (Allfrey, Faulkner, and Mirsky 1964). The research during the next one and half decade discovered the presence of methylation on lysine, arginine, and histidine amino-acids (reviewed in Paik, Paik, and Kim 2007). The direct evidence pertaining to the regulatory role of histone methylation began to appear in the late 1990s when the mechanism of transcriptional activation by nuclear receptors through transcriptional coactivator p160 was found to involve coactivator-associated arginine methyltransferase 1 (CARM1) (D. Chen et al. 1999). In another study, methylation patterns were analysed between Tetrahymena, yeast and HeLa cell line and it was revealed that 4^{th} lysine of H3 was a highly conserved site of methylation and its level of strongly correlated with the transcriptional activity in Tetrahymena nuclei (B. D. Strahl et al. 1999). Further research in the post-genomic era elucidated the indispensable and pivotal roles of histone methylation in gene expression, development and diseases (reviewed in Bannister and Kouzarides 2011; Greer and Shi 2012; M. Lawrence, Daujat, and Schneider 2016).

Methylation of the same lysine residue can exist in more than one state; i.e. mono, di, or trimethylation and each of the methylation state has a different biological function. For example,

H3K4me3 is strongly associated with the promoters of active genes, whereas H3K4me1 is more strongly associated with the enhancers and promoters of repressed genes (Kusch 2012). Similarly, methylation of different residues can have different outcomes. For example, H3K4me3 and H3K36me3 are active marks, while H3K9me3 and H3K27me3 are repressive marks (Dong and Weng 2013). Often the active and inactive marks are present at mutually exclusive genomic loci, H3K4me3 and H3K27me3 can co-exist on the same histone marking bivalent regulatory potential to the promoters of development associated genes in pluripotent or multipotent cells (Azuara et al. 2006; Bradley E. Bernstein et al. 2006; Mikkelsen et al. 2007). Therefore, the role of methylation in gene expression cannot be simply classified as activating or repressing and exhibits combinatorial complexity as compared to histone acetylation.

Histone methylation is carried out by a special class of enzymes called histone methyltransferases (HMTs) which were discovered and purified as early as the 1970s (Paik, Paik, and Kim 2007). HMTs belong to three different protein families; PRMT1 family, SET-domain family, and the non-SET-domain family (Wood and Shilatifard 2004). The members of PRMT1 methylates the arginine residue while SET/non-SET domains mainly methylate lysine residues of histone tails (Wood and Shilatifard 2004). For a detailed description about discovery, classification, and mechanisms of action of histone methyltransferases and demethylases, interested readers can refer to the following reviews (Bannister and Kouzarides 2011; Desjarlais and Tummino 2016; Kouzarides 2007; Marmorstein 2001).

1.3.3 **Histone phosphorylation**

Phosphorylation occurs at the serine, threonine or tyrosine amino acids present at the N terminal of histones (Bannister and Kouzarides 2011) and was known since the 1960s (Ord and Stocken 2015). After its initial discovery, H1 phosphorylation was found to be implicated in chromosome condensation and cell cycle regulation and its level was shown to oscillate according to the cell cycle stage peaking around M-phase when chromosomes are maximally condensed (E M Bradbury 1975; Gurley, Walters, and Tobey 1975; Hohmann, Tobey, and Gurley 1976). The histone phosphorylation witnessed a surge of interest when the MAP kinases and their downstream effectors were found to be involved in the rapid modulation of histone acetylation and phosphorylation of early response genes (Thomson, Mahadevan, and Clayton 1999). Cheung et al., in the year 2000, showed a synergistic coupling between acetylation and phosphorylation of H3 histones in response to the stimulation by epidermal growth factor (P. Cheung et al. 2000). The transcriptional roles and coupling of histone phosphorylation with acetylation were supported by a number of other reports which assayed the simultaneous presence of both these marks at the activated promoters (Brunmeir et al. 2010; Di Croce et al. 2010; Winter et al. 2008). Some reports also showed the involvement of the H3 Ser10 phosphorylation in the transcriptional activation of

genes involved in NF- κB pathway (review in Banerjee and Chakravarti 2011). These studies indicated that despite being the positive regulator of chromatin condensation during cell cycle progression, rapid phosphorylation induced by signaling cascades can play a positive role in the transcriptional activation (Sawicka and Seiser 2014).

In addition to transcription and chromosome condensation, phosphorylation of H2A.X is one of the key mediators of DNA damage response in eukaryotic organisms. It was shown, in the year 1998, that DNA double-stranded breaks induced the phosphorylation of specific serine residues in various different mammalian cell cultures (Rogakou et al. 1998). Later in 2003, a similar observation was made in budding yeast (Redon et al. 2003). Over the next few years, phosphorylation of H2A.X was extensively studied and has emerged as one of the main hallmarks of DNA damage in eukaryotic systems (Banerjee and Chakravarti 2011; Rossetto, Avvakumov, and Côté 2012; Sawicka and Seiser 2014). Phosphorylation and dephosphorylation are carried out by classes of enzymes broadly known as kinases and phosphatases respectively (Watson and Higgins 2016). The detailed description of these two categories of enzymes is beyond the scope of the present thesis and can be found elsewhere (Watson and Higgins 2016).

With the advent of modern molecular methods in the 21st century, more than 60 chemical modifications of histone have been described (mentioned in Kouzarides 2007). A summary of some commonly studied chromatin modifications in listed in presented in table 1.

Table 1 Chromatin modifications and their functions.

	His-tone	Residue	Species	Enzymes Responsible	Biological Functions	References
Acetylation	H2A	K4/5/7	*Hs*, *Sc*	K4:Esa1; K5:Tip60,p300/CBP K7:Hat1,Esa1	Transcriptional activation	(Altaf et al. 2010)
	H2B	K5/11/16/20	*Hs*, *Sc*	K5/20: p300/CBP, K11/16: Gcn5,Esa1	Transcriptional activation	(Allis et al. 2007; Myers et al. 2003)
	H3	K4/27/36	*Hs*, *Sc*	K4:Esa1; K27:p300/CBP; K36:Gcn5	Transcriptional activation	(Guillemette et al. 2011; Inagaki, Sakai, and Kajimura 2016; Morris et al. 2007)
		K9	*Hs*, *Sc*	GCN5, SRC-1	Histone deposition, transcriptional activation	(Gates et al. 2017; Karmodiya et al. 2012)
		K14	*Hs*, *Sc*	Gcn5, Tip60, Esa1, SRC-1, TAFII,Sas2, Sas3, p300	Transcriptional activation, elongation and DNA repair.	(M. R. Duan and Smerdon 2014; Kurdistani, Tavazoie, and Grunstein 2004)
		K18	*Hs*, *Sc*	Gcn5, p300,CBP	Transcriptional activation, DNA repair, DNA replication	(Pham et al. 2007; Tatsuoka, Liu, and Guo 2005)
	H4	K5	*Hs*, *Sc*	Hat1, Esa1, Tip60,ATF2, Hpa2,p300,NuA4	Histone deposition, transcriptional activation, DNA repair	(Dhar et al. 2017; Goudarzi et al. 2016; Rossetto et al. 2010)
		K8	*Hs*, *Sc*	GCN5,PCAF,Esa1,Tip60,ATF2 P300,Elp3	Transcriptional activation, elongation and DNA repair	(Dhar et al. 2017; Goudarzi et al. 2016)
		K12	*Hs*, *Sc*	Hat1, Esa1, Tip60, Hpa2, p300,Nua4	Histone deposition, transcriptional activation, DNA repair, telomere	(Dhar et al. 2017; Rossetto et al. 2010)

					silencing	
		K16	*Hs*, *Sc*	Gcn15, Esa1, Tip60, ATF2,Sas2	Transcriptional activation, DNA repair, Euchromatin	(Dhar et al. 2017; Shogren-Knaak et al. 2006; Tatsuoka, Liu, and Guo 2005)
Methylation	H2A	R3	*Hs*,	PRMT1/6, PRMT5/7	Transcriptional activation, repression	(Bedford and Clarke 2009; Hyllus et al. 2007)
		Q104/105	*Hs* (104), *Sc* (105)	Nop1	Ribosomal gene expression	(Tessarz et al. 2014)
	H3	R17/26/42	*Hs*	CARM1	Transcriptional activation	(Bauer et al. 2002; Blanc and Richard 2017; Di Lorenzo and Bedford 2011)
		R2	*Hs*, *Sc*	PRMT5, PRMT6	Gene Expression	(Guccione et al. 2007; Kirmizis et al. 2007)
		K4	*Hs*, *Sc*	Set1, Set7/9. MLL, ALL-1, Ash1	Transcriptional activation	(Ruthenburg, Allis, and Wysocka 2007)
		K9	*Hs*	Suv39h, Clr4, SETDB1, Ash1	Histone deposition, transcriptional repression heterochromatin	(Greer and Shi 2012; Hublitz, Albert, and Peters 2009)
		K27	*Hs*	Ezh2	Polycomb repression, transcriptional silencing	(Hublitz, Albert, and Peters 2009; Wiles and Selker 2017)
		K36	*Hs*, *Sc*	Set2	Transcriptional Elongation	(Wagner and Carpenter 2012)
		K79	*Hs*, *Sc*	Dot1	Transcriptional elongation, DNA damage response, telomere silencing	(Farooq et al. 2016)
	H4	R3	*Hs*, *Sc*	PRMT1/6	Transcriptional activation, repression	(Di Lorenzo and Bedford 2011)
		K/8/12	*Sc*	Set5	DNA Replication stress	(Green et al. 2012)
		K20	*Hs*, *Sc*	PR-Set7, Suv4-20h, Ash1, Set9	Transcriptional silencing of genes	(Kaimori et al. 2016; Kalakonda et al. 2008)
		K59	*Sc*	NA	Transcriptional silencing, formation of silent chromatin	(Kalakonda et al. 2008; L. Zhang et al. 2003)
Phosphorylation	H2A	S1	*Hs*, *Sc*	MSK1	Mitosis and chromatin assembly	(Barber et al. 2004; W.-L. Wang et al. 2013)
		S122/129/139	*Hs*(139), *Sc*(122,129)	Mec1/Tel1, ATR/ATM, DNA-PK	DNA repair	(Moore et al. 2007)
	H2B	S10	*Sc*	Ste20	Apoptosis	(Ahn et al. 2005)
		S14	*Hs*	Mst1	Apoptosis, DNA repair	(Fernandez-Capetillo, Allis, and Nussenzweig 2004)
		S33/36	*Sc*(33), *Hs*(36)	S33:TAF1; S36: AMPK	Transcriptional activation	(W. L. Cheung et al. 2003; Maile et al. 2004)
	H3	T3/6/11	*Hs*(6), *Sc*(11)	T3: Haspin; T6: PKCβ; D1K/ZIP	Mitosis	(Sawicka and Seiser 2012)
		S10	*Hs, Sc*	MSK1, MSK2, IKK-α, Snf1	Mitosis, Meiosis, transcriptional activation	(Sawicka and Seiser 2012)
		S28	*Hs , Sc*	MSK1, MSK2, Aurora-B kinase	Mitosis, immediate early genes induction	(Sawicka and Seiser 2012)
		Y41	*Hs, Sp*	JAK2	Transcriptional activation	(Dawson et al. 2009)
		Y45	*Hs*, *Sc*	PKCδ	Apoptosis, DNA replication	(Brehove et al. 2015)
	H4	S1	*Hs*, *Sc*	Casein kinase 2	DNA damage response	(W. L. Cheung et al. 2005)
Ub	H2A	K13/15/63	*Mm*	K13/15: Rnf168; K63: Rnf8	Part of DDR to DSBs	(Uckelmann and Sixma 2017; Vissers et al. 2008)

		K119,K123	*Sc*(123) *Hs* (119)	dRing, RING1B, RING2, Rad6-Bre1	Gene silencing, DDR	(Vissers et al. 2008)
	H2B	K34	*Hs*, *Sc*	MSL2	Transcriptional activation	(L. Wu et al. 2011)
		K120	*Hs, Sc*	RNF20/40, UbcH6	Cell cycle, transcriptional activation, DDR	(Cao and Yan 2012)
		K123	*Sc*	Rad6, Bre1 (E2, E3)	Telomeric silencing, transcriptional activation	(Cao and Yan 2012)
Sumoylation	H2A	K126	*Sc*	Ubc9	Transcriptional repression, blocks histone acetylation	(Nathan et al. 2006)
	H2B	K6/K7	*Sc*	Ubc9	Transcriptional repression	(Nathan et al. 2006)
		K16/K17	*Sc*	Ubc9?	Gene repression	(Nathan et al. 2006)
	H4	N-terminal	*Hs, Sc*	Ubc9	Transcriptional repression	(Shiio and Eisenman 2003)

Note. Hs: Homo sapiens, Sc: Saccharomyces cerevisiae, Sp: Schizosaccharomyces pombe, Mm: Mus musculus

1.3.4 **Combinatorial complexity of chromatin modifications and the histone code**

By the end of the 20th century, several different chromatin modifications were discovered and it was proposed that these modifications constitute a 'histone code' which stored the information about the activity state of a given gene (Brian D. Strahl and Allis 2000; Jenuwein and Allis 2001). An important aspect of this hypothesis was the combinatorial effects of chromatin modifications which were often difficult to be predicted from the individual modifications. For example, the phosphorylation of H3 histone was involved in decondensation of the chromatin by imparting a negative charge to nucleosomes. However, H3 phosphorylation at serine 10, along with phosphorylation at serine 28 was also involved in the chromosome condensation during mitosis as well as meiosis. This condensation could not be explained in the light of the role of histone phosphorylation in chromatin decondensation observed during the induction of immediate early genes (reviewed in Brian D. Strahl and Allis 2000). The combinatorial complexity of chromatin modifications was also exhibited by the transcriptionally poised genes which were associated with the bivalent marks, i.e. the simultaneous presence of both active (H3K4me3) and inactive (H3K27me3) chromatin modifications, as discussed earlier. The chromatin bearing bivalent marks is highly enriched in the promoters of the genes involved in early embryonic development (Azuara et al. 2006; Bradley E. Bernstein et al. 2006; Mikkelsen et al. 2007) as well as in maintenance of pluripotency (Harikumar and Meshorer 2015; Muers 2012; Voigt, Tee, and Reinberg 2013). To understand the broader aspects of combinatorial modifications, several studies used computational methods and elucidated that pattern of chromatin modifications can be classified in different chromatin classes and each class of modifications represented the distinct biological state of their target genomic locus (Ernst et al. 2011; Ernst and Kellis 2010, 2012; Filion et al. 2010; Jaschek and Tanay 2009; Kharchenko et al. 2011; Ram et al. 2011). For example, a study by Filion et al. showed that there were 5 major types of chromatin in *Drosophila,* which the authors called as Blue (Polycomb repressed heterochromatin with H3K27me3), Green (HP1

marked heterochromatin), Black (Repressed chromatin, Neither marked by HP1 nor Polycomb associated H3K27me3), Red (Active chromatin, but lacked H3K36me3), and Yellow (Active chromatin, had H3K36me3) (Filion et al. 2010). As noted by the Filion et al., 5 major categories of chromatin may have further subtypes and consistent with this idea, a more elaborate classification of chromatin states was also reported in *Drosophila* (Kharchenko et al. 2011), mouse (Yue et al. 2014) and humans (Ernst et al. 2011; Ernst and Kellis 2010). These studies highlighted the fact that chromatin modifications across genomes exhibit combinatorial patterns which increases the complexity of regulatory effect of these modifications.

1.4 ATP dependent chromatin remodeling enzymes

Besides covalent modifications, another kind of chromatin remodeling involves the use of ATP dependent remodeling enzymes which can displace, evict, and replace the histones. Past two decades of research have shown that such chromatin remodelers can be classified into four distinct families depending on the similarities and differences in their ATPase domain (Clapier et al. 2017; Flaus and Owen-Hughes 2011; Narlikar, Sundaramoorthy, and Owen-Hughes 2013)

a. SWI/SNF family remodelers
b. ISWI family remodelers
c. CHD family remodelers
d. INO80 family remodelers.

Each remodeling complex is a multi-subunit macromolecular assembly consisting of a conserved ATPase domain belonging to the SNF2 family flanked by variable domains (figure 1.5) which differ between different families (Gangaraju and Bartholomew 2007; Narlikar, Sundaramoorthy, and Owen-Hughes 2013). A brief description of four major classes of chromatin remodelers in budding yeast is presented in the following sections.

1.4.1 SWI/SNF family remodelers

The SWI/SNF complex contains 11 subunits with Snf2/Swi2 as core ATPase component and is one of the first chromatin remodeling complexes to be discovered (Clapier et al. 2017; Gangaraju and Bartholomew 2007; Hauhnar and Lalhruaitluanga 2016). SWI/SNF complex was originally shown to play important roles in mating type switching as well as growth on non-sucrose carbon sources, hence the name SWI/SNF (SWI: mating type SWItching and SNF: Sucrose Non-Fermenting) (mentioned in Gangaraju and Bartholomew 2007). The budding yeast SWI/SNF also assembles in an alternate form called RSC, which itself can exist into functionally two different complexes depending upon the presence of Rsc1 or Rsc2 (Cairns et al. 1999). Mechanistically, SWI/SNF complex was shown to displace the nucleosomes along the DNA molecule in an ATP dependent reaction (Whitehouse et al. 1999). SWI/SNF has been identified and characterized in *Drosophila* and humans too, where it is known to exist in two different forms; namely BAP and

PBAP in case of *Drosophila* and BAF and PBAF in case of humans (Mohrmann and Verrijzer 2005). The SWI/SNF has an HSA domain, which is involved in actin binding and a bromodomain at the C-terminal of Snf2, which binds the acetylated lysines of the histone tales and helps in targeting the acetylated nucleosomes by the SWI/SNF complex (Dhalluin et al. 1999; Ladurner et al. 2003). Other components of SWI/SNF complex such as Swi1 and Swi3 also contain specific domains (Swi1: ARID domain, Swi3: SWIRM domain and SANT domain), which are conserved in from yeast to *Drosophila* and humans and present in other ATP-dependent chromatin remodelers such as ISW1and RSC as well (Gangaraju and Bartholomew 2007).

Table 2 The chromatin remodeling complexes and their subunits

	Complexes in Yeast	Subunits
SWI/SNF Sub-Family	SWI/SNF	Swi2/snf2,Swi1,Swi3,Snf5,Swp82,Swp73,Rtt102,Snf6, Arp7, Arp9, Swp29,Snf11
	RSC	Sth1, Rsc1, Rsc2, Rsc4, Rsc5, Rsc6, Rsc7, Rsc8, Rsc9, Rsc10, Rsc13, Rsc14, Rsc15 Sfh1, Arp7, Arp9, Rtt102
ISWI Sub-Family	ISW1a	Isw1, Ioc3
	ISW1b	Isw1, Ioc2, Ioc4
	ISW2	Isw2, Itc1, Dbp4, Dls1
CHD1 Sub-Family	CHD1	Chd1
INO80 Sub-Family	INO80	Ino80, Arp4, Arp5, Arp8, Rvb1, Rvb2, Les1, Les2, Les3, Les4, Les5, Les6, Act1,Taf14, Nhp10
	SWR1	Swr1, Arp4, Arp6, Rvb1, Rvb2, Act1, Swc1, Swc2, Swc4, Swc5, Swc6, Yaf9, Bdf1

This table is based on (Clapier and Cairns 2009; Ehrenhofer-Murray 2004; Lusser and Kadonaga 2003)

1.4.2 **ISWI family remodelers**

The core catalytic subunit of ISWI (Imitation SWItching) complex is similar to SNF2 subfamily ATPases which are involved in mating type switching (Gangaraju and Bartholomew 2007; Hauhnar and Lalhruaitluanga 2016). It was first discovered in *Drosophila* embryos where it assembles into three complexes called NURF, ACF, and CHRAC (Hauhnar and Lalhruaitluanga 2016). In budding yeast, the core subunit of ISWI complex is encoded by two distinct genes; the ISW1 and ISW2, which were identified based on their extensive homology with dISW and was shown to assemble into two different forms: ISW1 and ISW2 (Hauhnar and Lalhruaitluanga 2016). The ISW1 complex was described as a four subunit complex having ATPase activity

stimulated by nucleosomes and Isw1p as core ATPase. Additionally, it was also found to have nucleosome disruption and spacing activity (Tsukiyama et al. 1999). The ISW2 complex, with Isw2p as core ATPase, was shown to lack any significant nucleosome disruption activity. Further, the ISW1 was shown to assemble into two distinct complexes; the ISW1a (with Isw1p, Ioc3p) and ISW1b (with Isw1p, Ioc2p, and Ioc4p) (Vary et al. 2002). ISW1p/2p complexes were implicated in transcriptional silencing by displacing the basal transcription machinery and (Mellor and Morillon 2004). In addition to the core ATPase, ISWI family members also contained several other domains such as SANT, SLIDE, HAND, and AID (Grüne et al. 2003). The DNA binding was found to be strongly dependent on the SLIDE domain, but the nucleosome binding required both the SANT and SLIDE domain and not by the single deletion of individual domains (Grüne et al. 2003).

1.4.3 **CHD family remodelers.**

The members belonging to CHD1 remodelers family are least characterized and were first described in murine lymphoid cell lines in the form of DNA binding protein containing a chromo-domain as well as SNF2-like helicase domain (Delmas, Stokes, and Perry 1993). Later studies showed that the Chd1 protein had preferential binding in long AT tracks on double-stranded DNA (Stokes and Perry 1995). However, unlike HP1, which is one of the major constituents of heterochromatin near centromeres, Chd1 lacked preferential binding for centromeric chromatin in spite of the presence of higher AT-rich tracks in centromeres (Stokes and Perry 1995). While *Drosophila* and mammals such as humans and mice were found to contain 2 to 4 genes for CHD1, the budding yeast had only one gene for this protein (Woodage et al. 1997). The budding yeast Chd1 protein was found to be a member of SAGA and SLIK (SAGA-Like) complexes and enhanced the acetylation of the methylated residues (Pray-Grant et al. 2005). Relatively recent studies have shown that budding yeast Chd1 is also important for maintaining the structural integrity of chromatin along with Isw1b by preventing the *trans*-histone exchange (Smolle et al. 2012).

1.4.4 **INO80/SWR1 family remodelers**

The INO80/SWR1 remodeling complex contains more than 15 subunits and plays important roles in a wide range of DNA templated processes such as transcription, DNA repair, and DNA replication (Conaway and Conaway 2009). In budding yeast, it was identified by Shen et.al as an open reading frame highly similar to *Drosophila* ISWI gene (Shen et al. 2000). The INO80 gene was also identified in a genetic mutant screen to be causing inositol auxotrophy (Ebbert, Birkmann, and Schüller 2002). The INO80 complex identified by Shen et al. contained Actin (Act1) as well as Arp4,5,&8, which were three actin-related proteins. Two other proteins which were discovered to be present in the INO80 complex were Rvb1 and Rvb2, each of which shared

homology with the Holliday junction DNA helicase of bacteria (Shen et al. 2000). Indeed it was confirmed that INO80 complex had the DNA helicase activity pertaining to the Rvb1 and Rvb2 subunits (Shen et al. 2000). More recent data have shown that INO80 is involved in the genome-wide distribution of a variant histone called H2A.Z by replacing the H2A/H2B dimer with nucleosomal H2A.Z/H2B (Papamichos-Chronakis et al. 2011; Tosi et al. 2013). The incorporation of H2A.Z/H2B into the nucleosomes by replacing the H2A/H2B is regulated by SWR1 complex which shares many subunits with INO80, including Arps and Rvb1/2 and has been found to be active at telomeres, centromeres and other intragenic regions (Gangaraju and Bartholomew 2007). SWR1 complex was identified in genetic screens, where three of its component were found to show synthetic lethality with the transcriptional elongation and chromatin metabolism-related proteins (Krogan et al. 2003).

As shown in figure 1.6 chromatin remodeling complexes are often composed of multi-domain proteins and exhibit long stretches of the intrinsically disordered region (Mahmoudabadi et al. 2013; Sandhu 2009). It is proposed that the presence of multiple domains and long disordered regions might allow diverse macro-molecular interactions of chromatin factors during transcriptional regulation, DNA repair, and replication (Sandhu 2009).

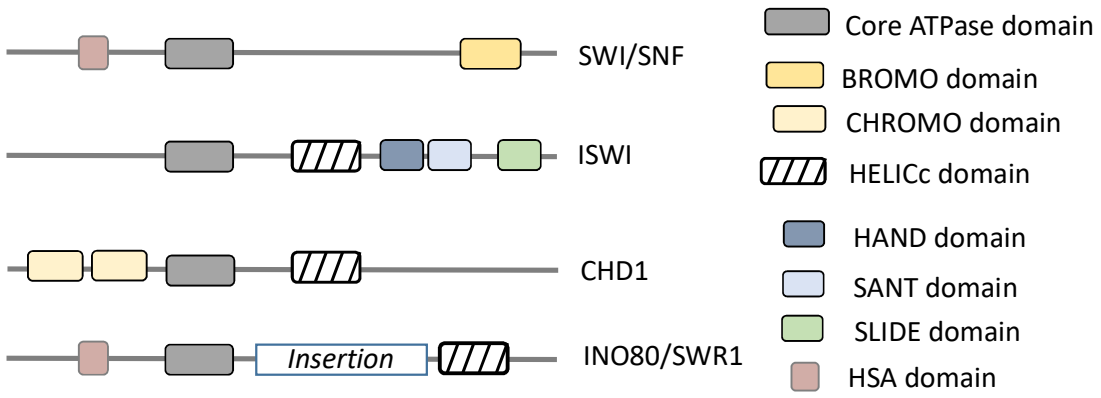

Figure 1.6 Protein domains of ATP dependent chromatin remodeling enzymes.

The bromo domains of SWI/SNF and chromo domains of CHD1 family are important in recognizing the acetylated and methylated histones, respectively.

1.5 Chromatin remodeling

The dense array of nucleosomes presents a significant barrier against the DNA accessibility by various factors involved in transcription, replication, DNA repair, etc. (Groth et al. 2007; Studitsky et al. 2016). Therefore, chromatin is actively remodeled during the process (Ehrenhofer-Murray 2004; Groth et al. 2007; B. Li, Carey, and Workman 2007; Petty and Pillus 2013). Some of the important changes accomplished in chromatin structure during transcription, replication and DNA repair are reviewed in this part of the introduction.

1.5.1 **Chromatin remodeling during transcription**

The process of transcription initiation involves the assembly of a multi-subunit macromolecular entity called pre-initiation complex (PIC) (Kubik, Bruzzone, and Shore 2017; Shandilya and Roberts 2012). The first step in this process is the loading of the TATA-binding protein (TBP) onto the promoters which is accomplished either by the SAGA or TFIID dependent pathway (Huisinga and Pugh 2004; Venters et al. 2011). This results in the recruitment of several general transcription factors (TFIIA/B/D/E/F/H/) and the subunits of mediator complex which leads to the recruitment of RNA-PolII resulting in the formation of the pre-initiation complex (Allen and Taatjes 2015; Kubik, Bruzzone, and Shore 2017; Poss, Ebmeier, and Taatjes 2013; Taatjes 2017). However, before these events can take place, chromatin undergoes active remodeling to alleviate the accessibility barrier imposed by the nucleosomes (figure 1.7).

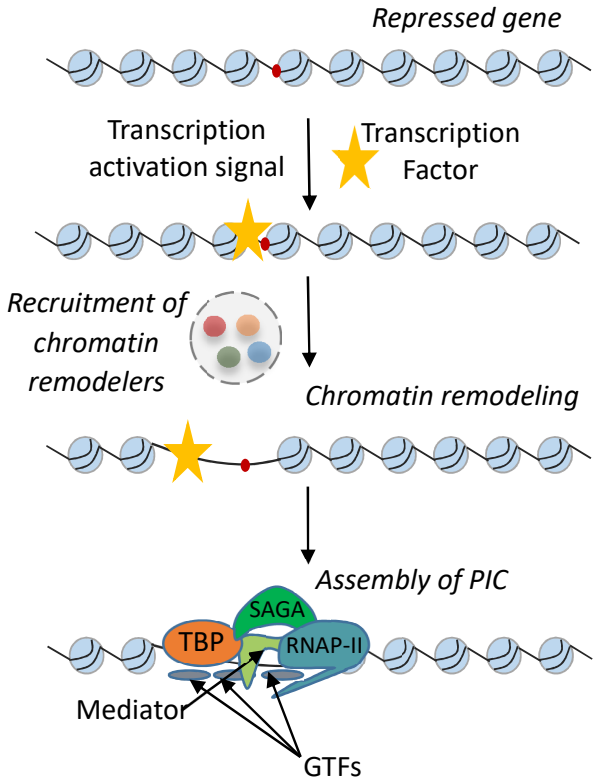

Figure 1.7 Transcription initiation at repressed gene.

Repressed genes have high nucleosome occupancy (blue circles) which restricts the access to transcriptional machinery. Following the transcriptional induction by transcription factors (yellow star), the chromatin modifiers (eg. histone acetyl transferases and histone demethylases) and ATP dependent chromatin remodelers (eg. SWI/SNF) create a long nucleosome free region near the transcription start site (TSS, red dot), where assembly of the pre-initiation complex takes place.

Although the precise mechanisms involved in the recruitment of chromatin modifiers at their target promoters are yet to be deciphered, some reports have indicated that the transcription

factors might play an important roles in directing the chromatin modifiers to the target genes, in particular, the pioneer transcription factors (Erdel et al. 2011; Swinstead et al. 2016; Varga-Weisz 2010; Voss and Hager 2014). The pioneer transcription factors play critical roles in activating the gene expression during embryonic development by inducing the permissible changes in the epigenetic landscape through the recruitment of chromatin modifiers (Swinstead et al. 2016) (Mayran and Drouin 2018; Zaret and Carroll 2011). A few examples of pioneer transcription factors include Yamanaka factors (such as Oct4, Sox2, Klf4, c-Myc) which are important for cellular reprogramming (Iwafuchi-Doi and Zaret 2014; Takahashi and Yamanaka 2006; Zaret and Carroll 2011)and GATA factors which are important for development of various organs (Tremblay, Sanchez-Ferras, and Bouchard 2018). A summary of chromatin modifications during the process of transcription is presented in the following text and summarised in figure 1.8, 1.9.

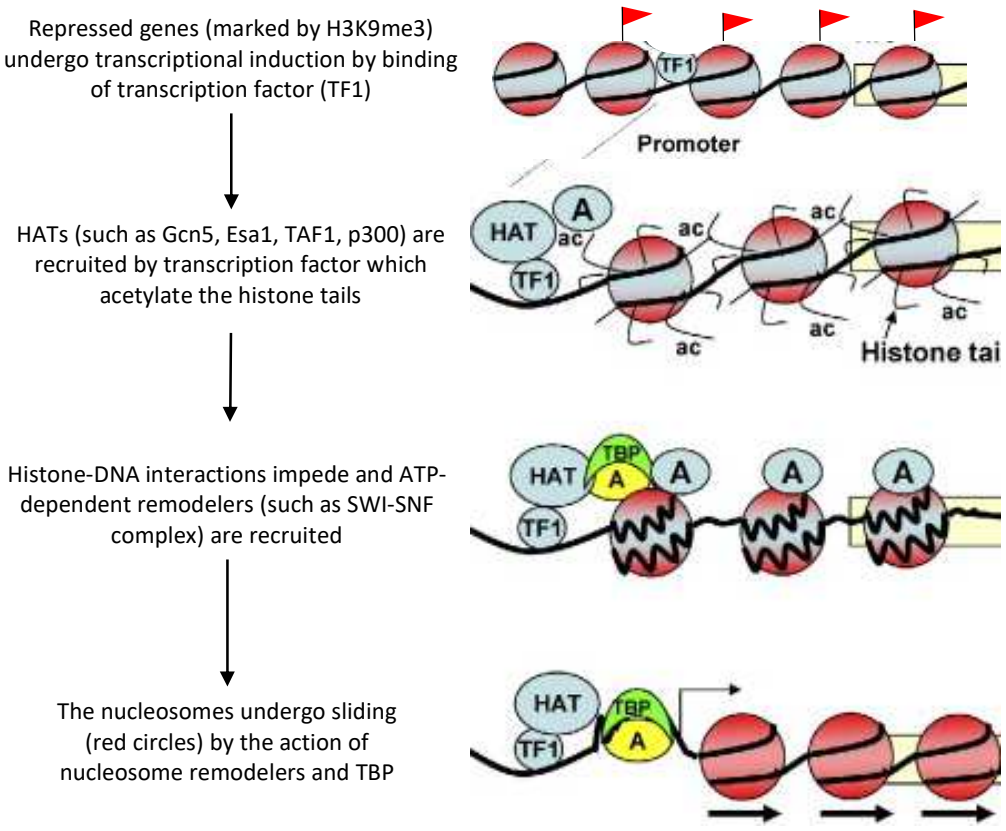

Figure 1.8 Chromatin remodeling at the site of transcription initiation.

Repressed genes have higher nucleosome occupancy and bears H3K9me3 mark. Upon induction of gene expression, the repressive chromatin is remodeled by the replacement of H3K9me3 with H3K9Ac by the action of HATs. The ATP dependent chromatin remodelers are recruited to the site of acetylated histones and cause the nucleosome displacement (indicated by red circles).

This figure is modified from Mellor (2005) and reused with permission.

The most important chromatin modification occurring during transcription is the acetylation of specific residues within the promoters of the transcribing genes. Both SAGA complex and TFIID possess HAT activity and can acetylate the specific lysine residues of histones in the core promoter region (T. I. Lee et al. 2000). The most common acetylation occurring in promoter regions is H3K9Ac and H3K14Ac (Guenther et al. 2007; Liang et al. 2004). Both these modifications are deposited by the action of one of the transcriptional co-activators such as Gcn5, TAF1/TAFII-250, and p300/CBP (Guenther et al. 2007; Sterner and Berger 2000). The acetylation of histones in the promoter region can neutralize the positive charge of histones and disrupt the DNA-nucleosome interactions which facilitate in alleviating the repressive barrier imposed by nucleosomes (figure 1.8) (Eberharter and Becker 2002; Gorisch 2005; Verdone et al. 2006). Another hallmark of transcriptional activation is the di-and trimethylation of H3K4 residues (B. E. Bernstein et al. 2002; Guenther et al. 2007; Schneider et al. 2004) and is brought about by the coordination of several different molecular events as described in the following paragraph.

Once the formation of the pre-initiation complex is complete, RNA-PolII can proceed into the elongation phase by the phosphorylation of the specific serine residues of the C-terminal domain (CTD) in its Rbp1 subunit (Phatnani and Greenleaf 2006). This phosphorylation is catalysed by the kinase activity of Cdk7 subunit of TFIIH (Phatnani and Greenleaf 2006). The phosphorylated RNA-PolII helps in the recruitment of Paf1 complex by the Bur1/2 kinase-dependent phosphorylation of spt4 and spt5 (Guenther et al. 2007; B. Li, Carey, and Workman 2007; Smolle and Workman 2013). Paf1 complex is one of the main elongation promoting complex that accompanies the RNA-PolII and from the promoters till the poly-A tails at the 3'end of the transcribing genes. The most important consequence of Paf1 complex is the ubiquitylation of lysines in H2A (H2AK123Ub) by the recruitment of ubiquitin-conjugating and ligase enzymes Rad9 and Bre1 (Wood et al. 2003). The ubiquitylated form of H2A favours the Set1 mediated di- as well as trimethylation of H3K4 in the histone octamer and also promoters the Dot1 dependent methylation of H3K79 (Wood et al. 2003). Interestingly, this role of Paf1 in inducing the methylation of specific residues near the transcribing RNA PolII is conserved in yeast and humans. During the late elongation phase RNA-PolII, the H3 histone undergoes Set2 dependent trimethylation at H3K36 residue (Venkatesh and Workman 2013). This methylation is recognized by an HDAC complex, Rpd3S, which then deacetylases the acetylated histones behind the elongating RNA PolII which is known to suppress the initiation of transcription from cryptic promoters within the gene body (Venkatesh and Workman 2013).

The covalent modifications associated with transcription also crosstalk with the ATP dependent chromatin remodeling enzymes (Petty and Pillus 2013). For example, the acetylated residues were found to be identified by SWI/SNF complex (Hassan, Neely, and Workman 2001) which can

perform the transcription associated nucleosome sliding (Steger and Workman 2005; Workman 2006). In addition to SWI/SNF, the yeast PHO5 gene has been shown to recruit INO80 complex upon transcriptional induction by transcription factor Pho4p, which contributes to the sliding or eviction of nucleosomes (Boeger et al. 2004; Steger et al. 2003). Chromatin remodeling involving the exchange of core histones with its variant form, (H2A with H2A.Z in yeast), has been proposed to destabilise the nucleosomes (Workman 2006) and assist in the process of transcription. The exchange with histone variant H2A.Z is known to be mediated *via* SWR1 complex (B. Li, Carey, and Workman 2007). FACT-complex, which is required for transcriptional elongation in yeast as well as mammals, was shown to cause the removal of one H2A-H2B dimer in order to allow passage for RNA-PolII (Belotserkovskaya et al. 2003). Therefore, the process of transcription involves the extensive changes in chromatin structure which are mediated through the concerted action of histone modifying enzymes and ATP-dependent chromatin remodeling complexes. A schematic of chromatin modifications during transcription is given in figure 1.8 and 1.9.

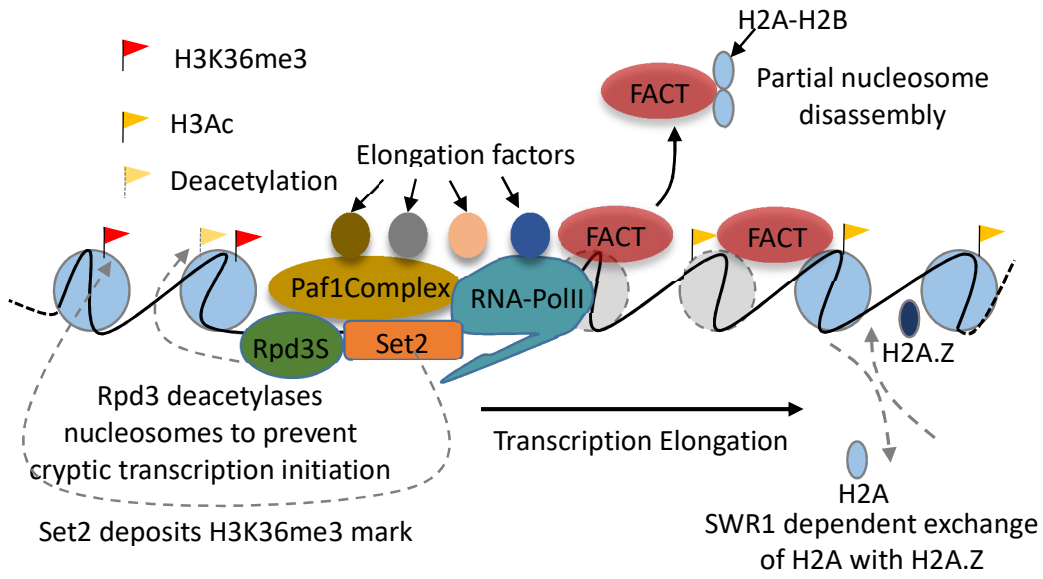

Figure 1.9 Summary of chromatin remodeling during transcription elongation.
The elongating RNA-PolII is associated with Paf1-complex, COMPASS, Rpd3, elongation factors such as PTEF-b, Ctk1 and FACT complex. The FACT complex accomplishes the partial disassembly of the nucleosomes, which is favored by the SWR1 dependent incorporation of H2A.Z in exchange of H2A into the nucleosomes. The Set2 (subunit of COMPASS) perform the H3K36me3, which helps in the deacetylation of H3 histones by Rpd3S complex, thereby preventing the cryptic transcription initiation inside the gene body.

1.5.1 **Chromatin remodeling in DNA repair**

DNA inside the nucleus is constantly exposed to various kinds of exogenous as well as endogenous mutagens which can result in the DNA lesions of various kinds and pose a threat to

the genome integrity. Multiple DNA repair pathways have evolved to counter these effects in a timely and efficient manner (Hakem 2008). Similar to transcription, the process of DNA repair also requires access to the underlying DNA sequence in the chromatin. The best characterized among these are the modifications which occur during the homologous recombination repair pathway in response to DNA double-strand breaks (DSBs) (Price and D'Andrea 2013; Tsukuda et al. 2005). Though relatively infrequent as compared to other lesions like mismatches, gaps or nicks in DNA, the DSBs pose a much more serious threat to the genome integrity and repair of DNA damage response involved in this pathway involves extensive chromatin remodeling (Lans, Marteijn, and Vermeulen 2012; Luijsterburg and Van Attikum 2011; Price and D'Andrea 2013).

At the heart of DNA damage response lies the role of ATM/ATR kinases which are the mammalian homologs of yeast Tel1/Mec1 kinases (Bakkenist and Kastan 2003; Kinner et al. 2008; Shiloh 2003; Uziel et al. 2003). These kinases are activated by the MRN (MRX in yeast) complex at the broken ends of DNA (J.-H. Lee and Paull 2004, 2005; Uziel et al. 2003) and results in the phosphorylation of serine 139 (serine 129 in yeast) of the H2A.X (H2A in yeast) subunit of the nucleosomes (Kinner et al. 2008). The phosphorylated form of these histones is called the ϒ-H2A.X and triggers the binding of DNA-damage checkpoint protein MDC1 at the site of DNA damage, which in turn, helps in the recruitment of 53BP1, BRCA1 and MRN at the ϒ-H2A.X sites (Stewart et al. 2003). The reinforced recruitment of MRN complex results in the phosphorylation of H2A.X (*via* ATM) within nearby nucleosomes which spreads up to several Kbs in yeast and Mbs in mammals (Foster and Downs 2005) and helps in the recruitment of more MDC1 (House, Koch, and Freudenreich 2014). Therefore, the phosphorylation of H2A.X and the recruitment of MDC1 operates by the positive feedback loop. (House, Koch, and Freudenreich 2014). This spreading results in the assembly of a large complex called 'repairsome' which contains many of the important DNA repair proteins including 53BP1, BRCA1 (reviewed in Walsh 2015), RAD51 (Davies et al. 2001), NBS1 (reviewed in Kobayashi et al. 2004) etc. It is worth noting that under normal conditions, the ϒ-H2A.X is more readily evicted from nucleosomes as compared to H2A.X by the activity of FACT complex (Heo et al. 2008). However, this is prevented at the damaged site by the PARP-1 dependent ADP-ribosylation of Spt16 subunit of FACT complex, which helps in the formation and propagation of ϒ-H2A.X foci (Heo et al. 2008; Lieber 2010). In addition to ϒ-H2A.X, researchers have discovered several other covalent modifications of histones which can assist in the process of DNA repair either by opening up the damaged chromatin or by help recruiting the other DNA repair proteins (Altaf, Saksouk, and Côté 2007; Jeggo, Downs, and Gasser 2017; Price and D'Andrea 2013). A list of different chromatin modifications and their associated functions is presented in table 1.

Phosphorylation of H2A.X is one single event which orchestrates and mediate a chain of

downstream events for recruitment of DNA repair proteins near the site of DSBs. However, this recruitment is useful only if the repair factors can access the underlying DNA sequences in order to carry out the typical steps of DNA repair which involves the exonuclease, polymerase and ligase activities (S. P. Jackson 2002). To facilitate these processes, the chromatin structure has to be physically remodeled by displacement or eviction of the nucleosomes at the site of DNA damage. While the exact nucleosome organization at the site of DNA breaks remains unknown (Lieber 2010), several studies have indicated the role of ATP dependent chromatin remodeling complexes in the faithful repair of DNA breaks (Foster and Downs 2005; House, Koch, and Freudenreich 2014). Importantly, the SWR1 complex results in the incorporation of H2A.Z in exchange of H2A at the damaged site (Morrison and Shen 2009). This exchange was shown to result in the formation of transiently open chromatin regions. H2A.Z has also been shown to be required for the acetylation and ubiquitylation of specific histones which promotes the loading of BRAC1 complex (Y. Xu et al. 2012). NuA4 dependent acetylation of H4 and H2A at their N-terminal tails, which resulted in the stimulation of Swr1 activity, also resulted in the increased incorporation of H2A.Z in the chromatin (Altaf et al. 2010). Consistently, several HATs such as NuA4 and TIP60, HDACs like NuRD have been shown to be recruited to the site of DNA breaks in order to repair the damaged sites (reviewed in Foster and Downs 2005; House, Koch, and Freudenreich 2014). The role of the SWR1 complex in DNA repair was also supported by a study in Arabidopsis where mutations in its subunits resulted in the failure of DNA repair via homologous recombination (Rosa et al. 2013). Another chromatin remodeling complex that has been shown to play DNA repair *via* homologous recombination pathway is INO80 (Gospodinov et al. 2011) which can place or evict nucleosomes along the chromatin. The INO80 binding at the site of DSBs was detected 1-2hr following the induction of DNA damage and some studies have proposed that INO80 dependent nucleosome eviction around DSBs promotes the binding of MRX, yKu80 and ATR kinase, as well as the processing of the broken DNA ends (Van Attikum, Fritsch, and Gasser 2007; Gospodinov et al. 2011; Tsukuda et al. 2005). The INO80 dependent eviction of nucleosomes was also required for the mobility of chromatin following DNA damage, which enhanced the rate of homologous recombination (Conaway and Conaway 2009; Hauer et al. 2017; also reviewed in Seeber, Dion, and Gasser 2013). Therefore, DNA damage response is accompanied by extensive chromatin remodeling. For a more detailed discussion of the same, the interested readers may find the following reviews useful (Jeggo, Downs, and Gasser 2017; Morrison and Shen 2009; Price and D'Andrea 2013; Smeenk and van Attikum 2013).

1.5.2 **Chromatin remodeling in DNA replication**

In addition to the transcription and DNA repair, another process that is closely interwoven in the chromatin context is the DNA replication. One key event during the DNA replication is the

disruption of nucleosomes near the replication fork. Several different chromatin remodeling complexes have been shown to be implicated in this process also (Vincent, Kwong, and Tsukiyama 2008). For example, ACF complex was found to affect the kinetics of replication and was required for replication of pericentromeric heterochromatin in mammalian systems (Collins et al. 2002; also reviewed in MacAlpine and Almouzni 2013). Another complex called WSTF has been implicated in the progression of transcription, replication and DNA repair (Barnett and Krebs 2011) and its removal led to increased chromatin compaction and was proposed to maintain chromatin structure in DNA replication (Poot et al. 2004). Both ACF and WSTF assemble as ISWI-type ATP dependent chromatin remodelers which are known to displace or evict the nucleosomes (Poot et al. 2005). Additionally, in yeast, INO80 complex was found to be localized at the origins of replication and its abundance increased upon DNA replication stress induced by hydroxyurea (HU) (Shimada et al. 2008). It was concluded that INO80 associated chromatin remodeling was particularly important for the resuming the replication at stalled replication forks (Shimada et al. 2008). Several studies have also indicated the role of INO80 in the progression of replication under the normal conditions in the absence of replication stress (Hur et al. 2010; H.-S. Lee et al. 2014; Vincent, Kwong, and Tsukiyama 2008).

The literature reviewed so far focused on the formation, structure, and function of chromatin inside the eukaryotic genome. In addition to chromatin remodeling, genome function is also influenced by the genomic neighbourhood of different genes along the chromosomal sequence as well as inside the 3D space of the nucleus. The detailed description of the genome organization and its influence on genome activity is given in the next part of this chapter.

1.6 Organization of genome in 1D

The non-random structure of a eukaryotic gene, as characterized by the presence of promoters, 5`UTRs, the alternative occurrence of exons and introns, and 3`UTRs indicates the sophistication and utility of the design in the eukaryotic genome. Besides the structure of the individual gene, past research showed that the eukaryotic genomes are highly structured and compartmentalized with respect to the composition of DNA bases along the chromosome. The genome architecture in 1D refers to the compositional compartmentalization and non-random arrangement of genes and other regulatory elements across the linear sequence of the DNA base pairs (Tarailo-Graovac and Chen 2013). Although the focus of this thesis is to understand the 3D organization and function of the yeast genome, the 3D organization is closely linked with the organization of genome in 1D (Naumova and Dekker 2010). Some of the important features of the 1D genome organization are presented in the following sections for the sake of interested readers.

1.6.1 Non-random distribution of GC content across the genome

One of the pioneering evidence for the ordered arrangement of functional elements along the

linear DNA came from the work of Bernardi et al. when they performed the NaCl based density gradient centrifugation experiments with the DNA isolated from vertebrates in 1985 (Bernardi et al. 1985). Due to the imperfect nature of isolation methods, the DNA would often break into random segments with sizes ranging up to hundreds of kilobases. In their work, Bernardi et al. showed the density gradient centrifugation of isolated genomic DNA often contained distinct bands which were separable based on their GC content and belonged to three major classes, which were termed as GC isochores. Further, the authors proposed that the distribution of genes, viral insertions, and other repeat elements was highly non-uniform across the vertebrate genomes (Bernardi et al. 1985). Although the existence of three types of GC isochores observed by Bernardi et al. was proved to be arbitrary upon the completion of human genome project, it was confirmed that GC content varies continuously and erratically across the human genome at the scale of ~300kb (Harris 2009) and that the large chunks of high and low GC content intersperse each other in the mammalian genome (figure 1.10).

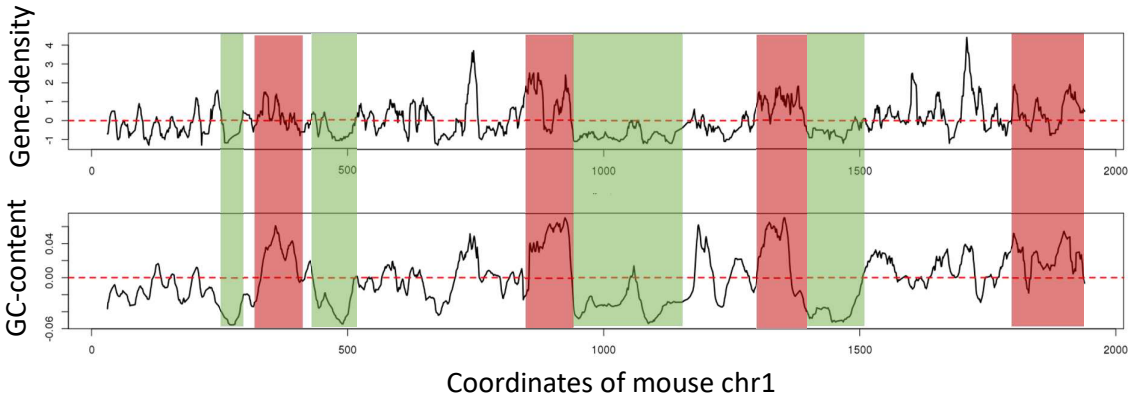

Figure 1.10 Distribution of GC content and gene-density across mouse chromosome 1.

The line plots show the variation of gene-density (upper panel) and GC content (lower panel). The red and green rectangles highlight the continuous regions of either high and low GC content or gene density, respectively. The lines were smoothened by using the moving average of 10 windows and GC content and gene density was median centered for visualization. Median values for both GC and gene-density correspond to the horizontal dotted red line.

1.6.2 **Gene dense and gene-poor regions**

Soon after the discovery of non-random distribution of GC content along the genome, it became apparent that the variation in GC content strongly correlated with the gene density of the local regions (Mouchiroud et al. 1991). In 1991, it was shown that 28% of the human genes reside within 3% of the human genome which represented the GC richest isochores (Mouchiroud et al. 1991). Further, it was found that the structures of genes present in different GC isochores were significantly different. By analyzing the sequences deposited in GenBank by 1995, it was revealed that genes present in GC poor regions coded for longer proteins and had longer introns as

compared to GC rich regions for many of the warm-blooded animals including humans, mouse, rat, cow, etc. (Duret, Mouchiroud, and Gautier 1995). Therefore, by the end of the 1990s, it became evident the non-random distribution of GC content was strongly correlated with the gene density creating the gene desert and gene-rich regions in the genome. An example of variation in GC-content and gene density across mouse chromosome 1 is given in figure 1.10.

1.6.3 **Repeat elements and their distribution in the genome**

Another important aspect of the 1D genome organization is the presence of repetitive DNA sequences (López-Flores and Garrido-Ramos 2012). Recent estimates have indicated that >65% of the human genome is composed of repetitive DNA sequences (de Koning et al. 2011). The evidence for the presence of repetitive DNA in the eukaryotic genomes came from the studies which investigated the re-association kinetics of DNA following its denaturation by high temperature (R. J. Britten and Kohne 1968; Roy J. Britten, Graham, and Neufeld 1974; Ohno 1972; Waring and Britten 1966). Though it was expected that mammalian genomes might take over months of time to re-associate because of the enormous amount of the DNA content (>2.5 billion DNA bases in mouse genome), certain regions were found to undergo surprisingly fast re-association for the case of the mouse genome (R. J. Britten and Kohne 1968). This observation hinted at the existence of repetitive DNA sequences which might have reduced the search space for the certain DNA strands undergoing re-association and consequently exhibited the faster re-association kinetics (R. J. Britten and Kohne 1968; Waring and Britten 1966). The initial analysis suggested the existence of about one million copies of repetitive DNA sequences in the mouse genome (Waring and Britten 1966). This analysis was termed as Cot-analysis as it involved the measurement of the concentration of DNA (C_0) and time (t) taken for the re-association of the dissociated DNA (Roy J. Britten, Graham, and Neufeld 1974; Peterson, Wessler, and Paterson 2002). Initially, the repetitive DNA sequences were believed to be non-functional (Ohno 1972). However, later studies have established that repetitive DNA elements have important roles in the regulation of gene expression (Shapiro and von Sternberg 2005). The repetitive DNA elements in the human genome were classified into two main types; Tandem repeats and Interspersed repeats (Biscotti, Olmo, and Heslop-Harrison 2015; Lander et al. 2001; López-Flores and Garrido-Ramos 2012; Richard, Kerrest, and Dujon 2008).

Tandem repeats. Tandem repeats consist of the DNA sequences which are consecutively arranged along the linear genome. The main constituent of tandem repeats is the satellite DNA and as mentioned in (Ellegren 2004), the origin of term ‘satellite’ DNA came from the investigations dating back to the 1960s when a distinct layer of DNA was observed in CsCl density-gradient centrifugation studies . Satellite DNA is further classified into three categories namely; satellites, microsatellites, and minisatellites (figure 1.11) depending upon the length of

repeat sequence (Biscotti, Olmo, and Heslop-Harrison 2015; Garrido-Ramos 2017; López-Flores and Garrido-Ramos 2012). The length of repeating subunit in microsatellites is <9bp and are also termed as short tandem repeats (Lander et al. 2001; López-Flores and Garrido-Ramos 2012). Microsatellites locus are highly polymorphic within the populations and exhibit several thousand-fold higher frequency of mutations as compared to the genome-wide average (López-Flores and Garrido-Ramos 2012). The expansion and contraction of the microsatellites are speculated to result from the replication slippage which occurs when elongating DNA-Polα encounters the stretches of direct repeats and might dissociate from the replicating DNA (Viguera, Canceill, and Ehrlich 2001). Such dissociation can result in the detachment of the newly synthesized strand from the template strand which might anneal to the nearby repeats having similar sequences, resulting in the expansion or contraction of the microsatellites through the mobilization of DNA repair machinery (Ellegren 2004; López-Flores and Garrido-Ramos 2012). Microsatellites mainly occur in the non-coding regions of the genome are considered to be non-functional but recent research indicates that microsatellites present near the coding regions might influence the disease phenotypes of various genetic disorders such as Huntington's disease (Gemayel et al. 2010; López-Flores and Garrido-Ramos 2012). The minisatellites have the repeat size of usually >9bp and their first discovery in the human genome came from the observation of restriction fragment length polymorphism (RFLP) within a 16kb fragment (Richard, Kerrest, and Dujon 2008; Wyman and White 1980). Such sequences were also termed as 'variable number tandem repeats' (VNTRs) (Richard, Kerrest, and Dujon 2008). DNA fingerprinting studies initially used the polymorphisms associated with VNTRs but later such studies started to use microsatellites owing to their greater abundance in the human genome (López-Flores and Garrido-Ramos 2012). The minisatellites in genomes of the most species including humans were mapped to the subtelomeric regions with the notable exception of budding yeast (Garrido-Ramos 2017; López-Flores and Garrido-Ramos 2012). Satellites consist of the longer repeat subunits (>100bp) and can be present in several hundred-thousand tandemly arranged copies in the human genome (López-Flores and Garrido-Ramos 2012). Satellites are mainly found in the centromeres and telomeres and are associated with repressed chromatin types (Nishibuchi and Déjardin 2017). The human α-satellites and murine ϒ-satellites are examples of satellite DNA (López-Flores and Garrido-Ramos 2012). In addition to satellites, rDNA and some paralogues genes also constitute a small proportion of tandem repeats in the mammalian genome (Biscotti, Olmo, and Heslop-Harrison 2015).

Interspersed repeats. The interspersed repeat elements are not clustered and are distributed throughout the genomes. The most common type of interspersed repeats are the transposable elements which were initially discovered by the Barbara McClintock when she observed that some genes could change their location in the chromosomes of Maize (Mcclintock 1956). The

existence of such mobile genetic element remained a highly controversial topic for a few decades after their initial discovery (Biémont 2010; Navarro 2017). However, the discovery of hybrid dysgenesis in *Drosophila*, which was characterized by the elevated mutation and recombination rates in its germline when wild-type males (P-type) were crossed with the females of the laboratory strain (M-type), reignited the interest in transposable elements within scientific community (M G Kidwell, Kidwell, and Sved 1977). Several studies established that transposition of *P*-elements was responsible for gene disruption leading to the hybrid dysgenesis (Ronsseray 1986; Rubin, Kidwell, and Bingham 1982). Hybrid dysgenesis was also observed to involve yet another transposable element called the *I* elements in *Drosophila* (Bucheton et al. 1984; Fawcett et al. 1986; Margaret G. Kidwell 1979).

The transposable elements have been classified into two classes; class-I and class-II (Feschotte and Pritham 2007; López-Flores and Garrido-Ramos 2012; Luning Prak and Kazazian 2000). *Drosophila I*-elements belong to the class-I transposons (RNA transposons) which move across the genome by copy-paste mechanism *via* RNA-intermediates and therefore, can amplify across the genome (Biémont 2010). On the other hand, the *P*-elements of Drosophila and transposons discovered by Barbara McClintock are the Class-II transposons (DNA-transposons) that moved across the genome by the cut-paste mechanism (Biémont 2010). Class-II transposons are present in ~350k copies in the human genome with MER1 and MER2 being their most frequent subtypes (Luning Prak and Kazazian 2000).

The class-I transposons can be of two types namely LTRs and non-LTRs (Kazazian H.H. 2000; Lander et al. 2001). The LTR-transposons are flanked by 'long terminal repeats' at their both ends and contain two genes which are similar to the *gag* and *pol* genes of retroviruses (Havecker, Gao, and Voytas 2004; Luning Prak and Kazazian 2000; Smit 1999; Xiong and Eickbush 1990). The first LTR-transposon to be discovered was the *Ty* element in the budding yeast genome which was present in 35 copies in the haploid genome with a consensus length of ~5.7kb (Cameron, Loh, and Davis 1979; Gafner and Philippsen 1980). In human genomes, most LTR-transposons have lost their ability to proliferate due to inactivation their associated genes and are termed as 'endogenous retroviruses' or the ERVs (Luning Prak and Kazazian 2000; Smit 1999). ERVs make up a total of 7% of the human genome with more than 240k copies distributed throughout the genome (Luning Prak and Kazazian 2000). The non-LTR transposons, on the other hand, are not flanked by long terminal repeats and are also referred to as retroposons (Lander et al. 2001; Luning Prak and Kazazian 2000). The mammalian retroposons are of two main types called 'long interspersed nuclear elements (LINEs) and 'short interspersed nuclear elements (SINEs) (Lander et al. 2001; Luning Prak and Kazazian 2000; Okada 1991). LINE elements constitute >20% of the human genome (Michel 2002) with more than a million copies scattered through the genome

and are enriched in the intergenic regions (Elbarbary, Lucas, and Maquat 2016; Medstrand, Van De Lagemaat, and Mager 2002). LINE-1 (L1) elements are the most common form of LINE elements and have the ability to transpose owing to the presence of reverse transcriptase genes (Kazazian H.H. 2000; Luning Prak and Kazazian 2000). Although the full-length L1 elements consist of 6.1kb of sequence, most of them are truncated at their 5` ends (Luning Prak and Kazazian 2000; Smit 1999). The SINE elements constitute almost 14% of the human genome with >1.7 million copies and their lengths vary between 100-400bp (Michel 2002). Further, the SINE elements are significantly enriched in the gene-proximal regions (Elbarbary, Lucas, and Maquat 2016; Medstrand, Van De Lagemaat, and Mager 2002). The most abundant SINE elements in the primate genomes are *Alu* elements (Schmid and Maraia 1992) which is believed to have originated from the 7SLRNA gene (Lander et al. 2001). Some of the *Alu* elements are still active in the human genome and can proliferate by borrowing the reverse transcriptase from the LINE-1 elements (Hancks and Kazazian 2016).

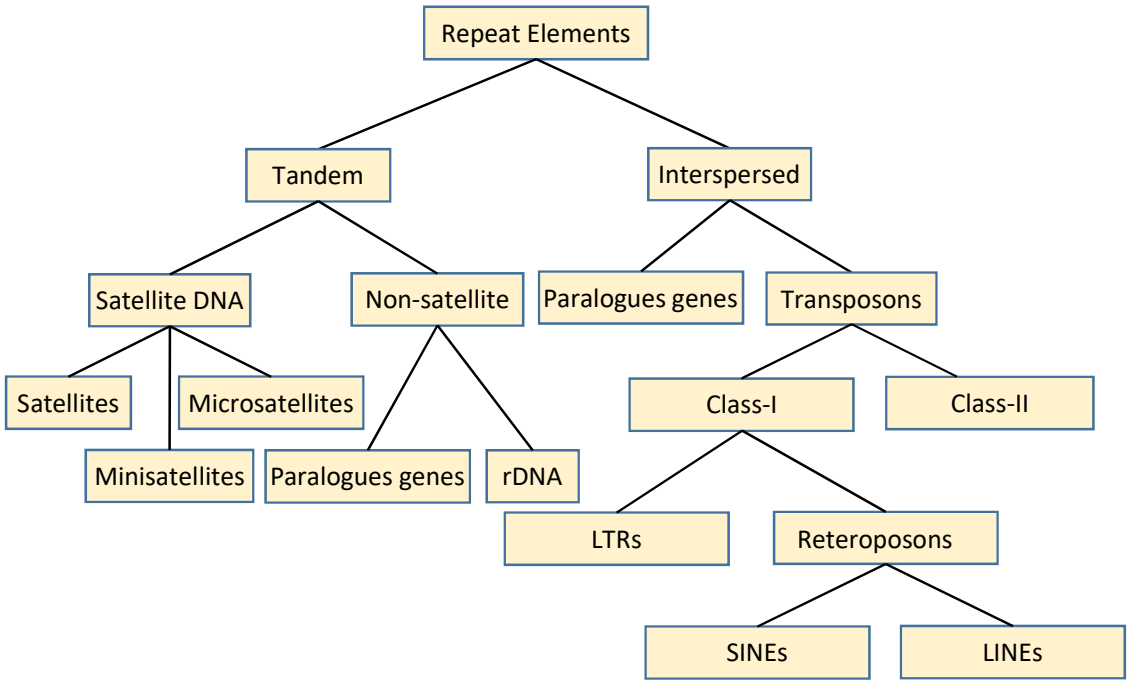

Figure 1.11 Classification of the repeat elements in the genome.

This figure is adopted from Richard, Kerrest, and Dujon (2008) with permission.

Together, the retroviral elements constitute almost ~45-50% of the human genome (Friedli and Trono 2015; Lander et al. 2001; Neidhart 2015). The most retroviral elements remain transcriptionally silent due to DNA methylation and their activation might cause the genomic instability by triggering mitotic recombination and random genomic insertions leading to various

genetic diseases including cancer (Hancks and Kazazian 2016; Kulis and Esteller 2010; Reilly et al. 2013). Some recent studies have also indicated that the transcriptional activity of transposable elements is tightly regulated during the early development and can influence the gene regulatory networks during mammalian development (Friedli and Trono 2015; Garcia-Perez, Widmann, and Adams 2016; Gifford, Pfaff, and MacFarlan 2013). Further, the transposable elements dependent alterations in the gene expression patterns were also proposed to be involved in the evolution of mammalian genomes (Cordaux and Batzer 2009; Franke et al. 2017; Mita and Boeke 2016; Platt, Vandewege, and Ray 2018).

1.6.4 **Distribution of CpG islands and DNA methylation**

The cytosine methylation was discovered back in 1948 when Hotchkiss obtained and analyzed the emission spectra of DNA bases (Hotchkiss 1948). More than a decade after the initial discovery, it was revealed that majority of the methylated cytosines were present at the CpG dinucleotide sites (Doskočil and Šorm 1962). Further the methylated cytosines were found to be highly enriched among the nuclease-resistant chromatin in various tissues of chicken (A Razin and Cedar 1977) and were significantly depleted near actively transcribed genes (Naveh-Many and Cedar 1981). Therefore it was believed that the distribution of methylated cytosines was highly non-random across the genome. The deployment of better molecular biology techniques over the following years revealed that non-methylated CpG dinucleotides were enriched near the active genes expressing in different tissues which hinted at the role of DNA methylation in gene repression (mentioned in Antequera and Bird 2018). It was revealed that almost 40% of the Sea urchin genome was resistant to the methylation-sensitive restriction enzymes (such as HpaII) while remaining 60% represented the non-resistant and methylated fraction of the genome (A. P. Bird, Taggart, and Smith 1979). Sea urchin genome was proposed to contain interspersed domains/compartments of methylated and unmethylated fraction (Antequera and Bird 2018; A. P. Bird, Taggart, and Smith 1979) in various tissues. But the vertebrate genomes lacked the compartmentalized pattern of methylated and unmethylated DNA and exhibited wide-spread methylation across their genomes (reviewed in Antequera and Bird 2018). However, the vertebrate genomes were found to contain a smaller fraction (~2%) of the genome with significantly enriched unmethylated CpG dinucleotides, which were present in short clusters of an average size of 120bp (D. N. Cooper, Taggart, and Bird 1983). Such regions were named as CpG islands (A. Bird et al. 1985; A. P. Bird 1987) and were derived from 1000bp long sequences which had very high (65%) GC content as compared to the genomic average of 40% in the mouse genome and were associated with widely expressed genes (A. Bird et al. 1985). Consistent with the role of CpG methylation in gene repression, the majority of CpG islands were found to be present near the 72% of the human genes which exhibited broader expression breadths (Saxonov,

Berg, and Brutlag 2006). The remaining 28% of genes were found to be present within the CpG depleted regions and expressed in a tissue-specific manner (Saxonov, Berg, and Brutlag 2006). Therefore, CpG content within human promoters exhibited bimodal distribution (Saxonov, Berg, and Brutlag 2006). The existence of CpG islands is likely to be an evolutionary selection against DNA methylation since the methylated cytosine can undergo deamination into thymine and can generate the dearth of CpG dinucleotide in a methylation majority genome (A. P. Bird 1980).

1.6.5 **Gene clusters**

In prokaryotes, it is well-known that many different genes can be controlled by a single promoter and are transcribed as a single unit in the form of polycistronic mRNA. Such co-regulated units of the bacterial genomes are known as operons and contain the genes involved in a single biological pathway. Although there are several examples of the polycistronic transcription in *C.elegans* (Spieth et al. 1993; Zorio et al. 1994) and some fungal species (Blumenthal 1998), most other higher eukaryotes have very few examples of genes organized as operons (Blumenthal 1998, 2004; Tarailo-Graovac and Chen 2013). However, most higher eukaryotes such as metazoans exhibit significant clustering of co-regulated genes along the linear genome (Hurst, Pál, and Lercher 2004). The presence of rDNA, clusters of HOX genes, Histone genes, α/β globulin, and several other genes are well-known examples of gene clustering involving the functionally related and spatiotemporally co-regulated genes (Elizondo et al. 2009; Hurst, Pál, and Lercher 2004). Therefore it was speculated the clusters of co-regulated genes might be a general feature of genome organization in higher eukaryotes.

The systematic investigations of gene clustering began by the end of the 1990s when it was reported that budding yeast had 25% of its cell-cycle regulated genes, which expressed in the same phase during the cell cycle, linearly adjacent to each other (Cho et al. 1998). Similar co-regulation dependent clustering was also noted for the genes which were specifically expressed in testis of *Drosophila* (Boutanaev et al. 2002) and for muscle-related genes in *C.elegans* (Roy et al. 2002). The clusters involving the co-expressed genes, with 10-30 members and average size of 100kb, were also reported by expression profiling in *Drosophila* (Spellman and Rubin 2002) and human genome (Caron et al. 2001; Dempsey et al. 2001; Gabrielsson, Carlsson, and Carlsson 2000; Y.-S. Yang et al. 2002). In particular, such clusters in the human genome were proposed to constitute ‘regions of increased gene expression’ (RIDGEs) having a high density of genes, SINE elements, and high GC content (Versteeg et al. 2003). In 2003, Lee and Sonnhammer showed that the genes involved in similar KEGG pathways were significantly more clustered in the genomes of five different species (J. M. Lee and Sonnhammer 2003).

Several studies have also attempted to explain the possible mechanisms underlying the emergence and maintenance of gene clusters. Many gene clusters were proposed to be comprised of

paralogues genes constituting a single gene family and were suggested to be evolved from whole genome and tandem duplication events (Demuth et al. 2006; Ferrier and Holland 2001; Lajoie, Bertrand, and El-Mabrouk 2010). But a majority of the linear gene clusters are not necessarily paralogues genes or even functionally related (Elizondo et al. 2009). The co-regulation dependent gene clustering might have resulted from the selective advantage provided by the concerted opening and closing of chromatin in these clusters which can assist in efficient gene regulation (de Wit et al. 2008) and was supported by the identification of large contiguous domains of chromatin bearing active and inactive marks (Ernst and Kellis 2010; Filion et al. 2010). Such concerted dynamics of chromatin was also shown to be important for co-regulation of developmental genes like HOX genes and α-globulin genes (van Heyningen and Bickmore 2013; Montavon and Duboule 2013). Concerted opening and closing of chromatin at gene clusters were also implicated in the reduction of gene expression noise by minimizing the stochastic fluctuations of chromatin domain boundaries (Batada and Hurst 2007; Wang, Lercher, and Hurst 2011). Therefore, gene clusters might have evolved due to the evolutionary selection against gene expression noise and not necessarily the co-expression of clustered genes (mentioned in Kustatscher, Grabowski, and Rappsilber 2017). This view was supported by the fact that co-expression of linearly proximal and bidirectional genes could be buffered and might not result in the correlated levels of protein abundance (Hurst 2017; Kustatscher, Grabowski, and Rappsilber 2017). Alternatively, the population genetics centric view of gene clustering asserted that evolutionary maintenance of linkage disequilibrium might have favored the formation of gene clusters because of selective advantage of linear proximity of particular genes (Hurst, Pál, and Lercher 2004; Sinervo and Svensson 2002). The examples of such cases (reviewed in Hurst, Pál, and Lercher 2004) include the essential gene clusters in yeast which exhibit low recombination rates (Pál and Hurst 2003), the mating type locus in *Chlamydomonas reinhardtii* where selection has promoted the linkage disequilibrium between chloroplast inheritance allele and mating type alleles (Ferris, Armbrust, and Goodenough 2002), and meiotic drive genes which have been shown to promote the inheritance of the genes/chromosomes in excess of the proportions predicted by the Mendelian laws (Lyttle 1991).

Therefore, the higher eukaryotes exhibit a widespread gene clustering (Elizondo et al. 2009; Fukuoka, Inaoka, and Kohane 2004; Hurst, Pál, and Lercher 2004; Michalak 2008) which might have evolved due to the enhancement of co-regulation, selection against gene expression noise, gene duplication events, or from the selective advantage of linkage disequilibrium between genes exhibiting epistatic effects between different alleles (Hurst, Pál, and Lercher 2004). It is worth noting that functionally related genes do not always exhibit the linear clustering and can be co-regulated even when located far apart along the linear genome or even on different chromosomes

(Hurst, Pál, and Lercher 2004; also reviewed in Lee and Sonnhammer 2003).

1.7 Genome organization in 3D

In addition to the non-random organization of genome in 1D, it is a well-established that the organization of chromatin in 3D is also non-random. To understand the organization of chromatin in 3D space of nucleus, extensive efforts have been put into the latter decades of the late 20th century and continue even today. This section summarizes the key features 3D genome organization and has been divided into two parts which are introduced below:

(i) *Microscopy-based studies,* which relied on high-resolution microscopy techniques such as electron microscope or confocal microscope to probe the 3D locations of whole chromosomes or different genomic regions.

(ii) *Chromosome conformation capture (3C) based studies,* which used a molecular biology-based technique called 'proximity ligation' coupled with NGS analysis.

3D genome organization in the pre-genomics era: From the lens of the microscope

1.7.1 Chromosome territories

The electron micrographs of interphase chromatin established that unlike the stiff and rod-shaped structures of chromosomes observed during metaphase of cell cycle, interphase chromatin is present in the form of highly flexible and loose threads of DNA resembling the 'beads-on-string' model (Olins and Olins 1974; Oudet, Gross-Bellard, and Chambon 1975; C. L. F. Woodcock, Safer, and Stanchfield 1976). However, the relative arrangement of chromatin of different chromosomes remained completely elusive, partly due to lack of chromosome-specific staining and visualization techniques. Back in 1885, Carl Rabl had proposed that each chromosome occupy only a discrete volume of the nucleus, thereby avoiding the intermingling of chromatin from other chromosomes. Following Carl Rabl's proposals of unique territory for each chromosome in the nuclear space, Theodor Boveri provided the initial evidence for territorial organization of interphase chromatin and coined the term chromosome territory (CT) (Boveri T, 1909; reviewed in Thomas Cremer & Cremer, 2010). In sharp contrast to the territorial organization, the chromatin appeared as randomly diffused polymer chain lacking any identifiable pattern (figure 1.12) in the early electron micrographs of the interphase nucleus (Wischnitzer 1973). These contrasting observations led to the dispute within the scientific community with regard to the 3D organization of chromosomes (reviewed in T. Cremer et al. 1993) and none the two possibilities was experimentally validated (figure 1.12).

The dispute was ultimately resolved with the landmark work of Cremer and Cremer (Cremer et al. 1982) where they stained the specific regions of the nucleus by supplying the radiolabelled nucleotides for repairing the localized DNA damage induced by a narrow beam of the laser. The

distribution of labeled nucleotides between different chromosomes strongly supported the idea that each chromosome occupies a discrete and fixed volume inside the nucleus, known as chromosomal territories (Cremer et al. 1982). During later years, chromosome territories were also successfully visualized using electron microscopy (Astrid E. Visser Françoise Jaunin and Aten 2000).

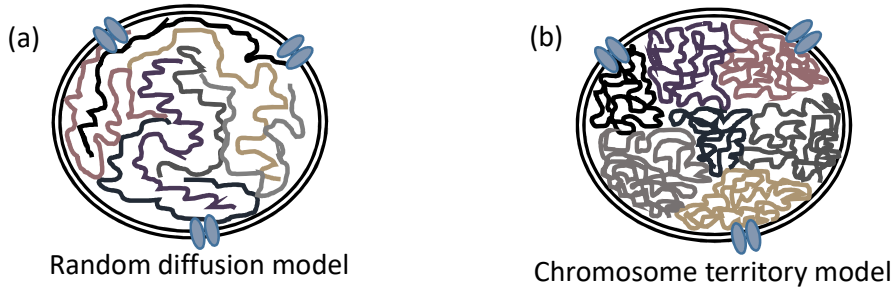

Figure 1.12 Random diffusion vs. territorial model of chromosomal organization.
Lines of different colours indicate 7 different chromosomes in a hypothetical nucleus. a). The random diffusion model: the chromatin fibres of different chromosomes are randomly scattered throughout the nuclear volume lacking any particular order. b). The territorial organization or CT model: Each chromosome folds within a specific sub-volume of the nucleus called chromosomal territory.

Therefore, it was unequivocally accepted that chromatin from different chromosomes is spatially segregated and each chromosome occupies its own unique territory (Cremer and Cremer 2001, 2010) allowing a limited, albeit significant amount of intermingling between different chromosomes (Branco and Pombo 2006).

Electron microscopy based visualization of chromatin organization was limited by the lack of chromosome-specific or gene-specific staining methods to probe the genome architecture of individual chromosomes or genes. This limitation was subjugated by the introduction of *in situ* hybridization in 1969 (Gall and Pardue 1969), which used sequence-specific RNA based probes to make RNA-DNA hybrids within the cellular nucleus. In the original method of *in situ* hybridization, the detection of RNA-DNA hybrids was performed using autoradiography. But the radiolabelled RNA probes were soon replaced by the fluorescent RNA probes for the detection of RNA-DNA hybrids; hence coupling the *in situ* hybridization with fluorescence microscopy. This technique was named as DNA FISH (an abbreviation for 'fluorescence *in situ* hybridization of DNA') (RUDKIN and STOLLAR 1977). Several studies began to use DNA FISH probes to paint the whole chromosomes or the large mega-base pair regions of chromosomes to investigate their arrangement in 3D space. Interestingly these studies reinstated the territorial view of chromatin organization (Branco and Pombo 2006; Cornforth et al. 2002; Parada et al. 2002). In addition to

recapitulating the territorial organization, the development of DNA FISH provided a powerful tool for probing the 3D location of genes in pursuit of questions pertaining to the relationship between the activity of genes and their position inside the chromosome territories. This led to many fundamental and useful insights about the arrangement of genes within CTs and are discussed in the following sections.

1.7.2 **Gene positioning within chromosomal territories**

The existence of the territorial organization of chromosomes during interphase led to the speculations that genes must be present at the surface of chromosome territories in order to be accessed by regulatory proteins. This hypothesis was also supported by the observed heterogeneity in the electron microscopic architecture of interphase nucleus (Shankar Narayan et al. 1967). Particularly, the interphase nucleus was found to contain the large and continuous network of fibrils and particles (Shankar Narayan et al. 1967). This network was largely composed of ribonucleoproteins, which was later found to constitute inter-chromatin space (Sui Huang and Spector 1991). Follow-up studies showed that the inter-chromatin space around the chromosome territories contained many of the splicing factors inside the sub-nuclear structures called nuclear speckles (Sui Huang and Spector 1991). Further, Zirbel et al. showed that the splicing machinery components such as snRNPs and transcripts of HPV genome were found to be excluded from the interior of the chromosome territories , (Zirbel et al. 1993), thereby leading to the proposals of inter-chromatin domain (ICD) organization of transcription and splicing or the CT-ICD model (Cremer et al. 2000). The CT-ICD model proposed that the transcription-related macromolecular complexes are exclusively present in the ICDs at the surface of CTs, and genes can be transcribed exclusively from the surface of chromosomes (Cremer et al. 2000; Kosak and Groudine 2002). Consistent with this model, the localization pattern of NLS-Vimentin fusion protein was strongly restricted at the surface of chromosome territories (Bridger et al. 1998; Reichenzeller et al. 2000). A pictorial representation of CT-ICD model of chromosome organization is presented in figure 1.11. This model was further supported by several DNA-FISH based studies which demonstrated that genes often looped out of their chromosomal territories in order to get transcribed (Beck et al. 1996; Dietzel et al. 1999; R. R. E. Williams et al. 2002). For example, the region containing Major histocompatibility complex (MHC) genes on human chromosome 6 looped out towards the surface of chromosome territories following the induction by interferon-gamma (Beck et al. 1996). Another study showed that active ANT2/ANT3 genes were more likely to present on the surface of chromosome territories(Dietzel et al. 1999). Similar observations were made for the DNA FISH studies of EDC gene in keratinocytes and gene-dense regions of high transcriptional activity, such as 11p15.5 at human chromosome 7, where actively transcribed chromatin frequently protruded towards the surface of chromosomal territory and in

the inter-chromatin space (Mahy, Perry, and Bickmore 2002; R. R. E. Williams et al. 2002). These studies collectively indicated that actively transcribed genes generally localize towards the periphery of chromosomes and often looped out of their chromosome territories. However, there were contradicting reports based on DNA FISH, which showed that the transcription statuses of genes were not linked with their positions w.r.t chromosome territories thereby making the interpretation of ICD models inconsistent and elusive (Abranches et al. 1998; Kurz et al. 1996). Later, Wendy Bickmore's group showed that active genes are distributed uniformly across chromosome territories and genes do not necessarily move out of chromosomal territories to get transcribed (Mahy et al. 2002; Morey et al. 2007; also reviewed in Thomas Cremer and Cremer 2010; & Heard and Bickmore 2007). These proposals were also consistent with the distribution of early and late replicating chromatin of chromosome 8 and X, which were found to be distributed throughout the volume of chromosomal territory, indicating that macromolecular complexes can be active even inside the chromosome territories (Astrid E. Visser Françoise Jaunin and Aten 2000).

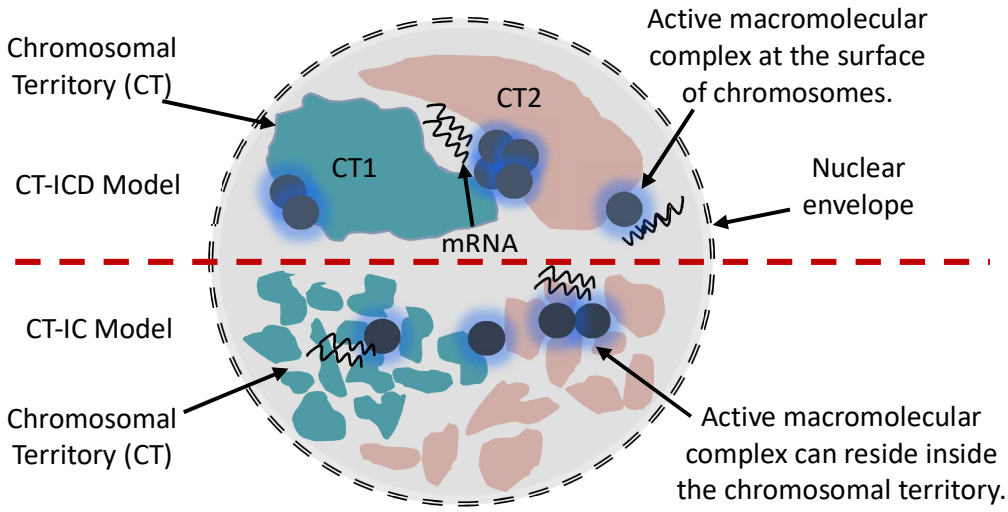

Figure 1.13 CT-ICD & CT-IC models of interphase chromosome organization.
Shown is the cartoon of the nucleus with two chromosomal territories (blue and pink). In the CT-ICD model (upper half), the macromolecular complexes exclusively reside at the surface of chromosomal territories inside the inter-chromatin domains. In CT-IC model, the chromosomal territories are invaginated by inter-chromatin compartments and transcriptional and other active complexes can reside even inside the chromosomal territories.
This figure is inspired from Kosak and Groudine (2002) and used with permission.

Revised CT-ICD model. The electron micrograph of the interphase nucleus presented by Fakan's group showed that ICDs were continuous structures that traverse through the volume of chromosome territories (Visser, Jaunin, and Fakan 2000). Interestingly, the distribution of TSSs

studied by immuno-staining of nascent RNA coupled with chromosome painting showed that active TSSs can be found distributed throughout the volume of chromosome territories (Verschure et al. 1999). Further, Verschure et al. found the presence of semi-continuous lightly stained substructures within chromosome territories, which had the diameters ranging from 300-450nm. Interestingly the nascent RNA was found to be exclusively localized in these lightly stained substructures within CTs (Verschure et al. 1999). Considering these observations in the light of Wendy Bickmore's work indicating the independence of gene transcription w.r.t its location in CT (Mahy et al. 2002), a revised model called CT-inter-chromatin compartment (CT-IC) was proposed (Kosak and Groudine 2002). This model stated that the invaginations present within the CTs can harbor the macromolecular complexes to carry out the genomic transactions (figure 1.13), and did not impose the strict requirement for the moment of active genes towards the surface of chromosomes (Cremer and Cremer 2001; Heard and Bickmore 2007; Kosak and Groudine 2002). But as noted by Heard and Bickmore , though the transcriptional machinery might not be excluded from inside of the chromosome territories, the concentration of transcription-related complexes might be particularly high at the surface of chromosome territories (Heard and Bickmore 2007). A recent high throughput FISH-based study, which used combinatorial labeling of TSSs in mouse, found that active TSSs were more likely to be present on the surface of chromosome territories (Shah et al. 2018). Therefore gene positioning with respect to chromosome territories is still a debatable topic and might require further research to assess its contribution in gene activation.

1.7.3 **Organization of heterochromatin**

In 1928, Emil Heitz proposed that chromosomes are made of two kinds of chromatin which differ in their staining intensity under the light microscope (Passarge 1979). Heitz referred these lightly stained regions of chromosomes as euchromatin and densely stained regions as heterochromatin. Extensive biochemical characterization established that heterochromatin is condensed and inactive form of chromatin with low gene density, late replication timing, lower levels of histone acetylation, and largely consists of repetitive DNA sequences (reviewed in Ekwall and Institutet 2011; Vanrobays, Thomas, and Tatout 2013). Additionally, heterochromatin was divided into two broad categories termed as constitutive heterochromatin (repetitive DNA sequences such as telomeres, centromeres, satellite DNA, transposons etc., marked by H3K9me3), and facultative heterochromatin (enriched at developmentally regulated genes, marked by H3K27me3) (Eissenberg and Elgin 2014; K. L. Jost, Bertulat, and Cardoso 2012). The constitutive heterochromatin was also marked by HP1 protein (Bannister et al. 2001; Hall et al. 2002; Lachner et al. 2001; Nakayama et al. 2001). HP1 is a chromo-domain protein and can bind with methylated form of H3K9 residues, and consequently recruits the histone methyltransferase SUV39H1 (mammalian homologue of yeast Clr4), which can further methylate the neighbouring

nucleosomes resulting in *cis*-spreading of heterochromatin (Bannister et al. 2001; Hall et al. 2002; Lachner et al. 2001; Nakayama et al. 2001) until blocked by boundary elements (Ekwall 2011; Kimura and Horikoshi 2004; Jiyong Wang et al. 2014). Another mechanism of heterochromatin formation that involves RNA interference (RNAi) has been described in fission yeast (Volpe et al. 2002). A few reports have also indicated the role of RNAi for heterochromatin formation in the mammalian genome as well (reviewed in Saksouk, Simboeck, and Déjardin 2015). The facultative heterochromatin, on the other hand, is deposited with polycomb group of proteins (PCG), which regulate the expression of the developmental genes (Dillon 2004; Gilbert et al. 2003; Politz, Scalzo, and Groudine 2013; Trojer and Reinberg 2007; Jiyong Wang et al. 2014; Jiyong Wang, Jia, and Jia 2016).

Mechanisms implicated in marking the heterochromatin boundaries have also been proposed. For example, nucleosome depletion at boundaries of heterochromatin can block the self-propagating wave of heterochromatin spreading (X. Bi et al. 2004; Xin Bi and Broach 1999). Further the RNA PolIII dependent transcription of tRNA genes was shown to serve as boundary elements for various heterochromatic regions in budding yeast, fission yeast, and in mammals (Cam et al. 2005; Donze and Kamakaka 2001; Ebersole et al. 2011; Raab et al. 2012; Scott, Merrett, and Willard 2006). Boundary formation has also been widely attributed to putative insulator protein named CTCF (Cuddapah et al. 2009; Moltó, Fernández, and Montoliu 2009; Saksouk, Simboeck, and Déjardin 2015; Jianrong Wang, Lunyak, and Jordan 2012), which is known for its role in blocking of the enhancer-promoter interactions (Phillips and Corces 2009). Some studies have also implicated the role of repetitive DNA elements in the demarcating the boundaries of heterochromatin (Elbarbary, Lucas, and Maquat 2016; Jianrong Wang, Lunyak, and Jordan 2012; Jiyong Wang et al. 2014). For example, the transcription of SINE element B2 in the mouse genome helped in the activation of tissue-specific genes by restricting the spread of heterochromatin (Lunyak et al. 2007). Several other mechanisms for the formation of chromatin boundaries have been proposed and are reviewed in the following papers (Kimura and Horikoshi 2004; Jiyong Wang et al. 2014).

Consistent with its distinct biochemical properties, the heterochromatin was shown to form a distinct compartment, typically near the nuclear periphery (K. L. Jost, Bertulat, and Cardoso 2012). The light microscopy studies of Emil Heitz showed that the densely stained heterochromatic regions of chromosome mainly localizes towards the nuclear periphery, or near the nucleolus assuming the more internal position (K. L. Jost, Bertulat, and Cardoso 2012; Vanrobays, Thomas, and Tatout 2013). These observations were also confirmed through the electron microscopic studies in the 1960s (R. Wang et al. 2015). Further, the centromeres and telomeres, which are the major constituents of constitutive heterochromatin, were found to be

preferentially localized near the nuclear periphery in yeast (reviewed in Zimmer and Fabre 2011) and required protein complexes like Sir2-Sir3 and yKu for this purpose (Zimmer and Fabre 2011). It must be noted that the peripheral arrangement of centromeres/telomeres (i.e. Rabl configuration, T. Cremer et al. 1993) was found to be common in plants (Cowan 2001) and fungal species (Idziak, Robaszkiewicz, and Hasterok 2015; Tiang, He, and Pawlowski 2012) but was less prominent in interphase organization of mammalian chromosomes which was dominated by the radial arrangement of CTs (Idziak, Robaszkiewicz, and Hasterok 2015). Nevertheless, heterochromatic regions of centromeres were shown to be preferentially localized near the nuclear periphery in human and mouse lymphocytes as well as in human embryonic stem cells (Weierich et al. 2003; Wiblin 2005). But the localization of telomeres towards nuclear periphery was somewhat disputed, with some reports showing that most human telomeres cluster towards nuclear interior (Amrichová et al. 2003; Weierich et al. 2003) or randomly scattered throughout the nucleus in mouse lymphocytes (Vourc'h et al. 1993) or can be attached to the nuclear matrix (Broccoli et al. 1997). A relatively recent report showed the enrichment of telomeres near the nuclear periphery using live cell imaging via time-lapse confocal microscopy in human fibroblasts (Crabbe et al. 2012). Likewise, the pericentromeric and telomeric heterochromatin in Drosophila was found to be concentrated at 'chromocenters' (Pimpinelli and Wakimoto 2003; Schulze et al. 2006) which were preferentially localized near the nuclear periphery (Marshall 2002).

Chromatin near the nuclear periphery involved the specific interactions with nuclear lamina, typically through B-type Lamins (Pickersgill et al. 2006). Such regions were termed as 'Lamina-Associated Domains' (LADs) and have been defined in *Drosophila,* mouse, and human (Guelen et al. 2008; Peric-Hupkes et al. 2010; Pickersgill et al. 2006). LADs were further classified as constitutive LADs (cLADs which remain invariable between different cell types) and facultative LADs (fLADs, which are shuffled between different cell types). Further, cLADs were found to contain A/T rich repetitive DNA sequences and repressed through H3K9me3 marks, while fLADs were found to contain tissue-specific genes, which were repressed through polycomb dependent H3K27me3 marks (Meuleman et al. 2013; van Steensel and Belmont 2017; Yáñez-Cuna and van Steensel 2017). Therefore, the nuclear periphery (or nuclear lamina) is typically viewed as repressive nuclear compartment (figure 1.14) whereas the nuclear interior is viewed as active nuclear compartment (Vanrobays, Thomas, and Tatout 2013). Nevertheless, there are some exceptions to this general rule. Firstly, nucleolus in mammals, which is placed towards the nuclear interior, is known to be surrounded by a significant amount of repressed chromatin (Padeken and Heun 2014). Nucleolus-Associated Domains (NADs) contain chromatin from centromeric and pericentromeric regions as well as repressed copies of rDNA genes (S. Huang 2012).

Secondly, some transcription activity, particularly for tissue-specific genes, has also been reported near the nuclear lamina (figure 1.14) particularly for the genes which are tethered near to the nuclear pore complex (C. R. Brown and Silver 2007; D'angelo 2018; Ibarra and Hetzer 2015; Ikegami and Lieb 2010; Starling 2017; Strambio-De-Castillia, Niepel, and Rout 2010). Thirdly, in the rod cells, heterochromatin has an exactly opposite configuration, where the euchromatin assumes the peripheral position and heterochromatin localizes towards the nuclear interior and it is considered to be an adaptation for vision (Solovei et al. 2009).

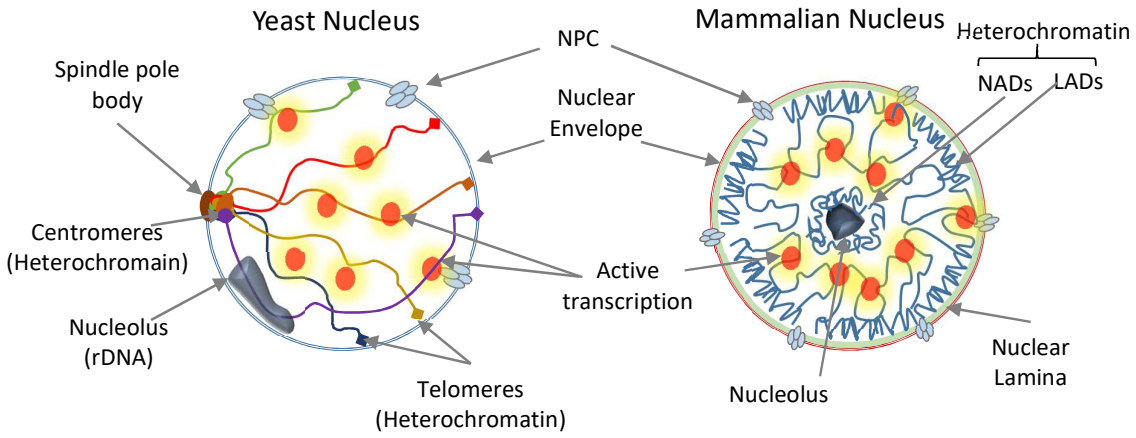

Figure 1.14 Organization of heterochromatin in yeast and mammals.
In yeast (left), the chromosomes adopt Rabl conformation and heterochromatic regions corresponding to centromeres and telomeres are localized near nuclear envelope, though on opposite sides. In mammals (right), the heterochromatic regions are either present near nuclear lamina or near the nucleolus. Though nuclear lamina is repressive compartment in general, active transcription (red dots) is observed at the nuclear pore complex (NPC).

Recent research has implicated the role of 'phase separation' in spatial compartmentalization of euchromatin and heterochromatin domains (Hyman, Weber, and Jülicher 2014). This proposal came from the observations that HP1 protein, which is a major constituent of heterochromatin, formed the phase-separated droplets through liquid-liquid demixing *in vitro* (Larson et al. 2017; Strom et al. 2017). Therefore, the compartmentalization of heterochromatin and euchromatin appears to be the outcome of biophysical properties of chromatin constituents under *in vivo* physiological conditions (Tatarakis, Behrouzi, and Moazed 2017).

1.7.4 **The positioning of chromosome territories w.r.t nuclear lamina**

In 1999, Wendy Bickmore and colleagues performed chromosome painting experiments to demonstrate that gene-rich human chromosome 19 is placed towards the center of the nucleus as compared to gene-poor human chromosome 18 (Croft et al. 1999). Later, the same group generalized these observations to other human chromosomes in lymphoblastoid and primary

fibroblast cell lines (Boyle 2001). Soon after this work, the comparative analysis of the spatial location of homologs of human chromosomes 18 and 19 in seven different primate species showed that gene-density dependent radial positioning of chromosomes was evolutionarily conserved (Tanabe et al. 2002), hinting at the strong evolutionary selection for gene-density dependent higher order chromosomal organization. These observations were further consolidated by later studies that focused on whole chromosomes (Bolzer et al. 2005) or megabase-sized chromatin domains (Shopland et al. 2006). Simultaneously, it was elucidated that GC-rich isochores in human chromosomes were localized towards the nuclear center as compared to GC-poor isochores which were localized towards the nuclear periphery (Saccone, Federico, and Bernardi 2002). Therefore, even for the same chromosome, GC rich regions exhibited preferential location towards the nuclear interior as compared to the periphery, which was consistent with the spatial segregation of euchromatin and heterochromatin as described in the previous section.

1.7.5 Gene repositioning to accommodate transcriptional changes

One of the first evidence for the dynamic localization of lineage-specific genes between different cell types of metazoans came from the DNA FISH-based studies of immunoglobin (Ig) loci, which become specifically activated in pro-B cells but remains repressed in pro-T cells. Kosak et al. showed that Igh and Ig-kappa loci, which were repressed in hematopoietic progenitors, were preferentially localized to the nuclear periphery in these cell types (Kosak et al. 2002). However upon differentiation, the Igh and Ig-kappa moved towards the center of the nucleus in Pro-B cells where these genes were specifically activated, but not in pro-T cells (Kosak et al. 2002). Further, it was found that switching in the expression and nuclear localization of Igh locus also resulted in the shift from late to early replication timing in pro-B cell lineage (J. Zhou et al. 2002).

Another example of such repositioning is the CFTR locus, which is present on the long arm of human chromosome 7. Three genes in this loci; GASZ, CFTR, and CORTBP2 are adjacent and functionally unrelated (Zink et al. 2004). Inactive CFTR and adjacent loci were found to be localized towards the nuclear periphery (figure 1.15). However, upon activation, these loci moved towards the nuclear interior into the euchromatic environment (Zink et al. 2004), thereby facilitating transcriptional activation (Figure 1.15). Transcription-dependent relocalization was also observed for CD4 and CD8 locus, which become stably repressed in T-cell lineage and moved towards nuclear periphery (Delaire et al. 2004), and for murine beta-globin locus which relocated to the nuclear interior following its activation in the erythroid lineage (Ragoczy et al. 2006).

Soon after the discovery of lineage-specific gene positioning in 3D, Tom Misteli's group published a comparative analysis of the position of subsets of chromosomes in different tissues of the mouse and found the tissue-specific spatial proximity between different chromosomes, which correlated with tissue-specific translocation frequency (Parada, McQueen, and Misteli

2004). These observations gave birth to the idea of tissue-specific genome organization governed by the transcription status of different genes and were proposed to have profound roles in chromosomal translocations observed in various cancers (Meaburn, Misteli, and Soutoglou 2007).

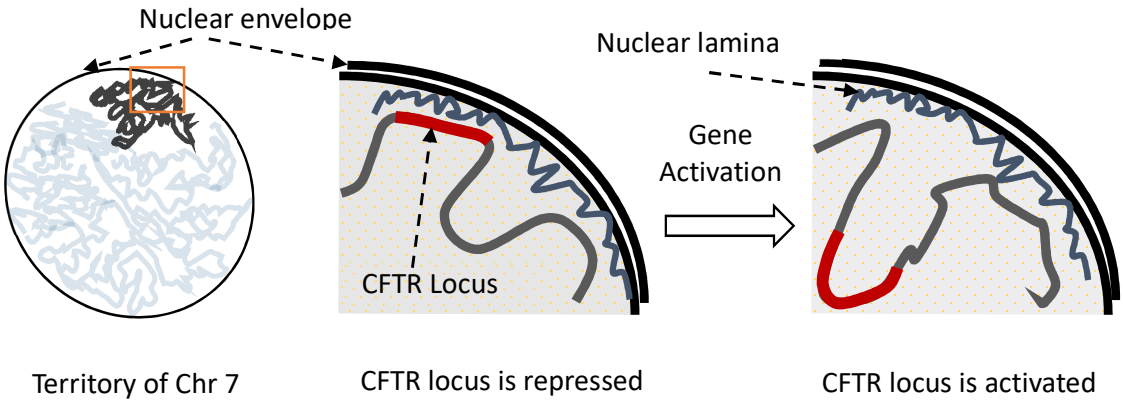

Figure 1.15 Dynamic gene positioning with respect to nuclear periphery.
The hypothetical territory of Chr 7 is highlighted by dark grey colour in the left panel. The inactive CFTR locus on human Chr 7 is localized near nuclear periphery, as shown in the middle figure. Upon activation, it relocates towards nuclear interior. The CFTR locus is indicated by red color.

Main highlights of 3D genome organization from microscopy-based methods

(i) Each chromosome is confined into a discrete nuclear volume called chromosomal territory.

(ii) Chromosomal territories are radially arranged in the nucleus, with gene dense and GC rich chromosomes preferentially located towards the nuclear interior.

(iii) Genes can be transcribed from the surface as well as from the inside of chromosomal territories.

(iv) Centromeres, telomeres, and other heterochromatic regions are preferentially located towards the nuclear periphery, and facultative chromosomal regions, such as those containing tissue-specific genes, are often dynamically placed in 3D space of nucleus in order to accommodate their changing transcriptional profiles.

(v) The gene positioning can be tissue-specific in metazoans.

3D genome organization in the post-genomics era: proximity ligation assay of chromatin

The research in 3D genome organization went through a paradigm shift when proximity ligation-based assays were employed to probe the architecture of interphase chromatin (Dekker et al. 2002). The original technique developed by Job Dekker and colleagues was termed as Chromosome Conformation Capture (3C, Dekker et al. 2002) and was soon modified to couple it with NGS analysis, thereby increasing its coverage and throughput which was unparalleled by

any of the microscopy-based methods. The development of 3C-based techniques and their impact on the understanding of 3D genome organization is described in the following two sections.

1.8 Development and evolution of 3C based techniques

The development of chromosome conformation capture based techniques to study the genome organization dates back to 2002 when Job Dekker and colleagues described an elegant method to test the physical proximity of known genomic loci (Dekker et al. 2002).

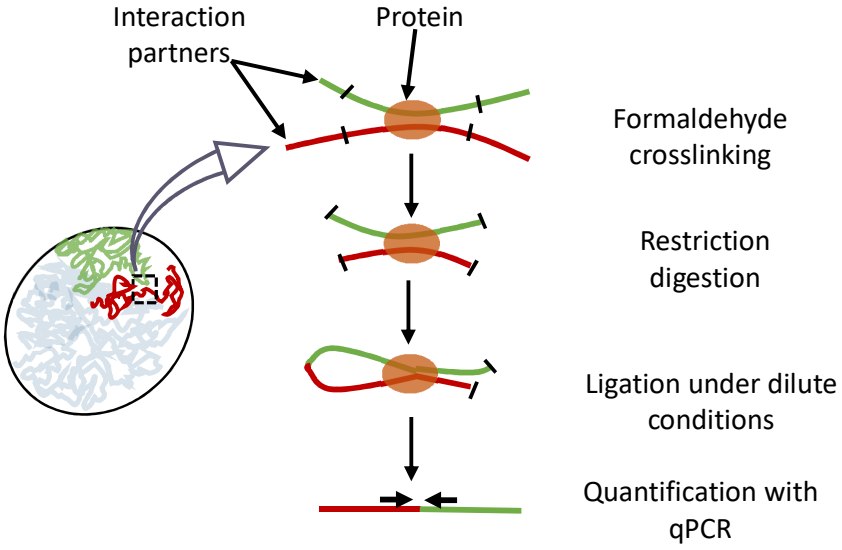

Figure 1.16 A schematic representation of 3C assay.

Shown on the left is the cartoon representation of a nucleus. Chromatin from two different chromosomes, indicated by red and green colours, is cross-linked with formaldehyde and subjected to restriction digestion. This is followed by ligation of the digested product under diluted conditions and ligated junctions are quantified using qPCR.

In their method, Job Dekker and colleagues used formaldehyde fixation to crosslink the spatially proximal chromosomal regions inside the nucleus. After fixing the architecture, the chromatin was subjected to restriction digestion, which generated spatially interacting pairs of restriction fragments. The restriction fragments were then subjected to intra-molecular ligation under diluted conditions, creating the ligated products. The ligated products created in this manner were purified after reversal of the formaldehyde crosslinks. The purified ligated products were then quantified through PCR using primers against known regions of interest (figure 1.16). Originally, Job Dekker and colleagues used 3C to probe the 3D organization in yeast chromosome III and recovered many of the known qualitative features of its chromosomal organization typically known from the microscopy-based methods (Dekker et al. 2002).

1.8.1 Circular Chromosome Conformation Capture (4C)

One limitation of 3C was that it required prior knowledge of candidate genomic regions to be probed, thereby restricting its scope to one-to-one interaction per assay. This limitation was partially alleviated by the development of 4C which represented the next generation version of 3C. The 4C was independently developed by two different groups (Simonis et al. 2006; Zhao et al. 2006) and is an acronym for two different versions of the technique: 'Circular-3C' and '3C-on-Chip' (table 3). However, both versions of 4C involved the formation of the circular DNA intermediates containing a known genomic region (bait or reference) and the unknown interacting partner. The unknown interacting partners in the circular DNA intermediates were amplified using the divergently oriented primers from the known genomic region. The amplified unknown interacting partner can either be hybridized to custom tiling arrays or sequenced directly using NGS (table 3, figure 1.18).

1.8.2 Chromosome Conformation Capture Carbon Copy (5C).

The scope of 3C-based techniques was further enhanced by the introduction of 5C, which could capture the spatial interactions among many different genomic regions (Dostie et al. 2006). In this technique, the single-stranded 3C DNA was annealed to a large number of sequence-specific primers from the ends of restriction fragments. These primers were flanked by universal overhangs (T7 and T3 sequences at 5' and 3' end respectively). The annealed primers got ligated to each other only when the complementary sites on 3C DNA were ligated generating the 'carbon-copies' of the 3C-library (figure 1.18). The resultant library was called 5C library and was quantified either through microarray or NGS.

1.8.3 HiC or the *'all vs. all'* method

The most popular among all 3C based method is the HiC (Lieberman-Aiden et al. 2009), which is often dubbed as the 'all-vs-all' method and can be used to capture all the possible chromatin interactions in any given fully sequenced species. In this technique, the restriction digested chromatin fragments were labeled with the biotin molecule which helped in the purification following the ligation of the digested ends (figure 1.17). The ligated DNA molecules were subjected to the sonication treatment which sheared the DNA and resulted in its random fragmentation. Size selection was performed on the sheared DNA fragments which were purified using biotin pull-down assay. The majority of the purified fragments were expected to bear the sequence from two different restriction fragments and hence, capturing the spatial interaction between two different genomic regions. The purified DNA fragments were then ligated to the universal sequencing adaptors *via* blunt end ligation following the repair of broken DNA ends. The adaptor-ligated DNA fragments were then sequenced using paired-end NGS, which are then aligned to the reference genome to infer genome-wide chromatin interaction frequencies. The sequencing reads, which originate from the self-ligated or un-ligated restriction fragments are

discarded during alignment by the most software used for HiC analysis as they do not contain information about spatial interactions (Wingett et al. 2015).

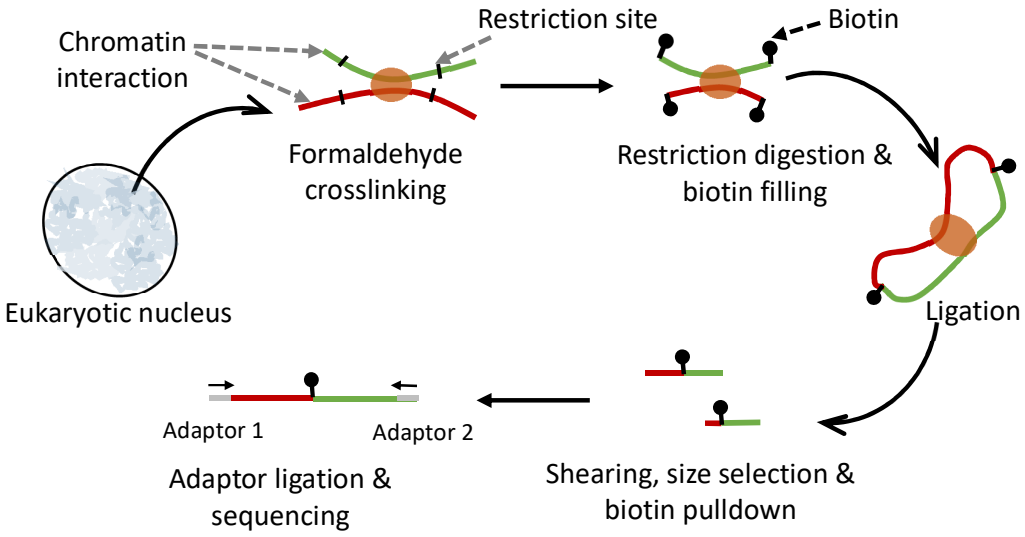

Figure 1.17 A schematic representation of HiC.

Chromatin is fixed using formaldehyde and subjected to restriction digestion. The digested ends are filled with biotin and subjected to ligation either *in-vivo* or in-dilution. The ligated product is then subjected to shearing, size selection and purification using biotin pull-down assay. The purified ligation junctions are then attached to the known sequencing adaptors for paired end sequencing.

In a HiC experiment, the average profile of genome-wide chromatin interactions is probed for a population of millions of cells. To probe and quantify the inter-cellular variability in chromatin organization, single-cell HiC method was developed by using the micro-pipetting to isolate the single cells followed by crosslinking, digestion, and ligation which were performed inside the intact nuclei (Nagano et al. 2013). Another variant of single cell HiC, which was termed as single-cell combinatorial indexed HiC (sciHi-C), made the use of unique barcoding of chromatin fragments from the individual cells by distributing the population of cells on a 96-well plate in a combinatorial fashion (Nagano et al. 2017; Ramani et al. 2017).

1.8.4 Chromatin Interaction Analysis using Paired-End Tag sequencing (ChIA-PET).

This technique was proposed almost simultaneously with HiC and used the ChIP experiment against specific antibodies after digesting the cross-linked chromatin fragments (Fullwood, Liu, Pan, Liu, Xu, et al. 2009). ChIP resulted in the enrichment of chromatin interactions anchored by a specific protein of interest (such as RNA-PolII or GTFs). The digested DNA fragments were then sonicated and the adaptors containing the restriction site of MmeI were incorporated at the open ends. These fragments were subjected to the ligation reaction generating the circular DNA molecules which were digested with MmeI. Since MmeI cuts the DNA 20bp away from the

restriction site, the digested products contained the sequence from both the interaction partners. The captured junctions were then sequenced using paired-end tag sequencing (figure 1.18).

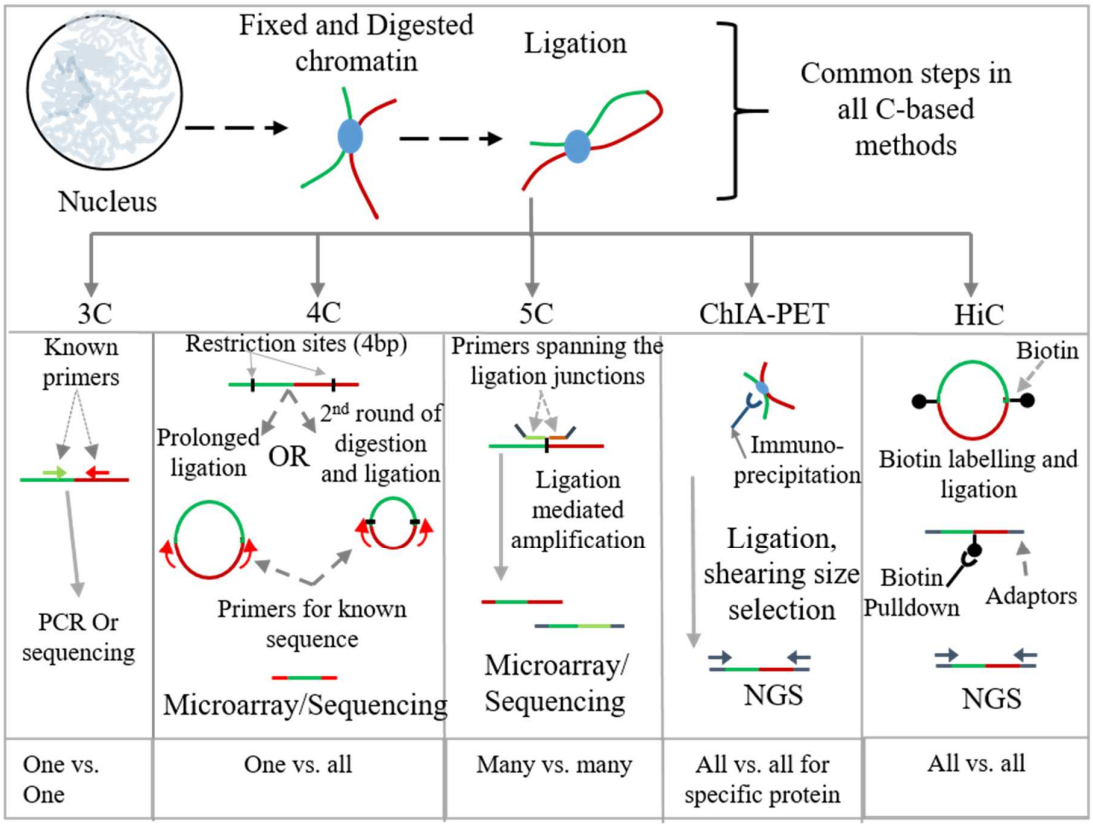

Figure 1.18 Summary of steps involved in commonly used C-based techniques.

This figure is redrawn from de Wit and de Laat (2012)

1.8.5 **Other variants of HiC**

Several other variants of 3C-based techniques were developed to probe the large scale organization of the genomes (Horike et al. 2005; Hsieh et al. 2015; Kalhor et al. 2012; Ma et al. 2014; Martin et al. 2015; Ramani, Shendure, and Duan 2016; Rodley et al. 2009). For example, a technique highly similar to HiC, called genome conformation capture (GCC), involved the spiking-in of the ligated chromatin interaction pairs with pUC19 plasmid to estimate the coverage and was first employed to study the genome-wide chromatin interaction landscape in budding yeast. Also, the GCC library was nebulized to generate the fragments of 150-180bp for the purpose of sequencing (Rodley et al. 2009). Another important variant of 3C was called capture-HiC and involved the enrichment of specific genomic regions of interest by using a microarray-based platform (Dryden et al. 2014). This technique was first used probe the long-range interactions between breast cancer susceptibility loci and their target genes and was subsequently used by many other studies (Jäger et al. 2015; Martin et al. 2015; Mifsud et al. 2015). A slightly

modified variant of capture HiC was recently developed to profile the genome-wide interaction landscape of genic promoters and was termed as 'promoter capture HiC' (Schoenfelder et al. 2018). In this technique, the usual HiC library was enriched for genomic fragments containing promoter sequences by the use of biotinylated RNA baits in solution which were then purified and sequenced (Schoenfelder et al. 2018). Besides these, DNAse HiC (Deng et al. 2015; Ma et al. 2014) and micro-HiC (Hsieh et al. 2015) are two another variants of HiC technique (Ma et al. 2014). DNAse-HiC used the DNAse-I enzyme to digest the cross-linked chromatin which improvised the resolution limit imposed by the distribution of restriction sites in the standard HiC experiment and was first used to map the interaction landscape of *cis*-regulatory elements at high resolution (Ma et al. 2014). Similarly, Micro-HiC used the MNase digestion to fragment the cross-linked chromatin which generated the mono-nucleosomes and the ligation products thereof, and was originally used to probe the 3D chromatin interactions in budding yeast at the nucleosome resolution (Hsieh et al. 2015). A brief description of these variations of 3C is included in table 3.

Table 3 A summary of 3C-based techniques

Name of Method	Scope	Quantification	Reference
Chromosome conformation capture (3C)	One vs. one	qPCR	(Dekker et al. 2002)
Chromosome conformation capture on chip (4C)	One vs. all	Array/NGS	(Simonis et al. 2006)
Circular Chromosome Conformation Capture (4C)	One vs. all	Array/NGS	(Zhao et al. 2006)
Chromosome Conformation Capture – Carbon Copy (5C)	Many vs. many	Array/NGS	(Dostie et al. 2006)
HiC (High throughput 'C')	All vs. all	NGS	(Lieberman-Aiden et al. 2009)
Chromatin Interaction Analysis using Paired End Tag sequencing (ChIA-PET)	All interactions anchored by specific protein	NGS	(Fullwood, Liu, Pan, Liu, Xu, et al. 2009)
Genome conformation capture (GCC)	All vs. all	NGS	(Rodley et al. 2009)
Capture-HiC (CHiC)	All vs. all for genomic regions of interest	NGS	(Dryden et al. 2014)
Promoter capture-HiC (PCHi-C)	Promoters vs. rest of the genome	NGS	(Schoenfelder et al. 2018)
DNAse-HiC	All vs. all	NGS	(Ma et al. 2014)
Micro-C	All vs. all	NGS	(Hsieh et al. 2015)

1.9 Impact of 'C' based methods on the understanding of 3D genome

1.9.1 3C experiments and chromatin looping

The enhancer activity of *cis*-acting sequences was first discovered in the studies involving the

transfection of mammalian systems with β-globin containing plasmid. It was found that the transcription of the beta-globin gene was significantly enhanced (~200 fold) with the co-transfection of SV40 DNA in the beta-globin plasmid (Banerji, Rusconi, and Schaffner 1981). The transcriptional enhancing activity was attributed to the presence of 72bp sequence present in *cis* (Banerji, Rusconi, and Schaffner 1981). Later, it was confirmed that 72bp sequence in SV40 can potentiate transcription from the native as well as substitute promoters (Wasylyk et al. 1983). Simultaneously, the enhancer activity was discovered within two different regions of the immunoglobin heavy chain region (Banerji, Olson, and Schaffner 1983; Gillies et al. 1983). A few years later, the enhancer activity was also reported in flanking DNA regions of the human beta-globin locus, which became active in a tissue-specific manner (Grosveld et al. 1987). A common theme that emerged from all these studies was that enhancers can be present several kilobases away from their target genes, and activated their targets in a tissue-specific manner (Rogers and Saunders 1986). However, the mode of action of the distal regulatory effect of enhancers remained to be determined. It was hypothesized that distal enhancer must come in close proximity to its target gene in order to exert its effects. In 2002, Wouter de Laat and colleagues used 3Cassay to test the conformation of the *cis*-acting locus control region (LCR) in the murine beta-globin locus, which was located 40-60kb away from its target genes (Tolhuis et al. 2002). The authors found that LCR of beta-globin locus came in close contact with its target genes while intervening chromatin looped out to facilitate the spatial proximity between the enhancer and target promoter (Tolhuis et al. 2002). Soon after this work, Palstra et al. published a detailed structure of active chromatin hub (ACH) consisting of β-globin genes along with multiple hypersensitive sites (HS) within its LCR in mice and humans (Palstra et al. 2003). It was revealed that multiple looping interactions occur between hypersensitive sites and active β-globulin genes, which got dynamically remodeled along the developmental trajectory following the activity status of β-globulin genes (Palstra et al. 2003). It was established that the formation of ACH for β-globulin locus required the multiple interactions with its LCR (Patrinos et al. 2004) and the presence of transcription factor EKLF (Drissen et al. 2004). Similar chromatin looping with distal regulatory elements was also observed between the mouse cytokine T_H2 locus and its LCR region located near the 3` end of the *Rad50* gene (Spilianakis and Flavell 2004). In another study employing 3C, the chromatin conformation around the *Ifng* gene was investigated during T-helper cell differentiation (Eivazova and Aune 2004). It was found that chromatin around *Ifng* locus is highly flexible and differed in its compaction between Th1 and Th2 subpopulations of T-helper cells (Eivazova and Aune 2004). Besides the globulin genes, the 3C assay was also used to explore the mechanism of *TNF* gene activation (Tsytsykova et al. 2007). In their investigation of *TNF* gene activation, authors found that two of the distal regulatory elements; HSS-9 and HSS4, which

were present a few kb upstream and downstream of *TNF* gene, respectively, were involved in specific looping with *TNF* gene upon T cell activation (Tsytsykova et al. 2007).

Over the next few years, several studies demonstrated the chromatin looping using the 3C-based assays (Ansari and Hampsey 2005; L. J. Burke et al. 2005; Gheldof, Tabuchi, and Dekker 2006; Kurukuti et al. 2006; Singh and Hampsey 2007; Teferedegne et al. 2006; Tsytsykova et al. 2007; de Wit and de Laat 2012; G.-L. Zhou et al. 2006). Some studies also attempted to explore the mechanism involved in loop formation and activity regulation of distal regulatory elements using the 3C technique (Drissen et al. 2004; Kurukuti et al. 2006; G.-L. Zhou et al. 2006). For example, Kurukuti et al., in 2006, showed that CTCF binding to imprinting control region (ICR) at *H19* resulted in the inactivation of maternal *Igf2* gene by looping it away from the enhancers located inside of the active chromatin hub (figure 1.19). A similar looping out event was also documented for the CFTR locus, where the DHSs located at the 3'end of CFTR gene moved in close physical proximity to its TSS by looping out the intervening chromatin (Blackledge et al. 2009).

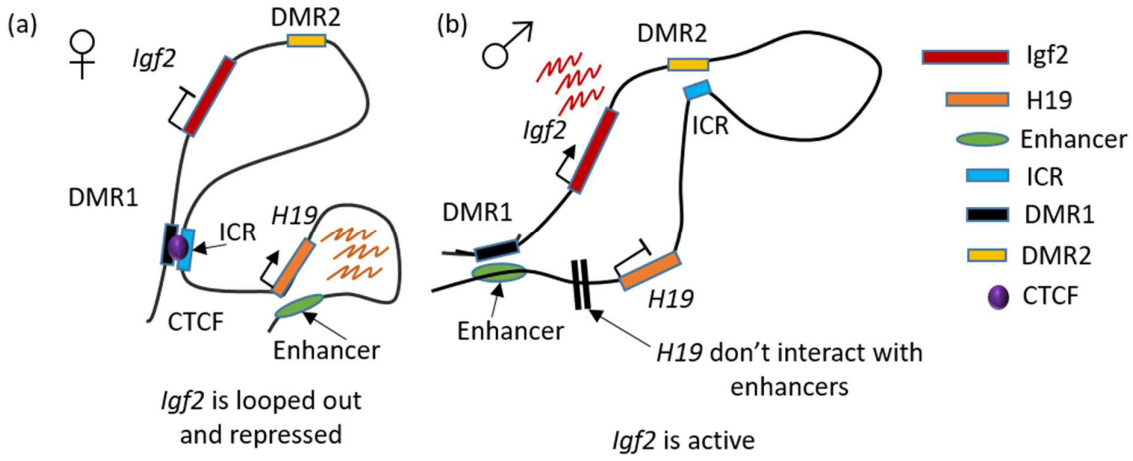

Figure 1.19 Chromatin looping involved at *H19/Igf2* locus.

a). At maternal locus, the binding of CTCF at ICR of *H19* brings it in close proximity to DMR1 and loops out *Igf2* making it inaccessible to the enhancer while *H19* remains proximal to enhancer. b) At the paternal locus, DMR2 interacts with the methylated ICR where CTCF is not bound which allows the enhancers to interact with DMR1 resulting in the activation of *Igf2*. *H19* remains silent due to methylation spread on and around promoter proximal ICR.

1.9.2 The 4C & the discovery of intra- and inter-chromosomal interaction hubs

In the pioneering 4C work carried out by Zhao et al., the authors investigated spatial organization of *H19* locus and found that many of the 4C interactions for the imprinting control region of the maternally inherited *H19* locus were specific (Zhao et al. 2006). Several differentially methylated regions (DMRs) interacted directly with the *H19* locus and the interactions were found to be linked with the epigenetic regulation of interacting loci in *trans* (Sandhu et al. 2009; Zhao et al.

2006). Further, the spatially interacting partners of *H19* locus were significantly reprogrammed upon the *in vitro* differentiation of ESCs. A key message from these studies was that a locus can *trans* regulate the epigenetic states of distant loci in an allele-specific manner through long-range chromatin interactions.

In the other version of 4C technology, Simonis et al. investigated the genomic neighborhood of β-globulin locus in mouse embryo at stage E14.5 in liver and brain samples (Simonis et al. 2006). They found that active β-globulin locus in mouse liver interacted with a completely different set of highly transcribed genes as compared to the inactive β-globulin locus in the mouse brain, which interacted with transcriptionally repressed genes. These findings implied that the eukaryotic genome was spatially segregated into active and inactive domains, consistent with the observed segregation of euchromatin and heterochromatin described in the previous section. Additionally, the Simonis et al. also observed that the spatial interactions of a housekeeping gene present on the chromosome 8, which interacted with active gene clusters in *cis* as well as *trans*, did not differ significantly between two tissues investigated. This finding suggested that the genes having conserved expression profiles between cell types might also have a conserved spatial organization in 3D. Taken together, both these observations indicated the specificity of 3D interaction partners in *cis* and *trans,* which could undergo significant remodeling to accommodate tissue-specific transcriptional programmes. Similar conclusions were also drawn from a later 4C study investigating the α and β-globin loci in erythroid tissues (Schoenfelder et al. 2010).

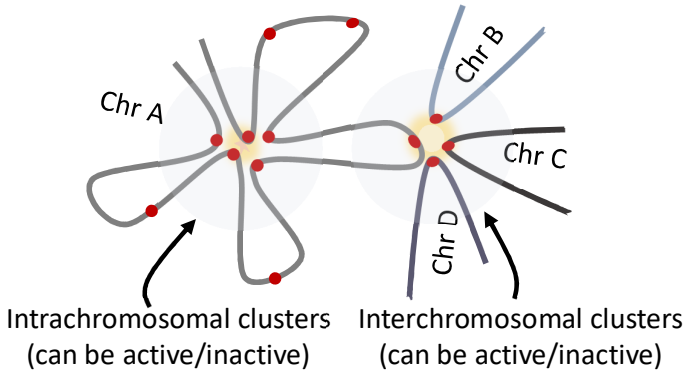

Figure 1.20 Chromatin interaction pattern uncovered by 4C.

Several different genes were often found to make intra- or interchromosomal spatial clusters. Such clusters can be comprised of actively transcribing genes or inactive genes. Red dots represent genes.

These observations supported the microscopy-based reports of the tissue-specific organization of 3D genome and established the importance of 3D-neighborhood of genes in regulating the thier function. Additionally, these two studies revealed the prevalence of inter and intrachromosomal interactions which were marked by the segregation of active and inactive genes as shown in figure

1.20. The observations that a locus can *trans*-regulate the epigenetic states through spatial contacts highlighted the network view of epigenome regulation (Sandhu et al. 2009).

1.9.3 HiC and the 'topologically associated domains' (TADs)

The non-random nature of chromatin interactions between *cis*-regulatory elements and their target genes and the intra-/inter-chromosomal clustering of different genes (figure 1.20) raised the speculations that such interactions might be widespread across the whole genome (A. Williams and Flavell 2008). However, none of 3C, 4C, and 5C techniques was suitable to measure the interaction frequencies among genes at the genome-wide scale. This technical limitation was alleviated with the introduction of an upgraded version of 3C technique which coupled the identification and quantification of interacting loci with NGS analysis. This technique was named as 'HiC' and did not required the prior knowledge of any of the interacting partners and was suitable to measure the all-to-all genome-wide interaction frequencies as described in the previous section (figure 1.17, table 3).

The sample space for possible interactions for HiC experiment in the mammalian genomes was billions of interacting pairs pertaining to the genome-scale measurements. In such a large sample space, the frequency of false positives due to random ligations and other technical reasons can significantly affect the quality of HiC data. Secondly, to get the faithful coverage for large genomes such as mammals, the sequencing depth of HiC experiment was required to be very high. Because of these reasons, the genomic coordinates of interacting regions obtained from HiC experiment were aggregated into bins of uniform size (for eg. 100kb) by summing up the observed interaction frequencies among the restriction fragment pairs mapping within a given bin pair (Lieberman-Aiden et al. 2009). The binned interaction frequencies were compiled into a contact map (figure 1.21). The size of the bins used was called the '*resolution of HiC data'* and ranged from 500kb to 1mb during the early days of HiC. However, during the subsequent years, the NGS sequencers achieved greater sequencing depth even for large genomes such as mouse and humans, which significantly enhanced the resolution of HiC data. Therefore, the 3D organization of genomes could be probed even at a kilobase resolution (Rao et al. 2014).

The first HiC map published by Lieberman-Aiden et al. 2009 recaptured the known feature of the large-scale 3D genome organization such as chromosomal territories, spatial clustering of small chromosomes, etc. In particular, it was proposed that the 3D genome is partitioned into two compartments namely; the 'A' compartment, which contained active chromatin, and the 'B' compartment, which contained repressed chromatin (Lieberman-Aiden et al. 2009). This was consistent with the microscopy-based observations pertaining to the segregation of euchromatin and heterochromatin. Soon after that, several studies employed HiC to probe the large scale 3D organization of genomes in various model systems. For example, in the year 2010, Duan et al.

used a technique highly similar to HiC to generate a genome-wide contact map of budding yeast and recovered many of the classically known features of yeast genome organization such as Rabl configuration, clustering of centromeres/tRNA genes, the spatial clustering of ribosomal genes and the clustering of early origins of replication (Z. Duan et al. 2010). Another group published the 3D map of fission yeast using a technique similar to HiC and recaptured the previously characterized features of fission yeast genome organization in addition to discovering functional relationships between spatially proximal genes (Tanizawa et al. 2010).

A major breakthrough in the understanding of large scale 3D genome was made when Bing Ren's group generated a genome-wide contact map in mESCs, hESCs and human fibroblasts (IMR90 cell line) (Dixon et al. 2012). In this work, authors observed that large megabase-sized consecutive regions of chromatin tend to preferentially engage in self-interactions and two neighboring domains were significantly depleted for inter-domain interactions (figure 1.21, 1.23).

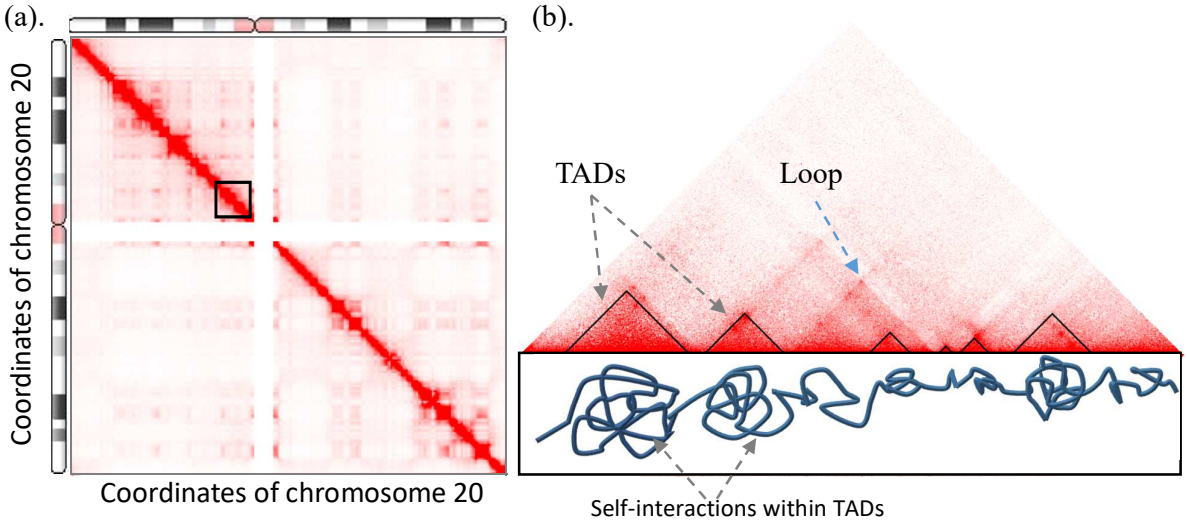

Figure 1.21 Contact map of HiC data.

a). Heatmap of HiC data of Chr20 in GM12878 cell line at 250kb resolution. X and Y axes are the coordinates of Chr20. Ideograms were placed approximately to emphasize the chromosomal positions and rendered from *https://www.ncbi.nlm.nih.gov/genome/tools/gdp/*. A dense pattern of interactions near diagonal indicates the presence of TAD and white streaks correspond to the un-mappable regions of the chromosome 20. b) Subsection of Chr20 (21mb-25mb) at 5kb resolution (zoomed image of highlighted box on the left) showing only upper triangle of HiC matrix. TADs are marked with the black lines. Loop (indicated by blue arrow) is characterized by strong HiC interaction frequency inspite of being far apart on linear genome. The cartoon below indicates that chromatin interactions are enriched within TADs and depleted between TADs. Heatmaps shown in a and b were plotted using online tool called *Juciebox* available at *http://aidenlab.org/juicebox*

Such domains were termed as Topologically Associated Domains (TADs) and were found to be

largely conserved between different cell types, and different species (Dixon et al. 2012). Several studies showed that TAD-borders were conserved between different cell types of the same species, and were evolutionary conserved among different species (Dixon et al. 2012; Krefting, Andrade-Navarro, and Ibn-Salem 2018; Rao et al. 2014; Vietri Rudan et al. 2015). However, a significant number of TADs switched between the A/B compartments during differentiation of human embryonic stem cells (hESC) (Dixon et al. 2015). Further, the chromatin interaction landscape within TADs was also suggested to undergo significant remodeling that correlated with the changes in the epigenetic profiles of the involved regions (Dixon et al. 2015).

The borders of TADs were also found to mark the transition between early and late replicating genomic regions in human and mouse cell types, which led to the proposal of TADs as stable units of DNA replication (Pope et al. 2014). To understand the developmental origin and maintenance of TADs, several recent studies investigated the 3D genome organization in gametes (Battulin et al. 2015; Ke et al. 2017) and tracked the changes in 3D genome from sperm/oocyte nuclei to embryo through zygote (Du et al. 2017; Flyamer et al. 2017; Ke et al. 2017). The findings from these studies suggested that mature sperm have TADs and compartments just like somatic cell types (Battulin et al. 2015; Ke et al. 2017). But the TADs were not well defined in the mature oocytes at the stage of meiosis II (Du et al. 2017; Ke et al. 2017). Another study, which employed the 'single nucleus HiC', showed the existence of weak TAD structure in oocytes (Flyamer et al. 2017). Collectively, the findings from these studies indicated that the TADs as well as compartments are obscure in the zygotic genome, but are gradually established during the early embryonic development at the 8-celled stage in mice and nuclear cycle 14 in *Drosophila* models (reviewed in Hug and Vaquerizas 2018).

An attractive, but presently debated model of TAD formation is the 'loop-extrusion' model which proposed that proteins like Cohesins can extrude the loops of chromatin. The extrusion of loops is blocked by the presence of architectural proteins, such as CTCF, at the boundaries of TADs (Fudenberg et al. 2016). This model is supported by the observation that Cohesin removal resulted in the elimination of all the loop domains (Rao et al. 2017). Alternatively, the formation of TADs was proposed to result from the transient interactions between the nucleosome rich inactive chromatin and the TAD boundaries were actively transcribing chromatin regions which separated the consecutive TADs (Ulianov et al. 2016). However, the presence of TADs in the transcriptionally silent sperm nuclei contradicts this proposal (Battulin et al. 2015). Further, the formation of TADs was also proposed to be dependent on DNA replication and not transcription (Ke et al. 2017). Therefore, more research is required to decipher and consolidate the exact mechanisms which might underlie the formation of TADs and their boundaries.

1.9.4 Systematic biases in HiC data

While HiC based studies were beginning to make an impact on the holistic understanding of the structure-function relationship of 3D genome, Yaffe and Tanay observed that the contact matrix derived from HiC data was loaded with different types of systematic biases (Yaffe and Tanay 2011). They described three distinct sources of biases that resulted from the inherent nature of the steps involved in the generation of HiC libraries (figure 1.22).

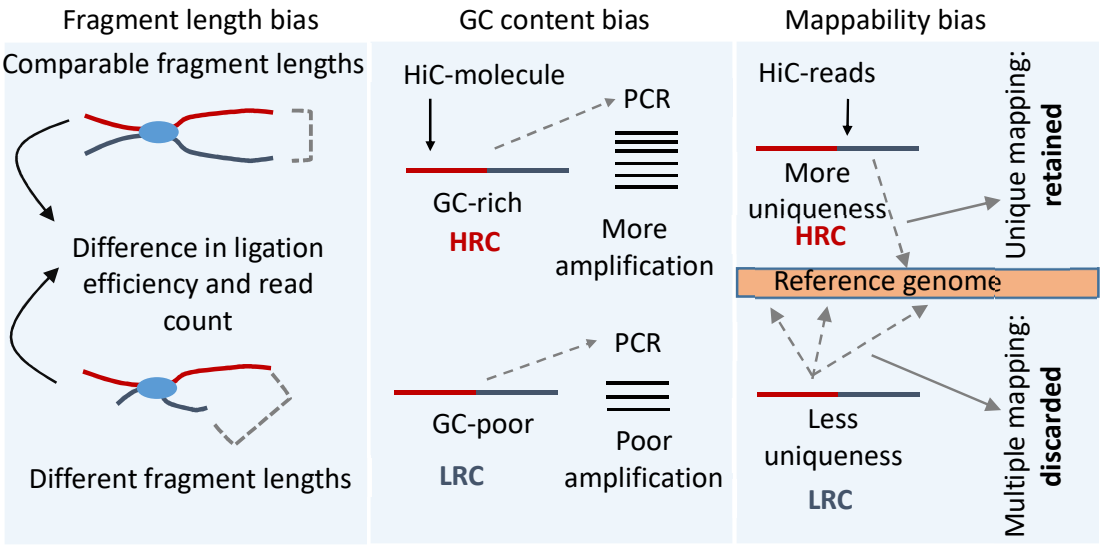

Figure 1.22 Representation of systematic biases in HiC data.

HRC and LRC indicate high read count and low read count, respectively. The restriction fragment pairs of different lengths are ligated with different efficiency (left). GC poor HiC molecules are poorly amplified in PCR as well as poorly sequenced which gives LRC as compared to GC-rich HiC molecules (middle). The reads which do not map uniquely are usually discarded, therefore the genomic regions from where they were originated are poorly represented in the HiC data (right).

The authors observed that differences in the lengths of restriction fragments undergoing ligation during the HiC protocol might impact the efficiency of ligation, thereby biasing the contact frequencies. Secondly, it is well-known that the GC-rich regions of DNA are efficiently amplified in the PCR reactions. Consequently, GC content of the restriction fragment ends can impact the relative interaction frequencies in the HiC data. Thirdly, the genomic regions with different genome mappability scores can also be a potential source of bias in the HiC interaction frequencies. Several computational methods were developed to normalize the HiC data (Cournac et al. 2012; M. Hu et al. 2012; Imakaev et al. 2012). One such method is HiCNorm, which was used to normalize systematic bias for the present study and is detailed in Materials and methods (Chapter II).

1.10 Comparison between 3C-methods and microscopy

In 3C-based studies, the spatial proximity of two distant genomic regions was assessed based on

their ability to ligate with each other (proximity ligation). However, some studies debated that whether the contact frequencies derived from proximity ligation-based assays could be reliably interpreted as spatial contact between different genomic elements (Andrew S. Belmont 2014; Dostie and Bickmore 2012; Fudenberg and Imakaev 2017). A brief summary of discrepancy and reconciliation between microscopy and 3C-based methods is presented here.

1.10.1 **Discrepancy.**

To compare the interaction frequency of 3C-based techniques with their spatial distance measured using microscopy, Wendy Bickmore's group investigated the conformation of *HoxD* locus using microscopy and 5C (Williamson et al. 2014). In their investigations where the mESCs were differentiated by 3 days long treatment of retinoic acid, they found that *HoxD* cluster underwent decompaction which was captured in both 5C and FISH experiments, suggesting a general agreement between these two approaches. However, while the FISH experiments suggested the *HoxD* to be the most decompact part of the 700kb region investigated in this study, the 5C data suggested that *HoxD* was more compact as compared to the flanking regions. These observations pointed towards the fact that high interaction frequency might not always mean spatial proximity. Further, they found that for the PRC1-null mESCs, where *HoxD* genes are derepressed, FISH experiments suggested open conformation of *HoxD* locus. But 5C experiments gave exactly opposite views and suggested the condensed conformation of *HoxD* locus even after its derepression in PRC1-null cells (Williamson et al. 2014). Besides these disagreements, the results of this study also found an agreement between 4C and FISH experiments for the organization of *Hoxd13* gene. In many of the 3C-based studies which validated their claims of spatial proximity between two genomic regions using DNA FISH, the spatial distance between FISH probes were often >500nm (reviewed in Belmont 2014; mentioned in Williamson et al. 2014). Therefore, the 3C-based inferences might not always agree with the microscopy and the spatial contacts between specific genomic regions should be carefully and jointly inferred from microscopy as well as 3C-based techniques (Dostie and Bickmore 2012; Fudenberg and Imakaev 2017).

1.10.2 **Reconciliation**

The discrepancy between 3C and microscopy might have more severe implications if one is interested in studying the 3D interactions between the specific regions of interest. For high throughput studies, which are aimed at studying the large scale behavior of chromatin organization, such differences might not impact the overall conclusion to the significant degree. In the past couple of years, several studies have employed the super-resolution microscopy to investigate the prominent features of large-scale 3D genome organization. For example, Wang et al. in 2016, used the multiplexed DNA-FISH to label all the TADs in many different chromosomes of the IMR90 cell line and investigated their 3D positions using super-resolution

microscopy. They found that TADs were broadly segregated into A/B compartments and relative distances among TADs had a striking correlation (0.91) with the inter-TAD HiC interaction frequency. Since TADs were identified from the HiC data, some studies speculated them to be mere statistical averages of HiC contacts across the population of cells rather than the true chromatin subunits (mentioned in Szabo et al. 2018). But investigations by Szabo et al. in 2018, using direct stochastic optical reconstruction microscopy (STORM), established the existence of TADs inside the single cells.

A few super-resolution microscopy studies using dSTORM and photo-activated localization microscopy (PALM) have also proposed the existence of compact chromatin domains which were formed by the interactions between nucleosomes (Nozaki et al. 2017) and the existence of higher order chromatin structures which had distinct epigenomic states (J. Xu et al. 2018).

1.11 Drifting away from proximity ligation.

As noted in the previous section, several studies have expressed concern regarding the interpretation of spatial proximities based on proximity ligation-based techniques such as HiC (Andrew S. Belmont 2014). To circumvent the need for proximity ligation, some alternative techniques were recently proposed which can be used to probe the large-scale 3D genome organization (table 4). A brief summary of two such techniques is given below.

1.11.1 Genome Architecture Mapping (GAM).

This technique was proposed by Beagrie et al. in 2017 and used the cryo-sectioning to generate the ultra-thin sections of the nucleus by using laser dissection at several different randomly chosen orientations in a population of nuclei. These sections were called ‘nuclear profiles’ and DNA contained in each section was identified by sequencing. The co-segregation of different genomic regions in the same nuclear profile provided evidence for their spatial proximity. The collection of a sufficiently large number of nuclear profiles enabled the authors of this work to construct a genome-wide matrix of proximities between different loci. Using this technique, the Beagrie et al. rediscovered the existence of TADs and wide-spread occurrence of spatial contacts which were dependent on gene expression. Further, this study also showed the existence of three-way contacts, which the authors claimed, remained undetected in previous 3C based methods (Beagrie et al. 2017).

1.11.2 Split-Pool Recognition of Interactions by TAG Extension (SPRITE)

This technique was invented by Quinodoz et al. in 2018 and is based on uniquely barcoding the spatially proximal chromatin fragments. Briefly, the chromatin was cross-linked and fragmented using shearing which generated the cross-linked chromatin complexes. These complexes were split across a 96-well plate and unique tags were ligated to all the DNA molecules within each chromatin complex. The procedure was repeated multiple times after which the tagged complexes

were pooled together, and contained a unique combination of tags called the barcodes. The sequences having the same barcodes provided evidence for their spatial proximities. Using this technique, Quinodoz et al. detected the extensive inter-chromosomal interactions which were anchored at subnuclear structures such as speckles and nucleolus (Quinodoz et al. 2018).

Besides GAM and SPRITE, another recent technique to probe the 3D genome organization without the use of formaldehyde crosslinking and proximity ligation is called DamC. This technique used the bacterial DNA adenine methyltransferase (Dam) enzyme to specifically methylate the adenine nucleotide in GATC sequences which were in spatial proximity to the genomic location of interest (called the viewpoint). In DamC, Dam enzyme is fused to reverse tetracycline receptor and is recruited to the viewpoints by the insertion of Tet operator sequences at targeted genomic locations. The specific methylation of GATC sites which are proximal to the viewpoint is detected by the combination of methylation-sensitive restriction enzyme DpnI and next-generation sequencing.

Table 4 Proximity ligation independent techniques to study 3D genome.

Name of technique	Reference	Principle	Scope	Quantification
Split pool recognition of interactions by TAG extension (SPRITE)	(Quinodoz et al. 2018)	Unique barcoding of spatially proximal genomic regions	All vs. all	NGS
Genome architecture mapping (GAM)	(Beagrie et al. 2017)	Cryo-sectioning of the nucleus to capture spatially proximal chromatin fragments	All vs. all	NGS
Dam-C	(Redolfi et al. 2019)	Specific methylation of adenine in GATC sequences contacting the viewpoint	One vs. all	NGS

Main highlights of 3D genome organization from 3C-based studies.

(i) Eukaryotic genomes are non-randomly organized in 3D space of the nucleus; while the large-scale organization appears to be conserved between species and cell lines, there are significant amount of species-specific differences.

(ii) Chromatin looping, involving distal regulatory elements and their target genes, is a highly dynamic and cell-type specific phenomenon which is actively remodeled following the gene expression alterations.

(iii) Metazoan genomes are compartmentalized into active and inactive compartments (A/B compartments).

(iv) TADs are structural and functional units of eukaryotic genomes which are largely conserved between different cell types but can change their epigenetic states in a cell-type specific manner.

(v) Wide-spread prevalence of long-range intra-chromosomal and inter-chromosomal gene clustering.

Some of the key features of large scale 3D genome organization learned from the 3C-based studies are summarized in figure 1.23.

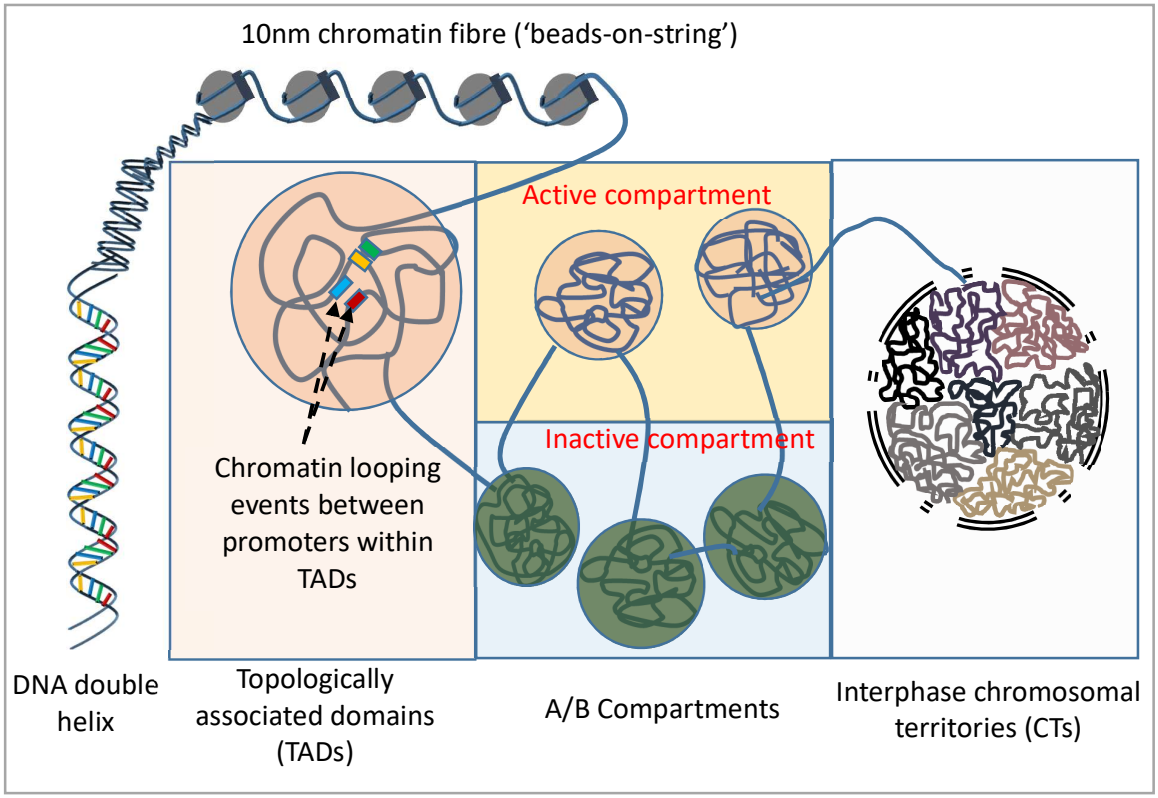

Figure 1.23 Summary of chromatin and genome organization in 3D.

The double helical DNA forms 10nm chromatin fibre by its association with histones. The chromatin is further organized into a series of self-interacting domains known as TADs. The TADs are distributed over active (or A) and inactive (or B) A/B compartments. The majority of functional interactions among gene promoters, between enhancers and their target genes occur within TADs. Each chromosome is organized into distinct chromosome territory as shown on the right most cartoon, where the lines of differed colors show different chromosomes.

This figure is modified from Ea et al. (2015) with permission.

1.12 Introduction to the project I

1.12.1 Spatial co-localization of genes

The main idea that emerged from the 3C-based studies was that different genes were often co-localized in the 3D space of the nucleus. These co-localization events were highly non-random and got remodeled during embryonic development (Simonis et al. 2006; Zhao et al. 2006). In particular, active genes were co-localized with active genes and inactive genes were co-localized with other inactive genes (figure 1.18). Series of genome-wide studies confirmed a widespread prevalence of constitutive and cell-type specific promoter-promoter interactions in 3D space (Y. Chen et al. 2014; Dostie et al. 2006; Z. Duan et al. 2010; Guoliang Li et al. 2012; Martin et al. 2015; Sanyal et al. 2012; Tanizawa et al. 2010). What are the underlying functional constraints that shape the spatial clustering of genes? Several possible mechanisms of gene clustering have been proposed which are discussed in the following paragraphs.

1.12.1.1 Transcription

The term transcription factory was originally coined by Jackson et al. in 1993 where they observed the presence of 300-400 RNA synthesis sites under the fluorescence microscopy in the HeLa cells (D. A. Jackson et al. 1993). The initial observations pertaining to the discrete sites of transcription were reconciled by several studies (reviewed in Rieder, Trajanoski, and McNally 2012). The detailed structure, function and the organization of transcription factories have been extensively reviewed elsewhere (Eskiw and Mitchell 2016; Papantonis and Cook 2013). It is worth noting that the transcription factory model was in a sharp contrast with the popular textbook representation of transcription, where RNA-Polz II was often depicted as molecular motor (Herbert, Greenleaf, and Block 2008) which can traverse the template DNA strand by sliding over it based on the in vitro evidence (Harada et al. 1999; Kabata et al. 1993). The transcription factory model proposed RNA-polII to be a stationary macromolecular complex located at definitive foci and genes much converge to the transcription factories in order to get transcribed. In the post-genomic era, the ‘transcription-factory’ dependent spatial co-localization of genes was supported by a number of studies. In 2004, the formation of the active chromatin hub at β-globulin locus was found to be dependent on transcription factor EKLF (Drissen et al. 2004). Although this report did not directly mention the word ‘transcription factory’ in their work, the involvement of transcription factors in mediating the chromatin looping observed in the formation of ACH indicated the transcription-dependent organization of chromatin loops. Peter Fraser’s group tested the transcription factory model and provided convincing evidence for spatial co-localization of actively transcribed genes at shared transcription foci using FISH as well as 3C based experiments (Osborne et al. 2004). A few years later, the results published from the same group showed that spatial co-localization of distal genes and their association with transcription factory were dependent on transcription initiation and persisted even in the absence of transcriptional

elongation (Mitchell and Fraser 2008). Later in 2010, Peter Fraser's group used a slightly modified form of 4C (called e4C) to investigate the organization of active mouse globin genes in erythroid lineage. This study concluded that co-regulated genes were clustered at a limited number of discrete foci that contained RNA-PolII which resembled the transcription factories (Schoenfelder et al. 2010). Further, a ChIA-PET study involving RNA-PolII uncovered the widespread prevalence multi-gene complexes which strongly overlapped with the RNA-PolII foci in DNA FISH experiments (Li et al. 2012). Particularly, the genes belonging to the same gene families, such as *HIST1H* of histones, formed 6 discrete clusters and involved 58 genes of chromosome 6. This study also observed that the genes which were not expressed along with *HIST1H* were not anchored in the multi-gene complex (representative figure 1.24). These studies collectively suggest the long-range spatial convergence of genes is mediated through transcription.

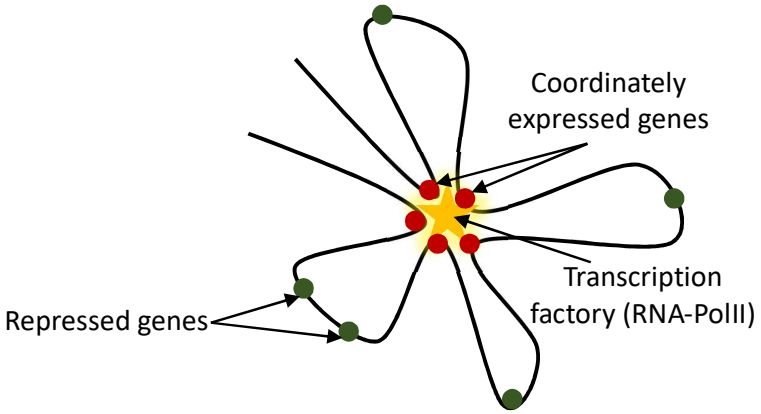

Figure 1.24 Interactions among genes anchored at transcription factory.
Cartoon representation of multi-gene complexes and 'transcription factory' model detected using ChIA-PET study, where coordinately expressed genes (red dots) are anchored at RNA-PolII foci while intervening genes, which are not expressed in the same tissue, are not anchored to the same factory.

1.12.1.2 Replication

Several studies from the late 80s to early 90s started to accumulate the evidence for the 'replication factories', which were proposed to be concentrated foci of DNA Pol-α (Nakamura, Morita, and Sato 1986). Using fluorescence microscopy, it was revealed that the process of replication occurs at specific 'granules' which were found to be discretely distributed throughout the nucleus (Nakayasu 2004). In electron microscopic studies, replication factories were found attached to the nuclear skeleton (Hozák et al. 1993) and in the year 2000, the existence of replication factories was also confirmed by the live cell imaging study of mouse myoblasts using 4D imaging (Leonhardt et al. 2000). Microscopy and molecular biology-based evidence

suggested that subunits of cohesin were involved in the formation and organization of chromatin loops around the DNA replication factories (Guillou et al. 2010). Additionally, in 2012, another study employed the 4C technique in human lymphoblastoid and PMBC cell lines and performed an integrative analysis by combining 4C with the DNA replication data (Moindrot et al. 2012). This study identified the megabase-sized self-interacting domains of chromatin and showed the organization of these domains was strongly related to the replication timing. Since TADs were also shown to mark the boundaries of transition from early to late replication timing (Pope et al. 2014), the replication domains identified by Moindrot et al. (2012) could be deemed similar to TADs. Moreover, the budding yeast TADs are considered similar to the replication factories (Eser et al. 2017). Therefore, it is reasonable to speculate that large scale spatial clustering of genes in *cis* and *trans* might be coordinated by their association with replication factories (Figure 1.25).

1.12.1.3 Other possible modes of gene clustering

Yet another model of gene clustering in the 3D nucleus vouched for the role of nuclear speckles in mediating the gene clustering. In a study undertaken by Brown et al. in 2008 using human erythropoiesis as a model system, it was found that 3D association of genes located on different chromosomes did not exhibit co-regulation, but were rather associated with nuclear speckles enriched in SC-35 (J. M. Brown et al. 2008). Further, the splicing-related 3D clustering of genes is also supported by the non-random spatial co-localization of genes which undergo *trans*-splicing, i.e. the splicing event where the exons from two different genes are joined in a single transcript (Hubbard et al. 2012; reviewed in Lei et al. 2016).

Some studies in yeast have shown that DNA repair is organized into discrete foci (Lisby et al. 2003). Following the DNA damage by irradiation, multiple damaged sites can be localized to the same repair focus containing Rad52 (Lisby, Mortensen, and Rothstein 2003). Though less common in mammals, such effects are widespread in yeast and can be influenced by the large scale nuclear organization of chromatin (reviewed in Misteli and Soutoglou 2009). Some studies have also proposed that spatial proximities between different genes might influence the frequency of genome-wide translocations (Engreitz, Agarwala, and Mirny 2012; Kenter et al. 2013; Meaburn, Misteli, and Soutoglou 2007; Yu Zhang et al. 2012). Such proposals are also supported by the observed non-random spatial proximity of mouse-human evolutionary breakpoints in the human nucleus (Bagadia, Singh, and Sandhu 2016; Véron et al. 2011).

1.12.2 Principles of large-scale organization of 3D genome are not fully understood

Considering the work described in the previous sections, one can argue that the large scale 3D organization of genomes is governed by the coordinated regulation of multiple genomic processes and each process might influence the genome organization to the certain degree (figure 1.25).

The most popular model of gene clustering is the transcription factory model. However, this

model has been put into question by several studies. For example, in their 4C work, Hakim et al. observed a strong co-localization even for the genes which had an opposite transcriptional response in glucocorticoid stimulation studies of mouse epithelial breast carcinoma cell lines (Hakim et al. 2011). Further, it has been proposed that the association of genes with transcription factories is not constrained and appears to be an outcome of limited numbers of the same (J. B. Lawrence and Clemson 2008). It is also possible that a large proportion of the gene-gene interactions observed in the case of 3C/4C based studies is the by-product of close packing inside the nucleus without any functional relevance beyond the segregation of euchromatin and heterochromatin (Eskiw and Mitchell 2016; J. B. Lawrence and Clemson 2008). Alternatively, the gene-gene contacts might represent the population average of association of different genes with a constrained number of splicing speckles or transcription factories inside the nucleus (see Lawrence and Clemson 2008). Therefore, a comprehensive understanding or large scale organization of 3D genome is missing and requires the thorough investigation of HiC datasets in relation to multiple functional genomic/epigenomic attributes.

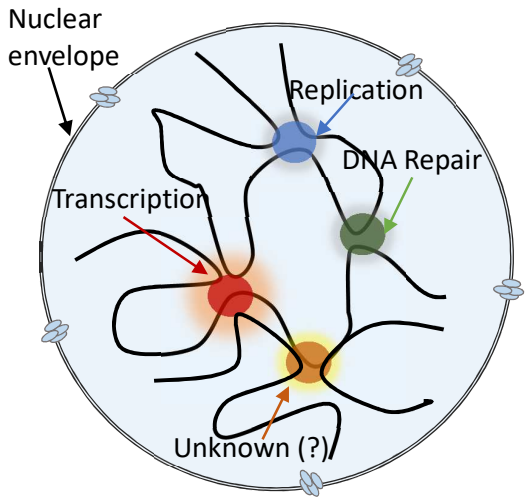

Figure 1.25 Possible modes of gene clustering proposed by different studies.

Different lines indicate the spatial convergence between different chromosomes in *cis* and *trans*. The colored dots drawn show the hypothetical representation of the various different macromolecular complexes related to the transcription (red), replication (blue), DNA repair (green), or those which remain yet to be deciphered (orange).

1.12.3 **The integrative analysis in 3D genomics**

The exploitation of HiC based methods to study 3D genome organization provided a unique opportunity to investigate various genomic and epigenomic attributes in the context of spatial interactions between genes. Several studies attempted to combine functional and epigenomic datasets in order to provide meaningful explanations to their observations in 3C-based

experiments. In the 4C work described by Zhao et al. in 2006, the authors combined the knowledge of differentially methylated regions (DMRs) and showed that direct interactions with DMRs were involved in epigenetic regulation in *trans* (Zhao et al. 2006). A couple of years later to this work, Stephan Baylin's group performed a 3C analysis of chromatin conformation around *GATA*-4 gene in the embryonic carcinoma (EC) cell lines and discovered the formation of the multi-loop complex at this locus (Tiwari et al. 2008). Further, the authors integrated the information of polycomb proteins (PcG) binding along with DNA methylation and deciphered the dynamic interplay between chromatin loop formation, DNA methylation, and gene activation. Similarly, Rödelsperger et al. integrated the genomic, functional and protein interaction data to understand and predict the spatial interactions between enhancers and target promoters. (Rödelsperger et al. 2011). Similar research, which involved the integrative analysis of genomic and epigenomic datasets, was undertaken by many other groups which employed HiC or ChIA-PET based techniques (Dixon et al. 2015; Fullwood, Liu, Pan, Liu, Xu, et al. 2009; G. Li et al. 2012; Lieberman-Aiden et al. 2009; Rao et al. 2014). For example, in Rao et al. in 2014, detected the five different sub-compartments in their HiC data of GM12878 cell line and established their unique epigenomic nature by incorporating the data of chromatin modifications. Likewise, the study by Fullwood et al. used the ChIP-seq data for H3K4me3 and RNA-PolII and revealed that ERα bound chromatin interactions were involved in the regulation of specific genes (Fullwood et al. 2009). The publically available chromosomal contact data generated from HiC make it possible for a broader scientific community to undertake the investigation of the 3D organization of genomes by combining it with various functional genomics studies.

1.12.4 **The 3D organization of yeast genome**

Yeast is one of the simplest eukaryotic model systems for genomics research. Ever since the availability of its genomic sequence in 1996, it was subjected to a number of genomics/transcriptomics studies using array-based (microarray, ChIP-chip) platforms which were later replaced by sequencing-based methods (such as RNA-Seq, ChIP-Seq). The 3D model of the yeast genome, published by Duan et al. captured many of the canonically known features of large scale organization of the yeast genome (Z. Duan et al. 2010). The tRNA genes of the yeast genome are known to reside inside a sub-nuclear compartment called nucleolus. This was also observed in Duan et al.'s HiC map where two discrete clusters of tRNA genes were found; one was associated with ribosomal DNA as per the nucleolar organization of tRNA genes, while the other was associated with centromeres. Furthermore, it was revealed that the centromeres formed a distinct cluster near the spindle pole body and telomeres localized near the nuclear envelope at the opposite end, consistent with the Rabl configuration. It was also observed that smaller chromosomes of yeast (eg. I, III, VI) were more likely to engage in interchromosomal

interactions as opposed to the self or intrachromosomal interactions. Although the authors found the evidence for the territorial organization of some chromosomal arm regions, the territorial behavior of whole chromosomes was not very strong and different chromosomes showed a significant amount of intermingling (Z. Duan et al. 2010).

1.12.5 Constraints underlying 3D genome organization of yeast genome

A few studies have attempted to systematically analyze the underlying constraints of large scale 3D organization in yeast. Using the HiC data generated from Duan et al., a study published in 2012 found the significant co-localization of genes regulated by similar transcription and chromatin factors and proposed the co-regulation to be an important aspect of the gene-gene clustering (Dai and Dai 2012). Another study, published in 2013, found somewhat similar results and proposed that spatially proximal genes had a greater correlation between their expression profiles as compared to the genomic-wide average correlation coefficient and tend to be functionally related (Homouz and Kudlicki 2013). Yet another study showed the strong correlation between the 3D distances among genes and similarity in their codon usage frequency for the case of budding yeast, fission yeast, Arabidopsis, mouse and human genomes (Diament, Pinter, and Tuller 2014). These observations conformed to the transcription factory model of gene clustering and suggested the co-regulation mediated distribution of genes inside the 3D space of the nucleus. On the contrary, an investigation by Witten and Noble has ruled out the spatial co-localization among co-expressed genes and found the evidence for co-localization of centromeres, DNA breakpoints, tRNA genes, and for early origins of replication in the yeast genome (Witten and Noble 2012). Importantly, all these studies focused on assessing the co-regulation dependent gene clustering (i.e. the transcription factory model) and did not focus much on assessing the co-localization mediated through replication factories. In addition to the conflicting reports, all the work in this direction suffered from the following two limitations:

(i) These studies were restricted by the number of functional variables analyzed and appeared biased towards analysing the co-expression of genes only.

(ii) Some of these studies lacked stringent and unbiased null models to test the significance of the observed association in their model.

1.13 Objectives of the project I

In the first part of this thesis, we have investigated the 3D organization of yeast genome by combining the chromosome contact data with different functional genomics datasets with an objective to decipher the underlying functional and evolutionary constraints which might have shaped the genome-wide landscape of chromosomal contacts among genes.

1.14 Introduction to project II

1.14.1 Investigation of histone code hypothesis

Ever since the proposal of the histone code hypothesis, extensive efforts were undertaken by the scientific community to obtain a holistic understanding of histone code. The development and maturation of microarray-based platforms such as tiling array (Dennis 2001; Huebert et al. 2006; Liu 2007; Shivaswamy and Iyer 2007) along with the development of increasingly more specific antibodies against various histone modifications and their effector molecules (Milne, Zhao, and Hess 2009; Venters et al. 2011) paved the way to investigate genome-wide distribution and effect of histone modifications. It was believed that these genome-wide maps will help in deciphering the logic of chromatin remodeling and histone code hypothesis. The strategies to map and understand the histone code must include the following experiments (Robyr et al. 2002):

(i) Genome-wide location analysis of chromatin modifications.

(ii) Genome-wide location analysis of chromatin modifiers.

(iii) Genome-wide transcriptional consequences of deletion of chromatin regulators.

The pioneering work in this field was performed on budding yeast as it was one of the simplest eukaryotic model organism. Secondly, its genome sequence became accessible to scientists in 1996, which made it possible to design various genome-wide studies. Thirdly, the genetic manipulations were easily done in yeast and a library of individual gene deletions of all the non-essential genes was available.

1.14.1.1 Genome-wide location of chromatin modifications

During the initial days of post-genomics era, several studies employed the genome-wide location analysis to understand the distributions of various chromatin remodelers and chromatin modifications in an attempt to understand their relationship with gene activity (B. E. Bernstein et al. 2002; Kurdistani et al. 2002; Lieb et al. 2001; Ng et al. 2002; Robyr et al. 2002; Roh et al. 2004). In 2004, Grunstein's group published a study, in which they had mapped the global histone acetylation patterns to understand the epigenetic regulation of gene expression (Kurdistani, Tavazoie, and Grunstein 2004). Interestingly, the authors of this work found that both hyper and hypoacetylation of lysines in histones were associated with transcriptional activation. Further, they found that yeast genes were partitioned into many different clusters. Each cluster had a different pattern of histone modifications, contained co-expressed genes, was involved in specific biological processes, and was regulated by specific transcription factors (Kurdistani, Tavazoie, and Grunstein 2004). Although the unique patterns of histone acetylation in different clusters of the genes were observed, the authors argued that a given set of modifications might not contain more information than an individual modification and, therefore, might not constitute a non-redundant histone code (Kurdistani, Tavazoie, and Grunstein 2004). The hallmark of histone code

is combinatorial complexity and authors claimed that the observed patterns of histone modifications were more similar to the signaling pathways and chromatin modifications served to signal the recruitment of various components of transcriptional machinery rather than constituting a combinatorial code (Kurdistani, Tavazoie, and Grunstein 2004; Schreiber and Bernstein 2002). Soon after this work, Dr. Richard Young's group used ChIP-on-chip technology on a tiling array platform to perform a genome-wide location analysis of different acetylation and methylation modifications of the histones by taking into account the relative nucleosome occupancy along the genome in order to normalize the observed levels of chromatin modifications (Pokholok et al. 2005). This study found that both acetylation and methylation of histones were correlated with transcriptional activation (Pokholok et al. 2005). In the year 2007, the ENCODE project (Encyclopaedia of DNA elements) finished its pilot phase and published an integrative analysis of transcription, replication, chromatin modifications, DHSs in 1% of the human genome (Birney et al. 2007). The findings presented in this work showed that a distinct pattern of chromatin modifications was strongly associated with the presence and activity of TSSs. Additionally, it was also elucidated that distal DHS sites had unique histone modification patterns, which were distinct from the promoter associated patterns. TSSs were strongly marked by the status of H3K4me2/3 and H3ac, the relative level of which differed significantly between active and inactive genes (Birney et al. 2007; Koch et al. 2007). The regions away from TSSs were enriched for H3K4me1 as opposed to H3K4me3 and H3Ac, which was consistent with the findings in budding yeast (Pokholok et al. 2005).

The investigations described in the preceding text did not find sufficient evidence for the combinatorial complexity of histone modifications despite their strong overall associations with the gene activation and repression. Therefore, the observed patterns of histone modifications were proposed to constitute a signaling pathway for recruitment of downstream effector molecules rather than a 'code' of histone modifications (Berger 2007; J. S. Lee, Smith, and Shilatifard 2010; Suganuma and Workman 2008). However, as described on page 14-15 of this dissertation, the implementation of computational methods on ChIP-seq datasets for different chromatin modifications in *Drosophila,* mouse, and humans supported the existence of combinatorial complexity in the genome-wide patterns of chromatin modifications.

1.14.1.2 Genome-wide location of chromatin factors

The genome-wide location analysis of histone modifications yielded direct insights into the functioning of the chromatin in the context of transcription. Additionally, mapping the genome-wide locations of the enzymes responsible for depositing these chromatin modifications constitute another pillar for gaining the systems level understanding of the language of chromatin modification. The pioneering work in this direction was done in 2002 with the generation of a

genome-wide binding map of a histone deacetylases enzyme named Rpd3 (Kurdistani et al. 2002). Further, the investigators from Richard Young's lab attempted to understand the transcriptional regulatory networks by performing the genome-wide location analysis of a large number of yeast transcription regulatory proteins (Harbison et al. 2004; T. I. Lee et al. 2002). However, the regulatory proteins analysed in these investigations did not contain the significant number of the chromatin modifying enzymes. To investigate the chromatin modifiers, Dr. Richard Young's lab performed the genome-wide location analysis of 4 different HATs and HDACs in the budding yeast genome (Robert et al. 2004). This study showed that while HATs (Gcn5, Esa1) were generally recruited to the active genes across the yeast genome, the HDACs (Hst1, Rpd3) were recruited to the subset of genes enriched in specific biological functions. In 2011, Venters et al. performed the genome-wide location analysis by including 202 different yeast chromatin and gene regulatory proteins (Venters et al. 2011). With this map covering 202 chromatin regulators, the Venters et al. recapitulated the regulatory distinction between SAGA dominated and TFIID dominated transcription activation pathways and showed the regulation associated with SAGA-dominated genes exhibited a greater specificity and diversity as compared to the TFII dominated genes. Further, the SAGA-dominated genes showed preferential binding at distal regions with respect to TSSs as compared to the TFIID dominated genes which exhibited the preferential binding proximal to the TSSs (Venters et al. 2011), thus deciphering the difference in the organization of chromatin regulators between these two categories of genes. Additionally, the authors found the extensive remodeling of chromatin factor occupancy in response to the heat shock at the 5' as well as 3' end of the genes (Venters et al. 2011).

1.14.1.3 Genome-wide transcriptional perturbation upon deletion of chromatin modifiers

The genome-wide studies described so far mainly relied upon the genome-wide location analysis of chromatin remodelers and various chromatin modifications. The gene expression measurement was often not performed in most of these studies and therefore, a direct causal link between the chromatin modifications and their downstream effects on transcription was missing (mentioned in Lenstra et al. 2011; Rando and Chang 2009). To directly measure the transcriptional consequences of chromatin modifiers and understand the cross-talks between different chromatin regulatory pathways, Frank Holstege's group, in the year 2011, published the genome-wide analysis of transcriptional consequences of 165 chromatin factors (Lenstra et al. 2011). The authors took the individual deletion strains of 165 chromatin regulators and performed the microarray experiments to measure the genome-wide mRNA levels in each of the deletion backgrounds. Importantly, this study recapitulated many of the interactions between chromatin regulators which were known to exist based on the functional or location analysis of individual genes. For example, the authors observed that SAGA and SWI/SNF were strongly connected in

the correlation network of chromatin factors generated in this study. This meant that many of the targets were shared between the SAGA and SWI/SNF as SWI/SNF complex is known to preferentially remodel the acetylated nucleosomes (Chandy et al. 2006; Lenstra et al. 2011). Further, the authors also found that the deletion of chromatin factors belonging to the same complex often did not result in the transcriptional perturbation of a similar set of genes (Lenstra et al. 2011). Nonetheless, the deletion of each chromatin factors exhibited the very specific effects even for the chromatin factors associated with more global chromatin modifications such as methylation of H3K4/36/79 (Lenstra et al. 2011).

1.14.2 **The disconnect between binding and effect of chromatin factors**

The genome-wide location analysis and genome-wide transcriptional effects of chromatin factors constituted two important pillars of the studies aimed towards understanding the logic of chromatin remodeling. However, as noted by Lenstra et al., the deletion of most chromatin factors resulted in the transcriptional perturbation of a very specific set of genes (Lenstra et al. 2011). Even for the factors which were known to establish the genome-wide chromatin marks (for eg. Set1, Set2, and Dot1 implicated in the methylation of H3K4/36/79, respectively.), Lenstra et al. observed that a very specific set of genes constituting 55, 72, and 2 genes were transcriptionally affected upon the deletion of Set1, Set2, and Dot1 respectively. But the genome-wide location maps of these factors (Venters et al. 2011) paralogues exhibit this specificity for their binding targets (mentioned in Lenstra et al. 2011). These observations indicated that the presence of the chromatin modifier at certain genomic loci does not always indicate the downstream regulatory activity and this phenomenon was termed as ‘disconnect’ or discrepancy between location and effect (Lenstra and Holstege 2012). Similar observations were made in several other studies which performed the genome-wide location and transcriptional perturbation analysis of individual chromatin modification and chromatin modifiers (reviewed in Lenstra and Holstege 2012; Rando and Chang 2009). For example, in the genome-wide location map of Rpd3 generated by Kurdistani et al. in 2002, it was observed that despite the significant binding of Rpd3 in the upstream regions of ribosomal genes, their expression was not upregulated upon the deletion of Rpd3 (Kurdistani et al. 2002; Robyr et al. 2002). Another example is Set1, which is known to deposit the H3K4me3 mark. This universal mark is positively correlated with the transcription. The deletion of Set1 is not lethal in yeast (mentioned in Rando and Chang 2009). Similarly, the depleting the levels of H1 histone cause the genome-wide changes in chromatin structure but only a smaller subset of genes was transcriptionally affected in mESCs (Fan et al. 2005), providing another example for the lack of consistency between location and effect analysis of chromatin modifications.

Exploiting the availability of genome-wide datasets, Lenstra et al. investigated the correlations between genome-wide binding and effect of 70 different chromatin regulators and found that most of the chromatin factors had insignificant correlations between their binding and effect (Fan et al. 2005). The strongest positive and negative correlations were observed for Ssn6 (~0.24) and Htz1(-0.15) respectively. Therefore, it became evident that almost all of the chromatin factors exhibit a widespread disconnect between their location and effect (figure 1.26).

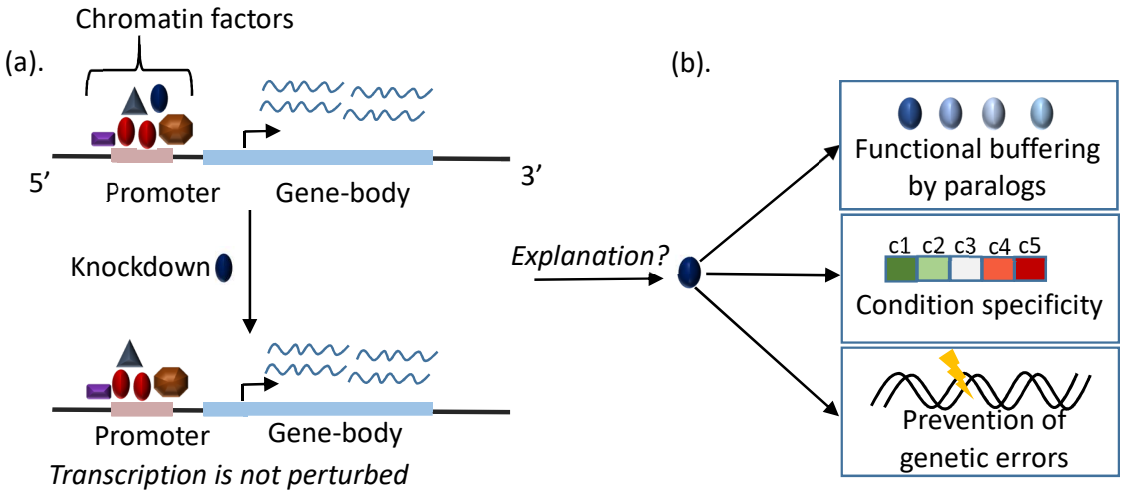

Figure 1.26 Disconnect between binding and effect of the chromatin modifiers.

a). Knockdown of certain chromatin factors often does not result in the perturbation of its binding targets. b). Possible explanations for the disconnect: Deletion of a factor might be buffered by the presence of paralogues genes (top). A factor might be bound at certain targets but might be only required in alternative environmental conditions (middle). The color bar indicate the hypothetical heatmap containing the gene expression values under 5 different conditions (c1-c5) for the binding targets of chromatin factors exhibiting disconnect. c) Factor binding might be involved in non-transcriptional activity such as prevention of genetic errors.

1.14.3 **Determinants of genome-wide disconnect are not fully understood**

What might explain the disconnect between location and effect of chromatin factors? This is an open question and no clear explanation is available to explain this discrepancy, which limits the understanding of the underlying logic of chromatin remodeling. Several speculations were raised by Lenstra et al. in 2012, when they quantified the disconnect between binding and effect of chromatin regulators in yeast. For example; H3K36me3 is positively associated with the transcription elongation and is recognised by Rpd3 enzyme, which deacetylases the histones bearing H3K36me3 mark in order to suppress the transcription initiation from cryptic promoters (Carrozza et al. 2005; also reviewed in Lenstra and Holstege 2012; Rando and Chang 2009). Therefore, it was proposed that merely correlations of histone modifications with transcription

should not be considered as the sign of causality and these modifications might have alternative roles which are coupled with the process of transcription (Rando and Chang 2009). In the case of loss of Rpd3, only the specific set of genes were transcriptionally perturbed, which were enriched for longer genes as longer genes might have more cryptic promoters (Lenstra and Holstege 2012; B. Li et al. 2007). These observations meant that the effect of removal of chromatin modifier was context-dependent (Lenstra and Holstege 2012). The case of H3K36me3 described above presents one such context. Another context, exemplified by Lenstra et al., was the preferential action of Htz1 near the telomeres, where H2A.Z is required to prevent the spreading of Sir complex (Lenstra and Holstege 2012). Therefore, in this case, only subtelomeric genes were preferentially affected upon the deletion of HTZ1. However, these proposals explained the 'connect', i.e. the genes which are likely to be affected upon deletion of a given chromatin modifier and not the disconnect *per se*. But the disconnect implies the lack of effect despite the binding of a given chromatin factor and proposals related to the susceptibility of some genes to the deletion of certain chromatin factor does not constitute a valid explanation for the disconnect.

The lack of effect can be ascribed to functional redundancy of chromatin factors. In fact, such an explanation have been employed to explain the disconnect between location and effect of transcription factors in budding yeast where the presence of paralogues genes was shown to buffer the impact of transcription factor deletions (Gitter et al. 2009; W.-S. Wu and Lai 2015). However, most chromatin modifying enzymes in yeast lacked the biochemically redundant partners and therefore the functional buffering might not have significant contribution in explaining the genome-wide disconnect between their location and effect (Lenstra and Holstege 2012). Another possibility is that the binding of a particular chromatin factor at certain genes is not involved in transcriptional regulation under the given set of conditions, but might become important for transcriptional regulation under stressful environmental conditions (figure 1.24). Such an explanation was proposed to explain the spurious transcription factor binding sites (Spivakov 2014), and even speculated to be the case for chromatin factors in yeast (Lenstra and Holstege 2012). But the evidence and comprehensive investigations of this hypothesis are lacking for the chromatin factors in budding yeast.

1.15 Objectives of project II

In the IInd part of the thesis, we have used the power of functional genomics datasets available in budding yeast to investigate and understand the determinants of genome-wide disconnect between location and effect of chromatin factors.

Chapter 2 Materials and methods

Most of the results presented in this thesis were produced by the statistical analysis of publically available datasets. A description of various statistical, computational, and experimental methods used in this work is presented below.

2.1 Datasets

The datasets used in this study were either downloaded from SGD, NCBI, or were obtained from supplementary files of the original publications. A list of dataset name and its source is given in table 5,6,7 and a brief overview of the datasets is given in Appendix A.

Table 5 Datasets for multivariate analysis

#	Name of dataset	Question to be addressed	Reference
1	Chromatin interaction data of yeast (240629 *trans* and 65683 *cis* interactions)	Which variable best explains the frequency of HiC data	(Z. Duan et al. 2010)
2	Functional similarity of yeast genes (fsm)	Do the spatially proximal genes exhibit functional similarity?	*gosimsem.*
3	Protein-protein interactions in yeast proteome (ppi, 50000 interactions)	Do the protein products of spatially proximal genes engage in protein-protein interactions?	(Kiemer et al. 2007)
4	Genome-wide map of histone acetylations (ace, 5993 genes)	Do the spatially proximal genes exhibit similarity in the level of histone acetylation?	(Kurdistani, Tavazoie, and Grunstein 2004)
5	Genome-wide map of histone methylations (met, 5383 genes)	Do the spatially proximal genes exhibit similarity in the level of histone methylation?	(Pokholok et al. 2005)
6	Replication time course (rep, 23964 genomic regions)	Do the spatially proximal genes have similar replication timing?	(Raghuraman et al. 2001)
7	Cell cycle gene expression time course (cct, 5754 genes)	Are the spatially proximal genes co-expressed during the cell cycle?	(Spellman et al. 1998)
8	Genome-wide co-fitness of yeast genes (cof, 4769 genes)	Do the spatially proximal genes have similar fitness defect?	(Hillenmeyer et al. 2008)
9	Genome-wide response to transcription factor perturbations (tfp, 4792 genes)	Do the spatially proximal genes exhibit a similar response to the deletion of transcription factors?	(Z. Hu, Killion, and Iyer 2007)
10	Genome-wide response to chromatin factor perturbations (cfp, 6123 genes)	Do the spatially proximal genes exhibit a similar response to the deletion of chromatin factors?	(Lenstra and Holstege 2012)
11	Evolutionary rate covariation of yeast genes (erc, 4459 genes)	Do the spatially proximal genes evolve at similar rates?	(Clark, Alani, and Aquadro 2012)

The primary dataset used for the project-I was the chromatin interaction data of budding yeast. As indicated in table 5, this dataset was taken from a previous study (Z. Duan et al. 2010). We downloaded the two available libraries of the data; the one generated using HindIII restriction enzyme, and the other using EcoR1 restriction enzyme. We have used the HindIII library for most of the analysis and the key observations were validated using EcoRI library. The datasets used in the multivariate analysis of chromatin interactions are given in table 5 and described in Appendix A. Other datasets used for the analysis of chromatin interactions are given in table 6 and a schematic of the workflow involved in this analysis is given in figure 2.1

Table 6 Other datasets for the analysis of chromatin interactions.

#	Name of dataset	Purpose	Reference
1	Origins of replication	To assess the spatial connectivity among origins of replication	(McCune et al. 2008)
2	Nucleosome occupancy	To check the preferential nucleosome occupancy at *cis* and *trans* interacting domains.	(Soriano et al. 2014)
3	Mutation rates of yeast genes	To assess the similarity in mutation rates of early and late replication domains	(Hirsh, Fraser, and Wall 2005)
4	Essential genes	To check the pattern of spatial interactions among essential genes.	(Winzeler et al. 1999) and PEC database
5	Chromatin mobility data.	To test the association between chromatin interaction frequency and chromatin mobility	(Albert et al. 2013; P. Heun et al. 2001)
6	Chromatin interactions network of the E.coli genome	To test the evolutionary conservation of noise minimization through *cis* interactions.	(Xie et al. 2015)
7	Chromatin interaction network of the mouse (mESCs) genome.	To test the evolutionary conservation of noise minimization through *cis* interactions.	(Tang et al. 2010)
8	RNA-PolII tethered chromatin interactions of mESCs	To test the mechanism of spatial tethering via *cis* interactions	(Schoenfelder et al. 2015)
9	Over-expression toxicity	To check if the *cis* interactions can minimize the expression noise of toxic genes.	(Yoshikawa et al. 2011)

For the analysis of disconnect between chromatin factor binding and effect, we have used two main datasets; namely genome-wide location analysis of chromatin factors (Venters et al. 2011), and genome-wide transcriptional perturbation dataset of chromatin factors (Lenstra et al. 2011). Source, purpose, and description of all the datasets used in this part of the work are given in table 7 and appendix A of this thesis.

Table 7 Datasets for the analysis of discrepancy between binding and effect of chromatin factors in yeast.

#	Name of the dataset	Purpose	Reference
1	Genome-wide location of chromatin factors	To obtain the significant binding targets of 67 chromatin factors.	(Venters et al. 2011)
2	Genome-wide chromatin factor perturbation dataset	To obtain the significantly affected targets upon the individual deletions of 67 chromatin factors.	(Lenstra et al. 2011)
3	Cell cycle time course of gene expression	To assess the cell cycle specificity of chromatin factors and their binding targets.	(Spellman et al. 1998)
4	Protein-protein interactions	To infer the function of chromatin factors	(Kiemer et al. 2007)
5	Single nucleotide mutations	To assess the genetic errors at binding targets	(Zhu et al. 2014)
6	ssDNA enrichment at yeast genes	To assess the genetic errors (double strand breaks) at binding targets	(Buhler, Borde, and Lichten 2007)
7	Genome-wide cryptic transcription	To assess the epigenetic errors at binding targets	(Neil et al. 2009)
8	*cis* variability of yeast genes.	To assess the genetic errors at binding targets	(Choi and Kim 2008)
9	Genetic interaction network of chromatin factors	To check the significance of chromatin factors with low EBRs.	(Costanzo et al. 2010)
10	Chromatin modification during the stress response	To test the condition-specific roles of chromatin factors.	(Weiner et al. 2015)
11	Dynamics of gene expression profile during stress response	To test the condition-specific roles of chromatin factors.	(Gasch et al. 2000)
12	Essential genes	To check the protected nature of unaffected binding targets.	(Winzeler et al. 1999)

2.2 Data processing and statistical methods

Quantile normalization of functional data sets

The multi-dimensional datasets used in this study were pre-processed with quantile normalization (Amaratunga and Cabrera 2001; Bolstad et al. 2003) in order to eliminate the global differences in the distributions of samples in each dataset. Such differences are often believed to be technical noise and might not represent the true biological variation across different samples (Y. H. Yang et al. 2002). Quantile normalization assumes and forces all the distributions to be the same by taking the average of each quantile across the samples. We have used the package **'DNAMR'** in **R** for the quantile normalization *(http://www.rci.rutgers.edu/~cabrera/DNAMR/).*

Elimination of systematic biases from chromatin interaction data

To correct for the systematic biases from the chromatin interaction dataset of yeast, we have used an **R**-based library called HiCNorm (M. Hu et al. 2012). HiCNorm used three predefined sources of bias (or genomic features) which included GC-content, effective length, and the mappability of the restriction fragments. We calculated length, GC-content and mappability of yeast restriction fragments by using the Perl and **R** scripts provided along with HiCNorm library. HiCNorm assumed that the read counts in the HiC data follows a Poisson distribution and employed the Poisson regression model to learn the expected distribution at the desired resolution of HiC data. The generalized linear model for estimating the rate of Poisson distribution in HiCNorm was defined according to the following equations:

$$\log\left(\theta_{jk}^{C}\right) = \beta_0^C + \beta_{len}^C \log\left(L_j^C L_k^C\right) + \beta_{gc}^C \log\left(G_j^C G_k^C\right) + \log\left(M_j^C M_k^C\right).$$

where, for a chromatin interaction pair *(j,k)* belonging to chromosome c, L_j^C, G_j^C, M_j^C represented the length feature, GC content, and mappability, respectively, of locus *j* and L_k^C, G_k^C, M_k^C represented the same for locus *k*. The terms β_0^C was the intercept and $\beta_{len}^C, \beta_{GC}^C$ represented the effective length, and GC content bias. The mappability was used as offset in the model.

Assuming that $\widehat{\beta}_0^c$, $\widehat{\beta}_{len}^c$ and $\widehat{\beta}_{GC}^c$ represented the estimated parameters of bias, the estimated rate of Poisson regression was defined as follows:

$$\widehat{\theta}_{jk}^{C} = exp\left\{\widehat{\beta}_0^c + \widehat{\beta}_{len}^c \log\left(L_j^c L_k^c\right) + \widehat{\beta}_{gc}^c \log\left(G_j^c C_k^c\right) + \log\left(M_j^c M_k^c\right)\right\}.$$

Where the $\widehat{\theta}_{jk}^{C}$ represent the expected rate of distribution of HiC data. The normalized read counts were obtained by the following relation:

$$e_{jk}^c = u_{jk}^c / \widehat{\theta}_{jk}^c,$$

Where $\boldsymbol{u_{jk}^c}$ is the observed read count for a given interaction between ***(j,k)*** of the HiC data. The aforementioned description of HiCNorm is adopted from M. Hu et al. (2012). A comparison of

raw and normalized contact frequencies showed the effectiveness of this approach (figure 2.1).

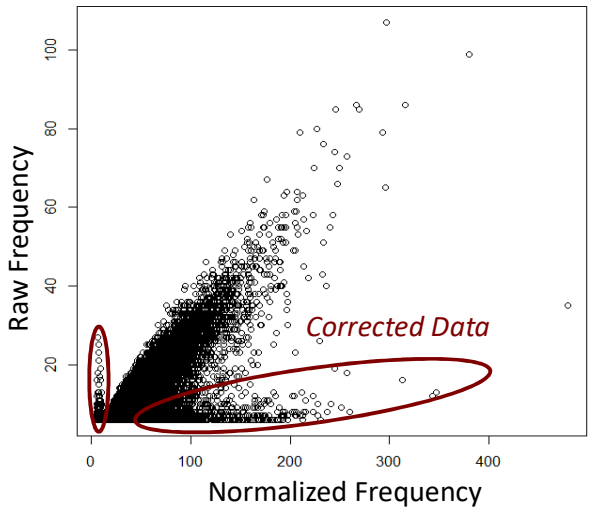

Figure 2.1 Comparison of raw and normalized HiC data.

Most data undergoes modest correction, while some data points undergo heavy correction (ellipse).

Calculation of similarities between genes

We downloaded the multidimensional functional genomics datasets, i.e., the datasets where the genome-wide measurements were performed at multiple time points or conditions. We performed quantile normalization of these datasets and computed the Pearson Correlation Coefficients (PCCs) to calculate the similarities between different genes for various genomic variables. The PCC was calculated using either **cor** or **cor.test** function in R. The functional similarity (fsm) between genes was measured by calculating the semantic similarity among the GO terms of the gene pairs using an **R**-package '*gosemsim*' (Yu et al. 2010). The genome-wide co-fitness values were taken from Hillenmeyer et al. (2008), which represented the PCCs between fitness values of genes across multiple drug conditions (see Appendix A). The ERC represented the similarity between evolutionary rates of different genes and were measured by calculating PCCs among the rates of sequence divergence of >4400 orthologous proteins in an 18 species phylogenetic tree (Clark, Alani, and Aquadro 2012). For protein interactions, the weights of PPI network downloaded from Kiemer et al. 2007 were taken as the strength of interactions. A description of these datasets can be found in Appendix A.

Partial least square regression (PLSR) analysis

To take care of multicollinearity among different variables and assess the independent contribution of each variable in explaining the variance of the response variable (HiC-frequency), we performed PLSR analysis, which decomposed the m-dimensional explanatory variables into m number of different components through a linear transformation. The linear transformation was

constrained by the criteria that the first component should explain the maximum variance of the response variable. The PLSR analysis was implemented using the package called PLS in **R** and data was Z-normalized for this purpose. The underlying model used in PLSR is described in Appendix C.

Mapping of genomic features to restriction fragments

To compute the interaction frequency between different genes, we assigned the genes to restriction fragments by mapping their TSS coordinates. To compute connectivity among origins of replication, we mapped the midpoints of origins onto the restriction fragments.

Binning the data for correlation and PLSR analysis

For each functional variable, we averaged the PCC values between different genes for a unit change in the interaction frequency of *cis* and *trans* interactions to smoothen the data for correlation and PLSR analysis.

Random sampling for calculation of significance of colocalization

We used the bootstrapping-based method to calculate the FDRs pertaining to the non-random spatial co-localization of genomic features via *cis* and *trans* interactions. For each feature under question, we first calculated its actual value; either the connectivity as in the case of origins of replication or the average interaction frequencies as in the case of essential genes. Then we perform the 1000 random samplings of genomic coordinates for either the replication origins or the essential genes while preserving the original chromosomal distribution. For each of these 1000 randomly sampled coordinates, we calculated the connectivity (for replication origins) or average interaction frequency (for essential genes) to generate a null distribution. Then we calculated the FDR according to the following relation.

$$p = \frac{1}{b}\left(1 + \sum\nolimits_{1}^{b} (1)_{k_o > k_r}\right)$$

Where b is the number of random samplings and the rest of the term counts the cases where the observed value (k_o) is greater than the randomized value (k_r).

Calculation of connectivity

The connectivity was calculated by taking the ratio of the observed number of interactions and the all possible number of interactions among replication origins. The possible number of interactions (Z_e) was calculated as per the following relation.

For *trans* interactions:

$$Z_e = \frac{1}{2} \times \left(\sum_{\substack{i,j=1 \\ i \neq j}}^{i,j=N} \left(X_i \times X_j\right) \right)$$

where N was the total number of chromosomes (N =16 for yeast) and X_i and X_j were the number of origins located on the i^{th} and j^{th} chromosomes, respectively.

For *cis* interactions:

$$Z_e = \sum_{i=1}^{i=N} \frac{1}{2} \times \left(X_i \times (X_i - 1)\right)$$

where N was the total number of chromosomes and X_i was the number of origins located on i^{th} chromosome. For Z_o number of observed interactions among replication origins in *cis* or *trans,* the relation Z_o/Z_e was used to calculate the connectivity; where Z_e was the all possible number of interactions among replication origins in *cis* or *trans.*

Randomization of the chromatin interaction network

To calculate the FDR for the observed associations of *trans* and *cis* interactions with the co-replication and the co-fitness respectively, we randomized the chromatin interaction network of yeast while conserving its topological degree distribution. We used an **R**-package called **igraph** for this purpose and performed 1000 randomizations. For each randomization, we repeated the correlation analysis to calculate FDR according to the following relation.

$$FDR = \frac{b+1}{m}$$

Where m is the total number of randomizations performed and b is the number of instances when the PCC obtained from randomized network exceeded the observed value of correlation.

Annotation of preferred genomic domains of either *cis* or *trans* interactions

For the comparison of preferred domains of *cis* and *trans* chromatin interactions, we used a sliding window approach with a window size of 20kb and calculated the average number of *cis* and *trans* interactions for each window across each 20kb window across the yeast genome. We then identified the genomic windows, where the number of either *cis* or *trans* chromatin interactions was at least 50% of the maximum value obtained in the whole genome. The genomic regions corresponding to such windows were annotated as *cis*-rich or *trans*-rich domains.

To establish the coordination of early and late replication through *trans* and *cis* interactions, respectively, we identified the early and late replication domains by taking the upper and lower quartile of the yeast replication timing data from a previous study (Raghuraman et al. 2001).

Other calculations for analysis of *cis* interactions

To investigate the relationship between *cis* interactions and gene essentiality, we first smoothened the HiC data into 20kb bins and calculated the number of essential genes mapping onto each bin. The average interaction frequency for the comparison of essential and non-essential gene clusters represented the average value of *cis* frequency for each 5kb bin across the yeast genome.

The high and low noise genes were calculated by taking the upper quartile and lower quartile of the abundance corrected expression noise (see below). The overlap ratio of HindIII and EcoRI libraries of HiC data was calculated by dividing the number of shared interactions between two libraries by the total number of interactions present in the HindIII library.

Mean and variance of interaction distance was calculated between the midpoints of either the 5kb bins (for essential genes) or restriction fragments (for expression noise) and their corresponding interaction partners in *cis.*

The ChIA-PET data for RNA-PolII mediated chromatin interactions in mESCs was taken from Tang et al. (2010). The mouse genomic bin-pairs having the ChIA-PET count of ≥3 were considered to be interacting.

Abundance correction of expression noise

The expression noise is usually taken as the coefficient of variation of gene-expression measurements from single cell studies or from replicates of a single experiment. However, a significant proportion of expression noise calculated by this method can be explained by the expression level of the genes as noted previously (J. Li et al. 2010; Newman et al. 2006). Therefore, it is necessary to correct the calculated values of expression noise of genes by their transcript abundance. For this purpose, we calculated the residuals of LOESS (locally estimated scatterplot smoothening) regression (Cleveland 1979) between mRNA abundance and coefficient of variation (i.e the raw noise) and used these values as abundance corrected noise. LOESS regression is a non-parametric form regression where the subsets of data are fitted using 1^{st} ot 2^{nd} order polynomials by using the weighted least squares (Cleveland 1979). We used the **R** function called '**loess**' to perform the LOESS regression.

Wilcoxon's rank sum test

To test for the difference in the distributions of a functional genomics variable between two groups of genes or genomic regions, we used Wilcoxon's rank sum test which is a non-parametric method to test the null hypothesis. The null hypothesis in Wilcox's rank sum test is that the median of two groups is same and rejection of null hypothesis required the p-values obtained in this test to be below a certain threshold, which is often 0.01 or o.05. We used the **R** function **wilcox.test** for this purpose.

Principal component analysis (PCA)

Principal component analysis is a dimensionality reduction technique which is widely used on genomics analysis to visualize the multi-dimensional dataset in a 2D representation. In the simplest form, PCA can be expressed as a linear transformation of n × m matrix **X** into another matrix **Y** according to the following relation:

$$\mathbf{Y} = \mathbf{PX}$$

where **P** is the linear transformation which transforms **X** into **Y** of the same dimensions.

The objective of PCA is to remove the redundancy from data to provide a meaningful low dimensional representation of the original data. Therefore, the columns in the resulting output matrix **Y** should co-vary as little as possible. This objective is achieved by selecting an orthogonal linear transformation **P**, of the matrix **X**, such that it diagonalizes the covariance matrix of **Y.** The diagonalization of the covariance matrix is achieved by the decomposition of the input matrix **X** into the linearly independent (orthogonal) components by the appropriate choice of the transformation matrix (appendix B). A brief mathematical treatment of PCA is presented in appendix B for the interested reader.

The most common way to interpret the results of PCA is to make a 'biplot' (Gabriel 1971). Biplots are made by simultaneously plotting the loadings of variables and scores of individuals, usually by considering the 1st and 2nd principal component of the PCA in a single 2-D representation. A biplot, therefore, summarizes the key relationship between different variables and individuals (or the data points) by considering the variance which is captured by the 1st and 2nd principal component.

We used the function **'PCA'** from the **R** package called **'FactoMineR'** for obtaining the principal components of GO-terms data of factors described above. Prior to running PCA, we scaled and centered the GO-terms dataset using the **R** function 'scale' with setting the value of 'scale' and 'center' field as TRUE. For generating the variable map of PCA analysis, we used the function **'fviz'** from package **'factoextra'**, by setting the value of **'element'** field as **'var'**. For inferring the chromatin factors associated with transcription and genome integrity related functions, we obtained the list of chromatin factors located in the first and second quadrant of PCA biplot based on the sign of their scores on the first and second principal component.

Identification of affected and unaffected binding targets of chromatin factors.

For binding maps of chromatin factors across the yeast genome, we obtained the genome-wide location data for 202 chromatin factors as provided in the supplementary tables of a ChIP-chip study by Venters et al. 2011. We defined the list of significantly bound genes for each factor as per the FDR criteria used in the original study. For measuring the genome-wide transcriptional effects upon deletion of chromatin factors, we used the microarray data from Lenstra et al. 2011, which measured expression profiles of 6123 yeast genes following the deletion of 165 chromatin factors. For identification of genes whose expression is sensitive to the perturbation of chromatin factors in this dataset, we used the default p-values cutoff of 0.05 with additional criteria of 1.5 fold change in their expression level. The datasets that were used for the analysis of affected and unaffected genes are given in table 7.

Calculation of effect-to-binding ratio (EBR) for chromatin factors

70 chromatin factors were shared between chromatin factors binding and chromatin factor deletion datasets described above, however, we used 67 chromatin factors in our analysis based on the availability of FDR cutoffs for significance for the genome-wide binding of chromatin factors. For quantifying 'direct' transcriptional consequences of chromatin factors deletion, we defined a measure called EBR for each chromatin factor as follows:

$$EBR(f) = \frac{(B_f \cap E_f)}{B_f}$$

Where, B_f was the number of significant binding targets of chromatin factor f, defined based on significance criteria, and E_f was the number of transcriptionally affected targets based on p-values and expression thresholds described above. The relation $B_f \cap E_f$ measured the common targets between the gene lists corresponding to E_f and B_f.

Classification of unaffected and affected (+/-) genes

We first pooled the affected and unaffected genes from the deletion of 67 chromatin factors into two separate lists. Then we subtracted the list of affected genes from the unaffected gene list reducing the size of the unaffected gene list. In this way, we retained only unique genes in the two lists. The affected gene list was further split into two parts corresponding to positively affected and negatively genes affected based on the sign of log2 fold change in Lenstra et.al dataset and retained only the uniquely called genes in each list. This processing left us with 3901 unaffected genes and 749 positively affected and 548 negatively affected genes.

Identification of paralogues of chromatin factors

We performed the blast search using 'blastp' option (https://blast.ncbi.nlm.nih.gov/Blast.cgi). All the unique BLAST hits with E-Value <=1e-05 were counted as paralogue for a given factor.

Gene ontology (GO) analysis of protein interaction partners of chromatin factors

We mapped the GO slim terms corresponding to the field 'biological process' onto the protein-protein interaction (PPI) network of chromatin factors obtained from BioGRID *(https://thebiogrid.org)*. For each factor, we calculated the proportion of its PPI partners having transcription term (GO:0006366) or genome integrity related terms such as DNA Repair, DNA Replication, DNA Recombination, DNA Damage Response, DNA Metabolic Process and (GO:0006281, GO:0006260, GO:0006310, GO:0006974 & GO:0051052). The resulting dataset was analysed using principal components analysis.

Gene ontology analysis of affected and unaffected targets

The GO analysis was performed using the Panther database *(www.pantherdb.org)*. The GO terms that corresponded to the terminal most node (most precise) of the graph were reported in this study.

Aggregation plots for single nucleotide variations (SNVs).
Data of spontaneously occurring SNVs in mutation accumulation lines of yeast was obtained from a previous study (Zhu et al. 2014). To check the enrichment of SNVs at affected and unaffected binding targets of chromatin factors, we counted SNVs within 500 bp windows in the 5kb region around TSSs of affected and unaffected. Average occurrence of SNVs was represented as an aggregate plot centered around TSS.

Mapping of cryptic unstable transcripts (CUTs)
We calculated the total SAGE counts of CUTs (from condition 1 in Neil et al. 2009) which mapped within 500bp promoter region upstream to each TSS. We normalized these counts by the expression level of its downstream gene which was calculated by mapping the total SAGE counts of mRNA (from condition 2 in Neil et al. 2009).

Statistical tests
All the statistical tests used in this study were performed in **R**. For determining the significance of the difference between two distributions, we used **'wilcox.test'** function in **R**, and for determining the significance of the correlation, we extracted the p-values given by **'cor.test'** function in **R**.

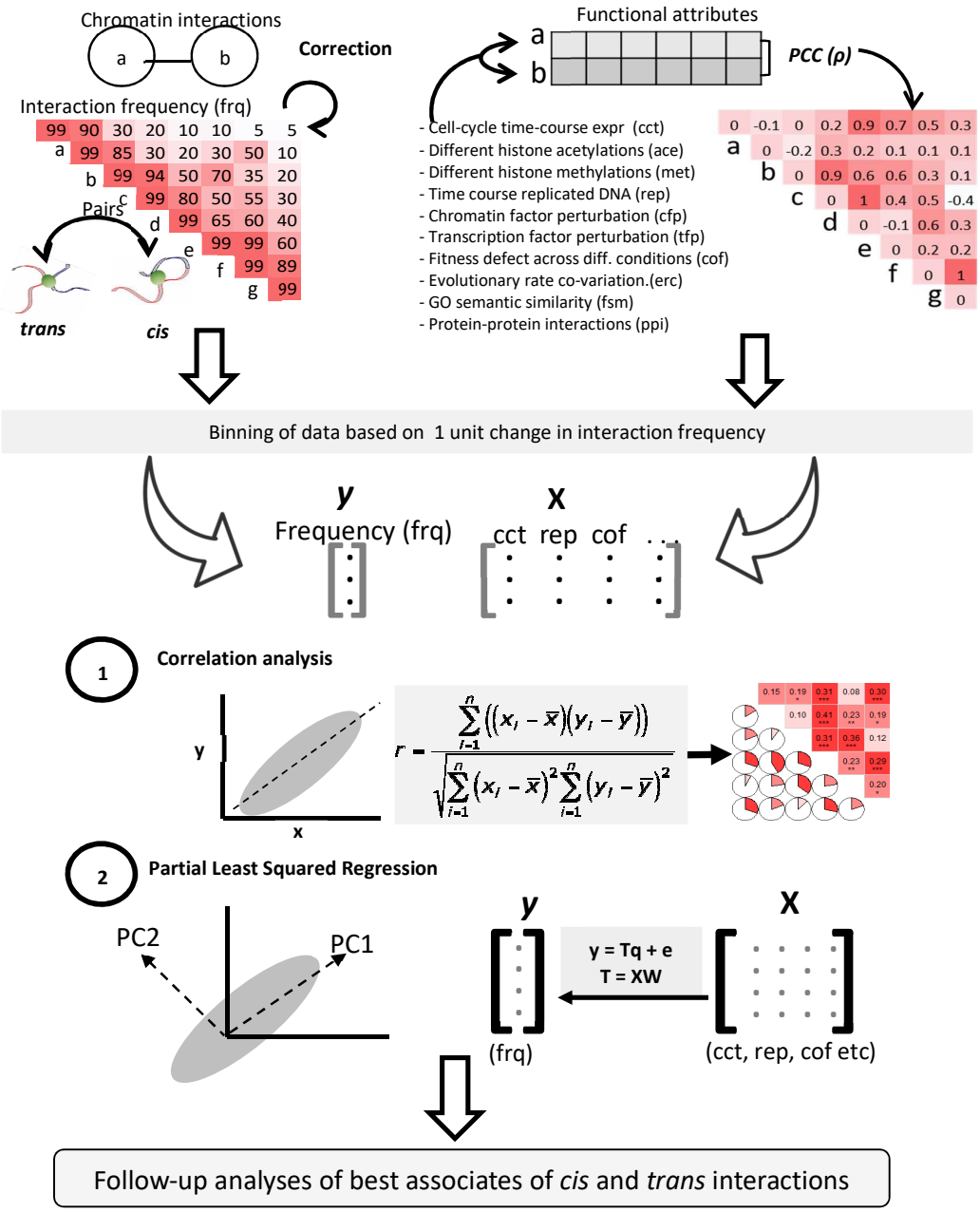

Figure 2.2 A schematic of workflow for the analysis of chromatin interaction data.

We first corrected the chromatin interaction dataset using HiCNorm and obtained the pair-wise interaction frequency values between the gene pairs. For each gene pair, we also computed the similarity in their functional attributes (eg. co-expression). Then we binned these two datasets for a unit change in the interaction frequency. The resulting vector of interaction frequency and matrix of functional similarities (**Y** and **X** respectively) were analyzed using the correlation and PLSR analysis. The best associates of *trans* and *cis* interaction frequency were further subjected to in-depth analyses.

2.3 Experimental methods

Yeast strains

Yeast strains used in the study were BY4741 (wild-type), *Δdot1* and *Δswc5* strains from Euroscarf *Saccharomyces cerevisiae* deletion library (*www.euroscarf.de*). The strains used for FOA assay were made using the transformation of the URA3 gene as described in the following sections. For the long term storage, the yeast strains were cultured overnight (16-18hrs) and the saturated culture was mixed with 50% (v/v) sterile glycerol in 1:3 proportion. These glycerol stocks were immediately stored at -80°C. The yeast strains were streaked on YPAD plates from the glycerol stocks as per need and were revived at 30°C for ~2 days before proceeding for the further experiments.

Molecular biology reagents

All the enzymes and reagents in PCR reactions, restriction enzymes used for cloning of the URA3 gene were purchased from New England Biolabs, Invitrogen, Sigma Aldrich, USA. The gel-extraction and the plasmid miniprep kits were obtained from Qiagen, USA, and Bioneer, Korea.

Culture media

Yeast strains were either cultured in YPAD-broth or YPAD-plates as per need. FOA assays were performed on SC-plates (synthetic complete) containing 1% FOA. All the media components, chemicals, and reagents were purchased from Sigma Aldrich (USA), HiMedia (India), Merck Ltd. (USA), Difco (USA), and Formedium, UK. 1000 ml of YPAD media contained 10g of yeast extract, 20g of peptone, 40 mg of adenine, and 20g of glucose. These components were dissolved in Millipore distilled water to the final volume of 1000ml. To make YPAD plates, we added 20gm of Agar to the components of YPAD. SC plates with FOA were made by adding 1gm of FOA to 6.8g of YNB, 2gm of dropout mixture of amino acids, 20gm of glucose, and 20gm of agar in 1000ml of Millipore distilled water. To make SC-ura plates, we used the dropout mixture without uracil and did not use FOA. Optionally, we added 1% Nat to YPAD plates which were used for the selection of positive transformants. The media used was sterilized by autoclaving at 15 lb/inch2 (psi) pressure at 121ºC.

Transformation of yeast strains

7µl of the construct (URA3-NatNT2) was used for transformation of 100µl of cells (competent cells). The cells were mixed in 600µl of PEG under sterile conditions. This was followed by a 30min incubation at 30°C and 25min incubation at 42°C, after which, the cells were immediately frozen in ice for 5min. After diluting the mixture of cells and PEG with ~1ml YPAD, the cells were pelleted down by centrifugation at 13000rpm for 1min. The supernatant was discarded and the pellet was resuspended in 2ml fresh YPAD media. The re-suspended pellet was transformed to 5ml glass tubes which were incubated at 30°C for 6hrs. The revived culture was platted on YPAD plates containing 1% Nat by pelleting down the 2ml culture and by re-suspending in ~100-

200µl of YPAD after discarding the supernatant.

Design of URA3-based reporter assay

For genomic URA3 insertion, a plasmid was made by inserting the URA3 gene, with its own promoter and terminator, in pFA6a-natNT2 vector (Janke et al. 2004) through SalI and BglII restriction sites. We used the PCR based strategy similar to gene deletion/tagging (Janke et al. 2004) to insert the URA3 gene with the NatNT2 cassette into a target locus. This cassette was flanked by 65 bp sequences which were homologous to the desired site of tagging in the budding yeast genome (figure 2.2). We transformed the amplified chromosomal tagging construct into the WT (BY4741-α) and Δdot1 or Δswc5 strains. The positively transformed colonies were selected by the acquisition of resistance against Nat and were confirmed using colony PCR.

The selection of the target site for FOA assay

The selection of the tagging site was based on the genomic coordinates of the binding peaks of either Dot1or Swc5 across the yeast genome. These binding coordinates were based on the mapping coordinates of Yeast Genome Oligo Set Version 1.1 (Operon Biotechnologies) and were obtained from Venters et al. (2011). For each Dot1 and Swc5, we selected a binding target site which gave the large value of ChIP-chip signal in Venters et.al data. The target selection was additionally constrained by the negligible change in its transcription upon the deletion of the corresponding chromatin factor in Lenstra et.al dataset. In addition to high-affinity binding targets, we also selected non-target loci for Dot1 and Swc5 as control sites.

FOA assay

5-FOA performs selection against uracil prototroph by causing toxicity through the action of ODCase on FOA. The accumulation of mutations in the transformed URA3 gene shall render it non-functional, thereby allowing it to grow on FOA plate (figure 2.3). We used FOA assay to score for the difference in mutagenesis at the selected sites between wild-type yeast strains and the deletion strains of Dot1 or Swc5. For this assay, the yeast cells were grown overnight to saturate the culture. Next day, the cells were inoculated in a fresh YPAD broth at an OD of 0.2, and were allowed to grow till the OD of 1-1.5. From this culture, the volume corresponding to 1OD of 1ml culture was used to plate the cells on FOA plates. Serial dilutions were used to prepare a 10k times diluted culture, which was platted on –URA plates to mark the proper functioning of the assay. The plates were incubated at 30°C for 5-7 days in order to score for the difference in colony count between WT and deletion strains of Dot1 and Swc5.

Primers used for the chromosomal tagging of the target and non-target binding sites

We designed the primers that contained a 65bp long stretch of homology with the intended site of chromosomal tagging. To confirm the tagging at an appropriate location, we designed an additional primer for each transformation which annealed ~150-200bp upstream of the site of tagging and used it along with the reverse primer from within the tagged construct. A list of the

primers used in this study is given in table 8. The primers are named as per "factor:purpose:direction" format, where factor can be either Dot1 or Swc5 and purpose can be the target, non-target (i.e tagging either the target or non-target sites for Dot1 or Swc5). The yeast strains made in this study are given in table 9.

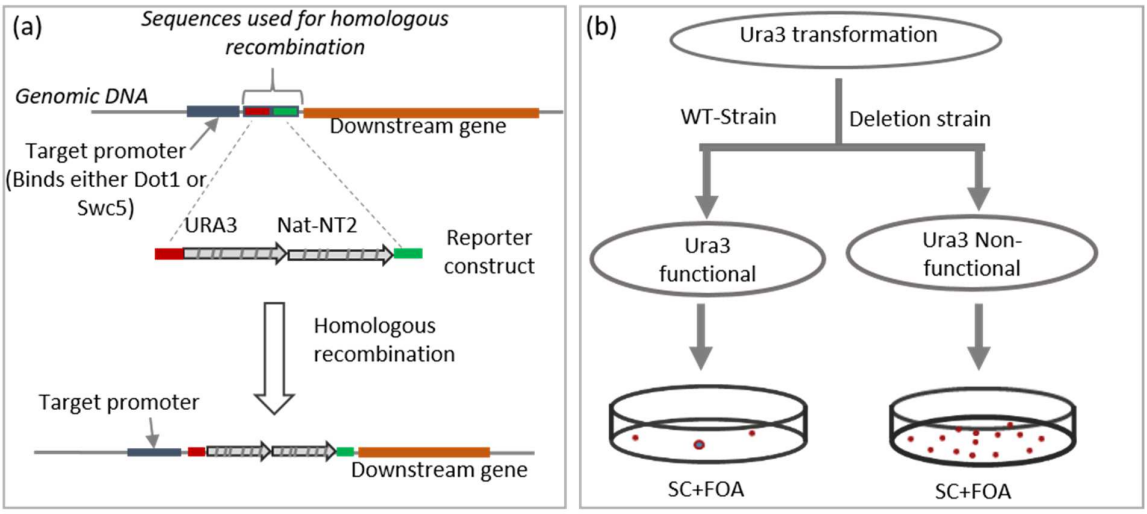

Figure 2.3 Schematic of URA3-based reporter and FOA assay.

a) Reporter construct (URA3-NatNT2) is flanked by sequences which are homologous to the desired site of chromosome tagging. The choice of these sites is described in above. The positively transformed colonies had the reporter construct inserted near the binding site of either Dot1 or Swc5. b). Ura3 gene was transformed into the WT and mutant strains of chromatin factors (*Δdot1* and *Δswc5*). Assuming the hypothesis that these two chromatin factors play a role in preventing the spontaneous DNA damage and considering the fact that only the cells which are uracil auxotroph can grow on FOA plates, fewer number of colonies were expected to grow on FOA plates for WT strains as compared to the deletion strains.

Table 8 List of primers used for chromosomal tagging

#	primer Name/description	Primer Sequence
1	Dot1:Target:Forward	5'GGAAACCTGAATGAACTTGAAGACCATTCCCAACTCCACTATTAATTAAACTTAGTACGCTCGAGCGTACGCTGCAGGTCGAC3'
2	Dot1:Target:Reverse	5'GTACGTGATTCTCCGATACCCGCACTCATTGCTTCGTCCCTGATTTGGTGGAATATTATCCTTGATCGATGAATTCGAGCTCG3'
3	Dot1:CheckTarget-Forward	5'CCAGGGCTTCTAACGATTTCC3'
4	Dot1:CheckTarget:Reverse	5'GAGGCAAGCTAAACAGATCTGG3'
5	Dot1:Non-Target:Forward	5'TTGTCTGTGGATTATACTTTTAATTGGATATGTTACTCCTCAATTTCTGTTATTTGCTATTCCTTCGTACGCTGCAGGTCGAC3'
6	Dot1:Non-Target:Reverse	TATTCTCTTCAACTCTCTAGATGCACGAAGATATAACGTACATACGGTATAATAAAGAGCGCATAATCGATGAATTCGAGCTCG
7	Dot1:CheckNonTarget:Forward	5'GTTACGTTTTCATAATGTCACGC3'
8	Dot1:CheckNonTarget:Reverse	5'GAGGCAAGCTAAACAGATCTGG3'
9	Swc5:Target:Forward	5'GTATATATAGATATATTATAGACGTTTAAAAGTTTTAAGGGATCGGATTTTCAGCCAGATATTTTCGTACGCTGCAGGTCGAC3'
10	Swc5:target:Reverse	5'GATATCTTACTCATCACGCAAGACAGCAGCAGTGAAAAATAAAAAAAAAAAATTAGAATAAAGTAAATCGATGAATTCGAGCTCG3'
11	Swc5:Check-Target:Forward	5'GTTGGAATTCTGATACAAAAAGCC3'
12	Swc5:CheckTarget:Reverse	5'GAGGCAAGCTAAACAGATCTGG3'
13	Swc5:Nontarget:Forward	CACCTGTATAGAGTGTTTAAAAAGAGATCCTTCTTGCTAGCTGAAAAAGTTTATGTCGAATAAAACGTACGCTGCAGGTCGAC
14	Swc5:Nontarget:Reverse	GCAGCAAGCGCATCCATCGCTGCATCTCCGCCGGAAAAACAGATGTTCAGTACATGAAAAGACGTATCGATGAATTCGAGCTCG
15	Swc5:Check-Target:Forward	5'GAGCTGAGTGCTTTGTGTGCG3'

Table 9 List of yeast strains used in this work.

#	Name	Genotype	Source
1	BY4741 (MAT a)	his3Δ1 leu2Δ0 met15Δ0 ura3Δ0	Euroscarf
2	*Δdot1*	his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 Δdot1::KanMX4	Euroscarf
3	*Δswc5*	his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 Δswc5::KanMX4	Euroscarf
4	WT-Target-Dot1	his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 $P^{BIM1:Ura3NatNT2}$	This work
5	WT-Target-Swc5	his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 $P^{ECM31:Ura3NatNT2}$	This work
6	WT-NonTarget-Dot1	his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 $P^{YHR054C:Ura3NatNT2}$	This work
7	WT-NonTarget-Swc5	his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 $P^{MIP1:Ura3NatNT2}$	This work
8	*Δdot1*-Target-Dot1	his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 Δdot1::KanMX4 $P^{BIM1:Ura3NatNT2}$	This work
9	*Δdot1*-NonTarget-Dot1	his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 Δdot1::KanMX4 $P^{YHR054C:Ura3NatNT2}$	This work
10	*Δswc5*-Target-Swc5	his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 Δswc5::KanMX4 $P^{ECM31:Ura3NatNT2}$	This work
11	*Δswc5*-NonTarget-Swc5	his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 Δswc5::KanMX4 $P^{MIP1:Ura3NatNT2}$	This work

Chapter 3 Results

3.1 Results of the project I

3.1.1 Genome-wide landscape of *cis* and *trans* interactions in the yeast genome

The territorial organization of chromosomes is well established in metazoans but remains elusive in fungal species like budding yeast. The territorial organization predicts that occurrence of *cis* (intra-chromosomal) interactions to be more abundant as compared to *trans* (inter-chromosomal) interactions due to the limited intermingling of chromosomal territories. To investigate the genome-wide landscape of *cis* and *trans* interactions in the yeast genome, we calculated and plotted the number of *cis* interactions of 20 kb genomic regions against the number of *trans* interactions of the same regions. The resulting scatter plot (figure 3.1 a) clearly showed that certain genomic regions were biased to be engaged in either the *cis* interactions or *trans* interactions. Similar inferences were also made when we plotted the average frequencies of *cis* vs. *trans* interactions of 20 kb genomic regions (figure 3.1b).

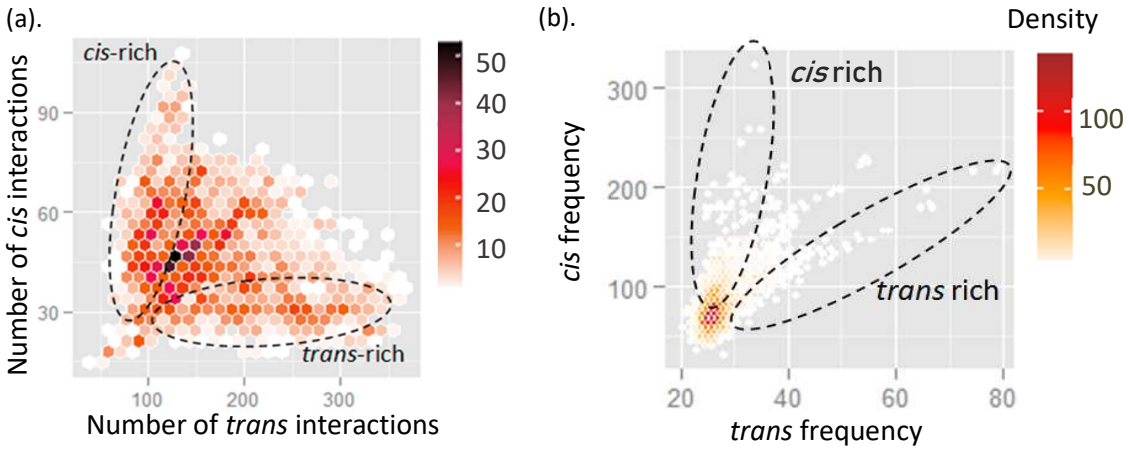

Figure 3.1 *cis* vs. *trans* interactions in yeast genome.

a) The scatter plots shows the Z- normalized number of *cis* and *trans* interactions of 20kb regions of yeast genome. The color bar indicates the density of data points in the scatter plot. The regions biased for *cis* or *trans* chromosomal interactions are indicated by dotted ellipses. b) The scatter plot shows the average frequency of *cis* interactions vs. *trans* interactions for each genomic region. The color bar indicates the density of data points in the scatter plot.

These results indicated that the yeast genome might have preferred domains of *cis* and *trans* interactions. In order to directly visualize these domains, we made the aggregation plots by using the moving average of the number of *cis* and *trans* interactions for 10 consecutive windows of 20kb across the entire genome of the yeast. As shown in figure 3.2, the red *(*for *cis)* and black *(*for *trans)* lines did not overlap across most of the yeast genome, which clearly showed the existence preferred domains of *cis* and *trans* interactions in the yeast genome.

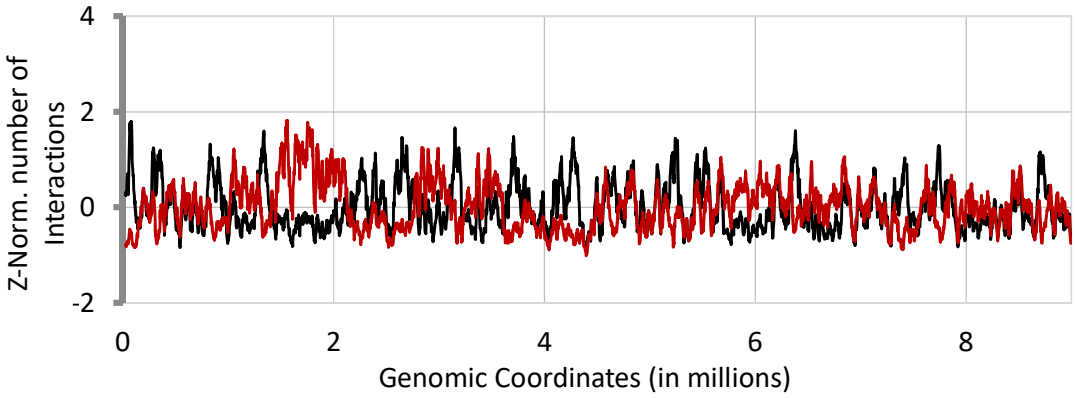

Figure 3.2 The number of *cis* and *trans* interactions as a function of genomic coordinates.
The red and black lines represent the Z-normalized number of *cis* and *trans* interactions calculated for 10 consecutive windows of 5kb across the yeast genome. Notice that broad domains across the genome, which are differentially enriched for either the *cis* or *trans* interactions.

To highlight the existence of preferred domains of *cis* and *trans* interactions, we also visualized the 3-dimensional model of yeast chromosome 4 which was generated by Duan et al in 2010. As shown in figure 3.3a-b, the regions enriched in *cis* interactions (figure 3.3b, dashed box) folded in a compact conformation in 3D-representation as compared to regions enriched in *trans* interactions (figure 3.3a)

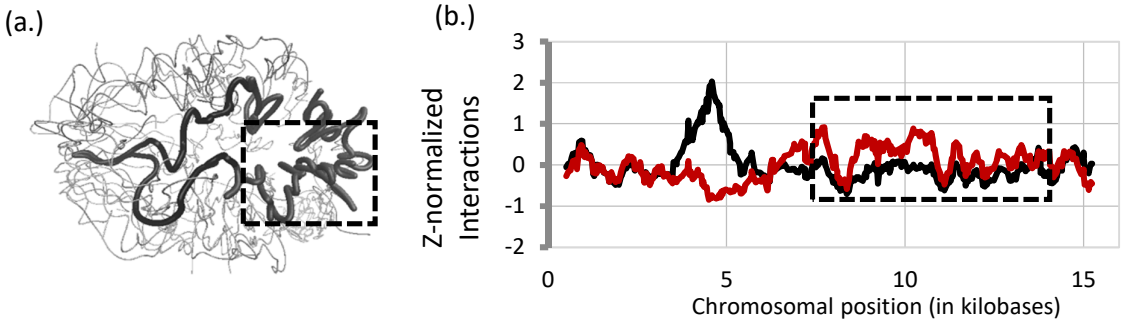

Figure 3.3 Preferred domains of *cis* and *trans* interactions in yeast chromosome 4.
a). The 3D model generated using Duan et al.'s method. Chromosome 4 is highlighted with thick grey line and thin lines represent rest of the chromosomes. (b) Line plot of Z-normalized number of *cis* (red) and *trans* (black) interactions which were smoothened by taking the moving average of 10 consecutive windows. The dotted box is the chromosomal region enriched in *cis* interactions.

The existence of preferred domains for *cis* and *trans* interactions raised the possibility that distinct functional and evolutionary constraints might underlie their spatial organization. For this reason, we performed a separate analysis of *cis* and *trans* interactions for the rest of the work (see figure 2.1 in Materials and methods).

3.1.2 Correlation analysis of *cis* and *trans* interactions

After establishing that the yeast genome has preferred domains engaging in *cis* and *trans* interactions, we sought to investigate the functional constraints which might have shaped the genomic landscape of these interactions. For this purpose, we performed the correlation analysis of *cis* and *trans* interaction frequency with several different functional variables of the yeast genome (table 5, Appendix A). Consistent with the notion of preferred domains of *cis* and *trans* interactions, we found that the functional correlates of the *cis* interactions were significantly different from that of *trans* interactions. In particular, we observed that the frequency of *trans* interaction frequency was most strongly correlated with the co-replication of engaged genes (figure 3.4). On the contrary, the *cis* interaction frequency showed the strongest correlation with co-fitness of the interacting genes (figure 3.4).

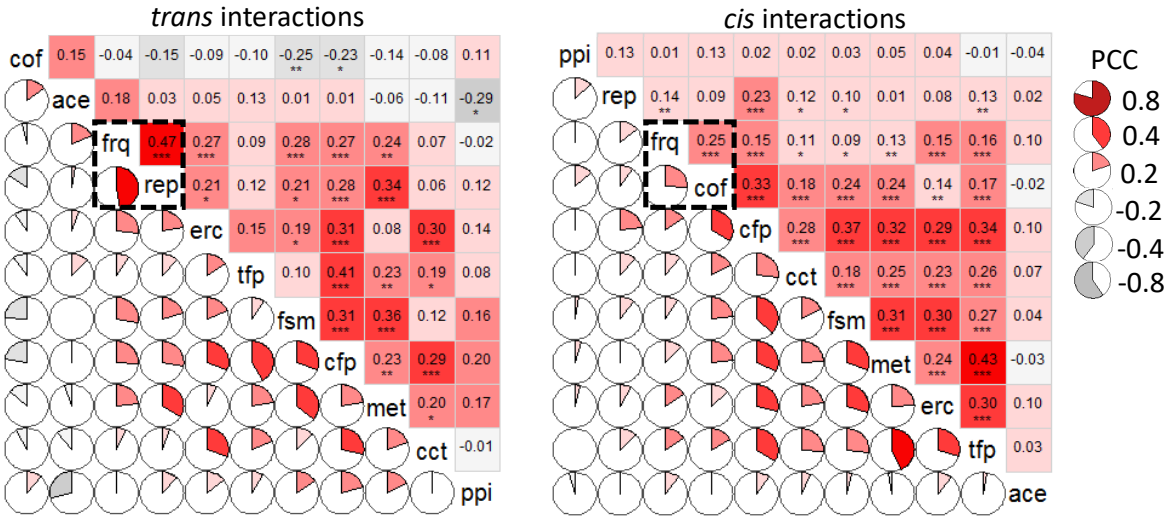

Figure 3.4 Correlation analysis of *trans* and *cis* interactions.

Pearson's correlation coefficients between different variables are represented as pie-charts in lower triangles and as heatmap in the upper triangle of the correlaograms for *trans* (left) and *cis* (right) interactions. The numbers in upper triangle show the value of correlation. Dotted box shows the variable which most strongly correlated with the interaction frequency. Name of variables are given along the diagonal in abbreviated form and expand as follows: frq (frequency), cof (co-fitness), ace (acetylation), met (methylation), erc (evolutionary rate co-variation), tfp (transcription factor perturbation), cfp (chromatin factor perturbation), cct (cell cycle time-course), ppi (protein-protein interactions), fsm (functional similarity), rep (replication).

The simplest interpretation of these results is that the genes which were spatially co-localized in *trans* had a similarity in their replication timing while the genes which were spatially co-localized in *cis* had a similarity in their fitness defects (i.e. co-fitness, which quantifies the similarity in

fitness defects of two genes when their individual deletion strains were grown in chemically different environmental conditions). A careful look at the correlation matrix described in figure 3.4 suggested that many of the functional genomic variables used in this study were correlated amongst each other. The presence of multiple correlated variables can delude the interpretation of the correlation analysis and often fails to deduce the true associates of the response variable (interaction frequency in this case). To decorrelate different functional variables (or the predictor variables) into different components of independent variables, we performed the partial least square regression (PLSR) analysis (Materials and methods).

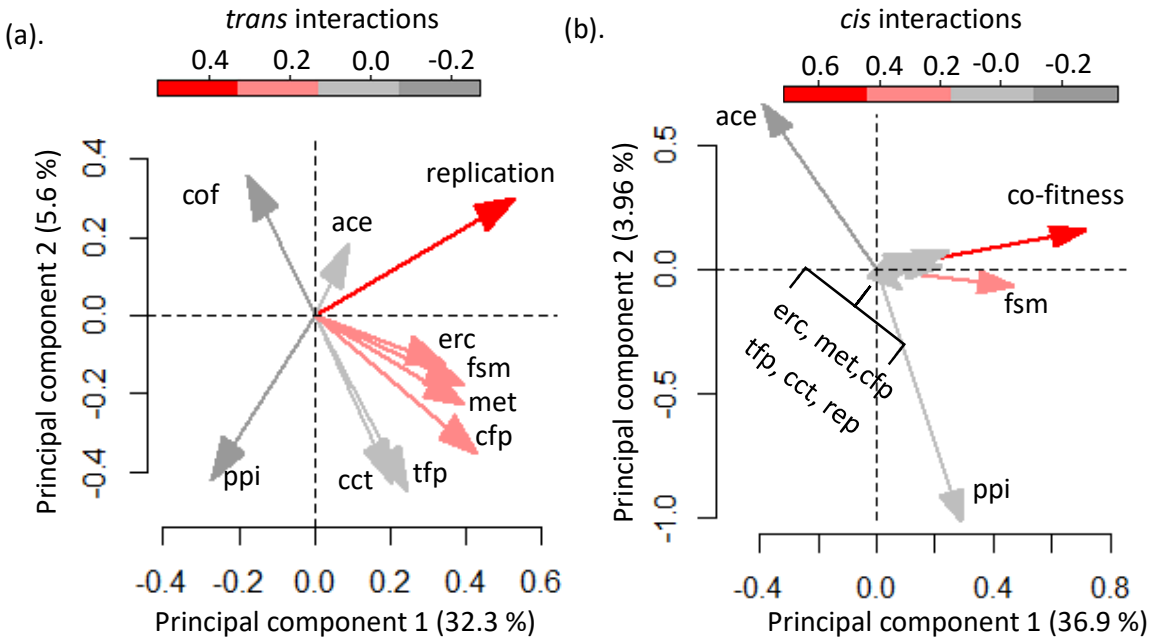

Figure 3.5 Variable map of PLSR analysis of the *trans* and *cis* interactions.
Arrows in the biplot indicate the relative orientation of loading vectors within Ist and IInd principal component. The color bar at the top indicates the loading of each variable onto the first component. The expansion for abbreviations of arrow names is as follows: frq (frequency), cof (co-fitness), ace (acetylation), met (methylation), erc (evolutionary rate co-variation), tfp (transcription factor perturbation), cfp (chromatin factor perturbation), cct (cell cycle time-course), ppi (protein-protein interactions), fsm (functional similarity), rep (replication). The numbers along-side the axis labels indicate the % variance of the interaction frequency explained by 1st and 2nd principal component.

To interpret the results of PLSR analysis, we made a biplot (Materials and methods) by using the first two components of PLSR analysis of *cis* and *trans* interactions. As shown in figure 3.5a, the largest component for *trans* interactions explained ~32.3% variance in the HiC interaction frequency and was maximally contributed by the co-replication index of the interacting genes. On the other hand, the largest PLSR component for the analysis of *cis* interactions explained

36.9% of the variance in the HiC interaction frequency and had co-fitness of interacting genes as a major contributor (figure 3.5b). Both the correlation and PLSR analysis showed that co-replication of the spatially proximal genes was the strongest associate of the *trans* interactions, while co-fitness of spatially proximal genes was the strongest associate for the *cis* interactions. To assess the robustness of observed distinction between the association of *cis* and *trans* interactions with co-fitness and replication, we repeated the correlation analysis of *cis* and *trans* interactions the using the EcoRI library of chromatin interaction data (Z. Duan et al. 2010). As shown in figure 3.6a, we observed that the replication and co-fitness were the strongest variables associated with the *trans* and *cis* interaction frequency, respectively.

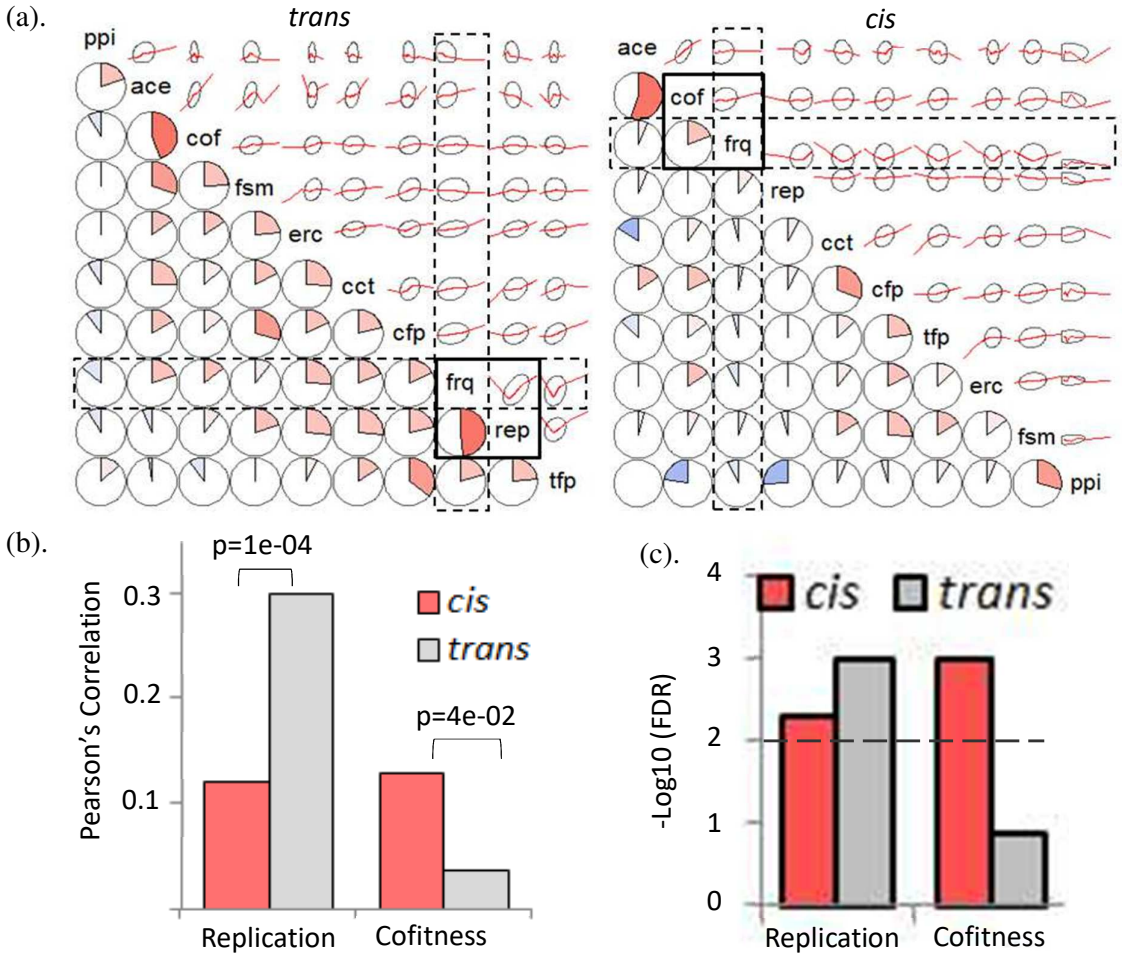

Figure 3.6 Robustness of the functional associates of *trans* and *cis* frequency.

a) Correlograms of *trans* (left) and *cis* (right) interactions by using the EcoRI library of chromatin interaction data. The solid rectangle highlights the strongest associates of *trans* and *cis* interaction frequency. The expansion of abbreviation along the diagonal same as in figure 3.4&3.5. b) The correlation of interaction frequency with replication or cofitness for *cis* and *trans* interactions using Rutledge et al.'s (2015) HiC data. c) Barplots of FDR calculated after 1000 rewiring of yeast ChIN. The horizontal dotted line corresponds to the FDR value of 1%.

Secondly, we also used the HiC data from yet another study in budding yeast (Rutledge et al. 2015) and found a similar association of *trans* and *cis* interactions with replication and co-fitness (figure 3.6b). Additionally, we also performed the 1000 controlled rewirings of the yeast chromatin interaction network by preserving its degree distribution (Materials and methods) and repeated the correlation analysis of rewired network to calculate the FDR for the correlation of *trans* and *cis* interaction frequency with the co-replication and co-fitness. We found that while co-fitness was significantly associated with *cis* interaction frequency, co-replication was associated with *trans* as well as *cis* interaction frequency through the significance for *trans* interaction frequency was stronger (figure 3.6c). We discern the observed significance of replication with *trans* as well as *cis* interaction frequency in the next sections. The robustness of our analysis motivated us to further investigate the underlying cause for the association between *trans* and *cis* interaction frequency with co-replication and co-fitness of spatially proximal genes.

3.1.3 Spatial co-localization of origins of replication

Several studies have shown that replication program of eukaryotic genome is spatially organized into discrete loci called replication factories and some studies have shown the gene looping events to be involved in the organization of replication factories (Chagin, Stear, and Cardoso 2010; Guillou et al. 2010). The process of replication begins at specific predefined loci across the genome which are called origins of replication. Based on the observed correlation of replication timing with *trans* interaction frequency in light of the replication factories, we tested if the origins of replication might spatially cluster in 3D space in order to coordinate the replication timing of the yeast genome.

To test this hypothesis, we downloaded the coordinates of early and late origins of replication and calculated the connectivity (Materials and methods) among origins of replication. To assess if the calculated value of connectivity among origins represented the non-random spatial clustering, we compared this value against the random null generated *via* bootstrapping (Materials and methods). We observed that the early origins of replication exhibited non-random spatial clustering specifically *via trans* but not *via cis* interactions (figure 3.7a). The late origins, on the other hand, exhibited the non-random spatial clustering specifically *via cis* but not *via trans* interactions (figure 3.7b). Further, the early and late origins of replication did not exhibit significant co-localization with each other either in *cis* or in *trans* (P=0.997 in *cis* and P=0.988 in *trans*).

These observations indicated that the non-random spatial interactions among early origins of replication might underlie the observed correlation of *trans* interaction frequency with the co-replication of spatially interacting genes. Additionally, we have observed a weaker albeit significant correlation between *cis* interaction frequency and co-replication in the correlation analysis (figure 3.4 &3.6c).

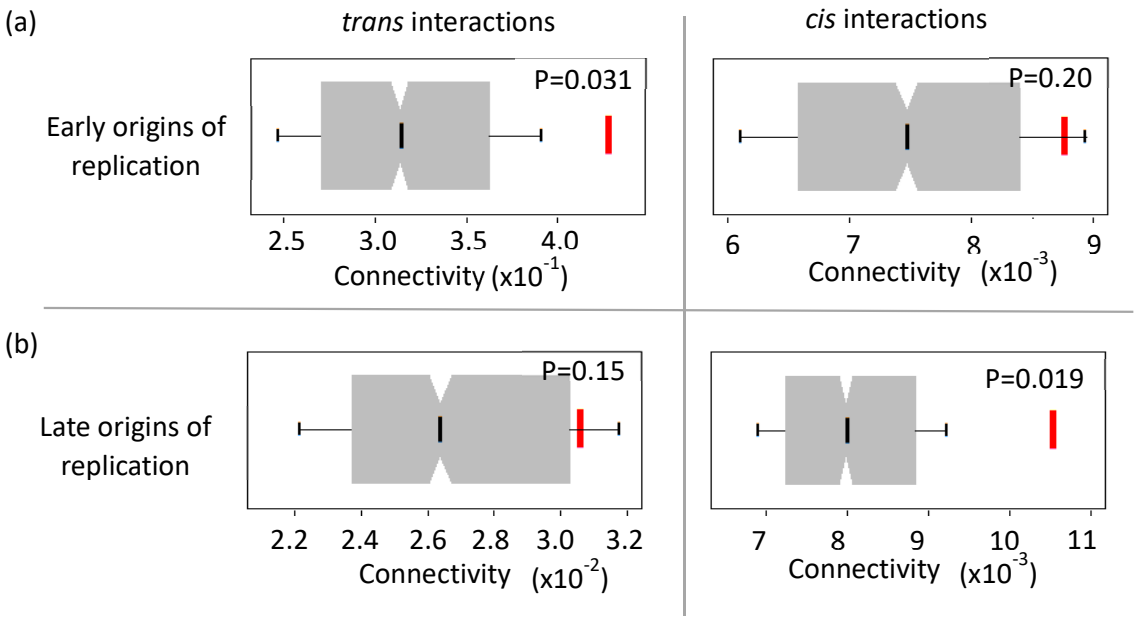

Figure 3.7 Spatial clustering of replication origins.

The Boxplots represent the mean and standard deviation of null distribution for the connectivity of for early (upper pane) and late (lower panel) origins through *trans (*left) and *cis* (right) interactions. The vertical red line indicates the observed value of connectivity for early origins of replication (a) and for late origins of replication (b).

The non-random spatial clustering of late origins of replication in *cis* indicated that this correlation might have resulted from the non-random spatial clustering of late origins of replication in *cis.* Taken together, these observations indicated that *trans* and *cis* interactions among genes might coordinate the early and late replication timing by the spatial segregation of early and late replication *via* non-random clustering of early and late origins, respectively.

3.1.4 Spatial segregation of early and late replication factories

To confirm the spatial segregation of early and late replication described in the previous section, we identified the early and late replication domains in the yeast genome based on the genome-wide replication timing data (Materials and methods). Upon comparing the HiC interaction frequency of early and late replication domains with their co-replication profiles, we found a stronger correlation between co-replication of early replicating domains with *trans* interaction frequency but not with the *cis* frequency (figure 3.8). On the other hand, the late replicating domains exhibited a significant correlation between interaction frequency and the co-replication for both the *cis* as well as *trans* interactions (figure 3.8). However, the strength of significance, as measured through –log10 transformed p-values, was greater for the *cis* interactions as compared

to *trans* interaction for late replication domains (figure 3.8). We reasoned that this difference in significance can be attributed to the greater number of data points in *cis* (711) as compared to *trans* (146) for the case of late replicating domains as the correlation coefficient and the p-values of significance decay with the increase in the sample size. Taken together, these results confirmed that *trans* and *cis* interactions coordinate early and late replication *via* non-random spatial clustering of early and late origins of replication, respectively.

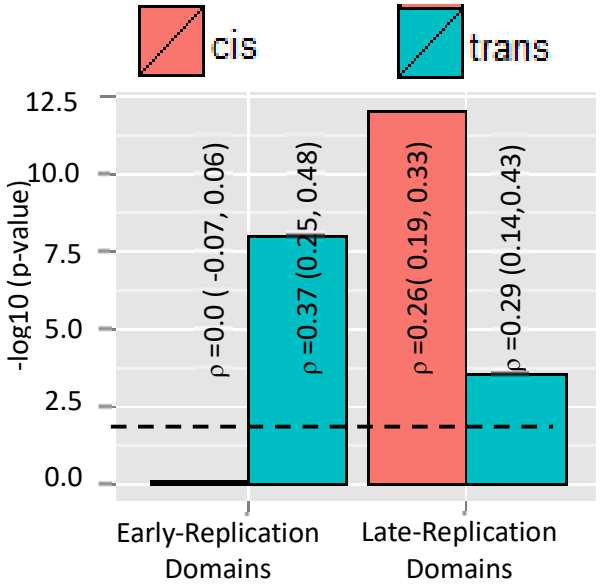

Figure 3.8 Spatial segregation of early and late replication.

The barplots for the significance of the correlation between co-replication and *cis*/ *trans* interaction frequency of early and late replicating domains. ρ is the Pearson's correlation coefficient between interaction frequency and co-replication with 95% confidence intervals in brackets. Dotted line shows significance of 0.05.

3.1.5 Association between mutation rates and the chromatin interactions

Since replication associated mutagenesis has been shown to be a major source of mutations in eukaryotic genomes (Lang and Murray 2011), we speculated that the spatially proximal gene pairs might also exhibit similarity in their mutation rates. Indeed, we observed that in the early replicating domains, the similarity in the mutation rates of gene-pairs (table 6) correlated strongly with the *trans* but not with *cis* interaction frequency (figure 3.9). But in the late replicating domains, the similarity in mutation rates of gene-pairs exhibited similar correlations with the *cis* as well as *trans* interaction frequency (Figure 3.9). However, the late replicating domains exhibited a greater significance of this correlation for *cis* as compared to the *trans* interaction frequency (figure 3.9). The co-varying mutation rates of genes *via trans* and *cis* interactions were consistent with the spatial segregation of early and late replication *via* trans and *cis* interactions

and highlighted the possible role of 3D genome in the evolution of spatially proximal genes.

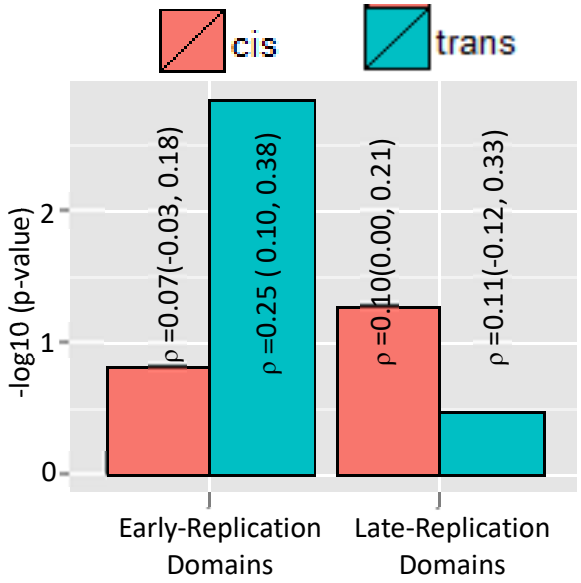

Figure 3.9 Covariance of mutation rate *via cis* and *trans* interactions.

The barplots for the significance of correlation of similarity in mutation rates with *cis* and *trans* interaction frequency for the gene pairs within early and late replication domains with. ρ is the Person's correlation coefficient between interaction frequency and co-replication profiles of gene pairs within early and late replication domains along with the 95% confidence intervals given in the brackets.

3.1.6 Association between nucleosome occupancy and chromatin interactions

The process of replication is intricately linked with the chromatin structure and nucleosome occupancy. Since we had observed the spatial segregation of early and late replication *via trans* and *cis* interactions, respectively, we sought to test if the *trans* and *cis* interactions also have a distinct preference for nucleosome occupancy. For this analysis, we made the scatter plots of the Z-normalized nucleosome occupancy against the number of *trans* and *cis* interactions of 3kb genomic bins (Materials and methods). We found that the nucleosome occupancy exhibited a negative correlation with the number of *cis* interactions and a positive correlation with the number of *trans* interactions (figure 3.10a-b, P<2.2e-16 for both *cis* and *trans*).

We also observed a negative correlation between the nucleosome occupancy and the *cis*-to-*trans* ratio of the number of interactions (figure 3.10c). These results clearly suggested that the *cis* and *trans* interactions have a distinct preference for nucleosome occupancy.

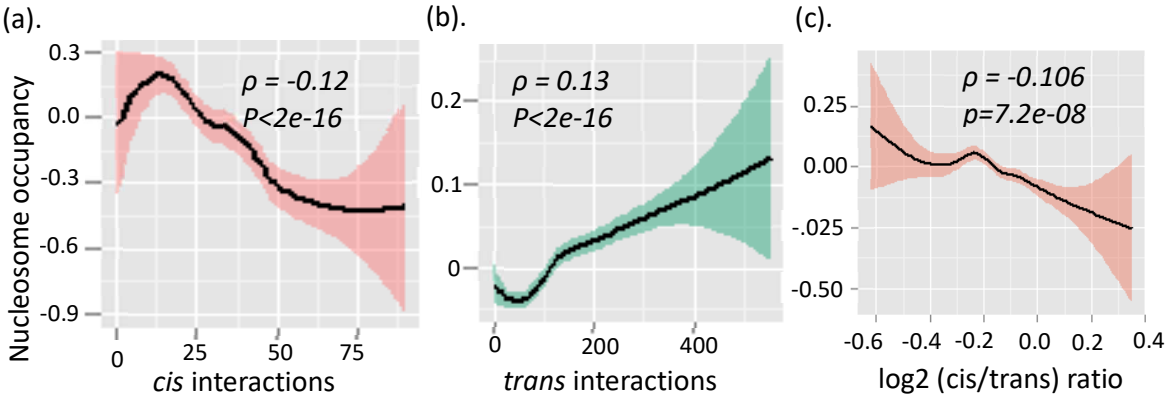

Figure 3.10 Association between nucleosome occupancy and chromatin interactions.
Shown are the scatter plots of Z-normalized nucleosome occupancy against the number of *cis* interactions (a), number of *trans* interactions (b), and the *cis*-to-*trans* ratio of number of interactions (c). ρ is Pearson's correlation coefficient. The p-values are the significance of correlations.

To directly visualize this distinction, we plotted the number of *cis* and *trans* interactions as a function of the genomic coordinates and annotated regions as *cis*-rich and *trans*-rich by the rolling window approach (Materials and methods). These regions were then compared for nucleosome occupancy. Interestingly, we observed a remarkable overlap between higher nucleosome occupancy, enrichment of *trans* interactions and depletion of *cis* interactions (figure 3.11a).

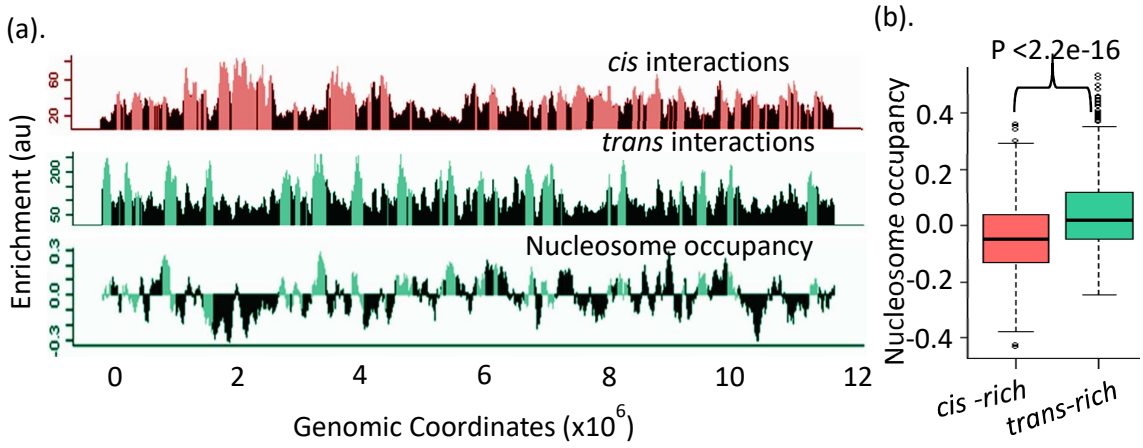

Figure 3.11 Nucleosome occupancy of *cis*-rich and *trans*-rich genomic domains.
a). The number of *cis* and *trans* interactions as a function of genomic position. All yeast chromosomes concatenated in chr1-16 order for simplification. Red and green colours represent *cis-rich* and *trans-rich* genomic domains respectively. b) The boxplots of nucleosome occupancy in the *cis*-rich and *trans*-rich genomic regions. P-values were calculated using Wilcoxon's test

The peaks of nucleosome occupancy showed a striking concordance with the peaks of *trans*

interactions. As evident in the boxplot, nucleosome occupancy of *trans*-rich regions was significantly greater as compared to *cis*-rich regions (figure 3.11b). These results, therefore, highlighted that *cis* and *trans* interactions have a distinct preference for nucleosome occupancy.

Next, we investigated the nucleosome occupancy at origins of replication. We observed an overall greater nucleosome occupancy at early as compared to late origins (figure 3.12a). This observation aligned well our earlier observations that the early origins interacted non-randomly in *trans* (figure 3.7) and *trans* interactions had a preference for higher nucleosome occupancy (figure 3.11). To discern the difference between nucleosome occupancy at the early and late origins, we mapped the nucleosome occupancy data from G1 phase of the on and around origins.

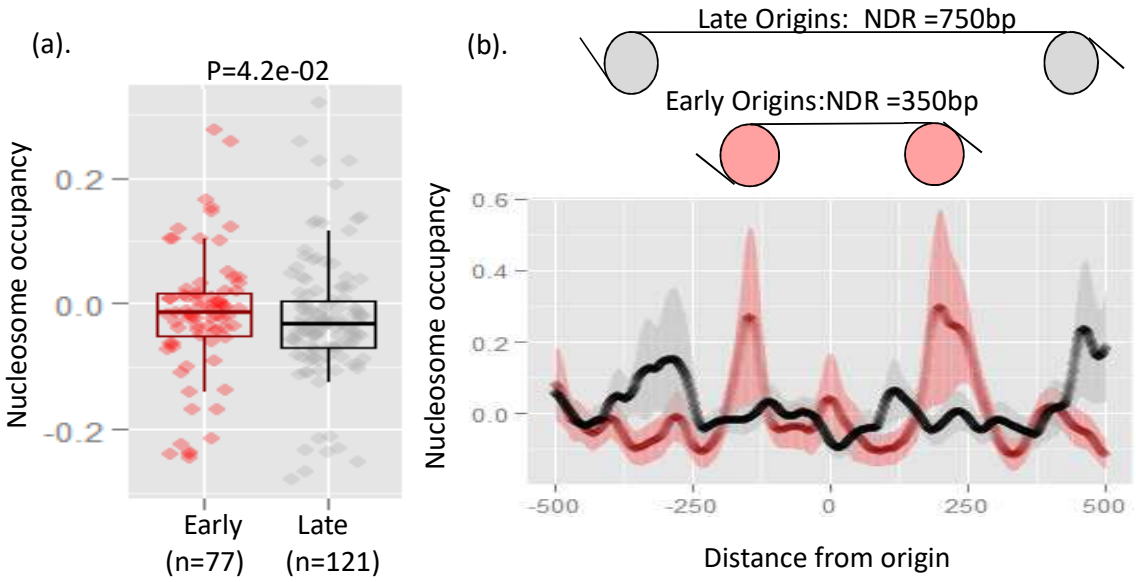

Figure 3.12 Nucleosome occupancy on and around early and late origins of replication.
(a) Boxplots of distribution of nucleosome occupancy around (10kb window) early and late origins of replication. The p-value was calculated using Wilcoxon's test. (b) Average nucleosome occupancy at +/-500 bp around early (red) and late (black) origins of replication. The shaded regions indicate the standard error margins calculated from nucleosome occupancy of all early (77) and late (121) origins.

We observed that both early and late origins of replication had a sharp window of nucleosome depletion regions (NDR, figure 3.12b). However, we found that the size of NDR for the late origins of replication was wider with an average size of 750 bp as compared to that of early origins replication, which had a narrow NDR of an average size of 350bp (figure 3.12b). We proposed that wider window of NDR in case of late origins might assist in the late replication by exposing the additional DNA sequence elements in order to efficiently assemble ORC, as replication factors are present in the limited amount towards the end of S-phase.

3.1.7 **3D interactions among essential genes**

Correlation and PLSR analysis indicated that *cis* interaction frequency had a strong association with the co-fitness of the interacting genes (figure 3.4, 3.5). Therefore, genes which exhibited similar fitness defect showed evidence for their spatial co-localization through *cis* interactions. Extrapolating this observation, we reasoned that genes with extreme fitness defect such as lethality (i.e. the essential genes) might also be associated with the interaction frequency in *cis* or *trans*. To establish the link between gene essentiality and their chromatin interactions, we used the random null model-based approach (Materials and methods) to survey the interaction profile of essential genes. The result of this analysis indicated that essential genes were engaged in the *cis* interactions of greater average frequency when compared to the null distribution (figure 3.13). Interestingly, this association was particularly more significant for *cis* interactions as compared to *trans* interactions.

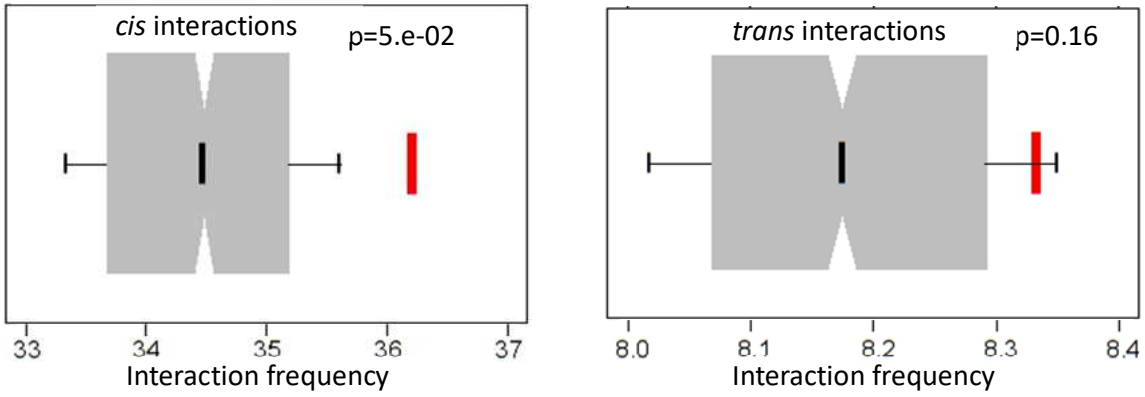

Figure 3.13 Spatial interactions among essential genes.

Boxplots represent the mean and standard deviation of random null distribution of average interaction frequency and the red bar indicates the observed values of average frequency of interactions among essential genes. The p-value was calculated as per materials and methods and section.

As the frequency of the HiC data indicates the strength of spatial co-localization among genes, this would mean that interactions of higher frequencies are present for a longer time in a given cell. This would imply that essential genes, which engage in *cis* interactions of greater frequency, might be spatially restrained. Since essential genes are known to exhibit lower expression noise, the spatially restrained state of essential genes might help to reduce their expression noise by minimization the stochastic fluctuations of the chromatin fiber. To test if the essential genes could be spatially restrained, we proposed that the spatially restrained loci should exhibit a short-range spatial clustering *via* high-frequency *cis* interactions. Consequently, we expected that essential genes should have greater interaction frequency and shorter mean and variation of genomic distances among interacting loci as compared to the non-essential genes (figure 3.14a).

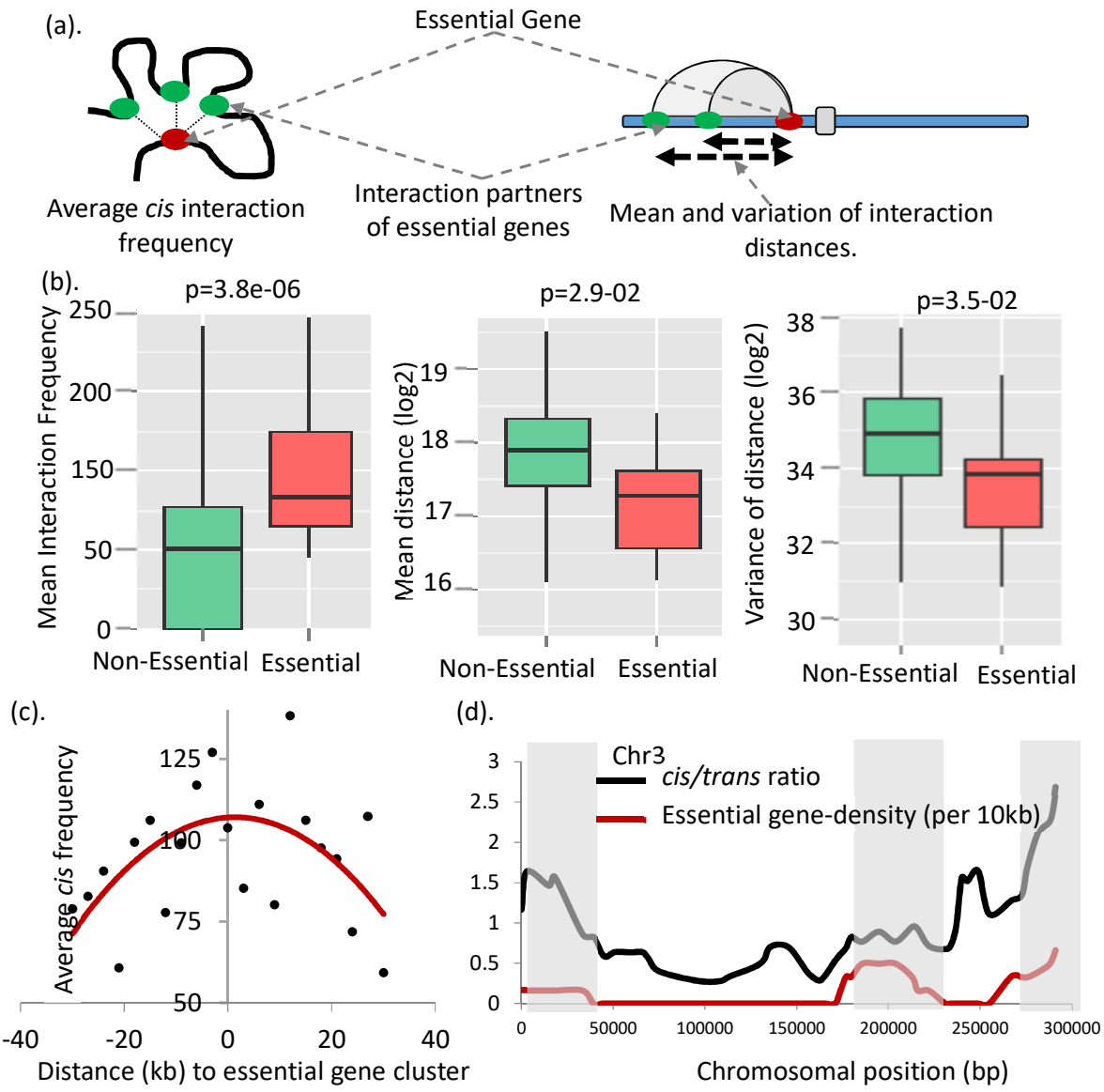

Figure 3.14 Patterns of *cis* interactions of essential genes clusters.

a) Cartoon representations of parameters calculated and plotted in the panel-b. Red dots indicate the essential genes and green dots indicate their spatial interaction partners. b) Boxplots of *cis* interaction frequency (left), mean interaction distance (middle), variance of interaction distance (left) for essential and non-essential genes. The p-values were calculated using Wilcoxon's test. c) Scatter plot average *cis* interaction frequency as a function of distance from the center of essential genes clusters. d). Line plot of essential gene density and *cis/trans* ratio of 10kb genomic bins across chromosome 3

Consistent with this proposal, we observed that genomic regions enriched in essential genes, i.e. the essential gene clusters (>2 essential genes in 5kb bins) indeed had a significantly greater average frequency of interactions as compared to the non-essential genes and the neighbouring

domains of the essential gene clusters (figure 3.14b-c). Further, we also observed that essential gene clusters were involved in the spatial interactions of shorter and less varying genomic distances as compared to the non-essential genes (figure 3.14b). Additionally, the average *cis* interaction frequency strongly increased near the essential genes clusters (figure 3.14c), and as the number of genes in a given cluster were increased (figure 3.14d). These observations suggested the strong influence of the gene essentiality in shaping the global landscape of *cis* interactions.

3.1.8 Gene expression noise and *cis* interaction frequency

Essential genes are present in the low nucleosome occupancy regions of the genome and their association with higher frequency *cis* interactions is consistent with our observation of lower nucleosome occupancy at genomic regions enriched in *cis* interactions (figure 3.11). The essential genes are also known to exhibit lesser gene expression noise and we speculated that spatially restrained state of essential genes might be one mechanism involved in the minimization of stochastic noise in their gene expression. To test the association between expression noise and their 3D interaction pattern, we calculated the gene expression noise for the yeast genes, performed the abundance correction, and identified the genes exhibiting higher and lower expression noise (Materials and methods, figure 3.15).

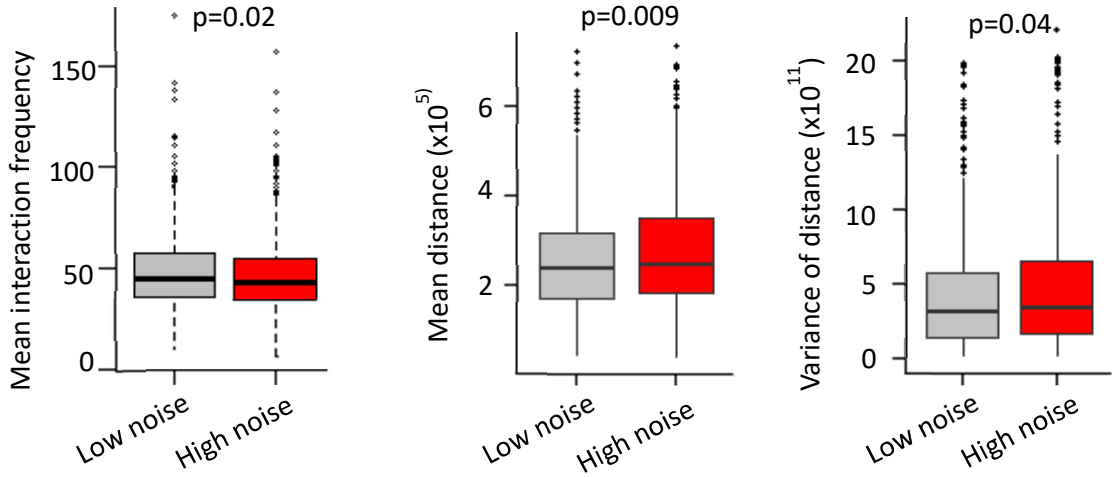

Figure 3.15 *cis* interactions and gene expression noise.

Boxplots of average *cis* interaction frequency (left), mean interaction distance (middle) and variation of interaction distance (right) for the yeast genes exhibiting low and high expression noise. Low and high noise genes were defined based on upper and lower quartiles of the abundance corrected noise. The p-values were calculated using Wilcoxon's rank sum test.

Interestingly, we found that the genes having low expression noise had relatively greater average *cis* interaction frequency, and had a lesser mean and variance of the genomic distances among interacting sites (figure 3.15). These results suggested the short-range spatial clustering of low

noise genes *via cis* interactions of higher average frequency, similar to essential genes (figure 3.14b).

To further support our observations, we hypothesized that spatially restrained chromatin would be more likely to show the similar chromatin interaction profile between different biological replicates of HiC data, but the relatively mobile chromatin might switch its interaction partners, and therefore, would be less likely to show the similar interaction profile between biological replicates. To test this, we calculated the overlap between two different libraries of HiC data; the HindIII and the EcoRI libraries. We found that chromatin interaction partners of genomic regions exhibiting lower expression noise indeed had a greater overlap between these two libraries as compared to genomic regions exhibiting higher expression noises (figure 3.16), consolidating our previous observation that genes having lower expression noise tend to be spatially restrained.

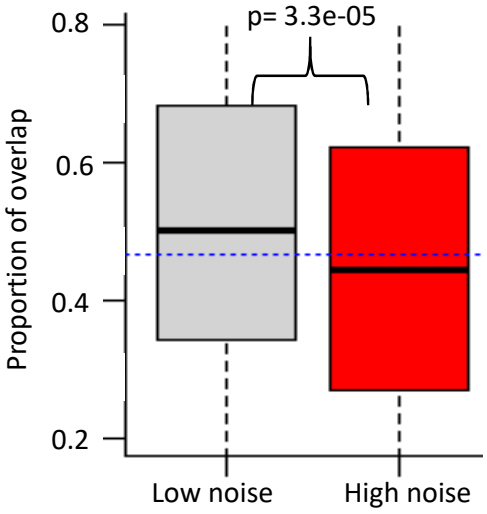

Figure 3.16 Overlap between HindIII and EcoRI libraries of HiC data.

Yeast genomic bins with lower expression noise (grey) have a greater overlap between two libraries of HiC data generated by Duan et al. Proportion of overlap was calculated by dividing the number of shared interactions between two libraries by the total number of interactions in the HindIII library. Dotted blue line is the genome-wide median of overlap.

The chromatin mobility analysis explained so far is based on the analysis of interaction frequencies, genomic spans, and data reproducibility. To directly test if the *cis* interactions frequency can minimize the chromatin fluctuations, we downloaded the data for experimentally determined chromatin mobility for several different loci of the yeast genome (table 6 & appendix A) and investigated its association with the *cis* interaction frequency. We found that the chromatin mobility negatively correlated with *cis* interaction frequency (figure 3.17a), and positively correlated with the variance of genomic distances among interacting sites (figure 3.17a).

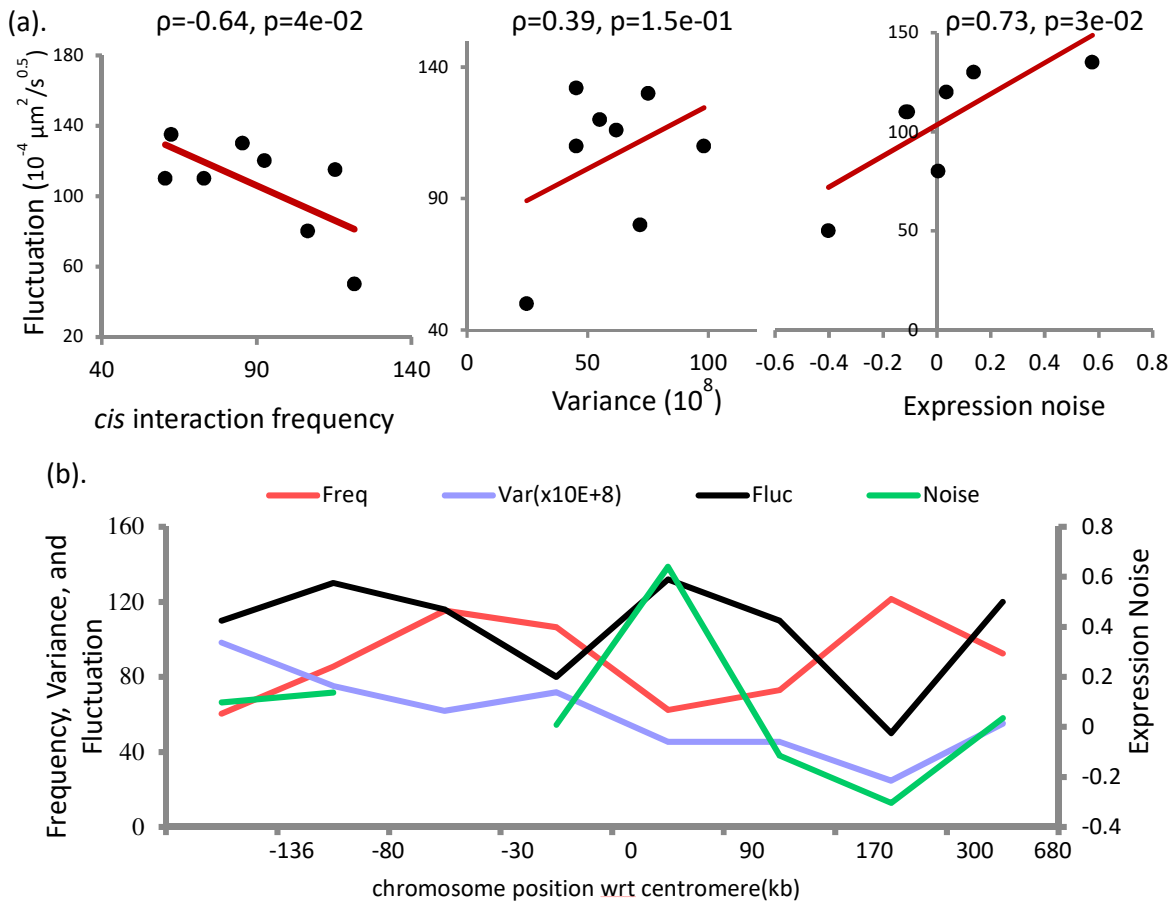

Figure 3.17 Chromatin mobility, *cis* interaction frequency and gene expression noise.

a) Scatter plot of chromatin mobility (Y-axes) against *cis* interaction frequency (left), and variance of interaction distances (middle), and the gene expression noise (right). b) Line plots for chromatin mobility (Fluc), *cis* interaction frequency (Freq), variance (Var) of interaction distances, and gene expression noise as a function of genomic position along chromosome 12.

These results confirmed our previous proposal that short-range *cis* interactions of higher *cis* interaction frequency can restrain the chromatin in a localized environment. We also found that chromatin mobility exhibited a strong positive correlation with gene expression noise (figure 3.17a), which provided a direct evidence that the spatially restrained chromatin can indeed minimize lower expression noise. These relations are also summarized in figure 3.17b.

To consolidate these observations, we also used another dataset (table 6), where the chromatin mobility was measured for autonomously replicating sequence (ARS908 and ARS1413) and the centromeres of chromosome 4 and 13 in yeast. We observed that lower chromatin mobility values were associated with increased *cis* interaction frequency (figure 3.18a-b) and higher chromatin mobility values were associated with decreased *cis* interaction frequency (figure 3.18a-b).

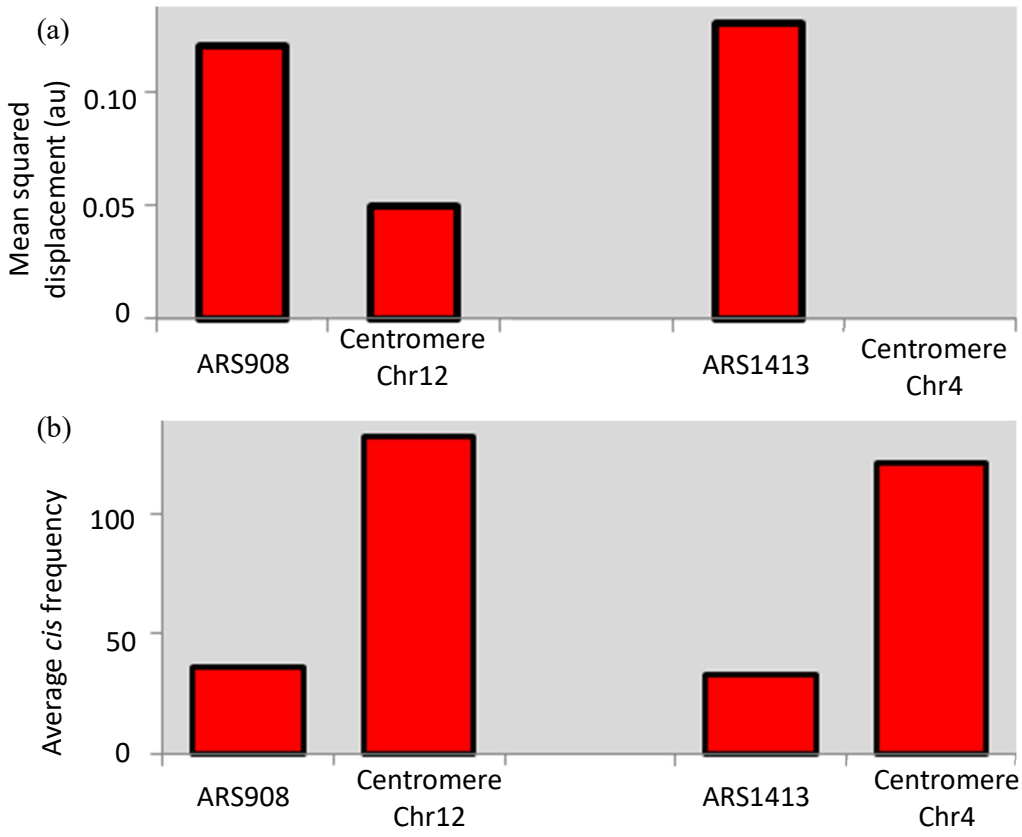

Figure 3.18 Validation of association between chromatin mobility and *cis* interaction frequency. a) Mean-squared displacement of autonomously replicating sequences (ARS908 and ARS1413) and centromeres of chr12 and chr4. b) Average *cis* interaction frequency of genomic regions plotted in a. Notice that inverse relationship of chromatin mobility and interaction frequency by comparing a and b.

Analysis of peaks in HiC data

We further adopted the converse approach and instead of selecting the low noise regions beforehand, we identified the yeast genomic regions which had a significantly greater frequency of *cis* interactions w.r.t rest of the genome. For this purpose, we made the density plot of the average *cis* interaction profile of the whole genome and observed a distinct population of genomic fragments which had significantly greater average *cis* interaction frequencies (figure 3.19a). Interestingly, we found that genes corresponding to the peaks of greater *cis* interacting pairs had a significantly lower expression noise as compared to the random null generated *via* bootstrapping (figure 3.19b). Therefore, the analysis described so far provides robust evidence for the involvement of *cis* interaction frequency in the minimization of gene expression noise.

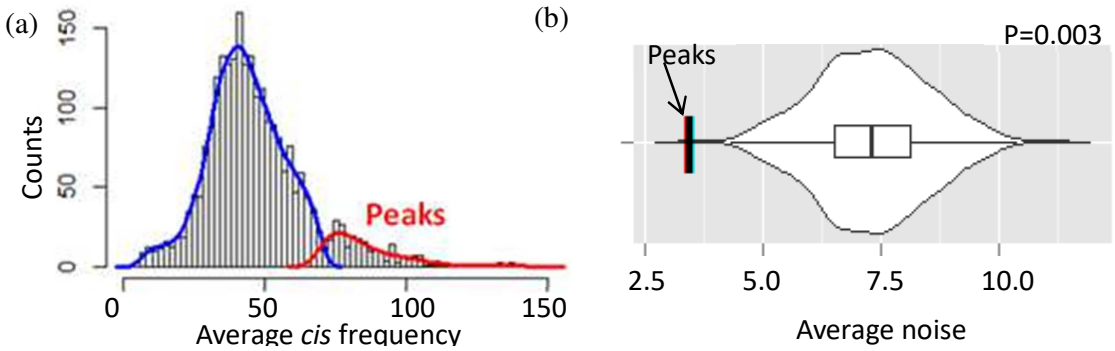

Figure 3.19 Peaks of higher average *cis* interaction frequency and the gene expression noise.
a). A density plot of average frequency of *cis* interactions across the yeast genome. The bars lined by red color correspond to the genomic regions having distinctly higher *cis* interaction frequency. (b). Comparison of the average expression noise of peak regions (thick black bar) with random null distribution. The p-value was calculated using bootstrap method.

3.1.9 *cis* interaction frequency and toxic genes

In addition to essential genes, the genes that can cause toxicity upon their over-expression are also known to exhibit lower expression noise (Batada and Hurst 2007). Therefore, we wondered if the same mechanism might be involved in the minimization of gene expression noise of toxic genes as was implied for essential genes. We did not find the significant enrichment or depletion of the *cis* interaction frequencies of toxic genes when compared to non-toxic genes (figure 3.20). These results implied that spatial tethering *via cis* interactions avoids the stochastic loss of expression but not the abrupt gain of expression.

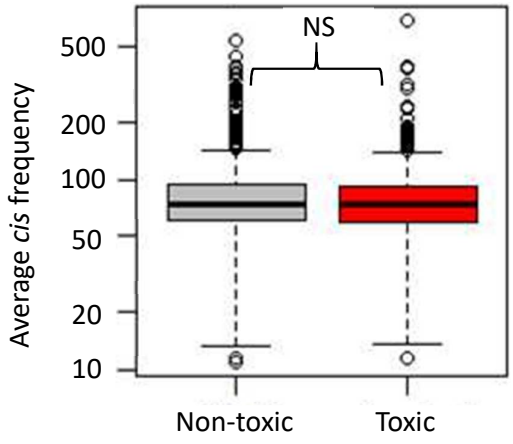

Figure 3.20 cis interactions and over expression toxicity.
Boxplots of average *cis* interaction frequencies of toxic (red) and non-toxic (grey) genes.

3.1.10 Evolutionary conservation of noise-minimization through chromatin interactions

Association of 3D genome organization with the mitigation of gene expression noise of essential loci suggested that such a property might be evolutionarily conserved. To test this, we investigated

the gene expression noise in relation to chromatin interactions in *E.coli* and mouse embryonic stem cells (mESCs) as the HiC and gene expression noise data were available for these two model systems (see table 6 and Appendix A). We obtained the list of essential genes in *E.coli* (table 6) which are also known to exhibit lower gene expression noise (Silander et al. 2012). Next, we checked if the essential genes were non-randomly associated with higher interaction frequency as compared to the random null. Indeed we observed that essential genes in *E.coli* genome were engaged in 3D interactions of higher average interaction frequencies as compared to the null distribution generated through boot-strapping (figure 3.21a) and showed positive scaling with the density of essential genes in genomic regions (figure 3.21b).

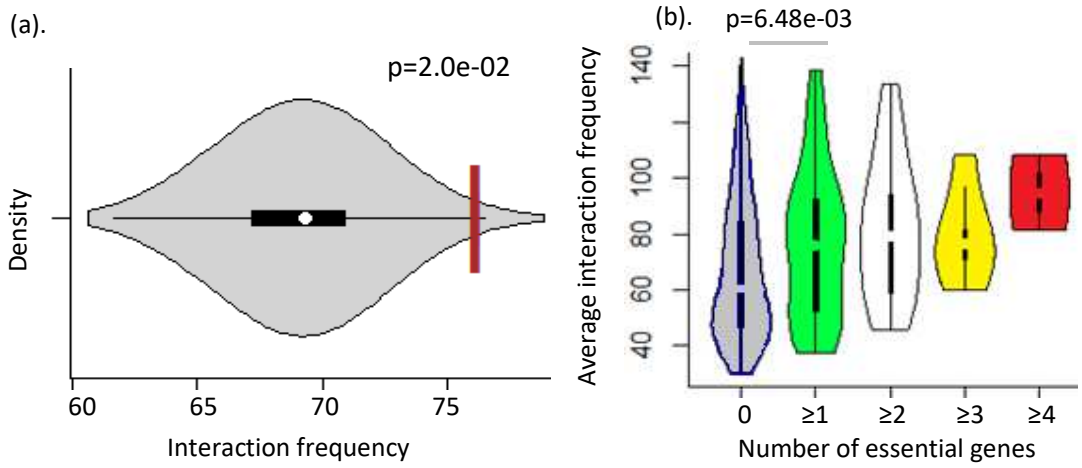

Figure 3.21 *cis* interaction profile of essential genes in *E.coli.*

a) Average *cis* interaction frequency (red bar) of essential genes as compared to the null distribution (grey area) in the *E.coli* genome. The p-value was calculated as per materials and methods. b) Violin plots of the distributions of average *cis* interaction frequencies as a function of number of essential genes. The p-value was calculated using the Wilcoxon's test.

We further analysed the genes having lower and higher expression noise in *E. coli* (table 6; Materials and methods). Similar to yeast, we observed that the genes exhibiting lower expression noise had significantly greater average *cis* interaction frequency as compared to genes exhibiting higher expression noise (figure 3.22, left). Next, we calculate the gene expression noise from single cell RNAseq data of mESCs and compared the *cis* interaction between high-noise and low-noise regions. Remarkably, we observed a trend similar to yeast and *E.coli,* where genes with lower expression noise had a higher average *cis* interaction frequency as compared to genes with higher expression noise (figure 3.22, right). Collectively, these observations indicated that the minimization of gene expression noise might be an evolutionarily conserved property which might have shaped the 3D genome organization.

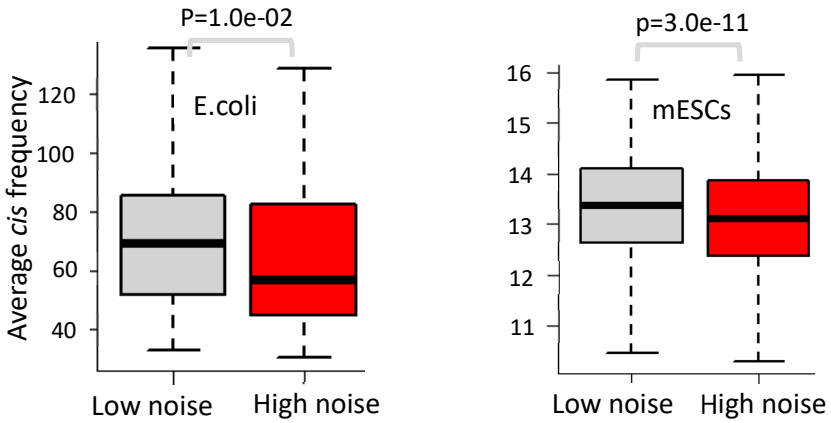

Figure 3.22 Expression noise and *cis* interaction frequency in *E.coli* and mESCs.

Boxplots represent the average *cis* interaction frequency between high and low noise genes in E.coli (left) and mESCs (right). The p-values are from Wilcoxon's test.

3.1.11 **The possible mechanism of noise minimization through spatial interactions**

How the spatial restraint of chromatin fiber might be achieved *via* chromatin interactions? We hypothesized that chromatin mobility can be influenced by the association of chromatin with sub-nuclear structures such as transcription factories. Therefore, it is likely that the genes which are stably recruited at transcription foci might exhibit lower chromatin mobility and consequently lower gene expression noise. To test this possibility, we downloaded the data for RNA-polII tethered chromatin interactions in mESCs measured by ChIA-PET technology (table 6 and appendix A). We observed a significantly higher number of RNA-polII tethered interactions for low-noise genes as compared to high-noise genes (figure 3.23).

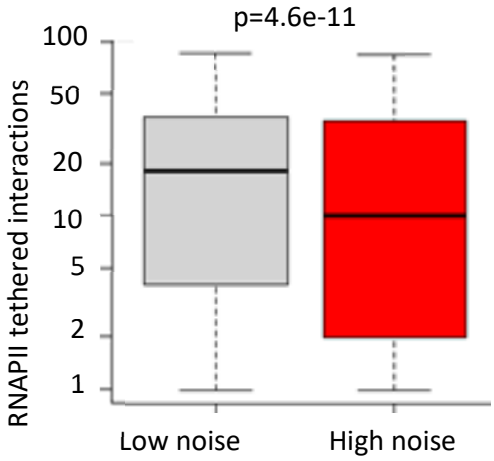

Figure 3.23 RNA-polII tethered interactions of high and low noise genes.

Low noise genes in mESCs exhibit a significantly greater number of RNA-PolII tethered interactions as compared to the high noise genes. The p-value was calculated using the Wilcoxon's test.

This result suggested that the stable localization of genes near the nuclear structures, such as transcription factories *via cis* interactions, might be a possible mechanism involved in the spatial tethering of genomic loci to minimize the gene expression noise.

3.2 Results of project II

3.2.1 Quantification of the disconnect between binding and effect of chromatin factors

To quantify the disconnect between binding and effect, we calculated the effect-to-binding ratio (EBR) as described in materials and methods. The distribution of EBR showed that most chromatin factors have EBRs of less than 0.1, which meant that less than 10% of the bound targets were transcriptionally affected upon the deletion of the respective chromatin factor (figure 3.24a).

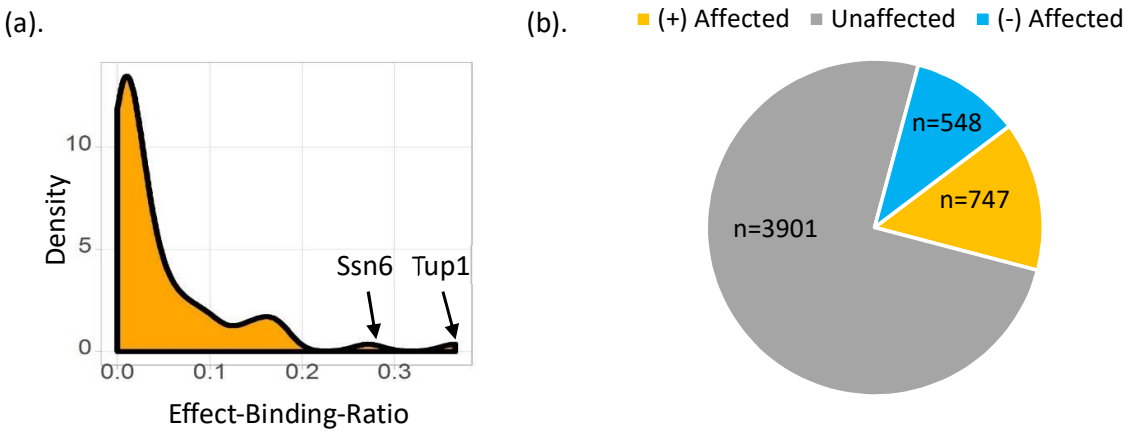

Figure 3.24 Quantifying the global disconnect between binding and effect of chromatin factors. (a). Density plot shows the distribution of EBR. (b). Pie chart shows the number of unique binding target genes which were affected (747 positively affected and 548 negatively affected) or remain unaffected (3901) upon the deletion of respective chromatin factor.

The highest of the EBR was observed for Tup1 (0.36) and Ssn6 (0.27), both of which function as part of a single complex and are known repressors of carbohydrate metabolism genes (Trumbly 1992). Interestingly, in spite of both these factors being part of the same complex, we observed that Tup1 was significantly bound at ~180 bound targets and the Ssn6 was significantly bound at more than 2100 targets. Therefore, the mutually exclusive binding events of these chromatin factors might indicate their important biological role beyond the canonically understood pathways. To investigate the transcriptionally non-functional binding of chromatin factors, we used the EBR as a measure of disconnect for the further analysis presented in this thesis. We further measured the number of genome-wide binding targets of 67 chromatin factors which were affected upon the deletion of each chromatin factor. As shown in the pie chart in figure 3.24b, we found that out of 5296 unique binding targets genes of the 67 chromatin factors, only 1295 unique targets were affected upon the deletion of the respective chromatin factor and a total of 3901 targets remained unaffected. The distribution of EBR and the pie-chart (figure 3.24 a-b) highlighted the widespread disconnect between binding and effect of chromatin factors in yeast. We also scrutinized the observed disconnect by calculating Pearson's correlation coefficients

between binding and effect. As shown in figure 3.25, most factors exhibit poor correlations between binding and effect, and the highest correlations were observed for Tup1 (0.14) and Ssn6 (0.13), respectively. In the subsequent analysis, we attempted to explore the determinants of disconnect in order to provide a plausible explanation for the same.

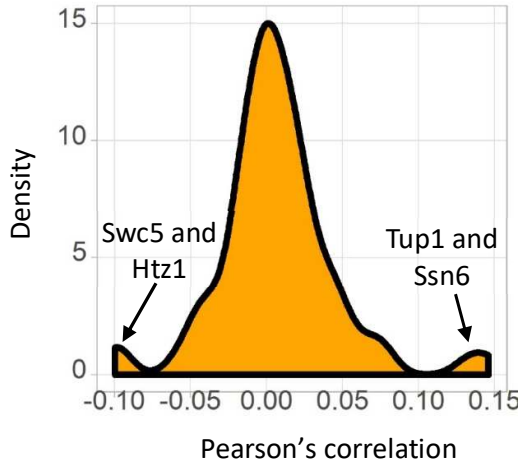

Figure 3.25 The distribution of Pearson's correlation coefficients between binding and effect of chromatin factors.

3.2.2 **The analysis of functional buffering**

The budding yeast has undergone two whole genome duplication and several small-scale duplications events during its evolutionary history (Fares et al. 2013; Kellis, Birren, and Lander 2004; Wolfe and Shields 1997). Therefore, it is expected that many of the chromatin factors might have paralogues genes in the yeast genome, which can buffer the functional impact of chromatin factor deletion. To explore this, we obtained the list of paralogues genes of each chromatin factor in the budding yeast genome using the BLASTp searches (Materials and methods). We reasoned that a chromatin factor having more number of paralogues genes would be more likely to be functionally buffered upon chromatin factor deletion as compared to the factor that had fewer or no paralogues genes. Therefore, a negative correlation between EBR and the number of paralogues genes was expected. But we observed a slight positive, albeit insignificant correlation between EBR and number of paralogues genes for chromatin factors (figure 3.26a).

Alternatively, we looked at the binding targets of chromatin factors that were simultaneously occupied by a chromatin factor as well as by its paralogues factor. We called this set of co-bound targets as the 'shared' set and the remaining targets as 'unshared' set. If the functional buffering was indeed playing a role in masking the effect of chromatin factor deletion, then the shared target genes would be expected to be affected less often as compared to the unshared target genes. However, we observed that the shared and the unshared targets had an insignificant difference in their EBRs (figure 3.26b). These two results ruled out the possibility that functional buffering by

paralogues genes can generate the widespread patterns of disconnect between chromatin factor binding and effect.

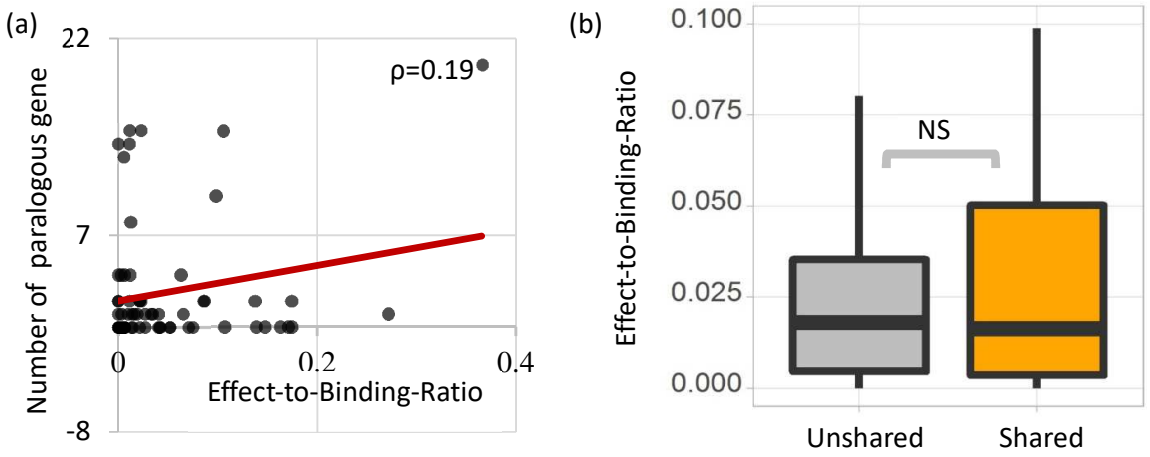

Figure 3.26 Association between disconnect and functional buffering.

a) Scatter plot between EBR and the number of paralogues genes of chromatin factors. ρ represents the Pearson's correlation coefficient. b) Boxplot distribution of EBR values for the 'shared' and 'unshared' binding targets of chromatin factors. The p-value calculated using the Wilcoxon's test was not significant (NS).

3.2.3 **Analysis of condition-specific roles of chromatin factors**

It is possible that the chromatin factors pre-occupy the genomic sites that are meant to be transcriptionally regulated during environmental stress. Therefore, the deletion of such factors under normal physiological conditions might not result in the transcriptional perturbation of its binding targets. To test this hypothesis, we calculated the mean and variance of expression change in the genome-wide transcriptional response to different environmental conditions in budding yeast (see table 7). If a chromatin factor with low EBR was having a condition specific role, then one would expect that its binding targets might exhibit greater stress responsiveness (i.e. higher mean or variance of expression across multiple environmental conditions) as compared to the factor that exhibited relatively higher EBR. Therefore, EBR was expected to be negatively correlated with the mean and the variance of expression of the binding targets across different environmental conditions. However, the EBR did not correlate negatively with the mean and variance of expression of the binding targets across the different environmental conditions (figure 3.27), implying that the chromatin factors with higher EBRs were as likely to exhibit the disconnect as were the chromatin factors with lower EBRs. Hence the widespread disconnect between binding and effect could not be ascribed to the condition-specific roles of the factors.

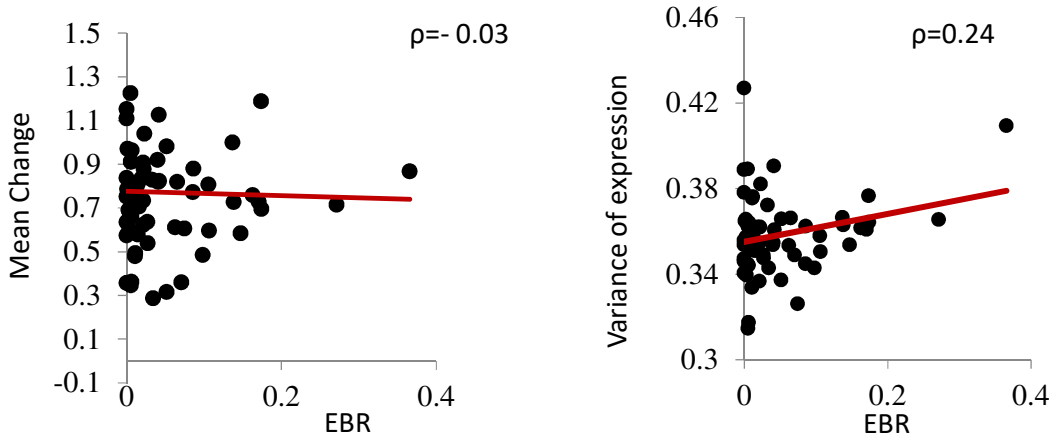

Figure 3.27 Stress responsiveness and the disconnect.

The scatter plot of EBR against mean change (left) and the variance of expression change (right) for the binding targets of chromatin factors under diverse environmental conditions. ρ is Pearson's correlation coefficient.

It should be noted that we did not conclude that the chromatin factors are not implicated in stress responsiveness. Instead, the variation in EBR could not be explained by the difference is the stress responsiveness of the binding targets of chromatin factors. In fact, previous studies have noted that genes which exhibit stress responsiveness and contain canonical TATA box in their promoters were more likely to be affected upon deletion of chromatin factors (Choi and Kim 2009). To cross-validate the previous studies, we compared the stress-responsiveness of affected and unaffected binding targets of chromatin factors.

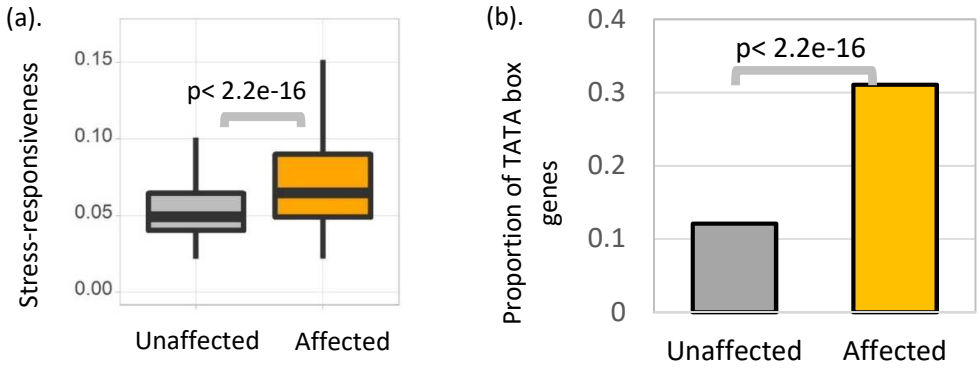

Figure 3.28 Stress responsiveness of affected and unaffected targets.

a) Boxplot for stress responsiveness. b) Barplots showing the proportion of TATA box containing promoters within affected genes as compared to unaffected genes. The p-values were calculated using Wilcoxon's test (a) and Fisher's exact test in (b).

The stress responsiveness values represented the variance in gene expression across multiple environmental conditions and were taken from a previous study (Choi and Kim 2009). Consistent with previous reports, we also observed that affected binding targets had greater stress responsiveness as compared to unaffected binding targets (figure 3.28a) and were significantly enriched for TATA-containing genes (figure 3.28b).

We further investigated the dynamics of 18 different chromatin modifications during the yeast stress response (see table 7). Interestingly, we found that both affected and unaffected binding targets exhibited a similar level of change for most chromatin modifications (figure 3.29a). The strongest change was observed for H3K4Me2 modification which had almost 10 percent change for the case of positively affected genes. We further confirmed that a significant majority of positively affected genes lost this modification as compared to the negatively affected genes during the yeast stress response (12% vs. 8%, Fisher's p–value=0.02).This decrease in H3K4me2 among the positively affected genes was consistent with its role in transcriptional repression in yeast (Kusch 2012; Margaritis et al. 2012)

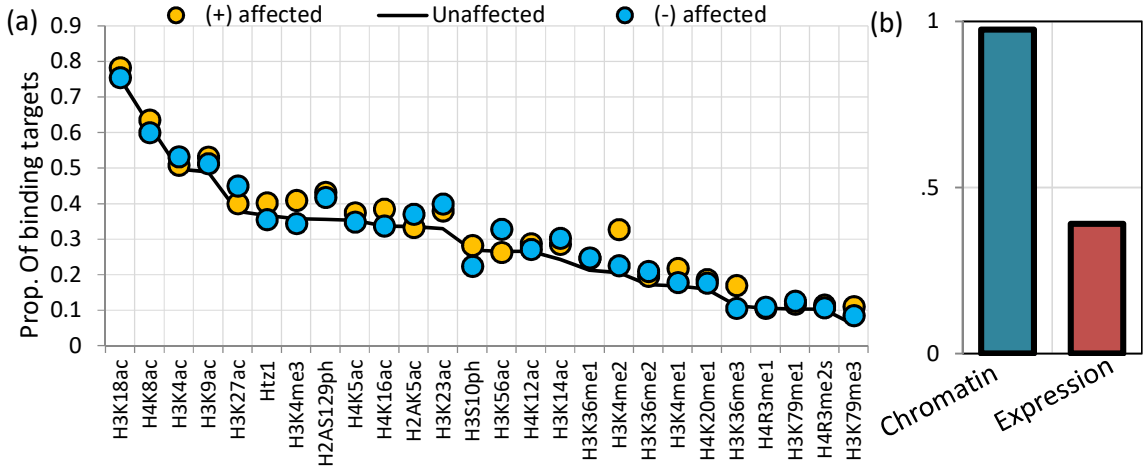

Figure 3.29 Chromatin modifications of binding targets during yeast stress response.

a) The x-axis shows the name of modification and Y-axis is the proportion of binding targets which exhibit the change of modifications for the case of unaffected (black line) and affected (blue and yellow dots) binding targets. b) The barplot showing the proportions of binding target either changing the chromatin modification level (blue) or expression level (red) during yeast stress response.

Additionally, we calculated the proportion of affected genes that underwent a significant change in at least one of the chromatin modifications and compared it with the proportion of affected genes which underwent significant change in expression in at least one of the environmental conditions. As shown in figure 3.29b, while ~98% of the genes exhibited a change in at least one of chromatin modifications, only 35% of the bound genes were significantly affected in at least

one of the alternative environmental conditions. These results indicated that most of the yeast genes indeed exhibited the changes in their chromatin modifications during the stress response, however, most changes did not translate into the expression effects (98% vs. 35%, figure 3.29b). Taken together, we concluded that wide-spread disconnect between binding and effect of chromatin factors cannot be ascribed to the condition-specific role of chromatin factors.

3.2.4 **Functional analysis of affected and unaffected binding targets**

As shown in the pie chart in figure 3.24 and, the deletion of 67 different chromatin factors resulted in the transcriptional perturbation of only a small number of the binding target genes. Secondly, the affected targets had greater stress responsiveness and were highly enriched for TATA box-containing genes. Therefore it was clear that a specific set of genes were more likely to be affected by the deletion of chromatin factors. This observation has also been previously emphasized (Lenstra and Holstege 2012). On similar lines, we also observed that the set of affected binding targets had a greater proportion of the yeast cell cycle-regulated genes as compared to the unaffected binding targets (figure 3.30).

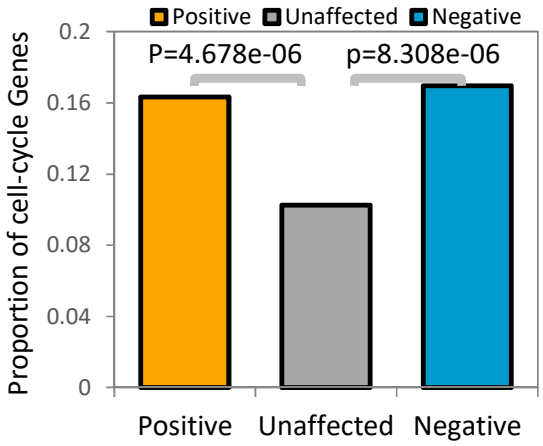

Figure 3.30 Enrichment of cell-cycle regulated genes among binding targets.

The barplot represents the proportion of affected (orange and blue bars) and unaffected (grey bar) genes that were identified to undergo cyclical regulation during cell cycle in Spellman's data. The p-values were calculated from Fisher's exact test.

Therefore, we tested if the chromatin factors exhibited the cell cycle specificity, which might explain their EBR. For this purpose, we downloaded and mapped the gene expression levels of each of the 67 chromatin factors from the yeast cell cycle data (table 7) and probed their expression pattern across the cell cycle. We observed that chromatin factors with and without effect had a similar pattern of expression across the yeast cell cycle (figure 3.31a) and has chromatin factors with affect had an insignificant increase (fold change of 1.04) in expression

towards the G2 phase. However, we observed that affected binding targets had a tendency to be enriched for G2 and M phase of cell cycle (figure 3.31b). On the other hand, unaffected targets showed a strong depletion towards G2 and M phase and were relatively enriched in towards G1 and S phase of the cell cycle (figure 3.31b). This observation highlighted an overall distinction between the cell cycle phase-specific functional roles of affected and unaffected binding targets, while the chromatin factors themselves were not cell-cycle phase-specific (figure 3.31a).

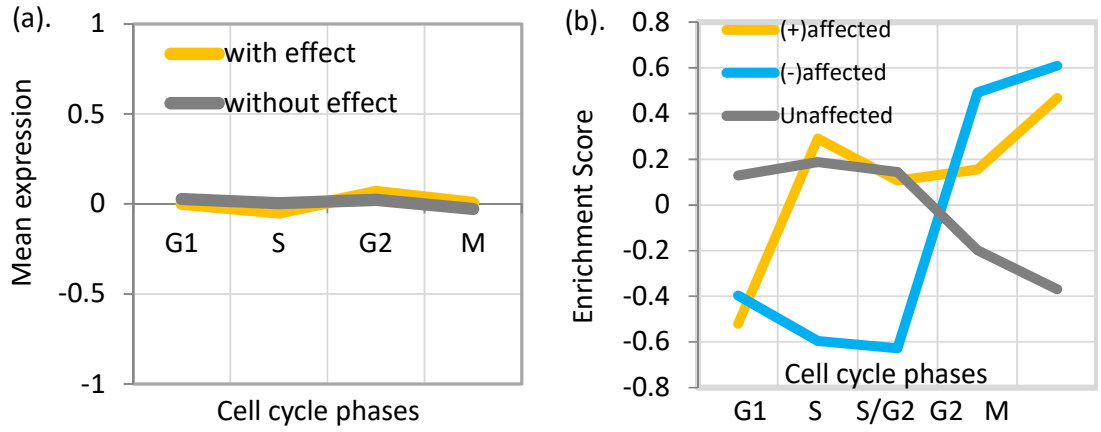

Figure 3.31 Analysis of cell cycle specificity of chromatin factors and their targets.

(a) Cell cycle expression pattern of chromatin factors with effect (yellow) and without effect (grey). (b) Proportion of affected and unaffected genes that were cell-cycle phase-specific. Enrichment score represents the log2 transformed observed-to-expected ratio of binding targets that were cell cycle phase specific.

To characterize the functional distinction between affected and unaffected binding targets, we performed their gene ontology analysis (Materials and methods). We observed that the affected binding target genes were significantly enriched for metabolism-related functions such as transmembrane transport, catabolism of sugars, and nucleotide biosynthesis (figure 3.32). Interestingly, metabolism is known to oscillate during the cell cycle and escalate around the G2/M phase (Papagiannakis et al. 2017; Z. Wang et al. 2014). This suggested that the chromatin remodeling at affected target genes might be required for the G2/M phase specific modulation of gene expression patterns. On the contrary, the unaffected binding targets had greater enrichment of G1-phase specific GO terms such as RNA processing, cytoskeleton organization, chromatin organization, etc., and were significantly depleted for G2 and M phase-specific functions (figure 3.32). Taken together, these results were suggestive of distinct functional roles of affected and unaffected binding targets.

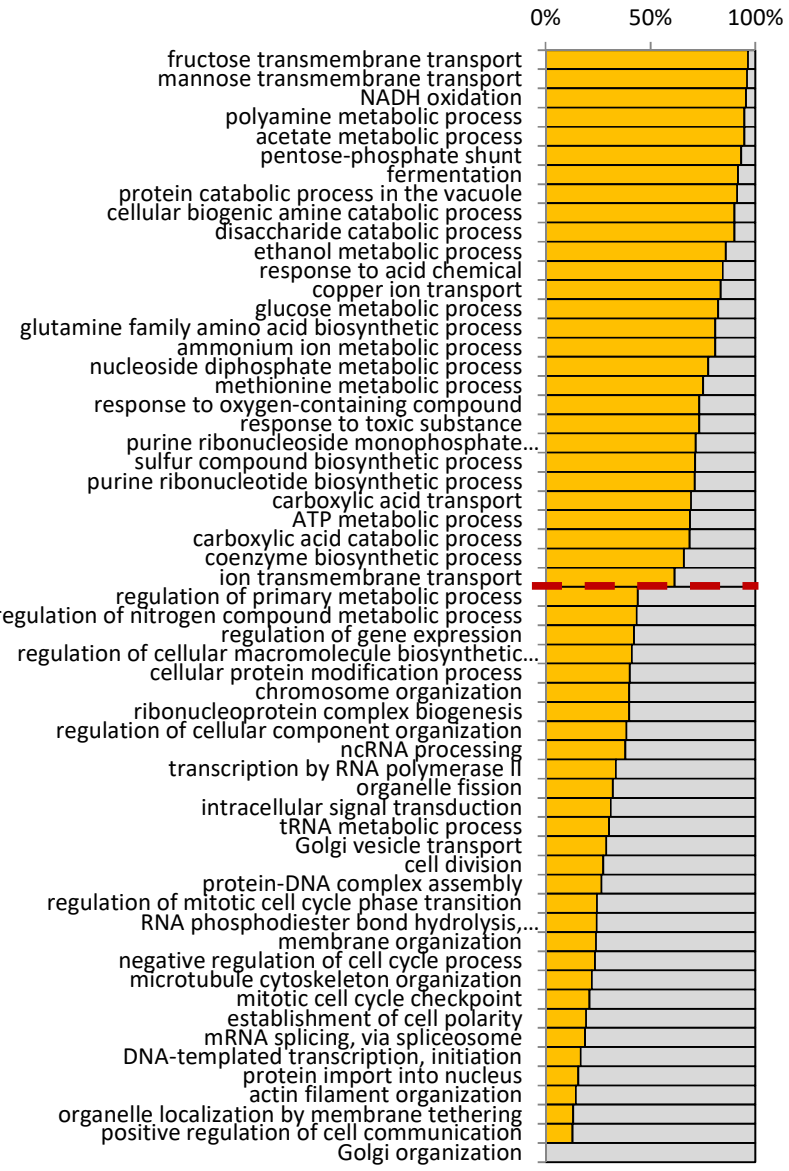

Figure 3.32 Gene ontology analysis of the binding targets.

The bars represent the relative proportion of a given gene ontology term (named along vertical axis) among affected (orange) and unaffected (grey) binding targets. The terms reported here had the enrichment p-value <0.05 after Bonferroni's correction. Terms above red dotted line are enriched among affected genes.

3.2.5 **Functional distinction of chromatin factors with high and low EBR**

Although the chromatin factors did not exhibit the cell cycle specificity in terms of their expression pattern, we sought to test if the chromatin factors with low and high EBR were functionally distinct by using different functional categories of chromatin factors as defined by Venters et al. (Venters et al. 2011). In this classification, the chromatin factors were divided into

four different categories namely; Access, Initiation, Elongation, and Orchestration depending upon their mechanistic roles in the process of chromatin remodeling and transcription. We observed that the chromatin factors related to initiation and access had a greater EBR as compared to the chromatin factors associated with the access and orchestration categories (figure 3.33), highlighting that the explanation to the variation in EBR is rooted in the mechanistic roles of chromatin factors.

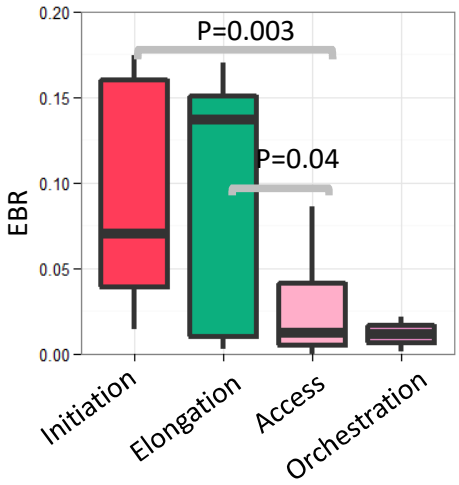

Figure 3.33 EBR of chromatin factors from Venters et.al's classification.

Boxplot distribution of chromatin factors performing different roles during the transcription cycle of the eukaryotic genes. The p-values are calculated using Wilcoxon's rank sum test.

To explore the observed distinction further, we further analyzed the protein interacting partners of chromatin factors. Although it is widely assumed that the genome-wide occupancy of chromatin modifiers at yeast promoters is involved in the transcription-related processes (Lenstra et al. 2011; Venters et al. 2011), the chromatins modifiers have a well-established roles in other DNA templated processes such as DNA repair, DNA replication, DNA recombination etc (Groth et al. 2007). Therefore, in addition to the transcriptional regulation, the genome-wide occupancy of chromatin factors might also be involved in the maintenance of genome integrity by mediating the processes such as DNA replication, DNA repair, DNA recombination, etc.

We obtained the protein-protein interaction (PPI) partners (see table 7 for reference to protein interactions dataset) of the chromatin factors and mapped the GO-slim terms related to transcription or the genome integrity (DNA repair, replication, recombination, DNA metabolism, and DNA damage response) onto the PPI partners of chromatin factors. We calculated the proportion of PPI partners belonging to each functional category considered in this analysis. We then performed the principal component analysis (PCA) of this data (Materials and methods) to assess the relative associations between different GO terms and EBR. Interestingly, we observed

that the genome integrity related terms, such as DNA replication, repair, etc. formed a distinct cluster away from EBR and transcription, which themselves clustered closely (figure 3.34a).

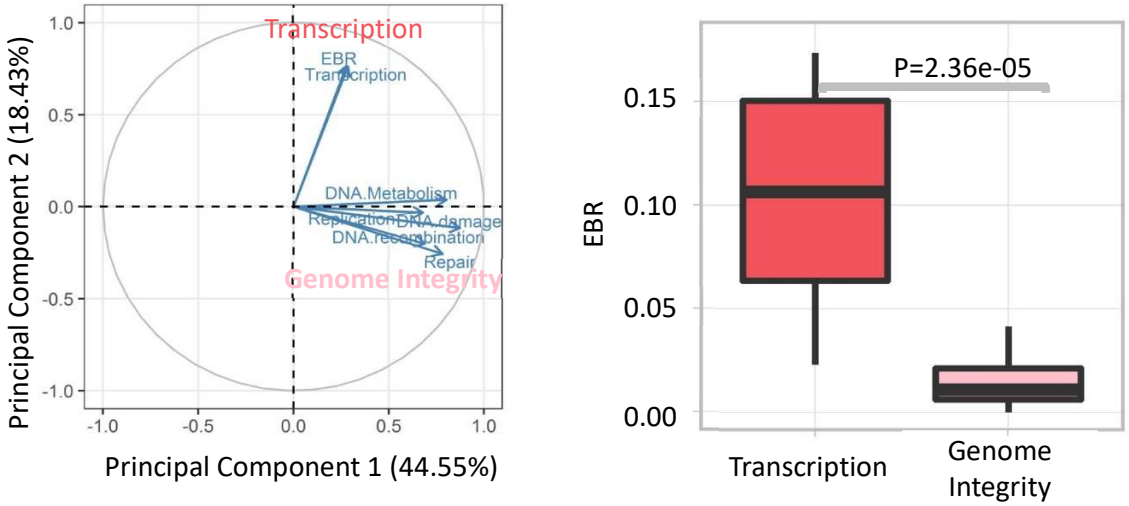

Figure 3.34 The GO analysis of protein interaction partners of chromatin factors.
The PCA factor map (left) shows the clear segregation of genome integrity related function from transcription and EBR. The arrows represent the loadings of GO terms and the percentage variance of the data explained by first two components is given alongside of axis labels. The boxplot distribution of EBR for chromatin factors related to transcription vs. genome integrity (right). The p-value was calculated from Wilcoxon's rank sum test.

We compared the EBR of chromatin factors belonging to either the 'transcription' or the 'genome-integrity' functions as defined based on PCA analysis presented above. We found that the genome-integrity related cluster had a significantly lesser EBR as compared to the transcription-related cluster (figure 3.34b). This result indicated that the chromatin factors with low and high EBRs might have distinct functional preferences. High-EBR factors were preferred for transcription regulation, while low-EBR factors were preferred for genome-integrity related functions in their respective protein interaction networks. Therefore, we speculated that the non-transcriptional roles of chromatin factors might underlie the observed discrepancy between binding and effect of chromatin factors. To test this hypothesis, we inspected the chromatin factors deposited at affected and unaffected binding targets as explained in the next section.

3.2.6 The distinction among chromatin factors deposited at affected and unaffected targets

We noted that the affected binding targets had a greater overall occupancy of chromatin factors as compared to unaffected binding targets (figure 3.35a). On somewhat similar lines, a previous study also showed that the genes having TATA-box in their promoter were regulated by a more number of transcription factors (L. Yang et al. 2014). Since affected target genes were enriched

in TATA-box genes (figure 3.28b), we also found significantly greater transcription factor occupancy at affected targets as compared to the unaffected targets (figure 3.35b). The observed differences in the chromatin factor occupancy at affected and unaffected binding targets highlighted the possibility that chromatin remodelers at these sites had distinct functional roles. To test this possibility, we investigated the GO terms of the chromatin and observed that chromatin factors deposited at unaffected binding targets had a significantly greater enrichment of DNA repair-related factors as compared to affected binding targets (figure 3.35c).

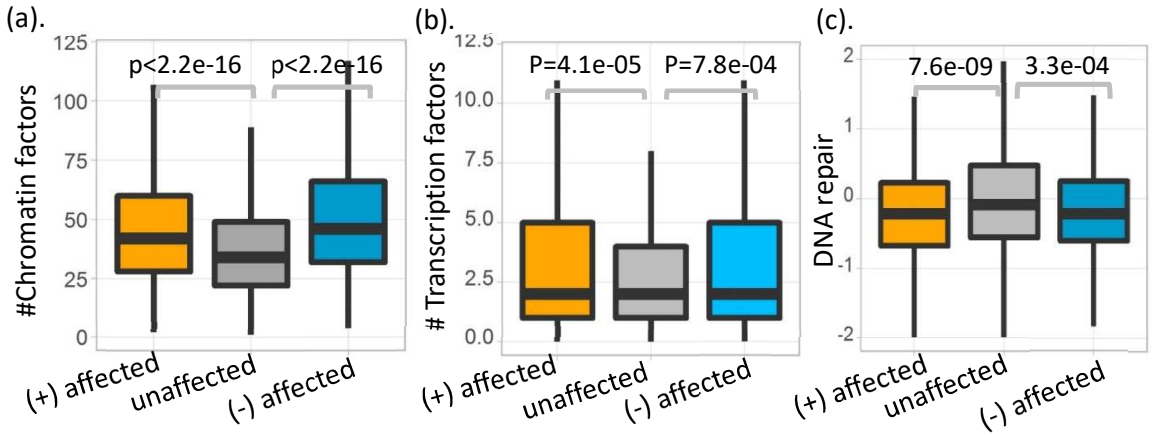

Figure 3.35 Distinction among factors deposited at positively-affected, negatively-affected and unaffected binding targets.

The boxplot distribution of total count of chromatin factors (a), transcription factors (b) and observed by expected number of chromatin factors associated with DNA repair at affected (orange and blue) and unaffected (grey) binding targets. The p-values were calculated using Wilcoxon's rank sum test.

This result not only validated the distinct functional preferences between chromatin factors with high and low EBR but also indicated that unaffected binding targets might be under greater genomic surveillance due to the enrichment of DNA repair-related factors (figure 3.35c). Consequently, unaffected binding targets were expected to bear fewer genetic or epigenetic errors as compared to affected binding targets as investigated in the succeeding text.

3.2.7 **Analysis of genetic errors at binding targets.**

To assess the occurrence of genetic errors at affected and unaffected genes, we obtained the data for single nucleotide mutations (SNMs) from yeast mutation accumulation lines, and DNA double-strand breaks (DSBs) (table 7 and appendix A). We mapped the occurrence of these errors on affected and unaffected binding targets. For SNMs, we calculated their cumulative occurrence near the transcription start site (TSS) of affected genes (Materials and methods) to plot the relative fold enrichment of SNMs at affected over unaffected targets (figure 3.36). For DSBs, we compared their distribution at affected and unaffected within their promoter regions (figure 3.36). Interestingly, we found that SNMs and DSBs were significantly enriched at the affected binding

targets as compared to the unaffected binding targets (figure 3.36 a-b).

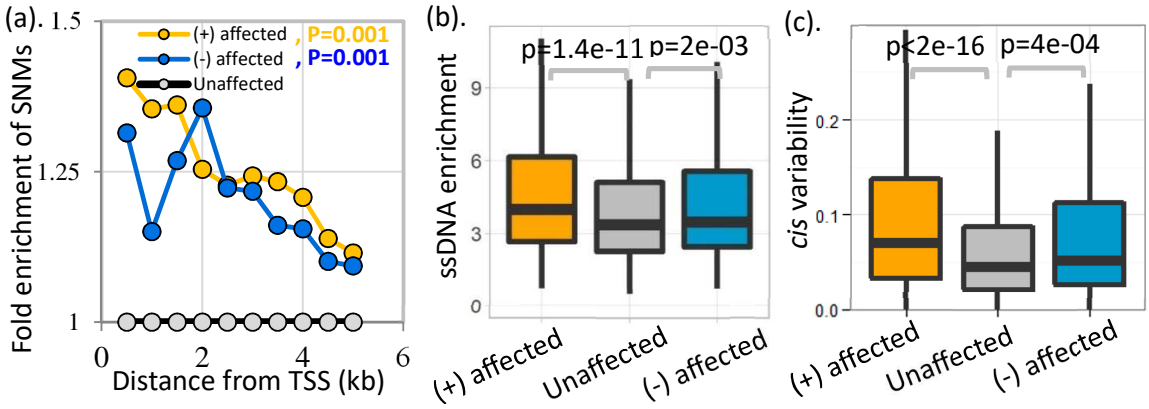

Figure 3.36 Analysis of the genetic errors at the binding targets.

a) Line plots of relative cumulative fold enrichment of SNMs around TSSs of affected (yellow and blue lines) targets over unaffected targets (black). The p-values is the minimum of the p-values obtained at all the distances from TSSs using the Wilcoxon's test. b) The boxplot distribution of DSBs at affected (yellow and blue) and unaffected (grey) binding target genes. c) The boxplot distribution of *cis* variability of the promoters of affected and unaffected binding targets. The p-values in b-c were calculated using Wilcoxon's test.

These results indicated that unaffected binding targets were protected against genetic errors. To further consolidate the relative lack of genetic errors at unaffected genes, we compared the *cis* variability of affected and unaffected binding targets of chromatin factors. *cis* variability indicates the inter-strain variability in gene expression which is attributed to variation in the promoter sequences of the genes between different strains (table 7 and appendix A). We found that affected binding targets exhibited the greater *cis* variability in their promoters as compared to unaffected binding targets (figure 3.36c). This result indicated that the sequences of promoters, and consequently, the expressions of downstream genes were more similar between different yeast strains for the case of unaffected binding targets as compared to that of affected binding targets. Collectively, these results implied that unaffected binding targets were indeed more robust against the accumulation of genetic errors and the divergence of their promoter sequences as compared to the affected binding targets.

3.2.8 **Analysis of cryptic transcription (epigenetic errors) at binding targets**

In addition to genetic errors, the genome integrity is also threatened by the cryptic unstable transcripts (CUTs, see table 7 and Appendix A), which are often produced from widespread bidirectional promoters in budding yeast (Neil et al. 2009). As discussed in chapter I, chromatin

modifications have important roles in the suppression of cryptic transcription initiation within the gene bodies during the elongation phase of RNA-PolII (Smolle and Workman 2013). Additionally, several ATP dependent chromatin remodelers have been implicated in the suppression of cryptic transcription near the promoter by their nucleosome remodeling activity (Hennig et al. 2012; Whitehouse et al. 2007). Therefore, we wondered if the unaffected binding targets were also protected against the cryptic transcription owing to the transcriptionally non-functional binding of chromatin factors near their promoters. To perform this analysis, we downloaded the genome-wide data of Cryptic Unstable Transcripts (CUTs) in budding yeast (table 7) and measured their occurrence within 500 base-pair upstream to each promoter. The enrichment of CUTs was further normalized by the total mRNA levels of the downstream gene (Materials and methods). As shown in figure 3.37a, we found that the unaffected binding targets were significantly depleted for the accumulation of CUTs as compared to the affected binding targets.

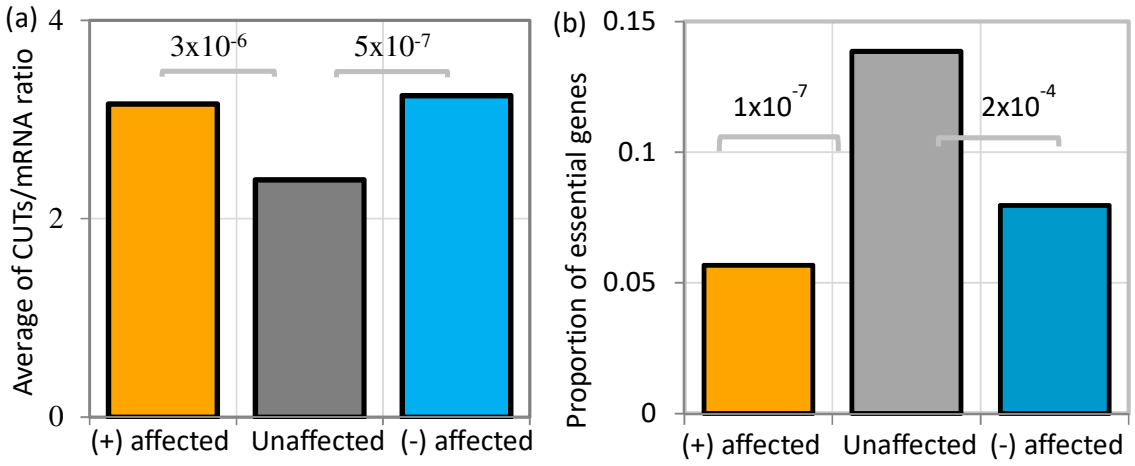

Figure 3.37 Analysis of cryptic transcription and gene essentiality at the binding targets.
(a). The barplots showing the level of cryptic transcription in 500 bp promoter upstream of TSS which was normalized by the mRNA level of the downstream gene. The p-values were calculated by using Wilcoxon's test. (b) The barplots for the proportion of essential genes among affected and unaffected binding targets. Yellow and blue bars are for affected (+vely and –vely, respectively) genes and grey bars are for unaffected genes. The p-values were calculated using Fisher's exact test.

This result indicated that in addition to the genetic errors, the unaffected binding targets were protected against the epigenetic errors too, possibly through the chromatin-modifying activities of the factors deposited at those promoters. Further, we found that the essential genes, which are known to be selected against genetic errors, were significantly over-represented among unaffected binding targets are compared to the affected binding targets (figure 3.37b). Collectively, these

observations suggested that chromatin factors might be preferably involved in the maintenance of the genome integrity at certain promoters, thereby causing a widespread disconnect between genome-wide location and transcription perturbation datasets of chromatin factors.

3.2.9 **Fitness defects of chromatin factor deletions**

It must be noted that all the chromatin factors considered in this study, despite being part of important macromolecular complexes, are non-essential for the survival of yeast. From the design point of view, this made us wonder why such non-essential factors exhibit genome-wide binding when their deletion neither lead to loss of survivability, nor the transcriptional perturbation of their bound target genes. A partial answer to this came when we investigated the fitness defect of these chromatin remodelers across >400 different environmental conditions (Materials and methods). As shown in figure 3.38a, we observed a strong negative correlation between mean fitness defect and the EBR of chromatin factors. A similar negative correlation of EBR was also observed with the number of negative genetic interactions of chromatin factors (figure 3.38b, see table 7 for source genetic interactions data).

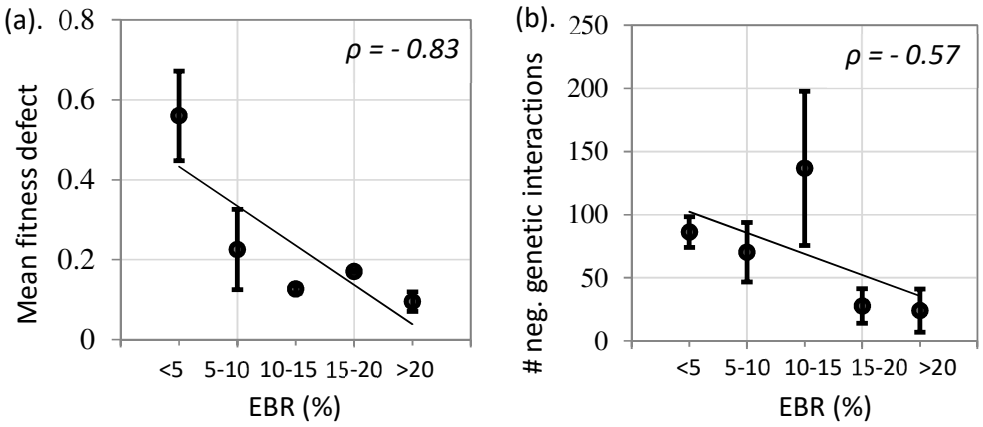

Figure 3.38 Fitness defects of chromatin factor deletions.

Scatterplots show the correlation of binned EBR (in percentage) against mean fitness defect (a), and number of negative genetic interactions (b). The vertical bars around each data point indicate the standard error bar. ρ is the Pearson's correlation coefficient.

These observations indicated that the chromatin factors with low EBR have a greater number of negative genetic interactions and exhibit a greater fitness defect when their deletion strains are grown under multiple drug conditions. Therefore, although the deletion of these chromatin factors is tolerated in the WT genetic and environmental backgrounds, these factors might adopt an important role in the survival of organism under genetically or environmentally perturbed conditions. Such roles might not necessarily involve the transcriptional regulation at first place, but can be also related to the surveillance of genome integrity as shown in our previous analysis

(figure 3.36, 3.37).

3.2.10 **Experiment to test the role of chromatin factors in genome integrity**

Next, we sought to perform a wet lab experiment to gain the evidence for the genome integrity related function of two of the chromatin factors; Dot1 and Swc5. We selected these two factors as they were associated with the 'genome-integrity' function in our PCA analysis (figure 3.34) and exhibited low EBRs despite their widespread binding. For each of these factors, we selected one promoter that exhibited high-affinity binding in Venters et al.'s (2011) ChIP-chip dataset (figure 3.39a). Additionally, we constrained the promoter choice based on the extremely mild or lack of transcriptional response upon the deletion of the respective chromatin factor in microarray dataset (Lenstra et al. 2011) (figure 3.39).

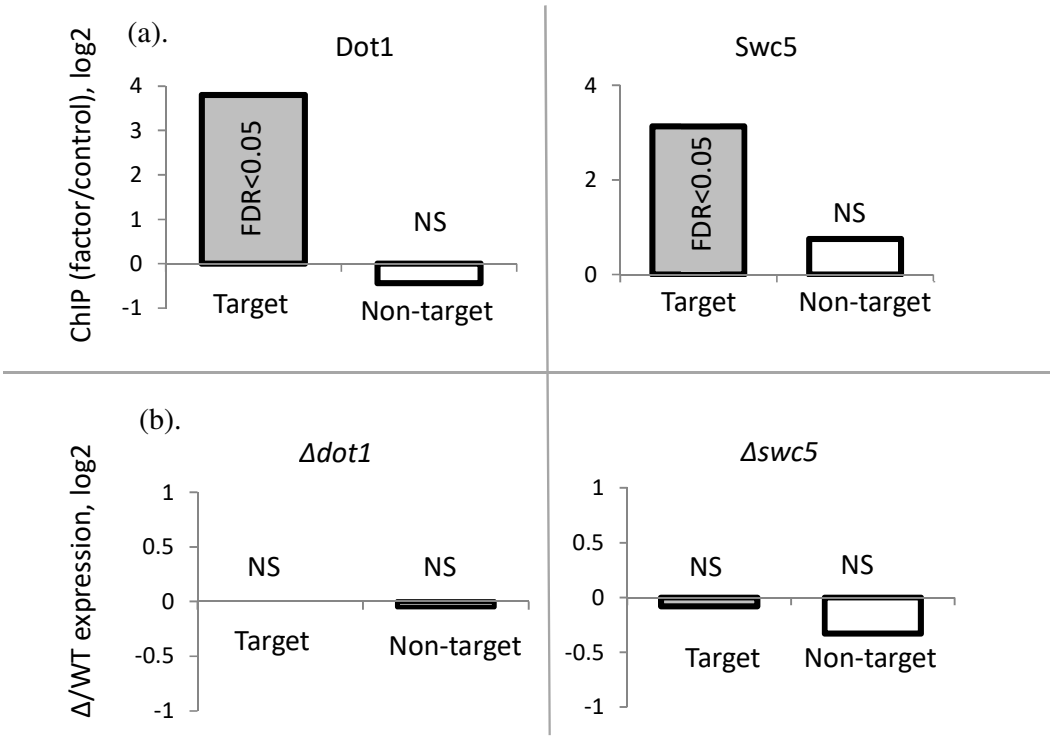

Figure 3.39 Selection of targets for FOA assay.

a) The binding values for Dot1 (left) and Swc5 (right) are indicated by grey bars for target and white bars for non-target loci. b) The choice of targets for FOA assay was also constrained by negligible transcriptional changes in the deletion strains of Dot1 (left) and Swc5 (right). NS indicates the non-significant binding or effect in each of the plots.

These two targets provided us with the promoters that exhibited highly significant binding of either Dot1 or Swc5, but were not transcriptionally affected upon the deletion of the respective factors. Our objective was to test whether the significantly bound but not affected targets showed a greater accumulation of mutations in the *Δdot1* or *Δswc5* strains as compared to the WT strains. To perform this test, we used the yeast strains deleted for either DOT1 or SWC5 and inserted the

URA3 reporter construct downstream of the selected binding site of either Dot1 or Swc5 by using homologous recombination in these strains (Materials and methods) and selected the recombinant yeast colonies (Materials and methods). We made similar recombinant strains in the wild-type (WT) backgrounds to compare it against the *Δdot1* or *Δswc5* recombinant strains. The transformation of URA3 made these recombinant strains uracil prototroph. If the binding of Dot1 or Swc5 was indeed important in keeping a check on genetic errors, the recombinant strains deleted for either the DOT1 or SWC5 should revert to the uracil auxotrophy more frequently as compared to the WT recombinant strains due to the accumulation of mutations in URA3 gene. Therefore, as compared to the transformants with WT backgrounds, transformants with deletion backgrounds (of either DOT1 or SWC5) were expected to show a greater number of colonies on FOA-containing culture plates as FOA selected against the uracil prototrophs (Materials and methods). Indeed, we found that the URA3 gene inserted near the binding target loci of Dot1 andSwc5 exhibited a greater frequency of mutagenesis in *Δdot1* and *Δswc5* strains as compared to the WT backgrounds (figure 3.40 a-b).

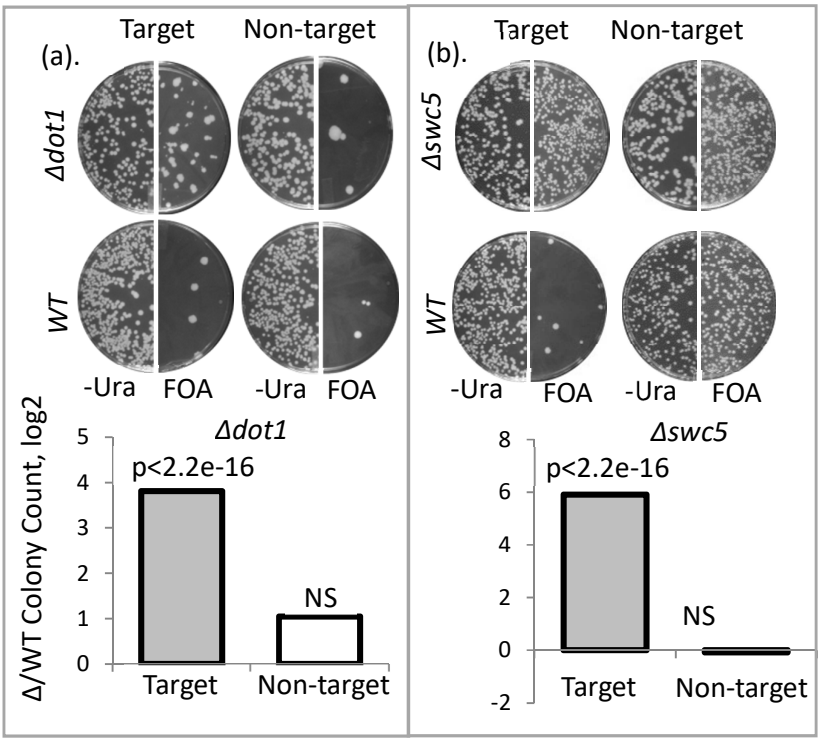

Figure 3.40 Results of FOA assay for Dot1 and Swc5.

(a) The comparison of colonies on FOA plates between WT and Δ*dot1* strains for the targets and non-target loci. Left half of the plate contains SC-Ura media and right half side contains SC+FOA media. The same comparison for Δ*Swc5* is presented in (b). The barplots indicate the log2 fold change in colony count between WT and deletion strains for target and non-target loci. The p-values are calculated using Fisher's exact test.

Similar experiments for the non-target sites of Dot1 and Swc5 did not exhibit a significant difference in the colony counts between WT and deletion strains (figure 3.40). These result indicated that the promoter-proximal binding sites of Dot1 and Swc5 indeed had greater genome integrity in the WT strains as compared to the *Δdot1* and *Δswc5*. Hence, these observations provided experimental evidence in the support of the role of non-transcriptional binding of chromatin factors in the maintenance of genome integrity in the budding yeast genome (figure 3.36).

The key message conveyed through these results is that genome-wide location analysis of chromatin factors should not always be interpreted to cipher or decipher the transcriptional regulatory code. Instead, a large proportion of this genome-wide binding might actually be involved in maintaining the genome integrity, which gives rise to the discrepancy between the genome-wide location and transcriptional perturbation data of chromatin factors.

Chapter 4 Discussion and conclusions

4.1 Functional and evolutionary constraints for 3D genome organization of the yeast

The non-random nature of the spatial genome organization is known to scientists for almost 3 decades. Several studies have shown the preferential locations of active and inactive genes in specific sub-nuclear compartments. In the post-genomic era, the innovations and the applications of 3C-based methods have fuelled an exponential growth of 3D genomics (de Wit and de Laat 2012). These methods have shown that the spatial co-localization of different genes exhibit a non-random pattern between different cell types and can be significantly remodeled during cellular differentiation (Dixon et al. 2015). But the underlying constraints of such gene-gene co-localization events remain poorly understood. The most common explanation proposed for such co-localization events is the co-regulation spatially proximal genes (Rajapakse et al. 2009; Szczepińska and Pawłowski 2013). Several reports have indicated the transcription to be the major driver for the spatial co-localization of subsets of genes associated with specific transcription factors or biological pathways (Fullwood, Liu, Pan, Liu, Han, et al. 2009; Schoenfelder et al. 2010; Di Stefano et al. 2013). In lower eukaryotes, few studies have attempted to use the genome-wide chromatin interactions datasets and proposed the co-expression of genes as a significant contributor for the spatial co-localization of genes *via* inter as well as intrachromosomal interactions (Dai and Dai 2012; Homouz and Kudlicki 2013). However, such studies were not comprehensive due to the lack of multiple functional variables incorporated in the analysis. Secondly, there had been reports that did not support the co-expression of genes as a major constraint of spatial co-localization of genes (Witten and Noble 2012). In this thesis, we performed a comprehensive analysis of multiple functional variables to identify the functional and evolutionary constraints that might have shaped the genome-wide interaction landscape of the yeast genome. The results obtained in this study suggested that yeast genome has preferred domains of *cis* and *trans* interactions, which exhibited different functional and evolutionary constraints. While the spatial coordination of early replication by the non-random spatial association of Clb5-independent (early) origins constrained the genome-wide landscape of *trans* interactions, the *cis* interactions were constrained by the spatial coordination of late replication through the non-random association of Clb5-dependent (late) replication origins. We found that the *cis* interactions were also constrained by the minimization of gene expression noise.

Importantly, our analysis revealed that the co-regulation of spatially proximal genes was not a significant constraint for the organization of *cis* or *trans* chromosomal interactions as the correlation between co-expression and interaction frequency was weak in magnitude. Such a weak correlation between spatial proximity and co-expression were also reported by previous studies and the correlation value of 0.09 was deemed significant by comparison the genome-wide average value of 0.07 (Homouz and Kudlicki 2013). We argued that comparison with genome-

wide average might have led to the erroneous conclusions because of the disproportionately larger size of all possible gene pairs across the yeast genome (Homouz and Kudlicki 2013). In present this analysis, we have used the bootstrap and controlled rewiring of chromatin interactions, which might serve as better control for assessing the significance of the correlations. The spatial co-localization of co-regulated genes has been questioned previously also (Witten and Noble 2012). More importantly, in the absence of other functional variables, it is not entirely justified to claim that co-expression is the major constraint of genome organization.

Our observations that *trans* and *cis* interactions were specifically associated with the spatial coordination of early and late replication supported the previous reports of spatial segregation of early and late replication factories (Dimitrova and Gilbert 2000; Göndör and Ohlsson 2009; Saner et al. 2013). Interestingly, recent research in budding yeast has shown that the replication origins, which were present in the same TAD, fired synchronously and assisted in the spatial coordination of DNA replication (Eser et al. 2017). Further, the investigations in mESCs have demonstrated that intact chromatin architecture is crucial for the faithful execution of the DNA replication program (Almeida et al. 2018). These two reports support our claim that replication is the predominant functional constraint underlying the 3D organization of the genome. In addition to the spatial segregation of early and late origins, we observed that the early origins had a narrow window of the nucleosome-depleted region, which was flanked by well-positioned nucleosomes on either side. Such an organization of nucleosomes was proposed to be important for the assembly of replication pre-initiation complex around the early origins of replication (Lipford and Bell 2001). In our study, we observed that nucleosome-depleted regions were wider for the case of late origins of replication. We proposed that wider window of nucleosome depletion might be involved in exposing the additional DNA sequence elements which might help to recruit the replication promoting factors because of their limiting availability during the late replication towards the end of S-phase. Alternatively, the late origins of replication were shown to frequently localize towards nuclear periphery (Patrick Heun et al. 2001) and the sequence elements exposed by wider nucleosome-depleted regions might help in the recruitment of late origins towards the nuclear periphery. Besides replication, we also found that the *cis* interaction frequency was strongly associated with the co-fitness of the spatially proximal. Since co-fitness is believed to indicate the functional similarity between genes (Hillenmeyer et al. 2008), the shared functionality might underlie the observed association of *cis* interaction frequency with co-fitness. The observation that the essential genes, which exhibit extreme fitness defect (lethality), were engaged in the *cis* interactions of higher average frequency led us to propose their spatially restrained state. Since essential genes exhibit lower gene expression noise, the spatially restrained state of essential genes due to the *cis* interactions of higher frequency indeed might be important

in minimization of their gene expression noise. Since we only considered the *cis* interactions which were at least 20kb apart on the linear genome, our observations were not impacted by the linear clustering of essential genes along the genome. The spatially restrained state of chromatin by its engagement in *cis* interactions of greater frequency was also confirmed for the genes exhibiting lower expression noise in *E.coli*, Yeast, and mESCs. The reciprocal relationship between long-range chromatin interaction and gene expression noise was proposed in a previous study also (McCullagh et al. 2010).

The mechanism underlying the reduction in chromatin mobility might be complex and could involve the role of multiple molecular players. It is possible that spatial tethering of chromatin with sub-nuclear can reduce chromatin mobility. A previous study, using ChIA-PET, discovered several multigene complexes and were proposed to resemble transcription factories (Li et al. 2012). The genes engaged in the formation of multi-gene complexes might indicate the spatially restrained chromatin. Indeed, we found that the genes having lower expression noise had a significantly greater number of RNA-polII associated 3D interactions in mESCs. The higher number of RNA-polII associated interactions would indicate the stable co-localization of different genes at the transcription factories and will be captured as interactions of higher frequency in 3C-based experiments. Since there is only a limited number of transcription factories present in a given cell (varies from 100-8000 as per Eskiw and Mitchell 2016), the stable co-localization of millions of gene-pairs near the transcription factories cannot imply their co-expression as the number of co-expressing gene-pairs are much lesser. In fact, a recent study has shown that the co-expression of spatially proximal genes could be buffering at the level of proteins (Kustatscher, Grabowski, and Rappsilber 2017), indicating that spatial proximity of genes might not be constrained by the strict requirement for their co-expression. We proposed that stable localization of genes near transcription factories can constrain their chromatin mobility as compared to the genes which are not tethered (i.e. the high noise genes) with the sub-nuclear structures. This proposal was also supported by our observation that the shared interactions between different HiC libraries tend to have higher average *cis* interaction frequency and were significantly enriched for regions having low expression noise. Therefore, the *cis* interactions originating from tethering of different genes with RNA-PolII might not necessarily reflect the co-expression of genes but could be important in minimization of gene expression noise. These observations reinforce our claim that co-expression of the spatially proximal genes might not be a potent functional constraint which might govern the spatial contacts between distant loci.

4.2 Analysis of genome-wide disconnect of chromatin factors in yeast

In this 2nd part of the thesis, we have performed the comprehensive analysis in order to understand the transcriptionally non-functional binding of chromatin factors in yeast and provide a simple

explanation to the genome-wide discrepancy between location and effect of chromatin factors. The role of chromatin in transcription is unquestionable (B. Li, Carey, and Workman 2007) and several studies have attempted to decipher the logic behind chromatin remodeling by using the genome-wide location analysis of chromatin modifiers and modifications (Kurdistani et al. 2002; Ram et al. 2011; Robyr et al. 2002; Venters et al. 2011). The findings of such studies were supplemented by the genome-wide transcriptional response analysis following the deletion of chromatin modifiers (Lenstra et al. 2011). However, stark discrepancy between genome-wide location and effect of chromatin modifiers have always puzzled the scientific community (Lenstra and Holstege 2012). Interestingly, such a discrepancy has also been observed for transcription factors in the case of yeast as well as mammalian systems (Cusanovich et al. 2014; Spivakov 2014; W.-S. Wu and Lai 2015). In the past, a few studies have attempted to explain the lack of transcriptional regulation at the binding locations of these factors. The most common explanation offered is that deletion of transcription factors is buffered by the presence of paralogues genes in case of yeast as well as human (Gitter et al. 2009; W.-S. Wu and Lai 2015). However, the widespread disconnect between chromatin factors was not comprehensively analyzed to assess the contribution of functional buffering for lack of transcriptional effects upon their deletion. Through analysis of chromatin factors as well as their binding targets, we did not find the substantial role of functional buffering by paralogues genes in masking the effects of chromatin factor deletion in yeast. From the evolutionary point of view, our observation was consistent with the idea of 'evolution by gene duplication' which asserted that duplicated genes, because of relaxed selective constraints, might undergo greater expression divergence and are more likely to undergo neo-functionalization (Khaladkar and Hannenhalli 2012; J. Zhang 2003). In fact, a recent study also reported the similar findings for the case of transcription factor binding data in the human genome where widespread disconnect could not be ascribed to the functional buffering by paralogues genes (Cusanovich et al. 2014).

We also tested if the wide-spread disconnect between chromatin factor binding and effect represent the condition-specific roles of chromatin factors through the pre-binding of chromatin regulators at the genomic sites which undergo expression changes under stressful conditions. Such an explanation was previously found to partially account for the discrepancy between binding and effect of transcription factors (Spivakov 2014). But our observation that variance in EBR cannot be ascribed to the stress responsiveness of their binding targets ruled out this possibility. Additionally, we observed that most of the changes in chromatin modifications did not translate into expression effects (98% vs. 35%) during the yeast stress response. Therefore, these results provided a hint for the non-transcriptional roles of these chromatin modifications and the chromatin modifiers involved in their deposition.

The PCA analysis of GO terms mapped to the PPI partners of chromatin factors indicated the clear segregation between transcription and genome integrity related functions of chromatin factors and showed that genome integrity related chromatin factors were more likely to exhibit the transcriptionally nonfunctional binding. This result is very interesting as it indicated that the occupancy of certain chromatin factors' at genic promoters is more important in the maintenance of genome integrity rather than the transcription *per se*. The genome integrity is constantly under threat by constant exposure to external or internal DNA damaging agents. The opening of chromatin during the process of transcription itself imposes a greater risk for the genome integrity (Gaillard and Aguilera 2016; Svejstrup 2010) and is known to associate with greater accumulation of DNA breakage, recombination and elevated levels of mutagenesis (Choi and Kim 2009; Park, Qian, and Zhang 2012; Schwer et al. 2016). Therefore, genome needs robust mechanisms to repair itself against these genetic errors. The role of chromatin modifiers in mediating the DNA damage response is well established and provides the access to underlying DNA bases by nucleosome remodeling via ATP dependent nucleosome eviction, displacement or chemical modifications of histones (House, Koch, and Freudenreich 2014; Lans, Marteijn, and Vermeulen 2012). Therefore, deposition of chromatin remodeling enzymes at important genomic elements such as promoters might be a mechanism to ensure the genome integrity without any delay in the process of DNA repair. Indeed we observed evidence for the greater genome integrity at the unaffected binding targets, which had a lesser accumulation of SNMs and DSBs as compared to affected binding targets. Further, our analysis of *cis* variability, and inter-strain as well as inter-species variation of yeast genes *(data not shown)* also suggested the protected nature of unaffected binding targets. To supplement these observations, our FOA assay demonstrated a significantly greater frequency of spontaneous mutagenesis in the URA3 gene inserted downstream to the binding sites of Dot1 or Swc5 in their deletion strains. Hence the chromatin factor binding indeed appears to be associated with the maintenance of genome integrity at promoter-proximal sites.

Additionally, the process of transcription is also known to cause wide-spread cryptic transcription (Neil et al. 2009; Z. Xu et al. 2009) which are quickly degraded by tightly regulated pathways and may pose a threat to genome integrity if left unchecked. Some studies have also shown the nucleosome remodeling associated with ATP dependent chromatin remodelers to be involved in the suppression of cryptic transcription across the yeast genes (Hennig et al. 2012; Whitehouse et al. 2007; Yadon et al. 2010). Interestingly we found that similar to SNMs and DSBs, the unaffected targets were protected against cryptic unstable transcripts also, which indicate the role of non-transcriptional binding in mitigation of genetic as well as epigenetic errors.

Taken together, these results demonstrated that the chromatin factors have non-transcriptional roles at promoter-proximal sites and the widespread disconnect between binding and effect of

chromatin factors can be explained by considering their role in the maintenance of genome integrity. It is worth mentioning that these results do not imply the mutual exclusiveness of the transcription and genome integrity related functions by chromatin factors. Instead, we propose that the causal role of chromatin factors in transcriptional regulation should be dealt with caution. Wide-spread occupancy of certain regulators might be more important in the maintenance of genome integrity than the transcription.

4.3 Conclusions

In project I, we concluded that the 3D organization of yeast genome is shaped by the spatial assortment of co-replicating and co-evolving genomic domains through interactions among replication origins. The short-range intra-chromosomal clustering is constrained by the transcription minimization of the gene expression noise, an observation that is evolutionary conserved from prokaryotes to higher eukaryotes. Importantly, our comprehensive analysis deflated the popular claim that the co-expression of spatially proximal genes is the major functional constraint shaping the 3D organization of yeast genome. Some of the key findings are summarized in figure 4.1

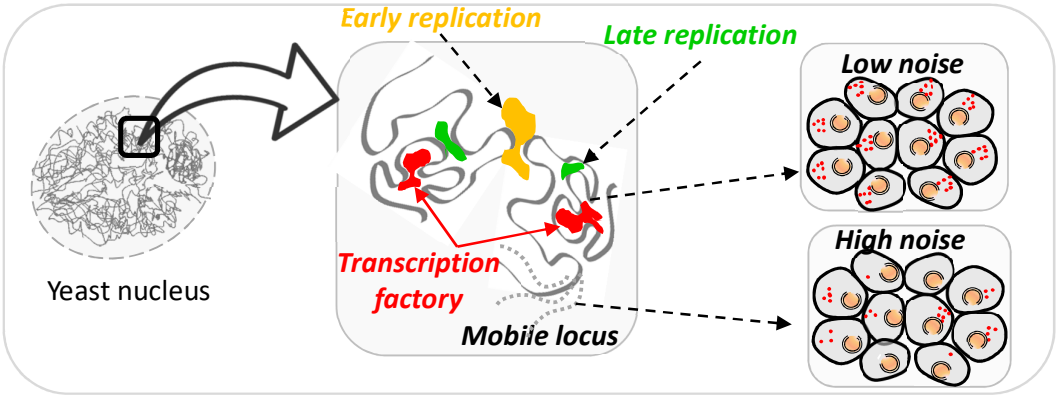

Figure 4.1 A model for functional constraints underlying the 3D organization of yeast genome. The large scale organization of *cis* and *trans* interactions has distinct functional constraints. *trans* interactions are constrained by the coordination of early replication through their association with early replication factories (yellow objects). The *cis* interactions are constrained by the coordination of late replication through their association with late replication factories (green objects) and minimization of expression noise *via* spatial tethering of chromatin with transcription factories.

In project II, we established that most chromatin regulators do not occupy their target promoters to regulate transcription, but to surveil the genome against genetic and epigenetic errors (figure 4.2), presenting a simple explanation to the global discrepancy between binding and effect of

chroamtin regulators. Our findings imply that the causal role of chromatin factors' occupancy in transcriptional regulation should be dealt with caution and support the much wider roles of chromatin factors in processes other than transcription.

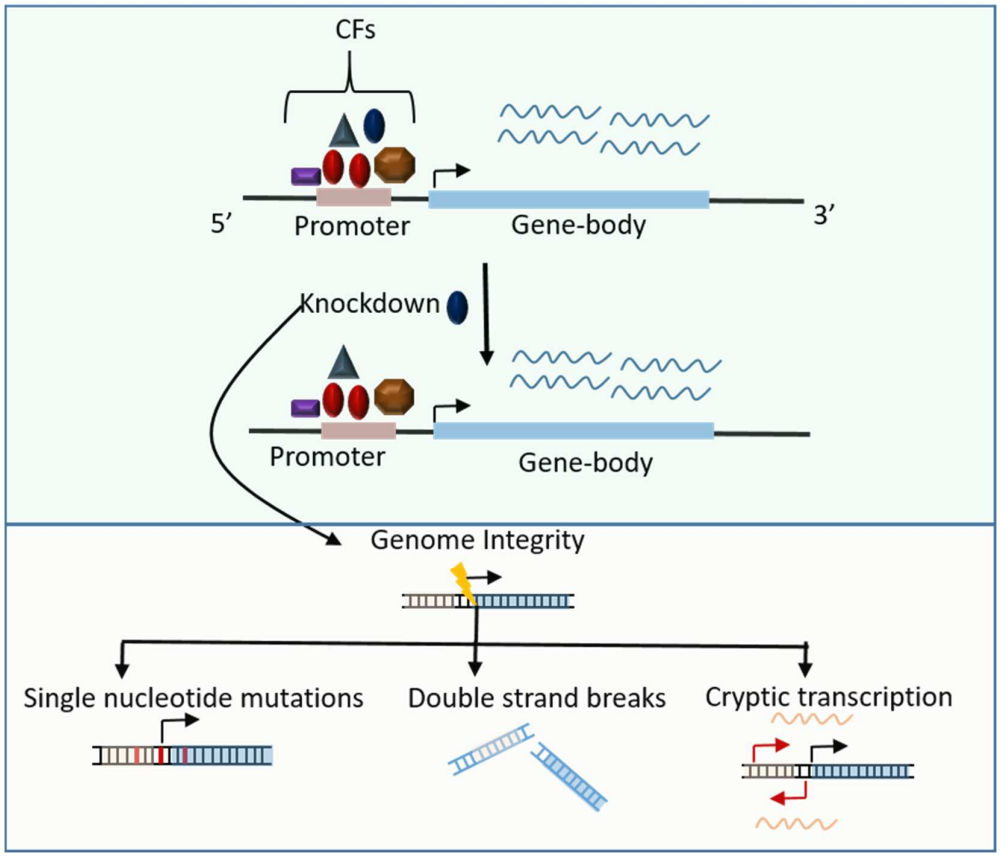

Figure 4.2 A model explaining the genome-wide disconnect of chromatin factors.
Deletions of certain chromatin factors do not affect the transcription at certain genomic loci, and instead, could be involved in the prevention of errors such as SNMs (red vertical lines), DSBs, and cryptic transcription.

Appendix A

A description of the datasets used in this study is given below.

Chromatin interaction data of yeast. The chromatin interaction data used in the analysis of functional and evolutionary constraints underlying the 3D organization of the yeast genome was taken from a previous study (Z. Duan et al. 2010). This dataset comprised the pairwise measurements of interaction frequencies between different restriction fragments of yeast genome generated by either HindIII or EcoRI restriction endonucleases and was downloaded from the supplementary link provided by the authors. A total of 240629 inter- and 65683 of intra-chromosomal interactions at FDR of 1% were obtained for the HindIII library. Prior to the analysis, we processed this data using HiCNorm (see chapter II) for removing the various systematic biases.

Cell cycle time course data. To calculate the genome-wide co-expression between yeast genes, we downloaded the data for the time course of mRNA measurement spanning through the different stages of cell cycle in yeast (Spellman et al. 1998). This data was generated by first synchronizing the yeast culture by one of the three methods namely; α-factor deletion, cdc5-deletion or elutriation and then performing the DNA microarray experiment at various time points corresponding to different phases of cell cycle. Since the different methods of synchronization gave identical results as reported by Spellman et al., we arbitrarily selected the data generated by α-factor deletion which represented the time course expression pattern of 5754 yeast genes at 16 different time points. In addition to the analysis of chromatin interactions, we used this dataset for the analysis of expression patterns of chromatin factors and their binding targets across the cell cycle (Project II).

Replication time course data. The replication of the eukaryotic genome follows the well-defined temporal order with certain genes undergoing early replication, while others undergoing late replication (W. Burke and Fangman 1975). To understand the relation between 3D genome organization and replication timing, we downloaded the replication time course data for yeast genome from a previous study (Raghuraman et al. 2001). This data was generated by oligonucleotide microarray hybridization to measure the proportion of replicated DNA for each of the 500 bp windows across the entire yeast genome (23964 regions) at eight different time points. The multiple time points in this data enabled us to calculate the co-replication index (in terms of PCCs between different windows) and perform the correlation analysis of replication timing with the spatial co-localization frequencies in HiC data.

Transcription factor perturbation data. Many of the transcription factor target genes are known to present in linear gene clusters as well as on specific chromosomes. In addition to linear clustering, a study by Dai and Dai in 2012 proposed that a number of transcription factors tend to

regulate the targets that are co-localized in nuclear space (Dai and Dai 2012). Further, they show that co-localized transcription factor targets show more co-regulation as compared to other transcription factor target genes. Therefore, we incorporated the data of transcription factor targets for the correlation analysis of chromatin interactions. We downloaded this data from SGD and consisted of genome-wide (4792 genes) microarray experiment for the mRNA measurement after the individual deletion of 263 different transcription factors (Z. Hu, Killion, and Iyer 2007). We calculated the PCC between gene pairs from this study to see if there is any association of transcription factor perturbation and 3D genome organization.

Chromatin factor perturbation data. Similar to transcription factors, chromatin factors are the proteins which target specific genes to carry out various DNA template processes such as transcription, replication and DNA repair, etc. To assess whether the spatially co-localized genes show similar response to the deletion of the chromatin factors we obtained the dataset generated by Lenstra et al. (Lenstra et al. 2011) which consisted of the genome-wide (6123 genes) microarray experiments for mRNA measurement after the individual deletions of 165 different chromatin and gene regulatory proteins. We used this data to calculate the PCC between gene expression profiles and incorporated it into the correlation analysis of chromatin interactions. We also used this dataset for the analysis of discrepancy between chromatin factor binding and effect where we inferred the significantly affected genes upon the deletion of each chromatin factor from this data (see chapter II).

Histone acetylation and methylation data. To assess the similarity in epigenetic states of spatially colocalized genes, we used the genome-wide profiles of several different histone methylations (Pokholok et al. 2005), and acetylations (Kurdistani, Tavazoie, and Grunstein 2004). Both these studies used the original or the modified form of ChIP-chip technique, which was used for yeast by Bernstein et.al in 2002, to measure the level of 16 different acetylations for 4993 genes and 5 different methylations for 5383 genes. To check the association of these epigenetic modifications with 3D genome in yeast, we calculated the PCC across all acetylation and methylations across yeast genome in for all available gene pairs in these datasets and incorporated it into the correlation analysis.

Functional similarity between genes. The previous reports of association of 3D genome with co-expression and transcription factor targets prompted us to look into the broader aspect of co-regulation i.e. functional similarity. Therefore, we incorporated the information of functional similarity between spatially co-localized genes for the analysis of *cis* and *trans* interactions. The functional similarity dataset used in this study was generated by calculating the semantic similarity between all vs. all pairs of gene pairs by mapping the GO terms (see chapter II). A

higher value of semantic similarity indicated the greater functional similarity between a given gene pair.

Co-fitness of genes. In addition to co-expression and functional similarity between genes, Hillenmeyer et., in the year 2008, defined co-fitness of genes which reflected the novel functional relationships among genes which were not evident in other metrics such as protein-protein interaction or genetic interaction (Hillenmeyer et al. 2008). Therefore we included the co-fitness of spatially interacting genes for the correlation analysis of chromatin interactions. Co-fitness was defined as the PCCs between the fitness values of heterozygous or homozygous deletion strains of yeast when these strains were grown in >400 different environmental backgrounds in a chemical genomics screen. Fitness data consisted of the fitness measurements of when 4,769 pooled homozygous *S. cerevisiae* single deletion strains were cultured in 418 different compounds in a chemical genomics screen. Co-fitness values for 8935890 gene pairs were derived by calculating the correlation between fitness values of genes across conditions. We also used this data to calculate the mean fitness defect of each chromatin factor when we investigated the discrepancy between chromatin factor binding and effect.

Protein-protein interaction (PPI) data. The Protein-protein interactions considered in this study were taken from biogrid *(https://thebiogrid.org/),* which is assembled and curated from different high throughput studies. The PPI data consisted of 50000 interactions and each interaction was weighted by socio-affinity (SA), which indicated the strength of a given interaction. Interactions through indirect evidence like synthetic lethality, synthetic growth defect, synthetic rescue, phenotypic enhancement, epistatic miniarray profile, dosage rescue, phenotypic rescue, colocalization, predicted through insilco approach were not considered. The SA scores were further scaled by a log likelihood score of a dataset against a literature-curated benchmarked dataset. We used this data to test if the products of spatially proximal genes engage in protein-protein interactions. We also used this data to infer the functions of chromatin factors when we analyzed the genome-wide discrepancy between chromatin factor binding and effect.

Evolutionary rate co-variation (ERC). The genes sharing functional similarities and those involved in protein-protein interactions are often constrained by similar selection pressure. Consequently, their rates of sequence divergence over the course of evolution appear to be correlated with each other due to parallel evolution. To check if the spatial proximity of different genes is constrained by the similar evolutionary rates of different genes, we used the dataset from a previous study (Clark, Alani, and Aquadro 2012) which has calculated the genome-wide ERC among 8.4 million pairs of proteins. ERC was calculated using 4459 yeast proteins across 18 different species. Each proteins phylogenetic branch length was calculated over 18 species, which was then transformed into a relative deviation from the expected length in the average proteome-

wide tree. 8.4 million Possible branch lengths were obtained and a coefficient of correlation was calculated for each pair of proteins.

Origins of replication. The previous study by Duan et.al has proposed the early origins of replication are non-randomly co-localized via chromatin interactions. By using the improved resampling-based approach (Witten and Noble 2012) as compared to hyper-geometric methods used by Duan et.al (Z. Duan et al. 2010), we decided to revisit the claims for the association of early and late origins of replication with 3D genome in yeast. For this, we downloaded the data for early and late origins of replication which were identified and characterized by McCune et.al based on their dependence on B-type Cyclin Clb5 (McCune et al. 2008). This data consisted of the genomic coordinates of 122 Clb5 dependent (late) and 78 Clb5 independent (early) origins of replication were obtained from McCune et al 2008.

Nucleosome Occupancy data. For testing the difference in the nucleosome occupancy of genomic regions enriched in *cis* or *trans* chromatin interactions, we downloaded the nucleosome occupancy data for G1 phase from a previous study (Soriano et al. 2014). This data was generated MNase-Seq technique which is based on the preferential digestion of nucleosome-free DNA with nucleosome DNA being protected from the digestion Micrococcal Nuclease. The mononucleosome DNA generated in this process was sequenced to infer the genome-wide nucleosome occupancy patterns at base pair resolution.

Transcriptional noise. The term transcriptional noise refers to the stochastic fluctuations in gene expression in time and space. To test the association between gene expression noise and 3D organization of 3D genome, we obtained the single cell protein abundance measurement data from a 2006 study (Newman et al. 2006). From this data, we calculated the coefficient of variation (standard deviation/mean) as a measure of noise. This data was then transformed into abundance-corrected noise. We also used the gene expression noise data for E.Coli (Silander et al. 2012) and mESCs (Tang et al. 2010)to test the evolutionary conservation of relationship between expression noise and 3D genome organization of these species.

Essential genes. Essential genes refer to the minimal set of genes which is necessary to ensure the survival of the organism under the optimal conditions. Yeast genome has more than 6000 annotated genes, but only ~10 % of all the genes are known to be critical for the survival of yeast under optimal conditions. The deletion of these genes results in the lethality (R. Zhang 2004). These genes were earlier identified from the gene deletion phenotypes generated in Saccharomyces Genome Deletion Project. These studies used a PCR based deletion strategy to knock out almost every single gene of yeast genome and analyzing their effect on phenotypes (Winzeler et al. 1999). To test the association of 3D chromatin organization with gene essentiality, we obtained a list of 694 genes which were annotated as essential in yeast. This list of essential

genes for the E.coli genome was taken from *https://shigen.nig.ac.jp/ecoli/pec/about.jsp.* Additionally, we also used this list of essential genes for yeast when we analyzed the difference in the genome integrity at affected and unaffected binding targets of chromatin factors in project II.

Overexpression toxicity. The protein products of some genes can cause toxicity to the cell if their amount is abruptly increased. This effect is known as overexpression toxicity. Such toxic genes were found to exhibit lower expression noise. To test if the minimization of noise for the case of toxic genes can be attributed to their chromatin interaction profile, we downloaded the data for >1300 genes for which toxicity status was studied by using their overexpression strains by a previous study (Yoshikawa et al. 2011).

Chromatin mobility data. To confirm the association between *cis* interaction frequency and minimization of chromatin mobility, and consequently of gene expression noise, we used the chromatin mobility values for the specifically tagged yeast loci. These values were measured either for 15 different equally spaced fluorescent probes within chromosome 12 by using high throughput live cell imaging (Albert et al. 2013), or for 4 different locations across the chromosome 5 and 12, which included the two centromeres and two ARSs, by LacZ based reporter (P. Heun et al. 2001).

Mutation rates of yeast. To estimate the mutation rates of genes or genomes, several methods have used the multiple sequence comparison to calculate the sequence divergence (G. M. Cooper et al. 2003). As the mutations acquired by the ancestors are subjected to selection pressure, the actual number of mutational events will be more than the calculations derived from the observed rate of divergence of a sequence, because of strong selection against non-synonymous or deleterious mutations. For these reasons, the mutational events at the synonymous sites are used to estimate the mutation rates. But past work in bacterial species has the synonymous mutations are not neutral because of the selection pressure against the synonymous sites (Begun and Aquadro 1992). So to get the true rates of mutation during evolution, Aaron et.al in 2004 calculated the mutation rate of budding yeast using the relationship between codon bias and synonymous divergence for four yeast species (Hirsh, Fraser, and Wall 2005). We downloaded and used this data to study the mutation rates of early and late replicating domains in relation to their *cis* and *trans* chromatin interaction profiles.

Chromatin interaction data of *E.coli* and mESCs. To test the evolutionary conservation of minimization of expression noise by *cis* interactions, we have used the HiC data of *E.coli* (Xie et al. 2015)and mESCs (Schoenfelder et al. 2015). The *E.coli* data consisted of 0.2 million DNA interactions among 462 bins of 10kb. The mouse data consisted of 0.5million promoter centered interactions in mESCs. Additionally, we also used the 0.26 million RNAPII tethered promoter-

promoter interactions (Yubo Zhang et al. 2013) which we used to propose a mechanism of spatial restraint of chromatin fiber *cis* interactions.

Stress-Responsive genes. Broadly, yeast genome has two classes of genes which are called growth genes and stress-responsive genes. While the growth genes encode for proteins involved in housekeeping functions such as ribosomal assembly, chromatin organization, cell division, and cytoskeleton reorganization, etc., the stress-responsive genes mediate the cellular response to diverse kind of environmental stress conditions. It has been shown that stress-responsive genes are typically associated with TATA-containing promoters and are regulated by SAGA transcriptional complex as opposed to growth genes, which lacks TATA box in their promoter and are regulated by TFIID. To check whether there was a difference between the stress responsiveness of affected and unaffected genes, we downloaded the data from the Choi and Kim (2009), who had previously calculated the stress responsiveness of yeast genes by using the microarray data of mRNA measurement across different environmental stresses.

Dynamics of the gene-expression profile during the yeast stress response. In order to assess the condition-specific roles of chromatin factors, we have used the genome-wide mRNA measurements of the expression pattern of 5848 yeast genes in response to 15 different environmental perturbations, like heat shock, hyperosmotic shock, chemical, amino acid starvation, nitrogen depletion, etc. This dataset was downloaded from SGD and was generated by a previous study (Gasch et al. 2000).

Genome-wide binding of chromatin factors. For the investigation of non-transcription binding of chromatin factors, we downloaded the Chip-chip data for the genome-wide location analysis of 202 factors (Venters et al. 2011). We used this data to obtain a list of significantly bound genes for each chromatin factor which yielded a list of the 231,704 protein-DNA interactions meeting the criteria of 5% FDR (Venters et al. 2011).

***cis*-variability.** The gene expression profile is an important property related to the survival, adaptation, and evolution of organisms and species. The gene expression level is dependent on the binding of regulatory proteins (*trans* elements), or the presence and composition of the regulatory sequence elements such as promoters (*cis* elements). Therefore, the gene-expression profiles of different individuals, populations, or species can be modulated by the changes in *cis* or *trans*-regulatory factors. Choi and Kim, in 2008, calculated the *cis* and *trans* variability of a large number of yeast genes (Choi and Kim 2008). Using the data from previous study by Brem et.al (Brem et al. 2002) where a cross between a laboratory and wild-type strain of yeast is performed (BY and RM respectively), the Choi and Kim, in their 2008 study, separated the population derived from this cross according to the inheritance at promoter, classifying the segregants as either BY or RM genotype based on the genetic markers flanking the given

promoter. The transcriptional variance within each group was called *trans* variability, and between the means of two groups was defined as *cis* variability.

Gene ontology terms and enrichment analysis. Gene ontology refers to the development and computational representation of controlled vocabulary for functional annotation of each gene across all the species (Ashburner et al. 2000; Stevens, Goble, and Bechhofer 2000). This major bioinformatics initiative was started in 1998, initially as a collaboration between Saccharomyces Genome Database (SGD), Mouse Genome Database (MGD) and FlyBase. With the advancements over the following years, 40000 biological concepts have been annotated to the genes across multiple species. All the combined ontologies come out as directed acyclic graph (DAG), where each node represent the GO term id of given biological concept and the edge between the nodes represent the parent-child relationship or the hierarchy of annotation. The vocabulary of GO terms is classified into three different classes; namely, the biological process, molecular function, and cellular component Many of the studies in past have used the GO annotations to study the functional specialization of a given set of genes. Such a method is called the GO term enrichment analysis (Subramanian et al. 2005). To understand the functional distinction between affected and unaffected binding targets of chromatin factors, we did the GO term enrichment analysis (See methods).

SNMs in the yeast genome. To test the differences in genome integrity at affected and unaffected binding targets of chromatin factors in project II, we downloaded the data for genome-wide accumulation of SNMs in mutation accumulation (MA) lines from the supplementary link provided by a previous study (Zhu et al. 2014)This study employed a whole genome sequencing-based approach and identified the accumulation of >800 SNMs over >300000 generations in 145 diploid mutation accumulation lines in yeast.

ssDNA enrichment. In addition to SNMs, we tested the relative occurrence of DNA double-strand breaks (DSBs) at affected and unaffected binding targets to assess the difference in genome integrity at these locations. This data was taken from a study which generated the genome-wide (6337 genes) measurements of single-stranded DNA (ssDNA) associated meiotic double-strand breaks in repair-deficient *dmc1*Δ and *dmc1*Δ *rad51*Δ mutants (Buhler, Borde, and Lichten 2007).

Genetic interactions. The genetic interactions among different genes are defined by the fitness analysis of double mutants. Double mutants often display the phenotypic effects (for eg. fitness or survivability) which could not be explained by simple product or summation of effects of individual gene deletions. Such genes are said to be genetically interacting and can be of phenotypically positive (alleviating, one gene alleviates the deletion of other genes), negative (exaggerating, more severe phenotype of double deletion as compared to individual deletions), or lethal (double mutants do not survive). In yeast, the landscape of genetic interactions has been

extensively studied several high throughput screening strategies such as Synthetic Genetic Array (SGA, Baryshnikova et al. 2010), Diploid-based Synthetic Lethality Analysis (dSLAM, Pan et al. 2007) and Epistatic Miniarray Profile (E-MAP, Schuldiner et al. 2005). We have used the data for genetic interactions in project II when we analyzed and commented upon the non-essentiality of certain chromatin factors in wild-type conditions. For this purpose, we downloaded the data from the supplementary link of a genome-scale Synthetic Genetic Array (SGA) study covering the 75% yeast genome and yielding 5.5 million pairs of synthetic interactions (Costanzo et al. 2010).

Appendix B

Mathematical treatment of PCA

For a data matrix **X** (n × m), the goal of PCA is to find a linear transformation

$$\mathbf{Y} = \mathbf{PX} \tag{i}$$

such that it diagonalizes the co-variance matrix of Y.

If $\mathbf{S_X}$ is the co-variance matrix of X and $\mathbf{S_Y}$ is that of Y, given by the following relations:

$$\mathbf{S_X} = \frac{1}{\mathrm{n}-1}\mathbf{XX^T} \tag{ii}$$

$$\mathbf{S_Y} = \frac{1}{\mathrm{n}-1}\mathbf{YY^T} \tag{iii}$$

Substituting equation (i) in equation (ii) we get:

$$\mathbf{S_Y} = \frac{1}{\mathrm{n}-1}(\mathbf{PX})(\mathbf{PX})^\mathbf{T}$$

Rearranging which:

$$\mathbf{S_Y} = \frac{1}{\mathrm{n}-1}\mathbf{PXX^TP^T}$$

$$\mathbf{S_Y} = \mathbf{P}\boldsymbol{S_X}\mathbf{P^T}$$

The next step of derivation makes the use of spectral theorem which states that a symmetric matrix **A** can be diagonalized by the orthogonal matrix of its eigenvectors **E** as per the following relation:

$$\mathbf{A} = \mathbf{EDE^T}$$

Where **D** is the diagonal matrix in the above relation and **E** is arranged as columns. If one selects the linear transformation $\mathbf{P} \equiv \mathbf{E^T}$, and write the $\boldsymbol{S_X}$ as per the spectral theorem:

$$\mathbf{S_Y} = \mathbf{P}\boldsymbol{S_X}\mathbf{P^T}$$

$$\mathbf{S_Y} = \mathbf{P}(\mathbf{P^TDP})\mathbf{P^T}$$

Setting $\mathbf{P^T} = \mathbf{P^{-1}}$ (for orthogonal matrices), we get:

$$\mathbf{S_Y} = \mathbf{D} \tag{iv}$$

Which provides a solution for the PCA. Therefore, the principal components of data matrix **X** are the eigenvectors of relation $\mathbf{XX^T}$ arranged as rows. The above description of PCA is taken from Shlens (2014) and Tanaka (1988).

Appendix C

Model representation in PLSR analysis.

The simple linear regression between explanatory variable **x** with n rows and response variable **y** with n rows is given by the following equation:

$$\mathbf{y} = \mathrm{a}\mathbf{x} + \mathrm{b} + \boldsymbol{\epsilon} \quad \textbf{(i)}$$

Where a is the slope and b is the intercept of the best fit line, and $\boldsymbol{\epsilon}$ is the vector of residuals. For multiple explanatory variables (i.e $\boldsymbol{x_1, x_2, x_3 \ldots x_p}$), the regression equation can be modified as follows:

$$\mathbf{y} = a_1\boldsymbol{x_1} + a_2\boldsymbol{x_2} + a_3\boldsymbol{x_3} + \cdots + a_p\boldsymbol{x_p} + \mathrm{b} + \boldsymbol{\epsilon} \quad \textbf{(ii)}$$

In the matrix notation, if **X** is the data matrix with n×p dimensions (i.e. the explanatory variables) and **y** is the response vector with n×1 dimensions, multiple linear regression can be written as per the following relation:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \quad \textbf{(iii)}$$

Where $\boldsymbol{\beta}$ is the vector of coefficients of regression with 'p' rows and $\boldsymbol{\epsilon}$ is the vector of residuals with 'n' rows. The least square estimation of $\boldsymbol{\beta}$ is given by the following relation:

$$\boldsymbol{\beta} = (\mathbf{X^T X})^{-1}\mathbf{X^T Y}$$

However, such a solution to multiple linear regression might have redundancy in the regression coefficient due to the correlated entires in $(\mathbf{X^T X})$, the problem which is known as multicollinearity in the regression analysis (Haenlein and Kaplan 2005; J Wang, Yang, and Cui 2001). To overcome this problem, one option is to split the matrix **X** into orthogonal components through linear transformation. The PCA described in the appendix B describes one such transformation where matrix **X** can be decomposed into matrix **S** of scores and **P** of loadings as per the following relation:

$$\mathbf{S} = \mathbf{X}\mathbf{P}$$

Since the columns of matrix **S** obtained from PCA of matrix **X** are supposedly orthogonal, using the matrix **S** instead of **X** might give the better estimates of response vector **Y** through the following equation.

$$\mathbf{Y} = \mathbf{S}\boldsymbol{\beta} + \boldsymbol{\epsilon} \quad \textbf{(iv)}$$

This is called the principal component regression (PCR). However, PCR considers only the matrix **X** for linear transformation and therefore, the components extracted from the PCA of **X** might not necessarily predict the response variable. In the other version, what is known as PLSR model (Haenlein and Kaplan 2005; Handb. Partial Least Squares 2010; Wehrens 2007; Wold et al. 2009), both input matrix **X** and the response variable **Y** are transformed through the linear transformation as per the following equations:

$$\mathbf{X} = \mathbf{TP^T} + \mathbf{E} \quad \textbf{(v)}$$

$$\mathbf{Y} = \mathbf{UQ^T} + \mathbf{F} \quad \textbf{(vi)}$$

Where **T** and **U** are the n × r matrices of extracted latent vectors and **P** (p × r) and **Q** (1 × r) are the loadings of **X** and **Y**. **E** (*n* × *p)* and **F** (*1* × *p)* are the matrices of residuals. Importantly, **T** and **U** are constrained to have maximum covariance as per the following relation:

$$\mathbf{U} = \mathbf{TB} + \mathbf{H} \quad \textbf{(vii)}$$

Where **B** (*r* × *r)* diagonal matrix containing the regression coefficients and **H** is a matrix of residuals. The regression model can be defined by substituting the value of **U** from equation **vii** into equation **vi** as follows:

$$\mathbf{Y} = (\mathbf{TB} + \mathbf{H})\mathbf{Q^T} + \mathbf{F}$$

$$\mathbf{Y} = (\mathbf{TBQ^T}) + (\mathbf{HQ^T} + \mathbf{F})$$

$$\mathbf{Y} = \mathbf{TC^T} + \mathbf{F}^* \quad \textbf{(viii)}$$

Where, $\mathbf{C^T} = \mathbf{BQ^T}$ and $\mathbf{F}^* = \mathbf{HQ^T} + \mathbf{F}$

References

Abranches, Rita, Alison F. Beven, Luis Aragón-Alcaide, and Peter J. Shaw. 1998. "Transcription Sites Are Not Correlated with Chromosome Territories in Wheat Nuclei." *Journal of Cell Biology* 143(1): 5–12.

Ahn, Sung Hee et al. 2005. "Sterile 20 Kinase Phosphorylates Histone H2B at Serine 10 during Hydrogen Peroxide-Induced Apoptosis in S. Cerevisiae." *Cell* 120(1): 25–36.

Albert, Benjamin et al. 2013. "Systematic Characterization of the Conformation and Dynamics of Budding Yeast Chromosome XII." *Journal of Cell Biology* 202(2): 201–10.

Allen, Benjamin L., and Dylan J. Taatjes. 2015. "The Mediator Complex: A Central Integrator of Transcription." *Nature Reviews Molecular Cell Biology* 16(3): 155–66.

Allfrey, V G, R Faulkner, and A E Mirsky. 1964. "Acetylation And Methylation Of Histones And Their Possible Role In The Regulation Of Rna Synthesis." *Proceedings of the National Academy of Sciences of the United States of America* 51: 786–94.

Allis, C. David et al. 2007. "New Nomenclature for Chromatin-Modifying Enzymes." *Cell* 131(4): 633–36.

Allis, C. David, and Thomas Jenuwein. 2016. "The Molecular Hallmarks of Epigenetic Control." *Nature Reviews Genetics* 17(8): 487–500.

Almeida, Ricardo et al. 2018. "Chromatin Conformation Regulates the Coordination between DNA Replication and Transcription." *Nature Communications* 9(1): 1590.

Altaf, Mohammed et al. 2010. "NuA4-Dependent Acetylation of Nucleosomal Histones H4 and H2A Directly Stimulates Incorporation of H2A.Z by the SWR1 Complex." *Journal of Biological Chemistry* 285(21): 15966–77.

Altaf, Mohammed, Nehmé Saksouk, and Jacques Côté. 2007. "Histone Modifications in Response to DNA Damage." *Mutation Research - Fundamental and Molecular Mechanisms of Mutagenesis* 618(1–2): 81–90.

Amaratunga, Dhammika, and Javier Cabrera. 2001. "Analysis of Data from Viral DNA Microchips." *Journal of the American Statistical Association* 96(456): 1161–70.

Amrichová, Jana, Emilie Lukášová, Stanislav Kozubek, and Michal Kozubek. 2003. "Nuclear and Territorial Topography of Chromosome Telomeres in Human Lymphocytes." *Experimental Cell Research* 289(1): 11–26.

Ansari, Athar, and Michael Hampsey. 2005. "A Role for the CPF 3′ -End Processing Machinery in RNAP II-Dependent Gene Looping." *Genes and Development* 19(24): 2969–78.

Antequera, Francisco, and Adrian Bird. 2018. "CpG Islands: A Historical Perspective." In *Methods in Molecular Biology*, , 3–13.

Ashburner, Michael et al. 2000. "Gene Ontology: Tool for the Unification of Biology." *Nature Genetics* 25(1): 25–29.

Astrid E. Visser Françoise Jaunin, Stanislav Fakan, and Jacob A Aten. 2000. "High Resolution Analysis of Interphase Chromosome Domains." *J. Cell Sci.* 113: 2585–93.

Van Attikum, Haico, Olivier Fritsch, and Susan M. Gasser. 2007. "Distinct Roles for SWR1 and INO80 Chromatin Remodeling Complexes at Chromosomal Double-Strand Breaks." *EMBO Journal* 26(18): 4113–25.

Azuara, Véronique et al. 2006. "Chromatin Signatures of Pluripotent Cell Lines." *Nature Cell Biology* 8(5): 532–38.

Bagadia, Meenakshi, Arashdeep Singh, and Kuljeet Singh Sandhu. 2016. "Three Dimensional Organization of Genome Might Have Guided the Dynamics of Gene Order Evolution in Eukaryotes." *Genome Biology and Evolution* 8(3): 946–54.

Bakkenist, Christopher J., and Michael B. Kastan. 2003. "DNA Damage Activates ATM through Intermolecular Autophosphorylation and Dimer Dissociation." *Nature* 421(6922): 499–506.

Bancaud, Aurélien et al. 2009. "Molecular Crowding Affects Diffusion and Binding of Nuclear Proteins in Heterochromatin and Reveals the Fractal Organization of Chromatin." *EMBO Journal* 28(24): 3785–98.

Banerjee, T., and D. Chakravarti. 2011. "A Peek into the Complex Realm of Histone Phosphorylation." *Molecular and Cellular Biology* 31(24): 4858–73.

Banerji, Julian, Laura Olson, and Walter Schaffner. 1983. "A Lymphocyte-Specific Cellular Enhancer Is Located Downstream of the Joining Region in Immunoglobulin Heavy Chain Genes." *Cell* 33(3): 729–40.

Banerji, Julian, Sandro Rusconi, and Walter Schaffner. 1981. "Expression of a β-Globin Gene Is Enhanced by Remote SV40

DNA Sequences." *Cell* 27(2 PART 1): 299–308.

Bannister, Andrew J. et al. 2001. "Selective Recognition of Methylated Lysine 9 on Histone H3 by the HP1 Chromo Domain." *Nature* 410(6824): 120–24.

Bannister, Andrew J., and Tony Kouzarides. 2011. "Regulation of Chromatin by Histone Modifications." *Cell research* 21(3): 381–95.

Barber, Cynthia M. et al. 2004. "The Enhancement of Histone H4 and H2A Serine 1 Phosphorylation during Mitosis and S-Phase Is Evolutionarily Conserved." *Chromosoma* 112(7): 360–71.

Barnett, Chris, and Jocelyn E Krebs. 2011. "WSTF Does It All: A Multifunctional Protein in Transcription, Repair, and ReplicationThis Paper Is One of a Selection of Papers Published in a Special Issue Entitled 31st Annual International Asilomar Chromatin and Chromosomes Conference, and Has Undergon." *Biochemistry and Cell Biology* 89(1): 12–23.

Barton, Sheila C., M. A H Surani, and M. L. Norris. 1984. "Role of Paternal and Maternal Genomes in Mouse Development." *Nature* 311(5984): 374–76.

Baryshnikova, Anastasia et al. 2010. "Chapter 7 - Synthetic Genetic Array (SGA) Analysis in Saccharomyces Cerevisiae and Schizosaccharomyces Pombe." In *Guide to Yeast Genetics: Functional Genomics, Proteomics, and Other Systems Analysis*, Academic Press, 145–79.

Batada, Nizar N., and Laurence D. Hurst. 2007. "Evolution of Chromosome Organization Driven by Selection for Reduced Gene Expression Noise." *Nature Genetics* 39(8): 945–49.

Battulin, Nariman et al. 2015. "Comparison of the Three-Dimensional Organization of Sperm and Fibroblast Genomes Using the Hi-C Approach." *Genome Biology* 16(1): 77.

Bauer, Uta Maria et al. 2002. "Methylation at Arginine 17 of Histone H3 Is Linked to Gene Activation." *EMBO Reports* 3(1): 39–44.

Beagrie, Robert A. et al. 2017. "Complex Multi-Enhancer Contacts Captured by Genome Architecture Mapping." *Nature* 543(7646): 519–24.

Beck, S. et al. 1996. "Evolutionary Dynamics of Non-Coding Sequences within the Class II Region of the Human MHC." *Journal of Molecular Biology* 255(1): 1–13.

Bedford, Mark T., and Steven G. Clarke. 2009. "Protein Arginine Methylation in Mammals: Who, What, and Why." *Molecular cell* 33(1): 1–13.

Bednar, J. et al. 1998. "Nucleosomes, Linker DNA, and Linker Histone Form a Unique Structural Motif That Directs the Higher-Order Folding and Compaction of Chromatin." *Proceedings of the National Academy of Sciences of the United States of America* 95(24): 14173–78.

Begun, David J., and Charles F. Aquadro. 1992. "Levels of Naturally Occurring DNA Polymorphism Correlate with Recombination Rates in D. Melanogaster." *Nature* 356(6369): 519–20.

Bell, Adam C., and Gary Felsenfeld. 2000. "Methylation of a CTCF-Dependent Boundary Controls Imprinted Expression of the Igf2 Gene." *Nature* 405(6785): 482–85.

Belmont, A. S., J. W. Sedat, and D. A. Agard. 1987. "A Three-Dimensional Approach to Mitotic Chromosome Structure: Evidence for a Complex Hierarchical Organization." *The Journal of cell biology* 105(1): 77–92.

Belmont, Andrew S. 2014. "Large-Scale Chromatin Organization: The Good, the Surprising, and the Still Perplexing." *Current Opinion in Cell Biology* 26(1): 69–78.

Belmont, Andrew S., and Kathy Bruce. 1994. "Visualization of G1 Chromosomes: A Folded, Twisted, Supercoiled Chromonema Model of Interphase Chromatid Structure." *Journal of Cell Biology* 127(2): 287–302.

Belotserkovskaya, Rimma et al. 2003. "FACT Facilitates Transcription-Dependent Nucleosome Alteration." *Science* 301(5636): 1090–93.

Berger, Shelley L. 2007. "The Complex Language of Chromatin Regulation during Transcription." *Nature* 447(7143): 407–12.

Bernardi, Giorgio et al. 1985. "The Mosaic Genome of Warm-Blooded Vertebrates." *Science* 228(4702): 953–58.

Bernstein, B. E. et al. 2002. "Methylation of Histone H3 Lys 4 in Coding Regions of Active Genes." *Proceedings of the National Academy of Sciences* 99(13): 8695–8700.

Bernstein, Bradley E. et al. 2006. “A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells.” *Cell* 125(2): 315–26.

Bi, X., Q. Yu, J. J. Sandmeier, and Y. Zou. 2004. “Formation of Boundaries of Transcriptionally Silent Chromatin by Nucleosome-Excluding Structures.” *Molecular and Cellular Biology* 24(5): 2118–31.

Bi, Xin, and James R. Broach. 1999. “UAS(Rpg) Can Function as a Heterochromatin Boundary Element in Yeast.” *Genes and Development* 13(9): 1089–1101.

Biémont, Christian. 2010. “A Brief History of the Status of Transposable Elements: From Junk DNA to Major Players in Evolution.” *Genetics* 186(4): 1085–93.

Bird, Adrian et al. 1985. “A Fraction of the Mouse Genome That Is Derived from Islands of Nonmethylated, CpG-Rich DNA.” *Cell* 40(1): 91–99.

Bird, Adrian P. 1980. “DNA Methylation and the Frequency of CpG in Animal DNA.” *Nucleic Acids Research* 8(7): 1499–1504.

Bird, Adrian P. 1987. “CpG Islands as Gene Markers in the Vertebrate Nucleus.” *Trends in Genetics* 3(C): 342–47.

Bird, Adrian P., Mary H. Taggart, and Barbara A. Smith. 1979. “Methylated and Unmethylated DNA Compartments in the Sea Urchin Genome.” *Cell* 17(4): 889–901.

Birney, Ewan et al. 2007. “Identification and Analysis of Functional Elements in 1% of the Human Genome by the ENCODE Pilot Project.” *Nature* 447(7146): 799–816.

Biscotti, Maria Assunta, Ettore Olmo, and J. S.(Pat) Heslop-Harrison. 2015. “Repetitive DNA in Eukaryotic Genomes.” *Chromosome Research* 23(3): 415–20.

Blackledge, Neil P., Christopher J. Ott, Austin E. Gillen, and Ann Harris. 2009. “An Insulator Element 3′ to the CFTR Gene Binds CTCF and Reveals an Active Chromatin Hub in Primary Cells.” *Nucleic Acids Research* 37(4): 1086–94.

Blanc, Roméo S., and Stéphane Richard. 2017. “Arginine Methylation: The Coming of Age.” *Molecular Cell* 65(1): 8–24.

Blumenthal, Thomas. 1998. “Gene Clusters and Polycistronic Transcription in Eukaryotes.” *BioEssays* 20(6): 480–87.

Blumenthal, Thomas. 2004. “Operons in Eukaryotes.” *Briefings in functional genomics & proteomics* 3(3): 199–211.

Boeger, Hinrich, Joachim Griesenbeck, J. Seth Strattan, and Roger D. Kornberg. 2004. “Removal of Promoter Nucleosomes by Disassembly Rather than Sliding in Vivo.” *Molecular Cell* 14(5): 667–73.

Bolstad, B. M., R. A. Irizarry, M. Åstrand, and T. P. Speed. 2003. “A Comparison of Normalization Methods for High Density Oligonucleotide Array Data Based on Variance and Bias.” *Bioinformatics* 19(2): 185–93.

Bolzer, Andreas et al. 2005. “Three-Dimensional Maps of All Chromosomes in Human Male Fibroblast Nuclei and Prometaphase Rosettes.” *PLoS Biology* 3(5): 0826–42.

Boutanaev, Alexander M., Alla I. Kalmykova, Yuri Y. Shevelyov, and Dmitry I. Nurminsky. 2002. “Large Clusters of Co-Expressed Genes in the Drosophila Genome.” *Nature* 420(6916): 666–69.

Boyle, S. 2001. “The Spatial Organization of Human Chromosomes within the Nuclei of Normal and Emerin-Mutant Cells.” *Human Molecular Genetics* 10(3): 211–19.

Bradbury, E M. 1975. “Chapter 1 Histones, Chromatin Structure, and Control of Cell Division.” In *Experimental Systems for Analysis of Cell Differentiation*, eds. A A Moscona and Alberto B T - Current Topics in Developmental Biology Monroy. Academic Press, 1–13.

Bradbury, E Morton. 1989. “K. E. Van Holde. Chromatin. Series in Molecular Biology. Springer-Verlag, New York. 1988. 530 Pp. $98.00.” *Journal of Molecular Recognition* 2(3): i–i.

Branco, Miguel R., and Ana Pombo. 2006. “Intermingling of Chromosome Territories in Interphase Suggests Role in Translocations and Transcription-Dependent Associations.” *PLoS Biology* 4(5): 780–88.

Brannan I, Camilynn, and Marisa S. Bartolomei. 1999. “Mechanisms of Genomic Imprinting.” *Current Opinion in Genetics and Development* 9(2): 164–70.

Brehove, Matthew et al. 2015. “Histone Core Phosphorylation Regulates DNA Accessibility.” *Journal of Biological Chemistry* 290(37): 22612–21.

Brem, Rachel B., Gaël Yvert, Rebecca Clinton, and Leonid Kruglyak. 2002. “Genetic Dissection of Transcriptional Regulation

in Budding Yeast." *Science* 296(5568): 752–55.

Bressmann, Tim. 2004. "Self-Inflicted Cosmetic Tongue Split: A Case Report." *Journal (Canadian Dental Association)* 70(3): 156–57.

Bridger, Joanna M, Harald Herrmann, Christian Münkel, and Peter Lichter. 1998. "Identification of an Interchromosomal Compartment by Polymerization of Nuclear-Targeted Vimentin." *Journal of cell science* 111: 1241–53.

Britten, R. J., and D. E. Kohne. 1968. "Repeated Sequences in DNA." *Science* 161(3841): 529–40.

Britten, Roy J., Dale E. Graham, and Berney R. Neufeld. 1974. "Analysis of Repeating DNA Sequences by Reassociation." *Methods in Enzymology* 29(C): 363–418.

Broccoli, Dominique, Agata Smogorzewska, Laura Chong, and Titia de Lange. 1997. "Human Telomeres Contain Two Distinct Myb-Related Proteins, TRF1 and TRF2." *Nature Genetics* 17(2): 235.

Brown, Christopher R., and Pamela A. Silver. 2007. "Transcriptional Regulation at the Nuclear Pore Complex." *Current Opinion in Genetics and Development* 17(2): 100–106.

Brown, Jill M. et al. 2008. "Association between Active Genes Occurs at Nuclear Speckles and Is Modulated by Chromatin Environment." *Journal of Cell Biology* 182(6): 1083–97.

Brownell, J. E. et al. 1996. "Tetrahymena Histone Acetyltransferase A: A Homolog to Yeast Gcn5p Linking Histone Acetylation to Gene Activation." *Cell* 84(6): 843–51.

Brunmeir, Reinhard et al. 2010. "Epigenetic Regulation of a Murine Retrotransposon by a Dual Histone Modification Mark." *PLoS genetics* 6(4): e1000927.

Bucheton, A. et al. 1984. "The Molecular Basis of I-R Hybrid Dysgenesis in Drosophila Melanogaster: Identification, Cloning, and Properties of the I Factor." *Cell* 38(1): 153–63.

Buhler, Cyril, Valérie Borde, and Michael Lichten. 2007. "Mapping Meiotic Single-Strand DNA Reveals a New Landscape of DNA Double-Strand Breaks in Saccharomyces Cerevisiae." *PLoS Biology* 5(12): 2797–2808.

Bunz, Uwe H.F. 2005. "How Are Alkynes Scrambled?" *Science* 308(5719): 216–17.

Burke, Les J. et al. 2005. "CTCF Binding and Higher Order Chromatin Structure of the H19 Locus Are Maintained in Mitotic Chromatin." *EMBO Journal* 24(18): 3291–3300.

Burke, Wylie, and Walton L Fangman. 1975. "Temporal Order in Yeast Chromosome Replication." *Cell* 5(3): 263–69.

Cairns, Bradley R. et al. 1999. "Two Functionally Distinct Forms of the RSC Nucleosome-Remodeling Complex, Containing Essential AT Hook, BAH, and Bromodomains." *Molecular Cell* 4(5): 715–23.

Cairns, Bradley R. 2009. "The Logic of Chromatin Architecture and Remodelling at Promoters." *Nature* 461: 193.

Cam, Hugh P. et al. 2005. "Comprehensive Analysis of Heterochromatin- and RNAi-Mediated Epigenetic Control of the Fission Yeast Genome." *Nature Genetics* 37(8): 809–19.

Cameron, John R., Elwyn Y. Loh, and Ronald W. Davis. 1979. "Evidence for Transposition of Dispersed Repetitive DNA Families in Yeast." *Cell* 16(4): 739–51.

Cao, Jian, and Qin Yan. 2012. "Histone Ubiquitination and Deubiquitination in Transcription, DNA Damage Response, and Cancer." *Frontiers in Oncology* 2: 26.

Caron, H. et al. 2001. "The Human Transcriptome Map: Clustering of Highly Expressed Genes in Chromosomal Domains." *Science* 291(5507): 1289–93.

Carrozza, Michael J. et al. 2005. "Histone H3 Methylation by Set2 Directs Deacetylation of Coding Regions by Rpd3S to Suppress Spurious Intragenic Transcription." *Cell* 123(4): 581–92.

Carrozza, Michael J., Rhea T. Utley, Jerry L. Workman, and Jacques Côté. 2003. "The Diverse Functions of Histone Acetyltransferase Complexes." *Trends in Genetics* 19(6): 321–29.

CHAFFEY, NIGEL. 2003. 91 Annals of Botany *Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K. and Walter, P. Molecular Biology of the Cell. 4th Edn*. Oxford University Press.

Chagin, Vadim O., Jeffrey H. Stear, and M. Cristina Cardoso. 2010. "Organization of DNA Replication." *Cold Spring Harbor Perspectives in Biology* 2(4): a000737–a000737.

Chandy, Mark, José L Gutiérrez, Philippe Prochasson, and Jerry L Workman. 2006. "SWI/SNF Displaces SAGA-Acetylated

Nucleosomes." *Eukaryotic Cell* 5(10): 1738–47.

Chen, Dagang et al. 1999. "Regulation of Transcription by a Protein Methyltransferase." *Science* 284(5423): 2174–77.

Chen, Yunqin, Yixue Li, Jia Wei, and Yuan-Yuan Li. 2014. "Transcriptional Regulation and Spatial Interactions of Head-to-Head Genes." *BMC Genomics* 15(1): 519.

Chepelev, Iouri et al. 2012. "Characterization of Genome-Wide Enhancer-Promoter Interactions Reveals Co-Expression of Interacting Genes and Modes of Higher Order Chromatin Organization." *Cell Research* 22(3): 490–503.

Cheung, Peter et al. 2000. "Synergistic Coupling of Histone H3 Phosphorylation and Acetylation in Response to Epidermal Growth Factor Stimulation." *Molecular Cell* 5(6): 905–15.

Cheung, Wang L. et al. 2003. "Apoptotic Phosphorylation of Histone H2B Is Mediated by Mammalian Sterile Twenty Kinase." *Cell* 113(4): 507–17.

Cheung, Wang L. et al. 2005. "Phosphorylation of Histone H4 Serine 1 during DNA Damage Requires Casein Kinase II in S. Cerevisiae." *Current Biology* 15(7): 656–60.

Cho, Raymond J. et al. 1998. "A Genome-Wide Transcriptional Analysis of the Mitotic Cell Cycle." *Molecular Cell* 2(1): 65–73.

Choi, Jung Kyoon, and Young Joon Kim. 2008. "Epigenetic Regulation and the Variability of Gene Expression." *Nature Genetics* 40(2): 141–47.

Choi, Jung Kyoon, and Young Joon Kim. 2009. "Intrinsic Variability of Gene Expression Encoded in Nucleosome Positioning Sequences." *Nature Genetics* 41(4): 498–503.

Chuang, Tony Chih Yuan et al. 2004. "The Three-Dimensional Organization of Telomeres in the Nucleus of Mammalian Cells." *BMC biology* 2: 12.

Clapier, Cedric R., and Bradley R. Cairns. 2009. "The Biology of Chromatin Remodeling Complexes." *Annual Review of Biochemistry* 78(1): 273–304.

Clapier, Cedric R., Janet Iwasa, Bradley R. Cairns, and Craig L. Peterson. 2017. "Mechanisms of Action and Regulation of ATP-Dependent Chromatin-Remodelling Complexes." *Nature Reviews Molecular Cell Biology* 18(7): 407–22.

Clark, Nathan L., Eric Alani, and Charles F. Aquadro. 2012. "Evolutionary Rate Covariation Reveals Shared Functionality and Coexpression of Genes." *Genome Research* 22(4): 714–20.

Clayton, Alison L., Catherine A. Hazzalin, and Louis C. Mahadevan. 2006. "Enhanced Histone Acetylation and Transcription: A Dynamic Perspective." *Molecular Cell* 23(3): 289–96.

Cleveland, William S. 1979. "Robust Locally Weighted Regression and Smoothing Scatterplots." *Journal of the American Statistical Association* 74(368): 829–36.

Collins, Nadine et al. 2002. "An ACF1-ISWI Chromatin-Remodeling Complex Is Required for DNA Replication through Heterochromatin." *Nature Genetics* 32(4): 627–32.

Conaway, Ronald C., and Joan Weliky Conaway. 2009. "The INO80 Chromatin Remodeling Complex in Transcription, Replication and Repair." *Trends in Biochemical Sciences* 34(2): 71–77.

Cooper, David N., Mary H. Taggart, and Adrian P. Bird. 1983. "Unmethlated Domains in Vertebrate DNA." *Nucleic Acids Research* 11(3): 647–58.

Cooper, Gregory M. et al. 2003. "Quantitative Estimates of Sequence Divergence for Comparative Analyses of Mammalian Genomes." *Genome Research* 13(5): 813–20.

Cordaux, Richard, and Mark A. Batzer. 2009. "The Impact of Retrotransposons on Human Genome Evolution." *Nature Reviews Genetics* 10(10): 691–703.

Cornforth, Michael N. et al. 2002. "Chromosomes Are Predominantly Located Randomly with Respect to Each Other in Interphase Human Cells." *Journal of Cell Biology* 159(2): 237–44.

Costanzo, Michael et al. 2010. "The Genetic Landscape of a Cell." *Science* 327(5964): 425–31.

Cournac, Axel et al. 2012. "Normalization of a Chromosomal Contact Map." *BMC Genomics* 13(1): 436.

Cowan, C. R. 2001. "The Polar Arrangement of Telomeres in Interphase and Meiosis. Rabl Organization and the Bouquet." *PLANT PHYSIOLOGY* 125(2): 532–38.

Cowley, Michael, and Rebecca J. Oakey. 2012. "Resetting for the Next Generation." *Molecular Cell* 48(6): 819–21.

Crabbe, Laure et al. 2012. "Human Telomeres Are Tethered to the Nuclear Envelope during Postmitotic Nuclear Assembly." *Cell Reports* 2(6): 1521–29.

Cremer, T. et al. 1993. "Role of Chromosome Territories in the Functional Compartmentalization of the Cell Nucleus." In *Cold Spring Harbor Symposia on Quantitative Biology*, , 777–92.

Cremer, T., and Marion Cremer. 2010. "Chromosome Territories." *Cold Spring Harbor Perspectives in Biology* 2(3): a003889–a003889.

Cremer, T et al. 1982. "Rabl's Model of the Interphase Chromosome Arrangement Tested in Chinise Hamster Cells by Premature Chromosome Condensation and Laser-UV-Microbeam Experiments." *Human Genetics* 60(1): 46–56.

Cremer, T et al. 2000. "Chromosome Territories, Interchromatin Domain Compartment, and Nuclear Matrix: An Integrated View of the Functional Nuclear Architecture." *Critical reviews in eukaryotic gene expression* 10(2): 179–212.

Cremer, T, and C. Cremer. 2001. "Chromosome Territories, Nuclear Architecture and Gene Regulation in Mammalian Cells." *Nature reviews. Genetics* 2(4): 292–301.

Di Croce, Luciano et al. 2010. "A Phosphorylation Switch Regulates the Transcriptional Activation of Cell Cycle Regulator P21 by Histone Deacetylase Inhibitors." *Journal of Biological Chemistry* 285(52): 41062–73.

Croft, Jenny A. et al. 1999. "Differences in the Localization and Morphology of Chromosomes in the Human Nucleus." *Journal of Cell Biology* 145(6): 1119–31.

Cuddapah, Suresh et al. 2009. "Global Analysis of the Insulator Binding Protein CTCF in Chromatin Barrier Regions Reveals Demarcation of Active and Repressive Domains." *Genome Research* 19(1): 24–32.

Cusanovich, Darren A., Bryan Pavlovic, Jonathan K. Pritchard, and Yoav Gilad. 2014. "The Functional Consequences of Variation in Transcription Factor Binding." *PLoS genetics* 10(3): e1004226.

D'angelo, Maximiliano A. 2018. "Nuclear Pore Complexes as Hubs for Gene Regulation." *Nucleus* 9(1): 142–48.

Dai, Zhiming, and Xianhua Dai. 2012. "Nuclear Colocalization of Transcription Factor Target Genes Strengthens Coregulation in Yeast." *Nucleic Acids Research* 40(1): 27–36.

van Dam, Joost W. et al. 2012. "Additive Pressures of Elevated Sea Surface Temperatures and Herbicides on Symbiont-Bearing Foraminifera" ed. Frank Seebacher. *PLoS ONE* 7(3): e33900.

Davies, Adelina A. et al. 2001. "Role of BRCA2 in Control of the RAD51 Recombination and DNA Repair Protein." *Molecular Cell* 7(2): 273–82.

Dawson, Mark A. et al. 2009. "JAK2 Phosphorylates Histone H3Y41 and Excludes HP1α from Chromatin." *Nature* 461(7265): 819–22.

Daxinger, Lucia, and Emma Whitelaw. 2012. "Understanding Transgenerational Epigenetic Inheritance via the Gametes in Mammals." *Nature Reviews Genetics* 13: 153.

Dekker, Job, Karsten Rippe, Martijn Dekker, and Nancy Kleckner. 2002. "Capturing Chromosome Conformation." *Science* 295(5558): 1306–11.

Delaire, Stephanie, Yina Hsing Huang, Shaio Wei Chan, and Ellen A. Robey. 2004. "Dynamic Repositioning of CD4 and CD8 Genes during T Cell Development." *The Journal of Experimental Medicine* 200(11): 1427–35.

Delcuve, Geneviève P., Dilshad H. Khan, and James R. Davie. 2012. "Roles of Histone Deacetylases in Epigenetic Regulation: Emerging Paradigms from Studies with Inhibitors." *Clinical epigenetics* 4(1): 5.

Delmas, V, D G Stokes, and R P Perry. 1993. "A Mammalian DNA-Binding Protein That Contains a Chromodomain and an SNF2/SWI2-like Helicase Domain." *Proceedings of the National Academy of Sciences* 90(6): 2414–18.

Dempsey, Adam A., Noel Pabalan, Hong Chang Tang, and Choong Chin Liew. 2001. "Organization of Human Cardiovascular-Expressed Genes on Chromosomes 21 and 22." *Journal of Molecular and Cellular Cardiology* 33(3): 587–91.

Demuth, Jeffery P. et al. 2006. "The Evolution of Mammalian Gene Families." *PloS one* 1(1): e85.

Deng, Xinxian et al. 2015. "Bipartite Structure of the Inactive Mouse X Chromosome." *Genome Biology* 16(1): 152.

Dennis, Carina. 2001. "Tiled Arrays for Gene Hunting." *Nature Reviews Genetics* 2: 161.

Desjarlais, Renee, and Peter J. Tummino. 2016. "Role of Histone-Modifying Enzymes and Their Complexes in Regulation of

Chromatin Biology." *Biochemistry* 55(11): 1584–99.

Dhalluin, Christophe et al. 1999. "Structure and Ligand of a Histone Acetyltransferase Bromodomain." *Nature* 399(6735): 491–96.

Dhar, Surbhi, Ozge Gursoy-Yuzugullu, Ramya Parasuram, and Brendan D. Price. 2017. "The Tale of a Tail: Histone H4 Acetylation and the Repair of DNA Breaks." *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 372(1731): 20160284.

Diament, Alon, Ron Y Pinter, and Tamir Tuller. 2014. "Three-Dimensional Eukaryotic Genomic Organization Is Strongly Correlated with Codon Usage Expression and Function." *Nature Communications* 5(1): 5876.

Dietzel, Steffen et al. 1999. "The 3D Positioning of ANT2 and ANT3 Genes within Female X Chromosome Territories Correlates with Gene Activity." *Experimental Cell Research* 252(2): 363–75.

Dillon, Niall. 2004. "Heterochromatin Structure and Function." *Biology of the Cell* 96(8): 631–37.

Dimitrova, Daniela S., and David M. Gilbert. 2000. "Temporally Coordinated Assembly and Disassembly of Replication Factories in the Absence of DNA Synthesis." *Nature Cell Biology* 2(10): 686–94.

Dixon, Jesse R et al. 2012. "Topological Domains in Mammalian Genomes Identified by Analysis of Chromatin Interactions." *Nature* 485(7398): 376–80.

Dixon, Jesse R et al. 2015. "Chromatin Architecture Reorganization during Stem Cell Differentiation." *Nature* 518: 331.

Djebali, Sarah et al. 2012. "Evidence for Transcript Networks Composed of Chimeric Rnas in Human Cells." *PLoS ONE* 7(1): e28213.

Doenecke, Detlef, and Peter Karlson. 1984. "Albrecht Kossel and the Discovery of Histones." *Trends in Biochemical Sciences* 9(9): 404–5.

Dong, Xianjun, and Zhiping Weng. 2013. "The Correlation between Histone Modifications and Gene Expression." *Epigenomics* 5(2): 113–16.

Donze, D., and R. T. Kamakaka. 2001. "RNA Polymerase III and RNA Polymerase II Promoter Complexes Are Heterochromatin Barriers in Saccharomyces Cerevisiae." *EMBO Journal* 20(3): 520–31.

Doskočil, J., and F. Šorm. 1962. "Distribution of 5-Methylcytosine in Pyrimidine Sequences of Deoxyribonucleic Acids." *BBA Specialized Section on Nucleic Acids and Related Subjects* 55(6): 953–59.

Dostie, Josée et al. 2006. "Chromosome Conformation Capture Carbon Copy (5C): A Massively Parallel Solution for Mapping Interactions between Genomic Elements." *Genome Research* 16(10): 1299–1309.

Dostie, Josée, and Wendy A. Bickmore. 2012. "Chromosome Organization in the Nucleus - Charting New Territory across the Hi-Cs." *Current Opinion in Genetics and Development* 22(2): 125–31.

Drissen, Roy et al. 2004. "The Active Spatial Organization of the β-Globin Locus Requires the Transcription Factor EKLF." *Genes and Development* 18(20): 2485–90.

Dryden, Nicola H. et al. 2014. "Unbiased Analysis of Potential Targets of Breast Cancer Susceptibility Loci by Capture Hi-C." *Genome Research* 24(11): 1854–68.

Du, Zhenhai et al. 2017. "Allelic Reprogramming of 3D Chromatin Architecture during Early Mammalian Development." *Nature* 547(7662): 232–35.

Duan, Ming Rui, and Michael J. Smerdon. 2014. "Histone H3 Lysine 14 (H3K14) Acetylation Facilitates DNA Repair in a Positioned Nucleosome by Stabilizing the Binding of the Chromatin Remodeler RSC (Remodels Structure of Chromatin)." *Journal of Biological Chemistry* 289(12): 8353–63.

Duan, Zhijun et al. 2010. "A Three-Dimensional Model of the Yeast Genome." *Nature* 465: 363.

Duret, Laurent, Dominique Mouchiroud, and Christian Gautier. 1995. "Statistical Analysis of Vertebrate Sequences Reveals That Long Genes Are Scarce in GC-Rich Isochores." *Journal of Molecular Evolution* 40(3): 308–17.

Durrin, Linda K., Randall K. Mann, Paul S. Kayne, and Michael Grunstein. 1991. "Yeast Histone H4 N-Terminal Sequence Is Required for Promoter Activation in Vivo." *Cell* 65(6): 1023–31.

Ea, Vuthy, Marie Odile Baudement, Annick Lesne, and Thierry Forné. 2015. "Contribution of Topological Domains and Loop Formation to 3D Chromatin Organization." *Genes* 6(3): 734–50.

Ebbert, Ronald, Alexander Birkmann, and Hans-Joachim Schüller. 2002. “The Product of the SNF2/SWI2 Paralogue INO80 of Saccharomyces Cerevisiae Required for Efficient Expression of Various Yeast Structural Genes Is Part of a High-Molecular-Weight Protein Complex.” *Molecular Microbiology* 32(4): 741–51.

Eberharter, Anton, and Peter B. Becker. 2002. “Histone Acetylation: A Switch between Repressive and Permissive Chromatin. Second in Review on Chromatin Dynamics.” *EMBO Reports* 3(3): 224–29.

Ebersole, Thomas et al. 2011. “TRNA Genes Protect a Reporter Gene from Epigenetic Silencing in Mouse Cells.” *Cell Cycle* 10(16): 2779–91.

Ehrenhofer-Murray, Ann E. 2004. “Chromatin Dynamics at DNA Replication, Transcription and Repair.” *European Journal of Biochemistry* 271(12): 2335–49.

Eissenberg, Joel C, and Sarah CR Elgin. 2014. “Heterochromatin and Euchromatin.” In *ELS*, Chichester, UK: John Wiley & Sons, Ltd, 1–7.

Eivazova, E R, and T M Aune. 2004. “Dynamic Alterations in the Conformation of the Ifng Gene Region during T Helper Cell Differentiation.” *Proceedings of the National Academy of Sciences* 101(1): 251–56.

Ekwall, Annelie Stralfors and Karl. 2011. “Heterochromatin and Euchromatin – Organization, Boundaries, and Gene Regulation.” 71(24): 1–19.

Elbarbary, Reyad A., Bronwyn A. Lucas, and Lynne E. Maquat. 2016. “Retrotransposons as Regulators of Gene Expression.” *Science (New York, N.Y.)* 351(6274): aac7247.

Elhamamsy, Amr Rafat. 2017. “Role of DNA Methylation in Imprinting Disorders: An Updated Review.” *Journal of Assisted Reproduction and Genetics* 34(5): 549–62.

Elizondo, Leah, Paymaan Jafar-Nejad, J. Clewing, and Cornelius Boerkoel. 2009. “Gene Clusters, Molecular Evolution and Disease: A Speculation.” *Current Genomics* 10(1): 64–75.

Ellegren, Hans. 2004. “Microsatellites: Simple Sequences with Complex Evolution.” *Nature Reviews Genetics* 5(6): 435–45.

Ellis, Leigh, Peter W Atadja, and Ricky W Johnstone. 2009. “Epigenetics in Cancer: Targeting Chromatin Modifications.” *Molecular Cancer Therapeutics* 8(6): 1409–20.

Eltsov, M. et al. 2008. “Analysis of Cryo-Electron Microscopy Images Does Not Support the Existence of 30-Nm Chromatin Fibers in Mitotic Chromosomes in Situ.” *Proceedings of the National Academy of Sciences* 105(50): 19732–37.

Engreitz, Jesse M., Vineeta Agarwala, and Leonid A. Mirny. 2012. “Three-Dimensional Genome Architecture Influences Partner Selection for Chromosomal Translocations in Human Disease” ed. Shawn Ahmed. *PLoS ONE* 7(9): e44196.

Erdel, Fabian, Jana Krug, Gernot Längst, and Karsten Rippe. 2011. “Targeting Chromatin Remodelers: Signals and Search Mechanisms.” *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms* 1809(9): 497–508.

Ernst, Jason et al. 2011. “Mapping and Analysis of Chromatin State Dynamics in Nine Human Cell Types.” *Nature* 473(7345): 43–49.

Ernst, Jason, and Manolis Kellis. 2010. “Discovery and Characterization of Chromatin States for Systematic Annotation of the Human Genome.” *Nature biotechnology* 28(8): 817–25.

Ernst, Jason, and Manolis Kellis. 2012. “ChromHMM: Automating Chromatin-State Discovery and Characterization.” *Nature Methods* 9(3): 215–16.

Eser, Umut et al. 2017. “Form and Function of Topologically Associating Genomic Domains in Budding Yeast.” *Proceedings of the National Academy of Sciences* 114(15): E3061–70.

Eskiw, Christopher, and Jenifer Mitchell. 2016. “Transcription Factories.” In *The Functional Nucleus*, Cham: Springer International Publishing, 469–89.

Fan, Yuhong et al. 2005. “Histone H1 Depletion in Mammals Alters Global Chromatin Structure but Causes Specific Changes in Gene Regulation.” *Cell* 123(7): 1199–1212.

Fares, Mario A. et al. 2013. “The Roles of Whole-Genome and Small-Scale Duplications in the Functional Specialization of Saccharomyces Cerevisiae Genes.” *PLoS genetics* 9(1): e1003176.

Farooq, Zeenat, Shahid Banday, Tej K. Pandita, and Mohammad Altaf. 2016. “The Many Faces of Histone H3K79 Methylation.” *Mutation Research - Reviews in Mutation Research* 768: 46–52.

Fawcett, D H, C K Lister, E Kellett, and D J Finnegan. 1986. “Transposable Elements Controlling I-R Hybrid Dysgenesis in D. Melanogaster Are Similar to Mammalian LINEs.” *Cell* 47(6): 1007–15.

Felsenfeld, Gary. 2014. “A Brief History of Epigenetics.” *Cold Spring Harbor Perspectives in Biology* 6(1): a018200–a018200.

Feng, Suhua et al. 2014. “Genome-Wide Hi-C Analyses in Wild Type and Mutants Reveal High-Resolution Chromatin Interactions in Arabidopsis.” *Molecular cell* 55(5): 694–707.

Fernandez-Capetillo, Oscar, C. David Allis, and André Nussenzweig. 2004. “Phosphorylation of Histone H2B at DNA Double-Strand Breaks.” *The Journal of Experimental Medicine* 199(12): 1671–77.

Ferrier, David E.K., and Peter W.H. Holland. 2001. “Ancient Origin of the Hox Gene Cluster.” *Nature Reviews Genetics* 2(1): 33–38.

Ferris, Patrick J., E Virginia Armbrust, and Ursula W. Goodenough. 2002. “Genetic Structure of the Mating-Type Locus Of.” *Molecular Biotechnology* 200(January): 181–200.

Feschotte, Cédric, and Ellen J. Pritham. 2007. “DNA Transposons and the Evolution of Eukaryotic Genomes.” *Annual Review of Genetics* 41(1): 331–68.

Filion, Guillaume J. et al. 2010. “Systematic Protein Location Mapping Reveals Five Principal Chromatin Types in Drosophila Cells.” *Cell* 143(2): 212–24.

Finch, J. T., and A. Klug. 1976. “Solenoidal Model for Superstructure in Chromatin.” *Proceedings of the National Academy of Sciences of the United States of America* 73(6): 1897–1901.

Flaus, Andrew, and Tom Owen-Hughes. 2011. “Mechanisms for ATP-Dependent Chromatin Remodelling: The Means to the End.” *FEBS Journal* 278(19): 3579–95.

Flyamer, Ilya M. et al. 2017. “Single-Nucleus Hi-C Reveals Unique Chromatin Reorganization at Oocyte-to-Zygote Transition.” *Nature* 544(7648): 110–14.

Foster, Elinor R., and Jessica A. Downs. 2005. “Histone H2A Phosphorylation in DNA Double-Strand Break Repair.” *FEBS Journal* 272(13): 3231–40.

Franke, Vedran et al. 2017. “Long Terminal Repeats Power Evolution of Genes and Gene Expression Programs in Mammalian Oocytes and Zygotes.” *Genome Research* 27(8): 1384–94.

Friedli, Marc, and Didier Trono. 2015. “The Developmental Control of Transposable Elements and the Evolution of Higher Species.” *Annual Review of Cell and Developmental Biology* 31(1): 429–51.

Fry, Christopher J., and Craig L. Peterson. 2001. “Chromatin Remodeling Enzymes: Who’s on First?” *Current Biology* 11(5): 185–97.

Fudenberg, Geoffrey et al. 2016. “Formation of Chromosomal Domains by Loop Extrusion.” *Cell Reports* 15(9): 2038–49.

Fudenberg, Geoffrey, and Maxim Imakaev. 2017. “FISH-Ing for Captured Contacts: Towards Reconciling FISH and 3C.” *Nature Methods* 14(7): 673–78.

Fukuoka, Yutaka, Hidenori Inaoka, and Isaac S. Kohane. 2004. “Inter-Species Differences of Co-Expression of Neighboring Genes in Eukaryotic Genomes.” *BMC genomics* 5(1): 4.

Fullwood, Melissa J., Mei Hui Liu, You Fu Pan, Jun Liu, Han Xu, et al. 2009. “An Oestrogen-Receptor-α-Bound Human Chromatin Interactome.” *Nature* 462(7269): 58–64.

Fullwood, Melissa J, Mei Hui Liu, You Fu Pan, Jun Liu, Xu Han, et al. 2009. “An Oestrogen Receptor α-Bound Human Chromatin Interactome.” *Nature* 462(7269): 58–64.

Gabriel, K. R. 1971. “The Biplot Graphic Display of Matrices with Application to Principal Component Analysis.” *Biometrika* 58(3): 453–67.

Gabrielsson, Britt L., Björn Carlsson, and Lena M.S. Carlsson. 2000. “Partial Genome Scale Analysis of Gene Expression in Human Adipose Tissue Using DNA Array.” *Obesity Research* 8(5): 374–84.

Gafner, Jürg, and Peter Philippsen. 1980. “The Yeast Transposon Ty1 Generates Duplications of Target DNA on Insertion.” *Nature* 286(5771): 414–18.

Gaillard, Hélène, and Andrés Aguilera. 2016. “Transcription as a Threat to Genome Integrity.” *Annual Review of Biochemistry* 85(1): 291–317.

Gall, J. G., and M. L. Pardue. 1969. “FORMATION AND DETECTION OF RNA-DNA HYBRID MOLECULES IN CYTOLOGICAL PREPARATIONS.” *Proceedings of the National Academy of Sciences* 63(2): 378–83.

Gangaraju, Vamsi K, and Blaine Bartholomew. 2007. “Mechanisms of ATP Dependent Chromatin Remodeling.” *Mutation research* 618(1–2): 3–17.

Garcia-Perez, Jose L., Thomas J. Widmann, and Ian R. Adams. 2016. “The Impact of Transposable Elements on Mammalian Development.” *Development* 143(22): 4101–14.

Garrido-Ramos, Manuel. 2017. “Satellite DNA: An Evolving Topic.” *Genes* 8(9): 230.

Gasch, A. P. et al. 2000. “Genomic Expression Programs in the Response of Yeast Cells to Environmental Changes.” *Molecular biology of the cell* 11(12): 4241–57.

Gates, Leah A. et al. 2017. “Acetylation on Histone H3 Lysine 9 Mediates a Switch from Transcription Initiation to Elongation.” *Journal of Biological Chemistry* 292(35): 14456–72.

Gayon, Jean. 2016. “De Mendel à l’épigénétique : Histoire de La Génétique.” *Comptes Rendus - Biologies* 339(7–8): 225–30.

Gemayel, Rita, Marcelo D. Vinces, Matthieu Legendre, and Kevin J. Verstrepen. 2010. “Variable Tandem Repeats Accelerate Evolution of Coding and Regulatory Sequences.” *Annual Review of Genetics* 44(1): 445–77.

Ghavifekr Fakhr, Mehrdad, Majid Farshdousti Hagh, Dariush Shanehbandi, and Behzad Baradaran. 2013. “DNA Methylation Pattern as Important Epigenetic Criterion in Cancer.” *Genetics Research International* 2013: 1–9.

Gheldof, N., T. M. Tabuchi, and J. Dekker. 2006. “The Active FMR1 Promoter Is Associated with a Large Domain of Altered Chromatin Conformation with Embedded Local Histone Modifications.” *Proceedings of the National Academy of Sciences* 103(33): 12463–68.

Gifford, Wesley D., Samuel L. Pfaff, and Todd S. MacFarlan. 2013. “Transposable Elements as Genetic Regulatory Substrates in Early Development.” *Trends in Cell Biology* 23(5): 218–26.

Gilbert, Nick et al. 2003. “Formation of Facultative Heterochromatin in the Absence of HP1.” *EMBO Journal* 22(20): 5540–50.

Gillies, Stephen D., Sherie L. Morrison, Vernon T. Oi, and Susumu Tonegawa. 1983. “A Tissue-Specific Transcription Enhancer Element Is Located in the Major Intron of a Rearranged Immunoglobulin Heavy Chain Gene.” *Cell* 33(3): 717–28.

Gitter, Anthony et al. 2009. “Backup in Gene Regulatory Networks Explains Differences between Binding and Knockout Results.” *Molecular Systems Biology* 5: 276.

Goldberg, Aaron D., C. David Allis, and Emily Bernstein. 2007. “Epigenetics: A Landscape Takes Shape.” *Cell* 128(4): 635–38.

Göndör, Anita, and Rolf Ohlsson. 2009. “Replication Timing and Epigenetic Reprogramming of Gene Expression: A Two-Way Relationship?” *Nature Reviews Genetics* 10(4): 269–76.

Gorisch, S. M. 2005. “Histone Acetylation Increases Chromatin Accessibility.” *Journal of Cell Science* 118(24): 5825–34.

Gospodinov, A. et al. 2011. “Mammalian Ino80 Mediates Double-Strand Break Repair through Its Role in DNA End Strand Resection.” *Molecular and Cellular Biology* 31(23): 4735–45.

Goudarzi, Afsaneh et al. 2016. “Dynamic Competing Histone H4 K5K8 Acetylation and Butyrylation Are Hallmarks of Highly Active Gene Promoters.” *Molecular Cell* 62(2): 169–80.

Green, Erin M. et al. 2012. “Methylation of H4 Lysines 5, 8 and 12 by Yeast Set5 Calibrates Chromatin Stress Responses.” *Nature Structural and Molecular Biology* 19(3): 361–63.

Greer, Eric L., and Yang Shi. 2012. “Histone Methylation: A Dynamic Mark in Health, Disease and Inheritance.” *Nature reviews. Genetics* 13(5): 343–57.

Grosveld, Frank, Greet Blom van Assendelft, David R. Greaves, and George Kollias. 1987. “Position-Independent, High-Level Expression of the Human β-Globin Gene in Transgenic Mice.” *Cell* 51(6): 975–85.

Groth, Anja, Walter Rocha, Alain Verreault, and Geneviève Almouzni. 2007. “Chromatin Challenges during DNA Replication and Repair.” *Cell* 128(4): 721–33.

Grüne, Tim et al. 2003. “Crystal Structure and Functional Analysis of a Nucleosome Recognition Module of the Remodeling Factor ISWI.” *Molecular Cell* 12(2): 449–60.

Guccione, Ernesto et al. 2007. “Methylation of Histone H3R2 by PRMT6 and H3K4 by an MLL Complex Are Mutually Exclusive.” *Nature* 449(7164): 933–37.

Guelen, Lars et al. 2008. “Domain Organization of Human Chromosomes Revealed by Mapping of Nuclear Lamina Interactions.” *Nature* 453(7197): 948–51.

Guenther, Matthew G. et al. 2007. “A Chromatin Landmark and Transcription Initiation at Most Promoters in Human Cells.” *Cell* 130(1): 77–88.

Guillemette, Benoit et al. 2011. “H3 Lysine 4 Is Acetylated at Active Gene Promoters and Is Regulated by H3 Lysine 4 Methylation.” *PLoS genetics* 7(3): e1001354.

Guillou, Emmanuelle et al. 2010. “Cohesin Organizes Chromatin Loops at DNA Replication Factories.” *Genes & Development* 24(24): 2812–22.

Gurley, L. R., R. A. Walters, and R. A. Tobey. 1975. “Sequential Phosphorylation of Histone Subfractions in the Chinese Hamster Cell Cycle.” *Journal of Biological Chemistry* 250(10): 3936–44.

Haenlein, Michael, and Andreas M. Kaplan. 2005. “A Beginner’s Guide to Partial Least Squares Analysis.” *Understanding Statistics* 3(4): 283–97.

Hakem, Razqallah. 2008. “DNA-Damage Repair; the Good, the Bad, and the Ugly.” *EMBO Journal* 27(4): 589–605.

Hakim, Ofir et al. 2011. “Diverse Gene Reprogramming Events Occur in the Same Spatial Clusters of Distal Regulatory Elements.” *Genome Research* 21(5): 697–706.

Hall, Ira M. et al. 2002. “Establishment and Maintenance of a Heterochromatin Domain.” *Science* 297(5590): 2232–37.

Hancks, Dustin C., and Haig H. Kazazian. 2016. “Roles for Retrotransposon Insertions in Human Disease.” *Mobile DNA* 7(1): 9.

Handel, Adam E., George C. Ebers, and Sreeram V. Ramagopalan. 2010. “Epigenetics: Molecular Mechanisms and Implications for Disease.” *Trends in Molecular Medicine* 16(1): 7–16.

Handy, Diane E., Rita Castro, and Joseph Loscalzo. 2011. “Epigenetic Modifications: Basic Mechanisms and Role in Cardiovascular Disease.” *Circulation* 123(19): 2145–56.

Harada, Yoshie et al. 1999. “Single-Molecule Imaging of RNA Polymerase-DNA Interactions in Real Time.” *Biophysical Journal* 76(2): 709–15.

Harbison, Christopher T. et al. 2004. “Transcriptional Regulatory Code of a Eukaryotic Genome.” *Nature* 431(7004): 99–104.

Harikumar, A., and E. Meshorer. 2015. “Chromatin Remodeling and Bivalent Histone Modifications in Embryonic Stem Cells.” *EMBO reports* 16(12): 1609–19.

Hark, Amy T et al. 2000. “CTCF Mediates Methylation-Sensitive Enhancer-Blocking Activity at the H19/Igf2 Locus.” *Nature* 405: 486.

Harris, Susan R. 2009. “Response from Editor-in-Chief.” *Physiotherapy Canada* 61(1): 51.

Hassan, Ahmed H, Kristen E Neely, and Jerry L Workman. 2001. “SWI / SNF Binding to Promoter Nucleosomes.” *Cell* 104: 817–27.

Hauer, Michael H. et al. 2017. “Histone Degradation in Response to DNA Damage Enhances Chromatin Dynamics and Recombination Rates.” *Nature Structural and Molecular Biology* 24(2): 99–107.

Hauhnar, Lalremruata, and H. Lalhruaitluanga. 2016. “ATP-Dependent Chromatin Remodeling Complexes.” *Science & Technology Journal* 4(1): 38–42.

Havecker, Ericka R., Xiang Gao, and Daniel F. Voytas. 2004. “The Diversity of LTR Retrotransposons.” *Genome biology* 5(6): 225.

Heard, Edith, and Wendy Bickmore. 2007. “The Ins and Outs of Gene Regulation and Chromosome Territory Organisation.” *Current Opinion in Cell Biology* 19(3): 311–16.

Hennig, Bianca P., Katja Bendrin, Yang Zhou, and Tamás Fischer. 2012. “Chd1 Chromatin Remodelers Maintain Nucleosome Organization and Repress Cryptic Transcription.” *EMBO Reports* 13(11): 997–1003.

Heo, Kyu et al. 2008. “FACT-Mediated Exchange of Histone Variant H2AX Regulated by Phosphorylation of H2AX and ADP-Ribosylation of Spt16.” *Molecular Cell* 30(1): 86–97.

Herbert, Kristina M., William J. Greenleaf, and Steven M. Block. 2008. "Single-Molecule Studies of RNA Polymerase: Motoring Along." *Annual Review of Biochemistry* 77(1): 149–76.

Heun, P. et al. 2001. "Chromosome Dynamics in the Yeast Interphase Nucleus." *Science* 294(5549): 2181–86.

Heun, Patrick, Thierry Laroche, M. K. Raghuraman, and Susan M. Gasser. 2001. "The Positioning and Dynamics of Origins of Replication in the Budding Yeast Nucleus." *Journal of Cell Biology* 152(2): 385–400.

van Heyningen, Veronica, and Wendy Bickmore. 2013. "Regulation from a Distance: Long-Range Control of Gene Expression in Development and Disease." *Philosophical Transactions of the Royal Society B: Biological Sciences* 368(1620): 20120372–20120372.

Hillenmeyer, Maureen E. et al. 2008. "The Chemical Genomic Portrait of Yeast: Uncovering a Phenotype for All Genes." *Science* 320(5874): 362–65.

Hirsh, Aaron E., Hunter B. Fraser, and Dennis P. Wall. 2005. "Adjusting for Selection on Synonymous Sites in Estimates of Evolutionary Distance." *Molecular Biology and Evolution* 22(1): 174–77.

Hohmann, P., R. A. Tobey, and L. R. Gurley. 1976. "Phosphorylation of Distinct Regions of F1 Histone. Relationship to the Cell Cycle." *Journal of Biological Chemistry* 251(12): 3685–92.

Holliday, Robin. 2006. "Epigenetics: A Historical Overview." *Epigenetics* 1(2): 76–80.

Homouz, Dirar, and Andrzej S. Kudlicki. 2013. "The 3D Organization of the Yeast Genome Correlates with Co-Expression and Reflects Functional Relations between Genes" ed. Arkady B Khodursky. *PLoS ONE* 8(1): e54699.

Horike, Shin Ichi et al. 2005. "Loss of Silent-Chromatin Looping and Impaired Imprinting of DLX5 in Rett Syndrome." *Nature Genetics* 37(1): 31–40.

House, Nealia C M, Melissa R. Koch, and Catherine H. Freudenreich. 2014. "Chromatin Modifications and DNA Repair: Beyond Double-Strand Breaks." *Frontiers in genetics* 5(SEP): 296.

Hozák, Pavel, A. Bassim Hassan, Dean A. Jackson, and Peter R. Cook. 1993. "Visualization of Replication Factories Attached to a Nucleoskeleton." *Cell* 73(2): 361–73.

Hsieh, Tsung Han S. et al. 2015. "Mapping Nucleosome Resolution Chromosome Folding in Yeast by Micro-C." *Cell* 162(1): 108–19.

Hu, Ming et al. 2012. "HiCNorm: Removing Biases in Hi-C Data via Poisson Regression." *Bioinformatics* 28(23): 3131–33.

Hu, Zhanzhi, Patrick J. Killion, and Vishwanath R. Iyer. 2007. "Genetic Reconstruction of a Functional Transcriptional Regulatory Network." *Nature Genetics* 39(5): 683–87.

Huang, S. 2012. "An MBoC Favorite: High-Resolution Whole-Genome Sequencing Reveals That Specific Chromatin Domains from Most Human Chromosomes Associate with Nucleoli." *Molecular Biology of the Cell* 23(11): 2016–2016.

Huang, Sui, and David L. Spector. 1991. "Nascent Pre-MRNA Transcripts Are Associated with Nuclear Regions Enriched in Splicing Factors." *Genes and Development* 5(12): 2288–2302.

Hublitz, Philip, Mareike Albert, and Antoine H.F.M. Peters. 2009. "Mechanisms of Transcriptional Repression by Histone Lysine Methylation." *International Journal of Developmental Biology* 53(2–3): 335–54.

Huebert, Dana J., Michael Kamal, Aisling O'Donovan, and Bradley E. Bernstein. 2006. "Genome-Wide Analysis of Histone Modifications by ChIP-on-Chip." *Methods* 40(4): 365–69.

Hug, Clemens B., and Juan M. Vaquerizas. 2018. "The Birth of the 3D Genome during Early Embryonic Development." *Trends in Genetics* 34(12): 903–14.

Huh, Iksoo, Jia Zeng, Taesung Park, and Soojin V. Yi. 2013. "DNA Methylation and Transcriptional Noise." *Epigenetics & Chromatin* 6(1): 9.

Huisinga, Kathryn L., and B. Franklin Pugh. 2004. "A Genome-Wide Housekeeping Role for TFIID and a Highly Regulated Stress-Related Role for SAGA in Saccharomyces Cerevisiae." *Molecular cell* 13(4): 573–85.

Hur, Shin Kyoung et al. 2010. "Roles of Human INO80 Chromatin Remodeling Enzyme in DNA Replication and Chromosome Segregation Suppress Genome Instability." *Cellular and Molecular Life Sciences* 67(13): 2283–96.

Hurst, Laurence D., Csaba Pál, and Martin J. Lercher. 2004. "The Evolutionary Dynamics of Eukaryotic Gene Order." *Nature Reviews Genetics* 5(4): 299–310.

Hurst, Laurence D. 2017. “It’s Easier to Get along with the Quiet Neighbours.” *Molecular Systems Biology* 13(9): 943.

Hyllus, Dawin et al. 2007. “PRMT6-Mediated Methylation of R2 in Histone H3 Antagonizes H3 K4 Trimethylation.” *Genes and Development* 21(24): 3369–80.

Hyman, Anthony A., Christoph A. Weber, and Frank Jülicher. 2014. “Liquid-Liquid Phase Separation in Biology.” *Annual Review of Cell and Developmental Biology* 30(1): 39–58.

Ibarra, Arkaitz, and Martin W. Hetzer. 2015. “Nuclear Pore Proteins and the Control of Genome Functions.” *Genes and Development* 29(4): 337–49.

Idziak, Dominika, Ewa Robaszkiewicz, and Robert Hasterok. 2015. “Spatial Distribution of Centromeres and Telomeres at Interphase Varies among Brachypodium Species.” *Journal of Experimental Botany* 66(21): 6623–34.

Ikegami, Kohta, and Jason D. Lieb. 2010. “Nucleoporins and Transcription: New Connections, New Questions.” *PLoS Genetics* 6(2): e1000861.

Imakaev, Maxim et al. 2012. “Iterative Correction of Hi-C Data Reveals Hallmarks of Chromosome Organization.” *Nature Methods* 9(10): 999–1003.

Inagaki, Takeshi, Juro Sakai, and Shingo Kajimura. 2016. “Transcriptional and Epigenetic Control of Brown and Beige Adipose Cell Fate and Function.” *Nature Reviews Molecular Cell Biology* 17(8): 480–95.

Iwafuchi-Doi, Makiko, and Kenneth S. Zaret. 2014. “Pioneer Transcription Factors in Cell Reprogramming.” *Genes and Development* 28(24): 2679–92.

Jackson, D.A., A.B. Hassan, R.J. Errington, and P.R. Cook. 1993. “Visualization of Focal Sites of Transcription within Human Nuclei.” *The EMBO Journal* 12(3): 1059–65.

Jackson, S. P. 2002. “Sensing and Repairing DNA Double-Strand Breaks.” *Carcinogenesis* 23(5): 687–96.

Jäger, Roland et al. 2015. “Capture Hi-C Identifies the Chromatin Interactome of Colorectal Cancer Risk Loci.” *Nature communications* 6(1): 6178.

Janke, Carsten et al. 2004. “A Versatile Toolbox for PCR-Based Tagging of Yeast Genes: New Fluorescent Proteins, More Markers and Promoter Substitution Cassettes.” *Yeast* 21(11): 947–62.

Jaschek, Rami, and Amos Tanay. 2009. “Spatial Clustering of Multivariate Genomic and Epigenomic Information.” In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, , 170–83.

Jeggo, Penny A., Jessica A. Downs, and Susan M. Gasser. 2017. “Chromatin Modifiers and Remodellers in DNA Repair and Signalling.” *Philosophical Transactions of the Royal Society B: Biological Sciences* 372(1731): 20160279.

Jenuwein, T., and C. D. Allis. 2001. “Translating the Histone Code.” *Science* 293(5532): 1074–80.

Jin, Bilian, Yajun Li, and Keith D. Robertson. 2011. “DNA Methylation: Superior or Subordinate in the Epigenetic Hierarchy?” *Genes and Cancer* 2(6): 607–17.

Jones, P. A., and D. Takai. 2001. “The Role of DNA Methylation in Mammalian Epigenetics.” *Science* 293(5532): 1068–70.

Jost, Daniel, and Cédric Vaillant. 2018. “Epigenomics in 3D: Importance of Long-Range Spreading and Specific Interactions in Epigenomic Maintenance.” *Nucleic Acids Research* 46(5): 2252–64.

Jost, K Laurence, Bianca Bertulat, and M Cristina Cardoso. 2012. “Heterochromatin and Gene Positioning: Inside, Outside, Any Side?” *Chromosoma* 121(6): 555–63.

Kabata, Hiroyuki et al. 1993. “Visualization of Single Molecules of RNA Polymerase Sliding along DNA.” *Science* 262(5139): 1561–63.

Kagohara, Luciane T. et al. 2018. “Epigenetic Regulation of Gene Expression in Cancer: Techniques, Resources and Analysis.” *Briefings in Functional Genomics* 17(1): 49–63.

Kaimori, Jun-Ya et al. 2016. “Histone H4 Lysine 20 Acetylation Is Associated with Gene Repression in Human Cells.” *Scientific Reports* 6(1): 24318.

Kalakonda, N. et al. 2008. “Histone H4 Lysine 20 Monomethylation Promotes Transcriptional Repression by L3MBTL1.” *Oncogene* 27(31): 4293–4304.

Kalhor, Reza et al. 2012. “Genome Architectures Revealed by Tethered Chromosome Conformation Capture and Population-

Based Modeling." *Nature Biotechnology* 30(1): 90–98.

Kanduri, Chandrasekhar et al. 2000. "Functional Association of CTCF with the Insulator Upstream of the H19 Gene Is Parent of Origin-Specific and Methylation-Sensitive." *Current Biology* 10(14): 853–56.

Karmodiya, Krishanpal et al. 2012. "H3K9 and H3K14 Acetylation Co-Occur at Many Gene Regulatory Elements, While H3K14ac Marks a Subset of Inactive Inducible Promoters in Mouse Embryonic Stem Cells." *BMC genomics* 13(1): 424.

Kayne, Paul S. et al. 1988. "Extremely Conserved Histone H4 N Terminus Is Dispensable for Growth but Essential for Repressing the Silent Mating Loci in Yeast." *Cell* 55(1): 27–39.

Kazazian H.H., Jr. 2000. "L1 Retrotransposons Shape the Mammalian Genome." *Science* 289(5482): 1152–53.

Ke, Yuwen et al. 2017. "3D Chromatin Structures of Mature Gametes and Structural Reprogramming during Mammalian Embryogenesis." *Cell* 170(2): 367-381.e20.

Kellis, Manolis, Bruce W. Birren, and Eric S. Lander. 2004. "Proof and Evolutionary Analysis of Ancient Genome Duplication in the Yeast Saccharomyces Cerevisiae." *Nature* 428(6983): 617–24.

Kenter, Amy L., Robert Wuerffel, Satyendra Kumar, and Fernando Grigera. 2013. "Genomic Architecture May Influence Recurrent Chromosomal Translocation Frequency in the Igh Locus." *Frontiers in Immunology* 4: 500.

Khaladkar, Mugdha, and Sridhar Hannenhalli. 2012. "Functional Divergence of Gene Duplicates – a Domain-Centric View." *BMC Evolutionary Biology* 12(1): 126.

Kharchenko, Peter V. et al. 2011. "Comprehensive Analysis of the Chromatin Landscape in Drosophila Melanogaster." *Nature* 471(7339): 480–86.

Kidwell, M G, J. F. Kidwell, and J. A. Sved. 1977. "Hybrid Dysgenesis in DROSOPHILA MELANOGASTER: A Syndrome of Aberrant Traits Including Mutation, Sterility and Male Recombination." *Genetics* 86(4): 813–33.

Kidwell, Margaret G. 1979. "Hybrid Dysgenesis in Drosophila Melanogaster: The Relationship between the P–M and I–R Interaction Systems." *Genetical Research* 33(3): 205–17.

Kiemer, Lars, Stefano Costa, Marius Ueffing, and Gianni Cesareni. 2007. "WI-PHI: A Weighted Yeast Interactome Enriched for Direct Physical Interactions." *Proteomics* 7(6): 932–43.

Kim, S., and W. K. Paik. 1965. "Studies on the Origin of Epsilon-N-Methyl-L-Lysine in Protein." *Journal of Biological Chemistry* 240(12): 4629–34.

Kimura, Akatsuki, and Masami Horikoshi. 2004. "Partition of Distinct Chromosomal Regions: Negotiable Border and Fixed Border." *Genes to Cells* 9(6): 499–508.

Kinner, Andrea, Wenqi Wu, Christian Staudt, and George Iliakis. 2008. "Gamma-H2AX in Recognition and Signaling of DNA Double-Strand Breaks in the Context of Chromatin." *Nucleic acids research* 36(17): 5678–94.

Kireeva, Natashe et al. 2004. "Visualization of Early Chromosome Condensation: A Hierarchical Folding, Axial Glue Model of Chromosome Structure." *Journal of Cell Biology* 166(6): 775–85.

Kirmizis, Antonis et al. 2007. "Arginine Methylation at Histone H3R2 Controls Deposition of H3K4 Trimethylation." *Nature* 449(7164): 928–32.

Kobayashi, Junya et al. 2004. "NBS1 and Its Functional Role in the DNA Damage Response." *DNA Repair* 3(8–9): 855–61.

Koch, Christoph M. et al. 2007. "The Landscape of Histone Modifications across 1% of the Human Genome in Five Human Cell Lines." *Genome Research* 17(6): 691–707.

de Koning, A. P Jason et al. 2011. "Repetitive Elements May Comprise Over Two-Thirds of the Human Genome" ed. Gregory P. Copenhaver. *PLoS Genetics* 7(12): e1002384.

Kornberg, Roger D. 1974. "Chromatin Structure: A Repeating Unit of Histones and DNA." *Science* 184(4139): 868–71.

Kosak, Steven T. et al. 2002. "Subnuclear Compartmentalization of Immunoglobulin Loci during Lymphocyte Development." *Science* 296(5565): 158–62.

Kosak, Steven T., and Mark Groudine. 2002. "The Undiscovered Country: Chromosome Territories and the Organization of Transcription." *Developmental Cell* 2(6): 690–92.

Kouzarides, Tony. 2007. "Chromatin Modifications and Their Function." *Cell* 128(4): 693–705.

Krefting, Jan, Miguel A. Andrade-Navarro, and Jonas Ibn-Salem. 2018. "Evolutionary Stability of Topologically Associating

Domains Is Associated with Conserved Gene Regulation." *BMC biology* 16(1): 87.

Krogan, Nevan J. et al. 2003. "A Snf2 Family ATPase Complex Required for Recruitment of the Histone H2A Variant Htz1." *Molecular Cell* 12(6): 1565–76.

Kubik, Slawomir, Maria Jessica Bruzzone, and David Shore. 2017. "TFIID or Not TFIID, a Continuing Transcriptional SAGA." *The EMBO Journal* 36(3): 248–49.

Kulis, Marta, and Manel Esteller. 2010. "DNA Methylation and Cancer." In *Advances in Genetics*, Academic Press, 27–56.

Kurdistani, Siavash K., Daniel Robyr, Saeed Tavazoie, and Michael Grunstein. 2002. "Genome-Wide Binding Map of the Histone Deacetylase Rpd3 in Yeast." *Nature Genetics* 31(3): 248–54.

Kurdistani, Siavash K., Saeed Tavazoie, and Michael Grunstein. 2004. "Mapping Global Histone Acetylation Patterns to Gene Expression." *Cell* 117(6): 721–33.

Kurukuti, S. et al. 2006. "CTCF Binding at the H19 Imprinting Control Region Mediates Maternally Inherited Higher-Order Chromatin Conformation to Restrict Enhancer Access to Igf2." *Proceedings of the National Academy of Sciences* 103(28): 10684–89.

Kurz, Anette et al. 1996. "Active and Inactive Genes Localize Preferentially in the Periphery of Chromosome Territories." *Journal of Cell Biology* 135(5): 1195–1205.

Kusch, Thomas. 2012. "Histone H3 Lysine 4 Methylation Revisited." *Transcription* 3(6): 310–14.

Kustatscher, Georg, Piotr Grabowski, and Juri Rappsilber. 2017. "Pervasive Coexpression of Spatially Proximal Genes Is Buffered at the Protein Level." *Molecular systems biology* 13(8): 937.

Lachner, M. et al. 2001. "Methylation of Histone H3 Lysine 9 Creates a Binding Site for HP1 Proteins." *Nature* 410(6824): 116–20.

Ladurner, Andreas G., Carla Inouye, Rajan Jain, and Robert Tjian. 2003. "Bromodomains Mediate an Acetyl-Histone Encoded Antisilencing Function at Heterochromatin Boundaries." *Molecular Cell* 11(2): 365–76.

Lajoie, Mathieu, Denis Bertrand, and Nadia El-Mabrouk. 2010. "Inferring the Evolutionary History of Gene Clusters from Phylogenetic and Gene Order Data." *Molecular Biology and Evolution* 27(4): 761–72.

Lander, Eric S. et al. 2001. "Initial Sequencing and Analysis of the Human Genome." *Nature* 409(6822): 860–921.

Lang, Gregory I., and Andrew W. Murray. 2011. "Mutation Rates across Budding Yeast Chromosome VI Are Correlated with Replication Timing." *Genome Biology and Evolution* 3(1): 799–811.

Lans, Hannes, Jurgen A. Marteijn, and Wim Vermeulen. 2012. "ATP-Dependent Chromatin Remodeling in the DNA-Damage Response." *Epigenetics & chromatin* 5(1): 4.

Larson, Adam G. et al. 2017. "Liquid Droplet Formation by HP1α Suggests a Role for Phase Separation in Heterochromatin." *Nature* 547(7662): 236–40.

Lawrence, Jeanne B., and Christine M. Clemson. 2008. "Gene Associations: True Romance or Chance Meeting in a Nuclear Neighborhood?" *Journal of Cell Biology* 182(6): 1035–38.

Lawrence, Moyra, Sylvain Daujat, and Robert Schneider. 2016. "Lateral Thinking: How Histone Modifications Regulate Gene Expression." *Trends in Genetics* 32(1): 42–56.

Le, Tung B K, Maxim V Imakaev, Leonid A Mirny, and Michael T Laub. 2013. "High-Resolution Mapping of the Spatial Organization of a Bacterial Chromosome." *Science (New York, N.Y.)* 342(6159): 731–34.

Lee, Daniel Y., Jeffrey J. Hayes, Dmitry Pruss, and Alan P. Wolffe. 1993. "A Positive Role for Histone Acetylation in Transcription Factor Access to Nucleosomal DNA." *Cell* 72(1): 73–84.

Lee, Han-Sae et al. 2014. "Stabilization and Targeting of INO80 to Replication Forks by BAP1 during Normal DNA Synthesis." *Nature Communications* 5(1): 5128.

Lee, Heather J, Timothy A Hore, and Wolf Reik. 2014. "Reprogramming the Methylome: Erasing Memory and Creating Diversity." *Cell Stem Cell* 14(6): 710–19.

Lee, Jennifer M., and Erik L.L. Sonnhammer. 2003. "Genomic Gene Clustering Analysis of Pathways in Eukaryotes." *Genome Research* 13(5): 875–82.

Lee, Ji-Hoon, and Tanya T Paull. 2004. "Direct Activation of the ATM Protein Kinase by the Mre11/Rad50/Nbs1 Complex."

Science 304(5667): 93 LP – 96.

Lee, Ji-Hoon, and Tanya T Paull. 2005. "ATM Activation by DNA Double-Strand Breaks Through the Mre11-Rad50-Nbs1 Complex." *Science* 308(5721): 551 LP – 554.

Lee, Jung Shin, Edwin Smith, and Ali Shilatifard. 2010. "The Language of Histone Crosstalk." *Cell* 142(5): 682–85.

Lee, Kenneth K., and Jerry L. Workman. 2007. "Histone Acetyltransferase Complexes: One Size Doesn't Fit All." *Nature Reviews Molecular Cell Biology* 8(4): 284–95.

Lee, Tong Ihn et al. 2000. "Redundant Roles for the TFIID and SAGA Complexes in Global Transcription." *Nature* 405(6787): 701–4.

Lee, Tong Ihn et al. 2002. "Transcriptional Regulatory Networks in Saccharomyces Cerevisiae." *Science* 298(5594): 799–804.

Leeuw, Jan de. 2007. "The Pls Package: Principal Component and Partial Least Squares Regression in R." *Journal of Statistical Software* 18(2): 1–23.

Lei, Quan et al. 2016. "Evolutionary Insights into RNA Trans-Splicing in Vertebrates." *Genome Biology and Evolution* 8(3): 562–77.

Lenay, Charles. 2000. "Hugo De Vries: From the Theory of Intracellular Pangenesis to the Rediscovery of Mendel." *Comptes Rendus de l'Academie des Sciences - Serie III* 323(12): 1053–60.

Lenstra, Tineke L et al. 2011. "The Specificity and Topology of Chromatin Interaction Pathways in Yeast." *Molecular cell* 42(4): 536–49.

Lenstra, Tineke L, and Frank C P Holstege. 2012. "The Discrepancy between Chromatin Factor Location and Effect." *Nucleus* 3(3): 213–19.

Leonhardt, Heinrich et al. 2000. "Dynamics of DNA Replication Factories in Living Cells." *Journal of Cell Biology* 149(2): 271–79.

Li, Bing et al. 2007. "Infrequently Transcribed Long Genes Depend on the Set2/Rpd3S Pathway for Accurate Transcription." *Genes and Development* 21(11): 1422–30.

Li, Bing, Michael Carey, and Jerry L. Workman. 2007. "The Role of Chromatin during Transcription." *Cell* 128(4): 707–19.

Li, Congru et al. 2018. "DNA Methylation Reprogramming of Functional Elements during Mammalian Embryonic Development." *Cell Discovery* 4(1): 41.

Li, Guohong, and Danny Reinberg. 2011. "Chromatin Higher-Order Structures and Gene Regulation." *Current Opinion in Genetics and Development* 21(2): 175–86.

Li, Guoliang et al. 2012. "Extensive Promoter-Centered Chromatin Interactions Provide a Topological Basis for Transcription Regulation." *Cell* 148(1–2): 84–98.

Li, J. et al. 2010. "Exploiting the Determinants of Stochastic Gene Expression in Saccharomyces Cerevisiae for Genome-Wide Prediction of Expression Noise." *Proceedings of the National Academy of Sciences* 107(23): 10472–77.

Liang, G. et al. 2004. "Distinct Localization of Histone H3 Acetylation and H3-K4 Methylation to the Transcription Start Sites in the Human Genome." *Proceedings of the National Academy of Sciences* 101(19): 7357–62.

Lieb, Jason D., Xiaole Liu, David Botstein, and Patrick O. Brown. 2001. "Promoter-Specific Binding of Rap1 Revealed by Genome-Wide Maps of Protein-DNA Association." *Nature Genetics* 28(4): 327–34.

Lieber, Michael R. 2010. "The Mechanism of Double-Strand DNA Break Repair by the Nonhomologous DNA End-Joining Pathway." *Annual Review of Biochemistry* 79(1): 181–211.

Lieberman-Aiden, Erez et al. 2009. "Comprehensive Mapping of Long Range Interactions Reveals Folding Principles of the Human Genome." *Science (New York, N.Y.)* 326(5950): 289–93.

Lim, Derek H K, and Eamonn R Maher. 2010. "DNA Methylation: A Form of Epigenetic Control of Gene Expression." *The Obstetrician & Gynaecologist* 12(1): 37–42.

Lipford, J. Russell, and Stephen P. Bell. 2001. "Nucleosomes Positioned by ORC Facilitate the Initiation of DNA Replication." *Molecular Cell* 7(1): 21–30.

Lisby, Michael, Adriana Antúnez de Mayolo, Uffe H. Mortensen, and Rodney Rothstein. 2003. "Cell Cycle-Regulated Centers of DNA Double-Strand Break Repair." *Cell cycle (Georgetown, Tex.)* 2(5): 479–83.

Lisby, Michael, Uffe H. Mortensen, and Rodney Rothstein. 2003. “Colocalization of Multiple DNA Double-Strand Breaks at a Single Rad52 Repair Center.” *Nature Cell Biology* 5(6): 572–77.

Liu, X. Shirley. 2007. “Getting Started in Tiling Microarray Analysis.” *PLoS Computational Biology* 3(10): e183.

Loidl, P. 1994. “Histone Acetylation: Facts and Questions.” *Chromosoma* 103(7): 441–49.

López-Flores, I., and Manuel A. Garrido-Ramos. 2012. “The Repetitive DNA Content of Eukaryotic Genomes.” In *Repetitive DNA*, , 1–28.

Di Lorenzo, Alessandra, and Mark T. Bedford. 2011. “Histone Arginine Methylation.” *FEBS Letters* 585(13): 2024–31.

Luijsterburg, Martijn S., and Haico Van Attikum. 2011. “Chromatin and the DNA Damage Response: The Cancer Connection.” *Molecular Oncology* 5(4): 349–67.

Luning Prak, Eline T., and Haig H. Kazazian. 2000. “Mobile Elements and the Human Genome.” *Nature Reviews Genetics* 1(2): 134–44.

Lunyak, Victoria V. et al. 2007. “Developmentally Regulated Activation of a SINE B2 Repeat as a Domain Boundary in Organogenesis.” *Science* 317(5835): 248–51.

Lusser, Alexandra, and James T. Kadonaga. 2003. “Chromatin Remodeling by ATP-Dependent Molecular Machines.” *BioEssays* 25(12): 1192–1200.

Lyon, Mary F. 1961. “Gene Action in the X-Chromosome of the Mouse (Mus Musculus L.).” *Nature* 190(4773): 372–73.

Lyttle, T. 1991. “Segregation Distorters.” *Annual Review of Genetics* 25(1): 511–57.

Ma, Wenxiu et al. 2014. “Fine-Scale Chromatin Interaction Maps Reveal the Cis-Regulatory Landscape of Human LincRNA Genes.” *Nature Methods* 12(1): 71–78.

MacAlpine, David M., and Geneviève Almouzni. 2013. “Chromatin and DNA Replication.” *Cold Spring Harbor perspectives in biology* 5(8): a010207.

Mahmoudabadi, Gita et al. 2013. “Intrinsically Disordered Proteins and Conformational Noise: Implications in Cancer.” *Cell cycle (Georgetown, Tex.)* 12(1): 26–31.

Mahy, Nicola L. et al. 2002. “Spatial Organization of Active and Inactive Genes and Noncoding DNA within Chromosome Territories.” *Journal of Cell Biology* 157(4): 579–89.

Mahy, Nicola L., Paul E. Perry, and Wendy A. Bickmore. 2002. “Gene Density and Transcription Influence the Localization of Chromatin Outside of Chromosome Territories Detectable by FISH.” *Journal of Cell Biology* 159(5): 753–63.

Maile, Tobias et al. 2004. “TAF1 Activates Transcription by Phosphorylation of Serine 33 in Histone H2B.” *Science* 304(5673): 1010–14.

Margaritis, Thanasis et al. 2012. “Two Distinct Repressive Mechanisms for Histone 3 Lysine 4 Methylation through Promoting 3′-End Antisense Transcription” ed. Hiten D. Madhani. *PLoS Genetics* 8(9): e1002952.

Marmorstein, Ronen. 2001. “Structure of Histone Deacetylases: Insights into Substrate Recognition and Catalysis.” *Structure* 9(12): 1127–33.

Marshall, Wallace F. 2002. “Order and Disorder in the Nucleus.” *Current Biology* 12(5): R185–92.

Martin, Paul et al. 2015. “Capture Hi-C Reveals Novel Candidate Genes and Complex Long-Range Interactions with Related Autoimmune Risk Loci.” *Nature communications* 6(1): 10069.

Mayran, Alexandre, and Jacques Drouin. 2018. “Pioneer Transcription Factors Shape the Epigenetic Landscape.” *Journal of Biological Chemistry* 293(36): 13795–804.

McClintock, B. 1956. “Controlling Elements and the Gene.” *Cold Spring Harbor Symposia on Quantitative Biology* 21: 197–216.

McCullagh, Emma, Anupama Seshan, Hana El-Samad, and Hiten D. Madhani. 2010. “Coordinate Control of Gene Expression Noise and Interchromosomal Interactions in a MAP Kinase Pathway.” *Nature Cell Biology* 12(10): 954–62.

McCune, Heather J. et al. 2008. “The Temporal Program of Chromosome Replication: Genomewide Replication in Clb5Δ Saccharomyces Cerevisiae.” *Genetics* 180(4): 1833–47.

McGrath, James, and Davor Solter. 1984. “Completion of Mouse Embryogenesis Requires Both the Maternal and Paternal Genomes.” *Cell* 37(1): 179–83.

Meaburn, Karen J., Tom Misteli, and Evi Soutoglou. 2007. "Spatial Genome Organization in the Formation of Chromosomal Translocations." *Seminars in Cancer Biology* 17(1): 80–90.

Medstrand, Patrik, Louie N. Van De Lagemaat, and Dixie L. Mager. 2002. "Retroelement Distributions in the Human Genome: Variations Associated with Age and Proximity to Genes." *Genome Research* 12(10): 1483–95.

Mellor, Jane. 2005. "The Dynamics of Chromatin Remodeling at Promoters." *Molecular Cell* 19(2): 147–57.

Mellor, Jane, and Antonin Morillon. 2004. "ISWI Complexes in Saccharomyces Cerevisiae." *Biochimica et Biophysica Acta - Gene Structure and Expression* 1677(1–3): 100–112.

Messerschmidt, Daniel M., Barbara B. Knowles, and Davor Solter. 2014. "DNA Methylation Dynamics during Epigenetic Reprogramming in the Germline and Preimplantation Embryos." *Genes and Development* 28(8): 812–28.

Meuleman, Wouter et al. 2013. "Constitutive Nuclear Lamina-Genome Interactions Are Highly Conserved and Associated with A/T-Rich Sequence." *Genome Research* 23(2): 270–80.

Michalak, Pawel. 2008. "Coexpression, Coregulation, and Cofunctionality of Neighboring Genes in Eukaryotic Genomes." *Genomics* 91(3): 243–48.

Michel, Uwe. 2002. "Non-Coding Ribonucleic Acids - A Class of Their Own?" *International Review of Cytology* 218(Pol II): 143–219.

Mifsud, Borbala et al. 2015. "Mapping Long-Range Promoter Contacts in Human Cells with High-Resolution Capture Hi-C." *Nature Genetics* 47(6): 598–606.

Mikkelsen, Tarjei S. et al. 2007. "Genome-Wide Maps of Chromatin State in Pluripotent and Lineage-Committed Cells." *Nature* 448(7153): 553–60.

Milne, Thomas A., Keji Zhao, and Jay L. Hess. 2009. "Chromatin Immunoprecipitation (ChIP) for Analysis of Histone Modifications and Chromatin-Associated Proteins." In *Methods in Molecular Biology (Clifton, N.J.)*, , 409–23.

Mishra, Arpit, and R. David Hawkins. 2017. "Three-Dimensional Genome Architecture and Emerging Technologies: Looping in Disease." *Genome Medicine* 9(1): 87.

Misteli, Tom, and Evi Soutoglou. 2009. "The Emerging Role of Nuclear Architecture in DNA Repair and Genome Maintenance." *Nature Reviews Molecular Cell Biology* 10(4): 243–54.

Mita, Paolo, and Jef D. Boeke. 2016. "How Retrotransposons Shape Genome Regulation." *Current Opinion in Genetics and Development* 37: 90–100.

Mitchell, Jennifer A., and Peter Fraser. 2008. "Transcription Factories Are Nuclear Subcompartments That Remain in the Absence of Transcription." *Genes and Development* 22(1): 20–25.

Mizuguchi, Takeshi et al. 2014. "Cohesin-Dependent Globules and Heterochromatin Shape 3D Genome Architecture in S. Pombe." *Nature* 516(7531): 432–35.

Mohrmann, Lisette, and C. Peter Verrijzer. 2005. "Composition and Functional Specificity of SWI2/SNF2 Class Chromatin Remodeling Complexes." *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression* 1681(2–3): 59–73.

Moindrot, Benoit et al. 2012. "3D Chromatin Conformation Correlates with Replication Timing and Is Conserved in Resting Cells." *Nucleic Acids Research* 40(19): 9470–81.

Moltó, Eduardo, Almudena Fernández, and Lluís Montoliu. 2009. "Boundaries in Vertebrate Genomes: Different Solutions to Adequately Insulate Gene Expression Domains." *Briefings in Functional Genomics and Proteomics* 8(4): 283–96.

Montavon, Thomas, and Denis Duboule. 2013. "Chromatin Organization and Global Regulation of Hox Gene Clusters." *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 368(1620): 20120367.

Moore, John D., Oya Yazgan, Yeganeh Ataian, and Jocelyn E. Krebs. 2007. "Diverse Roles for Histone H2A Modifications in DNA Damage Response Pathways in Yeast." *Genetics* 176(1): 15–25.

Morris, Stephanie A. et al. 2007. "Identification of Histone H3 Lysine 36 Acetylation as a Highly Conserved Histone Modification." *Journal of Biological Chemistry* 282(10): 7632–40.

Morrison, Ashby J., and Xuetong Shen. 2005. "DNA Repair in the Context of Chromatin." *Cell Cycle* 4(4): 568–71.

Chromatin Remodelling beyond Transcription. 2009. "Chromatin Remodelling beyond Transcription: The INO80 and SWR1 Complexes." *Nature Reviews Molecular Cell Biology* 10(6): 373–84.

Mouchiroud, Dominique et al. 1991. “The Distribution of Genes in the Human Genome.” *Gene* 100(C): 181–87.

Muers, Mary. 2012. “Bivalency Buffer Makes Pluripotency Connections.” *Nature Reviews Genetics* 13: 826.

Muller, H J, and E Altenburg. 1930. “The Frequency of Translocations Produced by X-Rays in Drosophila.” *Genetics* 15(4): 283–311.

Murray, K. 1964. “The Occurrence of ϵ-N-Methyl Lysine in Histones.” *Biochemistry* 3(1): 10–15.

Myers, Fiona A. et al. 2003. “Acetylation of Histone H2B Mirrors That of H4 and H3 at the Chicken β-Globin Locus but Not at Housekeeping Genes.” *Journal of Biological Chemistry* 278(38): 36315–22.

Nagano, Takashi et al. 2013. “Single-Cell Hi-C Reveals Cell-to-Cell Variability in Chromosome Structure.” *Nature* 502(7469): 59–64.

Nagano, Takashi et al. 2017. “Cell-Cycle Dynamics of Chromosomal Organization at Single-Cell Resolution.” *Nature* 547(7661): 61–67.

Nakamura, Hiromu, Toshiteru Morita, and Chikako Sato. 1986. “Structural Organizations of Replicon Domains during DNA Synthetic Phase in the Mammalian Nucleus.” *Experimental Cell Research* 165(2): 291–97.

Nakayama, J. et al. 2001. “Role of Histone H3 Lysine 9 Methylation in Epigenetic Control of Heterochromatin Assembly.” *Science* 292(5514): 110–13.

Nakayasu, H. 2004. “Mapping Replicational Sites in the Eucaryotic Cell Nucleus.” *The Journal of Cell Biology* 108(1): 1–11.

Narlikar, Geeta J J., Ramasubramanian Sundaramoorthy, and Tom Owen-Hughes. 2013. “Mechanisms and Functions of ATP-Dependent Chromatin-Remodeling Enzymes.” *Cell* 154(3): 490–503.

Nathan, Dafna et al. 2006. “Histone Sumoylation Is a Negative Regulator in Saccharomyces Cerevisiae and Shows Dynamic Interplay with Positive-Acting Histone Modifications.” *Genes and Development* 20(8): 966–76.

Naumova, N., and J. Dekker. 2010. “Integrating One-Dimensional and Three-Dimensional Maps of Genomes.” *Journal of Cell Science* 123(12): 1979–88.

Navarro, Caryn. 2017. “The Mobile World of Transposable Elements.” *Trends in Genetics* 33(11): 771–72.

Naveh-Many, T., and H. Cedar. 1981. “Active Gene Sequences Are Undermethylated.” *Proceedings of the National Academy of Sciences* 78(7): 4246–50.

Neidhart, Michel. 2015. “DNA Methylation and Viral Infections.” In *DNA Methylation and Complex Human Disease*, Academic Press, 81–102.

Neil, Helen et al. 2009. “Widespread Bidirectional Promoters Are the Major Source of Cryptic Transcripts in Yeast.” *Nature* 457(7232): 1038–42.

Newman, John R.S. et al. 2006. “Single-Cell Proteomic Analysis of S. Cerevisiae Reveals the Architecture of Biological Noise.” *Nature* 441(7095): 840–46.

Ng, Huck Hui, François Robert, Richard A. Young, and Kevin Struhl. 2002. “Genome-Wide Location and Regulated Recruitment of the RSC Nucleosome-Remodeling Complex.” *Genes and Development* 16(7): 806–19.

Nishibuchi, Gohei, and Jérôme Déjardin. 2017. “The Molecular Basis of the Organization of Repetitive DNA-Containing Constitutive Heterochromatin in Mammals.” *Chromosome Research* 25(1): 77–87.

Nozaki, Tadasu et al. 2017. “Dynamic Organization of Chromatin Domains Revealed by Super-Resolution Live-Cell Imaging.” *Molecular Cell* 67(2): 282-293.e7.

Ohno, S., W. D. Kaplan, and R. Kinosita. 1959. “Formation of the Sex Chromatin by a Single X-Chromosome in Liver Cells of Rattus Norvegicus.” *Experimental Cell Research* 18(2): 415–18.

Ohno, S. 1972. “So Much ‘Junk’ DNA in Our Genome.” *Brookhaven symposia in biology* 23: 366–70.

Okada, Norihiro. 1991. “SINEs: Short Interspersed Repeated Elements of the Eukaryotic Genome.” *Trends in Ecology and Evolution* 6(11): 358–61.

Olins, Ada L., and Donald E. Olins. 1974. “Spheroid Chromatin Units (v Bodies).” *Science* 183(4122): 330–32.

Ord, MG, and LA Stocken. 2015. “Metabolic Properties of Histones from Rat Liver and Thymus Gland.” *Biochemical Journal* 98(3): 888–97.

Osborne, Cameron S. et al. 2004. “Active Genes Dynamically Colocalize to Shared Sites of Ongoing Transcription.” *Nature*

Genetics 36(10): 1065–71.

Ou, Horng D. et al. 2017. "ChromEMT: Visualizing 3D Chromatin Structure and Compaction in Interphase and Mitotic Cells." *Science* 357(6349): eaag0025.

Oudet, P., M. Gross-Bellard, and P. Chambon. 1975. "Electron Microscopic and Biochemical Evidence That Chromatin Structure Is a Repeating Unit." *Cell* 4(4): 281–300.

Paik, Woon Ki, David C. Paik, and Sangduk Kim. 2007. "Historical Review: The Field of Protein Methylation." *Trends in Biochemical Sciences* 32(3): 146–52.

Pál, Csaba, and Laurence D. Hurst. 2003. "Evidence for Co-Evolution of Gene Order and Recombination Rate." *Nature Genetics* 33(3): 392–95.

Palstra, Robert Jan et al. 2003. "The β-Globin Nuclear Compartment in Development and Erythroid Differentiation." *Nature Genetics* 35(2): 190–94.

Pan, Xuewen et al. 2007. "DSLAM Analysis of Genome-Wide Genetic Interactions in Saccharomyces Cerevisiae." *Methods* 41(2): 206–21.

Papagiannakis, Alexandros, Bastian Niebel, Ernst C. Wit, and Matthias Heinemann. 2017. "Autonomous Metabolic Oscillations Robustly Gate the Early and Late Cell Cycle." *Molecular Cell*.

Papamichos-Chronakis, Manolis, Shinya Watanabe, Oliver J. Rando, and Craig L. Peterson. 2011. "Global Regulation of H2A.Z Localization by the INO80 Chromatin-Remodeling Enzyme Is Essential for Genome Integrity." *Cell* 144(2): 200–213.

Papantonis, Argyris, and Peter R. Cook. 2013. "Transcription Factories: Genome Organization and Gene Regulation." *Chemical Reviews* 113(11): 8683–8705.

Parada, Luis A., Philip G. McQueen, Peter J. Munson, and Tom Misteli. 2002. "Conservation of Relative Chromosome Positioning in Normal and Cancer Cells." *Current Biology* 12(19): 1692–97.

Parada, Luis A, Philip G McQueen, and Tom Misteli. 2004. "Tissue-Specific Spatial Organization of Genomes." *Genome biology* 5(7): R44.

Park, Chungoo, Wenfeng Qian, and Jianzhi Zhang. 2012. "Genomic Evidence for Elevated Mutation Rates in Highly Expressed Genes." *EMBO Reports* 13(12): 1123–29.

Passarge, E. 1979. "Emil Heitz and the Concept of Heterochromatin: Longitudinal Chromosome Differentiation Was Recognized Fifty Years Ago." *American journal of human genetics* 31(2): 106–15.

Patrinos, George P. et al. 2004. "Multiple Interactions between Regulatory Regions Are Required to Stabilize an Active Chromatin Hub." *Genes and Development* 18(12): 1495–1509.

Paulsen, Martina, and Anne C Ferguson-Smith. 2001. "DNA Methylation in Genomic Imprinting, Development, and Disease." *The Journal of Pathology* 195(1): 97–110.

Peric-Hupkes, Daan et al. 2010. "Molecular Maps of the Reorganization of Genome-Nuclear Lamina Interactions during Differentiation." *Molecular Cell* 38(4): 603–13.

Peterson, Daniel G., Susan R. Wessler, and Andrew H. Paterson. 2002. "Efficient Capture of Unique Sequences from Eukaryotic Genomes." *Trends in Genetics* 18(11): 547–50.

Petty, Emily, and Lorraine Pillus. 2013. "Balancing Chromatin Remodeling and Histone Modifications in Transcription." *Trends in Genetics* 29(11): 621–29.

Pham, Hung et al. 2007. "Modeling the Regulatory Network of Histone Acetylation in Saccharomyces Cerevisiae." *Molecular systems biology* 3: 153.

Phatnani, Hemali P., and Arno L. Greenleaf. 2006. "Phosphorylation and Functions of the RNA Polymerase II CTD." *Genes and Development* 20(21): 2922–36.

Phillips, Jennifer E., and Victor G. Corces. 2009. "CTCF: Master Weaver of the Genome." *Cell* 137(7): 1194–1211.

Pickersgill, Helen et al. 2006. "Characterization of the Drosophila Melanogaster Genome at the Nuclear Lamina." *Nature Genetics* 38(9): 1005–14.

Pimpinelli, Sergio, and Barbara T. Wakimoto. 2003. "Expanding the Boundaries of Heterochromatin." *Genetica* 117(2–3): 111–

16.

Platt, Roy N., Michael W. Vandewege, and David A. Ray. 2018. “Mammalian Transposable Elements and Their Impacts on Genome Evolution.” *Chromosome Research* 26(1–2): 25–43.

Pokholok, Dmitry K. et al. 2005. “Genome-Wide Map of Nucleosome Acetylation and Methylation in Yeast.” *Cell* 122(4): 517–27.

Politz, Joan C. Ritland, David Scalzo, and Mark Groudine. 2013. “Something Silent This Way Forms: The Functional Organization of the Repressive Nuclear Compartment.” *Annual Review of Cell and Developmental Biology* 29(1): 241–70.

Poot, Raymond A. et al. 2004. “The Williams Syndrome Transcription Factor Interacts with PCNA to Target Chromatic Remodelling by ISWI to Replication Foci.” *Nature Cell Biology* 6(12): 1236–44.

Poot, Raymond A. et al. 2005. “Chromatin Remodeling by WSTF-ISWI at the Replication Site: Opening a Window of Opportunity for Epigenetic Inheritance?” *Cell Cycle* 4(4): 543–46.

Pope, Benjamin D. et al. 2014. “Topologically Associating Domains Are Stable Units of Replication-Timing Regulation.” *Nature* 515(7527): 402–5.

Poss, Zachary C., Christopher C. Ebmeier, and Dylan J. Taatjes. 2013. “The Mediator Complex and Transcription Regulation.” *Critical Reviews in Biochemistry and Molecular Biology* 48(6): 575–608.

Pray-Grant, Marilyn G. et al. 2005. “Chd1 Chromodomain Links Histone H3 Methylation with SAGA- and SLIK-Dependent Acetylation.” *Nature* 433(7024): 434–38.

Price, Brendan D., and Alan D. D’Andrea. 2013a. “Chromatin Remodeling at DNA Double-Strand Breaks.” *Cell* 152(6): 1344–54.

Price, Brendan D, and Alan D D’Andrea. 2013b. “Chromatin Remodeling at DNA Double Strand Breaks.” *Cell* 152(6): 1344–54.

Pushparajan, J., and M. Sivasankaran Nair. 2015. “Synthesis of Some Organometallic Complexes of Ti(IV)/Zr(IV) Involving Sulfur Containing Schiff Base Ligands: Spectral, Electrochemical, Surface Morphology, and Antimicrobial Studies.” *Synthesis and Reactivity in Inorganic, Metal-Organic and Nano-Metal Chemistry* 45(10): 1549–57.

Quina, A. S., M. Buschbeck, and L. Di Croce. 2006. “Chromatin Structure and Epigenetics.” *Biochemical Pharmacology* 72(11): 1563–69.

Quinodoz, Sofia A. et al. 2018. “Higher-Order Inter-Chromosomal Hubs Shape 3D Genome Organization in the Nucleus.” *Cell* 174(3): 744-757.e24.

Raab, Jesse R. et al. 2012. “Human TRNA Genes Function as Chromatin Insulators.” *EMBO Journal* 31(2): 330–50.

Raghuraman, M. K. et al. 2001. “Replication Dynamics of the Yeast Genome.” *Science* 294(5540): 115–21.

Ragoczy, Tobias et al. 2006. “The Locus Control Region Is Required for Association of the Murine β-Globin Locus with Engaged Transcription Factories during Erythroid Maturation.” *Genes and Development* 20(11): 1447–57.

Rajapakse, I. et al. 2009. “The Emergence of Lineage-Specific Chromosomal Topologies from Coordinate Gene Regulation.” *Proceedings of the National Academy of Sciences* 106(16): 6679–84.

Ram, Oren et al. 2011. “Combinatorial Patterning of Chromatin Regulators Uncovered by Genome-Wide Location Analysis in Human Cells.” *Cell* 147(7): 1628–39.

Ramani, Vijay et al. 2017. “Massively Multiplex Single-Cell Hi-C.” *Nature methods* 14(3): 263–66.

Ramani, Vijay, Jay Shendure, and Zhijun Duan. 2016. “Understanding Spatial Genome Organization: Methods and Insights.” *Genomics, Proteomics and Bioinformatics* 14(1): 7–20.

Rando, Oliver J., and Howard Y. Chang. 2009. “Genome-Wide Views of Chromatin Structure.” *Annual Review of Biochemistry* 78(1): 245–71.

Rao, Suhas S.P. et al. 2017. “Cohesin Loss Eliminates All Loop Domains.” *Cell* 171(2): 305-320.e24.

Rao, Suhas S P et al. 2014. “A Three-Dimensional Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping.” *Cell* 159(7): 1665–80.

Rattner, J. B., and C. C. Lin. 1985. “Radial Loops and Helical Coils Coexist in Metaphase Chromosomes.” *Cell* 42(1): 291–96.

Razin, A, and H Cedar. 1977. “Distribution of 5-Methylcytosine in Chromatin.” *Proceedings of the National Academy of Sciences of the United States of America* 74(7): 2725–28.

Razin, Aharon, and Arthur D. Riggs. 1980. “DNA Methylation and Gene Function.” *Science* 210(4470): 604–10.

Redolfi, Josef et al. 2019. “DamC Reveals Principles of Chromatin Folding in Vivo without Crosslinking and Ligation.” *Nature Structural & Molecular Biology* 26(6): 471–80.

Redon, Christophe et al. 2003. “Yeast Histone 2A Serine 129 Is Essential for the Efficient Repair of Checkpoint-Blind DNA Damage.” *EMBO Reports* 4(7): 678–84.

Reichenzeller, Michaela, Annett Burzlaff, Peter Lichter, and Harald Herrmann. 2000. “In Vivo Observation of a Nuclear Channel-like System: Evidence for a Distinct Interchromosomal Domain Compartment in Interphase Cells.” *Journal of Structural Biology* 129(2–3): 175–85.

Reilly, M. T. et al. 2013. “The Role of Transposable Elements in Health and Diseases of the Central Nervous System.” *Journal of Neuroscience* 33(45): 17577–86.

Richard, G.-F., A. Kerrest, and B. Dujon. 2008. “Comparative Genomics and Molecular Dynamics of DNA Repeats in Eukaryotes.” *Microbiology and Molecular Biology Reviews* 72(4): 686–727.

Robert, François et al. 2004. “Global Position and Recruitment of HATs and HDACs in the Yeast Genome.” *Molecular Cell* 16(2): 199–209.

Robinson, P. J. J., L. Fairall, V. A. T. Huynh, and D. Rhodes. 2006. “EM Measurements Define the Dimensions of the ‘30-Nm’ Chromatin Fiber: Evidence for a Compact, Interdigitated Structure.” *Proceedings of the National Academy of Sciences* 103(17): 6506–11.

Robyr, Daniel et al. 2002. “Microarray Deacetylation Maps Determine Genome-Wide Functions for Yeast Histone Deacetylases.” *Cell* 109(4): 437–46.

Rödelsperger, Christian et al. 2011. “Integrative Analysis of Genomic, Functional and Protein Interaction Data Predicts Long-Range Enhancer-Target Gene Interactions.” *Nucleic Acids Research* 39(7): 2492–2502.

Rodley, C. D.M., F. Bertels, B. Jones, and J. M. O’Sullivan. 2009. “Global Identification of Yeast Chromosome Interactions Using Genome Conformation Capture.” *Fungal Genetics and Biology* 46(11): 879–86.

Rogakou, Emmy P. et al. 1998. “DNA Double-Stranded Breaks Induce Histone H2AX Phosphorylation on Serine 139.” *Journal of Biological Chemistry* 273(10): 5858–68.

Rogers, Bruce L., and Grady F. Saunders. 1986. “Transcriptional Enhancers Play a Major Role in Gene Expression.” *BioEssays* 4(2): 62–65.

Roh, Tae Young et al. 2004. “High-Resolution Genome-Wide Mapping of Histone Modifications.” *Nature Biotechnology* 22(8): 1013–16.

Ronsseray, Stéphane. 1986. “P-M System of Hybrid Dysgenesis in Drosophila Melanogaster: Thermic Modifications of the Cytotype Can Be Detected for Several Generations.” *MGG Molecular & General Genetics* 205(1): 23–27.

Rosa, M. et al. 2013. “The Arabidopsis SWR1 Chromatin-Remodeling Complex Is Important for DNA Repair, Somatic Recombination, and Meiosis.” *The Plant Cell* 25(6): 1990–2001.

Rossetto, Dorine, Nikita Avvakumov, and Jacques Côté. 2012. “Histone Phosphorylation: A Chromatin Modification Involved in Diverse Nuclear Events.” *Epigenetics* 7(10): 1098–1108.

Rossetto, Dorine, Andrew W. Truman, Stephen J. Kron, and Jacques Côté. 2010. “Epigenetic Modifications in Double-Strand Break DNA Damage Signaling and Repair.” *Clinical Cancer Research* 16(18): 4543–52.

Roy, Peter J., Joshua M. Stuart, Jim Lund, and Stuart K. Kim. 2002. “Chromosomal Clustering of Muscle-Expressed Genes in Caenorhabditis Elegans.” *Nature* 418(6901): 975–79.

Rubin, Gerald M., Margaret G. Kidwell, and Paul M. Bingham. 1982. “The Molecular Basis of P-M Hybrid Dysgenesis: The Nature of Induced Mutations.” *Cell* 29(3): 987–94.

RUDKIN, GEORGE T., and B. D. STOLLAR. 1977. “High Resolution Detection of DNA–RNA Hybrids in Situ by Indirect Immunofluorescence.” *Nature* 265(5593): 472–73.

Ruthenburg, Alexander J., C. David Allis, and Joanna Wysocka. 2007. “Methylation of Lysine 4 on Histone H3: Intricacy of

Writing and Reading a Single Epigenetic Mark." *Molecular Cell* 25(1): 15–30.

Rutledge, Mark T. et al. 2015. "The Yeast Genome Undergoes Significant Topological Reorganization in Quiescence." *Nucleic Acids Research* 43(17): 8299–8313.

Saccone, Salvatore, Concetta Federico, and Giorgio Bernardi. 2002. "Localization of the Gene-Richest and the Gene-Poorest Isochores in the Interphase Nuclei of Mammals and Birds." In *Gene*, , 169–78.

Saksouk, Nehmé, Elisabeth Simboeck, and Jérôme Déjardin. 2015. "Constitutive Heterochromatin Formation and Transcription in Mammals." *Epigenetics & Chromatin* 8(1): 3.

Saldanha, Sabita N., and Louis Patrick Watanabe. 2015. "Epigenetics and Aging." In *Epigenetics and Dermatology*, , 379–406.

Salmon, Octavio D. Rodriguez et al. 2016. "The Spin-3/2 Blume–Capel Model with Competing Short- and Long-Range Interactions." *Physica A: Statistical Mechanics and its Applications* 464: 103–14.

Sandhu, Kuljeet Singh. 2009. "Intrinsic Disorder Explains Diverse Nuclear Roles of Chromatin Remodeling Proteins." *Journal of Molecular Recognition* 22(1): 1–8.

Sandhu, Kuljeet Singh et al. 2009. "Nonallelic Transvection of Multiple Imprinted Loci Is Organized by the H19 Imprinting Control Region during Germline Development." *Genes & Development* 23(22): 2598–2603.

Sandhu, Kuljeet Singh. 2012. "Did the Modulation of Expression Noise Shape the Evolution of Three Dimensional Genome Organizations in Eukaryotes?" *Nucleus* 3(3): 286–89.

Sandhu, Kuljeet Singh et al. 2012. "Large-Scale Functional Organization of Long-Range Chromatin Interaction Networks." *Cell reports* 2(5): 1207–19.

Saner, Nazan et al. 2013. "Stochastic Association of Neighboring Replicons Creates Replication Factories in Budding Yeast." *Journal of Cell Biology* 202(7): 1001–12.

Sanyal, Amartya, Bryan R. Lajoie, Gaurav Jain, and Job Dekker. 2012. "The Long-Range Interaction Landscape of Gene Promoters." *Nature* 489(7414): 109–13.

Sawicka, Anna, and Christian Seiser. 2012. "Histone H3 Phosphorylation - A Versatile Chromatin Modification for Different Occasions." *Biochimie* 94(11): 2193–2201.

Sawicka, Anna, and Christian Seiser. 2014. "Sensing Core Histone Phosphorylation - A Matter of Perfect Timing." *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms* 1839(8): 711–18.

Saxonov, S., P. Berg, and D. L. Brutlag. 2006. "A Genome-Wide Analysis of CpG Dinucleotides in the Human Genome Distinguishes Two Distinct Classes of Promoters." *Proceedings of the National Academy of Sciences* 103(5): 1412–17.

Schalch, Thomas, Sylwia Duda, David F Sargent, and Timothy J Richmond. 2005. "X-Ray Structure of a Tetranucleosome and Its Implications for the Chromatin Fibre." *Nature* 436(7047): 138.

Schmid, Carl, and Richard Maraia. 1992. "Transcriptional Regulation and Transpositional Selection of Active SINE Sequences." *Current Opinion in Genetics and Development* 2(6): 874–82.

Schneider, Robert et al. 2004. "Histone H3 Lysine 4 Methylation Patterns in Higher Eukaryotic Genes." *Nature Cell Biology* 6(1): 73–77.

Schoenfelder, Stefan et al. 2010. "Preferential Associations between Co-Regulated Genes Reveal a Transcriptional Interactome in Erythroid Cells." *Nature Genetics* 42(1): 53–61.

Schoenfelder, Stefan et al. 2015. "The Pluripotent Regulatory Circuitry Connecting Promoters to Their Long-Range Interacting Elements." *Genome Research* 25(4): 582–97.

Schoenfelder, Stefan et al. 2018. "Promoter Capture Hi-C: High-Resolution, Genome-Wide Profiling of Promoter Interactions." *Journal of Visualized Experiments* 28(136): 57320.

Schreiber, Stuart L., and Bradley E. Bernstein. 2002. "Signaling Network Model of Chromatin." *Cell* 111(6): 771–78.

Schuldiner, Maya et al. 2005. "Exploration of the Function and Organization of the Yeast Early Secretory Pathway through an Epistatic Miniarray Profile." *Cell* 123(3): 507–19.

Schulze, Sandra R. et al. 2006. "Heterochromatic Genes in Drosophila: A Comparative Analysis of Two Genes." *Genetics* 173(3): 1433–45.

Schwer, Bjoern et al. 2016. "Transcription-Associated Processes Cause DNA Double-Strand Breaks and Translocations in

Neural Stem/Progenitor Cells." *Proceedings of the National Academy of Sciences* 113(8): 2258–63.

Scott, Kristin C., Stephanie L. Merrett, and Huntington F. Willard. 2006. "A Heterochromatin Barrier Partitions the Fission Yeast Centromere into Discrete Chromatin Domains." *Current Biology* 16(2): 119–29.

Sedat, J., and L. Manuelidis. 1977. "A Direct Approach to the Structure of Eukaryotic Chromosomes." *Cold Spring Harbor Symposia on Quantitative Biology* 42(1): 331–50.

Seeber, Andrew, Vincent Dion, and Susan M. Gasser. 2013. "Checkpoint Kinases and the INO80 Nucleosome Remodeling Complex Enhance Global Chromatin Mobility in Response to DNA Damage." *Genes and Development* 27(18): 1999–2008.

Seisenberger, Stefanie et al. 2012. "The Dynamics of Genome-Wide DNA Methylation Reprogramming in Mouse Primordial Germ Cells." *Molecular Cell* 48(6): 849–62.

Shah, Sheel et al. 2018. "Dynamics and Spatial Genomics of the Nascent Transcriptome by Intron SeqFISH." *Cell* 174(2): 363-376.e16.

Shandilya, Jayasha, and Stefan G E Roberts. 2012. "The Transcription Cycle in Eukaryotes: From Productive Initiation to RNA Polymerase II Recycling." *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms* 1819(5): 391–400.

Shankar Narayan, K., W. J. Steele, K. Smetana, and H. Busch. 1967. "Ultrastructural Aspects of the Ribonucleo-Protein Network in Nuclei of Walker Tumor and Rat Liver." *Experimental Cell Research* 46(1): 65–77.

Shapiro, James A., and Richard von Sternberg. 2005. "Why Repetitive DNA Is Essential to Genome Function." *Biological Reviews* 80(2): 227–50.

Shen, Xuetong, Gaku Mizuguchi, Ali Hamiche, and Wu Carl. 2000. "A Chromatin Remodelling Complex Involved in Transcription and DNA Processing." *Nature* 406(6795): 541–44.

Shiio, Y., and R. N. Eisenman. 2003. "Histone Sumoylation Is Associated with Transcriptional Repression." *Proceedings of the National Academy of Sciences* 100(23): 13225–30.

Shiloh, Yosef. 2003. "ATM and Related Protein Kinases: Safeguarding Genome Integrity." *Nature Reviews Cancer* 3(3): 155–68.

Shimada, Kenji et al. 2008. "Ino80 Chromatin Remodeling Complex Promotes Recovery of Stalled Replication Forks." *Current Biology* 18(8): 566–75.

Shivaswamy, Sushma, and Vishwanath R. Iyer. 2007. "Genome-Wide Analysis of Chromatin Status Using Tiling Microarrays." *Methods* 41(3): 304–11.

Shlens, Jonathon. 2014. "A Tutorial on Principal Component Analysis." April.

Shogren-Knaak, Michael et al. 2006. "Histone H4-K16 Acetylation Controls Chromatin Structure and Protein Interactions." *Science* 311(5762): 844–47.

Shopland, Lindsay S. et al. 2006. "Folding and Organization of a Contiguous Chromosome Region According to the Gene Distribution Pattern in Primary Genomic Sequence." *Journal of Cell Biology* 174(1): 27–38.

Silander, Olin K. et al. 2012. "A Genome-Wide Analysis of Promoter-Mediated Phenotypic Noise in Escherichia Coli" ed. Ivan Matic. *PLoS Genetics* 8(1): e1002443.

Simonis, Marieke et al. 2006. "Nuclear Organization of Active and Inactive Chromatin Domains Uncovered by Chromosome Conformation Capture-on-Chip (4C)." *Nature Genetics* 38(11): 1348–54.

Sinervo, B., and E. Svensson. 2002. "Correlational Selection and the Evolution of Genomic Architecture." *Heredity* 89(5): 329–38.

Singh, Badri Nath, and Michael Hampsey. 2007. "A Transcription-Independent Role for TFIIB in Gene Looping." *Molecular Cell* 27(5): 806–16.

Smeenk, Godelieve, and Haico van Attikum. 2013. "The Chromatin Response to DNA Breaks: Leaving a Mark on Genome Integrity." *Annual Review of Biochemistry* 82(1): 55–80.

Smit, Arian Fa. 1999. "Interspersed Repeats and Other Mementos of Transposable Elements in Mammalian Genomes." *Current Opinion in Genetics and Development* 9(6): 657–63.

Smith, Zachary D., and Alexander Meissner. 2013. "DNA Methylation: Roles in Mammalian Development." *Nature Reviews*

Genetics 14(3): 204–20.

Smolle, Michaela et al. 2012. “Chromatin Remodelers Isw1 and Chd1 Maintain Chromatin Structure during Transcription by Preventing Histone Exchange.” *Nature Structural and Molecular Biology* 19(9): 884–92.

Smolle, Michaela, and Jerry L. Workman. 2013. “Transcription-Associated Histone Modifications and Cryptic Transcription.” *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms* 1829(1): 84–97.

Solovei, Irina et al. 2009. “Nuclear Architecture of Rod Photoreceptor Cells Adapts to Vision in Mammalian Evolution.” *Cell* 137(2): 356–68.

Soriano, Ignacio et al. 2014. “Different Nucleosomal Architectures at Early and Late Replicating Origins in Saccharomyces Cerevisiae.” *BMC Genomics* 15(1): 791.

Spellman, Paul T. et al. 1998. “Comprehensive Identification of Cell Cycle–Regulated Genes of the Yeast Saccharomyces Cerevisiae by Microarray Hybridization” ed. Gerald R. Fink. *Molecular Biology of the Cell* 9(12): 3273–97.

Spellman, Paul T, and Gerald M Rubin. 2002. “Evidence for Large Domains of Similarly Expressed Genes in the Drosophila Genome.” *Journal of biology* 1(1): 5.

Spencer, M., W. Fuller, M. H.F. Wilkins, and G. L. Brown. 1962. “Determination of the Helical Configuration of Ribonucleic Acid Molecules by X-Ray Diffraction Study of Crystalline Amino-Acid-Transfer Ribonucleic Acid.” *Nature* 194(4833): 1014–20.

Spieth, John et al. 1993. “Operons in C. Elegans: Polycistronic MRNA Precursors Are Processed by Trans-Splicing of SL2 to Downstream Coding Regions.” *Cell* 73(3): 521–32.

Spilianakis, Charalampos G., and Richard A. Flavell. 2004. “Long-Range Intrachromosomal Interactions in the T Helper Type 2 Cytokine Locus.” *Nature Immunology* 5(10): 1017–27.

Spivakov, Mikhail. 2014. “Spurious Transcription Factor Binding: Non-Functional or Genetically Redundant?” *Bioessays* 36(8): 798–806.

Starling, Shimona. 2017. “The Nuclear Pore — a Scaffold for Inducible Genes.” *Nature Reviews Genetics* 18: 273.

Stedman, Edgar, and Ellen Stedman. 1950. “Cell Specificity of Histones.” *Nature* 166(4227): 780–81.

van Steensel, Bas, and Andrew S. Belmont. 2017. “Lamina-Associated Domains: Links with Chromosome Architecture, Heterochromatin, and Gene Repression.” *Cell* 169(5): 780–91.

Di Stefano, Marco et al. 2013. “Colocalization of Coregulated Genes: A Steered Molecular Dynamics Study of Human Chromosome 19” ed. Wyeth W. Wasserman. *PLoS Computational Biology* 9(3): e1003019.

Steger, David J. et al. 2003. “Regulation of Chromatin Remodeling by Inositol Polyphosphates.” *Science* 299(5603): 114–16.

Steger, David J, and Jerry L Workman. 2005. “Remodeling Chromatin Structures for Transcription: What Happens to the Histones?” *BioEssays* 18(11): 875–84.

Sterner, D. E., and S. L. Berger. 2000. “Acetylation of Histones and Transcription-Related Factors.” *Microbiology and Molecular Biology Reviews* 64(2): 435–59.

Stevens, R., C. A. Goble, and S. Bechhofer. 2000. “Ontology-Based Knowledge Representation for Bioinformatics.” *Briefings in bioinformatics* 1(4): 398–414.

Stewart, Grant S. et al. 2003. “MDC1 Is a Mediator of the Mammalian DNA Damage Checkpoint.” *Nature* 421(6926): 961–66.

Stokes, D G, and R P Perry. 1995. “DNA-Binding and Chromatin Localization Properties of CHD1.” *Molecular and cellular biology* 15(5): 2745–53.

Strahl, B. D., R. Ohba, R. G. Cook, and C. D. Allis. 2002. “Methylation of Histone H3 at Lysine 4 Is Highly Conserved and Correlates with Transcriptionally Active Nuclei in Tetrahymena.” *Proceedings of the National Academy of Sciences* 96(26): 14967–72.

Strahl, Brian D., and C. David Allis. 2000. “The Language of Covalent Histone Modifications.” *Nature* 403(6765): 41–45.

Strambio-De-Castillia, Caterina, Mario Niepel, and Michael P. Rout. 2010. “The Nuclear Pore Complex: Bridging Nuclear Transport and Gene Regulation.” *Nature Reviews Molecular Cell Biology* 11(7): 490–501.

Strom, Amy R. et al. 2017. “Phase Separation Drives Heterochromatin Domain Formation.” *Nature* 547(7662): 241–45.

Studitsky, Vasily M, Ekaterina V Nizovtseva, Alexey K Shaytan, and Donal S Luse. 2016. “Nucleosomal Barrier to

Transcription: Structural Determinants and Changes in Chromatin Structure." *Biochemistry & molecular biology journal* 2(2): 8.

Subramanian, Aravind et al. 2005. "Gene Set Enrichment Analysis: A Knowledge-Based Approach for Interpreting Genome-Wide Expression Profiles." *Proceedings of the National Academy of Sciences* 102(43): 15545–50.

Suganuma, Tamaki, and Jerry L. Workman. 2008. "Crosstalk among Histone Modifications." *Cell* 135(4): 604–7.

Surani, M. A.H., S. C. Barton, and M. L. Norris. 1984. "Development of Reconstituted Mouse Eggs Suggests Imprinting of the Genome during Gametogenesis." *Nature* 308(5959): 548–50.

Svejstrup, Jesper Q. 2010. "The Interface between Transcription and Mechanisms Maintaining Genome Integrity." *Trends in Biochemical Sciences* 35(6): 333–38.

Swinstead, Erin E., Ville Paakinaho, Diego M. Presman, and Gordon L. Hager. 2016. "Pioneer Factors and ATP-Dependent Chromatin Remodeling Factors Interact Dynamically: A New Perspective: Multiple Transcription Factors Can Effect Chromatin Pioneer Functions through Dynamic Interactions with ATP-Dependent Chromatin Remodeling Factors." *BioEssays* 38(11): 1150–57.

Szabo, Quentin et al. 2018. "TADs Are 3D Structural Units of Higher-Order Chromosome Organization in Drosophila." *Science advances* 4(2): eaar8082.

Szczepińska, Teresa, and Krzysztof Pawłowski. 2013. "Genomic Positions of Co-Expressed Genes: Echoes of Chromosome Organisation in Gene Expression Data." *BMC Research Notes* 6(1): 229.

Taatjes, Dylan J. 2017. "The Continuing SAGA of TFIID and RNA Polymerase II Transcription." *Molecular Cell* 68(1): 1–2.

Takahashi, Kazutoshi, and Shinya Yamanaka. 2006. "Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors." *Cell* 126(4): 663–76.

Tanabe, H. et al. 2002. "Evolutionary Conservation of Chromosome Territory Arrangements in Cell Nuclei from Higher Primates." *Proceedings of the National Academy of Sciences* 99(7): 4424–29.

Tanaka, Yutaka. 1988. "Sensitivity Analysis in Principal Component Analysis : Influence on the Subspace Spanned by Principal Components." *Communications in Statistics - Theory and Methods* 17(9): 3157–75.

Tang, Fuchou et al. 2010. "Tracing the Derivation of Embryonic Stem Cells from the Inner Cell Mass by Single-Cell RNA-Seq Analysis." *Cell Stem Cell* 6(5): 468–78.

Tanizawa, Hideki et al. 2010. "Mapping of Long-Range Associations throughout the Fission Yeast Genome Reveals Global Genome Organization Linked to Transcriptional Regulation." *Nucleic Acids Research* 38(22): 8164–77.

Tarailo-Graovac, Maja, and Nansheng Chen. 2013. "Gene Clustering in Eukaryotes." In *ELS*, Major Reference Works, Chichester, UK: John Wiley & Sons, Ltd.

Tatarakis, Antonis, Reza Behrouzi, and Danesh Moazed. 2017. "Evolving Models of Heterochromatin: From Foci to Liquid Droplets." *Molecular Cell* 67(5): 725–27.

Tatsuoka, Kay, Rongxiang Liu, and Xiang Guo. 2005. "Histone Acetylation and Transcriptional Regulation in the Genome of Saccharomyces Cerevisiae." *Bioinformatics* 22(4): 392–99.

Teferedegne, B., M. R. Green, Z. Guo, and J. M. Boss. 2006. "Mechanism of Action of a Distal NF- B-Dependent Enhancer." *Molecular and Cellular Biology* 26(15): 5759–70.

Tessarz, Peter et al. 2014. "Glutamine Methylation in Histone H2A Is an RNA-Polymerase-I-Dedicated Modification." *Nature* 505(7484): 564–68.

Thévenin, Annelyse, Liat Ein-Dor, Michal Ozery-Flato, and Ron Shamir. 2014. "Functional Gene Groups Are Concentrated within Chromosomes, among Chromosomes and in the Nuclear Space of the Human Genome." *Nucleic Acids Research* 42(15): 9854–61.

Thomas, J. O., and R. D. Kornberg. 1975. "An Octamer of Histones in Chromatin and Free in Solution." *Proceedings of the National Academy of Sciences* 72(7): 2626–30.

Thompson, Jeffrey S., Xuefeng Ling, and Michael Grunstein. 1994. "Histone H3 Amino Terminus Is Required for Telomeric and Silent Mating Locus Repression in Yeast." *Nature* 369(6477): 245–47.

Thomson, Stuart, Louis C. Mahadevan, and Alison L. Clayton. 1999. "MAP Kinase-Mediated Signalling to Nucleosomes and

Immediate-Early Gene Induction." *Seminars in Cell and Developmental Biology* 10(2): 205–14.

Tiang, C.-L., Yan He, and Wojtek P Pawlowski. 2012. "Chromosome Organization and Dynamics during Interphase, Mitosis, and Meiosis in Plants." *PLANT PHYSIOLOGY* 158(1): 26–34.

Tiwari, Vijay K. et al. 2008. "PcG Proteins, DNA Methylation, and Gene Repression by Chromatin Looping" ed. Peter B Becker. *PLoS Biology* 6(12): e306.

Tolhuis, Bas et al. 2002. "Looping and Interaction between Hypersensitive Sites in the Active β-Globin Locus." *Molecular Cell* 10(6): 1453–65.

Tollefsbol, Trygve O. 2011. "Epigenetics: The New Science of Genetics." In *Handbook of Epigenetics*, Academic Press, 1–6.

Tosi, Alessandro et al. 2013. "Structure and Subunit Topology of the INO80 Chromatin Remodeler and Its Nucleosome Complex." *Cell* 154(6): 1207–19.

Tremblay, Mathieu, Oraly Sanchez-Ferras, and Maxime Bouchard. 2018. "GATA Transcription Factors in Development and Disease." *Development* 145(20): dev164384.

Tremethick, David J. 2007. "Higher-Order Structures of Chromatin: The Elusive 30 Nm Fiber." *Cell* 128(4): 651–54.

Trojer, Patrick, and Danny Reinberg. 2007. "Facultative Heterochromatin: Is There a Distinctive Molecular Signature?" *Molecular Cell* 28(1): 1–13.

Trumbly, R. J. 1992. "Glucose Repression in the Yeast Saccharomyces Cerevisiae." *Molecular Microbiology* 6(1): 15–21.

Tsukiyama, Toshio et al. 1999. "Characterization of the Imitation Switch Subfamily of ATP-Dependent Chromatin-Remodeling Factors in Saccharomyces Cerevisiae." *Genes and Development* 13(6): 686–97.

Tsukuda, Toyoko, Alastair B. Fleming, Jac A. Nickoloff, and Mary Ann Osley. 2005. "Chromatin Remodelling at a DNA Double-Strand Break Site in Saccharomyces Cerevisiae." *Nature* 438(7066): 379–83.

Tsytsykova, Alla V et al. 2007. "Activation-Dependent Intrachromosomal Interactions Formed by the TNF Gene Promoter and Two Distal Enhancers." *Proceedings of the National Academy of Sciences* 104(43): 16850–55.

Tycko, Benjamin. 1997. "DNA Methylation in Genomic Imprinting." *Mutation Research - Reviews in Mutation Research* 386(2): 131–40.

Uckelmann, Michael, and Titia K. Sixma. 2017. "Histone Ubiquitination in the DNA Damage Response." *DNA Repair* 56: 92–101.

Ulianov, Sergey V. et al. 2016. "Active Chromatin and Transcription Play a Key Role in Chromosome Partitioning into Topologically Associating Domains." *Genome Research* 26(1): 70–84.

Unnikrishnan, Archana et al. 2019. "The Role of DNA Methylation in Epigenetics of Aging." *Pharmacology and Therapeutics* 195: 172–85.

Uziel, Tamar et al. 2003. "Requirement of the MRN Complex for ATM Activation by DNA Damage." *EMBO Journal* 22(20): 5612–21.

Vaissière, Thomas, Carla Sawan, and Zdenko Herceg. 2008. "Epigenetic Interplay between Histone Modifications and DNA Methylation in Gene Silencing." *Mutation Research - Reviews in Mutation Research* 659(1–2): 40–48.

Vanrobays, Emmanuel, Mélanie Thomas, and Christophe Tatout. 2013. 46 Annual Plant Reviews, Plant Nuclear Structure, Genome Architecture and Gene Regulation *Heterochromatin Positioning and Nuclear Architecture*.

Varga-Weisz, Patrick D. 2010. "Insights into How Chromatin Remodeling Factors Find Their Target in the Nucleus." *Proceedings of the National Academy of Sciences* 107(46): 19611–12.

Vary, J. C. et al. 2002. "Yeast Isw1p Forms Two Separable Complexes In Vivo." *Molecular and Cellular Biology* 23(1): 80–91.

Venkatesh, Swaminathan, and Jerry L. Workman. 2013. "Set2 Mediated H3 Lysine 36 Methylation: Regulation of Transcription Elongation and Implications in Organismal Development." *Wiley Interdisciplinary Reviews: Developmental Biology* 2(5): 685–700.

Venters, Bryan J et al. 2011. "A Comprehensive Genomic Binding Map of Gene and Chromatin Regulatory Proteins in Saccharomyces." *Molecular cell* 41(4): 480–92.

Verdone, Loredana, Eleonora Agricola, Micaela Caserta, and Ernesto Di Mauro. 2006. "Histone Acetylation in Gene Regulation." *Briefings in Functional Genomics and Proteomics* 5(3): 209–21.

Véron, Amélie S. et al. 2011. “Close 3D Proximity of Evolutionary Breakpoints Argues for the Notion of Spatial Synteny.” *BMC Genomics* 12(1): 303.

Verschure, Pernette J., Ineke Van Der Kraan, Erik M.M. Manders, and Roel Van Driel. 1999. “Spatial Relationship between Transcription Sites and Chromosome Territories.” *Journal of Cell Biology* 147(1): 13–24.

Versteeg, Rogier et al. 2003. “The Human Transcriptome Map Reveals Extremes in Gene Dentistry, Intron Length, GC Content, and Repeat Pattern for Domains of Highly and Weakly Expressed Genes.” *Genome Research* 13(9): 1998–2004.

Vettese-Dadey, Michelle et al. 1996. “Acetylation of Histone H4 Plays a Primary Role in Enhancing Transcription Factor Binding to Nucleosomal DNA in Vitro.” *The EMBO Journal* 15(10): 2508–18.

Vietri Rudan, Matteo et al. 2015. “Comparative Hi-C Reveals That CTCF Underlies Evolution of Chromosomal Domain Architecture.” *Cell Reports* 10(8): 1297–1309.

Viguera, Enrique, Danielle Canceill, and S. Dusko Ehrlich. 2001. “Replication Slippage Involves DNA Polymerase Pausing and Dissociation.” *EMBO Journal* 20(10): 2587–95.

Vincent, Jack A, Tracey J Kwong, and Toshio Tsukiyama. 2008. “ATP-Dependent Chromatin Remodeling Shapes the DNA Replication Landscape.” *Nature structural & molecular biology* 15(5): 477–84.

Vissers, Joseph Ha et al. 2008. “The Many Faces of Ubiquitinated Histone H2A: Insights from the DUBs.” *Cell Division* 3(1): 8.

Voigt, Philipp, Wee Wei Tee, and Danny Reinberg. 2013. “A Double Take on Bivalent Promoters.” *Genes and Development* 27(12): 1318–38.

Volpe, Thomas A. et al. 2002. “Regulation of Heterochromatic Silencing and Histone H3 Lysine-9 Methylation by RNAi.” *Science* 297(5588): 1833–37.

Voss, Ty C., and Gordon L. Hager. 2014. “Dynamic Regulation of Transcriptional States by Chromatin and Transcription Factors.” *Nature Reviews Genetics* 15(2): 69–81.

Vourc’h, Claire, Domenica Taruscio, Ann L. Boyle, and David C. Ward. 1993. “Cell Cycle-Dependent Distribution of Telomeres, Centromeres, and Chromosome-Specific Subsatellite Domains in the Interphase Nucleus of Mouse Lymphocytes.” *Experimental Cell Research* 205(1): 142–51.

WADDINGTON, C H. 1942. “CANALIZATION OF DEVELOPMENT AND THE INHERITANCE OF ACQUIRED CHARACTERS.” *Nature* 150: 563.

Wagner, Eric J., and Phillip B. Carpenter. 2012. “Understanding the Language of Lys36 Methylation at Histone H3.” *Nature Reviews Molecular Cell Biology* 13(2): 115–26.

Walsh, Christine S. 2015. “Two Decades beyond BRCA1/2: Homologous Recombination, Hereditary Cancer Risk and a Target for Ovarian Cancer Therapy?” *Gynecologic Oncology* 137(2): 343–50.

Wang, Guang Zhong, Martin J. Lercher, and Laurence D. Hurst. 2011. “Transcriptional Coupling of Neighboring Genes and Gene Expression Noise: Evidence That Gene Orientation and Noncoding Transcripts Are Modulators of Noise.” *Genome Biology and Evolution* 3(1): 320–31.

Wang, J, L Yang, and C Cui. 2001. “[The Clinical Effect of Tropisetron in the Prevention of Nausea and Vomiting Induced by Anti-Cancer Drugs].” *Zhonghua zhong liu za zhi [Chinese journal of oncology]* 23(3): 251–53.

Wang, Jianrong, Victoria V. Lunyak, and I. King Jordan. 2012. “Genome-Wide Prediction and Analysis of Human Chromatin Boundary Elements.” *Nucleic Acids Research* 40(2): 511–29.

Wang, Jiyong, Sharon T. Jia, and Songtao Jia. 2016. “New Insights into the Regulation of Heterochromatin.” *Trends in Genetics* 32(5): 284–94.

Wang, Jiyong, Stephanie T. Lawry, Allison L. Cohen, and Songtao Jia. 2014. “Chromosome Boundary Elements and Regulation of Heterochromatin Spreading.” *Cellular and Molecular Life Sciences* 71(24): 4841–52.

Wang, Renjie, Julien Mozziconacci, Aurélien Bancaud, and Olivier Gadal. 2015. “Principles of Chromatin Organization in Yeast: Relevance of Polymer Models to Describe Nuclear Organization and Dynamics.” *Current Opinion in Cell Biology* 34: 54–60.

Wang, Siyuan et al. 2016. “Spatial Organization of Chromatin Domains and Compartments in Single Chromosomes.” *Science*

353(6299): 598–602.

Wang, Wei-Lin et al. 2013. “Developmental Regulation of Histone H2A Deposition via Serine-1 Phosphorylation.” *Epigenetics & Chromatin* 6(1): P74.

Wang, Zhaoqing et al. 2014. “Cyclin B1/Cdk1 Coordinates Mitochondrial Respiration for Cell-Cycle G2/M Progression.” *Developmental Cell*.

Waring, Michael, and Roy J. Britten. 1966. “Nucleotide Sequence Repetition: A Rapidly Reassociating Fraction of Mouse DNA.” *Science* 154(3750): 791–94.

Wasylyk, B., C. Wasylyk, P. Augereau, and P. Chambon. 1983. “The SV40 72 Bp Repeat Preferentially Potentiates Transcription Starting from Proximal Natural or Substitute Promoter Elements.” *Cell* 32(2): 503–14.

Watson, N.A., and J.M.G. Higgins. 2016. “Histone Kinases and Phosphatases.” In *Chromatin Signaling and Diseases*, eds. Olivier Binda and Martin Ernesto B T - Chromatin Signaling and Diseases Fernandez-Zapico. Boston: Elsevier, 75–94.

Weierich, Claudia et al. 2003. “Three-Dimensional Arrangements of Centromeres and Telomeres in Nuclei of Human and Murine Lymphocytes.” *Chromosome Research* 11(5): 485–502.

Weiner, Assaf et al. 2015. “High-Resolution Chromatin Dynamics during a Yeast Stress Response.” *Molecular Cell* 58(2): 371–86.

Whitehouse, Iestyn et al. 1999. “Nucleosome Mobilization Catalysed by the Yeast SWI/SNF Complex.” *Nature* 400(6746): 784–87.

Whitehouse, Iestyn, Oliver J. Rando, Jeff Delrow, and Toshio Tsukiyama. 2007. “Chromatin Remodelling at Promoters Suppresses Antisense Transcription.” *Nature* 450(7172): 1031–35.

Wiblin, A. E. 2005. “Distinctive Nuclear Organisation of Centromeres and Regions Involved in Pluripotency in Human Embryonic Stem Cells.” *Journal of Cell Science* 118(17): 3861–68.

Wiles, Elizabeth T., and Eric U. Selker. 2017. “H3K27 Methylation: A Promiscuous Repressive Chromatin Mark.” *Current Opinion in Genetics and Development* 43: 31–37.

Williams, Adam, and Richard A Flavell. 2008. “The Role of CTCF in Regulating Nuclear Organization: Figure 1.” *The Journal of Experimental Medicine* 205(4): 747–50.

Williams, Ruth R.E., Simon Broad, Denise Sheer, and Jiannis Ragoussis. 2002. “Subchromosomal Positioning of the Epidermal Differentiation Complex (EDC) in Keratinocyte and Lymphoblast Interphase Nuclei.” *Experimental Cell Research* 272(2): 163–75.

Williamson, Iain et al. 2014. “Spatial Genome Organization: Contrasting Views from Chromosome Conformation Capture and Fluorescence in Situ Hybridization.” *Genes and Development* 28(24): 2778–91.

Wingett, Steven W. et al. 2015. “HiCUP: Pipeline for Mapping and Processing Hi-C Data.” *F1000Research* 4: 1310.

Winter, Stefan et al. 2008. “14-3-3 Proteins Recognize a Histone Code at Histone H3 and Are Required for Transcriptional Activation.” *EMBO Journal* 27(1): 88–99.

Winzeler, Elizabeth A. et al. 1999. “Functional Characterization of the S. Cerevisiae Genome by Gene Deletion and Parallel Analysis.” *Science* 285(5429): 901–6.

Wischnitzer, Saul. 1973. “The Submicroscopic Morphology of the Interphase Nucleus.” *International Review of Cytology* 34(C): 1–48.

de Wit, Elzo et al. 2008. “Global Chromatin Domain Organization of the Drosophila Genome.” *PLoS genetics* 4(3): e1000045.

de Wit, Elzo, and Wouter de Laat. 2012. “A Decade of 3C Technologies: Insights into Nuclear Organization.” *Genes & Development* 26(1): 11–24.

Witten, Daniela M., and William Stafford Noble. 2012. “On the Assessment of Statistical Significance of Three-Dimensional Colocalization of Sets of Genomic Elements.” *Nucleic Acids Research* 40(9): 3849–55.

Wold, Svante et al. 2009. “The PLS Model Space Revisited.” *Journal of Chemometrics* 23(2): 67–68.

Wolfe, Kenneth H., and Denis C. Shields. 1997. “Molecular Evidence for an Ancient Duplication of the Entire Yeast Genome.” *Nature* 387(6634): 708–13.

Wood, Adam et al. 2003. “The Paf1 Complex Is Essential for Histone Monoubiquitination by the Rad6-Bre1 Complex, Which

Signals for Histone Methylation by COMPASS and Dot1p." *Journal of Biological Chemistry* 278(37): 34739–42.

Wood, Adam, and Ali Shilatifard. 2004. "Posttranslational Modifications of Histones by Methylation." *Advances in Protein Chemistry* 67: 201–22.

Woodage, T. et al. 1997. "Characterization of the CHD Family of Proteins." *Proceedings of the National Academy of Sciences* 94(21): 11472–77.

Woodcock, C. 1984. "The Higher-Order Structure of Chromatin: Evidence for a Helical Ribbon Arrangement." *The Journal of Cell Biology* 99(1): 42–52.

Woodcock, C. L.F., J. P. Safer, and J. E. Stanchfield. 1976. "Structural Repeating Units in Chromatin. I. Evidence for Their General Occurrence." *Experimental Cell Research* 97(1): 101–10.

Worcel, Abraham, S Strogatz, and D Riley. 1981. "Structure of Chromatin and the Linking Number of DNA." *Proceedings of the National Academy of Sciences* 78(3): 1461–65.

Workman, Jerry L. 2006. "Nucleosome Displacement in Transcription." *Genes and Development* 20(15): 2009–17.

Wray, Gregory A. 2007. "The Evolutionary Significance of Cis-Regulatory Mutations." *Nature Reviews Genetics* 8(3): 206–16.

Wu, Lipeng et al. 2011. "The RING Finger Protein MSL2 in the MOF Complex Is an E3 Ubiquitin Ligase for H2B K34 and Is Involved in Crosstalk with H3 K4 and K79 Methylation." *Molecular Cell* 43(1): 132–44.

Wu, Wei-Sheng, and Fu-Jou Lai. 2015. "Functional Redundancy of Transcription Factors Explains Why Most Binding Targets of a Transcription Factor Are Not Affected When the Transcription Factor Is Knocked Out." *BMC systems biology* 9(Suppl 6): S2.

Wyman, A. R., and R. White. 1980. "A Highly Polymorphic Locus in Human DNA." *Proceedings of the National Academy of Sciences* 77(11): 6754–58.

Xie, Ting et al. 2015. "Spatial Features for Escherichia Coli Genome Organization." *BMC genomics* 16(1): 37.

Xiong, Y., and T. H. Eickbush. 1990. "Origin and Evolution of Retroelements Based upon Their Reverse Transcriptase Sequences." *The EMBO Journal* 9(10): 3353–62.

Xu, Jianquan et al. 2018. "Super-Resolution Imaging of Higher-Order Chromatin Structures at Different Epigenomic States in Single Mammalian Cells." *Cell Reports* 24(4): 873–82.

Xu, Ye et al. 2012. "Histone H2A.Z Controls a Critical Chromatin Remodeling Step Required for DNA Double-Strand Break Repair." *Molecular Cell* 48(5): 723–33.

Xu, Zhenyu et al. 2009. "Bidirectional Promoters Generate Pervasive Transcription in Yeast." *Nature* 457(7232): 1033–37.

Yadon, A. N. et al. 2010. "Chromatin Remodeling around Nucleosome-Free Regions Leads to Repression of Noncoding RNA Transcription." *Molecular and Cellular Biology* 30(21): 5110–22.

Yaffe, Eitan, and Amos Tanay. 2011. "Probabilistic Modeling of Hi-C Contact Maps Eliminates Systematic Biases to Characterize Global Chromosomal Architecture." *Nature Genetics* 43(11): 1059–65.

Yáñez-Cuna, J. Omar, and Bas van Steensel. 2017. "Genome–Nuclear Lamina Interactions: From Cell Populations to Single Cells." *Current Opinion in Genetics and Development* 43: 67–72.

Yang, Lei et al. 2014. "Characterization of TATA-Containing Genes and TATA-Less Genes in S. Cerevisiae by Network Topologies and Biological Properties." *Genomics* 104(6): 562–71.

Yang, Yee Hwa et al. 2002. "Normalization for CDNA Microarray Data: A Robust Composite Method Addressing Single and Multiple Slide Systematic Variation." *Nucleic acids research* 30(4): e15.

Yang, Yi-Sheng et al. 2002. "Chromosome Localization Analysis of Genes Strongly Expressed in Human Visceral Adipose Tissue." *Endocrine* 18(1): 57–66.

Yoshikawa, Katsunori et al. 2011. "Comprehensive Phenotypic Analysis of Single-Gene Deletion and Overexpression Strains of Saccharomyces Cerevisiae." *Yeast* 28(5): 349–61.

Yu, Guangchuang et al. 2010. "GOSemSim: An R Package for Measuring Semantic Similarity among GO Terms and Gene Products." *Bioinformatics* 26(7): 976–78.

Yue, Feng et al. 2014. "A Comparative Encyclopedia of DNA Elements in the Mouse Genome." *Nature* 515(7527): 355–64.

Zaret, Kenneth S., and Jason S. Carroll. 2011. "Pioneer Transcription Factors: Establishing Competence for Gene Expression."

Genes and Development 25(21): 2227–41.

Zhang, Jianzhi. 2003. “Evolution by Gene Duplication: An Update.” *Trends in Ecology and Evolution* 18(6): 292–98.

Zhang, Liwen, Ericka E. Eugeni, Mark R. Parthun, and Michael A. Freitas. 2003. “Identification of Novel Histone Post-Translational Modifications by Peptide Mass Fingerprinting.” *Chromosoma* 112(2): 77–86.

Zhang, R. 2004. “DEG: A Database of Essential Genes.” *Nucleic Acids Research* 32(Database issue): D271–72.

Zhang, Y., and D. Reinberg. 2001. “Transcription Regulation by Histone Methylation: Interplay between Different Covalent Modifications of the Core Histone Tails.” *Genes and Development* 15(18): 2343–60.

Zhang, Yu et al. 2012. “Spatial Organization of the Mouse Genome and Its Role in Recurrent Chromosomal Translocations.” *Cell* 148(5): 908–21.

Zhang, Yubo et al. 2013. “Chromatin Connectivity Maps Reveal Dynamic Promoter-Enhancer Long-Range Associations.” *Nature* 504(7479): 306–10.

Zhao, Zhihu et al. 2006. “Circular Chromosome Conformation Capture (4C) Uncovers Extensive Networks of Epigenetically Regulated Intra- and Interchromosomal Interactions.” *Nature Genetics* 38(11): 1341–47.

Zhou, G.-L. et al. 2006. “Active Chromatin Hub of the Mouse -Globin Locus Forms in a Transcription Factory of Clustered Housekeeping Genes.” *Molecular and Cellular Biology* 26(13): 5096–5105.

Zhou, Jie et al. 2002. “Replication and Subnuclear Location Dynamics of the Immunoglobulin Heavy-Chain Locus in B-Lineage Cells.” *Molecular and Cellular Biology* 22(13): 4876–89.

Zhu, Y. O., M. L. Siegal, D. W. Hall, and D. A. Petrov. 2014. “Precise Estimates of Mutation Rate and Spectrum in Yeast.” *Proceedings of the National Academy of Sciences* 111(22): E2310–18.

Zimmer, Christophe, and Emmanuelle Fabre. 2011. “Principles of Chromosomal Organization: Lessons from Yeast.” *Journal of Cell Biology* 192(5): 723–33.

Zink, Daniele et al. 2004. “Transcription-Dependent Spatial Arrangements of CFTR and Adjacent Genes in Human Cell Nuclei.” *Journal of Cell Biology* 166(6): 815–25.

Zirbel, R. M. et al. 1993. “Evidence for a Nuclear Compartment of Transcription and Splicing Located at Chromosome Domain Boundaries.” *Chromosome Research* 1(2): 93–106.

Zorio, Diego A.R., Niansheng Nick Cheng, Thomas Blumenthal, and John Spieth. 1994. “Operons as a Common Form of Chromosomal Organization in C. Elegans.” *Nature* 372(6503): 270–72.