Visualization of loop extrusion by nanoscale 3D DNA tracing in single human cells

K.S. Beckwith[†1], Ø. Ødegård-Fougner[†‡1], N.R. Morero[1], C. Barton[2#], F. Schueder[3,4], W. Tang[5], S. Alexander [1], J-M. Peters[5], R. Jungmann[3,4], E. Birney[2], J. Ellenberg*[1]

Affiliations:

[1] Cell Biology and Biophysics Unit, European Molecular Biology Laboratory, Heidelberg, Germany.

[2] European Bioinformatics Institute, European Molecular Biology Laboratory, Hinxton, United Kingdom.

[3] Faculty of Physics and Center for Nanoscience, Ludwig Maximilian University, Munich, Germany.

[4] Max Planck Institute of Biochemistry, Martinsried, Germany.

[5] Research Institute of Molecular Pathology, Vienna Biocenter, Vienna, Austria

[†] Contributed equally to this work.

* Corresponding author.

Present address: Department of Computer Science and Information Systems, Birkbeck, University of London, London, WC1E 7HX, UK.

‡ Present address: Department of Molecular Cell Biology, Institute for Cancer Research, The Norwegian Radium Hospital, Oslo, Norway.

Abstract

The spatial organization of the genome is essential for its functions, including gene expression, DNA replication and repair, as well as chromosome segregation. Loop extrusion has been proposed to create non-random chromatin structures such as topologically associating domains (TADs). Whether the 3D-folding of DNA in single cells is consistent with this mechanism has been difficult to address in situ. Here, we developed a workflow for nanoscale reconstruction of 3D genome architecture in structurally well-preserved nuclei of single human cells. Our data shows that chromatin behaves as a flexible random coil polymer in the absence of loop extruding enzymes such as Cohesin. Without its boundary factor CTCF, Cohesin forms loops that compact chromatin in an unstructured manner. In the presence of both Cohesin and CTCF, reproducible loop structures are clearly resolved in single cells. The loop architecture and variability of DNA folds we observe in single cells are quantitatively explained by a data-constrained computational model of progressive loop extrusion. Our analysis further reveals how the presence of multiple dynamic and thus comparatively sparse loops cumulatively explains the emergence of topological domains and boundaries across cell populations.

Introduction

The spatial organization of the human genome covers a large hierarchy of spatial scales, from the assembly of individual nucleosomes at the nanometer scale to supra-chromosomal compartments of several micrometers, and plays a central role in regulating genome functions (*1*, *2*). Biochemical chromosome conformation capture methods (3C/Hi-C) have revealed compartmentalization at the megabase (Mb) scale and topologically-associating domains (TADs) of several hundred kilobases (kb) as conserved features of chromatin organization (*3*, *4*). TADs have been shown to depend on the Cohesin complex, a molecular motor that can extrude loops in vitro (*5–8*), while TAD boundaries are defined by CCCTC-binding factor (CTCF) binding sites in the genome (*4*). Polymer simulations recapitulating HiC data have suggested that such structures could arise if chromatin-bound Cohesin would promiscuously extrude loops in chromosomal DNA until encountering CTCF-binding sites (*9–11*), and recent evidence from live cell imaging of fluorescently labelled CTCF-sites indeed suggested that Cohesin activity leads to a stabilization of such loop anchors (*12*). However, direct evidence for the structure of individual loops and their formation by progressive extrusion in situ in single cells is still lacking.

Providing direct structural evidence for loop extrusion in situ has been technically challenging. Genome-wide maps of contact frequencies between genomic loci after crosslinking and sequencing as done in Hi-C result from averages from large cell populations and do not directly measure the three-dimensional (3D) physical structure of chromatin (*13*). Single-cell Hi-C methods provide only modest genomic resolution and sampling, precluding reliable detection of individual loops (*14*, *15*). By contrast, microscopy-based methods such as fluorescence in situ hybridization (FISH) in principle allow direct, single-cell visualization of genome structure. Indeed, the combination of the oligopaint technology, for bar-coded labelling of many genomic loci, with their sequential imaging by super-resolution microscopy have started to enable the visualization of structural features of the genome inside cells, recapitulating the main features of HiC-based contact maps and highlighting single-cell variability (*16–26*).

In this study, we aimed to make use of the power of genome imaging to investigate the nanoscale folding of single chromatin fibres down to individual loops in situ. We therefore developed a scalable high-resolution chromatin tracing workflow combining non-denaturing labelling with oligopaint FISH probes with a resolution of better than 5 kb with automated sequential imaging with a precision of better than 25 nm. Using this approach, which we term "LoopTrace", we systematically investigate the 3D-folding of chromatin in multiple genomic regions from kb to Mb-scale in single human cells. We demonstrate that our approach offers near native preservation of nuclear and chromatin structure and enables precise and reliable 3D chromatin tracing. We directly visualize single loops and explore the relationship between random-coil behavior of the chromatin polymer and Cohesin-dependent active structuring processes from the level of single loops up to more complex long-range interactions in entire TADs. The topology and frequency of the observed chromatin loops we find is fully consistent with a computational model of progressive loop extrusion, and their cumulative effect explains the emergence of TAD-like features at the population level.

Results

LoopTrace enables high precision chromatin tracing in non-denatured cells

To investigate which FISH protocol would best preserve chromatin structure and yet allow the high labelling efficiency required for chromatin tracing (*27*), we compared previously used denaturing FISH approaches with enzymatic, non-denaturing FISH (CO-FISH/RASER-FISH; see Supplementary Note 1 and Supp. Fig. S1) (*28–30*). We found that non-denaturing FISH

combines good hybridization efficiency of oligopaint probes to genomic DNA with near-native preservation of chromatin and nuclear structure, making it a suitable method to investigate the folding of chromosomal DNA at the nanoscale in situ in single human cells.

We next developed a sequential DNA-FISH imaging workflow, LoopTrace, to achieve reliable high-resolution 3D-chromatin tracing at the scale of single DNA loops. To establish a baseline, we first imaged the non-coding genomic region upstream of the MYC gene on the right arm of chromosome 8 predicted to represent minimally constrained, open euchromatin (Supp. Fig. S2a) and containing no CTCF-binding sites (Fig. 1a). To resolve its structure, we targeted ten discrete 5-kb-loci spaced 5 kb apart with a set of oligopaint probes containing locus-specific barcodes. The primary probes were detected by sequentially hybridizing and imaging unique readily exchangeable fluorescent 12-bp imager strands matching each locus barcode (*31*). We found that our workflow provided a reliable and fully removable FISH signal (Fig. 1b; Supp. Fig. S2b) that could be localized with better than 25 nm 3D precision (Supp. Fig. S2c-e). "LoopTrace" thus allowed reliable reconstruction of chromatin paths in 3D and was highly efficient, with over 80% of sequential hybridizations detected in the majority of imaged cells (Supp. Fig. S2f,g). The rapid imager exchange and imaging cycles enabled measurements of more than one hundred loci in ~1000 cells in a typical overnight workflow.

A 100 kb open euchromatic region folds like a random coil

To get first structural insight into the 100 kb euchromatic region upstream of the MYC gene, we compared the 3D-traces between individual cells and found substantial cell-to-cell variability of chromatin folding (Fig. 1c) despite low variability between experimental replicates (Supp. Fig. S2h, average Spearman's r = 0.99). Combining the data from hundreds of cells into a consensus chromatin trace by general Procrustes analysis or pairwise median distance map of all loci (see methods) showed no reproducible substructure (Fig. 1d, Supp. Fig. S2i). As a polymer, the chromatin fibre would be predicted to fold like a random coil in the absence of structuring elements such as CTCF sites or other interactions (*32*). We therefore set up a computational model for an ideal random coil scaled to match our target region and compared it with our experimental data. Consensus traces or pairwise distance maps showed very close agreement between experiment and random coil model (Supp. Fig. 2i, r = 0.98). Fitting the measured 3D distances to a power law gave a scaling exponent $\nu = 0.39$ (Supp. Fig. S2j), lower than the value of 0.5 expected for an ideal random coil or 0.588 for a highly solvated

polymer (*32*), indicating that chromatin folds like a compacted random coil in this 100 kb region without CTCF binding sites.

LoopTrace resolves single chromatin loops between convergent CTCF sites in 3D

To investigate if LoopTrace could resolve chromatin loops in single cells we next targeted 100-300 kb-scale genomic regions on four chromosomes with different predicted looping architectures due to the presence of convergent CTCF sites (Supp. Fig S3a) (*4*). To resolve their structure, we tiled 25-30 FISH probes at 4 or 10 kb resolution and imaged them in diploid RPE-1 and the on average triploid (*33*) HeLa cancer cell line, respectively. Converting our single cell 3D traces into pairwise distance and contact frequency maps (defined as 3D distances <120 nm) qualitatively matched Hi-C experiments (compare Supp. Fig. S3a and b) and showed conserved genome structure between the two human cell lines ($r = 0.95 \pm 0.02$), further validating our approach. Inspection of single traces highlighted the heterogeneity of single cell folds that exhibited the prominent features of the population averages such as compacted loop-like structures (Supp. Fig. S3b).

To gain a detailed quantitative understanding of the physical 3D structure of a single chromatin loop, we visualized probes targeting a 120 kb region with a pair of convergent CTCF sites 100 kb apart near the MEOX1 gene on chromosome 17 with 10 kb genomic resolution (Fig. 2a). Consensus traces from more than 2000 single cells resolved loop features with the CTCF sites in close proximity in 3D (Fig. 2b) and were clearly distinct from the random coil like fold of the unstructured 100 kb region near the MYC gene on chromosome 8 (compare Fig. 1d-e). By dividing the data for the MEOX1 region into looping and non-looping subsets we found about 50% of RPE-1 and HeLa cells to contain a clear consensus loop fold with a distance between the CTCF sites of less than 120 nm. Inspection of single cell loop traces showed proximity of the CTCF sites as the main conserved features but revealed a highly variable internal structure (Fig. 2c) with 3D loop diameters that varied broadly around 150 nm (Fig. 2b). Consistent folds exhibiting clear conservation of loop structures with close proximity between convergent CTCF sites in ~50% of cells (39% in RPE-1, 55% in HeLa) and high variability within the loop was also found in a second 100 kb region on chromosome 9, which we traced at a genomic resolution of 4 kb (Supp. Fig. S3d). In summary, single 100 kb chromatin loops are present in about 50% of the cells between one pair of convergent CTCF sites. These loops consistently exhibit close proximity of the loop base sequences, but a high cell to cell variability inside the loop. The physical size of a 100 kb scale loop is on average 150 nm.

Computational ordering of 3D folds suggests that increasing loop formation drives compaction

With single cell traces revealing both looping and non-looping conformations, we next investigated if we could computationally sort the traces into structurally similar groups to gain a better understanding of their variation and potential transitions. To this end, we ordered the traces from the region near the MEOX1 gene on Chr17 in an unsupervised manner using spectral seriation (see methods). Inspection of the seriated difference matrix confirmed a large conformational variability (Fig. 2c). Interestingly, consensus traces from six consecutive seriation groups revealed a progressive transition from an open, random-coil like fold via progressively more compacted folds until the fully looped state was reached (Fig. 2c). A similar transition from open to looping folds occurred for the second 80 kb looping region on chromosome 9 (Supp. Fig. S4a). By contrast, simulated random coils showed lower conformational variability and no transitions into a looping state, with all seriation groups representing open, random-coil like conformations (Supp. Fig. S4b). Together, the seriation analysis reveals progressively more compact conformational states consistent with the idea of increased loop formation structuring an otherwise randomly coiling polymer.

Single loops are Cohesin and CTCF dependent

If the loop-like 3D folds that we found in 50% of the cells and that can be ordered consistent with a progressive formation process, indeed result from active extrusion, they should depend on the presence of the main loop extruder Cohesin and its positioning factor CTCF. To test this, we acutely depleted the Rad21 or CTCF in HeLa cells where these genes had been homozygously tagged with a degron (HeLa Rad21-EGFP-AID and HeLa CTCF-EGFP-AID (*5*)). Consistent with previous HiC analysis (*5*, *6*), our 3D tracing data clearly showed the loss of looping features in the traced regions after protein depletion (Fig. 2c, Supp. Fig. S3e). Cohesin depletion led to an almost complete loss of loop like compact folds (Fig. 2c), primarily by loss of proximity between loop anchors (Supp. Fig. S5a) and led to an increase in the scaling exponent v from 0.39±0.11 for WT cells to 0.49±0.05 for Cohesin depleted cells (Supp. Fig. S5b), indicating nearly ideal random coil scaling. Seriation analysis in Cohesin-depleted cells showed random coil-like folds similar to simulated random coils, with lower conformational variability than the same region in WT cells (Supp. Fig. S4c, compare to Fig. 2c). CTCF depletion also substantially reduced loop folds below 20% (Fig. 2c, Supp. Fig. S5a), however, overall, the traced regions remained significantly more compact than after Cohesin depletion (Supp. Fig. S5b, v = 0.35±0.05), likely due to continuous Cohesin activity without sequence

specific stabilization. Such promiscuous loop formation was further supported by tracing a more open region near the HIST1 gene cluster on chromosome 6 predicted to have only short range loops (Supp. Fig. S3a,b), where CTCF loss rather led to increased chromatin compaction and increased frequency of long-range contacts (Supp. Fig. S3e, S5b), presumably due to Cohesin being able to create longer loops without halting at CTCF sites. Overall, our single cell tracing data clearly shows that single chromatin loops depend on Cohesin and that chromatin up to the 300 kb scale closely resembles an ideal random coil without Cohesin. They further indicate that Cohesin based loop extrusion continues in the absence of CTCF, which leads to loss of positioning at loop anchors.

The 3D fold of TAD-scale genomic regions is characterized by sparse and open looping conformations

Given that we could capture an ensemble of variable and therefore likely dynamic (*12*, *34*) structural states of random coil as well as looping chromatin at the ~100 kb scale, we next explored if LoopTrace could be used to resolve the architecture of TAD-scale (1-2 Mb) regions predicted to be more compact by bulk biochemical approaches (*4*). To this end, we traced four 1.3 to 1.8 Mb sized regions with typical TAD-like features in available HiC data on chromosomes 2, 4 and 7 (Supp. Fig. S4a) reported to be either transcriptionally active (*35*) (chr2-A, chr7-HoxA), silent (chr4-B), or have mixed activity (chr7-SHH). We targeted these regions in HeLa cells with 10-12 probe sets spaced at around 100 kb targeted to validated CTCF sites and intermediate loci (Fig. 3a; Supp. Fig. S6a).

3D chromatin tracing of more than one thousand cells for each region identified close physical contacts, whose frequency was consistent with the available HiC data (Supp. Fig. S6b). Representing our single cell data in consensus traces highlighted compact structural features corresponding to long-range contacts bounded by convergent CTCF sites (Fig. 3b; Supp. Fig. S6c, d). These contacts spanned up to several hundred kb, and, very similar to single loops at the 100 kb scale, were also dependent on Cohesin and CTCF, as demonstrated by acute protein depletion and retracing (Fig. 3b; Supp. Fig. S6c, d). Interestingly, such long-range contacts could be further enhanced by acute depletion of WAPL (in HeLa WAPL-Halo-AID cells), which was often accompanied by a reduction of short-range contacts (Supp. Fig. S6). These structural findings are consistent with reports that WAPL depletion increases Cohesin's residence time and loop extrusion activity (*5*, *36*, *37*).

Averaging data from many single cell traces clearly reproduced overall compact TAD domains and boundaries (Supp. Fig. S6b). Exploring the 3D fold of single cell traces interactively revealed rich and informative structural features. Single cell folds consistently showed an overall rather open architecture of the megabase scale regions, with only sparse looping interactions (Fig. 3b). Quantifying the number of contacts closer than 150 nm per trace showed that a megabase scale region on average contained only ~3 close contacts between the sampled loci (Supp. Fig. S6e). This average contact number was significantly reduced to below 2 upon Cohesin depletion, and significantly increased to above 4 by WAPL depletion (Supp. Fig. S6e). By contrast, CTCF depletion did not affect the number of contacts but rather abolished their reproducible position, leading to similarly compact but more variable folds (Fig. 3b; Supp. Fig. S6c, e), consistent with our findings from the 100-300 kb regions (compare Fig. 2b; Supp. Fig. S3b). Even in WAPL-depleted cells, where contact number was maximal, single cell traces revealed multiple, extended looping features rather than overall highly compacted or collapsed "TADs" (Fig. 3c). Interestingly, most long-range contacts (over 500 kb) resulted from multiple stacked internal contacts, suggesting that the long-range contacts mainly emerge from combinations of several intermediate loops rather than single very large loops, which occur more rarely (Fig. 3b, Supp. Fig. S6f).

To mine the wealth of more complex TAD-scale 3D chromatin folds from thousands of single cells statistically, we performed unsupervised clustering of all single traces using a Bayesian Gaussian mixture model (see methods) to identify classes of 3D folds with similar overall structural features, using the Chr7-SHH locus as an example (Fig. 3c; Supp. Fig. S6g). Comparing consensus traces of different clusters revealed that they are enriched in specific pairwise combinations of contacts between loci that connected the CTCF site at the upstream TAD boundary of the SHH gene with a different downstream loop anchor in each cluster (Fig. 3a, c, Supp. Fig. S6g). The presence of these specific loops led to larger physical distances for the other loci (Supp. Fig. S6g), showing how a single strong anchor on one side of a loop can generate an effective TAD boundary and yet allow a large heterogeneity of internal TAD topologies.

3D folds of TADs are consistent with a quantitative model for progressive loop extrusion

The clear dependence of loop structures on the presence of Cohesin and CTCF and their increased size and abundance upon WAPL depletion indicated that these structures are most likely generated by active loop extrusion. To quantitatively test if the abundance and fold of

the 3D structures we report here are consistent with the loop extrusion capacity of a single human cell, we performed polymer simulations, similar to previous approaches used to model HiC data (*9*, *10*). We constrained the parameters of the model with the quantitative data available for the loop extrusion machinery (Supplemental Note 2; Supp. Fig. S7a), including cellular concentration (*38*), in vivo chromatin residence times (*38–40*) and estimated extrusion rate of Cohesins (*6*, *34*) as well as number and orientation of CTCF binding sites in the human genome (*4*, *38*). To estimate the remaining unknown parameters, especially the fraction of Cohesin complexes engaged in active loop extrusion, we used our traces of 3D chromatin structure after Cohesin depletion as a reference for the random coil polymer state and our traces of chromatin after CTCF depletion as a reference for promiscuously extruding Cohesins (Supp. Fig. S7b, c). Intriguingly, the experimental data were best reproduced by a model where only 20% of Cohesins, corresponding to the abundance of STAG1 containing complexes (*38*), were active. By contrast, the model predicted a much stronger than our observed compaction if all Cohesins (*38*) , were used to simulate the structural outcome. This is consistent with previous biochemical reports that STAG1-Cohesin alone can support loops and TADs (*39*).

Overall, the model fit our single cell 3D structural data with high accuracy (Supp. Fig. S7d-g, $r_{RPE-1} = 0.92 \pm 0.05$, $r_{HeLa} = 0.88 \pm 0.06$) and explained the comparatively sparse occurrence of loops, which is predictive of the fraction of Cohesins engaged in active loop extrusion. In more detail, the model predicts the heterogeneity of internal loop topologies we observed in our 3D traces (Fig. 2b; Supp. Fig. S7d,e), that ~100 kb sized loops are created by a single Cohesin and that longer range contacts result from the action of multiple Cohesins (Supp. Fig. S7i), explaining the stacked loops we found in our traces. Furthermore, the model also predicted the increased contact abundance and loop length upon increasing Cohesin's residence time by WAPL depletion (Supp. Fig. S7h). Overall, the model fits the data statistically and in rather surprising structural detail, assuming about 20% of active Cohesin. This allows to calculate that per Mb of genomic DNA on average about 4 Cohesins interact with a dynamically exchanging pool of on average about 40 CTCF molecules to generate the structural features we measured with the LoopTrace approach.

Discussion

The "LoopTrace" workflow we introduce here provides a precise, robust and scalable approach to investigate the nanoscale folding principles of the genome in situ in structurally well-preserved nuclei of single human cells. While our method does not suffer from the micrometer scale structural artefacts of denaturation-based FISH methods and shows good conservation of

TAD sized domains at the 100 nm scale, we note that even non-denaturing FISH methods show mild alterations of nuclear structure, such as slight swelling of nucleoli and moderate changes in nuclear shape. In the future, hopefully even more native methods to target sequence specific probes to genomic DNA can be developed.

3D tracing with up to 25 nm/5 kb resolution provides direct, in situ measured structural data, which is key to understanding of the folding principles of the genome. Based on indirect HiC data and computer simulations, Cohesin driven loop extrusion had been proposed to be the mechanism through which this SMC protein and its regulators drive TAD formation (*34*). However, a direct visualization of the proposed looping conformations of genomic DNA has so far been lacking. Our 3D tracing data in single cells now provides such direct evidence. It shows that in the absence of Cohesin, chromatin folds like a non-interacting random coil polymer. Onto this rather open and dynamic structure, loops are superimposed that depend on Cohesin activity and are stabilized at specific genomic positions by CTCF. Interestingly, in single cells the maximal available loop extrusion capacity is used at an intermediate level, and the cell could adjust this to both up and down, as illustrated by our depletions of WAPL and Rad21, to dynamically regulate genome structure. Our loop extrusion modelling predicts that only one fifth of Cohesins, corresponding in abundance and dynamic properties to the Cohesin-STAG1 isoform, would suffice to create the measured structures at the 100-2000 kb scale which is supported by biochemical approaches (*39*, *41*), and could suggest alternative functions for the STAG2 isoform.

We find that short-range (~100 kb) loops between pairs of convergent CTCF sites are present in about 50% of cells, consistent with estimates from single-cell HiC (*14*) and live cell imaging (*42*). Longer-range contacts are generally less frequent and sparse, around 1 per megabase region and result from stacking of around three intermediate loops. The complete set of TAD features, i.e. strong boundaries and high frequency of internal interactions thus does not exist in any single cell at any time point. Rather, TAD boundaries arise from the cumulative effect of dynamically positioned sparse looping interactions inside the TAD in cell population data.

We anticipate that the LoopTrace approach that combines near-native oligopaint FISH with scalable high-precision chromatin tracing will be a valuable tool to directly investigate the structure-function relationship of the genome at the nanoscale in single cells. Scaled to whole chromosomes and the whole genome (*27*), such methods should enable a deeper understanding

of the link between genome architecture and the functional state of individual cells in health and disease.

References

1. D. U. Gorkin, D. Leung, B. Ren, *Cell Stem Cell*. **14**, 762–775 (2014).
2. J. Dekker, L. Mirny, *Cell*. **164**, 1110–1121 (2016).
3. J. R. Dixon *et al.*, *Nature*. **485**, 376–380 (2012).
4. S. S. P. Rao *et al.*, *Cell*. **159**, 1665–1680 (2014).
5. G. Wutz, *The EMBO Journal*. **36**, 3573–3599 (2017).
6. S. S. P. Rao *et al.*, *Cell*. **171** (2017)
7. I. F. Davidson *et al.*, *Science*. **366**, 1338–1345 (2019).
8. Y. Kim, Z. Shi, H. Zhang, I. J. Finkelstein, H. Yu, *Science*. **366**, 1345–1349 (2019).
9. G. Fudenberg *et al.*, *Cell Reports*. **15**, 2038–2049 (2016).
10. A. L. Sanborn *et al.*, *PNAS*. **112**, E6456–E6465 (2015).
11. L. Vian *et al.*, *Cell*. **173**, 1165-1178.e20 (2018).
12. M. Gabriele *et al.*, *Science*. **376**, 496–501 (2022).
13. S. Sati, G. Cavalli, *Chromosoma*. **126**, 33–44 (2017).
14. T. Nagano *et al.*, *Nature*. **502** (2013)
15. T. J. Stevens *et al.*, *Nature*. **544**, 59–64 (2017).
16. B. J. Beliveau *et al.*, *PNAS*. **109**, 21301–21306 (2012).
17. B. J. Beliveau *et al.*, *Nature Communications*. **6** (2015)
18. Y. Markaki *et al.*, *BioEssays*. **34**, 412–426 (2012).
19. A. N. Boettiger *et al.*, *Nature*. **529**, 418–422 (2016).
20. S. Wang *et al.*, *Science*. **353**, 598–602 (2016).
21. B. Bintu *et al.*, *Science*. **362** (2018)
22. G. Nir *et al.*, *PLOS Genetics*. **14**, e1007872 (2018).
23. A. M. Cardozo Gizzi *et al.*, *Molecular Cell*. **74**, 212-222.e5 (2019).
24. L. J. Mateo *et al.*, *Nature*. **568**, 49–54 (2019).
25. J.-H. Su, P. Zheng, S. S. Kinrot, B. Bintu, X. Zhuang, *Cell*. **182**, 1641-1659.e26 (2020).
26. Y. Takei *et al.*, *Nature*. **590**, 344–350 (2021).

27. C. Barton *et al.*, *PLOS Computational Biology*. **14**, e1006002 (2018).

28. E. Goodwin, J. Meyne, *CGR*. **63**, 126–127 (1993).

29. J. M. Brown *et al.*, *Nature Communications*. **9**, 3849 (2018).

30. J. M. Brown, S. De Ornellas, E. Parisi, L. Schermelleh, V. J. Buckle, *Nat Protoc*, 1–26 (2022).

31. F. Schueder *et al.*, *Angewandte Chemie International Edition*. **56**, 4052–4055 (2017).

32. G. Fudenberg, L. A. Mirny, *Curr Opin Genet Dev*. **22**, 115–124 (2012).

33. J. J. M. Landry *et al.*, *G3 Genes|Genomes|Genetics*. **3**, 1213–1224 (2013).

34. I. F. Davidson, J.-M. Peters, *Nat Rev Mol Cell Biol*. **22**, 445–464 (2021).

35. K. Xiong, J. Ma, *Nat Commun*. **10**, 5069 (2019).

36. J. H. I. Haarhuis *et al.*, *Cell*. **169**, 693-707.e14 (2017).

37. S. Kueng *et al.*, *Cell*. **127**, 955–967 (2006).

38. J. Holzmann *et al.*, *eLife*. **8**, e46269 (2019).

39. G. Wutz *et al.*, *eLife*. **9**, e52091 (2020).

40. A. S. Hansen, I. Pustova, C. Cattoglio, R. Tjian, X. Darzacq, *eLife*. **6**, e25776 (2017).

41. V. Casa *et al.*, *Genome Res.* **30**, 515–527 (2020).

42. P. Mach *et al.*, *Nat Genet*, 1–12 (2022).

43. E. M. Darrow *et al.*, *PNAS*. **113**, E4504–E4512 (2016).

44. F. A. Ran *et al.*, *Cell*. **154**, 1380–1389 (2013).

45. A. Yesbolatova, T. Natsume, K.-I. Hayashi, M. T. Kanemaki, *Methods*. **164–165**, 73–80 (2019).

46. J. D. Gans, M. Wolinsky, *Nucleic Acids Res*. **36**, e74 (2008).

47. B. J. Beliveau *et al.*, *PNAS*. **115**, E2183–E2192 (2018).

48. J. N. Zadeh *et al.*, *Journal of Computational Chemistry*. **32**, 170–173 (2011).

49. I. Georgakopoulos-Soares, O. Yizhar-Barnea, I. Mouratidis, M. Hemberg, N. Ahituv, *Genome Biology*. **22**, 245 (2021).

50. A. Untergasser *et al.*, *Nucleic Acids Research*. **40**, e115 (2012).

51. Q. Xu, M. R. Schlabach, G. J. Hannon, S. J. Elledge, *PNAS*. **106**, 2289–2294 (2009).

52. J. R. Moffitt, X. Zhuang, in *Methods in Enzymology*, G. S. Filonov, S. R. Jaffrey, Eds. vol. 572 of *Visualizing RNA Dynamics in the Cell*, pp. 1–49.

53. L. J. Mateo, N. Sinnott-Armstrong, A. N. Boettiger, *Nat Protoc*. **16**, 1647–1713 (2021).

54. W. Xiang *et al.*, *The Journal of cell biology*. **217**, 1973–1984 (2018).

55. P. Virtanen *et al.*, *Nature Methods*. **17**, 261–272 (2020).

56. C. Stringer, T. Wang, M. Michaelos, M. Pachitariu, *Nature Methods*. **18**, 100–106 (2021).

57. A. Z. Politi *et al.*, *Nature protocols*. **13**, 1445–1464 (2018).

58. The pandas development team, pandas-dev/pandas: Pandas (2020)

59. S. van der Walt *et al.*, *PeerJ*. **2**, e453 (2014).

60. F. Pedregosa *et al.*, *Journal of Machine Learning Research*. **12**, 2825–2830 (2011).

61. Dask Development Team, Dask: Library for dynamic task scheduling (2016)

62. napari contributors, napari: a multi-dimensional image viewer for python. (2019)

63. E. Czech, B. A. Aksoy, P. Aksoy, J. Hammerbacher, *BMC Bioinformatics*. **20**, 448 (2019).

64. L. McInnes, J. Healy, J. Melville, *arXiv:1802.03426 [cs, stat]* (2020)

65. M. L. Waskom, *Journal of Open Source Software*. **6**, 3021 (2021).

66. K. S. Arun, T. S. Huang, S. D. Blostein, *IEEE Trans. Pattern Anal. Mach. Intell.* **9**, 698–700 (1987).

67. J. E. Atkins, E. G. Boman, B. Hendrickson, *SIAM J. Comput.* **28**, 297–310 (1998).

68. L. Lopez-Delisle *et al.*, *Bioinformatics*. **37**, 422–423 (2021).

69. E. Miron *et al.*, *Science Advances*. **6**, eaba8811 (2020).

70. F. Ochs *et al.*, *Nature*. **574**, 571–574 (2019).

71. I. Solovei *et al.*, *Experimental Cell Research*. **276**, 10–23 (2002).

72. T. Nozaki *et al.*, *Molecular Cell*. **67**, 282-293.e7 (2017).

73. P. E. Rouse, *J. Chem. Phys.* **21**, 1272–1280 (1953).

74. H. A. Shaban, R. Barth, L. Recoules, K. Bystricky, *Genome Biology*. **21**, 95 (2020).

75. A. Zidovska, D. A. Weitz, T. J. Mitchison, *Proceedings of the National Academy of Sciences*. **110**, 15555–15560 (2013).

76. V. Levi, Q. Ruan, M. Plutz, A. S. Belmont, E. Gratton, *Biophys J*. **89**, 4275–4285 (2005).

77. R. Barth, K. Bystricky, H. A. Shaban, *Science Advances*. **6**, eaaz2196 (2020).

78. L. Nanni, S. Ceri, C. Logie, *Genome Biology*. **21**, 197 (2020).

79. I. F. Davidson *et al.*, BioRxiv:507093 (2022)

80. Y. Li *et al.*, *Nature*. **578**, 472–476 (2020).

Acknowledgements: We thank Merle Hantsche-Grininger for developing the electroporation protocol for replication domain labelling and critical reading of the manuscript; Franziska Kundel, Jonas Ries and Virginie Uhlmann for discussions; Magda Bienko and Jill Brown for sharing of FISH protocols; the EMBL Advanced Light Microscopy Facility (ALMF) for microscope support; Christian Tischer and Jean-Karim Hériché for help with data/metadata handling, and the electronic workshop and mechanical workshop at EMBL for assistance with the microfluidics setup.

Funding: This work was supported by grants from the National Institutes of Health Common Fund 4D Nucleome Program (Grant U01 EB021223 / U01 DA047728) to J.E. and E.B., as well as by the The Paul G. Allen Frontiers Group through the Allen Distinguished Investigator Program to J.E. and R.J. K.S.B. was supported by the Alexander von Humboldt foundation and EMBL. Work in the laboratory of J.-M.P. has received funding from Boehringer Ingelheim, the Austrian Research Promotion Agency (Headquarter grant FFG-852936), the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreements No 693949 and No 101020558), the Human Frontier Science Program (grant RGP0057/2018) and the Vienna Science and Technology Fund (grant LS19-029).

Author contributions:

Conceptualization: JE and EB
Methodology: ØØF, KSB, NRM, CB, FS, TW
Software: KSB, CB
Validation: KSB, NRM, TW
Formal analysis: KSB
Investigation: ØØF, KSB, NRM
Data Curation: KSB, CB
Writing – original draft preparation: ØØF, KSB, JMP, JE
Writing – review and editing: All authors
Visualization: KSB
Supervision: SA, JMP, EB, RJ, JE
Funding acquisition: SA, JMP, EB, RJ, JE

Competing interests:

Authors declare no competing interests.

Data and materials availability: Source image data and processed data from this study are available on the EMBL EBI Bioimage Archive/Biostudies at the following website: https://www.ebi.ac.uk/biostudies/studies/S-BIAD59. Publicly available datasets used in this study were GSE71831 (RPE-1 HiC) (*43*), GSE63525 (HeLa HiC) (*4*), ENCSR000DVI (RPE CTCF CHIP-seq), ENCODE: ENCSR000EDE (HeLa RAD21 CHIP-seq) and ENCODE ENCSR000AOA (HeLa CTCF CHIP-seq). Custom code for image acquisition and microfluidic control is available online at git.embl.de/grp-ellenberg/tracebot. Image analysis and LoopTrace analysis software is available online at git.embl.de/grp-ellenberg/looptrace. The primary oligo FISH probe design software is available at gitlab.com/Chromotrace_utils/probe-design.

Supplementary Materials

Materials and Methods

Supplementary Text

Figs. S1 to S8

References (*44–80*)

Data S1 to S2

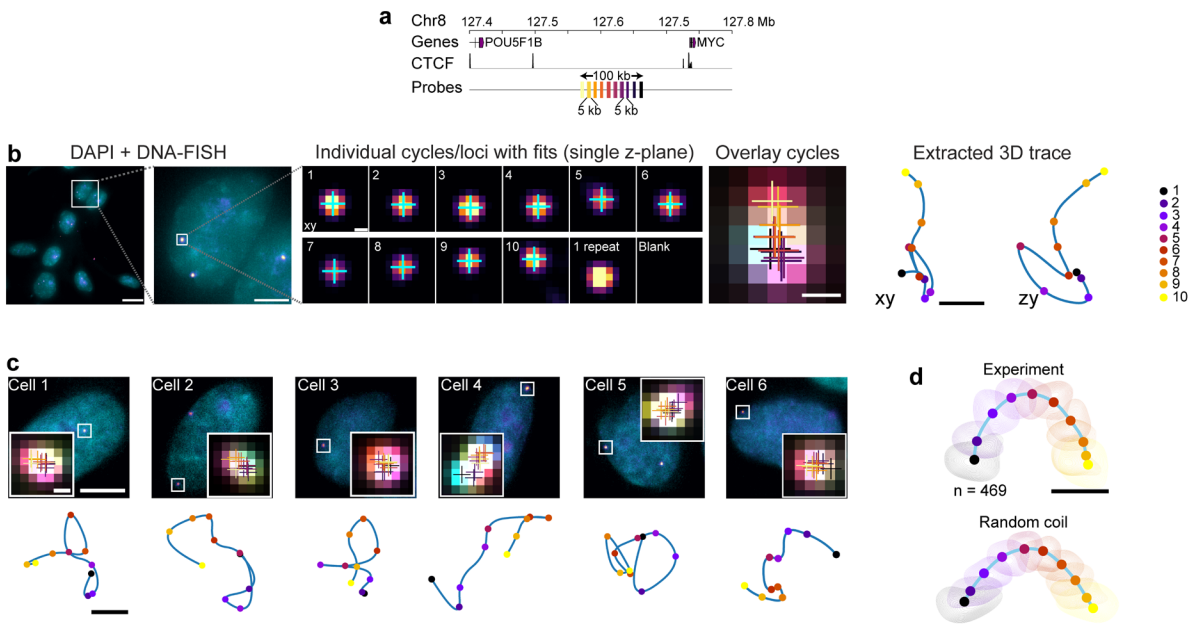

Fig. 1: LoopTrace enables high precision chromatin tracing in non-denatured cells. **a,** Overview of a non-coding region upstream of MYC gene selected for DNA tracing pipeline validation with genes and CTCF sites (RPE Chip-seq) highlighted. The probes cover 10 regions of 5 kb with 5-kb-gaps, spanning a total of 100 kb. **b,** Examples of FISH signal, projection of sequential imaging steps and centroid fits, and resulting 3D chromatin traces. **c**, Overlaid FISH signal with fits and resulting chromatin traces from six representative cells. **d,** 2D-projected consensus traces of experimental and simulated data by general Procrustes analysis, overlaid with kernel density estimates (KDE) of aligned single trace data (see methods). Data from n=469 traces from 3 independent experiments or n=469 simulated random walks. Scale bars: 20 (overview) and 5 µm (nuclei), 200 nm (insets) and 100 nm (traces).

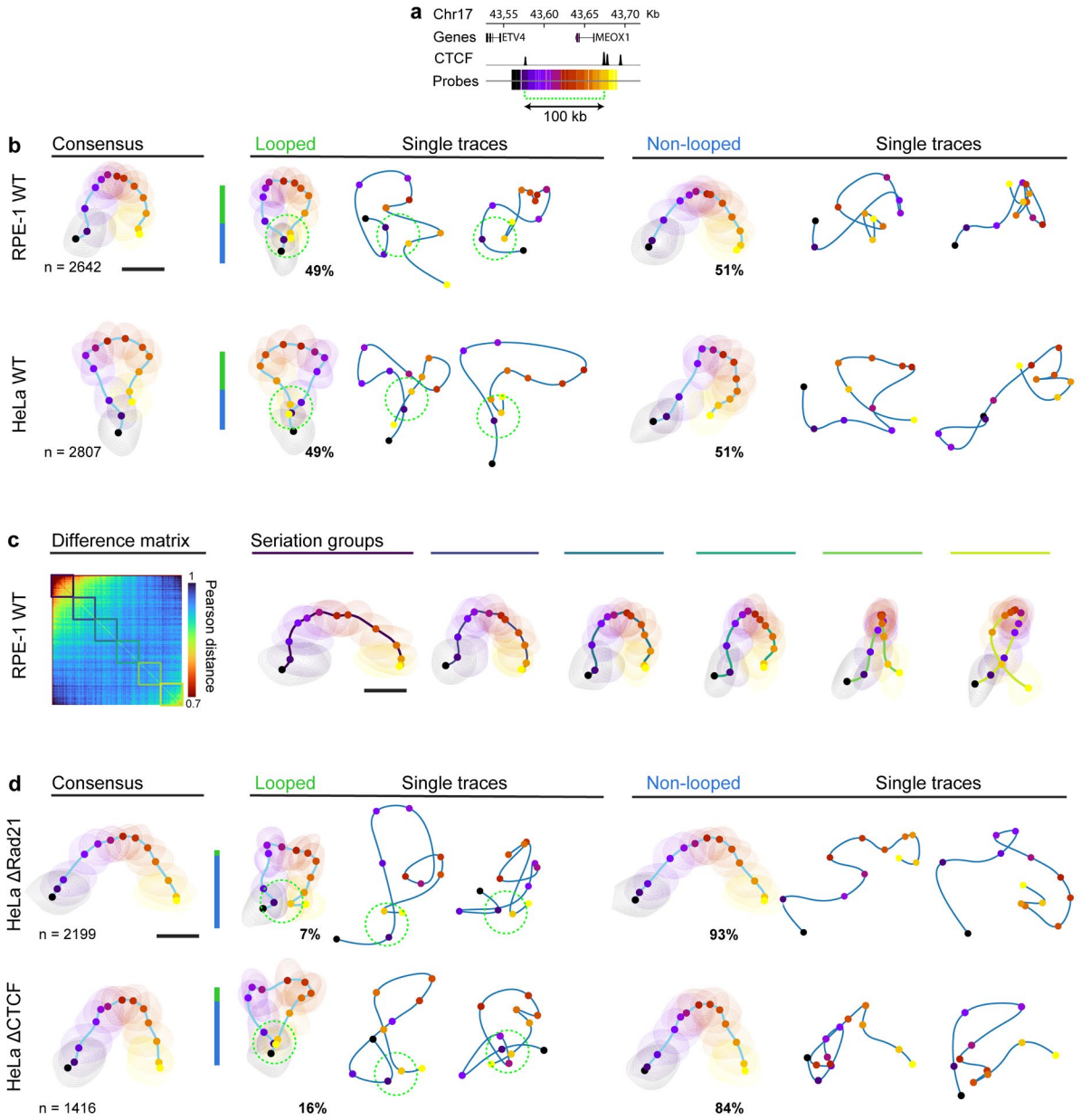

Fig. 2: LoopTrace resolves single CTCF and Cohesin-anchored loops.

a, Region targeted for tracing around the MEOX1 gene on chromosome 17. A pair of convergent CTCF sites predicted to form loops are indicated by the dashed line. **b,** Consensus traces, looping and non-looping subsets of traces and representative single trace examples from the region indicated in (**a**) from RPE-1, HeLa Kyoto wild type (HeLa WT) cells. Proximity (<120 nm) between the positions of the looping CTCF-sites indicated by dashed circles. **c**, Overall difference matrix measuring the structural heterogeneity among single traces after sorting by seriation (see methods), and the consensus traces from each seriation group indicated by colored squares on the difference matrix. n = 440 traces per group. **d**, Consensus traces of all traces and looping and non-looping fractions from HeLa Rad21-EGFP-AID or HeLa CTCF-EGFP-AID cells depleted of cohesin or CTCF for 2h by auxin treatment. The number of traces for each experimental condition (2 replicates) are indicated. Scale bars 100 nm.

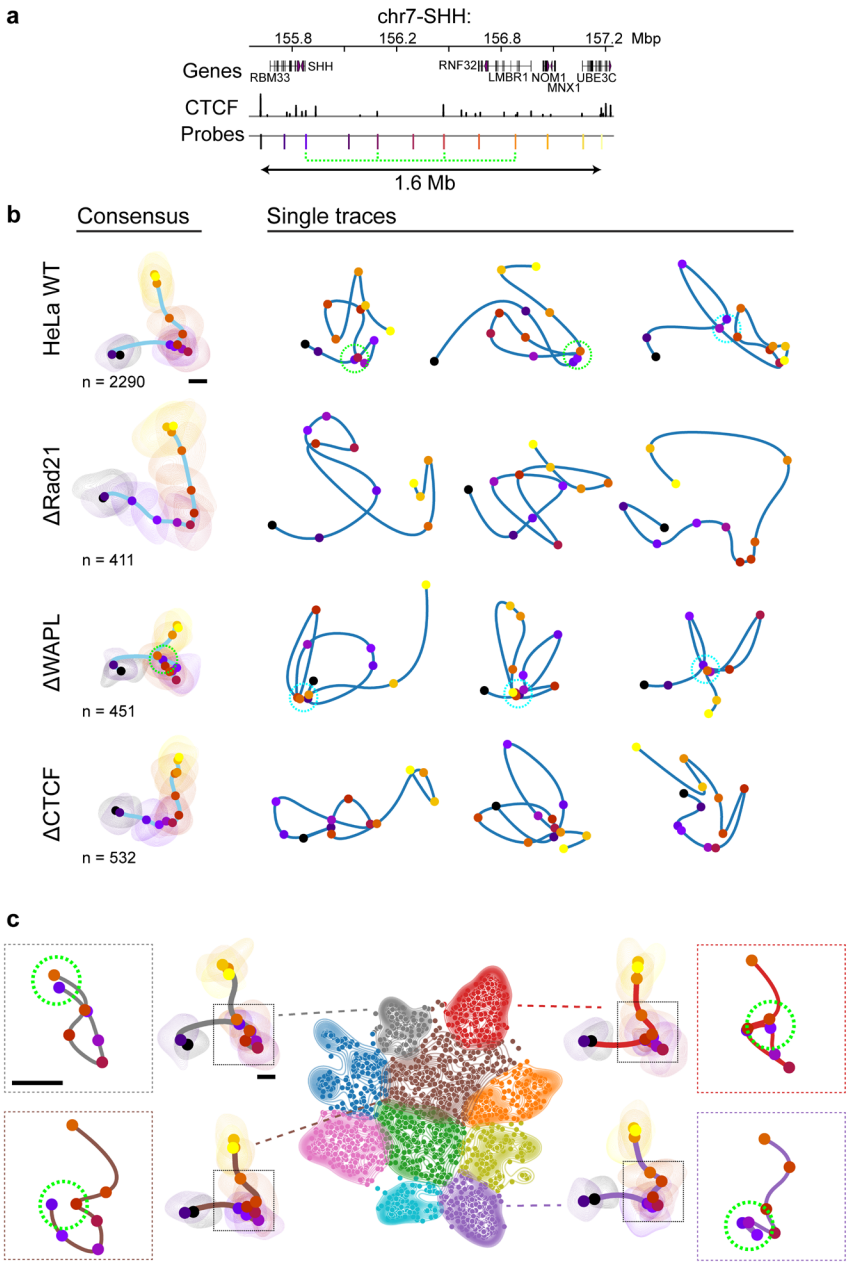

Fig. 3: TAD-scale chromatin architecture in single cells.
a, Overview of a genomic region (Chr7-SHH) targeted with 12 probes across 1.6 Mb. Each probe targets a ~10 kb-sized region. Major CTCF sites predicted to form loops are indicated by dashed green line. **b**, Consensus traces and single trace examples from HeLa WT or HeLa Rad21-EGFP-AID, CTCF-EGFP-AID or WAPL-Halo-AID treated for 2 h with auxin. Data from 3 (WT) or 2 (AID cells) replicates, number of traces per condition are indicated. Green dashed circles indicate <150 nm proximity between probe no. 3 at the TAD boundary and downstream CTCF sites, while cyan circles indicate multiple stacked contacts. **c,** UMAP with clustering of single traces by a Bayesian GMM (see methods) from the chr7-SHH locus in wild type HeLa cells. Consensus traces of exemplary clusters and insets highlighting the internal TAD architecture (3rd – 9th probe positions, dominating loops highlighted with green dashed

ring) as indicated by trace/cluster colour. n=2290 from 3 experiments, n=119-365 traces in each cluster. Scale bars are 100 nm.