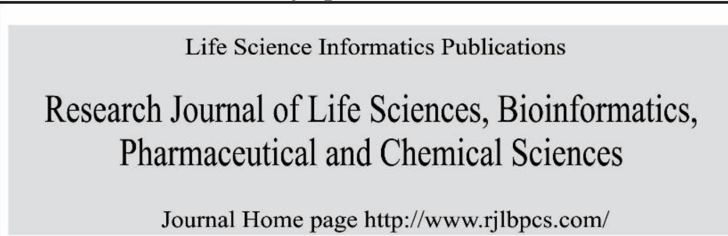
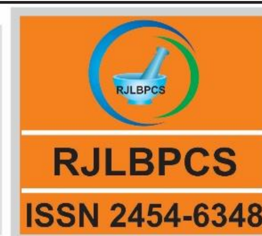




Original Research Article **DOI: 10.26479/2021.0703.05**

COMPUTATIONAL ANALYSIS OF HISTONE MODIFICATION AND TFBS THAT MEDIATES GENE LOOPING

Sudheer Menon

Department of Bioinformatics, Bharathiar University, Tamil Nadu, India.

ABSTRACT: Histones are essential proteins found in eukaryotic cells that structure DNA into units called nucleosomes. Histones are modified by transcription factor binding sites also known as TFBs which mediates gene looping/regulation. Gene loops are not static features of chromosomes but appear to form in response to the transcriptional status of the gene [1].in this study we did the analysis by using RNA seq, Chip-seq Hi-C, ChiA-PET (NGS) data. The Histone Modification data analysis was done and transcription factors like Pol II and CTCF were studied which revealed the role of Transcription Ends sites (TES) in gene looping[13].

Keywords: Histone modification, gene looping. Transcription factor binding sites.

Article History: Received: May 29, 2021; Revised: June 10, 2021; Accepted: June 25, 2021.

Corresponding Author: Prof. Sudheer Menon* Ph.D.

Department of Bioinformatics, Bharathiar University, Tamil Nadu, India.

Email Address: profsudhimenon@gmail.com

1.INTRODUCTION

Histones are essential proteins found in eukaryotic cells that structure DNA into units called nucleosomes. A nucleosome is a short strand of 146 ± 2bp of DNA wrapped by 1.65 left-handed super helical turns around a core of proteins called histones [2]. Histones are mostly made up of lysine and arginine. There are four different families of histones known as H1/H5, H2A, H2B, H3, and H4 from which H1/H5 are linkers and all others are core proteins. Histones generally

© 2021 Life Science Informatics Publication All rights reserved

transcribe and repress genes and also help in DNA repair. Methylation, phosphorylation and many other procedures help in histone modification, which then help in a covalent post-translational modification (PTM) of histone proteins.the structure of chromatin is modified by the histone PTMs which influence gene regulation [12]. A histone alters the gene in a way that neutralizes the positive charge on it by inserting acetyl groups on lysines in histone tails and loosens the grip of the nucleosome on DNA [17].This method allows the DNA to be accessed by transcriptional machinery, and genes are active. Other adjustments influence transcription in various ways. In gene alteration and gene looping, TFBs(transcriptional factor binding sites) also play a critical role.

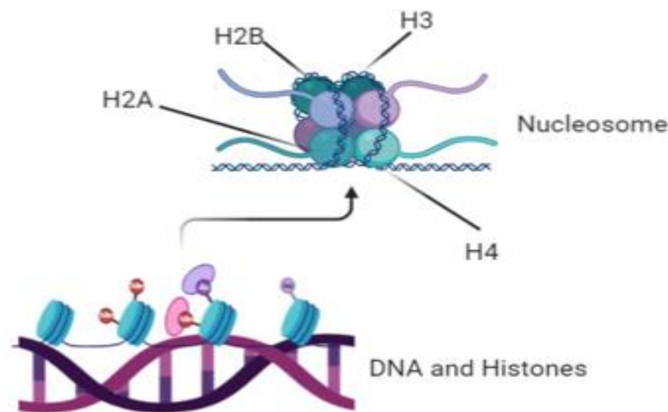
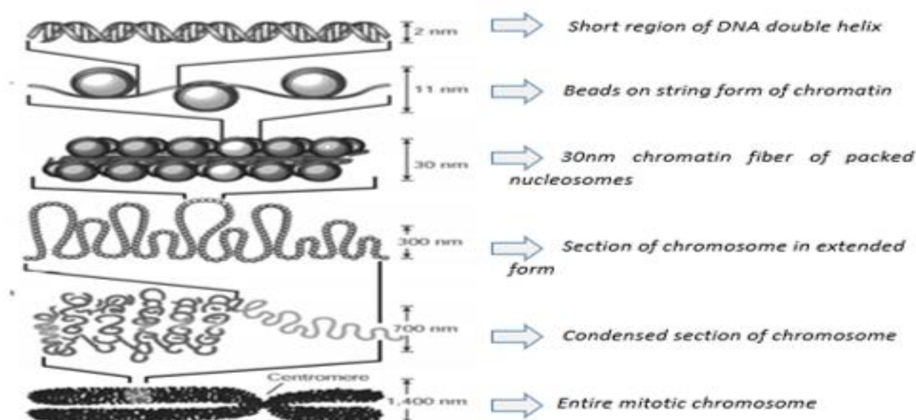


FIG 1 HISTONES AND ITS TYPES

Epigenetic modifications are known as modifications to chromatin, even if it's related to DNA or proteins that cannot change or modify the DNA sequence, and this can cause a visible effect on gene regulation[18]. Epigenetics is the investigation of aggregate changes that are not because of DNA succession modifications[14]. How epigenetic changes are reported through ensuing cell divisions stays questionable and whether epigenetic marks are the reason or result of memory.covalent adjustments of histone proteins happen on both histone tails and the globular histone place [3]. Eleven distinct classes of histone modifications are currently defined [4] and several enzymes are known to catalyze post-translational histone modifications [5]. enzyme complexes undergo mutations and cause malignancies in humans [15], Histone alterations are also a cause human cancers [6].

© 2021 Life Science Informatics Publication All rights reserved



A lot of studies were done in order to understand the histone modifications while time resolved the succession of epigenetics regulation during early mammalian development if we stick to mammals[20]. In early human and mouse embryos, complex landscapes of different layers of epigenetic details, including chromatin accessibility, histone modifications, DNA methylation, and higher-order chromatin structure, were only studied at 2-stage and 4-stage of the embryonic cell[6]. Accessibility to chromatin reflects TFBSs, many of which are found in enhancers. They found that in earlier embryonic development (2- to 8-cell stages), chromatin accessibility is closely correlated with gene expression, while the correlation decreases sharply in the ICM and mESC stages[19]. One plausible reason for this outcome is that early gene expression at that time is merely based on the state of chromatin [16]. Chromatin accessibility depends on further control by H3K4me3 in the 4-cell stage and spatial regulation by the 3D chromatin structure. Computational analysis of histone modifications show even better results that will be discussed further.

CHROMATIN MODIFICATIONS	MODIFIED FUNCTION	RESIDUE MODIFIED
Methylation	Transcription and repair	H3 and H4
phosphorylation	Transcription repair and condensation	H3, H2B and H4
ubiquitylation	Transcription and repair	H2A and H2B
sumoylation	transcription	H2A, H2B AND H4
Acetylation	Transcription repair and replication	H2A, H3 and H4
biotinylation	Gene silencing and mitotic condensation	H2A, H3 and H4

FIG 3 Role of Histone modifications in Gene regulation

© 2021 Life Science Informatics Publication All rights reserved
Peer review under responsibility of Life Science Informatics Publications

Basic concepts:

A comprehensive knowledge on transcription factors binding sites is necessary for the understanding of transcriptional regulations.following are some modifications done to the chromatin along with their functions after modification.

EXPERIMENTATION:

Different sorts of computational analysis were done on histone modifications and the transcriptional factors binding sites including:

1. yeast TFs
2. CRISPR
3. RNA seq
4. chIp seq
5. HI-C
6. 3C
7. CHiA-PET

And other new technologies are being used in order to better understand the process of gene regulation, histone modification and gene looping.

2. MATERIALS AND METHODS

Computational materials and methods usually require only machines and soft wares rather than wet labs and their equipment. To find out more about all the gene regulation in modernized and tech supportive studies following technologies and studies were used to analyze the process.

1. YEAST TFS:

There are two types of yeast TFs, one is sensitive and the other one is insensitive. Transcription factors regulate target gene expression by binding to specific genomic regions. In Saccharomyces cerevisiae, record factor (TF) restricting destinations (TFBSs) are regularly adjoining and upstream of target loci because of the reduced idea of the yeast genome. When binding, TFs associate with RNA polymerase II to repress and activate transcription. TFs likewise enroll chromatin alteration catalysts to actuate chromatin structure changes, which then influence the availability of components to genomic DNA areas. The target gene of a TF changes as per formative, physiological and extracellular natural conditions. Moreover, TFs associate with one another through combinatorial binding.

To contemplate this two separate histone change datasets were considered, the two of them had an open understanding edge (ORF) and the promoter region of binding targets for 37 yeast

© 2021 Life Science Informatics Publication All rights reserved

transcription factors. The two outcomes uncovered an unmistakable histone change design between the utilitarian protein-DNA restricting destinations and non-practical ones for practically 50% of all TFs tried. Such contrast is a lot more grounded at the ORF than at the advertiser locale. Furthermore, a protein-histone adjustment communication pathway must be construed from the functional protein restricting targets.the results propose that histone alteration data can be utilized to recognize the practical protein-DNA official from the non-useful, and that the guideline of different proteins is constrained by the change of various histone lysines, for example, the protein-explicit histone change levels.

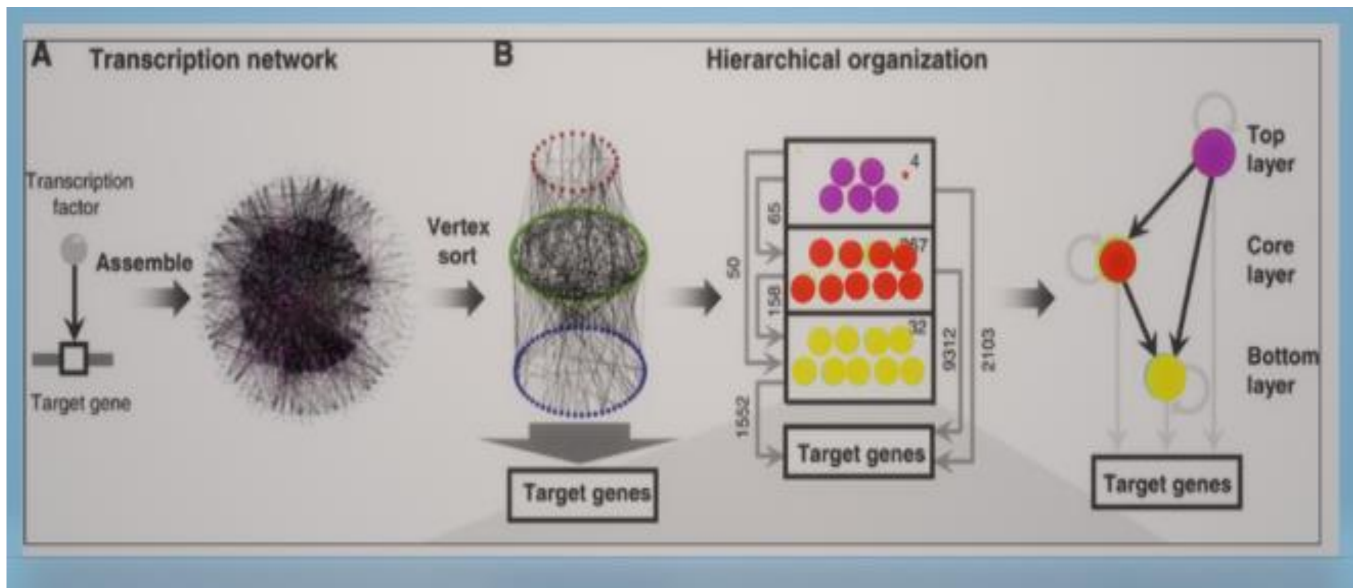


FIG 4 Hierarchical Organization Of Transcription

Studying the Genome-wide analysis of chromatin characteristics to identify histone change sensitive and insensitive transcription factors of yeast (chao cheng) and his associates proposed a strategy to anticipate the target of yeast transcription factor by coordinating the profiles of histone modification with transcription factor limiting binding motif details. In comparison to a binding motive only approach, it demonstrates enhanced prescient power. They discovered the classification of transcription factors into histone-sensitive.

© 2021 Life Science Informatics Publication All rights reserved

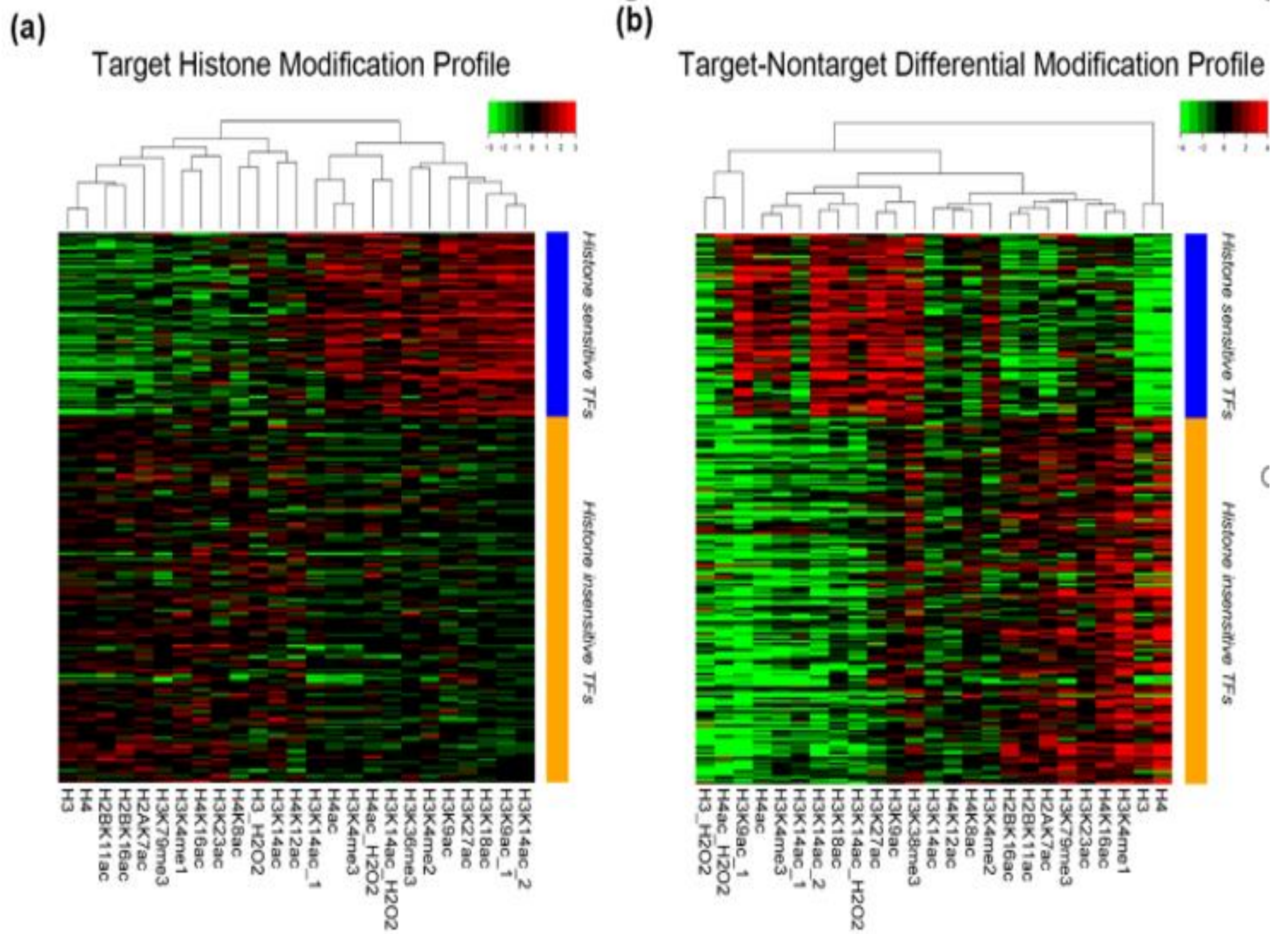


FIG 5: Target and non target histone modifications profile

© 2021 Life Science Informatics Publication All rights reserved

2. CRISPR CAS-9

CRISPR CAS-9 is short for clustered regularly interspaced short palindromic repeats and CRISPR-associated protein 9. The CRISPR-Cas9 framework has created interesting results in mainstream researchers since it is quicker, less expensive, more exact, and more productive than other existing genome altering system.CRISPR-Cas9 was adjusted from a normally occurring genome editing framework in microscopic organisms. The microscopic organisms catch pieces of DNA from attacking infections and use them to make DNA portions known as CRISPR arrays.

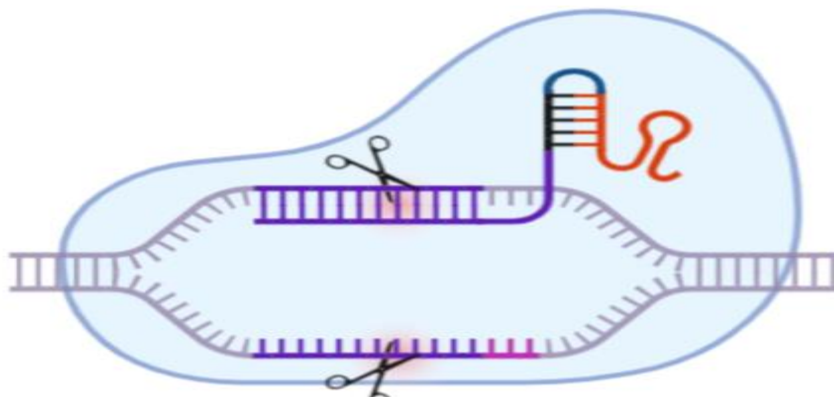


FIG 6 CRISPR CAS9 Scissors

A CRISPR interceded epigenetic altering system is utilized to regulate the gene expression by altered promoter methylation.

3. RNA seq: The presence of gene loops facing the promotor and terminator regions of genes with unusually long ORFs in yeast has been shown in recent studies. Here, we report that looping is not limited to long genes, but occurs with ORFs as short as 1 kb between the distal ends of genes [23]. We propose a model that suggests that TFIIB binds RNAP II to the terminator, which in turn binds to the scaffold promoter. RNA polymerase II (RNAP II) assembles with a transcription preinitiation complex (PIC) into a transcription preinitiation complex (PIC).Studies show that DNA loops are significant structures in mammalian cells that control gene expression. DNA looping juxtaposes promoter DNA in each case with enhancer elements that often lie far upstream [22]. For example, erythroid krüppel-like factor (EKLF) and CCCTC-binding factor (CTCF) are included in the mouse β-globin ACH, two gene-specific transcription factors needed for adult β-globin gene expression [21]. A similar structural structure has been identified for the interleukin genes in initiated mouse TH2 cells [7]. All things considered, transcriptional enactment is

© 2021 Life Science Informatics Publication All rights reserved

related with gene association into a progression of loops secured at each base to the nuclear scaffold protein special AT-rich binding protein 1 (SATB1) [8].

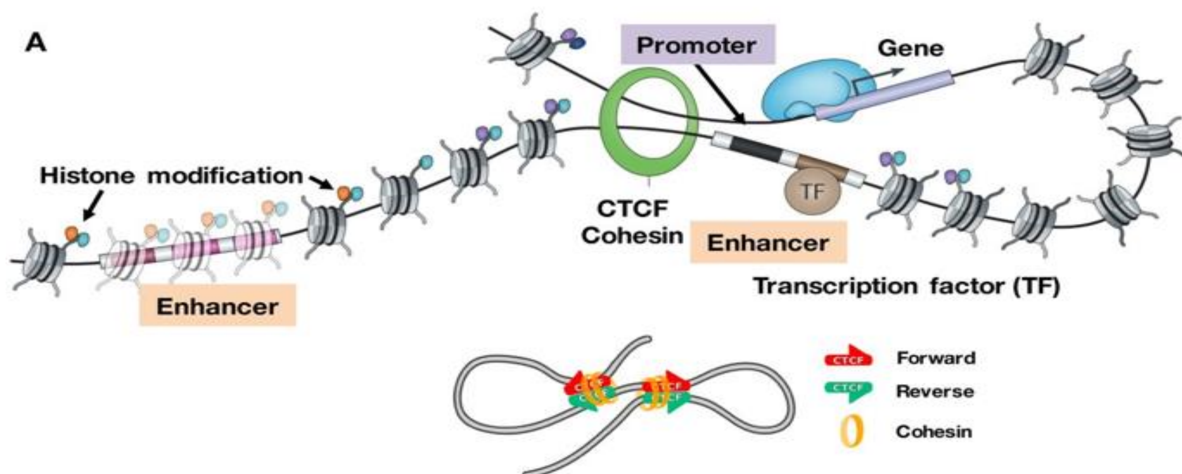


FIG 7 CTCF Binding

We demonstrated the existence of gene loops in the yeast Saccharomyces cerevisiae that precisely juxtapose the promoter and terminator regions of three genes [24]. For the SEN1 and BUD3 genes, loop formation was stated to be based on RNAP II and to involve the 3' end processing machinery components of Ssu72 and Pta1 [9]. These findings indicate that gene loops are complex structures that form following a previous round between promoter and terminator regions [25]. The general transcription factor TFIIB is an intended candidate for a protein that joins the promoter-terminator interaction. The genetic association with TFIIB was initially established by Ssu72 CTD phosphatase [10], and these two proteins interact directly in vitro [11]. We used "capturing chromosome conformation" (3C) and chromatin immunoprecipitation (ChIP) assays in the study mentioned here to show that looping is not prohibited to long genes but occurs at RNAP II-transcribed genes independent of gene length. In addition, we show that looping is based on TFIIB in a manner independent of its role in transcription. We propose a model that includes the interaction of TFIIB with RNAP II at the terminator, which then associates to form a gene loop with the promoter scaffold. Because of the technological advantages provided by their extremely long ORFs, these three genes were analyzed. We used the 3C assay to the SAC3 (3.9 kb), BLM10 (6.4 kb), GAL10 (2.1 kb), and HEM3 (1.0 kb) genes (Figures 1A–1D) to conclude whether looping is a more general occurrence and not restricted to long genes. We find that all four genes form loops and are detected regardless of whether a flanking gene

© 2021 Life Science Informatics Publication All rights reserved

or an intergenic region is located within the adjacent restriction site. We infer that gene looping is not idiosyncratic to long genes but can occur at genes as short as 1 kb in length.

3. RESULTS AND DISCUSSION

3.HI-C AND ChIP-seq:

A significant and technically challenging task is genome-wide mapping of three dimensional chromatin organization. We have developed a computational model incorporating Hi-C and histone mark ChIP-seq data to assist experimental effort and to understand the determinants of long-range chromatin interactions. The genome-wide distribution of histone modifications can be profiled using chromatin immunoprecipitation followed by high-throughput sequencing (ChIP-seq).Histone modifications typically serve as separate markers for transcriptional regulation and many other biological processes by regulating DNA accessibility and unique protein recruitment. Multiple Hi-C datasets have recently been created and stored in the public domain Different features such as chromatin compartments, topologically related domains (TADs), and chromatin loops are revealed by analyzing this knowledge. Mapping genome-wide chromatin interactions at high-resolution, however, remains difficult and expensive. In comparison, we have used ChIP-seq experiments that can be regularly performed at much lower cost by many laboratories, and a significant amount of data is already available in the public domain. It has been noted that interactions with chromatin are correlated with different patterns. By doing the computational analysis of this data, we predict that the genomic loci with frequent chromatin interactions serve as nucleation sites of looping and helps in gene regulation. This investigation demonstrated that chromatin interaction sites are enhanced with regulatory regions that help in histone alteration and gene looping. Histone marks are significant for pointing to the regulatory regions and TAD boundaries. And CTCF is the most instructive predictor for TAD boundaries.

© 2021 Life Science Informatics Publication All rights reserved

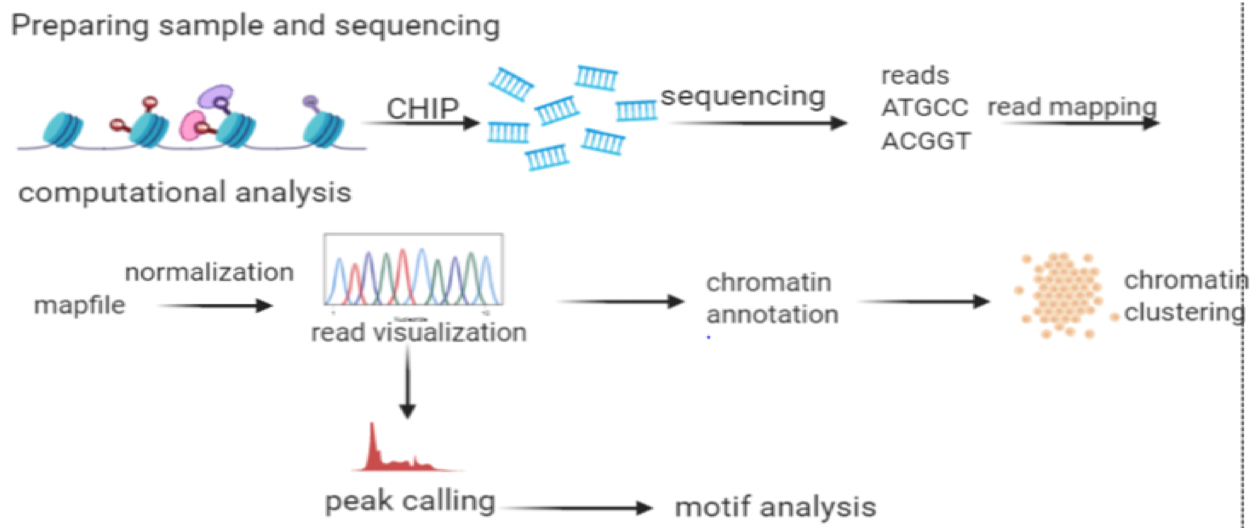


FIG 12 Transcription Models

4. ChiA-PET:

Long-range chromatin interactions need a huge transcription regulation feature. Chromatin Interaction Analysis with Paired-End-Tag Sequencing (ChIA-PET) is an emerging development that has a huge load of tendencies in the study of chromatin collaboration, thereby providing understanding in the transcription regulation examination of A million short sequences of tags. ChIA-PET provides a worldwide and unbiased questioning of high-order chromatin structures linked to particular protein factors by integrating Chromatin Immunoprecipitation (ChIP), proximity ligation, and high-throughput sequencing. In eukaryotes, cis administrative components and related executing factors control spatiotemporal quality articulation, which influences singular turn of events. Remarkably, *cis* regulatory elements such as promoters, enhancers, and insulators have distinct epigenetic features. The general process of doing chIp seq is shown below:

© 2021 Life Science Informatics Publication All rights reserved

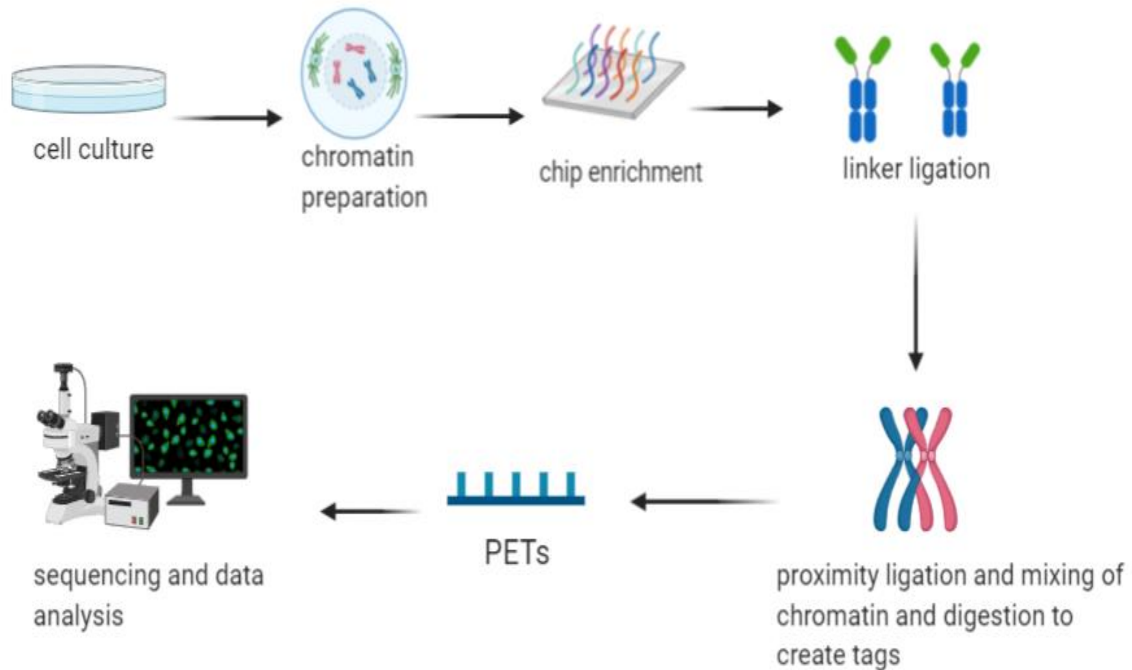


FIG 10 Mechanism of ChiA-PET

To study histone modifications through chiP sequencing technique we followed the main steps of the procedure that are:

© 2021 Life Science Informatics Publication All rights reserved

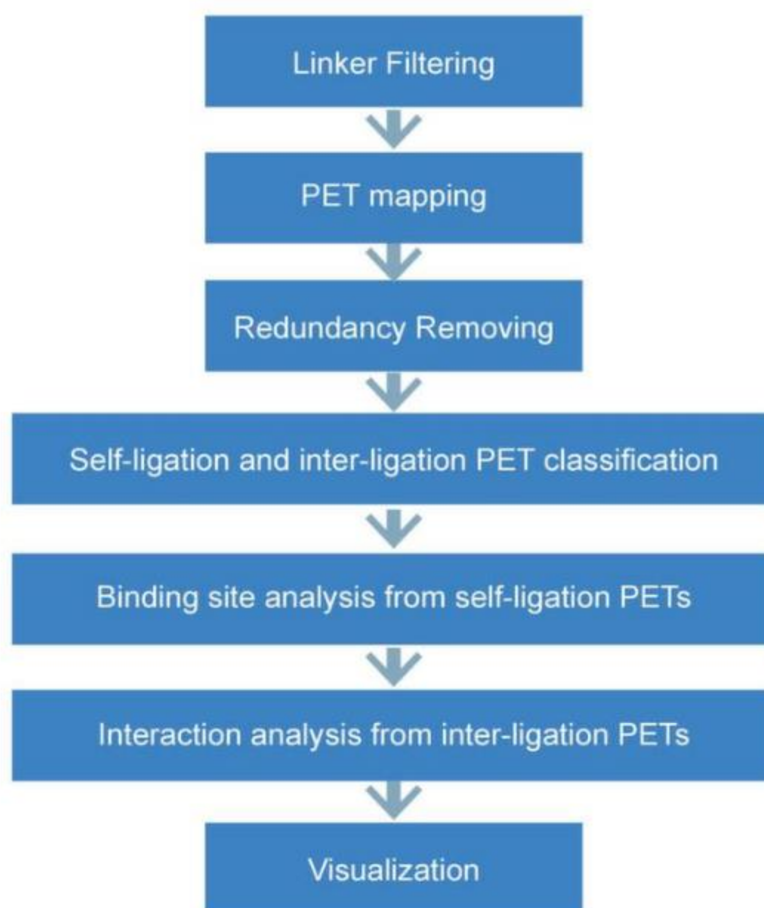


FIG 11 Steps of ChiA-PET

5. CTCF mediating transcription and gene regulation

CTCF are binding factors that mediate transcription in histones, adjust them and afterward control gene regulation.to study this We broke down CTCF chromatin immunoprecipitation sequencing, RNA sequencing, and Hi-C information, along with genotypes from a solid human accomplice, and measure factual relationship between singular changeability in CTCF binding and elective exon use. We show that CTCF-mediated chromatin loops are widespread between promoters and intragenic regions, and that CTCF binding connects with exons in spliced mRNA when exons are in real proximity to their promoters. The following figure shows CTCF along with its binding sites and respective factors which conclude the working of CTCF that is promoting and mediating gene regulation causing gene looping under the process of cohesion.

© 2021 Life Science Informatics Publication All rights reserved
Peer review under responsibility of Life Science Informatics Publications

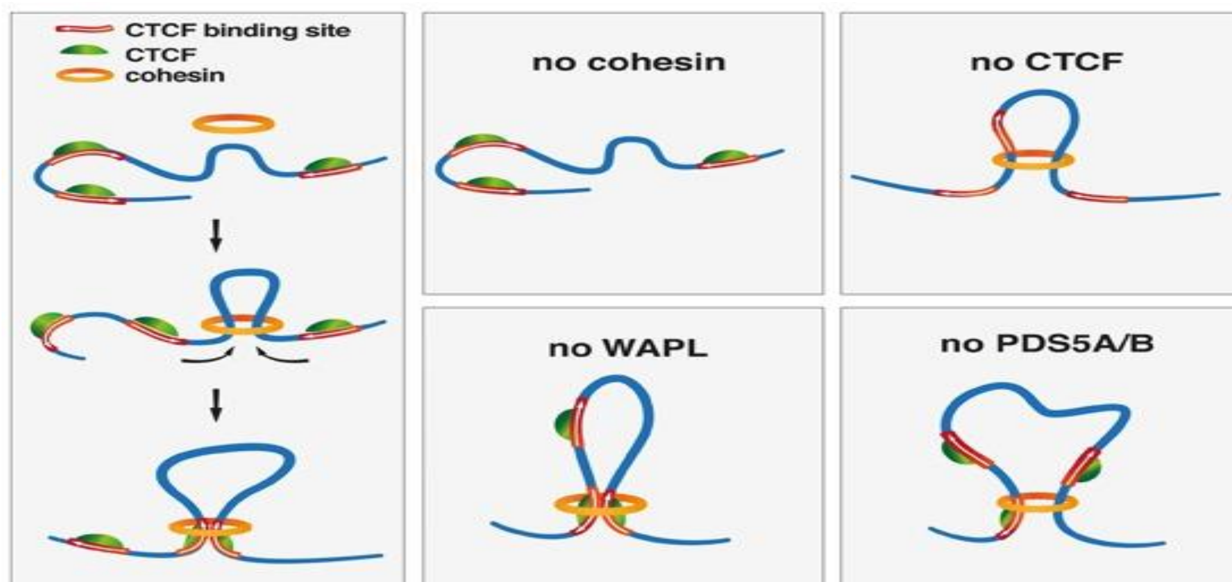


FIG 13 CTCF Regulating gene expressions

4. CONCLUSION

Histone modifications have always been involved in gene looping, evident from the studies earlier. and now By the computational analysis of the whole process using different computational tools and techniques like studying yeast TFs, using CRISPR tool, RNA,HI-C and chIp sequencing techniques, chIA-PET sequencing, and studying the CTCF mediating transcription and gene regulation we found out really intriguing and evident details about the procedure of gene looping and we conclude that histone modification under the influence of transcription factors plays the key role in the process of gene regulation and gene looping.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Not applicable.

HUMAN AND ANIMAL RIGHTS

No Animals/Humans were used for studies that are base of this research.

CONSENT FOR PUBLICATION

Not applicable.

AVAILABILITY OF DATA AND MATERIALS

The author confirms that the data supporting the findings of this research are available within the article.

FUNDING

None.

ACKNOWLEDGEMENT

We thank our anonymous referees for their useful suggestions.

© 2021 Life Science Informatics Publication All rights reserved
Peer review under responsibility of Life Science Informatics Publications

CONFLICT OF INTEREST

The author declares no conflict of interest in preparing this article.

REFERENCES

1. Bannister, A. J., & Kouzarides, T.. Regulation of chromatin by histone modifications. Cell Research, (2011), 21(3), 381–395.
2. Cai, C., Xu, C., Li, X., Ferguson, I., & Chen, K.. Accumulation of lignin in relation to change in activities of lignification enzymes in loquat fruit flesh after harvest. Postharvest Biology and Technology, (2006), 40(2), 163–169.
3. Santos-Rosa, H., & Caldas, C. Chromatin modifier enzymes, the histone code and cancer. European Journal of Cancer (Oxford, England : 1990), (2005) 41(16)
4. Singh, B. N., & Hampsey, M.A Transcription-Independent Role for TFIIB in Gene Looping. Molecular Cell, (2007). 27(5), 806–816.
5. Tan, M., Luo, H., Lee, S., Jin, F., Yang, J. S., Montellier, E., Buchou, T., Cheng, Z., Rousseaux, S., Rajagopal, N., Lu, Z., Ye, Z., Zhu, Q., Wysocka, J., Ye, Y., Khochbin, S., Ren, B., & Zhao, Y.. Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. Cell, (2011), 146(6), 1016–1028.
6. Uger, R. A., & Barber, B. H. Presentation of an influenza nucleoprotein epitope incorporated into the H-2Db signal sequence requires the transporter-associated with antigen presentation. The Journal of Immunology, (1997).158(2), 685 LP – 692.
7. Bannister, A. J., & Kouzarides, T. Regulation of chromatin by histone modifications. Cell Research, (2011) 21(3), 381–395.
8. Cai, C., Xu, C., Li, X., Ferguson, I., & Chen, K.. Accumulation of lignin in relation to change in activities of lignification enzymes in loquat fruit flesh after harvest. Postharvest Biology and Technology, (2006) 40(2), 163–169.
9. Santos-Rosa, H., & Caldas, C.. Chromatin modifier enzymes, the histone code and cancer. European Journal of Cancer (Oxford, England : 1990), (2005) 41(16), 2381–2402.
10. Singh, B. N., & Hampsey, M.. A Transcription-Independent Role for TFIIB in Gene Looping. Molecular Cell, (2007) 27(5), 806–816.
11. Tan, M., Luo, H., Lee, S., Jin, F., Yang, J. S., Montellier, E., Buchou, T., Cheng, Z., Rousseaux, S., Rajagopal, N., Lu, Z., Ye, Z., Zhu, Q., Wysocka, J., Ye, Y., Khochbin, S., Ren, B., & Zhao, Y.. Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. (2011) 11(12), 23–28

© 2021 Life Science Informatics Publication All rights reserved

12. Palstra RJ, Simonis M, Klous P, Brasset E, Eijkelkamp B, de Laat W. Maintenance of Long-RangeDNA Interactions after Inhibition of Ongoing RNA Polymerase II Transcription. PLoS ONE, 2008;3:e1661.
13. Buchenau P, Hodgson J, Strutt H, Arndt-Jovin DJ. The distribution of polycomb-group proteins duringcell division and development in Drosophila embryos: impact on models for silencing. J Cell Biol, 1998;141:469–81.
14. Grimaud C, Bantignies F, Pal-Bhadra M, Ghana P, Bhadra U, Cavalli G. RNAi components arerequired for nuclear clustering of Polycomb group response elements. Cell 2006;124:957–71.
15. Kagey MH, Melhuish TA, Wotton D. The polycomb protein Pc2 is a SUMO E3. Cell 2003;113:127–37.
16. Lanzuolo C, Roure V, Dekker J, Bantignies F, Orlando V. Polycomb response elements mediate theformation of chromosome higher-order structures in the bithorax complex. Nat Cell Biol, 2007;9:1167–74.
17. Engel JD, Tanimoto K. Looping, linking, and chromatin activity: new insights into beta-globin locus regulation. Cell 2000 ;100:499–502.
18. Bulger M, Groudine M. Looping versus linking: toward a model for long-distance gene activation.Genes Dev 1999;13:2465–77.
19. Martin DI, Fiering S, Groudine M. Regulation of beta-globin gene expression: straightening out thelocus. Curr Opin Genet Dev 1996;6:488–95.
20. Blackwood EM, Kadonaga JT. Going the distance: a current view of enhancer action. Science 1998;281:60–3.
21. Chambeyron S, Bickmore WA. Does looping and clustering in the nucleus regulate gene expression?Curr Opin Cell Biol 2004;16:256–62.
22. Gaszner M, Felsenfeld G. Insulators: exploiting transcriptional and epigenetic mechanisms. Nat RevGenet 2006;7:703–13.
23. Cremer T, Cremer M, Dietzel S, Muller S, Solovei I, Fakan S. Chromosome territories--a functionalnuclear landscape. Curr Opin Cell Biol 2006;18:307–16.
24. Dekker J, Rippe K, Dekker M, Kleckner N. Capturing chromosome conformation. Science 2002;295:1306–11.
25. Zhao Z, Tavoosidana G, Sjolinder M, Gondor A, Mariano P, Wang S, Kanduri C, Lezcano M, SandhuKS, Singh U, Pant V, Tiwari V, Kurukuti S, Ohlsson R. Circular chromosome conformation capture(4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomalinteractions. Nat Genet 2006;38:1341–7.
26. Hatzis P, Talianidis I. Dynamics of Enhancer-Promoter Communication during Differentiation-Induced Gene Activation. Mol Cell 2002;10:1467–77.

© 2021 Life Science Informatics Publication All rights reserved