

UC San Diego

UC San Diego Electronic Theses and Dissertations

Title

Multiplexed genome imaging and analysis

Permalink

<https://escholarship.org/uc/item/4d3441g8>

Author

Jia, Bojing Blair

Publication Date

2024

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA SAN DIEGO

Multiplexed genome imaging and analysis

A Dissertation submitted in partial satisfaction of the requirements
for the degree Doctor of Philosophy

in

Bioinformatics and Systems Biology

by

Bojing Blair Jia

Committee in charge:

Professor Bing Ren, Chair
Professor Albert Hsiao, Co-Chair
Professor Vineet Bafna
Professor Pallav Kosuri
Professor Gal Mishne
Professor Jin Zhang

2024

Copyright

Bojing Blair Jia, 2024

All rights reserved.

The Dissertation of Bojing Blair Jia is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2024

TABLE OF CONTENTS

<u>DISSERTATION APPROVAL PAGE</u>	iii
<u>TABLE OF CONTENTS</u>	iv
<u>LIST OF FIGURES</u>	vi
<u>ACKNOWLEDGEMENTS</u>	vii
<u>VITA</u>	xi
<u>ABSTRACT OF THE DISSERTATION</u>	xii
<u>Chapter 1 : A SPATIAL GENOME ALIGNER FOR RESOLVING CHROMATIN ARCHITECTURES FROM MULTIPLEXED DNA FISH</u>	1
<u>1.1 Abstract</u>	1
<u>1.2 Introduction</u>	2
<u>1.3 Results</u>	6
<u>1.4 Discussion</u>	25
<u>1.5 Methods</u>	29
<u>1.6 Figures</u>	44
<u>1.7 Supplementary Figures</u>	53
<u>1.8 Acknowledgements</u>	65
<u>1.9 References</u>	66
<u>Chapter 2 PRINCIPLES OF SPATIAL MULTIOMICS: INSIGHTS AND APPLICATIONS TO DERMATOLOGY</u>	74
<u>2.1 Abstract</u>	74
<u>2.2 Introduction</u>	75
<u>2.3 Principles of Spatial technologies</u>	77
<u>2.4 Multi-omic technologies</u>	85
<u>2.5 Future Directions and Applications of Spatial Multiomics to Dermatologic Research</u>	88
<u>2.6 Figures</u>	91

<u>2.7 Acknowledgements</u>	95
<u>2.8 References</u>	96

LIST OF FIGURES

<u>Figure 1.1: Spatial genome alignment of multiplexed DNA-FISH imaging data against a reference soft-polymer structural model of DNA.....</u>	<u>44</u>
<u>Figure 1.2: Spatial genome alignment of seqFISH+ chromatin imaging of mESCs at 25 kb and 1 Mb resolution.....</u>	<u>45</u>
<u>Figure 1.3: Polymer fiber karyotyping on interphase mES cells.....</u>	<u>46</u>
<u>Figure 1.4: Sister chromatid interactions in mESCs resolved by polymer fiber karyotyping.....</u>	<u>48</u>
<u>Figure 1.5: Sister chromatid interactions in compact homolog state.....</u>	<u>49</u>
<u>Figure 1.6: Copy number variations in mouse cortical neurons.....</u>	<u>51</u>
<u>Figure S1.1: Spatial genome alignment on 1 Mb seqFISH+ chromatin imaging of mESCs.....</u>	<u>53</u>
<u>Figure S1.2: Spatial genome alignment on sequential 5 kb M-DNA-FISH imaging of mESC Sox2 locus.....</u>	<u>54</u>
<u>Figure S1.3: Spatial genome alignment on simulated aneuploid cells.....</u>	<u>56</u>
<u>Figure S1.4: Polymer fiber karyotyping on simulated aneuploid cells.....</u>	<u>58</u>
<u>Figure S1.5: Assessment of misselection between tightly coupled sister chromatids.....</u>	<u>59</u>
<u>Figure S1.6: Spatial genome alignment locus detection in putative sister chromatids.....</u>	<u>60</u>
<u>Figure S1.7: Spatial genome alignment on simulated paired chromatin fibers.....</u>	<u>61</u>
<u>Figure S1.8: Haplotype-resolved single-cell sequencing of mouse cortical excitatory neurons.....</u>	<u>62</u>
<u>Figure S1.9: Effect of gene dosage on transcription in excitatory mouse cortical neurons.....</u>	<u>63</u>
<u>Figure S1.10: Polymer fiber karyotyping routine.....</u>	<u>64</u>
<u>Figure 2.1: Classification and illustration of mechanism among families of spatial genomic technologies.....</u>	<u>91</u>
<u>Figure 2.2: Contrastive advantages and weaknesses between imaging in situ and release-and-capture spatial technologies.....</u>	<u>92</u>
<u>Figure 2.3: Strategies for multiomics utilized by spatial technologies.....</u>	<u>93</u>
<u>Figure 2.4: Potential future applications of spatial multiomics in cutaneous research.....</u>	<u>94</u>

ACKNOWLEDGEMENTS

In writing this thesis, I feel the importance of recognizing not just the people involved, but also the exceptional circumstances under which they provided their support, because it makes them all the more remarkable. Not long after my PhD started in late 2019, a global pandemic soon shuttered all labs. Projects got upended. Alone for months, we couldn't fully collaborate. A supply chain shortage wreaked havoc on experiments. Incredulously I am graduating, and it is in enormous part thanks to the extraordinary support I have received from the following extraordinary people:

First and foremost I extend my deepest thanks to my advisor Bing Ren for his vision, kindness, and steadfast support. How he tolerated my “heretical” endeavors, as joked by other faculty, is a testament to his open-mindedness and unwavering belief in the potential of our work. He has a remarkable talent for spotting transformative technology. Bing emboldened me with a rare, shared vision that gave me new recognition for how truly exciting science can be – one that I often found myself waking up at 4 AM to, rising with bounding enthusiasm that I get to continue working on such an elegant research problem that became the spatial genome aligner. It's been such an honor to train with him, and I hope to pay it forward by mentoring in his style in my future scientific career.

My wife Xingzhao, my better half and a wellspring of scientific inspiration, first showed me the beauty of computational biology through stories (ie. Procrustes) that breathed life to abstract algorithms. She's always full of dynamism and optimism it is infectious, and my work and life have both been uplifted by her. She has been there for me through thick and thin, and during the many difficult times these past years I have leaned on her – sometimes literally when I

couldn't stand up – for support. Without her, I wouldn't be where I am today, and to her, I am forever grateful.

I owe immense gratitude to my parents and sister for their steadfast support in every aspect of my life. My father, my first scientific mentor, instilled in me from an early age an appreciation for rhetoric and for the art of problem solving, useful lessons for my scientific journey. My mother, the embodiment of determination and grit, made sacrifices (often too great) for my sister and me that have shaped my resolve during my PhD, because if she could overcome anything so could I. My sister, the even-keeled individual I strive to be, has supported both our parents and me in times I could not, a duty I feel should not have fallen on my baby sister. Their collective sacrifices have afforded me incredible opportunities, and to them, I am also forever grateful.

My committee members Pallav Kosuri, Gal Mishne, Vineet Bafna, Jin Zhang, and co-chair Albert Hsiao provided invaluable advice, with Albert taking me under his wing before the pandemic, Gal patiently guiding me as I tried to code dimensionality reduction algorithms from scratch, Vineet teaching me algorithms outside traditional bioinformatics that truly broadened my problem solving skillset, Jin providing me critical imaging expertise, and Pallav mentoring me over the pandemic – from practice talks to microscope building and even surfing together. An important footnote is that Pallav was one of the first to see my algorithm during the pandemic, when opportunities for feedback were scarce and I wasn't sure if my work was just a hallucination. When I presented to Pallav over Zoom, his understanding of biophysics and belief in my algorithm's value gave me tremendous confidence (and a much needed sanity check) that carried me through my Ph.D. Additionally, I want to convey incredible gratitude for the opportunity to collaborate with Vineet on ongoing projects deriving from spatial genome

alignment. While outside the scope of this thesis, our collaborative work has gifted me an immense sense of fulfillment and academic challenge, as I never would have predicted I would have the opportunity to learn from him the art of designing very elegant algorithms.

I will always cherish the stimulating discussions and fits of laughter in the Ren Lab. I'm grateful to Yang Xie, a good friend and wizard in his craft, for our invaluable late-night conversations filled with his insight and humor. Ethan Armand is an incredibly considerate soul whose thoughtful care, especially during stressful times, and his detailed feedback, have been incredibly supportive. I owe tremendous thanks to Zane Gibbs for guiding me in baby-steps my first imaging experiments and tolerating my enthusiastic ideas. On that note Brett Taylor and Jenny Weixiu Dong also deserve thanks for their patience with my bursts of excitement. I am also grateful for Jie Xu and Zhaoning Wang for teaching me how to do my first Hi-C experiments, Quan Zhu, Yuanyuan Han, Adam Jussila and Colin Kern for introducing me to chromatin imaging, Melodi Frey for helping with iPSC cell culture, Kseniya Malukhina and Mahsa Nafisi from the neighboring Bintu lab for helping me with hardware and molecular biology involved in multiplexed DNA-FISH, and for Jeffrey Chiu and Weronika Bartosik for helping with Hi-C library prep. My thanks extend to all past and present members of the lab for their contributions: Ramya Raviram, Chenxu Zhu, Yanxiao Zhang, Yang Li, Benson Chen, Dan Cunningham-Bryant, Lei Chang, Songpeng Zu, Bharath Saravanan, Luisa Amaral, Hannah Indralingam, Nathan Zemke, Kai Zhang, Sainath Mamde, Seoyeon Lee, Sam Kuan, Zhen Ye, Bernadeth Torres, Timothy Loe, Bin Li, and Jake Hocker.

Outside the lab, I'm grateful to Kian Kalhor for being such a great friend and for his great scientific advice – I'll always look back fondly on our walk-and-talks through the eucalyptus grove, making diagrams from fallen leaves and twigs. I'm also grateful to Siavash Raeisi

Dehkordi for being a wonderful collaborator and friend – I’m so glad we were able to innovate an incredible algorithm. Yuening Liu was always available to lend an ear and provided me with invaluable scientific and life advice (not to mention incredible dumplings), and I’m so glad our paths crossed in Pallav’s lab. My algorithm was made better by feedback from Eran Mukamel, Ming Hu, Arshad Desai, Xiaowei Zhuang, and Bogdan Bintu. I also want to thank my collaborators Ernest Lee, Bryan Sun, Pengtao Xie, Han Guo, Sai Somayajula, and Youwei Liang for our joint work on applications and future perspectives of spatial genome alignment. Pavel Pevzner and Melissa Gymrek designed incredible bioinformatics classes both rigorous and comprehensive, and I’m so much richer for their effort put into teaching – the algorithms they taught laid the foundations with which I was able to build the spatial genome aligner. Bruce Hamilton was kind enough to have me as part of the Genetics Training Program, and I especially loved his themed journal clubs (“The Right Stuff” was always my favorite). Importantly, I want to especially thank Brian Ross, who generously answered my cold email and then personally guided me through his spatial genome work, a work that is both breathtaking and prescient.

My journey has also been shaped by mentors beyond the lab. Sabrina Leslie and Claire Brown at McGill introduced me to single-molecule microscopy and cellular imaging, while Paul Cohen and his student Chan Hee Choi at Rockefeller guided me in pursuing an MD-PhD. I’m indebted to my MD-PhD directors Paul Insel and Neil Chi for bringing me to San Diego, my classmates in BISB who made grad school such a wonderful environment to learn in, as well as the entire BISB and UCSD MD-PhD office. Finally, a heartfelt thanks to my grandparents, who helped raise me and supported me from afar. Though separated by distance and the pandemic, I look forward to celebrating together soon.

Chapter 1 is, in full, a reformatted reprint of the material as it appears as “A spatial genome aligner for resolving chromatin architectures from multiplexed DNA FISH” in *Nature Biotechnology*, 2023. Bojing Blair Jia, Adam Jussila, Colin Kern, Quan Zhu and Bing Ren. The dissertation author was the primary investigator and author of this paper.

Chapter 2 is, in part, based on material currently being prepared for submission as “Principles of Spatial Multi-omics: Insights and Applications to Dermatology”. Bojing Blair Jia, Brian Sun, Ernest Lee, and Bing Ren. The dissertation author was the primary investigator and author of this paper.

VITA

- 2015 Bachelor of Science, McGill University
Honors, Biology (Quantitative Biology)
- 2024 Doctor of Philosophy, University of California San Diego
Bioinformatics & Systems Biology

PUBLICATIONS

Zu, S., Li, Y. E., Wang, K., Armand, E. J., Mamde, S., Amaral, M. L., Wang, Y., Chu, A., Xie, Y., Miller, M., Xu, J., Wang, Z., Zhang, K., **Jia, B.**, Hou, X., Lin, L., Yang, Q., Lee, S., Li, B., Kuan, S., Liu, H., Zhou, J., Pinto-Duarte, A., Lucero, J., Osteen, J., Nunn, M., Smith, K. A., Tasic, B., Yao, Z., Zeng, H., Wang, Z., Shang, J., Behrens, M. M., Ecker, J. R., Wang, A., Preissl, S. & Ren, B. Single-cell analysis of chromatin accessibility in the adult mouse brain. *Nature* 624, 378–389 (2023).

Liu, S., Zheng, P., Wang, C. Y., **Jia, B. B.**, Zemke, N. R., Ren, B. & Zhuang, X. Cell-type-specific 3D-genome organization and transcription regulation in the brain (2023).
doi:10.1101/2023.12.04.570024

Lee, L., Yu, H., **Jia, B. B.**, Jussila, A., Zhu, C., Chen, J., Xie, L., Hafner, A., Mishra, S., Wang, D. D., Strambio-De-Castillia, C., Boettiger, A., Ren, B., Li, Y. & Hu, M. SnapFISH: A computational pipeline to identify chromatin loops from multiplexed DNA Fish Data. *Nature Communications* 14, (2023).

Jia, B. B., Jussila, A., Kern, C., Zhu, Q. & Ren, B. A spatial genome aligner for resolving chromatin architectures from multiplexed DNA fish. *Nature Biotechnology* 41, 1004–1017 (2023).

Jia, B., Wee, T.-L. (Erika), Boudreau, C. G., Berard, D. J., Mallik, A., Juncker, D., Brown, C. M. & Leslie, S. R. Parallelized cytoindentation using convex micropatterned surfaces. *BioTechniques* 61, 73–82 (2016).

McFaul, C. M., Leith, J., **Jia, B.**, Michaud, F., Arsenault, A., Martin, A., Berard, D. & Leslie, S. Single-molecule microscopy using tunable nanoscale confinement. *SPIE Proceedings* (2013).
doi:10.1117/12.2021768

ABSTRACT OF THE DISSERTATION

Multiplexed genome imaging and analysis

by

Bojing Blair Jia

Doctor of Philosophy in Bioinformatics and Systems Biology

University of California San Diego, 2024

Professor Bing Ren, Chair
Professor Albert Hsiao, Co-Chair

Chromatin encapsulates not only genetic sequence information but also its structural organization. The physical layout of DNA is shaped by conserved chromosome folding patterns

that juxtapose distal functional elements of DNA in space, which can modulate gene expression. Directly visualizing chromosome architecture to study this grammar of chromosome folding has been a longstanding ambition in the field of epigenetics. An emergent imaging technology called multiplexed DNA-FISH has made it possible to localize the positions of thousands of genomic loci with nanometer precision inside single cells. Given a cell's ploidy, tracing chromosomes from this data initially assumed detecting a fixed number of bright signals commensurate with the cell's ploidy and connecting them in sequential genomic order. But during cell cycle chromosomes copies are subject to change; chromosomes are also subject to copy number variations such as amplifications or deletions; it is subjected further still to technical noise such as false positives appearing as true signal, or signal dropout. In my thesis, I sought to develop a different framework that accounts for various sources of noise that have gone unappreciated in multiplexed DNA imaging, to show that by embracing the complexity of noise we can in fact uncover new biology previously unseen.

First, I describe spatial genome alignment and its derivative polymer fiber karyotyping, methods for accurately resolving chromatin structures in a copy number agnostic fashion. We apply this method to multiplexed DNA-FISH data of mouse embryonic stem cells (mESC) and the mouse cerebral cortex. In dividing mESCs, we provide the first reconstructions of tightly intertwined sister chromatids decondensed in interphase – convoluted structures which previously could not be parsed by computer vision or by human eye. We go on to uncover unusual structures such as replicated homologs interacting in one single chromosome territory, suggestive of mitotic crossover. In the mouse cortex, we uncover tightly paired chromosomes in non-dividing neurons suggestive of sister chromatids in replication.

Second, I summarize the field of spatial multi-omics, describing where spatial DNA imaging fits within the technological landscape, as well as describing other emerging technologies. I illustrate the principles of two main methods for achieving spatial resolution: imaging *in situ* and sequencing-based “release-and-capture” techniques, and further provide a broad survey of different multiplexing strategies for indexing cellular contents. I compare and contrast the two main families of spatial technologies, delineating how the distinct advantages of each lends them to different experimental applications. Given the new capabilities we and others uncovered using spatial multi-omics, such as detecting copy number variations and visualizing histone modifications, that extend beyond transcriptomic profiling, I provide a future perspective on how spatial multi-omics and the many channels of information it captures could inform future clinical applications.

Chapter 1 : A SPATIAL GENOME ALIGNER FOR RESOLVING CHROMATIN ARCHITECTURES FROM MULTIPLEXED DNA FISH

1.1 Abstract

Multiplexed fluorescence in situ hybridization (FISH) is a widely used approach for analyzing 3D genome organization, but it is challenging to derive chromosomal conformations from noisy fluorescence signals, and tracing chromatin is not straightforward. Here we report a spatial genome aligner that parses true chromatin signal from noise by aligning signals to a DNA polymer model. Using genomic distances separating loci, our aligner estimates expected spatial distances and evaluates the physical probability observed signals belonging to these loci are connected, thereby tracing chromatin structures. We demonstrate that this spatial genome aligner can efficiently model chromosome architectures from DNA-FISH data across multiple scales and be used to predict chromosome ploidies de novo in interphase cells. Reprocessing of previous whole-genome chromosome tracing data with this method suggests the spatial aggregation of sister chromatids in S/G2 phase cells in asynchronous mouse embryonic stem cells and provides evidence for extranumerary chromosomes that remain tightly paired in post-mitotic neurons of the adult mouse cortex.

1.2 Introduction

Eukaryotic chromosomes undergo dramatic compaction and decompaction in the life cycle of a cell, and the dynamic chromosomal structure plays an integral role in a range of nuclear processes such as DNA replication, recombination, repair, and gene transcription¹⁻⁵. In interphase nuclei, different chromosomes generally occupy separate territories with limited intermingling⁶. Within each chromosomal territory, the chromatin fibers are organized into compartments and domains⁷⁻¹¹ driven in part by the ATP-dependent motor protein complex and loop extruder cohesin¹²⁻¹⁶. The complex chromatin structures enable juxtaposition of remote DNA in space and subsequent transcriptional activation of genes by distal enhancers^{5,17-19}. Disruption of chromatin structures underlies a score of pathologies ranging limb malformations, oncogenesis, and heart disease²⁰⁻²³. Delineating how chromatin fibers are folded in the nucleus is therefore of fundamental importance for study of gene regulation and other nuclear processes in health and disease.

Multiplexed DNA Fluorescence In-Situ Hybridization (M-DNA-FISH) is a widely used imaging technique for the study of chromatin structure in eukaryotic cells²⁴⁻⁴³. These technologies can identify tens to thousands of genomic loci in the nucleus, permitting potential folding patterns of chromosomes to be detected. They achieve this through the use of barcodes which, when decoded via serial hybridization of fluorescent probes, enable sequence-specific identification, and thus distinction, of genomic loci; in diffraction-limited imaging the chromatin fiber connecting them is not visualized and must be inferred. The inference of physical connection between two discrete signals is the most salient problem facing chromatin imaging to date. Early efforts to multiplex DNA-FISH often found one, three or four fluorescent signals

emanating from one genomic region in a diploid cell line expected to produce two signals^{35,36}. Biological copy number variation, chromosomal intermingling as well as poor probe hybridization have been acknowledged to explain missing signals^{31,32,35}, sister chromatids and aneuploidy as well as imaging noise have been acknowledged to explain extra signals^{31,34,35}. If both noise and biological variation can explain any observed scenario, chromatin fibers cannot be naïvely traced by connecting the first immediate spot. In fact, this uncertainty around imaging has led some to forgo tracing altogether and instead tabulate proximal pairs of imaged loci for bulk analysis³¹. When noise appears indistinguishable from true imaged genomic loci, and biological variation at the single cell level confounds expectation, reconstruction of chromatin fibers remains an intractable computational problem.

In the current benchmark for chromatin tracing, the tracing problem is simplified with assumptions about copy number and with emphasis on the optical quality of detected signals. Explicitly, an expectation-maximization algorithm (E-M) is first tasked to find k chromatin fibers corresponding to an k -ploid cell^{28,30}. Repeated k -times per cell, candidate fluorescence spots corresponding to a genomic region are scored based on signal intensity, proximity to a moving average of downstream selected spots as well as upstream selected spots, and proximity to a chromosome center (ie. the aggregate of many fluorescent loci) determined by k -means clustering^{28,30}. Implicit in this approach are two strong assumptions: that the brightness is a measure of detection confidence and that the copy number of a DNA segment is fixed and known beforehand. However, background fluorescence, non-specific probe binding, and even hot-pixels can frequently emit similarly intense focal signals indistinguishable from the true signal. Additionally, looking for a fixed number of chromosomes may fail to capture true

biological copy number variations and aneuploidy. To disambiguate these scenarios, we need a new framework for chromatin tracing that leverages yet unused information, improving on the above heuristics.

Here we present a spatial genome aligner that considers the above challenges. We reason that while the shape of chromatin fiber is highly variable, it is subject to spatial constraints dictated by polymer physics^{44,45}. In addition to considering optical quality of signals, our algorithm aims to select true fluorescence spots corresponding to a DNA locus from a number of candidates by picking the one that best conforms to a reference polymer model of chromatin. Briefly, these restrictions are the genomic distances between two discrete sequence-specific labeled loci, which should be proportional to the square of its spatial separation. We use the Gaussian chain polymer model to estimate an expected spatial distance given a genomic distance, and compare the observed spatial distance in imaging to this estimated spatial distance as a test of physical likelihood⁴⁴. We attempted to evaluate the accuracy of the spatial genome aligner by comparing pairwise distances discovered by tracing against pairwise contact frequencies discovered by Hi-C. We find that our spatial genome aligner can recapitulate patterns of chromatin organization found in Hi-C at multiple genomic length scales (5 kb, 25 kb, and 1 Mb). We note circumstances of chromatin folding may not completely adhere to this Gaussian chain model, and chromatin structure may be influenced by factors including but not limited to chromatin remodelers, gene expression, proximity to nuclear structures and certain chromatin loops that in part or in sum evade prediction by the Gaussian chain, despite our model's best effort to permit flexibility. The genomic footprint of the oligos tiling at a target site, as well as transcriptional state of the target site, may also influence measurement. Diffraction

limited imaging lacks the sensitivity to resolve single oligos and “merges” the cumulative signals of multiple oligos. Image smoothing may inadvertently fuse signals belonging to separate chromatin fibers; conversely, sample prep and large genomic footprint may potentially fragment one signal. We further note that extrapolation of persistence length of DNA to chromatin scale remain challenging; we removed one free variable by fixing the persistence length ($l_p = 150$ bp) and fitting a variable genomic-to-spatial distance scale (τ). Interestingly, we find our spatial genome aligner uncovers more chromatin fibers than previously reported in published datasets, and infer these extra fibers may in fact be sister chromatids. We show that each pair of putative sister chromatids usually resides in a spatially separate chromosome territory, but in $\sim 2\%$ of replicating cells both pairs of sister chromatids coalesce to interact in one convergent territory. We go on to apply our spatial genome alignment to previous chromatin tracing data generated from mouse cortical excitatory neurons, where we uncover patterns of spatial organization of extranumerary chromosomes inside the nucleus.

1.3 Results

Spatial genome alignment

Chromosomes are linear, flexible polymers that take on convoluted structures inside the nucleus. One simple but robust model for the spatial configuration of flexible polymers is a Gaussian chain⁴⁴. In this model, the polymer is represented as a chain of successive monomers, linked by bonds of approximately constant length b . Each successive monomer is allowed to freely rotate with respect to each other. Transitioning from one monomer to another along the polymer chain is to take one step in a three-dimensional random walk. For any two monomers i and j on this chain, the probability they are separated by a vector distance R_{ij} follows a Gaussian distribution (hence, Gaussian chain)⁴⁴:

$$P(R_{ij}, n) = \frac{1}{\left(\frac{2\pi}{3}nb^2\right)^{3/2}} e^{-\left(\frac{3R_{ij}^2}{2nb^2}\right)} \quad (\text{Eq. 1})$$

where n is the number of bonds, each of length b , separating two monomers i and j .

For a chain with N bonds, the likelihood of the entire chain (also known as the conformational distribution function; CDF) is the product of all bond probabilities on the chain⁴⁴:

$$\Psi(\{R_N\}) = \prod_{n=1}^N \frac{1}{\left(\frac{2\pi}{3}b^2\right)^{3/2}} e^{-\left(\frac{3R_n^2}{2b^2}\right)} \quad (\text{Eq. 2})$$

In multiplexed DNA-FISH experiments, we recognize entire chromosomes are labeled at discrete positions, analogous to discrete monomers on a Gaussian chain. Furthermore, these

discrete loci are interspaced by regular genomic intervals (eg. 1 Mb), akin to the constant bond length b that separate monomers on the model chain. We hypothesized that at large genomic length scales, we can model DNA conformation with a Gaussian chain in which the bond length b can be estimated from the genomic distance separating two loci^{44,45}:

$$nb^2 = 2l_p\tau L \quad (\text{Eq. 3})$$

where l_p is the persistence length of DNA, τ is genomic-to-spatial distance conversion factor (nanometers per base pair), and L is the genomic distance separating two loci, for $l_p \ll \tau L$. τL together represents the contour length along the DNA polymer separating two loci. τ is fit separately for each chromosome, as previous studies show their length scales differ³¹.

In a setting where multiple signals are detected for *each* of two genomic loci, it is ambiguous which pair of signals lies on the same chromatin fiber. This Gaussian chain model allows us to express the probability two discrete loci imaged are physically connected as a function of both its observed spatial distance and its expected spatial distance. Here, the expected spatial distance is derived from the known genomic distance separating two loci on a reference genome. In taking one step along the chromatin fiber, we can select or omit a fluorescence signal by identifying (if any) a pair of signals with unique sequence identities whose observed spatial separation is ideally congruent with its expected spatial separation. In tracing the entire chromatin fiber, the most likely polymer among imaged loci is one where the collective segment lengths along the chromatin fiber best aligns with its expected segment lengths. Our optimization

objective is therefore to find the sequence of spatially resolved genomic loci that maximizes the likelihood, or CDF, of the polymer traced.

Algorithmically, we first abstract imaged chromatin fluorescent signals as nodes in a directed acyclic graph (DAG). The topological order of nodes is determined by the order of loci on the reference genome (Fig. 1). We connect each node to the adjacent nodes on the linear genome, with each directed edge emulating a polymer segment. For each directed edge, we leverage known genomic distances separating the two imaged loci to estimate an expected spatial distance. We utilize both the expected spatial distance and observed spatial distance between the two imaged loci to calculate a bond probability, assigned as the edge weight. Traversing the graph from beginning to end is to find a potential chromatin fiber. Keeping track and multiplying the edge weights traversed, the score of one path reflects its physical likelihood (CDF).

Operationally, we transform the edge probabilities with a negative logarithm function into positive edge weights, such that the additive sum of edge weights reflects the polymer CDF. With this transformation, our optimization objective of maximizing likelihood is reframed as minimizing the sum of negative logarithm transformation of edge probabilities – in other words, we wish to find the shortest path through our graph representation of our polymer. Using dynamic programming, we find the shortest path through the adjacency matrix of our polymer graph not unlike traditional sequence alignment^{46,47}. To account for false positives and false negative imaged spots, all valid paths are explored with the option to “skip” a node permitted by a gap penalty. Since DNA loci from a chromosome must lie on the same chromatin fiber which

cannot branch, finding the shortest path is to find the most probable polymer without physical discontinuity discoverable from data⁴⁸.

We first benchmark our spatial genome aligner against the chromatin tracing strategy that connects adjacent genomic loci by converting tabulated distances into an ensemble contact frequency. We analyzed previously published DNA seqFISH+ genome-wide chromatin tracing on mouse embryonic stem cells (mESC), tracing every mouse chromosome at ~ 1 Mb resolution across 1160 single cells³¹. Whereas in published work, detected loci were binned and tabulated to convert distances into an ensemble contact frequency, our spatial genome aligner resolves single-molecule chromatin fibers at single-cell resolution across multiple genome scales. Indeed, our spatial genome aligner traces points whose median distance matrices are commensurate with KR-normalized bulk Hi-C contacts (1 Mb: Spearman corr = -0.9 ± 0.04 ; 25 kb: Spearman corr = -0.85 ± 0.04). We found our spatial genome aligner can resolve large chromatin compartments imaged at 1 Mb intervals (Supplemental Fig. 1A, B) as well as finer, single-cell chromatin domains imaged at 25 kb intervals (Fig. 2A, B). At 25-kb resolution, local chromatin structure is often nonlinearly organized into topologically-associating-domain-like patterns, which may have variable boundaries, variable domain numbers, and lifetimes^{26-34,39-43}. Because our polymer model is a freely rotating chain of flexible segments, we found it accommodates such abrupt changes in local topology not easily captured when tabulated in an ensemble fashion. (Fig 2A, B, C).

To evaluate the performance of spatial genome alignment on finer genomic length scales, and on data from other chromatin imaging protocols, we performed spatial genome alignment on

multiplexed DNA-FISH data of the Sox2 locus imaged at 5-kb resolution⁴⁹. In our previous work, we adapted a protocol based on sequential DNA-FISH²⁸ to label a 210-kb genomic region on mouse chr3, spanning both the Sox2 gene locus in the F123 hybrid mESC line and its super-enhancer 110 kb downstream. By sequentially imaging these loci and tracing the chromatin, we were previously able to visualize promoter-enhancer contacts corralled within a TAD. When we applied our spatial genome aligner to this fine 5-kb resolution chromatin imaging experiment, we found our spatial genome aligner can indeed recapitulate the TAD found at this region (Supplemental Fig. 2), faithfully capturing known promoter-enhancer interactions. We note the positional uncertainties, determined using Gaussian fitting for fluorophores, can vary across imaging platforms. This uncertainty is ~ 50 nm when drift and chromatic aberration are appropriately corrected³⁰.

We additionally benchmarked our spatial genome aligner with a published chromatin tracing algorithm⁴⁹ (Supplemental Fig. 2). Previously, chromatin tracing on multiplexed DNA-FISH emphasized the optical quality of a fluorescence spot, a metric incorporating (a) brightness (b) proximity to a chromosome center and (c) relative agreement to a moving average of preceding and subsequent spots. An expectation-maximization (E-M) procedure then sequentially selected one spot with the highest quality for each chromatin locus, while iteratively updating its quality scores. In contrast, our spatial genome aligner introduces another metric into spot selection – physical constraints dictated by polymer physics – as a decision criterion for selecting spots. Compared to previously published E-M spot selection algorithms (Spearman corr = -0.73), our spatial genome aligner produces a median distance matrix which achieves similar correlation with respect to Hi-C (Spearman corr = -0.76). The difference between the two

methods become larger when we tabulate distances into contacts against a range of distance thresholds, with spatial genome alignment achieving a Spearman correlation of -0.714 at a 150-nm threshold compared to the E-M algorithm's -0.651 (Supplemental Fig. 2F-H). Notably, our spatial genome aligner performs a global optimization that incorporates the relative positioning of all imaged loci rather than a local moving average. Considering all imaged loci may be more robust against noise compared to a moving average, in the event a starting point has especially low signal-to-noise and unduly affects the downstream moving average. Additionally, our spatial genome aligner traces the most likely polymer fibers in linearithmic time using dynamic programming, compared to the heuristic E-M approach which relies on iterative convergence. Taken together, we believe that the genomic distance separating imaged loci has the potential to disambiguate spot selection and that our spatial genome aligner can be used to model chromatin fibers at multiple lengths scales, on multiple datasets, and on different multiplexed FISH imaging modalities.

Polymer Fiber Karyotyping

A nucleus may have multiple copies of a chromosome. Finding all copies has traditionally relied on identifying compact clusters of imaged loci, aggregating by chromatin fiber. A k -means approach of clustering assumes the ploidy of a cell is known beforehand^{28,30}, and this approach is unable to accommodate copy number variations. For $k=2$ and ploidy $n=1$, k -means may inadvertently look for a non-existent second “phantom” chromosome. Conversely, for $k=2$ and ploidy $n=3$, k -means would fail to detect an entire chromosome altogether. A copy number agnostic approach of clustering, such as DBSCAN, relies on density of detected loci^{31,32}; however, its density neighborhood parameter is difficult to tune. A large density neighborhood

may inadvertently aggregate two spatially separable chromosomes as a single dense cluster, misassigning two separate homologs as one. A small density neighborhood may fracture an intact chromosome into separate partitions.

In contrast, our spatial genome aligner provides a density or ploidy independent framework for identifying chromatin fibers. We provide all detected spatial coordinates of a chromosome and a reference genome to our spatial genome aligner, tasking it to extend (if at all possible) the most likely path from chromosome start to end. Since the path length (CDF) of a putative polymer reflects the physical likelihood of a polymer, we reasoned that the copy numbers of chromosomes inside interphase cells can be obtained simply by counting all physically likely polymer fibers. First, we set a likelihood threshold by scrambling a simulated polymer model of a reference genome such that the observed spatial distances between genes no longer abides by the genomic intervals that separate them (Fig. 3B). Next, we iteratively apply spatial genome alignment, extending polymer paths from putative seeds and subtracting nodes visited by the shortest path before searching for next shortest path, until no physically likely polymer path can be discovered (Fig. 3A). In this manner, we produce orthogonal sets of coordinates belonging to contiguous chromatin fibers with likelihood scores below our threshold. We call this process polymer fiber karyotyping (PFK).

Using chromatin tracing data spanning the mouse genome at ~ 1 Mb intervals, we performed spatial genome alignment to discover all possible chromatin fibers in the mouse ES cells. Intuitively, a diploid cell should have half as many chromatin fibers as a tetraploid cell; we reasoned this should also reflect in the total number of loci detected in a cell. Comparing the total

detected fluorescence signals per chromosome in a cell to its assigned ploidy determined by PFK, we notice a linear relationship. Every incremental increase in ploidy is accompanied by a stepwise, multiplicative increase in the total number of detected loci (Fig. 3C). Building on this, we compare the agreement of copy numbers assigned by each chromosome. Hierarchical clustering of copy numbers assigned by each chromosome across 1160 cells show three distinct clusters of cells whose copy numbers are homogeneously congruent for all 19 autosomal chromosomes. Namely, we see cells proportionally falling into a 6:2:2 distribution of 2N:3N:4N: cells respectively, matching a replicative profile of highly dividing mES cells (Fig. 3D). Treating each chromosome as a separate agent for karyotyping, we quantified the copy number agreement between different chromosomes using Cohen's kappa test. Pairwise comparisons of each chromosome against another show demonstrably significant agreement ($\text{Kappa} \geq 0.3$), save chromosome X (Fig. 3E). Although the spatial genome aligner had every opportunity to find as many fibers for chromosome X as it did for other autosomal chromosomes (2.618 ± 1.114 copies; mean, standard error), it found fewer copies (1.699 ± 0.669 copies) of chromosome X in this male cell line. Some heterogeneity among copy numbers between cells exists in part due to inherent imaging labeling efficiency, detection efficiency, cell segmentation, as well as the stringent spatial fitting criteria of the spatial genome alignment. This gave us additional confidence that our spatial genome aligner produces accurate cell karyotypes without supervision, and that it can discriminate copy numbers in interphase where even the human eye cannot.

To assess our polymer fiber karyotyping's accuracy in counting interphase chromosomes, we simulated 1000 aneuploid cells. In each cell, we randomized copy numbers between 0 – 4 for

each of the 20 mouse chromosomes, such that the copy number of one chromosome cannot inform the copy number of another chromosome in the same cell (Supplemental Fig. 3A). To these simulated chromosomes for which we know the true localizations of loci, we added additional layers of noise including extra stray localizations (False Positive) and signal dropout (False Negative) (Supplemental Fig. 3A). Surveying a grid of 49 different combinations of false positive and false negative rates, we evaluated our polymer fiber karyotyping accuracy in the analysis of chromosome counting (Supplemental Fig. 4). We also evaluated the accuracy of spatial genome alignment in the classification of signal selection amid noise on these simulated aneuploid cells (Supplemental Fig. 3).

At the alignment level, we honed in on 50% False Negative Rate (FNR) and 0% False Positive Rate (FPR) – a condition most representative of detection efficiency and off-targets reported in DNA seqFISH+ experiments^{31,32}. At this noise condition, we achieve an average true coverage rate (True Positive $\rightarrow$ True Positive) of $27 \pm 13.2\%$ out of a maximum 50% true positive loci (Supplemental Fig. 3D). In fact, even in the setting of 0% FNR and 0% FPR (perfect information) our aligner captures $88 \pm 13.2\%$ true coverage rate out of a maximum 100% true positive loci (Supplemental Fig. 3D). This strongly suggests our aligner's requirement to fit spatial distances between loci to be congruent with its genomic distance is extremely stringent. Indeed, the most dominant error made by our aligner is the omission of true positives or Type III Error (True Positive $\rightarrow$ False Negative), across all chromosomes (Supplemental Fig. 3C) and across all noise conditions (Supplemental Fig. 3D). Explicitly, at 50% FNR 0% FPR our aligner averages a Type III (True Positive $\rightarrow$ False Negative) error of $18 \pm 13.2\%$ across all chromosomes (Supplemental Fig. 3D). In comparison, we report an average $6 \pm 6\%$ Type I error

rate (True Negative $\rightarrow$ False Positive), where we select a false positive spot in the absence of true signal (Supplemental Fig. 3D). This again suggests our aligner is making highly specific judgements in spot selection. The lowest source of error is Type II Error (True Positive $\rightarrow$ False Positive), where the selection of a false positives in the presence of true signal hovers around $5 \pm 4.8\%$ at 50% (Supplemental Fig. 3D). Since we set an upper limit for number of loci skipped, the omission of true positive loci may accumulate for especially high signal dropout conditions, leading to early termination of alignment and potentially missed chromosomes.

At the karyotyping level, we found our routine achieves near perfect precision-recall in the setting of perfect information (Supplemental Fig. 4A) but begins to miss chromosomes at increasing dropout rates (Supplemental Fig. 4B). We wondered if our aligner ever “alternates” between loci, leaving behind enough “residual points” that lead to an overcounting of chromosomes on second iteration. Instead, we found our algorithm may undercount but almost never overcount, across a spectrum of noise conditions (Supplemental Fig. 4A-C). Even when we added extra noise signals, our algorithm does not mistake these stray localizations for additional chromatin fibers (Supplemental Fig. 4C). Curiously, we found that chromosome detection efficiency slightly increases with added noise (Supplemental Fig. 4C). For instance, at 50% FNR 0% FPR, our polymer fiber karyotyping identifies 3243 of 3927 2N chromosomes. At 50% FNR 20% FPR, our routine identifies 3472 of 3927 2N chromosomes. We attribute this improved chromosome detection sensitivity to a “placeholder” effect that stems from the design of our aligner. Just as we have an upper limit to total number of loci skipped per chromatin fiber, our aligner also has an upper limit to the number of loci skipped at any given locus. Significant dropout without any noise may result in consecutive missing signals stretching beyond the

furthest allowable skip. Yet in the setting of signal dropout and added noise, our algorithm may potentially align to a placeholder “noise” spot, taking the penalty of an incongruent spatial segment in order to complete the alignment from chromosome start to end.

Aggregation of homologous chromosomes in tetraploid mESC cells

Of the putative 4N chromosomes predicted by our spatial genome aligner, we asked if these are polyploid cells with four separable chromosomes before replication, or diploid cells with two pairs of sister chromatids after replication (Fig. 4A, B, C). As sister chromatids are shown to be tightly paired in a parallel fashion⁵⁰⁻⁵³, we reasoned that if two chromatin fibers reside in the same spatial neighborhood, they are likely sister chromatids of the same homolog. We performed density-based clustering to assign fibers of every ploidy to homologs. Under a set density parameter, the majority of diploid cells had two spatially resolvable fibers singularly residing in different territories (Fig. 4B). Notably, under the same density parameter the majority of tetraploid cells also had not four spatially resolvable fibers but rather two clusters of paired fibers that cannot be parsed by eye or by known clustering algorithms (Fig. 4B).

To test if these paired fibers are indeed sister chromatids, we examined the *trans*-fiber loci distances relative to the *cis*-fiber loci distances. In agreement with published sister-chromatid sensitive Hi-C on *Drosophila* and human cell lines^{52,53}, paired fibers in mouse ES cells resolved by our spatial genome aligner are spatially coupled. Explicitly, a given locus of one sister chromatid is followed by the same locus on its attendant sister chromatid, faithfully “shadowing” each other (eg. Chr 1: $\mu = 1606.8$ nm separation; 95% CI [1540.3, 1734.5]) (Fig. 4D). Congruent with previously published work, the spatial distance between *cis*-fiber

interactions is closer for smaller genomic distances but which converges with *trans*-fiber interactions above 10 Mb (Fig. 4E). Given the parallel nature of pairing and recapitulation of sister-chromatid interactions found by sister chromatid sensitive Hi-C, we presumed the tetraploid cells were in fact replicated diploid cells exhibiting paired sister chromatids.

Because these sister chromatids are tightly coupled, it is plausible loci belonging to one sister are inadvertently selected by its attendant sister and vice versa. To assess this misselection event, we revisited all putative sister chromatid pairs of mESC chr 1. Examining a sliding window of three loci, we allowed sisters to exchange their selected spatial positions for these three loci. Next, we re-evaluated the resultant bond probabilities between three upstream loci to the three downstream exchanged loci (Supplemental Fig. 5A). If the new bond probabilities are more probable than the original on both sister chromatid fibers, we call this a misselection event (Supplemental Fig. 5B). A sequence of consecutive misselections of any length is called a cross-over event (Supplemental Fig. 5D). We found that our spatial genome aligner is susceptible to these errors when tracing paired sister chromatids, and has an average locus misselection rate of $5.06 \pm 1.84\%$. The mean cross-over per fiber is 5.08 ± 3.74 (Supplemental Fig. 5D). One source of misselections lies in the sparsity of data (Supplemental Fig. 6A, C). Rarely are two signals simultaneously detected for a locus (chr 1: 7.34%) and present to be selected by both the main sister and its attendant (Supplemental Fig. 6A, C). In fact, for chr 1 34.5% of all cases have zero signals are detected for a given locus. In the absence of robust signal, our polymer fiber karyotyping routine is greedy. Whichever sister is discovered first is incentivized to select as many detected loci as it sees fit, at the cost of potentially selecting signal that belong to the other sister. A signal cannot be re-selected by subsequent sisters. When only one signal is detected for

a given locus, we therefore see a disproportionate imbalance of the single locus signal assigned to the main sister traced first (34%) compared to the signal assigned to the attendant sister traced second (19.6%). In a controlled simulation of paired sister chromatid fibers, we find that the predominant cross-over error our aligner makes is a cross-over in the absence of signal on the present sister chromatid but presence of signal on the attendant sister (Type I; True Negative → False Positive; Supplemental Fig. 7D). In the presence of signal on both sister chromatids, our aligner rarely makes cross-over errors (Type II; True Positive → False Positive). These results hold true for a range of sister chromatid pairing strengths (Supplemental Fig. 7B, D). Future improvements in experimental detection efficiency, sister-chromatid specific labeling, and algorithmic flexibility permitting revisiting signals with higher intensity (indicative of coincident signals) may decrease this misselection error and improve sister chromatid tracing coverage.

Canonically, homologous chromosomes are divorced from each other in the nucleus and are widely acknowledged to reside in separate territories⁶. Yet, of the 207 cells tetraploid for chr 1 we find two predominant patterns: three-quarters (146/207 cells) bearing two spatially separable clusters presumed to be different homologs (sep-hom), and strangely, a quarter (49/207 cells) with all four fibers are coalesced (compact) (Fig. 4C). We note a marginal population of cells (12/207 cells) with three or more separable structures corresponding to separated sister chromatids (sep-sis). In case our clustering density parameter had inadvertently grouped two separable homologs together, we visually inspected each putative compact 4N chr 1. To our surprise, we find a significant proportion of the compact state (31/49 candidates), cumulatively 2.67% of total cell population, has four chromatin fibers spatially intermingling and which cannot be separated by eye.

Why would the newly replicated homologous chromosomes coalesce? While homologous chromosome generally lie in separate chromosome territories^{54,55}, some species such as *Drosophila* exhibit extensive homologous pairing in somatic cells⁵⁶. It is thought that communication between homologs, across a number of species⁵⁵⁻⁵⁷, underlies gene regulation⁵⁸, chromosome counting⁵⁹, DNA repair (with resultant loss-of-heterozygosity)⁶⁰⁻⁶³, and even chromosome topology⁶⁴. The *trans*-homologous interactions identified in mammalian cells under physiological conditions have been sparse^{65,66}. However, there have been reports of such events in the setting of human diseases^{67,68}, suggesting this rarity of homologous pairing may be either species specific or due to mechanisms in mammals that deliberately prevent such interactions⁵⁵. Here we suggest the possibility that multiplexed DNA-FISH and chromatin tracing may help better understand this potential for *trans*-homologous interactions.

To understand our observed compact chromosome territories, we began by parsing which two fibers belong to one homolog within the compact structure. Spatial proximity prohibits clustering from separating homologs and assigning pairs of fibers as sisters. Since true sister pairings should involve two fibers shadowing each other, we reasoned sisters can be assigned by proximity of a given locus on two fibers. There are two natural assignments: grouping the closest pairs by the starting locus (SA; mouse centromere, which, in mouse, is located at the end of each chromosome), and grouping the closest pairs by the end locus (EA; mouse telomere, noncentromeric mouse telomere) (Fig. 5A). While both are telomeres, we use centromeric and telomeric to denote locations – namely start and end locus of the chromosomes, respectively. We note spatial proximity may cause our spatial genome aligner to inadvertently select spots

belonging to other fibers. Therefore, we explored the centromere-centromere distances as well as telomere-telomere distances of the two remaining permutations. Specifically, these permutations correspond to the best possible alternate pairing (alt1) as well as the remaining pairing (alt2), ranked in this order.

When replicated homologs reside in separate chromosome territories, we find that the telomeres of putative sister chromatids grouped by their centromere (SA) are likely coupled (Fig 5D, left). The mean distance separating telomeres of SA sisters (mean = 3857.5 nm; 90% CI [3451.5, 4263.5]) is smaller than the distance between its centromere-telomere not known to interact (mean = 4474.9 nm; 90% CI [4295.8, 4654.1]). In contrast, the next best alternative pairing has a telomere separation (mean = 5767.4 nm; 90% CI [5342.2, 6192.6]), larger than two non-interacting loci. In the same manner, the centromeres of putative sisters grouped by their telomere (EA) are also tightly coupled. All other alternate pairings exhibit a spatial separation above that of two non-interacting loci.

When replicated homologs coalesce, we find that putative sisters grouped by their centromere may lose pairing at their telomeres. The mean distance separating telomeres of SA sisters is similar (μ = 3286.5 nm; 90% CI [2835.9, 3737.1]), compared to the distance between two non-interacting loci (μ = 4310.9 nm; 90% CI [4110.2, 4511.7]). Curiously, the next best alternative pairing has a telomere separation (μ = 2544.2 nm; 90% CI [2131.2, 2597.1]) closer than the SA assigned telomere distance. Should this be due to misassignment, then all three pairing scenarios should share a uniformly unpaired distance distribution with mean distances above coupling. Yet, there almost always exists an alternate pairing between putative homologs

that theoretically should not interact. The same analysis on EA sisters confers an ambiguous result, likely due to this loss of pairing.

The tendency for *cis*-homolog coupling decreases moving away from the centromere, resulting in a “flare-up” of putative *trans*-homolog interactions near the telomere of chromosomes (Fig. 5B). Since DNA seqFISH+ labels discrete genomic loci, we are not able to visualize the contiguous polymer physically linking imaged loci. Additionally, DNA seqFISH+ probes do not discriminate homologs. Our re-analysis of misselection events (Supplemental Fig. 5) and analysis of simulated sister chromatid pairs (Supplemental Fig. 7) suggest our aligner may inadvertently cross-over between loci that belong to other chromatin fibers, albeit at lower rates than true coverage. We are therefore unable to definitively show all spots traced by our aligner lie on the same contiguous fiber, and that our assignments by loci proximity are always sisters of the same homolog. However, our spatial alignment analysis in bulk suggests a possible loss of sister pairing towards the telomeric end and increased *trans*-homologous interaction within this compact 4N state. We ordered the cells along the same pseudotime axis as determined using previously published cell cycle markers (H4K20me1, H4K16ac) (Fig. 5C, D, E). We found that this compact 4N state is scattered throughout interphase leading up to S phase. Additionally, this compact 4N state is rarely synchronized across multiple chromosomes, and appears to occur stochastically. This suggests *trans*-homologous interactions between replicated chromosomes are likely uncoordinated and not initiated by a particular cell-cycle checkpoint.

Extranumerary paired chromosomes in mouse cortical neurons

Extranumerary chromosomes and more broadly aneuploidy have previously been reported in the brain, by both FISH and single-cell sequencing⁶⁹⁻⁷³. This copy number variation is thought to underlie a functional diversity adequately supporting neural complexity⁷⁰. In DNA seqFISH+ imaging of mouse cortex, we sought to evaluate such copy number variations with polymer fiber karyotyping. Because cryosectioning may cut through nuclei, leaving some chromosomes (in whole or in part) out of view, we analyzed 701 fully segmented mouse neurons³², all lying in the center z-sections of female mouse cortexes derived from 3 biological replicates. Of these intact nuclei, we honed in on excitatory neurons – the predominant cell type in this dataset (n = 458/701 neurons).

Our karyotyping routine reports $58.05 \pm 6.38\%$ (mean, standard deviation) of a given chromosome as 2N. It classifies $13.82 \pm 3.03\%$ of a given chromosome as 1N, and 19.68 ± 5.08 of a given chromosome as 3N (Fig. 6A). Unlike cell line imaging, karyotyping on multiplexed-FISH imaging of tissue is challenged by (a) poor probe labeling (b) higher background fluorescence (c) difficult cell segmentation and (d) aforementioned cryosectioning artefact that may collectively contribute to missed chromosomes (Fig. 6A). Additionally, our extensive analysis of polymer fiber karyotyping on simulated aneuploid cells suggest that, on data with low imaging detection efficiency, our aligner may potentially undercount chromosomes (Supplemental Fig. 4). Our aligner's criteria of fitting the spatial separation between each pair of loci to be congruent with its genomic distance is extremely specific in filtering out noise, sometimes at the cost of lost sensitivity. It currently has a set threshold for total number of skipped loci that qualify a valid polymer. It also has a restriction in the number of loci can be

skipped at a step in spatial genome alignment. Either successive dropout in a region of a genome (despite all other loci fully detected) or cumulative dropout across the chromosome may terminate our alignment, potentially missing chromosome. In this mouse brain imaging dataset, the detection efficiency hovers around $47.3 \pm 16.2\%$ (2326 ± 795 spots; 2460 labeled loci) for 1 Mb tracing³². At this sparsity, we therefore believe the number of chromosomes classified as 1N is at least partially contributed by a tendency to undercount in our algorithm. The potential to undercount, however, led us to take a closer look at extranumerary chromosomes (3N) in cortical neurons.

Due to data sparsity, many loci are absent on putative extranumerary chromosomes. To ascertain these are indeed three separate chromatin fibers and not two fibers parsed into three, we filtered all putative 3N chromosomes seeking fibers that are routed through three separate detections per locus for at least three different loci. We note that each detection of a locus under DNA seqFISH+ imaging is not a one-time detection, but in fact a temporal barcode wherein a locus is detected at least four out of five times, at the correct temporal sequence, while satisfying error correction. We also note that in our simulated aneuploid cells and in simulated sister-chromatid pairs, despite significant dropout and simulated noise, our aligner does not split one chromatin fiber into two (Supplemental Fig. 7F). Even still, we find that $8.72 \pm 4.18\%$ of each chromosome are predicted to be 3N under this more stringent criterion. We validated our karyotyping results by analyzing haplotype resolved single-cell Dip-C sequencing on mouse cortical neurons classified as excitatory⁷⁴ (Supplemental Fig. 8A). By counting total reads, as well as inspecting the relative fold change of reads assigned to the maternal vs. paternal haplotype, the majority of sequenced neurons have balanced haplotyped reads ($|\log_2|$ fold change

< 0.9). Interestingly, $10.82 \pm 1.31\%$ of every chromosome have twice ($|\log_2|$ fold change ≥ 0.9) as many reads of one haplotype as the other (Supplemental Fig. 8A, B). Not only may this reflect copy number variations prevalent in cortical neurons, it also suggests the extranumerary chromosome may potentially be contributed by one haplotype.

Because single-cell sequencing affords relative copy numbers of reads and fails to capture the nuclear organization of the aneuploid cells, we inspected the spatial organization of these copy number variations. Looking at chr X, where the inactive chromosome is distinguished from the active by RNA imaging of *Xist*, density-based clustering reveals that excitatory neurons with three chr X have predominantly two chromosome territories (Fig. 6B). One chromatin fiber is standalone, while the two remaining fibers are constituents of the same territory (Fig. 6C). Of the doubly occupied chromosome territories, two-thirds are devoid of any *Xist* signal, suggesting a preference for active chr X ($p = 0.049369$, one-sided binomial test; Fig. 6F). Despite different gene dosage, we detect no significant gene expression relative to cells with two detected copies of the X chromosome, irrespective of chr X activation status (Fig. 6E). The majority of labeled genes show no significant change relative to elevated gene dosage (Supplemental Fig 9).

Between fibers in the doubly-occupied chromosome territory, the active chr X fibers show some degree of pairing (Spearman $r = 0.26$) at a locus-to-locus level with its attendant fiber (Fig. 6D, G). This manifests as a strong diagonal in the *trans*-fiber pairwise distance matrix, as well as a spatial distance distribution mimicking that of *cis*-fiber pairwise distances. The same cannot be said of the inactive chr X, whose double constituents bear little resemblance (Spearman $r = 0.046$).

1.4 Discussion

Here we present a spatial genome aligner for multiplexed DNA-FISH data. We show that this framework resolves chromatin fibers from discretely labeled positions of genomic loci, amid noise and signal dropout. In our spatial genome aligner, each observed locus' spatial position is checked against a reference model of a polymer chain. This reference model, a Gaussian chain abstracting connections between imaged loci as bond probabilities, dictate that even a highly variable structure as DNA follows predictable patterns of distance separation between loci. Our model utilizes a fixed scaling factor to roughly estimate spatial distance from genomic distance. We recognize that DNA compaction is known to change throughout the genome, and our fixed scaling factor does not recapitulate the exact spatial distance separating each locus imaged. However, we hypothesized this estimated distance can inform soft-decisions in the selection of imaged DNA loci, in addition to considering optical qualities of signals detected, to trace chromatin structures. We found this framework captures chromatin compartments and domains also found in Hi-C, on multiple lengths scales and across different chromosomes. Future frameworks may fuse other channels of information, such as A/B compartments, chromatin modifications, gene expression levels, and association with nuclear structures to estimate different distance parameters throughout the genome during chromatin tracing.

Our algorithm falls into an early lineage of spatial genome aligners, including work by Ross et al. that abstracts connections between loci as polymer segments and whose edge weights are proportional to physical likelihood⁴⁵. Chiefly, whereas a reference polymer structure can reconcile each individual locus' most likely spatial position using the forwards-backwards algorithm^{45,48}, we demonstrate the utility of dynamic programming to find the most plausible

sequence of spatial positions discoverable⁴⁸. In other words, finding the shortest path in our graph representation is to find the most physically-likely polymer without any physical discontinuity. Our aligner builds connections between loci in the sequential order they appear on the reference genome. Currently, this prevents discovery of certain structural variants, such as inversions, translocations, and duplications. However, we believe finding the most likely contiguous polymer may be instrumental to uncovering copy number variations at the single-cell level. Through iterative subtraction of shortest paths to find all valid polymers, our data suggest a capacity to recover sister chromatids otherwise mistakenly grouped as one chromosome fiber. We therefore propose a new form of karyotyping called polymer fiber karyotyping that is density or clustering independent. Ascribing a physical likelihood to polymers allows us to estimate the copy number of chromosomes, paving way for the study of copy number variations in interphase for which the expected copy number is unknown. For instance, the study of oncogene amplification in the setting of cancer heterogeneity is currently limited by reliance on compact alignment of probes in metaphase spreads²². It is also held back by uncertainty in measurement due to unknown true copy number post-oncogene amplification.

Perhaps not unexpectedly, a potential to resolve discrete polymer fibers instead of tabulating chromosome positions can point to inter-chromosomal interactions.. Multiplexed FISH captures in high throughput native chromosome structures directly in intact nucleus among a spectrum of different replicative states. Chromosomes undergo transformative structural change throughout the cell-cycle, disassembling the interphase nucleus to condense into sister chromatids during mitosis – a process that has been historically studied using proximity ligation sequencing¹. And yet, to date sister chromatid level interactions have been difficult to resolve

from imaging. This is perhaps due to a lack of biochemical labeling that can discriminately label sisters which are identical in sequence. Here, we propose a computational model relying on statistical mechanics to resolve sister chromatid interactions from fluorescence imaging and although our spatial genome aligner awaits further field testing, it has enabled us to model sister chromatids from a welter of imaged loci difficult to be discerned by the human eye. While the majority of replicated homologs appear with our method to be divorced and reside in separate territories, our modeling suggests compact territories where all four sister chromatids of a given chromosome spatially aggregate. Whereas in separate territories each sister seems to be shadowed by its attendant sister, we believe that in the compact territory sister chromatids can lose sister pairing and might even pair with the other homolog. Such an interpretation is evocative of a crossover event, which is thought to occur in mitotic cells at exceedingly rare frequencies (~1-2%) consistent with our observation^{60,61}. This deserves further investigation to study whether genetic information is indeed being exchanged and may bring significant insight to the regulatory mechanisms of loss of heterozygosity underlying disease. Mechanistically, how a sister chromatid eschews pairing with its own homologous attendant and assiduously choose a sister from the other homolog is a mystery. We acknowledge that diffraction-limited widefield microscopy may “merge” two signals belonging to separate chromatin fibers. Multiplexed STORM (ie. Oligo-STORM)^{25,27}, which can precisely resolve chromatin organization with single-molecule sensitivity at sub-diffraction limit resolution, may provide complementary information on sister-chromatid pairing and global organization. Additionally, as homologous interactions are highly species specific (ie. Diptera)^{55,56-58} we believe this study would benefit from multiplexed DNA-FISH in multiple model systems to study the permeance of this phenomenon. Studies such as these will simultaneously test the robustness of our spatial genome

aligner. For example, to what extent might the inferences of copy numbers and sisters by the method reflect artefacts arising from suboptimal primary imaging data? Does the size of each targeted region, which if too large may upon imaging disperse into multiple spots, affect our results? Will open, actively transcribed regions exacerbate this potential? Such questions regarding our approach could be addressed through the application of our approach to various biological systems across many laboratories.

Multiplexed DNA-FISH affords an intimate look into the elusive inner realities of genomic mosaicism in the brain. We chronicle the intranuclear spatial organization of copy number variations, with single-cell sensitivity and at whole-genome scale, heretofore only reported as frequencies⁶⁹⁻⁷³. We show that in neurons with three copies of a chromosome, the extranumerary chromosome shares a chromosome territory with another chromatin fiber, preserving two chromosome territories in the nuclei. Within the doubly occupied territory, each chromatin fiber appears to shadow each other, evocative of sister chromatids previously imaged in dividing mESCs. In our study we focus on the X chromosome, previously annotated in this dataset by the detection or absence of *Xist* RNA signal. We note that the absence of *Xist* signal as a proxy for identifying the active X chromosome may not unequivocally identify the active X chromosome,. We also cannot rule out the possibility double occupancy itself induces loss of *Xist* expression, or expression of genes on other chromosomes, potentially explaining why we detect no gene dosage effect across copy number variations (Supplemental Fig. 9). Further, we cannot exclude the possibility that active state influences probe labeling efficiency and imaging detection errors. The role and origin of the extranumerary chromosome is unclear. One possibility is that the extranumerary chromosome is the derivative of a non-disjunction event,

occurring in a neuroprogenitor during development. Why these chromatin fibers did not dissociate post-division is unclear. Another possibility is that the extranumerary chromosome is a remnant of asynchronous replication, a vestige in a neuroprogenitor that failed to withdraw or complete its replication timing⁷⁵. It has been noted for olfactory receptors, which singularly and stochastically express one receptor gene from an array of over a thousand possibilities, must then also select an allele after a gene selection⁷⁶. The allele selection appears to hinge on asynchronous replication; the expressed allele is faithfully marked by early replication. Others have also noted asynchronous replication as a possible epigenetic mark for monoallelic expression⁷⁷⁻⁷⁹. In this spirit and along the emerging field of genome imaging, we believe our spatial genome aligner provides a novel perspective of chromatin imaging analysis with which we hope to unravel foundational biological principles.

1.5 Methods

Spatial genome alignment algorithm

Conceptually, we abstract each detected fluorescence signal from a 3D image stack as a 4-D node $v = (x, y, z, t)$. Here, (x, y, z) correspond to sub-pixel spatial coordinates of a genomic locus detected in imaging, with a resultant positional uncertainty $(\sigma_x, \sigma_y, \sigma_z)$ in each spatial axis discovered from 3D Gaussian fitting. The fourth dimension, t , corresponds to an order of the gene on the reference genome, ordered from 5' to 3' for every chromosome. For notation, we use $v_{t,i}$ to refer to a node i with spatial position (x_i, y_i, z_i) corresponding to order t on the reference genome; there may be as many as n_t detected nodes for a given gene order t .

1. Graph construction: We define a directed acyclic graph $G = (V, E)$ as follows:

- $V = \{v_{t;i}\}$, $1 \leq i \leq n_t$, $1 \leq t \leq T$ represents the set of nodes in the graph, for every candidate node i of a gene order t among n_t candidates, for all T genes on the reference genome.

We note that due to signal dropout, there may be genes for which $n_t = 0$, in which case no nodes for the given order t are populated.

- $E = \{w_{t;i}^{t+c;j} = (v_{t;i}, v_{t+c;j})\}$, $1 \leq i \leq n_t$, $1 \leq j \leq n_{t+c}$, $1 \leq t \leq T$, $1 \leq c \leq C$ represents the set of all edges connecting ordered pairs of nodes, between every candidate node i of a gene order t among n_t candidates, to every candidate node j of a gene order $t+c$ among n_{t+c} candidates, for allowable skips $1 \leq c \leq C$, for all T genes on the reference genome.

We disallow self-loops by enforcing a lower bound on the skip parameter $c \geq 1$, such that no edges propagate $3' \rightarrow 5'$, or more explicitly, no edges from a node of order t connect to any node with order less than or equal to $t+1$. This prevents discovery of certain structural variants, such as inversions, translocations, and duplications, but helps restrict solutions to strictly the reference genome. We permit nodes to “look ahead” to downstream genes by skipping up to a permissible upper bound $c \leq C$, scaled later by an affine gap penalty. This accounts for signal dropout resulting in false negative signals, in which case all nodes of a given order t maybe false positives and must be skipped.

2. Calculate bond probabilities: We weight the edges using a physical analogy of a polymer model of DNA. Namely, we utilize the freely jointed Gaussian chain model, wherein chemical bonds model connections between two monomers. Here, our discrete spatially resolved genomic locations are analogous to these monomers, connected on the same chromatin fiber. In this model, the spatial distance separating these two locations is modeled after a Gaussian distribution:

$$w_{t,i}^{t+c;j} = 4\pi(R_{t,i}^{t+c;j})^2 \frac{1}{(2\pi(S_{t,i}^{t+c;j})^2)^{3/2}} e^{-\left(\frac{(R_{t,i}^{t+c;j})^2}{2(S_{t,i}^{t+c;j})^2}\right)} \quad (\text{Eq. 4})$$

where $R_{t,i}^{t+c;j}$ is the distance in nanometers between the i th node with gene order t to the j th node with gene order $t+c$. $S_{t,i}^{t+c;j}$ is expanded as:

$$(S_{t,i}^{t+c;j})^2 = \sigma_{t,i}^2 + \sigma_{t+c,j}^2 + \frac{2}{3}l_p\tau L_{t,i}^{t+c;j} \quad (\text{Eq. 5})$$

where the positional uncertainties of both the start locus $\sigma_{t,i}^2$ and end locus $\sigma_{t+c,j}^2$ are appended to the second moment $\langle S^2 \rangle = \frac{2}{3}l_p\tau L_{t,i}^{t+c;j}$. We note these positional uncertainties are previously measured via a Gaussian fitting routine around each detected fluorophore, and done separately for each laser channel. While the positional uncertainties may differ across imaging platforms, we estimate based on Su et al.³⁰ that the uncertainty is ~ 50 nm when the drift and chromatic aberration are appropriately corrected for. Operationally, we permit the toggling of the

term arising from integrating the spherical differential volume $4\pi(R_{t,i}^{t+c;j})^2$. In our analysis, we did not use this term, instead opting for the following:

$$w_{t,i}^{t+c;j} = \frac{1}{(2\pi(S_{t,i}^{t+c;j})^2)^{3/2}} e^{-\left(\frac{(R_{t,i}^{t+c;j})^2}{2(S_{t,i}^{t+c;j})^2}\right)} \quad (\text{Eq. 6})$$

Here, l_p is the persistence length of DNA in nanometers, τ is a scaling factor that converts genomic distance in base pairs to spatial distance in nanometers, and $L_{t,i}^{t+c;j}$ is the genomic distance in base pairs that separate the start locus $v_{t,i}$ and end locus $v_{t+c;j}$. The positional uncertainty is calculated separately for the start and end locus to accommodate chromatic aberrations, in which case loci imaged using different laser channels may have different localization error. $\tau L_{t,i}^{t+c;j}$ together reflects the contour length of the DNA polymer. Collectively, this allows a comparison of the observed spatial distance $R_{t,i}^{t+c;j}$ that separate two loci with an estimated spatial distance parametrized as $S_{t,i}^{t+c;j}$.

In this present study, we estimate τ by fitting a power-law function through pairwise spatial distance and genomic interval data to estimate a length scale of each base-pair per nanometer, as had been done in previous literature^{26,27,31,42}. We do so separately for each chromosome, as previous studies show the length-scales differ across mouse chromosomes³¹. We fix the persistence length at $l_p = 150$ bp, and allow τ to scale both the contour length L and the persistence length l_p , reducing one free parameter. Our key assumption in this model relies on the conversion factor τ . DNA compaction is known to change throughout A/B compartments⁷ and highly-expressed genes⁸⁰ among other regions, and our model's genomic scale is fixed: it

does not recapitulate the exact spatial distance at each imaged locus. Its fidelity may also change at different genomic scales (ie. 1 Mb, 25 kb, 5 kb). However, our main aspiration is to utilize this estimated spatial distance to inform *soft-decisions* in the selection of spots, in addition to optical qualities of signals detected, in a bid to trace chromatin structures.

In this manner, a traversal from chromosome start to end along this graph would accumulate a sequence of bond probabilities whose product reflects the physical likelihood of the discovered polymer:

$$CDF = \prod_{h=1}^H w_{v_h}^{v_{h+1}}, p = \langle v_1, v_2 \dots v_H \rangle \quad (\text{Eq. 7})$$

for every node v visited on path p from source to sink.

Each edge weight $w_{t,i}^{t+1;j}$ is negative log normalized into positive edge weights whose additive sum is equivalent to the polymer CDF. We do so for several reasons: (a) this controls for numerical underflow in calculating the multiplicative product of small decimals (b) this reframes the optimization objective from maximizing the likelihood function to minimizing the negative log likelihood (c) the CDF is now computed as a sum of edge weights, permitting the use of existing dynamic programming shortest-path algorithms solving for additive edge weights. Below, we write edge weights w to represent negative log normalized bond probabilities. To permit nodes to skip potential false positive and “look ahead” to downstream genes, we apply penalize each bond skipped:

$$w_{t,i}^{t+c;j} = \gamma_{c-1} w_{t,i}^{t;j} \quad (\text{Eq. 8})$$

where γ_{c-1} is a gap penalty scaled for every skip c . Adjacent nodes ($c=1$) are not penalized. This gap penalty is an important regularization policy that makes sure a skipped edge never has a smaller cost than the cumulative cost of an equivalent number of consecutive edges. For example, if a segment of 1 Mb has a log-bond probability of 5013.34 (0.34 nm-per-bp, $l_p = 150$ bp, 1 pixel ~ 100 nm) and a segment of 2 Mb has a log-bond probability of 10014.38, the cost of “skipping” an edge (traversing a 2 Mb bond) may be preferred instead of traversing two consecutive 1 Mb edges ($2 \times 5013.34 = 10026.68$). We apply a multiplicative gap penalty of at least this ratio $\gamma_1 \geq \frac{10026.68}{10014.38}$ to prevent a preferential skip.

3. Initialize adjacency matrix: We append a single source and a single sink node in our graph to allow gaps in the start and end of polymer alignment. From our graph G with total $N = \sum_{t=1}^T n_t$ nodes, we construct an (N, N) adjacency matrix that is padded by an additional row with index 1 and column with index $N+2$ to an $(N+2, N+2)$ matrix. We initialize the first row of the adjacency matrix with “pseudo”-bonds that enable up to the first K genes to be skipped. These edges linking an imaginary starting position to an observed position are weighted as:

$$w_{1;i}^{k;j} = \frac{1}{\left(\frac{4\pi}{3} l_p \tau L_{t,i}^{t+c;j}\right)^{3/2}} e^{-\left(\frac{\alpha^2 \tau L}{3 l_p}\right)}, \quad 1 < k \leq K \quad (\text{Eq. 9})$$

where α scales an imaginary stretched genomic segment with an implicit skip penalty.

We also initialize the last column of the adjacency matrix with “pseudo”-bonds that enable up to the last K genes to be skipped. Similarly, the edges linking an observed position to an imaginary ending position is weighted as:

$$w_{N-k;i}^{N;j} = \frac{1}{\left(\frac{4\pi}{3}l_p\tau L_{t;i}^{t+c;j}\right)^{3/2}} e^{-\left(\frac{\alpha^2\tau L}{3l_p}\right)}, 1 < k \leq K \quad (\text{Eq. 10})$$

These initializations behave as “seeds” that permit late chromosome starts (from locus 1 – k) and early chromosome ends (from locus $N-k - N$) should the first or last few loci have poor labeling efficiency and need to be skipped.

4. Path finding: We find the shortest path from source to sink in our graph using a dynamic programming algorithm. As all edge weights in this directed graph are non-negative, and the sum of traversed edges equate to the CDF of the traversed polymer, we utilize Dijkstra’s shortest-path algorithm as a dynamic programming means for finding the most plausible polymer:

$$\min \sum_{h=1}^H w_{v_h}^{v_{h+1}}, p^* = \langle v_1, v_2 \dots v_H \rangle \quad (\text{Eq. 11})$$

for every node v visited on the shortest path p^* .

We estimate our worst-case time complexity to be:

$$\mathcal{O}(|E| + |V| \log|V|) = \mathcal{O}\left(\left|\sum_{t=1}^T n_t \sum_{t+1}^{t+C \leq T} n_t\right| + \left|\sum_{t=1}^T n_t\right| \log\left|\sum_{t=1}^T n_t\right|\right) \quad (\text{Eq. 12})$$

Polymer fiber karyotyping algorithm

We developed a routine that finds all possible polymers of a given chromosome on a cell-by-cell basis. Chiefly, we accept all polymers below a physical likelihood threshold. This

threshold can be derived by scrambling the genomic intervals separating each probed locus, such that the observed genomic distance no longer abides by the expected distances. We then perform an iterative search wherein nodes of each shortest path discovered are subtracted from graph G before searching for the next shortest path, until no likely paths below the physical likelihood threshold can be discovered (Supplemental Fig. 10).

We retrieved spatially resolved fluorescent signal information for mouse genome-wide DNA seqFISH+ probe sets at multiple length scales (1 Mb, 25 kb), and for multiple cell types (mESC: <https://zenodo.org/record/3735329>; mouse cortical neurons: <https://doi.org/10.5281/zenodo.4708112>). Congruent with previous publications^{26-32,42}, we fitted a power-law function through pairwise spatial-distances of observed loci, plotted against its genomic-distance separation. From this power-law function, we estimated a parameter corresponding to nanometers per base pair, for every cell-type, for every mouse chromosome. To evaluate the performance of our parameter in its ability to recapitulate true chromatin structure, we performed a hyperparameter search. Explicitly, we fit a power-function through pairwise spatial and genomic intervals, and with the power-function we estimated the distance scale (nanometers-per-basepair) for a range of different genomic resolutions (ie. 10kb, 100kb, 1Mb, 10 Mb ...) (Supplemental Fig. 3A). Using 10% of cells in each dataset, we performed spatial genome alignment and compared its median distance matrix to Hi-C or cell-type resolved Dip-C. We utilized the parameter that engendered the best fit for our final analysis.

With this spatial distance parameter, for every nucleus and for every chromosome, we iteratively performed spatial genome alignment until no physically plausible fibers could be

discovered. Finally, we counted the number of fibers discovered for every chromosome to assign a chromosome copy number, producing a cell karyotype.

mESC Hi-C data analysis

To evaluate the spatial genome alignment results of mESCs, we retrieved mESC Hi-C contact matrices from the 4DN Data Portal (experiment set 4DNESU4Y9CBF). Next, we utilized Straw (<https://github.com/aidenlab/straw>; v=0.0.6) to extract Knight-Ruiz normalized count matrices for every mouse. Whereas read counts can be evenly binned, the loci imaged by DNA seqFISH+ spanned irregular intervals. To compare Hi-C with DNA seqFISH+ imaging data, where the genomic distances separating each locus are irregular intervals, we performed the following normalization. For DNA seqFISH+ imaging data, we kept the imaged locus ordered 5' to 3' closest to each integer 1 Mb or 25 kb bin, dropping all other loci to calculate a final distance matrix. For the corresponding Hi-C matrix, we binned the reads at 1 Mb and 25 kb resolution respectively, dropping the same bins removed from the DNA seqFISH+ distance matrix. To assess spatial genome alignment accuracy, we compared the median imaging distance matrix to its corresponding KR-normalized Hi-C contact matrix using the Spearman correlation coefficient.

Excitatory mouse cortical neuron Dip-C data analysis

To evaluate the spatial genome alignment results of excitatory mouse cortical neurons, we retrieved cell-type resolved Dip-C contact matrices from NCBI GEO (accession GSE162511)⁷⁴. First, we note a dearth of multi-modal data that ideally allows concomitant cell type classification using one sequencing modality and proximity-ligation analysis on the same

cell. In Dip-C, neuronal cell types were resolved by co-projecting *NeuN*⁺ neurons with bulk Hi-C sequencing of multiple cell-types. Neurons co-clustering with *Slc17a7*⁺ cells delineating excitatory neurons were also classified as excitatory. It is possible some of these cells classified as excitatory may belong to broader cell types. In contrast, DNA seqFISH⁺ imaging resolved both RNA and DNA in the nuclei, which allowed excitatory neuron markers (ie. *Slc17a7*⁺, *Neurod2*) detected in RNA imaging to label cell types. Second, we note a difference in the mouse strains of the two datasets. Dip-C focused on an F1 cross with CAST/EiJ x C57BL/6J background while the DNA seqFISH⁺ imaging purely focused on C57BL/6J mice.

Amid these caveats, we utilized Straw to extract Knight-Ruiz normalized count matrices for every mouse chromosome. For the DNA seqFISH⁺ imaging data, we again kept the imaged locus ordered 5' to 3' closest to each integer 1 Mb bin, dropping all other loci to calculate a final distance matrix. For the corresponding Dip-C contact matrix, we binned the reads at 1 Mb resolution and dropped the same bins removed from the imaging distance matrix. To assess spatial genome alignment accuracy, we compared the imaging distance matrix to its corresponding Hi-C matrix using the Spearman correlation coefficient.

Using the same Dip-C dataset, we inspected haplotype-resolved reads – namely pre-processed `seg` files – to evaluate read counts assigned to each haplotype. We discarded any ambiguous or multi-contact read-pairs, and counted cells wherein one haplotype has nearly twice as many reads as the other, as a proxy for identifying copy number variations and the haplotype source of those copy number variations at the single neuron level.

Benchmarking against M-DNA-FISH spot selection algorithm

In conjunction, we analyzed M-DNA-FISH imaging of a 210-kb genomic region spanning the Sox2 locus (chr3:34,601,078-34,811,078) based on prior work⁴⁹. We considered all chromosome centers assigned to the 129 allele, which lacks the 7.5 kb tandem CTCF-binding sites (CBSs) inserted on the CAST allele. The previous expectation-maximization routine outlined in (28, 49) generates 10 candidate spots per locus, per putative chromosome in every nucleus, assuming a diploid cell line. Each candidate spot is assigned a score, derived as a combination of (a) fluorescence intensity (b) proximity to chromosome center (c) agreement with moving average of previous loci positions. A candidate spot with a score of -1.5 or more is considered a high-quality spot for the E-M routine, with more negative scores tracking with decreasing quality. For spatial genome alignment, we fed all candidate spots with a much lower quality threshold score of -4, such that for every high-quality spot there is also a low-quality spot. We included this extra noise and ignored the fluorescence intensity information to demonstrate the supreme utility and specificity of genomic distances as a spot selection criterion. We compared the spatial genome alignment results and E-M tracing results against Hi-C, which sequenced the control mouse chr 3 lacking the CBS insertion. We also calculated the Spearman correlation between the discovered pairwise median distances and pairwise KR-normalized Hi-C contact frequency for each of the two algorithms.

Homolog assignment and sister chromatid aggregation analysis

In diploid cells, we utilized density-dependent clustering algorithm DBSCAN⁸¹ (scikit-learn; v=1.0.1) to separate homologous chromosomes residing in separate chromosome territories.

In tetraploid cells, instead of utilizing density-based clustering to parse and assign homologs, we opted to take a different approach. We first assigned chromatin interaction patterns (ie. separate homologs, compact homologs, separated sisters in tetraploid cells) with DBSCAN to find spatially separable structures and classify tetraploid cells. To assign sister chromatids and in turn homologs, we paired chromatin fibers by the closest starting positions and assigned them as sisters of the same homolog. This allowed us to pair sisters as part of the same homolog in tetraploid cells which had only one spatially dense cluster (ie. compact homologs). In the setting of compact chromosomes where two homologs are spatially not separable, we accounted for alternative pairing scenarios, such as pairing by the telomeric ends. We analyzed the spatial separation of each chromosome starts and ends based on pairing by centromeric starts as well as pairing by telomeric ends, and all possible alternative pairings.

Sister chromatid misselection assessment

To identify potential loci belonging to one sister chromatid but misselected by the other, we examined all 402 putative sister chromatid pairs for mESC chr 1 (177 loci in length). First, we interpolated spatial coordinates for any loci not detected on a sister chromatid fiber. Next, using a sliding window of three loci beginning at locus 3 and sliding until locus 175, we exchanged spatial positions between the main sister and attendant sister for these three loci. Finally, we re-calculated the bond probabilities of three loci upstream to the three loci exchanged. If the multiplicative product of all recalculated probabilities is more likely than the product of original bond probabilities, we call this a misselection event. We also counted any consecutive sequence of misselections and called this a cross-over event.

Simulation of aneuploid cells and analysis

To study the impact our aligner has on chromosome counting in isolation from other sources of technical artefact, we utilized the polychrom package (<https://github.com/open2c/polychrom/>) to simulate chromosomes under nuclear confinement⁸². We randomized copy numbers for each of the 20 mouse chromosomes, generating 1000 such sets of aneuploid cells. We then performed polymer fiber karyotyping on each cell, repeating the process for different levels of added noise and signal dropout. Detailed descriptions of the parameters underlying the simulation, as well as descriptions of the analysis can be found under Supplementary Materials.

Simulation of sister-chromatid pairs and analysis

To study the impact our aligner has on parsing tightly paired chromatin fibers from other sources of technical artefact, we utilized the polychrom package to simulate tightly paired chromatin fibers. Simulating 100 pairs of sister chromatids corresponding to mouse chromosome 1, we varied the pairing strength between coupled fibers to generate five separate conditions. We then applied polymer fiber karyotyping on each of the 100 pairs, across the five conditions, fixing the signal dropout to 50% and added noise to 0%. Detailed descriptions of the parameters underlying the simulation, as well as descriptions of the analysis can be found under Supplementary Materials.

Normalization of chromatin levels for pseudotime analysis

Consistent with previously published work, we constructed a generalized linear model (GLM) for normalizing sequential immunofluorescence data. In this GLM are latent variables

controlling for and removing contributions from cell size, total fluorescence intensity over all chromatin marks, as well as batch effect from replicate ID and field of view (FOV). Specifically, we utilized a non-canonical log-link function:

$$\log Y_i \approx \beta_0 + \sum_j \beta_j X_j \quad (\text{Eq. 13})$$

where Y_i is the vector of total fluorescence intensity of chromatin mark i across all cells, and X_j are latent variables contributing to bias. Using the Pearson residuals of each fitting, we corrected the readout of two chromatin marks, H4K20me1 and H4K16ac, with which we constructed a principal curve to order cells across the cell cycle or pseudotime. Matching previously published results, H4K20me1 and H4K16ac peaked at opposite ends of the pseudotime axis, ordering cells from G2/M.

Gene dosage and transcriptional activity analysis

After polymer fiber karyotyping, we stratified cells by the copy numbers discovered for each chromosome and inspected the seqFISH+ RNA imaging corresponding to genes in the DNA seqFISH+ imaging. We performed pairwise t-tests to evaluate statistical significance for the few genes that lie on each mouse chromosome.

Data Availability

All traced chromatin structures with spatial genome alignment is available and hosted on 4D Nucleome Data Portal (accessions: 4DNFIXGTJBGU , 4DNFIFBXKXK9, 4DNFIYNWVJEP, 4DNFI7G3BWDF, 4DNFIVBL8AWT, 4DNFIU73OR5W, 4DNFIS6MLXGA). Published seqFISH+ datasets analyzed in this manuscript is hosted and available on Zenodo (mESC: <https://zenodo.org/record/3735329> ; mouse cortical neurons:

<https://doi.org/10.5281/zenodo.4708112>). Published M-DNA-FISH datasets analyzed in this manuscript is hosted and available on 4D Nucleome Web Portal (accessions 4DNESIGBIXQS, 4DNESC5PKTQ9, 4DNES51KSIZ9, 4DNESNEKOCAP and 4DNESOXQX1JT).

Code Availability

The spatial genome aligner is available at <https://github.com/b2jia/jie83>. The aligner and peripheral analysis relies on algorithms implemented in scipy (v=1.7.2)⁸⁴ and scikit-learn (v=1.0.1)⁸¹. A walkthrough of the intuition behind this algorithm can be found at <https://github.com/b2jia/jie/blob/main/jupyter/00-spatial-genome-alignment-walk-thru.ipynb>.

1.6 Figures

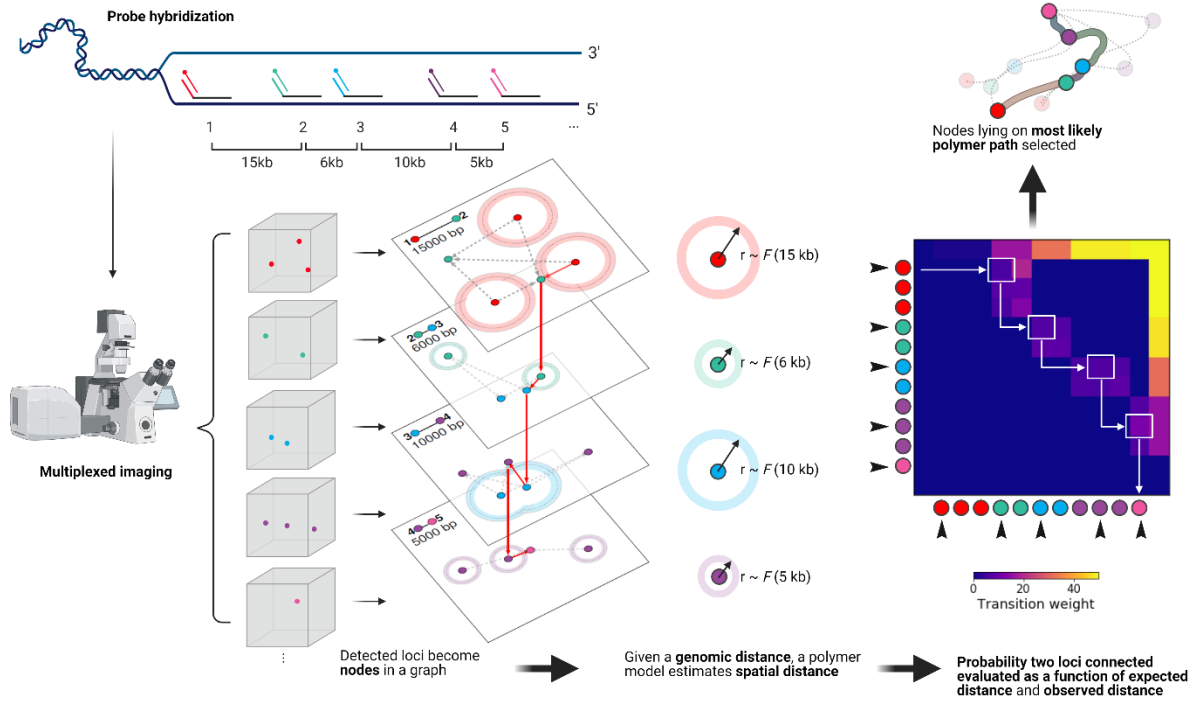


Figure 1.1: Spatial genome alignment of multiplexed DNA-FISH imaging data against a reference soft-polymer structural model of DNA. Schematic of spatial genome alignment. Spatial coordinates in three dimensions (x, y, z) of signal detected from each imaged loci are abstracted as nodes in a graph, ordered by their appearance on the reference genome. Utilizing a freely jointed Gaussian chain model, our aligner estimates an expected spatial distance based on the genomic distance separating two loci. The observed spatial distance is compared to the observed distance, and an edge between two loci are connected weighted proportionally to physical likelihood. We task the aligner to find the shortest path through our adjacency matrix, which returns the sequence of spatial positions whose path length equates to the most likely polymer.

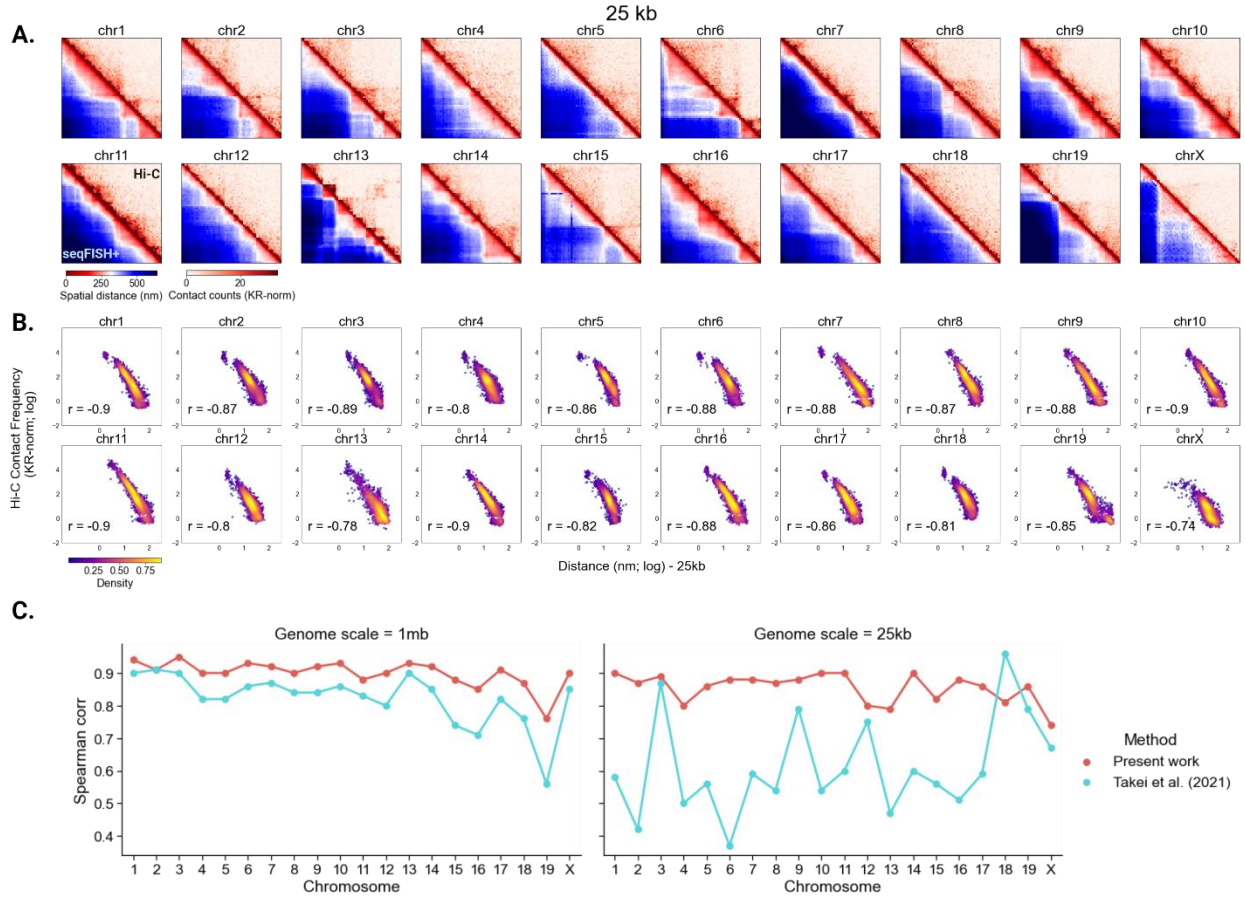
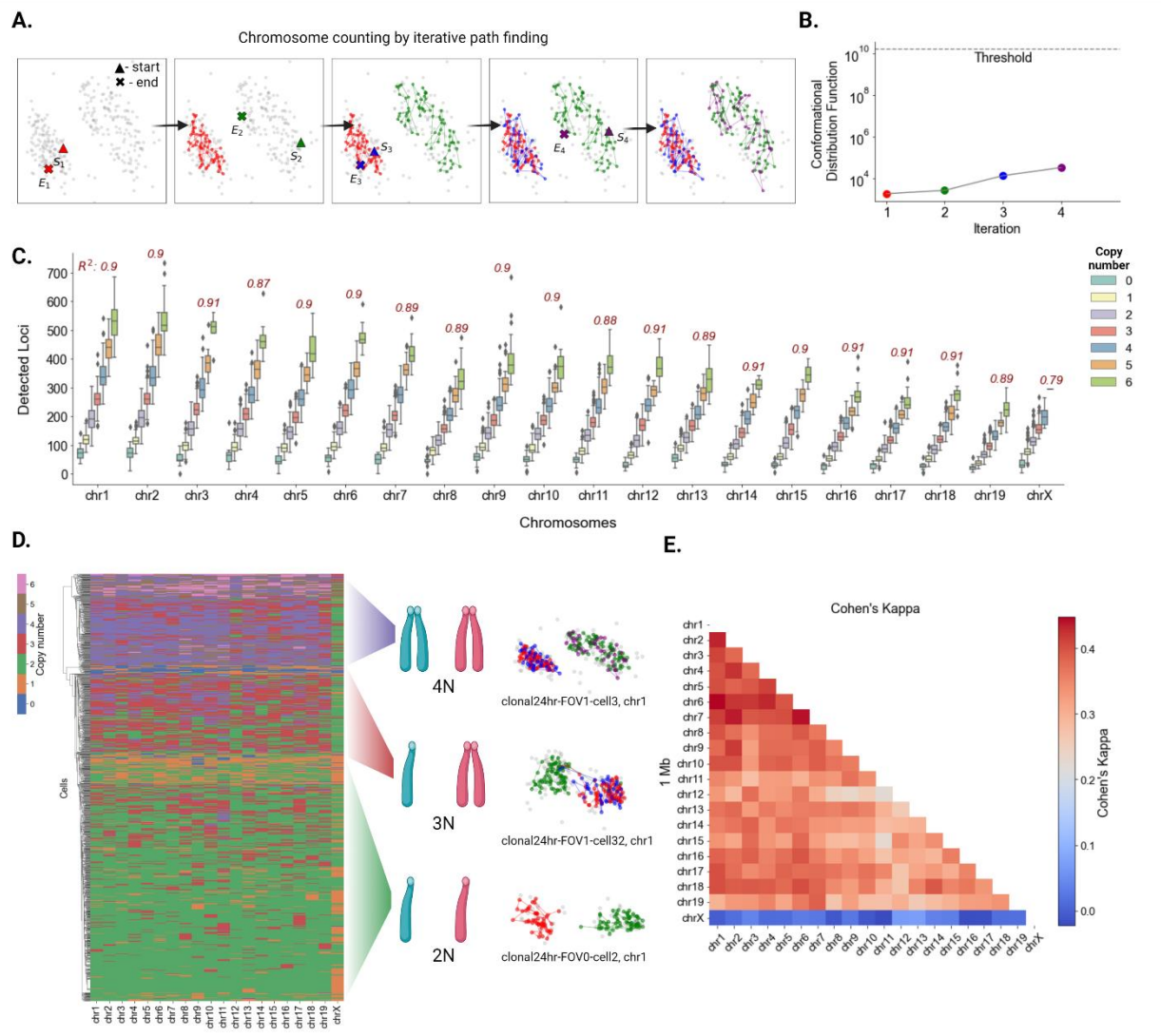


Figure 1.2: Spatial genome alignment of seqFISH+ chromatin imaging of mESCs at 25 kb and 1 Mb resolution. (A) Heatmaps of seqFISH+ chromatin imaging of mouse embryonic stem cells (mESCs) at 25 kb resolution (bottom left) juxtaposed to contact frequency from bulk proximity ligation assay or Hi-C binned at 25 kb (top right). The distance matrix is an untabulated, median distance matrix of all single-molecule chromatin fibers identified by spatial genome alignment for every chromosome across 1160 cells. Heatmaps for 1 Mb chromatin tracing imaging data are shown in Supplementary Fig. 4. (B) Spearman correlation between pairwise spatial distances (x-axis; log-normalized) imaged at 25 kb resolution against Hi-C contact frequency (y-axis; log-normalized) binned at 25 kb resolution. (C) Spatial-distance to proximity-ligation correlation comparison across methodologies, at 1 Mb seqFISH+ imaging resolution as well as 25 kb resolution. For 25 kb resolution where DNA domains organize nonlinearly, we note that single-molecule chromatin fiber tracing via spatial genome alignment captures structural variations more faithfully than tabulating all pairwise spatial distances within a specified radius into an ensemble structure.

Figure 1.3: Polymer fiber karyotyping on interphase mES cells. (A) Schematic of polymer fiber karyotyping by iterative path subtraction. For every plausible polymer found, all nodes visited on the polymer path are subtracted before a new graph is constructed for additional rounds of spatial genome alignment. Paths are extended until no physical plausible paths can be discovered. (B) The score of each path is recorded, interpreted as the conformational distribution function of the polymer, and compared to a physically unlikely threshold. Paths are extended for a given chromosome in a cell until either no more paths can be extended or a path extended have a score above threshold. (C) Boxplot (center line - median; box - 25 to 50%tile; bars - minima, maxima): of predicted copy number and total detected loci per chromosome (including spots omitted by spatial genome alignment) of mESC (n=1160 cells examined over 4 experiments). For every extra chromosome detected by polymer fiber karyotyping, we find a stepwise multiplicative increase in the total detected loci (eg. 1 chr - 100 points, 2 chr - 200 points, ...). Pearson correlation coefficient evaluates this trend of detected loci per increase in assigned ploidy. (D) Hierarchical clustering of mESCs by copy number similarity. Copy number of chromosomes is congruent across the mouse genome in a given cell with the exception of chr X in this male cell line. A dominant faction of cells is 2N in addition to a smaller faction of 3N cell with unsynchronized replication and a smaller faction of 4N cells post-replication. (E) Heatmap of pairwise comparisons of copy number assigned by different chromosomes, with agreement scored by Cohen's kappa.



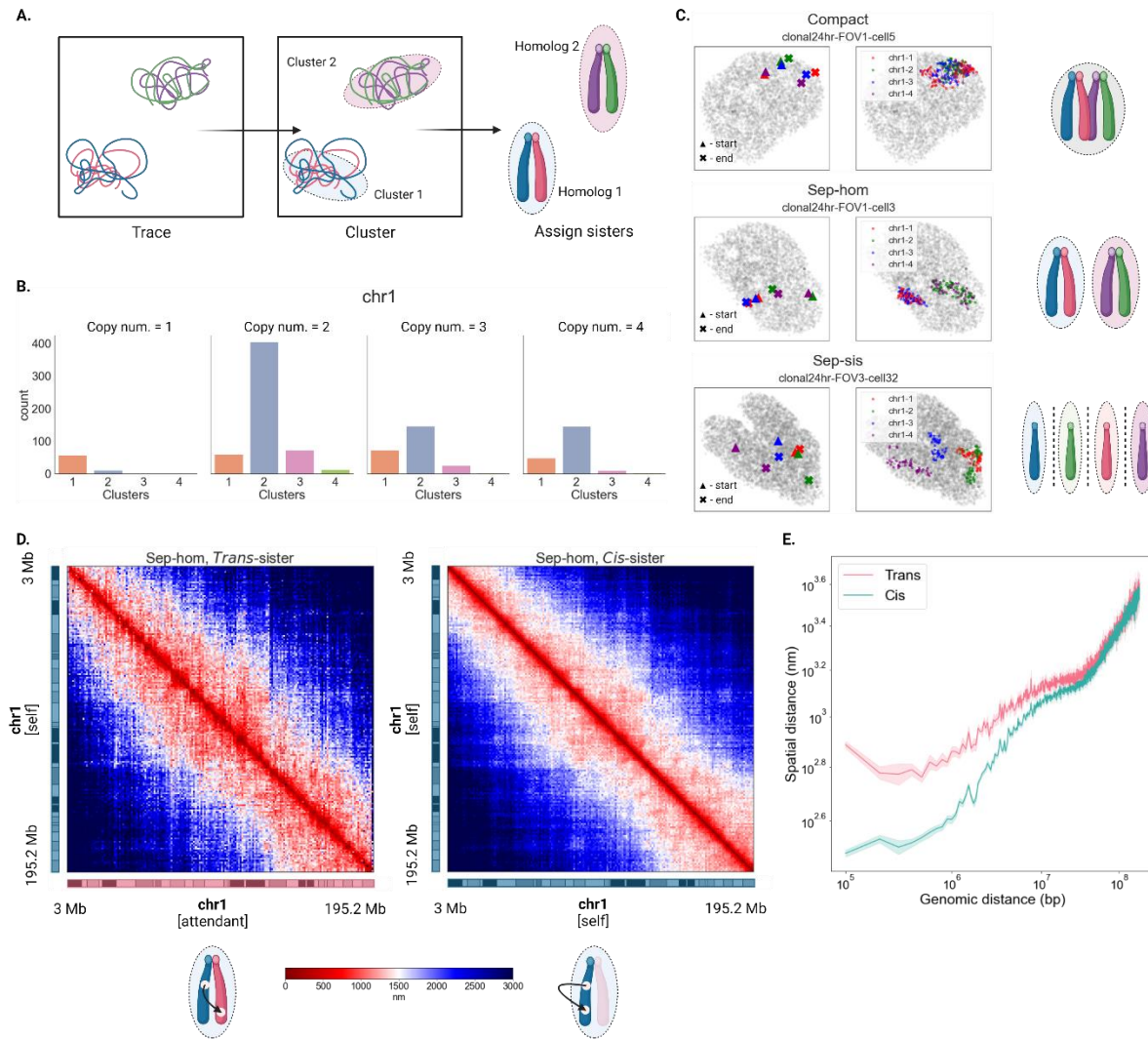
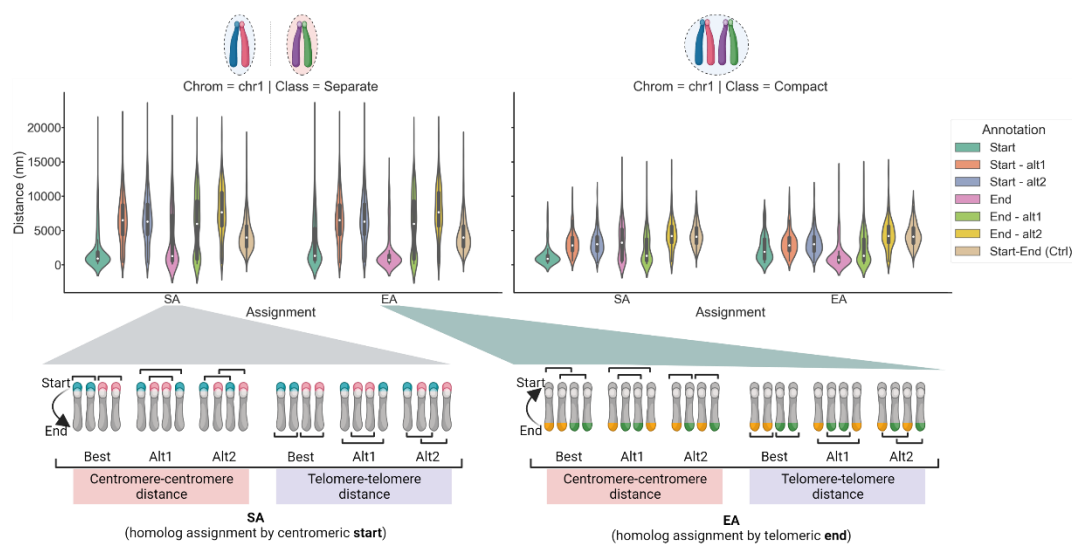


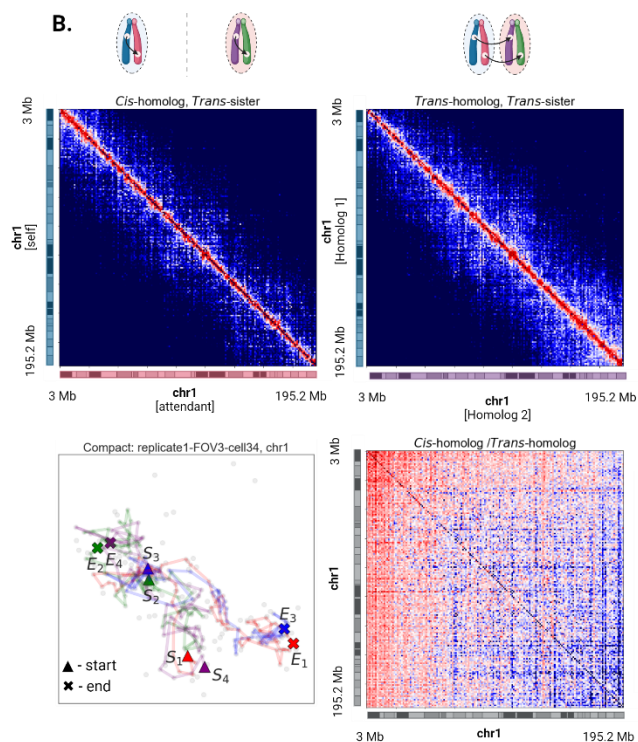
Figure 1.4: Sister chromatid interactions in mESCs resolved by polymer fiber karyotyping. (A) Schematic of homolog assignment. Spatially separable structures are clustered using DBSCAN and polymer fibers residing in the same spatial neighborhood are classified as sisters of the same homolog. (B) Histogram profile of chr 1 traced at 1 Mb resolution across 1N, 2N, 3N and 4N cells (n=1160 cells over 4 experiments). (C) Representative classes of homologous chromosome organization in 4N mESC cells. A dominant fraction of cells has two spatially separable homologs (sep-hom) residing in chromosome territories, while another fraction of cells has one spatial cluster wherein both pairs of sisters from different homologs interact (compact). A marginal population have three or more spatially separable structures showing separated sister chromatids (sep-sis). (D) Median distance matrices of intra-sister chromatin fiber vs inter-sister chromatin fiber interactions exhibiting tight locus pairing between sister chromatid of the same homolog. (E) Spatial distance separation (bolded line – mean; shaded band – 95% confidence interval) across genomic length scales between loci on the same sister chromatin fiber (cis) vs separation between loci across different sister chromatin fibers (trans).

Figure 1.5: Sister chromatid interactions in compact homolog state. (A) Violin plot (center dot – median; gray bar - quartiles) comparing spatial separation between centromeric starts or between telomeric ends of sister chromatids, stratified by sister chromatids in separate chromosome territories (n=146 cells across 4 experiments) and by sister chromatids in a compact territory (n=49 cells across 4 experiments). All 6 permutations, including sisters assigned by closest centromeric starts and its alternate pairings are shown, as well as sisters assigned by closest telomeric ends and its alternate pairings. Note that in mouse chromosomes, the centromere is located at the end of the chromosome. We refer to the centromere and centromeric to delineate the centromeric telomere; the telomere and telomeric to delineate the noncentromeric mouse telomere. The pairwise distance between a centromeric start and its telomeric end on the same fiber is presented as a control for unpaired loci. The colors of the chromosome cartoons do not imply phased labeling, and instead emphasize combinations of grouping two fibers. (B) Median distance matrix of cis-homolog trans-sister chromatids vs trans-homolog trans-sister chromatids. A representative structure is shown on the bottom left, with a heatmap showing cis-homolog to trans-homolog distance ratio on the bottom right. (C, D & E) A heatmap of assigned copy numbers (C) and tetraploid chromosome organization state (D) ordered along pseudotime determined as a function of cell cycle markers H4K20me1 and H4K16ac (E).

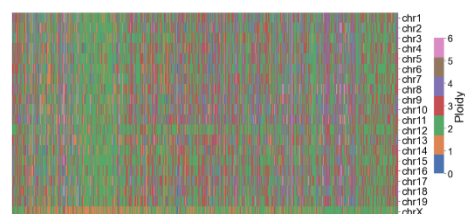
A.



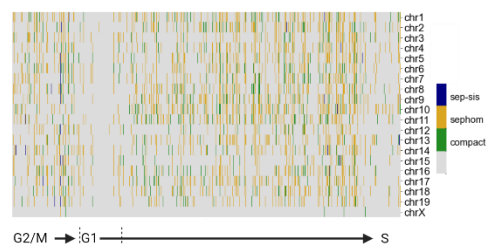
B.



C.



D.



E.

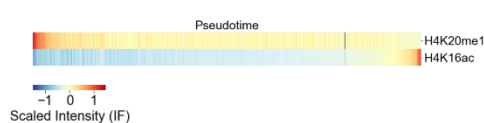
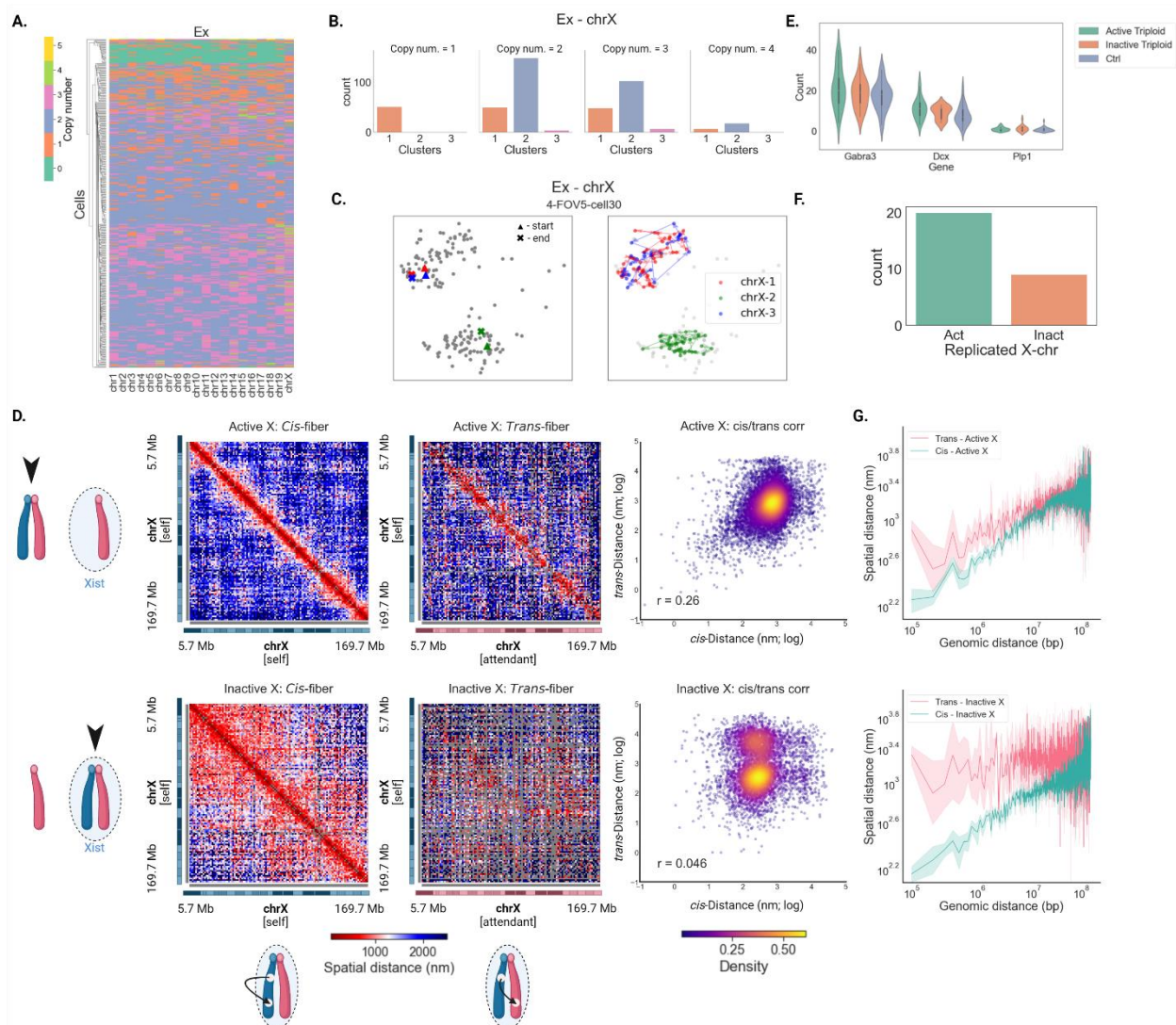


Figure 1.6: Copy number variations in mouse cortical neurons. (A) Hierarchical clustering of n=458 excitatory cortical neurons from female mice by copy number similarity. (B) Histogram profile of chr X traced at 1 Mb resolution across excitatory cortical neurons bearing 1, 2, 3, and 4 copies. (C) Representative image of 3N chr X organization in excitatory cortical neurons. Two chromosome territories are preserved despite harboring three chromatin fibers; two fibers reside in one territory while the third is divorced in a separate territory. (D) Median distance matrices comparing cis-fiber loci distances to trans-fiber loci distances on the active and inactive chr X. Log-normalized distance density heatmap (right) examines the correlation between cis-trans distances, on the active and inactive X chromosome. (E) Violin plot (center dot – median; gray bar – quartiles) of gene expression counts of three genes lying on chr X, comparing the effect of gene dosage on transcription, sub-classified by chr X activation status (active: n=20 cells across 4 experiments; inactive: n=9 cells across 4 experiments; control: n=20 randomly selected cells across 4 experiments). (F) Count plot of excitatory cortical neuron nuclei with three copies of chr X, classified by the activation status of the territory where two of the three chr X fibers colocalize. (G) Spatial distance separation (bolded line – mean; shaded band – 95% confidence interval) across genomic length scales between loci on the same chromatin fiber (cis) vs separation between loci across separate chromatin fibers (trans), sub-classified by chr X activation status.



1.7 Supplementary Figures

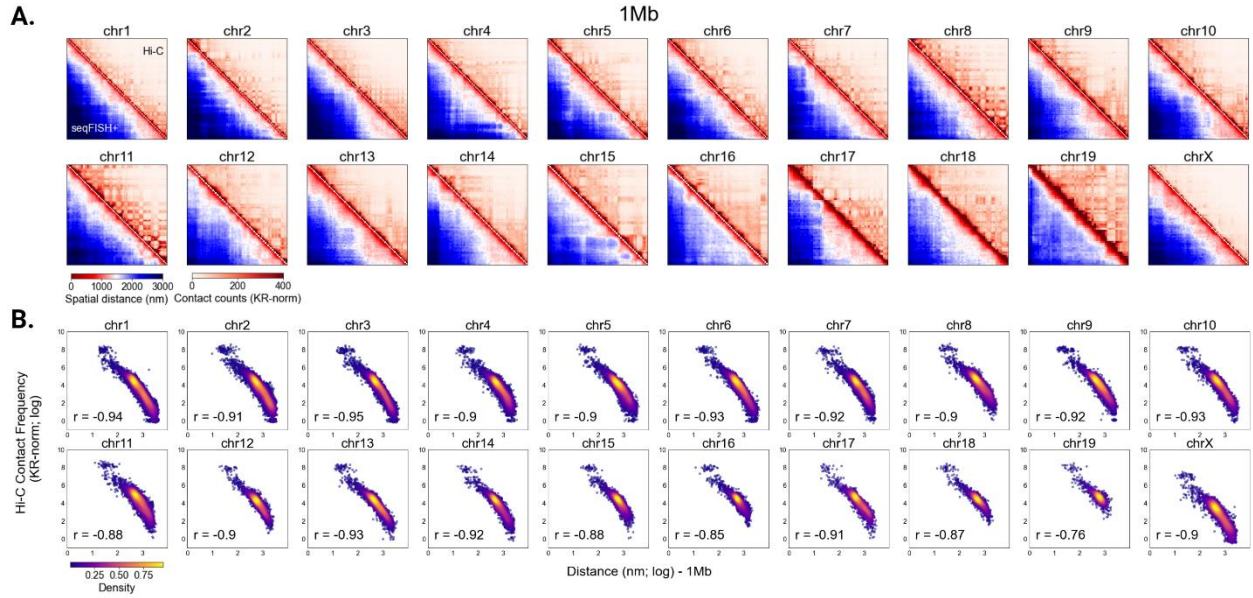


Figure S1.2: Spatial genome alignment on sequential 5 kb M-DNA-FISH imaging of mESC Sox2 locus. (A) Median distance matrix of spatial genome alignment resolved Sox2 locus (chr3:34601078-34811078), sequentially imaged at a finer 5 kb resolution, using a different imaging protocol based on multiplexed DNA-FISH. (B) Median distance matrix identified by traditional spot selection using an expectation-maximization routine previously used to analyze M-DNA-FISH datasets. (C) KR-normalized Hi-C contact matrix of the Sox2 locus, juxtaposed as a comparison to median distance matrices discovered by two chromatin tracing algorithms. (D & E) Correlation plot of pairwise genomic distances discovered by spatial genome alignment and an E-M spot selection routine respectively, plotted against KR normalized Hi-C contact frequency. (F) Heatmaps of tabulated distances (radius = 150 nm) from spatial genome alignment of M-DNA-FISH imaging (bottom left) juxtaposed to contact frequencies from Hi-C binned at 5 kb (top right). (G) Heatmaps of tabulated distances (radius = 150 nm) from E-M spot selection of M-DNA-FISH imaging (bottom left) juxtaposed to contact frequencies from Hi-C binned at 5 kb (top right). (H) Spearman correlation (y-axis) of tabulated distances from M-DNA-FISH tracing against bulk Hi-C, across a range of distance thresholds (x-axis) for tabulating distances into contacts, delineated for each of the tracing methods (spatial genome alignment and E-M spot selection).

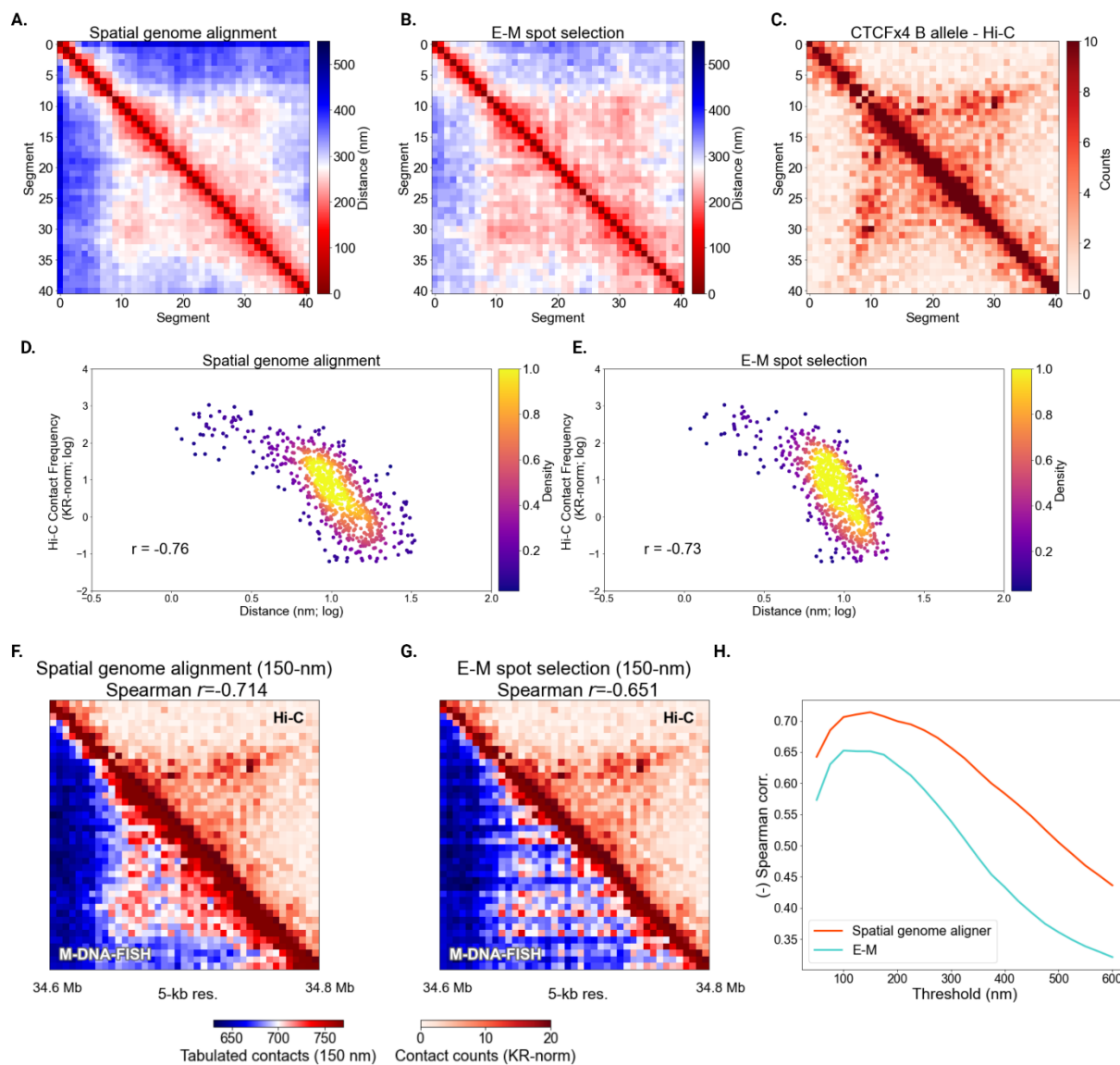
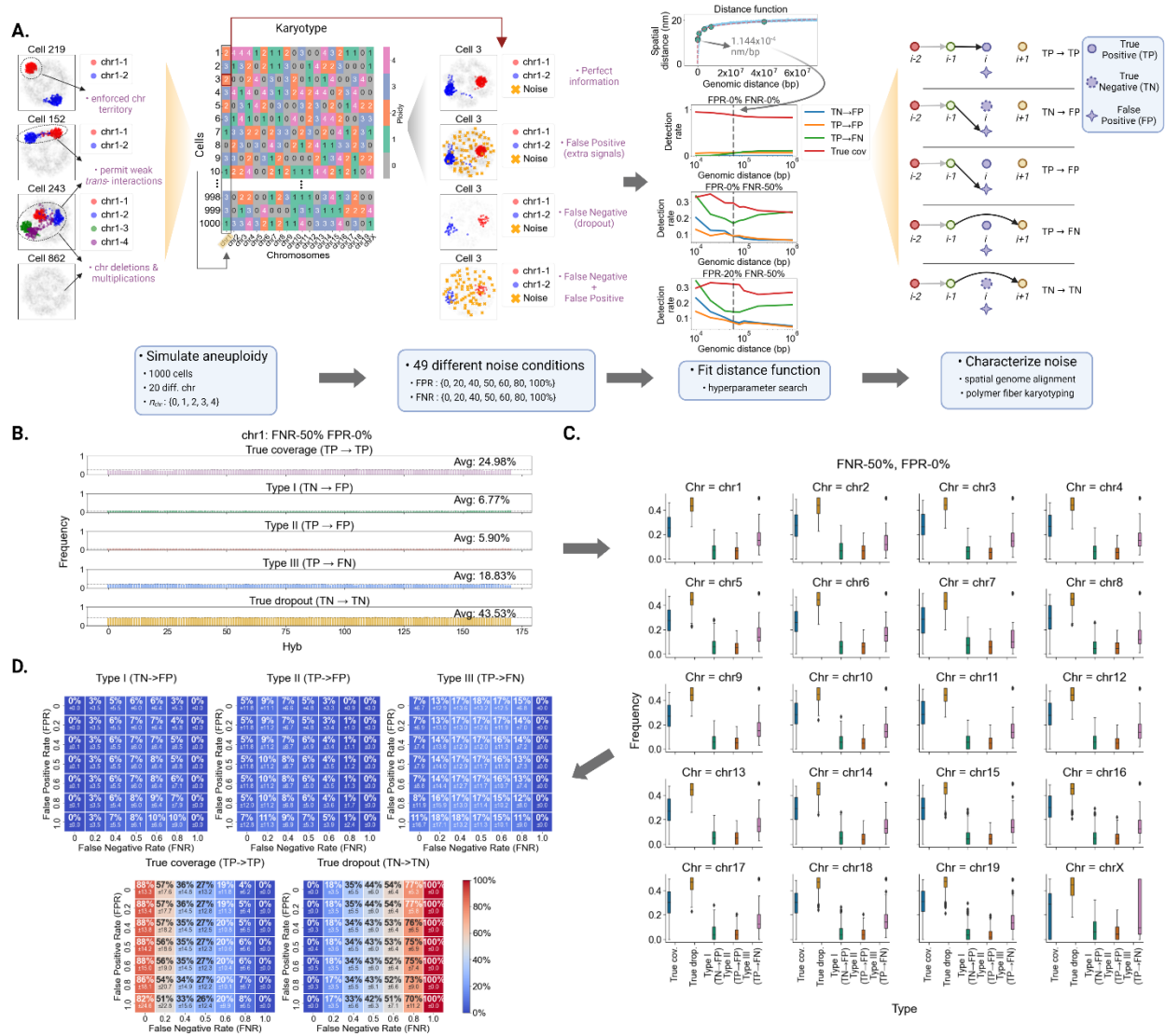


Figure S1.3: Spatial genome alignment on simulated aneuploid cells. (A) Schematic of simulation process generating aneuploid cells. Enforced chromosome territories segregated separate copy numbers, but permitted weak trans-interactions. Each aneuploid cell was subject to 49 different conditions of dropout and added noise. A power law fit to pairwise genomic-to-spatial distance allowed the estimation of expected spatial distance for a range of genomic distances (green dots, scatter plot). Subsequent alignment via polymer fiber karyotyping was parsed into 5 categories of signal selection. The best distance parameter – spatial distance divided by genomic distance – was selected based on the parameter that minimized the errors parsed, after alignment on 10 simulated cells on both perfect information (0% FNR 0% FPR) as well as imperfect information (50% FNR 0% FPR; 50% FNR 20% FPR). (B) Histogram of signal selection, delineated into 5 categories, for mouse chromosome 1 at 50% False Negative Rate (FNR) and 0% False Positive Rate (FPR). (C) Boxplot (center line – 50%tile; box – 25% to 75%tile; bar – 0 to 100%tile) of signal selection, delineated into 5 categories, for all mouse chromosomes (n=1000 simulated cells with an average copy number of 2 per chromosome) at 50% FNR and 0% FPR. (D) Table of signal selection, delineated into 5 categories, summarized for all mouse chromosomes across all 49 conditions of dropout and added noise.



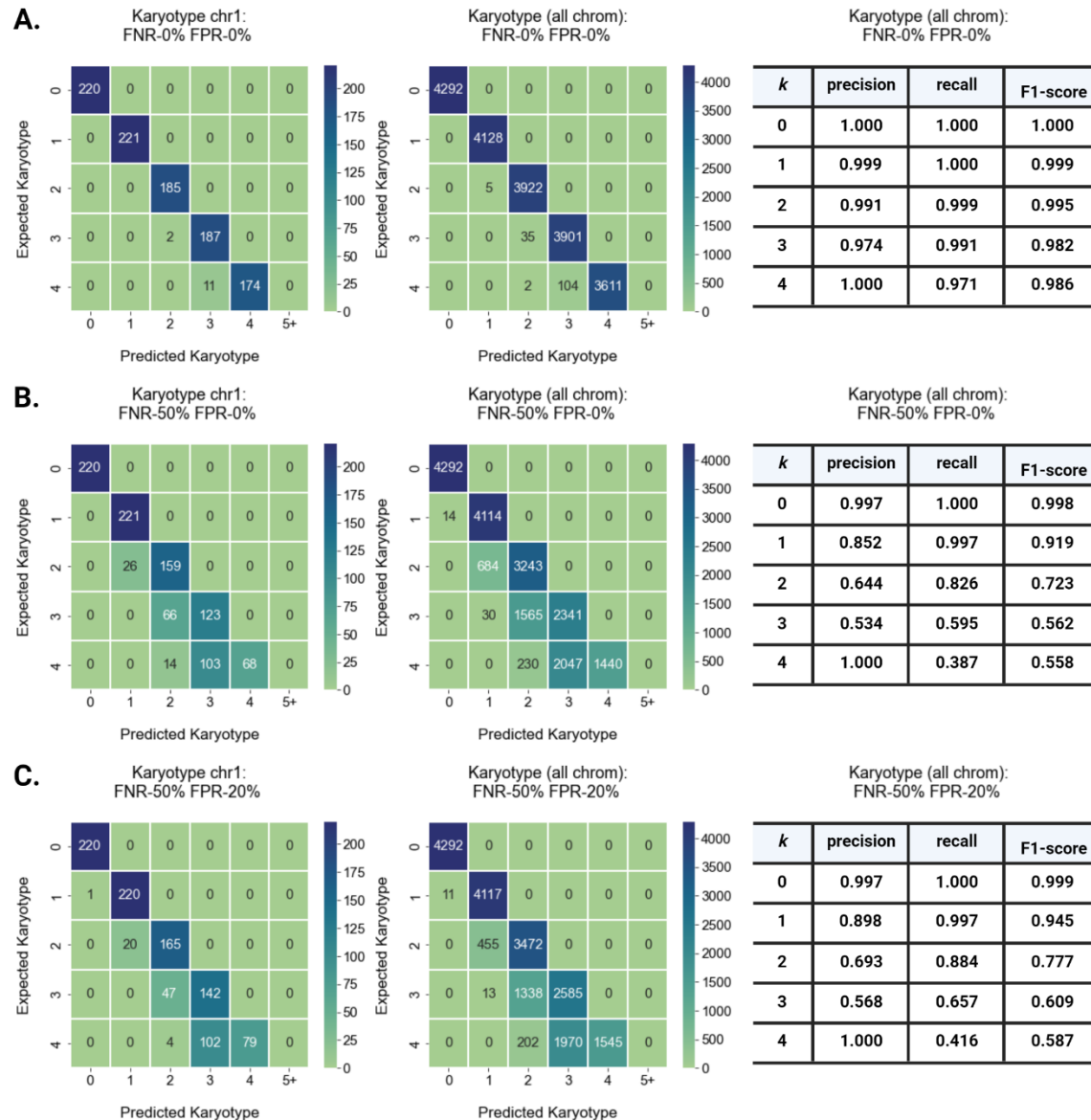


Figure S1.4: **Polymer fiber karyotyping on simulated aneuploid cells.** (A) Table of expected compared to predicted copy numbers, summarized for mouse chromosome 1 (left) and for all mouse chromosomes (middle), and its precision/recall statistics at 0% FNR 0% FPR. (B) Table of expected compared to predicted copy numbers, summarized for mouse chromosome 1 (left) and for all mouse chromosomes (middle), and its precision/recall statistics at 50% FNR 0% FPR. Table of expected compared to predicted copy numbers, summarized for mouse chromosome 1 (left) and for all mouse chromosomes (middle), and its precision/recall statistics at 50% FNR 20% FPR.

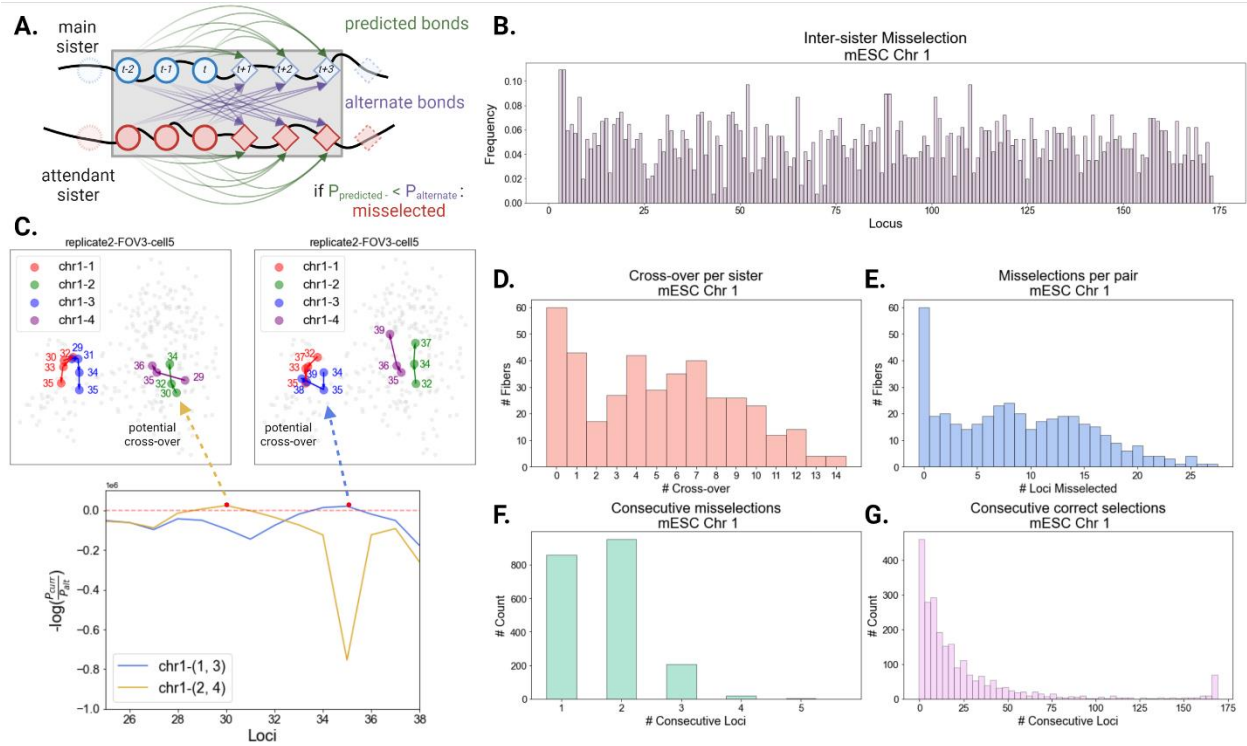


Figure S1.5: Assessment of misselection between tightly coupled sister chromatids. (A) Schematic of misselection assessment. For every tightly paired sister chromatid, a sliding window of three downstream loci are allowed to exchange with loci on the other sister. Bond probabilities between three upstream loci and three downstream loci are calculated. Original bond probabilities are compared to the cross-over probabilities to evaluate misselection. (B) Bar-plot of misselection frequency per locus on mESC chr 1. (C) Representative image of loci belonging to sister chromatids whose bond probabilities are more probable ($-\log(P_{\text{predicted}}/P_{\text{alternate}}) > 0$) if loci are exchanged between sisters. (D) Number of cross-over events, defined as sequences of misselections of any length, per sister chromatid. (E) Number of misselected loci per sister chromatid. (F) Histogram of length of consecutive misselections across tightly paired sister chromatids. (G). Histogram of length of consecutive correct selections across tightly paired sister chromatids.

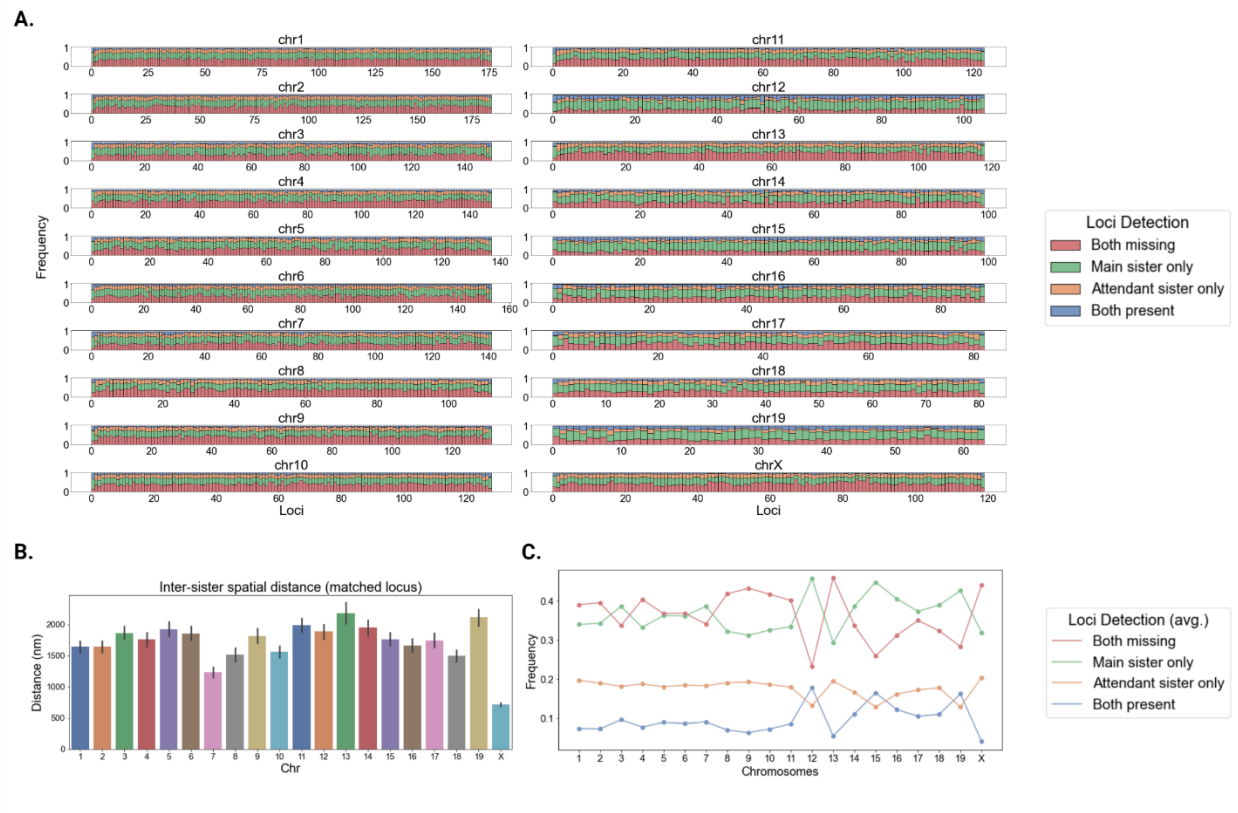


Figure S1.6: Spatial genome alignment locus detection in putative sister chromatids. (A) Frequency of loci detection on tightly coupled sister chromatids, stratified by detection on both sister chromatids, detection on main sister, detection on attendant sister, and failure of detection on either, for each of the mouse chromosomes. (B) Boxen plot (center line – 50%tile) of inter-sister spatial distance, for each matched locus, across all mouse chromosomes (n=1160 cells). (C) Detection rate averaged across all loci of a given chromosome, for each chromosome, stratified by detection on both sister chromatids, detection on main sister, detection on attendant sister, and failure of detection on either, for each of the mouse chromosomes.

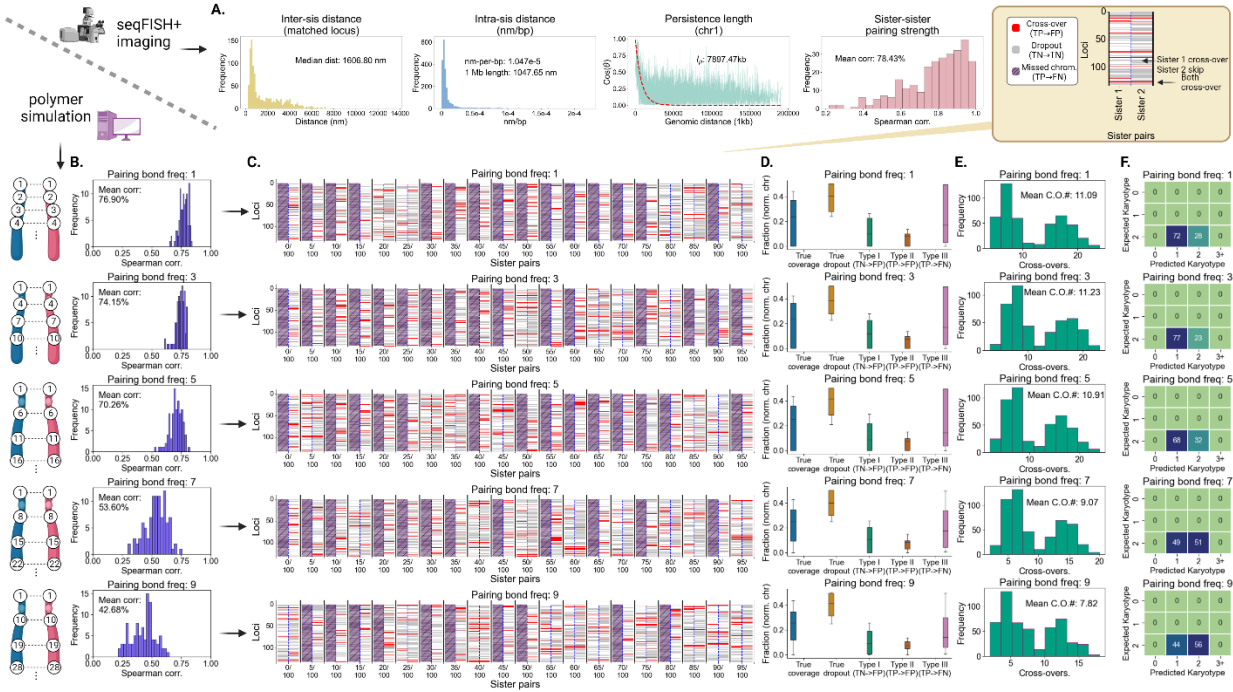


Figure S1.7: Spatial genome alignment on simulated paired chromatin fibers. (A) Summary of inter-sister locus distance, intra-sister locus distance, persistence length, and sister-chromatid pairing strength for DNA seqFISH+ imaging of mouse chromosome 1. (B) Spearman correlation plot of all 100 pairs of paired fibers, delineated by the frequency of pairing bonds. (C) Representative heatmap of twenty simulated pairs at different pairing strengths, traced via spatial genome alignment. Red delineates misselected loci (cross-over), while gray delineate loci dropped out as input to emulate 50% FNR 0% FPR. (D) Bar plot (center line – 50%tile; box – 25 to 50%tile; bar – 0 to 100%tile) of locus selection summarized for alignment on simulated pairs of different pairing strengths (n=100 pairs of fibers, per pairing strength), parsed into 5 categories of locus selection. (E) Frequency of cross-over during spatial genome alignment on simulated chromatin pairs of different pairing strengths. (F) Table of expected compared to predicted copy numbers for polymer fiber karyotyping on simulated chromatin pairs of different pairing strengths.

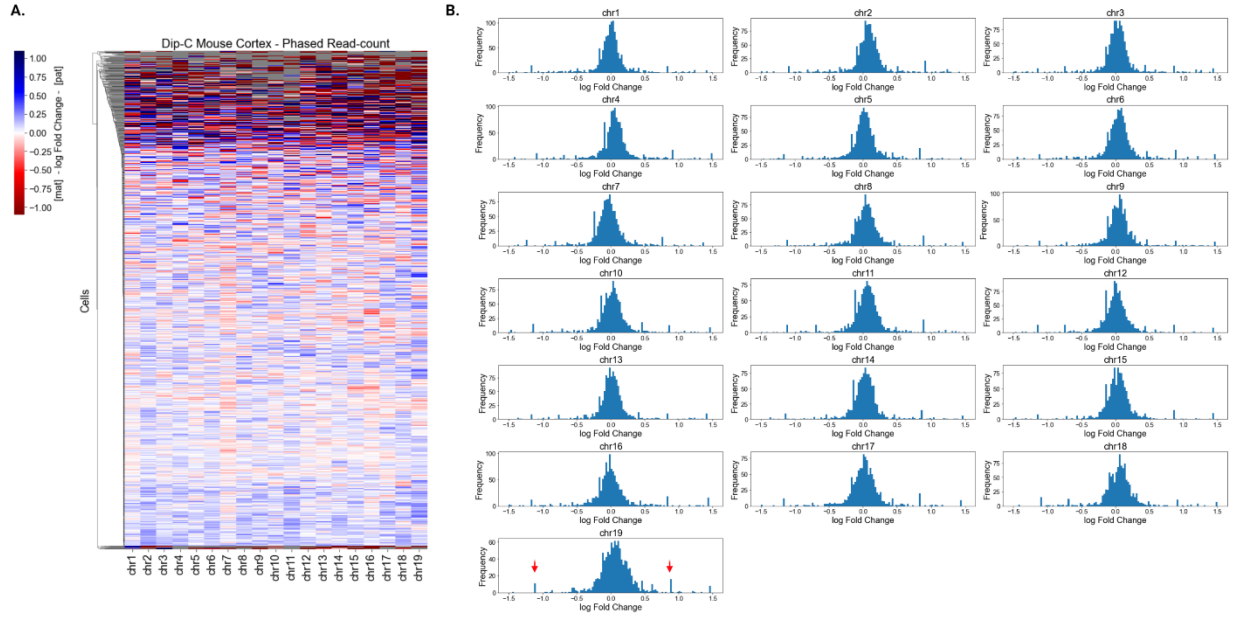


Figure S1.8: Haplotype-resolved single-cell sequencing of mouse cortical excitatory neurons. (A) Heatmap of log₂ fold change of haplotyped Dip-C reads assigned to the paternal vs maternal allele. (B) Histogram of log₂ fold change of haplotyped Dip-C reads. Most reads are balanced between the two haplotypes, while a distinct peak around 1 and -1 (red arrows) are present for every chromosome.

A.

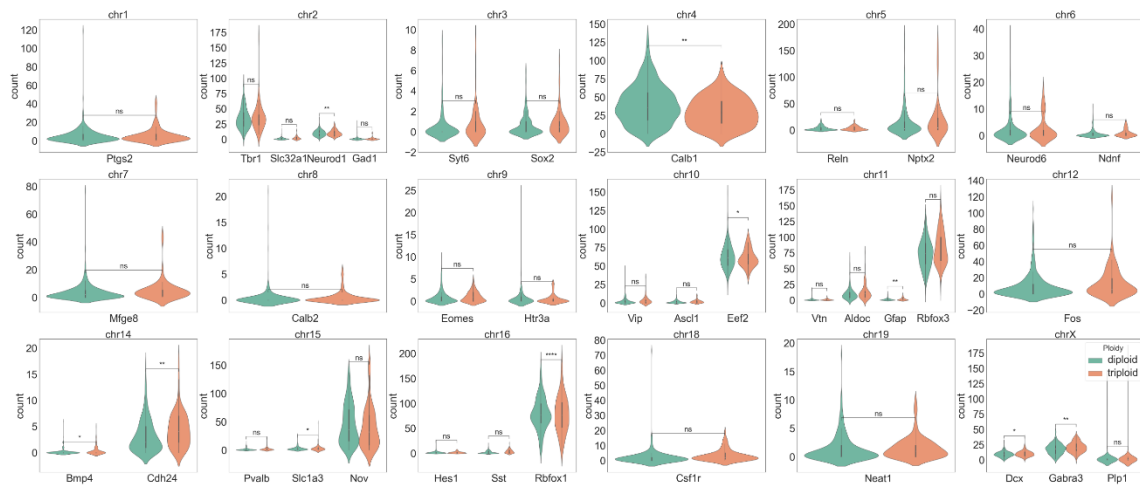


Figure S1.9: Effect of gene dosage on transcription in excitatory mouse cortical neurons. (A) Violin-plot (center dot – median; gray bar – quartiles) of RNA transcripts detected by seqFISH+ RNA imaging, corresponding to genes traced by seqFISH+ DNA imaging (excitatory neurons - n=458 cells across 3 experiments). RNA counts are stratified by gene dosage, delineated by mouse chromosome. Statistical significance and p-values evaluated by two-sample pairwise t-tests are labeled above, comparing 2N vs 3N nuclei of a given chromosome.

Polymer Fiber Karyotyping

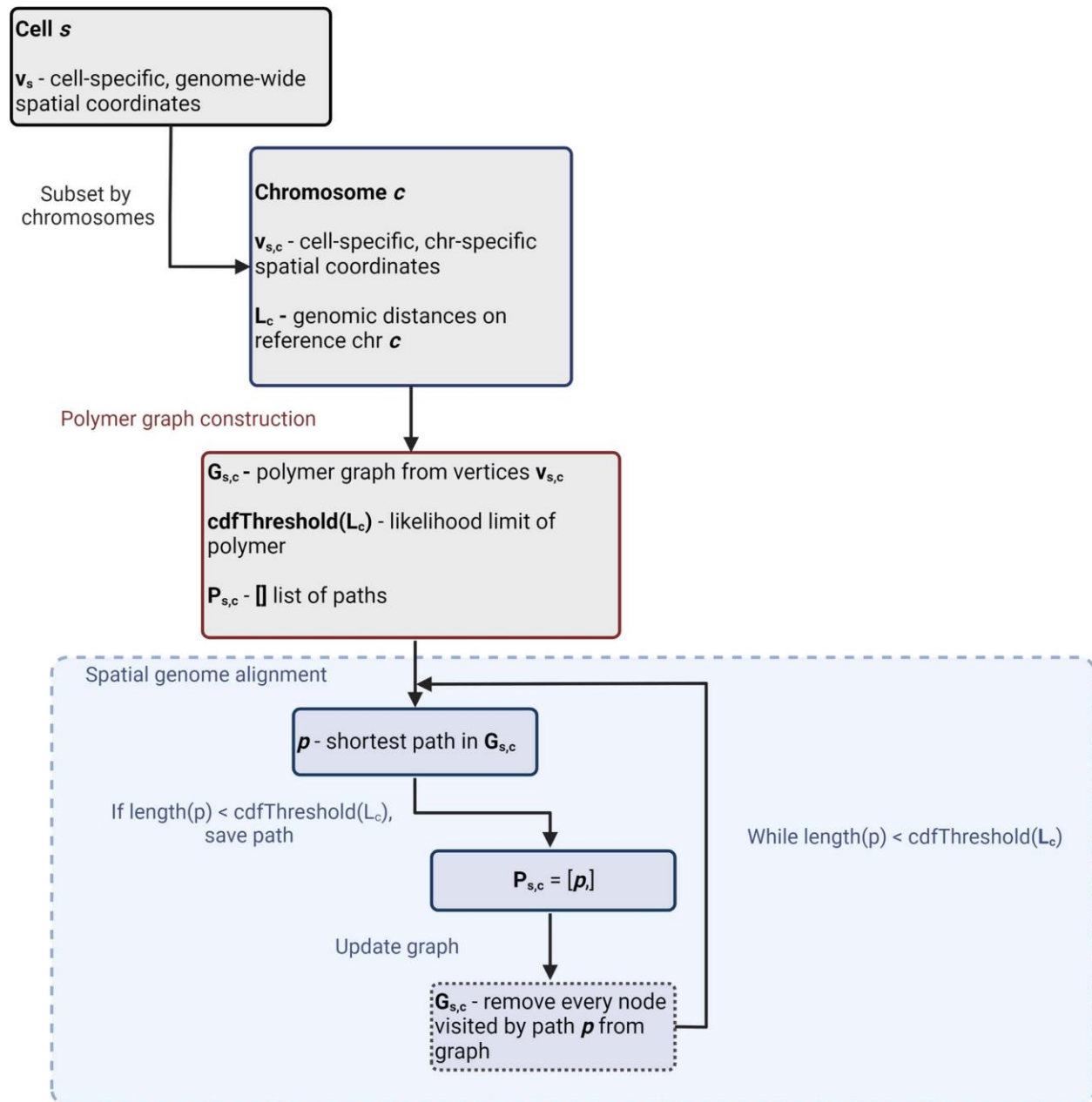


Figure S 1.10: **Polymer fiber karyotyping routine.** Workflow of iterative spatial genome alignment to uncover all physically likely fibers of a given chromosome in single cells. For any given cell, after subsetting detected loci by chromosomes, the aligner constructs a directed acyclic graph (DAG) from all signals belonging to a given chromosome. The aligner next finds the shortest path through this DAG using dynamic programming, where this path represents a putative polymer fiber and the path score its physical likelihood. For every putative fiber discovered from alignment, the aligner updates the DAG by subtracting all visited nodes while the discovered fiber has a likelihood score within threshold. This process is repeated until no polymer under threshold can be discovered, upon which the routine terminates alignment.

1.8 Acknowledgements

Chapter 1 is, in full, a reformatted reprint of the material as it appears as “A spatial genome aligner for resolving chromatin architectures from multiplexed DNA FISH” in *Nature Biotechnology*, 2023. Bojing Blair Jia, Adam Jussila, Colin Kern, Quan Zhu and Bing Ren. The dissertation author was the primary investigator and author of this paper.

We are grateful for B. Li for bioinformatic support. B.B.J. would like to thank P. Kosuri, V. Bafna, B. Ross, X. Wen, B. Bintu, X. Zhuang, Y. Han, Y. Xie, G. Mishne, H. Brandão, E. Mukamel, S. Zu, Y. Zhang, C. Zhu, Y.E. Li, P.B. Chen, D. Cunningham-Bryant, M. Frey, K. Zhang, and M. Hu for their invaluable feedback on the spatial genome aligner and polymer fiber karyotyping. The work was supported by NIH 4D Nucleome Common Fund project 1UM1HG011585 to B.R. B.B.J. by was supported, in part, by Ruth Kirschstein Institutional National Research Service Award T32 GM008666.

This study was conceived by B.R. and B.B.J. B.B.J. designed and implemented the spatial genome alignment and the polymer fiber karyotyping algorithm. B.B.J. analyzed chromatin tracing experiments. B.B.J., A.J., C.K. and Q.Z. benchmarked the spatial genome aligner. B.R. and B.B.J. wrote the manuscript.

1.9 References

1. Naumova, N., Imakaev, M., Fudenberg, G., Zhan, Y., Lajoie, B., Mirny, L. & Dekker, J. Organization of the Mitotic Chromosome. *Science* 342, 948-953 (2013).
2. Pope, B., Ryba, T., Dileep, V., Yue, F., Wu, W., Denas, O., Vera, D., Wang, Y., Hansen, R., Canfield, T., Thurman, R., Cheng, Y., Gülsoy, G., Dennis, J., Snyder, M., Stamatoyannopoulos, J., Taylor, J., Hardison, R., Kahveci, T., Ren, B. & Gilbert, D. Topologically associating domains are stable units of replication-timing regulation. *Nature* 515, 402-405 (2014).
3. Schuster-Böckler, B. & Lehner, B. Chromatin organization is a major influence on regional mutation rates in human cancer cells. *Nature* 488, 504-507 (2012).
4. Arnould, C., Rocher, V., Finoux, A., Clouaire, T., Li, K., Zhou, F., Caron, P., Mangeot, P., Ricci, E., Mourad, R., Haber, J., Noordermeer, D. & Legube, G. Loop extrusion as a mechanism for formation of DNA damage repair foci. *Nature* 590, 660-665 (2021).
5. Fulco, C., Nasser, J., Jones, T., Munson, G., Bergman, D., Subramanian, V., Grossman, S., Anyoha, R., Doughty, B., Patwardhan, T., Nguyen, T., Kane, M., Perez, E., Durand, N., Lareau, C., Stamenova, E., Aiden, E., Lander, E. & Engreitz, J. Activity-by-contact model of enhancer–promoter regulation from thousands of CRISPR perturbations. *Nature Genetics* 51, 1664-1669 (2019).
6. Cremer, T. & Cremer, C. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nature Reviews Genetics* 2, 292-301 (2001).
7. Lieberman-Aiden, E., van Berkum, N., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B., Sabo, P., Dorschner, M., Sandstrom, R., Bernstein, B., Bender, M., Groudine, M., Gnirke, A., Stamatoyannopoulos, J., Mirny, L., Lander, E. & Dekker, J. Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. *Science* 326, 289-293 (2009).
8. Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., Parrinello, H., Tanay, A. & Cavalli, G. Three-Dimensional Folding and Functional Organization Principles of the Drosophila Genome. *Cell* 148, 458-472 (2012).
9. Kalhor, R., Tjong, H., Jayathilaka, N., Alber, F. & Chen, L. Genome architectures revealed by tethered chromosome conformation capture and population-based modeling. *Nature Biotechnology* 30, 90-98 (2011).
10. Dixon, J., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J. and Ren, B., 2012. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, 485(7398), pp.376-380.
11. Nora, E., Lajoie, B., Schulz, E., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N., Meisig, J., Sedat, J., Gribnau, J., Barillot, E., Blüthgen, N., Dekker, J. and Heard, E., 2012. Spatial partitioning of the regulatory landscape of the X-inactivation

centre. *Nature*, 485(7398), pp.381-385.

12. Alipour, E. & Marko, J. Self-organization of domain structures by DNA-loop-extruding enzymes. *Nucleic Acids Research* 40, 11202-11212 (2012).
13. Fudenberg, G., Imakaev, M., Lu, C., Goloborodko, A., Abdennur, N. & Mirny, L. Formation of Chromosomal Domains by Loop Extrusion. *Cell Reports* 15, 2038-2049 (2016).
14. Goloborodko, A., Marko, J. & Mirny, L. Chromosome Compaction by Active Loop Extrusion. *Biophysical Journal* 110, 2162-2168 (2016).
15. Rao, S., Huntley, M., Durand, N., Stamenova, E., Bochkov, I., Robinson, J., Sanborn, A., Machol, I., Omer, A., Lander, E. & Aiden, E. A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell* 162, 687-688 (2015).
16. Davidson, I., Bauer, B., Goetz, D., Tang, W., Wutz, G. & Peters, J. DNA loop extrusion by human cohesin. *Science* 366, 1338-1345 (2019).
17. Jin, F., Li, Y., Dixon, J., Selvaraj, S., Ye, Z., Lee, A., Yen, C., Schmitt, A., Espinoza, C. & Ren, B. A high-resolution map of the three-dimensional chromatin interactome in human cells. *Nature* 503, 290-294 (2013).
18. Arnold, C., Gerlach, D., Stelzer, C., Boryń, Ł., Rath, M. & Stark, A. Genome-Wide Quantitative Enhancer Activity Maps Identified by STARR-seq. *Science* 339, 1074-1077 (2013).
19. Mifsud, B., Tavares-Cadete, F., Young, A., Sugar, R., Schoenfelder, S., Ferreira, L., Wingett, S., Andrews, S., Grey, W., Ewels, P., Herman, B., Happe, S., Higgs, A., LeProust, E., Follows, G., Fraser, P., Luscombe, N. & Osborne, C. Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. *Nature Genetics* 47, 598-606 (2015).
20. Lupiáñez, D., Kraft, K., Heinrich, V., Krawitz, P., Brancati, F., Klopocki, E., Horn, D., Kayserili, H., Opitz, J., Laxova, R., Santos-Simarro, F., Gilbert-Dussardier, B., Wittler, L., Borschiwer, M., Haas, S., Osterwalder, M., Franke, M., Timmermann, B., Hecht, J., Spielmann, M., Visel, A. & Mundlos, S. Disruptions of Topological Chromatin Domains Cause Pathogenic Rewiring of Gene-Enhancer Interactions. *Cell* 161, 1012-1025 (2015).
21. Morton, A., Dogan-Artun, N., Faber, Z., MacLeod, G., Bartels, C., Piazza, M., Allan, K., Mack, S., Wang, X., Gimple, R., Wu, Q., Rubin, B., Shetty, S., Angers, S., Dirks, P., Sallari, R., Lupien, M., Rich, J. & Scacheri, P. Functional Enhancers Shape Extrachromosomal Oncogene Amplifications. *Cell* 179, 1330-1341.e13 (2019).
22. Wu, S., Turner, K., Nguyen, N., Raviram, R., Erb, M., Santini, J., Luebeck, J., Rajkumar, U., Diao, Y., Li, B., Zhang, W., Jameson, N., Corces, M., Granja, J., Chen,

- X., Coruh, C., Abnoui, A., Houston, J., Ye, Z., Hu, R., Yu, M., Kim, H., Law, J., Verhaak, R., Hu, M., Furnari, F., Chang, H., Ren, B., Bafna, V. & Mischel, P. Circular ecDNA promotes accessible chromatin and high oncogene expression. *Nature* 575, 699-703 (2019).
23. Nora, E., Goloborodko, A., Valton, A., Gibcus, J., Uebersohn, A., Abdennur, N., Dekker, J., Mirny, L. & Bruneau, B. Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell* 169, 930-944.e22 (2017).
 24. Beliveau, B., Joyce, E., Apostolopoulos, N., Yilmaz, F., Fonseka, C., McCole, R., Chang, Y., Li, J., Senaratne, T., Williams, B., Rouillard, J. & Wu, C. Versatile design and synthesis platform for visualizing genomes with Oligopaint FISH probes. *Proceedings of the National Academy of Sciences* 109, 21301-21306 (2012).
 25. Beliveau, B., Boettiger, A., Avendaño, M., Jungmann, R., McCole, R., Joyce, E., Kim-Kiselak, C., Bantignies, F., Fonseka, C., Erceg, J., Hannan, M., Hoang, H., Colognori, D., Lee, J., Shih, W., Yin, P., Zhuang, X. & Wu, C. Single-molecule super-resolution imaging of chromosomes and in situ haplotype visualization using Oligopaint FISH probes. *Nature Communications* (2015).
 26. Wang, S., Su, J., Beliveau, B., Bintu, B., Moffitt, J., Wu, C. & Zhuang, X. Spatial organization of chromatin domains and compartments in single chromosomes. *Science* 353, 598-602 (2016).
 27. Boettiger, A., Bintu, B., Moffitt, J., Wang, S., Beliveau, B., Fudenberg, G., Imakaev, M., Mirny, L., Wu, C. & Zhuang, X. Super-resolution imaging reveals distinct chromatin folding for different epigenetic states. *Nature* 529, 418-422 (2016).
 28. Bintu, B., Mateo, L., Su, J., Sinnott-Armstrong, N., Parker, M., Kinrot, S., Yamaya, K., Boettiger, A. and Zhuang, X., 2018. Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. *Science*, 362(6413).
 29. Mateo, L., Murphy, S., Hafner, A., Cinquini, I., Walker, C. and Boettiger, A., 2019. Visualizing DNA folding and RNA in embryos at single-cell resolution. *Nature*, 568(7750), pp.49-54.
 30. Su, J., Zheng, P., Kinrot, S., Bintu, B. and Zhuang, X., 2020. Genome-Scale Imaging of the 3D Organization and Transcriptional Activity of Chromatin. *Cell*, 182(6), pp.1641-1659.e26.
 31. Takei, Y., Yun, J., Zheng, S., Ollikainen, N., Pierson, N., White, J., Shah, S., Thomassie, J., Suo, S., Eng, C., Guttman, M., Yuan, G. and Cai, L., 2021. Integrated spatial genomics reveals global architecture of single nuclei. *Nature*, 590(7845), pp.344-350.
 32. Takei, Y., Zheng, S., Yun, J., Shah, S., Pierson, N., White, J., Schindler, S., Tischbirek, C., Yuan, G. & Cai, L. Single-cell nuclear architecture across cell types in the mouse

brain. *Science* 374, 586-594 (2021).

33. Nguyen, H., Chatteraj, S., Castillo, D., Nguyen, S., Nir, G., Lioutas, A., Hershberg, E., Martins, N., Reginato, P., Hannan, M., Beliveau, B., Church, G., Daugharthy, E., Marti-Renom, M. and Wu, C., 2020. 3D mapping and accelerated super-resolution imaging of the human genome using in situ sequencing. *Nature Methods*, 17(8), pp.822-832.
34. Payne, A., Chiang, Z., Reginato, P., Mangiameli, S., Murray, E., Yao, C., Markoulaki, S., Earl, A., Labade, A., Jaenisch, R., Church, G., Boyden, E., Buenrostro, J. and Chen, F., 2021. In situ genome sequencing resolves DNA sequence and structure in intact biological samples. *Science*, 371(6532).
35. Shachar, S., Pegoraro, G. and Misteli, T., 2015. HIPMap: A High-Throughput Imaging Method for Mapping Spatial Gene Positions. *Cold Spring Harbor Symposia on Quantitative Biology*, 80, pp.73-81.
36. Shachar, S., Voss, T., Pegoraro, G., Sciascia, N. and Misteli, T., 2015. Identification of Gene Positioning Factors Using High-Throughput Imaging Mapping. *Cell*, 162(4), pp.911-923.
37. Cardozo Gizzi, A., Cattoni, D., Fiche, J., Espinola, S., Gurgo, J., Messina, O., Houbbron, C., Ogiyama, Y., Papadopoulos, G., Cavalli, G., Lagha, M. & Nollmann, M. Microscopy-Based Chromosome Conformation Capture Enables Simultaneous Visualization of Genome Organization and Transcription in Intact Organisms. *SSRN Electronic Journal* (2018). doi:10.2139/ssrn.3265728
38. Fields, B., Nguyen, S., Nir, G. & Kennedy, S. A multiplexed DNA FISH strategy for assessing genome architecture in *Caenorhabditis elegans*. *eLife* 8, (2019).
39. Nir, G., Farabella, I., Pérez Estrada, C., Ebeling, C., Beliveau, B., Sasaki, H., Lee, S., Nguyen, S., McCole, R., Chatteraj, S., Erceg, J., AlHaj Abed, J., Martins, N., Nguyen, H., Hannan, M., Russell, S., Durand, N., Rao, S., Kishi, J., Soler-Vila, P., Di Pierro, M., Onuchic, J., Callahan, S., Schreiner, J., Stuckey, J., Yin, P., Aiden, E., Marti-Renom, M. & Wu, C. Walking along chromosomes with super-resolution imaging, contact maps, and integrative modeling. *PLOS Genetics* 14, e1007872 (2018).
40. Rosin, L., Nguyen, S. & Joyce, E. Condensin II drives large-scale folding and spatial partitioning of interphase chromosomes in *Drosophila* nuclei. *PLOS Genetics* 14, e1007393 (2018).
41. Sawh, A., Shafer, M., Su, J., Zhuang, X., Wang, S. & Mango, S. Lamina-Dependent Stretching and Unconventional Chromosome Compartments in Early *C. elegans* Embryos. *Molecular Cell* 78, 96-111.e6 (2020).
42. Liu, M., Lu, Y., Yang, B., Chen, Y., Radda, J., Hu, M., Katz, S. & Wang, S. Multiplexed imaging of nucleome architectures in single cells of mammalian tissue. *Nature Communications* 11, (2020).
43. Cheng, Y., Liu, M., Hu, M. & Wang, S. TAD-like single-cell domain structures exist on

- both active and inactive X chromosomes and persist under epigenetic perturbations. *Genome Biology* 22, (2021).
44. Yamakawa, H. & Yoshizaki, T. Helical Wormlike Chains in Polymer Solutions. (2016). doi:10.1007/978-3-662-48716-7
 45. Ross, B. & Wiggins, P. Measuring chromosome conformation with degenerate labels. *Physical Review E* 86, (2012).
 46. Needleman, S. and Wunsch, C., 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of Molecular Biology*, 48(3), pp.443-453.
 47. Smith, T. and Waterman, M., 1981. Identification of common molecular subsequences. *Journal of Molecular Biology*, 147(1), pp.195-197.
 48. Rabiner, L. and Juang, B., 1986. An introduction to hidden Markov models. *IEEE ASSP Magazine*, 3(1), pp.4-16.
 49. Huang, H., Zhu, Q., Jussila, A., Han, Y., Bintu, B., Kern, C., Conte, M., Zhang, Y., Bianco, S., Chiariello, A., Yu, M., Hu, R., Tastemel, M., Juric, I., Hu, M., Nicodemi, M., Zhuang, X. and Ren, B., 2021. CTCF mediates dosage- and sequence-context-dependent transcriptional insulation by forming local chromatin domains. *Nature Genetics*, 53(7), pp.1064-1074.
 50. AlHaj Abed, J., Erceg, J., Goloborodko, A., Nguyen, S., McCole, R., Saylor, W., Fudenberg, G., Lajoie, B., Dekker, J., Mirny, L. and Wu, C., 2019. Highly structured homolog pairing reflects functional organization of the *Drosophila* genome. *Nature Communications*, 10(1).
 51. Erceg, J., AlHaj Abed, J., Goloborodko, A., Lajoie, B., Fudenberg, G., Abdennur, N., Imakaev, M., McCole, R., Nguyen, S., Saylor, W., Joyce, E., Senaratne, T., Hannan, M., Nir, G., Dekker, J., Mirny, L. and Wu, C., 2019. The genome-wide multi-layered architecture of chromosome pairing in early *Drosophila* embryos. *Nature Communications*, 10(1).
 52. Mitter, M., Gasser, C., Takacs, Z., Langer, C., Tang, W., Jessberger, G., Beales, C., Neuner, E., Ameres, S., Peters, J., Goloborodko, A., Micura, R. and Gerlich, D., 2020. Conformation of sister chromatids in the replicated human genome. *Nature*, 586(7827), pp.139-144.
 53. Oomen, M., Hedger, A., Watts, J. & Dekker, J. Detecting chromatin interactions between and along sister chromatids with SisterC. *Nature Methods* 17, 1002-1009 (2020).
 54. Heride, C., Ricoul, M., Kiêu, K., von Hase, J., Guillemot, V., Cremer, C., Dubrana, K. & Sabatier, L. Distance between homologous chromosomes results from chromosome positioning constraints. *Journal of Cell Science* 123, 4063-4075 (2010).

55. Joyce, E., Erceg, J. & Wu, C. Pairing and anti-pairing: a balancing act in the diploid genome. *Current Opinion in Genetics & Development* 37, 119-128 (2016).
56. McKee, B. Homologous pairing and chromosome dynamics in meiosis and mitosis. *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression* 1677, 165-180 (2004).
57. Esposito, M. Evidence that spontaneous mitotic recombination occurs at the two-strand stage. *Proceedings of the National Academy of Sciences* 75, 4436-4440 (1978).
58. Apte, M. & Meller, V. Homologue Pairing in Flies and Mammals: Gene Regulation When Two Are Involved. *Genetics Research International* 2012, 1-9 (2012).
59. Wu, C. & Morris, J. Transvection and other homology effects. *Current Opinion in Genetics & Development* 9, 237-246 (1999).
60. Stark, J. & Jasin, M. Extensive Loss of Heterozygosity Is Suppressed during Homologous Repair of Chromosomal Breaks. *Molecular and Cellular Biology* 23, 733-743 (2003).
61. Moynahan, M. & Jasin, M. Mitotic homologous recombination maintains genomic stability and suppresses tumorigenesis. *Nature Reviews Molecular Cell Biology* 11, 196-207 (2010).
62. Rong, Y. & Golic, K. The Homologous Chromosome Is an Effective Template for the Repair of Mitotic DNA Double-Strand Breaks in *Drosophila*. *Genetics* 165, 1831-1842 (2003).
63. LaFave, M. & Sekelsky, J. Mitotic Recombination: Why? When? How? Where?. *PLoS Genetics* 5, e1000411 (2009).
64. Joyce, E., Williams, B., Xie, T. & Wu, C. Identification of Genes That Promote or Antagonize Somatic Homolog Pairing Using a High-Throughput FISH-Based Screen. *PLoS Genetics* 8, e1002667 (2012).
65. Selvaraj, S., Dixon, J., Bansal, V. & Ren, B. Whole-genome haplotype reconstruction using proximity-ligation and shotgun sequencing. *Nature Biotechnology* 31, 1111-1118 (2013).
66. Arnoldus, E., Peters, A., Bots, G., Raap, A. & van der Ploeg, M. Somatic pairing of chromosome 1 centromeres in interphase nuclei of human cerebellum. *Human Genetics* 83, 231-234 (1989).
67. Luo, G., Santoro, I., McDaniel, L., Nishijima, I., Mills, M., Youssoufian, H., Vogel, H., Schultz, R. & Bradley, A. Cancer predisposition caused by elevated mitotic recombination in Bloom mice. *Nature Genetics* 26, 424-429 (2000).
68. Koeman, J., Russell, R., Tan, M., Petillo, D., Westphal, M., Koelzer, K., Metcalf, J.,

- Zhang, Z., Matsuda, D., Dykema, K., Houseman, H., Kort, E., Furge, L., Kahnoski, R., Richard, S., Vieillefond, A., Swiatek, P., Teh, B., Ohh, M. & Furge, K. Correction: Somatic Pairing of Chromosome 19 in Renal Oncocytoma Is Associated with Deregulated EGLN2-Mediated Oxygen-Sensing Response. *PLoS Genetics* 4, (2008).
69. Rehen, S., Yung, Y., McCreight, M., Kaushal, D., Yang, A., Almeida, B., Kingsbury, M., Cabral, K., McConnell, M., Anliker, B., Fontanoz, M. & Chun, J. Constitutional Aneuploidy in the Normal Human Brain. *Journal of Neuroscience* 25, 2176-2180 (2005).
 70. McConnell, M., Lindberg, M., Brennand, K., Piper, J., Voet, T., Cowing-Zitron, C., Shumilina, S., Lasken, R., Vermeesch, J., Hall, I. & Gage, F. Mosaic Copy Number Variation in Human Neurons. *Science* 342, 632-637 (2013).
 71. Cai, X., Evrony, G., Lehmann, H., Elhosary, P., Mehta, B., Poduri, A. & Walsh, C. Single-Cell, Genome-wide Sequencing Identifies Clonal Somatic Copy-Number Variation in the Human Brain. *Cell Reports* 8, 1280-1289 (2014).
 72. Knouse, K., Wu, J., Whittaker, C. & Amon, A. Single cell sequencing reveals low levels of aneuploidy across mammalian tissues. *Proceedings of the National Academy of Sciences* 111, 13409-13414 (2014).
 73. van den Bos, H., Spierings, D., Taudt, A., Bakker, B., Porubský, D., Falconer, E., Novoa, C., Halsema, N., Kazemier, H., Hoekstra-Wakker, K., Guryev, V., den Dunnen, W., Foijer, F., Colomé-Tatché, M., Boddeke, H. & Lansdorp, P. Single-cell whole genome sequencing reveals no evidence for common aneuploidy in normal and Alzheimer's disease neurons. *Genome Biology* 17, (2016).
 74. Tan, L., Ma, W., Wu, H., Zheng, Y., Xing, D., Chen, R., Li, X., Daley, N., Deisseroth, K. & Xie, X. Changes in genome architecture and transcriptional dynamics progress independently of sensory experience during post-natal brain development. *Cell* 184, 741-758.e17 (2021).
 75. Klein, K., Zhao, P., Lyu, X., Sasaki, T., Bartlett, D., Singh, A., Tasan, I., Zhang, M., Watts, L., Hiraga, S., Natsume, T., Zhou, X., Baslan, T., Leung, D., Kanemaki, M., Donaldson, A., Zhao, H., Dalton, S., Corces, V. & Gilbert, D. Replication timing maintains the global epigenetic state in human cells. *Science* 372, 371-378 (2021).
 76. Chess, A., Simon, I., Cedar, H. & Axel, R. Allelic inactivation regulates olfactory receptor gene expression. *Cell* 78, 823-834 (1994).
 77. Kitsberg, D., Selig, S., Brandels, M., Simon, I., Keshet, I., Driscoll, D., Nicholls, R. & Cedar, H. Allele-specific replication timing of imprinted gene regions. *Nature* 364, 459-463 (1993).
 78. Knoll, J., Cheng, S. & Lalande, M. Allele specificity of DNA replication timing in the Angelman/Prader-Willi syndrome imprinted chromosomal region. *Nature Genetics* 6, 41-46 (1994).

79. Blumenfeld, B., Masika, H., Farago, M., Yehuda, Y., Halaseh, L., Vardi, O., Rapoport, R., Levin-Klein, R., Cedar, H., Bergman, Y. & Simon, I. Chromosomal coordination and differential structure of asynchronous replicating regions. *Nature Communications* 12, (2021).
80. Winick-Ng, W., Kukalev, A., Harabula, I., Zea-Redondo, L., Szabó, D., Meijer, M., Serebreni, L., Zhang, Y., Bianco, S., Chiariello, A., Irastorza-Azcarate, I., Thieme, C., Sparks, T., Carvalho, S., Fiorillo, L., Musella, F., Irani, E., Torlai Triglia, E., Kolodziejczyk, A., Abentung, A., Apostolova, G., Paul, E., Franke, V., Kempfer, R., Akalin, A., Teichmann, S., Dechant, G., Ungless, M., Nicodemi, M., Welch, L., Castelo-Branco, G. & Pombo, A. Cell-type specialization is encoded by specific chromatin topologies. *Nature* 599, 684-691 (2021).
81. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. & Duchesnay, É. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12, 2825-2830 (2011).
82. Imakaev, M., Goloborodko, A., Fudenberg, G., and Brandao, H. mirnylab/polychrome: v0.1.0 [Internet] 2022. Available from <https://zenodo.org/record/3579473>
83. Jia, B.B. jie. Github. Available from <https://github.com/b2jia/jie> (2022)
84. Virtanen, P., Gommers, R., Oliphant, T., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S., Brett, M., Wilson, J., Millman, K., Mayorov, N., Nelson, A., Jones, E., Kern, R., Larson, E., Carey, C., Polat, İ., Feng, Y., Moore, E., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E., Harris, C., Archibald, A., Ribeiro, A., Pedregosa, F. & van Mulbregt, P. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods* 17, 352-352 (2020).

Chapter 2 PRINCIPLES OF SPATIAL MULTIOMICS: INSIGHTS AND APPLICATIONS TO DERMATOLOGY

2.1 Abstract

Molecular pathology, using high-throughput genomic and proteomic profiling, identifies precise disease targets but often loses valuable histology and spatial context. Emerging spatial multiomic technologies enable joint visualization of genomic, proteomic, and epigenomic targets within tissue. Initially focusing on RNA (spatial *transcriptomics*), spatial technologies can now concurrently capture DNA, genome accessibility, histone modifications, and proteins with single-cell resolution. This article outlines their principles, techniques, advantages, limitations, and potential for dermatologic research. By jointly profiling multiple molecular channels, spatial multiomics catalyze new studies such as copy number variations, clonal heterogeneity, and enhancer dysregulation, replete with spatial context, illuminating skin's complex heterogeneity.

2.2 Introduction

Tissues and organs are composed of organized networks of cells, where the location of each niche plays a pivotal role in physiology and homeostasis. The skin's laminar structure, with different layers performing contrasting and complementary functions such as pathogen defense, UV protection, sensory feedback, and injury repair, reflects its multifunctionality (Hsu et al., 2014). This multifunctionality is supported by a heterogeneity commensurate with the diverse roles it fulfills. Fibroblasts, melanocytes, and basal cells among other cell types have different transcriptional signatures and protein expressions that lend unique roles in meaningful spatial locales ¹⁻³. This cellular heterogeneity is further shaped by the non-coding genome and epigenetic influences, such as genome accessibility and histone modifications, which together orchestrate distinct responses in DNA repair, gene regulation and cell differentiation ⁴⁻⁶. In disease, disruptions in tissue architecture are detected through histology ⁷, while dysregulated molecular signatures are separately detected through genomics or proteomics ^{8,9}. Concurrent detection of spatial context and high-throughput molecular profiling has been a long standing challenge. An ideal tool for cutaneous research should capture both cellular spatial context and the diverse gene regulatory programs at high throughput.

Single-cell sequencing has unveiled the rich diversity of cell types within tissues by dissociating tissue into an emulsion of cells, each encapsulated in a droplet and uniquely barcoded to identify its DNA and/or RNA ^{10,11}. This technique reveals cell heterogeneity but loses invaluable spatial information like cell-cell interactions and clonal margins in tumors ¹²⁻¹⁵. In its wake, spatial technologies have been developed to capture both cellular heterogeneity and their spatial arrangement. These technologies build on single-cell sequencing by incorporating

fluorescence imaging principles, enabling the spatial localization of DNA, RNA, and proteins within single cells ^{12,13,16}. Imaging confers unique advantages. While RNA exists in multiple copies, making it easier to detect, DNA's presence as only two copies in a diploid cell poses detection challenges for sequencing that imaging can overcome with its single-molecule sensitivity ¹⁷. Furthermore, sequencing must often compromise detection quality in multiomics to accommodate distinct chemistries for RNA and DNA, whereas fluorescence imaging employs a unified chemistry for multiomic detection ^{16,18,19}. This allows for simultaneous visualization of DNA, RNA, and even histone modifications of DNA loci within single cells of complex tissues ^{20–22}. With these advantages, spatial technologies are poised to advance multiomics, particularly in studying spatially stereotyped tissues like the skin where structure informs function. This article will delve into how spatial technologies function, illustrate their advantages and limitations, their methodology in capturing multiomic information, and their potential impact on understanding cutaneous biology.

2.3 Principles of Spatial technologies

Spatial technologies have undergone rapid advancement, giving rise to a wide array of methodologies^{12,13}. Amid such rapid progress, this also presents a challenge in understanding the fundamental principles, capabilities and distinguishing features of each technology. Here we summarize three key axis that underpin the understanding and differentiation of spatial -omic technologies: (i) how they record the spatial positions of single cells through **positional encoding**; (ii) how they scale to detect thousands of targets (ie. all genes in the transcriptome) through **multiplexing**; and (iii) how they manage the simultaneous capture of multiple channels of information, such as the genome, transcriptome, epigenome, and proteome, collectively called **multiomics**.

Positional encoding

Positional encoding is often the first critical step in spatial -omics technologies, as it enables the precise and indispensable localization of a cell within a tissue cross section. One method involves direct imaging of molecular targets inside cells, a process known as 'Image in situ' (Figure 1). In this approach, tissue specimens are embedded in hydrogel to maintain the spatial integrity of cells and their internal components²¹⁻²⁹. This embedding crosslink nucleic acids and proteins to the gel, enabling the tissue to endure intensive washing while removing light-blocking elements like lipids. The hydrogel-treated tissue allows for the penetration of fluorescent probes, which, upon laser excitation, illuminate as bright dots, indicating the location of single molecules of nucleic acids and proteins with nanometer super-resolution^{21,23-27,29}. Simultaneous imaging of cell membranes ensures that gene and protein counts and identities are accurately assigned to their precise locations within cells, achieving true single-cell resolution.

Release-and-capture technologies offer an alternative yet innovative approach to positional encoding in spatial genomics, referred to as 'Release & Capture' (Figure 1). This strategy differs from direct imaging by placing cells on a DNA barcoded array, each array position designed to be smaller than a single cell ^{30–38}. The spatial location and DNA barcode of each array position are pre-recorded, forming a detailed map that links DNA sequence barcodes to their respective spatial coordinates. Upon cell permeabilization, internal contents are released, trickling down and becoming captured by the underlying array. The nucleic acids from each cell adhere to and incorporate the unique DNA barcodes, thus achieving spatial indexing. Crucially, after the cells are digested, their nucleic acids are pooled for Next Generation Sequencing (NGS), allowing for the comprehensive analysis of each cell's transcriptome and/or genome. Despite pooling the cells contents together, this technique maintains the spatial context of each sequence, by tracing sequences back to its original location in the tissue via the corresponding barcode and pre-recorded map position.

A third approach for positional encoding in spatial -omics is the 'Region-of-interest' method (Figure 1), which involves selecting specific microscopic tissue areas for downstream sequencing. This includes techniques such as Laser Capture Microdissection, which uses a UV or infrared laser to precisely dissect a cluster of cells or even a single cell under a microscope ¹³. These isolated clusters are then separately sequenced, providing a basic level of spatial encoding. More recently developed in this group are technologies employing Digital Mirror Devices (DMDs), commonly found in theater projectors ^{39–41}. DMDs consist of arrays of microscopic mirrors, each individually adjustable, allowing for the selective illumination of specific cell

patterns with near single-cell resolution. In techniques such as Light-seq, Zipseq or Nanostring GeoMx, cells illuminated are marked with a photoactivatable dye or sequence labels^{39–41}. During sequencing, these labels help in decoding and assigning cells back to their original locations within the tissue, thus achieving positional encoding.

Multiplexing

Following positional encoding, a critical next step is the unique labeling of each molecular target through a process called multiplexing. This is essential for distinguishing thousands of different targets, such as the entire range of genes in the genome, within single cells.

Subtypes within 'imaging *in situ*' spatial technologies primarily differ in their multiplexing approach, a necessity given the limited range of spectrally distinct fluorophores available for distinguishing a vast array of genetic targets (Figure 1).

Fluorescence In-Situ Hybridization (FISH) employs a unique method of temporal barcoding^{25,42–45}. Unlike methods that read nucleotide sequences directly, FISH uses multiple short DNA sequences called oligonucleotides to densely tile across a single gene^{46–48}. This approach does not decode the genetic sequence nucleotide by nucleotide. Instead, it relies on the combined fluorescence from these multiple oligonucleotides, which light up together and are interpreted as 'on' or 'off' signals over time. Techniques like MERFISH and seqFISH+ use this method to generate distinct temporal patterns, which are then analyzed across a series of images

taken at different timepoints^{24,42}. The identity of genomic or proteomic targets is deduced from these patterns, similar to reading a traditional black-and-white barcode.

In contrast, Sequencing-By-Ligation (SBL) directly deciphers the nucleotide sequence. Each of the four nucleotides – A, C, T, and G – is associated with a unique color, allowing the sequential identification of nucleotides in a target sequence through a series of images^{23,26,27,49,50}. This method offers a nucleotide-by-nucleotide readout, capturing the precise genetic sequence of targets within a cell. Techniques such as FISSEQ, ExSeq, and STARmap employ SBL for a variety of applications, including detailed chromosome structure analysis^{26,50} and transcriptomic mapping^{23,49}.

Release-and-capture methods in spatial genomics offer a distinct multiplexing approach from ‘imaging *in situ*’, leveraging traditional Next Generation Sequencing (NGS) chemistry to decode target sequences^{30–38,51}. The power of NGS chemistry is its broad capture of the whole genome and transcriptome, a multiplexing capacity unparalleled among spatial technologies^{12,13}. Unlike *in situ* methods where individual molecules are imaged and counted, release-and-capture can discriminate sequences but faces the challenge of quantifying copy numbers without direct visualization. To overcome this, these methods utilize Unique Molecular Identifiers (UMIs), short DNA sequences that distinctively tag each single RNA or DNA molecule regardless of their identity. In bead-based systems like Slide-seq and High-Definition Spatial Transcriptomics (HDST), thousands of DNA sequences anchored on beads capture nucleic acids released from cells^{35,38}. Each bead's barcode denotes its location while each captured molecule is tagged with a distinct UMI, allowing quantitation of multiple gene copies. Grid-based arrays, like DBiT-seq,

expose cells to positional encoding sequences in one orientation to create a pattern of stripes, followed by a 90-degree rotation to apply another layer of stripes, resulting a cross-hatched grid³⁴. During this process, UMIs are co-integrated with these cross-hatch positional barcodes into the cell's nucleic. Similarly, DNA array-based methods such as Spatial Transcriptomics (Visium), Stereo-seq, Pixel-seq, and Seq-scope attach UMIs during spatial barcode attachment, using a “lawn” of spatial barcodes and UMIs^{30–32,36,37}.

Comparative advantages and weaknesses

Imaging *in situ* and release-and-capture techniques represent the most popular of spatial -omic technologies to date. Their contrasting approaches to acquiring spatial information lends contrastive advantages and weaknesses, rendering them each more suitable for different research objectives (Figure 2).

For one, imaging *in situ* methods have high detection efficiency and single-molecule sensitivity^{42,45}. Because they detect targets directly inside intact cells, imaging *in situ* assigns detected genes unambiguously to their cells of origin and achieves true single-cell resolution. Release-and-capture methods, in contrast, is inherently lossy as some cell contents released may be lost during cell lysis^{30–32,37} (Figure 2). Additionally, some cell contents may diffuse laterally during release and be incorrectly captured by neighboring spatial barcodes in a process called lateral drift^{30,34,37} (Figure 2). Indeed a number of release-and-capture works detect transcripts in areas devoid of tissue, indicating significant lateral drift.

Another caveat of release-and-capture is that its array boundaries often lack natural alignment with cell borders, causing cell signals to overlap on beads or grid positions^{30,32,35–37}. This results in the loss of true single-cell resolution (Figure 2). In an attempt to compensate, release-and-capture methods use computational algorithms to try and separate mixed sequences from multiple cells sharing the same capture point, albeit with some error^{35,52–55}. Imaging *in situ* methods therefore excel in applications requiring accurate copy number determination and single-cell resolution, including karyotyping and studying copy number variations^{17,26}. Indeed, the majority of spatial -omics studies studying DNA – present as usually two copies in a diploid cell instead of the thousands of copies for RNA – has been imaging *in situ* methods^{20–22,50,56–61}. Because imaging *in situ* methods also has super-resolution, it has been able to study differentially expressed genes at subcellular locales, such as neuronal spines and dendrites^{23,24,28}. It has further been used to trace chromatin architectures inside nuclei with nanometer resolution, to study the grammar of DNA folding on gene regulation^{18,19,62,63}.

Release-and-capture's design necessitates thin tissue sections (~10 µm) to allow cellular content to percolate down and react with the barcoded array^{30–32,34,35,37,38,51} (Figure 2). Some reports have noted that cells in the thicker upper layers often escape detection, suggesting a bias for nucleic acids in immediate array contact³². Notably, release-and-capture methods records only the planar location (x, y-axis), omitting the vertical origin (z-axis) of molecules¹². In contrast, imaging *in situ* methods report successful detection of transcripts in as thick as 150 µm²⁷. These imaging *in situ* techniques enable the full 3-D distribution of RNA and DNA loci to be visualized inside intact cells.

Where release-and-capture methods suffer from lateral drift and lack of true single-cell resolution, they excel in the robust capture of the whole genome and transcriptome. Adapting robust and proven Next Generation Sequencing chemistry, release-and-capture methods are able to identify an extensive range of mRNA and DNA species, approaching whole genome or transcriptome coverage^{12,13} (Figure 2). The broad detection of genomic targets allows detailed transcriptomic and genomic profiling, enabling studies such as cell-ligand interactions⁶⁴ and detecting clonal heterogeneity in tumors⁶⁵. Furthermore, their sequence-reading capability makes them sensitive to isoforms and mutations, facilitating the study of alternative splicing in mRNA²³, haplotype phasing and detecting SNPs in DNA²⁶. Conversely, imaging *in situ* methods, while offering precise localization, face limitations in target diversity due to optical signal crowding^{12,13}; the emitted signals from densely labeled molecules can overlap in adjacent voxels, making it difficult to distinguish individual targets when scaling to tens of thousands of genes.

An important limitation of imaging *in situ* methods is that they are constrained by the necessity of designing fluorescent probes for known targets, which, due to factors like optical crowding, limits the selection to a few hundred to a few thousand targets^{12,13} (Figure 2). This selection requires prior knowledge such as cell-type markers from single-cell sequencing of the same tissue¹². This further precludes a comprehensive and unbiased genomic and transcriptomic profiling. Additionally, an extensive number of fluorescent oligos is often needed to amplify signals for visibility multiplies the target library size, requiring large oligo pools numbering tens to hundreds of thousands of sequences^{20–22,24,28} (Figure 2). Release-and-capture technologies are not subject to these scaling limitations, allowing broader genomic profiling. This scalability has

made release-and-capture methods particularly accessible for recent cutaneous research applications, such as profiling cell type composition changes in keloid formation ⁶⁶, detecting perturbations in immune signaling at leading edges of melanoma ⁶⁷, and analyzing tumor responses to checkpoint blockade therapy ⁶⁸.

2.4 Multi-omic technologies

Gene regulation, in health and disease, is orchestrated by multifactorial influences. They range from chromatin accessibility shaping transcription factor binding ⁶⁹, histone modifications influencing gene activation ⁵, and chromatin architecture aligning distal regulatory elements such as enhancers ⁷⁰. Gene transcription and ultimately protein expression, reflecting these regulatory activities, are critical for identifying disease states and cell types ⁷¹. A comprehensive understanding of cutaneous biology should ideally integrate this dynamic landscape: genomics, epigenomics, transcriptomics, and proteomics. Historically, joint profiling has been difficult and separate datasets, each employing a different -omic channel, are often computationally integrated ⁷². Spatial multiomics advances this integration, allowing simultaneous profiling of these aspects in their spatial context, creating great potential for dermatopathology and related fields.

Recent efforts to capture epigenomic information such as chromatin accessibility borrow principles of Assay for Transposable-Accessible Chromatin sequencing (ATAC-seq). Spatial ATAC-seq and more recently Spatial ATAC-RNA-seq, both based on DBiT-seq, employs a transposase enzyme (Tn5) to insert barcodes into accessible genome regions ^{73,74}. These barcodes are later decoded through DNA sequencing. Spatial ATAC-RNA-seq simultaneously captures mRNAs by targeting their poly-A tails and reverse transcribing them into DNA, converting them into the same substrate as the DNA sequences collected for chromatin accessibility. These mRNAs are distinguished from true chromatin DNA by a poly-T sequence, allowing joint profiling.

In another application of DBiT-seq, the transposase targets not only accessible regions but also specific histone marks. By tethering to antibodies against one histone modification – either H3K27me3, H3K27ac, or H3K4me3 – the transposase inserts barcodes into chromatin regions marked by these modifications ⁷⁴. As with previous methods, mRNAs are captured by targeting poly-A sequences. This technique, known as Spatial Cut&Tag, or Spatial Cut&Tag-RNA when jointly profiled with RNA, profiles potential enhancers and regulatory elements with the potential of capturing transcriptional data, offering detailed insights into the interplay between epigenomics and gene regulation^{74,75}. A similar approach of localizing transposase to specific histone marks have been adapted for imaging *in situ* methods, in the form of epigenomic MERFISH ⁷⁶.

Instead of localizing an enzyme to histone marks, imaging *in situ* approaches directly label such histone modifications with fluorescent antibodies. In seqFISH+, these histone marks appear as “globules”, decorating sharp puncta reporting the positions of mRNA and DNA ^{21,22}. Because each histone mark can be assigned a different barcode, as many as 17 different modifications have been simultaneously captured along with DNA and RNA positions inside mouse embryonic stem cells (mESCs) and mouse cortical neurons ^{21,22}.

Joint profiling of DNA and RNA has proven more straightforward than joint profiling of nucleic acids and proteins, which are fundamentally different mediums. In order to incorporate co-detection of proteins along with RNA, techniques such as CODEX, SM-Omics, SPOTs, and Nanostring *GeoMx* attach oligo-conjugated antibodies called antibody-derived DNA tags (ADTs) to target specific proteins ^{38,41,77–80} (Figure 3). These ADTs transform the protein

detection challenge into a sequence detection task amenable to Next Generation Sequencing, allowing the quantification of protein expression within cells upon oligo cleavage. This sequence-based approach offers a stark advantage over the labor-intensive serial immunofluorescence (ie. 4i, IBEX, CyCIF), which requires time-consuming cycles of antibody stripping and reapplication to image multiple targets^{81–83}. ADTs, by converting protein detection into a sequencing-compatible format, significantly streamline multiplexing. Current state-of-the-art joint profiling of RNA and proteins have profiled more than 160 different proteins⁷⁹.

2.5 Future Directions and Applications of Spatial Multiomics to Dermatologic Research

With emerging capabilities of probing multiple channels of -omic information with spatial context, we believe spatial multiomics has the potential to make profound impact on skin research. We particularly believe it will unlock new insights underlying discoveries made with NGS, providing integrative molecular information and spatial context where none was possible before.

Recent advances in artificial intelligence have accelerated tremendous progress in digital pathology ⁸⁴. Deep learning vision models now report dermatologist-level accuracy in classifying skin cancers ⁸⁵. These models are trained on standard H&E stained tissue specimens, with recent incorporation of molecular biomarkers through immunohistochemistry ⁸⁴. While morphometric features such as cell size, shape, and color inform disease staging and classification, these are decidedly nonspecific and lack a direct mechanistic connection to underlying disease. These deep learning models could therefore greatly benefit from spatial multiomic data, which offers a richer visualization of genomic, transcriptomic, and proteomic information. This data is not only more precise but often actionable, identifying potential druggable targets that transcend the limited informational scope of morphometric features. For instance, the visualization of oncohistone distribution within a tumor with spatial CUT&Tag, or copy number variations of a specific oncogene (ie. EGFR) in the form of extrachromosomal amplifications detected through DNA-MERFISH, may reform the way we stage and classify tumors.

Spatial multiomics also lends itself to the study of intrinsic mosaicism present in the skin, as well as genetic skin diseases. For one, spatial multiomics may help elucidate the genetic and

epigenetic basis of mosaic pattern formation, from Blaschko lines to Phylloid hypomelanosis^{86,87}. In as much as our genome has been sequenced, recent studies of the skin has revealed incredible “patchwork” of heterogeneity among cellular DNA. They found hundreds of competing clones within a few square millimeters of skin each bearing different mutational burdens⁹. This clonal architecture further differ across ancestry, and engender different susceptibilities to skin cancer, potentially through different inflammatory responses⁸⁸. Identifying *spatial* clones and quantitating their mutational burden through spatial DNA sequencing, while characterizing cell-type and cell-ligand interactions with joint RNA and proteomic cell surface receptor profiling, may yield key mechanisms underlying different skin cancer susceptibilities. Spatial multiomics may also help narrow down potential genetic targets of mosaic diseases⁸⁷. As the technology continue to scale and capture larger surface area in the foreseeable future, it maybe possible to assay both an affected and an adjacent but unaffected region of skin in one field of view. With this we can ask: what cell types and mutations predominate in affected niches? What cell-ligand interactions are lost or gained in the context of disease? What potential non-coding regions such as enhancers may become dysregulated and even hijacked within lesions (Figure 4)? The spatial and multiomic nature of these technologies naturally drive such comparative studies that help filter and identify specific signatures underlying the disease which by signature have spatial arrangements. Recent advances in NGS has enabled rapid sparse sequencing and sequencing analysis less than an hour after resection during surgery, and has proven to become a potential real-time tumor classification tool during surgery⁸⁹. By analogy, we are excited by the prospect where with increasing speed, spatial multiomics can directly empower physicians to visualize molecular markers *in situ* to change their management or operation.

Spatial technologies mark an exciting era for tissue research with important implications for digital pathology, including dermatopathology. In the context of the skin, they may help characterize the remarkable complexity that gives it its multifunctionality, potentially reform disease grading and classification with multiomic biomarker information, and in success, become a foundational tool for understanding cutaneous biology.

2.6 Figures

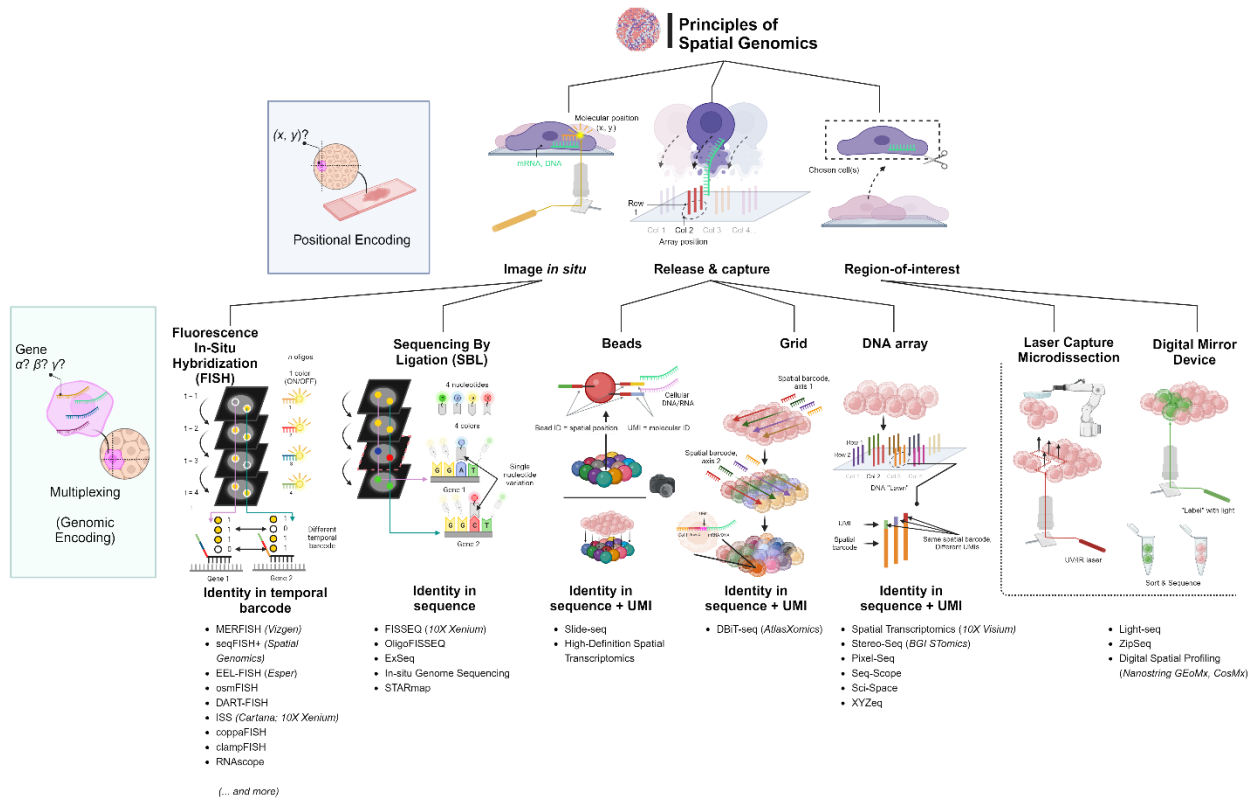


Figure 2.1: Classification and illustration of mechanism among families of spatial genomic technologies.

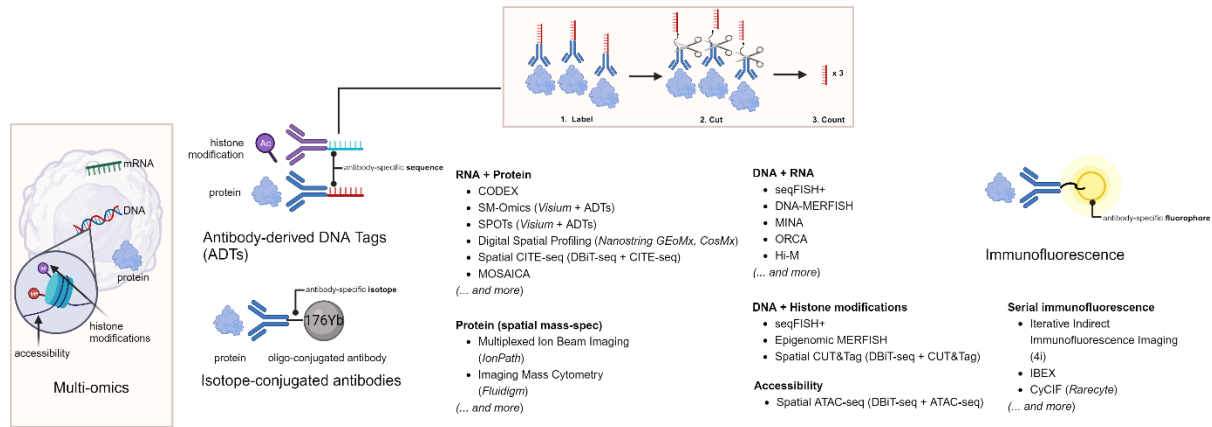


Figure 2.3: Strategies for multiomics utilized by spatial technologies.

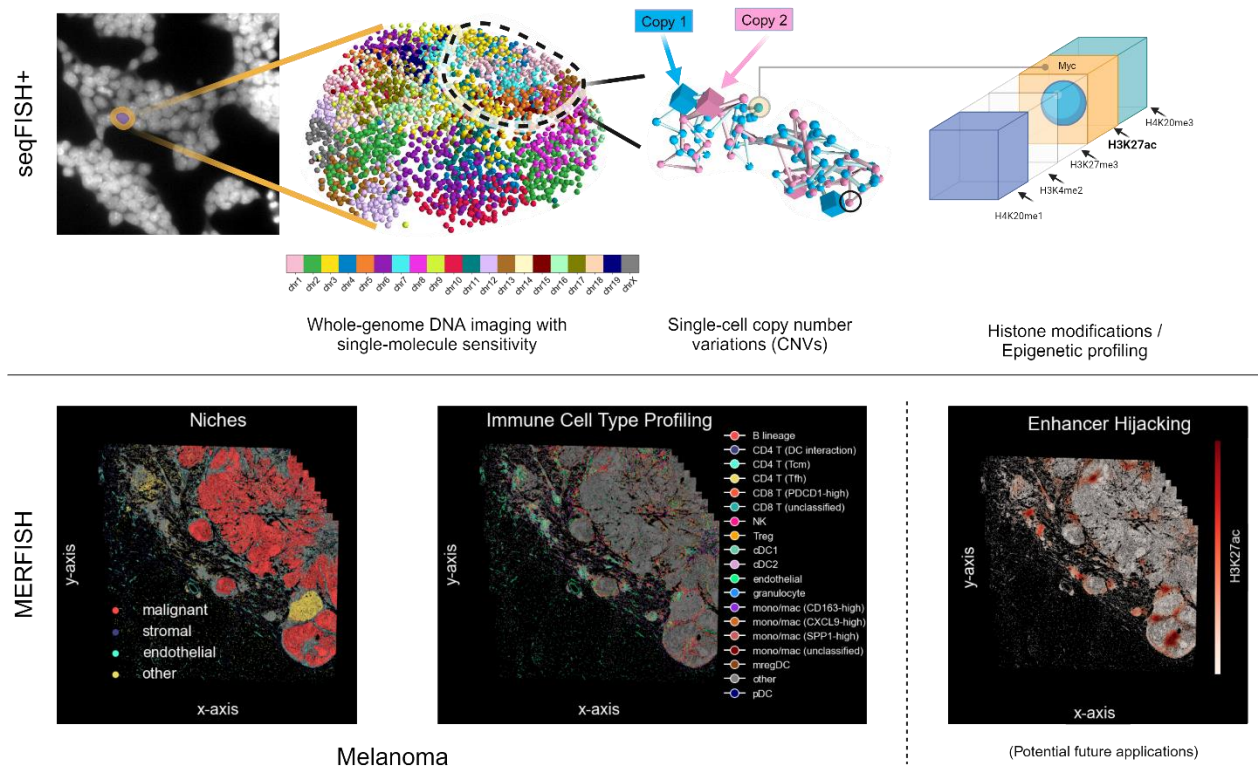


Figure 2.4: Potential future applications of spatial multiomics in cutaneous research.

2.7 Acknowledgements

Chapter 2 is, in part, based on material currently being prepared for submission as “Principles of Spatial Multi-omics: Insights and Applications to Dermatology”. Bojing Blair Jia, Brian Sun, Ernest Lee, and Bing Ren. The dissertation author was the primary investigator and author of this paper.

2.8 References

1. Dyring-Andersen, B., Løvendorf, M. B., Coscia, F., Santos, A., Møller, L. B. P., Colaço, A. R., Niu, L., Bzorek, M., Doll, S., Andersen, J. L., Clark, R. A., Skov, L., Teunissen, M. B. M. & Mann, M. Spatially and cell-type resolved quantitative proteomic atlas of healthy human skin. *Nature Communications* 11, 5587 (2020).
2. Wang, S., Drummond, M. L., Guerrero-Juarez, C. F., Tarapore, E., MacLean, A. L., Stabell, A. R., Wu, S. C., Gutierrez, G., That, B. T., Benavente, C. A., Nie, Q. & Atwood, S. X. Single cell transcriptomics of human epidermis identifies basal stem cell transition states. *Nature Communications* 11, 4239 (2020).
3. Zou, Z., Long, X., Zhao, Q., Zheng, Y., Song, M., Ma, S., Jing, Y., Wang, S., He, Y., Esteban, C. R., Yu, N., Huang, J., Chan, P., Chen, T., Izpisua Belmonte, J. C., Zhang, W., Qu, J. & Liu, G.-H. A Single-Cell Transcriptomic Atlas of Human Skin Aging. *Developmental Cell* 56, 383-397.e8 (2021).
4. Naik, S., Larsen, S. B., Gomez, N. C., Alaverdyan, K., Sendoel, A., Yuan, S., Polak, L., Kulukian, A., Chai, S. & Fuchs, E. Inflammatory memory sensitizes skin epithelial stem cells to tissue damage. *Nature* 550, 475–480 (2017).
5. Strahl, B. D. & Allis, C. D. The language of covalent histone modifications. *Nature* 403, 41–45 (2000).
6. Tessarz, P. & Kouzarides, T. Histone core modifications regulating nucleosome structure and dynamics. *Nature Reviews Molecular Cell Biology* 15, 703–708 (2014).
7. Janowczyk, A. & Madabhushi, A. Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases. *Journal of Pathology Informatics* 7, 29 (2016).
8. Betancourt, L. H., Gil, J., Kim, Y., Doma, V., Çakır, U., Sanchez, A., Murillo, J. R., Kuras, M., Parada, I. P., Sugihara, Y., Appelqvist, R., Wieslander, E., Welinder, C., Velasquez, E., De Almeida, N. P., Woldmar, N., Marko-Varga, M., Pawłowski, K., Eriksson, J., Szeitz, B., Baldetorp, B., Ingvar, C., Olsson, H., Lundgren, L., Lindberg, H., Oskolas, H., Lee, B., Berge, E., Sjögren, M., Eriksson, C., Kim, D., Kwon, H. J., Knudsen, B., Rezeli, M., Hong, R., Horvatovich, P., Miliotis, T., Nishimura, T., Kato, H., Steinfelder, E., Oppermann, M., Miller, K., Florindi, F., Zhou, Q., Domont, G. B., Pizzatti, L., Nogueira, F. C. S., Horvath, P., Szadai, L., Tímár, J., Kárpáti, S., Szász, A. M., Malm, J., Fenyő, D., Ekedahl, H., Németh, I. B. & Marko-Varga, G. The human melanoma proteome atlas—Defining the molecular pathology. *Clinical and Translational Medicine* 11, e473 (2021).
9. Martincorena, I., Roshan, A., Gerstung, M., Ellis, P., Van Loo, P., McLaren, S., Wedge, D. C., Fullam, A., Alexandrov, L. B., Tubio, J. M., Stebbings, L., Menzies, A., Widaa, S., Stratton, M. R., Jones, P. H. & Campbell, P. J. High burden and pervasive positive selection of somatic mutations in normal human skin. *Science* 348, 880–886 (2015).

10. Macosko, E. Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A. R., Kamitaki, N., Martersteck, E. M., Trombetta, J. J., Weitz, D. A., Sanes, J. R., Shalek, A. K., Regev, A. & McCarroll, S. A. Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* 161, 1202–1214 (2015).
11. Stuart, T. & Satija, R. Integrative single-cell analysis. *Nature Reviews Genetics* 20, 257–272 (2019).
12. Moffitt, J. R., Lundberg, E. & Heyn, H. The emerging landscape of spatial profiling technologies. *Nature Reviews Genetics* 23, 741–759 (2022).
13. Moses, L. & Pachter, L. Museum of spatial transcriptomics. *Nature Methods* 19, 534–546 (2022).
14. Piñeiro, A. J., Houser, A. E. & Ji, A. L. Research Techniques Made Simple: Spatial Transcriptomics. *Journal of Investigative Dermatology* 142, 993-1001.e1 (2022).
15. Tekkela, S., Theocharidis, G., McGrath, J. A. & Onoufriadis, A. Spatial transcriptomics in human skin research. *Experimental Dermatology* 32, 731–739 (2023).
16. Baysoy, A., Bai, Z., Satija, R. & Fan, R. The technological landscape and applications of single-cell multi-omics. *Nature Reviews Molecular Cell Biology* 24, 695–713 (2023).
17. Jia, B. B., Jussila, A., Kern, C., Zhu, Q. & Ren, B. A spatial genome aligner for resolving chromatin architectures from multiplexed DNA FISH. *Nature Biotechnology* 41, 1004–1017 (2023).
18. Boettiger, A. & Murphy, S. Advances in Chromatin Imaging at Kilobase-Scale Resolution. *Trends in Genetics* 36, 273–287 (2020).
19. Hu, M. & Wang, S. Chromatin Tracing: Imaging 3D Genome and Nucleome. *Trends in Cell Biology* 31, 5–8 (2021).
20. Su, J.-H., Zheng, P., Kinrot, S. S., Bintu, B. & Zhuang, X. Genome-Scale Imaging of the 3D Organization and Transcriptional Activity of Chromatin. *Cell* 182, 1641–1659.e26 (2020).
21. Takei, Y., Zheng, S., Yun, J., Shah, S., Pierson, N., White, J., Schindler, S., Tischbirek, C. H., Yuan, G.-C. & Cai, L. Single-cell nuclear architecture across cell types in the mouse brain. *Science* 374, 586–594 (2021).
22. Takei, Y., Yun, J., Zheng, S., Ollikainen, N., Pierson, N., White, J., Shah, S., Thomassie, J., Suo, S., Eng, C.-H. L., Guttman, M., Yuan, G.-C. & Cai, L. Integrated spatial genomics reveals global architecture of single nuclei. *Nature* 590, 344–350 (2021).
23. Alon, S., Goodwin, D. R., Sinha, A., Wassie, A. T., Chen, F., Daugharthy, E. R., Bando,

- Y., Kajita, A., Xue, A. G., Marrett, K., Prior, R., Cui, Y., Payne, A. C., Yao, C.-C., Suk, H.-J., Wang, R., Yu, C.-C. (Jay), Tillberg, P., Reginato, P., Pak, N., Liu, S., Punthambaker, S., Iyer, E. P. R., Kohman, R. E., Miller, J. A., Lein, E. S., Lako, A., Cullen, N., Rodig, S., Helvie, K., Abravanel, D. L., Wagle, N., Johnson, B. E., Klughammer, J., Slyper, M., Waldman, J., Jané-Valbuena, J., Rozenblatt-Rosen, O., Regev, A., IMAXT Consortium, Church, G. M., Marblestone, A. H., Boyden, E. S., Ali, H. R., Al Sa'd, M., Alon, S., Aparicio, S., Battistoni, G., Balasubramanian, S., Becker, R., Bodenmiller, B., Boyden, E. S., Bressan, D., Bruna, A., Burger, M., Caldas, C., Callari, M., Cannell, I. G., Casbolt, H., Chornay, N., Cui, Y., Dariush, A., Dinh, K., Emenari, A., Eyal-Lubling, Y., Fan, J., Fatemi, A., Fisher, E., González-Solares, E. A., González-Fernández, C., Goodwin, D., Greenwood, W., Grimaldi, F., Hannon, G. J., Harris, O., Harris, S., Jauset, C., Joyce, J. A., Karagiannis, E. D., Kovačević, T., Kuett, L., Kunes, R., Küpcü Yoldaş, A., Lai, D., Laks, E., Lee, H., Lee, M., Lerda, G., Li, Y., McPherson, A., Millar, N., Mulvey, C. M., Nugent, F., O'Flanagan, C. H., Paez-Ribes, M., Pearsall, I., Qosaj, F., Roth, A. J., Rueda, O. M., Ruiz, T., Sawicka, K., Sepúlveda, L. A., Shah, S. P., Shea, A., Sinha, A., Smith, A., Tavaré, S., Tietscher, S., Vázquez-García, I., Vogl, S. L., Walton, N. A., Wassie, A. T., Watson, S. S., Weselak, J., Wild, S. A., Williams, E., Windhager, J., Whitmarsh, T., Xia, C., Zheng, P. & Zhuang, X. Expansion sequencing: Spatially precise in situ transcriptomics in intact biological systems. *Science* 371, eaax2656 (2021).
24. Eng, C.-H. L., Lawson, M., Zhu, Q., Dries, R., Koulana, N., Takei, Y., Yun, J., Cronin, C., Karp, C., Yuan, G.-C. & Cai, L. Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH+. *Nature* 568, 235–239 (2019).
 25. Kalhor, K., Chen, C.-J., Lee, H. S., Cai, M., Nafisi, M., Que, R., Palmer, C., Yuan, Y., Zhang, Y., Song, J., Knoten, A., Lake, B. B., Gaut, J. P., Keene, D., Lein, E., Kharchenko, P. V., Chun, J., Jain, S., Fan, J.-B. & Zhang, K. Mapping Human Tissues with Highly Multiplexed RNA in situ Hybridization. (*Genomics*, 2023). doi:10.1101/2023.08.16.553610
 26. Payne, A. C., Chiang, Z. D., Reginato, P. L., Mangiameli, S. M., Murray, E. M., Yao, C.-C., Markoulaki, S., Earl, A. S., Labade, A. S., Jaenisch, R., Church, G. M., Boyden, E. S., Buenrostro, J. D. & Chen, F. In situ genome sequencing resolves DNA sequence and structure in intact biological samples. *Science* 371, eaay3446 (2021).
 27. Wang, X., Allen, W. E., Wright, M. A., Sylwestrak, E. L., Samusik, N., Vesuna, S., Evans, K., Liu, C., Ramakrishnan, C., Liu, J., Nolan, G. P., Bava, F.-A. & Deisseroth, K. Three-dimensional intact-tissue sequencing of single-cell transcriptional states. *Science* 361, eaat5691 (2018).
 28. Xia, C., Fan, J., Emanuel, G., Hao, J. & Zhuang, X. Spatial transcriptome profiling by MERFISH reveals subcellular RNA compartmentalization and cell cycle-dependent gene expression. *Proceedings of the National Academy of Sciences* 116, 19490–19499 (2019).
 29. Zhang, M., Eichhorn, S. W., Zingg, B., Yao, Z., Cotter, K., Zeng, H., Dong, H. &

Zhuang, X. Spatially resolved cell atlas of the mouse primary motor cortex by MERFISH. *Nature* 598, 137–143 (2021).

30. Chen, A., Liao, S., Cheng, M., Ma, K., Wu, L., Lai, Y., Qiu, X., Yang, J., Xu, J., Hao, S., Wang, X., Lu, H., Chen, X., Liu, X., Huang, X., Li, Z., Hong, Y., Jiang, Y., Peng, J., Liu, S., Shen, M., Liu, C., Li, Q., Yuan, Y., Wei, X., Zheng, H., Feng, W., Wang, Z., Liu, Y., Wang, Z., Yang, Y., Xiang, H., Han, L., Qin, B., Guo, P., Lai, G., Muñoz-Cánoves, P., Maxwell, P. H., Thiery, J. P., Wu, Q.-F., Zhao, F., Chen, B., Li, M., Dai, X., Wang, S., Kuang, H., Hui, J., Wang, L., Fei, J.-F., Wang, O., Wei, X., Lu, H., Wang, B., Liu, S., Gu, Y., Ni, M., Zhang, W., Mu, F., Yin, Y., Yang, H., Lisby, M., Cornall, R. J., Mulder, J., Uhlén, M., Esteban, M. A., Li, Y., Liu, L., Xu, X. & Wang, J. Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned arrays. *Cell* 185, 1777-1792.e21 (2022).
31. Cho, C.-S., Xi, J., Si, Y., Park, S.-R., Hsu, J.-E., Kim, M., Jun, G., Kang, H. M. & Lee, J. H. Microscopic examination of spatial transcriptome using Seq-Scope. *Cell* 184, 3559-3572.e22 (2021).
32. Fu, X., Sun, L., Dong, R., Chen, J. Y., Silakit, R., Condon, L. F., Lin, Y., Lin, S., Palmiter, R. D. & Gu, L. Polony gels enable amplifiable DNA stamping and spatial transcriptomics of chronic pain. *Cell* 185, 4621-4633.e17 (2022).
33. Lee, Y., Bogdanoff, D., Wang, Y., Hartoularos, G. C., Woo, J. M., Mowery, C. T., Nisonoff, H. M., Lee, D. S., Sun, Y., Lee, J., Mehdizadeh, S., Cantlon, J., Shifrut, E., Ngyuen, D. N., Roth, T. L., Song, Y. S., Marson, A., Chow, E. D. & Ye, C. J. XYZeq: Spatially resolved single-cell RNA sequencing reveals expression heterogeneity in the tumor microenvironment. *Science Advances* 7, eabg4755 (2021).
34. Liu, Y., Yang, M., Deng, Y., Su, G., Enniful, A., Guo, C. C., Tebaldi, T., Zhang, D., Kim, D., Bai, Z., Norris, E., Pan, A., Li, J., Xiao, Y., Halene, S. & Fan, R. High-Spatial-Resolution Multi-Omics Sequencing via Deterministic Barcoding in Tissue. *Cell* 183, 1665-1681.e18 (2020).
35. Rodriques, S. G., Stickels, R. R., Goeva, A., Martin, C. A., Murray, E., Vanderburg, C. R., Welch, J., Chen, L. M., Chen, F. & Macosko, E. Z. Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. *Science* 363, 1463–1467 (2019).
36. Srivatsan, S. R., Regier, M. C., Barkan, E., Franks, J. M., Packer, J. S., Grosjean, P., Duran, M., Saxton, S., Ladd, J. J., Spielmann, M., Lois, C., Lampe, P. D., Shendure, J., Stevens, K. R. & Trapnell, C. Embryo-scale, single-cell spatial transcriptomics. *Science* 373, 111–117 (2021).
37. Ståhl, P. L., Salmén, F., Vickovic, S., Lundmark, A., Navarro, J. F., Magnusson, J., Giacomello, S., Asp, M., Westholm, J. O., Huss, M., Mollbrink, A., Linnarsson, S., Codeluppi, S., Borg, Å., Pontén, F., Costea, P. I., Sahlén, P., Mulder, J., Bergmann, O., Lundberg, J. & Frisén, J. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* 353, 78–82 (2016).

38. Vickovic, S., Eraslan, G., Salmén, F., Klughammer, J., Stenbeck, L., Schapiro, D., Äijö, T., Bonneau, R., Bergensträhle, L., Navarro, J. F., Gould, J., Griffin, G. K., Borg, Å., Ronaghi, M., Frisén, J., Lundeberg, J., Regev, A. & Ståhl, P. L. High-definition spatial transcriptomics for in situ tissue profiling. *Nature Methods* 16, 987–990 (2019).
39. Hu, K. H., Eichorst, J. P., McGinnis, C. S., Patterson, D. M., Chow, E. D., Kersten, K., Jameson, S. C., Gartner, Z. J., Rao, A. A. & Krummel, M. F. ZipSeq: barcoding for real-time mapping of single cell transcriptomes. *Nature Methods* 17, 833–843 (2020).
40. Kishi, J. Y., Liu, N., West, E. R., Sheng, K., Jordanides, J. J., Serrata, M., Cepko, C. L., Saka, S. K. & Yin, P. Light-Seq: light-directed in situ barcoding of biomolecules in fixed cells and tissues for spatially indexed sequencing. *Nature Methods* 19, 1393–1402 (2022).
41. Merritt, C. R., Ong, G. T., Church, S. E., Barker, K., Danaher, P., Geiss, G., Hoang, M., Jung, J., Liang, Y., McKay-Fleisch, J., Nguyen, K., Norgaard, Z., Sorg, K., Sprague, I., Warren, C., Warren, S., Webster, P. J., Zhou, Z., Zollinger, D. R., Dunaway, D. L., Mills, G. B. & Beechem, J. M. Multiplex digital spatial profiling of proteins and RNA in fixed tissue. *Nature Biotechnology* 38, 586–599 (2020).
42. Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S. & Zhuang, X. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* 348, aaa6090 (2015).
43. Codeluppi, S., Borm, L. E., Zeisel, A., La Manno, G., van Lunteren, J. A., Svensson, C. I. & Linnarsson, S. Spatial organization of the somatosensory cortex revealed by osmFISH. *Nat Methods* 15, 932–935 (2018).
44. Ke, R., Mignardi, M., Pacureanu, A., Svedlund, J., Botling, J., Wählby, C. & Nilsson, M. In situ sequencing for RNA analysis in preserved tissue and cells. *Nat Methods* 10, 857–860 (2013).
45. Lubeck, E., Coskun, A. F., Zhiyentayev, T., Ahmad, M. & Cai, L. Single-cell in situ RNA profiling by sequential hybridization. *Nature Methods* 11, 360–361 (2014).
46. Beliveau, B. J., Joyce, E. F., Apostolopoulos, N., Yilmaz, F., Fonseka, C. Y., McCole, R. B., Chang, Y., Li, J. B., Senaratne, T. N., Williams, B. R., Rouillard, J.-M. & Wu, C. Versatile design and synthesis platform for visualizing genomes with Oligopaint FISH probes. *Proceedings of the National Academy of Sciences* 109, 21301–21306 (2012).
47. Femino, A. M., Fay, F. S., Fogarty, K. & Singer, R. H. Visualization of Single RNA Transcripts in Situ. *Science* 280, 585–590 (1998).
48. Raj, A., Van Den Bogaard, P., Rifkin, S. A., Van Oudenaarden, A. & Tyagi, S. Imaging individual mRNA molecules using multiple singly labeled probes. *Nat Methods* 5, 877–879 (2008).
49. Lee, J. H., Daugharthy, E. R., Scheiman, J., Kalhor, R., Yang, J. L., Ferrante, T. C., Terry, R., Jeanty, S. S. F., Li, C., Amamoto, R., Peters, D. T., Turczyk, B. M.,

- Marblestone, A. H., Inverso, S. A., Bernard, A., Mali, P., Rios, X., Aach, J. & Church, G. M. Highly Multiplexed Subcellular RNA Sequencing in Situ. *Science* 343, 1360–1363 (2014).
50. Nguyen, H. Q., Chatteraj, S., Castillo, D., Nguyen, S. C., Nir, G., Lioutas, A., Hershberg, E. A., Martins, N. M. C., Reginato, P. L., Hannan, M., Beliveau, B. J., Church, G. M., Daugharthy, E. R., Marti-Renom, M. A. & Wu, C. -ting. 3D mapping and accelerated super-resolution imaging of the human genome using in situ sequencing. *Nat Methods* 17, 822–832 (2020).
 51. Stickels, R. R., Murray, E., Kumar, P., Li, J., Marshall, J. L., Di Bella, D. J., Arlotta, P., Macosko, E. Z. & Chen, F. Highly sensitive spatial transcriptomics at near-cellular resolution with Slide-seqV2. *Nat Biotechnol* 39, 313–319 (2021).
 52. Biancalani, T., Scalia, G., Buffoni, L., Avasthi, R., Lu, Z., Sanger, A., Tokcan, N., Vanderburg, C. R., Segerstolpe, Å., Zhang, M., Avraham-Davidi, I., Vickovic, S., Nitzan, M., Ma, S., Subramanian, A., Lipinski, M., Buenrostro, J., Brown, N. B., Fanelli, D., Zhuang, X., Macosko, E. Z. & Regev, A. Deep learning and alignment of spatially resolved single-cell transcriptomes with Tangram. *Nat Methods* 18, 1352–1362 (2021).
 53. Dries, R., Zhu, Q., Dong, R., Eng, C.-H. L., Li, H., Liu, K., Fu, Y., Zhao, T., Sarkar, A., Bao, F., George, R. E., Pierson, N., Cai, L. & Yuan, G.-C. Giotto: a toolbox for integrative analysis and visualization of spatial expression data. *Genome Biology* 22, 78 (2021).
 54. Elosua-Bayes, M., Nieto, P., Mereu, E., Gut, I. & Heyn, H. SPOTlight: seeded NMF regression to deconvolute spatial transcriptomics spots with single-cell transcriptomes. *Nucleic Acids Research* 49, e50 (2021).
 55. Lopez, R., Li, B., Keren-Shaul, H., Boyeau, P., Kedmi, M., Pilzer, D., Jelinski, A., Yofe, I., David, E., Wagner, A., Ergen, C., Addadi, Y., Golani, O., Ronchese, F., Jordan, M. I., Amit, I. & Yosef, N. DestVI identifies continuums of cell types in spatial transcriptomics data. *Nat Biotechnol* 40, 1360–1369 (2022).
 56. Bintu, B., Mateo, L. J., Su, J.-H., Sinnott-Armstrong, N. A., Parker, M., Kinrot, S., Yamaya, K., Boettiger, A. N. & Zhuang, X. Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. *Science* 362, eaau1783 (2018).
 57. Gizzi, A. M. C., Cattoni, D. I., Fiche, J.-B., Espinola, S. M., Gurgo, J., Messina, O., Houbon, C., Ogiyama, Y., Papadopoulos, G. L., Cavalli, G., Lagha, M. & Nollmann, M. Microscopy-Based Chromosome Conformation Capture Enables Simultaneous Visualization of Genome Organization and Transcription in Intact Organisms. *Molecular Cell* 74, 212–222.e5 (2019).
 58. Huang, H., Zhu, Q., Jussila, A., Han, Y., Bintu, B., Kern, C., Conte, M., Zhang, Y., Bianco, S., Chiariello, A. M., Yu, M., Hu, R., Tastemel, M., Juric, I., Hu, M., Nicodemi, M., Zhuang, X. & Ren, B. CTCF mediates dosage- and sequence-context-dependent

transcriptional insulation by forming local chromatin domains. *Nat Genet* 53, 1064–1074 (2021).

59. Liu, M., Yang, B., Hu, M., Radda, J. S. D., Chen, Y., Jin, S., Cheng, Y. & Wang, S. Chromatin tracing and multiplexed imaging of nucleome architectures (MINA) and RNAs in single mammalian cells and tissue. *Nat Protoc* 16, 2667–2697 (2021).
60. Mateo, L. J., Sinnott-Armstrong, N. & Boettiger, A. N. Tracing DNA paths and RNA profiles in cultured cells and tissues with ORCA. *Nat Protoc* 16, 1647–1713 (2021).
61. Nir, G., Farabella, I., Estrada, C. P., Ebeling, C. G., Beliveau, B. J., Sasaki, H. M., Lee, S. D., Nguyen, S. C., McCole, R. B., Chatteraj, S., Erceg, J., Abed, J. A., Martins, N. M. C., Nguyen, H. Q., Hannan, M. A., Russell, S., Durand, N. C., Rao, S. S. P., Kishi, J. Y., Soler-Vila, P., Pierro, M. D., Onuchic, J. N., Callahan, S. P., Schreiner, J. M., Stuckey, J. A., Yin, P., Aiden, E. L., Marti-Renom, M. A. & Wu, C. -ting. Walking along chromosomes with super-resolution imaging, contact maps, and integrative modeling. *PLOS Genetics* 14, e1007872 (2018).
62. Flores, V., Farabella, I. & Nir, G. Genome-wide tracing to decipher nuclear organization. *Current Opinion in Cell Biology* 82, 102175 (2023).
63. Zhang, Y., Boninsegna, L., Yang, M., Misteli, T., Alber, F. & Ma, J. Computational methods for analysing multiscale 3D genome organization. *Nat Rev Genet* 1–19 (2023). doi:10.1038/s41576-023-00638-1
64. Ji, A. L., Rubin, A. J., Thrane, K., Jiang, S., Reynolds, D. L., Meyers, R. M., Guo, M. G., George, B. M., Mollbrink, A., Bergenstr hle, J., Larsson, L., Bai, Y., Zhu, B., Bhaduri, A., Meyers, J. M., Rovira-Clav , X., Hollmig, S. T., Aasi, S. Z., Nolan, G. P., Lundeberg, J. & Khavari, P. A. Multimodal Analysis of Composition and Spatial Architecture in Human Squamous Cell Carcinoma. *Cell* 182, 497-514.e22 (2020).
65. Zhao, T., Chiang, Z. D., Morriss, J. W., LaFave, L. M., Murray, E. M., Del Priore, I., Meli, K., Lareau, C. A., Nadaf, N. M., Li, J., Earl, A. S., Macosko, E. Z., Jacks, T., Buenrostro, J. D. & Chen, F. Spatial genomics enables multi-modal study of clonal heterogeneity in tissues. *Nature* 601, 85–91 (2022).
66. Shim, J., Oh, S. J., Yeo, E., Park, J. H., Bae, J. H., Kim, S.-H., Lee, D. & Lee, J. H. Integrated Analysis of Single-Cell and Spatial Transcriptomics in Keloids: Highlights on Fibrovascular Interactions in Keloid Pathogenesis. *Journal of Investigative Dermatology* 142, 2128-2139.e11 (2022).
67. Thrane, K., Eriksson, H., Maaskola, J., Hansson, J. & Lundeberg, J. Spatially Resolved Transcriptomics Enables Dissection of Genetic Heterogeneity in Stage III Cutaneous Malignant Melanoma. *Cancer Research* 78, 5970–5979 (2018).
68. Curry, J., Alnemri, A., Philips, R., Fiorella, M., Sussman, S., Stapp, R., Solomides, C., Harshyne, L., South, A., Luginbuhl, A., Tuluc, M., Martinez-Outschoorn, U., Argiris, A., Linnenbach, A. & Johnson, J. CD8+ and FoxP3+ T-Cell Cellular Density and

Spatial Distribution After Programmed Death-Ligand 1 Check Point Inhibition. *The Laryngoscope* 133, 1875–1884 (2023).

69. Klemm, S. L., Shipony, Z. & Greenleaf, W. J. Chromatin accessibility and the regulatory epigenome. *Nat Rev Genet* 20, 207–220 (2019).
70. Yu, M. & Ren, B. The Three-Dimensional Organization of Mammalian Genomes. *Annual Review of Cell and Developmental Biology* 33, 265–289 (2017).
71. Rao, A., Barkley, D., França, G. S. & Yanai, I. Exploring tissue architecture using spatial transcriptomics. *Nature* 596, 211–220 (2021).
72. Vandereyken, K., Sifrim, A., Thienpont, B. & Voet, T. Methods and applications for single-cell and spatial multi-omics. *Nat Rev Genet* 24, 494–515 (2023).
73. Deng, Y., Bartosovic, M., Ma, S., Zhang, D., Kukanja, P., Xiao, Y., Su, G., Liu, Y., Qin, X., Rosoklija, G. B., Dwork, A. J., Mann, J. J., Xu, M. L., Halene, S., Craft, J. E., Leong, K. W., Boldrini, M., Castelo-Branco, G. & Fan, R. Spatial profiling of chromatin accessibility in mouse and human tissues. *Nature* 609, 375–383 (2022).
74. Zhang, D., Deng, Y., Kukanja, P., Agirre, E., Bartosovic, M., Dong, M., Ma, C., Ma, S., Su, G., Bao, S., Liu, Y., Xiao, Y., Rosoklija, G. B., Dwork, A. J., Mann, J. J., Leong, K. W., Boldrini, M., Wang, L., Haeussler, M., Raphael, B. J., Kluger, Y., Castelo-Branco, G. & Fan, R. Spatial epigenome–transcriptome co-profiling of mammalian tissues. *Nature* 616, 113–122 (2023).
75. Deng, Y., Bartosovic, M., Kukanja, P., Zhang, D., Liu, Y., Su, G., Enninfu, A., Bai, Z., Castelo-Branco, G. & Fan, R. Spatial-CUT&Tag: Spatially resolved chromatin modification profiling at the cellular level. *Science* 375, 681–686 (2022).
76. Lu, T., Ang, C. E. & Zhuang, X. Spatially resolved epigenomic profiling of single cells in complex tissues. *Cell* 185, 4448–4464.e17 (2022).
77. Ben-Chetrit, N., Niu, X., Swett, A. D., Sotelo, J., Jiao, M. S., Stewart, C. M., Potenski, C., Mielinis, P., Roelli, P., Stoeckius, M. & Landau, D. A. Integration of whole transcriptome spatial profiling with protein markers. *Nat Biotechnol* 41, 788–793 (2023).
78. Goltsev, Y., Samusik, N., Kennedy-Darling, J., Bhate, S., Hale, M., Vazquez, G., Black, S. & Nolan, G. P. Deep Profiling of Mouse Splenic Architecture with CODEX Multiplexed Imaging. *Cell* 174, 968–981.e15 (2018).
79. Williams, M., Bonnardel, J., Haest, B., Vanderborght, B., Wagner, C., Remmerie, A., Bujko, A., Martens, L., Thoné, T., Browaeys, R., De Ponti, F. F., Vanneste, B., Zwicker, C., Svedberg, F. R., Vanhalewyn, T., Gonçalves, A., Lippens, S., Devriendt, B., Cox, E., Ferrero, G., Wittamer, V., Willaert, A., Kaptein, S. J. F., Neyts, J., Dallmeier, K., Geldhof, P., Casaert, S., Deplancke, B., Ten Dijke, P., Hoorens, A., Vanlander, A., Berrevoet, F., Van Nieuwenhove, Y., Saeys, Y., Saelens, W., Van Vlierberghe, H.,

- Devisscher, L. & Scott, C. L. Spatial proteogenomics reveals distinct and evolutionarily conserved hepatic macrophage niches. *Cell* 185, 379-396.e38 (2022).
80. Liu, Y., DiStasio, M., Su, G., Asashima, H., Enniful, A., Qin, X., Deng, Y., Nam, J., Gao, F., Bordinon, P., Cassano, M., Tomayko, M., Xu, M., Halene, S., Craft, J. E., Hafler, D. & Fan, R. High-plex protein and whole transcriptome co-mapping at cellular resolution with spatial CITE-seq. *Nat Biotechnol* 41, 1405–1409 (2023).
 81. Gut, G., Herrmann, M. D. & Pelkmans, L. Multiplexed protein maps link subcellular organization to cellular states. *Science* 361, eaar7042 (2018).
 82. Lin, J.-R., Izar, B., Wang, S., Yapp, C., Mei, S., Shah, P. M., Santagata, S. & Sorger, P. K. Highly multiplexed immunofluorescence imaging of human tissues and tumors using t-CyCIF and conventional optical microscopes. *eLife* 7, e31657 (2018).
 83. Radtke, A. J., Kandov, E., Lowekamp, B., Speranza, E., Chu, C. J., Gola, A., Thakur, N., Shih, R., Yao, L., Yaniv, Z. R., Beuschel, R. T., Kabat, J., Croteau, J., Davis, J., Hernandez, J. M. & Germain, R. N. IBEX: A versatile multiplex optical imaging approach for deep phenotyping and spatial analysis of cells in complex tissues. *Proc. Natl. Acad. Sci. U.S.A.* 117, 33455–33465 (2020).
 84. van der Laak, J., Litjens, G. & Ciompi, F. Deep learning in histopathology: the path to the clinic. *Nat Med* 27, 775–784 (2021).
 85. Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M. & Thrun, S. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 542, 115–118 (2017).
 86. Kouzak, S. S., Mendes, M. S. T. & Costa, I. M. C. Cutaneous mosaicisms: concepts, patterns and classifications. *An Bras Dermatol* 88, 507–517 (2013).
 87. Lim, Y. H., Moscato, Z. & Choate, K. A. Mosaicism in Cutaneous Disorders. *Annual Review of Genetics* 51, 123–141 (2017).
 88. King, C., Fowler, J. C., Abnizova, I., Sood, R. K., Hall, M. W. J., Szeverényi, I., Tham, M., Huang, J., Young, S. M., Hall, B. A., Birgitte Lane, E. & Jones, P. H. Somatic mutations in facial skin from countries of contrasting skin cancer risk. *Nat Genet* 55, 1440–1447 (2023).
 89. Vermeulen, C., Pagès-Gallego, M., Kester, L., Kranendonk, M. E. G., Wesseling, P., Verburg, N., de Witt Hamer, P., Kooi, E. J., Dankmeijer, L., van der Lugt, J., van Baarsen, K., Hoving, E. W., Tops, B. B. J. & de Ridder, J. Ultra-fast deep-learned CNS tumour classification during surgery. *Nature* 622, 842–849 (2023).