


FISHnet: detecting chromatin domains in single-cell sequential Oligopaints imaging data

Received: 18 June 2024

Accepted: 26 March 2025

Published online: 12 May 2025

Rohan Patel^{1,2,3}, Kenneth Pham^{1,2,3}, Harshini Chandrashekar^{1,2,3} & Jennifer E. Phillips-Cremins^{1,2,3} 

Sequential Oligopaints DNA FISH is an imaging technique that measures higher-order genome folding at single-allele resolution via multiplexed, probe-based tracing. Currently there is a paucity of algorithms to identify 3D genome features in sequential Oligopaints data. Here, we present FISHnet, a graph theory method based on optimization of network modularity to detect chromatin domains in pairwise distance matrices. FISHnet sensitively and specifically identifies domains and boundaries in both simulated and real single-allele imaging data and provides statistical tests for the identification of cell-type-specific domains-like folding patterns. Application of FISHnet across multiple published Oligopaints datasets confirms that nested domains consistent with TADs and subTADs are not an emergent property of ensemble Hi-C data but also observable on single alleles. We make FISHnet code freely available to the scientific community, thus enabling future studies aiming to elucidate the role of single-allele folding variation on genome function.

A decade of technology development leading to chromosome-conformation-capture sequencing assays^{1–4} has revealed that the mammalian genome is folded into A and B compartments, topologically associated domains (TADs), subTADs, and loops^{5–8}. Using bulk sequencing assays and perturbative experiments, leading models assert that TADs, subTADs, and their boundaries regulate gene expression by restricting enhancers' physical search space for their distal target genes and preventing ectopic enhancer–promoter interactions (reviewed in ref. 9). Boundary disruption has been linked to gene expression dysregulation in several models of human disease^{10–12}. Moreover, boundaries created by cohesin-mediated loop extrusion stalling at high-density arrays of CTCF binding sites have been functionally linked to the placement of replication initiation zones in early S-phase¹³. A major question that has emerged from bulk Hi-C-based studies is whether domain-like folding patterns can be detected in single cells. To further elucidate the structure–function relationship of the genome, there is a need for

technologies that can quantitatively detect TAD and subTAD folding patterns at single-cell resolution.

Multiplexed sequential DNA FISH Oligopaint imaging technologies have recently enabled single-allele imaging of genome folding at kilobase resolution across megabase (Mb)-sized sections of the genome^{14–19}. Sequential Oligopaints imaging experiments use tiled probes and sequential imaging steps to generate spatial localization of individual loci that are transformed into dense matrices of pairwise distances between genomic loci (Fig. 1a)¹⁵. Key challenges have slowed advances in computational methods to detect domain-like patterns in single-allele imaging data. First, the signal-to-noise ratio is low within pairwise distance matrices, which makes sensitive domain detection while minimizing false positives particularly challenging.

Second, sequential Oligopaints experiments involve tiling of probes in a stepwise bin-by-bin manner and are prone to a high number of dropouts owing to technical biases, such as low probe accessibility

¹Department of Bioengineering, University of Pennsylvania, Philadelphia, PA, USA. ²Epigenetics Institute, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. ³Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA.

✉ e-mail: jcremins@seas.upenn.edu

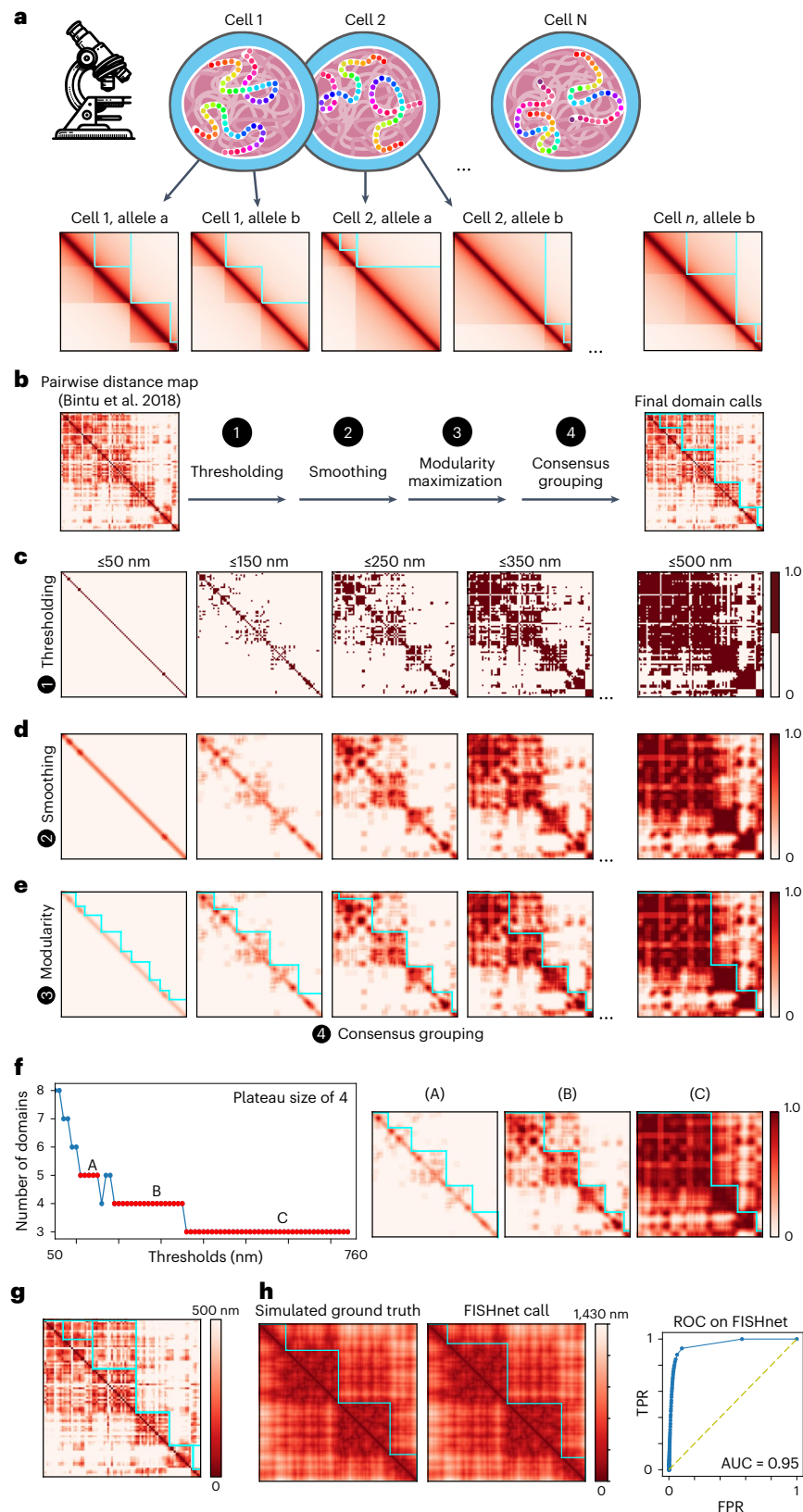


Fig. 1 | Detection of domain-like genome folding patterns in single-cell sequential Oligopaints imaging data with FISHnet. **a**, Cartoon schematic of sequential Oligopaints data and pairwise distance matrices, with FISHnet chromatin domain calls overlaid. **b**, FISHnet methodological steps. The pairwise distance matrix of a 30-kb resolution trace spanning chr21:34.6 Mb–37.1 Mb on a single-allele from HCT116 cells¹⁵. **c**, Binary matrices resulting from a sweep of distance thresholds on the input pairwise distance matrix. Color bar indicates a binary value of 1 or 0. **d**, Smoothing of binary matrices. **e**, Domain calls per each smoothed, binarized matrix by optimizing network modularity (see Methods).

Color bar indicates arbitrary value between 0–1. **f**, Number of domains as a function of distance threshold. Red points indicate domain calls that plateau for at least four consecutive thresholds resulting in three plateau groups: A, B, and C. Consensus domain calls within plateau groups A, B, or C are shown overlaid on averaged smoothed, binarized matrices from each distance-thresholded group. **g**, Final non-redundant domain calls overlaid on the pairwise distance matrix. **h**, ROC curve measuring FISHnet's performance on simulated pairwise distance matrices. The yellow dotted line represents $y = x$.

or low coverage due to repetitive regions of the genome. The dropouts are often interpreted by algorithms as boundaries, increasing false-positive domain calls. By contrast, bulk Hi-C data allow for more leniency regarding missed proximity ligation events because a missed ligation junction in one cell can be captured in another cell. Additionally, bias due to poor coverage can be computationally corrected in Hi-C with Knight–Ruiz matrix balancing^{8,20}. However, explicit modeling of bin-to-bin intrinsic differences influencing probe density and binding efficiency in single-allele imaging data has not yet been achieved. Overall, the challenges inherent to sequential Oligopaints data underscore the need for specialized analytical approaches tailored to the unique characteristics of single-allele resolution datasets distinct from those of bulk Hi-C technologies.

Results

Detection of domain-like structures in sequential Oligopaints data with FISHnet

To detect domain-like structures in sequential Oligopaints data, we developed FISHnet as an open-source graph-theory-based method with four steps: binarization with thresholding, smoothing, network modularity maximization, and consensus grouping (Fig. 1b). FISHnet begins by thresholding a single two-dimensional array of pairwise distance measurements between genomic loci from one allele into multiple binarized matrices (Fig. 1c). Thresholding converts entries in a pairwise distance matrix into binary values by marking those with distances less than or equal to a specified value. FISHnet uses multiple distance thresholds to sensitively detect a wide range of domain structure sizes in these matrices. We determined the range of distance thresholds by computing the minimum and maximum pairwise distances in the matrix, using a gradient of 10-nm steps. We discovered that small distance thresholds, such as <150 nm, yield binarized arrays with small domain-like patterns and large distance thresholds, such as ≥500 nm, yield larger domain-like patterns (Extended Data Fig. 1). Thus, by using a range of thresholds, FISHnet effectively captures a spectrum of domain-like patterns across different distance scales.

We next smoothed the binary matrices by applying a 2 × 2 signal averaging window to the binarized matrices (Fig. 1d). Smoothing converts the binary matrices into matrices with continuous values from 0 to 1 and improves FISHnet's ability to call domains (Extended Data Fig. 2a–c). We apply a small smoothing window because larger windows result in poor performance, likely owing to boundary shifting (Extended Data Fig. 2a–c). Hereafter, we refer to the binarized and smoothed arrays as adjacency matrices. We hypothesized that chromatin domains can be identified in adjacency matrices through a community detection method based on the maximization of network modularity (Q):

$$Q = \frac{1}{m} \sum_{i,j} \left[A_{i,j} - \frac{k_i k_j}{m} \right] \delta(g_i, g_j)$$

where A is the adjacency matrix, $A_{i,j}$ is the edge weight representing the interaction between nodes i and j , k_i is the sum of all of the edge weights for node i , and m is the sum of all non-diagonal edge weights within the network A . Nodes i and j are assigned to communities g_i and g_j , respectively. The Kronecker delta, $\delta(g_i, g_j)$, ensures that only nodes that are in the same group when maximizing modularity are considered. It is 1 if $g_i = g_j$ and 0 otherwise.

To evaluate whether two nodes are interacting, independent, or not interacting, we computed an observed minus expected relationship between nodes using the modularity equation. The observed value is computed as $A_{i,j}$, and the expected value is computed as $\frac{k_i k_j}{m}$. For values of modularity, a positive non-zero difference between the observed and expected values signifies two interacting nodes, indicating that their members are physically close to each other but distant from members of other domains. A zero difference represents nodes that

are independent of each other and thus exhibit similar intra- and inter-domain distances. A negative non-zero difference represents two nodes that do not interact. Thus, a high modularity metric is conceptually intuitive as a metric for quantifying chromatin domains because they consist of genomic loci that fold in physical proximity to other loci within versus outside the domain-like structures⁹.

We applied a Louvain-like algorithm²¹ (see Methods) to the adjacency matrices to maximize modularity (Fig. 1e). To address stochasticity in the maximization process, the algorithm was executed 20 times for each adjacency matrix. For each nanometer distance threshold, we computed a consensus set of domain calls from 20 runs of modularity maximization using the adjusted RAND score metric^{22,23} (see Methods). We then plotted the number of domains as a function of distance threshold. When the plot plateaus at four or more adjacent thresholds, we compute a consensus partition representing the plateau and then merge all consensus partitions into final domain calls (Fig. 1f,g). We provide a plateau parameter to enforce consistency across adjacent thresholds for robust and stable domain calls that minimize false positives by overcoming the low signal-to-noise inherent in Oligopaints data. Further discussion on how the choice of smoothing window, the distance step between thresholds, and plateau sizes affect FISHnet's performance are provided in the Methods and Extended Data Figure 2.

FISHnet sensitively and specifically identifies both domains and boundaries in simulated and real single-allele Oligopaints data

We assessed FISHnet's accuracy in calling domain-like structures by implementing FISHnet on simulated ground-truth data for single-allele microscopy using the established strings and binders (SBS) model (Methods and Extended Data Fig. 3a)²⁴. We applied FISHnet to more than 600 simulations of Oligopaints data and computed an area under the curve (AUC) of 0.95 from the receiver operating characteristic (ROC) curve (Fig. 1h). Most sequential Oligopaints datasets include pairwise distance matrices in which many genomic loci have dropouts owing to probe inefficiencies. We simulated these matrices with bin dropout rates ranging from 5% to 100% and observed that FISHnet domain calls are robust, with an AUC of 0.91 with datasets with up to 40% dropouts (Extended Data Fig. 3b,d). FISHnet's performance improved to an AUC of 0.88 with up to 80% dropouts when we used linear imputation as previously published²⁵ (Extended Data Fig. 3c, d). Thus, FISHnet sensitively and specifically calls domain-like structures in simulated Oligopaints data and is robust to technical dropout of genomic loci owing to probe inefficiencies.

We also benchmarked FISHnet boundary calls against a published insulation score method, which uses a sliding-window approach to sum signals and identify domain boundaries in pairwise distance matrices¹⁵ (Extended Data Fig. 4). For boundary feature detection, FISHnet outperforms the insulation score method in all scenarios: no dropout, dropout, and dropout with interpolation (Extended Data Fig. 4a). In addition, FISHnet facilitates the detection of both nested TAD or subTAD-like structures and domain boundaries, a capability that the insulation score method lacks (Extended Data Fig. 4b). Our data demonstrate that FISHnet is sensitive and specific for calling both domains and boundaries in single-allele sequential Oligopaints data, even amid substantial missing data, and support the use of linear interpolation as a solution for addressing dropouts in sequential Oligopaints data.

We next assessed the role of dropouts in the stability of FISHnet domain calls on real Oligopaints data, given the wide range of dropouts in these datasets (Extended Data Fig. 5a). We analyzed $n = 595$ 0%-dropout pairwise distance matrices from human HCT116 cells at 30-kb resolution¹⁵ (Table 1) and artificially introduced dropouts ranging from 5% to 80% (Extended Data Fig. 5b). We then compared FISHnet domains from 0% dropout matrices with those with increasing dropout percentages, both with and without linear interpolation (Extended Data Fig. 5b,c). We quantified the true positive rate (TPR) and false positive

Table 1 | Table of publicly available datasets used within this study

| Dataset | Description | Link | Identifier |
|---|--------------------------------|---|--|
| Multiplexed sequential Oligopaints imaging data | HCT116-RAD21-AID auxin induced | https://github.com/BogdanBintu/ChromatinImaging/tree/master/Data | File name: HCT116_chr21-34-37Mb_6h auxin.csv |
| Multiplexed sequential Oligopaints imaging data | HCT116-RAD21-AID control | https://github.com/BogdanBintu/ChromatinImaging/tree/master/Data | File name: HCT116_chr21-34-37Mb_untreated.csv |
| Multiplexed sequential Oligopaints imaging data | mESC | https://zenodo.org/records/3735329 | Folder name: DNaseqFISH+.zip |
| Multiplexed sequential Oligopaints imaging data | Mouse brain tissue | https://zenodo.org/records/4708112 | File name: TableS8_brain_DNaseqFISH_25kb_voxel_coordinates_2762cells.csv |
| Multiplexed sequential Oligopaints imaging data | K562 | https://github.com/BogdanBintu/ChromatinImaging/tree/master/Data | File name: K562_chr21-28-30Mb.csv |
| Multiplexed sequential Oligopaints imaging data | IMR90 | https://github.com/BogdanBintu/ChromatinImaging/tree/master/Data | File name: IMR90_chr21-28-30Mb.csv |
| Multiplexed sequential Oligopaints imaging data | A549 | https://github.com/BogdanBintu/ChromatinImaging/tree/master/Data | File name: A549_chr21-28-30Mb.csv |
| Multiplexed sequential Oligopaints imaging data | BX-C 10-kb resolution | https://data.4dnucleome.org/experiment-set-replicates/4DNESWM4U8RX/ | 4DN accession: 4DNESWM4U8RX |
| Multiplexed sequential Oligopaints imaging data | BX-C 2-kb resolution | https://data.4dnucleome.org/experiment-set-replicates/4DNESEC7D1AH/ | 4DN accession: 4DNESEC7D1AH |
| In situ Hi-C | HCT116-RAD21-AID control | https://data.4dnucleome.org/experiment-set-replicates/4DNES3QAGOZZ/ | 4DN accession: 4DNES3QAGOZZ |
| Hi-C | mESC | https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE96107 | GEO accession: GSE96107 |
| In situ Hi-C | IMR90 | https://data.4dnucleome.org/experiment-set-replicates/4DNES1ZEJNRU/ | 4DN accession: 4DNES1ZEJNRU |
| In situ Hi-C | K562 | https://data.4dnucleome.org/experiment-set-replicates/4DNES17DEJTM/ | 4DN accession: 4DNES17DEJTM |

rate (FPR) as they relate to the dropout fraction. We found that TPR decreases and FPR increases linearly with increased dropout fraction in the absence of interpolation (Extended Data Fig. 5d, left). Linear interpolation limits the FPR to approximately 5.5% at dropouts greater than 50% (Extended Data Fig. 5d, right). Our results demonstrate that linear interpolation can effectively mitigate false positive domain calls in FISHnet at high dropout levels, highlighting its utility in preserving data accuracy under challenging experimental conditions.

We next tested the extent to which FISHnet performance corresponds to boundaries present in ensemble Hi-C data. We ran FISHnet on publicly available sequential Oligopaints data from human HCT116 cells at 30-kb resolution¹⁵ and mouse embryonic stem cells (mESCs) at 25-kb resolution¹⁸ (Table 1). We ran FISHnet with linear imputation on all single alleles with pairwise distance matrices at more than 60% coverage. Representative pairwise distance matrices demonstrate FISHnet’s ability to identify domain-like structures (Fig. 2a). Using bulk Hi-C data^{26,27} and ensembled sequential Oligopaints data, we confirm that the highest frequency FISHnet boundary calls correspond to TAD or subTAD boundaries in bulk Hi-C data (Fig. 2b). We further confirmed that FISHnet accurately identifies domain-like structures that are visually apparent in high-resolution sequential Oligopaints data from fruit fly embryos at 10-kb and 2-kb resolution¹⁷ (Table 1; Extended Data Fig. 6). Our work confirms that domains called by FISHnet in single-allele imaging data correspond to TADs and/or subTADs in ensemble Hi-C data.

Finally, we sought to test FISHnet’s performance in an established perturbative model system with known disruptions to TADs and subTADs in ensemble Hi-C data. We ran FISHnet on sequential Oligopaints data generated in human HCT116 cells engineered with an auxin-inducible degron system tagging RAD21 (ref. 15) (Table 1). Following auxin treatment to remove RAD21, we observed an equal probability of FISHnet boundary calls across the locus¹⁵ (Fig. 2c). The strongest boundary was present in approximately 10% of alleles in

control HCT116 cells, decreasing to 4% of alleles upon RAD21 depletion. This finding is consistent with other reports^{15,28,29} (Fig. 2c). Our results demonstrate that FISHnet quantification of domains can recapitulate known patterns of boundary disruption due to cohesin knock-down in ensemble Hi-C data^{26,30,31}.

FISHnet enables statistical testing of single-cell heterogeneity unique to cell type and cellular state

We next assessed FISHnet’s capacity to detect domain and boundary positions across a population of single cells in primary tissue. Using published imaging data from mouse brain tissue¹⁹ (Table 1), we ran FISHnet on sequential Oligopaints data from $n = 2,065$ excitatory neurons (Fig. 3a) and $n = 130$ microglia (Fig. 3b). We equally filtered maps with 20–80% coverage from both conditions, applied linear imputation, ran FISHnet, and plotted the frequency of boundary calls across single alleles compared with ensembled imaging matrices (Fig. 3a,b). We clustered neurons and microglia into subsets with similar boundary structure (see Methods). We found $n = 49$ and $n = 8$ independent clusters of single alleles with similar folding patterns in excitatory neurons (Fig. 3c and Extended Data Fig. 7) and microglia (Fig. 3d and Extended Data Fig. 8). In each cluster, we observed a strong correlation between the visual location of boundaries and the frequency of FISHnet domain calls in single alleles. Our data reveal that FISHnet has the ability to resolve single-allele differences in boundary locations in the same cell type from sequential Oligopaints imaging data.

We next explored the distribution of single-cell variation in genome folding between two cell types at a given boundary location. We formulated a statistical test to identify boundaries that differed significantly between two cell types (see Methods). To compare a known boundary between excitatory neurons and microglia, we applied a chi-squared test to FISHnet data and computed the probability of obtaining a single-allele boundary frequency across a population of cells that was equal to or more extreme than what would be expected

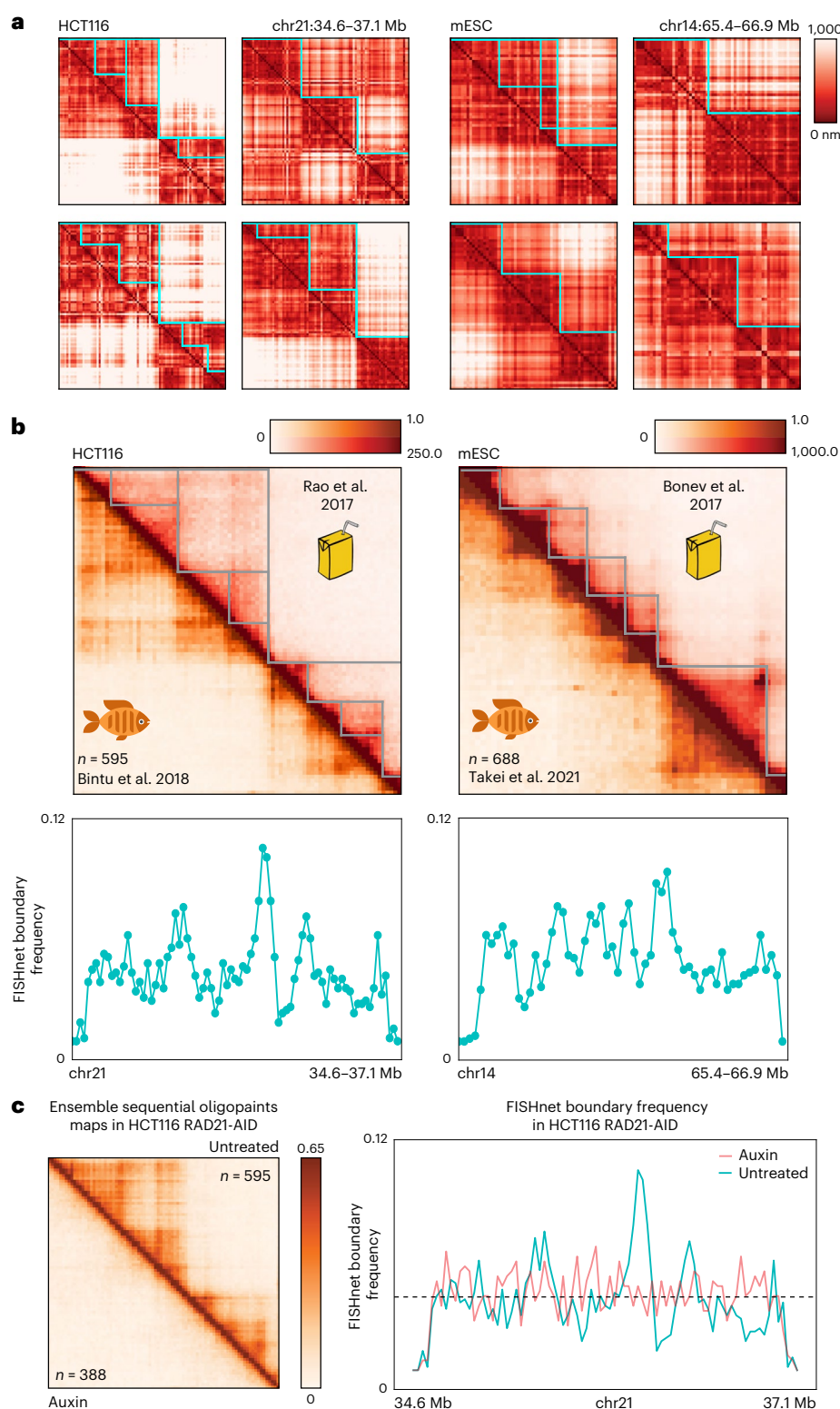


Fig. 2 | FISHnet chromatin domain calls correspond with TADs and subTADs called in ensemble Hi-C. a, Individual pairwise distance matrices with FISHnet calls on published sequential Oligopoints imaging data in mESC alleles and HCT116 alleles. The color bar indicates the pairwise distances between genomic loci. mESC alleles were traced at 25-kb resolution, and HCT116 alleles were traced at 30-kb resolution. **b**, The results of Hi-C with TAD and subTAD calls using 3DNetMod³² in mESCs (25-kb resolution) and HCT116 cells (30-kb resolution) are provided in the top half of the heatmaps. The ensemble frequency matrices of sequential Oligopoints imaging data from the same cell type are provided in the bottom half of the maps ($n = 595$ alleles, HCT116 and $n = 688$ alleles, mESCs). Juice box indicates Hi-C data and cartoon fish indicates ensemble imaging data. The

color bar indicates the normalized interaction frequency (Hi-C) and the number of alleles with distances less than 250 nm, normalized by the total number of alleles (DNA FISH). Blue line plots indicate the frequency of FISHnet boundary calls across $n = 595$ HCT116 alleles and $n = 688$ mESC cell alleles. **c**, Ensemble frequency matrices of sequential Oligopoints imaging data in HCT116 RAD21-AID cells in the untreated ($n = 595$ alleles) and auxin-treated ($n = 388$ alleles) conditions at 30-kb resolution. The color bar indicates the number of alleles with distances less than 250 nm, normalized by the total number of alleles. Adjacent to the heatmap are line plots representing the frequency of FISHnet boundary calls in single cells from untreated and auxin-treated conditions. The black dotted line signifies the average frequency of a boundary in the auxin condition.

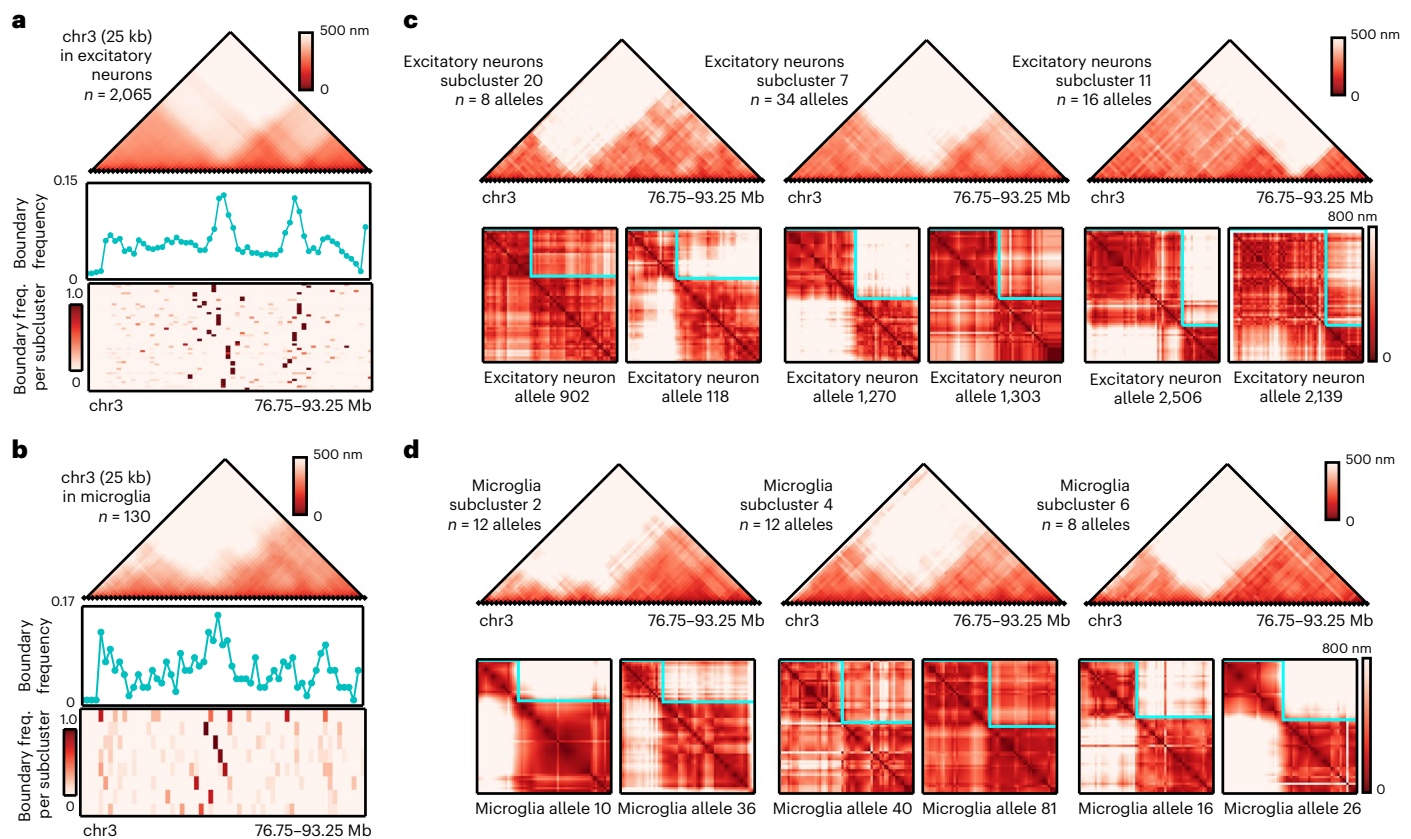


Fig. 3 | FISHnet detects heterogeneous domain calls within single alleles. **a, b**, Ensemble averaged matrices of all sequential Oligopaints imaging data in excitatory neurons ($n = 2,065$ alleles) (**a**) and microglia ($n = 130$ alleles) (**b**). Blue line plots representing the frequency of FISHnet boundary calls in single cells from excitatory neurons and microglia. The heatmap showcases the boundary frequency per each subcluster found within excitatory neurons and microglia, respectively. **c, d**, Subclusters of alleles with different boundary structures. **c**, Three excitatory neuron ensemble averaged maps from subclusters 20, 7, and

11 ($n = 8, 34$, and 16 alleles, respectively). **d**, Three microglia ensemble averaged maps from subclusters 2, 4, and 6 ($n = 12, 12$, and 8 alleles, respectively). Below each ensemble map are single-allele example heatmaps of sequential Oligopaints imaging data representing that particular domain pattern. Domain calls on single-allele pairwise distance matrices are based on the largest nanometer plateau calls. The color bar indicates the nanometer pairwise distances between genomic loci. Oligopaints data for excitatory neurons and microglia were traced at 25-kb resolution.

if the null hypothesis, that the boundary occurs in similar proportions of alleles in both cell type A and cell type B, was true. We demonstrated that the chi-squared test effectively identifies boundaries that are significantly more likely to be present in a higher proportion of single alleles in either microglia or excitatory neurons (Fig. 4a). Our first statistical test suggests that FISHnet can identify genomic locations of statistically significant cell-type-specific boundaries.

To provide users with a rigorous independent method for the identification of cell-type-specific boundaries, we developed an empirical permutation test to identify microglia-specific boundaries after accounting for single-allele variation. We formulated an empirical test statistic representing the proportion of alleles with a specific FISHnet boundary (see Methods). We computed a one-tailed, right-tailed empirical P value by comparing the test statistic from $n = 130$ single microglia alleles to a null distribution representing 10,000 draws of $n = 130$ alleles from excitatory neuron Oligopaints data. Using our permutation test, we confirmed that the boundaries identified as cell-type-specific by the chi-squared test were also present in a significantly higher proportion of alleles in microglia than would be expected by chance from the excitatory neuron null distribution (Fig. 4b). Visually apparent neuron-specific boundaries did not result in low P values as expected from the one-tailed, right-tailed test. We also chose two random boundaries that were not significant in the chi-square test and verified their non-significance with the permutation test (Fig. 4b). Together, our statistical tests demonstrate that FISHnet can reproducibly identify cell-type-specific differences in the distributions of single-allele boundary positions.

FISHnet boundary calls can be used to distinguish cell types

We sought to determine whether FISHnet boundaries can be used to discern cell types. We applied FISHnet to published sequential Oligopaints imaging data from $n = 2,101$, $n = 2,060$, and $n = 1,762$ single alleles from IMR90, K562, and A549 cells, respectively¹⁵ (Table 1). Plotting the ensemble single-allele data revealed that the visual location at which specific boundaries are present in IMR90 cells and absent in K562 and A549 cells correlates with a high FISHnet boundary single-allele frequency only in IMR90 cells (Fig. 4c). Principal component analysis using the pairwise distance matrices could not distinguish cell types (Fig. 4d). However, principal component analysis using the single-allele FISHnet boundary calls alone could readily separate IMR90 from K562 and A549 cells (Fig. 4e). Altogether, our data underscore FISHnet's ability to identify cell-type-specific boundaries across a range of cell lines and its ability to distinguish cell-type-relevant genome folding features more effectively than the raw data.

FISHnet provides capability for the detection of hierarchical domain-like structures consistent with nested subTADs within TADs originally detected in bulk Hi-C data

Early studies in ensemble Hi-C data uncovered nested subTADs within TADs^{7,32}. Two leading open questions are whether nested domain-like structures might exist on single alleles and if TADs and subTADs are observable on single alleles or are only an emergent property of bulk Hi-C data. We transformed the single-allele FISHnet domain calls from HCT116, K562, and IMR90 cell lines into an integer mask that

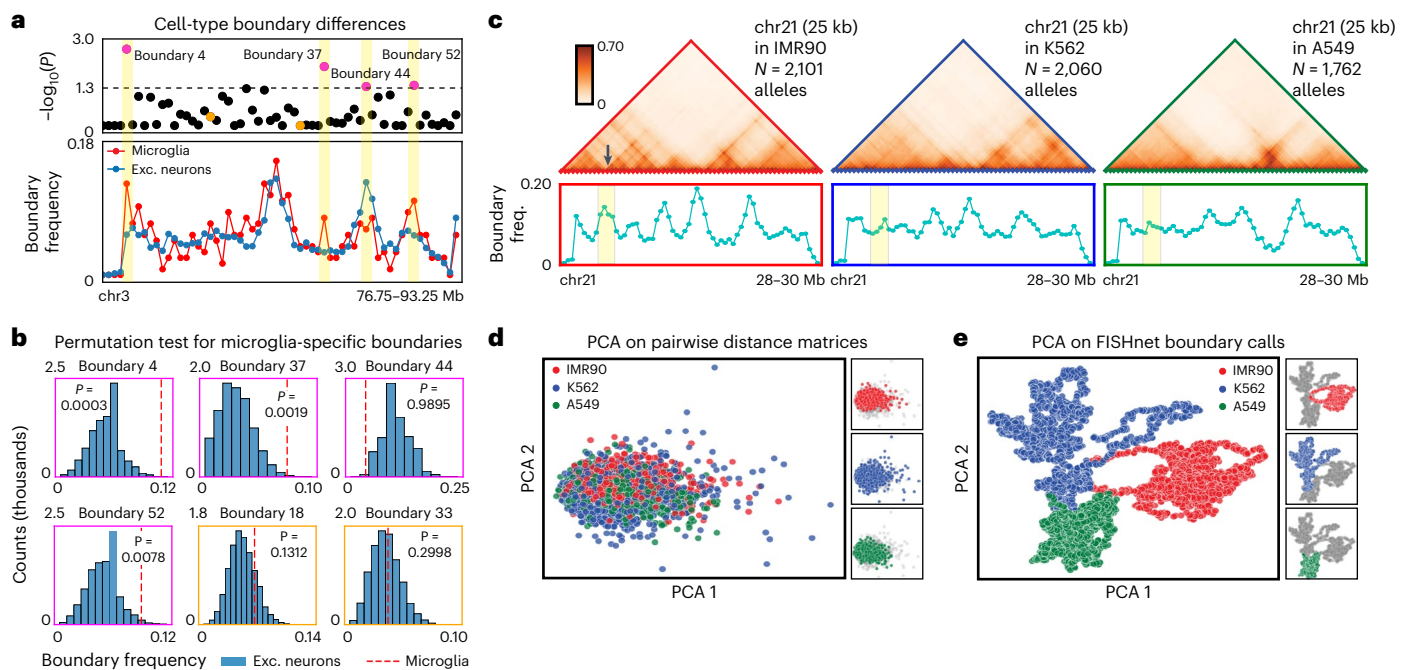


Fig. 4 | FISHnet identifies cell-type-specific boundaries. **a**, Top, P values for differential boundaries for every bin within excitatory neurons and microglia, determined using the chi-squared test. Bottom, line plots representing the frequency of FISHnet boundary calls for excitatory neurons ($n = 2,065$ alleles) and microglia ($n = 130$ alleles). **b**, For significant boundaries determined by the chi-square test in **a**, a permutation test through bootstrapping is shown. Two negative permutation tests are also shown in orange. **c**, Ensemble frequency

matrices of sequential Oligopaints imaging data in IMR90 ($n = 2,101$), K562 ($n = 2,060$), and A549 ($n = 1,762$) alleles. The color bar indicates the number of alleles with distances less than 250 nm, normalized by the total number of alleles. Blue line plots represent the frequency of FISHnet boundary calls for each cell type. The yellow vertical highlights show the IMR90 cell-type-specific boundary. **d**, **e**, Principal component analysis on the pairwise distance matrices from three human cell lines (**d**) and their FISHnet boundary calls (**e**).

represents domain structures (see Methods). To create an ensemble of FISHnet domain calls, we summed all masked pairwise distance matrices (Fig. 5a). Each entry (i, j) in the ensemble FISHnet domain mask matrix represents the number of times that the two genomic loci (i and j) were present within the same FISHnet-called domain (Fig. 5a,b). We observed a strong correlation between the ensemble FISHnet domain mask counts and ensemble Hi-C data in the same locus and cell type (Pearson's correlation coefficient 0.90, 0.88, and 0.91 for HCT116, IMR90, and K562, respectively) (Fig. 5b and Extended Data Fig. 9). These data suggest that the frequency of FISHnet domain calls in single-allele imaging data can be used to recapitulate TADs and subTADs in ensemble Hi-C data.

We tested how the nanometer distance between genomic fragments in single-allele imaging data might be linked to the size of TADs or subTADs in ensemble Hi-C data and ensemble masked imaging data. Using single-allele sequential Oligopaints data from HCT116, we utilized FISHnet to detect chromatin domains, with thresholds of <150 nm, 150–500 nm, and >500 nm. Upon application of the FISHnet domain mask (Fig. 5a,b), we observed a strong enrichment for classic small subTAD-like structures in single alleles when using thresholds of <150 nm (Fig. 5c). Moreover, for thresholds of 150–500 nm or >500 nm pairwise distances, we observed a strong bias toward detection of medium to large TAD-like structures in single alleles (Fig. 5c). We quantitatively confirmed that increasing distance scales results in larger and un-nested FISHnet domain calls (Extended Data Fig. 10). We note that FISHnet detects both TAD and nested subTAD-like domains on the same single-allele when we use matrices thresholded at different nanometer pairwise distances (Fig. 5d), which is consistent with the established mechanisms of cohesin-mediated loop extrusion and compartmentalization^{9,26,33}. Our data confirm previous reports that nested TAD- and subTAD-like structures are present in single alleles in sequential Oligopaints imaging data^{15,34}, therefore suggesting that

domain-like folding patterns are not an emergent property of ensemble Hi-C data but are also observable on single alleles.

Discussion

Here, we present FISHnet, a graph-theory-based algorithm designed to detect chromatin domains within pairwise distance matrices from sequential Oligopaints data. We demonstrate that FISHnet sensitively and specifically detects domains and their boundaries in real and simulated single-allele imaging data. We applied FISHnet on multiple published Oligopaints datasets, we discovered nested domains consistent with TADs and subTADs on single alleles, thus suggesting that nested chromatin domains exist on single alleles in a moment in time and are not just an emergent property of ensemble Hi-C data. We demonstrate FISHnet's implementation on nine experimental datasets ranging from 2-kb to 30-kb resolution and covering loci ranging from 130 kb to 2.5 Mb, across three model systems, highlighting the robustness of the algorithm (Table 1). FISHnet quantifies the frequency of boundary placement across a population of single cells and can evaluate statistically significant cell-type-specific domains and boundaries upon comparison of data sets from two independent conditions. We have made the FISHnet code freely available to the scientific community, thus enabling future studies aiming to elucidate the role for single-allele folding variation on genome function (<https://github.com/Rohanpate-IUpenn/FISHnet>).

FISHnet uses multiple strategies to overcome technical challenges that represent a barrier. It uses thresholds across a range of nanometer distances between probes to call domains at different sizes. It implements linear interpolation to overcome the high frequency of dropout events that plagues sequential Oligopaints data. It also uses a plateau parameter to ensure high-confidence, reproducible domain calls across a range of nanometer distances, thus preventing high false positive rates. To our knowledge, there is a paucity of algorithms published

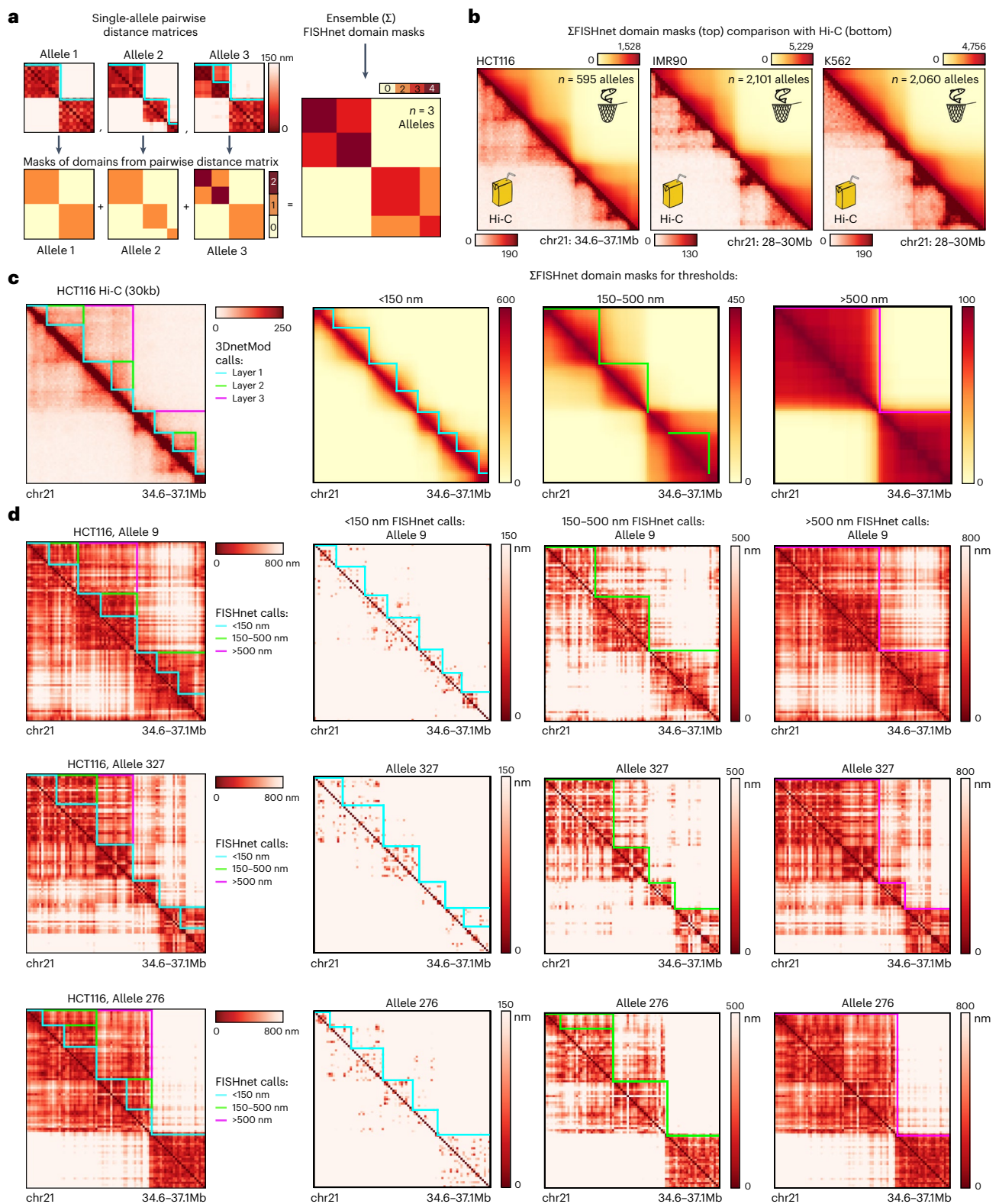


Fig. 5 | FISHnet uncovers nested structures consistent with TADs and subTADs on single alleles in sequential Oligopaints data. **a**, Schematic of the method for implementing an ensemble FISHnet domain mask. **b**, Ensemble FISHnet domain mask matrices (top) compared with ensemble Hi-C data (bottom) for HCT116, IMR90, and K562 alleles. Juice box indicates Hi-C data while the fish-in-net cartoon indicates ensemble FISHnet domain mask matrices. The color bar indicates normalized interaction frequency (Hi-C) and ensemble FISHnet domain mask counts (DNA FISH). **c**, Hi-C data for HCT116 cells with 3dNetMod subTAD and TAD calls showcasing different hierarchical structures from nested

layer 1 to encompassing layer 3 (left). Ensemble FISHnet domain mask matrices at different distance scales, with layer 1, layer 2, and layer 3 3dNetMod calls overlaid. Layer 1 (cyan), layer 2 (green), and layer 3 (magenta) are shown. **d**, Single-allele examples of hierarchical chromatin folding domains dependent on different distance scales. FISHnet calls at different threshold ranges (<150 nm (cyan), 150–500 nm (green), and >500 nm (magenta)) are overlaid on the pairwise distance matrices (left). The same pairwise distance matrices were replotted at three different distances cutoffs—150 nm, 500 nm, and 800 nm—with the FISHnet domain calls from the corresponding threshold ranges (right).

for calling domains in single-allele imaging data; therefore, FISHnet provides a unique and impactful functionality. The insulation score method built for calling boundaries in ensemble Hi-C data has recently been applied to Oligopaints data¹⁵. We demonstrate here that FISHnet provides more sensitive and specific boundary identification than does the insulation score method. Together, our data demonstrate FISHnet's utility for calling nested domains and boundaries across genomic length scales in single-allele sequential Oligopaints data.

We plan to improve and extend FISHnet in future freely available versions. Currently, FISHnet works on symmetric pairwise distance matrices ranging from small to medium sizes (<150 bins in matrix size). We used published data sets with high signal-to-noise to build the FISHnet algorithm. Applying FISHnet on larger matrices or noisier data will increase computational time, raising cost and data storage complexity. Currently, 20 repetitions are applied to optimize the modularity solution, and a consensus partition is used to lower the risk of overfitting. In the future, we can adjust the number of partitions to address differences in quality and size of new datasets. To improve domain calling accuracy in data with higher noise, we can test the utility of preprocessing steps such as low-pass or Gaussian filters or Fourier transforms. We use imputation in the first version of FISHnet to smooth bin-to-bin dropouts and technical variation in signal. It will be important in the future to develop new computational methods to model and correct for technical biases, such as low probe accessibility or low bin coverage. FISHnet was written in a modular manner with usage instructions to clarify parameters that can be adjusted by the user for their specific data type and quality.

Application of FISHnet revealed significant variation in boundary positioning in a population of single cells representing one cell type, as well as between populations of excitatory neurons and microglia. One possible model that can support the existence of allele-to-allele domain heterogeneity is the loop extrusion model, where domains are created by active ATP-dependent cohesin-mediated extrusion of chromatin loops and thus placed randomly at a given point in time by asynchronous cohesin movement along the genome in a population of cells^{26,30,35–37}. Using published Oligopaints data, we confirmed previous reports that domains remain in HCT116 cells even after cohesin degradation¹⁵. Similar results have been observed in budding yeast, in which a mutant lacking loop extrusion might also have chromatin domains³⁸. Alternative mechanisms that might play a role in domain formation could include phase separation, condensation, compartmentalization, and/or unknown molecular motors that extrude the DNA. Together, these data highlight the power of FISHnet when used in combination with genetic perturbations to elucidate the mechanisms that regulate, maintain, and create domains at single-allele resolution.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41592-025-02688-1>.

References

- Lieberman-Aiden, E. et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009).
- Fullwood, M. J. et al. An oestrogen-receptor- α -bound human chromatin interactome. *Nature* **462**, 58–64 (2009).
- Beagrie, R. A. et al. Complex multi-enhancer contacts captured by genome architecture mapping. *Nature* **543**, 519–524 (2017).
- Quinodoz, S. A. et al. Higher-order inter-chromosomal hubs shape 3D genome organization in the nucleus. *Cell* **174**, 744–757 (2018).
- Dixon, J. R. et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).
- Nora, E. P. et al. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381–385 (2012).
- Phillips-Cremins, J. E. et al. Architectural protein subclasses shape 3D organization of genomes during lineage commitment. *Cell* **153**, 1281–1295 (2013).
- Rao, S. S. et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).
- Beagan, J. A. & Phillips-Cremins, J. E. On the existence and functionality of topologically associating domains. *Nat. Genet.* **52**, 8–16 (2020).
- Lupianez, D. G. et al. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* **161**, 1012–1025 (2015).
- Sun, J. H. et al. Disease-associated short tandem repeats co-localize with chromatin domain boundaries. *Cell* **175**, 224–238 (2018).
- Franke, M. et al. Formation of new chromatin domains determines pathogenicity of genomic duplications. *Nature* **538**, 265–269 (2016).
- Emerson, D. J. et al. Cohesin-mediated loop anchors confine the locations of human replication origins. *Nature* **606**, 812–819 (2022).
- Wang, S. et al. Spatial organization of chromatin domains and compartments in single chromosomes. *Science* **353**, 598–602 (2016).
- Bintu, B. et al. Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. *Science* **362**, eaau1783 (2018).
- Nir, G. et al. Walking along chromosomes with super-resolution imaging, contact maps, and integrative modeling. *PLoS Genet.* **14**, e1007872 (2018).
- Mateo, L. J. et al. Visualizing DNA folding and RNA in embryos at single-cell resolution. *Nature* **568**, 49–54 (2019).
- Takei, Y. et al. Integrated spatial genomics reveals global architecture of single nuclei. *Nature* **590**, 344–350 (2021).
- Takei, Y. et al. Single-cell nuclear architecture across cell types in the mouse brain. *Science* **374**, 586–594 (2021).
- Knight, P. A. & Ruiz, D. A fast algorithm for matrix balancing. *IMA J. Numer. Anal.* **33**, 1029–1047 (2012).
- Blondel, V. D., Guillaume, J.-L., Lambiotte, R. & Lefebvre, E. Fast unfolding of communities in large networks. *J. Stat. Mech.* P10008 (2008).
- Hubert, L. & Arabie, P. Comparing partitions. *J. Classifi.* **2**, 193–218 (1985).
- Red, V., Kelsic, E. D., Mucha, P. J. & Porter, M. A. Comparing community structure to characteristics in online collegiate social networks. *SIAM Rev.* **53**, 526–543 (2011).
- Conte, M. et al. Loop-extrusion and polymer phase-separation can co-exist at the single-molecule level to shape chromatin folding. *Nat. Commun.* **13**, 4070 (2022).
- Rajpurkar, A. R., Mateo, L. J., Murphy, S. E. & Boettiger, A. N. Deep learning connects DNA traces to transcription to reveal predictive features beyond enhancer–promoter contact. *Nat. Commun.* **12**, 3423 (2021).
- Rao, S. S. P. et al. Cohesin loss eliminates all loop domains. *Cell* **171**, 305–320 (2017).
- Bonev, B. et al. Multiscale 3D genome rewiring during mouse neural development. *Cell* **171**, 557–572 (2017).
- Gabriele, M. et al. Dynamics of CTCF- and cohesin-mediated chromatin looping revealed by live-cell imaging. *Science* **376**, 496–501 (2022).
- Mach, P. et al. Cohesin and CTCF control the dynamics of chromosome folding. *Nat. Genet.* **54**, 1907–1918 (2022).

30. Schwarzer, W. et al. Two independent modes of chromatin organization revealed by cohesin removal. *Nature* **551**, 51–56 (2017).
31. Wutz, G. et al. Topologically associating domains and chromatin loops depend on cohesin and are regulated by CTCF, WAPL, and PDS5 proteins. *EMBO J.* **36**, 3573–3599 (2017).
32. Norton, H. K. et al. Detecting hierarchical genome folding with network modularity. *Nat. Methods* **15**, 119–122 (2018).
33. Nora, E. P. et al. Targeted degradation of CTCF decouples local insulation of chromosome domains from genomic compartmentalization. *Cell* **169**, 930–944 (2017).
34. Szabo, Q. et al. Regulation of single-cell genome organization into TADs and chromatin nanodomains. *Nat. Genet.* **52**, 1151–1157 (2020).
35. Fudenberg, G. et al. Formation of chromosomal domains by loop extrusion. *Cell Rep.* **15**, 2038–2049 (2016).
36. Vian, L. et al. The energetics and physiological impact of cohesin extrusion. *Cell* **173**, 1165–1178 (2018).
37. Davidson, I. F. et al. Rapid movement and transcriptional re-localization of human cohesin on DNA. *EMBO J.* **35**, 2671–2685 (2016).
38. Guerin, T. M., Barrington, C., Pobegalov, G., Molodtsov, M. I. & Uhlmann, F. An extrinsic motor directs chromatin loop formation by cohesin. *EMBO J.* **43**, 4173–4196 (2024).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, 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 licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence 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 licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025

Methods

Hi-C binning, bias correction, and 3DNetMod

We downloaded Hi-C matrices for HCT116 (ref. 26), mESC²⁷, K562, and IMR90 cell lines from the 4DN portal and the Gene Expression Omnibus (Table 1). We binned the Hi-C matrix for HCT116, K562, and IMR90 cells at 30-kb resolution and for mESCs at 25-kb resolution. All Hi-C matrices underwent Knight–Ruiz balancing for bias correction. We called TADs and subTADs using 3DNetMod, as we have previously reported with minor modifications^{11,32,39–42}.

Community detection through modularity maximization (Louvain-like algorithm)

We have previously described the details of the methods in another manuscript³², and they are again described in detail here to ensure reproducibility. Mathematical terms and equations should be expected to be similar due to reproducibility of the methodological steps. To partition binarized and smoothed pairwise distance matrices (also referred to as adjacency matrices) into communities, we utilized a Louvain-like, locally greedy algorithm to maximize modularity. We first calculated the modularity matrix (M):

$$M_{xy} = \left[A_{xy} - \frac{k_x k_y}{m} \right] / m$$

where M is the modularity matrix with size $C \times C$, and C is the number of communities. M_{xy} is the normalized interaction between communities x and y , k_x is the sum of all edge weights for community x and m is the sum of all interactions in the adjacency matrix (A), excluding the diagonal. In this equation, we see that A_{xy}/m is the normalized edge strength connecting communities x and y and $\frac{k_x k_y}{m^2}$ is the expected normalized edge strength at communities x and y . We computed modularity (Q) on M :

$$Q = \sum_{x,y} [M_{x,y}] \delta_{x,y}$$

where $\delta_{x,y}$ ensures that only edge weights within communities are added to the summation by being 1 if x and y are assigned to the same community ($x = y$) and 0 otherwise ($x \neq y$).

The Louvain-like algorithm works by using an iterative approach to rapidly converge on a local modularity maximum without comprehensively examining the entire search space. For each iteration (itr), individual nodes are given the opportunity to move into a new community placement that yields a locally maximal gain in ΔQ according to the following equation:

$$\Delta Q = Q_{\text{itr}} - Q_{\text{itr}-1} = \left(\sum_{x,y} [M_{x,y}] \delta_{x,y} \right)_{\text{itr}} - \left(\sum_{x,y} [M_{x,y}] \delta_{x,y} \right)_{\text{itr}-1}$$

For the first iteration of the Louvain-like algorithm, every individual node is its own community. In this $\text{itr} = 0$ circumstance, the indices for communities x and y correspond to the indices for nodes i and j , and M is the same size as the adjacency matrix A . Before the end of the itr, each node has the chance to merge to form a new community. If $\text{itr} = 1$, modularity is recalculated and if ΔQ is greater than or equal to 1×10^{-10} , the algorithm advances to the next iteration and adjusts the dimensions of M to $C \times C$ (where C represents the number of communities computed in the previous iteration, and $(M_{x,y})_{\text{itr}}$ is the sum of all previous $(M_{x,y})_{\text{itr}-1}$ constituent edges that are merged into $(M_{x,y})_{\text{itr}}$). The algorithm terminates when no further single community merge leads to an improvement in modularity; that is, ΔQ is less than 1×10^{-10} . The Louvain algorithm is inherently non-deterministic owing to (1) the randomness in the order in which nodes merge during the modularity optimization phase and (2) ties in decision making. The algorithm looks at each node in a certain order, but the order is random. If the algorithm picks nodes in a different order each time, it might end up with different communities. Moreover, a node might have multiple choices

that would lead to the same modularity gain. When this happens, the algorithm might randomly pick one of these options, which can change the outcome slightly. To counteract the potential convergence on local maxima inherent in modularity-maximization algorithms, the Louvain-like algorithm underwent 20 iterations. To ensure a diverse exploration of the landscape and avoid bias towards specific solutions, we randomly selected a node with a distinct seed value for each of the 20 partitions. Finally, a consensus partition is selected by taking the partition that is most similar to the 19 other partitions through the use of adjusted RAND score (see below).

Consensus partition through the adjusted RAND score

We have previously described the details of the methods in our other manuscript³², and they are again described in detail here to ensure reproducibility. Mathematical terms and equations are similar owing to reproducibility of the methodological steps. FISHnet uses the adjusted RAND score to find a consensus partition when running the Louvain-like algorithm 20 times to prevent fixation on a local maximum as well as when finding a consensus partition within a plateau sweep. The consensus partition is a partition that is most similar to the other partitions within its group. We find the consensus partition using a similarity metric called the adjusted RAND (aRAND) index, derived from the sklearn toolbox `sklearn.metrics.adjusted_rand_score`. The aRAND index yields unity for perfectly matched partitions and tends towards 0 or slightly negative values for partitions that are no more similar than expected by random chance. We computed aRAND for all pairs of partitions within the set. We chose the partition with the highest average aRAND as the consensus partition. The consensus partition represents the most consistent set of domain calls out of the 20 FISHnet runs. We use a consensus partition that is most similar to all other partitions rather than simply using the partition with the highest modularity to attenuate bias to a particular randomization event and lower the risk of overfitting.

Creation of Ensemble counts matrix

To compute an ensemble counts matrix we started by binarizing and thresholding all of the pairwise distance matrices within a dataset by the value of less than or equal to 250 nm. We then summed all of the binarized maps which yields the ensemble counts matrix.

Creation of Ensemble frequency matrix

To compute an ensemble frequency matrix, as seen in Figs. 2b,c and 4c, we started by computing the ensemble counts matrix. Then to normalize for dropouts and number of pairwise distance matrices, we divided each entry in the ensemble counts matrix by the number of non-Nan elements in that entry for all of the pairwise distance matrices. This yields the ensemble frequency matrix. Frequency matrices are shown in orange heatmaps.

Creation of Ensemble average matrix

To compute an ensemble average matrix, as seen in Fig. 3, we began by adding all of the pairwise distance matrices within a dataset and dividing the value by the total number of pairwise distance matrices in the dataset.

Simulating pairwise distance matrices using the string and binders model

We used a published strings and binders (SBS) model to simulate pairwise distance matrices of chromatin folding structures²⁴. We created simulations using default parameters as stated in settings_SBS_IMR90_30kb.py, modifying polymer size from $n = 750$ to $n = 302$. To create variable domains within the SBS simulation, we modified the polymer_IMR90.bed file to have semi-random locations of binding domains throughout the polymer. The semi-random locations are to ensure capturing the complexity of the chromatin structures

found within the sequential Oligopaints data. We varied the number of domains per map from 2–5, varied domain sizes from 50–150 monomers and incorporated nested structures (Extended Data Fig. 3a).

Receiver operator curve creation for FISHnet and boundary caller

We computed a ROC using calls from FISHnet and the previously published boundary caller¹⁵ to test the sensitivity and specificity of the algorithms. First, we ranked each FISHnet boundary call by varying the plateau parameter from 0 to 50. Having a small plateau value enables the majority of FISHnet calls to pass regardless of frequency; having a large plateau value allows only the most consistent calls to pass. We ran FISHnet on the simulated data set with a size exclusion of 10 monomers, no merge, and step size of 10 nm. For the boundary caller, we ranked each boundary call by varying their cutoff_max parameter from 0 to 6.1 in step sizes of .01. We used the default parameters for the boundary caller method: su = 4, sl = 2, valley = 3, gb = 0.75, func = np.nanmean (ref. 15). For both FISHnet and the boundary caller, we created a confusion matrix for all of the pairwise distance matrices in the simulated dataset ($n = 606$). In the confusion matrices, we counted boundaries within 3 monomers of a ground truth boundary as a true positive (TP). We classified a FISHnet boundary as a false positive (FP) when FISHnet called a boundary with no ground truth boundary within three monomers. Then, we count a false negative (FN) when FISHnet failed to call a boundary at a ground truth boundary. Finally, a true negative (TN) occurs when neither FISHnet called a boundary nor a ground truth boundary exists. The same logic was applied for the boundary caller in determining TPs, FPs, TNs, and FNs. We calculated the false positive rate calculated as $FP / (FP + TN)$ and the true positive rate as $TP / (TP + FN)$. We plotted the ROC curve with the true positive rate on the x axis and the false positive rate on the y axis.

Parameter sweep of smoothing window, distance step, and plateau size

To assess FISHnet's generalizability, we conducted a parameter sweep assessing how different smoothing windows, distance steps, and plateau sizes affect FISHnet's performance (Extended Data Fig. 2). First, we used $n = 606$ simulated pairwise distance matrices from the SBS model to assess how different sizes of smoothing windows affect FISHnet's ability to call ground-truth domains while holding plateau size and nanometer step constant. We observe that small smoothing windows tend to correlate best with bona fide domain calls and larger windows tend to skew the data through over smoothing (Extended Data Fig. 2a,b). AUC quantification of FISHnet's performance demonstrates improved sensitivity and specificity for domain calls upon smoothing with window sizes less than 3% of the map. (Extended Data Fig. 2c).

Second, we assessed how distance step size affects FISHnet's performance. We swept nanometer step sizes from 1 nm–2,500 nm while holding the smoothing window and plateau size constant. AUC analysis revealed that increasing step size leads to reduced AUC. Step sizes of less than 200 nm resulted in superior AUC performance, with an optimum step size of 5 nm (Extended Data Fig. 2d). We note that lowering the nanometer step size results in drastic increases in computation time for only marginal gain in AUC performance.

Finally, we assessed how different plateau sizes affect FISHnet's performance on calling domains. To determine how plateau sizes interact with different distance steps, we calculated a true positive rate (TPR) and false positive rate (FPR) heatmap, with plateau sizes varying from 0 to 50 and distance steps from 1 nm to 200 nm (Extended Data Fig. 2e,f). We observe that plateau size has an indirect relationship with TPR and FPR. Lower plateau size increases TPR and FPR, whereas higher plateau size decreases TPR and FPR. The distance step size also has an indirect relationship with TPR and FPR. When using smaller step sizes, users can sample the pairwise distance matrix frequently (higher plateau size) to prevent increases in FPR. Moreover, when using larger

step sizes, users can sample the pairwise distance matrix less frequently (lower plateau size) to ensure a non-zero TPR.

For the analyses in this manuscript, we selected a plateau size of 4 and nanometer step size of 10, and the resultant TPR and FPR is highlighted by the yellow box in Extended Data Figure 2e,f. Overall, FISHnet's domain calling performance is optimized by mitigating high FPRs with small distance steps by using larger plateau values, and retaining high true positive rates with large distance steps by using smaller plateau values all while keeping a low smoothing window size.

FISHnet excitatory neuron and microglia subcluster classification

To identify subclusters within the excitatory neuron and microglia population as shown in Fig. 3a,b, we used FISHnet calls associated with the largest nanometer plateau calls. We constructed an $N \times M$ matrix for both excitatory neurons and microglia, where N represents a genomic bin, and M corresponds to a single allele. For each allele in the $N \times M$ matrix, a value of 1 indicates the presence of a boundary at genomic bin i in allele j (that is, at position $[i, j]$ in the matrix), while a value of 0 indicates its absence. Then, we computed the correlation matrix to determine the Pearson correlation coefficients between each pair of alleles, transforming the $N \times M$ matrix into an $M \times M$ correlation matrix. Using Python's 'networkx package', we applied the 'greedy_modularity_maximization' module to identify clusters within the correlation matrix. We used the 'networkx.algorithms.community.greedy_modularity_communities' command, where G is the $M \times M$ correlation matrix thresholded at either 0.75 or 0.5 for excitatory neurons or microglia, respectively. We retained clusters comprising more than three pairwise distance matrices and discarded the rest.

Chi-square test for identifying cell type-specific boundaries

To identify boundaries that are significantly different between two cell types, we conducted a chi-square test of independence for FISHnet calls from each genomic locus using Oligopaints data from microglia and excitatory neurons (Fig. 4a). For a given boundary, we constructed a 2×2 contingency table in which we classified and tallied each microglia and neuron single allele as either having or not having the boundary as measured by FISHnet. We used the Python package Scipy to run `scipy.stats.chi2_contingency(contingency_table)` to test the null hypothesis that the boundary occurs in similar proportions of alleles in both cell type A and cell type B. The alternative hypothesis is that the boundary does not occur in similar proportions across alleles in both cell type A and cell type B.

Permutation test for identifying cell type-specific boundaries

We formulated a permutation test to identify microglia-specific boundaries (cell-type-A-specific boundaries), as shown in Fig. 4b. We selected a test statistic of the proportion of alleles with a FISHnet called boundary at a particular genomic locus. Using the microglia and neuron data, we ran a one-tailed, right-tailed test for the null hypothesis that the proportion of microglia and excitatory neuron alleles with a boundary is unchanged. The alternative hypothesis is that the proportion of microglia alleles with a boundary is higher than the proportion of neuron alleles with a boundary. We compute the test statistic using the microglia data and compare it with a null distribution formulated by 10,000 random samplings, without replacement, of $n = 130$ alleles from the excitatory neuron single allele Oligopaints data. We compute a one-tailed, right-tailed empirical P value as the AUC of the null distribution to the right of the microglia test statistic.

Pairwise distance matrix PCA for IMR90, K562, and A549 alleles

We performed principal component analysis on the pairwise distance matrices in Fig. 4d using $n = 2,101$, 2,060, and 1,762 pairwise distance matrices for IMR90, K562, and A549, respectively. We kept pairwise

distance matrices with >95% coverage, resulting in $n = 2,082, 2,032$, and $1,700$ alleles for IMR90, K562, and A549, respectively. We then extracted the distances from each pair of genomic loci and created a matrix where each row stores the distances between genomic loci. Thus, in our matrix sized $R \times P$, rows (R) represent individual alleles and columns (P) contain the distances between two genomic loci for each allele. We used `sklearn.preprocessing.scale` to zero mean scale the matrix, calculated the covariance matrix, and computed the eigenvectors and eigenvalues using `numpy.linalg.eigh`. We ranked the eigenvalues and selected the eigenvectors corresponding to the two largest eigenvalues and created a reduced eigenvector matrix. Finally, we reduced the scaled matrix by taking the dot product with the transposed scaled matrix and the reduced eigenvector matrix and plotted the first two principal components.

FISHnet PCA for IMR90, K562, and A549 alleles

We conducted principal component analysis on a masked domain representation array from FISHnet domain calls in Fig. 4e using $n = 2,101, 2,060$, and $1,762$ pairwise distance matrices for IMR90, K562, and A549 cells, respectively. For this analysis, we used FISHnet domain calls with the smallest nanometer plateau to focus on nested domains. We created a $N \times M$ matrix as described in the 'FISHnet excitatory neuron and microglia subcluster classification' section for IMR90, K562, and A549 single alleles. Next, we convolved each column in the $N \times M$ matrix, where M stores the FISHnet boundary calls for a given allele (kernel = 150). This convolution retains contextual boundary information, allowing statistical methods to give boundaries that are close to each other more weight than those that are farther apart when comparing two alleles. We then concatenated each of the $N \times M$ matrices together creating a matrix that contains the boundary calls for each cell type for all of the alleles present in the IMR90, K562, and A549 datasets. We used `sklearn.preprocessing.scale` to zero mean scale the matrix, calculated the covariance, and computed the eigenvectors and eigenvalues using `numpy.linalg.eigh`. We ranked the eigenvalues and selected the eigenvectors corresponding to the two largest eigenvalues and created a reduced eigenvector matrix. Finally, we reduced the scaled matrix by taking the dot product with the transposed scaled matrix and the reduced eigenvector matrix and plotted the first two principal components.

Ensemble FISHnet domain mask creation

We first transformed each individual pairwise distance matrix by creating an integer mask based on each map's FISHnet domain calls (Fig. 5a). The resultant mask transforms the pairwise distances into an integer encoding how many domains FISHnet finds each pair of genomic loci (Fig. 5a, bottom left). If a pixel is not in any FISHnet domain, its value in the mask is 0. If a pixel is in only one FISHnet domain, its value in the mask is 1. In cases where a pixel is in two FISHnet domains, such as one nested within another larger one, its value is 2. To create an ensemble FISHnet domain mask, we summed individual masks, resulting in a matrix in which each pixel represents the count of domains FISHnet found for each pair of genomic loci (Fig. 5a, right).

FISHnet Settings for sequential Oligopaints datasets

For all sequential Oligopaint datasets, we ran FISHnet run with a plateau sweep of 4, size exclusion of 3, merge of 3, and smoothing window size of 2×2 . We used size exclusion to require communities to be greater than three. The merge parameter dictated the averaging of boundaries within three bins of each other.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All public datasets used in this study are defined in Table 1.

Code availability

FISHnet code is provided publicly at the following GitHub link: <https://github.com/RohanpatelUpenn/FISHnet>.

References

39. Zhang, H. et al. Chromatin structure dynamics during the mitosis-to-G1 phase transition. *Nature* **576**, 158–162 (2019).
40. Beagan, J. A. et al. Three-dimensional genome restructuring across timescales of activity-induced neuronal gene expression. *Nat. Neurosci.* **23**, 707–717 (2020).
41. Malachowski, T. et al. Spatially coordinated heterochromatinization of long synaptic genes in fragile X syndrome. *Cell* **186**, 5840–5858 (2023).
42. Maury, E. A. et al. Schizophrenia-associated somatic copy-number variants from 12,834 cases reveal recurrent NRXN1 and ABCB1 disruptions. *Cell Genom.* **3**, 100356 (2023).

Acknowledgements

We thank members of the Cremins lab for helpful discussions. The fish and net art used was created by Laymik from the Noun Project. Funding: NIH National Institute of Mental Health (1R01MH120269; 1DP1MH129957; J.E.P.C.); NIH NINDS (5-R01-NS114226; J.E.P.C.); 4D Nucleome Common Fund grants (1U01DK127405, 1U01DA052715; J.E.P.C.); NSF CAREER Award (CBE-1943945; J.E.P.C.); Chan Zuckerberg Initiative Neurodegenerative Disease Pairs Awards (2020-221479-5022; DAF2022-250430; J.E.P.C.); NIH F31 Fellowship (F31HG014082; R.P.); NIH F30 Fellowship (F30HD114405; K.P.).

Author contributions

Conceptualization: R.P., J.E.P.-C. Methodology: R.P., J.E.P.-C. Investigation: R.P., K.P., H.C., J.E.P.-C. Visualization: R.P., K.P., H.C., J.E.P.-C. Funding acquisition: J.E.P.-C. Project administration: J.E.P.-C. Supervision: J.E.P.-C. Writing and editing: R.P., K.P., J.E.P.-C.

Competing interests

The authors declare no competing interests.

Additional information

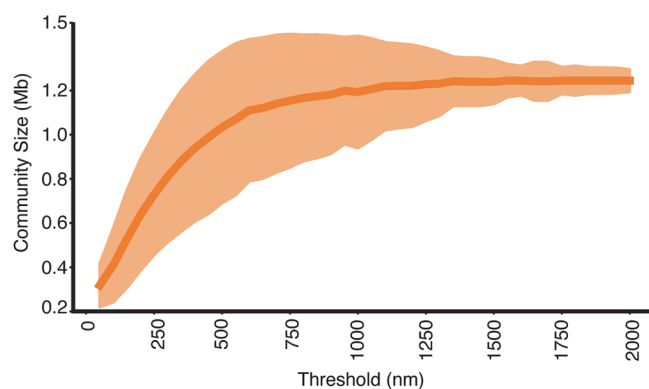
Extended data is available for this paper at <https://doi.org/10.1038/s41592-025-02688-1>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41592-025-02688-1>.

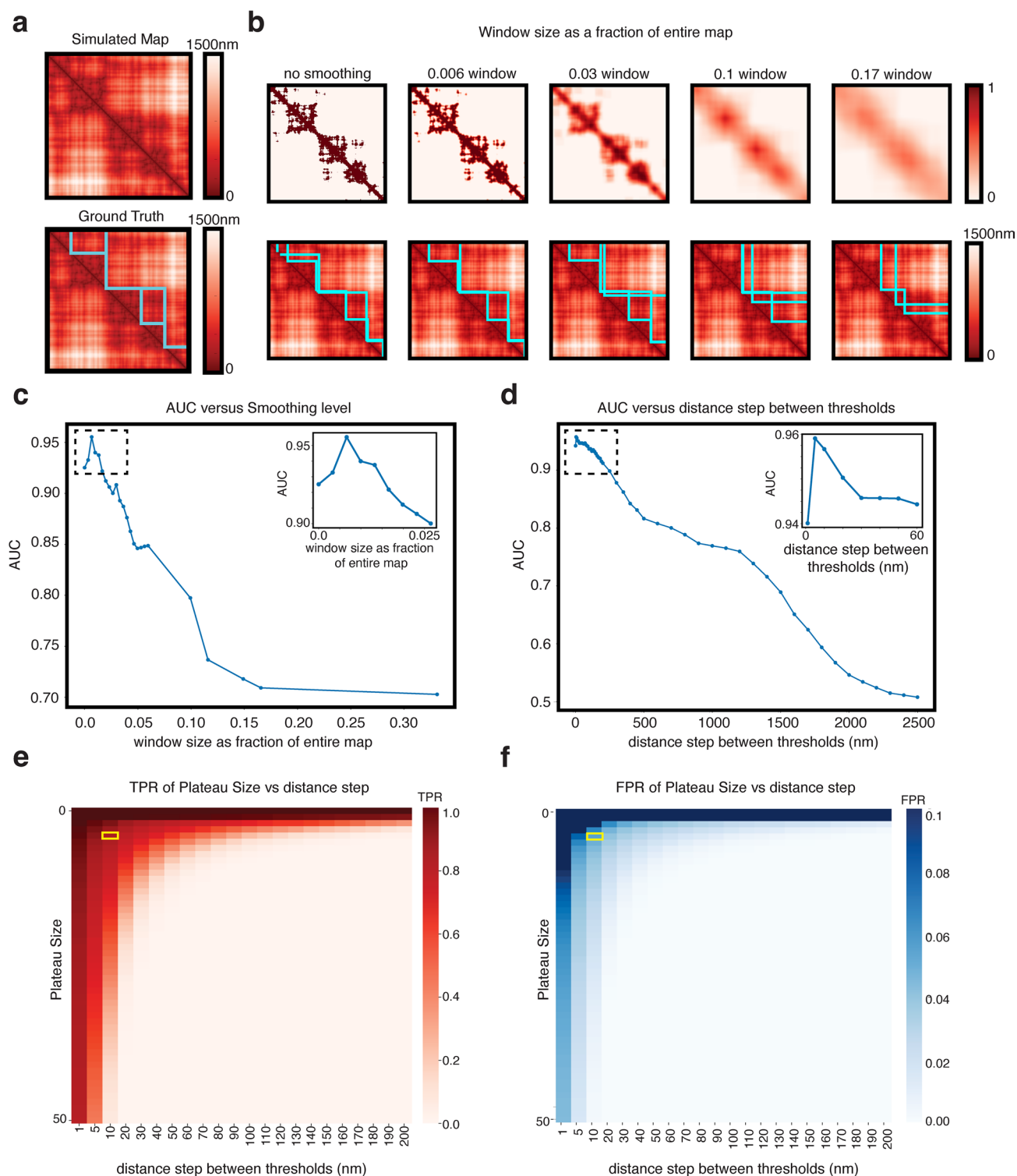
Correspondence and requests for materials should be addressed to Jennifer E. Phillips-Cremins.

Peer review information *Nature Methods* thanks the anonymous reviewers for their contribution to the peer review of this work. Primary Handling Editor: Rita Strack, in collaboration with the *Nature Methods* team.

Reprints and permissions information is available at www.nature.com/reprints.

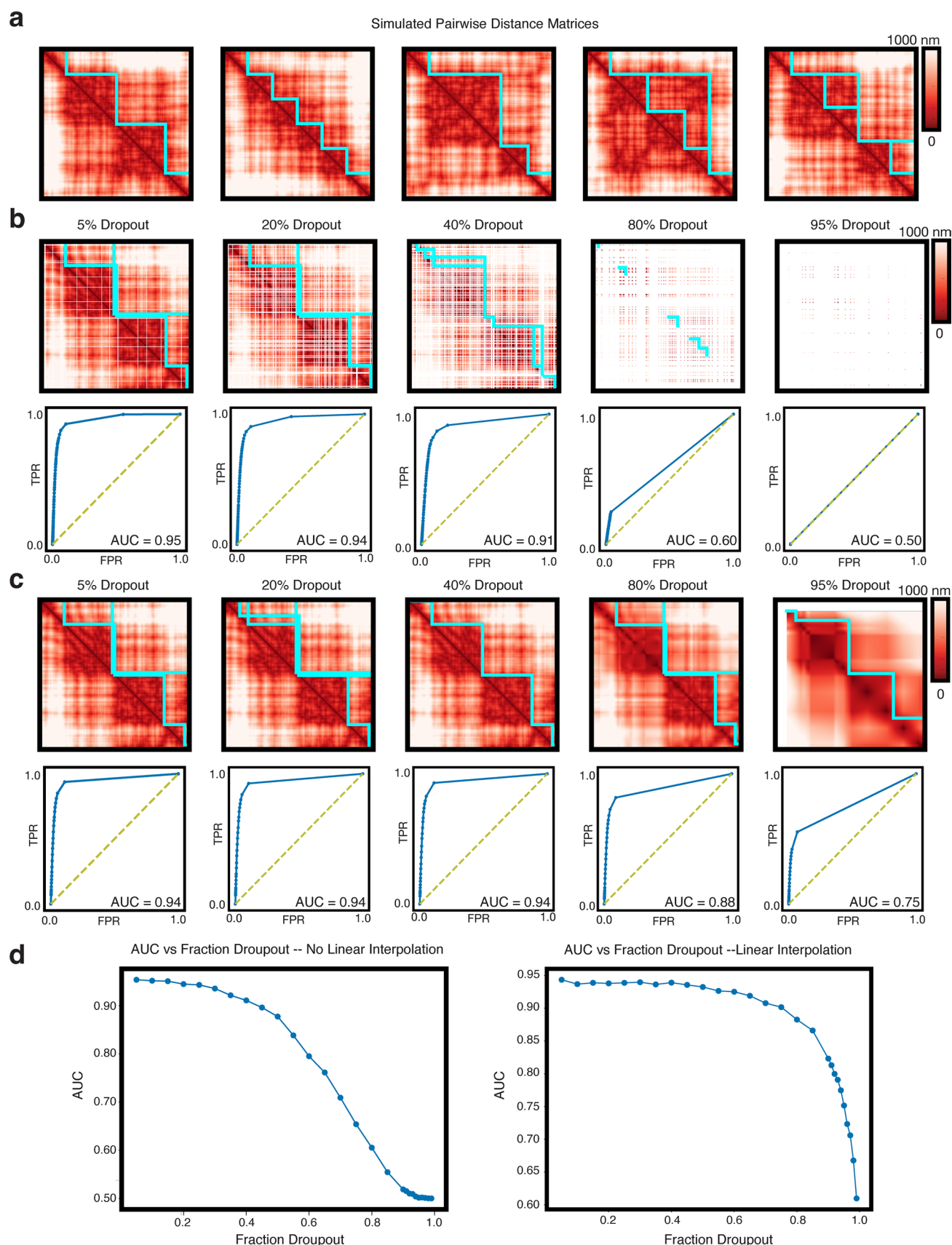


Extended Data Fig. 1 | Use of nanometer distance threshold as a resolution parameter. Line plot of average domain call sizes at different nm distance threshold values for a chromatin tracing data in HCT116 cells (Chr21: 34.6–37.1 Mb at 30 kb resolution)¹⁵. N = 595 pairwise distance maps were used. Shadow indicates average domain call \pm one standard deviation.



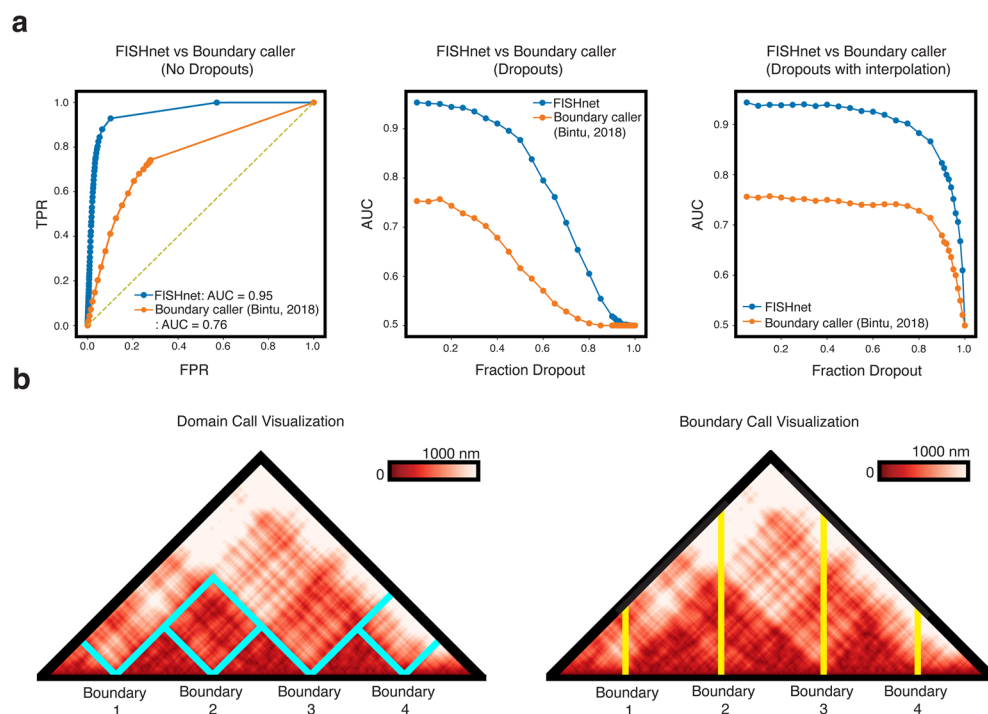
Extended Data Fig. 2 | Effect of smoothing window, distance step between thresholds, and plateau size on FISHnet's performance. (a) Example of a pairwise distance matrix of chromatin folding from Strings and Binders (SBS) model²⁴ (top) with ground truth domains calls overlaid (bottom). (b) The pairwise distance matrix from panel (a) thresholded at 250 nm and different window sizes of smoothing applied to it (top). Window size is reported as a fraction of the entire map. The corresponding FISHnet domain calls with the respective window size of smoothing (bottom). (c) Area under the curve (AUC) as

a function of different window size of smoothing applied. (d) AUC as a function of different distance step between thresholds. (e-f) Heatmap of TPR (e) and FPR (f) as plateau size and distance step between thresholds. Yellow box indicates the plateau size and nanometer step size used on the sequential Oligopaints data within this paper. The distance step between thresholds and plateau size were held constant in panel (d). Window size was held constant in panel (e)-(f) N = 606 pairwise distance matrices of chromatin folding from a SBS model was used in panels (c)-(f).



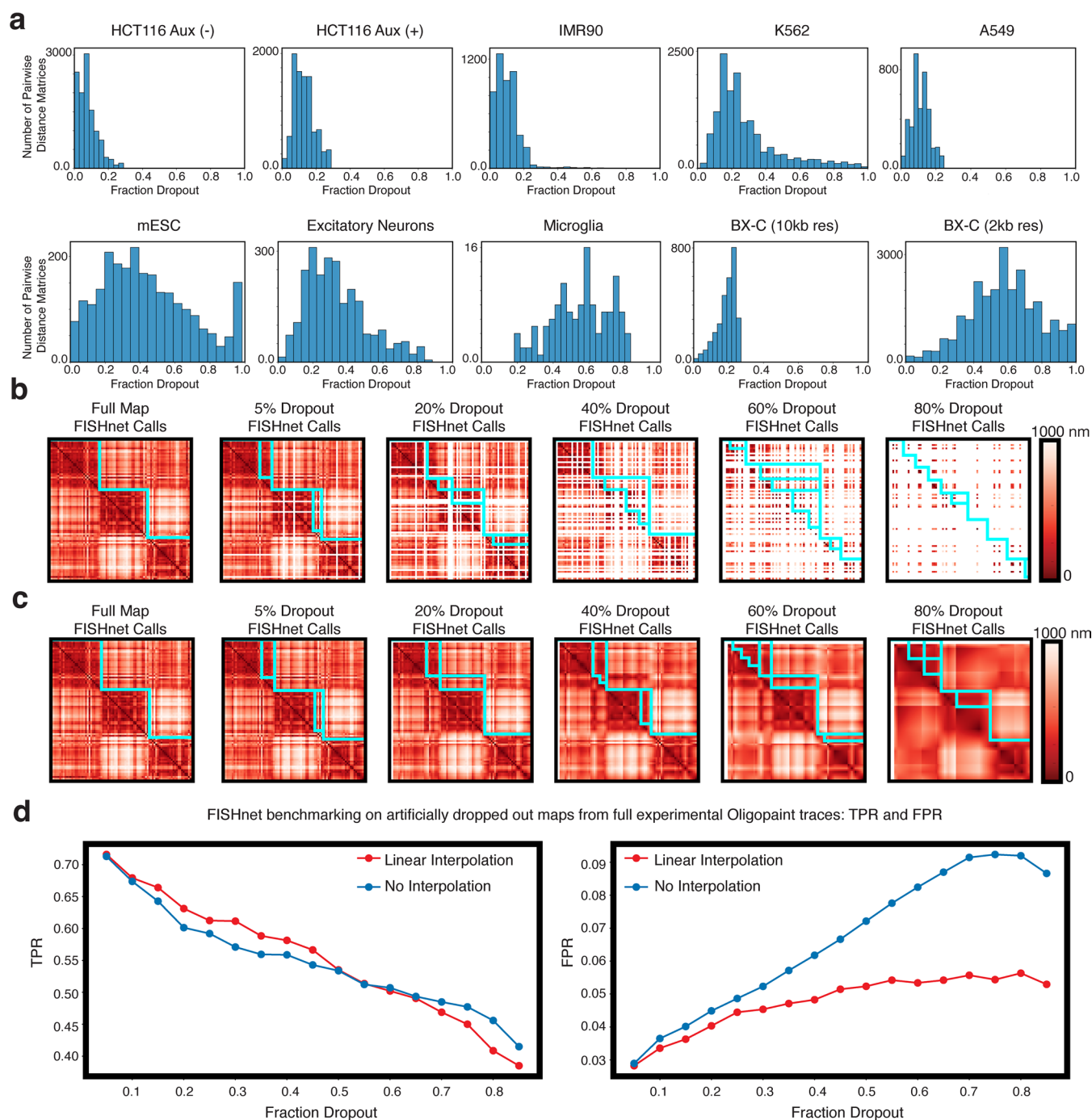
Extended Data Fig. 3 | Assessment of FISHnet's sensitivity and specificity across a range of loci dropout and with and without linear interpolation. (a) Representative examples of SBS pairwise distance matrices with ground truth domain calls. (b) Visualization of dropouts with receiver operating curve (ROC) for FISHnet's performance without linear interpolation. White lines indicate

data with dropouts. (c) Visualization of dropouts with ROCs for FISHnet's performance with linear interpolation. Yellow dotted line showcases $y = x$. (d) AUC versus fraction dropout without (left) and with (right) linear interpolation. $N = 606$ pairwise distance matrices of chromatin folding from a SBS model was for this analysis.



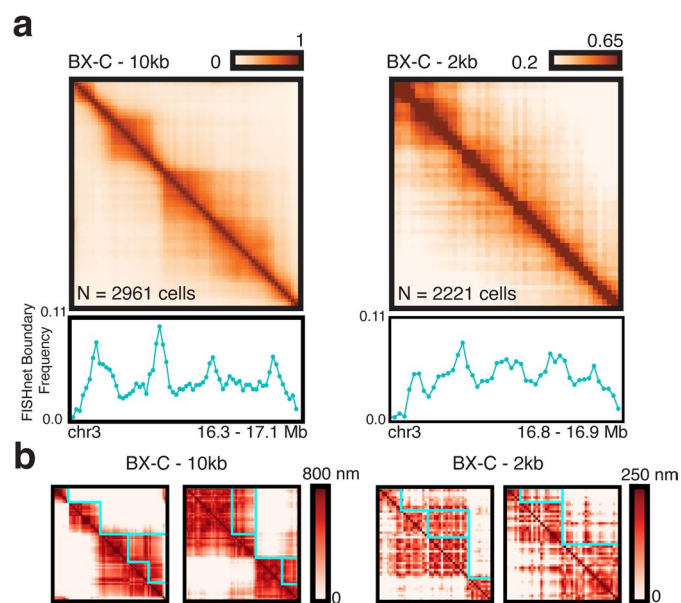
Extended Data Fig. 4 | FISHnet comparison with boundary caller. (a) ROC measuring performance on simulated pairwise distance matrices of chromatin folding from SBS model (left). FISHnet AUC = 0.95, boundary caller AUC = 0.76. AUC versus fraction dropout on simulated pairwise distance matrices of chromatin folding from SBS model (middle). AUC versus fraction dropout with linear interpolation on simulated pairwise distance matrices of chromatin

folding from SBS model (right). N = 606 pairwise distance matrices of chromatin folding data from the SBS model were used. (b) Visualization of domain calls versus boundary calls. Domain calls enable identification of hierarchy within the domain structures (domain from boundary 1 to boundary 3 encompasses domains from boundary 1 to 2, and 2 to 3).



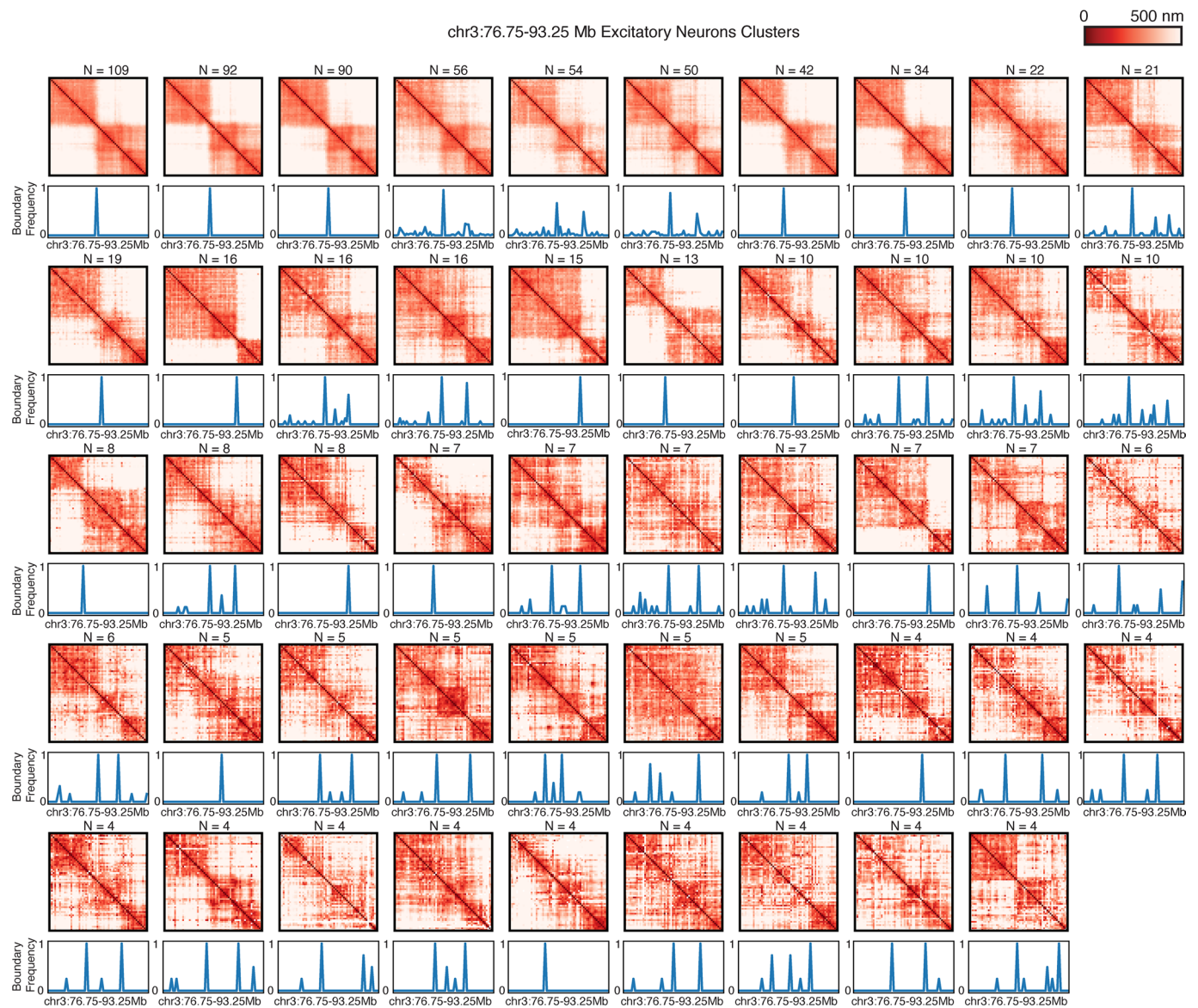
Extended Data Fig. 5 | FISHnet's performance on artificially dropped out maps. (a) Histograms showing the range of dropouts in experimental Oligopaints data. **(b–c)** A single allele pairwise distance matrix with 0% dropouts from HCT116 cells, chr21: 34.6–37.1 Mb at 30 kb resolution¹⁵ with FISHnet domain call (left). To the right, progressive artificial dropouts with FISHnet domains calls without

linear interpolation **(b)** and with linear interpolation **(c)**. **(d)** Quantification of TPR and FPR as a function of dropout with and without interpolation, using the 0% dropout FISHnet calls as the ground truth. $N = 595$ pairwise distance matrices from HCT116 cells, chr21: 34.6–37.1 Mb at 30 kb resolution with 0% dropouts were used for panel **(d)**.

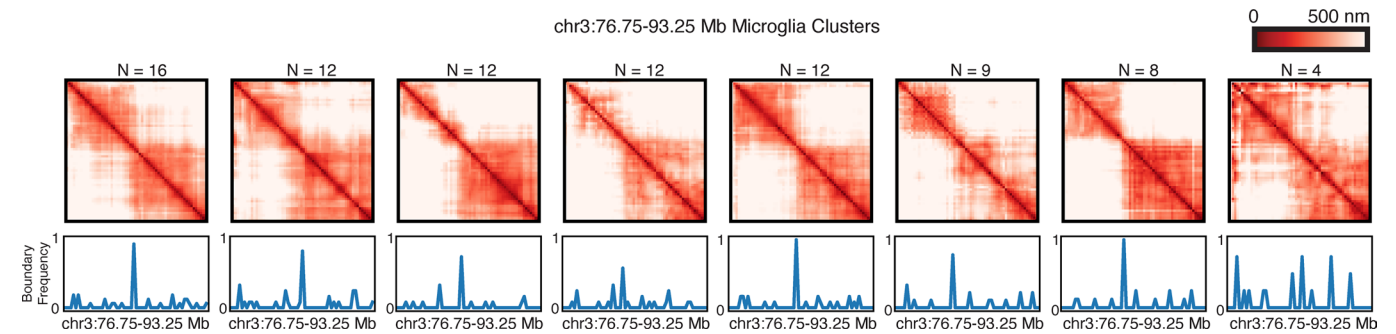


Extended Data Fig. 6 | FISHnet on 10 kb and 2 kb data. (a) Ensemble frequency matrices of sequential Oligopaints imaging data in Bithorax Complex (BX-C) region of *Drosophila melanogaster* embryos 10 kb resolution (left) and 2 kb resolution (right)¹⁷. Blue line plots indicate the frequency of FISHnet boundary

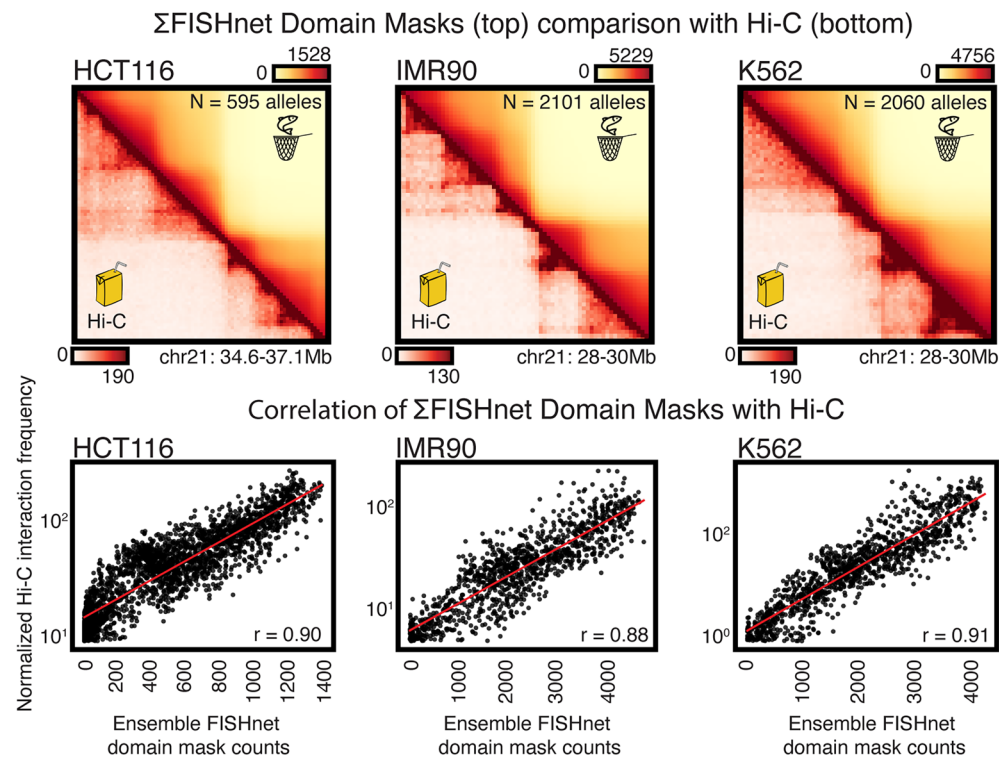
calls across N = 2961 and N = 2221 cells in 10 and 2 kb, respectively. **(b)** Individual pairwise distance matrices with FISHnet calls on published sequential Oligopaints imaging data at 10 kb resolution (left) and 2 kb (right).



Extended Data Fig. 7 | Excitatory neuron clusters created using FISHnet boundary calls. Mean pseudo-ensemble bulk matrices for each cluster along with FISHnet frequency domain calls underlaid underneath each cluster. N indicates the number of pairwise distance matrices present in each cluster.

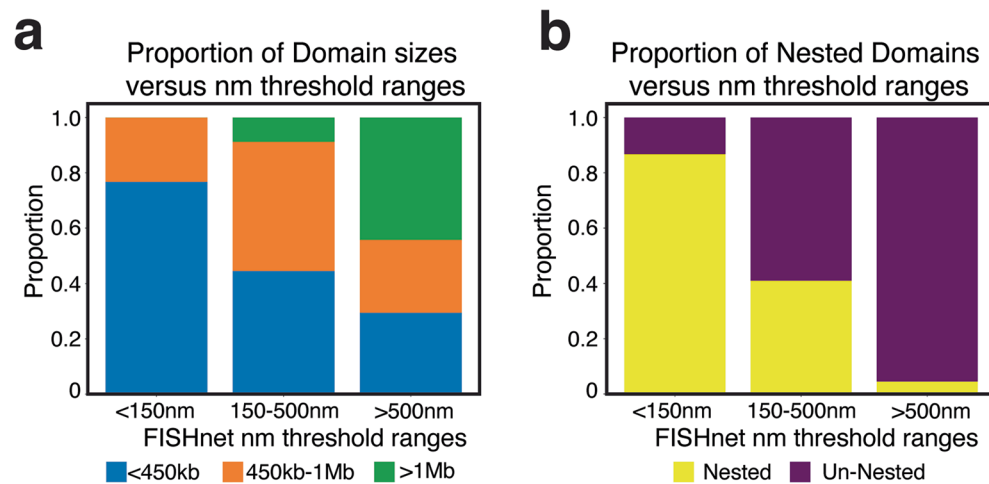


Extended Data Fig. 8 | Microglia clusters created using FISHnet boundary calls. Mean psuedo-ensemble bulk matrices for each cluster along with FISHnet frequency domain calls underlaid underneath each cluster. N indicates the number of pairwise distance matrices present in each cluster.



Extended Data Fig. 9 | Correlation of FISHnet domain mask with Hi-C data.
Visualization of FISHnet domain masks with Hi-C data in HCT116, IMR90, and K562 cells (top). Scatterplots of the normalized Hi-C interaction frequency and

the ensemble FISHnet domain mask counts for each cell type are shown. Red line indicates the linear line of best fit, and Pearson correlation values are indicated for each scatterplot.



Extended Data Fig. 10 | Proportions of domain sizes and nested domains at different threshold ranges. (a) Proportion of domain sizes less than 450 kb, between 450 kb and 1 Mb, and greater than 1 Mb across three different threshold

ranges. **(b)** Proportion of nested domains and un-nested domains in three different threshold ranges. (Chr21: 34.6–37.1 Mb at 30 kb resolution)¹⁵. N = 595 pairwise distance maps were used.

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

| | |
|-------------------------------------|--|
| n/a | Confirmed |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> The exact sample size (<i>n</i>) for each experimental group/condition, given as a discrete number and unit of measurement |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> The statistical test(s) used AND whether they are one- or two-sided <i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i> |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> A description of all covariates tested |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> For null hypothesis testing, the test statistic (e.g. <i>F</i> , <i>t</i> , <i>r</i>) with confidence intervals, effect sizes, degrees of freedom and <i>P</i> value noted <i>Give P values as exact values whenever suitable.</i> |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> Estimates of effect sizes (e.g. Cohen's <i>d</i> , Pearson's <i>r</i>), indicating how they were calculated |

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

| | |
|-----------------|--|
| Data collection | No software was used for data collection |
| Data analysis | Python 3 was used. Specifically we used these packages: matplotlib>=3.0.0 numpy>=1.6.0 pandas>=0.11.0 scikit-learn>=0.14 scipy=1.13.1 |

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

All data used in this manuscript is publicly available and listed in Table 1.

Human research participants

Policy information about [studies involving human research participants and Sex and Gender in Research](#).

Reporting on sex and gender

N/a

Population characteristics

N/a

Recruitment

N/a

Ethics oversight

N/a

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size

We used sample sizes in the permutation test. We determined the sample size for the Excitatory Neuron population by matching it with the N in which it was compared to i.e. microglia population which was N=130. All other indications of "N" refer to the number of pairwise distance matrices used.

Data exclusions

We excluded pairwise distance matrices with high dropouts. We have mentioned the thresholds used within the main text.

Replication

We only used publicly available data.

Randomization

We only used publicly available data, and no human experiments were done.

Blinding

We only used publicly available data, and no human experiments were done.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

| | |
|-------------------------------------|--|
| n/a | Involved in the study |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Antibodies |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Eukaryotic cell lines |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Palaeontology and archaeology |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Animals and other organisms |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Clinical data |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Dual use research of concern |

Methods

| | |
|-------------------------------------|---|
| n/a | Involved in the study |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> ChIP-seq |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Flow cytometry |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> MRI-based neuroimaging |