Supporting Information

Marie et al., "Integrated view of genome structure and sequence of a single DNA molecule in a nanofluidic device"

SI Appendix

Terminology

Local denaturation of stained DNA makes a pattern appear on the DNA, a *denaturation pattern*. If denaturation was induced by heating, this denaturation pattern can also be referred to as a *melting pattern*. Since this pattern is sequence-specific and highly reproducible, it is coarse-grained information about the sequence of the DNA that was denatured partially. We improve the signal-to-noise ratio of this information by averaging over the frames of a movie of a patterned DNA molecule. We refer to the average as a *denaturation map* or, if heat-induced, a *melting map*. In the present work, this kind of pattern is fixed—as in chemical fixing of a developed photographic print—by renaturing the DNA. We refer to such a fixed pattern/map as a *denaturation-renaturation pattern/map*, in short *DR pattern* and *DR map*.

A shorter piece of such a DR map is called a "barcode" [1] because this is what it resembles. The patterned piece of DNA in the field-of-view of a microscope with sufficient magnification to resolve the DR pattern is a *raw barcode* [1]. Its time-average calculated from the raw barcodes in each frame of a movie is an *averaged barcode*. DR maps are assembled from overlapping averaged barcodes. A DNA molecules that have been DR mapped, is aligned within the genome by aligning shorter pieces (*barcodes*) of its DR map—a staggered sequence of them in which consecutive barcodes overlap each other. This is why we call it the *Sliding Window* method.

This alignment is done by using theory on the known sequence of the genome to compute an *in silico DR map* within which experimental barcodes are aligned. This is done by comparing them with segments of the *in silico* DR map of the same length. Such a segment also looks like a barcode, so it is referred to as a *theoretical barcode*.

Experimental methods

Chip design. Previous devices for stretching single DNA molecules rely on three kind of designs: nanochannels that provide two dimensions of confinement [2, 1, 3, 4], nanoslits that provide only one degree of confinement [5], and flow-through nanochannels in which DNA is stretched in an elongation flow [6]. Nanochannels and nanoslits provide a stretching of the order of 50% in common buffer conditions. At this stretching, density fluctuations due to the free oscillation of the DNA are observed along the DNA strand and can thus challenge the mapping of fluorescence features along the DNA. Recent work has shown that by carefully reducing the ionic strength, higher stretching can be achieved [7]. This requires the staining/labeling of the DNA to be performed off-chip before the buffer is exchanged to lower the ionic strength. Flow-through stretching elongate molecules while they pass through a nanochannel. This could limit the perspectives of mapping genomic features at high base pair resolution, as the time available for imaging is limited by the velocity of the molecules. Our nanofluidic device adds two stretching mechanisms: confinement in a nanoslit and hydrodynamic drag. The device is basically a cross-shaped nanoslit connecting four microchannels [Fig. S1A].

The microchannels terminate in millimeter-size loading pads, which are accessed from the backside of the device through sandblasted holes. At constant pressure in two opposite microchannels, the device creates a constant hydrodynamic drag on DNA that has been placed in the nanoslit as shown in Fig. S1B. This drag force stretches the DNA beyond its initial elongation caused by its confinement in the 85 nm-high slit (Fig. S1C). When the DNA is placed in the chip, it is balanced at a stagnation point, i.e., in an unstable equilibrium, the dynamics of which is so slow, however, that it is negligible.

Lab-on-a-chip device fabrication. Silica devices were fabricated by depositing a 2 μm thick poly-silicon layer, used as etching mask to define microchannels. The pattern of microchannels and inlet structures was defined by reactive ion etching, first in the poly-silicon layer used as a mask, then in the silica at a final depth of 5 μm. Then the poly-silicon layer was etched away and the cross-shaped slit connecting the four microchannels was etched to a depth of 85 nm in the silica substrate. Alternatively, silicon devices with similar dimensions were fabricated using a process described in [8]. Inlet holes were made by powder blasting from the backside of the device. Individual chips were diced out of the four-inch wafer. Finally, individual devices were sealed by fusion bonding to a 170 μm thick silica lid (Valley Design, USA).

Degree of stretching and tension in the DNA. A DNA molecule that spans the two nanoslit branches with outflow and reaches into the two microchannels, as shown in Fig. S1B–C, experiences three contributions to its stretching in the nanoslit:

1. the hydrodynamic drag on it in the nanoslit,
2. an entropic recoil force at each nanoslit-microchannel transition, and
3. the fluctuating pulling forces from the hydrodynamic drag forces on the two parts of the molecule in the microchannels.

In the absence of flow in the microchannels, the free ends of the molecule there can coil thermally to a degree not possible in the nanoslit. This causes a so-called *entropic recoil* [9], which pulls at both ends of the DNA confined in the nanoslit. As obvious from the fact that the flow in the microchannel stretches the DNA there, except towards its ends, at our experimental conditions the entropic recoil is negligible compared to the pull from the hydrodynamic drag force in the microchannel.

In the nanoslit, the tension in the DNA grows in the upstream direction. It grows as the integral over the hydrodynamic drag force on the downstream part of the DNA. So the DNA is stretched the most where the four branches of the nanoslit meet. There, and most of the way downstream as well, the observed fluctuations of the DNA are caused solely by thermal noise acting on the same DNA. The fluctuations in the pull from the microchannels do not contribute to the DNA's motion well inside the nanoslit; there, only the average pull from the microchannels is felt.

These considerations can be summarized as follows: The tension at a point s along the nanoslit can be expressed as

$$\tau(s) = \left[\tau_{\mathrm{micro}} + \tau_{\mathrm{nano}} \times \left(1 - \frac{|s|}{L_0/2}\right)\right] \quad . \tag{1}$$

for $-L_0/2 < s < L_0/2$ with L_0 being the length of the slit. That is, the tension is maximal at the center (with the value $\tau_{\mathrm{micro}} + \tau_{\mathrm{nano}}$), and decays linearly towards the ends (with values τ_{micro}).

The tension can be converted to a degree of stretching using the Marko-Siggia interpolation formula for polymers, which relates the tension τ to the stretching q $(0 < q < 1)$ [10]

$$\tau = \frac{k_{\mathrm{B}}T}{L_p}\left[q + \frac{1}{4(1-q)^2} - \frac{1}{4}\right] \quad . \tag{2}$$

Here L_p is the persistence length of the DNA. Throughout the analysis we assume a persistence length for the stained DNA molecule of $L_p = 60\,\mathrm{nm}$ [11]. Combing Eqs. (1,2) provides an expression for the stretching in the slit versus position, which is needed when analyzing the DR maps with the Sliding Window method (see below).

Why and how to produce a denaturation-renaturation map (DR map). The patterning of the DNA was obtained with a modification of the method of melting mapping (or denaturation mapping) of [1]. This method by Reisner *et al.* patterns DNA into single and double stranded regions according to local base-composition, with double stranded regions retaining intercalating dye and hence fluorescing (visibly light), while single stranded regions lose intercalating dye and hence do not fluoresce (appear visibly dark). This is achieved by a uniform staining of the DNA prior to insertion in the chip, followed by a gradual heating of the DNA on the chip. As adenosine-thymine (AT) base pair are less stable thermodynamically than guanine-cytosine (GC) base pairs, AT-rich regions melt before GC-rich regions. This creates a 'barcode-like' pattern on the DNA: bright(dark) areas in double(single)-stranded DNA for GC(AT)-rich regions. In [1] the DNA molecules were imaged between increases in temperature, resulting in images taken at elevated temperatures of DNA with alternating double- and single-stranded segments.

The DR method presented here also heats stained dsDNA to denature AT-rich regions to produce a melting pattern. But after heating and consequential loss of stain in denatured regions, the DNA is renatured by cooling. This causes the denaturation pattern of light and dark regions to remain fixed in the renatured double-stranded form. This has several advantages:

1. our pattern is fixed—as in chemical fixing of a developed photographic print—while the method of [1] allows dye to leak from stained regions, which makes them shrink, because loss of dye reduces the persistence length of the double-stranded regions. Movies of melting patterns produced with the method of [1] show a shrinking length of partially extended DNA, so data-taking is literally a matter of shooting a moving target;
2. our pattern is fixed with the stained regions maximally extended, because maximal stain gives maximal length of the dsDNA;
3. since unstained regions are renatured in our method, they are more clearly visible, because dsDNA has longer persistence length than ssDNA, which is what the unstained regions consist of in the method of [1];
4. since unstained regions are renatured in our method, we do not have the helicity of the DNA as a fitting parameter. Less parameters fitted means our interpretation of data is more easily falsified, and if it stands, remaining parameters fitted are more precisely determined;
5. the thermal motion of dsDNA is simpler than the motion of a chain of segments of dsDNA connected by segments of two strands of ssDNA. Since our DR-processed DNA has simpler thermal motion, it is easier modeled, and data extraction is consequently more reliable.

How to produce barcodes from the DR pattern. For further analysis of the images, it is crucial that the DNA is kept still for so long that averaging over several images does not introduce a significant smear. Single DNA molecules were stretched in the device through confinement and hydrodynamic drag, resulting in almost full stretching, significantly more than previously reported for DNA melting mapping of DNA in nanochannnels, where the stretching is ∼50% of the contour length [1]. Our approach optimizes resolution in three ways:

1. By optimal stretching of the imaged barcode, as detailed above;
2. by eliminating local Brownian motion of the DNA along its length (longitudinal fluctuations of position); and
3. by eliminating transverse Brownian fluctuations of the DNA except for its longest wavelengths.

The long-wavelength thermal fluctuations of the stretched DNA do not affect resolution, since they only displace the DNA transversely between frames in movies of the nanoslit. This transverse displacement does not affect results, since we see the DNA clearly in the image and work with the pixels covering it, wherever they are. A time-trace of a segment of the image cut around the DNA is shown in Fig. S2. The lack of motion of the pattern in the direction of the DNA over a 5 second time-interval shows the excellent stability of the DNA's configuration in the nanoslit. Molecules that were significantly longer than the $440\,\mu$m of the nanoslit, were translated sideways in order to image their full length (see below).

To improve the signal-to-noise ratio in images of a given field-of-view, each pixel was time-averaged over 2.5 seconds = 50 frames and also averaged spatially, over a 1×9 strip of pixels ($\sim 180\,\mathrm{nm} \times 1.6\,\mu\mathrm{m}$) chosen orthogonal to the direction of the DNA and such that the DNA is seen to cut across this strip in all frames of the movie. A time-independent background—spatially non-uniform, but slowly varying—is identified and subtracted. The right-most part of Fig. S2 shows an averaged image, which contains a clear pattern. This is the sought DR map for the given field-of-view.

A DR map spanning the whole nanoslit is produced by 'stitching' together DR maps from overlapping fields-of-view with least squares fitting; see Fig. S3A. Fig. S3B shows that the match between adjacent regions is perfect, with no rescaling or morphing needed to create a match. Note in Fig. S3B how the ten steep sides of the five large peaks in the overlap region coincide pairwise between the two overlapping DR maps. This is what we mean by a perfect match. The small intensity differences also seen in Fig. S3B between the two overlapping DR maps are of the same magnitude as intensity fluctuations due to photon shot-noise plus excess noise from the electron-multiplication process of the EMCCD camera (see [12] Supplementary Note 1, p. 26), i.e., they are due to the finite statistics of the two signals compared here.

In summary, the chip design keeps the DNA at a fixed position for several seconds, eliminates longitudinal Brownian motion of the DNA, and leaves transverse Brownian motion trivial to cope with. Thus there is no need to compensate for Brownian motion of the DNA with morphing of the images, as was done in [1].

How to produce theoretical DR maps

DR maps based on hg18. Theoretical melting patterns of the reference genome (hg18, build 36) were computed using the software *Bubblyhelix* (http://bubblyhelix.org); see also [13]. Salt concentration and temperature are input parameters of this computation. As previously reported [14], the TBE buffer system to which 2-mercaptoethanol (BME) was added, has a complex ionic strength. Although the overall ionic strength can be calculated, the calculation is unreliable at the high temperatures used in the present work (76 °C) because the pKa-values of the components are known only close to room temperature; see, e.g., [15] p. 18. We therefore fitted the ionic strength, using lambda-DNA to produce this benchmark: At an ionic strength of 27.5 mM, there is optimal agreement between its theoretical melting pattern and its experimental map obtained with our procedure of melting followed by rehybridization (denaturation-renaturation) before imaging.

The output from the *Bubblyhelix* software is a theoretical probability map which gives the probability for each basepair to remain closed through the denaturation-renaturation cycle (i.e., the probability to appear bright on the DR map). The theoretical melting pattern was compressed to match the DNA in the experiment, then the point-spread function of the microscope was applied to it (i.e., the melting pattern was convoluted with a Gaussian with a standard deviation of 94 nm in accordance with the properties of the optical imaging system), and the resulting melting profile was binned into 179 nm-wide pixels, like our imaging system does. The result is our theoretical DR map.

Construction of NA18507-specific DR maps. The structural variations reported for the NA18507 genome by Li *et al.* [16] were incorporated into the reference genome (hg18, build 36) in order to compare the experimental barcodes obtained from denatured-renatured NA18507 chromosomal DNA with a NA18507-specific assembly of the genome sequence (compiled by Alexander Kanapin, Wellcome Trust Centre for Human Genetics, University of Oxford). Using the table of structural variants provided by Li *et al.* (`NA18507.final.updata` file from item 15 at: http://yh.genomics.org.cn/download.jsp), an assembly was created by incorporating the base sequence of insertions reported by Li *et al.* into the hg18 reference and removing the deletions reported. In a few instances, the length of insertions reported varied from the length of the sequence given for the insertion; in those cases, only the actual sequence reported was inserted. Theoretical DR maps were produced as described above. Additional structural variations reported in separate sequencing studies by McKernan *et al.* [17] and Kidd *et al.* [18] were taken into consideration separately since they did not provide the sequence of the insertions reported.

Localizing a single DNA-molecule within the genome

A key feature of our method of DR mapping is its ability to determine unambiguously the chromosomal origin of a single imaged DNA-molecule and then retain such a molecule for further investigation if this is desired. Below it is demonstrated how to align single molecules within the genome and prove that the match is unique.

How to align individual molecules to a reference genome. We quantify the match between experimental and theoretical *in silico* barcodes following [1] by introducing the figure-of-merit function χ^2: Let $I_{\rm exp}(j)$ denote the experimentally measured intensity in pixel j, and let

$$\hat{\mu}_{\rm exp}(i) = \frac{1}{N}\sum_{j=i}^{i+N} I_{\rm exp}(j) \ , \tag{3}$$

$$\hat{\sigma}^2_{\rm exp}(i) = \frac{1}{N}\sum_{j=i}^{i+N} [I_{\rm exp}(j) - \hat{\mu}_{\rm exp}(i)]^2 \tag{4}$$

be the local mean and variance, respectively, for a barcode of length N starting at pixel i. Here i can take the values $1,\ldots,n-N$, where n denotes the number of pixels in the whole DR map of a molecule in the slit. The rescaled intensity is

$$\tilde{I}_{\rm exp}(i,j) = \frac{I_{\rm exp}(j) - \hat{\mu}_{\rm exp}(i)}{\hat{\sigma}_{\rm exp}(i)} \ , \tag{5}$$

where $i \le j \le i+N$. The theoretical DR map $P_{\rm th}(j)$ is renormalized identically, and the result denoted by $\tilde{P}_{\rm th}(i,j)$.

The figure-of-merit function χ^2_i for position i is defined as

$$\chi^2(i) = \frac{1}{2N}\sum_{j=i}^{i+N}\left[\tilde{I}_{\rm exp}(i,j) - \tilde{P}_{\rm th}(i,j)\right]^2 \ . \tag{6}$$

The renormalized barcode (and its reverse image) was compared to the renormalized theoretical melting pattern by calculating χ^2 of their difference for every 100 base pairs (1/5th of a pixel, a distance across which the theoretical barcode changes). χ^2-values are found to scatter around 1, which is the expected value for χ^2 of two random patterns; a perfect match would correspond to $\chi^2 = 0$; a maximal mismatch would correspond to $\chi^2 = 2$. Neither are achieved experimentally, but the minimum in χ^2-values indicates where the barcode maps to the genome, if its χ^2-value is highly improbable according to the extremum-statistics of the distribution of mismatching values.

We identified a barcode's absolute minimum of χ^2 across the genome with its location in the genome. Different barcodes from the same molecule mapped to the same chromosome at distances that agree with their distances on the DR map of the DNA. In this manner we located 21 DR-mapped DNA-molecules within the known genome, covering at least 27.3 Mb in total; see Table S1. An example of a typical alignment, Molecule n from Table S1, is shown in Fig. S4.

The deepest minima for the χ^2-value for whole molecules were found with an overall stretching value of 96%-99% (with some exceptions), which amounts to yet another experimental determination of the degree of stretching. It confirms the value found by analyzing the DNA's transverse Brownian motion, but the true value is a few per cent lower: It is obtained by first correcting the DNA's contour length for the extension caused by staining with YOYO-1 [11]. The best alignment gives $\chi^2_{\rm min} = 0.08$. Note that a barcode that overlaps with a structural variation, will not give a statistically significant match to a reference genome without that structural variation. When no match is found, another barcode is selected from the DR map. The length of the DR map ($\sim$1 Mb) compared to the length of the barcode ($\sim$125 kb), allows us eight independents searches for a unique match. This is one of several advantages of working with molecules that are many times longer than the length of barcode necessary to ensure a unique match within the genome. We used this for two of the 21 molecules from Yoruban cells.

How to prove that an alignment is unique. When aligning a barcode covering only $\sim$125 kb within the whole human

genome, it is crucial to rule out that the minimal value for χ^2 could have occurred by chance. After all, the whole genome is 3.2 Gb long, and we attempt to align in two directions for every 0.1 kb. That gives $n = 2 \times 3.2\,\text{Gb}/0.1\,\text{kb} = 6.4 \times 10^7$ attempts—a large number of chances.

To prove that our alignment did not happen by chance, we assume that it did (our *null hypothesis*) and then use the mathematical simplicity following from this assumption to prove absurd consequences. The null hypothesis is thus rejected, which means that its complement (the *alternate hypothesis*) must be true: Our alignment is not a chance event.

To illustrate this reasoning quantitatively, we use Molecule n from Table S1 as example. Its alignment is shown in Fig. S4 and achieved $\chi^2_{\text{min}} = 0.19$. We first determine the probability p of finding $\chi^2 \leq 0.19$ in a single, *random* alignment. We base this calculation on the experimentally determined probability density function of χ^2-values that result from *random* alignment attempts. Then we determine the number of statistically independent alignments that are possible across the whole genome. This calculation we base on the correlation function for χ^2-values found in alignment attempts along the theoretical melting map. From these two numbers we calculate the probability P of finding *any* χ^2-values ≤ 0.19 by *random* alignment in the number of alignment attempts we made, i.e., in a genome of the length we consider. This is the so-called P-value of our null hypothesis: The probability of doing worse than we did, under the assumption that the null hypothesis is true.

The experimentally determined probability density function of χ^2-values is modeled as follows. A histogram of frequencies of χ^2-values that occur in Fig. S4C (excluding the part of the chromosome with unknown sequence), is shown in Figs. SS4D and SS5A. The fitted blue curve is a Gaussian with a quartic correction term, as discussed in the SI of [1]. This blue curve defines our probability density function of χ^2-values. It is used for downwards extrapolation of the experimental distribution of χ^2-values of mismatches. After the whole distribution has been normalized, its integral from $-\infty$ to χ^2_{min} is the probability p of getting a value as low as χ^2_{min} or lower from a single, random alignment. We found $p = 5{\cdot}10^{-16}$.

Our value for p is sensitive to our choice of downwards extrapolation, the blue curve in Figs. SS4D and SS5A. However, considering the distribution of our χ^2-values, it is a reasonable approximation, and that is fully sufficient for its use to calculate p, because what is crucial, is not the actual value of p, but its smallness. Any different, but realistic, description of the lower tail of this probability distribution function will *also* give an extremely low value for p, which is the relevant fact here.

A low value for p was expected: Considering that we did a lot of alignments and now focus on the lowest χ^2-value that we encountered. So according to the null hypothesis, p should equal approximately one over *the number of independent alignment attempts* we have made. The latter number is lower than the total number of alignment attempts we have made, because we moved the barcode only 0.1 kb between successive attempts, and DR maps are correlated on a scale longer than that. In the next subsection, we find the correlation length of DR maps. We do this for the record, so the reader has the procedure, should he or she need it more than we do.

We do not need the correlation length, because we can reject the null hypothesis without this number. An inequality is fully sufficient: We can treat *all* our $n = 6.4 \times 10^7$ attempts at alignment as independent. Picking that many χ^2-values at random and independently from any distribution, results in a smallest value that has probability $1/n \approx 10^{-8}$ for another draw to yield a lower value. This probability is substantially larger than our value $p = 5 \cdot 10^{-16}$.

Phrased differently: when the probability of getting a χ^2-value below χ^2_{min} in *one* random alignment attempt is p, then the probability of *not* getting any χ^2-values below χ^2_{min} in n independent random alignment attempt is $(1-p)^n$. For our case of $p \ll 1$, this number does not differ from $\exp(-np)$, not even from $1 - np$, since also $np \ll 1$. Thus we have for the probability P of getting any χ^2-values below χ^2_{min} in n independent random alignment attempt that $P = np = 3 \cdot 10^{-8}$. With this statistical support for the null hypothesis, it is rejected even if we are willing to accept it with unreasonably low levels of statistical significance. To the extent that anything can be proven statistically, we have just proven that our low value for χ^2_{min} is *not* a random result. It is an extremely significant alignment.

How to determine the number of independent alignments within a genome for a given barcode. From the sequence of χ^2-values shown in Fig. S4C, we calculated the correlation function

$$C(\delta) = \frac{1}{N-\delta-L} \sum_{i=1}^{N-\delta-L} \left[\chi^2(i+\delta) - \bar{\chi^2}\right]\left[\chi^2(i) - \bar{\chi^2}\right] \quad . \tag{7}$$

Here L is the length of the barcode, N is the length of the chromosome, i the position on it, and $\bar{\chi^2} \approx 1$. The correlation function $C(\delta)$ decreases exponentially with a correlation length $\lambda \sim 3$ kb; Fig. S5. Consequently, the number of statistically independent alignments possible over the whole genome is $n_{\text{indep}} = \frac{2\times 3.2\text{Gb}}{2\lambda + 0.5\text{kb}} \sim 10^6$ (see [19], Eq. (31)). Here, the factor two in the numerator accounts for the two read directions, and the 0.5 kb in the denominator is one bin-width. We see that the number of *independent* attempts of alignments which we have done, is a factor seven smaller than the total number of attempts: $\frac{2\lambda + 0.5\text{kb}}{0.5\text{kb}} \sim 7$.

How to apply the Sliding Window method to experimental data

Sliding Window Analysis. Sliding window analysis is a piecewise comparison done by repeating, for all pixel positions on the experimental DR map (e.g., in steps of 1 pixel (∼500bp)), the localization of a 150-pixel barcode onto the theoretical DR map (Fig. S6II–IV). This yields a figure-of-merit, $\chi^2_{\text{min},i}$, for each position i on the DR map to which we can assign a position b_i on the reference genome (Fig. S6V), and it indicates how well the single molecule DR map agrees with the reference genome. Agreement between the experimental and the theoretical maps is signaled by low values for $\chi^2_{\text{min},i}$, so low that they are highly significant statistically. Conversely, a stretch of *high* values for $\chi^2_{\text{min},i}$ reveal a local mismatch between the molecule compared to the reference genome, which may indicate a structural variation.

Correction for local and global stretching. When the Sliding Window method described /changeabove is applied to experimental DR maps, two effects complicate the simple picture outlined: (i) the position-dependent stretching of the DNA in the nanoslit, and (ii) local variations in the degree of stretching of the DNA caused by the local variations in the degree of staining.

First, as discussed above, the tension of the DNA in the nanoslit is non-uniform due to the hydrodynamic drag causing the extension. The tension has a maximum at the center

and decreases as a first-degree polynomial towards the ends, which gives rise to position-dependent stretching $0 < q_i < 1$ (i being the pixel number on the experimental DR map). Here 0 corresponds to unstretched DNA, while 1 is for DNA stretched to its full contour length. We refer to this stretching as global stretching: Global vs local refers to the cause of the stretching, and not its degree. The piece-wise linear tension of the DNA in the nanoslit can be convert to a degree of stretching using the Marko-Siggia interpolation formula; see **??**Eqs. 1]EqMarkoSiggia. If the model for the tension is correct, the optimal tension q_i for the sliding window centered at position i and localized to position b_i in the genome will fall on a curve which can be fitted to the theoretical expression for the stretching.

Secondly, staining of the DNA with the intercalating dye YOYO-1 enhances the contour length of the DNA in a manner that depends on the staining ratio [11]. After the denaturation-renaturation process of the DNA, the DNA will only be stretched by YOYO-1 where this is YOYO-1, i.e., in its bright parts. This local stretching can be incorporated in the theoretical map by scaling it for each basepair with a factor $1 + r_{\rm ds}P_{\rm ds}$. Here $P_{\rm ds}$ is the probability that a given basepair remains closed through the heating cycle. Since it is a physical property of the DNA that is modeled here, it is important that this local basepair-wise stretching is done before the global stretching of the DNA is applied to the theoretical map, and before the theoretical map is folded with the point-spread-function of the optics. The result is a staining-corrected theoretical map, in which basepairs no longer are equidistant. Since we do not know the degree of local stretching of the segments of DNA that remain stained after a DR cycle, we use a value for $r_{\rm ds}$, which best fits the stretching curve based on the piecewise linear tension and the Marko-Siggia interpolation formula; Eqs. 1,2.

We found that the best fits are obtained with a stretching factor $r_{\rm ds} = 0.13$, but our results are not sensitive to variations of $r_{\rm ds}$ in the interval $0.1 < r_{\rm ds} < 0.2$. This regime is also in good agreement with the expected value based on the initial staining ratio of the DNA and the results in [11].

An algorithm for the Sliding Window method. When applying the Sliding Window method to experimental data, a successive reduction of the search space is performed, as outlined below:

1. From an experimental DR map covering the whole slit (∼2300 pixels∼1.3 Mb), a smaller segment (e.g. 240 pixels∼125 kb) is used to localize the molecule in the genome as described above.
2. As the length of a molecule covering the slit is ∼1.3 Mb, the search space on the theoretical melting pattern is reduced to an interval of a few Mb covering the molecule, and the theoretical melting pattern is corrected for staining (see above). Now a Sliding Window calculation is performed where the figure-of-merit function $\chi^2(i)$ is also minimized with respect to the stretching q_i. This step is denoted the 'unrestricted' search. The position on the theoretical melting pattern, b_i, and the corresponding stretching q_i is recored [blue points in Fig. S7A–B].
3. If the molecule is located correctly in the genome, the majority of the points (i, b_i) will fall on a diagonal (with some clear outliers), and the stretching can be fitted to a function based on a piecewise-linear tension and the Marko-Siggia formula for the degree of stretching; see above. This gives a fitted stretching for each location of the probe $q_i^{\rm fit}$, see Fig. S7B. In all fits a persistence length for the DNA of $L_p = 60$ nm is used.
4. Finally a new sliding window calculation is performed—referred to as 'restricted'—where the search space on the theoretical melting pattern is reduced to a range of, e.g., ±50 kb around the diagonal [red dashed lines in Fig. S7A], and the stretching at each placement of the probe is fixed at $q_i^{\rm fit}$. The corresponding positions on the theoretical melting pattern is shown in Fig. S7A with red symbols.

The corresponding values for $\chi^2_{{\rm min},i}$ for the 'unrestricted' and 'restricted' sliding windows are shown in Fig. S7C. The advantage of the procedure outlined above is that the $\chi^2_{{\rm min},i}$ values for the 'restricted' sliding window better reflect the quality of a match between the experimental and theoretical melting patterns than the search with 'unrestricted' sliding windows. For the latter method unphysical tensions and jumps in position can create too low $\chi^2_{{\rm min},i}$-values. In addition, by reducing the search space and fixing the stretching, step (iv) can be repeated for shorter lengths of the sliding window $L_{\rm bc}$, e.g. down to around 40 kb. This increases the resolution of the method, allowing for detection of isolated structural variations separated by even shorter distances.

Detecting structural variations

Below we elaborate on the detection of structural variation using the same molecules as in the main text as examples, but providing more technical details.

Detecting large scale variations—inversion. As an example of a structural variation greater than 0.5 Mpbs, a match of an experimental DR map on chromosome 16 around 22 Mb is used; Fig 1D in the main article. The structural variation is a segment of the DNA from further up the contour which has been cut (or copied), reversed and inserted at a new position, i.e., an inversion. An overlapping region spanning approximately 140 kb is clearly observed.

Figure S8 shows the results of 'unrestricted' Sliding Window calculations for the $\chi^2_{{\rm min},i}$-values, and the optimal stretchings q_i for both the 'direct' search direction (blue) and 'reverse' (red). For the positions on the genome, the 'direct' direction does not match up to a breakpoint around pixel 1150 and the positions scatters randomly. After this point they match (except for a possible minor structural variation near pixel 1700). The optimal stretchings are also unsystematic up to this point, and then falls on a smoother curve. The opposite holds for the 'reverse' search direction. The difference in the qualities of the matches on each side of the center region are further enhanced by the 'restricted' Sliding Window calculation for $\chi^2_{{\rm min},i}$; compare Figs. 7B and 7E. An overlapping region appears around pixel 1025 to 1275, where the experimental DR map matches at two different locations and in different directions. These regions approximately span 21.31–21.45 and 22.33–22.47 Mb, respectively. The theoretical (blue) and experimental (red) DR profiles for these intervals are plotted in Fig. S9 for both orientations of the experimental DR map. This directly visualizes the good match.

This example clearly shows that the DR map combined with the Sliding Window method can be used to detect large-scale structural variation as large as Mb.

Detecting large scale variations—translocation. Where a single molecule that we have mapped partly matches to one chromosome and partly to another, it may be due to insertion of an extra copy of a segment of the genome (copy number variation, CNV) or its transfer to a new location (translocation). We have demonstrated that the DR map can detect SV that is due to either a translocation or a copy number gain, covering more than 100 kb.

Figure S10A shows a 'restricted' Sliding Window calculation for an experimental DR map, which has been located to chromosome 19 around ∼61.5–63 Mb. While smaller structural variations spanning a few kpbs cause sharp peaks in the $(i, \chi^2_{\min,i})$-plots, the present molecule shows a large plateau between pixels 1250 to 1600. The $\chi^2_{\min,i}$-values indicate a poor match; see Fig. S10B. The search across the whole genome in Fig. S10E shows that a 250 pixels∼132 kb segment of the barcode covering a part of this interval (marked with grey in Fig. S10) matched on chromosome 9 centered at 22.183 Mb. The match is significant with a P-value for a search over the whole genome equal to 0.04 ($p = 4{\cdot}10^{-8}$ for a single match and correlation length $\lambda = 3.7$ kb). Figure 9C–D show a Sliding Window calculation for this region. A plot of the experimental (red) and theoretical (blue) DR profiles at the match is shown in Fig. S10F.

This shows that the Sliding Window method can detect structural variations in the form of translocations. The main limitation is that the segment transferred from one chromosome to another has to be sufficiently long to create a unique match across the entire genome. In practice it may set a lower limit around 100 kb, but it depends on the specific sequence.

Detecting small structural variations. Figure 3A–C in the main text illustrates how the value of $\chi^2_{\min,i}$ will increase and then decrease as the barcode 'slides' over the structural variation. The increase/decrease on each side of the variation will be linear with a slope equal to $\pm 1/L_{\rm bc}$ ($L_{\rm bc}$ being the probe length) as the expectation value of χ^2 is 1 for two random sequences. The intersection of the lines marks the position of the structural variation on the experimental map. The distance on the x-axis between two pairs of lines is $L_{\rm bc}$ for a deletion, and $L_{\rm bc} + \Delta_{\rm ins}$ for an insertion; $\Delta_{\rm ins}$ being the size of the insertion. An insertion (deletion) in the experimental map relative to the theoretical map will shift the (i, b_i)-line downwards (upwards); the size of this shift reveals the size of the variation.

Some complications: Several structural variations present within a distance shorter than the probe length $L_{\rm bc}$ cannot be distinguished, but they will still be defined as a small SV (in contrast to the large scale structural variations discussed above). The shift of the (i, b_i)-lines will be the sum of the insertions and deletions, and the width of the peak in $\chi^2_{\min,i}$ will equal the region on the experimental DR map spanned by the variations. In addition, the number of base pairs per pixel is not constant across the slit as the global and local stretching is non-uniform. This results in a curved line for (i, b_i). The analysis of small structural variations presented here is simplified if this bending is removed: Correcting the theoretical map for the effect of staining using a scale factor $r_{\rm ds}$, and the experimental map for the non-uniform stretching across the slit by dividing each pixel with the stretching factor q_i, the pairs (i, b_i) will fall on a straight line.

As an example, the molecule shown in Fig. 3B in the main text and in Fig. S7 is used. Only the large peak marked in the grey box in Fig. S7 is analyzed. In this region deletions were observed by McKernan *et al.* [17] on the genome NA18507, and by Kidd *et al.* [18] on NA19129. Figure S11A shows the value of $\chi^2_{\min,i}$ versus the position in the slit, and Fig. S11B the corresponding position in the genome. The center of the peak is located at 42.491 Mb. The slopes of the solid red lines and dashed grey lines in Fig. S11A are $\pm 1/L_{\rm bc}$. The red(grey) lines corresponds to a variation covering a region of 15(7) kb, respectively. In Fig. S11C the axes are rescaled to compensate for the staining and the stretching as described above. The straight lines with equal slopes are fit to the data in the interval covered by them, and the dashed grey line marks the center of the peak. Finally, Fig. S11D shows the vertical distance between the average of the fitted diagonals in Fig. S11B and the (i, b_i)-line for points away from the center of the peak. If the theoretical and experimental maps matched completely, Fig. S11D would show only horizontal plateaus and vertical steps.

The offset between the diagonals in Fig. S11C indicates a ∼7 kb insertion, in contrast to the deletions detected in [17] and [18]. However, the width of the peak indicated by the red lines in Fig. S11A marks a 15 kb region, which clearly match better than the 7 kb region marked by the dashed grey lines. This implies that the structural variation is not a *single*, isolated structural variation, but rather several variations adding up to a ∼7 kb insertions over a 15 kb interval. The same analysis with a probe length of $L_{\rm bc} = 150$ pixels showed a ∼6 kb insertion over a ∼25 kb interval centered at 42.497 Mb (Fig. S12), which illustrates the improved localization by applying a shorter probe length. However, Fig. S11B–C also demonstrate that close to the peak more matches occur away from the diagonal.

As mentioned above, the main limitation of the method, when it comes to detection of small SVs, is that SVs separated by less than the length of the probe ($L_{\rm bc}$) will show up as a single peak in $\chi^2_{\min}$ with a dip in the middle; see, e.g., the peak in Fig. S11A. The difference between plateau values in Fig. S11D will then correspond to the sum of the insertion/deletions. Better resolution can be achieved by using shorter probes, but it also increases the number of spurious matches far from the diagonal as shorter probes contain less information. This effect can be reduced by reducing the search range in the 'restricted' Sliding Window method, but with the present experimental setup, the limit of the probe length is around 40 kb. Another limitation is that only structural variations large enough to change the thermodynamical properties of the DNA (i.e., altering the melting profile) can be detected.

Reproducibility of the $\chi^2_{\min}$-signal. The Sliding Window method relies on detecting matches and mismatches between experimental and theoretical melting patterns. As only a single, non-amplified molecule is imaged, it is fair to ask whether the mismatch observed (i.e., high values for the figure-of-merit function) are caused by true variations in the underlying genomic structure, or stems from the experimental setup. Figure S13 shows a 'restricted' Sliding Window calculation for a molecule later localized to chromosome 21 around 40 Mb. The molecule was imaged over the the slit, then shifted halfway through the slit by applying a pressure, and imaged again. For each of the two distinct maps, the 'restricted' Sliding Window method was applied as described in Sec. . Plots of position in the slit i versus $\chi^2_{\min,i}$ are found in Fig. S13A–B. When $\chi^2_{\min}$ is plotted against the position on the genome in Fig. S13C, we find almost perfect agreement, without any fitting, in the overlapping region (∼38.8–39.3 Mb) for which the DNA was imaged at different locations in the slit (i.e., also at different stretchings). This demonstrates that the present chip design can be used to image, localize, and analyze experimental DR maps reproducibly. It emphasizes that the mismatches observed between the experimental and theoretical melting patterns are not caused by experimental conditions like Brownian motion of the DNA, artifacts of the chip or in the imaging system. Finally, it also support the vision that with the present technology, it is possible to spool the molecules across the slit, image, and analyze them to get a coherent view of molecules spanning tens of Mb.

Sequencing Data Analysis

A key feature of the LoC design and DR mapping protocol is the ability to recover the individual DNA molecules that have been mapped in order to integrate the map with conventional genomic analysis and reconcile genetic variation at multiple length scales. To this end, we rescued the DR mapped molecule from the LoC device, amplified it using multiple-displacement amplification (MDA), and prepared an aliquot for Illumina paired-end sequencing.

Data from the sequencing run was mapped to the hg19 reference assembly using *stampy* [20] in order to determine the genomic origin of the single DNA molecule "rescued" from the chip. The data showed a number of anticipated artifacts [21, 22]. In particular, genome coverage appeared to punctuate, as is known to be a feature of MDA from a small amount of material; here we amplified ∼2,000x less material than a single cell. Additionally, the stochastic nature of MDA from single molecule templates is known to bias read depth [23]; we therefore aimed to use the footprint of the sequencing reads. Using a custom bash script, the human genome was subdivided into 1 Mb bins, and each bin was further divided into 1,000 1-kb regions. The number of these smaller regions with more than 51% mean coverage was counted per larger 1 Mb bin using *samtools* [24]. A single peak was observed on chromosome 9, centered on 110 Mb, as shown in Fig. S14**a**, below.

To analyse the peak observed on chromosome 9 in greater detail, a read density map was constructed by counting the number of reads mapping to each kb of the genome. We observed discrete clusters of well-covered regions, which we termed "islands." Only on chromosome 9 did we observe a clustering of these islands, which we termed an "archipelag"—other minor peaks on the footprint map were consistently composed of single, isolated islands (Fig. S14**b**,**c**). This "archipelago" pattern is consistent with the read density patterns of other long single-molecule templates amplified by multiple-displacement-amplification that have been reported in the literature [23]. We constructed a read density map of these previously reported results to provide a direct visual comparison (Fig. S14**d**) by obtaining the raw sequencing reads from the NCBI SRA (accession no. SRR077303) and remapping them to the hg19 reference genome as described below, using an insert size window of 0-1000 bp. Perhaps not coincidentally, the width of the islands in both cases is on the same order as the processivity of the phi29 DNA polymerase used in the MDA protocol [25].

In order to confirm that this hit was not due to varying "mappability" of the genome due to gaps in the reference assembly or repeat regions, we mapped a simulated set of reads to the hg19 reference genome to determine an "expected" number of reads per chromosome, and then compared it to the "actual" number of reads we obtained [22]. We used the *Metasim* software package [26] to generate 1 million simulated mate-pairs corresponding to 10 million 51bp Illumina sequence reads from chr1-Y of the hg19 reference assembly for a human with a genotype of 2nXXYY. Empirical error profiles for 80 bp were used as supplied by the authors (http://ab.inf.uni-tuebingen.de/software/metasim/) except with the length trimmed to 51 bp. The reads were then aligned using *bowtie* [27] with the command:

```
/Applications/seq/bowtie/bowtie --chunkmbs 256 -f -S -p8
-m 1 --best --strata -I
$insertMin -X $insertMax $reference -1 $readsForward -2
$readsReverse |
samtools view -S -b - | samtools sort - ${readsForward}_bowtie.bam
```

Read density maps were constructed as above (Fig. S14**e**); no archipelago pattern is present.

Lastly, in order to automate detection of archipelagos in our data set, clusters of nearby (<20 kb away) well-covered regions (1 kb regions with >51% coverage) were grouped into islands. Islands longer than 5 kb were selected for further analysis and those within 1 Mb of one another were then collected together as a single archipelago using a custom algorithm. The observed archipelagos in a given data set were reported along with their overall length and constituent number of large islands to determine their significance; uniquely, reads on chromosome 9 clustered together in a large archipelago.

The output of the algorithm also provides a non-subjective manner in which to estimate the length of the long DNA strand using the edges of the archipelago, plus half the mean distance between the islands. In the case of the archipelago on chromosome 9, its extent is is 3.5±0.1 Mb—consistent with the visual inspection of the DNA molecule spanning the nanofluidic portion of the chip and beyond—from 108.3 to 111.6 Mb.

Heterozygosity Analysis

To confirm the isolation of a single DNA molecule, we then sought to compare the rate of heterozygosity of our rescued DNA molecule. After isolating total DNA from Jurkat cells in mammalian tissue culture, we carried out genotyping using the Illumina Infinium assay using a Human Omni1-Quad (v1.0) chip. Genotyping calls were made using Illumina software (Illumina, Cambridge, UK) and cross-matched with well-covered (depth >19 reads) positions in the sequencing data of the single DNA molecules: While all 70/70 relevant SNPs were called as homozygous from the sequencing data of the rescued DNA molecule, genotyping of bulk Jurkat DNA yielded just 23/70 homozygous positions.

To extend the analysis beyond these SNPs, we determined the degree of heterozygosity for all well-covered positions in our sequencing data by calculating the degree of discordance amongst the reads:

```
samtools mpileup -f chr9.fa -r 'chr9:108285000-111569000'
mol5.bam | awk
'{if($4>19) {split($5,x,","); split($5,y,".");
nummatch=length(x)+length(y)-2; if(nummatch/$4<0.75 &&
nummatch/$4>0.25)
print $1,$2,$3,$5,$4, nummatch}}' > mol5NoMatch.txt
```

Overall, 99.97% of all well-covered positions were homozygous, comparable to the rate of homozygosity present in sequencing data previously reported for physically isolated single DNA molecules [23] (99.99%, NCBI Accession Number SRR077303).

mFISH Analysis

See Fig. S15.

1. Reisner W, et al. (2010) Single-molecule denaturation mapping of dna in nanofluidic channels. Proceedings of the National Academy of Sciences of the United States of America 107:13294–13299.
2. Das SK, et al. (2010) Single molecule linear analysis of dna in nano-channel labeled with sequence specific fluorescent probes. Nucleic Acids Research 38:Article No.: e177.
3. Fang Lim S, et al. (2011) DNA methylation profiling in nanochannels. Biomicrofluidics 5:034106–034106–8.
4. Nyberg LK, et al. (2012) A single-step competitive binding assay for mapping of single DNA molecules. Biochemical and Biophysical Research Communications 417:404–408.
5. Jo K, et al. (2007) A single-molecule barcoding system using nanoslits for dna analysis. Proceedings of the National Academy of Sciences of the United States of America 104:2673–2678.
6. Cipriany BR, et al. (2010) Single molecule epigenetic analysis in a nanofluidic channel. Anal. Chem. 82:2480–2487.
7. Kim Y, et al. (2011) Nanochannel confinement: DNA stretch approaching full contour length. Lab on a Chip 11:1721.
8. Rasmussen KH, et al. (2011) A device for extraction, manipulation and stretching of DNA from single human chromosomes. Lab on a chip 11:1431–1433 PMID: 21350789.
9. Sakaue T (2006) Dna electrophoresis in designed channels. The European Physical Journal E 19:477–487.
10. Marko J, Siggia E (1995) Stretching dna. MACROMOLECULES 28:8759–8770.
11. Günther K, Mertig M, Seidel R (2010) Mechanical and structural properties of YOYO-1 complexed DNA. Nucleic Acids Research 38:6526–6532.
12. Mortensen KI, Churchman LS, Spudich JA, Flyvbjerg H (2010) Optimized localization analysis for single-molecule tracking and super-resolution microscopy. Nature Methods 7:377–381.
13. Tostesen E, Liu F, Jenssen T, Hovig E (2003) Speed-up of DNA melting algorithm with complete nearest neighbor properties. Biopolymers 70:364–376.
14. Reisner W, et al. (2007) Nanoconfinement-enhanced conformational response of single dna molecules to changes in ionic environment. Physical Review Letters 99:058302–1–4.
15. Perrin DD, Dempsey B (1979) Buffers for PH and Metal Ion Control (John Wiley & Sons, Incorporated).
16. Li Y, et al. (2011) Structural variation in two human genomes mapped at single-nucleotide resolution by whole genome de novo assembly. Nature Biotechnology 29:723–730.
17. McKernan KJ, et al. (2009) Sequence and structural variation in a human genome uncovered by short-read, massively parallel ligation sequencing using two-base encoding. Genome Research 19:1527–1541.
18. Kidd JM, et al. (2008) Mapping and sequencing of structural variation from eight human genomes. Nature 453:56–64.
19. Flyvbjerg H, Petersen HG (1989) Error estimates on averages of correlated data. The Journal of Chemical Physics 91:461–466.
20. Lunter G, Goodson M (2011) Stampy: A statistical algorithm for sensitive and fast mapping of illumina sequence reads. Genome Research 21:936–939.
21. Blainey PC, Quake SR (2011) Digital MDA for enumeration of total nucleic acid contamination. Nucleic Acids Research 39:e19–e19.
22. Navin N, et al. (2011) Tumour evolution inferred by single-cell sequencing. Nature 472:90–94.
23. Fan HC, Wang J, Potanina A, Quake SR (2011) Whole-genome molecular haplotyping of single cells. Nature Biotechnology 29:51–57.
24. Li H, et al. (2009) The sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079.
25. Blanco L, et al. (1989) Highly efficient DNA synthesis by the phage phi 29 DNA polymerase. symmetrical mode of DNA replication. Journal of Biological Chemistry 264:8935–8940.
26. Richter DC, Ott F, Auch AF, Schmid R, Huson DH (2008) MetaSim—A sequencing simulator for genomics and metagenomics. PLoS ONE 3:e3373.
27. Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biology 10:R25.

Table S1. Mapping of 21 different NA18507 molecules (a–u) to the hg18 reference with a barcode length of 240 pixels$\sim$125 kb taken near the center of the slit. Barcodes from the same molecule mapped to the same chromosome at a nearby position emphasizing the robustness of the method. Notice that the optimal stretching is consistently 0.96–0.99, except for the apparent full/over-stretching of the molecules f, k, l1, p1, r and j1. The latter is due to the lack of local stretching caused by staining.

Molecule	Chr. no.	Position [Mb]	$\chi^2_{\min}$	Stretching
a	1	89.561	0.18	0.97
b	2	64.115	0.22	0.98
c	4	53.074	0.18	0.98
d1	4	62.759	0.24	0.98
d2	4	63.526	0.17	0.96
e	5	42.931	0.17	0.97
f	6	17.306	0.15	1.00
g	6	17.899	0.16	0.99
h	6	31.457	0.18	0.98
i	7	4.477	0.26	0.97
j	7	42.274	0.16	0.98
k	7	45.659	0.09	1.00
l1	7	75.424	0.22	1.00
l2	7	76.100	0.08	0.98
m	14	75.681	0.12	0.99
n	15	28.878	0.19	0.97
o	16	16.591	0.16	0.99
p1	16	21.492	0.10	1.02
p2	16	21.913	0.08	0.98
q1	17	47.876	0.23	0.97
q2	17	49.149	0.20	0.97
r	17	74.096	0.28	1.00
s1	19	18.000	0.37	1.00
s2	19	18.495	0.30	0.99
t	19	62.049	0.19	0.98
u1	21	38.772	0.24	0.98
u2	21	39.462	0.18	0.97

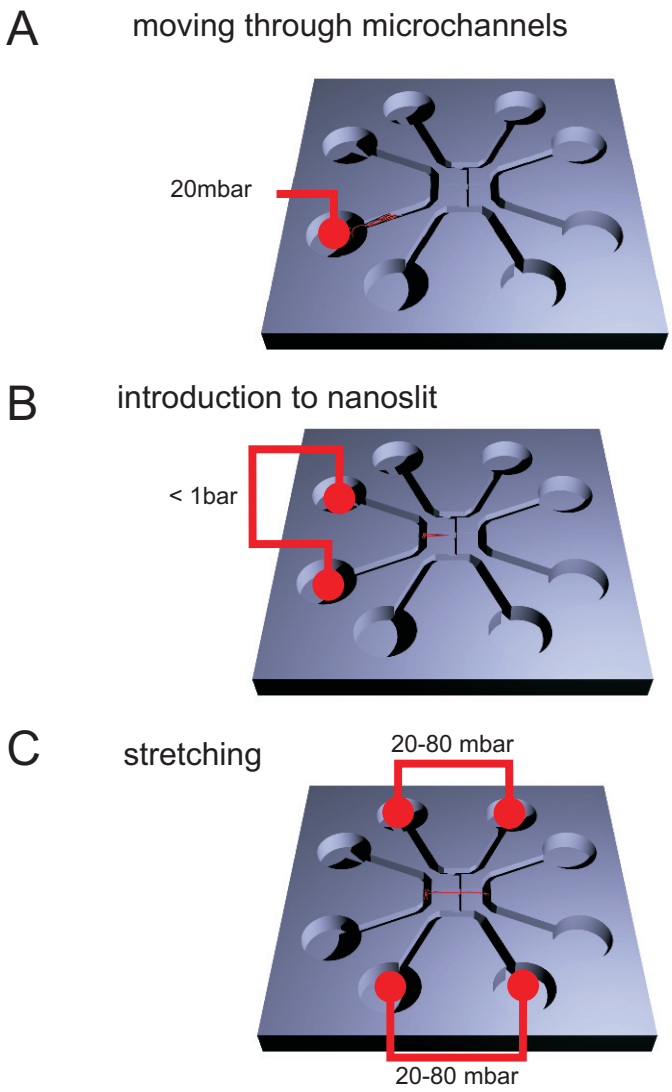


Fig. S1. Operating the device design for stretching mega base pair lengths of DNA. A dark field optical image of the silica device with its cross-shaped slit connecting four microchannels is shown in the main text. Excess pressure P is applied in opposite microchannels, which causes a flow into the two vertical arms of the slit towards the center, and then out the other two horizontal arms (arrows). This flow outwards from the centre will stretch a DNA molecule spanning the two horizontal arms. (A) DNA is moved through the microchannels at pressures below 20 mbar, (B) then forced into the nanoslit at pressures up to 1 bar. (C) When DNA is arranged in the nanoslit, a presure drop up to 80 mbar creates a flow in the nanoslit that stretches the DNA up to ∼95% of its crystallographic length.

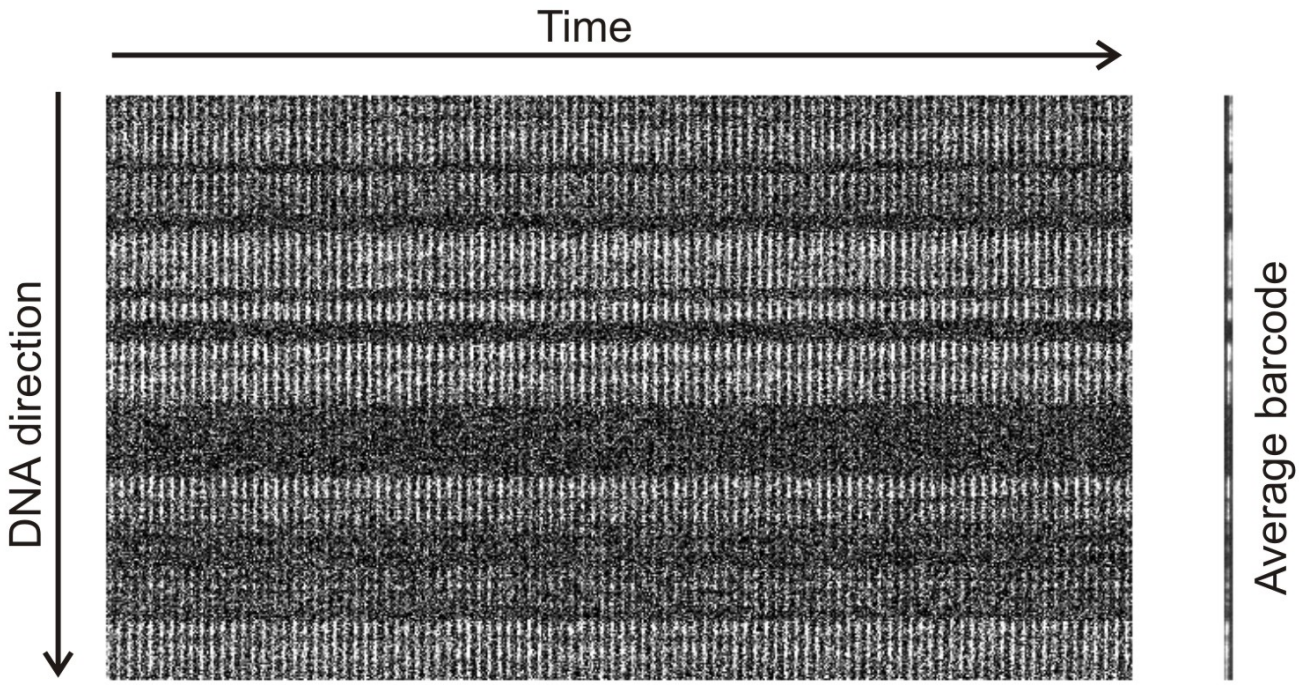


Fig. S2. "Barcode" pattern of single molecule from movie of a single field-of-view. Left: Due to near-total stretching, the melting pattern of a stretched molecule barely moves in 100 frames = 5 s shown here. Longitudinal drift is faintly visible. Simple time-averaging of the frames produces an image, which was averaged over rows of pixels that cover the DNA-image in its transverse direction. This produces a DR map of the molecule.

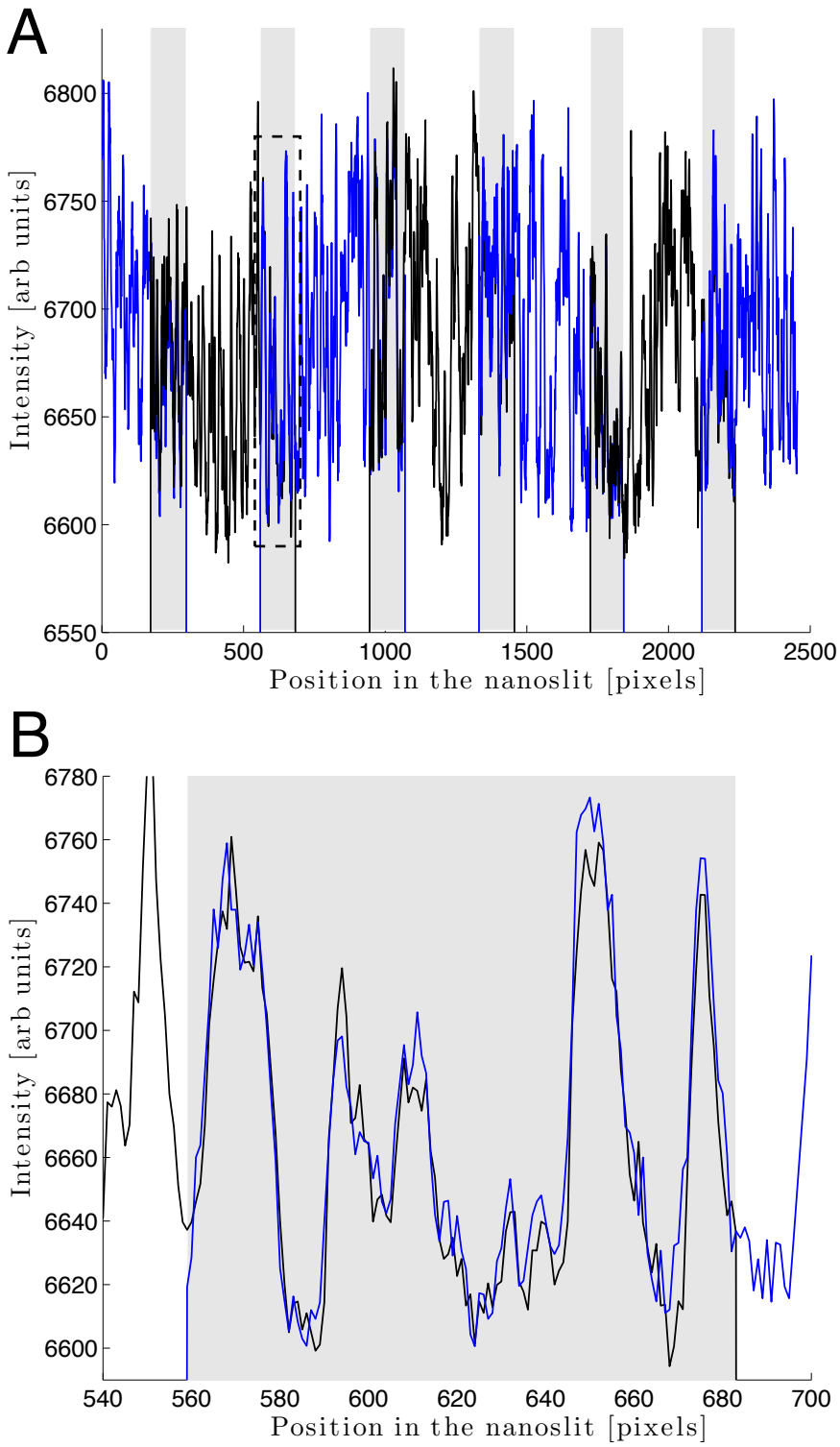


Fig. S3. 'Stitching' together DR maps from overlapping fields-of-view. (A) The intensities averaged over 50 frames (2.5 seconds) and 9 pixels ($\sim$1.6 μm) across the DNA for each of the seven fields of view covering the 440 μm of the nanoslit. Each field of view is 512 pixels $\sim$92 μm long. The six gray rectangles marks the overlapping regions. (B) Zoom of the dashed box in (A) shows the perfect match of overlapping parts of DR maps. No morphing and/or re-scaling was applied to the DR maps to achieve this match.

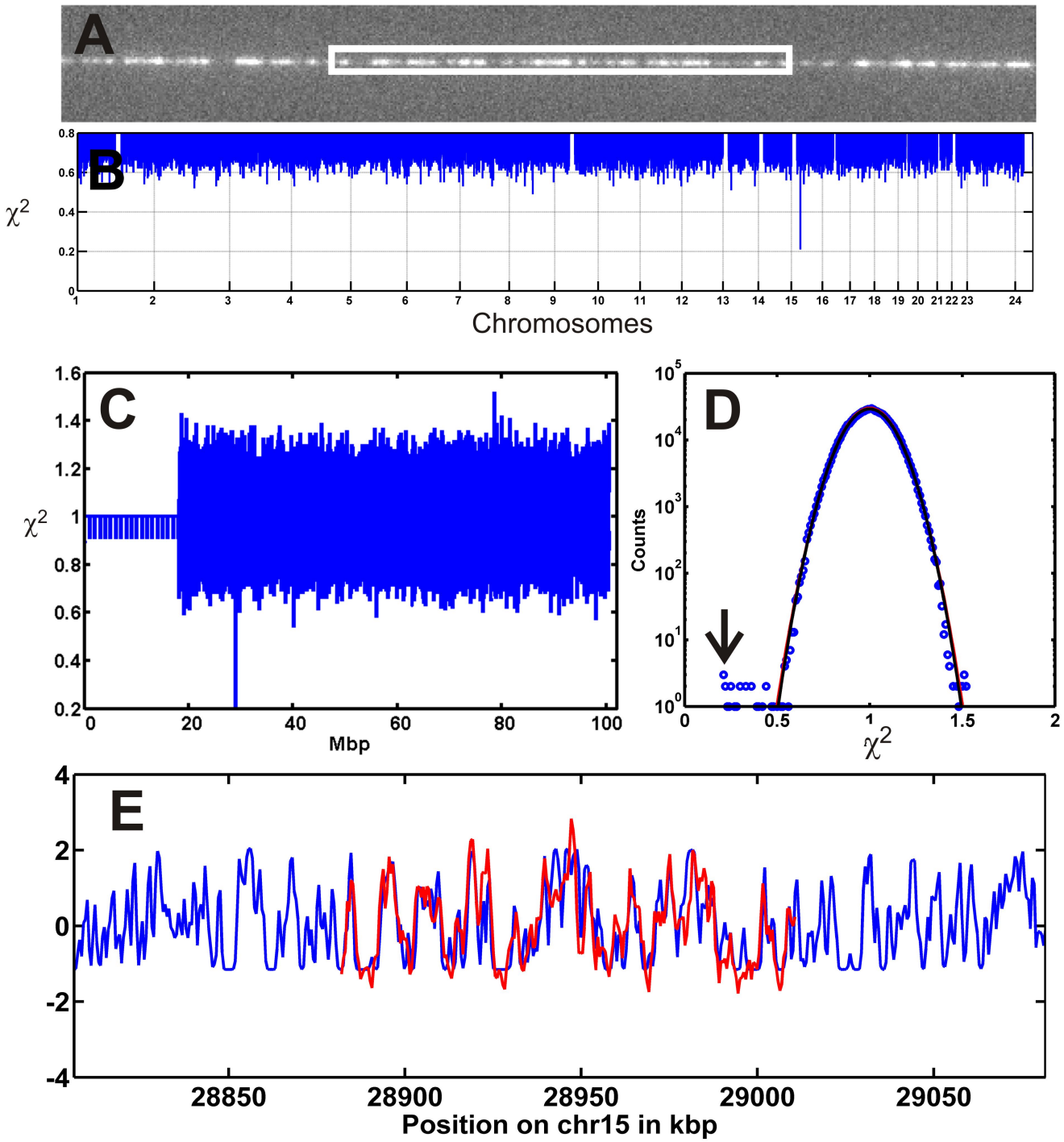


Fig. S4. Alignment of a DNA-molecule within the human reference genome (hg18, build 36). (A) DR map of Molecule n from Table S1. The rectangle indicates the region of the DNA used as barcode. (B) χ^2-values for the barcode at positions across the theoretical DR map. (C) The barcode maps to chr15 at position 28,878 kb. (D) Histogram of χ^2-values for the barcode found across chr15, excluding values from unknown part of the sequence (i.e., the acrocentric arm of chr15). The arrow indicates the improbably low values of χ^2, clear outliers relatively to the distribution of χ^2-values for mismatches, which is well described by a Gaussian with a quartic correction and shown fitted to the measured distribution of χ^2-values of mismatches. (E) Alignment of the experimental barcode from Molecule n (red) and theoretical melting pattern of chr15 (blue) centered at position 28,878 kb.

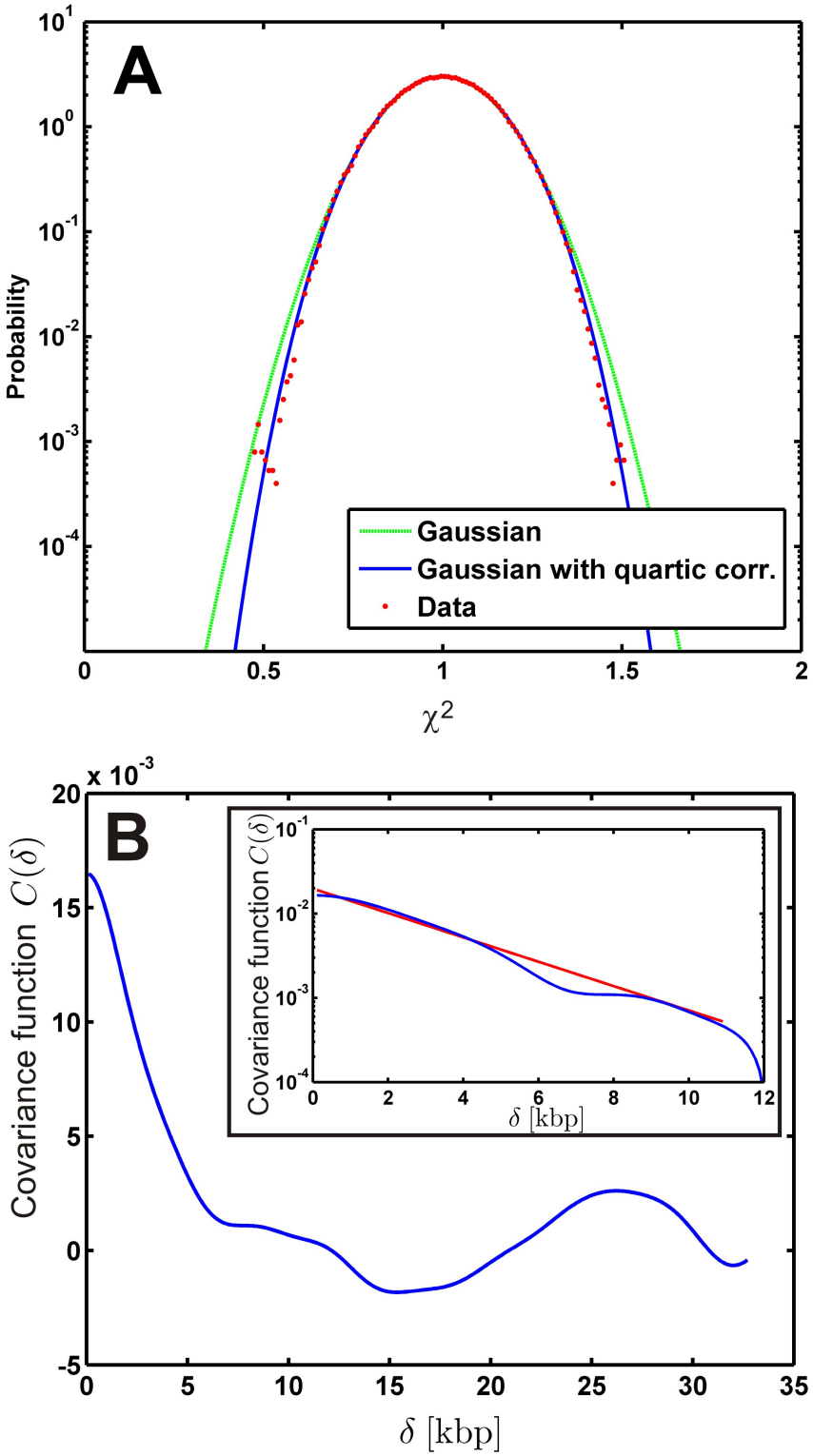


Fig. S5. Probability distribution and correlation length of χ^2 for the same data as in Fig. S4. (A) Red points: Frequency distribution of experimental χ^2-values, excluding the match. Green curve: Gaussian distribution fitted to red points. Blue curve: Gaussian with quartic correction fitted to red points. (B) Blue curve in lin-lin plot and lin-log inset: Covariance function for χ^2-values. Red curve in inset: Exponential function fitted to blue covariance function, with fitted characteristic length of 3 kb.

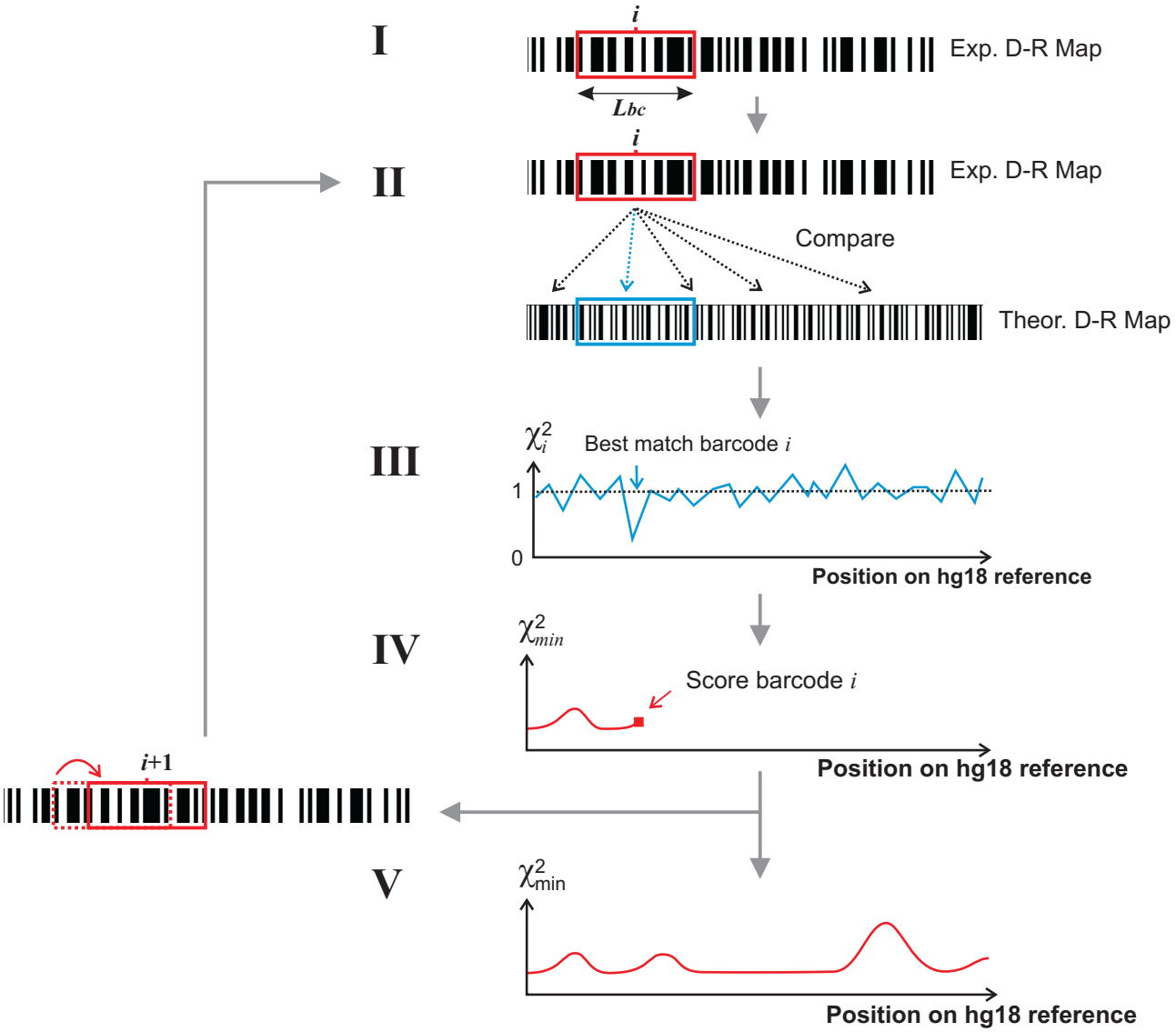


Fig. S6. Barcode localization and Sliding Window analysis. An experimental DR map covering the single molecule in the slit is assembled. (I) A segment of length L_{bc} centered and fixed at pixel i on the experimental DR map (red box) selected, (II) and compared to all positions on the theoretical melting pattern (e.g., the blue box), which has been corrected for staining and stretching. (III) For each comparison, the value of the figure-of-merit function χ^2_i is calculated [see Eq. 6], and (IV) the minimum value $\chi^2_{\mathrm{min},i}$ of the figure-of-merit function is recorded together with the corresponding position b_i. For the Sliding Window analysis the selected segment is shifted ("slid") one pixel and the procedure (II–IV) is repeated. (V) The resulting plot of $(b_i, \chi^2_{\mathrm{min},i})$ for all values of i.

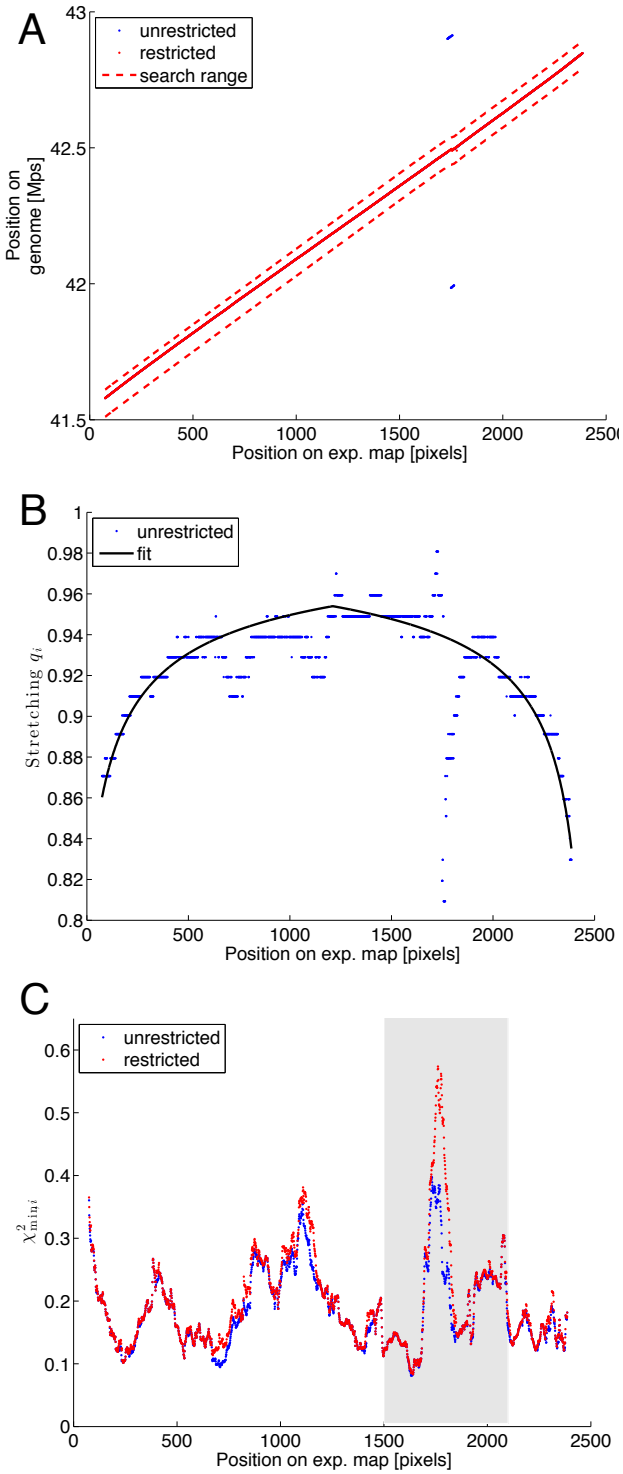


Fig. S7. Sliding Window method applied on experimental data with a probe length of $L_{\mathrm{bc}} = 150$ pixels∼80 kb for a molecule localized on chromosome 7 around 42–43 Mb. (A) shows the pairs (i, b_i) of the location b_i for the minimum of the figure-of-merit for position i of the probe. The blue/red line is for an 'unrestricted'/'restricted' search; see the text. The search range for the 'restricted' search is marked by the dashed red lines. (B) The blue dots are the optimal stretchings q_i for the positions i of the probe for the 'unrestricted' search. The solid black line is the fitted stretching q_i^{fit}. (C) Minimal values $\chi^2_{\mathrm{min},i}$ for the 'unrestricted' (blue) and 'restricted' (red) searches. Several distinct peaks are visible. The large one in the grey area is used as example in Fig. 3B in the main text.

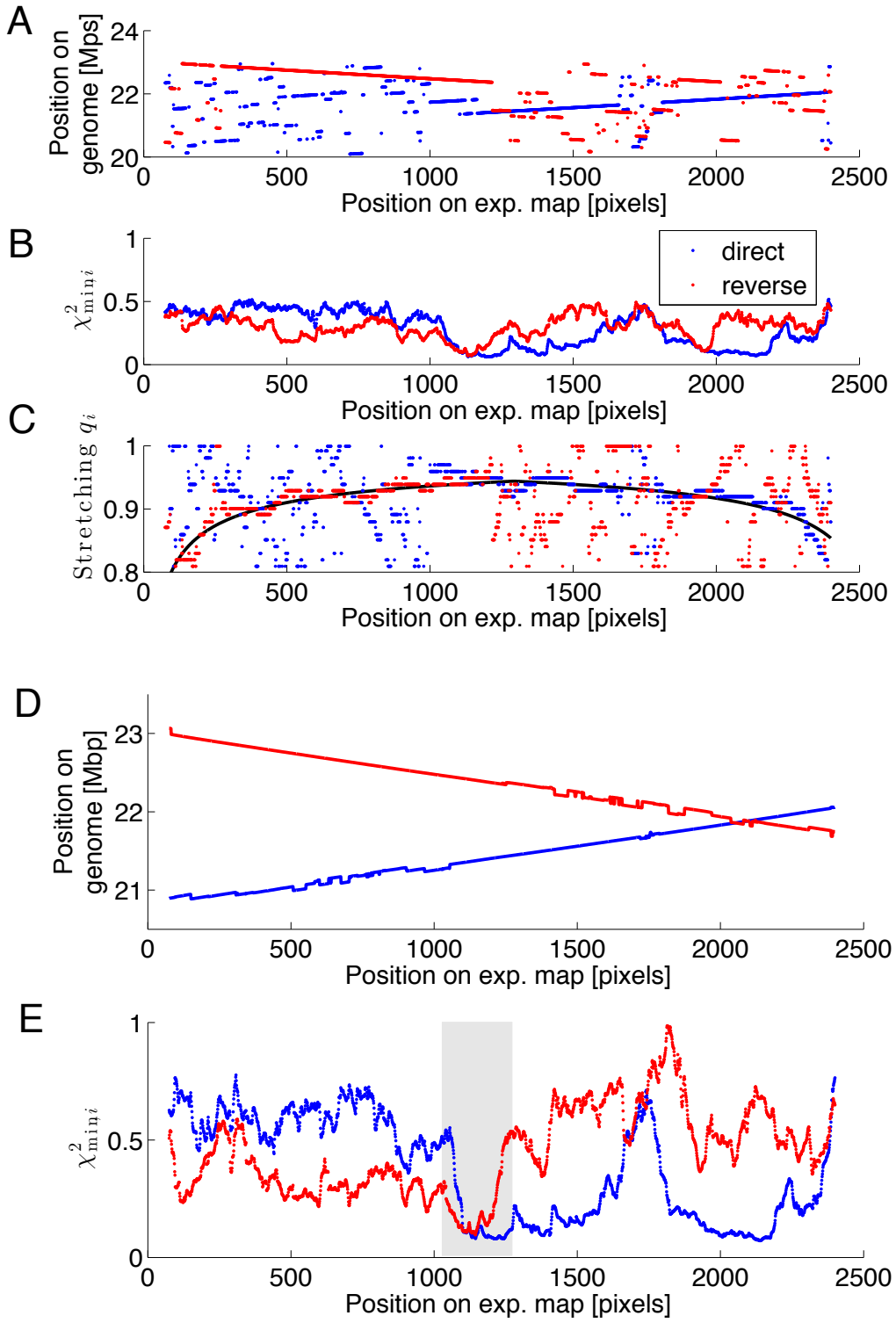


Fig. S8. Sliding Window method applied to a molecule with large-scale inversion (I): The molecule was localized on chromosome 16 from 21 to 23 Mb, and an 'unrestricted' Sliding Window calculation (A-C) detected a large inversion (≥ 0.5 Mb). In all panels, the blue points are for the 'direct' search direction and the red are for 'reverse'. (A) The position on the theoretical melting pattern shows a breakpoint in the region around pixel 1025 to 1275. (B) Values for $\chi^2_{\mathrm{min},i}$ for the two search directions, showing a significant increase/decrease when crossing the breakpoint. (C) Corresponding optimal stretchings q_i with a best fit, q_i^{fit} (solid black line). (D-E) A 'restricted' Sliding Window calculating with the fitted tension q_i^{fit} from the data in (C). (D) Positions in the genome for the two different search directions, and (E) the corresponding values for $\chi^2_{\mathrm{min},i}$. At the region between pixel 1025 to 1275, the experimental DR map fit the reference genome at two different places, as shown by the low $\chi^2_{\mathrm{min},i}$-values. This indicates an overlapping region.

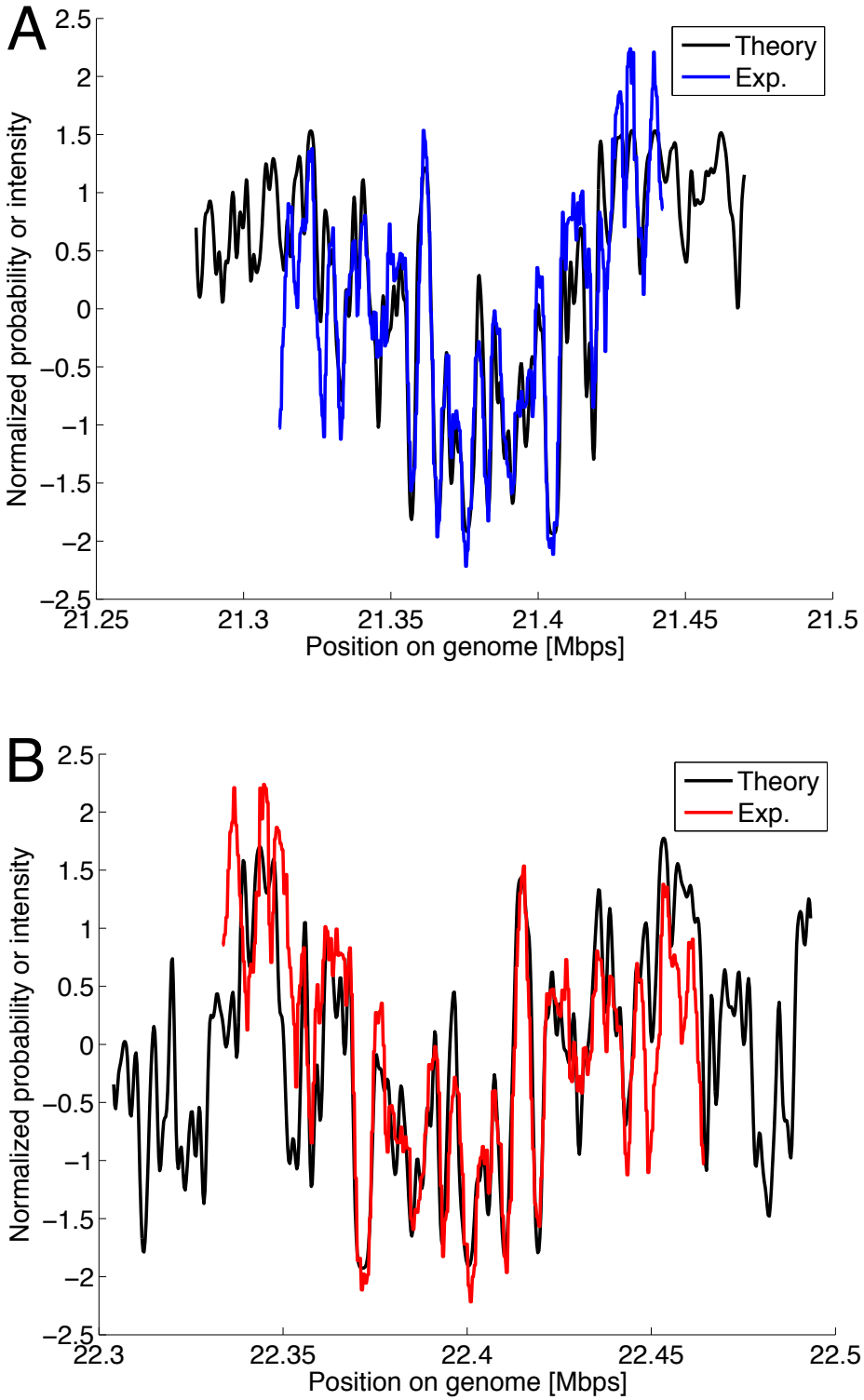


Fig. S9. Sliding Window method applied to a molecule with large scale inversion (II): same molecule as in Fig. S8. The experimental and theoretical melting patterns for the overlapping region between pixel 1025 to 1275, corresponding to 21.31-21.45 and 22.33-22.47 Mb for the 'direct'(A) and 'reverse' (B) search direction, respectively.

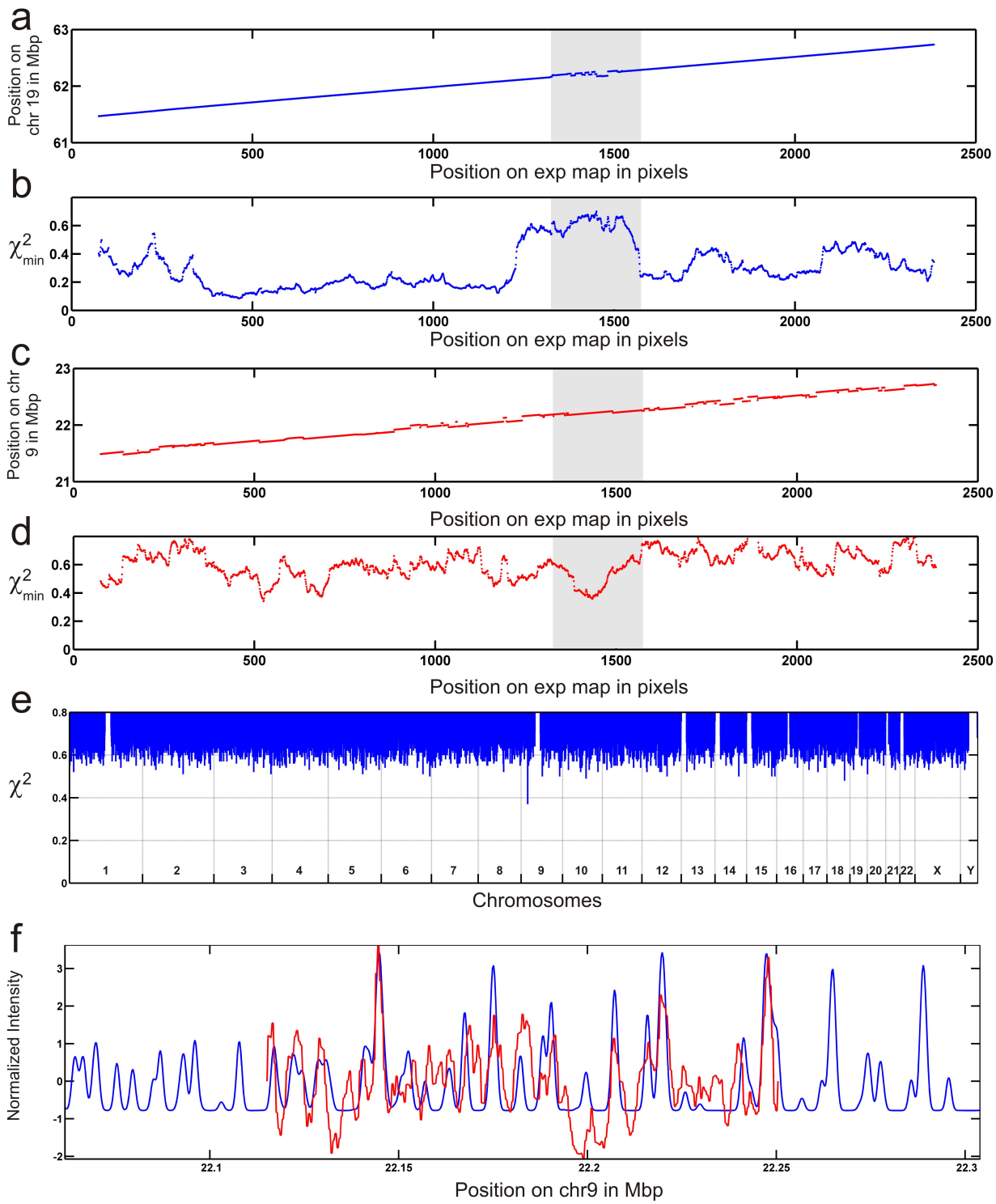


Fig. S10. Sliding Window method applied to a molecule with large-scale translocation: The 'restricted' Sliding Window calculation (probe length $L_{\mathrm{bc}} = 150$ pixels∼80 kb) for the position and $\chi^2_{\mathrm{min},i}$-values on chromosome 19 (A-B), and on chromosome 9 (C-D)]. The grey area from pixel 1325 to 1575 (∼132 kb) marks the translocation, i.e., where the match on chromosome 19 is poor, and a match on chromosome 9 is found. (E) Whole genome search for the translocation. The minimum $\chi^2_{\mathrm{min},i}$-value is 0.37 and the barcode is centered at 22.183 Mb on chromosome 9. The p-value for a single match is $5 \cdot 10^{-8}$, and $P = 0.04$ for a search over the whole genome (correlation length $\lambda = 3.7$ kb). Finally, (F) shows the experimental and theoretical DR profiles at the match.

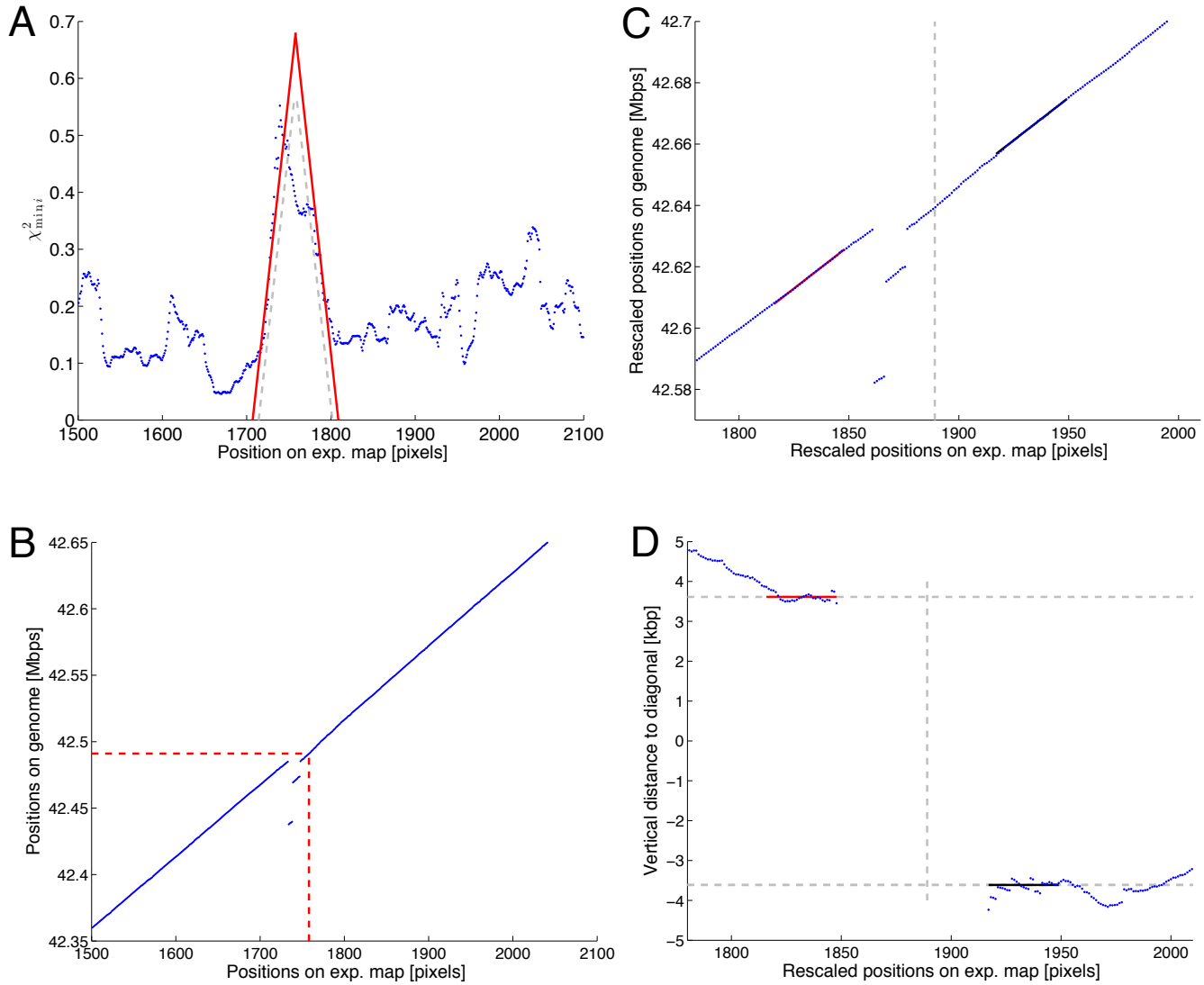


Fig. S11. Experimental demonstration of the Sliding Window method using the same experimental DR map as in Fig. 2B in the main text and in Fig. S7. A probe with length $L_{\mathrm{bc}} = 75$ pixels ($\sim$13.5 μm $\sim$40 kb) is 'slid' over the genome. (A) Minimum value of the χ^2-function recorded for each pixel. (B) corresponding position on the genome. In (A) lines with slopes $\pm 1/L_{\mathrm{bc}}$ are placed on top of the peak in the data as described in the text. The peak indicates a structural variations. (C) Fits to the diagonal after correction the experimental and theoretical melting patterns for global and local stretching, respectively. (D) Distance between the averages of the lines fitted on both sides of the peak (in regions corresponding to the black and red line) and the position in the genome. The gap shows that the variation is a $\sim$7 kb insertion.

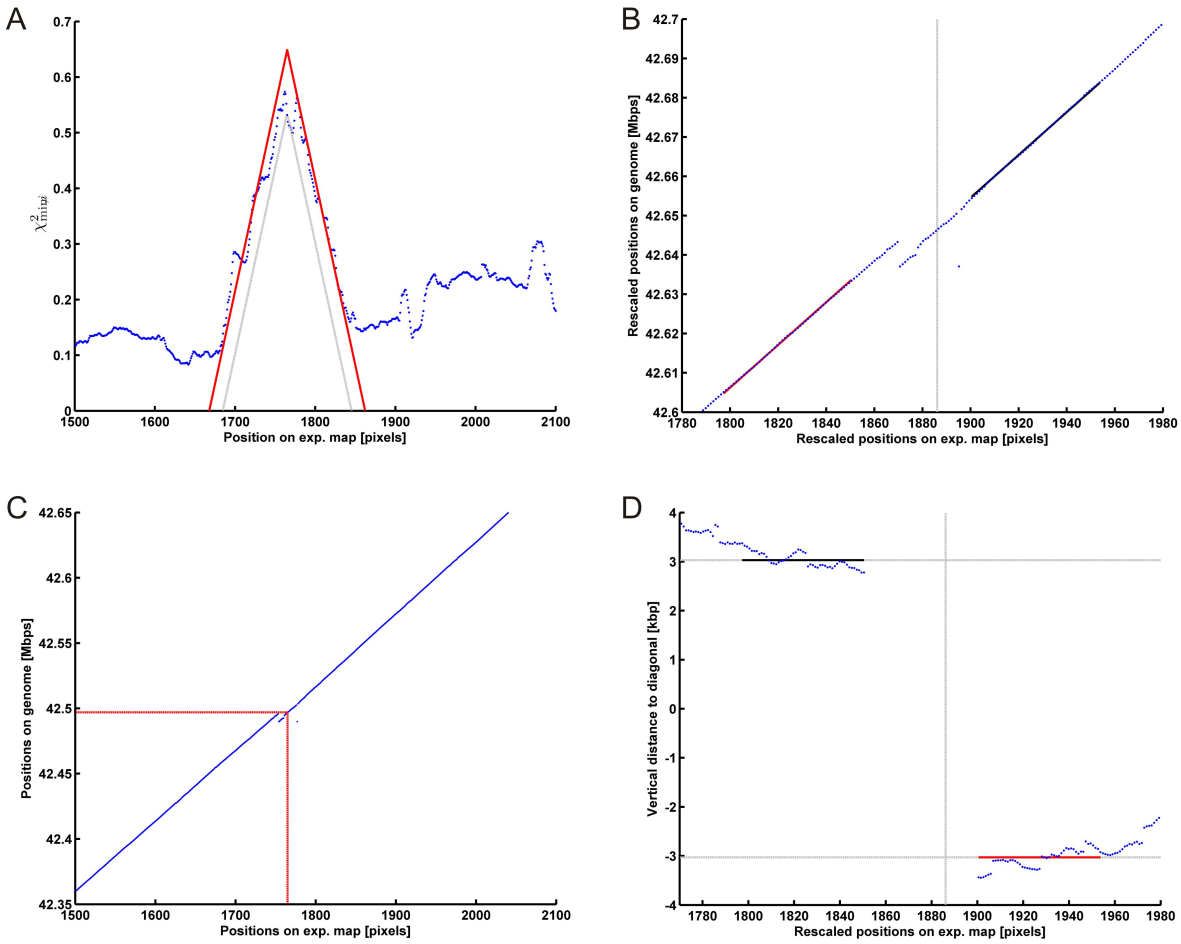


Fig. S12. Experimental demonstration of the Sliding Window method for the same as experimental data as in Fig. S11 but for a probe length $L_{\mathrm{bc}} = 150$ pixels ($\sim$27 μm $\sim$80 kb). The contents of the panels corresponds to the similar panels in Fig. S11. The gap shows that the variation is a $\sim$6 kb insertion.

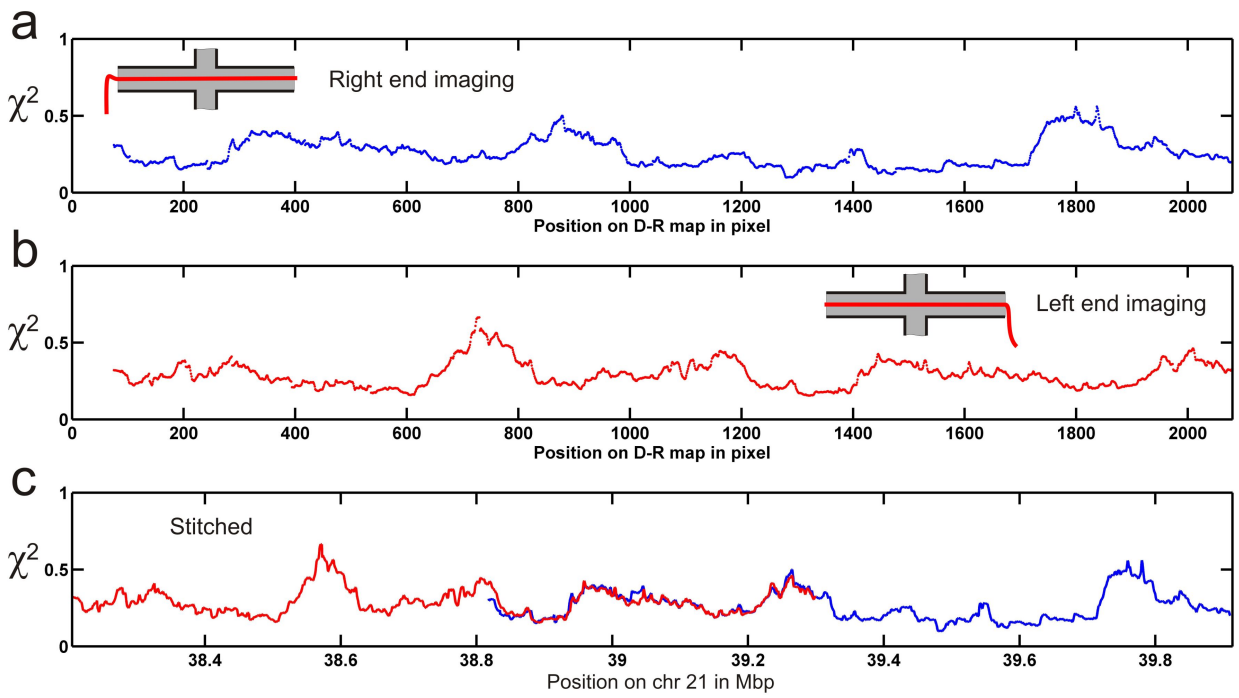
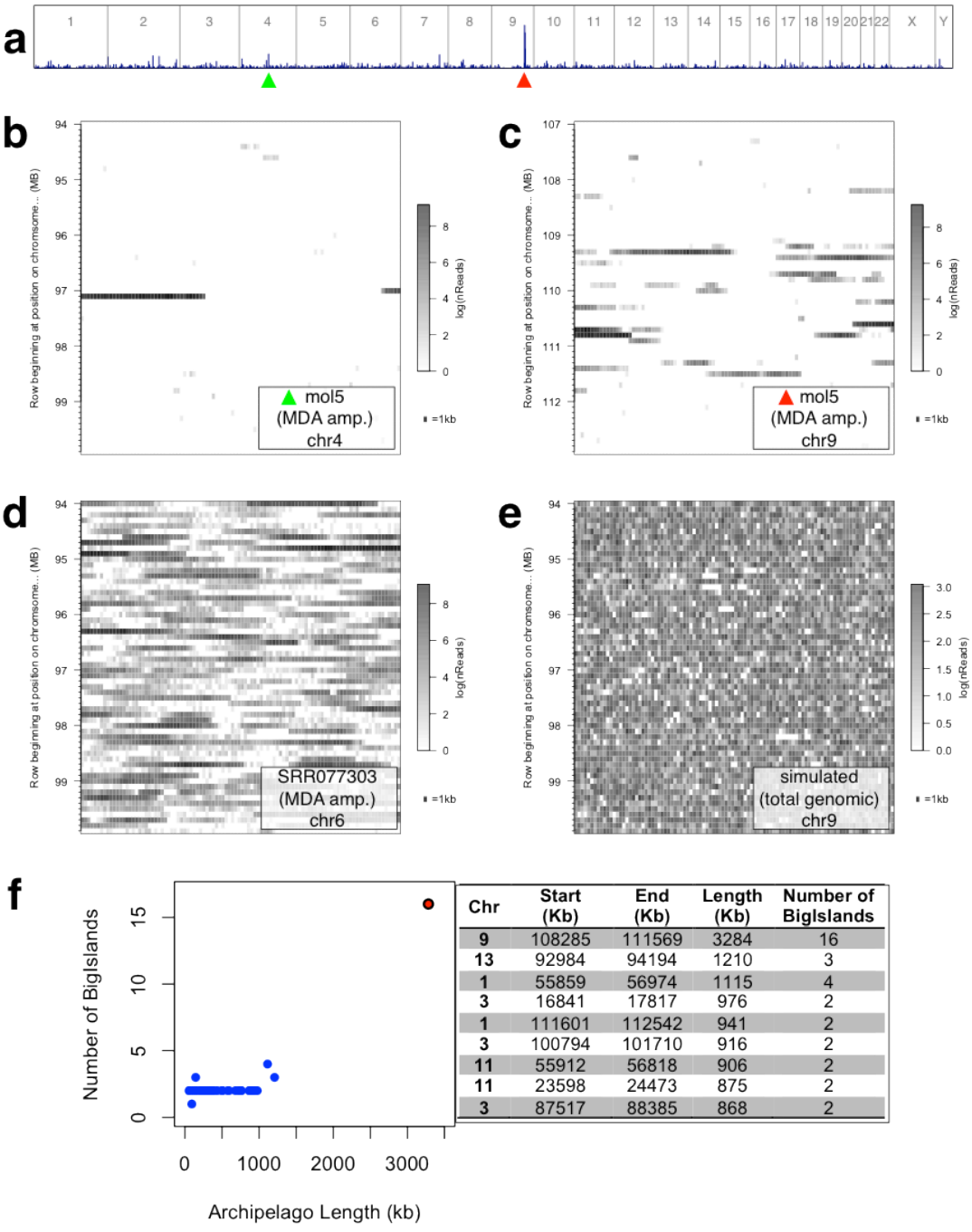


Fig. S13. Reproducibility of the Sliding Window method. An experimental DR map—later found to match on chromosome 21 near 40 Mb—was imaged over the nanoslit, then moved halfway across the nanoslit and imaged again. Thus two distinct, but overlapping, experimental DR maps are obtained. For each map, a 'restricted' Sliding Window calculation was performed; the corresponding $\chi^2_{\min,i}$-values are shown in (A-B). From the relationship between positions in the genome and the pixel numbers, a plot of the position in the genome versus $\chi^2_{\min,i}$ can be made for each of the two experimental DR maps. The result is found in (C) and shows almost perfect agreement without any fitting in the overlapping region (∼38.8–39.3 Mb) between the two DR maps. This is very important for the use of the Sliding Window method for detecting small structural variations.



Chr	Start (Kb)	End (Kb)	Length (Kb)	Number of BigIslands
9	108285	111569	3284	16
13	92984	94194	1210	3
1	55859	56974	1115	4
3	16841	17817	976	2
1	111601	112542	941	2
3	100794	101710	916	2
11	55912	56818	906	2
11	23598	24473	875	2
3	87517	88385	868	2

Fig. S14. Sequencing of a Rescued Single DNA Molecule. After mapping sequencing reads to the hg19 reference genome, we constructed a "footprint" map, **a**, to visualise regions of the genome which had a high degree of coverage by sequencing reads. A single peak was observed on chromosome 9 centred at 110 Mb (red ▲); other minor peaks (e.g. green ▲) were also observed. We then constructed read density maps, **b**-**e**, to "zoom-in" on regions of interest: each 1 kb of the reference genome was represented as a single block increasing in ascending order from left to right, with the number of reads mapping to each block represented by its shading. Each row contains 100 blocks, representing 100 kb of the reference genome beginning at position indicated on the vertical axis. Reads putatively from the rescued DNA molecule, **c**, clustered together in islands, which uniquely formed an archipelago on chromosome 9 around 110 Mb. This archipelago pattern was not present anywhere else in the genome, for example, **b**, on chromosome 4 at 97 Mb. Moreover, this archipelago effect is characteristic of single molecule templates amplified by multiple displacement amplification, such as reads from single amplified chromosomes [23], **d**, and is not due to varying mappability of the human genome, as we do not observe this pattern by mapping simulated reads, **e**, from total genomic DNA. We were able to automate detection of archipelagos, **f**, and observed a single, large archipelago (red •) on chromosome 9 from 108.3 to 111.6 Mb—the genomic origin of our rescued DNA molecule, as confirmed by FISH.

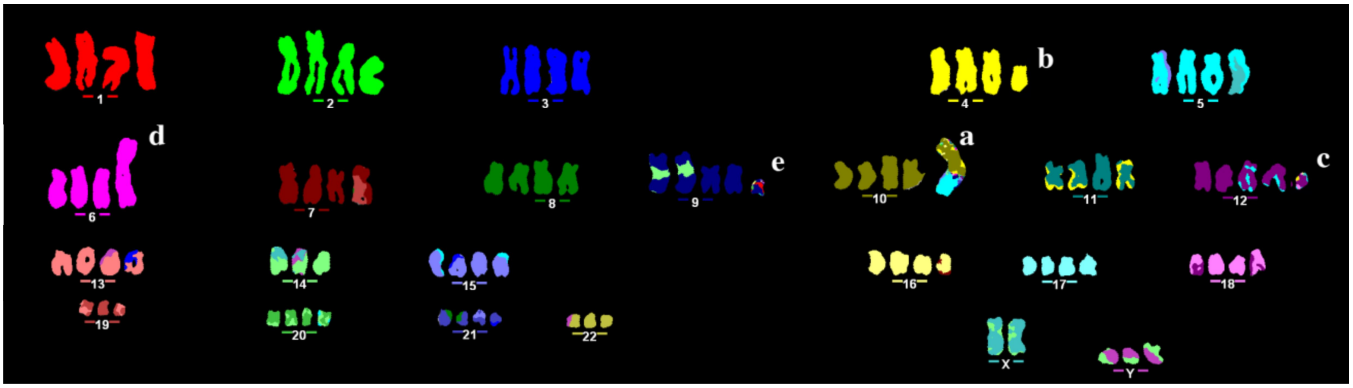


Fig. S15. Multicolor-FISH (mFISH) of the Jurkat cell line shows mosaicism comprising cells with different chromosomal number and chromosomal alterations on a tetraploid background: 42.3% of cells are 4n with numerical aneuploidies; 23% are 4n + derivative 10/5 (a); 3.85% are 4n +10/5(a)+derivative 4(b); 3.85% are 4n + derivative 10/5(a)+ derivative 12(c); 19.2% are 4n+ derivative 6(d); 7.7% are 4n+ derivative 9(e).