

HD-FISH, a glimpse into chromatin organization

Emiliano Izquierdo

— written on Monday, July 01, 2013

Abstract: Chromosomes change their topology throughout the cell cycle varying its position and degree of compaction. This features have been observed to influence gene expression. Furthermore, repeated patterns in chromatin organization have been observed, suggesting a direct link between the three-dimensional architecture of the genome and gene activity. Recent approaches to chromatin organization start to combine both experimental and computational models, allowing a more holistic view of the variability in structure and function of the genome. New advancements to microscopy methods like DNA-FISH allow for the direct measurement of the nuclear three-dimensional topology. In this work we advance this recent improvements to DNA-FISH by developing a semi-automated workflow that facilitates the identification of genomic regions accurately in the three-dimensional space of the nucleus. Finally we assess this workflow to determine the organization of the human chromosome 17.

Table of Contents

1. Introduction

- Theoretical approaches to chromatin organization
- Experimental techniques capturing chromatin structure
- Motivation
- Objectives
- System under investigation

2. Experimental Approach

- Cell culture and fluorescence *in situ* hybridization.
- Image acquisition
- Image analysis

3. Results and discussion

- Precise spot localization
- Chromatic aberration corrections
- SpotterUI
- 4 loci experiment

4. Concluding remarks

5. Acknowledgements

6. References and relevant literature

Introduction

The control of gene expression in eukaryotes can be viewed as the integrated response of mechanisms working at different hierarchical levels. Gene transcription is regulated from the sequence level to the chromatin level, to the nuclear level (van Driel et al. 2003) (see **Figure 1. Top**). At this top-most layer, the higher order topology of the genome has proven to play an important role, as meaningful patterns of organization have been identified; after mitosis, chromosomes decondense into fibers which remain confined into well defined and isolated compartments called *chromosome territories* (Cremer et al. 2010). Generally gene-poor chromosomes are more frequently positioned in the outer regions of the nucleus, whereas gene-rich chromosomes adopt a more internal localization (Croft et al. 1999, see **Figure 1-1**). Moreover, motion of genes along the radial axis of the nucleus has been observed upon gene activation, movements either toward the nuclear interior or toward the nuclear periphery, however, other observations show that the former is not exclusively a repressive location (Andrusis et al. 1998, Casolari et al. 2005, Meister et al. 2010). Alongside, distributed throughout the nucleus, sets of activated genes are spatially arranged into discrete foci termed *transcription factories*—dedicated to the expression of specific combinations of genes, even from distinct chromosomes—(Spilianakis et al. 2005, Osborne et al. 2007, Schoenfelder et al. 2010). These particular arrangements indicate that the associations at the transcriptional hot spots does not happen at random, but genes have preferential partners, commonly in a tissue specific manner (Osborne et al. 2007, see **Figure 1-2**). In addition, significant changes in spatial organization of genomes are triggered during cellular differentiation reflecting a dramatic change in gene expression (Kim et al. 2004).

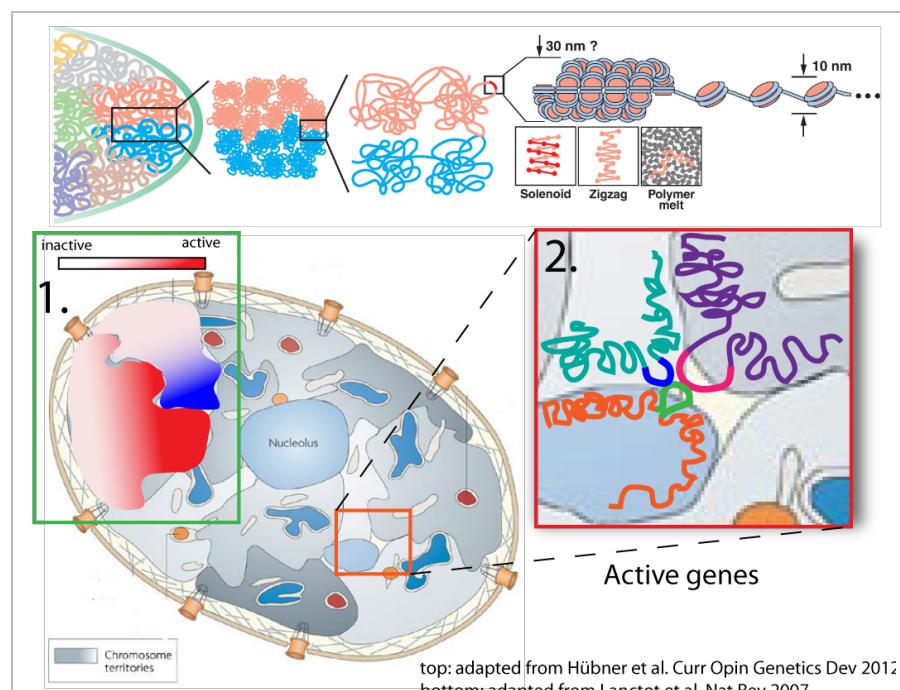


Figure 1. Top: Schematic illustration of the different layers of organization in the mammalian nucleus. From left to right: Chromosomes are organized in chromosome territories which are composed of smaller globular structures that intermingle at some extent. These globules are compacted loops of a 30 nm fiber with a solenoid, zigzag or polymer melt configuration. At the lowest level, chromatin is organized in a 10 nm, *beads on a string* fiber consisting of nucleosomes. **Bottom: 1.** A highly compartmented nucleus with non-random chromatin organization. Chromosomes are organized into territories in the

interphase nucleus. They are further arranged such that active regions are more commonly localized towards the nuclear interior whereas inactive portions tend to adopt a more external position **2**. Actively transcribed genes associate at the periphery of chromosome territories. Intra and interchromosomal contacts bridge coregulated genes as well as enhancer and promoter pairs.

The higher-order structure of chromatin offers an extra level of epigenetic regulation. Many factors may be involved in the modulation of nuclear conformations. Both local and long-range biochemical interactions between specific genomic regions, as well as the anchoring of chromatin domains to the nuclear lamina, shape and constrain the available topologies for the fiber to fold into, emphasizing the link between the linear structure and the spatial chromatin arrangement (Marshall et al. 1997, Chubb et al. 2002, Bickmore & van Steensel 2013).

Theoretical approaches to chromatin organization

The overall nuclear structure can be characterized by a set of features that range from cell cycle stage, spatial compartmentalization, degree of compaction, radial positioning within the nucleus, extent of intermingling, genomic to physical distance scaling, among others. These characteristics, together with the view of the chromatin fiber as an ensemble of nucleosomes —assuming a non-random spatial configuration — are suitable to be approached with polymer models, a unifying quantitative framework (Tark-Dame et al. 2011). These models represent chromosome fibers made up of monomer units connected by a flexible connector, this *beads-on-a-string* representation is *coarse-grained* at the three-dimensional level, where descriptions for the small scale details are eliminated in favor of an overall structure (Heermann et al. 2012).

Computer simulations from polymer models demonstrate that some features of organization can be reproduced, such as the organization into chromosome territories (Lieberman-Aiden et al. 2009). However, other model predictions deviate from experimental observations (i.e. the leveling-off of the linear to spatial scaling or the different scaling regimes at varying genomic lengths). This discrepancies provide evidence for additional molecular mechanisms contributing to chromatin folding which can be taken into account by accordingly extending the model assumptions.

One characteristic feature of a polymer model is the mean-squared spatial end-to-end distance $\langle R^2 \rangle$ which as a function of its genomic distance g scales $\langle R^2 \rangle \sim g^{2\nu}$, the exponent ν varies given the dynamics of each polymer representation, such as random walk, self-avoiding walk, or globular state (Tark-Dame et al. 2012); R is the spatial distance between two genomic regions linearly separated by g (see **Figure 2-1. Top**). For example, in the most simple random-walk model, the physical distance scales monotonically with increasing genomic separation, therefore, unable to match any of the experimentally observed nuclear patterns of organization (see **Figure 2-2. Top**). More recent polymer models already started to integrate further improvements to extend rather basic polymer models as are the *swollen coil* or the *equilibrium globule* (see **Figure 2-3. Top**). Models like the *fractal globule* (Lieberman-Aiden et al. 2009), the *random loop* model (Mateos-Langerak et al. 2009), and the recent *string and binders switch* model (Barbieri et al. 2013) are able to reproduce different aspects of the observed nuclear organization.

To be tested for prediction accuracy, all these approximations require precise measurements of the scaling relation between the genomic and the physical distance. Hence, it is worth stressing out the importance of better and more extensive experimental techniques to acquire quantitative data and to supply support to adapt the previously mentioned theoretical representations.

Experimental techniques capturing chromatin structure

Recently, the development and refinement of techniques that allow the mapping of genomic regions onto the spatial context of the nucleus—both in bulk and up to the single cell level—has allowed to have quantitative measures to describe the principles of chromatin organization. High-coverage methods derived from 3C (Chromatin Conformation Capture, Dekker et al. 2002) use formaldehyde crosslinking snapshots on the bulk to quantify distance between genomic loci, interpreted from contact probabilities; closely localized regions are found to be crosslinked more frequently.

In a more direct strategy, single-cell distance measurements can be acquired from the imaging of DNA-FISH experiments (DNA fluorescence *in situ* hybridization), where fluorescently labeled probes are designed to target specific genomic loci. Their positions are identified by means of 3D fluorescence-based microscopy, and subsequently, the distances between loci are measured (**see Figure 2 bottom panel**).

Both approaches are complementary, and both provide useful information in terms of nuclear topology which can be unified and reconciled with theoretical models. One advantage of DNA-FISH experiments is that cell-to-cell variability can be captured in a population of fixed cells as snapshots of different chromatin configurations crucial for characterizing the differences in chromosomal conformations and locations.

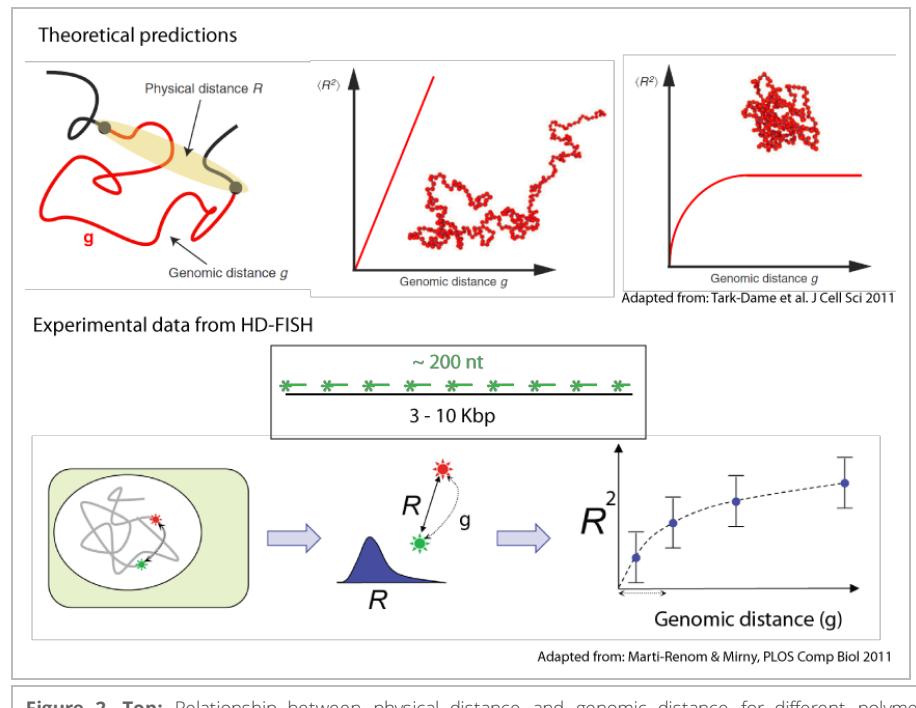


Figure 2. Top: Relationship between physical distance and genomic distance for different polymer models. Schematic of the relation between R which is the physical distance between two points on an interphase chromosome, and the genomic distance g . Random walk model, predicts that the mean squared distance $\langle R^2 \rangle$ increases linearly with increasing genomic distance g . While the looped chromatin model, in contrast, predicts that $\langle R^2 \rangle$ reaches a plateau at large genomic distances. **Bottom:** Light microscopy by fluorescent *in situ* hybridization; in our approach, *HD-FISH*, genomic regions are targeted by labeled amplicons of about 200 nt covering a region between 3 and 10 Kbp. In FISH experiments, chromatin folding shows a behavior where $\langle R^2 \rangle$ reaches a plateau indicating the folding of chromosomes into territories.

Motivation

In a recent advancement, named *HD-FISH*, Bienko and collaborators (Bienko et al. 2012) reduced by one order of magnitude the effective target size of DNA-FISH probes

thereby increasing the resolution up to the diffraction limit, this *spotting* strategy allows the position localization to be estimated with higher accuracy.

Here we propose the refinement of the current 3D DNA-FISH techniques and its extension by Bienko and collaborators by accurately localizing single genomic regions up to the sub-pixel resolution. This comprehensive measurements on the landscape of interacting genomic regions will further help us to gain deeper insight into the heterogeneity and functional organizing principles of the genome.

Objectives

The main objective of this study was to use images generated from *HD-FISH* experiments to precisely measure the distance between various genetically separated loci up to the sub-pixel resolution. More specifically:

1. Generate a data analysis pipeline in MATLAB for feature localization and image processing of *HD-FISH* experiments
2. Once the coarse localization of the loci is identified, the effect of chromatic aberrations should be corrected, allowing the precise localization of the targeted regions
3. Distances are then calculated from corrected loci positions

System under investigation

DNA has to be optimally folded to fit in the reduced volume of the nucleus conserving functional three-dimensional structure. This arrangement of chromosomes shows to be highly dynamic, changes arrangement throughout the cell cycle but despite the variability in organization among cells, chromatin preserves common spatial patterns at some layers of organization. The chromosomes change from a highly condensed structures during mitosis, to form the more or less condensed *chromosome territories* during interphase.

Here we approached the distribution of distances between genomic regions during interphase regardless of the cell type, in this manner obviating tissue specific differences.

For a population of cells we performed *HD-FISH*, image acquisition of several *fields-of-view* and carried out image and statistical data analysis. Moreover, we integrated the data analysis workflow in a custom based semi-automated graphical implementation using MATLAB.

Experimental Approach

Cell culture and fluorescence in situ hybridization.

We grew both *human mammary epithelial cells* (hTERT-HME1 mammary) and *primary human foreskin fibroblasts* (ATCC CRL 2097) in Dulbecco's modified Eagle's medium with Glutamax (DMEM, Life Technologies) supplemented with penicillin/streptomycin and 10% FBS. To fix the cells, we followed the protocol of Raj et al. 2008. We fixed the cells for 10 min at room temperature using 4% formaldehyde/10% formalin in 1× phosphate buffered saline solution (PBS) and followed the fixation by two rinses in 1× PBS, after which we permeabilized the cells with 70% EtOH and stored them at 4 °C overnight.

The *HD-FISH* probe sets used here were generated by real-time PCR reactions from genomic DNA with primers designed as previously described by Bienko et al. 2013,

delimiting amplicons of length 200–220 nucleotides. The resulting total of 50 to 80 unique fragments per region were subsequently labeled with the fluorophore of interest with the commercial kit for *Universal Linkage System*.

To perform DNA-FISH, 20 ng of *HD-FISH* probe were ethanol precipitated using 20 µg of glycogen as carrier and dissolved in 20 µl of 1.7× SSC, 70% formamide, 50 mM Na₂HPO₄/NaH₂PO₄, 10% dextran sulfate and 5× Denhardt's solution, pH 7.5. The three-dimensional *HD-FISH* procedure on nuclei was adapted from Solovei et al. 2010 preserving the three-dimensional nuclear structure. After hybridization, cover glasses were washed twice at 30 °C for 30 min in wash buffer, additionally 20 ng/ml DAPI were added to the second wash. After washes, the cover glasses were rinsed with 2× SSC.

For microscopy, samples were covered with a mounting solution containing 2× SSC buffer, 10 mM Tris, 0.4% glucose, 100 µg/ml catalase, 37 µg/ml glucose oxidase and 2 mM Trolox.

Image acquisition

All images were acquired at 100× magnification (oil immersion, high numerical aperture Nikon objective) on an inverted epi-fluorescence microscope (Nikon) equipped with a high-resolution charge-coupled device (CCD) camera (Paxis, Princeton Instruments) and controlled by MetaMorph software.

We sequentially acquired three-dimensional stacks of fluorescence images in four different fluorescence channels using filter sets for DAPI, TMR, Cy5, and Alexa 594. Our exposure times were roughly 2–3 s for most of the dyes except for DAPI (which we exposed for ~100 ms). The pixel size was 0.125µm, and the spacing between consecutive planes in our stacks was 0.25 µm.

Image analysis

Custom MATLAB software was implemented for the serial semi-automated identification of *HD-FISH* fluorescent spots (see Results and discussion), the general pipeline included the below described steps.

Nuclei segmentation

Using DAPI staining fluorescence signal, two criteria were used for segmentation. First the nuclei were segmented using Otsu's method for finding the optimal threshold which minimizes the intra-class variance for intensity pixel values (under the assumption of a bimodal distribution), resulting in a binary image where most of the nuclei were correctly segmented (see **Figure 10**). Following this step, a second criterium was taken into account for regions with an area above the average nuclei area (undersegmentation). For these regions, a subsequent segmentation using the watershed method was applied, this allowed us to recover the overlapping nuclei. Furthermore, nuclei adjacent to the edges of the *field-of-view* were discarded. Each of the resulting regions was dilated, suitably preventing segmentation beneath the nuclei edges. The resulting binary mask served to extract different nuclear features: area, mean intensity, perimeter, center of mass, etc. as well as for assigning the spots to cells.

Spot identification

For each fluorescence-channel image stack, non-uniform background was removed, and signal-to-noise ratio was enhanced by applying 3D Laplacian of Gaussian convolution. The optimal size of this filter (corresponding to the width of the Gaussian) depends on the size of the observed particle and was empirically adjusted to maximize

the signal of the particles.

Due to residual uniform noise, further use of a threshold is needed. First the number of spots is calculated as a function of monotonically increasing thresholds, then a single threshold was manually chosen from a normally observed plateau corresponding to a range of thresholds over which the total number of detected spots in the field-of-view does not vary. Thresholds chosen in this plateau yielded spot detections that matched the spots identified by eye (see **Figure 10**). Finally, the volume, mean intensity and *coarse center position* (as the center of mass of the intensities in the identified connected component) was determined for individual spots.

Results and discussion

Precise spot localization

Once the *coarse center position* of the spots was determined, we proceeded with the estimation of the precise position of the spots. The 2D image intensity (x, y) profile can be well represented by a 2D Gaussian function (**eq. (1)**).

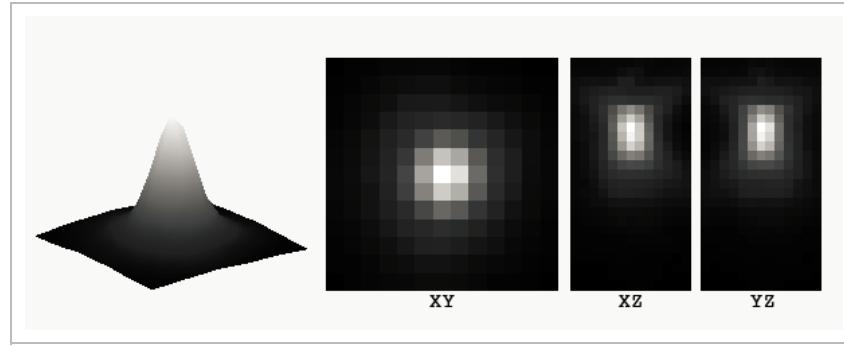


Figure 3. Distribution of image intensities can be approached with Gaussian functions. In the figure, a smoothed surface of fluorescence intensities in 2D, and 2D images of maximum projection across the depth axis for different planes

Based on the eq. (1), the 2D intensity model parameters are obtained to fit the xy -plane maximum-projection image intensity distribution in a window of a 6 pixels radius from the calculated coarse position (see **Figure 4-A,C**); A is the maximum intensity, x and y are the center of the 2D Gaussian and σ_* is the width of the spots, additionally we can also estimate the intensity background b for which we later get an average and correct for the whole slide.

$$f(x, y) = A \exp\left(-\left(\frac{(x - x_o)^2}{2\sigma_x^2} + \frac{(y - y_o)^2}{2\sigma_y^2}\right)\right) + b \quad (1)$$

After getting the precise (x, y) position, the intensity profile of this position along the z-axis was used to fit the following 1D Gaussian function (**see figure 4B**):

$$f(z) = A \exp\left(-\frac{(z - z_o)^2}{2\sigma_z^2}\right) \quad (2)$$

Both models are fitted using maximum likelihood estimates. The starting values for the model parameters are determined as follows: A as the maximum intensity value in the spot window, x, y, z from the coarse center positions, σ fixed to 1.5. We assume that the spot is spherical in 3D, thus, σ is equally initialized in all dimensions.

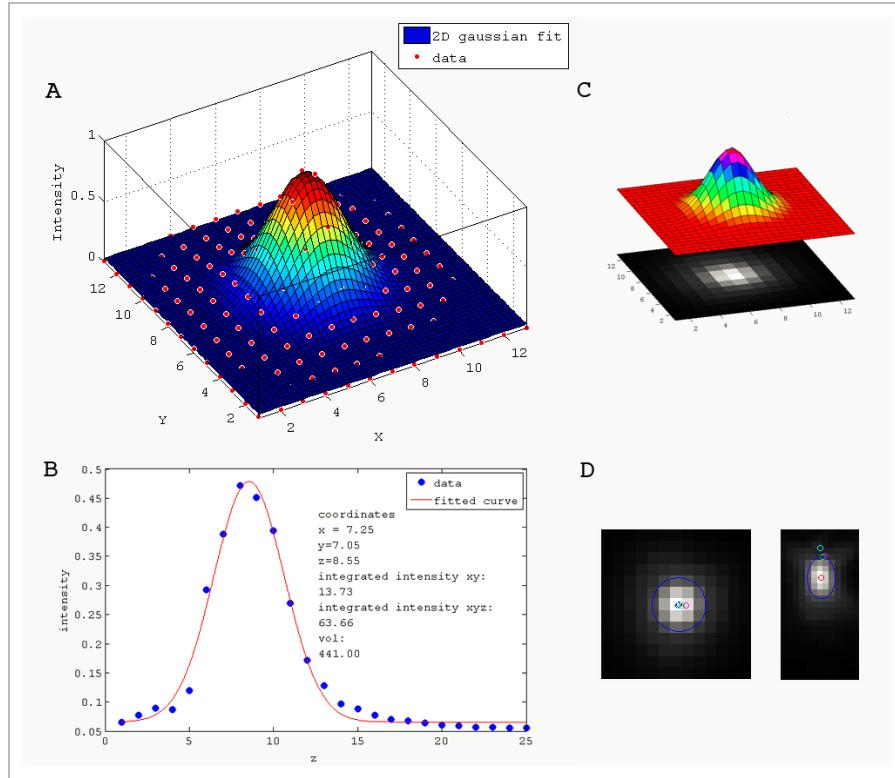


Figure 4. Precise localization is estimated from parametric fitting of 2D and 1D Gaussian functions for individual spots. **A.** Surface fit from the maximum projection over the xy-plane z-depth intensity profile. **B.** Gaussian fit from the maximum projection over the xz-plane y-depth intensity profile. **C.** Overlay of the resulting Gaussian surface over the image. **D.** Precise spot localization is shown in red while the blue circle shows the estimated spot size with a diameter corresponding to the *FWHM* (full width half maximum) of the Gaussian function; additionally, position of the same bead in other channels is shown.

As a result, we got new coordinates for the precise spot localization, background corrected intensity distribution, and estimated spot size from the *FWHM* (full width half maximum). Additionally we used the *goodness of fit* to infer for the appearance of multiple spots within one $13\text{px} \times 13\text{px}$ window; those occurrences were discarded for subsequent analysis; however, further improvement can be achieved by fitting a mixture of multi-variate Gaussians to estimate the localization of multiple spots within a spot window.

Despite that these measurements enhance the accuracy for the target position estimation, due to time constraints, an estimation of the error and extent of precision was not performed. From previous experience we know that the resolution is within 20 and 30 nm (Semrau et al. 2011), however this might vary across microscope setups. The experiment for the precision assessment would be as follows: Iterative imaging and estimation of the precise spot localization should yield a distribution of x, y, z localization coordinates for which the variance would serve as a measure for resolution.

Chromatic aberration corrections

The wavelength-dependency of the refraction indexes in optical systems involves *chromatic aberrations*: one object point is not projected on exactly one image point on the sensor plane, but dispersed depending on its wavelength. When precise estimation of the three-dimensional position of a point is necessary —as in HD-FISH experiments — the distortions have to be compensated.

To this end, commercial fluorescent beads where imaged and detected in different channels. For each channel, the acquired image stack of 15 planes in depth was filtered and thresholded to automatically detect the *coarse position* of the single beads in a 3D space ($1024\text{px} \times 1024\text{px} \times 15$ planes). The threshold was set automatically for a reference channel as the intensity value for which the inverse of the coefficient of variation was the highest. The threshold for the remaining channels was selected such that the detected beads was close in number, if not the same, to the number of beads detected on the reference channel. For each channel, the detected coarse positions (x, y, z) were enhanced for precision by fitting Gaussian functions as previously described in precise spot localization section.

Registration and shift measurement

Since the corresponding intensity signals of a bead in different channels do not co-localize, first we had to find the signals corresponding to the same single bead. This might be trivial for low density of beads in the *field-of-view*, however it proved to be difficult on high density of beads due to mis-assignment and to other factors as debris or bead movement between channel acquisition rounds. We approached this by assigning correspondences between the selected reference bead and its closest neighboring intensity signal regardless of the channel, under the assumption that the distance between channel signals is minimal, thus, the closest signals should correspond to the same bead. Then we used this information to build an undirected adjacency matrix where an edge was assigned every time we found a neighboring connection between two bead signals. This criteria allowed us to build independent and completely connected graphs called *cliques*, which correspond to single beads (see **Figure 5. Top**). High density of beads allowed to better capture the shift distribution within the *field-of-view*.

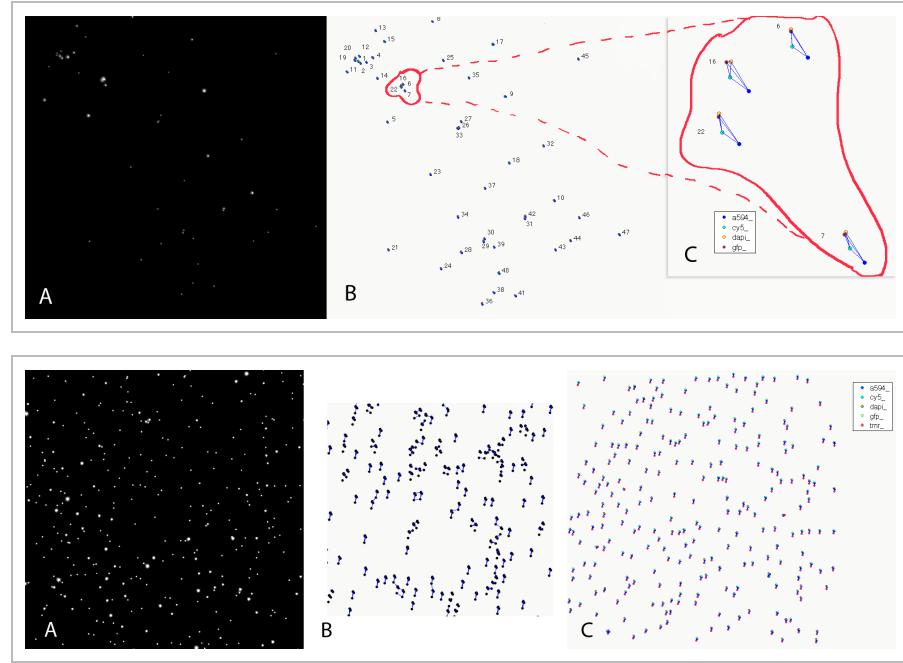


Figure 5. Beads registration. **Top:** Commercial beads are imaged in different channels, and processed for their exact locations, then assigned to a single bead (*registration*). **A.** Fluorescence intensity image of one channel for low density of beads within a *field-of-view*. **B.** Labeled beads with it corresponding identified signals in different channels. **C.** Each registered bead is the result of finding complete connected components (*cliques*) of degree four (corresponding to the number of imaged channels) in a graph. **Bottom:** **A.** Fluorescence intensity image of one channel for high density of beads. **B.** (detail) From a reference channel, neighboring signals are assigned as potential corresponding fluorescent signals for the same bead (blue edges connecting fluorescent signals, in black). Connections are

recorded in an adjacency matrix. **C.** Bead registration is solved by searching *cliques* in the adjacency matrix.

With this method we achieved the correct identification of most of the beads; outliers for the distance between signals associated with a single bead were discarded for further analysis. Once the beads where registered, the shift Δ (extent of *chromatic aberration*) from a reference channel ref , was calculated in each direction (x, y or z) for each bead i in each channel ch as follows:

$$\begin{aligned}\Delta_{xi}^{ch} &= x_i^{ch} - x_i^{ref} \\ \Delta_{yi}^{ch} &= y_i^{ch} - y_i^{ref} \\ \Delta_{zi}^{ch} &= z_i^{ch} - z_i^{ref}\end{aligned}$$

Then for each channel we formulated the shift Δ as a function of the three-dimesional localization:

$$\begin{aligned}\Delta_x^{ch} &= f(x, y, z) \\ \Delta_y^{ch} &= f(x, y, z) \\ \Delta_z^{ch} &= f(x, y, z)\end{aligned}$$

As expected, we observed that the extent of distortions was wavelength dependent, and the shift magnitudes vary for distinct fluorescence channel (see figure 6)

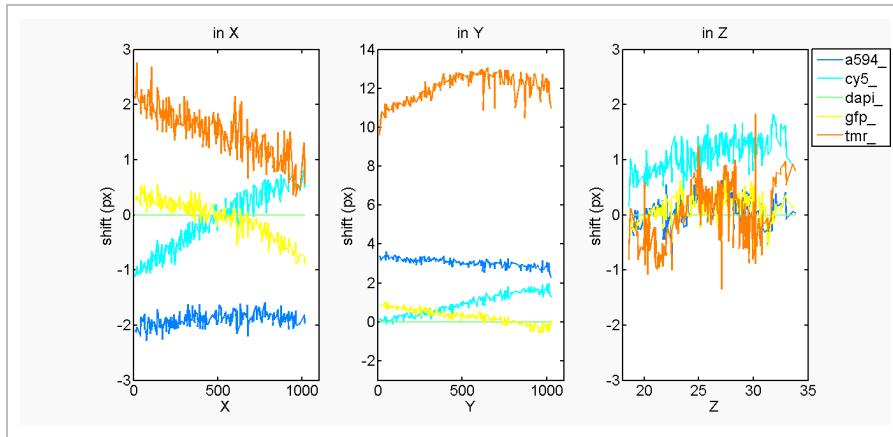


Figure 6. Once the fluorescence intensity signals are registered, (before correction) the shift Δ is measured in each direction x, y, z . For visualization proposes, the shift — y-axis [px] — for each channel (A594, Cy5, GFP and TMR) from a reference channel (DAPI) is plotted as a function of a single dimension in one of each three directions (x, y and z) — x-axis [px] —. The magnitude of the shift varies from channel to channel at the different dimensions, showing the wavelength dependency of the *chromatic aberration*.

Shift correction model fitting

To interpolate the shift for the whole volume we selected polynomial functions due to their smoothness, thereby reducing the irregularities due to thermal noise which could still be captured by using other interpolating approaches. Along this line, fifth-order polynomials were fit to the measured shifts (separately for each dimension, see eq. 3). The order of the polynomial was empirically selected such that the residual shift was the least.

$$\Delta_{dim}^{ch}(x, y, z) = a_{n,n,n}x^n y^n z^n + a_{n,n,n-1}x^n y^n z^{n-1} + a_{n,n-1,n}x^n y^{n-1} z^n + \dots + a_{1,0,0}x + a_{0,1,0}y +$$

The resulting polynomial functions were at some extent similar for different channels. For both \mathbf{x} and \mathbf{y} directions, the shift magnitudes were mostly in a lateral gradient, being positive on one extreme, and negative on the other (see **Figure 7**). Meanwhile, the interpolating function corresponding to the \mathbf{z} direction was observed to be focal, with a greater displacement at the outer regions of the *field-of-view*.

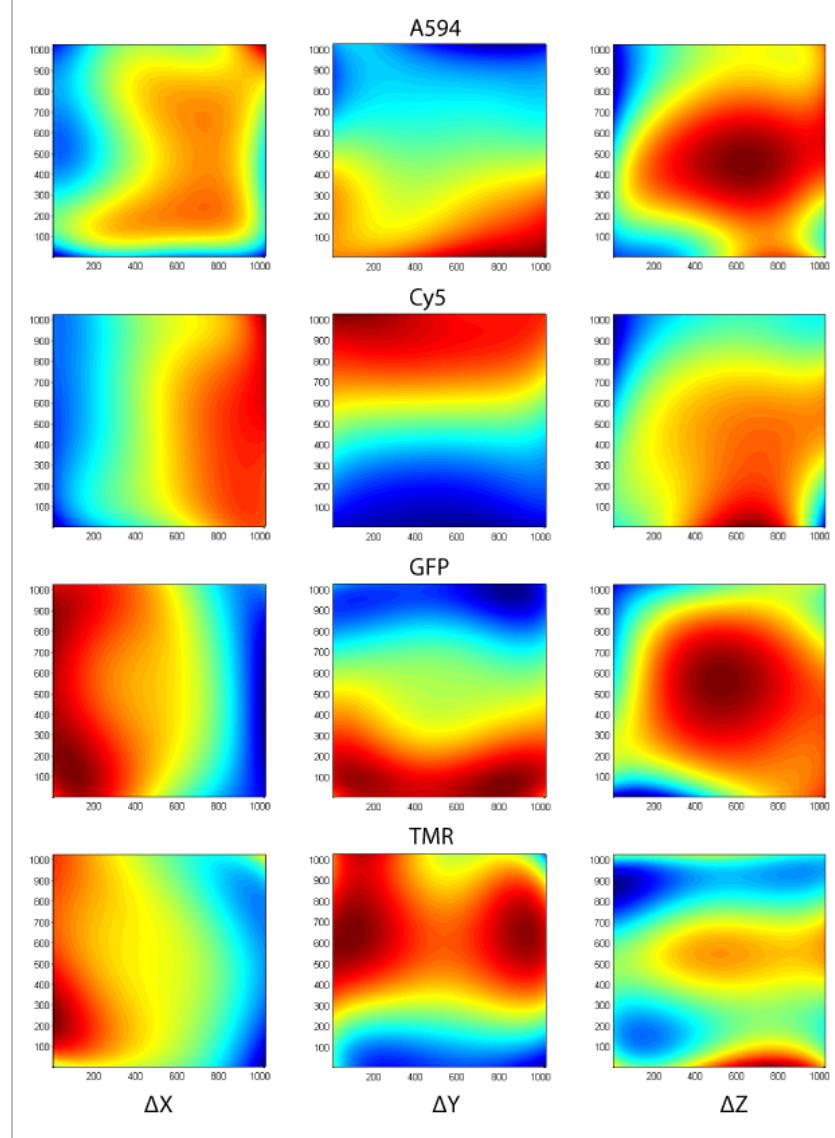


Figure 7. Color plots showing the extent of the shift Δ interpolated using polynomial functions. For a single plane at $z = 5$, the shift for different positions at $(\mathbf{x}, \mathbf{y}, z = 5)$ is plotted for different channels (A594, Cy5, GFP and TMR) at the different directions (\mathbf{x} , \mathbf{y} and \mathbf{z}).

By using fluorescent beads, in this first approach for interpolating the shift, we failed to capture the contribution of the distortion along the \mathbf{z} dimension since the totality of the beads were localized at the same z -depth (data not shown). Alternatively, and more optimally, we performed a second experiment where we used *primary human foreskin fibroblasts* to capture the range in z -depth at which the genomic regions could be localized. To this end, interleaved probes distinctively label were designed to target a single genomic region. To achieve a high density of localizations, after hybridization, the coverslip was imaged by moving the *field-of-view*, thus a single region was imaged at different positions throughout one single focal area and afterwards, shift

magnitudes were calculated as described on the previous section. The resulting interpolating functions for all channels in the different directions are shown above on **Figure 7..**

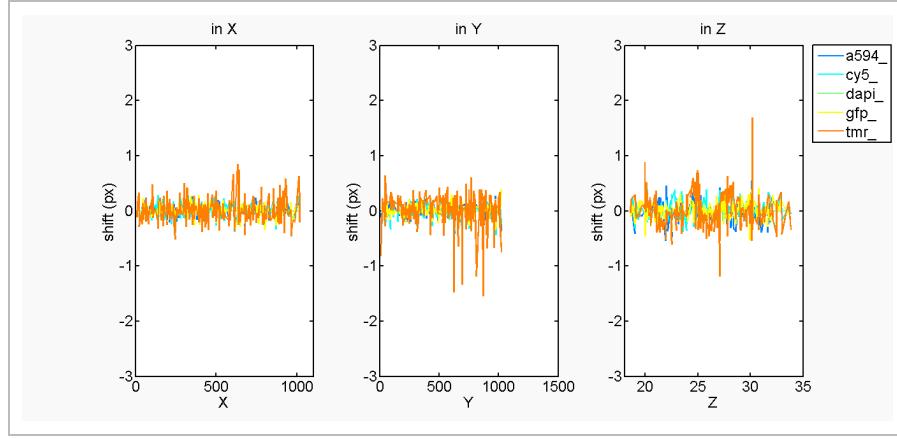


Figure 8. Shift measurements after correction with polynomial functions. The magnitude of the shift is corrected to be less than 1px. Some variation is still observed due to slight variations in movement between acquisition rounds, additionally to the fact that it is a projection and the three-dimensional shift is captured in a single dimension.

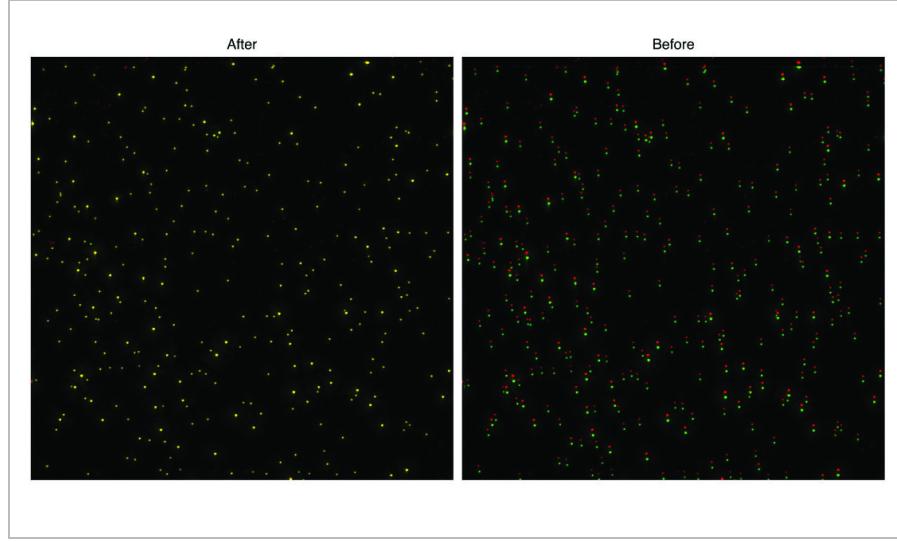


Figure 9. Overlay of fluorescent beads imaged in different channels, before and after correction.

Our approach resulted in a set of polynomial functions which served to correct the shift for each fluorescent channel, in every direction as a function of its three-dimensional position_ (see **Figure 6..**). A reduction in the shift up to less than 1px was achieved. We observed a better result when correcting for shifts in the planar directions than in the depth direction (see **Figure 8..**).

One drawback of the correction for *chromatic aberrations* is that they depend on the experimental setup, where small changes in variables, as are the imaging medium and slight movements of the objective are prone to change the shift pattern, accordingly, the measurement of the shift magnitudes have to be performed on an individual experiment basis, as a consequence, increasing the time and labor required for each image acquisition step.

SpotterUI

The previously described approaches to the precise localization of genomic regions were integrated into a custom Graphical User Interphase (*SpotterUI*) written in MATLAB (see **Figure 10.**). This semi-automated software includes various features that simplify the identification and estimation of the precise spot localizations from *HD-FISH* experiments. Automatic and manual segmentation, annotations for single nuclei, and overlayed figure creation are some of the general tools featured in the program. Moreover, quantitative measures like nuclei section area, DAPI average intensity measures, background correction and also measurements for individual spots are estimated. Among the few settings that have to be tuned and specified in a configuration file are the width and size of the filters, number of thresholds for finding the threshold plateau, average nuclei sectional area, voxel size, etc.

Within the same user interface, the shift-interpolating functions can be generated from images with fiducial probes like fluorescent beads or multi-labeled genomic regions. These polynomial functions are stored in a file which can be later used to batch correct actual experiment spot locations.

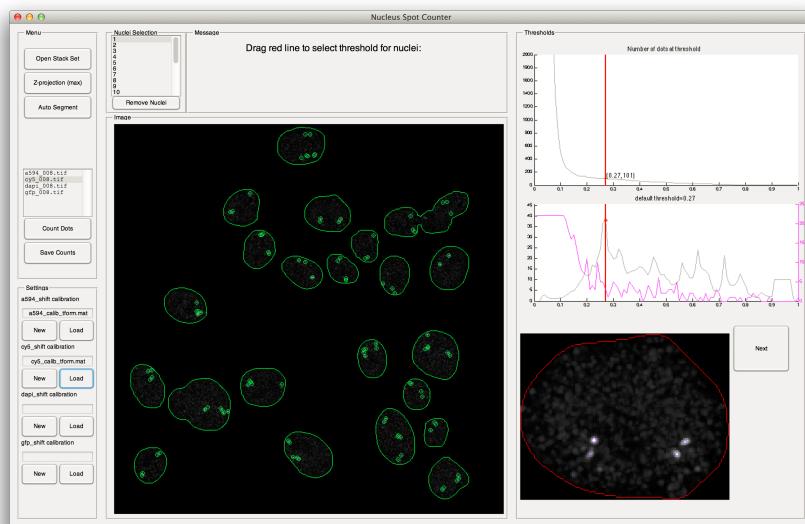
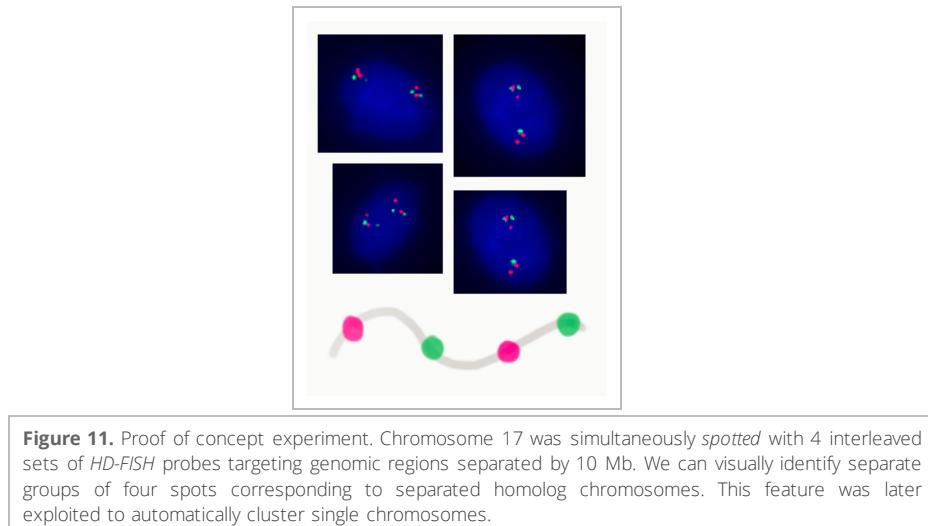


Figure 10. *SpotterUI* is a Graphical User Interface that integrates the workflow for the analysis of *HD-FISH* experiments into a software in a semi-automated manner. Configuration settings are given in a configuration file, while a simple user interface allows for the automatic segmentation and the interactive identification of the spots. Interaction with the *regions-of-interest* allows for selecting and annotating or deleting single cells, the interaction with single spot also allows to correct for misidentification. Polynomial functions can be re-used for batch correction of several *fields-of-view* as they are included from MATLAB native data files as an input setting. The workflow returns at the end several quantitative measures including the precise position of the *HD-FISH* spots for further statistical analysis.

4 loci experiment

In a proof of concept experiment, *HD-FISH* probes were designed to evenly target 4 regions within the chromosome 17 of *HME* cells. *Spotted* regions were visualized by genetically-equidistant set of fluorescently-labeled probes. Both Alexa594 and Cy5 labeled probes were interleaved with a distance of about 10Mb between each other (see **Figure 11.**).



After hybridization, we acquired 21 fields-of-view comprising a total of 324 nuclei in varying cell-cycle stages (see **Figure 12.**). With the help of the SpotterUI we segmented, identified, corrected, and processed the set of acquired image stacks.

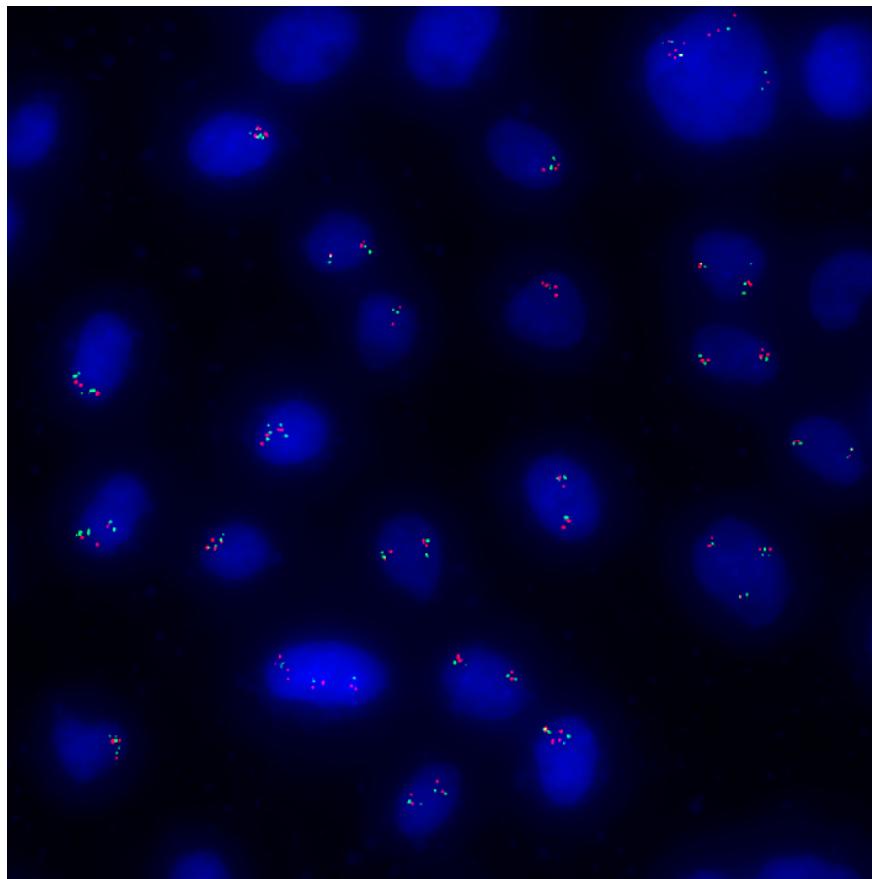
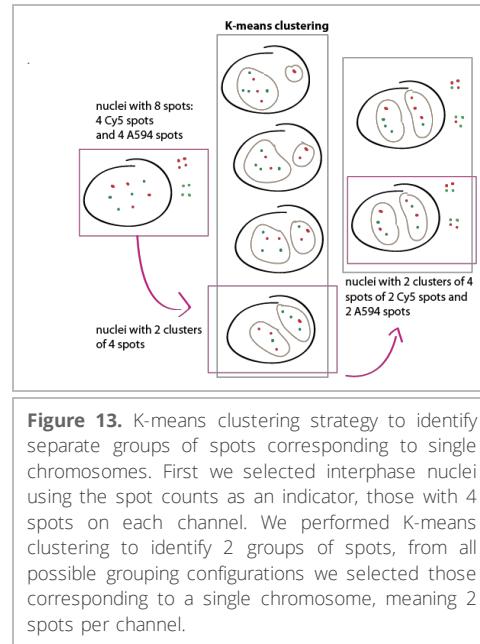


Figure 12. Sample overlay of the 4 loci experiment. Image of DAPI in the nuclei (blue) of HME cells and the identified spots in both Cy5 (green) and Alexa594 (magenta).

In most of the nuclei, we can visually identify separate groups of spots from which we assumed they belong to single chromosomes (see **Figure 12.**). To exploit this feature

of organization, we first filtered for nuclei in which the total number of spots was 8, corresponding to only 2 copies of the chromosome 17 during interphase. Furthermore we used K-means ($k=2$) clustering algorithm to group the single copies of chromosome 17. Among all the possible spot configurations within a cluster, we selected for those clusters with 4 spots (2 Cy5 spots and 2 Alexa594 spots, each (see **Figure 13.**). Roughly 33% of the nuclei showed to be in interphase, as we found 107 nuclei with interphase spot counts, and 105 satisfying the further required configuration, thus, yielding 105 spot clusters. On these clusters we proceeded to measure the three different euclidean distances between them: i) distance between Alexa594 spots, ii) distance between Cy5 spots and iii) distance between Alexa594 and Cy5 spots (see **Figure 14.**).



For a single spot cluster there is a single distance between two probes of the same channel, which corresponds to a distance of 20 Mb. However, for the 4 possible distances separating spots of different channels, we cannot directly identify the physical correspondence to the genomic separation, i.e. distance between a genomic separation of 10 Mb and 30 Mb. We approached the estimation of each distance by fitting a mixture of two Gaussians, where each distribution corresponds to one distance at each genomic separation (see **Figure 14. Bottom right**) under the assumption of normally distributed distances.

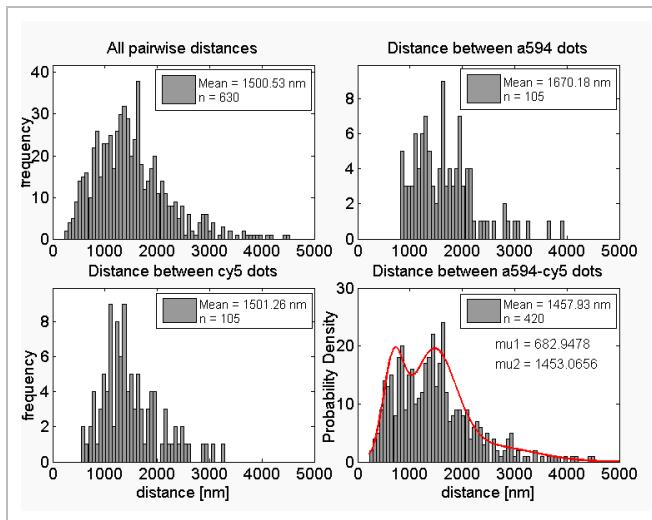


Figure 14. Histograms showing the distribution of calculated physical distances for varying genomic distances. **Top left:** All measured pairwise distances between spots. **Top right/Bottom left:** Distances corresponding to 20 Mb (spots from the same channel). **Bottom right:** Distribution of physical pairwise distances measured between spots from different channels corresponding to a genomic separation of 10 and 30 Mb approached by mixture of Gaussians.

With this proof of concept experiment we were able to estimate the physical distance for three equidistant genomic regions. Likewise, our results meet the previous observations where the leveling-off of the linear to spatial scaling can already be observed for genomic distances of 10Mb (Mateos-Langerak et al. 2009, see **Figure 15.**). As a result, we demonstrate the scope of our methodology to capture the chromosome topology.

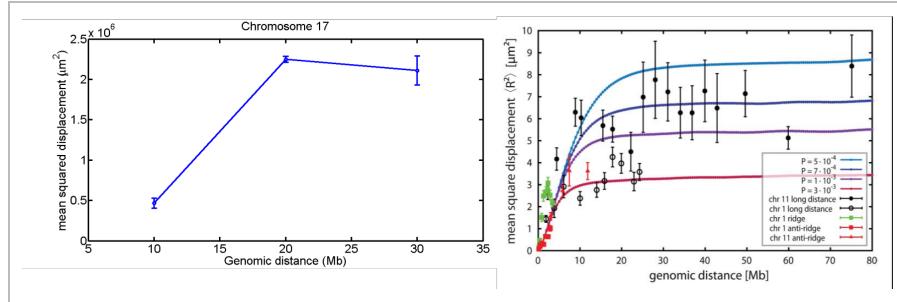


Figure 15. Leveling-off of the linear to spatial scaling in the organization of the genome. In agreement with previous experimental measurements, our results also show the leveling-off of the physical distance for genomic regions separated by more than 10Mb. This observations support the view of interphase chromatin organizing in globular structures termed *chromosome territories*.

For the sake of simplicity and to demonstrate the applications of our methodology, we limited our probe set to simply target 4 regions interleaved between two fluorophores. However in follow up experiments we are planning to use combinations of fluorophores to target and identify individual regions without having to struggle with the fitting mixtures of Gaussians. Moreover we can also decrease the separation between regions up to the Kb resolution, thus having a better quantitative measures for the scaling behavior of the mammalian nucleus.

Concluding remarks

Model ensembles of polymer conformations together with DNA-FISH experimental measures, provide a natural framework for understanding the nature of the emerging link between structure, composition, and function; they also serve as a starting point for understanding cell-to-cell variability in gene expression (Ozbudak et al. 2002, Elowitz et al. 2002), (i.e. enhancer-dependent gene activation would be promoted from transient spatial proximity of two genetically distant loci).

To this end, *HD-FISH* provides the required scope by capturing variations within chromatin conformations by the measurements in single cells. Our developed methodology and tools allow for the rapid and precise quantification of FISH experiments targeted to the understanding of genome organization. Furthermore, *HD-FISH* can be coupled with *single molecule RNA-FISH* to gain more insight into the link between structure and function.

Acknowledgements

This project was closely and jointly developed with Magda Bienko — without her enthusiasm, dedication and experimental dexterity, none of the experiments above could have been realized —. We would like to further acknowledge Stefan Semrau for fruitful discussion for the correction of the *chromatic aberrations*. Moreover we would like to thank Alexander van Oudenaarden for his support and for sharing space, time and expertise in addition to guidance. Finally we thank the van Odenaarden Lab members which suggestions, small talk, and daily interaction made a great deal of addition to the development of this project.

References and relevant literature

1. Andrulis E D, Neiman a M, Zappulla D C, Sternglanz R. (1998) Perinuclear localization of chromatin facilitates transcriptional silencing. *Nature* 394: 592-5.
2. Barbieri Mariano, Chotalia Mita, Fraser James, Lavitas Liron-Mark, Dostie Josee, Pombo Ana, Nicodemi Mario. (2012) Complexity of chromatin folding is captured by the strings and binders switch model. *Proceedings of the National Academy of Sciences of the United States of America* 109: 16173-8.
3. Bickmore Wendy a, van Steensel Bas. (2013) Genome architecture domain organization of interphase chromosomes. *Cell* 152: 1270-84.
4. Bienko Magda, Crosetto Nicola, Teytelman Leonid, Klemm Sandy, Itzkovitz Shalev, van Oudenaarden Alexander. (2012) A versatile genomescale PCRbased pipeline for highdefinition DNA FISH. *Nature methods* : 1--5.
5. Casolari Jason M, Brown Christopher R, Drubin David A, Rando Oliver J, Silver Pamela A. (2005) Developmentally induced changes in transcriptional program alter spatial organization across chromosomes. *Genes and Development* : 1188--1198.
6. Chubb Jonathan R, Boyle Shelagh, Perry Paul, Bickmore Wendy A. (2002) Chromatin Motion Is Constrained by Association with Nuclear Compartments in Human Cells. *Current Biology* 12: 439 - 445.
7. Cremer Thomas, Cremer Marion. (2010) Chromosome territories. *Cold Spring Harbor perspectives in biology* 2: a003889.
8. Croft J a, Bridger J M, Boyle S, Perry P, Teague P, Bickmore W a. (1999) Differences in the localization and morphology of chromosomes in the

- human nucleus. *The Journal of cell biology* 145: 1119--31.
9. Dekker Job, Rippe Karsten, Dekker Martijn, Kleckner Nancy. (2002) Capturing chromosome conformation. *Science (New York, N.Y.)* 295: 1306--11.
 10. Elowitz Michael B, Levine Arnold J, Siggia Eric D, Swain Peter S. (2002) Stochastic gene expression in a single cell. *Science* 297: 1183--6.
 11. Heermann Dieter W, Jerabek Hansjoerg, Liu Lei, Li Yixue. (2012) A model for the 3D chromatin architecture of pro and eukaryotes. *Methods (San Diego, Calif.)* : .
 12. Kim S H, McQueen P G, Lichtman M K, Shevach E M, Parada L a, Misteli T. (2004) Spatial genome organization during Tcell differentiation. *Cytogenetic and genome research* 105: 292--301.
 13. Lieberman-Aiden Erez, van Berkum Nynke L, Williams Louise, Imakaev Maxim, Ragoczy Tobias, Telling Agnes, Amit Ido, Lajoie Bryan R, Sabo Peter J, Dorschner Michael O, Sandstrom Richard, Bernstein Bradley, Bender M a, Groudine Mark, Gnirke Andreas, Stamatoyannopoulos John, Mirny Leonid a, Lander Eric S, Dekker Job. (2009) Comprehensive mapping of longrange interactions reveals folding principles of the human genome. *Science (New York, N.Y.)* 326: 289--93.
 14. Marshall W.F, Straight A, Marko J.F, Swedlow J, Dernburg A, Belmont A, Murray A.W, Agard D.A, Sedat J.W. (1997) Interphase chromosomes undergo constrained diffusional motion in living cells. *Current Biology* 7: 930 - 939.
 15. Mateos-Langerak Julio, Bohn Manfred, de Leeuw Wim, Giromus Osdilly, Manders Erik M M, Verschure Pernette J, Indemans Mireille H G, Gierman Hinco J, Heermann Dieter W, van Driel Roel, Goetze Sandra. (2009) Spatially confined folding of chromatin in the interphase nucleus. *Proceedings of the National Academy of Sciences of the United States of America* 106: 3812--7.
 16. Meister Peter, Towbin Benjamin D, Pike Brietta L, Ponti Aaron, Gasser Susan M. (2010) The spatial dynamics of tissuespecific promoters during C elegans development. *Genes and Development* : 766--782.
 17. Osborne Cameron S, Chakalova Lyubomira, Mitchell Jennifer a, Horton Alice, Wood Andrew L, Bolland Daniel J, Corcoran Anne E, Fraser Peter. (2007) Myc dynamically and preferentially relocates to a transcription factory occupied by Igf. *PLoS biology* 5: e192.
 18. Ozbudak Ertugrul M, Thattai Mukund, Kurtser Iren, Grossman Alan D, van Oudenaarden Alexander. (2002) Regulation of noise in the expression of a single gene. *Nature genetics* 31: 69--73.
 19. Raj Arjun, Bogaard Patrick Van Den, Rifkin Scott A, Oudenaarden Alexander Van, Tyagi Sanjay. (2008) Imaging individual mRNA molecules using multiple singly labeled probes. *Nature Methods* 5: 877--879.
 20. Schoenfelder Stefan, Sexton Tom, Chakalova Lyubomira, Cope Nathan F, Horton Alice, Andrews Simon, Kurukuti Sreenivasulu, Mitchell Jennifer a, Umlauf David, Dimitrova Daniela S, Eskwi Christopher H, Luo Yanquan, Wei Chia-Lin, Ruan Yijun, Bieker James J, Fraser Peter. (2010) Preferential associations between coregulated genes reveal a transcriptional interactome in erythroid cells. *Nature genetics* 42: 53--61.
 21. Semrau Stefan, Pezzarossa Anna, Schmidt Thomas. (2011) Microsecond singlemolecule tracking musSMT. *Biophysical journal* 100: L19--21.
 22. Solovei Irina, Cremer Marion. (2010) 3DFISH on Cultured Cells Combined with Immunostaining. *Methods in Molecular Biology: Fluorescence in situ Hybridization (FISH)* 659: 117-126.
 23. Spilianakis Charalampos G, Lalioti Maria D, Town Terrence, Lee Gap Ryol, Flavell Richard a. (2005) Interchromosomal associations between alternatively

- expressed loci. *Nature* 435: 637--45.
24. Tark-Dame Mariliis, van Driel Roel, Heermann Dieter W. (2011) Chromatin folding from biology to polymer models and back. *Journal of cell science* 124: 839--45.
 25. van Driel Roel, Fransz Paul F, Verschure Pernette J. (2003) The eukaryotic genome a system regulated at different hierarchical levels. *Journal of cell science* 116: 4067--75.