Biased visibility in HiC 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], Kuljeet Singh Sandhu[*,1]

[1] Department of Biological Sciences
Indian Institute of Science Education and Research (IISER) – Mohali
Knowledge City, Sector 81, SAS Nagar 140306, India

[2]Department of Animal Biology
School of Life Sciences
University of Hyderabad (UoH)
Prof. CN Rao Road, P O, Central University
Gachibowli, Hyderabad, Telangana 500046, India

[$]Equal contribution
[*]Co-corresponding authors (*skurukuti@uohyd.ac.in*, *sandhuks@iisermohali.ac.in*)

Abstract

Proximity ligation based techniques, like HiC, involve restriction digestion followed by ligation of formaldehyde cross-linked chromatin. We showed that the HiC datasets have significant bias due to differential accessibility of chromatin domains to restriction endonucleases. Through analysis of lamina-associated domains (LADs), inactive X-chromosome in mammals and polytene bands in fly, we established that the DNA in condensed chromatin had lesser accessibility to nucleases used in HiC as compared to that in decondensed chromatin. The observed bias was independent of known systematic biases attributed to length and GC content of the restriction fragments and was not appropriately corrected by computational methods. We identified novel condensed domains outside LADs, which were bordered by insulators and were dynamically associated with the developmentally regulated epigenetic and transcriptional states. Our observations suggested that the corrected read counts of existing HiC datasets can be reliably repurposed to study the dynamics of chromatin condensation and decondensation and that the existing HiC datasets should be interpreted with cautions.

Keywords

HiC, 3D genome, chromatin condensation, Lamina associated domains, CTCF

Introduction

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 several decades ago, 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 HiC(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 were 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 for evolutionary breakpoints within (15,16). Fudenberg et al proposed that 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). CTCF binding is transiently lost during pro-metaphase, which coincides with the loss of TAD structures during M-phase (18,19) (Oomen et al 2018, biorxiv). Systematic depletion of CTCF and Cohesins also leads to de-insulation and partial disruption of TADs (20,21). An array of studies has shown that TADs function as basic units of 3D genome that dynamically associate with the epigenetic states of genes, including replication timing, during development and differentiation (22-30). How these dynamical epigenetic states of TADs are regulated is not entirely clear. One of the ways, this can be achieved is through chromatin condensation and decondensation, implying inactive and active states of TADs respectively (2,31-34) (Benabdallah et al 2018, Biorxiv). While it is established that the gene-poor and transcriptionally inactive domains locate towards nuclear periphery, and remain stably condensed during differentiation with local gene-specific alterations(35), the dynamics of chromatin condensation and decondensation in the other regions of the genome largely remains under-explored. Condensation and decondensation of chromatin is generally studied through microscopic methods. In this study, we demonstrated that the condensed and decondensed states of chromatin domains could be directly identified from one dimensional HiC read counts.

Yaffe and Tanay have shown that HiC datasets have systematic bias due to differential ligation efficiency of restriction fragments of different lengths, the differential amplification of fragments with GC rich ends and differential mappability of sequences(36). Several methods have since been developed to normalize the aforementioned biases. These

methods can be broadly categorized into two classes, the ones that define the aforementioned biases explicitly in the algorithm and the ones which do not define the source of bias and instead adopt an implicit approach based on fractal folding of the chromatin and the equal visibility of all genomic loci(36-40). We showed that the differential visibility of genomic loci to the restriction endonucleases used in HiC protocols caused potential bias in HiC data, which was not addressed in the explicit methods of HiC correction. HiC datasets were significantly depleted for the interactions impinging from condensed heterochromatin domains. By repurposing the observed bias, we first demonstrated that the bias in one-dimensional read counts of HiC datasets reliably marked the known condensed and decondensed domains in the genome and then highlighted the developmentally regulated dynamics of condensed and decondensed states of chromatin.

Methods

We did not use any statistical method to predetermine sample size. We did not randomized any experiment. We were not blinded to allocations during experiments and outcome assessment. We mentioned the sample sizes and statistical tests wherever applicable. Source of each dataset is listed in Table S1

Processing of HiC datasets and correction of 1D read counts

Wherever processed HiC data was not available, we obtained the SRA files and converted into fastq files using NCBI SRA toolkit. We implemented HiCUP package (http://www.bioinformatics.babraham.ac.uk/projects/hicup/) to process the HiC reads. We also used an in-house pipeline to process the HiC reads through bowtie2 (http://bowtie-bio.sourceforge.net/bowtie2/index.shtml). In brief, we mapped the paired-end reads on indexed mm10 assembly of mouse genome. Reads were then paired and filtered for invalid and duplicated reads. To make one-dimensional chromosomal tracks of HiC data, we mixed together the HiCUP processed paired-end reads and calculated the read counts for 10kb bins of genome. To normalize the known systematic biases in HiC contact maps, we used HiC-Norm (http://www.people.fas.harvard.edu/~junliu/HiCNorm/) and ICE packages (https://github.com/mirnylab). Table S1 marks the details of HiC datasets and their processing details. We obtained the bedgraph files (mm9) for *in-situ* digested chromatin and *in-solution* digested naked DNA from GSE51821. The files were converted into mm10 assembly using UCSC's liftover utility (http://genome.ucsc.edu/cgi-bin/hgLiftOver). Sequence reads were binned into 10kb bins genome-wide. Relative density of restriction

enzyme sites might influence the 1D read counts obtained after HiCUP processing. We, therefore, removed the RE density associated bias in the read counts. We used the residuals of read counts after loess regression against RE density of genomic bins. 1D read counts are also biased due to varying GC content of genomic regions. GC rich domains are readily captured in sequencing reactions as compared to GC-poor regions. We removed this bias by calculating residuals of RE-corrected read counts through loess regression against GC content of corresponding genomic bins. The final corrected read count had no scaling against RE density and GC content of the genomic bins as shown in the Figure S1a-b.

Domain calling

We scaled the 1D HiC read counts using Z-score and identified the condensed and decondensed domains using the strategy given by Guelen et al (41). In short, read counts were first binarized as +1 and -1 depending upon whether the values were positive of negative in the Z-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 through 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 regions, we demarcated condensed and decondensed domains. We set the minimal proportion of either positive or negative values to 0.8 in order to classify the domains as decondensed and condensed respectively.

Analysis of constitutive LADs and constitutive inter-LADs

We downloaded 3843 LAD regions of mouse ES cells, neuronal progenitor cells and astrocytes from GSE17051. We classified LADs as constitutive LADs (cLADs) and constitutive inter-LADs (ciLADs) by comparing the genome coordinates of LADs in mESC, NPC and astrocytes. Similarly, 605 cLADs in human were obtained by comparing LAD coordinates in IMR90 and heterochromatin domain coordinates in h1ESC and K562 cell-lines from ENCODE (https://www.encodeproject.org/comparative/chromatin/). To analyse the alteration in the conformation of cLAD domains, we obtained the Lamin KO HiC data for mESCs, binned the reads into 10 kb bins and used log ratio of Lamin KO to WT for our analyses.

Analysis of polytene and normal diploid HiC in Drosophila

We downloaded HiC SRA files from GSE72510 (polytene, dm6) and GSE63518 (normal diploid Kc167, dm6) and processed using HiCUP pipeline. We binned the reads at 5kb

resolution for RE-density and GC correction of the read counts. We downloaded polytene TADs from Eagen et al (2015) and lifted over the coordinates to dm6 assembly. We mapped 5kb bins to polytene TADs and considered those inside TADs as polytene band bins and rest as inter-band. We generated raw contact maps at a 5kb resolution and normalized using HiC-norm and ICE using python *mirnylab* package.

Allele-specific HiC analysis of X-chromosome

We downloaded allele-specific valid HiC pairs of brain and patski cells from GSE68992. We removed the reads mapping to both references and binned the allele-specific reads at 20kb resolution to obtain one dimensional read counts. We corrected the read counts for GC content using loess regression as earlier. To visualize the interactions, we generated raw and ICE corrected contact maps at 100kb. HiCNorm does not suit to DNase-HiC data due to the usage of DNase instead of restriction endonuclease, and hence not used.

Analysis of histone modification and CTCF ChIP-seq datasets

Source of ChIP-Seq datasets are given in Table S1. Then, we binned the reads at 10kb resolution, and quantile normalized using the R package *preprocessCore (https://github.com/bmbolstad/preprocessCore)*. We generated median aggregate plots by aligning all boundaries that have at least 200kb long open/close on their either sides. We used R-package *ggplot2 (https://ggplot2.tidyverse.org/)* to scatterplot the data.

Results

Biased visibility in HiC data marks condensed and decondensed chromatin domains

We first showed that the *in-situ* restriction digestion of chromatin was not uniform in the genome. Towards this, we obtained the sequencing data for *in-situ* restriction digested chromatin and *in-solution* restriction digested naked DNA of mouse embryonic stem cells (mESC) (42). We calculated the read counts for 10 kb bins of the mouse genome and normalized by the total reads. We further corrected the read counts for restriction site density (RE-density) of the bins and the GC content of the bins using loess regression, in that order (Methods, Figure S1a-b). The scatter-plot of restriction digested naked DNA and *in-situ* digested chromatin showed skew towards naked DNA axis marking the inefficient digestion of certain genomic regions in chromatin but not in naked DNA (Figure 1a). This suggested that chromatin structure had influenced its own digestibility. The likely explanation was that the decondensed chromatin was readily digested while heterochromatin domains had

limited accessibility to restriction endonuclease. To assess this hypothesis, we obtained the Lamina Associated Domains (LADs), which are known heterochromatin domains attached to the nuclear periphery in condensed form(31,43). We calculated the raw and corrected read counts in the constitutive LADs (cLADs) and constitutive inter-LADs (ciLADs) in mESC. As shown in the figure 1b-c, cLADs exhibited significantly less raw read counts as compared to ciLADs in *in-situ* digested chromatin as well as in *in-solution* digested naked DNA, suggesting that the reads from digested naked DNA had bias likely due to varying densities of restriction sites and the relative GC content of cLADs and ciLADs (Figure 1b-c, $p<2.2e-16$). The read counts corrected for RE-density and GC content, however, exhibited bias only in the *in-situ* digested chromatin and not in the naked DNA, highlighting that the cLADs were relatively inaccessible to restriction endonuclease likely due to condensed nature of the chromatin (Figure 1b-c, $p<2.2e-16$). We further identified the chromatin domains significantly enriched (decondensed) or depleted (condensed) in corrected read counts (Methods, Figure S1c). Overall, 77% of length covered by condensed domains were within cLADs and 23% mapped to ciLAD regions, marking the condensed domains which were not associated with the nuclear lamina (Figure S1d).

We then expanded our analyses to 40 HiC datasets (21 *in-situ* HiC, 11 *in-solution* HiC and 8 scHiC) and obtained the processed reads (Figure 1d-f, Figure S2, & Table S1). One-dimensional read counts were corrected for the density of restriction sites and GC content as earlier. Through analysis of mESC data, we observed that the reads from *in-situ* HiC had a significant correlation with the reads from *in-situ* digested chromatin, but exhibited skewed scaling towards digested naked DNA (Figure 1d). This suggested that the *in-situ* HiC reads exhibited bias similar to the one in *in-situ* digested chromatin. As shown in the figure 1e-f and figure S2, the corrected read counts exhibited enrichment in ciLADs and depletion in cLADs ($p<2.2e-16$). Again, 70% of length covered by the condensed domains were within cLADs and 30% was within ciLAD regions, marking the condensed domains other than LADs. (Figure S1d). Our observations with cLADs and ciLADs were consistent with different *in-situ* HiC datasets, including single-cell HiC, generated using distinct restriction enzymes (Figure 1d-i and S2, $p<2.2e-16$). We illustrated the examples of condensed domains that mapped to cLADs, to ciLADs and the ones that exhibited cell-type specificity in the figure 1f and S2.

We also showed that the observed association with the cLAD and ciLADs was not due to HiC-specific processing of the sequencing data. We observed the bias in reads simply processed

through bowtie too (Figure S3a, p<2.2e-16, Methods). Further, the biased visibility was not the property of *in-situ* HiC only, but was also observed in *in-solution* HiC too (Figure 1g-l, S3b, p<2.2e-16). As shown in the figure S3c, the corrected reads from *in-situ* HiC exhibited good correlation with those from *in-solution* HiC in the same cell-type (mouse fetal liver) from the same study(44). This suggested that the visibility bias was not affected by the method of ligation and that the source of bias was the difference in accessibility to the restriction endonucleases, not the difference in ligation.

HiCNorm, an explicit method of HiC correction, failed to remove the bias in the read counts, supporting that the observed bias is independent of known systematic biases of HiC data (Figure 2a & S4, p<2.2e-16). Iterative correction, an implicit method, normalized the read counts attributing to its intrinsic nature of polishing the HiC matrices for equal visibility of all loci without defining the bias at first place (Figure 2a & S4). Data obtained from Genome architecture mapping (GAM)(45), which directly obtains the co-localized DNA segments through large number of thin nuclear sections and does not involve any restriction and ligation steps, did not exhibit any bias in the read counts. By comparing GAM and ICE-corrected HiC data, we further observed that ICE merely lifted the background and the obscure signals in the contact matrices. In the process of lifting the obscure signals in the poorly digested condensed regions, ICE inadvertently lifted the long-range background interactions impinging from condensed domains as shown in the figure 2b and S4. We illustrated the examples of condensed and decondensed domains in the raw HiC, normalized HiC and GAM datasets in the figure 2c-d. Depletion of HiC signals in the condensed domains was apparent in the raw and the HiCNorm-corrected contact matrices (Figure 2b-c). ICE lifted the obscure signals in a uniform and pattern-less manner, while GAM matrices clearly suggested structured sub-TAD and other types of interactions in the condensed domains. These observations hghlighted that the lack of proper HiC signals in the condensed domains could not be appropriately corrected by computational approaches. Instead, our results emphasized that HiC method itself needs to be improved by incorporating better fragmentation of chromatin.

To further scrutinize the differential digestion of condensed and decondensed domains, we obtained the HiC data for the polytene chromosome of Drosophila, which is a typical example of spatially condensed (polytene bands) and decondensed (inter-bands) domains (46). The HiC reads were mapped and corrected as earlier. The analysis suggested that the

polytene bands had lesser enrichment of corrected reads as compared to inter-band regions on both the polytene chromosome and the normal diploid chromosome (Figure 3a, $p<2.2e-16$). We illustrated our observations through examples in the figure 3b. On similar lines, we analysed the DNase-HiC data for active and inactive X-chromosomes in brain and patski cells (47). As shown through the scatter plots in figure 3c and examples in figure 3d, the X-chromosome had regions that were more visible in active X-chromosome and less visible in inactive X-chromosome. This suggested that the bias due to differential chromatin accessibility existed in both restriction endonuclease digested and DNase digested HiC datasets.

These observations highlighted that the observed variations in corrected HiC read counts was independent of known systematic biases of HiC and captured the condensed and decondensed states of chromatin domains reliably.

Dynamics of condensed and decondensed domains

To assess if the condensed and decondensed domains identified from restriction digestion bias in the ciLAD regions had functional significance, we analysed their dynamics during mouse embryonic stem cell (mESC) differentiation to neuronal progenitor cells (NPC) to cortical neurons (CN). As shown in the figure S5a, the differentiation from mESC to NPC exhibited greater overall change in corrected read counts as compared to NPC to CN differentiation. We, therefore, focussed on mESC to NPC differentiation to assess the developmental regulation of chromatin condensation and decondensation. We first mapped the histone modification and CTCF binding data around boundaries of domains by scaling all decondensed domains upstream and all condensed domains downstream to the domain boundaries (Figure 4a & S5b). We observed enrichment of active and inactive histone marks in decondensed and condensed domains respectively with transitions around boundaries that were marked with CTCF, RAD21, YY1, TOP2b, MIR and simple repeat elements (Figure 4a-b & S5c, $p=4.5e-05$ to $2.2e-16$). Total 27.7% of condensed ciLAD domains in mESC were decondensed in NPC and 13.5% decondensed domains in mESC were condensed in NPC, suggesting significant cell-type specificity of domains identified through biased visibility in HiC data (Figure S5d). Genes exhibiting condensation during differentiation switched to repressed state and the ones showing decondensation switched to active state (Figure 4c, $p=6.6e-13$ & $2.2e-16$). Through scatter plots of active and inactive chromatin types between mESC and NPC cells, we observed that the condensation of open chromatin domains during

differentiation associated with the coherent change of active to inactive chromatin states (Figure 4d). Similarly, the domains that exhibited decondensation during differentiation switched to active states from inactive states (Figure 4d). Figure 4e and S5e-g presented a few examples of chromatin domains that were constitutive (left panel) or cell-type specific (right panel) in mESC and NPC. These observations not only highlighted the developmental regulation of chromatin domains identified in the study, but also argued strongly against the dismissal of restriction digestion bias merely as an artefact.

Chromatin condensation and decondensation can be induced by knocking out certain factors like Lamins. We, therefore, tested if the such experimentally induced decondensation of LADs can be captured through analysis of HiC reads of Lamin knock out cells. We obtained the HiC data for WT and Lamin (*Lmb1, Lmb2, Lmna*) KO mouse embryonic stem cells from Zheng et al (48). As shown in the figure 5a, the cLADs in Lamin KO cells exhibited relatively more reads as compared to that in WT ($p<2.2e-16$). We illustrated this observation through examples in figure 5b-c. Our observations highlighted that the HiC read-counts alone captured the decondensation of LAD domains after lamin deletion.

Discussion

To study the dynamics of gene regulatory activity in response to environmental and developmental clues, mapping chromatin accessibility remained an important task over decades. While DNase-I had been a preferred choice to digest the chromatin due to relatively lesser sequence specificity of the enzyme, only few studies had attempted restriction endonucleases to identify accessible regions in the chromatin (42,49,50). On the contrary, restriction endonucleases have been extensively used to digest the chromatin, presumably in an unbiased manner, in proximity ligation based techniques like 3C, 4C, 5C and HiC (6-9). In techniques like 3C, which has limited scope in terms of regions to be tested for spatial interactions, the efficiency of restriction digestion of the regions of interest can be tested and controlled. For high throughput assays like HiC, data on genome-wide assessment of digestion efficiency is rarely seen in the research articles and the associated supplementary materials. It is presumed that prolonged incubation with restriction endonuclease ensures the equal visibility all genomic loci. We claim that the condensed nature of heterochromatin limits its visibility to restriction endonucleases and this introduces significant non-uniformity in HiC data. As a result, interactions within heterochromatin and the ones impinging onto heterochromatin are under-represented in

the current HiC datasets. We showed that this bias was not corrected by the methods that define the potential sources of bias explicitly. Widely used iterative correction method scales the contact matrices for equal visibility of each locus, without defining the source of bias at first place(38). We observed that the iterative correction method balanced the HiC signals in condensed and decondensed regions of the contact matrix locally. However, there were two issues with that: 1) It only raised the obscure signals in the HiC matrix uniformly and did not capture the structured pattern, which was observed through GAM; 2) It also raised the long-range signals, which otherwise were infrequent, impinging from the condensed domains. It can be argued that condensed inactive domains should exhibit more interaction frequencies due to its compact nature and density of HiC reads should ideally favour condensed than decondensed domains. Indeed, some of the examples shown in the figure 2 had more intense HiC signals in the condensed regions as compared to decondensed ones. Therefore, the presumption of the equal visibility of all genomic domains might not be appropriate either. These observations suggested that computational methods do not appropriately correct the depleted signals in the condensed chromatin domains and the bias needs to be corrected at experimental level itself.

Williamson et al had earlier reported discrepancy between 5C/HiC and the DNA FISH results concerning the condensation and decondensation of Hoxd locus during embryonic stem cell differentiation. While 5C/HiC largely showed that the locus remained condensed, DNA FISH clearly suggested that the locus decondensed upon differentiation (33). We suggest that the biased efficiency of restriction digestion of condensed and decondensed forms of the locus might underlie such interpretations.

We propose that unless the restriction digestion bias is corrected experimentally, it is advantageous not to erase this bias from the data. Differential visibility of chromatin domains help inferring dynamics of condensed and decondensed states of chromatin. Towards this, our analysis on developmental regulation of condensed and decondensed chromatin domains serve as a proof of principle. Coordinated changes in restriction enzyme accessibility and the epigenetic states of the genes contended against the arguments dismissing the observed bias merely as GC content bias or random variation in the data. The active chromatin marks exhibited shift towards the axis that represented the decondensed state of the involved domain in mESC or NPC, however the repressive H3K9me3 mark showed relatively subtle shift only. This was coherent with the earlier reports that suggested

only subtle changes in H3K9 methylation profiles during mESC differentiation (51,52). We instead observed that the enrichment of polycomb associated proteins Suz12 and Ezh2 exhibited shift towards the axis that represented condensed state of chromatin domains (Figure S5h), suggesting that the non-LAD condensed domains uncovered in this study were likely representative of polycomb repressed chromatin. Dynamical changes in the restriction enzyme accessibility were also observed during synthetic manipulation of the genome organization. Lamin deletion in the genome is associated with decondensation of LADs (Zheng et al, biorxiv 2018). We readily captured the decondensation of LADs in Lamin knock out cells from the HiC reads itself.

To identify the sequence features of domain boundaries, we tested the enrichment of several different genomic and epigenomic attributes (Figure 4b, S5c). We observed the enrichment of CTCF, RAD21, TOP2b, YY1 binding sites and repeat elements like MIR at boundaries, which are well known boundary elements of chromatin domains (11,14,53,54). These observations collectively reinforce our claim on biological authenticity of condensed and decondensed domains identified outside the LAD regions.

A major criticism of our observations could be the proposal that potential source of the observed bias was differential restriction digestion. It can be argued that restriction digestion could be uniform, but the differential ligation could have caused the biased visibility. We do not entirely rule out the possibility that the restriction endonuclease nicked the DNA in condensed domains, but ligase enzyme failed to ligate the digested DNA due to stiffness and steric hindrance caused by nucleosomes and heterochromatin proteins in the condensed domains. However, we counter-argue that this may not be the case because when we plotted the HiC read counts against those of restriction digested chromatin, we observed a linear scaling, while there should have been bias towards the axis of restriction digested chromatin because of presumed efficient digestion and poor ligation in HiC experiments. We also proposed that inefficient ligation of digested condensed chromatin can largely be a property of *in-situ* HiC and not the *in-solution* HiC, which involves dilution of digested chromatin and condensed chromatin is likely to exhibit relatively loose conformation owing to the usage of detergent and the heat during dilution step. We, therefore, compared the read counts of *in-situ* HiC and *in-solution* HiC datasets obtained from the same study. As shown in the figure S3c, we found good correlation between the two, while bias should have been reflected towards the axis representing reads from *in-*

solution HiC assuming efficient digestion but poor ligation *in-situ* HiC. We thus proposed that the biased visibility in HiC datasets is caused by biased restriction digestion and not the ligation.

Conclusion

Taken together, we highlighted a significant bias in the visibility of condensed and decondensed chromatin domains in HiC datasets. The computational methods failed to correct this bias appropriately. We claimed that the source of this bias is non-uniform digestion of chromatin through restriction endonuclease and DNase –I enzymes. Finally, we showed that the repurposing of observed bias itself was sufficient to study the developmental dynamics of chromatin condensation and decondensation.

Data availability

The bedgraph files for the corrected read counts of all the HiC datasets are available at following link:

https://bitbucket.org/ken_at_keerthivasan/compaction_from_hic/downloads/

Funding

This work was supported by Department of Biotechnology (DBT-India), (BT/PR13596/GET/119/31/2015; BT/PR8688/AGR/36/755/2013; BT/PR16366/BID/7/598/2016), Indian Council of Medical Research (F.No.90/09/2012/SCRT(TF)/BMS) and SERB (EMR/2015/001681).

Conflict of interest

The authors declare that they do not have any conflict of interests.

Acknowledgements

KRC and PKG acknowledge ICMR and CSIR respectively for their fellowship support.

References

1. Bonev, B. and Cavalli, G. (2016) Organization and function of the 3D genome. *Nature reviews. Genetics*, **17**, 772.
2. Therizols, P., Illingworth, R.S., Courilleau, C., Boyle, S., Wood, A.J. and Bickmore, W.A. (2014) Chromatin decondensation is sufficient to alter nuclear organization in embryonic stem cells. *Science*, **346**, 1238-1242.

3. Dekker, J. and Mirny, L. (2016) The 3D Genome as Moderator of Chromosomal Communication. *Cell*, **164**, 1110-1121.
4. Gonzalez-Sandoval, A., Towbin, B.D., Kalck, V., Cabianca, D.S., Gaidatzis, D., Hauer, M.H., Geng, L., Wang, L., Yang, T., Wang, X. *et al.* (2015) Perinuclear Anchoring of H3K9-Methylated Chromatin Stabilizes Induced Cell Fate in C. elegans Embryos. *Cell*, **163**, 1333-1347.
5. Garcia-Nieto, P.E., Schwartz, E.K., King, D.A., Paulsen, J., Collas, P., Herrera, R.E. and Morrison, A.J. (2017) Carcinogen susceptibility is regulated by genome architecture and predicts cancer mutagenesis. *The EMBO journal*, **36**, 2829-2843.
6. Dostie, J., Richmond, T.A., Arnaout, R.A., Selzer, R.R., Lee, W.L., Honan, T.A., Rubio, E.D., Krumm, A., Lamb, J., Nusbaum, C. *et al.* (2006) Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements. *Genome Res*, **16**, 1299-1309.
7. Dekker, J., Rippe, K., Dekker, M. and Kleckner, N. (2002) Capturing chromosome conformation. *Science*, **295**, 1306-1311.
8. Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O. *et al.* (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, **326**, 289-293.
9. Zhao, Z., Tavoosidana, G., Sjolinder, M., Gondor, A., Mariano, P., Wang, S., Kanduri, C., Lezcano, M., Sandhu, K.S., Singh, U. *et al.* (2006) Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nature genetics*, **38**, 1341-1347.
10. Simonis, M., Klous, P., Splinter, E., Moshkin, Y., Willemsen, R., de Wit, E., van Steensel, B. and de Laat, W. (2006) Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nature genetics*, **38**, 1348-1354.
11. Uuskula-Reimand, L., Hou, H., Samavarchi-Tehrani, P., Rudan, M.V., Liang, M., Medina-Rivera, A., Mohammed, H., Schmidt, D., Schwalie, P., Young, E.J. *et al.* (2016) Topoisomerase II beta interacts with cohesin and CTCF at topological domain borders. *Genome Biol*, **17**, 182.
12. Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S. and Ren, B. (2012) Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, **485**, 376-380.
13. Dixon, J.R., Gorkin, D.U. and Ren, B. (2016) Chromatin Domains: The Unit of Chromosome Organization. *Mol Cell*, **62**, 668-680.
14. Heidari, N., Phanstiel, D.H., He, C., Grubert, F., Jahanbani, F., Kasowski, M., Zhang, M.Q. and Snyder, M.P. (2014) Genome-wide map of regulatory interactions in the human genome. *Genome Res*, **24**, 1905-1917.
15. Harmston, N., Ing-Simmons, E., Tan, G., Perry, M., Merkenschlager, M. and Lenhard, B. (2017) Topologically associating domains are ancient features that coincide with Metazoan clusters of extreme noncoding conservation. *Nature communications*, **8**, 441.
16. Krefting, J., Andrade-Navarro, M.A. and Ibn-Salem, J. (2018) Evolutionary stability of topologically associating domains is associated with conserved gene regulation. *BMC biology*, **16**, 87.

17. Fudenberg, G., Imakaev, M., Lu, C., Goloborodko, A., Abdennur, N. and Mirny, L.A. (2016) Formation of Chromosomal Domains by Loop Extrusion. *Cell Rep*, **15**, 2038-2049.
18. Agarwal, H., Reisser, M., Wortmann, C. and Gebhardt, J.C.M. (2017) Direct Observation of Cell-Cycle-Dependent Interactions between CTCF and Chromatin. *Biophysical journal*, **112**, 2051-2055.
19. Nagano, T., Lubling, Y., Varnai, C., Dudley, C., Leung, W., Baran, Y., Mendelson Cohen, N., Wingett, S., Fraser, P. and Tanay, A. (2017) Cell-cycle dynamics of chromosomal organization at single-cell resolution. *Nature*, **547**, 61-67.
20. Nora, E.P., Goloborodko, A., Valton, A.L., Gibcus, J.H., Uebersohn, A., Abdennur, N., Dekker, J., Mirny, L.A. and Bruneau, B.G. (2017) Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell*, **169**, 930-944 e922.
21. Schwarzer, W., Abdennur, N., Goloborodko, A., Pekowska, A., Fudenberg, G., Loe-Mie, Y., Fonseca, N.A., Huber, W., C, H.H., Mirny, L. *et al.* (2017) Two independent modes of chromatin organization revealed by cohesin removal. *Nature*, **551**, 51-56.
22. Pope, B.D., Ryba, T., Dileep, V., Yue, F., Wu, W., Denas, O., Vera, D.L., Wang, Y., Hansen, R.S., Canfield, T.K. *et al.* (2014) Topologically associating domains are stable units of replication-timing regulation. *Nature*, **515**, 402-405.
23. Neems, D.S., Garza-Gongora, A.G., Smith, E.D. and Kosak, S.T. (2016) Topologically associated domains enriched for lineage-specific genes reveal expression-dependent nuclear topologies during myogenesis. *Proc Natl Acad Sci U S A*, **113**, E1691-1700.
24. Le Dily, F., Bau, D., Pohl, A., Vicent, G.P., Serra, F., Soronellas, D., Castellano, G., Wright, R.H., Ballare, C., Filion, G. *et al.* (2014) Distinct structural transitions of chromatin topological domains correlate with coordinated hormone-induced gene regulation. *Genes Dev*, **28**, 2151-2162.
25. Fraser, J., Ferrai, C., Chiariello, A.M., Schueler, M., Rito, T., Laudanno, G., Barbieri, M., Moore, B.L., Kraemer, D.C., Aitken, S. *et al.* (2015) Hierarchical folding and reorganization of chromosomes are linked to transcriptional changes in cellular differentiation. *Molecular systems biology*, **11**, 852.
26. Ke, Y., Xu, Y., Chen, X., Feng, S., Liu, Z., Sun, Y., Yao, X., Li, F., Zhu, W., Gao, L. *et al.* (2017) 3D Chromatin Structures of Mature Gametes and Structural Reprogramming during Mammalian Embryogenesis. *Cell*, **170**, 367-381 e320.
27. Flyamer, I.M., Gassler, J., Imakaev, M., Brandao, H.B., Ulianov, S.V., Abdennur, N., Razin, S.V., Mirny, L.A. and Tachibana-Konwalski, K. (2017) Single-nucleus Hi-C reveals unique chromatin reorganization at oocyte-to-zygote transition. *Nature*, **544**, 110-114.
28. Kaaij, L.J.T., van der Weide, R.H., Ketting, R.F. and de Wit, E. (2018) Systemic Loss and Gain of Chromatin Architecture throughout Zebrafish Development. *Cell Rep*, **24**, 1-10 e14.
29. Bonev, B., Mendelson Cohen, N., Szabo, Q., Fritsch, L., Papadopoulos, G.L., Lubling, Y., Xu, X., Lv, X., Hugnot, J.P., Tanay, A. *et al.* (2017) Multiscale 3D

Genome Rewiring during Mouse Neural Development. *Cell*, **171**, 557-572 e524.
30. Boya, R., Yadavalli, A.D., Nikhat, S., Kurukuti, S., Palakodeti, D. and Pongubala, J.M.R. (2017) Developmentally regulated higher-order chromatin interactions orchestrate B cell fate commitment. *Nucleic acids research*, **45**, 11070-11087.
31. Ciabrelli, F. and Cavalli, G. (2015) Chromatin-driven behavior of topologically associating domains. *Journal of molecular biology*, **427**, 608-625.
32. Rafique, S., Thomas, J.S., Sproul, D. and Bickmore, W.A. (2015) Estrogen-induced chromatin decondensation and nuclear re-organization linked to regional epigenetic regulation in breast cancer. *Genome Biol*, **16**, 145.
33. Williamson, I., Berlivet, S., Eskeland, R., Boyle, S., Illingworth, R.S., Paquette, D., Dostie, J. and Bickmore, W.A. (2014) Spatial genome organization: contrasting views from chromosome conformation capture and fluorescence in situ hybridization. *Genes Dev*, **28**, 2778-2791.
34. Chambeyron, S. and Bickmore, W.A. (2004) Chromatin decondensation and nuclear reorganization of the HoxB locus upon induction of transcription. *Genes Dev*, **18**, 1119-1130.
35. Peric-Hupkes, D., Meuleman, W., Pagie, L., Bruggeman, S.W., Solovei, I., Brugman, W., Graf, S., Flicek, P., Kerkhoven, R.M., van Lohuizen, M. *et al.* (2010) Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation. *Mol Cell*, **38**, 603-613.
36. Yaffe, E. and Tanay, A. (2011) Probabilistic modeling of Hi-C contact maps eliminates systematic biases to characterize global chromosomal architecture. *Nature genetics*, **43**, 1059-1065.
37. Hu, M., Deng, K., Selvaraj, S., Qin, Z., Ren, B. and Liu, J.S. (2012) HiCNorm: removing biases in Hi-C data via Poisson regression. *Bioinformatics*, **28**, 3131-3133.
38. Imakaev, M., Fudenberg, G., McCord, R.P., Naumova, N., Goloborodko, A., Lajoie, B.R., Dekker, J. and Mirny, L.A. (2012) Iterative correction of Hi-C data reveals hallmarks of chromosome organization. *Nature methods*, **9**, 999-1003.
39. Cournac, A., Marie-Nelly, H., Marbouty, M., Koszul, R. and Mozziconacci, J. (2012) Normalization of a chromosomal contact map. *BMC genomics*, **13**, 436.
40. Rao, S.S., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S. *et al.* (2014) A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*, **159**, 1665-1680.
41. Guelen, L., Pagie, L., Brasset, E., Meuleman, W., Faza, M.B., Talhout, W., Eussen, B.H., de Klein, A., Wessels, L., de Laat, W. *et al.* (2008) Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature*, **453**, 948-951.
42. Chen, P.B., Zhu, L.J., Hainer, S.J., McCannell, K.N. and Fazzio, T.G. (2014) Unbiased chromatin accessibility profiling by RED-seq uncovers unique features of nucleosome variants in vivo. *BMC genomics*, **15**, 1104.

43. van Steensel, B. and Belmont, A.S. (2017) Lamina-Associated Domains: Links with Chromosome Architecture, Heterochromatin, and Gene Repression. *Cell*, **169**, 780-791.
44. Nagano, T., Varnai, C., Schoenfelder, S., Javierre, B.M., Wingett, S.W. and Fraser, P. (2015) Comparison of Hi-C results using in-solution versus in-nucleus ligation. *Genome Biol*, **16**, 175.
45. Beagrie, R.A., Scialdone, A., Schueler, M., Kraemer, D.C., Chotalia, M., Xie, S.Q., Barbieri, M., de Santiago, I., Lavitas, L.M., Branco, M.R. *et al.* (2017) Complex multi-enhancer contacts captured by genome architecture mapping. *Nature*, **543**, 519-524.
46. Eagen, K.P., Hartl, T.A. and Kornberg, R.D. (2015) Stable Chromosome Condensation Revealed by Chromosome Conformation Capture. *Cell*, **163**, 934-946.
47. Deng, X., Ma, W., Ramani, V., Hill, A., Yang, F., Ay, F., Berletch, J.B., Blau, C.A., Shendure, J., Duan, Z. *et al.* (2015) Bipartite structure of the inactive mouse X chromosome. *Genome Biol*, **16**, 152.
48. Zheng, X., Hu, J., Yue, S., Kristiani, L., Kim, M., Sauria, M., Taylor, J., Kim, Y. and Zheng, Y. (2018) Lamins Organize the Global Three-Dimensional Genome from the Nuclear Periphery. *Mol Cell*, **71**, 802-815 e807.
49. Ohkawa, Y., Mallappa, C., Vallaster, C.S. and Imbalzano, A.N. (2012) An improved restriction enzyme accessibility assay for analyzing changes in chromatin structure in samples of limited cell number. *Methods in molecular biology*, **798**, 531-542.
50. Gargiulo, G., Levy, S., Bucci, G., Romanenghi, M., Fornasari, L., Beeson, K.Y., Goldberg, S.M., Cesaroni, M., Ballarini, M., Santoro, F. *et al.* (2009) NA-Seq: a discovery tool for the analysis of chromatin structure and dynamics during differentiation. *Developmental cell*, **16**, 466-481.
51. Filion, G.J. and van Steensel, B. (2010) Reassessing the abundance of H3K9me2 chromatin domains in embryonic stem cells. *Nature genetics*, **42**, 4; author reply 5-6.
52. Lienert, F., Mohn, F., Tiwari, V.K., Baubec, T., Roloff, T.C., Gaidatzis, D., Stadler, M.B. and Schubeler, D. (2011) Genomic prevalence of heterochromatic H3K9me2 and transcription do not discriminate pluripotent from terminally differentiated cells. *PLoS genetics*, **7**, e1002090.
53. Wang, J., Vicente-Garcia, C., Seruggia, D., Molto, E., Fernandez-Minan, A., Neto, A., Lee, E., Gomez-Skarmeta, J.L., Montoliu, L., Lunyak, V.V. *et al.* (2015) MIR retrotransposon sequences provide insulators to the human genome. *Proc Natl Acad Sci U S A*, **112**, E4428-4437.
54. Cuddapah, S., Jothi, R., Schones, D.E., Roh, T.Y., Cui, K. and Zhao, K. (2009) Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains. *Genome Res*, **19**, 24-32.

Figure legends

Figure 1. Biased visibility of chromatin domains in *in-situ* HiC datasets. (a) Scatter plots of raw and corrected read counts (per Mb) in *in-situ* digested chromatin vs. *in-solution* digested naked DNA. (b) Distribution of raw and corrected read counts of *in-situ* digested chromatin and in-solution digested naked DNA in cLADs and ciLADs. P-values were calculated using

two-tailed Mann-Whiney U tests. (c) Illustrative example of raw and corrected read counts of *in-situ* digested chromatin and *in-solution* digested naked DNA along chr4: 20-40Mb region. (d) Top: scatter plots of raw and corrected read counts in *in-situ* HiC and *in-situ* digested chromatin. Bottom: scatter plots of raw and corrected read counts in *in-situ* HiC and *in-solution* digested naked DNA. *'ρ'* represents the Pearson's correlation coefficient. (e) Distribution of raw and corrected read counts of *in-situ* HiC datasets in cLADs and ciLADs. P-values were calculated using two-tailed Mann-Whiney U test. (f) Illustrative examples of corrected read counts of *in-situ* HiC datasets along chr7: 100-120 Mb. Regions (i) and (ii) mark constitutively condensed and decondensed regions respectively. Regions (iii)-(v) mark cell-type specific condensed and decondensed states. (ii) (g-i) Same as d-f, but for *in-solution* HiC data obtained from Fraser et al.

Figure 2. Bias in explicitly and implicitly normalized HiC, and GAM datasets. (a) Distribution of 1D read-counts (0-1) of decondensed and condensed domains in raw, HiCNorm-corrected, ICE-corrected HiC and GAM datasets of mESC. Values were scaled from 0 to 1. P-values were calculated using two-tailed Mann-Whiney U tests. (b) Ratio of interaction frequencies of decondensed-to-decondensed and condensed-to-condensed interactions as a function of genomic distance in raw, HiCNorm-corrected, ICE-corrected and GAM datasets. (c) Illustrative examples of raw, HiCNorm corrected, ICE corrected, and GAM read counts and the contact maps.

Figure 3. Low visibility of polytene bands and inactive X-chromosome. (a) Distribution of raw and corrected read counts in band and inter-band regions of polytene chromosome and the corresponding regions in diploid chromosome. P-values were calculated using two-tailed Mann-Whitney U tests. (b) Illustrative examples of read counts and contact maps in band and inter-band regions of polytene and diploid chromosome. Band regions are marked as horizontal line below the line plots. (c) Scatter plots of raw and corrected DNase-HiC read counts of active vs. inactive x-chromosomes in Brain and Patski cells. (d) Illustrative examples of corrected read counts and contact maps of chrX: 36-44 Mb region in active and inactive X-chromosome.

Figure 4. Developmental dynamics of chromatin condensation and decondensation. (a) Aggregation plots of histone modifications +/- 1Mb around the boundary of decondensed and condensed domains in mESC and NPC. 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) Enrichment of CTCF, RAD21, YY1, TOP2b binding, MIR and simple repeats +/- 1Mb around domain boundaries (red) and around domain centres (grey). (c) Boxplots representing change in gene expression in the chromatin domains that were constitutively present in mESC and NPC, the ones that switched to condensed state in NPC from decondensed state in mESC and *vice-versa*. P-values were calculated using two-tailed Mann-Whitney U tests of RPKM values in mESC and NPC. (d) Scatter plots of histone modifications in domains that remained unchanged in mESC and NPC, and the ones that switched from decondensed to condensed or *vice-versa* in mESC and NPC. (e) Examples of decondensed and condensed domains that remained consistent in mESC and NPC (left), and a decondensed region in mESC that switched to condensed state in NPC.

Figure 5. Capturing chromatin decondensation in Lamin KO cells . (a) Boxplots representing change in HiC read counts in the LAD domains after Lamin knock out in mESC cells. P-value was calculated using two-tailed Mann-Whitney U test. (b-c) Examples representing change in HiC read counts and contact matrices in WT and Lamin KO cells. The bottom panel of matrices represent the Lamin KO to WT fold-change in interaction frequencies.

Figure 1

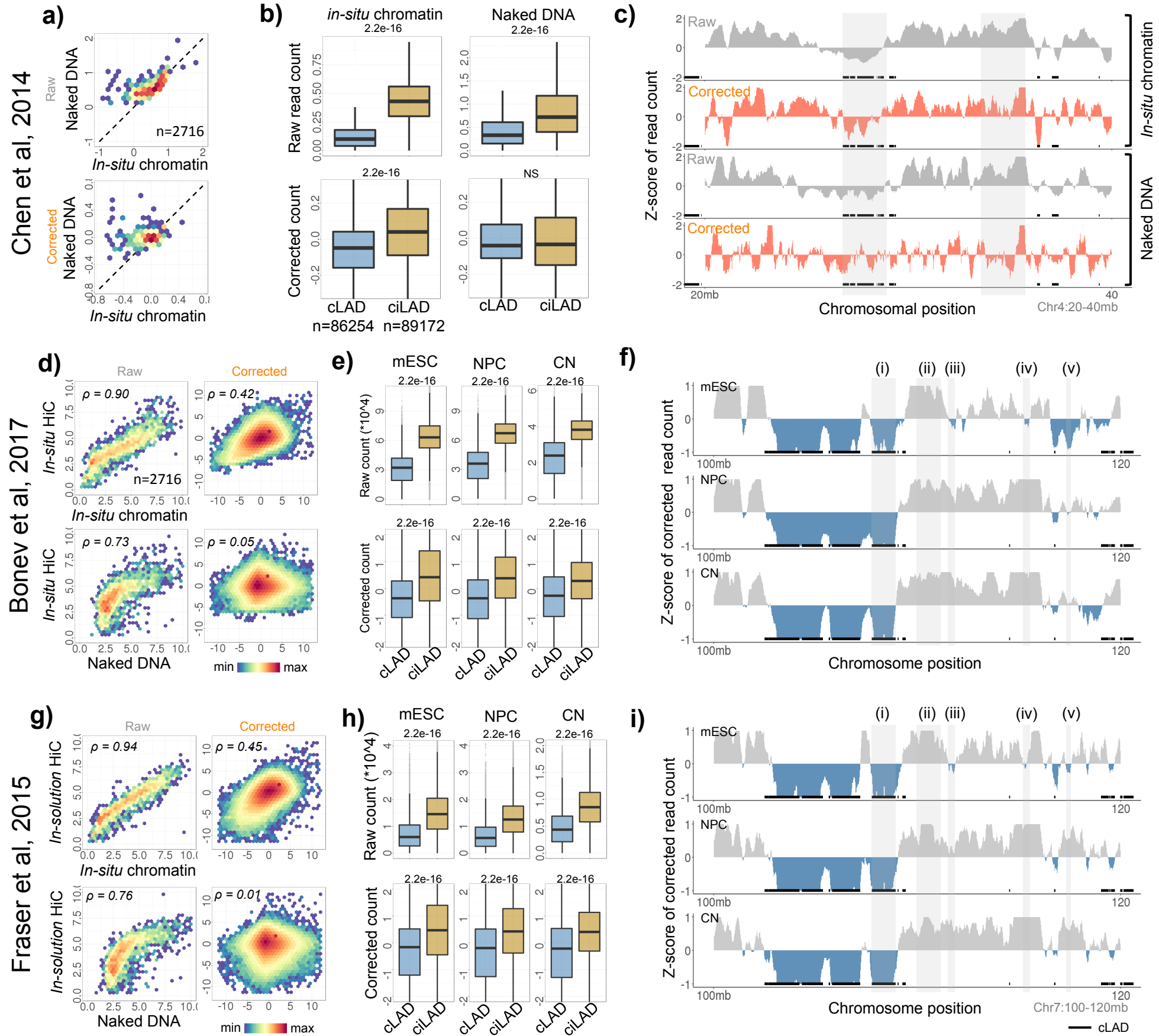

a)
Chen et al, 2014
Raw
Naked DNA
n=2716
In-situ chromatin
Corrected
Naked DNA
In-situ chromatin
b)
In-situ chromatin
Naked DNA
2.2e-16
2.2e-16
Raw read count
2.2e-16
NS
Corrected count
cLAD
ciLAD
n=86254
n=89172
c)
Raw
Corrected
Raw
Corrected
Z-score of read count
In-situ chromatin
Naked DNA
20mb
40
Chromosomal position
Chr4:20-40mb
d)
Raw
Corrected
ρ = 0.90
ρ = 0.42
In-situ HiC
n=2716
In-situ chromatin
ρ = 0.73
ρ = 0.05
In-situ HiC
Naked DNA
min
max
Bonev et al, 2017
e)
mESC
NPC
CN
2.2e-16
Raw count (*10^4)
Corrected count
cLAD
ciLAD
f)
(i)
(ii)
(iii)
(iv)
(v)
mESC
NPC
CN
Z-score of corrected read count
100mb
120
Chromosome position
g)
Raw
Corrected
ρ = 0.94
ρ = 0.45
In-solution HiC
In-situ chromatin
ρ = 0.76
ρ = 0.01
In-solution HiC
Naked DNA
min
max
Fraser et al, 2015
h)
mESC
NPC
CN
2.2e-16
Raw count (*10^4)
Corrected count
cLAD
ciLAD
i)
(i)
(ii)
(iii)
(iv)
(v)
mESC
NPC
CN
Z-score of corrected read count
100mb
120
Chromosome position
Chr7:100-120mb
cLAD

Figure 2

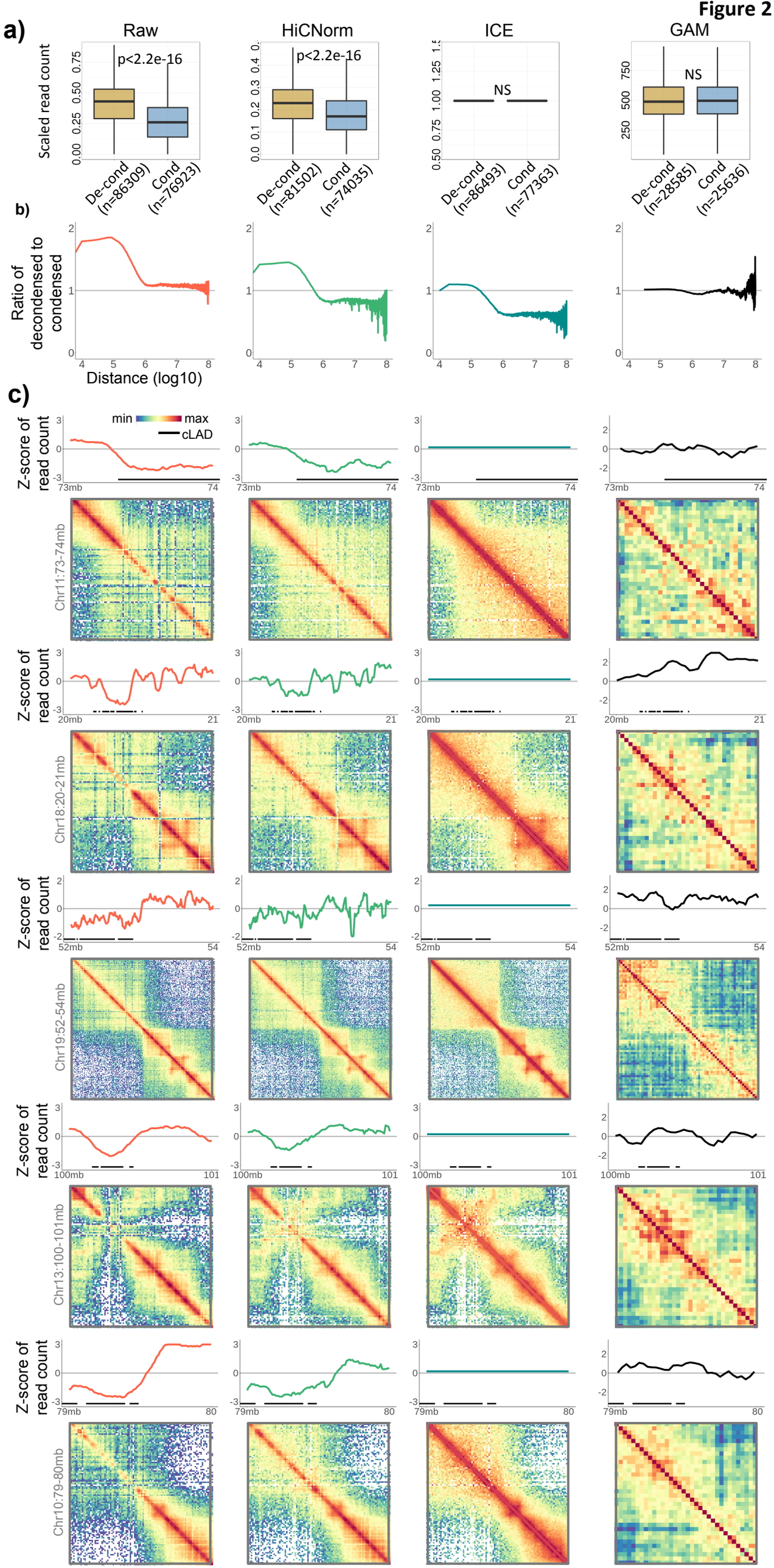

a)
Raw
HiCNorm
ICE
GAM
Scaled read count
p<2.2e-16
p<2.2e-16
NS
NS
De-cond (n=86309)
Cond (n=76923)
De-cond (n=81502)
Cond (n=74035)
De-cond (n=86493)
Cond (n=77363)
De-cond (n=28585)
Cond (n=25636)
b)
Ratio of decondensed to condensed
Distance (log10)
c)
min
max
cLAD
Z-score of read count
Chr11:73-74mb
Chr18:20-21mb
Chr19:52-54mb
Chr13:100-101mb
Chr10:79-80mb

Figure 3

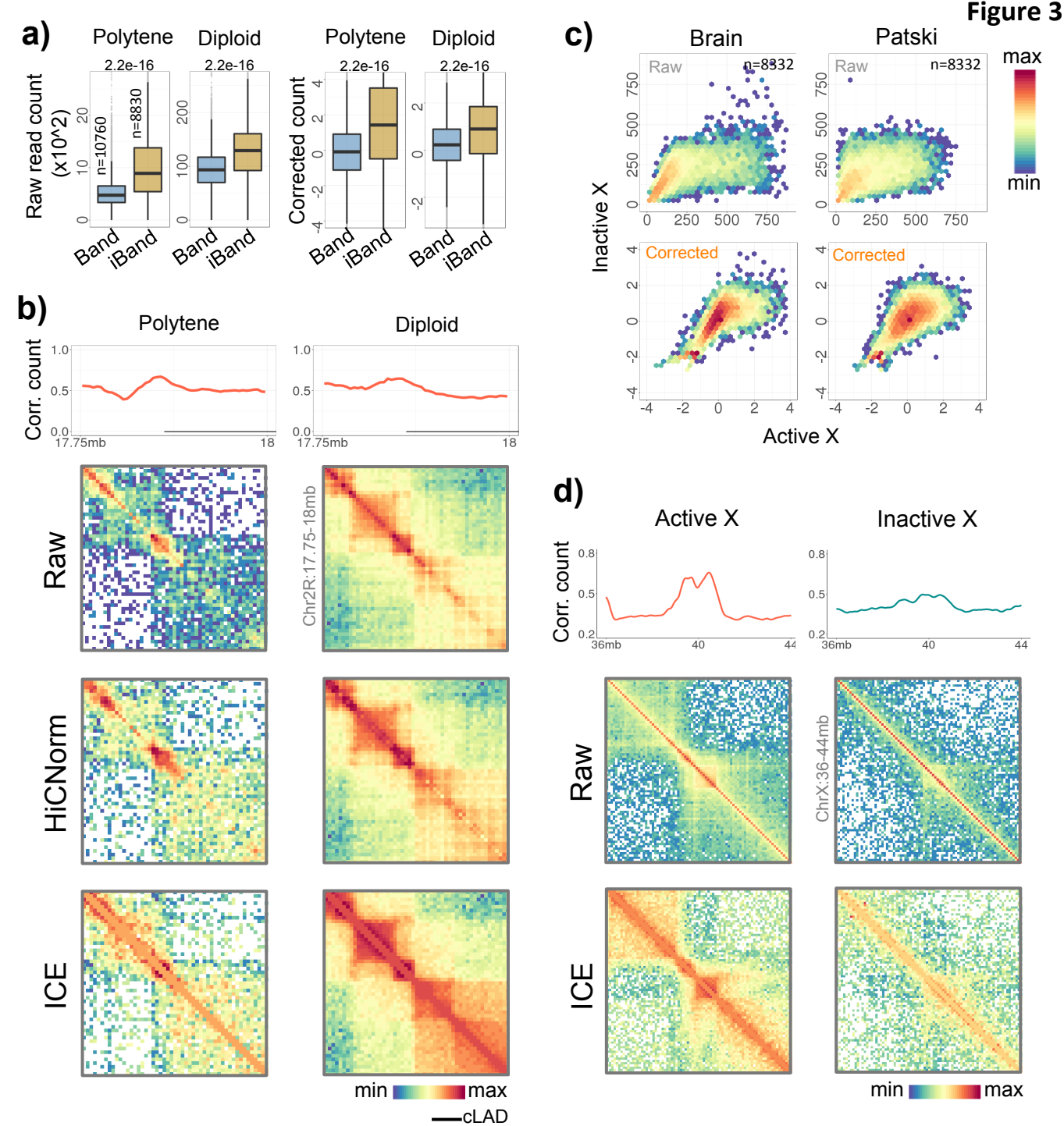

a)
Polytene
Diploid
Polytene
Diploid
2.2e-16
2.2e-16
2.2e-16
2.2e-16
Raw read count (x10^2)
n=10760
n=8830
Corrected count
Band
iBand
b)
Polytene
Diploid
Corr. count
17.75mb
18
Raw
HiCNorm
ICE
Chr2R:17.75-18mb
min
max
cLAD
c)
Brain
Patski
Raw
n=8332
Corrected
Inactive X
Active X
max
min
d)
Active X
Inactive X
Corr. count
36mb
40
44
Raw
ICE
ChrX:36-44mb
min
max

Figure 4

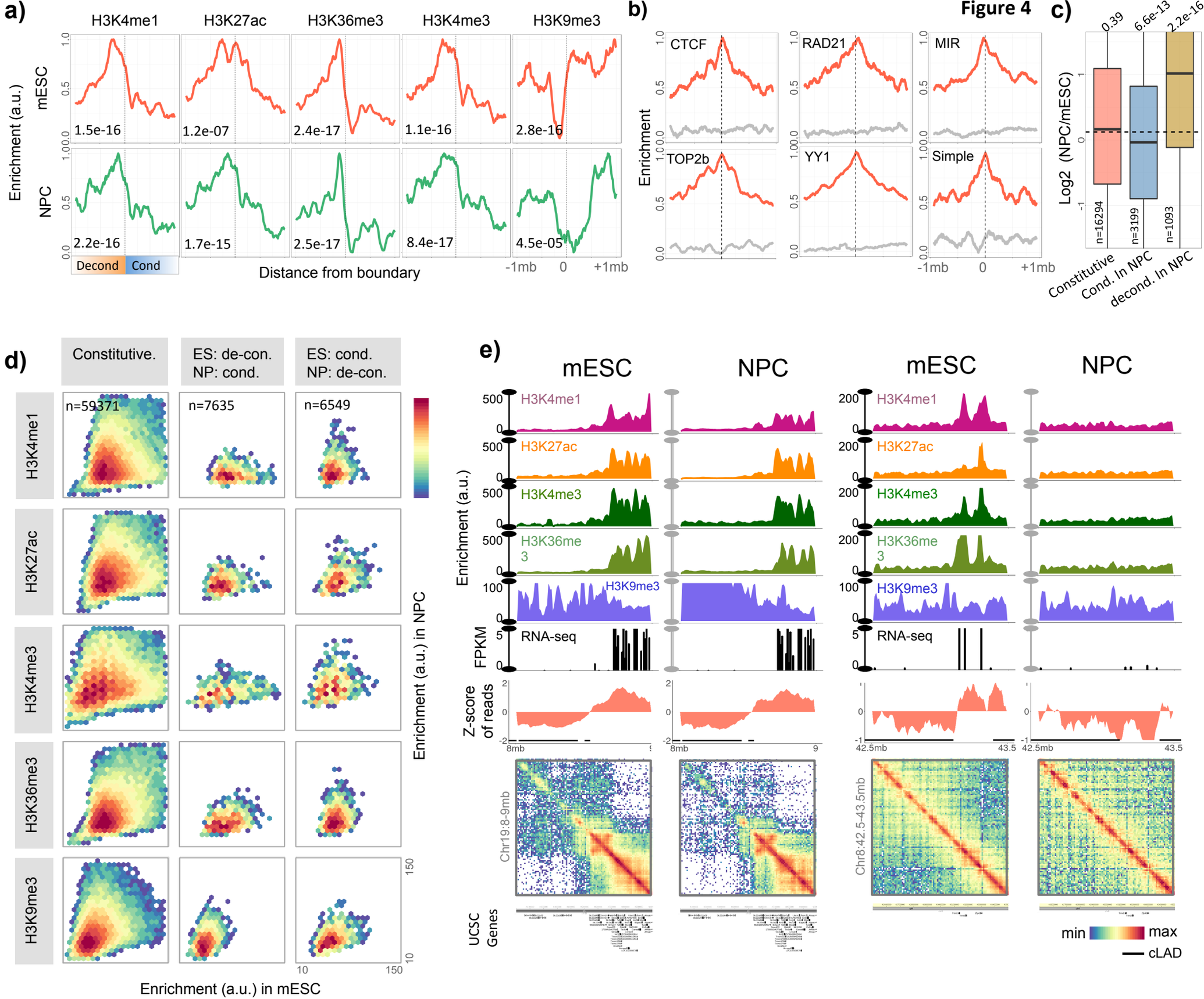
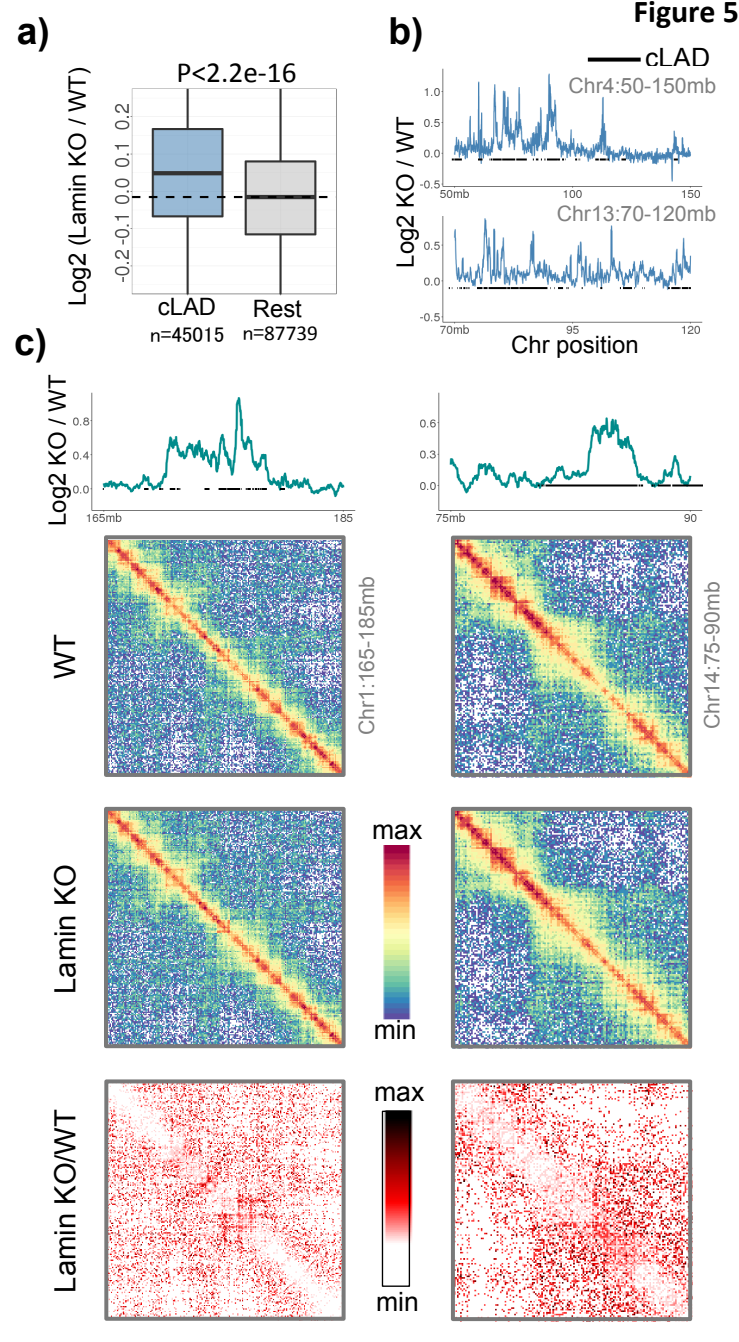

a)
H3K4me1
H3K27ac
H3K36me3
H3K4me3
H3K9me3
Enrichment (a.u.)
mESC
NPC
1.5e-16
1.2e-07
2.4e-17
1.1e-16
2.8e-16
2.2e-16
1.7e-15
2.5e-17
8.4e-17
4.5e-05
Decond
Cond
Distance from boundary
-1mb
0
+1mb
b)
CTCF
RAD21
MIR
TOP2b
YY1
Simple
Enrichment
c)
0.39
6.6e-13
2.2e-16
Log2 (NPC/mESC)
n=16294
n=3199
n=1093
Constitutive
Cond. In NPC
decond. In NPC
d)
Constitutive.
ES: de-con.
NP: cond.
ES: cond.
NP: de-con.
n=59371
n=7635
n=6549
H3K4me1
H3K27ac
H3K4me3
H3K36me3
H3K9me3
Enrichment (a.u.) in NPC
Enrichment (a.u.) in mESC
e)
mESC
NPC
mESC
NPC
H3K4me1
H3K27ac
H3K4me3
H3K36me 3
H3K9me3
RNA-seq
Enrichment (a.u.)
FPKM
Z-score of reads
8mb
9
42.5mb
43.5
Chr19:8-9mb
Chr8:42.5-43.5mb
UCSC Genes
min
max
cLAD

Figure 5


a)
P<2.2e-16
Log2 (Lamin KO / WT)
-0.2 -0.1 0.0 0.1 0.2
cLAD
Rest
n=45015
n=87739
b)
cLAD
Chr4:50-150mb
Chr13:70-120mb
Log2 KO / WT
1.0
0.5
0.0
-0.5
50mb
100
150
70mb
95
120
Chr position
c)
Log2 KO / WT
0.8
0.4
0.0
165mb
185
0.6
0.3
0.0
75mb
90
WT
Lamin KO
Lamin KO/WT
Chr1:165-185mb
Chr14:75-90mb
max
min
max
min