Spatiotemporal Dynamics of Chromatin Condensation During Embryonic Stem Cells to Neuronal Differentiation

Thesis submitted to University of Hyderabad for the award of Ph.D degree in

Department of Animal Biology

By

G Prashanth Kumar

13LAPH17

Department of Animal Biology
School of Life Sciences
University of Hyderabad
Hyderabad-500046
Telangana, India

March 2022

University of Hyderabad
(A Central University by an act of Parliament)
Department of Animal Biology
School of Life Sciences
P.O. Central University, Gachibowli, Hyderabad-500046

CERTIFICATE

This is to certify that the thesis entitled **"Spatiotemporal Dynamics of Chromatin Condensation During Embryonic Stem Cells to Neuronal Differentiation"** submitted by **G Prashanth kumar** bearing registration number **13LAPH17** in partial fulfillment of the requirements for award of Doctor of Philosophy in the **School of Life Sciences** is a bona fide work carried out by him under my supervision and guidance.

This thesis is free from plagiarism and has not been submitted previously in part or in full to this or any other University or Institution for award of any degree or diploma. Parts of this thesis had been **Published in following publications:**

1. Chandradoss KR*, Prashanth kumar Guthikonda*, Kethavath S, Dass M, Singh H, Nayak R, Kurukuti S, Sandhu KS. Biased visibility in Hi-C datasets marks dynamically regulated condensed and decondensed chromatin states genome-wide. **BMC genomics**. 2020 Dec;21(1):1-5. (* Co First-author)

2. Sornapudi TR, Nayak.R, Guthikonda PK, Pasupuleti AK, Kethavat S, Uppada V, Mondal S, Yellaboina S and Kurukuti S **(2018)**. Comprehensive profiling of transcriptional networks specific for lactogenic differentiation of HC11 mammary epithelial cells. ***Scientific Reports***. 8:11777.

Presented in the following conferences:

1. EMBO India conference on Regulatory Epigenomics, 2019 held at Chennai.
2. All India Cell Biology Conference, 2015 held at Trivandrum.
3. Genome Architecture and cell fate regulation conference, 2014 held at Hyderabad,

Further, the student has passed the following courses towards fulfillment of coursework requirement for Ph.D.

Course Code	Title of the Course	Credits	Pass/Fail
1. AS 801	Analytical Techniques	4	PASS
2. AS 802	Research Ethics, Data Analysis and Biostatistics	3	PASS
3. AS 803111	Lab Work and seminar	5	PASS

K. Sreenivasulu

Supervisor

Dr. K. Sreenivasulu
Professor
Department of Animal Biology
School of Life Sciences
University of Hyderabad
Gachibowli, Hyderabad-500 046. T.S.

K. Sreenivasulu

Head of the Department

अध्यक्ष / HEAD
जंतु जैविकी विभाग
Department of Animal Biology

Dean of the School

DEAN
School of Life Sciences
University of Hyderabad
Hyderabad-500 046.

UNIVERSITY OF HYDERABAD

Central University (P.O.), Hyderabad-500046, INDIA

DECLARATION

I hereby declare that the results of the study incorporated in the thesis entitled **"Spatiotemporal Dynamics of Chromatin Condensation During Embryonic Stem Cells to Neuronal Differentiation"** has been carried out by me under the supervision of **Prof. Sreenivasulu Kurukuti** and this work has not been submitted for any degree or diploma of any other university earlier.

Dated: 09.03.2022

G. Prashanth kumar

G Prashanth kumar

13LAPH17

Acknowledgements

- First and foremost, I am extremely grateful to my supervisor Prof. Sreenivasulu Kurukuti for his invaluable supervision, great support and encouragement throughout my work. His immense knowledge and plentiful experience have helped me a lot in troubleshooting and developing new ideas in the work, without whom I wouldn't have completed my thesis.
- I express my sincere gratitude towards my Doctoral committee members Dr. A Bindu Madhava Reddy and Prof. S Rajagopal for monitoring and guiding my work regularly.
- I thank Head, Dept of Animal Biology, Prof. Sreenivasulu kurukuti, and previous HOD's Prof. Anita Jagota, Prof. Jagan Pongubala, and Prof. Senthil kumaran for the departmental facilities.
- I thank the Dean, School of Life Sciences, Prof. S Dayananda, and former Deans Prof. K Ramaiah, Prof. P Reddanna, Prof. A.S. Raghavendra, Prof Aparna Dutta Gupta for providing the central facilities at the School of Life Sciences
- I thank our collaborators Dr. Sailu Yellaboina, Prof. Ranjith Padinhateeri, Dr. Kuljeet Sandhu for their valuable guidance in my work.
- I also thank the students of our collaborators Dr. Keerthivasan, Dr. Gaurav Bajpai, Bindia and Manisri for their help in my work.
- I thank my current labmates kethavath Sreenivas, Yuvasri Golivi, Netrika Tiwari, Satyanarayana, Sharmistha, Aditya and previous labmates Trinadharao Sornapudi, Rakhee Nayak and Sukalpa Mondal for providing a supportive lab environment and also helping me in my work and thesis.
- I especially thank kethavath Sreenivas for his help in coding throughout my work.
- I thank my university friends Rohith konada, Narahari, Damuka Naresh and Ram Gopal, and my close friends Kishore, Sagar, Anil, Govardhan, Amardeep and Siva Sankar for their all-rounded support.
- The financial support from DBT, NIAB, CSIR, ICMR, DST, UGC and IoE is highly acknowledged.
- Fellowship support from CSIR-UGC JRF and SRF is highly acknowledged.
- My family deserves endless gratitude: my parents for having immense belief in me and making me who I am today, my life partner Poornima for her love, support and for being my side in any adversity, my lovely son Chiku for his mischievous plays and unconditional love, and, my sister Swathi for morally supporting me at all times
- Last but never the least, I owe my sincere thanks to the Almighty.

- Prashanth

Table of contents

Abbreviations

1D : 1-Dimensional

3D : 3-Dimensional

CT: Chromosome Territory

PcG : Polycomb Group

PRCs : Polycomb Repressive Complexes

NL : Nuclear Lamina

µM: micrometre

nm : nanometer

3C : Chromosome Conformation Capture

4C: Chromosome Conformation Capture-on-Chip

5C: Chromosome Conformation Capture Carbon Copy

Hi-C : High-throughput Chromosome Conformation Capture

CTCF : CCCTC-binding factor

ICE : Iterative Correction and Eigen-value decomposition

CCDD : Contact Correction through Distance Decay plots

CC map: Chromosome Compaction map

DNase-seq : DNase digestion followed by sequencing

MNase-seq: Micrococcal Nuclease digestion followed by sequencing

NGS : Next Generation Sequencing

ATAC-seq: Assay for Transposase-Accessible Chromatin using sequencing

PCR : Polymerase Chain Reaction

REs : Restriction Enzymes

ESC: Embryonic Stem Cells

NPC : Neural Progenitor cells

CN: Cortical Neurons

NSC : Neural Stem Cells

AST : Astrocytes

FLiv : fetal Liver

ALiv : adult Liver

RES : Restriction Enzyme Sensitivity

ChIP-seq : Chromatin Immunoprecipitation followed by sequencing

cLAD: conserved Lamina Associated Domains

ciLAD: conserved inter Lamina Associated Domains

CTN: Chromosome Territory Neighborhood

RED-seq: Restriction Enzyme Digestion followed by sequencing

SRA : Sequence Read Archive

GEO : Gene Expression Omnibus

LOESS : Local Polynomial Regression

KO : Knocked-Out

WT: Wild Type

GFP: Green Fluorescent Protein

FPKM: Fragments Per Kilobase of transcript per Million mapped reads

GO: Gene Ontology

UCSC: University of California Santa Cruz

PCA: Principal Component Analysis

Chr: chromosomes

FISH: Fluorescence in situ Hybridization

FBS: Foetal Bovine Serum

Mb: Megabase

Kb: Kilobase

A.U: arbitrary units

TAD: Topologically Associated Domain

TSS: Transcription Start Site

miRNA : micro RNA

Rg : Radius of gyration

1. Introduction:

The life of every multicellular organism starts with a single-cell zygote. The zygote divides and forms 2 cell, 4 cell and 8 cell stages, then forms morula, blastocyst, Gastrula and into an adult organism. Hence, essentially all the cells of an organism originate from a single cell and thus have identical genomes (Gilbert., 2000). Despite this fact, there exists a vast diversity in the cellular phenotypes and functions they perform in a single organism. This is possible by selective and tight control of expression of genes which are turned on and off during development to form various cell types (Hoopes., 2008). This strict and controlled gene regulation happens at multiple levels. At the level of genes, regulation happens primarily during transcription and most of the transcriptional control happens at transcription initiation (Cooper., 2000). There are multiple modes of regulation both at genetic and epigenetic levels. Genetically, the promoter sequences and binding of transcription factors play a role. Epigenetically, post-translational modifications of histone proteins such as methylation and acetylation, and DNA modifications such as DNA methylation play a significant role in controlling gene expression (El-Osta et al., 2000). At post-transcriptional level, RNA splicing, microRNAs, Long Non-Coding RNAs, RNA binding proteins play a major role in gene regulation. At translational level, there are many factors such as translational initiation, elongation and post-translational modification. All of the above factors are synergistically orchestrated and regulated differentially in each type of cells, or during disease or in aging to generate different phenotypes. All of these factors are more or less well studied and the knowledge has been accumulating since many decades. Another layer of gene regulation uncovered and has been gaining attention, which is regulation of gene expression at pre-transcriptional, which include factors such as chromatin accessibility, chromatin condensation, 3D genome organization, replication timing (Tsompana et al., 2014). '3D organization of the genome' refers to the systematic packaging and topology of the DNA inside nucleus in three dimensions and its dynamics in space and time. It has been well established in the recent years that the 3D organization of genome plays a major role in regulating cell type specific gene expression, DNA repair, differentiation, embryonic development, several diseases and aging (Gorkin et al., 2014; Zheng et al., 2019; Babu et al., 2015). Studying the principles behind formation, functioning and dynamics of 3D genome architecture helps us understand various life processes and disease. The relation between chromatin condensation and 3D genome organization has been poorly studied and remains elusive.

1.1 The Cell Nucleus and Nuclear structure:

Eukaryotic nucleus is typically characterized by presence of double membrane, membrane-less organelles, tightly packed genome, chromocenters, nucleolus and actin meshwork (Nature Education., 2010). The nucleus is considered as the brain of the cell and many important functions take place inside the cell nucleus. It is the seat of the genetic material of the cell, which dictates the whole life processes and passes from generation to generation. In a typical electron micrograph of mammalian cell nucleus one can observe the presence of electron dense regions, which are condensed heterochromatic regions, and electron poor regions, which are decondensed euchromatic regions. Generally, heterochromatin is predominantly seen at the nuclear periphery often associated with nuclear lamina. Some heterochromatin is also present

in the inside of the nucleus which is mostly facultative in nature. Euchromatin is predominantly present in the interior of the nucleus (Peric-hupkes et al., 2010). Also, can be seen the presence of electron dense, highly condensed nucleolus, which is transcriptionally highly active and is the site of rRNA synthesis and ribosome subunit assembly (Pederson et al., 1998; Boisvert et al., 2007). Hence, although condensed heterochromatin is predominantly present at the periphery and decondensed euchromatin is in the interior of the nucleus, facultative heterochromatin and condensed nucleolus are present in the interior of the nucleus (Peric-hupkes et al., 2010). Chromocenters are formed by the association of pericentric heterochromatin which was shown to maintain the integrity of chromosome territories (Jagannathan et al., 2018; Jagannathan et al., 2019).

1.1.1 Chromosome Territories:

It is well known that mitotic chromosomes are highly condensed and can be seen as distinctive chromatids under a microscope (Ohta et al., 2011). It was long thought that the chromosomes in the interphase nuclei are randomly distributed throughout the volume of the nucleus. But with the advent of the techniques such as chromosome painting and other molecular studies, it became evident that the chromosomes occupy discrete locations inside the interphase nucleus termed chromosome territories (CT) (Cremer and Cremer., 2010). The chromosome territory and its neighborhood positioning is cell type specific and aids in the cell-type specific gene expression patterns (Cremer et al., 2001).

1.1.2 Nucleus is structurally and functionally compartmentalized:

Chromatin inside the nucleus is broadly compartmentalized into heterochromatin which is present at the nuclear periphery and euchromatin which is in the interior of the nucleus. Not only the chromatin is compartmentalized inside the nucleus, also various functions inside the nucleus are compartmentalized (Gavrilov et al., 2015). These specific functions such as splicing, assembly of small nuclear ribonucleoproteins (snRNPs), regulation of Long non-coding RNAs are carried at nuclear domains such as PML nuclear bodies, histone locus bodies, nuclear speckles, peri nucleolar compartment, polycomb bodies, cajal bodies, which are the sites for various nuclear functions (Spector et al., 2001). The formation of these nuclear domains was shown to be facilitated by liquid-liquid phase separation phenomenon (Lesne et al., 2019). For instance, RNA Pol II is not distributed all along the nucleus but are found clustered at specific locations termed transcriptional factories. Genes from linked and unlinked chromosomes coalesce along with DNA binding proteins to form transcription condensates for coordinated control of multiple genes (Osborne et al., 2004; Cho et al., 2018)

1.2 Epigenetic Characteristic Features of Active and Silent Genes in the Genome

If a gene is repressed or in OFF state, the DNA is often associated with methylation at cytosine residues. This is called DNA methylation (Robertson et al., 2005). Studies involving DNA methylation at genome level have given substantial understanding regarding regulation of ESC, differentiation of hematopoietic cells and pathways of cellular reprogramming (Law et al., 2010; Ji, H. et al., 2010). The part of the genome is tightly wrapped around the nucleosomes and is often associated with repressive histone modifications such as H3K9me3, marker for constitutive heterochromatin, and H3K27me3, a mark of polycomb repressors. Heterochromatin is often characterized by hypoacetylation of histones, H3K9me3 and DNA cytosine methylation (5mC) (Tamaru et al., 2010). H3K9me3 interacts with heterochromatin protein 1 (HP1) and forms heterochromatin (Nakayama et al., 2001). Some regions such as telomeres are always heterochromatinised whereas there are some regions which are in heterochromatin form in a cell type specific manner and these regions are referred to as facultative heterochromatin. H3K27me plays a major role in the polycomb-mediated silencing pathway and formation of facultative heterochromatin (Cao et al., 2002; Cao et al., 2005). Polycomb group of proteins (PcG) repress transcription and play a crucial role in gene regulation. They are in the form of multi-protein complexes, known as polycomb repressive complexes (PRCs) (Croce et al., 2013). PcG proteins are comprised of PRC1, PRC2, and PhoRC. Importantly, PRC2 has subunits such as EED, EZH2 and SUZ12 and participates in the recruitment of H3K27me3 (Wang et al., 2015). It is often highly condensed and is present at the nuclear periphery associating with nuclear lamina. H3K9me establishment and binding of HP1/Swi6 proteins to H3K9me is crucial for the assembly of heterochromatic structures (Nakayama et al., 2001). This binding also establishes a platform for further recruitment of other chromatin remodelers such as HDACs (Yamada et al., 2005). RNAi also plays a significant role in the formation of heterochromatin (Martienssen et al., 2015).

On the other hand, if a gene is active or in ON state, it is often not methylated. The part of the genome is loosely wrapped around the nucleosomes and is often associated with active histone marks such as H3K4me3, H3K4me1 and H3K36me3. It is usually decondensed and present in the interior of the nucleus. Enrichment of H3K4me1 is primarily considered as the mark of enhancers in the human cells (Heintzman et al., 2007). High levels of trimethylation (H3K4me3) predominantly mark active or poised promoters (Calo et al., 2013). H3K36me3 levels are generally well correlated with gene expression levels. Recent studies have noted that expressed exons were significantly enriched for H3K36me3 (Ong et al., 2011) compared with introns. Recently, Tom Misteli's group showed that the splicing of FGFR2 gene was altered by varying the H3K4me3 and H3K36me3 levels (Luco et al., 2010).

1.3 Lamina Associated Domains:

Another important feature of the mammalian nucleus is the presence of lamina associated domains (LADs). These are the chromatin regions associated with nuclear lamina, especially with Lamin proteins (Guelen et al., 2008). There are three types of lamins: lamin A, lamin C and lamin B, the latter having two subtypes: laminB1 and B2 (Adam et al., 2012). LADs are

often heterochromatic in nature and are associated with repressive epigenetic marks such as H3K4me2. Most of the LAD regions are relatively conserved across many cell types and they are called constitutive LADs (cLADs). The inter-LAD regions are also conserved across many cell types are called constitutive inter LAD regions (ciLADs). Facultative LADs are cell type specific and are not conserved. LADs are dynamically switched between the cell types during the course of differentiation (Peric-hupkes et al., 2010). The nuclear lamina (NL) facilitates a large surface area that functions as an anchoring platform for chromosomes (Chubb et al., 2002; Van Steensel and Dekker, 2010). Interactions between nuclear lamina and the LADs dictate significant constraints on the tethering and positioning of chromosomes. Changes in the spatial positioning of LADs occurs after mitosis i.e., relocation of LADs towards and away from the lamina requires cell cycle progression. Upon cell division, the position of the LADs is not inherited but believed to be rearranged stochastically. The genes physically located in the LADs are believed to be transcriptionally silent and often associated with H3K9me2 (Kind et al., 2013).

1.4 Genome Packaging and Chromatin Condensation

The DNA in the eukaryotic genome is organized and packed in such a way that nearly 3.4 billion base pairs (nearly 2 meter in length) fit into a volume of the nucleus of approx. 10µM in diameter (Olins et al., 2003). Though it is densely packed, DNA needs to be accessed by many DNA binding proteins and other enzymes for nuclear processes like transcription, replication and recombination etc. Moreover, the DNA should not get tangled or knotted in the process of dense packaging. DNA is compactly packed inside the nucleus and is hierarchically organized (Lieberman et al., 2009). Not only the linear sequence of the DNA is important, but also the 3-Dimensional arrangement and organization of the DNA inside the nucleus dictate differential gene expression programs (Gorkin et al., 2014). The 3D genome architecture is also associated with many physiological conditions, aging, cellular response to stress, cancer (Zheng et al., 2019; Babu et al., 2015).

At the very basic level, the double-stranded DNA is wrapped around histone proteins resulting in formation of nucleosome. Each nucleosome consists of two each of the four histones H2A, H2B, H3, and H4, which together form a histone octamer (Thomas & Kornberg, 1975). Approximately 1.7 turns or nearly 146 base pairs of DNA is wrapped around the histone octamer. (Van Holde, 1988; Wolffe, 1999). A linker DNA of approx. 20 base pair length runs between nucleosomes finally giving a beads on a string appearance (Annunziato et al., 2008). The chromatin is further folded into what is often referred as ‘30 nm fiber’, which is approx. 30nm in width (Woodcock, 2005). Various models have been proposed to explain the formation of secondary structure of chromatin, all of which can be broadly classified into two classes. One is Single-start helices model, which are solenoids, and another one is two start helices model (Luger et al., 2012). The solenoid model proposes that DNA is tightly packed around nucleosomes and further into a helix and that the linker DNA is bent inside the fiber (woodcock et al 1984; Dorigo et al 2004). The two-start helices model proposes that straight linker DNA segments connect between two helices of nucleosomes forming two helices

running parallel to each other. The Nucleosome-nucleosome interactions further strengthen the compaction and stability of the chromatin. However, recent studies by using genome-wide interaction studies have highlighted the fractal geometric organization of chromatin at various levels (Lieberman et al., 2009; Rao et al., 2014).

3-Dimensionally, chromatin is hierarchically organized in a fractal globule manner. Broadly, each chromosome territory is segregated into active compartment A and inactive compartment B. The Hi-C intra-chromosomal contact maps first showed the presence of a plaid pattern in the correlation heatmap. By further performing PCA (Principal Component Analysis) or eigen vector decomposition, the length of the chromosome could be divided into compartment A and B based on the conventional positive and negative values, respectively. Hi-C matrix shows that interactions of the regions within compartments are enriched and between compartments A and B are depleted (Lieberman et al, 2009). DNA sequences in compartment A generally consists of active, transcribed genes and euchromatic in nature. Similarly, compartment B consists of inactive genes and is generally heterochromatic (Lieberman et al., 2009; Kalhor et al., 2012; Sexton et al., 2012). Later, Rao et al in 2014, based on higher resolution of the Hi-C data, sub-divided compartments A and B further into A1-A2 and B1-B3 based on histone modifications and replication timing (Rao et al., 2014). Nonetheless, it is widely accepted that each chromosome territory is broadly segregated into active compartment A and inactive compartment B (Lieberman et al., 2009).

Further down the hierarchy, the genome is further organized into self-interacting domains termed topologically associated domains (TADs). These are generally of size of around 800kb -1.2Mb in length (Dixon et al., 2012). In Hi-C matrix, the TADs show up as regions which have higher interactions among itself than with the neighboring or other regions. Though first observed in mouse cell lines, TADs were also observed in other mammalian cell lines (Nora et al., 2012) and non-mammalian organisms like Drosophila (Sexton et al., 2012) and Zebrafish (Gomez-marin et al., 2015) etc. TADs are considered to be conserved across cell types and even among different species (Dixon et al., 2012; Dixon et al., 2015). TADs are considered as functional unit of genome organization (Dekker and Heard 2015). TADs facilitate interactions between promoters and enhancer elements within the same TAD over the interactions between different TADs (Dixon et al 2015, Zhan et al., 2017). TAD boundaries are enriched with transcription start sites, housekeeping genes, H3K4me3, H3K36me3, CTCF binding sites (Dixon et al., 2016). Downregulation of CTCF by RNAi technique resulted in enhanced inter-TAD interactions suggesting the role of CTCF in the formation and maintenance of TADs (Nora et al 2017). Recently, loop extrusion model has been put forward and been widely accepted as the driving principle behind the formation of TADs (Fudenburg et al., 2016; Sunburn et al., 2015; Ganji et al., 2018). Disruption to the TAD boundaries may lead to gene dysregulation and disease (Harewood et al., 2017). TADs are further divided into sub-TADs, the boundaries of which are also enriched with CTCF and cohesion (Rao et al., 2014; Cremins et al., 2013). Further down the lane, at sub-megabase scale, high-throughput conformation data signifies the presence of long-range interactions primarily between enhancers and promoters. These interactions are essential for the establishment of cell type specific gene regulatory programs.

1.5 Chromatin Accessibility:

Understanding the mechanisms by which the genome folds and compacts inside the nucleus is one of the most fundamental questions. Chromatin accessibility is the ability of the chromatin to be available for access to or binding of different proteins and enzymes for transcription, replication, DNA repair, enzyme digestion (Radman-Livaja et al., 2010; John et al., 2011). Generally, heterochromatin is tightly packed and is less accessible, whereas euchromatin is loosely packed and is more accessible (Duggan and Tang, 2010). They are bound to nucleosomes with lesser frequency. Thus, regulatory DNA could be assessed by studying open/accessible regions in the chromatin (Tsompana et al., 2014). The nucleosome positioning across the genome plays important regulatory role by restricting the accessibility of the DNA to various transcription factors (TFs). This consequently influences many processes such as replication and DNA repair, apart from transcription (Radman-Livaja et al., 2010). Most of the TFs that are studied in the ENCODE project tend to bind exclusively to open and accessible chromatin albeit with the exception of a few TFs that are seen enriched in either facultative or constitutive heterochromatin (Thurman et al., 2012). Recently, it has also been shown that the chromatin accessibility differs between nucleosome variants and depends on the factors required for their deposition (Chen et al., 2014).

Changes in the genome architecture have implications in disease and aging, which could be owed to mutations in chromatin modifiers that alter the nucleosome positioning on the DNA (Gaspar-Maia et al., 2009, Hargreaves et al., 2011, Schwartzentruber et al., 2012**)** and by disruption of chromosome neighborhood in activation of proto-oncogenes (Hnisz et al., 2016) and other diseases (Anania et al., 2020; Norton et al., 2017). Therefore, lot of emphasis has been put on determining genome-wide chromatin accessibility data, and locating corresponding epigenetic changes during cellular differentiation, disease and aging. One of the huge projects, ENCODE (Bernstein et al., 2012) has been a major contributing part towards this end.

Traditionally, Sucrose gradient sedimentation was used to separate the decondensed and condensed chromatin, and CsCl density gradient centrifugation to analyze the chromatin (Allan et al., 1981; Caplan et al., 1987; Gilbert et al., 2001), albeit with very low resolution. Further, procedures which make use of DNase I and MNase to digest the chromatin were one of the first to provide evidences that transcriptionally active parts of chromatin correlate well with the chromatin accessibility (Keene et al., 1981; Levy et al., 1981).

Recent advances in next generation sequencing (NGS) have given rise to many high-throughput assays to assess chromatin accessibility based on NGS. To study chromatin accessibility many methods have been developed based on the shielding of nucleosome bound DNA from endonuclease digestion. The endonuclease, DNase I favorably digests nucleosome depleted open DNA (Wu et al., 1980). Thus, genomic regions which are prone to digestion by DNase I, also known as DNase I hypersensitive sites (DHSs), could be deciphered by sequencing (DNase-seq) (Boyle et al., 2007). FAIRE-seq (Formaldehyde-Assisted Isolation of Regulatory Elements followed by deep sequencing) is another method used to identify accessible chromatin, in which chromatin is crosslinked with formaldehyde and nucleosome-

depleted DNA fragments are isolated and sequenced (Wai et al., 2011). Majority of TFs bind to the open DNA and thus DNase-seq and FAIRE-seq results often show enrichment of regulatory regions viz., the promoters and enhancers. On the contrary, digesting the DNA with micrococcal nuclease (MNase) followed by high-throughput sequencing of the regions shielded from getting digested (MNase-seq) provides the data on the occupancy levels and positions of nucleosomes (Hoeijmakers et al., 2018). When compared with nucleosome position maps, both FAIRE-seq and DNase-seq could be used to decipher nucleosome-free regions which generally are in range of 100-300 bp. Consequently, quantitative variations in accessibility of the DNA that are located around or within nucleosomes are hard to identify using the above methods (Chen et al., 2014).

ATAC-seq is a recent method developed for identifying accessible chromatin. This method involves fragmentation and integrating of DNA transposase into regulatory regions by hyperactive Tn5 (Goryshin et al., 1998; Adey et al., 2010; Buenrostro et al., 2013). In ATAC-seq, a population of nuclei are tagged with sequencing adapters *in vitro* using hyperactive mutant Tn5 transposase. Most of the adapters get integrated into open chromatin and are then sequenced. But ATAC-seq could only potentially identify nucleosome positioning in the regions which are close to open chromatin

For more than 30 years, restriction enzymes (RE) are being utilized to assess chromatin accessibility (Liberator et al., 1984; Almer et al., 1986; Ohkawa et al., 2006). The principal advantage with RE is that they cut the DNA at specific locations called restriction sites (RS). Thus, using PCR, one could detect and measure the cell-type specific variations in accessibility. In theory, the chromatin accessibility at any given loci could be measured since restriction sites are present almost all over the genome. Recently, Chen et al. (2014) employed a method termed RED-seq (Restriction Enzyme Digestion followed by sequencing) to identify open chromatin in different cell types. They found that RED-seq could quantitatively measure the cell type specific differences in chromatin accessibility. Importantly, RED-seq was also able to measure differences in the chromatin accessibility even within nucleosomes.

1.6 Epigenetic Memory of Chromatin Compartments and Neighborhood are Passed through Mitosis

The cell type specific epigenetic information is passed to the daughter cells through mitosis. Raphaël Margueron and Danny Reinberg. (2010) highlighted the mechanisms which are required for effective transmission of epigenetic information which are mostly associated with histones and chromatin organization in the nucleus (Margueron et al., 2010).

The above-mentioned condensation and decondensation states of chromatin, and their epigenetic memory are passed down to the daughter cells through mitosis. During prometaphase of the cell cycle, the chromatin condenses to from chromatids. They are aligned along metaphase plate during metaphase. The two sister chromatids separate during anaphase and the nuclear envelope reforms during telophase which is followed by the cytokinesis (Mitchison et al., 2001). In early G1 phase, the chromatin is decondensed but it is in the late

G1 phase the chromatin is fully compartmentalized into heterochromatin which moves to the periphery of the cell nucleus and euchromatin which resides in the interior of the cell nucleus.

1.7 Spatial reorganization of genome during cellular differentiation

The Embryonic stem cells are differentiated into different types of cells which is accompanied by lineage restriction, significant gene expression changes, gene repositioning within the nuclei, major changes in A/B compartments and extensive reorganization of 3D nuclear architecture. One of the first studies regarding gene repositioning during cellular differentiation showed that Beta-globin gene is relocated from nuclear periphery to the interior in the later stages of erythroid differentiation (palstra et al., 2003). This repositioning was seen to be accompanied by increase in the long-range interactions between B-globin gene and locus control regions (Krivega and Dean, 2012; Krivega et al., 2014). Another study showed that de novo long-range chromatin interactions were established during differentiation of B-cells due to repositioning between enhancer and VDG gene cluster (Guo et al., 2011). Hi-C is a method to measure the genome wide chromatin interactions which would allow one to assess the interactions between gene-gene and gene-regulatory elements across the genome (Lieberman et al., 2009). Hi-C analysis of ESC differentiation revealed dynamic switching of compartments between different cell types (Dixon et al., 2015). Bonev et al. (2017) showed that during differentiation of mESCs to cortical neurons, there was an overall increase in the size of compartment, a significant spike of interactions among B compartments and decline in interactions among A compartments. Analysis of TADs showed that most of the TAD boundaries during differentiation were conserved between cell types, although a minor portion of the TADs are dynamic and reorganize during differentiation. These changes in TADs correspond to both intra and inter-TAD interactions are also associated with corresponding epigenetic and transcriptional changes (Dixon et al., 2015). It was also shown that hierarchical folding of TADs is dynamically reorganized during differentiation of mESC to cortical neurons (Fraser et al., 2015). HOX genes, which are master regulators of embryonic development, provide valuable insights into the 3D genome reorganization during differentiation. During development of limb, HOXD genes are transcriptionally repressed in ESC and are sequentially activated. It was noted that silenced HOXD clusters are located in a single repressed chromatin compartment which contains all the genes. Developmental signaling in HOXD clusters is accompanied by corresponding changes in gene expression and long-range interactions between and within the chromatin compartments (Andrey et al., 2013).

Differentiation of ESC to neural progenitor cells (NPC) and further to cortical neurons (CN) represent a very robust *in vitro* differentiation system (Gaspard et al., 2008; Gaspard et al., 2009). This differentiation system has two main advantages. First of all, this *in vitro* differentiation system is well established and very robust. There are lot of genome-wide datasets available from public repositories pertaining to this model system (Edgar et al., 2002). Secondly, this differentiation model captures a ground ESC state, an intermediate NPC and then finally a terminally differentiated Neuronal state i.e., ground to terminal differentiation in

just three cell types. Thus, many aspects of the differentiation could be studied through this model system.

In stem-cell based therapies, generation of neurons is the key step towards treating the neurological disorders or damaged tissue. Because of their proliferative and pluripotent nature, embryonic stem cells (ESCs) have a huge potential in regenerative medicine and could replace fetal or adult-derived central nervous system (CNS) tissue (Cai et al., 2007).

1.8 Methods to study chromatin interactions

Having said the importance of studying 3D organization, scientists from decades have been trying to determine the interactions between genomic elements inside the cell nucleus. Traditionally microscopic techniques such as FISH (Fluorescence in situ Hybridization) have been used to decipher the genomic interactions (Edelmann et al., 2001). But studying the genome wide 3D organization using FISH is limited by its low resolution, low throughput and specificity of probe sequence. First direct evidence to show long range chromatin interactions were deciphered for mouse Beta-globin locus vs LCR interaction by using a method called "RNA Traps" (Carter et. al., 2002). However, with the advent of chromosome conformation capture (3C) method by Job Dekker et al., in 2002 it has become easier to elucidate the 3D organization of genome in an unprecedented way. Delaat group has extensively worked on the 3D organization of mouse beta globin locus by using 3C and showed existence of 3-dimentional active chromatin hubs orchestrated by gene promoters and enhancers.

The principle behind 3C to decipher genomic architecture is based on measuring the contact frequency between DNA segments in population of cells (Dekker et al., 2002). Various derivatives of 3C were developed to further study the 3D chromatin architecture in detail. These include 4C (Simonis et al., 2006), 5C (Dostie et al., 2006), Hi-C (Lieberman et al., 2009) and ChIA-PET (Li et al., 2010). Recently, single-cell Hi-C was developed to decipher the chromatin interactions on a single-cell level, instead of cell populations (Nagano et al., 2013).

Hi-C is the widely utilized and prominent variant of 3C which uses high-throughput next generation sequencing technologies enables one to determine the spatial organization of entire genome, chromosomes and sub-chromosomal domain within the cell nucleus. Hi-C read pairs give interaction data on all-vs-all genomic segments basis. One can retrieve intra as well as inter-chromosomal interactions based on Hi-C data. Hi-C protocol begins by fixing the cells in formaldehyde to crosslink the DNA and proteins. The cross-linked chromatin is digested with restriction enzyme of choice and the cut ends are filled with biotin-linked nucleotides. This is followed by blunt-end ligation where the far away regions in the genome are ligated together due to close proximity to generate hybrid junctions. The DNA is purified, sheared, size selected and the biotinylated junctions are pulled down using streptavidin beads which is followed by high throughput paired-end sequencing and further analysis using computational and statistical tools (Lieberman et al., 2009). The resulted sequenced reads are aligned to the reference genome, filtered and contact matrices are generated.

The genomic regions are divided into equal sized bins based on the sequencing depth and objective of the study, which is referred to as resolution. The combined interaction read pairs of all the regions between two bins is calculated as interaction frequency between those two bins. The pairwise interaction frequencies between all the bins are represented as a symmetric matrix if it's intra-chromosomal or asymmetric matrix if it's inter-chromosomal. This matrix is called a contact matrix or contact map. The contact matrix is represented as a heatmap, where, generally the intensity of the colour is proportional to the interaction frequency between two regions or bins. Thus, self-interacting domain like structures could be seen as squares along the diagonal depending upon the length of the domains.

1.9 Hi-C Data Normalization methods:

The sequencing data from the Hi-C experiment comes in the form of raw fastq files. Generally, the raw sequences that are obtained by different sequencing technologies suffer from technical biases (Imakaev et al., 2012). The raw reads undergo a number of quality measures and are trimmed subsequently (Bolger et al., 2014). Then they are aligned to the reference genome with required parameters and relaxations. There are several pipelines published in literature which take care of the pre-processing of the raw reads to the post-processing of the aligned reads like Juicer, HiC-Pro (Servant et al., 2015), Hi-C pipe (Castellano et al., 2015), and HiCUP (Wingett et al., 2015). Apart from the post alignment processing steps such as removing commonly encountered Hi-C artifacts and putative PCR duplicates, there exists several other biases in the Hi-C data itself and these need to be normalized as well. These biases include GC content of the reads, mappability, length of the restriction fragment, library size as well as other unknown factors. Downstream analysis like interpreting the 3D genomic structure and comparison between samples would be erroneous without normalizing these biases. Yaffe and Amos. (2011) first addressed these biases and proposed a potential way to normalize for these biases. Subsequently, many others have come up with new methods and algorithms to correct for these biases (Hu et al., 2012; Schmitt et al., 2016; Cournac et al., 2012). These different normalization methods can be broadly divided into two categories; Explicit and Implicit.

Explicit normalization is a type of normalization, the biases are defined explicitly such as GC content, mappability, RE site density etc. and are corrected for these biases and are accounted for in the statistical model. Probabilistic models are used to correct for these biases. There are several explicit normalization tools developed such as HiCNorm (Hu et al., 2012). Implicit normalization is a type of normalization in which the biases are not defined separately but the sequencing coverage of particular region is considered as a function of all the biases (known and unknown) combined. This normalization mostly uses matrix-balancing algorithms and regression models. Some of the implicit normalization methods include KR Normalization (Knight et al., 2013), Vanilla coverage Normalization (Rao et al., 2014), Iterative correction and eigen-value decomposition (ICE) (Imakev et al., 2012).

However, current normalization methods seem to be very biased in deciphering the interaction data and subsequent depiction of spatial organization of chromatin. Moreover, there exists lot

of inconsistency in terms of Hi-C derived interactions vs true in vivo interactions measured by in situ hybridization methods (Williamson et al., 2014). Hi-C data could not be able to decipher the occurrence of nucleolus, chromocenters and others, suggesting that still there requires normalization of Hi-C data. However, efforts in this direction are lacking. One major drawback of the above-mentioned normalization methods is that neither of them addresses the compaction status of heterochromatin and decompacted state of Euchromatin. In this current work, we have identified hitherto unidentified normalization step based on the restriction enzyme digestibility of a given chromatin region based on the chromatin compaction. In our analysis (Chandradoss et al., 2020), we found that heterochromatic regions yield less no. of sequencing reads than euchromatic regions. This would be due to the fact that heterochromatin is less accessible to restriction enzymes than euchromatin in the Hi-C protocol. This issue has not been addressed in any normalization methods so far. We try to address this problem by a new method named CCDD (Contact Correction through Distance Decay plots) in this work.

1.10 3D modeling of Chromatin using Hi-C data:

Hi-C data could be potentially used to model 3-Dimensional conformation of chromatin using in silico polymer simulations. There are many types of polymer models that could predict 3D structure of chromatin from the Hi-C contact matrix. They are Homopolymer model, Co-polymer model and Loop extrusion model. Each model has its own advantages and limitations. None of the above models explicitly study organization of heterochromatin and euchromatin.

As discussed earlier, the no. of reads from a given genomic bin depends on its compaction status and since the homopolymer model assumes each bin to be bead of equal size, we, together with our collaborators, propose a novel heteropolymer model in which the size of each bead on the string is a function of no. of 1D reads coming from that bin (genomic length) i.e., beads in the heterochromatin region are smaller in size compared to the beads of euchromatin. This model takes into account the compaction/decompaction status of the chromatin.

In this current study, we took advantage of Hi-C methodology to ask if repurposing of Hi-C data can be used to determine the chromatin compaction status, simply by read counts per unit of chromatin regions of one dimensional Hi-C datasets. By using repurposed Hi-C data, we deciphered the chromosome compaction status and generated chromosome compaction maps (CC map) of all the mouse chromosomes. Further, by studying the CC Map of ESC and differentiated cell types, we showed the dynamics of chromosome condensation and its relationship with gene expression and histone modifications. Further, we assessed the developmental dynamics of chromatin compaction in in vitro Embryonic stem cell (ESC) differentiation to Neural progenitors (NPC) and Cortical neuron (CN).

2. Materials and Methods

Unless otherwise stated, two-sided wilcoxon rank sum test (Mann-whitney test) was used to determine statistical significance. All the statistics are performed in R.

1. **Visualization of DNase-seq data:** DNase-seq is performed to determine regulatory regions in the genome. DNase-seq data for mouse Embryonic stem cells was downloaded from ENCODE (accession number: ENCSR000CMW) (Vierstra et al., 2014) to compare with 1D read counts of Hi-C and RED-seq data. We obtained BED files of mapped and processed reads of DNase-seq from (https://www.encodeproject.org/eperiments/ENCSR000CMW/). The data consisted of genomic coordinates of DNase 1 HSS and were visualized in UCSC genome browser (kent et al., 2002).

2. **RED-seq data processing:** RED-seq stands for Restriction Enzyme Digestion followed by sequencing, a technique used by Chen at al. (2014) to determine the accessible regions in the genome. To prove our hypothesis of biased restriction digestion, we utilized the data generated from sequencing of in-situ restriction enzyme digested chromatin of cross-linked cells and compared it with that of in-solution restriction enzyme digested naked DNA (RED-Seq). First, we collected the bedgraph files (mm9 version) of mapped and processed reads of in-situ restriction digested chromatin and in-solution restriction digested naked DNA of mESC from GEO series number GSE51821. The files were then converted to mm10 version assembly 'liftover' tool (http://genome.ucsc.edu/cgi-bin/hgLiftOver) from using UCSC genome browser. The reads were binned into 10 kb bins i.e., all the number of read counts within a 10kb bin were summed. The binned reads were then corrected for the of restriction site density and the GC content one after the other, as described in the next section. To crosscheck if our observations are not the artefacts of lifting over, we also mapped the raw reads (SRA files) to mm10 version genome assembly. 'SRA' files are first converted to 'fastq' format using 'fastq-dump' function of 'SRA toolkit'. Reads were mapped to mouse genome assembly (mm10) using bowtie2. The mapped reads were sorted using 'sort' function and then PCR duplicates were removed using 'markdup' function of SAMtools. The results from mapping with bowtie2 were consistent with the results obtained through lifting over the downloaded bedgraph data.

3. **Hi-C data processing using HiCUP:**
   Hi-C sequenced reads from any experiment are compressed and deposited in the NCBI in the form of SRA (Sequence Read Archive). We obtained the SRA files of Bonev et al. (2017) from GSE96107 and converted to fastq files using NCBI SRA toolkit. We utilized HiCUP pipeline from Babraham Institute, Cambridge to process the Hi-C reads. The forward and reverse reads were mapped separately to the mouse reference genome assembly mm10. Invalid and duplicated read pairs are further filtered out in HiCUP. The generated BAM files contain both forward and reverse read pairs. Then we summed the read counts into 10kb bins and these were referred to 1-dimensional (1D) raw Hi-C reads.

4. **Correction of 1D read counts and generation of RZ scores:**
   We used loess regression to correct the 1D Hi-C reads for the biases. To correct the 1D read counts for the NGS sequencing machine bias (where it is known that GC regions of genome are more represented that AT regions of the genome) the parameters of Loess regression were first optimized in such a way that the reads in naked DNA corrected with Loess form a straight line against GC content. First, we corrected the 1D reads for RE density in the 10kb bins, which varies from bin to bin. We derived the residuals of read counts after performing loess regression against RE density of 10kb bins. Next, we corrected for the GC content bias by deriving residuals of RE-corrected read counts through loess regression against GC content of corresponding 10kb genomic bins. These corrected 1D reads were then normalized to the total number of sequenced reads genome-wide for that sample. We further converted these values to Z-scores (measure of no. of standard deviations above or below the population mean a value is) using median centering and were referred to as RZ-Scores.

5. **Derivation of Condensed and decondensed domains:**
   We derived the decondensed and condensed domains using the strategy given by Guelen et al. (2008). In summary, the bins were first defined as + 1 and − 1 depending upon whether the values were +ve or -ve in the RZ-scores. The domain boundaries were identified by subtracting the average of 20 windows on either side of uniformly distributed (per 10 kb) reference points. We determined a cut-off on this value by randomization of the read counts in the genome and keeping the false discovery rate to < 5%. By calculating the relative proportion of positive and negative values in each inter-boundary region, we demarcated condensed and decondensed domains. We set the minimal proportion of either positive or negative values to 0.8 in order to define the domains as decondensed and condensed respectively.

6. **Analysis of cLADs and ciLADs :**
   LADs (lamina associated domains) are genomic regions which are attached to nuclear lamina. cLADs (constitutive LADs) and ciLADs (constitutive inter-LAD regions) are the regions which are conserved across the cell types. We downloaded cLAD and ciLAD boundaries from GSE17051. To scrutinize our claims of biased accessibility in cLAD domains, we collected the WT and the lamin knockout (KO) Hi-C data of mESCs (Zheng et al., 2018), corrected for RE density and GC content, summed the reads into 10 kb bins and used log ratio of Lamin KO to WT for comparing the change in the accessibility.

7. **Analysis of histone modification (ChIP-seq) data:**
   Sources of ChIP-Seq datasets are provided in Table 1. We analyzed H3K4me1, H3K27ac, H3K36me3, H3K4me3 to study active histone modifications and H3K9me3 to study silent/repressive histone modifications. We obtained pre-processed ChIP-seq data from the source, binned the reads at 10 kb resolution, combined into a table and quantile normalized the values using 'normalize.quantiles' function of preprocessCore from R-package (https://github.com/bmbolstad/preprocessCore). To characterize the condensed and decondensed domains derived using our method and see how various histone marks are associated with these regions, we have taken genome wide average boundary (+/- 1mb) of decondensed and condensed regions and derived the genome wide average histone marks associated with them. The left side of the border is decondensed and right side is condensed. We used ggplot2 from R-package (https://ggplot2.tidyverse.org/) to scatterplot the data.

8. **Analysis of polytene and diploid Hi-C data of Drosophila:**
   Drosophila polytene chromosome is a classic example of presence of condensed and decondensed chromatin. Eagen et al. (2015) has performed Hi-C on polytene chromosome. We obtained Hi-C data from GSE72510 (polytene, dm6) and GSE63518 (normal diploid Kc167, dm6) and further processed using HiCUP pipeline. We summed the reads at 5 kb bins and corrected for RE density and GC content as before. We also downloaded polytene TADs and lifted over the coordinates to dm6 assembly. Since the authors showed that polytene bands correspond to TADs, we mapped 5 kb bins to polytene TADs and considered those inside TADs as polytene band bins and rest as inter-band bins.

9. **Allele-specific Hi-C analysis of mouse X-chromosome:**
   Mouse female cells' nucleus contain one active and one condensed inactive X-chromosome. We wanted to see if we can recapitulate the condensed state of inactive X chromosome. DNase Hi-C is a Hi-C variant in which DNase I is used for chromatin fragmentation instead of Restriction Enzymes. Since we have two X chromosomes, allele-specific Hi-C data is used to distinguish between them. We obtained the allele-specific valid DNase Hi-C read pairs of brain and patski cells (nephron cells) from GSE68992. We ignored the reads which are mapping to both the references and binned the allele-specific reads at 20 kb resolution bins to get 1D read counts. We corrected the 1D read counts for GC content using loess regression as before.

10. **Normalization of Hi-C data:**
    To normalize for known systematic biases like mappability, RE density and GC content in Hi-C contact maps, we used HiCNorm (http://www.people.fas.harvard.edu/~junliu/HiCNorm/). For implicit correction, we used iterative correction and eigen vector decomposition (ICE) packages (https://github.com/mirnylab).

11. **Generation of compressed Chromosome condensation maps: Restriction enzyme sensitivity map:**

    1D HI-C reads were obtained as discussed previously using HICUP. The reads were normalized to the restriction enzyme bias. To generate peak representation, reads falling on every 100kb were allotted to first four bases of each bin thus generating chromosome wide view of peaks. For scaling, average of all intensities was considered from that sample and accordingly plotted and scale has been defined. The reads were visualized in UCSC genome browser and were referred to as restriction enzyme sensitivity (RES) map. A vertical flip of the image is appended to the original image to obtain vertically compressed chromosome condensation (CC) maps. To generate a realistic in vivo condensation state of chromosomes, horizontal condensation maps were generated. The width of each bin was taken as a direct function of number of 1D reads from that bin, indicating the degree of condensation horizontally by giving the colour gradient from red to yellow, where red represents compacted regions and yellow represents decompacted regions.

12. **ESC culture and Neural differentiation:**

    Monolayer differentiation of mouse embryonic stems cells was done based on the protocol as published Gaspard et al. (2009). Feeder independent mESCs were cultured in ESC medium (10% FBS) upon 0.1% gelatin-coated plates up to 40-60% confluency. Differentiation was initiated by the addition of DDM media containing N-2 supplement (Gibco#17502048) and is counted as day 0. From day 2, 1X Cyclopamine (Merck#C4116) treatment was initiated and continued till day 10. On day 12, cells were dissociated with 0.05% trypsin (diluted freshly with PBS and pre-heated at 37 degrees Celsius) and proceeded for centrifugation with 10% PBS (PBS with 10% FBS). Cell pellets were dissociated properly and plated onto Poly-L-Ornithine and Poly-Laminin coated plate with N2B27 medium without vitamin A. FGF2 was also added to media for an increase in survivability of progenitor cells. Neuronal Progenitor cells were harvested on day 14, after 2 days of plating. For Cortical Neuron differentiation, cells were continued to be cultured in N2B27-A culture medium till day 21

13. **Validation of gene expression by Real Time Polymerase chain reaction (RT-PCR):**

    Total RNA was isolated by TRIZOL reagent (Ambion ref no. 15596026). cDNA synthesis was done using 1ug of input RNA by using cDNA synthesis kit (Biorad Iscript cDNA synthesis kit). RT-PCR assay were performed in total of 10ul reaction volume with 20ng of cDNA used as input. For SYBER green detection, we used KAPA SYBER GREEN universal mix (KAPA#KK4618) and PCR reaction conditions were followed according to the manufacturer. Gapdh was used as a housekeeping gene for neurogenesis All the data were analyzed by using the -2 $\Delta\Delta$CT method (Livak and Schmittgen, 2001).

14. **RNA-Seq data analysis:**
    Raw RNA-seq data fastq files were downloaded from GSE96107. Raw sequences were checked for quality using FASTQC, trimmed using Trimmomatic and then mapped to mouse reference genome mm10 version using TopHat 2.0 with parameters set to default. Fragments Per Kilobase of transcript per Million mapped reads (FPKM) values are derived using "Cufflinks".

15. **Calculating correlation between gene expression and condensation:**
    The FPKM values of the genes present in every single 10kb region has been averaged to generate the expression values at 10kb bins. For the genes which span between two bins, the gene is considered into the bin in which more than 50% of the gene body lies. Between ESC and NPC, the bins which have both +ve or –ve values from differential gene expression and differential RZ scores were considered positively correlated and the bins which have one +ve and one –ve value in differential gene expression and differential RZ scores were considered negatively correlated. The negative correlated genes were further classified based on the upregulation/downregulation status of H3K27ac and H3K9me3. The GO terms associated with those genes were derived using DAVID (https://david.ncifcrf.gov/summary.jsp).

16. **Generation and Normalization of Hi-C contact maps:**
    The BAM files generated using HICUP are split into forward and reverse files using SAMtools. Then a directory was made in HOMER using the sam files. Contact Hi-C contact matrices were generated using HOMER at required resolution. The contact matrices were generated as raw, coverageNorm (previously SimpleNorm) and iteratively balanced using the commands in the source. The heatmaps are visualized using Java Treeview.

17. **3D polymer modeling of Hi-C data:**
    Nearly all polymer models that study Hi-C-based 3D chromatin structure assume that all beads have same size. But this is contradictory to what we know about chromatin. Different regions of chromatin have different compaction and hence it can have different sizes. Hence, we modelled the chromatin assuming chromatin as a heteropolymer and size of monomers of this polymer are according to their compactness. Our model explains nature of heterochromatin and euchromatin domains in different cell types. We performed this analysis and modelling in collaboration with Prof. Ranjith Padinhateeri and his student Dr. Gaurav Bajpai from IIT, Bombay. The following methods and formulas were adapted from Dr. Gaurav Bajpai's Ph.D thesis. let the diameter of the i^{th} bead be σ_i. Each bead would have an "equilibrium" distance of $r_i^{(s)}$. Total stretching energy of the chromosome was calculated by

$$E_s = \sum_i k_s(\mathbf{r}_i - r_i^{(s)})^2$$

Where r_i is the distance between i^{th} and $(i + 1)^{th}$ beads. k_s and $r_i^{(s)}$ are spring stiffness constant and equilibrium distance between two neighboring beads, respectively. We took

$$r_i^{(s)} = \frac{\sigma_i + \sigma_{i+1}}{2}$$

To account for steric hindrance, we used Lennard-Jones potential which is given by

$$E_{LJ} = 4\epsilon \sum_{\substack{i,j \\ j>i}} \left[\frac{\sigma_{ij}^{12}}{\mathbf{r}_{ij}^{12}} - \frac{\sigma_{ij}^{6}}{\mathbf{r}_{ij}^{6}} \right]$$

where $r_{ij} < 2^{1/6}\sigma_{ij}$ and zero otherwise. Here r_{ij} is the distance between i^{th} and j^{th} beads. σij is the effective LJ size parameter that depends on size parameters of i^{th} and j^{th} beads and have relation $\sigma_{ij} = \frac{\sigma_i + \sigma_j}{2}$

18. **Usage of Hi-C experimental data to generate 3-Dimensional model:** we converted contact frequency matrix data into contact probability matrix (CPM). We used upper triangular matrix of CPM in our simulation. After excluding diagonal and upper diagonal numbers, CPM number was represented by P^c_{ij} . To calculate contact energy, we generated a random number (r_{rand}) between 0 to 1. If $r_{rand} < P^c_{ij}$, we inserted a bond and computed contact energy using spring energy formula given by

$$E_c = \sum_{ij} k_c (\mathbf{r}_{ij} - r_{ij}^c)^2$$

If there is no bond ($r_{rand} > P^c_{ij}$) , this energy will be zero. k_c and r^c_{ij} are spring stiffness constant and equilibrium distance between i^{th} and j^{th} beads. r^c_{ij} is calculated same as σ_{ij} . We repeat this process many times and we do many simulations. Total energy of the system is calculated by

$$E_{tot} = E_s + E_c + E_{LJ}$$

This system is simulated using Brownian dynamics by solving the Langevin equation given by

$$\mathbf{r}_i(t + \Delta t) = \mathbf{r}_i(t) - \frac{\Delta t}{\gamma m} \nabla_{\mathbf{r}_i} E_{tot}(t) + \sqrt{\frac{6 k_B T \Delta t}{\gamma m}} \boldsymbol{\xi}_i(t)$$

where m is bead mass, Δt is time-step, Υ is damping contact and ξ is the thermal noise experienced by each bead.

19. **Parameters used in Heteropolymer model simulation:** To model chromatin as a heteropolymer, we consider the diagonal elements of the Hi-C contact matrix, obtained from experiments. First, we assumed that each bead will have a size which is related to the number of contacts measured in experiments; hence we decide the size of a bead for the heteropolymer using the following formula

$$\tilde{\sigma}_i = \frac{1 + \mathrm{e}^{(1-\langle D_i \rangle / D_i)}}{2}$$

where D_i is the diagonal value of Hi-C experiment matrix (M_{ij}) and ‹Di › is the average of all diagonal elements. We performed brownian dynamics simulation for 10^6 -time steps where t is taken 65s.

20. **Generation of genome wide chromosome neighborhood maps:**
Intra-chromosomal contacts were removed in raw genome-wide matrices at 1Mb resolution. Then vanilla coverage normalization was used to normalize inter-chromosomal contact matrices within sample. All the elements of the matrix corresponding to an inter-chromosomal pair were summed up and divided by the product of lengths of the corresponding chromosomes of that pair to account for contact frequency biases due to different lengths of the chromosomes. This normalized matrix (20 X 20) was termed chromosome territory neighborhood (CTN) map. Subtraction matrices were visualized using GI Tools where mean was taken as middle value during heat map generation. chromosomes 1-8 were considered to be larger chromosomes and chromosomes 9-19 were considered as smaller chromosomes. Clustering plots were plotted as box plots where interactions among larger chromosomes and among smaller chromosomes were plotted. Similarly, clustering in the subtraction matrices were plotted where change in the interactions among smaller chromosomes, among larger chromosomes and between smaller and larger chromosomes are plotted and unpaired wilcoxon test was used to calculate the statistical significance. P-value less than 0.05 was considered statistically significant.

21. **Generation of distance decay plots**:
All the elements in the intra-chromosomal contact matrix generated using SimpleNorm at 100kb resolution from ESC, NPC and CN were antilog2 transformed and then log10 transformed since HOMER outputs log2 transformed values. The mean interaction frequency of all the contacts at a given distance (in multiples of 100kb) was calculated and was plotted as line plots to get the distance decay plots. ‘Geom' smooth function in ‘ggplot2' package in R was used to smoothen the lines, and the smoothening method was put to default. To avoid extreme outliers in every matrix, the contact frequencies between bins separated up to 8Mb were drawn separately from the ones which are above 8Mb apart and were plotted separately. For genome wide, the mean interaction frequencies of all the chromosomes were plotted at each distance and the scatter plot was smoothened using above method.

22. **Dissection of interactions into various distance ranges**

First, the 10kb resolution SimpleNorm contact matrices of the three cell types were generated using HOMER. The coordinates of the cLADs and ciLADs were rounded off to the nearest 10kb bin. Then the interactions between cLAD-cLAD and ciLAD-ciLAD were separated. For cLAD, the range of interactions were divided into 6 ranges depending on the distance between them. They are <100kb, 100kb-500kb, 500kb- 2Mb, 2Mb-20Mb, 20-40Mb and >40Mb. The sum of all intra-range interactions was calculated for each range separately for each cell type and were plotted as bar charts. The mean of all the chromosomes was calculated for each range to represent the genome wide average bar chart for all the three cell types. Similarly, for ciLAD-ciLAD interactions, the above analysis of different ranges was performed and plotted as bar chart.

3. Results

Objective 1

Development of robust method to measure and characterize chromatin condensation state genome-wide

Accessibility is the ability of the chromatin to get digested. Generally, heterochromatin is highly condensed and is less accessible to digestion. On the other hand, euchromatin is in relatively decondensed form and is more accessible to digestion. Chromatin accessibility data is one of the keys to understand the regulatory mechanisms of gene expression. There are several methods to assess chromatin accessibility like MNase-seq, DNase-seq, ATAC-seq etc., which give information about whether a chromatin region is accessible or not. These methods are specially designed to determine regulatory elements in the genome. Recently, a technique called RED-Seq, which is based on restriction enzyme digestion, has been developed. It is advantageous over other methods that it assesses the amount of accessibility at each region of the genome. Also, since it involves digestion with restriction enzyme which cuts DNA at specific sites, lesser no. of sequenced reads is required to analyze the accessibility. RED-seq experiment was done on in-situ chromatin and naked DNA.

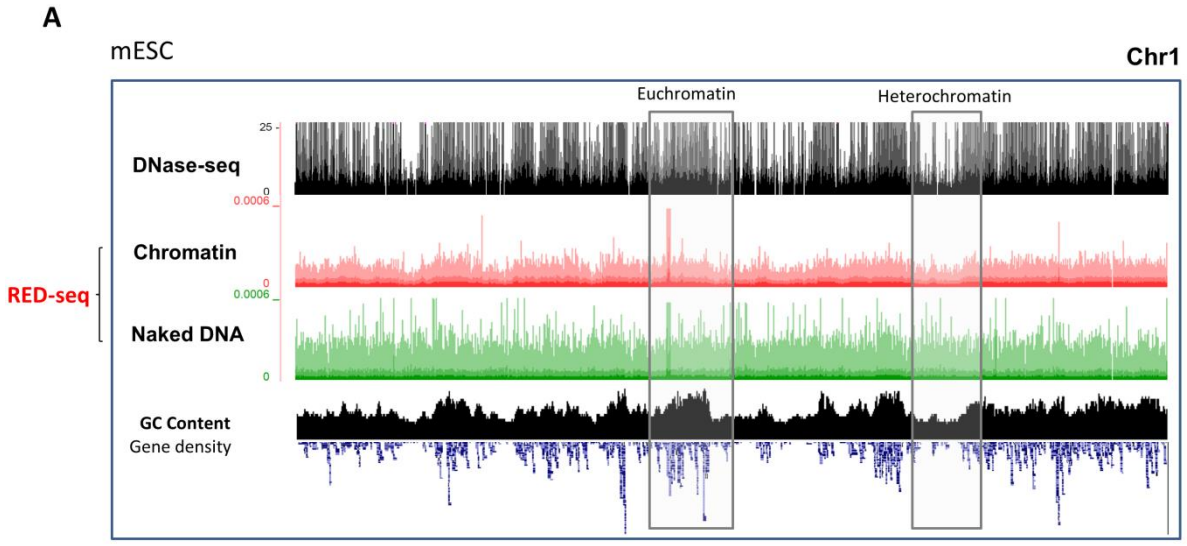
Most of the above methods use high-throughput sequencing to sequence the reads from chromatin. Since condensed chromatin is less accessible to digestion with restriction enzymes, we hypothesized that the no. of sequenced reads mapping to the condensed chromatin should be less than that of decondensed euchromatin. To test if this is true, we divided the Chr1 into equal sized 10kb bins and derived mapped 1D reads from in situ chromatin and naked DNA of ESC from RED-seq experiment and overlaid them along with DNase-seq data of ESC (Fig.1A).

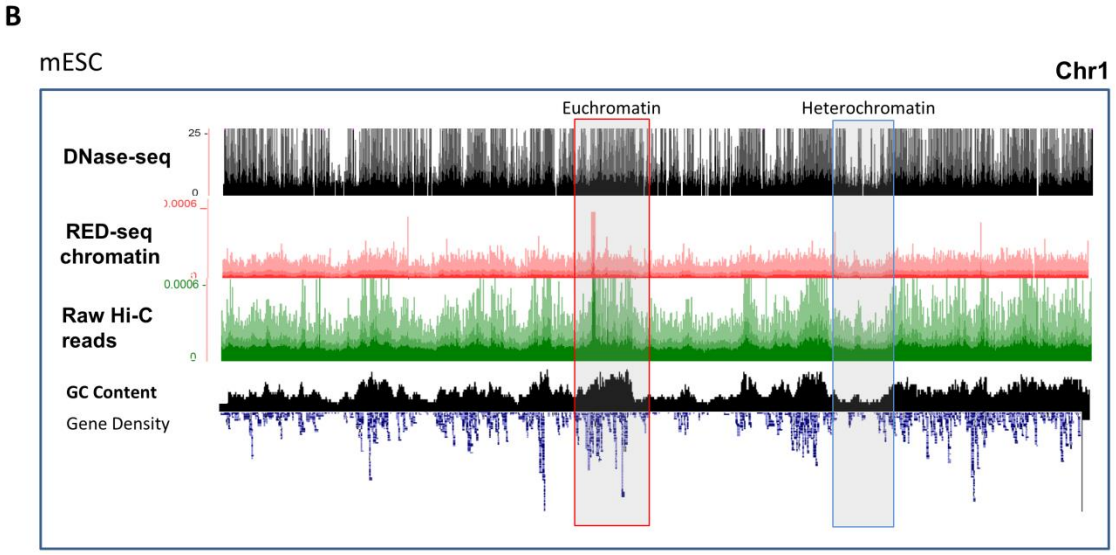
We could clearly observe that gene rich regions with denser reads in DNase-seq show more reads in chromatin than in that of gene poor regions. But this is not the case with Naked DNA where there is no bias in the reads between gene rich and gene poor regions. This gives an initial representation that 1D read count could be used to derive condensed heterochromatin regions and decondensed Euchromatin regions.

Then incidentally we hypothesized that, since Hi-C uses restriction enzymes to digest the chromatin, the Hi-C reads could be potentially repurposed to assess the degree of chromatin accessibility, similar to RED-seq. To test the hypothesis, we analyzed the in situ Hi-C data of mouse ESC from Bing Ren's Lab. From the reference genome mapped BAM/SAM files, we derived number of reads per 10kb regions from Chr1 and plotted them. The reads from in situ chromatin and raw Hi-C reads were plotted along with DNase-seq data of mouse ESC and GC content

As we see in Fig.1B, we can clearly observe that the gene rich, higher GC content regions, which represent decondensed chromatin, correspond to higher reads in Hi-C also, and the gene poor, lower GC content condensed regions have lower no. of reads.

Then, we wanted to see if this observation is cell type specific or a general cellular phenomenon. We analyzed in situ Hi-C data of mouse Neural Stem cells (NSC) and Astrocytes (AST) from Suzanne Hudjur's lab (Sofueva et al., 2013) and generated 1-Dimensional profiles. From the Fig.1C, one could clearly observe both in NSC and AST that gene rich, GC rich genomic regions has higher number of reads than gene poor, GC poor regions. This gives a pictorial representation that Hi-C reads can be used to derive condensed and decondensed regions.


A
mESC
Chr1
Euchromatin
Heterochromatin
DNase-seq
Chromatin
RED-seq
Naked DNA
GC Content
Gene density


B
mESC
Chr1
Euchromatin
Heterochromatin
DNase-seq
RED-seq chromatin
Raw Hi-C reads
GC Content
Gene Density

Contd..

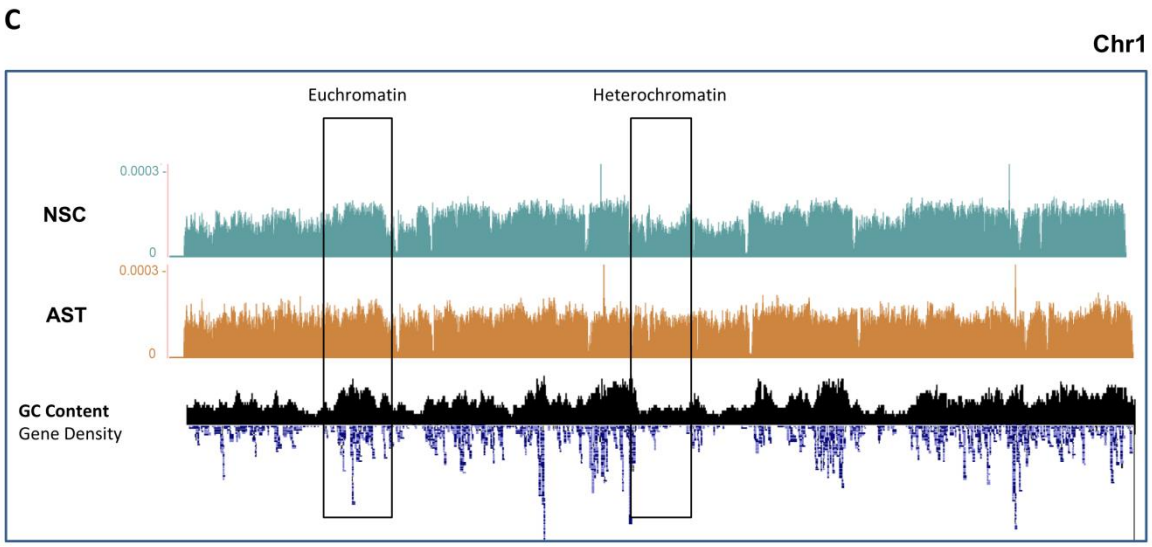


Fig 1: Biased reads in sequenced reads between Euchromatin and Heterochromatin

A. 1-D reads of RED-seq chromatin and naked DNA from Chr1 overlaid on GC content and gene density along with DNase-seq data. Representative Heterochromatin and Euchromatin regions are highlighted in boxes. Note that reads in euchromatin are more than that of heterochromatin. **B**. 1-D reads from Raw Hi-C reads shows higher no. of reads in GC rich gene dense Euchromatin than in GC poor gene poor Heterochromatin. RED-seq chromatin and DNase-seq data are overlaid for comparison. **C**. 1-D raw Hi-C reads from Neural Stem Cells (NSC) and Astrocytes (AST) from Chr1 shows similar bias in the reads between euchromatin and heterochromatin as seen in ESC.

Since now we have worked with raw data without any normalizations. One need to normalize for different biases in the Hi-C data and sequenced data in general. The sequenced reads should be normalized for GC content of the genomic region, because reads from GC rich regions are more represented than GC poor regions (Yaffe and Tanay., 2011). Conventionally the reads are normalized to the GC content of the genomic region, which may not be the true bias which comes from sequencing machine. Since naked-DNA from RED-seq experiment has only machine bias, we reasoned that correcting in situ chromatin and Hi-C reads with naked DNA would more appropriately correct for the biases.

Working Principle: The raw Hi-C reads should be corrected first for various biases. Since different bins have different number of restriction sites, 1D Hi-C reads should be normalized to restriction enzyme density. We used LOESS (locally weighted smoothing) regression to correct for the bias. LOESS is a non-parametrical regression model which uses least squares regression. First, the LOESS parameters were standardized in such a way that the reads in naked DNA corrected with LOESS form a straight line against GC content. This step corrects for the machine-specific bias. Then these LOESS parameters are used to correct the Hi-C reads for both GC content and RE density of the bins and then Z-scoring is applied to derive positive and negative values and we referred them as RZ scores (Fig.2A).

To assess the nature of the RZ scores, we overlaid them along with gene expression data derived from RNA-seq experiments and H3K27ac ChIP-seq data, both of which were publicly available (Fig.2B). We could clearly observe that the positive RZ values are correlated with higher gene expression and higher H3K27ac occupancy. On the other hand, negative RZ scores are correlated with lower gene expression and lower H3k27ac occupancy. Also, we observed that the positive RZ scores correlated with ciLADs boundaries, which are euchromatin, and, negative RZ scores correspond to heterochromatic cLADs. This has allowed one to use LOESS correction to obtain decondensed and condensed regions from Hi-C data.

Quantification of bias in reads:

The above results show the qualitative nature of Hi-C reads. Next, we wanted to quantify the above observations statistically. Supplementary Figure 1 shows the scatterplot of raw and corrected reads for each bias at a time i.e., RE density, GC percent. The final output is GC and RE corrected reads Normalized to total number of sequenced reads.

Next, we wanted to show the biased visibility of reads between euchromatin and heterochromatin statistically. Towards this end, we calculated reads from cLAD and ciLAD regions from both *in situ* chromatin and naked DNA, corrected them using LOESS for the biases as shown in the figure (Fig.3A). Then we plotted also raw read counts from *in situ* chromatin and naked DNA as boxplots, and we could observe that cLADs have significantly lesser reads than ciLADs in both. But when corrected reads were analyzed, cLADs have lesser number of reads than ciLADs in in situ chromatin but this is not seen in the naked DNA.

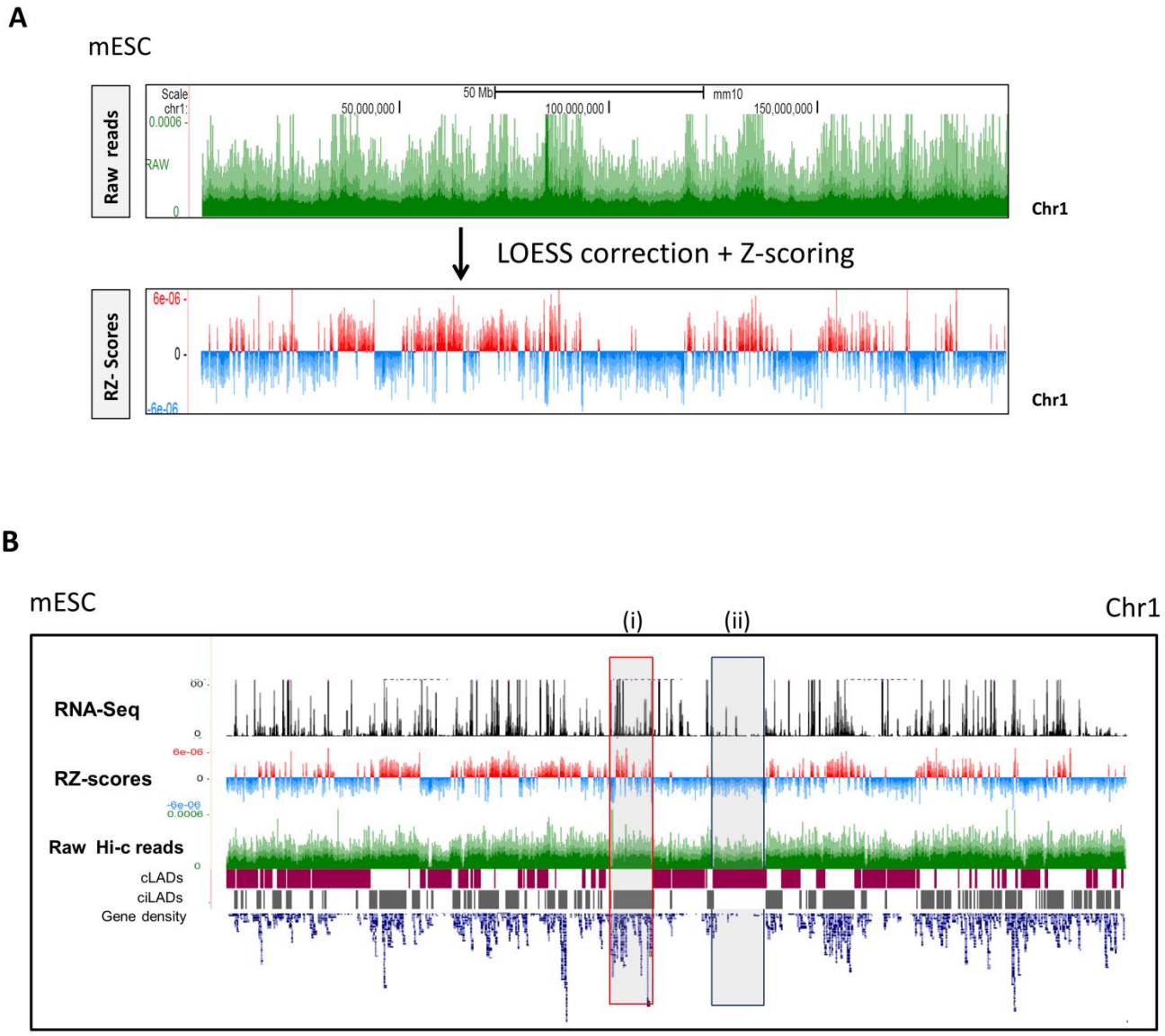


Fig 2: Normalization of the 1-D Hi-C reads

A. The raw 1-D Hi-C reads are Normalized for GC content and RE site density using LOESS regression whose parameters were optimized for correction of machine bias. This is followed by Z-scoring to obtain positive and negative values called RZ scores. **B**. Raw Hi-C reads overlaid along with RZ scores and RNA-seq data to decipher the nature of RZ scores. Positive RZ scores are correlated with higher gene expression, gene density and ciLADs. Negative RZ-scores are correlated with lower gene expression, gene density and cLADs.

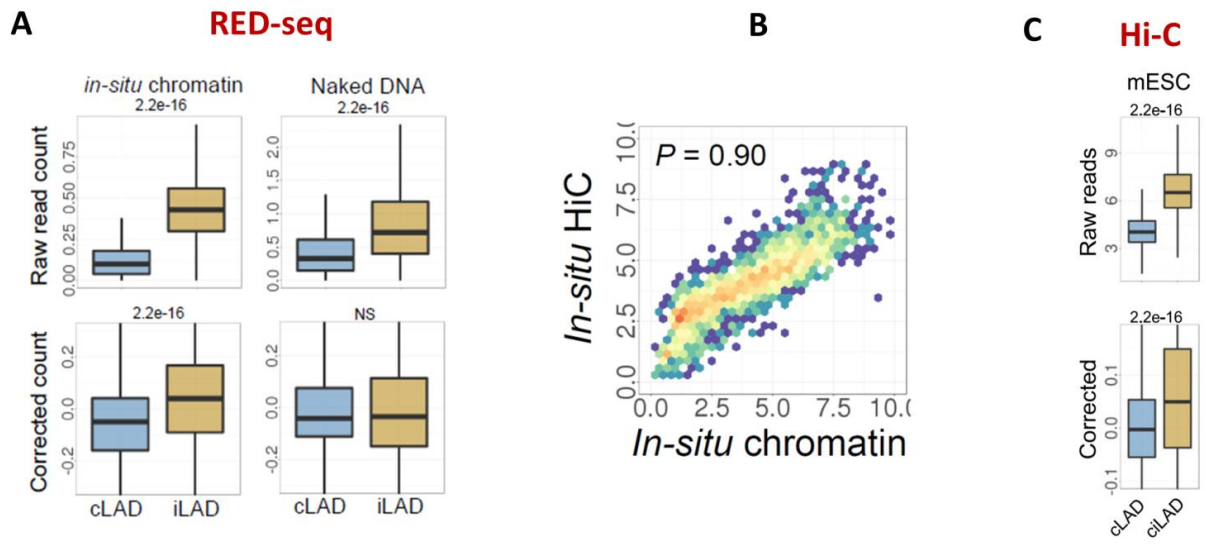


Fig 3: Quantification of the bias in reads.

A. Upper panel shows raw read counts in cLAD and ciLAD of both in in-situ chromatin and naked DNA. Lower panel shows corrected read counts in cLAD and ciLAD of both in-situ chromatin and naked DNA. P-values were shown above the boxplots all of which are significant except in corrected reads of naked DNA. Corrected naked DNA shows no bias in the read counts between cLADs and ciLADs whereas the raw reads exhibit the bias. **B**. Scatterplot of in-situ Hi-C and in-situ chromatin which shows high degree of correlation (spearman correlation coefficient p=0.9). **C**. Quantification of 1-D Hi-C reads in cLADs and ciLADs before and after correction shows bias in the reads which proves that Hi-C reads could be repurposed to derive condensed and decondensed regions

This shows the presence of accessibility bias between cLADs and ciLADs. In supplementary figure 2, a region of a chromosome has been taken to illustrate these biases. The picture shows presence of negative RZ scores which is condensed is still negative after correction in in situ chromatin but the negative region becomes random in the naked DNA and hence no bias.

To show that similar bias could be observed in Hi-C datasets, we calculated correlation between the 1D read counts of in situ chromatin and in situ Hi-C. The correlation between corrected reads has been drawn as a scatterplot (Fig.3B), which shows that they are highly correlated. Hence, we concluded that statistically similar bias exists in Hi-C reads and 1D Hi-C read counts could be potentially used to derive condensed and decondensed domains in the genome (Fig.3C).

Derivation of decondensed and condensed domains

Next, we wanted to derive condensed and decondensed domains using the obtained RZ scores. Continuous stretches of positive RZ scores were defined as decondensed domains and stretches of negative RZ scores defined as condensed regions with a false discovery rate <0.05 (Fig.4A). To verify the method, we matched the condensed regions with known boundaries of cLAD regions and could observe that nearly 70% of the condensed regions matches with boundaries of cLADs (Fig.4B). This is for the first time one has derived condensed and decondensed regions based on 1-Dimensional Hi-C reads.

Characterization and validation of the method:

We next wanted to characterize the condensed and decondensed domains derived using our method and see how various histone marks are associated with these regions. To this end, we have taken genome wide average boundary (+/- 1mb) of decondensed and condensed regions and derived the genome wide average histone marks associated with them (Fig.5A). The left side of the border is decondensed and right side is condensed. As we can see, the active histone marks such as H3K4me1, H3K27ac, H3K36me3, H3K4me3 have higher enrichment in decondensed region and low enrichment in condensed region. Whereas the repressive chromatin marks such as H3K9me3 are enriched in condensed region and vice versa. This shows that the regions derived from our method are in concordant with the known histone marks.

To further validate our method, we wondered if we can recapitulate the condensation in inactive X-chromosome. As we know, in human female cells, one of the two X-chromosome is inactive and condensed. Allele specific DNase Hi-C was performed by Deng et al. We obtained, analyzed and corrected the Hi-C data using our method. Here is a scatterplot (Fig.5B) showing reads of active X-chromosome in X-axis and of inactive X-chromosome in Y- axis. We can clearly see that condensed inactive X-chromosome has lesser number of reads than that of active X-chromosome. On the right side, we can see a small representative region of a chromosome where active X-chromosome has more number of reads than inactive

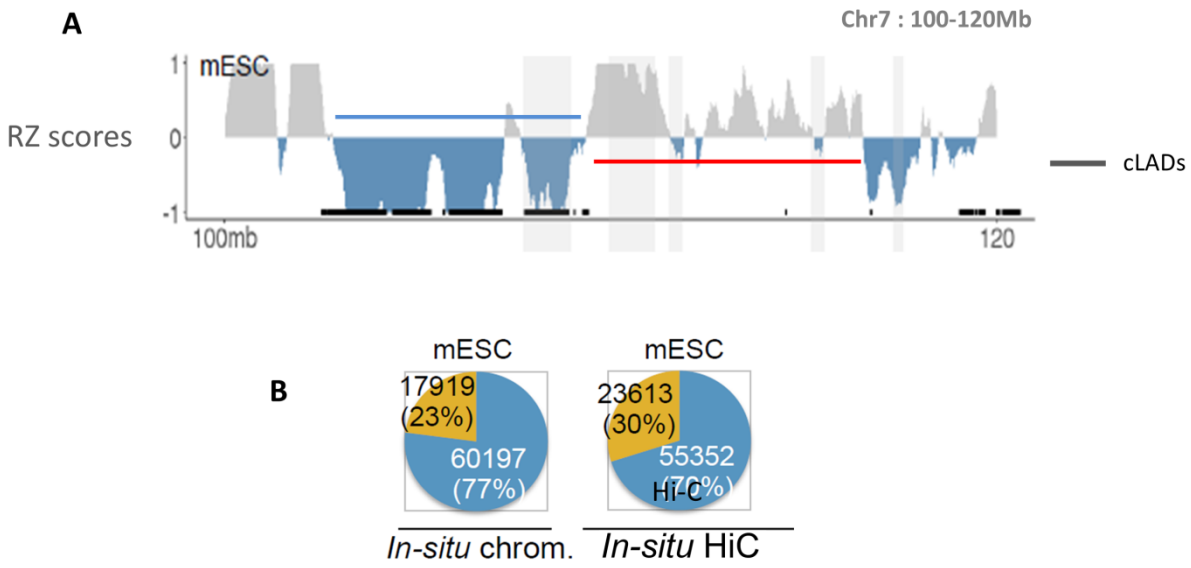


Fig 4: Derivation of condensed and decondensed domains

A. Condensed and decondensed domains were derived using continuous stretches of negative and positive values, respectively. A region of the chromosome (Chr7: 100-120Mb) showing derived condensed domains (blue line) and decondensed domains (red line) from the RZ scores. Black lines at the bottom are cLAD boundaries. Condensed domains are correlated with cLADs and decondensed domains are correlated with inter-LAD regions. **B**. Pie chart showing how much percentage and bins of condensed domain are matching with the boundaries of cLAD in both in-situ chromatin and in-situ Hi-C. Most of the decondensed domains derived are matching well with the known boundaries of cLADs.

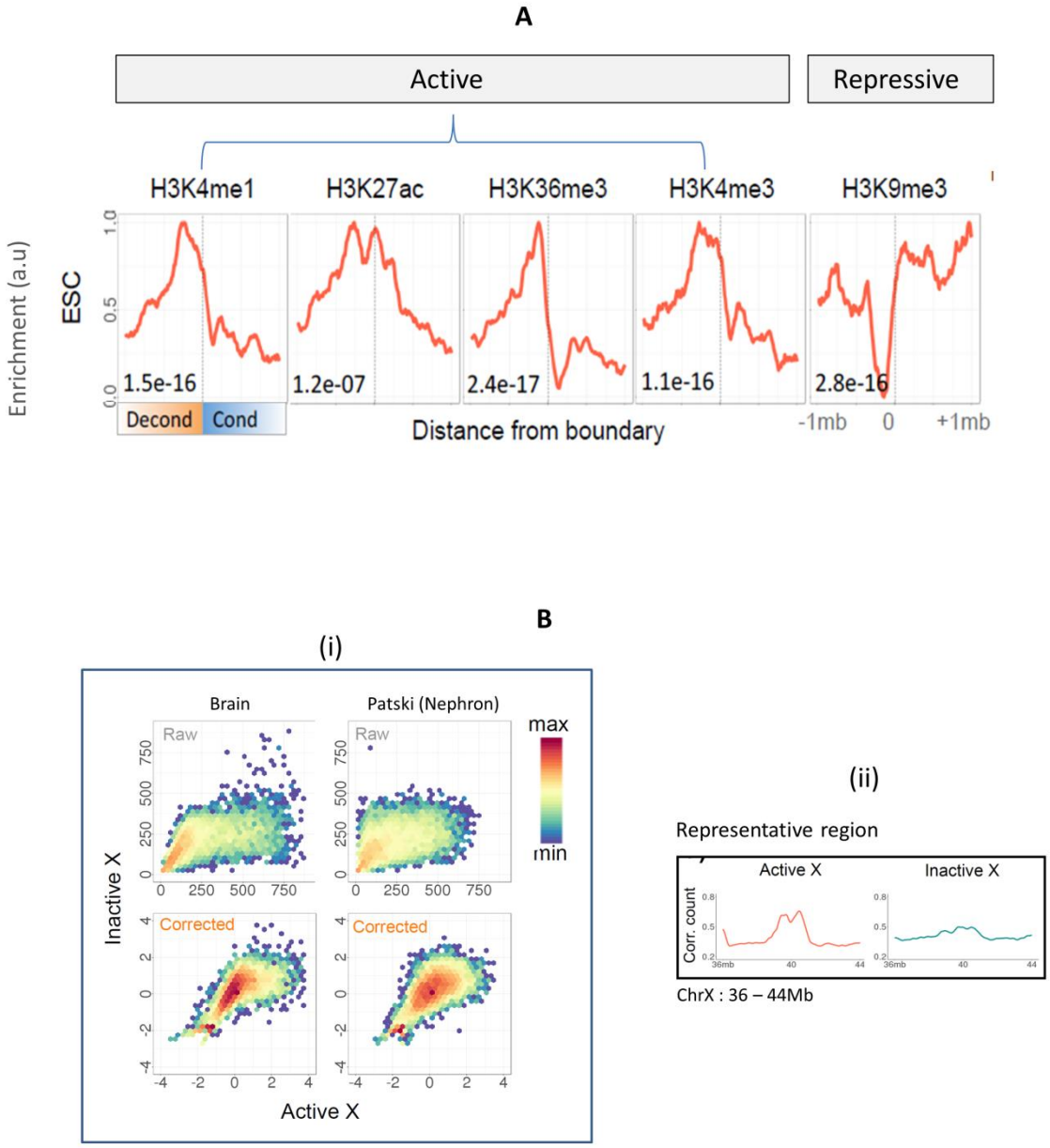


Fig 5: Characterization and validation of the method

A. Enrichment of active and repressive histone marks around (+/-1Mb) regions of average boundary between decondensed and condensed domains. Active marks shows enrichment at decondensed side and repressive marks at the condensed side of the boundary. P-values were calculated using two-tailed Mann Whitney U tests by comparing mean enrichment values in the bins of condensed and decondensed domains. **B.(i)** Scatterplots of raw and corrected DNase-HiC read counts of active vs. inactive X-chromosomes in Brain and Patski cells **(ii)** Illustrative examples of corrected read counts and contact maps of chrX: 36-44 Mb region in active and inactive X-chromosome.

X-chromosome. This shows that our method could potentially recapitulate the condensation of inactive X-chromosome.

To further scrutinize our method, we analyzed Hi-C data available on Drosophila polytene chromosome. Drosophila polytene chromosome is a classic example of presence of condensed and decondensed chromatin. The dark condensed regions are called bands and the lighter decondensed regions are called inter-bands (ibands). We wanted to see whether we can recapitulate the difference in the read counts between bands and ibands using our method. We downloaded and normalized both polytene and diploid drosophila chromosome Hi-C data as previously done, and we could see that band regions show significantly lower reads than ibands both in polytene and diploid chromosomes (Fig.6A and B).

Then we wondered if we could recapitulate the dynamics of LADs upon lamin knockout. One such experiment was done where lamin B was knocked out of ESC and Hi-C was performed (Deng et al., 2015). We analyzed and corrected the data using our method and could see that no. of reads coming from KO cLADS is higher than that of wild type, whereas the rest of the regions didn't get affected (Fig.6C). On the right side, we can see a region of the chromosome where cLAD regions are represented as dashed lines, the cLAD regions of lamin KO have enrichment of reads over WT (Fig.6D). This proves unequivocally that Hi-C reads can be efficiently repurposed to derive condensed and decondensed regions.

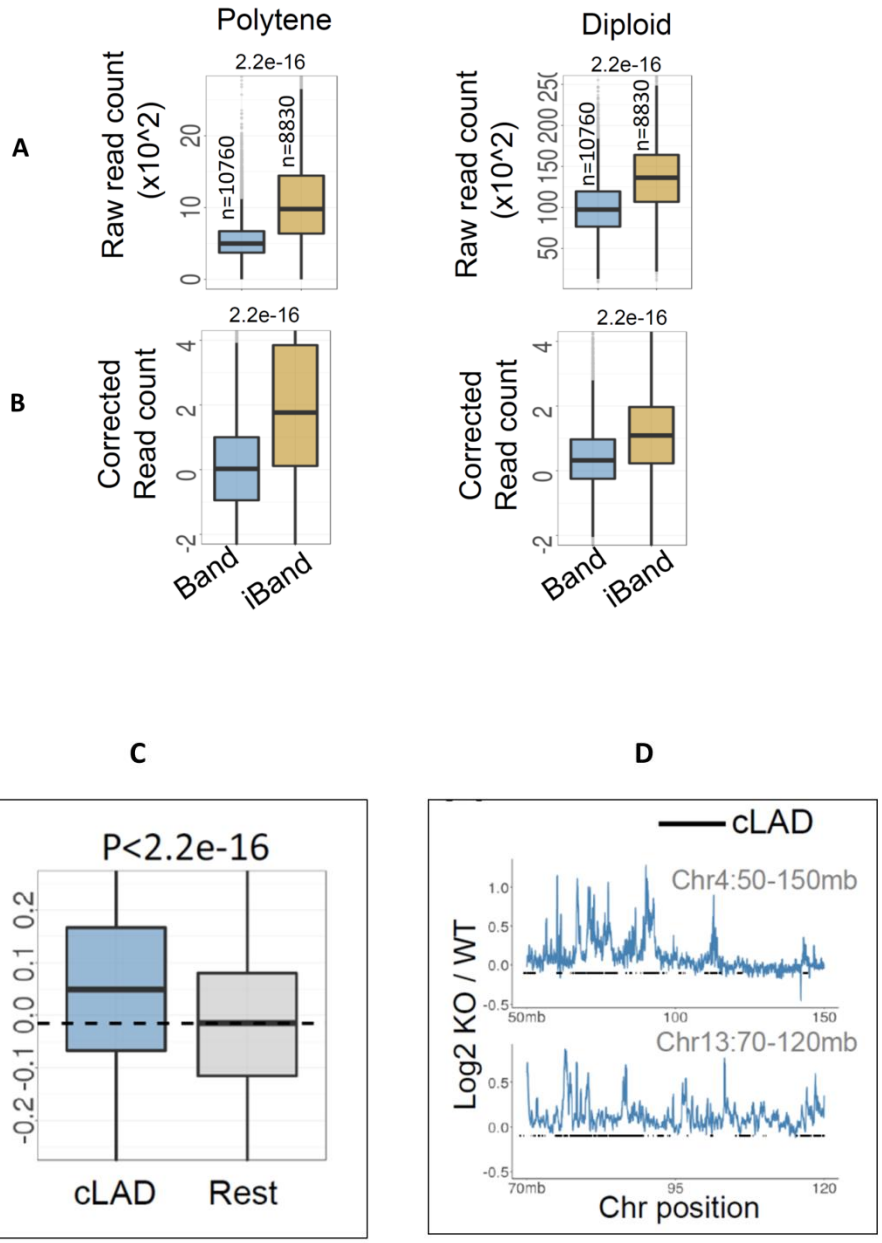


Fig 6: Further scrutinization of the method

Distribution of raw(**A**) and corrected(**B**) read counts in band and inter-band regions of polytene chromosome and the corresponding regions in diploid chromosome. Bias in the 1D read counts between bands and i-bands could be clearly seen in raw reads and corrected reads in both polytene and diploid chromosome **C**. Boxplots representing change in Hi-C read counts in the LAD domains after Lamin knock out in mESC cells. Read counts from cLADs increased upon lamin KO but change in the rest of the genome **D**. Example representing change in Hi-C read counts and contact matrices in WT and Lamin KO cells. P-values were calculated using two-tailed Mann-Whitney U tests.

Derivation of Chromosome Condensation Maps from Chromatin Accessibility Datasets

Next, we wanted to derive Chromosome Condensation maps using the chromatin accessibility data. Towards this end, we derived RES maps (see methods). By vertically mirroring the RES map, we could derive vertically compressed Chromosome Condensation map, where the compressed regions correspond to heterochromatin/LADs and decompressed regions correspond to euchromatin/ciLADs (Fig.7iii). Then we derived horizontally compressed chromosome condensation maps, where the length of the chromatin segment is directly proportional to the accessibility (Fig.7i). That is heterochromatin regions are horizontally compressed than euchromatin and the level of compression is depicted with color code where red represents highly compressed and yellow represents decompressed chromatin. This is for the first time ever one has derived both horizontally and vertically compressed Chromosome condensation maps, which actually represent the true in vivo chromosomes.

Then we wanted to see the difference between the chromosome condensation maps of other cell types as well. To this end, we analyzed 1-Dimensional Hi-C reads of Chr6 of Neural Stem Cells (NSC), Astrocytes (AST), fetal Liver (fLiv) and adult Liver (ALiv) and derived both vertically and horizontally compressed Chromosome condensation maps (Fig.8). As we can see, the Chr6 condensation maps look differently in different cell types. In vertically compressed condensation maps (Fig.8A), the difference is clearly seen in terms of compression and decompression of various regions of the same chromosome. The difference is even more pronounced in horizontally compressed condensation maps. In ESC, the Chr6 is relatively shorter, but in NSC they become longer and in AST they become further longer (Fig.8B). Similar kind of dynamics can be seen between the condensation maps of Chr6 in ESC, fLiver and ALiv (Fig.8C).

Similarly, the horizontally compressed chromosome condensation maps of all the chromosomes of the above cell types were compiled on a similar scale and we could see the dynamics of lengths of various chromosomes in various cell types during differentiation (Fig.9 A and B).

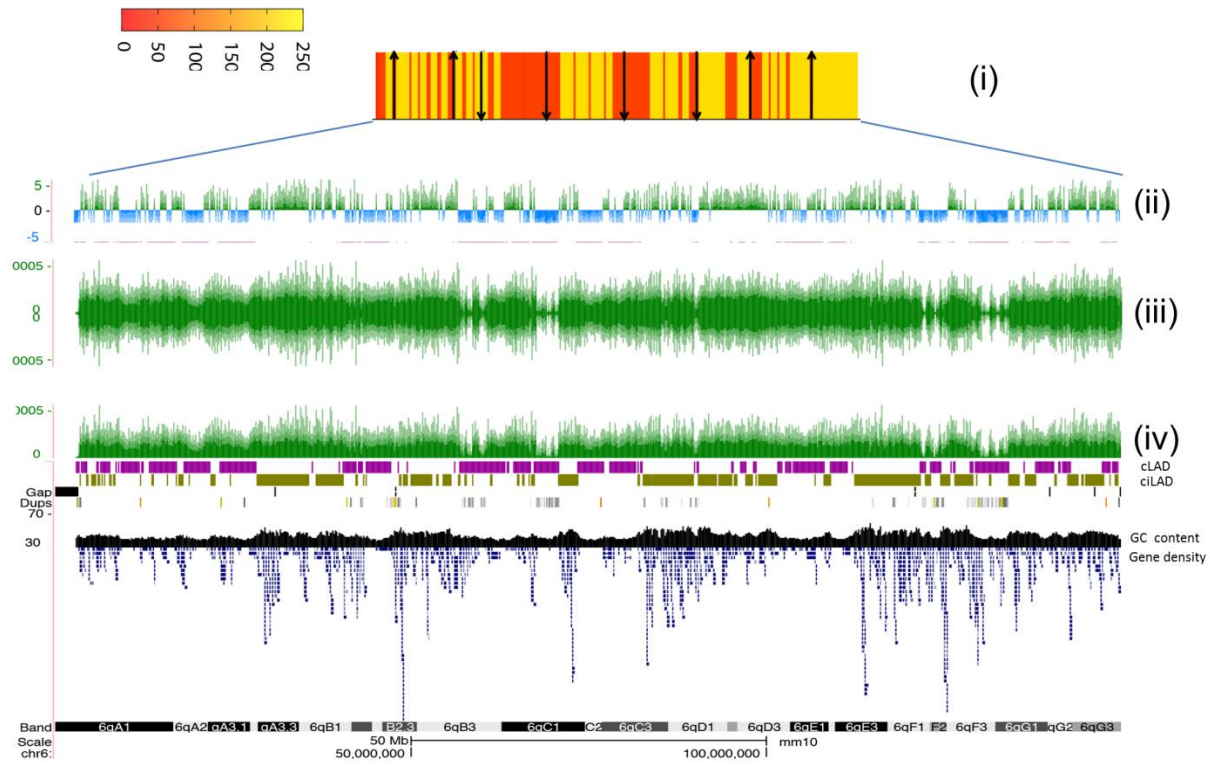


Fig 7: Chromosome Condensation maps

(i) Horizontally compressed chromosome condensation map of Chr6 in which the horizontal length of a bin is a function of no. of reads present in the bin. Red represents lower no. of reads and thus heterochromatin and Yellow represents higher no. of reads and thus Euchromatin. Up arrow represents ciLADs and Down arrow represents cLADs. **(ii)** Corresponding RZ scores of whole Chr6 derived from Hi-C data using our method. **(iii)** Vertically compressed chromosome condensation map where compressed regions represent heterochromatin and decompressed regions represent euchromatin. **(iv)** RES map which is the original map from which the vertically compressed map is derived, overlaid on cLAD and ciLAD positions.

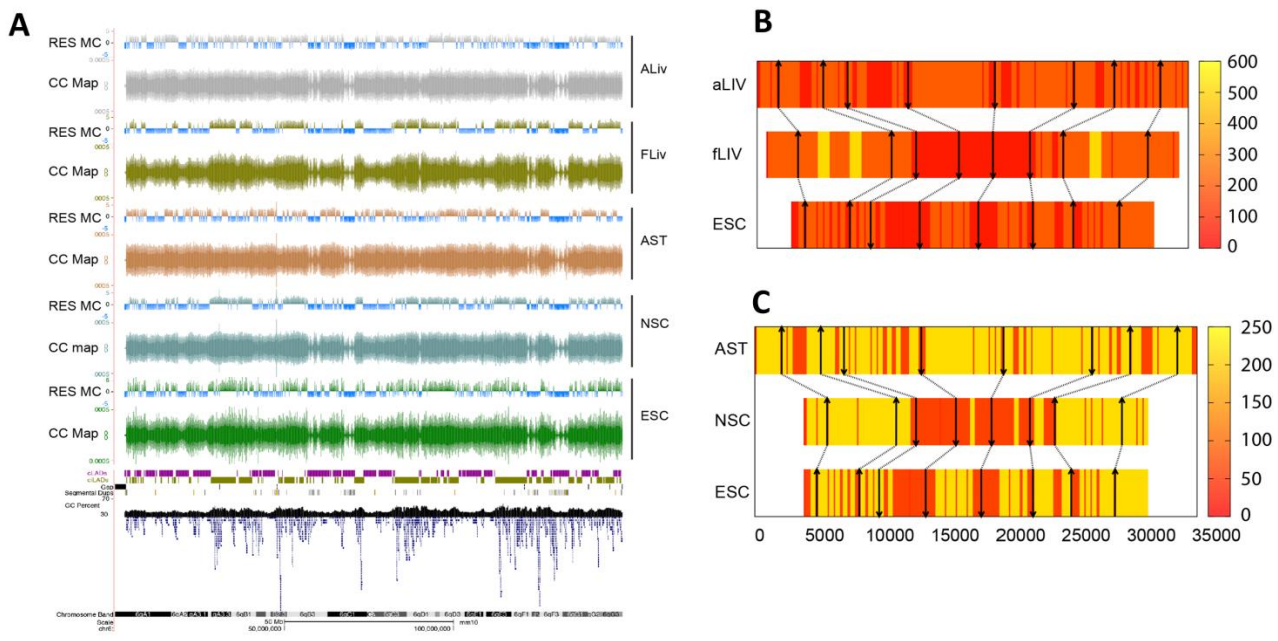


Fig 8: Dynamics of Chromosome Condensation maps during differentiation

A. Developmental dynamics of vertically compressed chromosome condensation maps during differentiation (from bottom to top) of ESC to NSC and AST. Similar dynamics could be seen in ESC to fetal Liver cells and adult liver cells. **B**. Dynamics of horizontally compressed chromosome condensation maps during differentiation of ESC to fetal Liver cells and then to adult liver cells. **C**. Dynamics of horizontally compressed chromosome condensation maps during differentiation of ESC to NSC and then to AST. Scale bar represents the no. of reads (a.u) from each bin(100kb). Up arrow represents ciLADs and Down arrow represents cLADs.

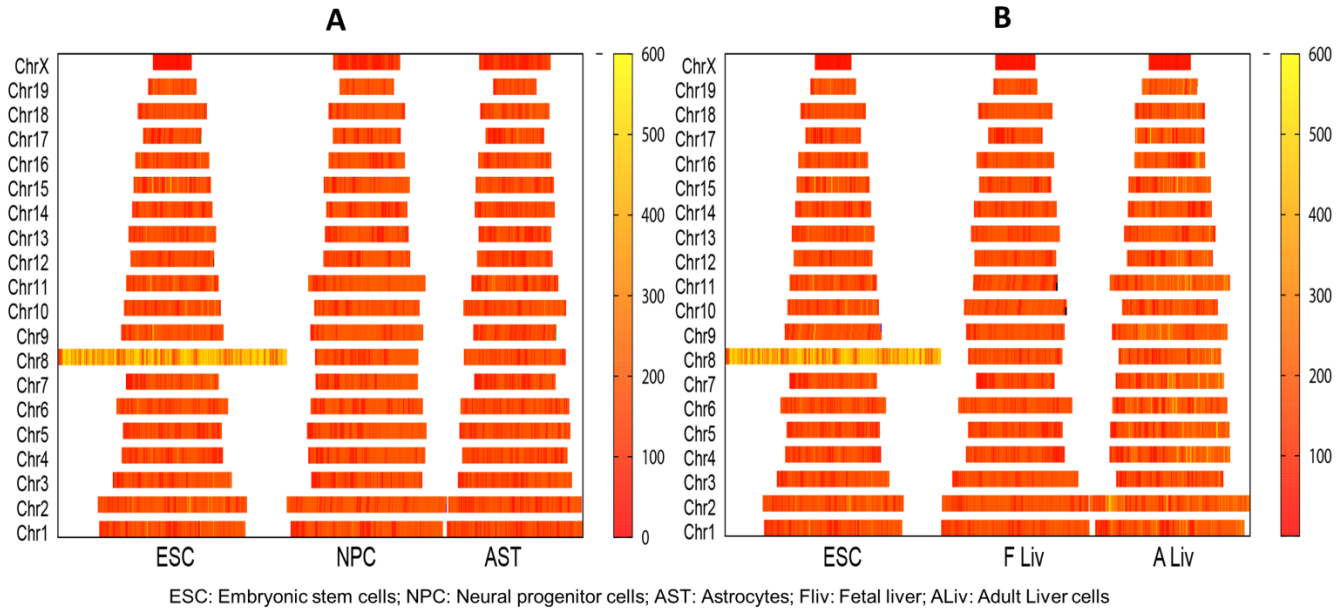
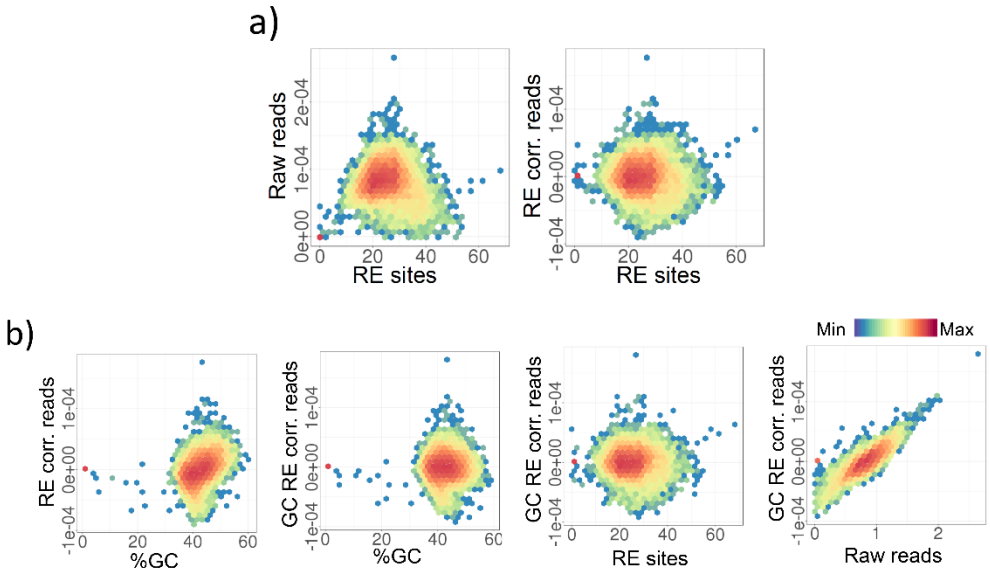


Fig 9: Ensemble of horizontally compressed maps of all chromosomes in different cell types

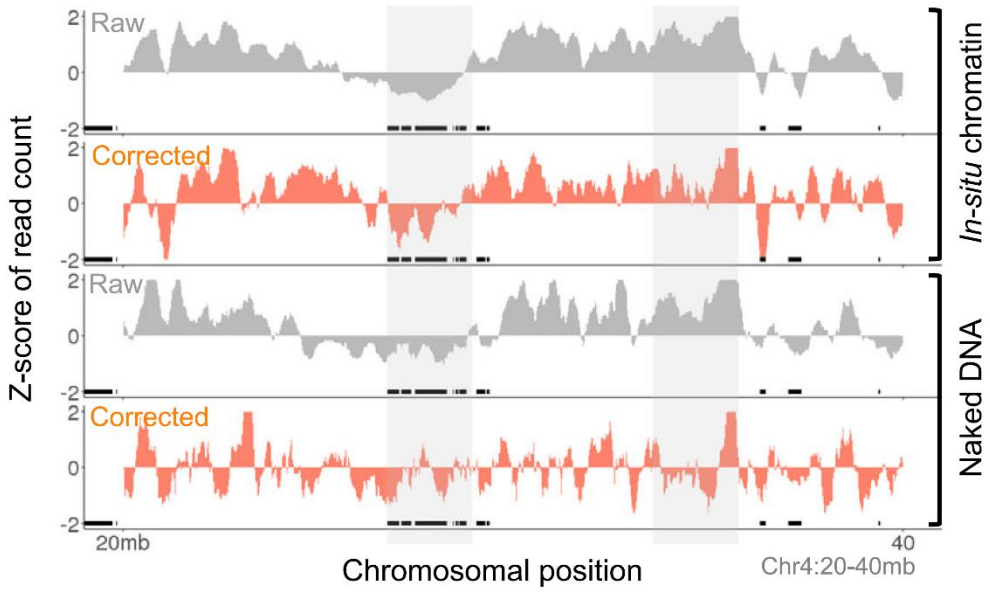
A. Dynamics of horizontally compressed chromosome condensation maps of all chromosomes during differentiation of ESC to NPC and then to AST. Note that the different lengths of the chromosomes is due to different level of condensation. **B**. Dynamics of horizontally compressed chromosome condensation maps of all chromosomes during differentiation of ESC to Fetal liver cells and then to adult liver cells. Scale bar represents the no. of reads (a.u) from each bin(100kb).

Supplementary Figures



Supplementary Figure 1: Loess correction of raw 1D Hi-C reads

(a) Loess correction for the negative scaling of raw read counts against the restriction site (RE) density in 10Kb genomic bins. Left panel represents data before loess correction and right panel after loess correction of read counts against RE density. **(b)** Loess correction for the positive scaling of RE-corrected read counts against the GC content of 10Kb genomic bins. First panel shows scatter plot of GC content vs. Recorrected read counts. Second panel shows scatter plot of GC content vs. GC- and recorrected read counts. Third panel represents RE-density vs. GC and RE corrected read counts. Fourth panel shows scaling of GC and RE corrected read counts against the raw read counts.



Supplementary Figure 2: Illustrative example of raw and corrected read counts.

An example showing raw and corrected read counts of in situ digested chromatin and in solution digested naked DNA along chr4: 20-40 Mb region. The picture shows presence of negative RZ scores which is condensed is still negative after correction in in situ chromatin but the negative region becomes random in the naked DNA and hence no bias.

Objective 2

Developmental dynamic role of chromatin condensation during neuronal differentiation

Having established the method to derive condensed and decondensed regions from Hi-C data, we next wanted to see the dynamics of condensed and decondensed regions during differentiation of Embryonic Stem Cells (ESC) to Cortical Neurons (CN) via Neural Progenitor Cells (NPC) (Fig.10A).

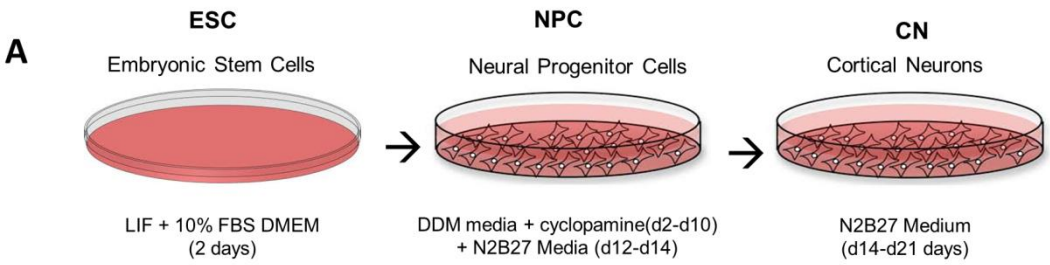
We chose this model system for two broad reasons. Firstly, this in vitro differentiation system is well established and very robust. There are lots of datasets available from public repositories pertaining to this model system. Secondly, this differentiation model captures a ground Embryonic stem cell state, an intermediate NPC and then finally a terminally differentiated Neuronal state i.e., ground to terminal differentiation in just three cell types. Thus, many aspects of the differentiation could be studied through this model system.

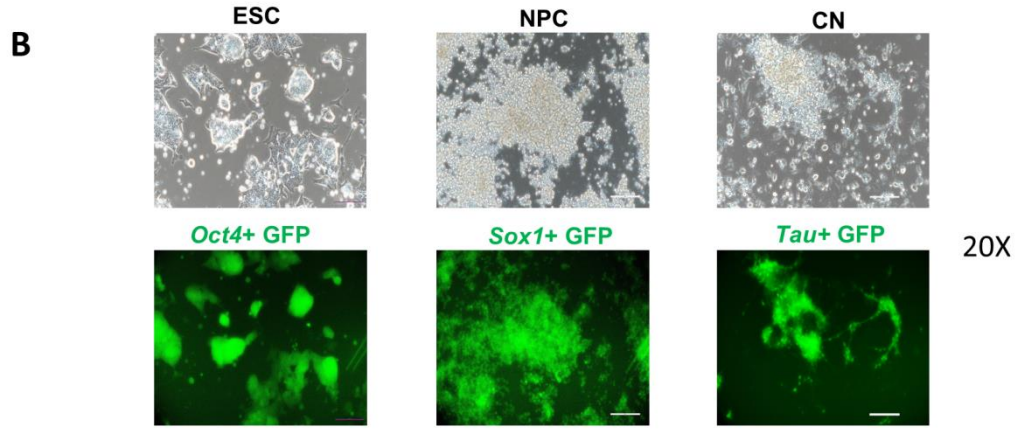
46C cell line of mouse ESC was cultured and differentiated into Neural Progenitors and then to Cortical Neurons (Fig.10B). To confirm the cellular identity during differentiation, cell type specific GFP tagged proteins were expressed in the three cell types. ESC specific *Oct4,* NPC specific *Sox1* and CN specific *Tau* were seen expressed along with GFP tags in the respective cell types (Fig.10B). To further characterize the cell types, total RNA was extracted from each cell type, converted to cDNA using RT-PCR and then semi-quantitative PCR and qPCR was performed to check the presence of cell type specific markers. In semi-quantitative PCR expression of *Oct4, Sox2, Nanog* in ESC was seen. *Nestin, Mapt, Pax6* and *Sox1* expression in NPC, and, expression of *Tbr1* and *Syp* was seen in CN (Fig.10C). In quantitative PCR *Nanog, Sox2, Gata4* and *Gata5* are specifically expressed in ESC. In NPC *Nestin, Neurogenin* and *Trkb* are specifically expressed. *Tuj1* and *Tau* are specifically expressed in CN (Fig.10D).

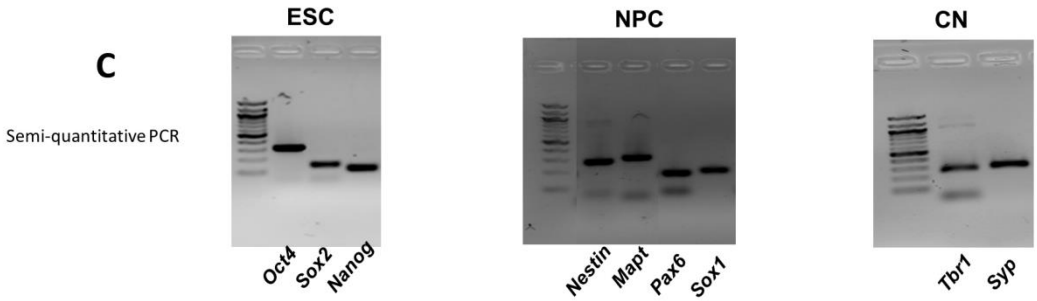
While we were working on this, Cavalli group (Bonev et al., 2017) from France published a work with ultra-deep sequenced Hi-C data along with gene expression and many ChIP-seq datasets. Though they studied different aspects of TADs and other important factors, none of our objectives were overlapping with their work. Hence, we took advantage of the study, downloaded and reanalyzed the *in situ* Hi-C and ChIP-seq datasets to address our objectives.

To see whether the accessibility bias exists in these cell types, first we derived 1D tracks of Hi-C data at 10kb resolution for all the three cell type viz., ESC, NPC and CN. Then we normalized the raw reads using our LOESS method and plotted the no. of reads in cLADs and ciLADs in all the three cell types (Fig.11A). We could clearly see that cLADs have lesser no. of reads than in ciLADs in the three cell types in both raw and corrected reads. This shows us that the accessibility bias exists in these datasets and thus the data could be potentially repurposed to derive condensed and decondensed domains.

We also wanted to see whether the observed bias in the visibility is specific to *in situ* Hi-C. To this end we analyzed data from another study by J Fraser (Nagano et al., 2015), where they have performed *in solution* Hi-C in ESC, NPC and CN. We analyzed the data, corrected and no. of reads form cLADs and ciLADs were plotted (Fig.11B). Interestingly, we could also observe the bias in in solution Hi-C, which gives us hint that the bias which we have been observing is due to differential digestibility by restriction enzymes but not because of difference in the ligation events. Then we derived condensed and decondensed domains as done


A
ESC
Embryonic Stem Cells
NPC
Neural Progenitor Cells
CN
Cortical Neurons
LIF + 10% FBS DMEM
(2 days)
DDM media + cyclopamine(d2-d10)
+ N2B27 Media (d12-d14)
N2B27 Medium
(d14-d21 days)


B
ESC
NPC
CN
Oct4+ GFP
Sox1+ GFP
Tau+ GFP
20X


C
Semi-quantitative PCR
ESC
NPC
CN
Oct4
Sox2
Nanog
Nestin
Mapt
Pax6
Sox1
Tbr1
Syp

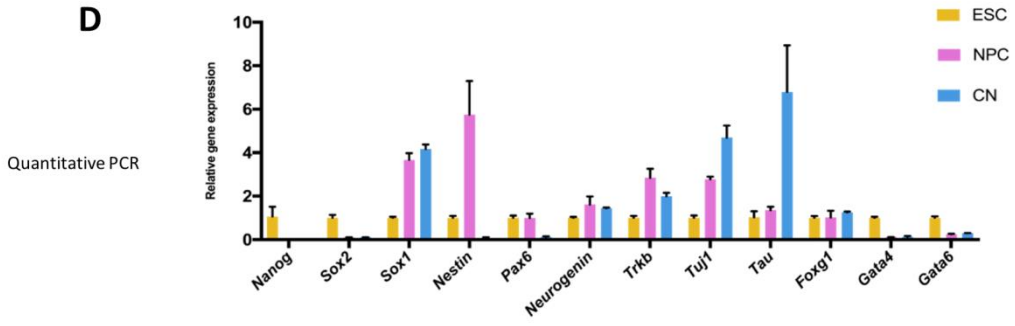

D
Quantitative PCR
Relative gene expression
ESC
NPC
CN
Nanog
Sox2
Sox1
Nestin
Pax6
Neurogenin
Trkb
Tuj1
Tau
Foxg1
Gata4
Gata6

Fig 10. Neuronal Differentiation of ESC and validation through semi-qPCR and qPCR.

A. Cartoon showing the dynamic system of differentiation of ESC to NPC and further to CN and the corresponding growth conditions. **B**. Upper row: Light microscope images of ESC, NPC and CN at 20X zoom. Clear colonies can be observed in the ESC whereas neurons forming dendrites could be seen in CN ; Lower row: Flourescent microscope images of ESC, NPC and CN at 20X zoom. Cells were tagged with stage-specific GFP markers which could be clearly observed in green colour and confirms cellular identity. **C**. Agarose gel pictures showing the expression of stage specific cellular marker genes in ESC, NPC and CN by semi-quantitative PCR. **D**. qPCR results show stage specific quantitative expression of marker genes in ESC, NPC and CN.

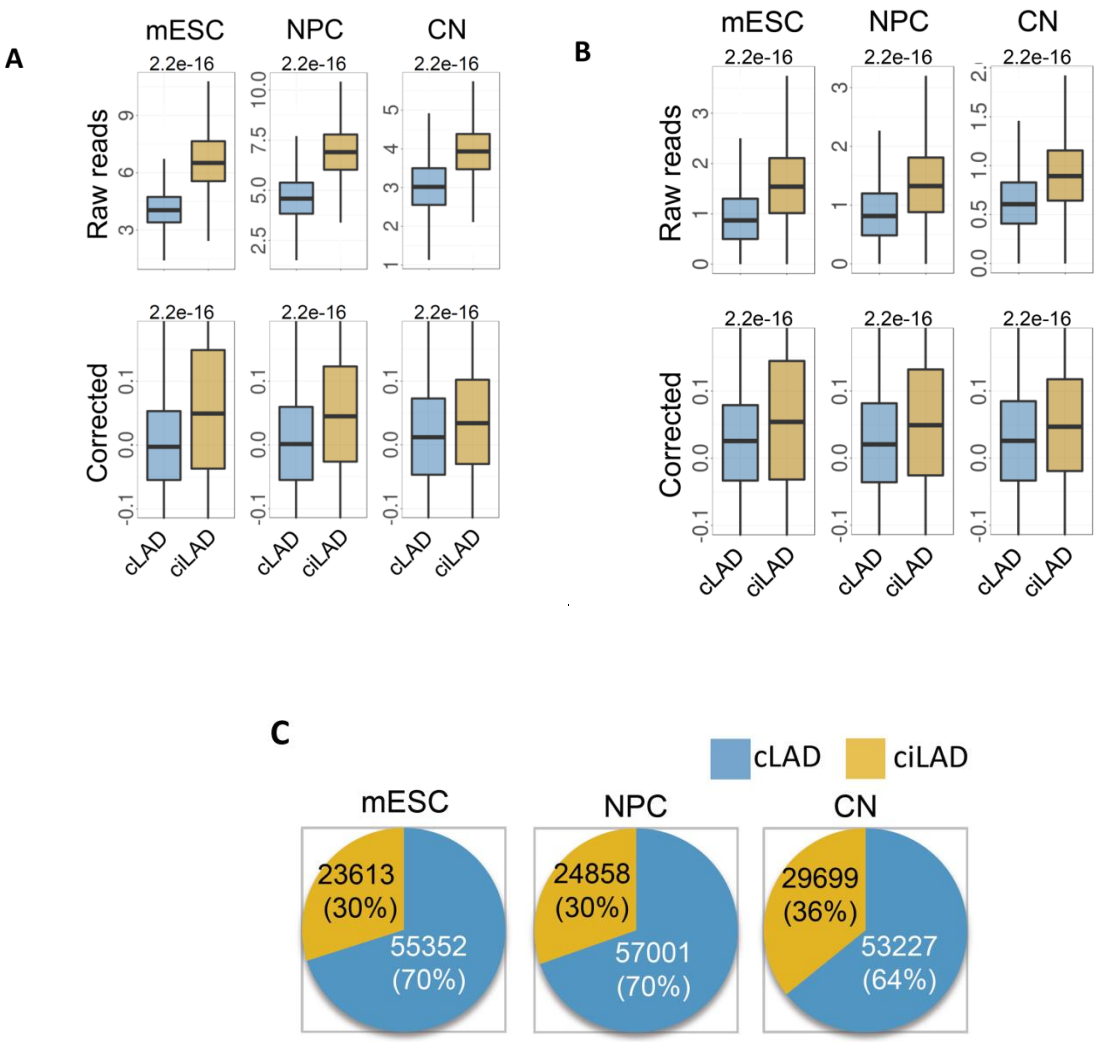


Fig 11. Bias in the visibility in Hi-C data of ESC differentiation system

A. Boxplots showing accessibility bias between cLADs and ciLADs in all the three cell types i.e., ESC, NPC and CN derived from in-situ Hi-C data (Bonev et al., 2017). The reads coming from cLADs are lesser than that of ciLADs in both raw and corrected. **B**. boxplots showing accessibility bias between cLADs and ciLADs in all the three cell types i.e., ESC, NPC and CN derived from in-solution Hi-C data (Fraser et al., 2015). Similar type of bias is seen between reads of cLADs and ciLADs as that of in situ Hi-C data. **C**. Pie charts showing the percentage and bins of derived condensed domains from loess corrected in-situ Hi-C data matching with cLAD boundaries. This shows that the decondensed domains derived using our method are in well agreement with the known boundaries of cLADs.

previously and wanted to see what percentage of condensed domains matched with known coordinates of cLADs. In fact, we could see that almost 70% of them are matching (Fig.11C).

Dynamics of Condensed and Decondensed Chromatin Domains during ESC to CN Differentiation

To study the dynamics of condensed and decondensed domains during neuronal differentiation, we plotted correlation scatterplots of the RZ scores between ESC and NPC, and between NPC and CN. ESC and NPC has a correlation of 0.90 and NPC and CN has a correlation of 0.97 (Fig.12A). Since most of the differences were seen significantly during ESC to NPC differentiation, we further restricted our downstream analysis on ESC to NPC differentiation.

First, we wanted to see how much condensed and decondensed regions are conserved between the two cell types. To this end, we calculated the no of regions conserved between ESC and NPC and plotted as relative percentage. We could see that around 65-70% of the regions which are condensed in ESC are also condensed in NPC, whereas around 80-85% of the regions which are decondensed in ESC are decondensed in NPC (Fig.12B). This is in well accordance with the previous studies that most of the regions are conserved between cell types in terms of condensed/decondensed state. Next, we wanted to see whether switch in the regions between the cell types is associated with corresponding change in gene expression. We calculated the gene expression changes in the constitutive and switching regions. We could clearly see that a switch from decondensed in ESC to condensed in NPC is accompanied by overall decrease in gene expression, whereas switch from condensed in ESC to decondensed in NPC is accompanied by drastic increase in the gene expression (Fig.12C).

Next, we wanted to see the status of histone marks associated with the regions during switching. On the left of Fig.13A, we show an example representation of effect of switching on well-known active mark H3K4me1. X-axis represents enrichment in ESC and Y-axis represents enrichment in NPC. When a chromatin region switches from decondensed to condensed from ESC to NPC, the plot show enrichment towards ESC, suggesting that condensation is accompanied by decrease in this active methylation mark. Similarly, ESC to NPC decondensation is accompanied by increase in the active mark. This is shown in the right panel for four active histone marks, whereas the repressive marks H3K9me3 increases during condensation and vice-versa (Fig.13A). Also examples of constitutive and switching regions were shown. Left panel shows constitutive region where all the histone marks remain unaffected including gene expression (Fig.13B). Right side, there is an example of switching region where the region with positive RZ score (decondensed) in ESC switches to a region with negative RZ score in NPC, which is accompanied by corresponding changes in histone marks and gene expression (Fig.13C).

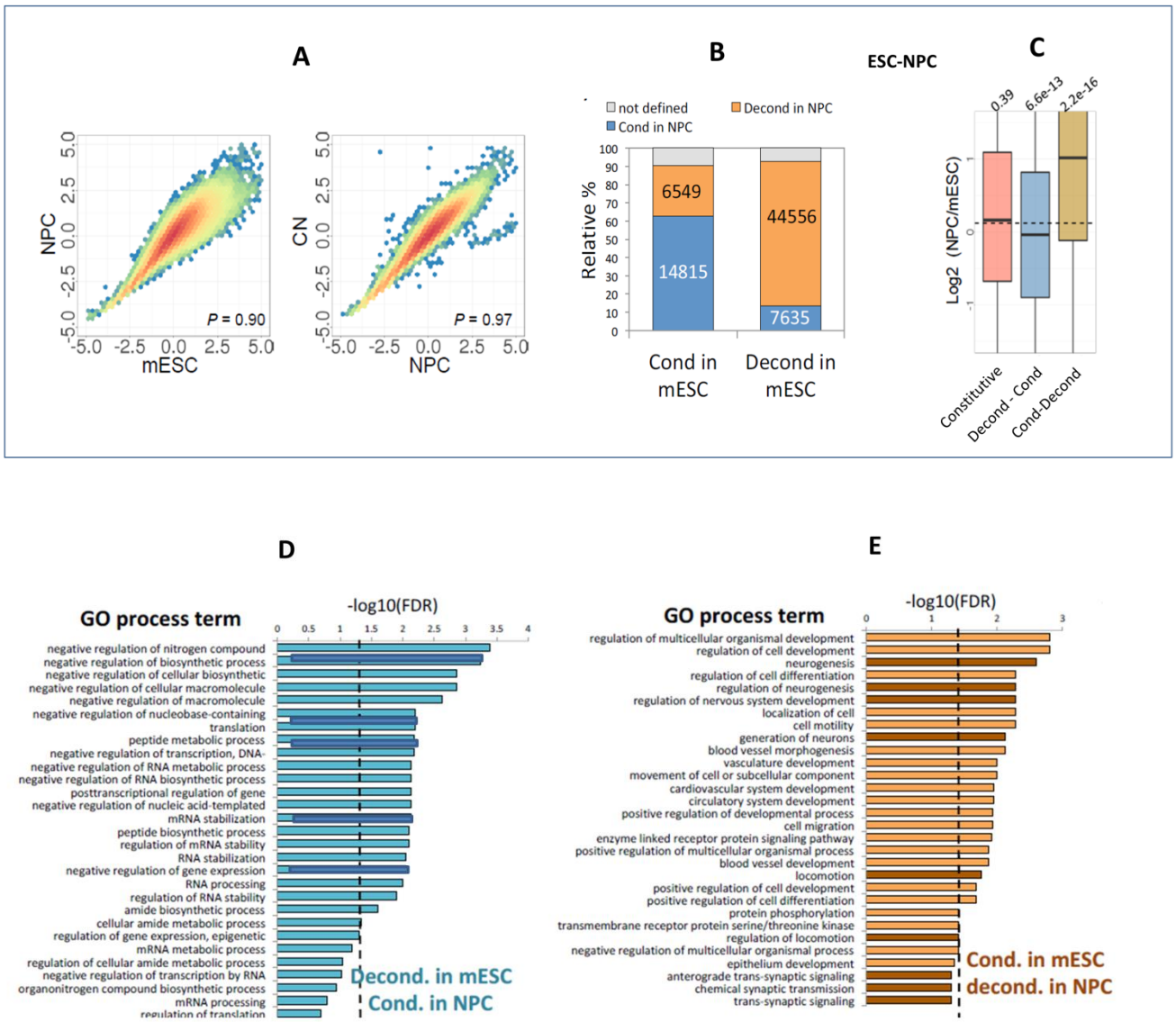


Fig 12. Developmental dynamics of condensed and decondensed domains

A. Scatterplots of corrected read counts in ESC vs. NPC and in NPC vs. CN. The correlation between NPC and CN (p=0.97) is greater than the correlation between mESC and NPC. (p=0.90). **B**. Distribution of condensed and decondensed states of chromatin domains during ESC to NPC differentiation. **C**. Enrichment of Gene Ontology Process terms among the genes exhibiting condensation (left) and decondensation (right) during ESC-to-NPC transition. Shown are the top 30 terms through ToppGene Suite. Nervous system associated terms are highlighted in brown colour.

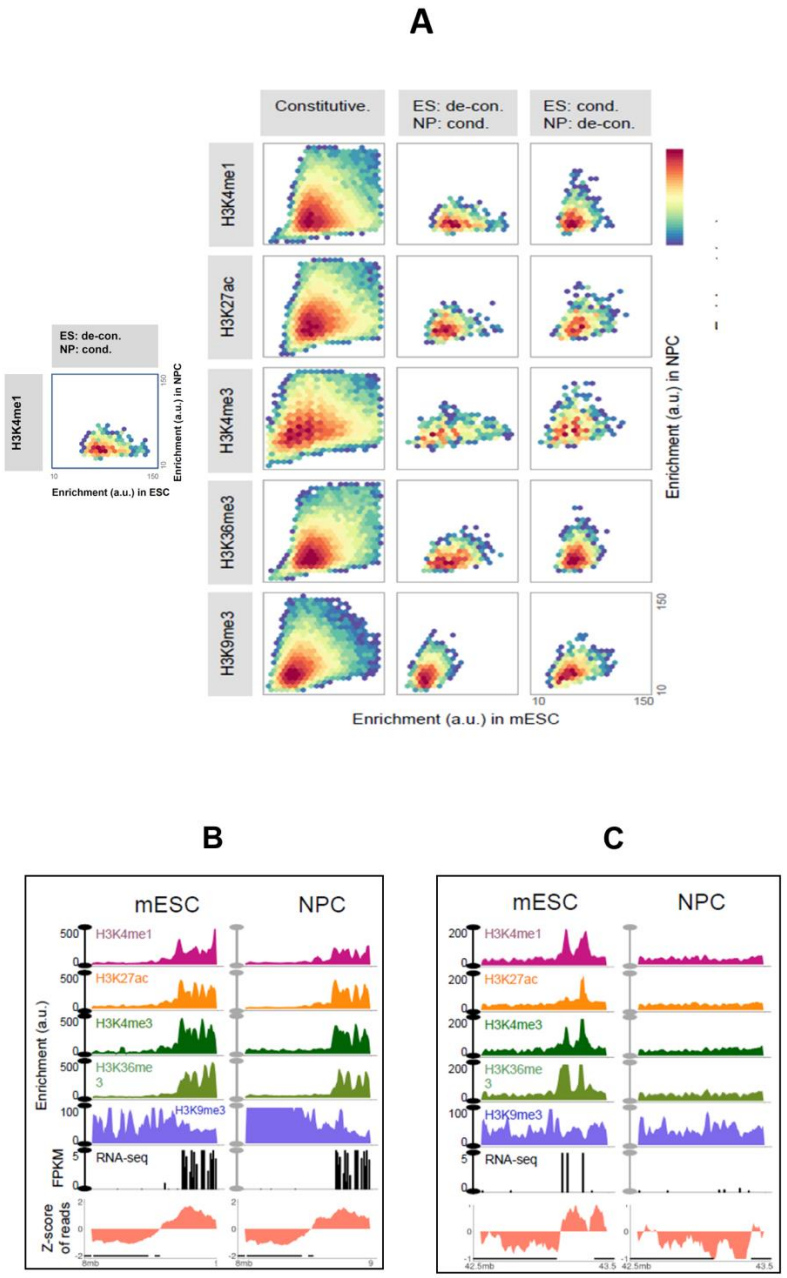


Fig 13. Epigenetic marks associated with constitutive and switching chromatin regions during neuronal differentiation

A. Scatterplots of histone modifications in domains that remained unchanged in mESC and NPC, and the ones that switched from decondensed to condensed or condensed to decondensed in mESC and NPC. **B&C.** Examples of decondensed and condensed domains that remained constutitive in mESC and NPC (left), and a decondensed region in mESC that switched to condensed state in NPC (right). Note that the gene expression and RZ scores remained similar in constitutive regions but they change their profiles during switching.

Non-LAD condensed regions associated with PcG

Though that the scatterplots of the active histone marks are tilted reasonably towards the X-axis which represents decondensation from ESC to NPC, we observed that the scatterplots of H3K9me3 are just slightly tilted towards the condensation. Previous studies suggest that H3K9me3 has just slight changes during differentiation of mouse ESC. Interestingly, we could see that the scatterplots of H3K27me3 and PcG proteins occupancy has a significant tilt towards the axis representing condensation from ESC to NPC (Fig.14A). We also observed that the expression levels of the genes targeted by PcG proteins were changed correspondingly. Together, these indicate that the non-LAD regions which were identified using our method may be suggestive of PcG repressed chromatin. Since facultative heterochromatin is associated with H3K27me3, these non-LAD condensed regions are suggestive of facultative heterochromatin.

Condensation Dynamics of stage specific Gene Loci

Here are the cell type specific genes' RZ profiles. As we can see, Oct4 is highly expressed in ESC and is drastically downregulated in NPC and further in CN. Subsequently, the RZ scores show that the domain encompassing Oct4 is decondensed in ESC, condensed in NPC and further condensed in CN (Fig.15A). The *Wnt5b* gene is specifically highly expressed in NPC and shows very little to no expression in ESC and CN. The region encompassing the gene is condensed in ES and CN, but eventually decondensed in NPC (Fig.15B). We took another example of CN cell type specific gene Pcdh9. This gene has very low expression in ESC, slightly increases in NPC and is highly expressed in CN. Subsequently, the domain encompassing the Pcdh9 gene is condensed in ESC, decondenses in NPC and further decondenses in CN (Fig.15C). Thus, we could show cell type specific examples of genes and condensed/decondensed state of the regions associated with them. The domains containing these genes are decondensed only in specific cell type where it has high expression and condensed in other cell types where its expression is very low.

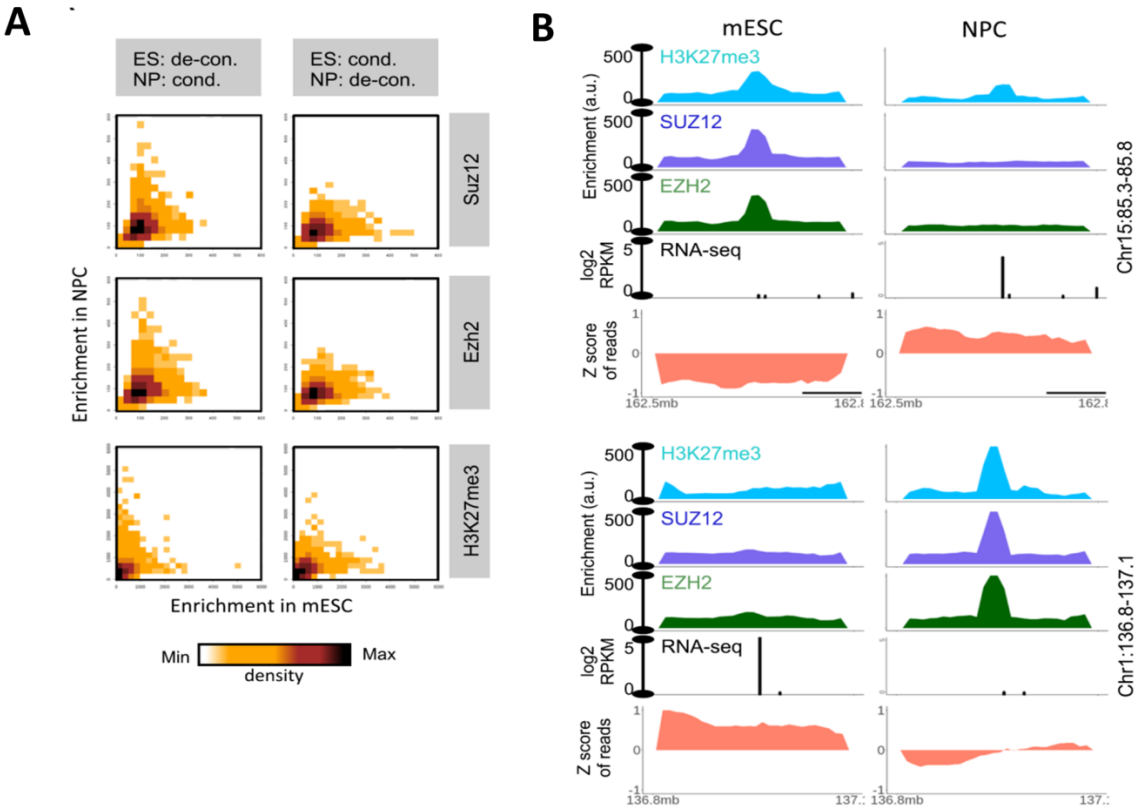


Fig 14. Polycomb association of non-LAD condensed domains identified through visibility bias.

A. Scatterplots showing enrichment of Suz12, Ezh2 and H3K27me3 in domains that switched from decondensed to condensed and condensed to decondensed during differentiation of ESC to NPC. **B**. Examples of decondensed and condensed domains that switched their status in ESC and NPC. Upper panel shows ESC to NPC decondensation and lower panel shows ESC to NPC condensation. Corresponding change in the gene expression (FPKM) and RZ scores has been shown.

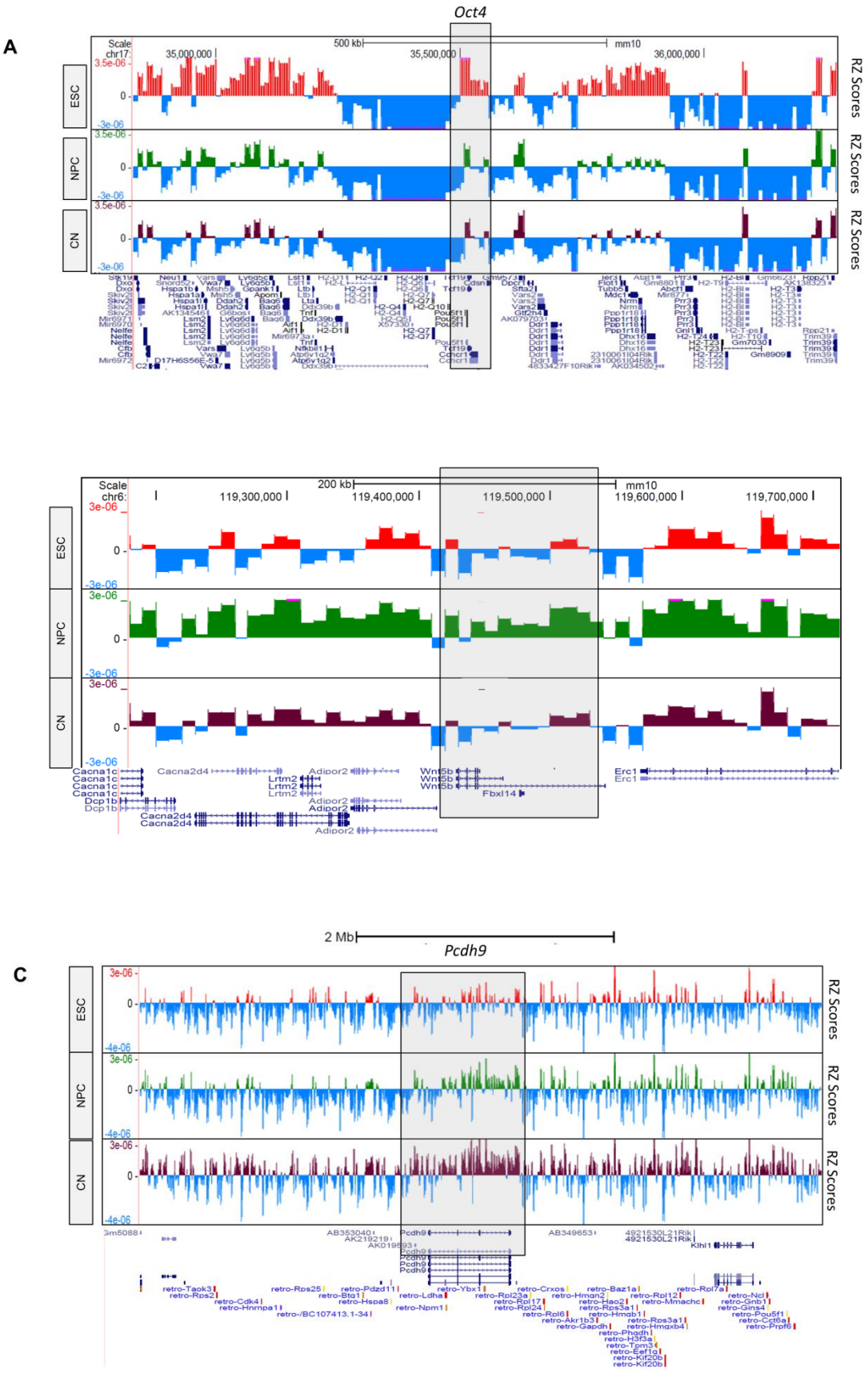
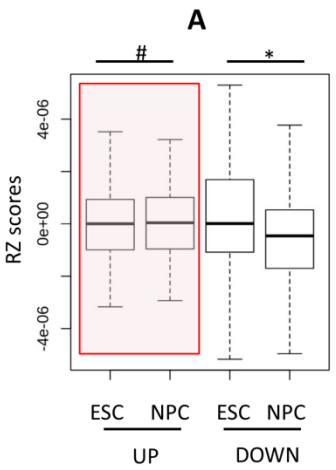
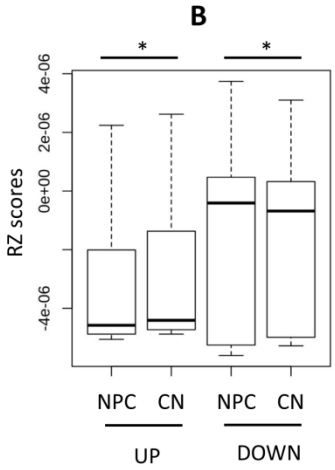
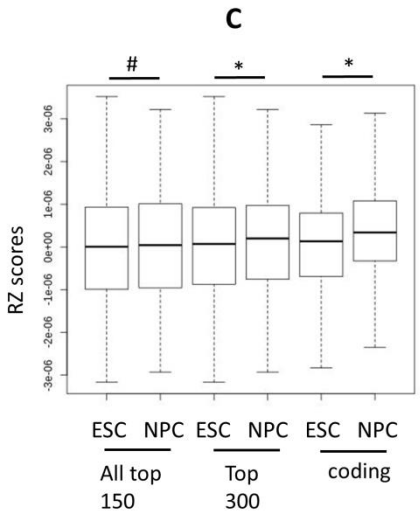

A
Oct4
Scale
chr17:
35,000,000
500 kb
35,500,000
mm10
36,000,000
3.5e-06
0
-3e-06
ESC
NPC
CN
RZ Scores
B
Scale
chr6:
119,300,000
200 kb
119,400,000
119,500,000
mm10
119,600,000
119,700,000
3e-06
0
-3e-06
ESC
NPC
CN
Cacna2d4
Adipor2
Lrtm2
Wnt5b
Fbxl14
Erc1
Dcp1b
C
2 Mb
Pcdh9
4e-06
AB353040
AK219219
AK019593
AB349653
4921530L21Rik
Klhl1
retro-Taok3
retro-Rps25
retro-Pdzd11
retro-Ldha
retro-Ybx1
retro-Crxos
retro-Baz1a
retro-Rpl7a
retro-Rps2
retro-Btg1
retro-Rpl23a
retro-Hmgn2
retro-Rpl12
retro-Ncl
retro-Cdk4
retro-Hspa8
retro-Rpl17
retro-Hac2
retro-Mmachc
retro-Gnb1
retro-Hnrnpa1
retro-Npm1
retro-Rpl24
retro-Rps3a1
retro-Gins4
retro-/BC107413.1-34
retro-Rpl6
retro-Hmgb1
retro-Pou5f1
retro-Akr1b3
retro-Rps3a1
retro-Cct6a
retro-Gapdh
retro-Hmgxb4
retro-Prpf6
retro-Phgdh
retro-H3f3a
retro-Tpm3
retro-Eef1g
retro-Kif20b

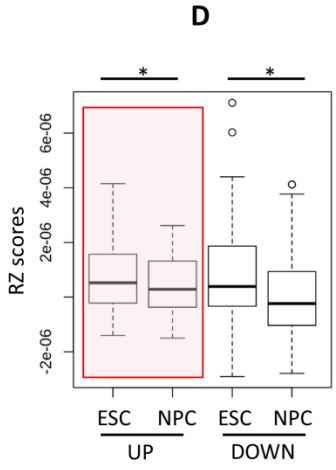
Fig 15. Condensation Dynamics of stage specific Gene Loci

A. RZ scores of domain encompassing *Oct4* locus, which has higher expression in ESC and very low expression in NPC and CN, showing decompaction in ESC but compaction in NPC and CN. **B**. RZ scores of domain encompassing *Wnt5b* locus, which has higher expression only in NPC, showing decompaction in NPC but compaction in ESC and CN. **C**. RZ scores of domain encompassing *Pcdh9* locus, which has higher expression in CN, showing decompaction in CN but compaction in ESC and CN


A
#
*
RZ scores
4e-06
0e+00
-4e-06
ESC
NPC
ESC
NPC
UP
DOWN


B
*
*
RZ scores
4e-06
2e-06
0e+00
-4e-06
NPC
CN
NPC
CN
UP
DOWN


C
#
*
*
RZ scores
3e-06
2e-06
1e-06
0e+00
-1e-06
-2e-06
-3e-06
ESC
NPC
ESC
NPC
ESC
NPC
All top 150
Top 300
coding


D
*
*
RZ scores
6e-06
4e-06
2e-06
-2e-06
ESC
NPC
ESC
NPC
UP
DOWN

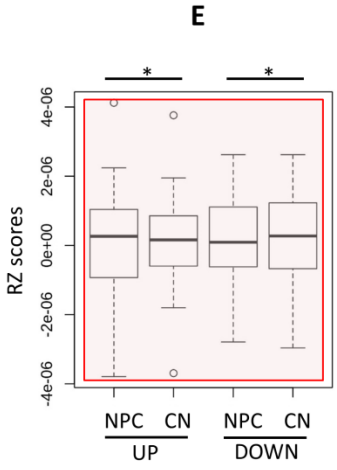

E
*
*
RZ scores
4e-06
2e-06
0e+00
-2e-06
-4e-06
NPC
CN
NPC
CN
UP
DOWN

Fig 16. Compaction dynamics of differentially regulated genes

A. Boxplots showing the RZ scores of top 150 differentially expressed genes between ESC and NPC. Both ESC to NPC upregulated and downregulated genes' RZ scores were shown. Highlight shows the anomalous trend. **B**. Boxplots showing the RZ scores of top 150 differentially expressed genes between NPC and CN. Both NPC to CN upregulated and downregulated genes' RZ scores were shown. **C**. Boxplots showing RZ scores of top 150 and top 300 upregulated genes along with RZ scores of only coding genes in top 300. **D**. Boxplots showing RZ scores of all miRNA genes which are differentially regulated between ESC and NPC. **E.** Boxplots showing RZ scores of all miRNA genes which are differentially regulated between NPC and CN.

represents p-value >0.05; * represents p-value <0.05. Statistical significance was calculated using Mann-witney U test.

Relationship between top differentially regulated genes and compaction

Although it is clear that higher gene expression is associated with decondensation and lower gene activity with condensation, we wanted to see if this is true for top differentially regulated genes between ESC, NPC and CN. Top 150 differentially expressed genes between ESC - NPC and NPC - CN were taken and the compaction status of their TSS was compared. ESC-NPC upregulated genes show no statistical correlation with the RZ scores, whereas ESC-NPC downregulation of the genes is accompanied by corresponding compaction of the TSS of the genes (Fig.16A). During NPC-CN upregulation, as expected, the genes were decondensed and during NPC-CN downregulation they condensed (Fig.16B).

Comparison between coding and non-coding genes

To probe the reason behind having no correlation between highly unregulated genes and decondensation during ESC-NPC, we looked into the nature of the top genes. We could observe that the top 150 genes mostly consist of predicted genes (whose official symbol is starting with Gm), which contains many non-coding genes and miRNAs. Then we listed out and checked the nature of top 300 upregulated genes from ES-NP. The top 300 upregulated genes consist of many coding genes and they are decondensed from ESC-NPC, which was expected. Also, when only coding genes are taken and analyzed, the decondensation is much more pronounced (Fig.16C). Hence, we could infer that non-coding and miRNA genes in the top upregulated genes were having this anomalous behavior.

Compaction dynamics of miRNA genes

Since many miRNAs were involved in anomalous behavior, we analyzed the condensation status of only miRNAs. During ESC-NPC, the upregulation of miRNAs is accompanied by condensation of the TSS of these genes. But the downregulation of miRNAs during ESC-NPC is also accompanied by condensation (Fig.16D). During NPC-CN, upregulation of the miRNAs is accompanied by condensation and downregulation by decondensation (Fig.16E). Thus, we show that condensation and decondensation status of the miRNAs is inversely related except during ESC-NPC downregulation.

Reverse Correlation between Condensation and gene expression:

It is well established in this study as well as previous ones that condensed chromatin has lower gene expression than decondensed chromatin. But on the contrary, we observed the presence of regions in which decondensation and gene expression are inversely correlated in miRNA genes. Hence, we wanted to see the presence of such regions genome wide. To this end, we calculated percentage of total genes with positive and negative correlation between gene expression and condensation status. In ESC to NPC, we found out that nearly 89% of the genes

are having positive correlation and nearly 11% of the total genes have negative correlation between gene expression and condensation status (Fig.17A). To further probe into inversely correlated regions, we divided them into two types. (1) ESC-NPC condensed domains showing Gene Upregulation**.** (2) ESC-NPC de-condensed domains showing Gene downregulation**.** To further characterize them, we probed the active histone mark H3K27ac and repressive histone mark H3K9me3 associated with these regions and further classified them (Fig.17B).

The first type of regions are further classified as

(1a) The regions which are condensed during ESC-NPC differentiation but show gene upregulation, and upregulation of active histone mark and downregulation of repressive marks (Fig.18A). These regions include genes which are involved in pathways such as Alcoholism, systemic lupus erythematosus, protein processing in ER, pathways in cancer etc. This observation could be due to formation of transcriptional condensates.

(1b) The regions which are condensed during ESC-NPC differentiation but show gene upregulation, and downregulation of active marks and upregulation of repressive marks (Fig.18B). The genes present in these regions are involved in pathways such as transcriptional regulation, dopaminergic neuron differentiation, dendrite formation, cytoskeleton etc. This behaviour is probably due to the presence of Imprinted genes

The second type of regions are also further classified as

(2a) The regions which are decondensed during ESC-NPC differentiation but show gene downregulation, and downregulation of active marks and upregulation of repressive marks (Fig.18C). These regions consist of genes which are involved in pathways such as metabolic pathways, arginine biosynthesis, carbon metabolism, biosynthesis of amino acids etc. This observation could be due to the presence of Facultative heterochromatin.

(2b) The regions which are decondensed during ESC-NPC differentiation but show gene downregulation, and upregulation of active marks and downregulation of repressive marks (Fig.18D). These regions harbour genes involved in pathways such as metabolic pathways, metabolism of xenobiotics, drug metabolism - cytochrome P450, chemical carcinogenesis. This behaviour might be due to presence of bivalent poised genes.

Thus, reverse correlation between decondensation and gene expression exists in the genome and they could be further classified based on other epigenetic marks.

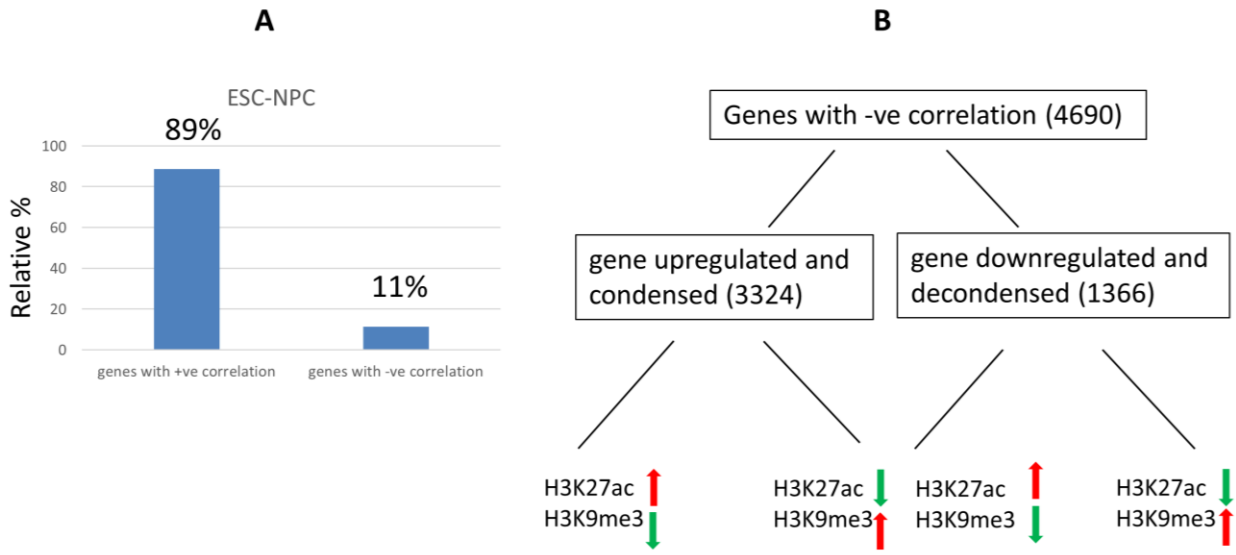


Fig 17. Reverse correlation between chromatin condensation and gene expression

A. Bar chart showing relative percentage of total genes which show direct correlation and reverse correlation between chromatin condensation and gene expression. Majority of the genes have positive correlation between condensation and gene expression**. B**. Flowchart showing classification of reverse correlated genes based on active and repressive histone marks associated with them along with no. of genes in brackets within each category.

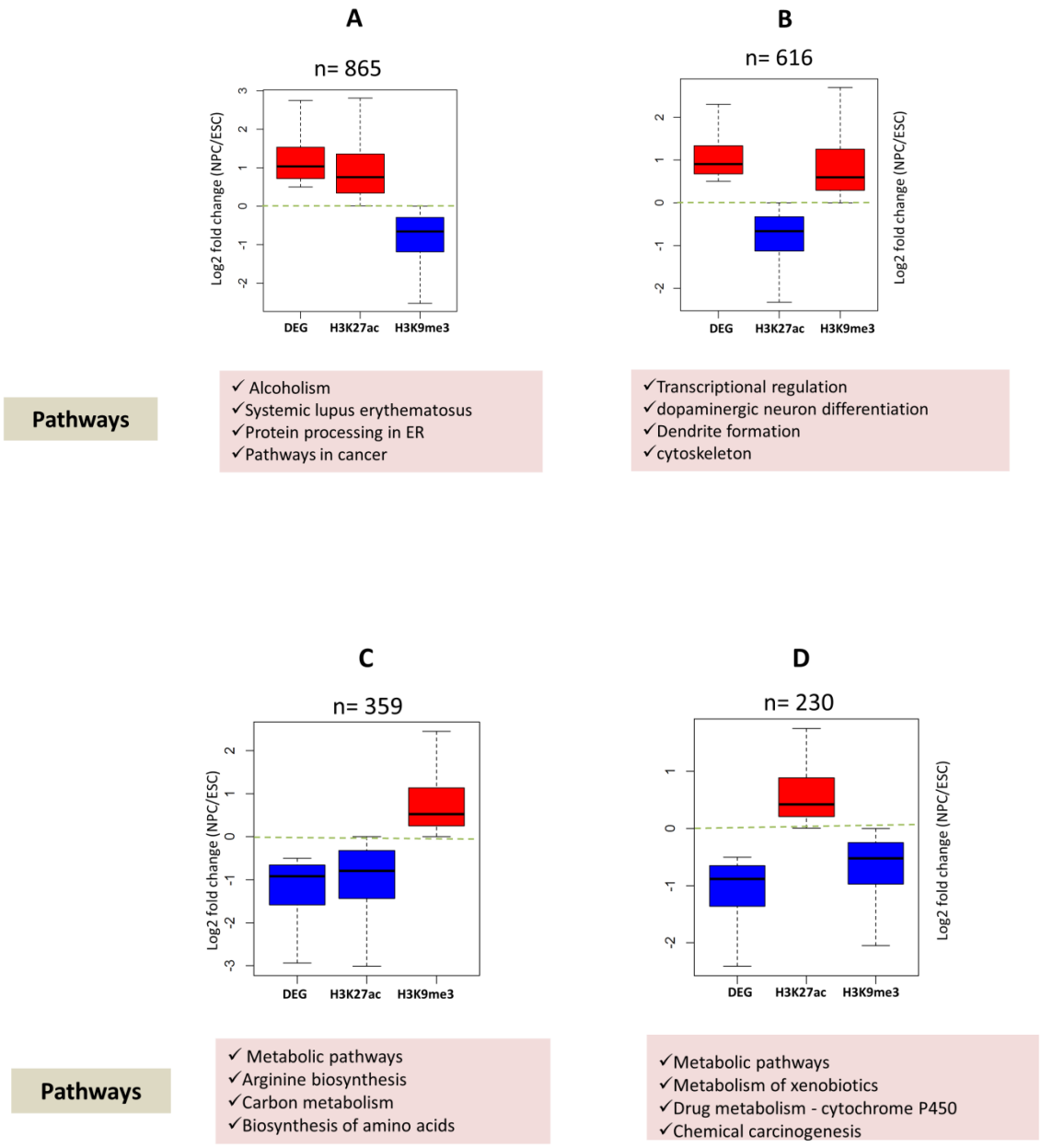
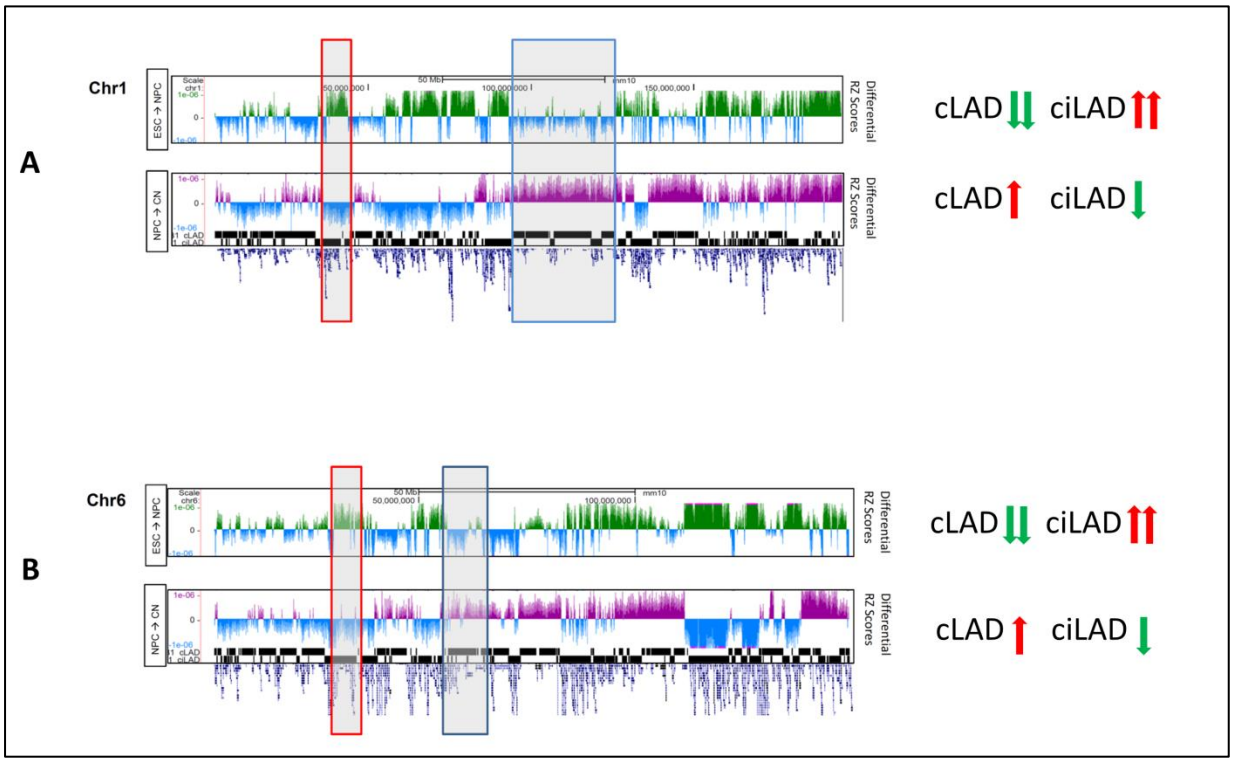


Fig 18. Classification of reverse correlated genes based on histone marks during ESC to NPC

A. Genes with upregulated expression, upregulated H3K27ac, downregulated H3K9me3 and are condensed from ESC-NPC. **B**. Genes with upregulated expression, downregulated H3K27ac, upregulated H3K9me3 and are condensed from ESC-NPC. **C**. Genes with downregulated expression, downregulated H3K27ac, upregulated H3K9me3 and are decondensed from ESC-NPC **D**. Genes with downregulated expression, upregulated H3K27ac, downregulated H3K9me3 and are decondensed from ESC-NPC.

Accessibility dynamics during differentiation:

Then we wondered what is the trend of the accessibility at chromosomal level between the cell types. To determine this, we made differential (subtraction) RZ score plots of NPC-ESC and CN-NPC for Chr1 (Fig.19A&B). We observed that differentiation of ESC to NPC is accompanied by loss of reads in cLADS and gain in ciLADs i.e., the cLADs condense from ESC-NP and then decondense from NPC-CN. Whereas ciLADS decondense from ESC-NPC and condense from NPC-CN (Fig.19A). We checked whether the above trend is true for all chromosomes and it turned out that the trend was true for all the chromosomes. Here is another example of Chr6 showing the differential RZ scores (Fig.19B). Then we wondered if this alternative loss and gain of RZ scores in cLADs and ciLADs is accompanied by corresponding change in the gene expression. Shown here is an example of ciLAD region from Chr6 (Fig.19C). This ciLAD region decondenses in NPC with a corresponding gain in gene expression and condenses in CN with concomitant loss of gene expression.


A
Chr1
ESC → NPC
NPC → CN
Differential RZ Scores
cLAD
ciLAD
cLAD
ciLAD
B
Chr6
ESC → NPC
NPC → CN
Differential RZ Scores
cLAD
ciLAD
cLAD
ciLAD

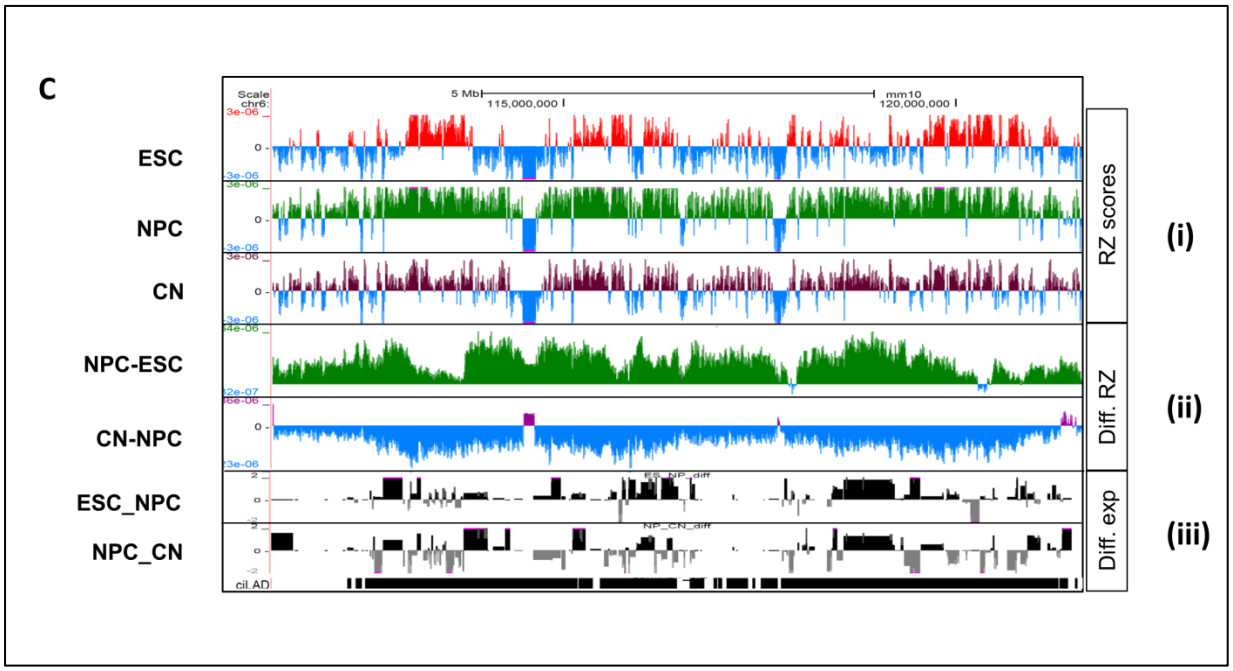

C
Scale
chr6:
5 Mb
115,000,000
mm10
120,000,000
ESC
NPC
CN
NPC-ESC
CN-NPC
ESC_NPC
NPC_CN
ciLAD
RZ scores
Diff. RZ
Diff. exp
(i)
(ii)
(iii)

Fig 19. **Accessibility dynamics during differentiation**

A. Whole Chr1 RZ scores showing condensation of cLADs and decondensation of ciLAD during ESC to NPC. Double arrows represent intensity of condensation/decondensation, while cLADs decondense and ciLADs condense during NPC to CN. **B**. Whole Chr6 RZ scores showing condensation of cLADs and decondensation of ciLAD. Double arrows represent more intensity of codensation/decondensation than single arrows, while cLADs decondense and ciLADs condense during NPC to CN. **C**. Association between Accessibility dynamics and gene expression showing RZ scores(i), differential RZ scores(ii) and differential gene expression from ESC to NPC and NPC to CN.

Objective 3

Dynamics of quantitative spatial genome architecture during neuronal differentiation

a) Accessibility problem in 2D Hi-C maps

b) Remodeling of chromatin architecture during differentiation

a) Accessibility problem in 2D Hi-C maps

The dynamics of 3D nuclear organization during differentiation of ESC to Cortical Neurons is not well studied and there exist many gaps in the knowledge. To pursue further, we have analyzed the ultra-deeply sequenced Hi-C data, which was also used in the previous objectives. The interactions between the genomic loci were determined using HOMER and contact matrices were generated at different resolutions both genome and chromosome wide. Basically, each chromosome or a region of the genome is divided into equal sized bins (of base pair lengths) and the interactions between each pair of regions is obtained as a contact matrix. The contact matrix is then drawn as a heatmap, where the color of the interactions represents degree of interactions.

Restriction Enzyme Accessibility Problem:

As seen from previous observations, the condensed chromatin is less accessible to restriction enzymes than the decondensed chromatin. As Hi-C is a proximity-based assay, the interactions represent how close two regions are in the nucleus are. But since condensed chromatin gives less no. of reads, the interactions coming from condensed chromatin are less represented in Hi-C matrix than the decondensed ones and gives an interpretation that the regions in the condensed chromatin are farther from each other. We refer this problem as **Restriction Enzyme (RE) accessibility problem**. This potential bias should be normalized to obtain the true proximity-based values.

To address the above problem, we have selected a genomic region which encompasses *Nanog* gene (Fig.20). This was referred as Nanog domain whose genomic coordinates are Chr6:122240000-123200000. The domain is roughly 1Mb long. We derived the RZ scores for this region using LOESS normalization for all the three cell types i.e., ESC, NPC and CN. We can clearly observe that the whole domain condenses from ESC to NPC and further in CN. This is in accordance with the fact that *Nanog* is highly and specifically expressed in ESC, gets downregulated in NPC and CN. From the above figure, it is clear that the Nanog domain condenses from ESC to NPC and then in CN.

On the other hand, we overlaid the contact matrices of the Nanog domain in ESC and NPC with active histone mark H3K27ac and silent histone mark H3K9me3. We observed that there is decrease in the active histone mark H3K27ac around the *Nanog* gene from ESC to NP. Concomitantly, there is increase of silent histone mark H3K9me3 at the *Nanog* locus (Fig.21A). On one hand it is clear that Nanog domain is condensed form ESC to NPC, and on the other hand it is associated with corresponding change in the histone marks. Hence, in theory, the genomic regions at the *Nanog* locus come closer to each other in NPC and correspondingly the interactions should increase. But, when we generated differential contact matrix (NPC-ESC), we observed that there is loss in interactions along the diagonal, which is represented by green colour in the heatmap (Fig.21A third panel). This discrepancy is due to RE accessibility problem.

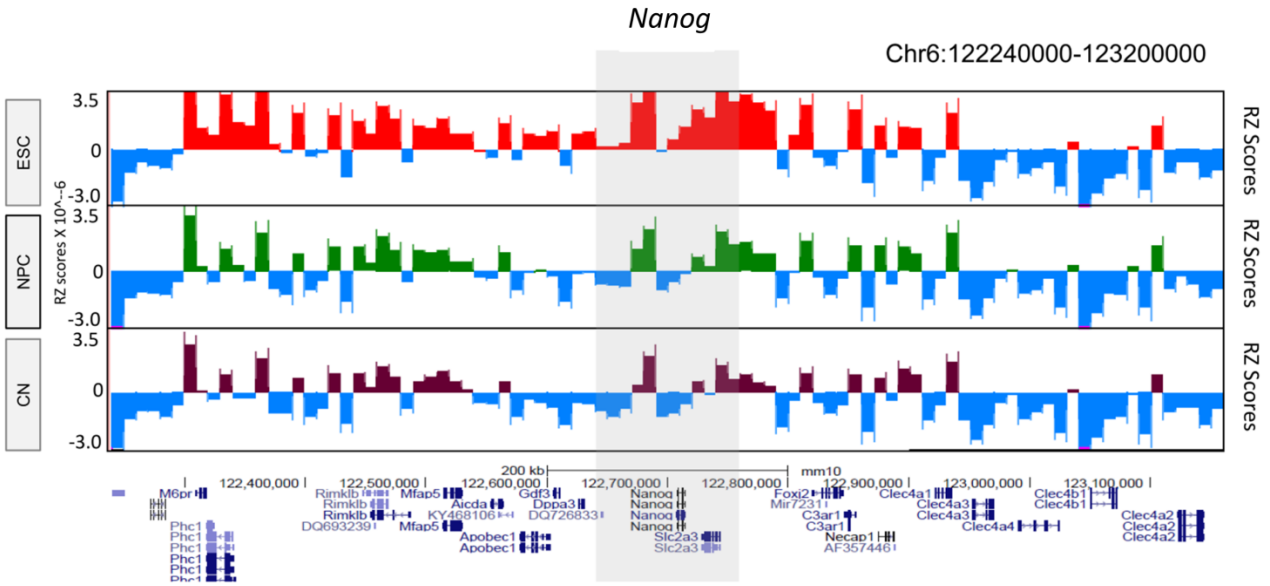


Fig 20. Condensation status of domain encompassing Nanog

Since Nanog has higher expression in ESC but lower expression in NPC and CN, RZ scores of Nanog domain showing that the whole domain has high RZ scores (decondensed) in ESC but lower RZ scores (condensed) in NPC and CN. This shows that not only the Nanog locus but the adjacent region is decondensed in ESC. Genomic coordinates of the domain were given in the upper right corner and the region just representing *Nanog* gene is highlighted

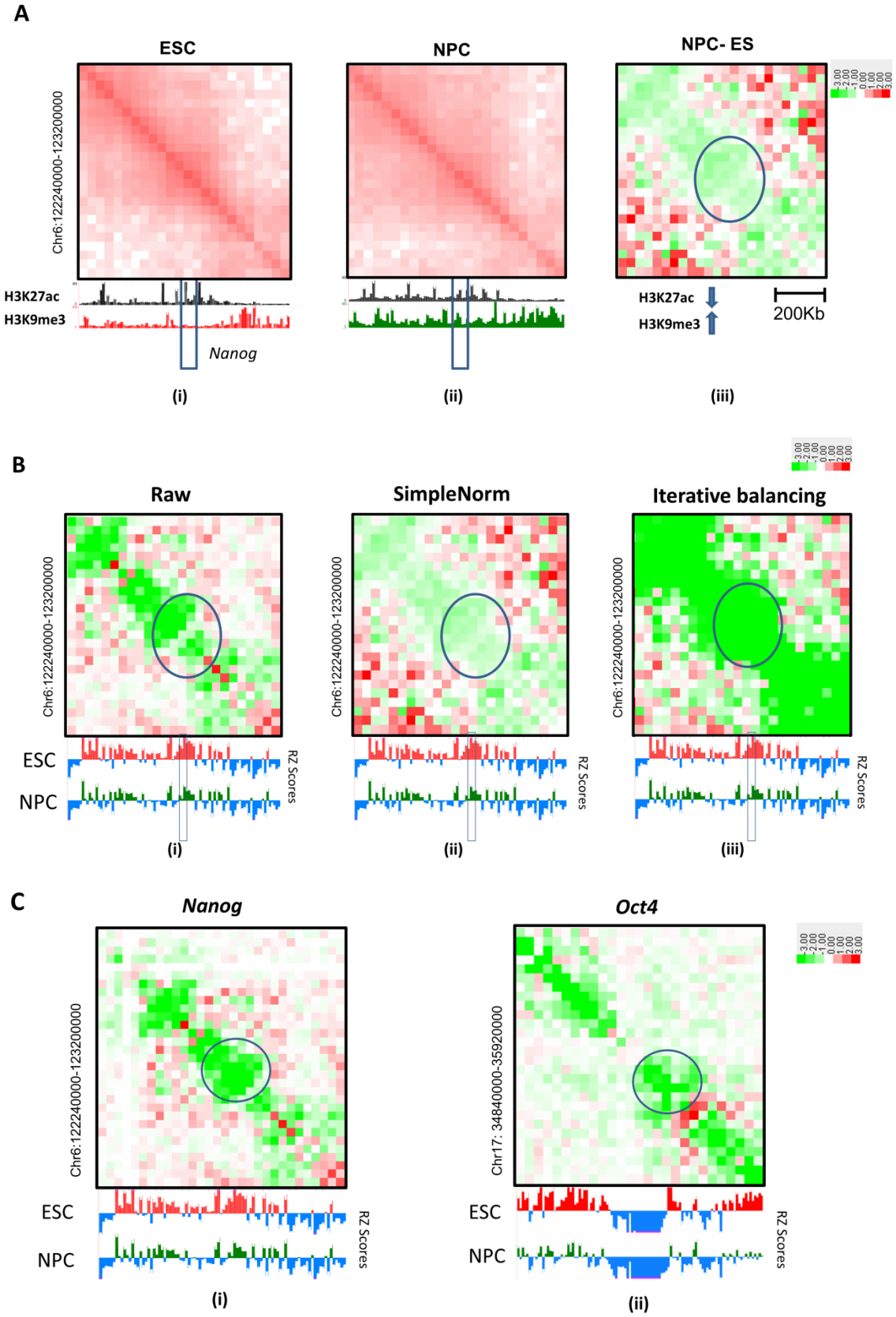

A
ESC
NPC
NPC- ES
Chr6:122240000-123200000
H3K27ac
H3K9me3
Nanog
200Kb
(i)
(ii)
(iii)
B
Raw
SimpleNorm
Iterative balancing
ESC
NPC
RZ Scores
(i)
(ii)
(iii)
C
Nanog
Oct4
Chr6:122240000-123200000
Chr17: 34840000-35920000
ESC
NPC
RZ Scores
(i)
(ii)

Fig 21. Restriction Accessibility problem in Nanog and Oct4 domains

A. (i) Heatmap of Nanog domain in ESC overlaid on H3K27ac amd H3K9me3. **(ii)** Heatmap of Nanog domain in NPC overlaid on H3K27ac amd H3K9me3. **(iii)** Subtraction matrix of NPC-ESC. Red shows gained interactions and green shows lost interactions. Region representing Nanog gene is rounded. **B. (i,ii** and **iii)** Subtraction matrices of NPC-ESC showing Nanog domain over different Normalization methods ie., Raw, SimpleNorm and Iterative balancing respectively. RZ scores of ESC and NPC are shown for comparision. **C**. Accessibility problem shown in Nanog and Oct4 domains where encircled region shows the gene locus and green colour along the diagonal shows loss of interactions.

To verify that whether this problem is intrinsic to the normalization method we used (SimpleNorm), we Normalized the raw contact matrix with other Normalization methods such as Iterative Correction and could see similar results (Fig.21B). Thus, it became apparent that the RE accessibility problem was not intrinsic to any type of Normalization method, but is the artifact of the Hi-C experiment itself. We also checked many known regions and the same held true for all of them (Fig.21C).

Principle for correction:

As we can see in the differential matrix of the Nanog domain, only the regions close to diagonal are exhibiting this discrepancy but off the diagonal interactions are in accordance with the change in condensation. This means, up to a size limit, the actual increase in the interactions and thus proximity values show inversely in the Hi-C data due to RE accessibility problem, and beyond the limit the accessibility problem doesn't show much of its effects. Thus, first the threshold should be determined and then the correction should be done only for the near-diagonal elements within the threshold zone (Fig.22A).

For this we generated distance decay powerplots of the differential matrix to determine the average length of the threshold where the curve starts getting positive values (Fig.22B). We made these powerplots for both Nanog and Oct4 and could see that on an average, the threshold is around 200kb in length (Fig.22C&D).

Then, all the values within this belt along the diagonal were corrected using the formula

$$CCDD(Mij) = max(belt) * min(belt) / mij$$

This method was termed as **Contact Correction through Distance Decay plots (CCDD)**. This formula ensures that the interaction values within the belt are inversely corrected but the average, max and min of the values of the belt remain same.

We applied this CCDD method to matrices of ESC and NPC and then derived the differential matrix. After Correction, we could clearly see that the interactions around the Nanog gene are increasing from ESC to NPC and these values mimic the true proximity based interactions (Fig.23A). Similarly, we generated the differential matrix based on CCDD corrected values for Oct4 domain and could see that the interactions around the Oct4 gene now increase from ESC to NPC (Fig.23B).

Thus, by accounting for bias due to restriction accessibility problem, we corrected the contact matrices based on the distance decay plots. These corrected matrices now truly represent the true proximity-based values.

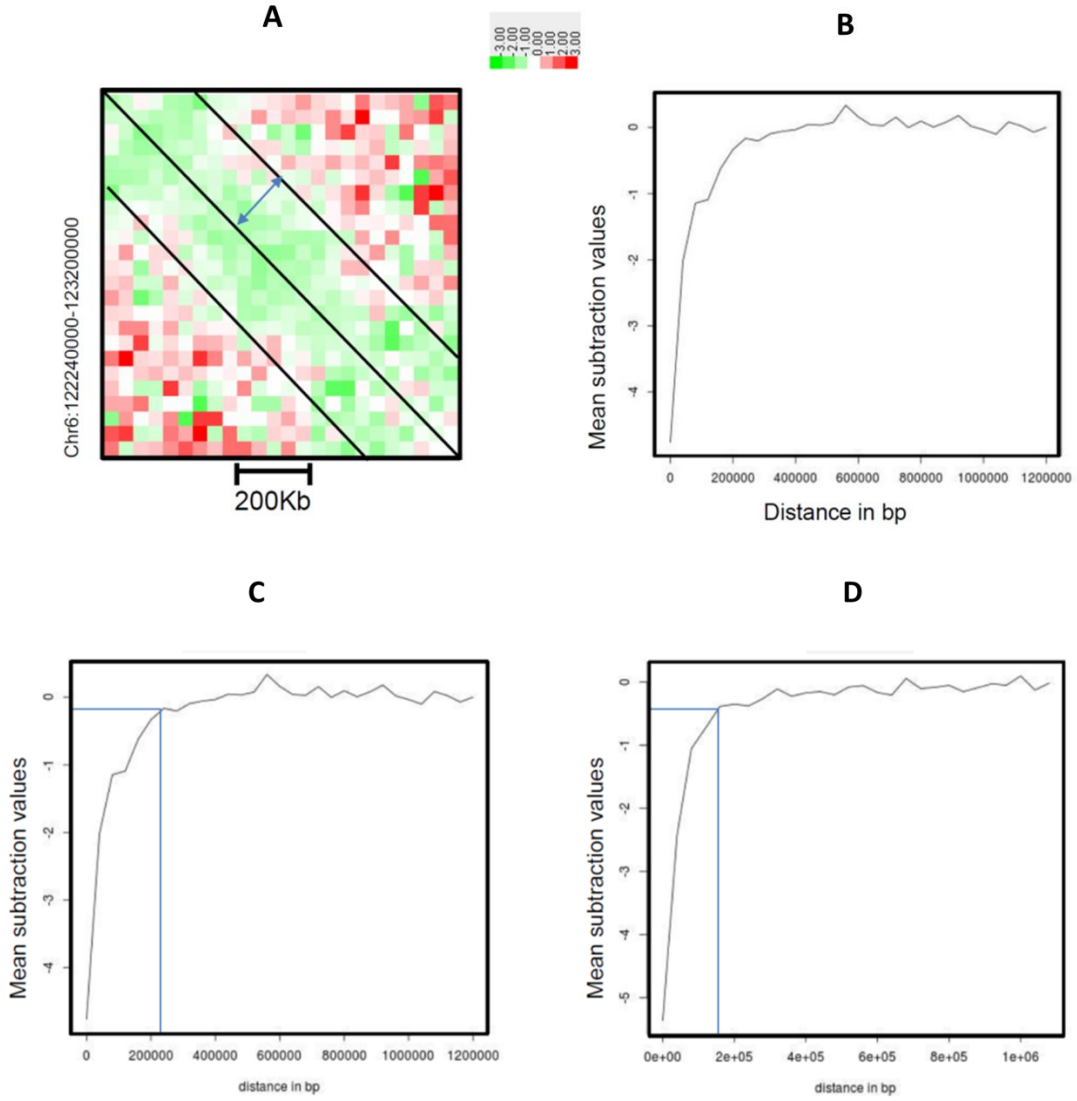


Fig 22. Identifying the threshold for the correction

A. Subtraction matrix of NPC-ESC of Nanog domain showing diagonal elements enriched in green color indicating loss of interactions. The limit of the lost interactions off the diagonal has been marked with black lines. **B**. A typical distance decay plot of mean subtraction values shows that the loss of interactions decreases as we move from the diagonal. **C&D**. Distance decay plots of Nanog(C) and Oct4(D) domains along with identified and marked threshold. This threshold is almost similar in both the loci averaging at about 200kb.

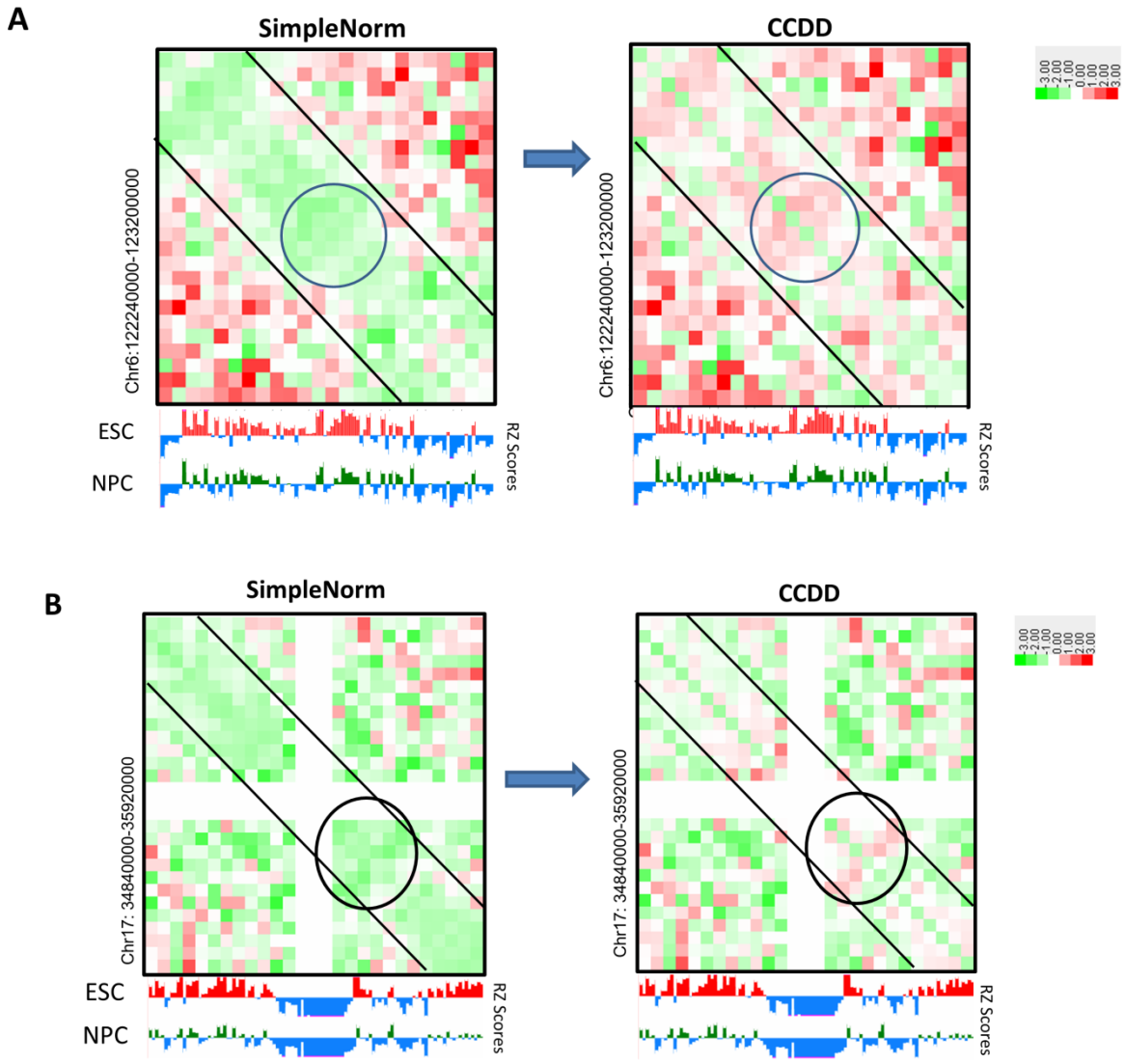


Fig 23. Contact Correction through Distance Decay plot

A. Correction of RE accessibility problem of Nanog domain using CCDD method. Threshold and the gene locus have been marked. RZ scores have been provided for reference. The reversal of the interactions and gain of interactions along the diagonal of the Nanog domain represented in red color could be clearly seen **B**. Correction of RE accessibility problem of Oct4 domain using CCDD method. Threshold and the gene locus have been marked. The reversal of the interactions and gain of interactions along the diagonal of the Oct4 domain represented in red color could be clearly seen.

Application in 3D polymer modeling:

The Hi-C contact matrices could be used to simulate 3D polymer modeling and obtain 3D conformations of the chromatin. These simulations are based on the contact frequency from the Hi-C data. Higher frequency between two regions is associated with higher springs between the beads and hence closer they would be. Lower contact frequency represents farther regions in the 3D conformations.

A well ignored fact in all the simulations done so far is that regions in the condensed chromatin are closer to each other and generate less no. of Hi-C contacts, as discussed earlier. Hence the lesser no. of interactions when simulated give the impression that the chromatin is decondensed. Consequently, decondensed euchromatin regions with higher no. of interactions would provide packed chromatin, when simulated. To address this problem, we, along with our collaborators (Prof. Ranjit Padinhateeri, IITB, India), proposed a bead and spring model with heteropolymer beads, where the size of the bead depends on the no. of 1-Dimensional Hi-C reads generated from that bin (Fig.24). That means, a 100kb bin with more no. of reads, which is decondensed (Fig.24ii), will have bigger volume as a bead than a 100kb bin bead with lesser no. of reads (Fig.24i). Essentially, the size of the bead is directly proportional to the no. of Hi-C reads obtained from that bin. We referred this as Heteropolymer model (Fig.24iv).

To test the model, we took Chr6 and Nanog domain as an example. As we have seen earlier, the Nanog domain gets condensed from ESC to NPC and further in CN. We modelled Chr6 and Nanog domain using both homopolymer and heteropolymer models. The volume of the Nanog domain was represented as Radius of gyration (Rg) (Fig.25A). As we can see, one can clearly observe that in Homopolymer model, the Rg of the Nanog domain is decreased from ESC to NPC and increases during NPC to CN (Fig.25B). But, in heteropolymer model, the Rg decreases from ESC to NP and further decreases from NP to CN. This serves as the proof of principle of the heteropolymer model (Fig.25C). We performed similar simulations using heteropolymer model for Chr17 along with Oct4 domain and could clearly observe that Rg of Oct4 domain decreases from ESC to NP and further to CN (Fig.26A & B).

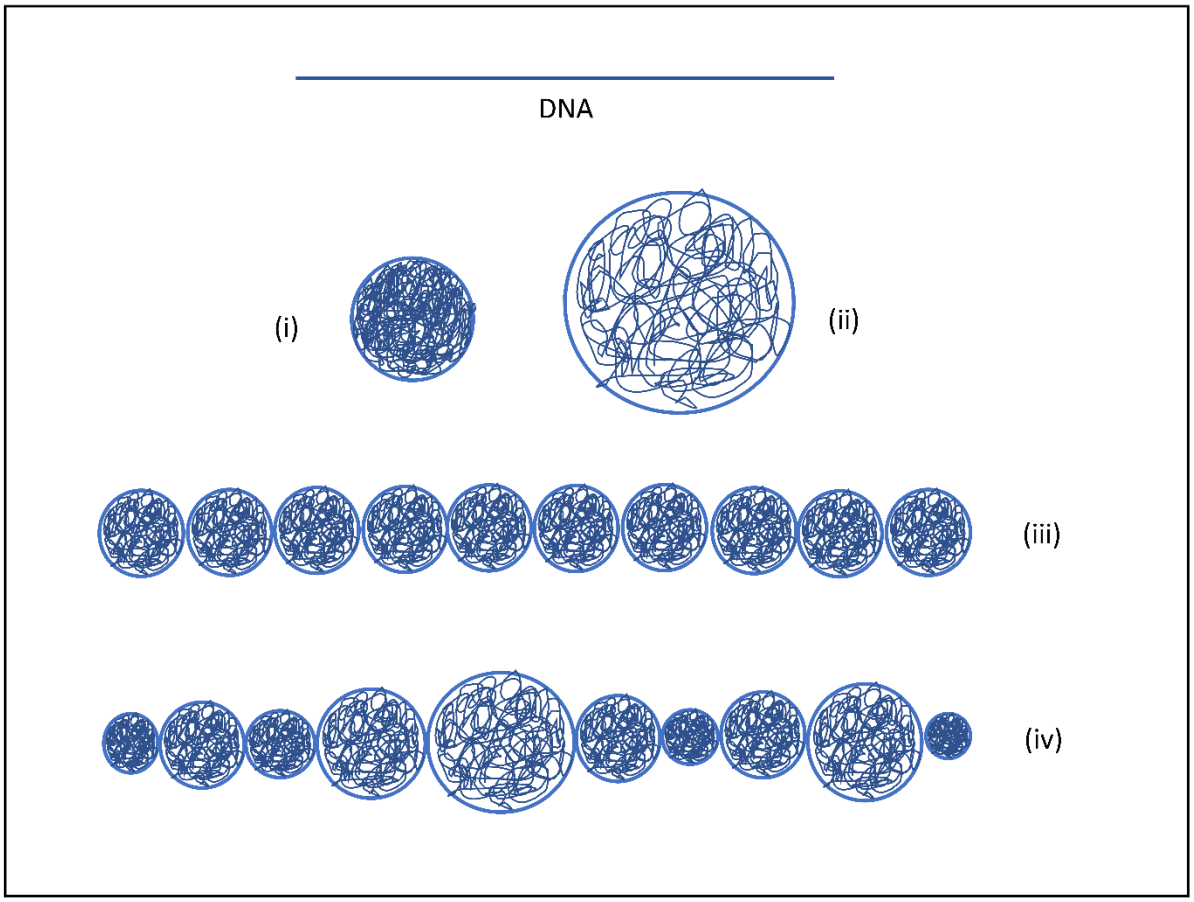


Diagram adapted from Dr. Ranjith Padinhateeri Lab

Fig 24. Comparison of Homopolymer and Heteropolymer model

The Diagram is adapted from Dr. Ranjith Padinhateeri's Lab **(i)** Example showing certain length of a DNA segment packed compactly into monomer of smaller radius. **(ii)** The same length of DNA packed relatively loosely into monomer of bigger radius. **(iii)** General Homopolymer modeling showing all monomers of equal radius irrespective of their compaction status. **(iv)** Novel Heteropolymer showing different monomer sizes depending of the condensation status of the DNA. Decompacted DNA is represented with bigger monomers and compacted DNA with smaller monomers, depending upon the compaction status of the DNA.

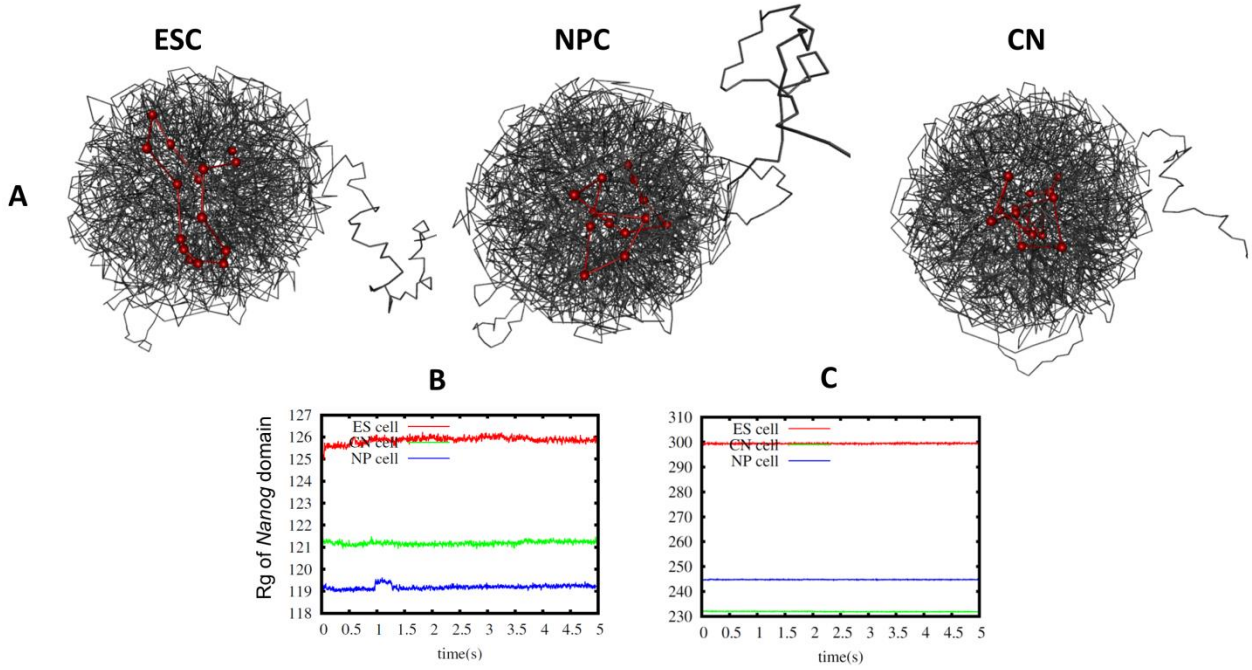


Fig 25. 3D modeling of Chr6 using Hi-C data

A. 3D models of the whole chromosome (chr6) simulated using Hi-C data and heteropolymer modeling. Red color represents Nanog domain's conformation and black line represents the remaining chromosome. **B**. Radii of gyration (Rg) of Nanog domain using Hi-C contact matrix without CCDD correction and simulated using homopolymer model which shows decrease in Rg from ESC-NPC but increase in Rg from NPC-CN. **C**. Radii of gyration of Nanog domain using Hi-C contact matrix corrected using CCDD and simulated using heteropolymer model showing that the Rg decreases from ESC to NPC and further in CN

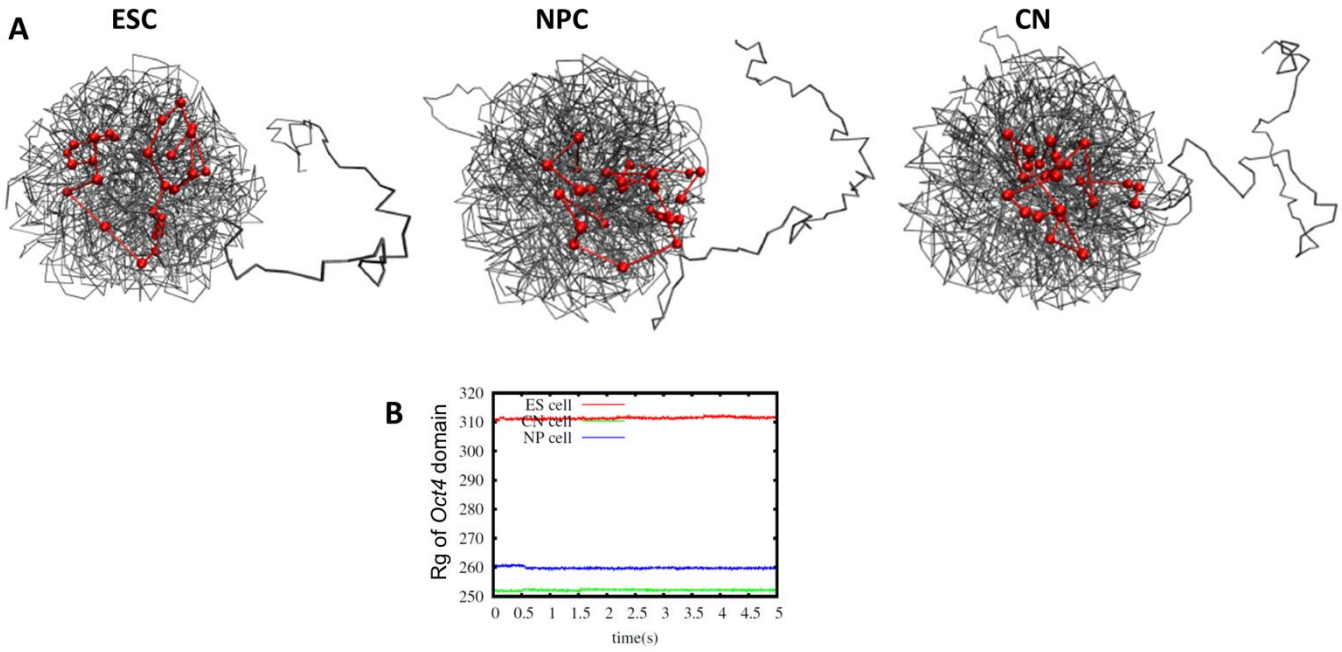


Fig 26. 3D modeling of Chr17 using Hi-C data

A. 3D models of the whole chromosome (chr17) simulated using Hi-C data and heteropolymer modeling. Red color represents Oct4 domain's conformation and black line represents the remaining chromosome. **B**. Radii of gyration of Oct4 domain using Hi-C contact matrix corrected using CCDD and simulated using heteropolymer model showing that the Rg decreases from ESC to NPC and further in CN

b) Remodeling of chromatin architecture during differentiation

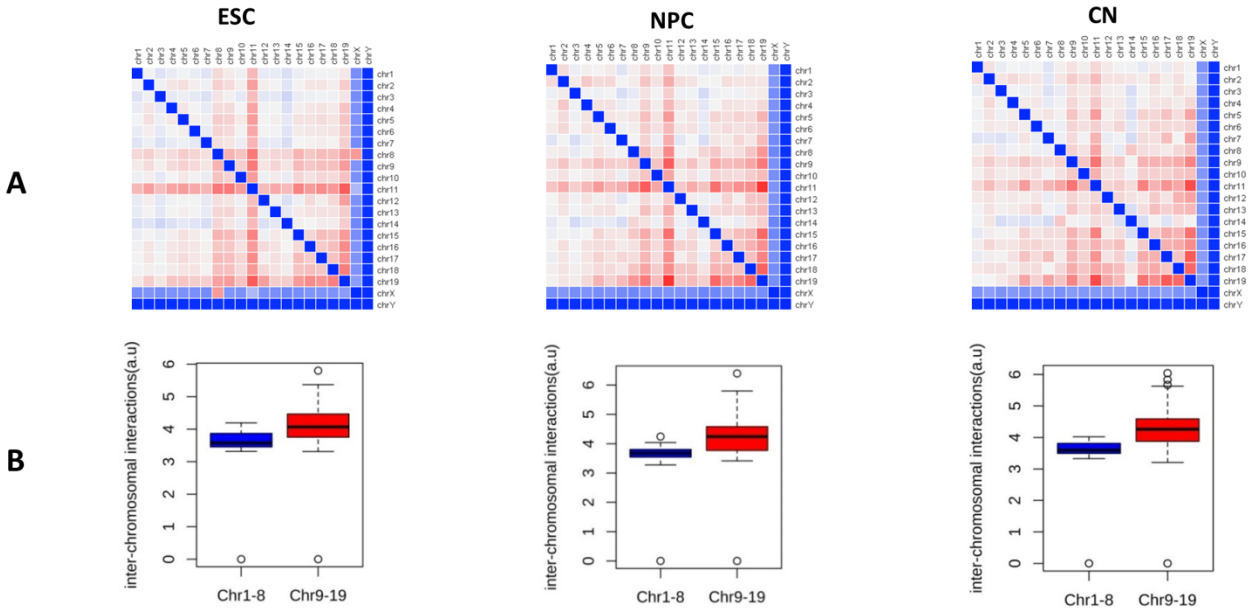
Chromosome Territory Neighborhood and Clustering:

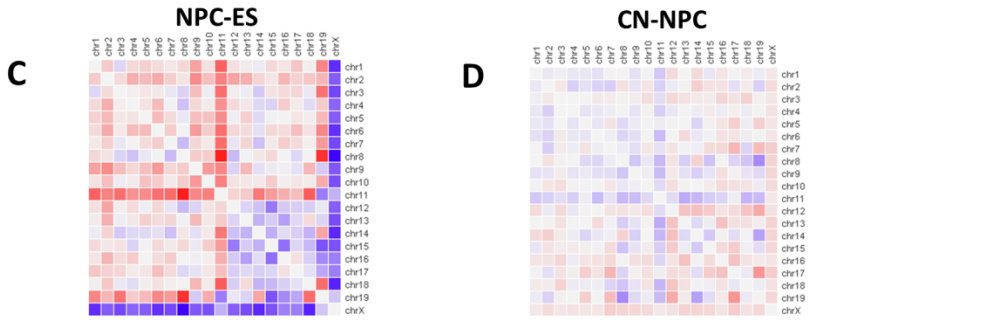
As chromosome neighborhood is cell type specific and reorganization happens during differentiation, we wanted to study the reorganization of chromosome territory Neighborhood (CTN) in the context of ESC differentiating into NP and then to CN. We obtained whole genome contact matrices of the three cell types, normalized them and obtained chromosome neighborhood maps (see methods section). The maps were represented as heatmaps. Red represents higher contact frequency and blue represents lower contact frequency (Fig.27A).

The CT neighborhood map shows how close two chromosomes are in the 3D nucleus. It is clear in all the three cell types that smaller chromosomes are clustered, which is in accordance with previous studies (Liebermann et al., 2008). The clustering of the chromosomes can be quantitatively seen as boxplots which shows smaller chromosomes (Chr9-19) clustering among themselves is more than the clustering of larger chromosomes (Chr1-8). We can also observe that chromosome 9 and 11 are having higher interactions with all the chromosomes, although it differs between the cell types (Fig.27B).

To see the reorganization of the CT neighborhood between the cell types, we generated subtraction matrices of NPC-ES (Fig.27C) and CN-NPC (Fig.27D). The red colour here represents gain in the interactions and blue represents loss in the interactions. From ESC to NPC, there is gain in the interactions within larger chromosomes which is also clear in the box plot (Fig.27E). There is an overall increase in the interactions within smaller chromosomes but specifically interactions within chromosomes 12-19 decrease from ESC to NPC (Fig.27F). This can be clearly seen in the subtraction CT map of NPC-ESC. The interactions between smaller and larger chromosomes increase from ESC to NPC. Whereas from NPC to CN, as seen from the map, very subtle changes are seen both in smaller and larger chromosomes (Fig.27G). Statistically, there is no significant change in the interactions within smaller chromosomes and between smaller and larger chromosomes. But there is slight decrease in the interactions between larger chromosomes from NPC to CN.

These results suggest that CT neighborhood reorganizes during this differentiation. Smaller chromosomes cluster together. Most of the reorganization happens during ESC to NPC transition and very subtle changes are seen from NPC to CN.


A
ESC
NPC
CN
B
inter-chromosomal interactions(a.u)
Chr1-8
Chr9-19


C
NPC-ES
D
CN-NPC

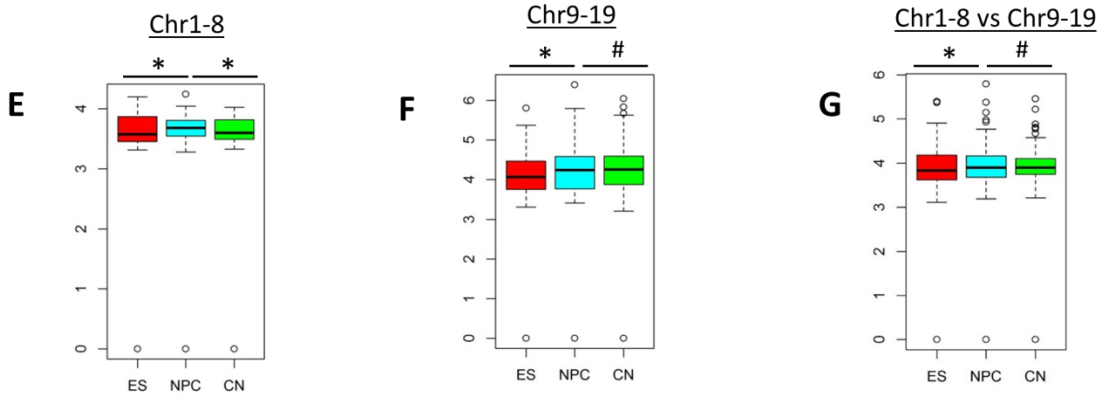

E
Chr1-8
F
Chr9-19
G
Chr1-8 vs Chr9-19
ES
NPC
CN

Fig 27. Chromosome Territory Neighborhood and it's dynamics during differentiation

A. Chromosome territory neighborhood maps of ESC, NPC and CN. Color gradient is from blue to red- Blue indicates lower interactions and red indicates higher interactions. Intra-chromosomal interactions were deleted and hence diagonals are blue. **B**. Interactions within larger chromosomes (Chr9-19) are lesser than the interactions within smaller chromosomes (Chr1-8) in all the three cell types. **C & D**. Subtraction CTN map of NPC-ESC(C) and CN-NPC(D). Red represents gain of interactions and blue represents loss of interactions. **E**. Comparison of Interactions among larger chromosomes across three cell types. **F**. Comparison of Interactions among smaller chromosomes across three cell types. **G.** Comparison of Interactions between larger and smaller chromosomes across the three cell types.

* represents p-value <0.05; # represents p-value >0.05

Dynamics of Intra-chromosomal Interactions

Then we wanted to see the dynamics of the intra-chromosomal interactions during neuronal differentiation. To observe that, we generated intra-chromosomal maps of Chr6 SimpleNorm matrices. As we can see, most of the interactions of the intra-chromosomal map are concentrated along diagonal and the interactions decrease as the distance increases between the bins i.e., off the diagonal (Fig.28A). Superficially, there seems no difference between the interaction maps of ESC, NPC and CN. To clearly see the differences, we generated subtraction matrices of NPC-ESC and CN-NPC.

In the subtraction matrices, from ESC to NPC, most of the short-range interactions (very close to diagonal) are increased and represented by red color along the diagonal. Mid-range interactions are decreased which is represented by green colour and then long-range interactions are increased (Fig.28B). These trends are similar in NPC to CN transition. Since, one cannot clearly see these dynamics with respect to genomic distances in the heatmaps, we sought to generate distance decay plots. The mean interaction frequency of the matrix as a function of genomic distance is plotted as a power law. To study the dynamics clearly, we plotted two decay plots: one within 8Mb distance, and another above 8Mb distances.

In the distance decay plots of below 8Mb distances, all the cell types i.e., ESC, NPC and CN follow similar trajectory as the distance increases. If we see further, the CN has the highest interactions and NPC has the lowest initially. After some distance, ESC will have the lowest interactions but CN continues with the highest interaction frequency up to about 7Mb. Then at the end, NPC has the highest frequency followed by CN and ESC (Fig.28C).

In the plot of above 8Mb, the trend starts with what has ended in below 8Mb i.e., NPC followed by CN and then ESC. At about 20 Mb, ESC makes a shift to the top and remains atop up to nearly 40 Mb. After 40 Mb, CN will have the highest frequency, followed by NPC and then ESC. So essentially, ESC has lot of interactions at the range of 20-40 Mb than NPC and CN. CN has the lowest frequency from around 10Mb to 40 Mb range (Fig.28D).

To further dissect the dynamics of the intra-chromosomal interactions, we divided the interaction distances into 6 ranges. i.e., <100kb, 100kb-500kb, 500kb-2Mb, 2Mb-20Mb, 20-40Mb, >40Mb. We calculated the mean interaction frequencies at all these ranges for all the chromosomes and the mean intra-chromosomal ranges have been determined to evaluate the general trend in all the chromosomes. We also divided these interactions based on whether these interactions are within cLADs or ciLADs (Fig.29A). As we observed previously, up to 500 kb, NPC has the highest interaction frequency followed by ESC and NPC has the lowest in ciLADs. But in cLADs, ESC has the highest frequency followed by NPC. From 500 kb to 20 Mb, ESC has the lowest interactions and NPC has the highest in both ciLADs and cLADs. From 20 Mb to 40 MB, ESC has the highest frequency and NPC has the lowest in both ciLADs and cLADs. But above 40Mb, the trend shows similar to that of 500kb to 20Mb in both ciLADs and cLADs , albeit with very low frequency. The initial lower interaction frequency of less than 500kb in both cLADs and ciLADs than the following ranges is due to the fact that these regions are so close that the interactions have not been captured in Hi-C in the first place.

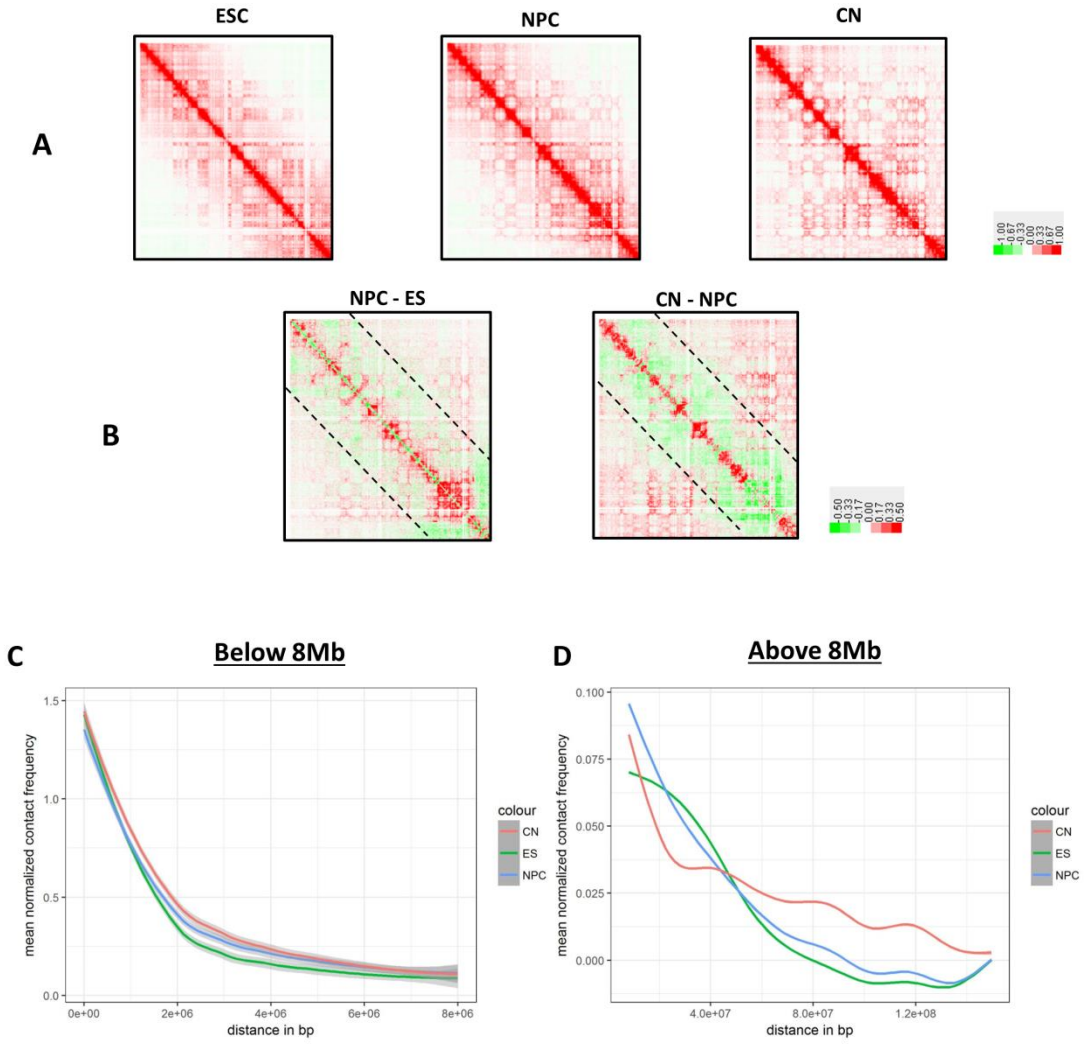


Fig 28. Dynamics of intra-chromosomal interactions during differentiation

A. SimpleNorm contact matrices of Chr6 showing intra-chromosomal interactions across ESC, NPC and CN. Interactions enriched along the diagonal shows that the proximal regions in the DNA have more interactions than with the distant regions. **B**. Subtraction matrices of intra-chromosomal interactions of NPC-ESC and CN-NPC. Red represents gain of interactions and green represents loss of interactions. The distance at which there is significant change in interactions is marked with dotted lines. **C**. Distance decay plot of interactions below 8Mb of ESC, NPC and CN. **D**. Distance decay plot of interactions above 8Mb of ESC, NPC and CN.

Dynamics of short-range interactions after CCDD:

Since the very short-range interactions don't represent the true proximity-based interaction values as we discussed previously, we corrected the intra-chromosomal matrices using CCDD for all the chromosomes and plotted the mean interaction frequencies. In cLADs, after CCDD, ESC has the highest frequency up to 500kb. But there seems no much difference between NPC and CN (Fig.29B). Whereas in ciLADs, after CCDD, although ESC has slightly larger frequency than NPC and CN, it seems that there is no difference in the frequencies between the cell types (Fig.29C). This suggests that during ESC to NPC transition, cLADs undergo reorganization at <500 kb distances but not ciLADs. Whereas during NPC to CN transition, there is no much reorganization either in cLADs or ciLADs.

Nature of genes and gene regulation in cLADs :

Lamina associated domains are often associated with repressive marks, gene poor and considered as heterochromatin. But we observed that cLADs in fact harbor many important genes and they tend to differentially regulate during differentiation of ESC to CN.

First to study the nature of the genes physically present in cLADs, we derived the list of the genes' coordinates which are within the boundaries of cLADs. We determined the Gene Ontology terms and pathways associated with these genes. Interestingly, many genes are involved in important pathways such as alcoholism, drug addiction etc. This suggests that although cLADs are gene poor, they harbor very important genes which are repressed in most of the cells.

Next, we wanted to see if any developmentally regulated genes involved in neuronal differentiation are present in cLADs. First, we derived the genes which have non-zero FPKM values and drew a venn diagram showing unique and commonly expressed genes in all the three cell types (Fig.30Ai). There are 937 LAD genes which are commonly expressed in all the three cell types. If uniquely expressed genes are concerned, ESC has the highest no. (553) of genes. This is in accordance with the fact that ESC genome is hyper-dynamic and genes present in cLADs are more accessible than in other types of cells. When we put a cutoff of FPKM >1, which is a standard way of looking at the expressing genes, we saw 340 LAD genes are commonly expressed in all the three cell types (Fig.30Aii). Then we wanted to see the expression levels of the genes present in cLADS in all the three cell types. When compared with the expression levels of all the genes from ciLADs and cLADs, the overall expression of genes in ciLADS is obviously lot higher than that of in clADs (Fig.30Aiii). But when the expression of only expressed genes is plotted, the values are comparable with that of the ciLAD genes (Fig.30Aiv). This suggests that cLADs harbor important regulatory genes whose expression is as high as genes in ciLADs.

Then we derived differentially regulated LAD genes between ESC, NPC and CN (see methods). During ESC to NPC, genes involved in GO terms such as membrane, Transmembrane which are crucial for neurogenesis are upregulated (Fig.30Bii) and during

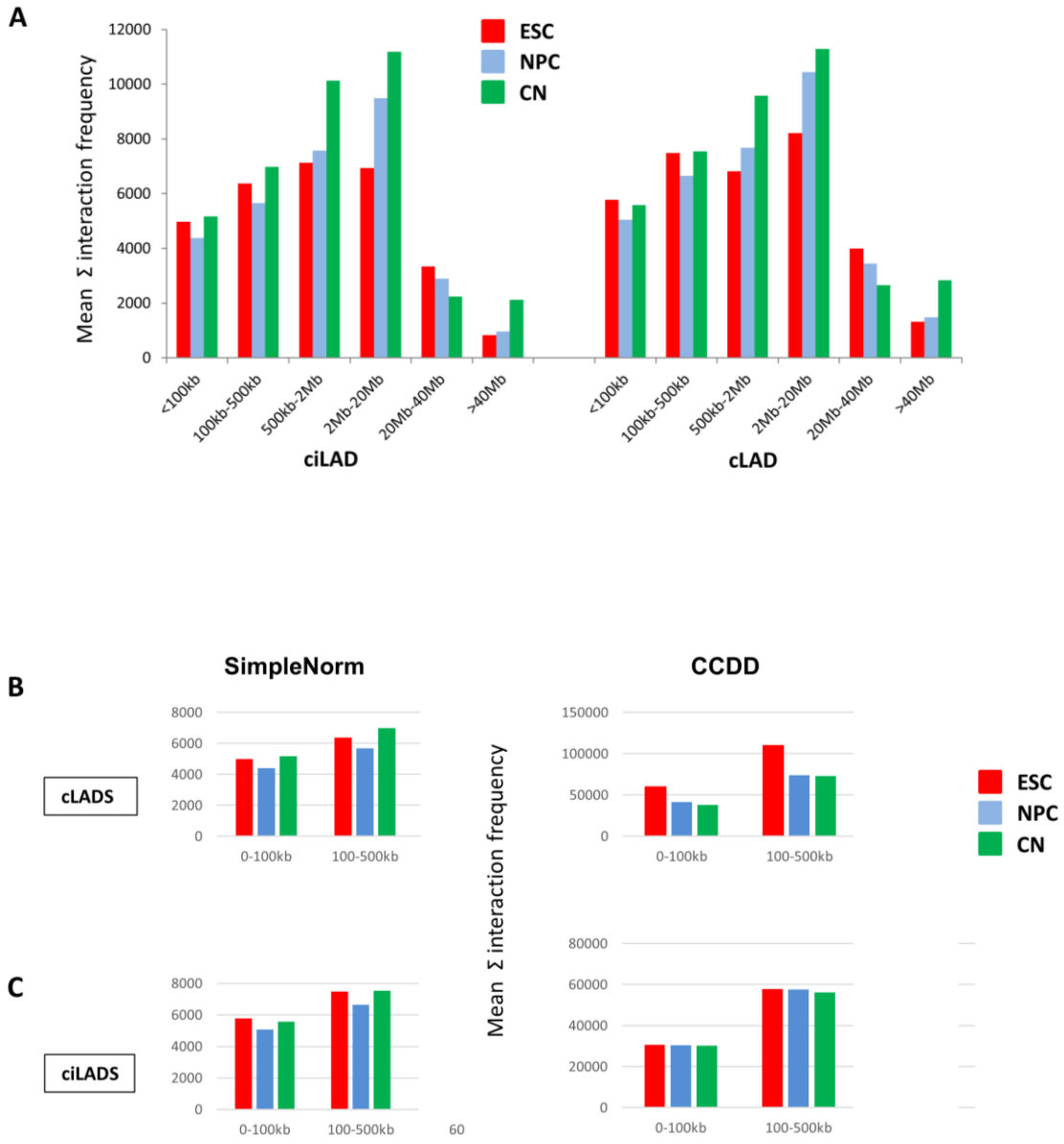


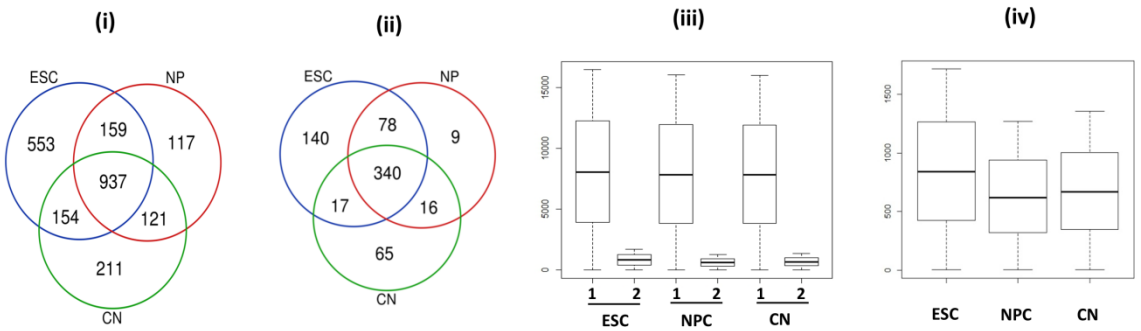
Fig 29. Dynamics of different distance ranges of interactions

A. Mean Σ of interaction frequencies at <100kb, 100kb-500kb, 500kb-2Mb, 2Mb-20Mb, 20-40Mb, >40Mb of the three cell types genome wide among ciLADs and cLADs. Note that interactions from both cLAD nd ciLADs show similar trends among various distance ranges. **B**. Mean Σ of interaction frequencies at 0-100kb and 100-500kb distances of three cell types before (SimpleNorm) and after correction among cLADs. **C**. Mean Σ of interaction frequencies at 0-100kb and 100-500kb distances of three cell types before (SimpleNorm) and after correction among ciLADs.

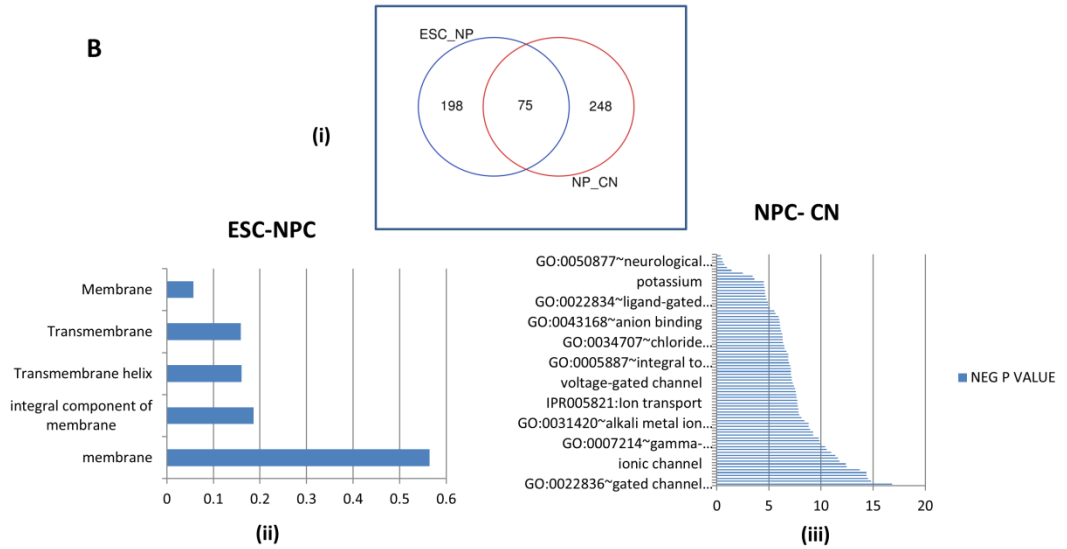
NPC to CN, genes with GO terms like ligand gated, anion binding, voltage gated channel are upregulated which are very important in neuronal formation (Fig.30Cii). Pathways downregulated in NPC from ESC include glycolysis, gluconeogenesis which are important for ESC maintenance (Fig.30Biii). From NPC to CN, genes involved in ribosomes, translation are downregulated (Fig.30Ciii) .

These results suggest that cLADs harbor many important developmentally regulated genes which are essential for differentiation of ESC into Neurons.

A


(i)
ESC
NP
CN
553
159
117
937
154
121
211
(ii)
ESC
NP
CN
140
78
9
340
17
16
65
(iii)
15000
10000
5000
0
1 2 1 2 1 2
ESC NPC CN
(iv)
1500
1000
500
0
ESC NPC CN

B


(i)
ESC_NP
198
75
248
NP_CN
ESC-NPC
Membrane
Transmembrane
Transmembrane helix
integral component of membrane
membrane
0 0.1 0.2 0.3 0.4 0.5 0.6
(ii)
NPC- CN
GO:0050877~neurological...
potassium
GO:0022834~ligand-gated...
GO:0043168~anion binding
GO:0034707~chloride...
GO:0005887~integral to...
voltage-gated channel
IPR005821:Ion transport
GO:0031420~alkali metal ion...
GO:0007214~gamma-...
ionic channel
GO:0022836~gated channel...
0 5 10 15 20
NEG P VALUE
(iii)

C

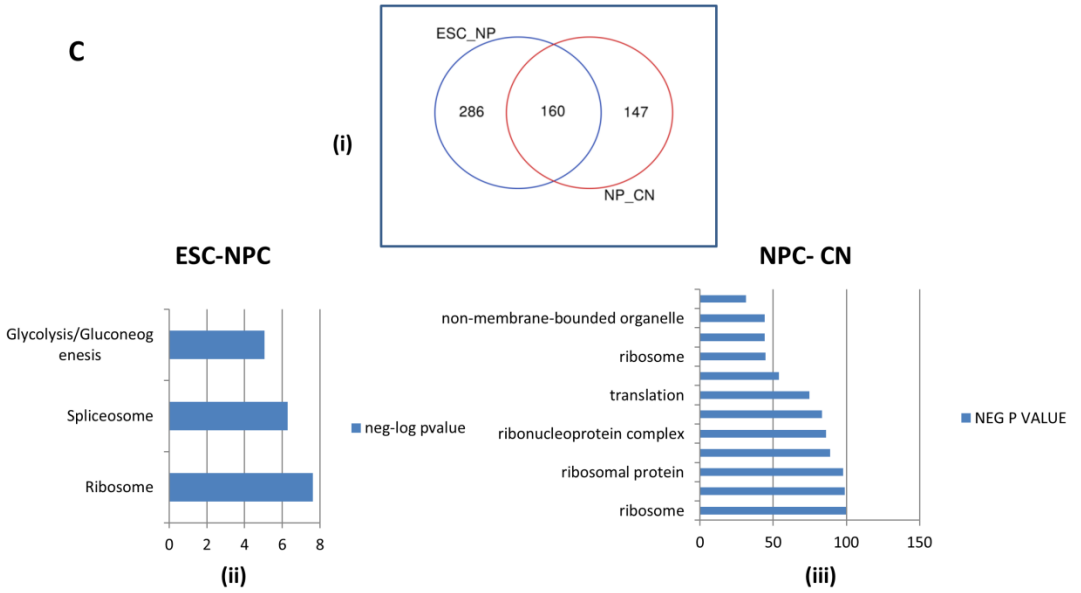

(i)
ESC_NP
286
160
147
NP_CN
ESC-NPC
Glycolysis/Gluconeogenesis
Spliceosome
Ribosome
0 2 4 6 8
neg-log pvalue
(ii)
NPC- CN
non-membrane-bounded organelle
ribosome
translation
ribonucleoprotein complex
ribosomal protein
ribosome
0 50 100 150
NEG P VALUE
(iii)

Fig 30. Nature of the genes present in the cLAD regions

A. **(i)** Venn diagram showing the no. of expressed genes (FPKM>0) from cLADs in ESC, NPC and CN. **(ii)** Venn diagram showing expressed genes (FPKM>1) from cLADs in ESC, NPC and CN. **(iii)** Boxplots showing expression levels of the genes present in ciLADs(1) and cLADs(2) in three cell types. **(iv)** Boxplot showing distribution of expressed genes from cLADs in the three cell types. **B. (i)** Venn diagram showing the no. of upregulated genes between ESC-NPC and NPC-CN in cLADs. **(ii)** Pathways upregulated during ESC-NPC in cLADs. **(iii)** Pathways upregulated during NPC-CN in cLADs. **C. (i)** Venn diagram showing the no. of downregulated genes between ESC-NPC and NPC-CN in cLADs. **(ii)** Pathways downregulated during ESC-NPC in cLADs. **(iii)** Pathways downregulated during NPC-CN in cLADs

4. Discussion

Objective 1: Development of robust method to measure and characterize chromatin condensation state genome-wide

Chromatin accessibility data has been very much useful in deriving and studying regulatory regions in the genome. The accessible genome accounts for nearly 2–3% of total DNA though it is able to capture more than 90% of sequences bound by transcription factors (Thurman, R. E. et al. 2012). DNase-I remained the preferred option to digest the open accessible chromatin due to its lower sequence specificity, albeit only few studies used restriction endonucleases to decipher the accessible regions in the chromatin (Chen et al., 2014; Ohkawa et al., 2012; Gargiulo et al., 2009]. Since there were lacunae in the methods traditionally used for deriving chromatin accessibility, we proposed the need for a more robust method. We proposed that Hi-C data, which is originally used for deriving the interactions between genomic elements, could be potentially repurposed to derive accessibility data and there by condensed and de-condensed regions in the genome. Since Hi-C uses restriction enzymes to cut the DNA at specific sites, lesser no. of sequenced reads are enough to analyze the accessibility. First, we observed that gene poor heterochromatin regions have lesser number of reads than gene-rich euchromatic regions in chromatin reads but not in naked DNA from the RED-seq experiment. Then we showed that 1Dimensional Hi-C reads representing heterochromatin and euchromatin visually as well as statistically. We used LOESS statistical Normalization to normalize the 1D reads to the GC content and Restriction enzyme site density. The highlight of the Normalization is that we corrected the 1D reads for machine bias as well, which is first of its kind. And for the first time, we were able to derive condensed and decondensed domains based on the normalized 1D Hi-C reads and also their correlation with LADs and ciLADs respectively.

Then we showed that RZ score derived Condensed and decondensed regions of chromatin are highly correlated with silent and active histone modifications respectively. Also, RZ score based measurement of chromatin condensation states accurately validates the condensed inactive X-chromosome vs decondensed active X-chromosome in brain and kidney female cells. We further authenticated our method in Polytene chromosome of Drosophila where the Polytene Bands have lesser reads compared to inter-Bands in both polytene and diploid chromosomes. Finally, we validated our method in dynamics of chromatin decondensation under Lamin KO Condition in ESC. Thus, our method to repurpose 1-dimensional in-situ Hi-C reads to derive condensed and decondensed domains is thoroughly validated. Before digesting the DNA with restriction enzyme, in-situ Hi-C protocol includes treatment of nuclei with 0.3–0.5% of sodium dodecyl sulphate (SDS) at 62 °C for nearly10 min, which is followed by quenching of SDS with Triton X-100 (Rao et al., 2014). The treatment with detergents and heat are expected enough to increase the accessibility of the chromatin. From our studies it is clear that regardless of these treatments, the accessibility of the condensed chromatin to restriction digestion is greatly constrained. Consequently, the interactions involving heterochromatin are under-represented in the available Hi-C datasets. Williamson et al., (2008) noted that there exists inconsistency between Hi-C data and the DNA FISH results regarding the decondensation and condensation of HoxD loci during differentiation of ESC. Although 5C or Hi-C mainly suggested that the HoxD locus remained in condensed state, the results of

DNA FISH showed that the locus instead got decondensed during differentiation. On the contrary, Kundu et al. has recently corroborated the 5C and DNA-FISH results on HoxA and HoxD locus. We infer that the differential digestion of decondensed and condensed forms of Hox loci by REs may underlie the inconsistencies.

Since the chromosome exists as a collation of condensed and decondensed regions, the traditional representation of chromosome as a mere chromatid may not hold true. Hence, we derived horizontally and vertically condensed chromosome maps using the 1D Hi-C reads which represent true in-vivo chromosome. Also, comparative CC maps of ESC and other differentiated cells reveal developmental regulation of chromatin condensation. As per our knowledge, we are the first to generate such condensation maps which truly represent the in vivo chromatin. These have profound implications in studying the chromatin organization and could also be potentially used as diagnostic markers.

Objective 2: Developmental dynamic role of chromatin condensation during neuronal differentiation

The neuronal differentiation of Embryonic stem cells seemed much more reliable/suitable to study the dynamics of chromatin condensation developmental. This three-point differentiation system is quite adequate to capture the dynamics of chromatin accessibility during differentiation. We cultured mouse ESC cell line to derive Neural Progenitors and then Cortical Neurons in a 21-day protocol. We validated the cell differentiation system through stage specific GFP markers and also gene expression through Real Time PCR. Then we took advantage of a published study from Cavalli's group from France, which has ultra-deep sequenced in-situ Hi-C and epigenetic datasets, to address our objectives. Although they studied Dynamics of CTCF insulation, promoters, TAD boundaries and gene expression along with other factors regulating genome architecture, they didn't study chromatin accessibility per se.

It was very striking that the ESC and NPC has more variation than NPC and CN in terms of chromatin accessibility, which suggests that most of the chromatin reorganization in terms of condensation and decondensation happens from ESC to NPC. This is in agreement with the fact that NPC and CN are lineage committed and ESC is still pluripotent. One could also infer from these results that the lineage commitment is accompanied by large scale changes in chromatin condensation. Switching of chromatin condensed states during differentiation is predominantly accompanied with corresponding changes in both gene expression and epigenetic marks. Thus, repurposed Hi-C datasets robustly deciphered the developmental dynamic role of chromatin condensation and decondensation during Neuronal differentiation. Interestingly, we found that non-LAD condensed regions derived using our method are associated with binding of polycomb group (PcG) of proteins. PcG proteins are often associated with H3K27me3 marks and also facultative heterochromatin (Bernstein et al., 2006). Thus, our method could be potentially used to identify facultative heterochromatin as well.

Though we showed that most of the changes in condensation are accompanied by corresponding change in the gene expression, a small percentage of genes have reverse correlation between condensation and gene expression. Interestingly, in top differentially regulated genes, we found that these include mostly miRNAs except in ESC-NPC downregulated genes. Thus, the chromatin with miRNA genes tends to condense during active transcription and Decondense during silencing. This peculiar behavior could be explained by the formation of transcriptional condensates through liquid-phase separation phenomenon. It has been shown that the retention of pri- microRNA transcripts at transcription sites resulted in enhanced production of microRNA (Pawlicki et al.,2008). Thus, the presence of pri-miRNA transcripts at the transcription site may have minimized the ability of restriction enzymes to cleave the DNA, which might have resulted in lesser number of Hi-C reads. Since there are little to no studies reporting this observed phenomenon, further experimental validation and investigation is required for understanding the mechanisms.

Since we found some top genes with negative correlation between gene expression and condensation, we extended the work to see globally. Genes showing negative correlation could

be subcategorized into different groups based on histone marks and mechanistic principles associated with them. Formation of transcriptional condensates, presence of imprinted genes, establishment of facultative heterochromatin and presence of bivalent genes could rationally provide explanations for the observations seen in these small percentage of genes. Further investigations are needed to gain insights into the mechanistic details of the observations.

We also observed for the first time that ESC to NPC differentiation is accompanied by larger decondensation in cLADS and condensation in ciLADs in a chromosome wide manner. This pattern just reverses from NPC to CN with lesser intensity. This alternative reciprocal behavior might have a role in gene regulation.

Objective 3: Dynamics of quantitative spatial genome architecture during neuronal differentiation

The key aspect of this objective was to address a previously unaddressed problem in the Hi-C analysis which arises due to the difference in the visibility of the condensed and decondensed regions. From the accessibility data, we could clearly see from the example that the part of the genome around Nanog locus is condensed from ESC to NPC and further to CN, and is supported by epigenetic data as well. But the subtraction matrices show otherwise, indicating that Hi-C contacts are lost up to certain length. We also showed that this was independent of normalization method. Calculating this threshold became very important in normalizing these short-range contacts. Hence, we employed a power law based technique, where the range at which the curve gets flattened indicates the threshold upto which these short range contacts are showing otherwise. To retain the mean, minimum and maximum values in the belt, we designed a formula and named the method as CCDD. This correction was also applied for other loci and got similar results proving the overall robustness of the technique. The technique could be further improved by correcting the reads non-linearly using other statistical methods. This is for the first time one has addressed the accessibility problem in Hi-C reads and has made an attempt to correct. Many further studies and analyses are needed to completely address the accessibility problem.

One of the applications of the accessibility problem corrected matrices is in the 3D modeling of chromosomes in silico. Without considering this problem, the 3D modeling of the chromosome would give false interpretations due to changes in the short -range interactions. We also proposed a novel Heteropolymer model in which the size of the bead depends on the accessibility of that particular genomic bin. Together, we could model the corrected matrices using heteropolymer model to obtain a closer-to-true picture of the chromosomes, which also captured the previously mentioned loci. Our study also suggests that one should always use hetero-polymer models (as opposed to popular homopolymer models in literature) to obtain accurate predictions of packaging and compaction.

Then we analyzed the remodeling of chromatin architecture during Neuronal differentiation using our CCDD corrected Hi-C matrices. First, we derived chromosome territory neighborhood maps and could see that smaller chromosomes are more clustered than the larger chromosomes, as shown in previous studies (Lieberman et al., 2009). Subtraction CTN maps clearly showed that a lot of reorganization happens from ESC to NPC and only subtle changes happen from NPC to CN in terms of chromosome neighborhood. This indicates that most of the chromosomal rearrangement in terms of neighborhood happens during transition from ESC to the next cellular state and least happens at the terminal differentiation. This is also in accordance with our other studies (not yet published) from our Lab. Moreover, we were able to show that interactions among and between both small and large chromosomes increases from ESC to NPC, whereas interactions among larger chromosomes decrease in CN. Then using subtraction intra-chromosomal maps, we showed that NPC and CN show gain of long-range interactions. This is important because while most ESC are mitotically active, their chromosomes lack long-range interactions, which are then gained during differentiation. We further dissected the interactions at various ranges and could find some important trends,

especially interactions at 20-40Mb range between them are lost during the process of neuronal differentiation. This might be unique to neuronal differentiation, as we observed a different pattern during other differentiation system (results yet to be published). Lastly, we could show the difference between trends in cLADs and ciLADs using our corrected matrices.

We hope that the results and observations from our study contribute significantly towards understanding the underlying principles of higher genome organization and trigger many future studies.

5. Summary

Objective 1:

We derived 1D read counts of Hi-C data and overlaid them with DNase-seq data and in situ chromatin of RED-seq and could clearly see that euchromatin regions showed higher no. of reads than heterochromatin regions, which gives an initial representation that Hi-C reads could be re-purposed to derive condensed and de-condensed domains. The raw Hi-C reads should be corrected to systemic biases. To this end, we used LOESS regression to first standardize the loess parameters to normalize for the machine bias and then to correct GC content and RE site density and then z-scored them to obtain RZ scores. The positive RZ scores corresponded to euchromatin and negative RZ to heterochromatin. Then quantitatively, we showed that the cLADs have significantly lesser corrected reads than ciLADs in in situ chromatin but not in naked DNA. Then we calculated the correlation between the 1D reads of in situ chromatin and in situ HI-C reads which showed a high correlation implicating that similar results could be observed in *in situ* Hi-C. In fact, we showed that corrected Hi-C reads are more represented in ciLADS than in cLADS. Then we derived condensed and decondensed regions based on RZ scores with FDR <0.05 and could show that the condensed regions derived using our method matched well with known cLADs. To characterize the derived domains, we assessed the histone marks associated with the domains and could show that active marks such as H3K4me1, H3K27ac etc are enriched in decondensed domains and repressive marks such as H3K9me3 are enriched in condensed domains. To validate our method, we analyzed allele specific Hi-C data of mouse female X-chromosome and could show that condensed inactive X-chromosome has lesser number of reads than that of active X-chromosome. To further scrutinize our method, we analyzed the Hi-C data of Drosophila polytene chromosome and showed that band regions show significantly lower reads than i-bands both in polytene and diploid chromosomes. Then we analyzed Hi-C data of lamin B knocked out ESC and could show that that no. of reads coming from KO cLADS is higher than that of wild type, whereas the rest of the regions didn't get affected. These studies unequivocally demonstrated that Hi-C reads can be efficiently repurposed to derive condensed and decondensed regions genome wide. Next, we derived chromosome condensation maps based on RZ scores before Z scoring. We derived both horizontally and vertically compressed chromosome condensation maps. These maps truly represent the chromosomes in their in vivo condition. To see the dynamics of these maps during differentiation, we generated the condensation maps from ESC, fetal liver and adult liver. We also generated these maps using Hi-C data of neural stem cells and Astrocytes. Finally, we created an ensemble of horizontally compressed maps of all chromosomes in the above cell types.

Objective 2:

We cultured ESC and differentiated them into Neural Progenitors and Cortical Neurons. Then we characterized and validated the cell types using RT-PCR and qPCR. While we were working on this, Cavalli group from France published a work (Bonev et al., 2017) on the same differentiation system with ultra-deep sequenced Hi-C data along with gene expression and many ChIP-seq datasets. Though they studied different aspects of TADs and other important

factors, none of our objectives were overlapping with their work. Hence, we reanalyzed their datasets to address our objectives. First, we showed that the accessibility bias existed in all the three cell types. As we see higher variation between ESC and NPC than between NPC and CN, we restricted most of the further studies in this objective to ESC-NPC. We showed that during ESC-NPC transition, most of the condensed and decondensed regions are conserved. The switching regions exhibited corresponding change in the gene expression. Next, we showed that condensation in NPC is associated with downregulation of pluripotent pathways and decondensation is associated with upregulation of neuronal pathways. Then we analyzed epigenetic marks associated with the constitutive and switching regions from ESC-NPC and showed that while constitutive regions retained the corresponding enrichment, the switching regions were accompanied by corresponding change in the enrichment of the histone marks. We also showed that the condensed non-LAD regions are enriched in polycomb group of proteins. When we analyzed relationship between top differentially regulated genes and condensation, we could observe that some genes showed no relation between gene expression and condensation. Upon further study, we found that miRNA genes showed reverse correlation between condensation and gene expression during differentiation except for ESC-NPC downregulated ones. We also found that a small percentage of genes globally tend to show this reverse correlation. To study those genes, we further classified them based on histone marks and could derive pathways associated with such genes. Finally, we wanted to see the dynamics of condensation at whole chromosome level. To this end, we made differential (subtraction) RZ score plots of NPC-ESC and CN-NPC. We observed that differentiation of ESC to NPC is accompanied by loss of reads in cLADS and gain in ciLADs i.e., the cLADs condense from ESC-NP and then decondense from NPC-CN. Whereas ciLADS decondense from ESC-NPC and condense from NPC-CN globally.

Objective 3:

3a: We first tested whether this accessibility problem is intrinsic to the normalization method used to normalize the contact matrix and showed that this problem exists in all major known normalization methods. We also showed that this problem is not specific to Nanog domain either. We observed that, this discrepancy is seen to a certain length off the diagonal which needs to be corrected. We employed a novel CCDD (contact correction through distance decay plots) method to first define the threshold of the inverse values and then correct the values using novel formula. Then we showed that the RE accessibility problem is corrected in Nanog and Oct4 domains. Then we wanted to use these corrected contact matrices in 3D modelling of the chromosome using HI-C data. We used brownian simulations to generate 3D polymer models of chromosome. Instead of using usual homopolymer model, we used a novel heteropolymer model where the bead size is dependent on the no. of reads coming from that particular bin, which essentially means condensed chromatin have smaller bead sizes whereas decondensed chromatin have larger bead sizes. We simulated Chr6 and Chr17 and showed dynamics of the Rg of Nanog and Oct4 domains respectively.

3b: We derived chromosome territory neighborhood maps from whole genome contact matrices of Hi-C data and could see that smaller chromosomes are more clustered than the larger chromosomes, as shown in previous studies (Lieberman et al., 2009). Subtraction CTN maps clearly showed that a lot of reorganization happens from ESC to NPC and only subtle changes happen from NPC to CN in terms of CTN. Then we showed that interactions among and between both small and large chromosomes increases from ESC to NPC, whereas interactions among larger chromosomes decrease in CN. Next, we wanted to study the dynamics of the intra-chromosomal interactions. By generating cis subtraction matrices, we could show that from ESC to NPC, most of the short-range interactions (very close to diagonal) are increased along the diagonal, mid-range interactions are decreased and then long-range interactions are increased. We generated two power law plots for each chromosome and observed that interactions are dynamic at various distances above 8Mb. To study the nature of interactions, we further dissected them into various distance ranges and could find some important trends, especially interactions at 20-40Mb range between them are lost during the process of neuronal differentiation. Then we plotted interactions below 500kb which were corrected using CCDD and showed that there exist differences between interactions of cLADs and ciLADs, and other trends. At the end, we analyzed the nature of genes present in the cLADs and showed that cLADs harbor many important developmentally regulated genes which are differentially regulated during the course of differentiation and some genes are expressed as high as the genes in ciLADs.

Table 1

Accession	Cell-type	Experiment	RE	Processing
GSE51821	mESC	RED-seq	Sau96I	Pre-processed
GSE96107	mESC, NPC, CN	In-situ Hi-C	DpnII	HiCUP/bowtie
GSE59027	mESC, NPC, CN	In-solution Hi-C	NcoI	HiCUP/bowtie
GSE72510	Polytene	Tethered Hi-C	DpnII	HiCUP/bowtie
GSE89520	mESC (laminKO)	In-situ Hi-C	BgIII	Pre-processed
GSE68992	Brain, Patski	DNase-Hi-C	DNase	Pre-processed
ENCSR032JUI	mESC	H3K4me1	ChIP-seq	Pre-processed
ENCSR000CGQ		H3K27ac	ChIP-seq	Pre-processed
ENCSR000CGO		H3K4me3	ChIP-seq	Pre-processed
ENCSR253QPK		H3K36me3	ChIP-seq	Pre-processed
ENCSR857MYS		H3K9me3	ChIP-seq	Pre-processed
ENCSR059MBO		H3K27me3	ChIP-seq	Pre-processed
GSE96107	NPC	H3K4me1	ChIP-seq	Pre-processed
		H3K27ac	ChIP-seq	Pre-processed
		H3K4me3	ChIP-seq	Pre-processed
		H3K36me3	ChIP-seq	Pre-processed
		H3K9me3	ChIP-seq	Pre-processed
		H3K27me3	ChIP-seq	Pre-processed
	CN	H3K4me1	ChIP-seq	Pre-processed
		H3K27ac	ChIP-seq	Pre-processed
		H3K4me3	ChIP-seq	Pre-processed
		H3K36me3	ChIP-seq	Pre-processed
		H3K9me3	ChIP-seq	Pre-processed
		H3K27me3	ChIP-seq	Pre-processed

Table 1. Details of the datasets used in the study.

References

1. Gilbert, S. F. (2000). Developmental biology. Sunderland, Mass: Sinauer Associates.
2. Hoopes, L. (2008) Introduction to the gene expression and regulation topic room. Nature Education 1(1):160
3. Cooper, Geoffrey M. The Cell: A Molecular Approach. 2nd Edition. : Sinauer Associates, 2000
4. El-Osta, A. S. S. A. M., and Alan P. Wolffe. "DNA methylation and histone deacetylation in the control of gene expression: basic biochemistry to human development and disease." Gene Expression The Journal of Liver Research 9.1-2 (2001): 63-75.
5. Gorkin, David U., Danny Leung, and Bing Ren. "The 3D genome in transcriptional regulation and pluripotency." Cell stem cell 14.6 (2014): 762-775.
6. Zheng, Hui, and Wei Xie. "The role of 3D genome organization in development and cell differentiation." Nature Reviews Molecular Cell Biology 20.9 (2019): 535-550.
7. Babu, Deepak, and Melissa J. Fullwood. "3D genome organization in health and disease: emerging opportunities in cancer translational medicine." Nucleus 6.5 (2015): 382-393.
8. https://www.nature.com/scitable/ebooks/essentials-of-cell-biology-14749010/118237915/
9. Pederson, Thoru. "The plurifunctional nucleolus." Nucleic acids research 26.17 (1998): 3871-3876.
10. Boisvert, François-Michel, et al. "The multifunctional nucleolus." Nature reviews Molecular cell biology 8.7 (2007): 574-585.
11. Peric-Hupkes, D., and B. van Steensel. "Role of the nuclear lamina in genome organization and gene expression." Cold Spring Harbor symposia on quantitative biology. Vol. 75. Cold Spring Harbor Laboratory Press, 2010.
12. Jagannathan, Madhav, Ryan Cummings, and Yukiko M. Yamashita. "A conserved function for pericentromeric satellite DNA." Elife 7 (2018): e34122.
13. Jagannathan, Madhav, Ryan Cummings, and Yukiko M. Yamashita. "The modular mechanism of chromocenter formation in Drosophila." Elife 8 (2019): e43938.
14. Ohta, Shinya, et al. "Building mitotic chromosomes." Current opinion in cell biology 23.1 (2011): 114-121.
15. Cremer, Thomas, and Marion Cremer. "Chromosome territories." Cold Spring Harbor perspectives in biology 2.3 (2010): a003889.
16. Cremer, Thomas, and Christoph Cremer. "Chromosome territories, nuclear architecture and gene regulation in mammalian cells." Nature reviews genetics 2.4 (2001): 292-301.
17. Peric-Hupkes, Daan, et al. "Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation." Molecular cell 38.4 (2010): 603-613.
18. Gavrilov, A. A., and S. V. Razin. "Compartmentalization of the cell nucleus and spatial organization of the genome." Molecular Biology 49.1 (2015): 21-39.
19. Spector, David L. "Nuclear domains." Journal of cell science 114.16 (2001): 2891-2893.
20. Lesne, Annick, et al. "Exploring mammalian genome within phase-separated nuclear bodies: experimental methods and implications for gene expression." Genes 10.12 (2019): 1049.

21. Cho, Won-Ki, et al. "Mediator and RNA polymerase II clusters associate in transcription-dependent condensates." Science 361.6400 (2018): 412-415.
22. Robertson, Keith D. "DNA methylation and human disease." Nature Reviews Genetics 6.8 (2005): 597-610.
23. Law, Julie A., and Steven E. Jacobsen. "Establishing, maintaining and modifying DNA methylation patterns in plants and animals." Nature Reviews Genetics 11.3 (2010): 204-220.
24. Ji, Hong, et al. "Comprehensive methylome map of lineage commitment from haematopoietic progenitors." Nature 467.7313 (2010): 338-342.
25. Tamaru, Hisashi. "Confining euchromatin/heterochromatin territory: jumonji crosses the line." Genes & development 24.14 (2010): 1465-1478.
26. Zeng, Weihua, Alexander R. Ball Jr, and Kyoko Yokomori. "HP1: heterochromatin binding proteins working the genome." Epigenetics 5.4 (2010): 287-292.
27. Cao, Ru, Yu-ichi Tsukada, and Yi Zhang. "Role of Bmi-1 and Ring1A in H2A ubiquitylation and Hox gene silencing." Molecular cell 20.6 (2005): 845-854.
28. Cao, Ru, et al. "Role of histone H3 lysine 27 methylation in Polycomb-group silencing." Science 298.5595 (2002): 1039-1043.
29. Wang, Wei, et al. "Polycomb group (PcG) proteins and human cancers: multifaceted functions and therapeutic implications." Medicinal research reviews 35.6 (2015): 1220-1267.
30. Nakayama, Jun-ichi, et al. "Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly." Science 292.5514 (2001): 110-113.
31. Martienssen, Robert, and Danesh Moazed. "RNAi and heterochromatin assembly." Cold Spring Harbor perspectives in biology 7.8 (2015): a019323.
32. Di Croce, Luciano, and Kristian Helin. "Transcriptional regulation by Polycomb group proteins." Nature structural & molecular biology 20.10 (2013): 1147.
33. Heintzman, Nathaniel D., et al. "Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome." *Nature genetics* 39.3 (2007): 311-318.
34. Calo, Eliezer, and Joanna Wysocka. "Modification of enhancer chromatin: what, how, and why?." Molecular cell 49.5 (2013): 825-837.
35. Ong, Chin-Tong, and Victor G. Corces. "Enhancer function: new insights into the regulation of tissue-specific gene expression." Nature Reviews Genetics 12.4 (2011): 283-293.
36. Luco, Reini F., et al. "Regulation of alternative splicing by histone modifications." Science 327.5968 (2010): 996-1000.
37. Guelen, Lars, et al. "Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions." Nature 453.7197 (2008): 948-951.
38. Adam, Stephen A., and Robert D. Goldman. "Insights into the differences between the A-and B-type nuclear lamins." Advances in biological regulation 52.1 (2012): 108.
39. Kind, Jop, et al. "Single-cell dynamics of genome-nuclear lamina interactions." Cell 153.1 (2013): 178-192.
40. Olins, Donald E., and Ada L. Olins. "Chromatin history: our view from the bridge." Nature reviews Molecular cell biology 4.10 (2003): 809-814.

41. Annunziato, A. "DNA packaging: nucleosomes and chromatin." Nature education 1.1 (2008): 26.
42. Woodcock, C. L., L-LY Frado, and J. B. Rattner. "The higher-order structure of chromatin: evidence for a helical ribbon arrangement." The Journal of cell biology 99.1 (1984): 42-52.
43. Dorigo, Benedetta, et al. "Nucleosome arrays reveal the two-start organization of the chromatin fiber." Science 306.5701 (2004): 1571-1573.
44. Lieberman-Aiden, Erez, et al. "Comprehensive mapping of long-range interactions reveals folding principles of the human genome." science 326.5950 (2009): 289-293.
45. Luger, Karolin, Mekonnen L. Dechassa, and David J. Tremethick. "New insights into nucleosome and chromatin structure: an ordered state or a disordered affair?." Nature reviews Molecular cell biology 13.7 (2012): 436-447.
46. Rao, Suhas SP, et al. "A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping." *Cell* 159.7 (2014): 1665-1680.
47. Dixon, Jesse R., et al. "Topological domains in mammalian genomes identified by analysis of chromatin interactions." *Nature* 485.7398 (2012): 376-380.
48. Nora, Elphège P., et al. "Spatial partitioning of the regulatory landscape of the X-inactivation centre." *Nature* 485.7398 (2012): 381-385.
49. Sexton, Tom, et al. "Three-dimensional folding and functional organization principles of the Drosophila genome." *Cell* 148.3 (2012): 458-472.
50. Dekker, Job, and Edith Heard. "Structural and functional diversity of Topologically Associating Domains." FEBS letters 589.20 (2015): 2877-2884.
51. Dixon, Jesse R., et al. "Chromatin architecture reorganization during stem cell differentiation." Nature 518.7539 (2015): 331-336.
52. Dixon, Jesse R., David U. Gorkin, and Bing Ren. "Chromatin domains: the unit of chromosome organization." Molecular cell 62.5 (2016): 668-680.

53. Zhan, Yinxiu, et al. "Reciprocal insulation analysis of Hi-C data shows that TADs represent a functionally but not structurally privileged scale in the hierarchical folding of chromosomes." Genome research 27.3 (2017): 479-490.
54. Nora, Elphège P., et al. "Targeted degradation of CTCF decouples local insulation of chromosome domains from genomic compartmentalization." Cell 169.5 (2017): 930-944.
55. Fudenberg, Geoffrey, et al. "Formation of chromosomal domains by loop extrusion." Cell reports 15.9 (2016): 2038-2049.
56. Ganji, Mahipal, et al. "Real-time imaging of DNA loop extrusion by condensin." Science 360.6384 (2018): 102-105.
57. Harewood, Louise, et al. "Hi-C as a tool for precise detection and characterisation of chromosomal rearrangements and copy number variation in human tumours." Genome biology 18.1 (2017): 1-11.
58. Phillips-Cremins, Jennifer E., and Victor G. Corces. "Chromatin insulators: linking genome organization to cellular function." Molecular cell 50.4 (2013): 461-474.
59. Radman-Livaja, Marta, and Oliver J. Rando. "Nucleosome positioning: how is it established, and why does it matter?." Developmental biology 339.2 (2010): 258-266.

60. John, Sam, et al. "Chromatin accessibility pre-determines glucocorticoid receptor binding patterns." Nature genetics 43.3 (2011): 264-268.
61. Duggan, N. M. & Tang, Z. I. (2010) The Formation of Heterochromatin. Nature Education 3(9):5
62. Tsompana, Maria, and Michael J. Buck. "Chromatin accessibility: a window into the genome." Epigenetics & chromatin 7.1 (2014): 1-16.
63. Thurman, Robert E., et al. "The accessible chromatin landscape of the human genome." Nature 489.7414 (2012): 75-82.
64. Chen, Poshen B., et al. "Unbiased chromatin accessibility profiling by RED-seq uncovers unique features of nucleosome variants in vivo." BMC genomics 15.1 (2014): 1-18.
65. Gaspar-Maia, Alexandre, et al. "Chd1 regulates open chromatin and pluripotency of embryonic stem cells." Nature 460.7257 (2009): 863-868.
66. Hargreaves, Diana C., and Gerald R. Crabtree. "ATP-dependent chromatin remodeling: genetics, genomics and mechanisms." Cell research 21.3 (2011): 396-420.
67. Schwartzentruber, Jeremy, et al. "Driver mutations in histone H3. 3 and chromatin remodelling genes in paediatric glioblastoma." Nature 482.7384 (2012): 226-231.
68. Hnisz, Denes, et al. "Activation of proto-oncogenes by disruption of chromosome neighborhoods." Science 351.6280 (2016): 1454-1458.
69. Anania, Chiara, and Darío G. Lupiáñez. "Order and disorder: abnormal 3D chromatin organization in human disease." Briefings in functional genomics 19.2 (2020): 128-138.
70. Norton, Heidi K., and Jennifer E. Phillips-Cremins. "Crossed wires: 3D genome misfolding in human disease." Journal of Cell Biology 216.11 (2017): 3441-3452.
71. Bernstein, B. E., et al. "Consortium EP. An integrated encyclopedia of DNA elements in the human genome." Nature 489.7414 (2012): 57-74.
72. Allan, James, et al. "Regulation of the higher-order structure of chromatin by histones H1 and H5." The Journal of cell biology 90.2 (1981): 279-288.
73. Caplan, Avrom, et al. "Perturbation of chromatin structure in the region of the adult beta-globin gene in chicken erythrocyte chromatin." Journal of molecular biology 193.1 (1987): 57-69.
74. Gilbert, Nick, and James Allan. "Distinctive higher-order chromatin structure at mammalian centromeres." Proceedings of the National Academy of Sciences 98.21 (2001): 11949-11954.
75. Keene, Michael A., and Sarah CR Elgin. "Micrococcal nuclease as a probe of DNA sequence organization and chromatin structure." Cell 27.1 (1981): 57-64.
76. Levy, Abraham, and Markus Noll. "Chromatin fine structure of active and repressed genes." Nature 289.5794 (1981): 198-203.
77. Wu, Carl. "The 5′ ends of Drosophila heat shock genes in chromatin are hypersensitive to DNase I." Nature 286.5776 (1980): 854-860.
78. Boyle, Alan P., et al. "High-resolution mapping and characterization of open chromatin across the genome." Cell 132.2 (2008): 311-322.
79. Giresi, Paul G., et al. "FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin." Genome research 17.6 (2007): 877-885.

80. Waki, Hironori, et al. "Global mapping of cell type–specific open chromatin by FAIRE-seq reveals the regulatory role of the NFI family in adipocyte differentiation." PLoS Genet 7.10 (2011): e1002311.
81. Goryshin, Igor Yu, and William S. Reznikoff. "Tn5 in vitro transposition." Journal of Biological Chemistry 273.13 (1998): 7367-7374.
82. Adey, Andrew, et al. "Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density in vitro transposition." Genome biology 11.12 (2010): 1-17.
83. Buenrostro, Jason D., et al. "Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position." Nature methods 10.12 (2013): 1213.
84. Liberator, Paul A., and J. B. Lingrel. "Restriction endonuclease accessibility of the developmentally regulated goat gamma-, beta C-, and beta A-globin genes in chromatin. Differences in 5'regions which show unusually high sequence homology." Journal of Biological Chemistry 259.24 (1984): 15497-15501.
85. Almer, A., and W. Hörz. "Nuclease hypersensitive regions with adjacent positioned nucleosomes mark the gene boundaries of the PHO5/PHO3 locus in yeast." *The EMBO journal* 5.10 (1986): 2681-2687.
86. Ohkawa, Yasuyuki, Concetta GA Marfella, and Anthony N. Imbalzano. "Skeletal muscle specification by myogenin and Mef2D via the SWI/SNF ATPase Brg1." *The EMBO journal* 25.3 (2006): 490-501.
87. Margueron, Raphaël, and Danny Reinberg. "Chromatin structure and the inheritance of epigenetic information." Nature Reviews Genetics 11.4 (2010): 285-296.
88. Mitchison, T. J., & Salmon, E. D. Mitosis: A history of division. Nature Cell Biology 3, E17–E21 (2001) doi:10.1038/35050656
89. Palstra, Robert-Jan, et al. "The β-globin nuclear compartment in development and erythroid differentiation." Nature genetics 35.2 (2003): 190-194.
90. Krivega, Ivan, and Ann Dean. "Enhancer and promoter interactions—long distance calls." Current opinion in genetics & development 22.2 (2012): 79-85.
91. Krivega, Ivan, Ryan K. Dale, and Ann Dean. "Role of LDB1 in the transition from chromatin looping to transcription activation." Genes & development 28.12 (2014): 1278-1290.
92. Guo, Chunguang, et al. "CTCF-binding elements mediate control of V (D) J recombination." Nature 477.7365 (2011): 424-430.
93. Dixon, Jesse R., et al. "Chromatin architecture reorganization during stem cell differentiation." Nature 518.7539 (2015): 331-336.
94. Bonev, Boyan, et al. "Multiscale 3D genome rewiring during mouse neural development." Cell 171.3 (2017): 557-572.
95. Fraser, James, et al. "Hierarchical folding and reorganization of chromosomes are linked to transcriptional changes in cellular differentiation." Molecular systems biology 11.12 (2015): 852.
96. Smallwood, Andrea, and Bing Ren. "Genome organization and long-range regulation of gene expression by enhancers." Current opinion in cell biology 25.3 (2013): 387-394.
97. Gaspard, Nicolas, et al. "An intrinsic mechanism of corticogenesis from embryonic stem cells." Nature 455.7211 (2008): 351-357.

98. Gaspard, Nicolas, et al. "Generation of cortical neurons from mouse embryonic stem cells." Nature protocols 4.10 (2009): 1454-1463.
99. Edgar, Ron, Michael Domrachev, and Alex E. Lash. "Gene Expression Omnibus: NCBI gene expression and hybridization array data repository." Nucleic acids research 30.1 (2002): 207-210.
100. Andrey, Guillaume, et al. "A switch between topological domains underlies HoxD genes collinearity in mouse limbs." Science 340.6137 (2013).
101. Cai, Chunyu, and Laura Grabel. "Directing the differentiation of embryonic stem cells to neural stem cells." Developmental dynamics: an official publication of the American Association of Anatomists 236.12 (2007): 3255-3266.
102. Edelmann, Peter, et al. "Morphology and dynamics of chromosome territories in living cells." Biochimica et Biophysica Acta (BBA)-Reviews on Cancer 1551.1 (2001): M29-M39.
103. Dekker, Job, et al. "Capturing chromosome conformation." science 295.5558 (2002): 1306-1311.
104. Simonis, Marieke, et al. "Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture–on-chip (4C)." Nature genetics 38.11 (2006): 1348-1354.
105. Dostie, Josée, et al. "Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements." Genome research 16.10 (2006): 1299-1309.
106. Li, Guoliang, et al. "ChIA-PET tool for comprehensive chromatin interaction analysis with paired-end tag sequencing." Genome biology 11.2 (2010): 1-13.
107. Nagano, Takashi, et al. "Single-cell Hi-C reveals cell-to-cell variability in chromosome structure." Nature 502.7469 (2013): 59-64.
108. Oluwadare, Oluwatosin, Max Highsmith, and Jianlin Cheng. "An overview of methods for reconstructing 3-D chromosome and genome structures from Hi-C data." Biological procedures online 21.1 (2019): 1-20.
109. Imakaev, Maxim, et al. "Iterative correction of Hi-C data reveals hallmarks of chromosome organization." Nature methods 9.10 (2012): 999-1003.
110. Bolger, Anthony M., Marc Lohse, and Bjoern Usadel. "Trimmomatic: a flexible trimmer for Illumina sequence data." Bioinformatics 30.15 (2014): 2114-2120.
111. Servant, Nicolas, et al. "HiC-Pro: an optimized and flexible pipeline for Hi-C data processing." Genome biology 16.1 (2015): 1-11.
112. Castellano, Giancarlo, et al. "Hi-Cpipe: a pipeline for high-throughput chromosome capture." (2015).
113. Wingett, Steven, et al. "HiCUP: pipeline for mapping and processing Hi-C data." F1000Research 4 (2015).
114. Yaffe, Eitan, and Amos Tanay. "Probabilistic modeling of Hi-C contact maps eliminates systematic biases to characterize global chromosomal architecture." Nature genetics 43.11 (2011): 1059.
115. https://www.ncbi.nlm.nih.gov/sra/docs/toolkitsoft/
116. http://bowtie-bio.sourceforge.net/bowtie2/index.shtml
117. http://www.bioinformatics.babraham.ac.uk/projects/hicup/

118. https://github.com/samtools/samtools
119. Hu, Ming, et al. "HiCNorm: removing biases in Hi-C data via Poisson regression." Bioinformatics 28.23 (2012): 3131-3133.
120. Schmitt, Anthony D., Ming Hu, and Bing Ren. "Genome-wide mapping and analysis of chromosome architecture." Nature reviews Molecular cell biology 17.12 (2016): 743-755.
121. Cournac, Axel, et al. "Normalization of a chromosomal contact map." BMC genomics 13.1 (2012): 1-13.
122. Knight, Philip A., and Daniel Ruiz. "A fast algorithm for matrix balancing." IMA Journal of Numerical Analysis 33.3 (2013): 1029-1047.
123. Williamson, Iain, et al. "Spatial genome organization: contrasting views from chromosome conformation capture and fluorescence in situ hybridization." Genes & development 28.24 (2014): 2778-2791.
124. Chandradoss, Keerthivasan Raanin, et al. "Biased visibility in Hi-C datasets marks dynamically regulated condensed and decondensed chromatin states genome-wide." BMC genomics 21.1 (2020): 1-15.
125. Ohkawa, Yasuyuki, et al. "An improved restriction enzyme accessibility assay for analyzing changes in chromatin structure in samples of limited cell number." Myogenesis. Humana Press, Totowa, NJ, 2012. 531-542.
126. Gargiulo, Gaetano, et al. "NA-Seq: a discovery tool for the analysis of chromatin structure and dynamics during differentiation." Developmental cell 16.3 (2009): 466-481.
127. Williamson, Iain, et al. "Spatial genome organization: contrasting views from chromosome conformation capture and fluorescence in situ hybridization." Genes & development 28.24 (2014): 2778-2791.
128. Bernstein, Emily, et al. "Mouse polycomb proteins bind differentially to methylated histone H3 and RNA and are enriched in facultative heterochromatin." *Molecular and cellular biology* 26.7 (2006): 2560-2569.
129. Pawlicki, Jan M., and Joan A. Steitz. "Primary microRNA transcript retention at sites of transcription leads to enhanced microRNA production." *The Journal of cell biology* 182.1 (2008): 61-76.

Publications

Chandradoss et al. BMC Genomics (2020) 21:175
https://doi.org/10.1186/s12864-020-6580-6

BMC Genomics

RESEARCH ARTICLE Open Access

Biased visibility in Hi-C datasets marks dynamically regulated condensed and decondensed chromatin states genome-wide

Keerthivasan Raanin Chandradoss[1†], Prashanth Kumar Guthikonda[2†], Srinivas Kethavath[2], Monika Dass[1], Harpreet Singh[1], Rakhee Nayak[2], Sreenivasulu Kurukuti[2*] and Kuljeet Singh Sandhu[1*]

Abstract

Background: Proximity ligation based techniques, like Hi-C, involve restriction digestion followed by ligation of formaldehyde cross-linked chromatin. Distinct chromatin states can impact the restriction digestion, and hence the visibility in the contact maps, of engaged loci. Yet, the extent and the potential impact of digestion bias remain obscure and under-appreciated in the literature.

Results: Through analysis of 45 Hi-C datasets, lamina-associated domains (LADs), inactive X-chromosome in mammals, and polytene bands in fly, we first established that the DNA in condensed chromatin had lesser accessibility to restriction endonucleases used in Hi-C as compared to that in decondensed chromatin. The observed bias was independent of known systematic biases, was not appropriately corrected by existing computational methods, and needed an additional optimization step. We then repurposed this bias to identify novel condensed domains outside LADs, which were bordered by insulators and were dynamically associated with the polycomb mediated epigenetic and transcriptional states during development.

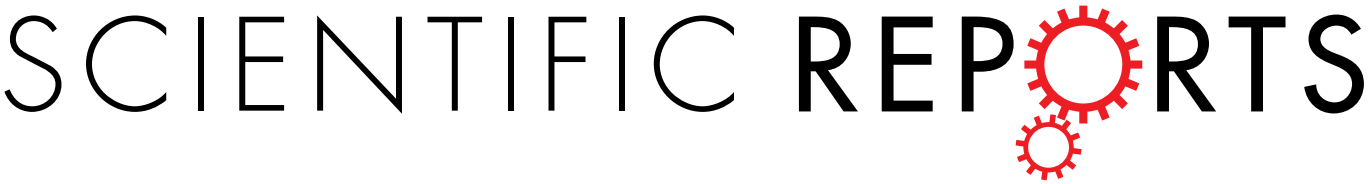
Conclusions: Our observations suggest that the corrected one-dimensional read counts of existing Hi-C datasets can be reliably repurposed to study the gene-regulatory dynamics associated with chromatin condensation and decondensation, and that the existing Hi-C datasets should be interpreted with cautions.

Keywords: Hi-C, 3D genome, Chromatin condensation, Lamina associated domains, CTCF

Background

The three-dimensional genome organization is tightly linked with the regulation of essential genomic functions like transcription, replication and genome integrity [1–5]. While the significance of genome organization has been realized for decades, the comprehensive evidence emerged somewhat recently through the advent of proximity ligation based techniques like Chromosome Conformation Capture (3C), Circular-3C (4C), 3C-Carbon-Copy (5C) and High-throughput 3C (Hi-C) [6–10]. It is recognized that the eukaryotic genome is hierarchically organized into self-interacting topologically associated domains (TADs), which can have distinct chromatin states that are insulated from neighbourhood through boundaries marked with CCCTC-binding factor (CTCF), Cohesins, ZNF143 and TOP2b factors [11–14]. The TADs are ancient genomic features and are depleted in evolutionary breakpoints inside [15, 16]. It is proposed that chromatin extrudes through the ring formed by the Cohesins until the chromatin encounters the CTCF insulator, a model known as 'loop extrusion' model [17–20]. CTCF binding is transiently lost during pro-metaphase, which coincides with the loss of TAD structures during M-phase [21–23].

* Correspondence: skurukuti@uohyd.ac.in; sandhuks@iisermohali.ac.in
†Keerthivasan Raanin Chandradoss and Prashanth Kumar Guthikonda contributed equally to this work.
[2]Department of Animal Biology, School of Life Sciences, University of Hyderabad (UoH), Central University, Prof. CN Rao Road, P O, Gachibowli, Hyderabad, Telangana 500046, India
[1]Department of Biological Sciences, Indian Institute of Science Education and Research (IISER) – Mohali, Knowledge City, Sector 81, SAS Nagar 140306, India



OPEN

Comprehensive profiling of transcriptional networks specific for lactogenic differentiation of HC11 mammary epithelial stem-like cells

Received: 13 December 2017
Accepted: 13 July 2018
Published online: 06 August 2018

Trinadha Rao Sornapudi[1], Rakhee Nayak[1], Prashanth Kumar Guthikonda[1], Anil Kumar Pasupulati[3], Srinivas Kethavath[1], Vanita Uppada[1], Sukalpa Mondal[1], Sailu Yellaboina[2,4] & Sreenivasulu Kurukuti[1]

The development of mammary gland as a lactogenic tissue is a highly coordinated multistep process. The epithelial cells of lactiferous tubules undergo profound changes during the developmental window of puberty, pregnancy, and lactation. Several hormones including estrogen, progesterone, glucocorticoids and prolactin act in concert, and orchestrate the development of mammary gland. Understanding the gene regulatory networks that coordinate proliferation and differentiation of HC11 Mammary Epithelial stem-like Cells (MEC) under the influence of lactogenic hormones is critical for elucidating the mechanism of lactogenesis in detail. In this study, we analyzed transcriptome profiles of undifferentiated MEC (normal) and compared them with Murine Embryonic Stem Cells (ESC) using next-generation mRNA sequencing. Further, we analyzed the transcriptome output during lactogenic differentiation of MEC following treatment with glucocorticoids (primed state) and both glucocorticoids and prolactin together (prolactin state). We established stage-specific gene regulatory networks in ESC and MEC (normal, priming and prolactin states). We validated the top up-and downregulated genes in each stage of differentiation of MEC by RT-PCR and found that they are comparable with that of RNA-seq data. HC11 MEC display decreased expression of *Pou5f1 and Sox2*, which is crucial for the differentiation of MEC, which otherwise ensure pluripotency to ESC. *Cited4* is induced during priming and is involved in milk secretion. MEC upon exposure to both glucocorticoids and prolactin undergo terminal differentiation, which is associated with the expression of several genes, including *Xbp1* and *Cbp that* are required for cell growth and differentiation. Our study also identified differential expression of transcription factors and epigenetic regulators in each stage of lactogenic differentiation. We also analyzed the transcriptome data for the pathways that are selectively activated during lactogenic differentiation. Further, we found that selective expression of chromatin modulators (*Dnmt3l*, *Chd9*) in response to glucocorticoids suggests a highly coordinated stage-specific lactogenic differentiation of MEC.

The events of cellular differentiation, which lead the transition of a primary cell into lineage-restricted phenotype, are accompanied by activation of specific gene regulatory networks instead of activation of a single or a few genes[1,2]. The cell fate transitions experience a global transcriptional activation and repression and are regulated by spatiotemporal expression of both transcription factors[3–5] (TFs) and epigenetic regulators[6] (ERs). The transition

[1]Department of Animal Biology, School of Life Sciences, University of Hyderabad, Hyderabad, 500046, India. [2]CR Rao Advanced Institute of Mathematics, Statistics and Computer Sciences, University of Hyderabad campus, Gachibowli, Hyderabad, 500046, India. [3]Department of Biochemistry, School of Life Sciences, University of Hyderabad, Hyderabad, 500046, India. [4]Nucleome Informatics Private Limited, 2nd Floor, Genome Block, Plot No 135, Mythrinagar Phase I, Madinaguda, Hyderabad, 500049, India. Correspondence and requests for materials should be addressed to S.Kurukuti (email: skurukuti@uohyd.ac.in)

Sornapudi *et al. BMC Res Notes (2018) 11:241*
https://doi.org/10.1186/s13104-018-3351-4

BMC Research Notes

DATA NOTE

Open Access

RNA sequencing of murine mammary epithelial stem-like cells (HC11) undergoing lactogenic differentiation and its comparison with embryonic stem cells

Trinadha Rao Sornapudi[1], Rakhee Nayak[1], Prashanth Kumar Guthikonda[1], Srinivas Kethavath[1], Sailu Yellaboina[2] and Sreenivasulu Kurukuti[1*]

Abstract

Objectives: Understanding of transcriptional networks specifying HC11 murine mammary epithelial stem cell-like cells (MEC) in comparison with embryonic stem cells (ESCs) and their rewiring, under the influence of glucocorticoids (GC) and prolactin (PRL) hormones, is critical for elucidating the mechanism of lactogenesis. In this data note, we provide RNA sequencing data from murine MECs and ESCs, MECs treated with steroid hormone alone and in combination with PRL. This data could help in understanding temporal dynamics of mRNA transcription that impact the process of lactogenesis associated with mammary gland development. Further integration of these data sets with existing datasets of cells derived from various stages of mammary gland development and different types of breast tumors, should pave the way for effective prognosis and to develop therapies for breast cancer.

Data description: We have generated RNA-sequencing data representing steady-state levels of mRNAs from murine ESCs, normal MECs (N), MECs primed (P) with hydrocortisone (HC) alone and in combination with PRL hormone by using Illumina sequencing platform. We have generated ~ 58 million reads for ESCs with an average length of ~ 100 nt and an average 115 million good quality mapped reads with an average length of ~ 150 nt for different stages of MECs differentiation.

Keywords: Mammary epithelial cells, HC11 cells, Embryonic stem cells, Transcriptome, RNA sequencing, Cellular differentiation, Glucocorticoid signaling, Prolactin signaling, Lactogenesis

Objective

HC11 cells are PRL responsive epithelial cell clone, derived from the COMMA1D cells and originated from the mammary gland tissue of a pregnant BALB/c mouse and are widely used model system to study the lactogenic differentiation in vitro [1]. Undifferentiated state of MECs is maintained in the presence of Insulin and epidermal growth factor (EGF). They are stimulated to differentiate by withdrawal of EGF and supplemented initially with insulin, GC and later in combination with PRL [2]. Glucocorticoids binds to cytosolic glucocorticoid receptor (GR) and functions via genomic and non-genomic pathways to accompany differential gene expression [3]. Further, PRL, a peptide hormone, upon binding to PRL receptor (PRLr) on plasma membrane initiates cascade of events which ultimately leads to the cytosolic dimerization and nuclear internalization of Stat5a/b, to promote differential expression of genes [4]. Dissecting the gene regulatory networks that act in cohort and orchestrate mammary epithelial cells differentiation under the influence of lactogenic hormones is critical for elucidating the mechanism of lactogenesis in the context of mammary gland development and differentiation. Previous studies

*Correspondence: skurukuti@uohyd.ac.in
[1] Department of Animal Biology, School of Life Sciences, University of Hyderabad, Gachibowli, Hyderabad 500046, India
Full list of author information is available at the end of the article

Anti-Plagiarism Certificate

Spatiotemporal dynamics of chromatin condensation during embryonic stem cells to neuronal differentiation

by Guthikonda Prashanth Kumar

Submission date: 09-Mar-2022 04:06PM (UTC+0530)
Submission ID: 1780165518
File name: Guthikonda_Prashanth_Kumar.pdf (510.83K)
Word count: 19159
Character count: 104465

Spatiotemporal dynamics of chromatin condensation during embryonic stem cells to neuronal differentiation

ORIGINALITY REPORT

4%	4%	4%	0%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.ncbi.nlm.nih.gov Internet Source	2%
2	bmcgenomics.biomedcentral.com Internet Source	<1%
3	Miao Yu, Bing Ren. "The Three-Dimensional Organization of Mammalian Genomes", Annual Review of Cell and Developmental Biology, 2017 Publication	<1%
4	genome.cshlp.org Internet Source	<1%
5	Gilbert, Scott F.. "Developmental Biology", Oxford University Press Publication	<1%
6	Kind, Jop, Ludo Pagie, Havva Ortabozkoyun, Shelagh Boyle, Sandra S. de Vries, Hans Janssen, Mario Amendola, Leisha D. Nolen, Wendy A. Bickmore, and Bas van Steensel.	<1%

"Single-Cell Dynamics of Genome-Nuclear Lamina Interactions", Cell, 2013.
Publication

---

7 Kaifang Pang, Li Wang, Wei Wang, Jian Zhou, Chao Cheng, Kihoon Han, Huda Y. Zoghbi, Zhandong Liu. "Co-expression enrichment analysis at the single-cell level reveals convergent defects in neural progenitor cells and their cell-type transitions in neurodevelopmental disorders", Cold Spring Harbor Laboratory, 2020 <1 %
Publication

---

8 www.biorxiv.org <1 %
Internet Source

---

9 Daijing Sun, Jie Weng, Yuhao Dong, Yan Jiang. "Three-dimensional genome organization in the central nervous system, implications for neuropsychological disorders", Journal of Genetics and Genomics, 2021 <1 %
Publication

---

10 Keerthivasan Raanin Chandradoss, Prashanth Kumar Guthikonda, Srinivas Kethavath, Monika Dass et al. "Biased visibility in HiC datasets marks dynamically regulated condensed and decondensed chromatin states genome-wide", Cold Spring Harbor Laboratory, 2019 <1 %
Publication

11	nbn-resolving.de Internet Source	<1 %
12	Kai Kruse, Clemens B. Hug, Juan M. Vaquerizas. "FAN-C: A Feature-rich Framework for the Analysis and Visualisation of C data", Cold Spring Harbor Laboratory, 2020 Publication	<1 %
13	Submitted to MCAST Student Paper	<1 %
14	Weizhi Ouyang, Zhilin Cao, Dan Xiong, Guoliang Li, Xingwang Li. "Decoding the plant genome: From epigenome to 3D organization", Journal of Genetics and Genomics, 2020 Publication	<1 %
15	Elliott, David, Ladomery, Michael. "Molecular Biology of RNA", Molecular Biology of RNA, 2015 Publication	<1 %
16	Nitasha Sehgal, Andrew J. Fritz, Kristen Morris, Irianna Torres, Zihe Chen, Jinhui Xu, Ronald Berezney. "Gene density and chromosome territory shape", Chromosoma, 2014 Publication	<1 %
17	Poshen B Chen, Lihua J Zhu, Sarah J Hainer, Kurtis N McCannell, Thomas G Fazzio.	<1 %

"Unbiased chromatin accessibility profiling by RED-seq uncovers unique features of nucleosome variants in vivo", BMC Genomics, 2014
Publication

18	Submitted to Queen Mary and Westfield College Student Paper	<1%
19	Submitted to University of Arizona Student Paper	<1%

Exclude quotes	On	Exclude matches	< 14 words
Exclude bibliography	On		