

Article

Not peer-reviewed version

FISH-Dist: An Automated Pipeline for Accurate 3D Genomic Spatial Distance Quantification in FISH Imaging

[Benoit Aigouy](#)*, [Emmanuelle Caturegli](#), [Bernard Charroux](#), [Carla Silva-Martins](#), Thomas Gregor, [Benjamin Prud'homme](#)*

Posted Date: 30 January 2026

doi: 10.20944/preprints202601.2343.v1

Keywords: fluorescence in situ hybridization (FISH); Oligopaint probes; chromatic aberration correction; 3D genome organization; short-range genomic interactions; distance quantification; sub-pixel localization; confocal microscopy; computational image analysis



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

FISH-Dist: An Automated Pipeline for Accurate 3D Genomic Spatial Distance Quantification in FISH Imaging

Benoit Aigouy ^{1,*} , Emmanuelle Caturegli ¹, Bernard Charroux ¹ , Carla Silva-Martins ¹, Thomas Gregor ^{2,3}  and Benjamin Prud'homme ^{1,*} 

¹ Aix-Marseille Université, CNRS, IBDM, Campus de Luminy Case 907, 13288 Marseille Cedex 9, France
² Lewis-Sigler Institute for Integrative Genomics, and Joseph Henry Laboratories of Physics, Princeton University, Princeton NJ 08544
³ Department of Developmental and Stem Cell Biology UMR3738, Institut Pasteur, 75015 Paris, France
* Correspondence: benoit.aigouy@univ-amu.fr (B.A.); benjamin.prudhomme@univ-amu.fr (B.P.)

Abstract

Accurate quantification of spatial distances between fluorescent signals in multi-channel 3D microscopy is essential for understanding genomic organization and gene regulation. However, chromatic aberration introduces systematic spatial offsets between channels that significantly bias distance measurements, particularly at short genomic distances. We present FISH-Dist, an automated computational pipeline for precise distance measurements in 3D fluorescence in situ hybridization (FISH) experiments acquired on standard confocal microscopes. Our method combines deep learning-based spot segmentation, 3D Gaussian fitting for sub-pixel localization, and two complementary chromatic aberration correction approaches, affine (ACC) and linear (LCC). We validated measurement accuracy using DNA origami nanorulers and systematically evaluated FISH probe design parameters, including probe spacing, density, and target sequence length. The pipeline achieves sub-pixel accuracy in signal detection and substantially reduces inter-channel distance measurement errors. This enables robust quantification of spatial relationships in 3D FISH datasets. Unlike existing tools optimized for long-range chromosomal interactions or requiring super-resolution microscopy, FISH-Dist specifically addresses the technical challenges of standard confocal imaging at short genomic distances, where chromatic aberration has a proportionally greater impact on measurement accuracy.

Keywords: fluorescence in situ hybridization (FISH); Oligopaint probes; chromatic aberration correction; 3D genome organization; short-range genomic interactions; distance quantification; sub-pixel localization; confocal microscopy; computational image analysis

1. Introduction

The three-dimensional organization of the genome is fundamental to gene regulation, as spatial proximity between regulatory elements and their target genes directly influences transcription [1,2]. Fluorescence in situ hybridization (FISH), including modern Oligopaint-based approaches, remains a gold-standard method for visualizing these spatial relationships with single-cell resolution within the native nuclear context [3–5].

While existing FISH studies and computational tools, sometimes complemented by Hi-C-based maps, have significantly advanced our understanding of large-scale chromosomal structures and long-range interactions spanning hundreds of kilobases to megabases [4,6–10], many critical regulatory interactions—such as enhancer–promoter loops and local chromatin contacts—occur at much finer genomic scales (<10–100 kb) [11]. At these scales, technical limitations, including chromatic aberration, can introduce measurement errors that are comparable to, or even exceed, the biological distances being investigated.

Despite the biological significance of short-range genomic measurements, many studies still rely on user-dependent distance measurements, which are prone to error [12–15]. Existing computational tools are typically not designed or rigorously validated for high-precision quantification at these scales. Moreover, the relationship between FISH probe design parameters—such as probe density and target sequence length—and measurement accuracy remains poorly characterized for short-distance applications.

Here we present FISH-Dist, an automated computational pipeline specifically designed for accurate 3D distance quantification in FISH imaging, with a focus on short-range genomic measurements. FISH-Dist integrates deep learning–based spot detection, sub-pixel 3D Gaussian fitting, and complementary chromatic aberration correction strategies for high-resolution measurements. We validate the pipeline using multiple orthogonal approaches, including colocalization experiments and calibrated nanorulers. Using FISH-Dist, we also systematically quantify how probe design parameters, such as labeling density and target sequence length, influence measurement accuracy, providing practical guidance for experimental planning. By addressing technical limitations in standard confocal FISH, FISH-Dist enables robust and reproducible quantification of spatial relationships at scales most relevant to gene regulation.

2. Materials and Methods

2.1. Oligopaint Probe Synthesis

Forty-five–nucleotide single-stranded DNA Oligopaint probes targeting specific DNA sequences were designed by Daicel Arbor Biosciences and synthesized from myTags Immortal Libraries (Daicel Arbor Biosciences) according to the manufacturer’s protocol [16]. Briefly, myTags Immortal Libraries were amplified by PCR, and the resulting products were column-purified. In vitro transcription was then performed to generate RNA, which was column-purified and used as a template for reverse transcription (RT). When RNA concentration was insufficient for RT, an additional ethanol precipitation was performed. RT was carried out using fluorescent primers (coupled at the 5’ end with a single ATTO 565 or ATTO 633 fluorescent dye) purchased from IDT. Unincorporated primers were degraded, and RNA:DNA hybrids were column-purified. RNA fragments were subsequently removed by RNase treatment, and the resulting fluorescent ssDNA probes were column-purified and stored at -20 °C. The fraction of successfully dye-conjugated probes was quantified using a Nanodrop spectrophotometer to ensure an appropriate labeling ratio.

2.2. Oligopaint Probe Design for Spatial Resolution Calibration in Multi-Color FISH

Oligopaint probes were designed to target a ~10 kb endogenous genomic region (X:6,760,094–6,770,369) adjacent to the attP18 landing site on the X chromosome of *Drosophila melanogaster*. A total of 142 Oligopaint probes were designed to hybridize exclusively to the + DNA strand of this locus. For dual-color experiments, the same probe set was labeled with either ATTO 565 or ATTO 633 fluorophores.

2.3. Construction and Oligopaint Targeting of Synthetic Reporter Sequences

Two synthetic reporter sequences (R1 and R6) were designed to be orthogonal to the *Drosophila melanogaster* genome by Daicel Arbor Biosciences and synthesized by Genewiz. Each reporter consists of a 2 kb DNA sequence absent from the *D. melanogaster* genome, with similar overall GC content but distinct sequence composition. Each reporter was targeted by a set of 87 Oligopaint probes tiled across the entire sequence and both DNA strands. Transgenes were integrated at the ZH-2A landing site on the X chromosome using site-specific recombination.

2.4. Transgene Design for Defined 3D Genomic Distance Measurements

To evaluate the accuracy of 3D distance measurements under biological conditions, we generated transgenic constructs containing two distinct Oligopaint-tagged sets targeting R1 and R6, labeled with ATTO 565 and ATTO 633 and separated by defined spacer lengths of 0 kb, 3 kb, or 10 kb. In all constructs, the spacer is flanked by the synthetic reporter tags R1 and R6, which serve as independent

Oligopaint target sites. 3kb spacer consists of randomly generated DNA sequences designed to minimize binding of known *Drosophila* transcription factors. The 10kb spacer is a cherry-tagged *yellow* gene of *Drosophila biarmipes*. All constructs were integrated at the ZH-2A landing site.

2.5. Oligopaint Target Length Variants for Assessing Spatial Resolution

To examine how FISH target sequence length influences spatial resolution, a series of transgenic constructs was generated in which progressively shorter segments of the R6 synthetic reporter were used as Oligopaint targets. Three target lengths were tested—2 kb, 1 kb, and 500 bp—preserving sequence composition while varying target size. All constructs were integrated at the ZH-2A landing site, ensuring a constant genomic context across conditions.

2.6. Modulation of Oligopaint Probe Density

To assess the impact of Oligopaint probe density on measurement performance, the fraction of fluorescently labeled probes targeting a fixed 2 kb R6 genomic region was systematically varied while maintaining a constant total probe concentration. This was achieved by mixing fluorescently labeled probes with unlabeled (“cold”) probes targeting the same sequence, thereby preserving hybridization conditions while progressively reducing the number of fluorescent probes contributing to the detected signal.

Three labeling conditions were examined. In the fully labeled condition (~100%), the 2 kb R6 locus could be covered by up to 87 labeled probes and no unlabeled probes; ATTO 565- and ATTO 633-labeled probe sets were each used at a final concentration of 3 ng/μl, resulting in a total labeled probe concentration of 6 ng/μl. In the intermediate labeling condition (~66%), the locus could be covered by up to 57 labeled probes and 30 unlabeled probes; each labeled probe set was used at 2 ng/μl and supplemented with 2 ng/μl of unlabeled probes. In the lowest labeling condition (~50%), the locus could be covered by up to 43 labeled probes and 43 unlabeled probes; here, each labeled probe set was used at 1.5 ng/μl and supplemented with 3 ng/μl of unlabeled probes. In all cases, the total probe concentration was maintained at 6 ng/μl. Hybridization, imaging, and signal detection were performed under standard conditions.

2.7. Fluorescence In Situ Hybridization on late Pupal Wings

At 70 h after puparium formation (APF), male flies -carrying a single X chromosome, and therefore a single set of each probe- were immobilized on adhesive tape, removed from their pupal cases, and gently dragged across the tape surface to remove residual cuticle shed during pupal development. Specimens were transferred to a small dish containing deionized water for approximately 1 min to allow wing unfolding. Wings were fixed in 4% formaldehyde in PBTT (1× PBS, 0.1% Tween-20, 0.3% Triton X-100) for 30 min at room temperature (RT), then washed twice in PBS. At 70h APF, wings are enclosed by non-permeable cuticle and therefore require mechanical separation of the dorsal and ventral epithelial layers prior to probe hybridization. To achieve this, wings were transferred in a droplet of deionized water onto a non-adhesive substrate, typically the protective liner of double-sided tape. Excess water was removed, and wings were arranged flat without overlap. A strip of Tesa ECO & ULTRA tape (66 m × 50 mm) —selected for minimal autofluorescence, as it remains attached throughout subsequent staining, mounting, and imaging steps— was gently pressed onto the wings and lifted, ensuring wing adhesion to the tape. The taped wings were overlaid with a second tape strip and pressed gently to preserve structural integrity. The two tape layers were then carefully separated, thereby splitting the dorsal and ventral wing epithelial layers. Finally, the separated wing layers, still tape-mounted, were post-fixed for 20 min at RT in 4% formaldehyde in PBTT.

Wings were then briefly rinsed twice by rapid addition and removal of 500 μL freshly prepared PBS with 0.1% Tween-20 (PTw), followed by two 10 min washes in the same buffer. Wings were incubated for 1 h at 37°C with RNase A diluted at 1:100 in PTw, then transferred to a 0.05 M HCl solution for 5 min to denature DNA and enhance probe penetration.

Wings were washed sequentially for 15 min each in 4:1, 1:1, 1:4, and 0:1 mixtures of PTw and FISH wash buffer (Fwb; 35% formamide, 0.1% Tween-20, 4X SSC). Fwb was prepared fresh before use and stored at 4°C for a maximum of 48 hours.

Wings were incubated with probes diluted to a final concentration of 3ng/μl each in hybridization buffer (50% formamide, 5X SSC, 0.1% Tween-20, 100 μg/mL denatured herring sperm DNA, 50 μg/mL heparin). The samples were heated at 87°C for 20 min in a thermocycler, then cooled gradually to 37°C over 2 h. Hybridization was performed overnight for 12–16 h at 37°C.

Following hybridization, wings were rinsed briefly, then washed twice for 1 h each at 37°C in pre-warmed Fwb. Nuclei were stained with 20 mM Hoechst diluted 1:500 in PTw for 5 min, followed by four 10 min washes in PTw at RT. Wings were mounted in Vectashield, flattened under a coverslip using gentle pressure applied with magnets, and sealed with nail polish.

2.8. Nanorulers Calibration Standards

DNA origami nanorulers (GATTAquant) were employed as ground truth standards for validation of distance measurements. These nanostructures consist of self-assembled DNA origami scaffolds with precisely positioned fluorescent dye attachment sites labeled with 40 ATTO 565 and 40 ATTO 633. Nanorulers with defined inter-dye separations of 120 nm and 160 nm were used to assess distance measurement accuracy, while a design with intermingled 40 ATTO 565 and 40 ATTO 633 dyes, effectively representing zero separation, served as a colocalization standard to evaluate measurement precision. Mounted slides were provided directly by the manufacturer. For each nanoruler type, images were acquired under conditions identical to those used for *Drosophila* wing samples.

2.9. Confocal Imaging

Hoechst, ATTO 565, and ATTO 633 channels were acquired on a Zeiss LSM 780 confocal microscope using a Plan-Apochromat 40× oil immersion objective (NA 1.4). To minimize inter-channel displacement, line switching mode was employed instead of frame switching. Images were acquired with laser power of 0.2% (405 nm), 15% (565 nm), and 27% (633 nm); dwell time of 1.57 μs; voxel size of 103.8 nm × 103.8 nm × 250 nm; 16-bit depth. The pinhole was set to 1 Airy unit (AU) for the 488 nm wavelength.

2.10. Deep Learning–Based Detection of Nuclei and Spots

Nuclei and Oligopaint spots were detected using two independently trained deep learning models applied to individual 2D planes of 3D image stacks, producing volumetric segmentations. Models were trained using EPySeg [17], with data augmentation (rotations, flips, intensity scaling). Probabilistic output maps were thresholded to yield binary masks for nuclei and point coordinates for spots.

2.11. Subpixel 3D Spot Centroid Extraction via Gaussian Fitting

Subpixel 3D coordinates of detected spots were refined using the BigFISH Python library[18], providing initial coordinates and estimated spot sizes for a 3D Gaussian fit. This procedure yielded precise centroids even in the presence of moderate background noise.

2.12. Spot Pairing Across Channels

Before applying chromatic aberration correction, corresponding Oligopaint spots were paired across channels based on spatial proximity. For each image, detected 3D spot coordinates were used to identify the nearest neighbor in the other channel. When applicable, only spots located within segmented nuclei were considered for pairing, thereby excluding non-genomic or non-specific signals. The resulting paired centroids serve as the input for chromatic aberration corrections.

2.13. Linear Chromatic Correction (LCC)

To correct for chromatic aberrations and ensure unbiased distance measurements, we applied a method we named the Linear Chromatic Correction [11,19,20]. The principle underlying this correction is that, for paired FISH Oligopaint centroids, the mean difference between corresponding points should

be zero in the absence of chromatic aberration—the pairs should have no preferred orientation—regardless of whether the points are colocalized or distant [11,19].

The correction procedure begins by computing the difference between each pair of centroids. The mean of these differences is subtracted to center the distribution at zero. For each spatial axis, a linear regression model is fitted to estimate the scaling bias. One set of points is then corrected using these per-axis scaling factors. Finally, any residual global translation is removed to ensure that the corrected differences remain centered at zero.

2.14. Affine Chromatic Correction (ACC)

In addition to LCC, we implemented an independent global method to correct chromatic aberrations by aligning spot coordinates across channels using a 3D affine transformation. This approach compensates for systematic inter-channel discrepancies, including global translations, anisotropic scaling, rotations, and shear.

The procedure operates on paired, colocalized Oligopaint centroids extracted from multiple images sharing the same chromatic aberration profile, acquired on the same microscope within a limited time window. To reduce the influence of mismatched pairs, pairwise distances are filtered using the median and interquartile range (IQR), removing extreme outliers that could strongly bias the affine transformation. Filtered paired centroids from all images are then aggregated to estimate a single global affine transformation matrix.

The affine transformation is estimated by minimizing the squared displacement between corresponding points, allowing for translation, rotation, anisotropic scaling, and shear. The transformation matrix is computed using an implementation adapted from a library by C. Gohlke [21], which solves the resulting least-squares point-set alignment problem using a singular value decomposition (SVD)-based approach. Once estimated, the transformation is applied to all points in one channel to generate corrected coordinates.

The resulting affine transform is stored and can be reapplied to images of non-colocalized spots acquired under the same conditions and within short time intervals, allowing consistent correction across datasets.

2.15. Distance Measurements and Resolution Assessment

Euclidean distances between registered, colocalized spots were calculated in 3D to assess spatial resolution. Distances for non-colocalized spots were similarly computed to quantify inter-locus spacing. Distributions were visualized using histograms and violin plots in Matplotlib and Seaborn [21,22], and robust statistics (median and interquartile range, IQR) were used to limit the influence of outliers and assess chromatic correction precision and spatial organization.

3. Results

3.1. Quantifying and Correcting Chromatic Aberration to Establish the Baseline Resolution

To establish the baseline resolution of our imaging system and quantify the impact of chromatic aberration on distance measurements, we performed colocalization experiments in which the same ~10 kb endogenous genomic region adjacent to the attP18 landing site was labeled with two distinct fluorophores using Oligopaint probes (Figure 1; see Materials and Methods). Because both probe sets target the identical locus, any observed separation reflects technical artifacts rather than true biological distances.

Analysis of 1917 dual-labeled spots revealed substantial systematic offsets between channels before chromatic aberration correction (Figure 2A). The distribution of measured distances showed a median value of 172.9 nm (IQR=135.2-212.3nm), with some spot pairs exhibiting apparent separations exceeding 500 nm. Chromatic aberration was highly anisotropic: lateral (XY) components showed median offsets of 29.7 nm in x and 27.3 nm in y (Figure 2B,C), whereas the axial (Z) component exhibited a median offset of 162.5 nm (Figure 2D). This reflects wavelength-dependent refractive index effects and the lower axial resolution of confocal microscopy.

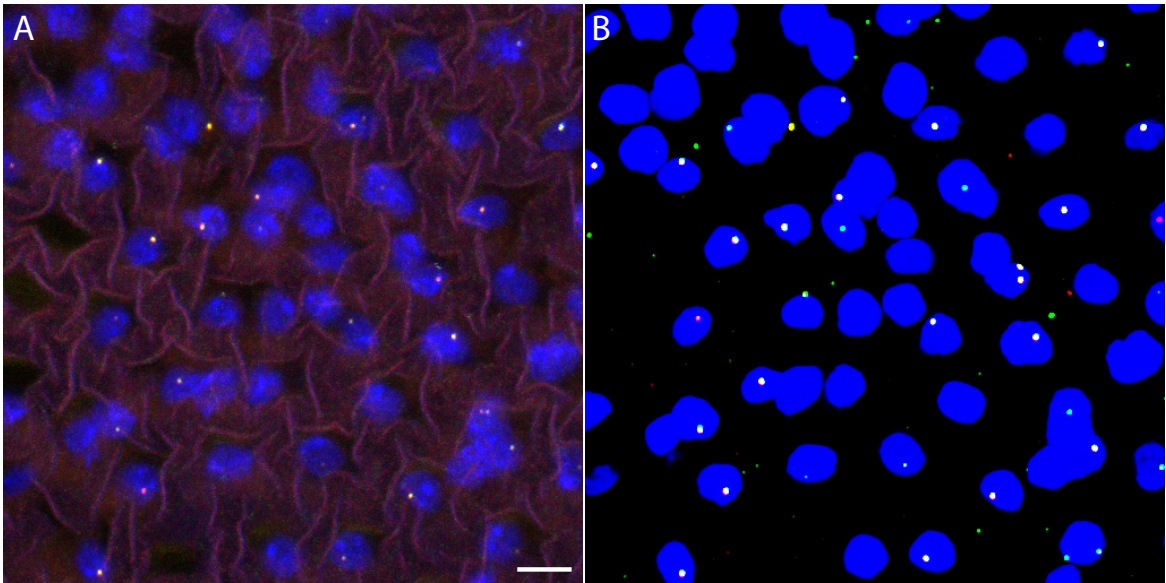


Figure 1. Colocalization control dataset for quantifying chromatic aberration in multi-color Oligopaint FISH. (A) Maximum-intensity projection of a cropped nuclear region from a representative 3D confocal Oligopaint FISH image, in which the same genomic locus is labeled with two fluorophores (ATTO 565, red; ATTO 633, green), serving as a colocalization control. Nuclear DNA is stained with Hoechst (blue). Under ideal conditions, the two FISH signals should spatially overlap. Scale bar, 5 μm . (B) Corresponding maximum-intensity projection of the deep learning-based segmentation output for the same cropped region, showing automated detection of the nucleus and individual Oligopaint FISH spots while preserving channel color coding. This comparison illustrates the ability of the deep learning models to accurately identify nuclear boundaries and colocalized FISH signals in multi-channel 3D microscopy data.

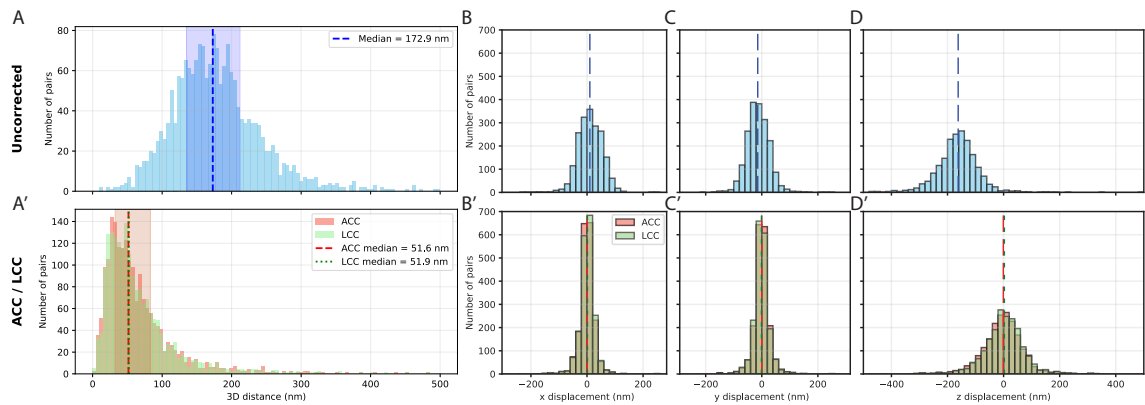


Figure 2. Chromatic aberration in multi-color FISH and its correction. (A) Histogram of 3D inter-channel distances between dual-labeled Oligopaint probes targeting the same genomic locus. Uncorrected measurements show substantial systematic offsets, with a median separation of 172.9 nm (IQR: 135.2–212.3 nm). (A') Same distances after chromatic aberration correction. Affine (ACC, salmon) and linear chromatic aberration correction (LCC, green) both reduce the median 3D separation to ~52 nm, with corresponding interquartile ranges shown as shaded areas. Medians are indicated by dashed (ACC) and dotted (LCC) lines. (B–D) Component-wise displacement histograms for uncorrected data along the x (B), y (C), and z (D) axes. Lateral displacements are small (median ~30 nm), whereas axial displacement is larger (median 162.5 nm). (B'–D') Component-wise displacements after correction using ACC (salmon) and LCC (green). Both methods reduce systematic offsets, with residual median displacements of ~13 nm in x and y and ~40 nm in z, reflecting the practical resolution limit of the system. Histograms are normalized to the number of spot pairs; shaded regions indicate the interquartile range. Data are shown for $n = 1917$ spot pairs.

Application of chromatic aberration correction substantially reduced inter-channel distances. Both the affine (ACC) and linear (LCC) correction methods decreased the median 3D distance from 172.9 nm (uncorrected) to 51.6 nm (IQR: 32.2–82.9 nm) for ACC and 51.9 nm (IQR: 32.2–82.7 nm) for LCC, indicating a marked reduction in systematic offsets (Compare Figures 2A and 2A'). Component-wise median displacements after correction were similar between methods: 12.8–13.1 nm in x, 13.6–13.7 nm in y, and 39.7–40.5 nm in z (Figures 2B'–D'). The close agreement between methods confirms both accurately model chromatic aberration. Residual distances likely reflect localization uncertainty, probe accessibility differences, and stochastic variation in probe positioning.

Analysis of residual chromatic displacements after affine correction revealed no obvious position-dependent variation (Figure S1A,B), indicating that the global correction captures the majority of systematic offsets. Decomposition of the 3D displacement vectors confirmed strong anisotropy, with lateral (XY) components showing no preferential in-plane orientation bias, whereas the axial (Z) component dominated (Figure S1C). The consistent agreement between correction methods throughout subsequent experiments further provides internal validation of measurement accuracy.

3.2. Validation of 3D Distance Measurements Using Genomic Standards and DNA Origami Nanorulers

To evaluate the accuracy of 3D distance measurements under biological conditions, we used a transgene containing two Oligopaint target sites, R1 and R6, labeled with ATTO 565 and ATTO 633, separated by 0 kb, 3 kb, or 10 kb spacers. Chromatic aberrations in these images were corrected using an affine transformation derived from a parallel colocalization experiment acquired under identical imaging conditions. After correction, the measured 3D distances between the two labeled regions increased systematically with spacer length (Figure 3A). For the 0 kb spacer, median distances were 102.8 nm and 101.1 nm for the ACC and LCC methods, respectively. For the 3 kb spacer, median distances increased to 169.0 nm with both methods, and for the 10 kb spacer, they reached 208.1 nm and 209.3 nm.

While the dual-targeted transgene provides qualitative insight into chromatic aberration and distance measurement precision, definitive ground-truth validation of absolute distances was achieved using GATTAquant DNA origami nanorulers. Three types of nanorulers were employed, including one with intermixed ATTO 565 and ATTO 633 fluorophores suitable for registration and resolution assessment, and two with fixed inter-fluorophore separations of 120 nm and 160 nm. For each nanoruler type, we imaged over 4000 individual structures and measured 3D distances after chromatic aberration correction using both ACC and LCC methods. Due to technical constraints, calibration could not be performed using the intermixed nanoruler; instead, registration was derived from dual-color colocalization experiments on wing samples.

For the intermixed nanoruler, the median measured 3D distance was 40.8 nm (IQR: 26.2–66.8 nm) using LCC and 46.1 nm (IQR: 28.6–73.8 nm) using ACC (Figure 3B). For the 120 nm nanoruler, LCC yielded a median distance of 134.5 nm (IQR: 110.3–174.8 nm), while ACC measured 145.9 nm (IQR: 113.1–195.2 nm). For the 160 nm nanoruler, LCC produced a median distance of 175.4 nm (IQR: 142.5–304.8 nm), with ACC yielding 183.3 nm (IQR: 142.7–335.1 nm) (Figure 3B). Both correction approaches accurately reproduced expected distances –and distance differences– across nanoruler sizes, confirming measurement robustness.

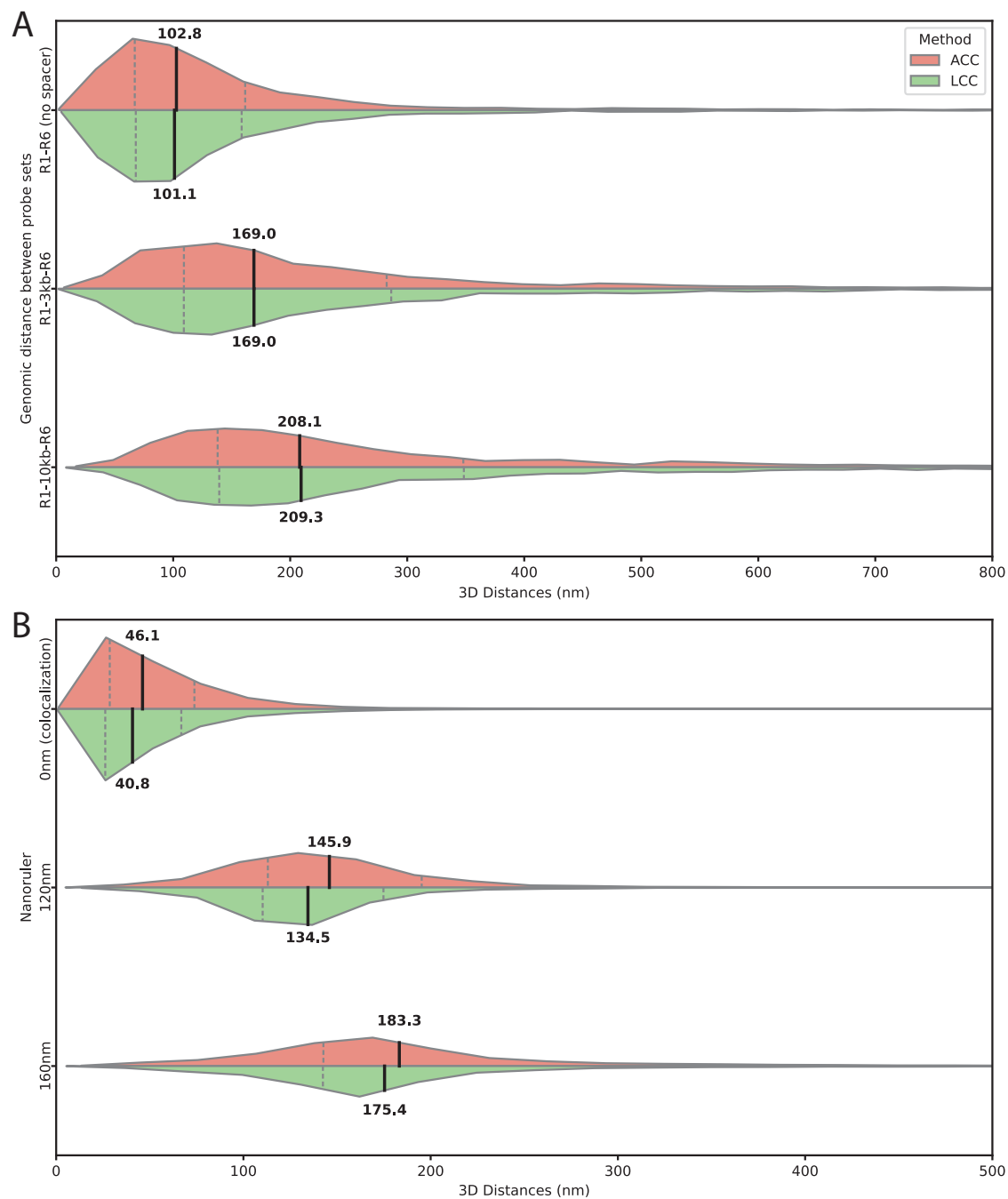


Figure 3. Validation of 3D distance measurements using genomic spacers and DNA origami nanorulers. (A) Measured 3D distances between two Oligopaint probe sets targeting R1 and R6 sequences flanking defined spacers of 0 kb, 3 kb, and 10 kb. Distances were measured after chromatic aberration correction using either the affine (ACC, salmon) or the linear chromatic aberration correction (LCC, green). Horizontal split violin plots show the distribution of distances for each spacer length and correction method, with medians and interquartile ranges indicated by internal lines. Distances increase systematically with spacer length, and the close agreement between methods demonstrates the accuracy of both chromatic aberration corrections. Numbers of analyzed spot pairs were $n = 2667$ (0 kb), $n = 2363$ (3 kb), and $n = 2606$ (10 kb). Throughout this and subsequent figures, x axes are truncated for visualization purposes; measurements outside the displayed range are included in the analysis. (B) 3D distances measured for GATTAquant DNA origami nanorulers with defined fluorophore separations of intermixed 0 nm (colocalization control), 120 nm, and 160 nm. Horizontal split violin plots show distance distributions after chromatic aberration correction using ACC (red) and LCC (green). Both methods accurately reproduce the expected distances, confirming that chromatic aberration correction reliably captures absolute spatial separations. Numbers of analyzed spot pairs were $n = 17,239$ (0 nm), $n = 6,460$ (120 nm), and $n = 4,290$ (160 nm).

3.3. Target Sequence Length And Composition Affect Spatial Resolution

To assess how target sequence length –at constant probe density– affects resolution, we used transgenes containing the full 2 kb R6 region and truncated versions of 1 kb and 500 bp, co-labeled with ATTO 565 and ATTO 633. For the 2 kb construct, median residual distances were 47.7-48.7 nm (Figure 4A). The 1 kb construct showed substantially increased distances of 76.3-84.9 nm (Figure 4B), while the 500 bp construct yielded 65.8-69.0 nm (Figure 4C).

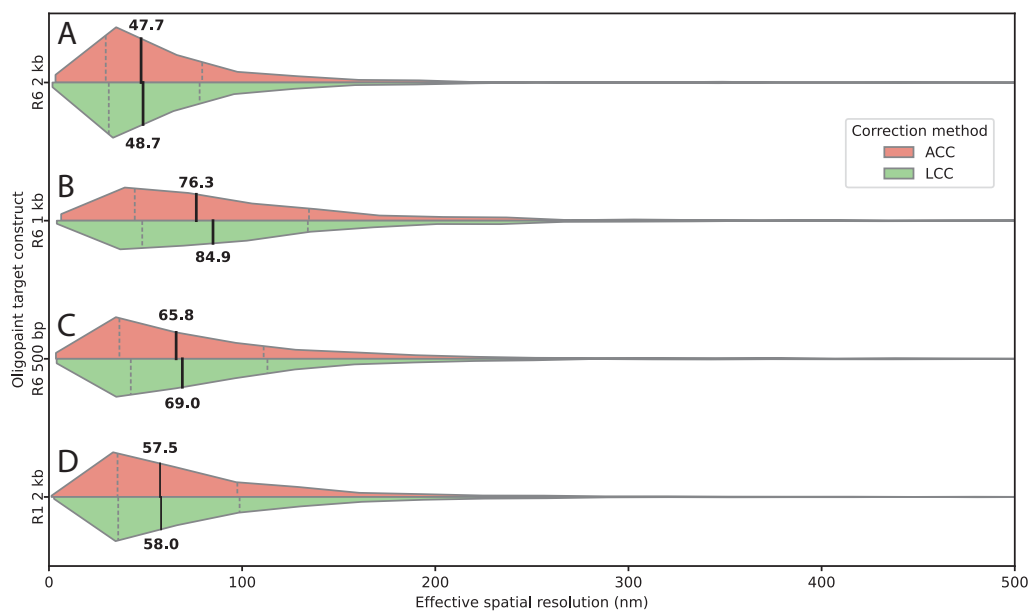


Figure 4. Target sequence length and composition determine effective spatial resolution in FISH. (A–D) Horizontal split violin plots of 3D inter-channel distances measured for Oligopaint probes targeting the same genomic locus, after affine (ACC, salmon) and linear chromatic aberration correction (LCC, green). (A) R6 2 kb; (B) R6 1 kb; (C) R6 500 bp; (D) R1 2 kb at the same locus but with a different sequence. Shortening the target sequence increases the measured inter-channel distance, indicating reduced effective spatial resolution. The alternative 2 kb sequence (R1) yields slightly larger distances compared to R6, highlighting sequence-dependent effects on localization precision. Medians and interquartile ranges are indicated on the plot. Numbers of analyzed spot pairs were: R1 2 kb, $n = 5,409$; R6 1 kb, $n = 803$; R6 2 kb, $n = 1,042$; R6 500 bp, $n = 641$.

Target length strongly affected detection yield. Average detected spots per image decreased from 190.5 spots (2kb) to 57.4 (1 kb) and 34.0 (500bp) (Table 1), reflecting reduced probe binding sites and signal intensity.

To assess whether factors other than target length contribute to measurement performance, we analyzed an additional 2 kb construct, R1, inserted at the same genomic locus but exhibiting a different DNA sequence compared to R6. Importantly, this construct was targeted by the same number of Oligopaint probes as the original 2 kb construct, ensuring comparable probe density and labeling strategy. Despite these similarities, the alternative sequence yielded fewer detected pairs per image (128.8 versus 190.5, Table 1) and slightly larger residual inter-channel distances (compare Figures 4A and 4D). This indicates that sequence-dependent effects —potentially related to hybridization efficiency or local chromatin organization— can modulate measurement performance even when target length and probe number are constant.

Table 1. Average number of detected pairs per image for different genomic constructs.

Construct	Target size	Detected pairs per image
R6	2 kb	190.5
R6	1 kb	57.4
R6	500 bp	34
R1	2 kb	128.8

3.4. Probe Density Influences Detection Efficiency and Spatial Resolution

To evaluate the impact of probe density on measurement performance, we systematically varied the fraction of fluorescently labeled Oligopaint probes targeting a fixed 2 kb genomic region by adjusting the ratio of labeled to unlabeled probes while keeping the total probe mass constant. This approach preserves hybridization conditions while progressively reducing the number of fluorescent probes contributing to the detected signal.

Three labeling conditions were examined, corresponding to approximately 100%, 66%, and 50% fluorescent labeling. Under these conditions, the 2 kb R6 locus could be covered by up to 87, 57, or 43 fluorescently labeled Oligopaint probes, with the remainder of the locus filled by 0, 30, or 43 unlabeled Oligopaint probes. Fully labeled conditions yielded an average of 190.5 spots per image (Table 2). Reducing the labeled fraction to ~66% decreased detection to 133.2 spots per image, and further reducing it to ~50% resulted in only 51.1 spots per image (Table 2). Reducing the fraction of labeled probes lowers fluorescence signal, diminishing both detection efficiency and localization precision, highlighting the strong dependence of measurement performance on the number of fluorescent probes present.

Spatial resolution also worsened with decreasing labeling fraction, although less sharply than detection efficiency. After affine chromatic aberration correction, the median residual distance increased from 45.4 nm at 100% labeling to 59.8 nm at ~66% labeling and 77.3 nm at ~50% labeling (Figure 5). Comparable values were obtained using linear chromatic aberration correction, with median residual distances of 46.8 nm, 60.4 nm, and 80.8 nm, respectively (Figure 5), demonstrating that this trend is robust across correction methods.

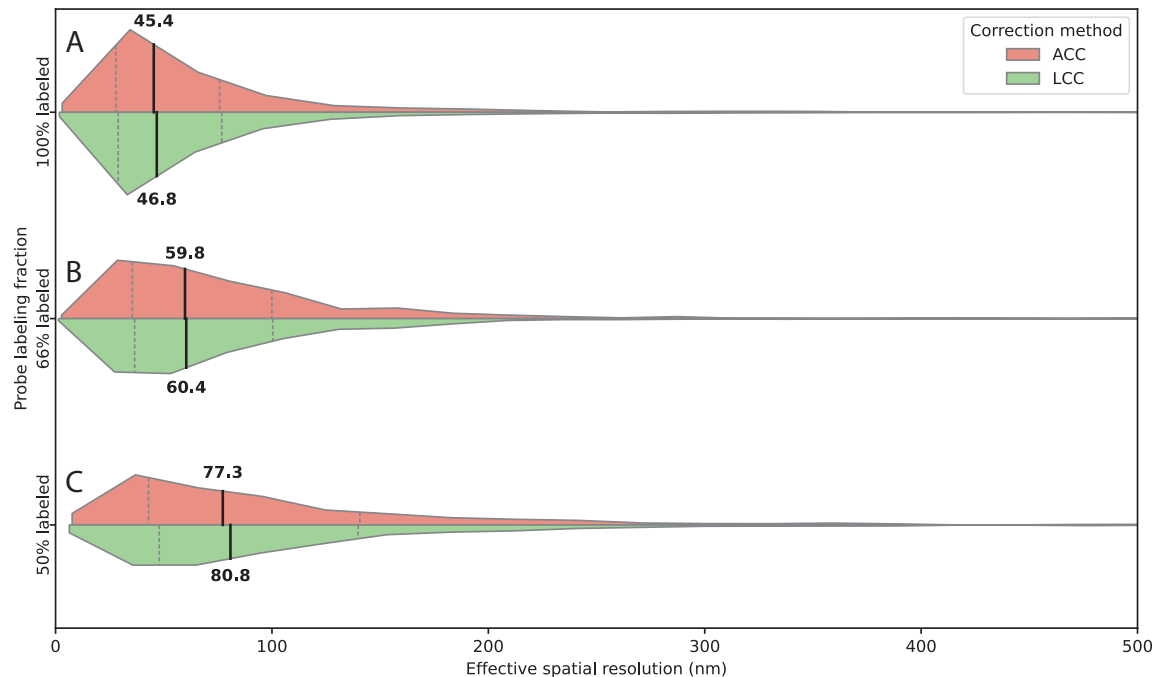


Figure 5. Probe density modulates detection efficiency and effective spatial resolution. (A–C) Horizontal split violin plots showing distributions of 3D inter-channel distances measured after affine (ACC, salmon) and linear chromatic aberration correction (LCC, green) for decreasing fractions of labeled probes targeting the same 2 kb genomic region. (A) 100% labeled probes, (B) ~66% labeled probes, (C) ~50% labeled probes. Medians and interquartile ranges are indicated by the internal quartile lines. Reducing the fraction of fluorescently labeled probes lowers effective spatial resolution, as shown by the systematic increase in median distances with decreasing labeling fraction. This trend is consistent across both correction methods, demonstrating the robustness of the measurements. Numbers of analyzed spot pairs were: 100% labeled, $n = 2,667$; 66% labeled, $n = 2,131$; 50% labeled, $n = 971$.

Table 2. Detection efficiency as a function of probe labeling fraction.

Labeled probe fraction	Fluorescent / unlabeled oligonucleotides per locus	Detected pairs per image
100%	87 / 0	190.5
66%	57 / 30	133.2
50%	43 / 43	51.1

4. Discussion

4.1. FISH-Dist Enables Accurate Short-Range Genomic Distance Measurements

We developed FISH-Dist, an automated pipeline optimized for precise 3D distance quantification in FISH imaging at the kilobase scale. FISH-Dist combines deep learning-based spot detection, sub-pixel 3D localization, and dual chromatic aberration correction methods (affine and linear), with accuracy validated using calibrated standards such as DNA origami nanorulers.

While many existing FISH tools focus on large-scale chromosomal structures (hundreds of nanometers to micrometers), biologically critical interactions—such as enhancer–promoter contacts, local chromatin loops, regulatory clustering within topologically associating domains, DNA damage foci, and homologous chromosome pairing—occur at much shorter distances (10–50 kb). At these scales, chromatic aberration and localization precision are limiting factors. In our system, uncorrected inter-channel offsets exceeded 170 nm, with axial components approaching 160 nm —comparable to or larger than many true biological separations. FISH-Dist reduces these errors to approximately 50 nm for colocalized targets.

Additionally, we demonstrate that 2 kb probe targets offer an optimal balance between detection efficiency and spatial compactness, achieving ~45 nm resolution. These advances enable robust, quantitative measurement of short-range genomic interactions that are typically inaccessible to conventional FISH analysis pipelines.

4.2. Probe Design Guidelines Emerge from Systematic Parameter Testing

Our systematic evaluation of probe design parameters yielded several actionable insights for experimental optimization. The relationship between target sequence length and resolution reveals a critical trade-off: longer sequences improve detection efficiency but compromise localization precision. Our 2 kb construct achieved optimal 45 nm resolution with robust detection (190.5 spots/image). Shortening to 1 kb degraded resolution to 75–80 nm despite reduced target size. The 500 bp construct showed slightly better resolution (~70 nm) than 1 kb but both exhibited markedly reduced detection efficiency.

These findings indicate that competing effects exist, with shorter targets providing compact labeling but reduced signal intensity and fewer probe binding sites, both of which compromise localization. The 2 kb length represents a favorable compromise. Importantly, sequence composition influences performance even at fixed length, as the R1 construct showed reduced efficiency despite having the same length and number of probes as R6. Factors such as GC content, secondary structure, or chromatin accessibility likely modulate hybridization efficiency.

Probe density experiments showed reducing labeling from 100% to 66% caused moderate degradation (45 nm to 60 nm resolution, 30% detection loss). Further reduction to 50% caused substantial degradation (80 nm resolution, 75% detection loss). Even with probe excess, reducing fluorescent fraction diminishes photon counts, degrading centroid localization. For high spatial accuracy, we recommend maintaining at least 57 oligos/fluorophores per probe set spanning 2kb (corresponding to ≥ 66% labeling in our conditions).

Our optimization was conducted in specific genomic and cellular contexts. Absolute values likely depend on chromatin accessibility, nuclear architecture, and cell type. However, general principles—optimal target length balancing signal and compactness, sequence-dependent performance, sensitiv-

ity to probe density— should apply broadly. We encourage pilot optimization in specific biological systems.

4.3. Methodological Considerations and Limitations

Validation with DNA origami nanorulers provides ground-truth accuracy under idealized conditions, while transgene-based measurements in fixed biological samples demonstrate accuracy in realistic contexts where chromatin organization and nuclear architecture are preserved. This combination represents a strength of our approach, as it bridges artificial calibration standards and biologically relevant measurements. Nevertheless, probe accessibility can vary with chromatin state, DNA sequence, and cell cycle stage. Although fixation stabilizes chromatin structure, labeled DNA retains conformational heterogeneity reflecting the ensemble of chromatin configurations present at fixation, such that measured distances represent spatial averages rather than rigid, single conformations.

Our chromatic aberration correction assumes a stable optical transformation between channels during image acquisition, and performance may degrade in samples with substantial axial drift or stage movement. In addition, the deep learning-based spot detection relies on training data representative of the target application. Although the trained model generalizes well across the range of probes, labeling densities, and imaging conditions tested here, users working with substantially different probe types or imaging modalities may benefit from retraining or fine-tuning the model. Due to their size, the training datasets and scripts are not publicly distributed but are available upon request.

Throughout validation and optimization, we employed two conceptually distinct chromatic aberration correction approaches, affine and linear, which consistently produced nearly identical corrected distances across all tested conditions, typically differing by only a few nanometers. This close agreement indicates that both methods accurately model the chromatic aberration in our imaging system and provides an internal consistency check that strengthens confidence in measurement accuracy.

Although the two approaches differ primarily in workflow rather than performance, maintaining both in the analysis pipeline offers important diagnostic value. Agreement between methods, as observed here, confirms optical stability and the adequacy of global correction models, whereas substantial divergence would indicate optical instability or model inadequacy and prompt re-evaluation of imaging conditions. Consistent with this interpretation, spatial analysis of residual displacements revealed relatively uniform chromatic aberration fields across the imaging volume, indicating that global correction models capture most systematic offsets in our system. Users are nevertheless encouraged to assess spatial uniformity in their own setups to determine the most appropriate correction strategy.

4.4. Comparison with Existing Methods

Several computational tools exist for FISH image analysis, each with distinct design priorities and validation strategies. Classical approaches based on watershed segmentation or intensity thresholding remain widely used, but they often require manual parameter tuning or segmentation, which is error-prone and can lead to substantial over-segmentation, particularly under variable background conditions or in regions with densely packed spots.

Particle tracking frameworks such as TrackMate [23] provide flexible and sophisticated spot detection and localization, capable of handling multiple spots per frame. However, in noisy images such as ours, they can generate excessive false-positive detections, and while post-processing can mitigate this, the procedure is slow and not optimized for high-throughput FISH distance measurements. TrackMate also does not explicitly address chromatic aberration correction, which is critical for accurate short-range distance quantification.

FISH-quant [18] is the most directly comparable tool. However, it was developed primarily for RNA FISH and transcript quantification rather than precise genomic distance measurements and does not include chromatic aberration correction. In contrast, FISH-Dist is specifically designed for short-range genomic distances, implements two complementary chromatic aberration correction meth-

ods (linear and affine), and has been rigorously validated using both transgenic constructs and DNA origami nanorulers. Its systematic evaluation of probe design parameters provides increased confidence in quantitative accuracy compared with single-method or less extensively validated pipelines.

5. Conclusions

FISH-Dist has been developed with accessibility and reproducibility in mind. The complete analysis pipeline is available as open-source software, accompanied by comprehensive documentation, example datasets, and tutorials to facilitate adoption. The tool is provided as a command-line application and includes pre-trained deep learning models that support many standard use cases, while also allowing retraining or fine-tuning for specialized experimental conditions.

More broadly, FISH-Dist provides a validated and automated framework for accurate 3D distance measurements in genome organization studies. By integrating chromatic aberration correction, robust spot detection, and quantitative distance analysis into a unified workflow, the pipeline lowers technical barriers to precise spatial measurements and enables reproducible, high-throughput analysis across diverse experimental systems.

Supplementary Materials: The following supporting information can be downloaded at the website of this paper posted on Preprints.org, Figure S1: Spatial distribution of residual chromatic displacements after correction.

Author Contributions: Conceptualization: B.P., T.G.; Software: B.A.; Formal Analysis: B.A., E.C., B.C., and C.S.M.; Investigation: E.C., B.C., and C.S.M.; Data Curation: E.C., B.C., and C.S.M.; Writing: B.A.; Review & Editing: B.A., B.P., E.C., B.C., and C.S.M.; Supervision: B.P.; Project Administration: B.P.; Funding Acquisition: B.P., T.G.

Funding: This work was supported by the Fondation pour la Recherche Médicale (Equipe FRM EQU202103012586, B.P.), the ANR (ANR-22-CE12-0017, B.P., T.G.), and the France 2030 programme (ANR-25-CHBS-0014, B.P.)

Data Availability Statement: The FISH-Dist pipeline is available as open-source software under the BSD license at [the FISH-Dist GitHub repository](#). Pre-trained TensorFlow deep learning models are automatically downloaded upon first use. Comprehensive tutorials and documentation are available at [the FISH-Dist documentation website](#). Due to their large size, example and training datasets are not publicly distributed but can be made available upon request.

Acknowledgments: We thank the imaging facility at IBDM, member of the National Infrastructure France-BioImaging (<https://ror.org/01y7vt929>) supported by the French National Research Agency (ANR-24-INBS-0005 FBI BIOGEN); and all members of the lab for helpful discussions and technical support.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Popay, T.M.; Dixon, J.R. Coming full circle: On the origin and evolution of the looping model for enhancer–promoter communication. 298. Publisher: Elsevier, <https://doi.org/10.1016/j.jbc.2022.102117>.
2. Kim, S.; Shendure, J. Mechanisms of Interplay between Transcription Factors and the 3D Genome. 76, 306–319. <https://doi.org/10.1016/j.molcel.2019.08.010>.
3. Beliveau, B.J.; Boettiger, A.N.; Avendaño, M.S.; Jungmann, R.; McCole, R.B.; Joyce, E.F.; Kim-Kiselak, C.; Bantignies, F.; Fonseca, C.Y.; Erceg, J.; et al. Single-molecule super-resolution imaging of chromosomes and in situ haplotype visualization using Oligopaint FISH probes. 6, 7147. <https://doi.org/10.1038/ncomms8147>.
4. Bolzer, A.; Kreth, G.; Solovei, I.; Koehler, D.; Saracoglu, K.; Fauth, C.; Müller, S.; Eils, R.; Cremer, C.; Speicher, M.R.; et al. Three-Dimensional Maps of All Chromosomes in Human Male Fibroblast Nuclei and Prometaphase Rosettes. 3, e157. <https://doi.org/10.1371/journal.pbio.0030157>.
5. Fields, B.D.; Nguyen, S.C.; Nir, G.; Kennedy, S. A multiplexed DNA FISH strategy for assessing genome architecture in *Caenorhabditis elegans*. 8, e42823. <https://doi.org/10.7554/eLife.42823>.
6. Shirley, J.W.; Ty, S.; Takebayashi, S.I.; Liu, X.; Gilbert, D.M. FISH Finder: a high-throughput tool for analyzing FISH images. 27, 933–938. <https://doi.org/10.1093/bioinformatics/btr053>.
7. Lieberman-Aiden, E.; van Berkum, N.L.; Williams, L.; Imakaev, M.; Ragozy, T.; Telling, A.; Amit, I.; Lajoie, B.R.; Sabo, P.J.; Dorschner, M.O.; et al. Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. 326, 289–293. <https://doi.org/10.1126/science.1181369>.

8. Mateo, L.J.; Murphy, S.E.; Hafner, A.; Cinquini, I.S.; Walker, C.A.; Boettiger, A.N. Visualizing DNA folding and RNA in embryos at single-cell resolution. *568*, 49–54. <https://doi.org/10.1038/s41586-019-1035-4>.
9. Harke, J.; Lee, J.R.; Nguyen, S.C.; Arab, A.; Rakowiecki, S.M.; Hugelier, S.; Paliou, C.; Rauseo, A.; Yunker, R.; Xu, K.; et al. Multiple allelic configurations govern long-range Shh enhancer-promoter communication in the embryonic forebrain. *84*, 4698–4710.e6. <https://doi.org/10.1016/j.molcel.2024.10.042>.
10. Gizzi, A.M.C.; Cattoni, D.I.; Fiche, J.B.; Espinola, S.M.; Gurgo, J.; Messina, O.; Houbbron, C.; Ogiyama, Y.; Papadopoulos, G.L.; Cavalli, G.; et al. Microscopy-Based Chromosome Conformation Capture Enables Simultaneous Visualization of Genome Organization and Transcription in Intact Organisms. *74*, 212–222.e5. <https://doi.org/10.1016/j.molcel.2019.01.011>.
11. Barinov, L.; Ryabichko, S.; Bialek, W.; Gregor, T. Transcription-dependent spatial organization of a gene locus. [2012.15819].
12. Urban, E.A.; Chernoff, C.; Layng, K.V.; Han, J.; Anderson, C.; Konzman, D.; Johnston, R.J. Activating and repressing gene expression between chromosomes during stochastic fate specification. *42*, 111910. <https://doi.org/10.1016/j.celrep.2022.111910>.
13. Galouzis, C.C.; Prud'homme, B. Transvection regulates the sex-biased expression of a fly X-linked gene. *371*, 396–400. <https://doi.org/10.1126/science.abc2745>.
14. Delker, R.K.; Munce, R.H.; Hu, M.; Mann, R.S. Fluorescent labeling of genomic loci in Drosophila imaginal discs with heterologous DNA-binding proteins. *2*. <https://doi.org/10.1016/j.crmeth.2022.100175>.
15. Viets, K.; Sauria, M.E.G.; Chernoff, C.; Viales, R.R.; Echterling, M.; Anderson, C.; Tran, S.; Dove, A.; Goyal, R.; Voortman, L.; et al. Characterization of Button Loci that Promote Homologous Chromosome Pairing and Cell-Type-Specific Interchromosomal Gene Regulation. *51*, 341–356.e7. <https://doi.org/10.1016/j.devcel.2019.09.007>.
16. Arbor, A. Labeling Protocol for myTags Immortal Libraries.
17. Aigouy, B.; Cortes, C.; Liu, S.; Prud'Homme, B. EPySeg: a coding-free solution for automated segmentation of epithelia using deep learning. *147*. <https://doi.org/10.1242/dev.194589>.
18. Imbert, A.; Ouyang, W.; Safieddine, A.; Coleno, E.; Zimmer, C.; Bertrand, E.; Walter, T.; Mueller, F. FISH-quant v2: a scalable and modular tool for smFISH image analysis. *28*, 786–795. <https://doi.org/10.1261/rna.079073.121>.
19. Brückner, D.B.; Chen, H.; Barinov, L.; Zoller, B.; Gregor, T. Stochastic motion and transcriptional dynamics of pairs of distal DNA loci on a compacted chromosome. *380*, 1357–1362. <https://doi.org/10.1126/science.adf5568>.
20. Leiwe, M.N.; Fujimoto, S.; Imai, T. Post hoc Correction of Chromatic Aberrations in Large-Scale Volumetric Images in Confocal Microscopy. *15*. <https://doi.org/10.3389/fnana.2021.760063>.
21. Waskom, M.L. seaborn: statistical data visualization. *Journal of Open Source Software* **2021**, *6*, 3021. <https://doi.org/10.21105/joss.03021>.
22. Hunter, J.D. Matplotlib: A 2D graphics environment. *Computing in Science & Engineering* **2007**, *9*, 90–95. <https://doi.org/10.1109/MCSE.2007.55>.
23. Ershov, D.; Phan, M.S.; Pylvänäinen, J.W.; Rigaud, S.U.; Le Blanc, L.; Charles-Orszag, A.; Conway, J.R.W.; Laine, R.F.; Roy, N.H.; Bonazzi, D.; et al. Bringing TrackMate into the era of machine-learning and deep-learning. <https://doi.org/10.1101/2021.09.03.458852>.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.