

Euchromatin forms condensed domains with short active regions on the surface

Corresponding Author: Dr Bin Zhang

Version 0:

Decision Letter:

17th Nov 2025

Dear Bin,

Your Article, entitled "Euchromatin forms condensed domains with short active regions on the surface", has now been seen by 3 referees. You will see from their comments below that while they find your work of interest, some important points are raised. We are interested in the possibility of publishing your study in Nature Genetics, but would like to consider your response to these concerns in the form of a revised manuscript before we make a final decision on publication.

We invite you to revise your manuscript taking into account all reviewer comments. Please highlight all changes in the manuscript text file. At this stage we will need you to upload a copy of the manuscript in MS Word .docx or similar editable format.

We are committed to providing a fair and constructive peer-review process. Do not hesitate to contact me if there are specific requests from the reviewers that you believe are technically impossible or unlikely to yield a meaningful outcome.

When revising your manuscript:

*1) Include a "Response to referees" document detailing, point-by-point, how you addressed each referee comment. If no action was taken to address a point, you must provide a compelling argument. This response will be sent back to the referees along with the revised manuscript.

*2) If you have not done so already please begin to revise your manuscript so that it conforms to our Article format instructions, available

[here](http://www.nature.com/ng/authors/article_types/index.html).

Refer also to any guidelines provided in this letter.

*3) Include a revised version of any required Reporting Summary: <https://www.nature.com/documents/nr-reporting-summary.pdf>

It will be available to referees (and, potentially, statisticians) to aid in their evaluation if the manuscript goes back for peer review.

A revised checklist is essential for re-review of the paper.

Please be aware of our [guidelines](https://www.nature.com/nature-research/editorial-policies/image-integrity) on digital image standards.

EXTENDED DATA FIGURES

When re-submitting your manuscript, please ensure that any supplementary figures and tables that are crucial to the manuscript's conclusions are converted into Extended Data figures and tables to increase visibility of these data. Extended Data figures and tables are online-only (present in the online PDF and full-text HTML versions of the paper), peer-reviewed display items that provide essential background to the article but are not included in the main article due to space constraints. A maximum of ten Extended Data display items (figures and tables) is permitted.

Please use the link below to submit your revised manuscript and related files:

Link Redacted

Note: This URL links to your confidential home page and associated information about manuscripts you may have submitted, or that you are reviewing for us. If you wish to forward this email to co-authors, please delete the link to your homepage.

Please do not hesitate to contact me if you have any questions or would like to discuss these revisions further.

Nature Genetics is committed to improving transparency in authorship. As part of our efforts in this direction, we are now requesting that all authors identified as 'corresponding author' on published papers create and link their Open Researcher and Contributor Identifier (ORCID) with their account on the Manuscript Tracking System (MTS), prior to acceptance. ORCID helps the scientific community achieve unambiguous attribution of all scholarly contributions. You can create and link your ORCID from the home page of the MTS by clicking on 'Modify my Springer Nature account'. For more information please visit please visit www.springernature.com/orcid.

We look forward to seeing the revised manuscript and thank you for the opportunity to review your work.

Sincerely,

Tiago

Tiago Faial, PhD
Chief Editor
Nature Genetics
<https://orcid.org/0000-0003-0864-1200>

Reviewers' Comments:

Reviewer #1 (Remarks to the Author):

Paggi and Zhang set out to leverage the nucleosome scale resolution of RCMC data, interpreted through polymer models, to improve our understanding of the 3D genome organization. In the process, they introduce a revised framework for normalizing RCMC and Micro-C data. From this modeling, they propose that euchromatin and heterochromatin have similar densities at the nanoscale, with euchromatin having more frequent short linker regions that tend to appear on the surface of the globular interface created by the more compacted regions. This analysis is supported by prior (rather lower resolution) microscopy data. The explanation offers a structural interpretation to the interesting patterns of domains that emerge at the nucleosome scale in the RCMC data which are not apparent in earlier Hi-C data. The argument for a correction of the normalization approach appears solid, and after some minor issues are addressed I think this will be also a very important consideration for the field going forward in accurate interpretation of micro-C and RCMC contact patterns – techniques that are seeing increasing use. Collectively, the observations in this work have substantial implications for the understanding of 3D genome organization and will likely be of broad interest to the community.

I. Primary concern

I think a little more clarity on how the authors plan to use the model - specifically whether it should be interpreted in a mechanistic sense to explain the 3D structures (why the DNA folds this way), or purely in a structural sense (a prediction of what the underlying 3D polymer ensemble behind these maps look like). Both approaches are used in the field, and if the reader misunderstands how the model should and should not be interpreted, it may create substantial problems.

For example - the authors show their model recapitulates many of the features (e.g. TADs) of the Sox2 locus from Hafner et al. It does not appear however that the model has any explicit representation of loop-extruding cohesin, or CTCF, and in that work there is also ORCA data from the cohesin and CTCF depleted cells that substantially alter the 3D distances across the locus, patterns which those authors argued were largely consistent with a simulation based on loop extrusion and extrusion blocking. Are we to interpret by the agreement to the domain patterns with a non-extrusion model the current authors are arguing for a mechanism other than loop extrusion to explain the structures? (This reviewer and I expect many readers would find this a problematic argument in need of deeper evidence, given available data supporting the role of cohesin loop extrusion in forming these structures). Or should we understand only that the aim is to get a nanoscale structural view of loci? This would circumvent the problematic implications for our understanding of cohesin function, but in the current text, is not clearly presented.

This is also relevant for how we should interpret the models' agreement with the RCMC data. RCMC reveals many patterns that are seen in Hi-C, and some new patterns of small boxes-at-the-diagonal and dots that are not seen in Hi-C. Whereas the patterns common to these two methods (e.g. TADs and CTCF loops) have also been validated by microscopy, and moreover have been linked to physical explanations and to biological behaviors (e.g. TADs arise from loop extrusion, and they enrich contacts within the TAD, including any E-P interactions within the TAD, without enriching interactions between TADs) – the additional boxes in micro-C don't have to my knowledge a clear acknowledged biophysical origin (e.g. they

don't line up with a common epigenetic mark, they appear largely to be the gaps between the accessible chromatin), nor do they have any association to my knowledge with a biological function (E-P elements in different boxes-at-the-diagonal which are unique to micro-C/RCMC have not to my knowledge been shown to have any 'insulation effect'). The authors current model could be interpreted as a mechanistic origin of these patterns based on density packing considerations resulting from nucleosome spacing. However, mechanistic interpretations of the model are again a bit problematic if they are contradicted by other mechanistic experiments like the cohesin depletion or CTCF depletion discussed above.

The mechanistic vs. ad hoc structural nature of the model would also be clearer if the authors would discuss a little more in the main text about how the energies are chosen and if they are specified on an all-to-all monomer-to-monomer basis or by a smaller parameter space.

II. Technical concerns and suggestions

1. Can the authors validate their normalization method on toy data from a simulation in which there are dramatic variations in the polymer density? A prime conclusion of the work is that euchromatin and heterochromatin have similar densities. Yet the authors have also shown how the normalization of ICE amounts to an a priori assumption of uniform density, and their revised method is a modest adjustment to ICE. The recommended simulation could help test if the normalization had a tendency to diminish true variation.

2. Can the authors explore / validate the effect of the observation window size on the inferred density distribution of the polymer after normalization?

Given the importance of the balancing method to the paper as a whole, some discussion of the effect of applying this on a *local region* as opposed to a full length polymer might help justify the method better. In the original work justifying the balancing approach in the first place, it was noted that as all parts of the polymer are present in equal abundance, all parts should be near something else if everything is measured. If the analysis considers only part of a polymer - for example a portion which contains a large self interacting domain, and an adjacent region that extrudes from this region and interacts with a part of the polymer not sampled, the contact map will lack reads for the extruded part. From the perspective of ICE, this is just like an undersampled (underamplified region), and its contacts with the rest of the polymer will be exaggerated. If a different size region were used, one which captured the distal part of the polymer in which our hypothetical local extruded domain is buried, this over-correction would not occur. Neighboring balancing does not seem to correct for this issue of ICE in the context of RCMC, but it seems problematic if technical differences (in the choice of region sampled) result in different answers for the underlying biophysical realities.

3. Could the authors comment on the limitations or implicit assumptions in adopting a maximum entropy approach, with respect to its impact on the types of interactions and structures that arise in the resulting polymer ensembles?

It seems any method that tries to generate an ensemble from knowing only the average has some implicit bias to certain types of ensembles (as well as a degree of solving a properly indeterminate problem, and making something from nothing). The maximum entropy method for inferring an ensemble of 3D structures seems to be affected by this too, and it would help to discuss the limitation. For example: Consider a hypothetical polymer ensemble in which 3 monomers are special (lets call them A, B, C, and lets assume they are spaced out along the polymer and in that order. A and C both compete to bind B, but only one can bind at a time, and also A can bind C and exclude B. The sticky binding interactions are weak enough that they are regularly disrupted by the thermal motion of the polymer, giving the other bond a chance to form next. Now consider a different hypothetical ensemble, in which A binding B increases the probability of also binding C in a 3 way hub, though all the individual pairwise binding events are weaker/ rarer. With the right choice of affinities, these ensembles may have identical pairwise contact maps (e.g. RCMC data), despite their different interaction rules and different physical 3D structures. A maximum entropy method - or any approach of producing an ensemble from this average, can not get both ensembles correct, and must through its inherent assumptions favor certain types of interactions (perhaps a lack of sensitivity to recognize competitive interactions?)

4. The authors may wish to use the 5 kb resolution Sox2 data from Mateo 2019 or better the deeper data from Ren 2021 and Ren 2025.

III. Text suggestions:

1. The introduction or discussion could be bolstered with some further references on prior work measuring local variation in chromatin density on the scales of 10s of kb to 1 Mb.

- For example, Boettiger et al Nature 2016, report some interesting length dependent and euchromatin/heterochromatin dependent variation in chromatin densities using STORM that I think connect well to the conclusions from this work.

- From the intro: ". However, recent evidence challenges this dichotomy, suggesting that euchromatin also contains largely condensed domains [1–4]." Similarly, some evidence suggests heterochromatin is not distinctly 'condensed' – Legant ~2023, Murphy Nat Gen 2024.

- It could also be interesting to compare the organization of such heterochromatin regions from micro-C if sufficient resolution exists. If the best available mESC micro-C data lacks the needed resolution and the available RCMC data lacks coverage of Polycomb heterochromatin domains, the Drosophila Micro-C data from the Levine lab has very high coverage owing to the small genome and deep sequencing, and might be a very informative place to apply this approach to varying types of heterochromatin, or compare the inferred organization with that from microscopy.

2. Line 110 " In MEI, one begins with a baseline model and then adds auxiliary energy terms to make the system recapitulate experimental data (see Methods)." I appreciate the authors choice to not clutter the text with excessive technical detail here.

However, this 1 line explanation I think is not so useful, and offers no explanation of what I see as the fundamental process – developing an ensemble of distinct states that will average to contact map (presumably in a maximal entropy sense). It would help to give the reader a brief sense of the baseline model (e.g. is it a simple crumpled polymer, or a simple freely jointed chain, or is it a polymer whose intra-monomer interaction energies are defined by the RCMC heatmap?

3. I think it would clarify the presentation to clarify what is meant by the 'no free parameters' -> I take it the different contact maps (per region, or of the same region in different perturbations) are considered inputs, not parameters? This issue is potentially confusing as the loop-extrusion approach has been described as a minimal set of parameters, where the aim is to explain and reproduce the contact map from polymer properties that make minimal reference to the contact map as input (instead using non 3D structural data, like CTCF ChIP-seq and CTCF site directions, as input.

Reviewer #1 (Remarks on code availability):

The neighbor-balance code appears to be complete and have detailed documentation.

The maximum entropy inversion code was not provided and could not be assessed.

Reviewer #2 (Remarks to the Author):

This manuscript discusses a novel computational framework to reconstruct chromatin 3D structures at nucleosome resolution using Region-Capture Micro-C (RCMC) data combined with a method of maximum entropy inversion (MEI), exploiting "neighbor balancing" to correct biases in contact maps while preserving biological variation.

By use of the new method, the authors show that euchromatin, traditionally viewed as generically "open", actually folds into compact "packing domains" containing dense nucleosome clutches. They also show that kilobase-scale active promoters and enhancers tend to protrude outward from these domains, making them accessible for protein binding and enhancer–promoter communication. Their simulations reproduce experimental contact maps, distance distributions from chromatin tracing, and structural motifs from imaging, thus providing a broad, multiscale view of chromatin organization. Overall, the manuscript concludes that euchromatin is not just "open", but rather composed of condensed domains with local accessible regions that spatially compartmentalize regulatory elements. This picture does clarify the current simplified view of chromatin organisation.

Recent imaging studies have suggested that euchromatin is more compact than previously believed, and that chromatin is organized into clutches and nanodomains. However, these methods lacked locus-specific resolution. This manuscript advances the field by providing nucleosome-resolution conformation exploration; by bridging structural observations across scales; and returning a plausible mechanistic model where condensed packing domains coexist with accessible regulatory protrusions.

The manuscript includes interesting methodological innovations. The introduction of "neighbor balancing" does offer an important improvement over standard ICE normalization, recovering meaningful biological variation in contact density. The integration of RCMC with MEI achieves nucleosome-level resolution, to help bridging population-averaged contact maps with single-cell imaging observations. Additionally, the model is validated against different experimental data, including ORCA chromatin tracing and super-resolution imaging, which strengthens confidence in its results. Finally, the finding that active enhancers and promoters tend to protrude from otherwise compact domains provides a plausible structural explanation for their accessibility and communication despite local chromatin compaction.

Nevertheless, the author could try to address some weak points. While their simulations reproduce experimental data well, there is limited experimental confirmation (e.g., imaging of specific protrusions corresponding to predicted enhancer/promoter sites). The study is restricted to mouse embryonic stem cells; whether the observed euchromatin architecture generalizes to differentiated cell types remains unanswered. The modeling depends on a broad set of parameters and assumptions such as nucleosome occupancy and capture radius; it'd be very important to assess the robustness of the findings and conclusions. Finally, direct links to gene expression outcomes or transcriptional activity are inferred rather than demonstrated.

Overall, this is a well-executed, technically valid and novel manuscript that makes a valuable contribution to the understanding of chromatin architecture. The main conclusions are well supported, though the manuscript would definitely benefit from additional analyses to address the weak points highlighted above.

Reviewer #3 (Remarks to the Author):

In the paper by Paggi and Zhang entitled "Euchromatin forms condensed domains with short active regions on the surface", the authors introduce a new computational tool that aims at simulating chromatin at "nucleosome" resolution taking as consideration micro-C datasets. This new approach is accompanied by a new micro-C normalization procedure (neighbor

balancing) that takes into consideration local packing of chromatin as observed by micro-C. Next, the authors validate the models comparing them to mainly microscopy-based data to finally apply it to a series of genomic regions. Finally, the authors provide a structural analysis of the models to propose that euchromatin is formed by dense groups of continuous “clutches” and decorated by open, less dense and protruding DNA containing promoters and enhancers. The authors conclude that this spatial organization allows to understand how promoters and enhancers can act at distance providing additional mechanisms to the long studied “loop extrusion” model. Overall, this is a well written manuscript with novel computational approaches that, when applied, provide substantial biological insights. However, the manuscript could benefit for further work specially in validating the models, making clear the limitations of the computational approaches, and further characterizing how the data and parameters affect the simulations. Next are some more specifics placed in order of relevance to this reviewer:

- To really complement the data, I would suggest the authors simulate a very dense heterochromatin region with little or no ATAC-seq peaks. Do the boundaries (protruded regions) still appear? If so, why?
- There is a need to more rigorously quantify that protruding regions contain more than expected regulatory elements.
- I appreciate the neighbor balancing approach proposed here. It apparently can “detect” ATAC signal in the micro-C (otherwise, the models would not show correlation to ATAC peaks). Can the authors elaborate why this is the case? How that can be captured by micro-C and how the new balancing does not remove this data?
- In Figure 3b/d is clear that the resulting models are not at the same scale than the imaging data. This does not seem to be the case when the data is plotted differently in Figures 3a/c/e. Why?
- Contrary to the statement in the text (page 7), the underestimation of distances in the models is also for long genomic separations.
- Similarly, can the authors demonstrate that the choice of contact indicator function affects only the scale of the models and not its topology/conformation?
- In Figure 2, the models clearly show a zig-zag organization of the beads. This usually happens when the size of the beads is inconsistent (too large) with the (too small) imposed restraints during simulation between consecutive beads. Is this a feature seen in all models? If so, do the authors consider re-visiting such parameters? (BTW, the code for simulation was not available to the reviewers).
- I am not sure I would call these models “nucleosome” resolution. To certain extent this is misleading as the beads do not explicitly represent nucleosome position. The models are 100s bp resolution indeed but do not take any data in consideration as to nucleosome positioning. This would separate the approach from other more explicit simulations (coarse-grained DNA simulations) that indeed simulate nucleosomes.
- There is an entire section on how to avoid knots in the models. Interestingly, most modeling methods that use Hi-C or related datasets result in no-knots in their models even though there are no specific restraints avoiding them. Why you see knots in your models? can those be avoided without explicitly using connecting beads or extreme beads?
- Can the authors quantify the effects on the results and conclusions of the two main parameters for DBSCAN?
- Similarly, can the authors quantify the “appears” in the sentence: “The domains identified by DBSCAN appear to correspond to experimentally observed packing (or TAD-like) domains”?
- From the models, could the authors try to obtain simulated image data to directly compare to the results in Ricci et al?
- Why 25.6kb resolution was used for balancing micro-C? Any reasoning?
- The authors could also place their work in context to the recent BioRxiv paper from the Maeshima Lab (<https://www.biorxiv.org/content/10.1101/2025.08.27.672592v1>).
- The abstract may need to be simplified down to the key points.
- I would suggest the authors explain briefly the “condensability score” (page 3) for the readers who are not familiar with the topic and do not want to read the reference #54.
- I would suggest using kde plots for Figure 4h/i.

Reviewer #3 (Remarks on code availability):

The main code is not available. Only partial code is and was not reviewed as somehow trivial code.

Version 1:

Decision Letter:

Our ref: NG-A69939R

31st Mar 2026

Dear Bin,

Thank you for submitting your revised manuscript entitled “Euchromatin forms condensed domains with short active regions on the surface” (NG-A69939R). It has now been seen by the original referees and their comments are below. The reviewers find that the paper has improved in revision, and therefore we'll be happy in principle to publish it in Nature Genetics, pending minor revisions to comply with our editorial and formatting guidelines.

The current version of your manuscript is in a PDF format. Please email us a copy of the file in an editable format (Microsoft Word) - we can not proceed with PDFs at this stage.

We will then be performing detailed checks on your paper and will send you a checklist detailing our editorial and formatting

requirements soon. Please do not upload the final materials and make any revisions until you receive this additional information from us.

Thank you again for your interest in Nature Genetics. Please do not hesitate to contact me if you have any questions.

Congratulations!

Sincerely,

Tiago

Tiago Faial, PhD
Chief Editor
Nature Genetics
<https://orcid.org/0000-0003-0864-1200>

Reviewer #1 (Remarks to the Author):

The authors have addressed my concerns, I support publication.



Reviewer #2 (Remarks to the Author):

The authors have answered the remarks raised by the Reviewers.

Reviewer #3 (Remarks to the Author):

I do appreciate the effort made by the authors to address my comments. The manuscript is, in my opinion, stronger now.

Open Access This Peer Review File is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

In cases where reviewers are anonymous, credit should be given to 'Anonymous Referee' and the source.

The images or other third party material in this Peer Review File are included in the article's Creative Commons license,

unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

Title: *Euchromatin forms condensed domains with short active regions on the surface*

Manuscript ID: NG-A69939

Authors: Joseph M. Paggi, Lawrence Y. Long, Bin Zhang

We thank the reviewers for their support and insightful commentary, which guided the revisions we made to our manuscript. We've made significant additions, including new simulations for HCT116 cells supporting the generality of our proposed euchromatin architecture, new analysis supporting that our conclusions are robust to changes in model parameters, and text changes to better place our work in the literature.

In this letter, we respond to each of the reviewers' comments. We support responses by referring to the revised manuscript and, where necessary, by including additional descriptions and analyses that we consider extrinsic to the manuscript itself. In this letter, all changes to the manuscript are shown in blue. As noted, the displayed figures have been added to the manuscript or the SI. We believe that we have successfully addressed all comments and that the quality of our paper has improved significantly.

We also updated a [GitHub repository](#) with all data and code used in this study to promote transparency and reproducibility.

RESPONSE TO REFEREE 1

Summary: Paggi and Zhang set out to leverage the nucleosome scale resolution of RCMC data, interpreted through polymer models, to improve our understanding of the 3D genome organization. In the process, they introduce a revised framework for normalizing RCMC and Micro-C data. From this modeling, they propose that euchromatin and heterochromatin have similar densities at the nanoscale, with euchromatin having more frequent short linker regions that tend to appear on the surface of the globular interface created by the more compacted regions. This analysis is supported by prior (rather lower resolution) microscopy data. The explanation offers a structural interpretation to the interesting patterns of domains that emerge at the nucleosome scale in the RCMC data which are not apparent in earlier Hi-C data. The argument for a correction of the normalization approach appears solid, and after some minor issues are addressed I think this will be also a very important consideration for the field going forward in accurate interpretation of micro-C and RCMC contact patterns – techniques that are seeing increasing use. Collectively, the observations in this work have substantial implications for the understanding of 3D genome organization and will likely be of broad interest to the community.

Response: We appreciate the reviewer's positive comments and thoughtful critique of our manuscript. Based on the reviewer's comments, we've made substantial efforts to clarify our modeling procedure and implicit assumptions. We've also performed additional simulations and analysis to characterize polycomb-repressed regions and compare to imaging studies.

I. Primary concern I think a little more clarity on how the authors plan to use the model - specifically whether it should be interpreted in a mechanistic sense to explain the 3D structures (why the DNA folds this way), or purely in a structural sense (a prediction of what the underlying 3D polymer ensemble behind these maps look like). Both approaches are used in the field, and if the reader misunderstands how the model should and should not be interpreted, it may create substantial problems.

For example - the authors show their model recapitulates many of the features (e.g. TADs) of the Sox2 locus from Hafner et al. It does not appear however that the model has any explicit representation of loop-extruding cohesin, or CTCF, and in that work there is also ORCA data from the cohesin and CTCF depleted cells that substantially alter the 3D distances across the locus, patterns which those authors argued were largely consistent with a simulation based on loop extrusion and extrusion blocking. Are we to interpret by the agreement to the domain patterns with a non-extrusion model the current authors are arguing for a mechanism other than loop extrusion to explain the structures? (This reviewer and I expect many readers would find this a problematic argument in need of deeper evidence, given available data supporting the role of cohesin loop extrusion in forming these structures). Or should we understand only that the aim is to get a nanoscale structural view of loci? This would circumvent the problematic implications for our understanding of cohesin function, but in the current text, is not clearly presented.

This is also relevant for how we should interpret the models' agreement with the RCMC data. RCMC reveals many patterns that are seen in Hi-C, and some new patterns of small boxes-at-the-

diagonal and dots that are not seen in Hi-C. Whereas the patterns common to these two methods (e.g. TADs and CTCF loops) have also been validated by microscopy, and moreover have been linked to physical explanations and to biological behaviors (e.g. TADs arise from loop extrusion, and they enrich contacts within the TAD, including any E-P interactions within the TAD, without enriching interactions between TADs) – the additional boxes in micro-C don't have to my knowledge a clear acknowledged biophysical origin (e.g. they don't line up with a common epigenetic mark, they appear largely to be the gaps between the accessible chromatin), nor do they have any association to my knowledge with a biological function (E-P elements in different boxes-at-the-diagonal which are unique to micro-C/RCMC have not to my knowledge been shown to have any 'insulation effect'). The authors current model could be interpreted as a mechanistic origin of these patterns based on density packing considerations resulting from nucleosome spacing. However, mechanistic interpretations of the model are again a bit problematic if they are contradicted by other mechanistic experiments like the cohesin depletion or CTCF depletion discussed above.

The mechanistic vs. ad hoc structural nature of the model would also be clearer if the authors would discuss a little more in the main text about how the energies are chosen and if they are specified on an all-to-all monomer-to-monomer basis or by a smaller parameter space.

Response: Thank you for pointing out that the mechanistic implications of our modeling were unclearly presented. Our model is intended to be interpreted strictly as a structure inference model, not a mechanistic one. Its goal is to gain a nano-scale structural view of the loci derived from the RCMC data (and, less directly, from chromatin tracing data).

The interaction energies are fit (one for each monomer-monomer pair) so that the simulated system reproduces the RCMC data. The learned energies are therefore a combination of all relevant biophysical phenomena, including but not limited to the effects of loop extrusion and nucleosome-nucleosome interactions. As such the agreement with RCMC data confirms that the algorithm is succeeding at its goal, but does not support a particular mechanism.

We have updated the results section to make this more clear and to specify in more depth how the energies are chosen. The updated text reads

While the RCMC contact maps and contact densities from neighbor balancing begin to suggest what distinguishes active from inactive chromatin, they do not provide the complete picture. We therefore used maximum entropy inversion (MEI) to recover a 3D conformational ensemble that together recapitulates the experimental contact map (Fig. 1) [1]. In MEI, one begins with a baseline model and then adds auxiliary energy terms that result in the system recapitulating the experimental data. The auxiliary energy terms are fit individually for each monomer-monomer pair by iteratively simulating the system, computing a simulated micro-C contact map, and updating the interaction energies according to the deviation from the micro-C contact map. This procedure is repeated until the simulated contact map is in close agreement with micro-C contact map (see Methods). The auxiliary energy terms should be viewed as an effective equilibrium approximation that captures the collective effects of chromatin-folding mechanisms, including but not limited to loop extrusion and nucleosome-nucleosome interactions.

That said, inferred structures can be combined with perturbation experiments to assess consistency with proposed mechanisms, without implying mechanistic causality from the model itself. For example, using RCMC data from cohesin-depleted cells, we find that clutch-scale structures (corresponding to small boxes of ~ 10 – 20 kb along the diagonal) remain present, whereas TAD-scale structures are disrupted (Figure R1). This observation originates from the RCMC experiments themselves, which report similar short-scale but altered large-scale organization under cohesin depletion (Figure R1c). This suggests that clutch-scale organization is not strictly dependent on cohesin and loop extrusion. Although the maximum entropy framework does not provide a mechanism for clutch formation, prior studies have implicated inter-nucleosome interactions in local compaction and accessible chromatin regions (e.g., ATAC-seq peaks) in boundary formation [2]. The precise molecular basis remains unresolved.

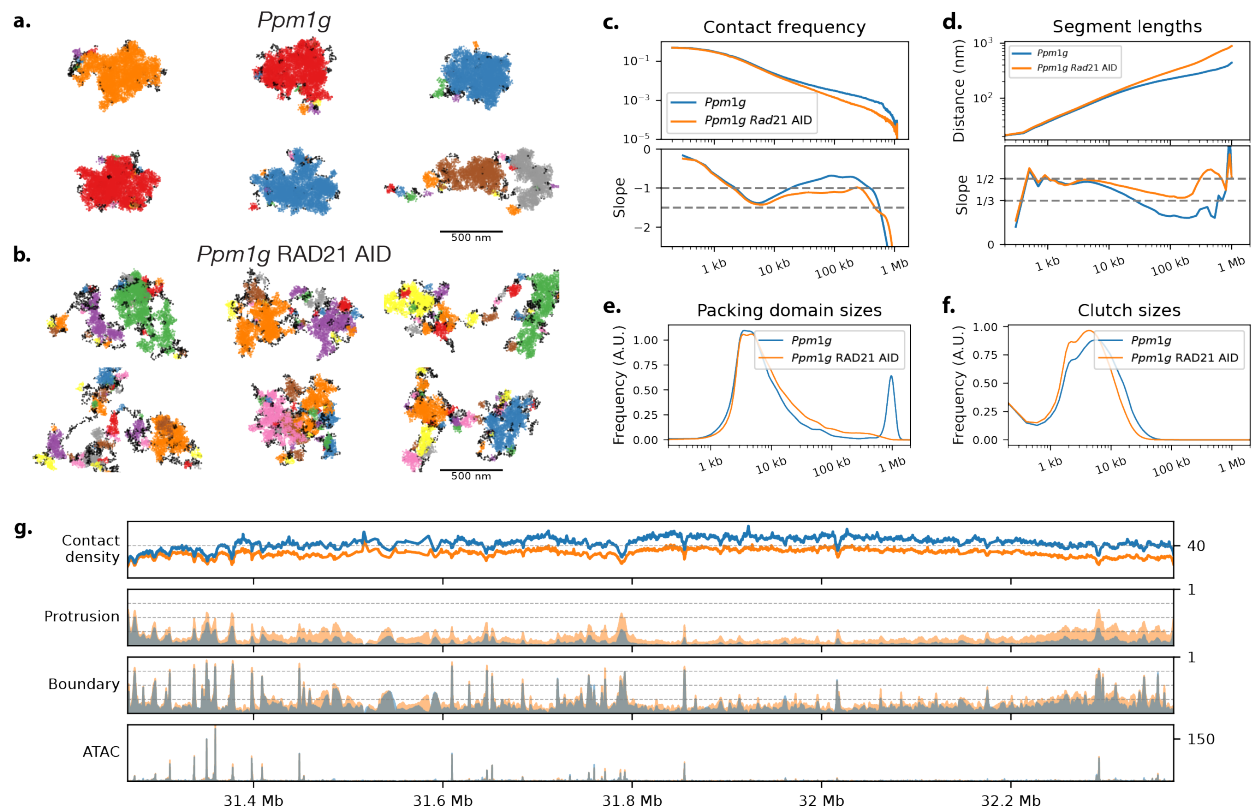


Figure R1: Upon cohesin depletion chromatin expands at large scales but local structure is maintained. Representative simulation frames for the *Ppm1g* system using (a) WT RCMC data and (b) RCMC data in which RAD21 was depleted using an auxin inducible degron (AID) [3]. (c) Neighbor balanced RCMC contact frequency scaling, (d) segment length scaling, (e) packing domain sizes, and (f) clutch sizes for the two conditions. (g) Upon cohesin depletion, contact densities, protrusion probabilities, and boundary probabilities generally decrease but remain correlated. The boundary probability peak heights change only modestly indicating preservation of microdomain scale structure. ATAC-seq for cohesin-depleted and non-treatment are shown [4].

II. Technical concerns and suggestions

1. Can the authors validate their normalization method on toy data from a simulation in which there are dramatic variations in the polymer density? A prime conclusion of the work is that euchromatin and heterochromatin have similar densities. Yet the authors have also shown how the normalization of ICE amounts to an a priori assumption of uniform density, and their revised method is a modest adjustment to ICE. The recommended simulation could help test if the normalization had a tendency to diminish true variation.

Response: The results of such a simulation are presented in Figure 1b and reproduced in [Figure R2a](#). In this simulated system, the contact density varies by approximately a factor of 4. Neighbor balancing correctly captures the true variation in contact density. To match this synthetic experiment to real-world data, we include multiplicative biases, i.e. randomly drawn weights corrupting entire rows/columns, mimicking technical effects, like MNase digestibility. While the neighbor balancing procedure is simple, the changes to the contact map are not necessarily “modest”; entries in the contact map change by a factor of 2 or more.

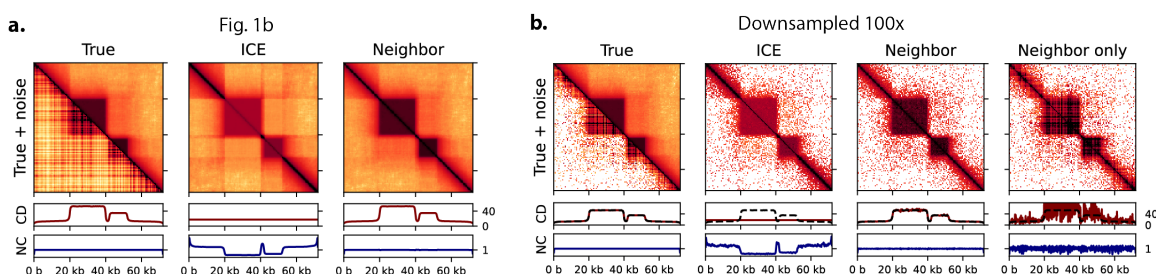


Figure R2: Neighbor balancing does not diminish variation in contact density. (a) Contact maps, contact densities (CD), and neighbor contact frequencies (NC) for a simulated test system. Before balancing, the true contact map is corrupted by multiplicate biases (lower triangle of left panel). This is a reproduction of Figure 1b. (b) Equivalent, but the read counts in the initial contact map are randomly downsampled by a factor of 100. In the column labeled “Neighbor only”, the ICE balancing step was omitted and the neighbor update was performed directly on the raw contact map (with multiplicative biases).

Neighbor balancing does not diminish true variation (at least low frequency ones, see below) because the update to each window is based only on the contact frequencies to its neighboring windows. Consider a case where there are two broad regions and the second one is denser by a factor of d . To make the contact densities uniform, ICE will systematically multiply all the contact frequencies inside the second region by a factor of $1/d$. This would include the neighbor contact frequencies, so the neighbor balancing will increase all contacts by the appropriate factor, exactly re-introducing the correct variation.

Why then do we do the ICE balancing step? The raw contact map is noisy, both due to sampling and technical effects. Performing neighbor balancing directly on the raw read counts generally fails to bring the neighbor contacts close to one and yields visually noisy contact maps and contact densities ([Figure R2b](#)). In theory, multiple neighbor updates could be performed to bring the neighbor

contacts closer to one but this is unstable because each weight depends on only two entries in the contact map. There is a failure mode where the weights alternate between high and low values (e.g. 2, 1/2, 2, ...). ICE balancing is much more robust in this regard, as each weight depends on the sum over an entire row. Adding this step makes the neighbor contact frequencies smoother, allowing neighbor balancing to successfully bring them close to one. This procedure serves to stabilize high frequency variation, but does not diminish low frequency variation. We have clarified this in the Supplementary Information.

With real data, we additionally need to smooth the contact frequencies between neighboring windows before using them to balance the contact map. Specifically, we applied gaussian smoothing with a standard deviation of 1 window. This acts to further smooth high frequency variation in contact densities—perhaps removing some genuine variation—but does not influence the overall compaction of broad euchromatin or heterochromatin regions.

*2. Can the authors explore / validate the effect of the observation window size on the inferred density distribution of the polymer after normalization? Given the importance of the balancing method to the paper as a whole, some discussion of the effect of applying this on a *local region* as opposed to a full length polymer might help justify the method better. In the original work justifying the balancing approach in the first place, it was noted that as all parts of the polymer are present in equal abundance, all parts should be near something else if everything is measured. If the analysis considers only part of a polymer - for example a portion which contains a large self interacting domain, and an adjacent region that extrudes from this region and interacts with a part of the polymer not sampled, the contact map will lack reads for the extruded part. From the perspective of ICE, this is just like an undersampled (underamplified region), and its contacts with the rest of the polymer will be exaggerated. If a different size region were used, one which captured the distal part of the polymer in which our hypothetical local extruded domain is buried, this over-correction would not occur. Neighboring balancing does not seem to correct for this issue of ICE in the context of RCMC, but it seems problematic if technical differences (in the choice of region sampled) result in different answers for the underlying biophysical realities.*

Response: Applying ICE balancing to subsets of the genome is indeed somewhat problematic (though the deviations are minor enough that it would not impact the original RCMC study's conclusions), but neighbor balancing corrects these issues. In [Figure R3a,b](#), we compute contact maps either using the full ~400 kb Nanog region and using only a 150 kb subset of the region. Entries in the ICE balanced contact map differ by less than a factor of 2, and there are no meaningful differences in the neighbor balanced contact map. In [Figure R3c](#), we compute the contact densities for these contact maps and find that if we only consider the entries in the 150 KB region, we indeed get the same contact densities in both neighbor balanced maps.

If we consider the example provided by the reviewer, ICE balancing will indeed inflate the contact frequencies in the extruded region. Critically, it will inflate the neighbor contact frequencies by the same factor, so neighbor balancing will correctly remove this deviation.

These results show that the neighbor balanced contact frequencies are invariant to cropping. However, only considering a subset of the genome naturally influences the total contact density. As such the contact densities summed over the cropped region differ from those computed using the full

map, especially near the borders of the 150 kb region (compare cyan to blue lines in [Figure R3c](#)). Analogously, some of the context is missing in our simulations, and we should imagine that the simulated system is embedded in the broader context of the nucleus. Considering whole genome micro-C data, over 75% of cis contacts are generally included in the RCMC regions (and over 50% of total contacts). Moreover, the in region contact densities are highly correlated to those computed across the whole genome ([Figure R3d](#)). B compartment regions do generally form more long range and therefore out-of-region contacts (Fig. E1a). We suspect that this corresponds to domains from distal B compartment regions coming together to form larger packing domains at a higher rate than for A compartment regions. We've added the following statement to the Discussion to point out this limitation of simulating a subset of the genome.

In our simulations, the density inside of packing domains is largely consistent between regions, despite differences in epigenetic state. This perhaps surprising result requires four clarifications. [...] [Third, B compartment regions form more long range contacts, outside of the simulated regions \[5\]](#) (Fig. E1a).

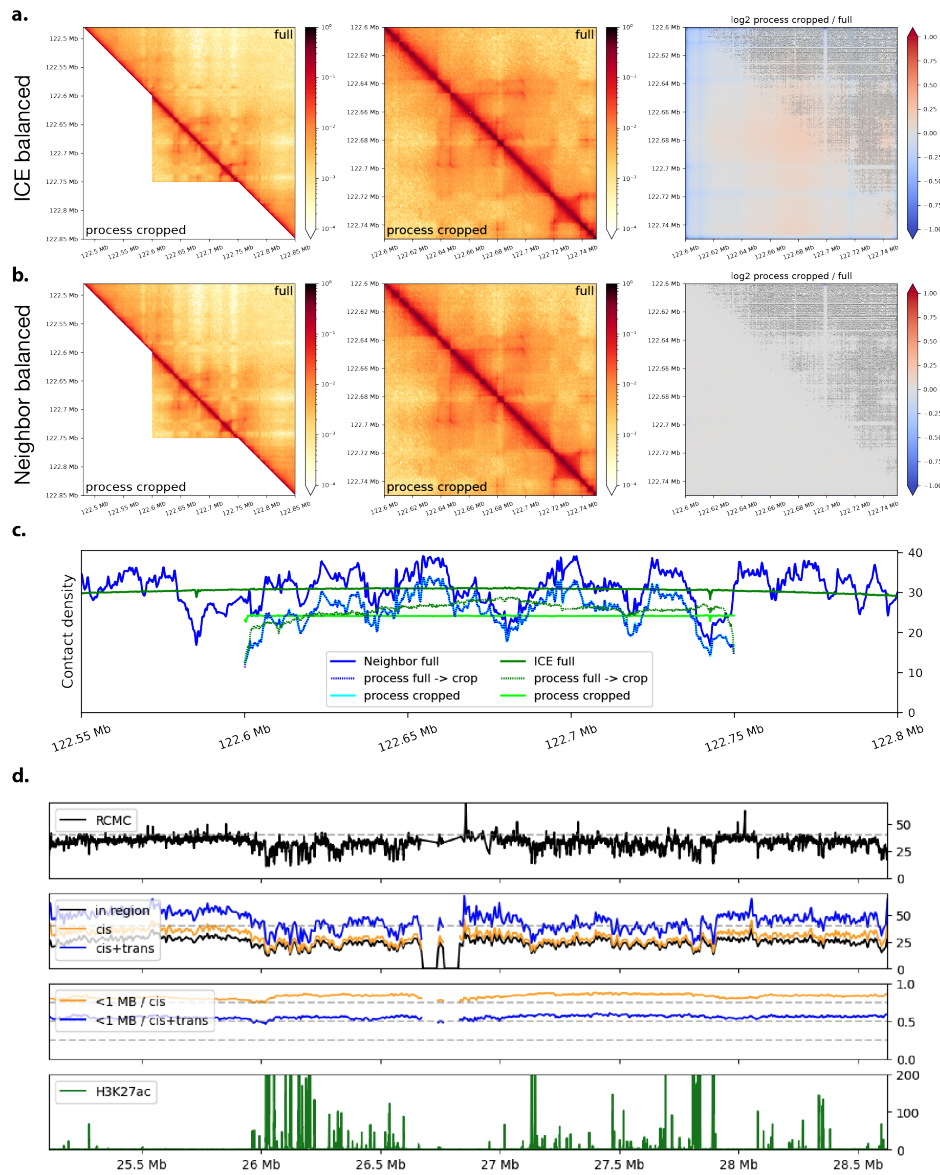


Figure R3: Considering a subset of the genome has limited impact on neighbor balanced contact maps. (a, left) ICE balanced contact maps computed using the entire the full ~ 400 kb Nanog region in the upper triangle and a contact map computed using only the shown 150 kb in the lower triangle. (a, center) A zoom in to only the the sub-region. (a, right) The log2 fold change between the contact maps computed using the full region and only the cropped region. The lower triangle is smoothed. (b) Equivalent for neighbor balanced contact maps. (c) Contact densities computed for the full map region (full), for only the cropped region but using the contact map computed with the full region (process full $\rightarrow$ crop), and for the cropped region using the contact map computed only using the cropped region. (d) Comparison of (top) contact densities computed with RCMC in HCT116 cells for region 1, (middle) contact densities computed with genome-wide micro-C considering the whole genome, only cis interactions, and only in region interactions, and (bottom) the fraction of short range interactions.

3. Could the authors comment on the limitations or implicit assumptions in adopting an maximum entropy approach, with respect to its impact on the types of interactions and structures that arise in the resulting polymer ensembles? It seems any method that tries to generate an ensemble from knowing only the average has some implicit bias to certain types of ensembles (as well as a degree of solving a properly indeterminate problem, and making something from nothing). The maximum entropy method for inferring an ensemble of 3D structures seems to be affected by this too, and it would help to discuss the limitation. For example: Consider a hypothetical polymer ensemble in which 3 monomers are special (lets call them A, B, C, and lets assume they are spaced out along the polymer and in that order. A and C both compete to bind B, but only one can bind at a time, and also A can bind C and exclude B. The sticky binding interactions are weak enough that they are regularly disrupted by the thermal motion of the polymer, giving the other bond a chance to form next. Now consider a different hypothetical ensemble, in which A binding B increases the probability of also binding C in a 3 way hub, though all the individual pairwise binding events are weaker/ rarer. With the right choice of affinities, these ensembles may have identical pairwise contact maps (e.g. RCMC data), despite their different interaction rules and different physical 3D structures. A maximum entropy method - or any approach of producing an ensemble from this average, can not get both ensembles correct, and must through its inherent assumptions favor certain types of interactions (perhaps a lack of sensitivity to recognize competitive interactions?)

Response: We agree that this is an important consideration. As the reviewer correctly notes, inferring an ensemble from averaged pairwise data is under-determined: a fundamental problem for data-driven approaches.

Under the maximum entropy framework, no cooperativity is introduced explicitly, but arises naturally as required from the experimental data and properties of the system. In the example provided, maximum entropy inversion would likely infer a favorable interaction energy between A-B, B-C, and A-C. At a local level, formation of an A-B-C hub would be favored by virtue of it including all three favorable interactions but inhibited by A and C competing for the limited physical space around B. Depending on the precise context this would have a different net effect. At a higher-level, the polymer structure will induce correlations. For example, if B and C are close in sequence, forming an A-B-C hub will be more entropically favorable and more common.

The maximum entropy approach is thereby in some sense the least biased way to recover a 3D ensemble, in that it does not utilize unwarranted assumptions not supported by the data. It represents the simplest estimate given the specified prior and the available experimental data [6]. But this does not mean that it is correct. This is why we took care to validate our structures against chromatin tracing data. The reproduction of the distance scaling and fluctuations in pairwise distances from chromatin tracing suggest that these simple assumptions provide an reasonable model of at least most chromatin regions.

In our initially submitted manuscript, we took care to not comment on phenomena that would be sensitive to cooperative interactions. Most notably, we find only weakly cooperative binding between microcompartment anchors in our simulations. It seems plausible that these interactions are more cooperative than would be inferred from just the properties of the system. For instance, due to joint recruitment to a transcriptional hub. We've added the below text in the Results section to clarify the implicit assumptions inherent to the maximum entropy approach and resulting limitations.

Recovering a 3D ensemble from pairwise contact data is fundamentally an underdetermined problem. MEI addresses this by identifying the ensemble that is most similar to the baseline model of the system while matching the experimental data [6]. In this way, no structural correlations are included unless they are necessitated by the contact map or the baseline model. The agreement with chromatin tracing data, and other imaging modalities as discussed in the next section, suggests that this produces an accurate model of at least most chromatin regions. However, it is possible that some sites interact cooperatively in excess of this least biased solution, for example by joint recruitment to transcriptional hubs.

4. The authors may wish to use the 5 kb resolution Sox2 data from Mateo 2019 or better the deeper data from Ren 2021 and Ren 2025.

Response: We have compared our simulation to the 5 kb resolution Sox2 data from Huang, ..., Ren 2021 and Tastemal, ..., Ren, 2025. The download link for the Mateo 2019 mESC data did not work, so we did not consider this dataset. For both datasets, we used the data from the 129 allele (not containing engineered CTCF sites) and merged all control conditions. We estimated measurement uncertainties following the same procedure we used for the Hafner 2022 Sox2 data. Overall, we observed high correlations between simulated and experimental pairwise distances with the correlation coefficients not much below the correlation between the two experiments (Figure R4b). However, our analysis revealed two inconsistencies.

First, our simulations only match the distance scaling with genomic separation in the Huang 2021 dataset. The two experimental datasets have different distance scaling, so it would not have been possible for our simulations to match both. The discrepancy can not be explained by differing levels of measurement uncertainty. The Tastemal 2025 data has overall higher median distances and the distance scaling curve has a larger slope. At least under the uncorrelated noise model we adopt here, more noise should yield higher median distances and a smaller slope because short distances are disproportionately increased. It is possible that the discrepancy is due to the CTCF and RAD21 AID tag added in the Tastemal 2025 data, but we are unsure.

Second, adding measurement noise to our simulations washes out the Sox2 to Sox2-control-region (SCR) loop and associated TAD, but these features are apparent in the experimental median distance maps and somewhat visible in the contact maps. We suspect that this is due to inaccuracies in our noise model: the Sox2-SCR loop is prominent in the simulations before noise is added (Figure R4) and the simulations accurately reproduce the Sox2-SCR loop distance distribution observed through live cell imaging (Fig. E2d). However, it is of course possible that our simulations genuinely underestimate the frequency of this interaction.

We have added the comparison to Huang 2021 as Fig. E2e and the below text to the Results section. We have chosen to omit the Tastemal 2025 comparison as we are unable to make sense of the discrepancy from the Huang 2021 data.

We additionally compared to live cell imaging [7] and 5 KB resolution chromatin tracing data [8] covering the loop between the *Sox2* promoter and its enhancer. Our simulations closely reproduce the distribution of promoter–enhancer distances observed in live cell imaging (Fig. E2d). Our simulations appear to underestimate the strength of this same interaction in chromatin tracing data, after accounting for measurement uncertainty (Fig. E2e), but overall there is a strong correlation between median pairwise distances ($r=0.89$).

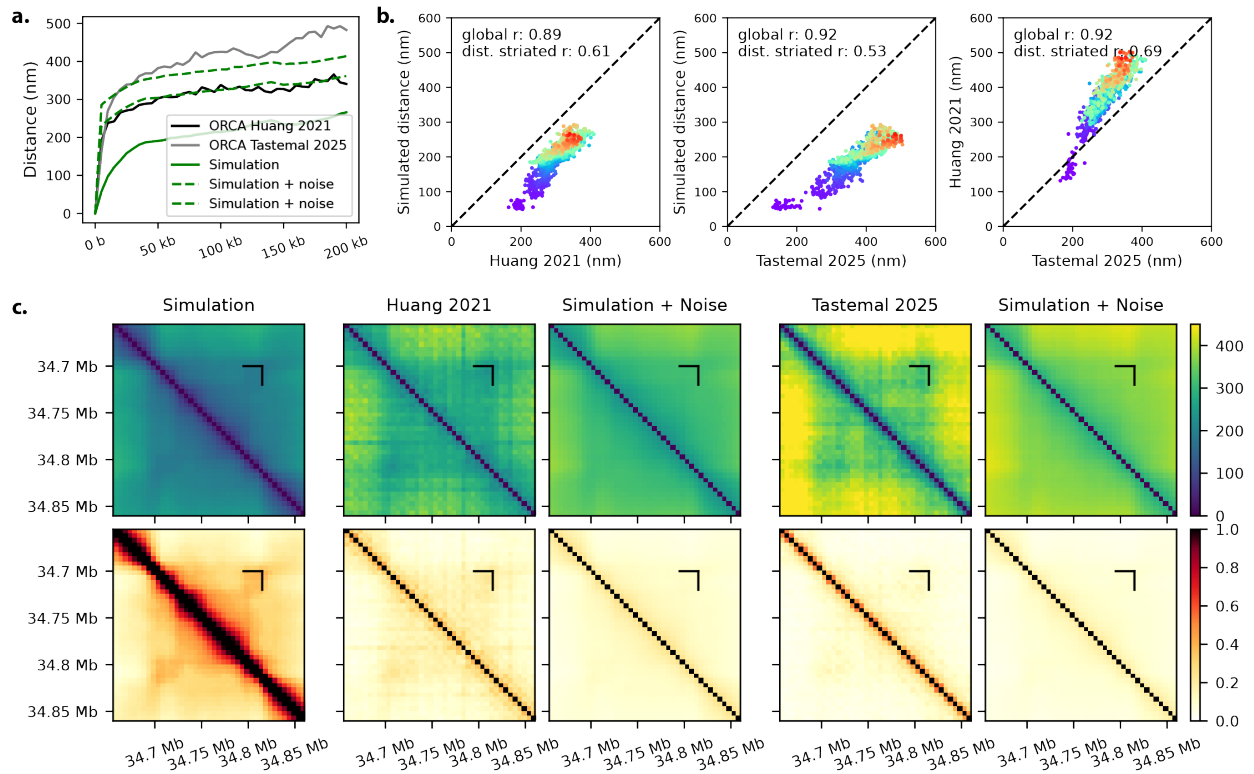


Figure R4: Comparison to Huang 2021 and Tastemal 2025 chromatin tracing. (a) Distance scaling with genomic distance for the two chromatin tracing experiments, our simulations, and our simulations with noise added to the coordinates. The two “Simulation + noise” lines represent noise calibrated at a genomic separation of 20 KB for either dataset. (b) Direct comparison of pairwise median distances. The distance stratified correlation coefficient is defined as the average Pearson correlation for each genomic separation. (c) Median distances (top) and contact maps computed with a distance threshold of 150 nm (bottom). The two “Simulation + Noise” entries are our simulations corrupted by the noise fit to either dataset.

III. Text suggestions:

1. The introduction or discussion could be bolstered with some further references on prior work measuring local variation in chromatin density on the scales of 10s of kb to 1 Mb.

- a. For example, Boettiger et al Nature 2016, report some interesting length dependent and euchromatin/heterochromatin dependent variation in chromatin densities using STORM that I think connect well to the conclusions from this work.
- b. From the intro: “. However, recent evidence challenges this dichotomy, suggesting that euchromatin also contains largely condensed domains [1–4].” Similarly, some evidence suggests heterochromatin is not distinctly ‘condensed’ – Legant 2023, Murphy Nat Gen 2024.
- c. It could also be interesting to compare the organization of such heterochromatin regions from micro-C if sufficient resolution exists. If the best available mESC micro-C data lacks the needed resolution and the available RCMC data lacks coverage of Polycomb heterochromatin domains, the Drosophila Micro-C data from the Levine lab has very high coverage owing to the small genome and deep sequencing, and might be a very informative place to apply this approach to varying types of heterochromatin, or compare the inferred organization with that from microscopy.

Response:

Thank you for pointing out these studies. We agree that they connect well with our work and have added the below text to the Discussion section

Second, gene-rich regions are broken into more, smaller packing domains (Fig. 4i) and thereby adopt more extended conformations on large length scales (Fig. 5e, Fig. E7c). This is consistent with the more extended conformations of active regions observed in chromatin tracing experiments [9–11] and the more clustered structure of broadly dilute nuclear regions observed in super-resolution imaging [12].

We’ve also performed additional simulations for HCT116 cells, including a polycomb-repressed region. This allowed us to compare the distribution of radius of gyration (R_g) for active, inactive, and polycomb-repressed chromatin, as done in Boettiger 2016 and Murphy 2024. We found that the polycomb-repressed region indeed has the smallest R_g s, though there is high overlap in the distributions as observed in Murphy 2024 (new Fig. E8c reproduced here as Figure R5c). We’ve added a new results section entitled “The observed chromatin organization generalizes across diverse cell-types and epigenetic states” featuring this comparison prominently.

We appreciate the insightful suggestion to analyze Drosophila micro-C data. We looked into the data from [13]. Despite the smaller genome, the available micro-C data is still roughly 10 times lower coverage than RCMC. This data does facilitate computing smooth genome-wide contact densities using neighbor balancing at much higher resolution than in mESCs. It would be interesting to use these data to study whole genome structure in the future.

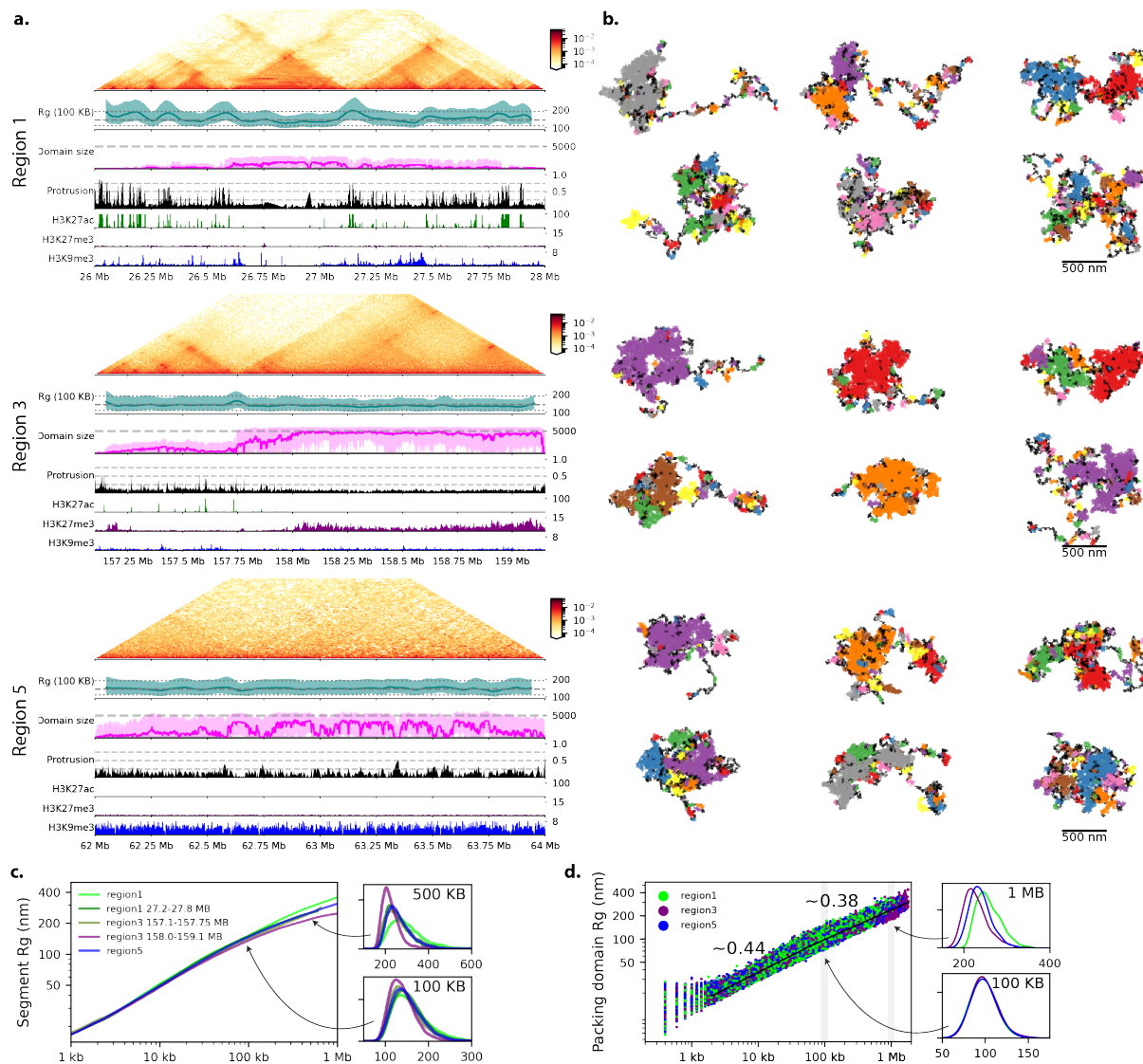


Figure R5: HCT116 cell chromatin is also largely composed of condensed domains (a) Quantitative results of MEI simulations in HCT116 cells. From top to bottom, the tracks show the contact maps from RCMC used in MEI, the median radius of gyration (Rg) for 100 kb windows centered on each position (10-90 percentile shaded), the median domain size (25-75 percentile shaded), protrusion probability, and histone ChIP-seq tracks from ENCODE. (b) Randomly selected example simulation frames colored by DBSCAN domain assignments. Protrusions are colored black. (c) Comparison of the scaling of median Rg with segment length for the different genomic regions. The sub-regions were chosen so that Region 1: 27.2-27.8 contains a single TAD, Region 3: 158.0-159.1 contains a polycomb-repressed region, and Region 3: 157.1-157.75 contains a neighboring gene-rich region. The two right panels highlight that there is high overlap in the distributions of Rgs for 500 KB windows (top) and 100 KB windows (bottom). (d) The Rg of individual packing domains compared to their genomic content. The right panels show the Rg distribution for packing domains containing 0.9-1.1 MB (top) and 90-110 KB (bottom). The 0.9-1.1 MB packing domains in region 3 have smaller Rg than the other regions, suggesting that polycomb-repressed packing domains are particularly dense.

2. Line 110 “In MEI, one begins with a baseline model and then adds auxiliary energy terms to make the system recapitulate experimental data (see Methods).” I appreciate the authors choice to not clutter the text with excessive technical detail here. However, this 1 line explanation I think is not so useful, and offers no explanation of what I see as the fundamental process – developing an ensemble of distinct states that will average to contact map (presumably in a maximal entropy sense). It would help to give the reader a brief sense of the baseline model (e.g. is it a simple crumpled polymer, or a simple freely jointed chain, or is it a polymer whose intra-monomer interaction energies are defined by the RCMC heatmap?

Response: Thank you for letting us know that this is unclear. We’ve expanded the first paragraph of the Results section to describe the maximum entropy approach in more detail. (The updated text highlighted in our response to comment 1.)

We have also added more details on the baseline model to the Results section. The updated text reads

To meet these two criteria, we employed a simple nucleosome-level model that matches key characteristics of chromatin while facilitating large-scale simulations. Nucleosomes are treated as spherical beads that attract at short range to capture intrinsic nucleosome interactions [14] and positioned uniformly. Small “DNA” beads are included between the nucleosomes to prevent knotting [15]. Prior to optimization, this baseline model adopts an unknotted globular conformation. See Methods for a detailed description of our model and the impact of its parameters on the final ensemble.

3. I think it would clarify the presentation to clarify what is meant by the ‘no free parameters’ -> I take it the different contact maps (per region, or of the same region in different perturbations) are considered inputs, not parameters? This issue is potentially confusing as the loop-extrusion approach has been described as a minimal set of parameters, where the aim is to explain and reproduce the contact map from polymer properties that make minimal reference to the contact map as input (instead using non 3D structural data, like CTCF ChIP-seq and CTCF site directions, as input).

Response: Yes, we mean that our modeling decisions, e.g. choice of baseline model or contact indicator function, do not specifically control individual pairwise distances. We agree that this sentence is confusing. We have simply removed this sentence from the Results section and updated the similar sentence in the Methods to read

While the overall magnitude of average distances was used to determine the model and thereby cannot be considered validation, none of our modeling decisions directly control the shape of the distance scaling curves, the individual pairwise distances, nor

the full distributions of individual pairwise distances. Close agreement of these three quantities therefore validates our model.

Remarks on code availability:

The neighbor-balance code appears to be complete and have detailed documentation. The maximum entropy inversion code was not provided and could not be assessed.

Response: We have made the maximum entropy inversion code available at the [GitHub Repository](#).

RESPONSE TO REFEREE 2

Summary:

This manuscript discusses a novel computational framework to reconstruct chromatin 3D structures at nucleosome resolution using Region-Capture Micro-C (RCMC) data combined with a method of maximum entropy inversion (MEI), exploiting "neighbor balancing" to correct biases in contact maps while preserving biological variation.

By use of the new method, the authors show that euchromatin, traditionally viewed as generically "open", actually folds into compact "packing domains" containing dense nucleosome clutches. They also show that kilobase-scale active promoters and enhancers tend to protrude outward from these domains, making them accessible for protein binding and enhancer-promoter communication. Their simulations reproduce experimental contact maps, distance distributions from chromatin tracing, and structural motifs from imaging, thus providing a broad, multiscale view of chromatin organization. Overall, the manuscript concludes that euchromatin is not just "open", but rather composed of condensed domains with local accessible regions that spatially compartmentalize regulatory elements. This picture does clarify the current simplified view of chromatin organisation.

Recent imaging studies have suggested that euchromatin is more compact than previously believed, and that chromatin is organized into clutches and nanodomains. However, these methods lacked locus-specific resolution. This manuscript advances the field by providing nucleosome-resolution conformation exploration; by bridging structural observations across scales; and returning a plausible mechanistic model where condensed packing domains coexist with accessible regulatory protrusions.

The manuscript includes interesting methodological innovations. The introduction of "neighbor balancing" does offer an important improvement over standard ICE normalization, recovering meaningful biological variation in contact density. The integration of RCMC with MEI achieves nucleosome-level resolution, to help bridging population-averaged contact maps with single-cell imaging observations. Additionally, the model is validated against different experimental data, including ORCA chromatin tracing and super-resolution imaging, which strengthens confidence in its results. Finally, the finding that active enhancers and promoters tend to protrude from otherwise compact domains provides a plausible structural explanation for their accessibility and communication despite local chromatin compaction.

Nevertheless, the author could try to address some weak points. ([See below.](#))

Overall, this is a well-executed, technically valid and novel manuscript that makes a valuable contribution to the understanding of chromatin architecture. The main conclusions are well supported, though the manuscript would definitely benefit from additional analyses to address the weak points highlighted above.

Response: We thank the reviewer for taking the time to understand our work, as evidenced by the accurate summary, and recognition of its value. Below we provide a point-by-point response for how we have addressed the weak points identified by the reviewer. We appreciate these insightful comments and believe that addressing them has significantly improved our manuscript.

While their simulations reproduce experimental data well, there is limited experimental confirmation (e.g., imaging of specific protrusions corresponding to predicted enhancer/promoter sites).

Response: We are now co-submitting our work with an experimental paper that provides complementary insights [doi: 10.1101/2025.08.27.672592]. This paper, which was posted online shortly after our manuscript submission, provides strong orthogonal validation of our conclusions.

Briefly, in this study, the Maeshima group applied super-resolution imaging of the nucleus using euchromatin-specific labeling of histone H3.3. Consistent with our observations (Fig. 4), the imaging results show that euchromatic regions form compact domains (Fig. R6A,B).

To further distinguish regulatory elements from gene bodies, the Maeshima group performed immunostaining using antibodies against the C-terminal domain (CTD) of RNA polymerase II (RNAP II) and its phosphorylated forms at serine 5 (Ser5P) and serine 2 (Ser2P). These phosphorylated forms correspond to transcriptionally active paused and elongating RNAP II, respectively, and are commonly used as proxies for regulatory elements and gene bodies. The imaging results revealed that the different RNAP II forms occupy distinct spatial niches, with transcription initiation preferentially localized at domain peripheries (Fig. R6C). The more peripheral localization of paused RNAP II relative to elongating RNAP II is consistent with our structural model, in which regulatory elements preferentially protrude to the surface of compact domains.

To help connect with this work, we've added Fig. 6d (reproduced here as Figure R6c), illustrating that regulatory elements but not gene bodies are enriched in protrusions.

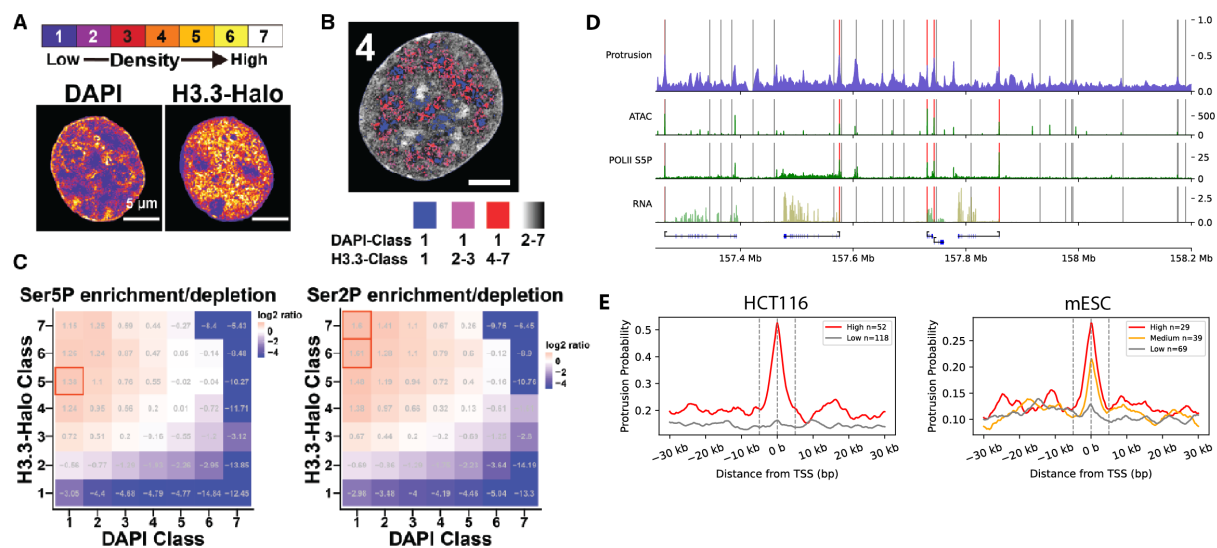


Figure R6: Condensed euchromatin with regulatory protrusions revealed by 3D-SIM. (A) 3D-SIM images of DAPI-stained chromatin (left) and H3.3-HaloTag-labeled nucleosomes (right), colored by signal intensity. (B) Regions with low DAPI but high H3.3 signal (red) support the presence of compact euchromatin domains. (C) Enrichment of Ser5- (initiation, promoter region) and Ser2-phosphorylated (elongation, gene body region) RNAP II across regions classified by DAPI and H3.3 intensity. (D) Protrusion probabilities are enriched at active promoters (with high Ser5P ChIP-seq signal) but not gene bodies. (E) Protrusion probabilities are enriched at expressed but not silent promoters, though there are exceptions to this rule, for example the promoters near 157.65 MB in panel D.

We've added the following summary of the above to the Discussion section

A recent study that performed super-resolution imaging of euchromatin-specific histone H3.3 provides particularly strong support to this model [16]. Consistent with our observations (Fig. 4), the imaging results show that euchromatin regions form compact domains. To further distinguish regulatory elements from gene bodies, the Maeshima group performed immunostaining against the serine 5 and serine 2 phosphorylated forms of the C-terminal domain of RNA polymerase II (RNAP II). These phosphorylated forms correspond to transcriptionally active paused and elongating RNAP II, respectively, and are commonly used as proxies for regulatory elements and gene bodies. The serine 5 form preferentially localized to domain peripheries and the serine 2 form to domain bodies, consistent with euchromatin forming condensed domains with only regulatory elements preferentially (Fig. 6d).

In addition, we generated simulated ChromSTEM images to directly compare with experiments reported in Li et al. [17] (see Figure R16). The simulated structures again closely reproduce the experimental values of packing domain radii.

We also compared our simulation to the 5 kb resolution Sox2 data from Huang et al. [8]. Overall, we observed high correlations between simulated and experimental pairwise distances with the correlation coefficients (Figure R4).

These additional results further support the biological relevance of the reported chromatin organization.

The study is restricted to mouse embryonic stem cells; whether the observed euchromatin architecture generalizes to differentiated cell types remains unanswered.

Response: We now include simulations for HCT116 cells and analysis of contact densities determined by neighbor balancing for HCT116, GM12878, K562, and H1 ESCs. In addition to new cell lines, these data also cover more diverse epigenetic states, including an exceptionally active histone locus, a polycomb-repressed region, and H3K9me3 enriched heterochromatin region.

The results suggest that our findings generalize to these other cell lines and chromatin states. As in mESCs, the contact densities are largely consistent apart from sharp dips at narrow active loci and are highly correlated with condensability scores in H1 ESCs and GM12878 cells (Fig. R7). In the HCT116 cell simulations, all of the regions are mostly arranged into condensed domains with protrusions occurring preferentially at narrow active regions (compare protrusion probabilities to H3K27ac in Fig. R8a). As observed in our mESC simulations, gene rich regions are fragmented into more, smaller packing domains, whereas the inactive regions generally form larger domains. Interestingly, we found that large packing domains in the polycomb-repressed regions have lower radius of gyration than large packing domains in active regions. This deviates from our findings in mESCs, where there were no notable differences in packing domain compaction. The difference is modest (median values of 223 nm and 252 nm), but nonetheless suggests that in certain heterochromatin regions in differentiated cells may form denser domains.

We present this analysis in a new Results section entitled “The observed chromatin organization generalizes to diverse cell-types and epigenetic states” and included two new Supplementary Figures, which we reproduce here for convenience.

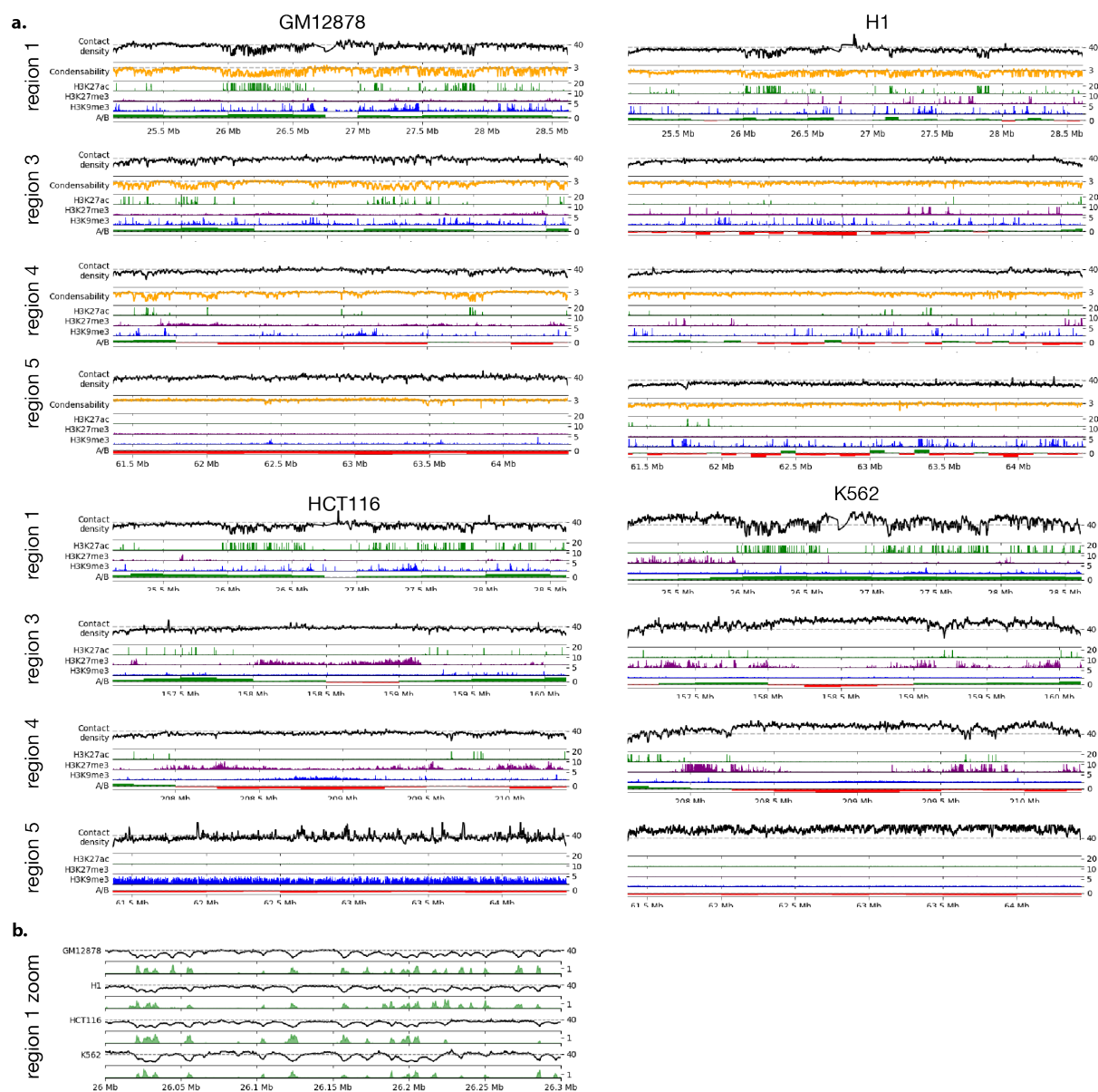


Figure R7: Contact densities derived from neighbor balancing for four human cell lines. (a) Comparison of contact densities computed using neighbor balancing for RCMC data from [18] to condensability scores [19] and histone ChIP-seq data from ENCODE. We suspect that the higher contact densities in K562 cells are due to batch effects in the RCMC data. **(b)** Cell-type specific H3K27ac peaks (green) are associated with cell-type specific dips in contact density (black).

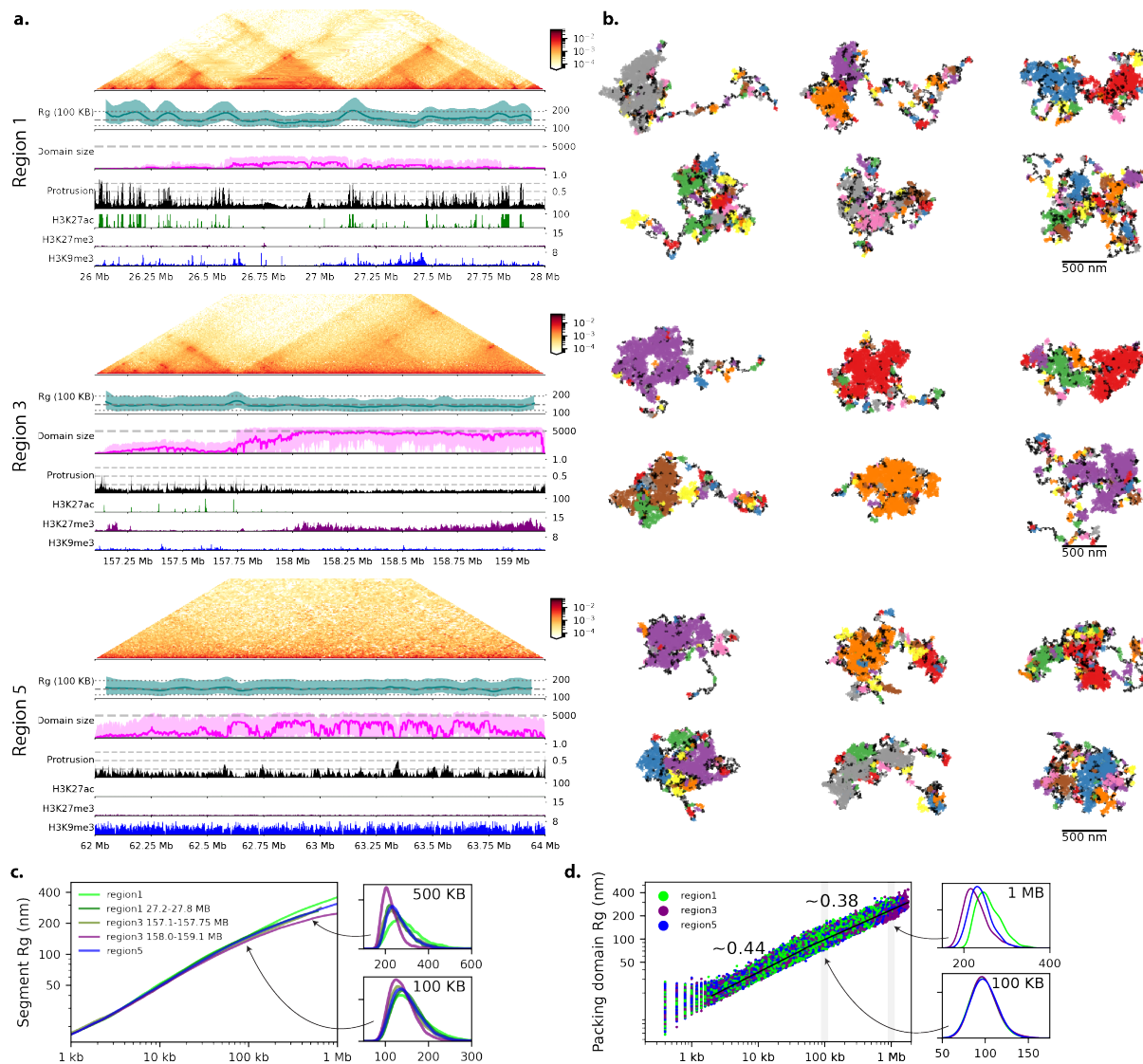


Figure R8: HCT116 cell chromatin is also largely composed of condensed domains (a) Quantitative results of MEI simulations in HCT116 cells. From top to bottom, the tracks show the contact maps from RCMC used in MEI, the median radius of gyration (Rg) for 100 kb windows centered on each position (10-90 percentile shaded), the median domain size (25-75 percentile shaded), protrusion probability, and histone ChIP-seq tracks from ENCODE. (b) Randomly selected example simulation frames colored by DBSCAN domain assignments. Protrusions are colored black. (c) Comparison of the scaling of median Rg with segment length for the different genomic regions. The sub-regions were chosen so that Region 1: 27.2-27.8 contains a single TAD, Region 3: 158.0-159.1 contains a polycomb-repressed region, and Region 3: 157.1-157.75 contains a neighboring gene-rich region. The two right panels highlight that there is high overlap in the distributions of Rgs for 500 KB windows (top) and 100 KB windows (bottom). (d) The Rg of individual packing domains compared to their genomic content. The right panels show the Rg distribution for packing domains containing 0.9-1.1 MB (top) and 90-110 KB (bottom). The 0.9-1.1 MB packing domains in region 3 have smaller Rg than the other regions, suggesting that polycomb-repressed packing domains are particularly dense.

The modeling depends on a broad set of parameters and assumptions such as nucleosome occupancy and capture radius; it'd be very important to assess the robustness of the findings and conclusions.

Response: We agree that characterizing the dependence of our conclusions on modeling parameters and assumptions is of the utmost importance. We now include additional analysis supporting that our conclusions hold qualitatively if the capture radius is set well above and below the range supported by chromatin tracing data. We added a new supplementary figure with this analysis (Fig. S13, reproduced here as [Figure R9](#)) and the below text to the Method sub-section “Selecting the contact indicator function”.

Our conclusions are robust to changes in the contact indicator function across the range supported by comparison to chromatin tracing experiments. In particular, increasing or decreasing the capture radius by 20% similarly leads to conformations that are a mix of packing domains and protrusions for the euchromatin nanog region (Fig S13d,e). The position dependence of protrusion probability is preserved (Fig S13d). Increasing the capture radius by 20% does result in a significantly higher fraction of the system forming protrusions, which perhaps is in conflict with our conclusion that euchromatin is mostly domains. However, this value results in pairwise distances being overestimated by over 50% compared to ORCA and is therefore well outside of the biologically relevant range. The slope parameter has a minimal impact. We increased and decreased it by a factor of 2 and adjusted the capture radius to maintain the effective volume enclosed by the contact indicator function, i.e. to preserve $\int_0^{\text{inf}} r^2 f(r) dr$. The distance scaling, protrusion probabilities, domain sizes, and qualitative appearance of the structures changed little (compare results labeled “1/21” and “4/21” to “42 nm” in (Fig. S13).

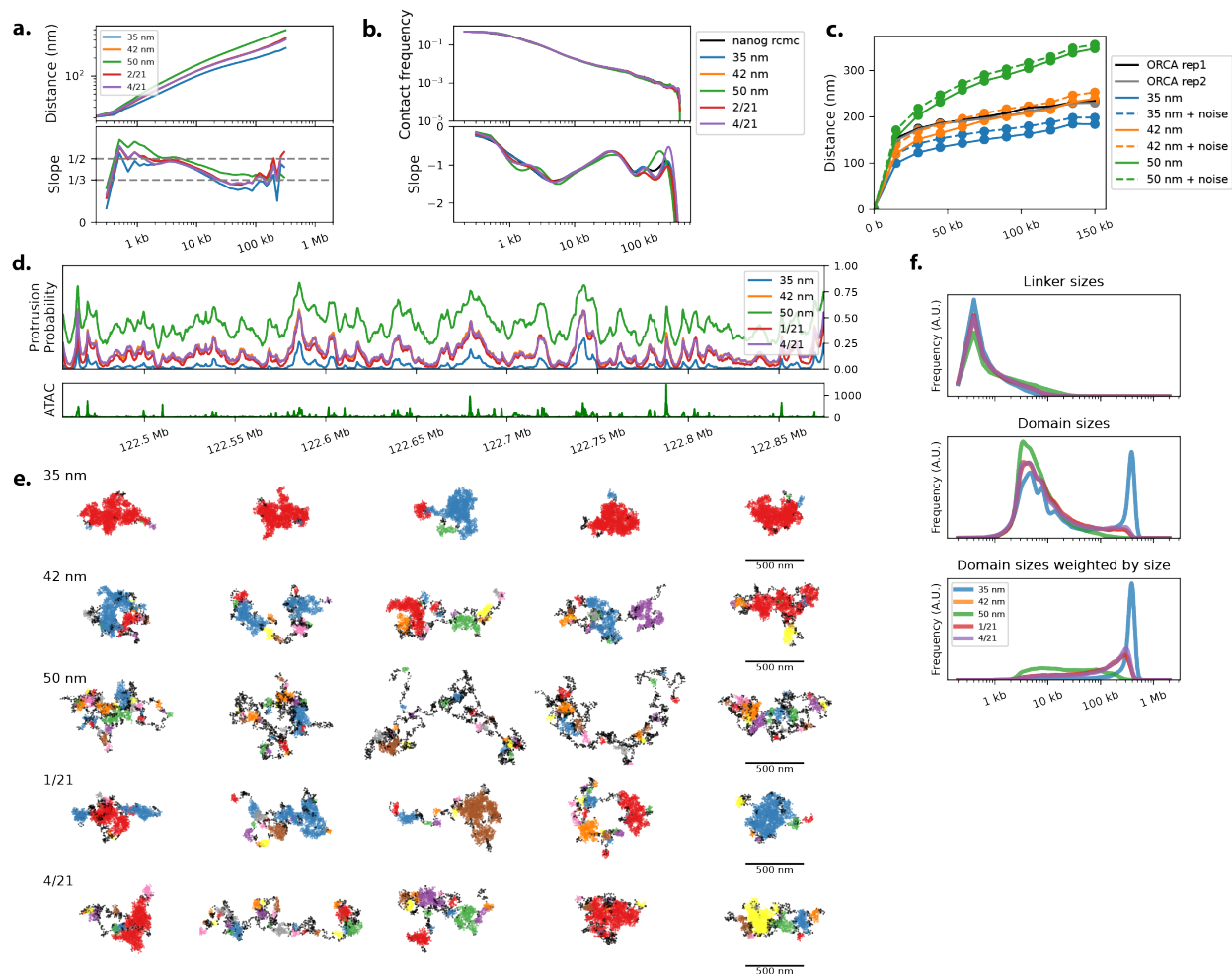


Figure R9: Condensed domains and protrusions are present well beyond the range of contact indicators supported by chromatin tracing data. (a) Segment end-to-end distances for the optimized *Nanog* system with a range of capture radii. The results in the main text all use a capture radius of 42 nm. (b) Simulated $P(s)$ curves after optimization, each computed using the indicated contact indicator. This is included to show that the optimizations are converged. (c) Comparison to chromatin tracing data (ORCA) shows that a capture radius of 50 nm (35 nm) results in a system that is on average too large (small). (d) Representative simulation frames colored by packing domain assignments for different nucleosome interaction energies. Protrusions are shown in black. (e) Increasing the capture radius generally increases the probability of a position occurring in a protrusion. (f) Size of domains and protrusions for different capture radii.

In our framework, it would be challenging to directly assess the impact of nucleosome occupancy. However, the data-driven nature of our model largely mitigates the impact of nucleosome occupancy assumptions, as compared to in physics-based model. The optimized energies implicitly incorporate the effects of nucleosome positioning on the structural ensemble. For example, some evidence supports that nucleosome free regions (NFR) favor locally extended structures. As long as this is reflected in the contact map, our model will fit repulsive energies along the chain, favoring extended

structures. We indeed see peaks in local extension (quantified by R_g for 5 nucleosome windows) at ATAC sensitive sites (Figure R10). This treatment does, however, underestimate the increased displacement: three of our nucleosome beads in a straight line will extend roughly 48 nm (two times the bond length), whereas 200 bp of naked DNA could extend as much as 65 nm. Therefore, our model likely provides a conservative estimate of the extension of clutch boundaries and protrusions. This limitation is mentioned in the Results. We have added the highlighted clause to stress that this could be particularly important at regulatory elements.

The only substantial disagreement between our simulated ensemble and the chromatin tracing data is that the simulations underestimate median distances at short genomic separations (Fig. 3a). We suspect this occurs because our simulations assume full nucleosome occupancy, whereas in reality at least some nucleosomes are missing, especially around promoters and enhancers, which would locally expand the system.

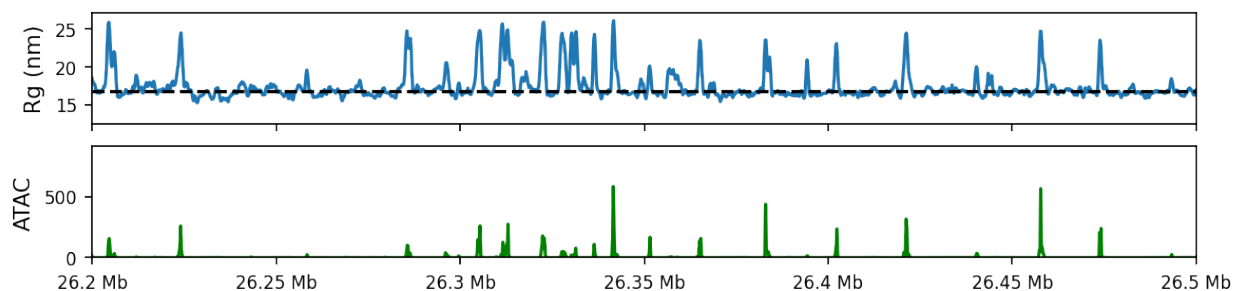


Figure R10: Simulations capture local extension favored by NFRs The top track shows the median radius of gyration (R_g) of 1 KB segments, and the bottom track shows ATAC seq coverage (ENCODE file ENCFF624HRW). These data are for a section of region 1 in the newly added HCT116 cell simulations, though similar trends are present in all simulations.

Finally, direct links to gene expression outcomes or transcriptional activity are inferred rather than demonstrated.

Response: The primary goal of our study is to infer nanoscale structures of chromatin and observe how they relate to epigenetic features. We agree that our attempts to connect these observations to transcriptional outcomes are based on correlations and we do not show causality. For instance, it is possible that promoters forming protrusions at a high rate is a result rather than cause of transcriptional activity. Due to this uncertainty, we restrict such comments to the Discussion.

We have attempted to at least make the correlations with gene expression more direct. We've added Fig. 6d (Figure R6C) depicting increases in protrusion probability at active promoters and Fig. S15b (Figure R6D) showing the average protrusion probability aggregated over all active and inactive promoters. Active promoters generally have higher protrusion probabilities than inactive ones, however there are a small subset of inactive promoters associated with peaks in protrusion probability (for example those near 157.65 MB in Figure R6).

RESPONSE TO REFEREE 3

Summary:

In the paper by Paggi and Zhang entitled "Euchromatin forms condensed domains with short active regions on the surface", the authors introduce a new computational tool that aims at simulating chromatin at "nucleosome" resolution taking as consideration micro-C datasets. This new approach is accompanied by a new micro-C normalization procedure (neighbor balancing) that takes into consideration local packing of chromatin as observed by micro-C. Next, the authors validate the models comparing them to mainly microscopy-based data to finally apply it to a series of genomic regions. Finally, the authors provide a structural analysis of the models to propose that euchromatin is formed by dense groups of continuous "clutches" and decorated by open, less dense and protruding DNA containing promoters and enhancers. The authors conclude that this spatial organization allows to understand how promoters and enhancers can act at distance providing additional mechanisms to the long studied "loop extrusion" model. Overall, this is a well written manuscript with novel computational approaches that, when applied, provide substantial biological insights. However, the manuscript could benefit for further work specially in validating the models, making clear the limitations of the computational approaches, and further characterizing how the data and parameters affect the simulations. Next are some more specifics placed in order of relevance to this reviewer:

Response: We thank the reviewer for the careful and constructive evaluation of our manuscript. We appreciate the recognition of our proposed computational approaches and biological insights.

We have performed substantial additional analyses in response to the reviewer's comments, including simulations of new regions (in HCT116 cells), more rigorous statistical testing, characterizing the effects of varying simulation and analysis parameters, and directly comparing to electron tomography images. The results proved quite interesting, and we believe strengthen our work substantially. We have also clarified parts of our manuscript and made the limitations of our modeling more explicit.

- To really complement the data, I would suggest the authors simulate a very dense heterochromatin region with little or no ATAC-seq peaks. Do the boundaries (protruded regions) still appear? If so, why?

Response:

We have expanded our analysis to include several additional heterochromatin regions in HCT116 cells, including a H3K9me3-enriched heterochromatin region and polycomb-repressed region (new Fig. E8 reproduced here as [Figure R11](#)). In these regions (all of region 5 and 158 MB onwards in region 3, respectively), we still observe protrusions. However, we do not see the sharp peaks in protrusion probability present at open chromatin sites in active regions.

Overall, our simulations show that most regions form protrusions at a fairly consistent baseline rate due to thermal fluctuations. This suggests that nucleosome-nucleosome interactions (both intrinsic and extrinsic) are strong enough to induce domain condensation, but not strong enough to completely counteract the entropic forces driving nucleosomes to exit from the domains. Heterochro-

matin specifically existing in this intermediate regime is supported by recent cryo-ET images of heterochromatin showing significant internal structure, as opposed to an entirely collapsed domain [20, 21] and chromatin tracing data showing overlapping distributions of radius of gyration for polycomb-repressed and control regions [10].

Interestingly, contact densities from neighbor balancing in these heterochromatin regions are not substantially higher than those from broadly defined euchromatin regions (Figure R7). This initially appears to diverge from chromatin tracing experiments, such as [9], showing that heterochromatin regions have lower radius of gyration (R_g), but our simulations reconcile this distinction. Heterochromatin regions form larger domains—and would in this way appear more compact in a chromatin tracing experiment—but the internal density of these domains is not substantially higher. In this way, the heterochromatin regions can have a comparable nanoscale compaction (which correlates closely with contact density) but higher mesoscale compaction (which correlates closely with R_g).

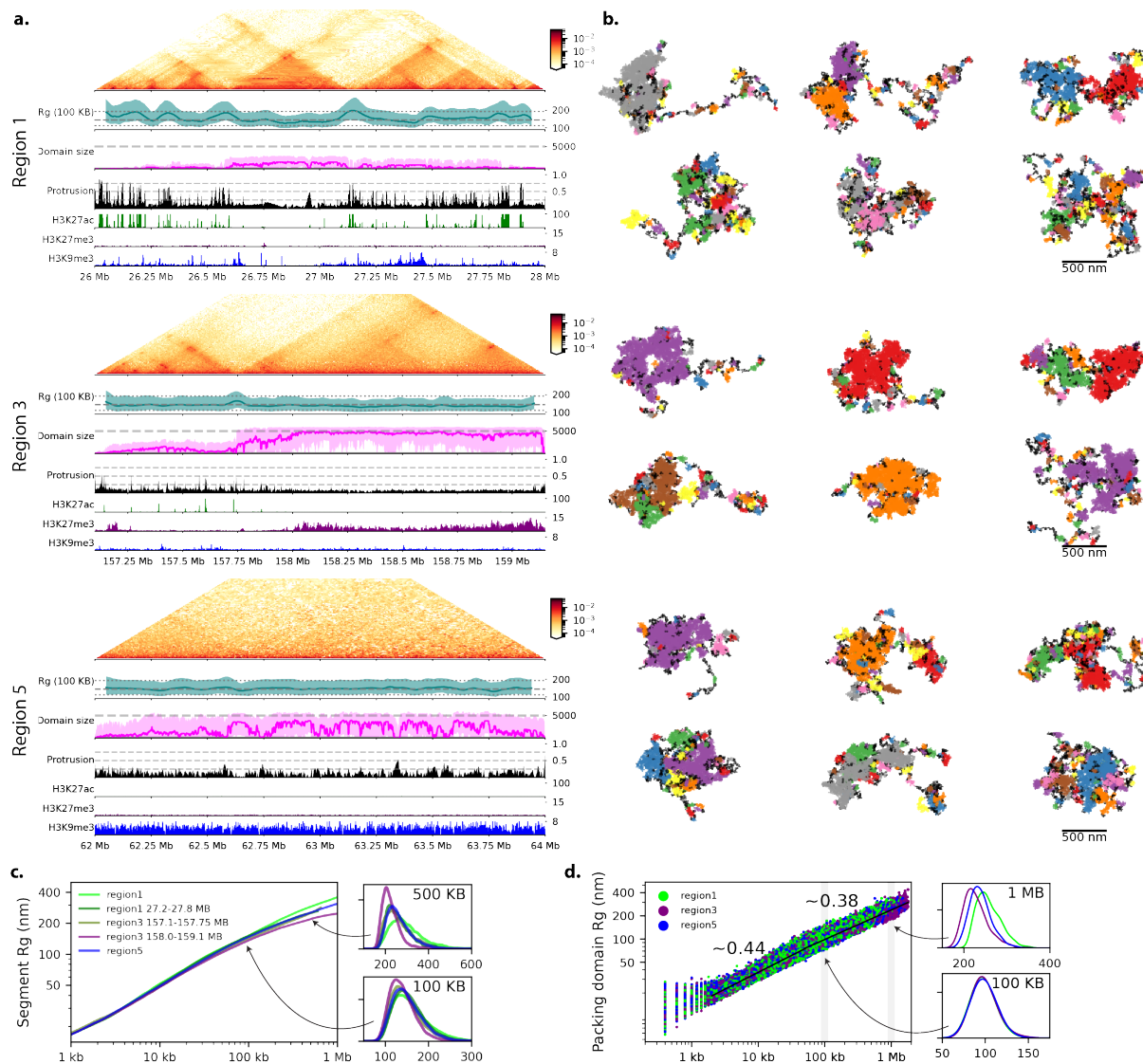


Figure R11: HCT116 cell chromatin is also largely composed of condensed domains (a) Quantitative results of MEI simulations in HCT116 cells. From top to bottom, the tracks show the contact maps from RCMC used in MEI, the median radius of gyration (Rg) for 100 kb windows centered on each position (10-90 percentile shaded), the median domain size (25-75 percentile shaded), protrusion probability, and histone ChIP-seq tracks from ENCODE. (b) Randomly selected example simulation frames colored by DBSCAN domain assignments. Protrusions are colored black. (c) Comparison of the scaling of median Rg with segment length for the different genomic regions. The sub-regions were chosen so that Region 1: 27.2-27.8 contains a single TAD, Region 3: 158.0-159.1 contains a polycomb-repressed region, and Region 3: 157.1-157.75 contains a neighboring gene-rich region. The two right panels highlight that there is high overlap in the distributions of Rgs for 500 KB windows (top) and 100 KB windows (bottom). (d) The Rg of individual packing domains compared to their genomic content. The right panels show the Rg distribution for packing domains containing 0.9-1.1 MB (top) and 90-110 KB (bottom). The 0.9-1.1 MB packing domains in region 3 have smaller Rg than the other regions, suggesting that polycomb-repressed packing domains are particularly dense.

- There is a need to more rigorously quantify that protruding regions contain more than expected regulatory elements.

Response: We have confirmed that protruding regions contain more than expected regulatory elements using a permutation test ([Figure R12](#)). Promoters were defined as RefSeq annotated gene starts with H3K4me3 signal p-value greater than 10. Regulatory elements more generally were defined as any position with an ATAC-seq coverage greater than 50. For both features, we computed the risk ratio, defined as the probability of protruding monomers being a regulatory element divided by the probability of all other monomers being a regulatory element. For all regions, we found that both promoters and accessible regions generally are more likely to form protrusions (risk ratios ranging from 1.92 to 5.01). We used a permutation test to assess the statistical significance. To create a null distribution, we kept the regulatory annotations fixed but shifted the protrusion labels by a random shift, wrapping the entries going past the end of the array back to the start. This null distribution maintains correlations along the genome. We repeated this for 1000 trials and the observed risk ratio was higher than nearly all of these samples ($p < 0.005$ for all regions). The opposite statistic, i.e. the probability of a regulatory element forming a protrusion divided by the probability of other sites forming a protrusion is also significantly enriched by the same test (data not shown).

We have added the above text and [Figure R12](#) to the Supplementary Information.

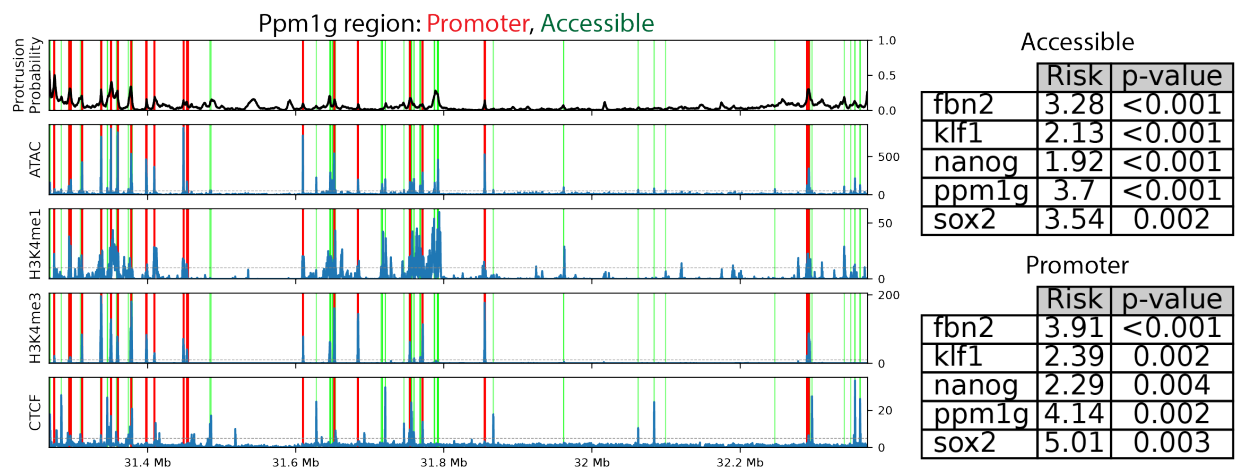


Figure R12: Enrichment regulatory elements in protrusions is statistically significant. The left panel shows an example region showing protrusion probabilities and several epigenetic tracks. Vertical red lines indicate active promoters and green lines show accessible regions. Some green lines are obscured by red lines. The horizontal lines in the ATAC and H3K4me3 indicate the thresholds used to define accessible sites and active promoters, respectively. The right panel shows risk ratios and p-values for the two sets of features being enriched for protrusions.

- I appreciate the neighbor balancing approach proposed here. It apparently can “detect” ATAC signal in the micro-C (otherwise, the models would not show correlation to ATAC peaks). Can the authors elaborate why this is the case? How that can be captured by micro-C and how the new balancing does not remove this data?

Response: Neighbor balancing re-weights each row/column of the contact map using the contact frequency between neighboring windows. The assumption of the method is that neighboring windows are always “in contact” and thereby should have the same contact frequency. Regions with high ATAC signal tend to have a rapid decline in contact frequency with distance (Figure 1d and Figure R13 for convenience). Balancing the map such that the neighbor contact frequencies are close to one thereby results in these regions having lower contact densities.

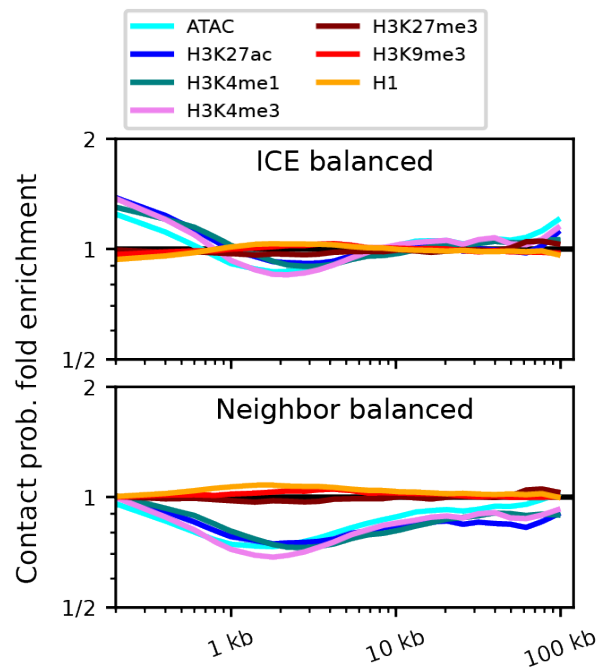


Figure R13: **Distance-dependence of contact frequencies stratified by epigenetic marks.** Enrichment of contact probabilities for windows in the top 5% of the indicated epigenetic feature over windows in the bottom 50%.

- In Figure 3b/d is clear that the resulting models are not at the same scale than the imaging data. This does not seem to be the case when the data is plotted differently in Figures 3a/c/e. Why?

Response: In Fig. 3b/d, the raw distances from simulation are shown. This corresponds to the lines labeled “Simulation” in panels a and c, which indeed show that the distances are underestimated. This mismatch arises because the simulated coordinates are exact, whereas distances inferred from chromatin tracing experiments are broadened by localization uncertainty. To enable a comparison, we therefore add noise to the simulated coordinates matched to the experimental measurement

uncertainty. After this adjustment (lines labeled “Simulation + noise in panels a, c, and d), the simulated and experimental distances are on the same scale.

- Contrary to the statement in the text (page 7), the underestimation of distances in the models is also for long genomic separations.

Response: We believe the reviewer is referring to the Sox2 region specifically (Fig. 3c). Pairwise distances are indeed underestimated for large genomic separations in this region. We did not comment on this specifically because the discrepancy is minor and there is a high overlap between the distribution of chromatin tracing and simulated (+ noise) distances (Fig. E2g right column). It is not clear there is a significant discrepancy, and we do not want to over-interpret it.

- Similarly, can the authors demonstrate that the choice of contact indicator function affects only the scale of the models and not its topology/conformation?

Response: The choice of contact indicator function indeed impacts the conformations of the models, not just their scale in a strict sense. We have replaced “overall scale” with “conformations” in the sentence “The choice of contact indicator function strongly influences the overall scale of the optimized system”.

We now include additional analysis supporting that our conclusions hold qualitatively if the capture radius is set well above or below the range supported by chromatin tracing data. We added a new supplementary figure with this analysis (Fig. S13, reproduced here as [Figure R14](#)) and the below text to the Method sub-section “Selecting the contact indicator function”.

Our conclusions are robust to changes in the contact indicator function across the range supported by comparison to chromatin tracing experiments. In particular, increasing or decreasing the capture radius by 20% similarly leads to conformations that are a mix of packing domains and protrusions for the euchromatin nanog region (Fig S13d,e). The position dependence of protrusion probability is preserved (Fig S13d). Increasing the capture radius by 20% does result in a significantly higher fraction of the system forming protrusions, which perhaps is in conflict with our conclusion that euchromatin is mostly domains. However, this value results in pairwise distances being overestimated by over 50% compared to ORCA and is therefore well outside of the biologically relevant range. The slope parameter has a minimal impact. We increased and decreased it by a factor of 2 and adjusted the capture radius to maintain the effective volume enclosed by the contact indicator function, i.e. to preserve $\int_0^{\text{inf}} r^2 f(r) dr$. The distance scaling, protrusion probabilities, domain sizes, and qualitative appearance of the structures changed little (compare results labeled “1/21” and “4/21” to “42 nm” in (Fig. S13).

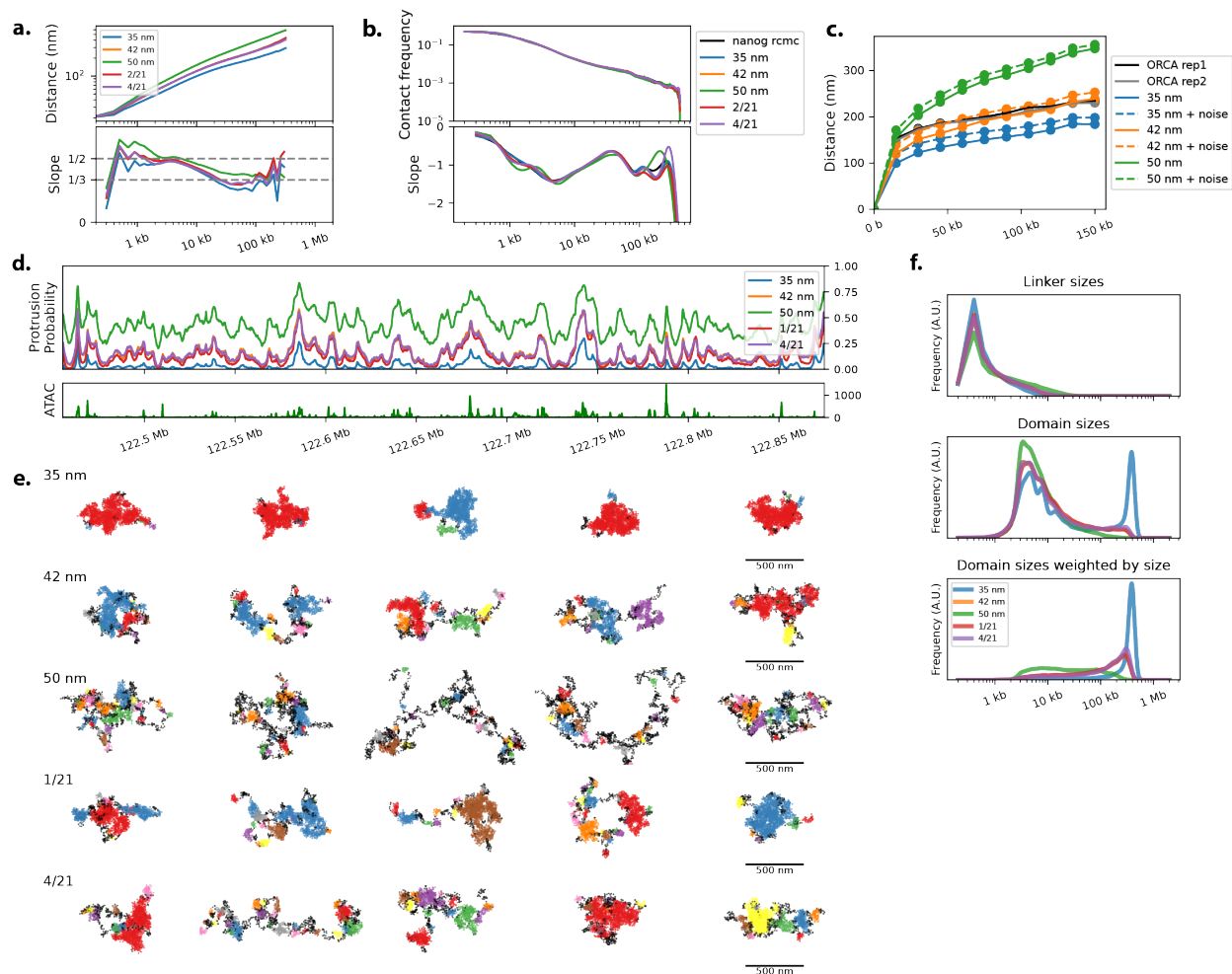


Figure R14: Condensed domains and protrusions are present well beyond the range of contact indicators supported by chromatin tracing data. (a) Segment end-to-end distances for the optimized *Nanog* system with a range of capture radii. The results in the main text all use a capture radius of 42 nm. (b) Simulated P(s) curves after optimization, each computed using the indicated contact indicator. This is included to show that the optimizations are converged. (c) Comparison to chromatin tracing data (ORCA) shows that a capture radius of 50 nm (35 nm) results in a system that is on average too large (small). (d) Representative simulation frames colored by packing domain assignments for different nucleosome interaction energies. Protrusions are shown in black. (e) Increasing the capture radius generally increases the probability of a position occurring in a protrusion. (f) Size of domains and protrusions for different capture radii.

- In Figure 2, the models clearly show a zig-zag organization of the beads. This usually happens when the size of the beads is inconsistent (too large) with the (too small) imposed restraints during simulation between consecutive beads. Is this a feature seen in all models? If so, do the authors consider re-visiting such parameters? (BTW, the code for simulation was not available to the

reviewers).

Response: This zig-zag pattern is expected given that the beads represent nucleosomes (see below response). We in fact add an angle potential to favor such conformations. Zig-zag structures between nucleosomes are supported by cryo-EM and cryo-ET structures [21, 22] as well as the similar value of $i-i+1$ and $i-i+2$ interactions observed in micro-C.

We apologize for not previously providing the simulation code. It is now publicly available at <https://github.com/ZhangGroup-MITChemistry/maxentnuc>.

- I am not sure I would call these models “nucleosome” resolution. To certain extend this is misleading as the beads do not explicitly represent nucleosome position. The models are 100s bp resolution indeed but do not take any data in consideration as to nucleosome positioning. This would separate the approach from other more explicit simulations (coarse-grained DNA simulations) that indeed simulate nucleosomes.

Response: Our model does capture some of the essential properties of nucleosomes, including their excluded volume, the distance between nucleosomes, and their tendency to form acute angles between three successive nucleosomes. We, at least, would refer to this as being “nucleosome resolution” and would think of coarse-grained DNA simulations as being “sub-nucleosome” or “near-atomistic” resolution. Respectfully, we would prefer to continue to describe the model as “nucleosome-resolution”.

We agree that it is a limitation of our approach that we don’t consider non-uniform nucleosome positioning. To some extent our model is able to account for nucleosome free regions by fitting the energies to increase the angle between nucleosomes. However, this does underestimate the increased displacement introduced by NFRs. This limitation is mentioned in the Results and, we have added the highlighted clause to stress that this could be particularly important at regulatory elements.

The only substantial disagreement between our simulated ensemble and the chromatin tracing data is that the simulations underestimate median distances at short genomic separations (Fig. 3a). We suspect this occurs because our simulations assume full nucleosome occupancy, whereas in reality at least some nucleosomes are missing, especially around promoters and enhancers, which would locally expand the system.

- There is an entire section on how to avoid knots in the models. Interestingly, most modeling methods that use Hi-C or related datasets result in no-knots in their models even though there are no specific restraints avoiding them. Why you see knots in your models? can those be avoided without explicitly using connecting beads or extreme beads?

Response: Not explicitly connecting the nucleosome beads does indeed result in knots. Distal segments of the polymer frequently tunnel through each other in what appears to be an unrealistic way.

This is more of a problem in our study than in previous Hi-C-based modeling approaches due to the finer resolution of our model and our attempts to match nucleosome features. Polymer simulations typically use beads with a diameter equal to the bond length. Here we use beads with the diameter (11 nm) and bond length (21 nm) of a nucleosome. In this way, the excluded volume explicitly included in our model is nearly 7-fold less than in typical polymer models and is insufficient to prevent segments of the polymer from overlapping in space. Explicitly connecting the beads induces an effective excluded volume effect, preventing segments from overlapping in space [23]. We believe that this is the most natural way to model the system and are not aware of another way to avoid the knots.

We've added the following summary of the above to the methods section

Interestingly, previous data-driven simulation studies generally do not observe a high degree of knotting, despite not taking explicit measures to exclude knots as we do here. These previous studies have typically used beads with a diameter equal to the bond length, whereas the diameter of our beads (11 nm) is less than the bond length (21 nm). This results in nearly 7-fold less excluded volume, which appears insufficient to prevent segments from overlapping, leading to entanglement.

- Can the authors quantify the effects on the results and conclusions of the two main parameters for DBSCAN?

Response: We first note that DBSCAN is generally formulated with a radius and a minimum points parameter (“epsilon” and “min_points”). Here we specify a nucleosome concentration and radius and then select the minimum points that gives the desired concentration for the given radius. For example, a radius of 30 nm and a concentration of 200 μM , yields 14 minimum points.

We've performed additional analysis of the impact of changing the DBSCAN parameters. The results are included in the newly added Fig. S14 (reproduced here as [Figure R15](#)) and discussed in the supplementary information sub-section “Assignment of packing domains using DBSCAN”.

We chose the epsilon and concentration (min samples) parameters to classify visually coherent groups of nucleosomes as domains. Increasing the required concentration results in some groups that appear spatially coherent being broken into multiple domains, whereas lowering the concentration seems to merge visually separated domains (Figure R15a). Admittedly, this choice is subjective.

These parameters impact our quantitative results, but the features relevant to our conclusions are maintained. In Figure R15, we compare how changing the DBSCAN parameters impacts the results for the most (klf1) and least (fbn2) active regions we studied. First, increasing the concentration or radius results in more nucleosomes being called protrusions. Despite the changes in absolute protrusion probability, the values are correlated with each other and with ATAC-seq (Figure R15c,d). Second, increasing

either parameter right shifts the peak of small domain sizes (Figure R15e, around 10 kb) and generally decreases the frequency of large domains. These small domain sizes have a similar distribution between the two regions in all cases, highlighting a consistent local structure. For all parameters, we also reach a similar conclusion that the fbn2 region more often forms large domains. Finally, the average protrusion probabilities change proportionately for the two systems for all parameters we considered (Figure R15b). In this way, regardless of the chosen parameters, we reach a similar conclusion that euchromatin regions are only modestly more likely to form protrusions.

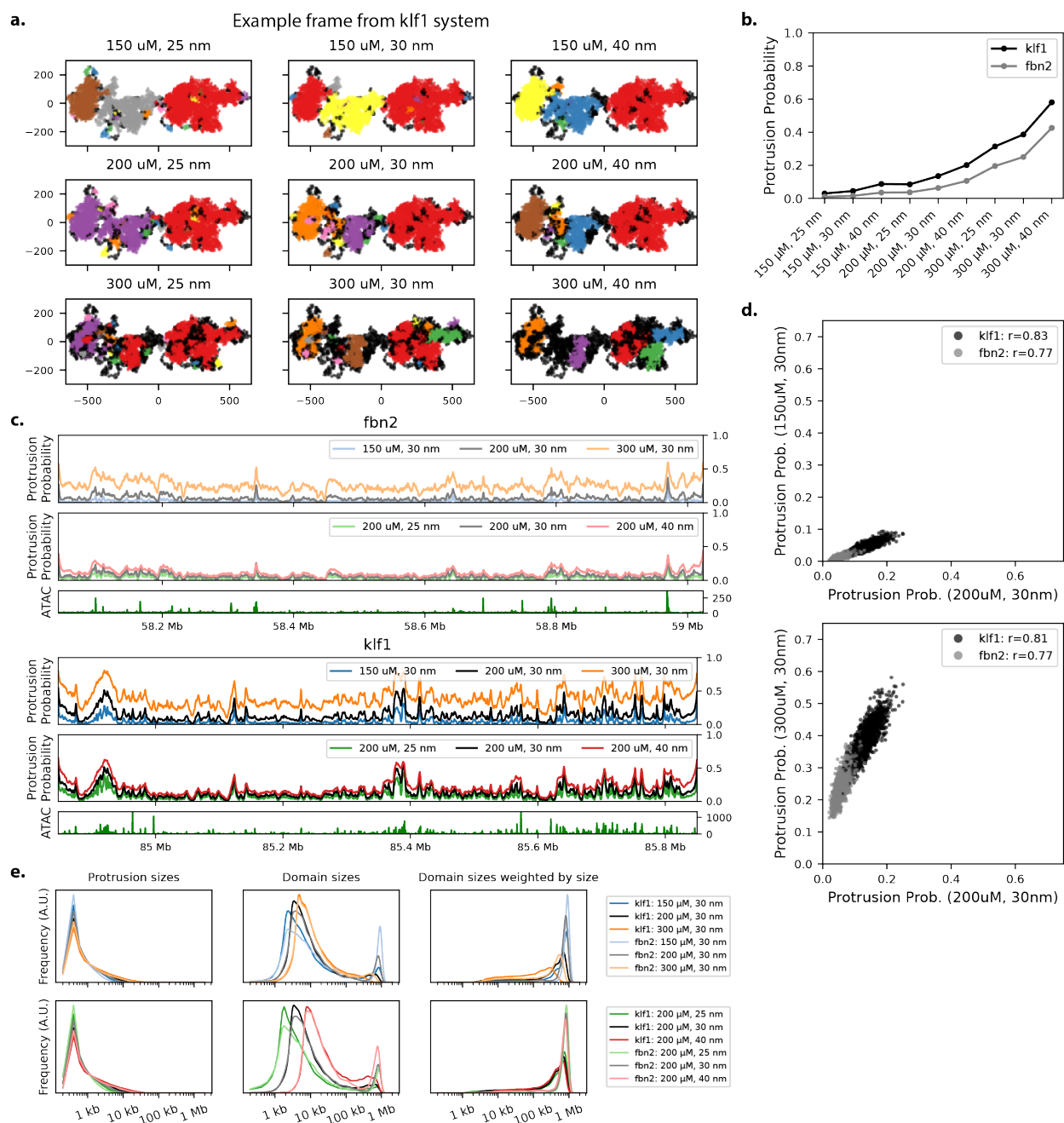


Figure R15: Sensitivity of conclusions to DBSCAN parameters. (a) An example simulation frame showing DBSCAN clustering results with a range of parameters. The center panel (200 μM , 30 nm) shows the clustering used in the main text figures. (b) Average protrusion probabilities for the *klf1* (euchromatin) and *fbn2* (heterochromatin) for a range of parameters. Increasing the concentration or radius increases the protrusion probability. Importantly, these changes are proportional between the two regions. (c) Protrusion probabilities as a function of genomic position for a range of concentrations (top panels) and range of radii (middle panels). (d) Direct comparison of protrusion probabilities with different parameters. The legend shows the Spearman correlation coefficient. (e) Protrusion and domain sizes for a range of concentrations (top) and range of radii (bottom).

- Similarly, can the authors quantify the “appears” in the sentence: “The domains identified by DBSCAN appear to correspond to experimentally observed packing (or TAD-like) domains”?

Response:

We agree that this point would be strengthened by quantitative comparisons. Quantitatively comparing across modalities is in general challenging because the results depend on the approach for translating from structures to experiment output, which is often non-trivial. However, electron tomography with DNA specific labeling (ChromSTEM) presents a good opportunity to compare more quantitatively to experiments [17]. Setting aside potential technical biases due to fixation and incomplete labeling, there is a relatively straightforward relationship between simulated nucleosome coordinates and ChromSTEM images.

We generated synthetic tomograms by extracting a 100 nm thick slice of a conformation, projecting the coordinates on the x-y plane, and representing each nucleosome as a Gaussian density with a standard deviation of 5 nm. This yields images that visually match the ChromSTEM images with broad “packing domains”, which themselves contain finer structures.

We implemented the algorithm proposed by [17] to identify and characterize packing domains. In their approach, they identify packing domains by smoothing and enhancing the contrast of the image and then identifying local maxima. To characterize each domain they consider how the density changes in concentric circles away from the peak center. In the real data, as well as our simulated data, peaks generally begin with a power-law mass scaling that then decays when the radius exceeds the size of the domain. The radius of the domain is thereby identified by the radius at which the density deviates from the initial regime by 5%. Two additional criteria are also applied, though they rarely determine the domain size: an increase in density or increase of the mass scaling exponent above 3. Domain sizes in our structures closely match those from ChromSTEM with a peak between 50 and 100 nm and a tail extending to several hundred nm.

The packing domains identified through this analysis are generally smaller and more numerous than those identified by DBSCAN. This analysis method identifies the shortest dimension of the domain (the minor axis) because when the shortest dimension is exceeded the density begins to drop. Additionally, this analysis places a domain on all local maxima, even if they are connected by continuously high density, so more domains are detected.

We also found that the initial scaling exponents which indicate the fractal dimension of the packing domains (labeled “D” in [Figure R16d](#)) to be in agreement. We find a value of about 2.5 suggesting the presence of internal structure inside the domains. Our simulations suggest that this internal structure is due to the packing domains being formed by nucleosome clutches.

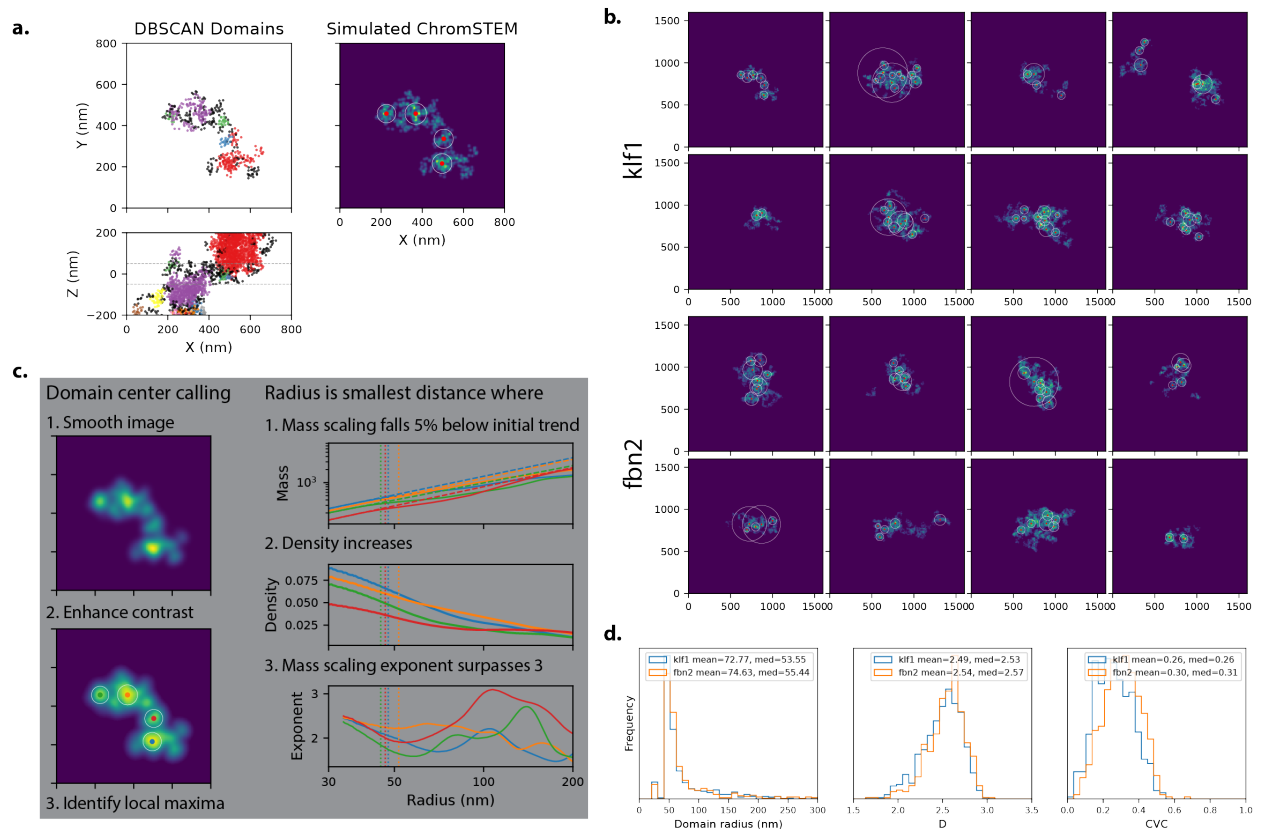


Figure R16: Analysis of simulated ChromSTEM images. (a) A 100 nm thick slice of a simulation frame, colored by DBSCAN domains, and a corresponding simulated ChromSTEM image. The ChromSTEM image has red dots marking the center of domains and white circles enclosing the radius of each domain. (b) An example of domain center calling and radius determination. (c) Example ChromSTEM images for the klf1 and fbn2 systems. (d) Distribution of domain radii, scaling exponents, and chromatin volume concentrations.

We've added [Figure R16](#) as Fig. E3, much of the above response to the Supplementary Information, and the following summary to the Results section

To compare directly to experiments, we generated simulated ChromSTEM images and ran the packing domain characterization approach proposed by [17] (Fig. E3, see Supplementary Information). In brief, their approach classifies packing domains as local maxima in a 2D projection of the chromatin density and identifies the domain radius as the distance from domain center where the density begins to decrease (Fig. E3b). The computed radii are largely between 50 and 100 nm with a tail extending to several hundred nm, consistent with experiments (Fig. E3d). These radii are smaller than for the domains identified by DBSCAN as the imaged slice sometimes does not capture the full domain and the algorithm generally identifies the shortest axis of the

- From the models, could the authors try to obtain simulated image data to directly compare to the results in Ricci et al?

Response: We attempted to generate 2D single molecule localization microscopy images to directly compare to Ricci 2015, but found that the results depend heavily on image generation parameters, most notably the assumed label efficiency. One of their key metrics, the nearest neighbor distance between clutches, is stable across parameters. However, other metrics, like the number of nucleosomes per clutch, are extremely sensitive. Therefore, we found these results hard to interpret and have chosen not to include these results in our revised manuscript. We describe our methods and results in detail below.

We adopt the following approach for simulating SMLM localizations. For each nucleosome, we sample a number of bound antibodies from a Poisson distribution with a specified mean (1/2 by default, justified below). The fluorophore attached end of each antibody is positioned 10 nm from the nucleosome center at a random orientation. For each antibody, a number of localizations is drawn from a geometric distribution. The rate parameter is set to match the mean localizations per antibody observed in calibration experiments (Figure R17b, blue lines). These localizations are further sampled depending on the z position of the antibody according to a Bernoulli distribution with number of trials given by the initial number of localizations and rate given by $\exp(-0.5(z/(w/2))^2)$. Here w is set to 400 nm giving a full width at half maximum of 470 nm, which is typical in SMLM images. With the distribution of localizations per antibody fixed, the the number of antibodies per nucleosome was fit to reproduce the 1, 12, and 24 nucleosome array calibration data (Figure R17b). This resulted in generally good agreement, though we were not able to identify parameters that reproduce the heavy tail observed for the 12 and 24 nucleosome arrays.

The images have blobs of density qualitatively similar to the images reported in Ricci 2015. To compare quantitatively, we implemented the clustering algorithm described in Ricci 2015 (see caption of Figure R17). We find that the distributions of nearest neighbor distances between clusters is similar to those observed experimentally. However, the intensity of clusters (note that scale bars have different maximum values), number of nucleosomes/localizations per cluster, and cluster density are too high by roughly a factor of 10.

This is potentially explained by the mean antibodies per nucleosome being lower in vivo than in the calibration experiments, which seems plausible given the crowded nuclear environment. Decreasing the mean number of antibodies per nucleosome brings the number of nucleosomes/localizations closer to the experimental values, but has only a modest effect on the nearest neighbor distances. This is consistent with a similar set of underlying clusters being detected regardless of the parameters, but more of the nucleosomes in each cluster being detected with higher label efficiencies.

In Ricci 2015, they find that hyper-acetylation generally decreases clutch sizes. Consistent with this, we observe that the heavily acetylated *klf1* region forms smaller clutches than the *fbn2* region, regardless of the chosen parameters.

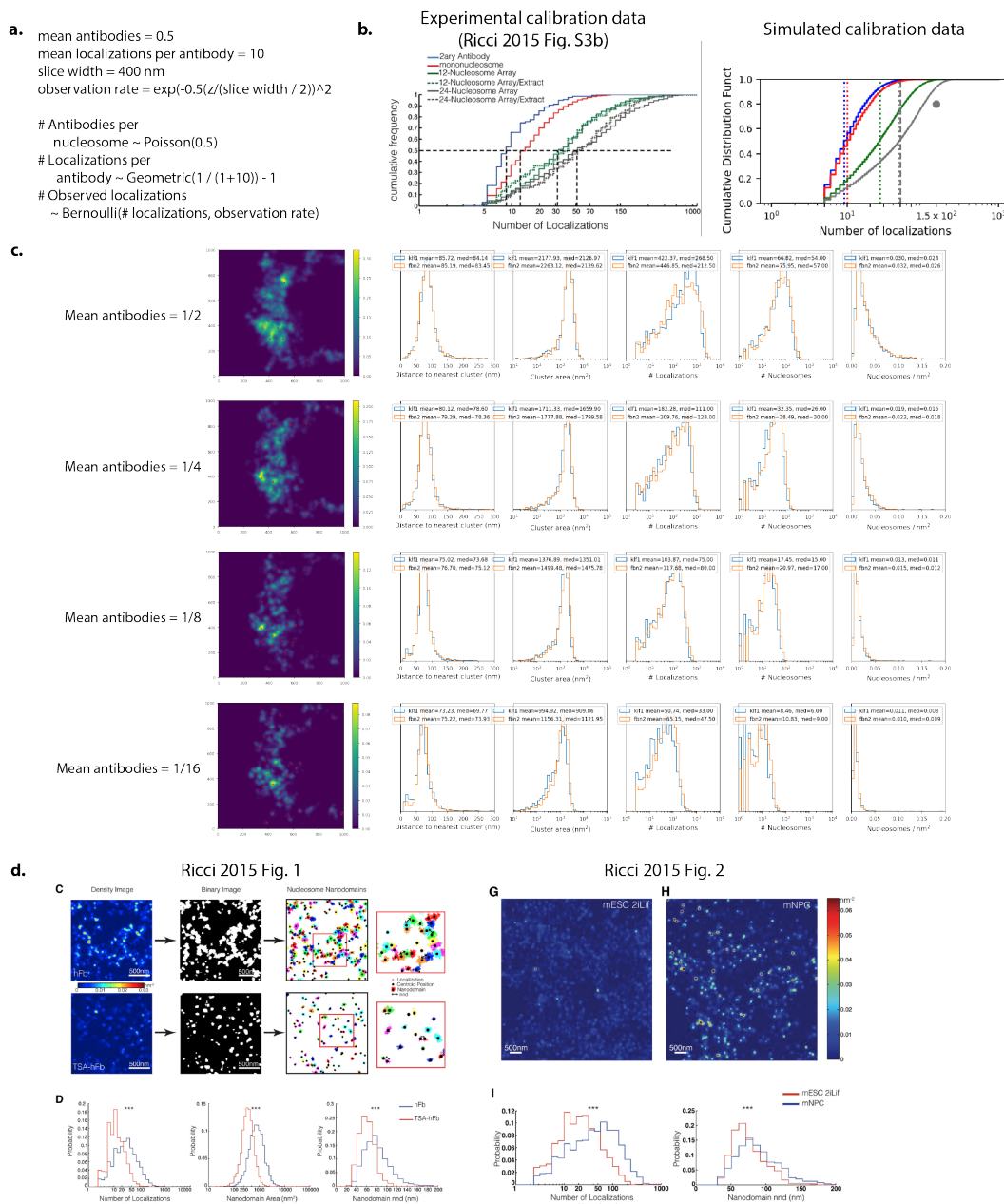


Figure R17: Analysis of simulated single molecule localization microscopy images. (a) Cartoon of approach for generating synthetic single molecule microscopy (SMLM) images. (b) Comparison of experimental calibration data (left) to synthetic data. These data were generated with a mean number of antibodies per nucleosome equal to 1/2. (c) Synthetic SMLM images (left) and cluster metrics (right). In the images, each localization is represented as a Gaussian density with a 9 nm standard deviation as in Ricci 2015. Clustering and cluster quantification were carried out following the methods from Ricci 2015, except that the number of nucleosomes per cluster was computed by mapping each localization to its corresponding nucleosome. The metrics were computed over 160 evenly sampled simulation frames per system. The z coordinates were randomly shifted by a value in (-1000, 1000) to simulate a random slice being in the focal frame in the experimental data. (d) Experimentally determined SMLM images and cluster metrics from Ricci 2015.

- Why 25.6kb resolution was used for balancing micro-C? Any reasoning?

Response: This arbitrary seeming number is $2^7 * 200$. We have clarified this in the text.

The authors could also place their work in context to the recent BioRxiv paper from the Maeshima Lab (<https://www.biorxiv.org/content/10.1101/2025.08.27.672592v1>).

Response: We agree with the reviewer that the manuscript, which was posted online after our manuscript submission, is highly relevant. In fact, many of the reported results provide strong support to our conclusions. Indeed, we are now co-submitting our work with this paper.

Briefly, in this study, the Maeshima group applied super-resolution imaging of the nucleus using euchromatin-specific labeling of histone H3.3. Consistent with our observations (Fig.4), the imaging results show that euchromatic regions form compact domains (Fig. R18A,B).

To further distinguish regulatory elements from gene bodies, the Maeshima group performed immunostaining using antibodies against the C-terminal domain (CTD) of RNA polymerase II (RNAP II) and its phosphorylated forms at serine 5 (Ser5P) and serine 2 (Ser2P). These phosphorylated forms correspond to transcriptionally active paused and elongating RNAP II, respectively, and are commonly used as proxies for regulatory elements and gene bodies. The imaging results revealed that the different RNAP II forms occupy distinct spatial niches, with transcription initiation preferentially localized at domain peripheries (Fig. R18C). The more peripheral localization of paused RNAP II relative to elongating RNAP II is consistent with our structural model, in which regulatory elements preferentially protrude to the surface of compact domains.

To help connect with this work, we've added Fig. 6d (reproduced here as Figure R6c), illustrating that regulatory elements but not gene bodies are enriched in protrusions.

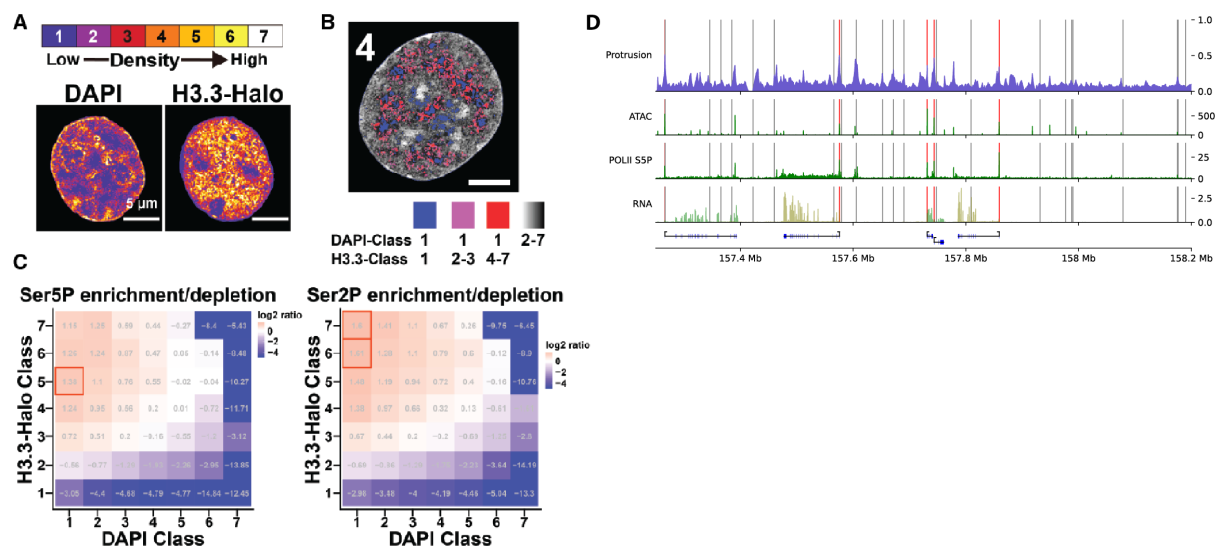


Figure R18: Condensed euchromatin with regulatory protrusions revealed by 3D-SIM. (A) 3D-SIM images of DAPI-stained chromatin (left) and H3.3-HaloTag-labeled nucleosomes (right), colored by signal intensity. (B) Regions with low DAPI but high H3.3 signal (red) support the presence of compact euchromatin domains. (C) Enrichment of Ser5- (initiation, promoter region) and Ser2-phosphorylated (elongation, gene body region) RNAP II across regions classified by DAPI and H3.3 intensity. (D) Protrusion probabilities are enriched at active promoters (with high Ser5P ChIP-seq signal) but not gene bodies.

We've added the following summary of the above to the Discussion section

A recent study that performed super-resolution imaging of euchromatin-specific histone H3.3 provides particularly strong support to this model [16]. Consistent with our observations (Fig. 4), the imaging results show that euchromatin regions form compact domains. To further distinguish regulatory elements from gene bodies, the Maeshima group performed immunostaining against the serine 5 and serine 2 phosphorylated forms of the C-terminal domain of RNA polymerase II (RNAP II). These phosphorylated forms correspond to transcriptionally active paused and elongating RNAP II, respectively, and are commonly used as proxies for regulatory elements and gene bodies. The serine 5 form preferentially localized to domain peripheries and the serine 2 form to domain bodies, consistent with euchromatin forming condensed domains with only regulatory elements preferentially protruding (Fig. 6d).

The abstract may need to be simplified down to the key points.

Response: Thank you, we agree that we let the abstract get too long. We have shortened it to approximately 150 words.

I would suggest the authors explain briefly the “condensability score” (page 3) for the readers who are not familiar with the topic and do not want to read the reference #54.

Response: Thank you for the suggestion. The updated text reads

These dilute positions are highly correlated with condensability scores, which reflect the propensity for nucleosomes to aggregate, ...

I would suggest using kde plots for Figure 4h/i.

Response: Thank you for the suggestion. We’ve updated Figure 4h,i, Figure 5f, and Figure S17e,f to use KDE plots. We agree that this looks much better.

Reviewer #3 (Remarks on code availability):

The main code is not available. Only partial code is and was not reviewed as somehow trivial code.

Response: The simulation code is now available at the [GitHub Repository](#).

References

- [1] Jaynes, E. T. Information Theory and Statistical Mechanics. II. *Physical Review* **1957**, *108*, 171–190.
- [2] Oberbeckmann, E.; Quililan, K.; Cramer, P.; Oudelaar, A. M. In vitro reconstitution of chromatin domains shows a role for nucleosome positioning in 3D genome organization. *Nature Genetics* **2024**, 1–10.
- [3] Goel, V. Y.; Huseyin, M. K.; Hansen, A. S. Region Capture Micro-C reveals coalescence of enhancers and promoters into nested microcompartments. *Nature Genetics* **2023**, *55*, 1048–1056.
- [4] Xie, L.; Dong, P.; Qi, Y.; Hsieh, T.-H. S.; English, B. P.; Jung, S.; Chen, X.; De Marzio, M.; Casellas, R.; Chang, H. Y.; Zhang, B.; Tjian, R.; Liu, Z. BRD2 compartmentalizes the accessible genome. *Nature Genetics* **2022**, *54*, 481–491.
- [5] Spracklin, G.; Abdennur, N.; Imakaev, M.; Chowdhury, N.; Pradhan, S.; Mirny, L. A.; Dekker, J. Diverse silent chromatin states modulate genome compartmentalization and loop extrusion barriers. *Nature Structural & Molecular Biology* **2023**, *30*, 38–51.
- [6] Roux, B.; Weare, J. On the statistical equivalence of restrained-ensemble simulations with the maximum entropy method. *The Journal of Chemical Physics* **2013**, *138*, 084107.
- [7] Li, J.; Hsu, A.; Hua, Y.; Wang, G.; Cheng, L.; Ochiai, H.; Yamamoto, T.; Pertsinidis, A. Single-gene imaging links genome topology, promoter–enhancer communication and transcription control. *Nature Structural & Molecular Biology* **2020**, *27*, 1032–1040.
- [8] Huang, H.; Zhu, Q.; Jussila, A.; Han, Y.; Bintu, B.; Kern, C.; Conte, M.; Zhang, Y.; Bianco, S.; Chiariello, A. M.; Yu, M.; Hu, R.; Tastemel, M.; Juric, I.; Hu, M.; Nicodemi, M.; Zhuang, X.; Ren, B. CTCF mediates dosage- and sequence-context-dependent transcriptional insulation by forming local chromatin domains. *Nature Genetics* **2021**, *53*, 1064–1074.
- [9] Boettiger, A. N.; Bintu, B.; Moffitt, J. R.; Wang, S.; Beliveau, B. J.; Fudenberg, G.; Imakaev, M.; Mirny, L. A.; Wu, C.-t.; Zhuang, X. Super-resolution imaging reveals distinct chromatin folding for different epigenetic states. *Nature* **2016**, *529*, 418–422.
- [10] Murphy, S. E.; Boettiger, A. N. Polycomb repression of Hox genes involves spatial feedback but not domain compaction or phase transition. *Nature Genetics* **2024**, *56*, 493–504.
- [11] Takei, Y.; Yun, J.; Zheng, S.; Ollikainen, N.; Pierson, N.; White, J.; Shah, S.; Thomassie, J.; Suo, S.; Eng, C.-H. L.; Guttman, M.; Yuan, G.-C.; Cai, L. Integrated spatial genomics reveals global architecture of single nuclei. *Nature* **2021**, *590*, 344–350, Number: 7845.
- [12] Daugird, T. A.; Shi, Y.; Holland, K. L.; Rostamian, H.; Liu, Z.; Lavis, L. D.; Rodriguez, J.; Strahl, B. D.; Legant, W. R. Correlative single molecule lattice light sheet imaging reveals the dynamic relationship between nucleosomes and the local chromatin environment. *Nature Communications* **2024**, *15*, 4178.

- [13] Dolsten, G. A.; Cofer, E. M.; Bing, X. Y.; Brack, B.; Curlin, M.; Theesfeld, C. L.; Troyanskaya, O. G.; Levine, M. S.; Pritykin, Y. 3D chromatin structures precede genome activation in *Drosophila* embryogenesis. *Cell Genomics* **2025**, *5*.
- [14] Xingcheng, L.; Bin, Z. Explicit Ion Modeling Predicts Physicochemical Interactions for Chromatin Organization. *eLife* **2023**, *12*.
- [15] Tavares-Cadete, F.; Norouzi, D.; Dekker, B.; Liu, Y.; Dekker, J. Multi-contact 3C reveals that the human genome during interphase is largely not entangled. *Nature Structural & Molecular Biology* **2020**, *27*, 1105–1114.
- [16] Iida, S.; Shimazoe, M. A.; Minami, K.; Tamura, S.; Ashwin, S. S.; Higashi, K.; Nishiyama, T.; Kanemaki, M. T.; Sasai, M.; Schermelleh, L.; Toyoda, A.; Kurokawa, K.; Maeshima, K. Cohesin prevents local mixing of condensed euchromatic domains in living human cells. 2025; <https://www.biorxiv.org/content/10.1101/2025.08.27.672592v1>, ISSN: 2692-8205 Pages: 2025.08.27.672592 Section: New Results.
- [17] Li, Y.; Agrawal, V.; Virk, R. K. A.; Roth, E.; Li, W. S.; Eshein, A.; Frederick, J.; Huang, K.; Almassalha, L.; Bleher, R.; Carignano, M. A.; Szleifer, I.; Dravid, V. P.; Backman, V. Analysis of three-dimensional chromatin packing domains by chromatin scanning transmission electron microscopy (ChromSTEM). *Scientific Reports* **2022**, *12*, 12198.
- [18] Hong, C. K. Y.; Feng, F.; Ramanathan, V.; Liu, J.; Hansen, A. S. Genome structure mapping with high-resolution 3D genomics and deep learning. 2025; <https://www.biorxiv.org/content/10.1101/2025.05.06.650874v1>, Pages: 2025.05.06.650874 Section: New Results.
- [19] Park, S.; Merino-Urteaga, R.; Karwacki-Neisius, V.; Carrizo, G. E.; Athreya, A.; Marin-Gonzalez, A.; Benning, N. A.; Park, J.; Mitchener, M. M.; Bhanu, N. V.; Garcia, B. A.; Zhang, B.; Muir, T. W.; Pearce, E. L.; Ha, T. Native nucleosomes intrinsically encode genome organization principles. *Nature* **2025**, *643*, 572–581.
- [20] Kreysing, J. P.; Cruz-León, S.; Betz, J.; Penzo, C.; Majtner, T.; Schreiber, M.; Turoňová, B.; Lusic, M.; Hummer, G.; Beck, M. Molecular architecture of heterochromatin at the nuclear periphery of primary human cells. 2025; <https://www.biorxiv.org/content/10.1101/2025.04.09.647790v1>, Pages: 2025.04.09.647790 Section: New Results.
- [21] Hou, Z.; Nightingale, F.; Zhu, Y.; MacGregor-Chatwin, C.; Zhang, P. Structure of native chromatin fibres revealed by Cryo-ET in situ. *Nature Communications* **2023**, *14*, 6324.
- [22] Li, Y.; Zhang, H.; Li, X.; Wu, W.; Zhu, P. Cryo-ET study from in vitro to in vivo revealed a general folding mode of chromatin with two-start helical architecture. *Cell Reports* **2023**, *42*, 113134.
- [23] Bohn, M.; Heermann, D. W. Topological interactions between ring polymers: Implications for chromatin loops. *The Journal of Chemical Physics* **2010**, *132*, 044904.