PNAS
www.pnas.org

# Supplementary Information for

## Single-Molecule DNA-Mapping and Whole-Genome Sequencing of Individual Cells

**Rodolphe Marie, Jonas N. Pedersen, Loic Bærlocher, Kamila Koprowska, Marie Pødenphant, Céline Sabatel, Maksim Zalkovskij, Andrej Mironov, Brian Bilenberg, Neil Ashley, Henrik Flyvbjerg, Walter Bodmer, Anders Kristensen and Kalim U. Mir**

**Rodolphe Marie.**
**E-mail: rodolphe.marie@nanotech.dtu.dk**

**This PDF file includes:**

## Supporting Information Text

### 1. Fragment length distribution from fragmentation protocol

Optical mapping of mega-base long genomic DNA molecules extracted from a single cell on a chip is enabled by a controlled fragmentation of the genomic DNA. The YOYO-1 stained DNA from the cell's nucleus is kept in the cell trap in a buffer that does not contain the anti-nicking agent $\beta$-Mercaptoethanol (BME). Random fragmentation is obtained by illuminating the DNA for a fixed time interval at constant intensity to induce photo-nicking. To characterise the fragment length distribution as a function of the illumination time, we perform a separate study where DNA fragments are stretched by confinement in nanochannels. First, we calibrate the extension along the nanochannel versus contour length with a molecule of known contour length ($\lambda$-phage DNA), and then we measure the extension of DNA fragments extracted from a single cell.

**A. Stretching calibration.** We use a device where DNA extracted from a single cell and fragmented can be sized in nanochannels (**Fig. S7**). The device and its operation (**Fig. S7c-f**) are similar to the device discussed in the main text, except that the flow-stretch area is replaced by an array of 127x154 $(\mathrm{nm})^2$ nanochannels. With the nanochannel array it is possible to measure the extension of DNA fragments down to 22 kb (**Fig. S7b**).

We calibrate the stretching of dsDNA in the nanochannels as follows: $\lambda$-phage DNA (48 kb) was diluted in 0.5xTBE + 1% v/v BME + 0.5% triton-X100 at 1 $\mu$g/mL. Molecules were introduced in the nanochannels via electrophoresis with a voltage of 5-10 V, and then imaged for 5 seconds at 20 frame per seconds. A box-car fit of the intensity profile in each frame gives the extension of the DNA molecule along the nanochannel $L(t)$ at time $t$, and the equilibrium extension of the molecule is $L = \langle L(t) \rangle$ (1). The average extension of 28 molecules is $(9.3 \pm 0.4)\mu$m.

As the extension of DNA in nanochannels scales linearly with the contour length (1), we get from the calibration experiment with $\lambda$-phage DNA that the ratio between the contour length $l_c$ and the extension in the nanochannel $L$ is $\frac{l_c}{L} = \frac{48\,\mathrm{kb}}{(9.3\pm0.4)\,\mu\mathrm{m}} = (5.2 \pm 0.2)\,\frac{\mathrm{kb}}{\mu\mathrm{m}}$. So if the extension of a molecule in the nanochannel is $L'$, we estimates its contour length $l'_c$ as $l'_c = \left[(5.2 \pm 0.2)\,\frac{\mathrm{kb}}{\mu\mathrm{m}}\right] L'$.

**B. Nicking probability.** DNA molecules from a single cell are illuminated for a fixed time interval $\Delta t_{\mathrm{illu}}$ to induce photonicking. Assume that the nicking probability per base pair $p_{\mathrm{nick}}$ is constant over the entire genome, and the nicking of base pairs is a Bernoulli process. The fragment lengths then follow a geometric distribution $P_k = p_{\mathrm{nick}}(1-p_{\mathrm{nick}})^k = p_{\mathrm{nick}} \exp(k \ln[1-p_{\mathrm{nick}}])$, where $k$ is the number of un-nicked base pairs before the next nick. For $p_{\mathrm{nick}} \ll 1$, it holds that $k \ln[1-p_{\mathrm{nick}}] \approx -k p_{\mathrm{nick}}$, and thus $P_k \approx p_{\mathrm{nick}} \exp(-k p_{\mathrm{nick}}) = \exp(-k/\lambda)/\lambda$ where $\lambda = 1/p_{\mathrm{nick}}$ is the mean of the exponential distribution. If distances are measured in base pairs and $a$ is the length of a base pair, then $x = ka$ is the length of a fragment.

**C. Truncated exponential distribution.** Only fragments longer than a cut-off $K$ can be imaged and sized in nanochannels. So the length distribution of the DNA fragments measured in the nanochannels becomes a truncated exponential distribution

$$p_{\mathrm{trunc}}(x) = \begin{cases} 0 & \text{if } x < K, \\ \frac{1}{\lambda} e^{-(x-K)/\lambda} & \text{if } x \geq K, \end{cases} \quad [1]$$

where $x$ is the length of a fragment and $\lambda$ is the mean of the distribution.

If we measure the lengths of $n$ DNA molecules $[x_i]_{i=1}^n$, the maximum likelihood estimator for the parameter $\lambda$ is

$$\hat{\lambda} = \frac{1}{n}\sum_{i=1}^{n} x_i - K = \bar{x} - K \ , \quad [2]$$

and from the Fisher Information we get the estimated standard deviation of $\hat{\lambda}$

$$\sigma_{\hat{\lambda}} = \frac{\bar{x} - K}{\sqrt{n}} \ . \quad [3]$$

**D. Illumination time versus mean fragment size.** DNA extracted from four individual cells was illuminated with the same light intensity, but for three different time intervals $\Delta t_{\mathrm{illu}} = 150\,\mathrm{s}, 210\,\mathrm{s}$, and $300\,\mathrm{s}$. Histograms of the fragment lengths are shown in **Fig. S8**. Red curves are the truncated exponential distributions from Eq. (1) with the estimated parameter $\hat{\lambda}$ from Eq. (2) and a common cut-off $K = 22\,\mathrm{kb}$. The estimated values for $\lambda$, their uncertainties, and the $P$-values for the fits are shown in **Supplementary Table S2**. **Fig. S9** shows the nicking probability versus illumination time. Data are consistent with a straight-line through the origin (red curve, $P$-value equal to 0.5).

**E. Fraction of the genome on DNA fragments longer than $L_{\mathrm{min}}$.** A limitation of the flow-stretch device with dimensions used in our study (*i.e.* the total length of the nanoslit) is that fragments shorter than $L_{\mathrm{min}} = 1\,\mathrm{Mb}$ cannot be captured and imaged accurately. With the 150 s illumination time used in experiments, these fragments make up a large fraction of the whole genome as demonstrated below.

If we assume that DNA from the whole genome is fragmented so fragment sizes follow an exponential distribution with a mean $\lambda$, then the fraction $Q_>(L_{\min})$ of DNA with a length above the length $L_{\min}$ is *

$$Q_>(L) = \frac{\int_{L_{\min}}^{\infty} dx\, x\, p(x)}{\int_0^{\infty} dx\, x\, p(x)} = e^{-L_{\min}/\lambda}\left(1 + \frac{L_{\min}}{\lambda}\right) \quad . \qquad [4]$$

**Fig. S10** shows $Q_>(L_{\min})$ for $\lambda$-values corresponding to the illumination times used in our experiments. Less than 1% of the genome is on DNA fragments longer than 1 Mb for $\Delta t_{\text{illu}} = 150$ s ($\lambda = 130$ kb).

Similarly, we can consider the case of a nanochannel device where DNA molecules as short as 22 kb are imaged. The fraction of DNA with a length below the cut-off $L_{\min}$ is denoted $Q_<(L_{\min})$:

$$Q_<(L_{\min}) = 1 - Q_>(L_{\min}) \quad . \qquad [5]$$

**Fig. S11** shows $Q_<(L_{\min})$ for the illumination times used in our experiments. The red curve is a plot of the expression in Eq. (5) with $\lambda$ from the straight-line fit of the nicking probability (see **Fig. S9**). Even with a cut-off at $L_{\min} = 22$ kb, potentially up to 99% of the DNA in a cell can be imaged in nanochannels when the DNA is fragmented using the nicking protocol outline above.

## 2. D-R mapping

We here describe how D-R maps are aligned to a reference genome, and we also briefly discuss the 'Sliding Window' analysis. Finally we comment on the number of unmapped molecules to expect in an experiment.

**A. Mapping to the reference genome.** The method for mapping a consensus map of D-R maps (i.e. an average over several single-molecule maps) to a reference genome was developed in (2). In Ref. (3, 4) the same method was used for aligning D-R maps from single molecules from human DNA to a reference genome. In Ref. (3) further methods were applied for detection of, e.g., structural variations (see in particular the Supplementary Information to Ref. (3)). For completeness, we here sketch the methods.

When aligning a D-R map to the *in silico* melting profile of a reference genome, we start from a segment, e.g., 150-200 pixels, of the experimentally obtained D-R map. This corresponds to 75-100 kb of DNA if the DNA is fully stretched †. This segment is compared with a theoretical D-R map computed from the entire human reference genome (GRCh37) using the software Bubblyhelix (www.bubblyhelix.org) (5). We calculate a dimensionless match score, $\chi^2$, for each position in the genome the segment is compared to (see, e.g., Fig. 4 in the Supplementary Information to Ref. (3)). The segment is said to match at a specific position in the genome if the match score is statistically significant, i.e., if it is improbable that the score occurred by chance (for a detailed description of statistical analysis, see the Supplementary Information to Ref. (3)). Table S1 shows an overview of the localized molecules and their positions in the human genome.

Each match was then checked with the 'Sliding Window'-method (3). Here the whole D-R map from a molecule is divided into overlapping segments ('windows'), e.g., shifted by a single pixel. Each segment is mapped to its location in the genome (see, e.g., Fig. 6 in the Supplementary Information to Ref. (3)). If the full experimental D-R map perfectly matches the *in silico* D-R map computed from the human reference genome, the matches found for the individual segments from the experimental D-R map fall on a straight line. The slope equals the number of base pairs per pixel (~0.5 kb) when the position of the segment on the reference genome is plotted against the position of the segment on the D-R map (see Figs. S12-S20). The sign of the slope is positive or negative, depending on the orientation of the D-R map relative to the reference genome. In the alignment procedure, we also take into account that the DNA is not uniformly stretched in the nanoslit (3).

Figs. S12-S20 show a 'Sliding Window' analysis for all localized molecules in the present study. The search for the match position of each segment is here not over the whole reference genome but only over a 2 Mb interval containing the match due to the computational costs. This method is denoted an 'unrestricted' search, and the results are shown with blue dots in Figs. S12-S20. The upper panels show the positions on the reference genome (GRCh37) versus the position on the D-R map. Straight lines clearly appear in all plots of the positions, although some of the lines are broken (especially for Molecules 5, 8, 12, 16, and 17). This could be due to experimental errors or mismatches between the experimental D-R maps and the *in silico* map. They are thus candidates for further investigation of structural variations. The thin black lines are guidelines that indicate the positions of the diagonals. The middle panels shows the match scores ($\chi^2$-values) for the best match at each position. Here $\chi^2 = 0$ corresponds to a perfect match, and a maximal mismatch would correspond to $\chi^2 = 2$ (see (3) for details). The lower panels show the degrees of stretching for the best match for each position of the window.

* This is not to be confused with the probability to draw a random fragment with a length above the cut-off. This probability is $P_> = \int_{L_{\min}}^{\infty} dx\, p(x) = 1 - e^{-L_{\min}/\lambda}$. From our fitted $\lambda$-values this probability is between 15% and 30%.

† In the flow-stretch device used in the experiment, the degree of stretching varies along the molecules, e.g. from 85-98%, see, e.g., Fig. 7b in the Supplementary Information to Ref. (3).

**B. Structural variation detection by D-R mapping.** The 'Sliding Window'-method can also be used for detection of structural variations, e.g. deletions, insertions, translocations or inversions (see Ref. (3), in particular Fig. 2). Here the search interval is restricted to a 50 kb range on each side of the diagonal (marked by dashed, red lines in Figs. S12-S20), and the degree of stretching is for each position of the window on the D-R map fixed at the value marked with a red dot in the lower panels in Figs. S12-S20. This is denoted a 'restricted' search. The corresponding $\chi^2$-values are marked with red dots in the middle panels.

In the present study, we identified several molecules with discontinuities in the plots of the positions on the reference genome versus the positions on the D-R maps. The 'Sliding Window' method can detect structural variations (insertions/deletions) down to a few kb, but only if the distance between the SVs is greater than the length of the Sliding Window (here ∼75 kb) and the D-R map and the reference genome match well (i.e. low $\chi^2$-values) between the SVs. So in order to pinpoint the nature of a SV, we need a discontinuity of the position plot to coincide with a peak of the $\chi^2$ plot in a region of good agreement with the reference genome. Of the molecules that were candidates for SV detection, a few fulfilled this prerequisite. We thus find a complex region in molecule 17 (see Fig. 4A in main text), two missing segments in molecule 12 (Fig. S23), as well as a 20 kb missing segment on a molecule from a different cell (see Fig. 5A in main text).

Finally, if there are several structural variations within a window, the shift of the lines in plots like Fig. 5A in the main text will correspond to the sum of the insertions and deletions. Thus the peak in the $\chi^2$-values will become broader than for an isolated insertion or deletion. For a detailed discussion, see the SI Appendix to [(3)].

**C. Unmapped molecules.** Four of the 21 molecules could not be aligned to the reference genome. We here argue that two of the molecules had a significantly different D-R pattern than the rest. Of the remaining 19 imaged molecules, two molecules could not be mapped to the human reference genome although their D-R maps did not differ significantly from the rest. We show that this number is consistent with 6.9% of the reference genome consisting of gaps larger than the lengths of the D-R maps.

First, Figure S21 shows D-R maps of the two molecules with a significantly different D-R pattern than the other molecules and which could not be aligned to the reference genome. For comparison, we also show a D-R map which could be aligned. The two unmapped molecules have significantly less features in the intensity signal, less contrast, and a higher average fluorescence. We suspect that this is due to imperfect melting of the DNA during the heat cycle. Whatever its cause, the lack of features has as a consequence that these molecules cannot be mapped to the reference genome. These molecules could potentially be mapped if the melting maps are taken at multiple temperatures.

The remaining 19 D-R maps are aligned to the human reference genome (GRCh37/hg19). The size of this reference genome is 3.137 Gb [‡]. A D-R map might be aligned to the melting profile of the reference genome even if it contain small gaps, but we assume that alignment is not possible for gaps comparable to the size of the D-R map itself, i.e. approximately 1 Mb. The size distribution of the gaps in the human genome GRCh37 are shown in Fig. S27 [§]. There are 357 gaps which cover 234.3 Mb. Of these are 35 gaps larger than 1 Mb, and they amount to 215.4 Mb, i.e. a fraction $p = 215.4 \cdot 10^6/3.137 \cdot 10^9 = 0.069 = 6.9\%$ of the reference genome.

If $n = 19$ is the total number of molecules we try to map and $p = 0.069$ is the probability that a molecule cannot be mapped to the reference genome due to larger gaps, then the probability to get exactly $k$ molecules we cannot map follows a binomial distribution $B(k|n,p) = \binom{n}{k} p^k (1-p)^{n-k}$, where $0 \leq k \leq n$ and $\binom{n}{k} = n!/[k!(n-k)!]$. Thus the probability to find $k$ or more molecules which cannot be mapped is the sum over $B(k'|n,p)$ for $k' \geq k$, i.e. $\sum_{k'=k}^{n} B(k'|n,p)$. For $n = 19$, $k = 2$ and $p = 0.069$, the summed probabilities equal 0.38. This is larger than the 5%-significance level (one-tailed test).

Other factors might influence the number of unmapped molecules, e.g. structural and copy number variations between the reference genome and actual molecules or experimental artefacts (e.g. due to the molecule drifting during imaging), nevertheless the statistics suggest that gaps in the reference genome is a probable cause for the observed number of molecules that produce high quality experimental D-R maps that cannot be mapped to the reference genome.

**D. Comparison of D-R maps and the reference genome.** Figures S28-S30 show plots similar to Fig. 3E in the main text. For each molecule from SI Appendix Table S1 a section of the D-R map from pixel 900 to pixel 1100 was selected and matched to the reference genome. The degree of stretching is equal to the one used for the 'restricted Sliding Window' method at pixel 1000 in Figs. S12-S20.

**E. Discussion of the genome's GC-content influence on D-R mapping.** The paper 'The Human Genomic Melting Map' (6) and the corresponding web-page (http://meltmap.uio.no/) provide a complete theoretical melting profile of the human genome (hg17 and hg18). Table 1 in that paper shows that although the average melting temperature is ∼70°C, it can vary between

[‡] From https://www.ncbi.nlm.nih.gov/assembly/GCF_000001405.13/ [accessed: 30-April-2018].

[§] Retrieved from http://hgdownload.cse.ucsc.edu/goldenPath/hg19/database/ [accessed: 30-April-2018].

∼50°C and ∼88°C for 1 kb intervals. And the standard deviations for such intervals are up to 11°C.

For our D-R maps, we use the melting properties of Mb-long DNA segments. So we used the data for the melting temperature for hg18 (http://meltmap.uio.no/), and calculated the mean and standard deviation of the melting temperatures for all 1 Mb intervals in steps of 10 kb. We found that the minimum average melting temperature in a 1 Mb interval can be as low as 66°C and the maximum as high as 77°C. The standard deviation is ∼3-4°C in each 1 Mb interval. Regions with extreme average melting temperatures are candidates for molecules that do not have sufficient denaturation pattern to allow for alignment to a reference profile. It is a potentially interesting research project to link the melting properties (and thus the AT/GC content) of a DNA molecule to the information content in the D-R map. The result would determine which portion of the whole genome that can be aligned using a single melting temperature in the chip.

Notice that Veal et al. (7) discuss the role of GC-content for regions of the genome that are hard to amplify and thus give poor sequencing coverage [denoted 'Thermodynamically Ultra-Fastened' (TUF) DNA]. So the problem of incomplete melting might not only be relevant for D-R mapping, but also for sequencing protocols.

## 3. Simulations of kb-size structural variations

We here align experimental D-R maps to modified version of the reference genome. The modifications are artificial structural variations (SVs) introduced by us. We do this in order to demonstrate theoretically that the D-R mapping technique is capable of detecting SVs in a DNA molecule extracted from a single cell. The artificial aspect of this demonstration makes it more objective because we know the true answer.

For the analysis, we selected four molecules (#8, #9, #10, and #13) that all align very well to the reference genome. That is, they show an almost unbroken line in plots, where the experimental D-R maps are compared to the reference genome using the 'Sliding Window' method (see Figs. S15, S16 and S18).

In each of the four molecules, we selected one position at or near the center of the experimental D-R map and recorded the corresponding position in the human reference genome (GRCh37) from Figs. S15, S16 and S18. The pairs of positions at the D-R maps and at the reference genome are shown with gray lines in Figs. S31–S40.

An *insertion* (*deletion*) is a SV consisting of an excess (deficit) of genetic material in the D-R map relatively to the reference genome. We simulate insertions (deletions) by removing (inserting) sequences from (into) the reference genome. *Mismatches* we simulate by replacing sequences in the reference genome by other sequences of identical lengths. Finally, *inversions* we simulate by inverting sequences in the reference genome. From a single-molecule D-R mapping point-of-view, a *translocation* appears as a mismatch. A translocation may also induce a gain or loss of genetic material if the translocated fragments have different length i.e. an insertion or a deletion may be detected by D-R mapping at the break point(s). We consequently do not treat translations as a separate case.

Next, the theoretical melting profiles of these modified reference genomes were calculated with the software Bubblyhelix (www.bubblyhelix.org) (5) and compared with our experimental D-R map, which aligned well with the unmodified reference genome. We used the exact same procedure as when aligning to the original reference genome (see Sec. 2, 'D-R mapping').

The analysis done to detect SVs is the same as the one done to detect SVs in, e.g., Molecule #12 (see Sec. 2B, 'Structural variation detection by D-R mapping' and Fig. S23), So here we only show the results of the 'restricted Sliding Window' method. The window size is 150 pixels, i.e. ∼75 kb.

We compared each of the four molecules to twenty modified versions of the reference genome (four different sizes of insertions, four different sizes of deletions with mouse DNA inserted in the reference genome, four different sizes of deletions with lambda phage DNA inserted in the reference genome, four different sizes of mismatches with mouse DNA replacing sequences in the reference genome, and four different sizes of inversions in the reference genome). The sizes of the SVs are in the range from 1 to 30 kb.

Thus, in total we tested 80 different artificial SVs. Larger SVs could have been included in the analysis. However, all SVs equal to or larger than 30 kb are easily detected, so we chose not to consider them here. In order to facilitate the graphical presentation of our results, the structural variations are placed at only four different positions in the genome, one position per molecule.

Based on our simulations discussed below, we conclude that with D-R mapping it is possible to detect structural variations down to 5 kb in size and to locate them to within ∼10 kb from their true positions in single molecules extracted from single cells when the structural variations are deletions or insertions. Structural variations without a gain or loss in genetic material (mismatches and inversions) cause a signal in the $\chi^2$-value and can in this manner be detected if larger than 10 kb and 30 kb, respectively, but the type of SV (mismatch or inversion) cannot be determined.

**A. Simulations of the detection of kb-size insertions.** We simulated the detection of insertions in an experimental D-R map relative to the reference genome by removing 1, 5, 10 or 30 kb from the reference genome at the selected positions. The experimental D-R maps were aligned to these modified reference genomes. Figures S31–S32 show the results of the 'restricted Sliding Window' analysis for all four molecules.

As an example, consider Fig. S31, which shows the results for Molecule #8. The top panels show the result of the 'restricted Sliding Window' analysis, where the D-R map is aligned to the original, unmodified genome ('0 kbp') and to modified versions

of the reference genome. As in Fig. S23, dashed red lines correspond to the search range for the 'restricted Sliding Window' analysis (see Sec. 2A, 'Mapping to the reference genome'). If an insertion is present in the experimental D-R map, a shift of the data relatively to the unmodified case ('0 kbp') is clearly visible. That is, some positions on the experimental D-R map are *not* assigned to any position on the reference genome. This is even clearer in the zoomed-in versions in the right panel. The shift corresponds exactly to the size of the sequence removed from the reference genome, as indicated with a dashed line (same color code as for the other data sets).

The lower panels show the corresponding $\chi^2$-values. A slight increase in the $\chi^2$-values is present even for a 1 kb insertion, while insertions larger than or equal to 5 kb cause a peak structure. The positions of the insertions are best determined from the peak positions. These are within ∼20 pixels, i.e. ∼10 kb, from the true positions of the insertions.

For all four molecules, the results for the simulation of 1 kb insertions are hard to distinguish from the results for the original reference genome, but insertions larger than 5 kb are easily detected from the shift of the lines. The positions of the SVs can be located within ∼10 kb.

**B. Simulations of the detection of kb-size deletions.** We simulated the detection of deletions in the experimental D-R map relative to the reference genome by adding sequences in the reference genome. So we cut out sequences of 1, 5, 10 or 30 kb from the same chromosomes and positions in the mouse reference genome (GRCm38/mm10) and inserted them in the human reference genome. The analysis used to detect these SVs is identical to the one used to detect insertions and presented above.

Figures S33–S34 show the results of the 'restricted Sliding Window' analysis. The results are similar to those for insertions, i.e., there is a shift in the alignment plots and an increase in the $\chi^2$-values near the deletion. Here the shift is due to some positions on the reference genome not being covered by positions on the experiment D-R map.

We also created more radically modified reference genomes; specifically, by inserting enterobacteria phage lambda DNA (GenBank: J02459.1); see Figs. S35–S36. The results are similar to those for artificial deletions simulated with sequences from the mouse genome.

**C. Simulations of the detection of kb-size mismatches.** Here we introduce mismatches in the genome without changing the amount of genetic material. Sequences of 1, 5, 10 or 30 kb from the same chromosomes and positions in the mouse reference genome (GRCm38/mm10) now replace sequences in the human reference genome. Figures S37–S38 show the results of the 'restricted Sliding Window' analysis. No shifts in the alignment of the D-R map to the modified reference genome are observed. This is expected as, contrary to deletions or insertions, there is no loss or gain of genetic material. The poorer agreement between the experimental D-R map and the modified genome at the position of the artificial SV is, however, clearly seen in the plots of the $\chi^2$-values.

The amplitude of the peak in the $\chi^2$-plot induced by a given size mismatch varies from molecule to molecule because the change in $\chi^2$-value depends strongly on how different the melting profile of the mismatch sequence is from the original reference sequence. So the detection sensitivity depends on the actual sequence of the mismatch. However, for a 10 kb-mismatch, a clear peak is detected for all molecules.

**D. Simulations of the detection of kb-size inversions.** In Ref. (3) it was demonstrated that a large-scale ($> 0.5$ Mb) inversion could be detected with D-R mapping. Here we simulate kb-size inversions by inverting sequences of 1, 5, 10 or 30 kb in the human reference genome.

Figures S39–S40 show the results of the 'restricted Sliding Window' analysis. Like for the mismatches, no shifts in the alignment of the D-R map to the modified reference genome are observed as the amount of genetic material is conserved. Clear peaks in the plots of the $\chi^2$-values are only observed for the 30 kb inversion.

As for mismatch, the amplitude of the peak in the $\chi^2$-plot induced by a given size of inversion varies from molecule to molecule, and for the same reason. We also note that an inversion can go undetected if the D-R pattern of the inverted sequence is symmetric. As for the mismatch case, the detection sensitivity depends on the actual sequence of the inversion. However, for a 30 kb-inversion, a clear peak in the $\chi^2$-values is detected for all molecules.

**Table S1. Positions of the molecules localised on the human reference genome GRCh37.**

| Molecule | Chromosome | Start [Mb] | End [Mb] |
|---|---|---|---|
| 1 | 2 | 70.151 | 71.234 |
| 2 | 2 | 74.605 | 75.722 |
| 3 | 2 | 80.108 | 81.248 |
| 4 | 2 | 52.841 | 53.935 |
| 5 | 3 | 32.742 | 33.903 |
| 6 | 3 | 56.403 | 57.492 |
| 7 | 7 | 3.3738 | 4.4865 |
| 8 | 7 | 87.431 | 88.442 |
| 9 | 8 | 14.440 | 15.559 |
| 10 | 8 | 35.734 | 36.820 |
| 11 | 8 | 47.853 | 48.985 |
| 12 | 8 | 50.843 | 51.948 |
| 13 | 8 | 75.428 | 76.531 |
| 14 | 12 | 50.612 | 51.726 |
| 15 | 14 | 29.422 | 30.454 |
| 16 | 16 | 4.3558 | 5.4807 |
| 17 | 19 | 8.6151 | 9.6799 |

**Table S2. Overview for each cell of the number of fragments, the total length of DNA analyzed, the estimate of the characteristic length scale $\lambda$, and the $P$-value for the fit to the truncated exponential distribution with $K = 22$ kb [see Eq. (1)]. For the illumination time $\Delta t_{\mathrm{illu}} = 300$ s, a $P$-value could not be calculated due to the low number of fragments.**

| $\Delta t_{\mathrm{illu}}$ [s] | no. of. fragments $n$ | total length [Mb] | $\hat{\lambda}$ [kb] | P-value | $Q_{>}(L_{\mathrm{min}})$ |
|---|---|---|---|---|---|
| 150 s | 33 | 4.4 | $130 \pm 20$ | 0.27 | 0.4% |
| 150 s | 45 | 6.4 | $140 \pm 20$ | 0.10 | 0.64% |
| 210 s | 43 | 3.8 | $90 \pm 10$ | 0.27 | 0.018% |
| 300 s | 24 | 4.4 | $70 \pm 10$ | NA | 9.5 $10^{-4}$% |

**Table S3. Remapping of the quality reads from the ROI on chromosome 4 (49.127–49.147 Mb). Using the LS174T bulk sequencing data, we select the reads that originally mapped to the ROI on chromosome 4 with a high quality (Q⩾10) and remap to a version of hg18 where the ROI is masked. We list the regions where the reads are remapped. All reads can be remapped with high quality. The remapping region on chromosome 4 is a larger complex region (564 kb) that encompass the ROI (20 kb).**

| Chromosome | Start | End | Size |
|---|---|---|---|
| chr2 | 89,857,375 | 89,879,116 | 21,741 |
| chr4 | 49,094,349 | 49,659,115 | 564,766 |
| chr10 | 38,778,094 | 38,806,341 | 28,247 |
| chr10 | 42,355,506 | 42,815,376 | 459,870 |
| chr16 | 33,866,659 | 33,874,047 | 7,388 |
| chr20 | 29,811,928 | 29,829,677 | 17,749 |
| chr21 | 10,775,779 | 10,856,836 | 81,057 |
| chrY | 13,141,009 | 13,869,803 | 728,794 |
| chrUn - gl000216 | 1,461 | 80,663 | 79,202 |

**Table S4. Result of deletion/amplification detection on Chr19 of the LS174T cell line. The analysis was performed on Affymetrix 500K SNP data for all 23 chromosomes (data generated in 2007, reference genome used was hg18).**

| Chr. | start | end | length (bp) | Number of regions | Average copy number from segmentation result | Heterozygous rate from LOH result | Description |
|---|---|---|---|---|---|---|---|
| 19 | 23,661,800 | 24,084,419 | 422,620 | 13 | 1.99906 | 0.0361446 | Copy-Neutral LOH |
| 19 | 32,026,335 | 32,441,437 | 415,103 | 17 | 1.99906 | 0.038835 | Copy-Neutral LOH |
| 19 | 36,871,295 | 37,670,469 | 799,175 | 10 | 1.99906 | 0.0136986 | Copy-Neutral LOH |

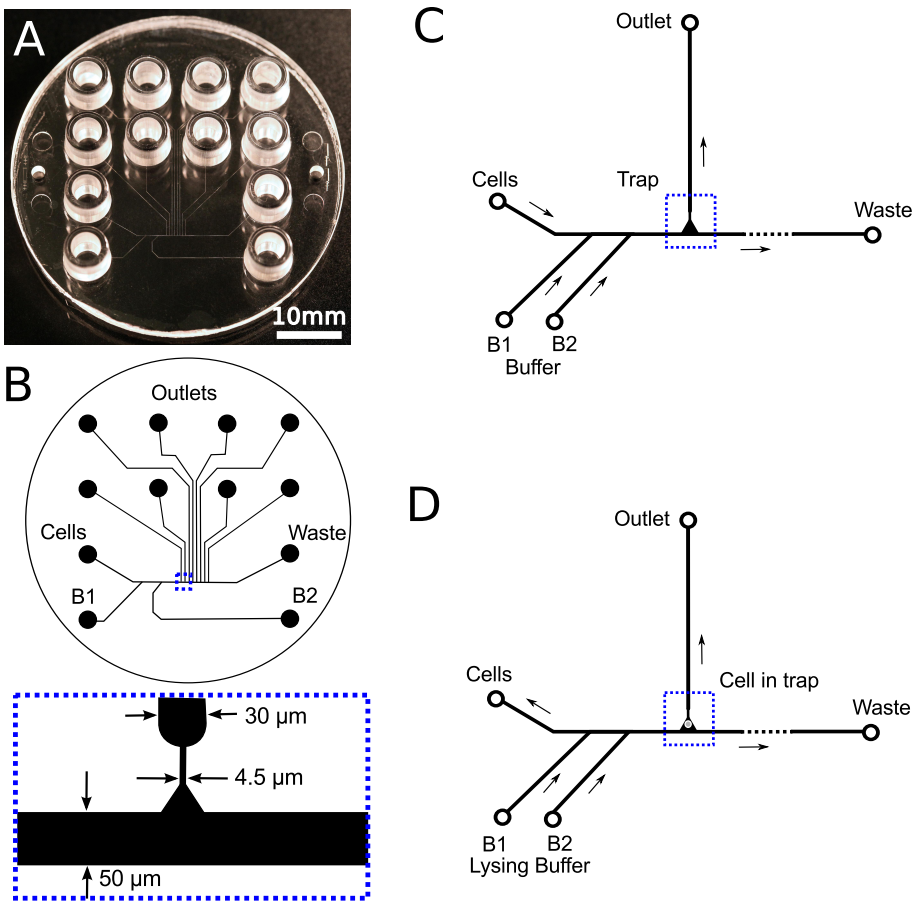


**Fig. S1.** Device layout used for single cell DNA sequencing experiments. (A) Picture of the device. (B) The device comprises eight parallel cell traps with individual outlets. Each cell trap is a 4.5 $\mu$m wide constriction connecting the main channel with the outlet channel. The device has three inlets connecting the main channel, one for cells, two for buffer (B1 and B2) and a waste. (C) For cell capture, cells are introduced from the cell inlet and aligned against the wall of the channel by the incoming flow of buffer from inlet B1 and B2. (D) Lysing buffer is introduced simultaneously from the buffer inlets (B1 and B2). Each trap is connected to a separate outlet where the cell lysate can be collected.

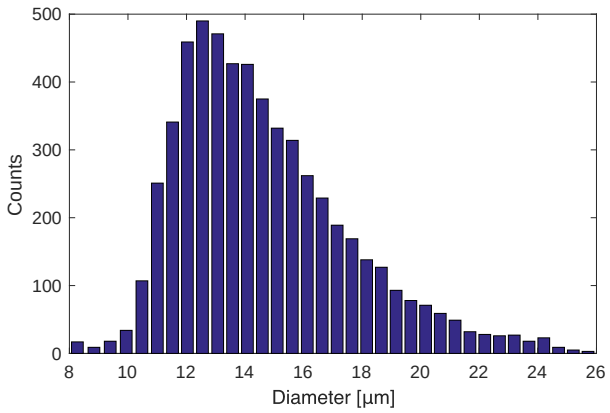


**Fig. S2.** Distribution of diameters of cells from the population of LS174T cells measured as individual cells were fed into an Orflo cell counter in FACSflow buffer. The average diameter is 14 $\mu$m with a standard deviation of 2 $\mu$m. This population may contain cell doublets.

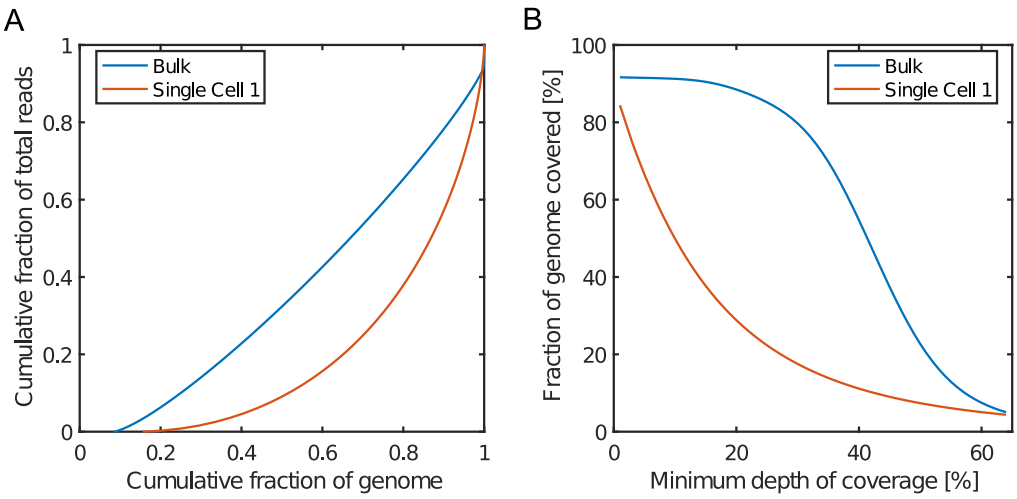


**Fig. S3.** Sequencing metrics. (A) Lorentz and (B) Coverage plots for the LS174T bulk sequencing and the single cell #1 shown in Fig. 4 of the main text.

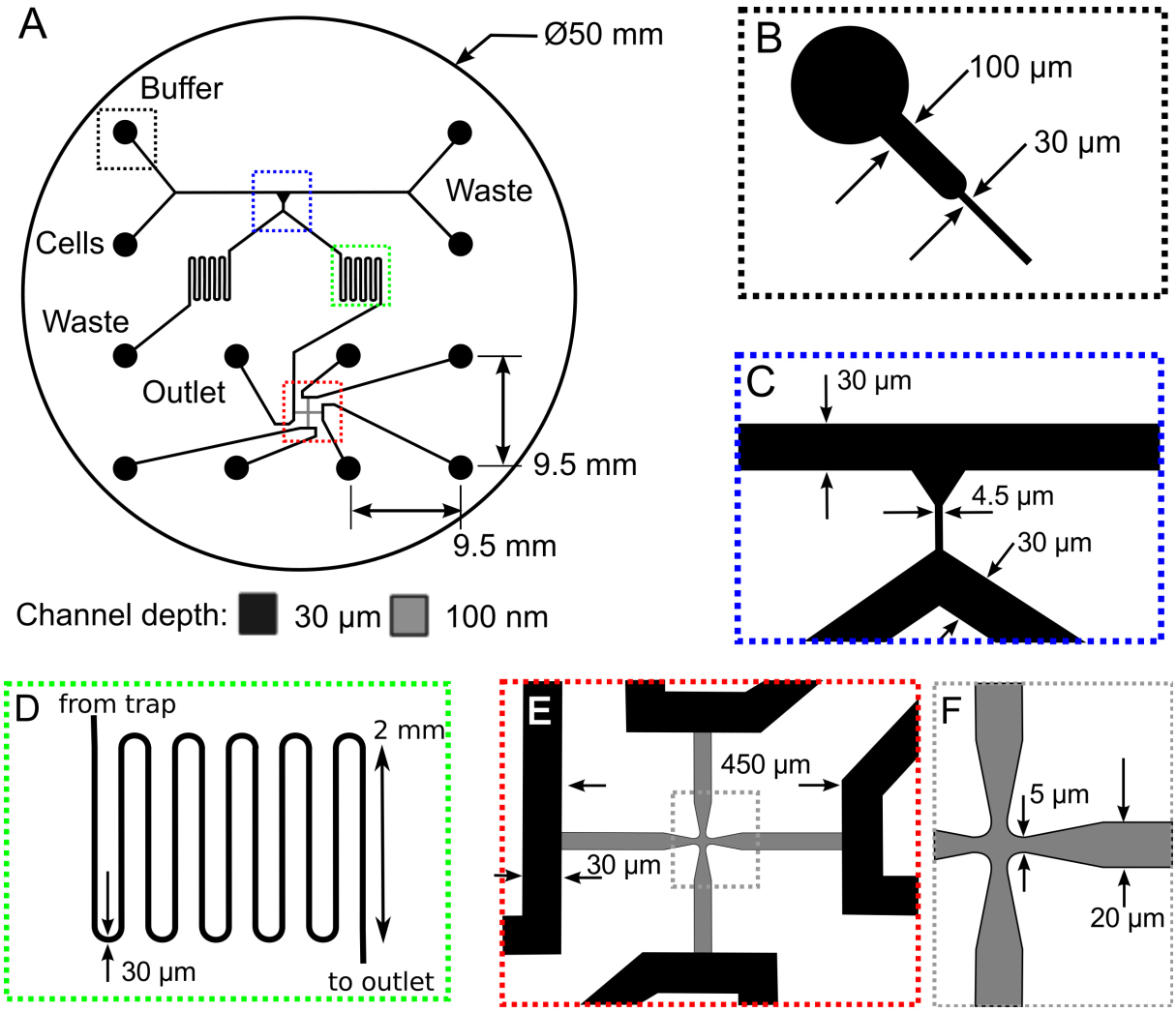


**Fig. S4.** Design of D-R mapping device. (A) The device comprises a main channel with inlets for cells and buffer, a cell trap connecting the main channel to the outlets. On the outlet side, there is meandering channel and a flow stretch device. (B) All channels are 30 $\mu$m-wide, and deep. Near inlets and outlets the channel width is larger. (C) The cell trap is a 4.5 $\mu$m-wide constriction. (D) The meandering channel is about 20 mm-long. (E) The nanoslit is 100 nm-deep, 20 $\mu$m-wide and 450 $\mu$m across. (F) Detail of the nanoslit at the center of the device.

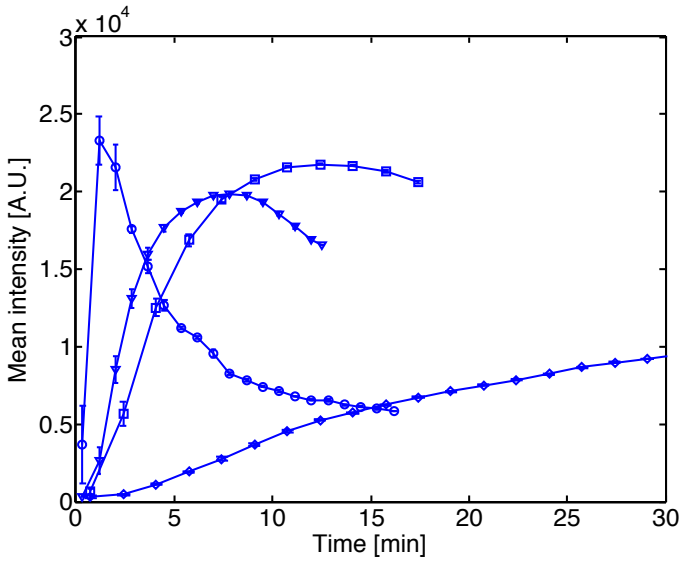


**Fig. S5.** Nucleus DNA intensity during staining at various YOYO-1 concentrations. Mean fluorescence intensity of the nucleus of cells during lysis and staining at YOYO-1 concentration of 5 $\mu$M (circles, 4 cells), 1 $\mu$M (triangles, 5 cells averaged), 0.5 $\mu$M (squares, 5 cells), and 0.1 $\mu$M (diamonds, 5 cells) in 0.5xTBE + 0.5% v/v Triton-X100.

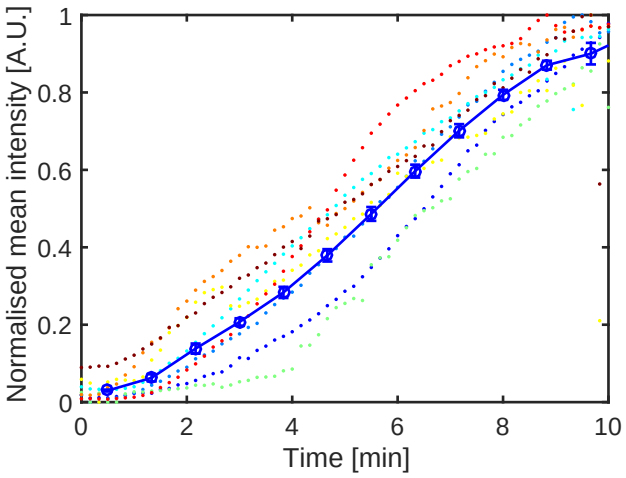


**Fig. S6.** Intensities from nucleic DNA during staining at 0.1 $\mu$M YOYO-1. Normalised YOYO-1 fluorescence intensity of eight individual cells lysed and stained at 0.1 $\mu$M YOYO-1 in 0.5xTBE + 0.5% v/v Triton-X100 (dots). Mean of all eight cells (circles).

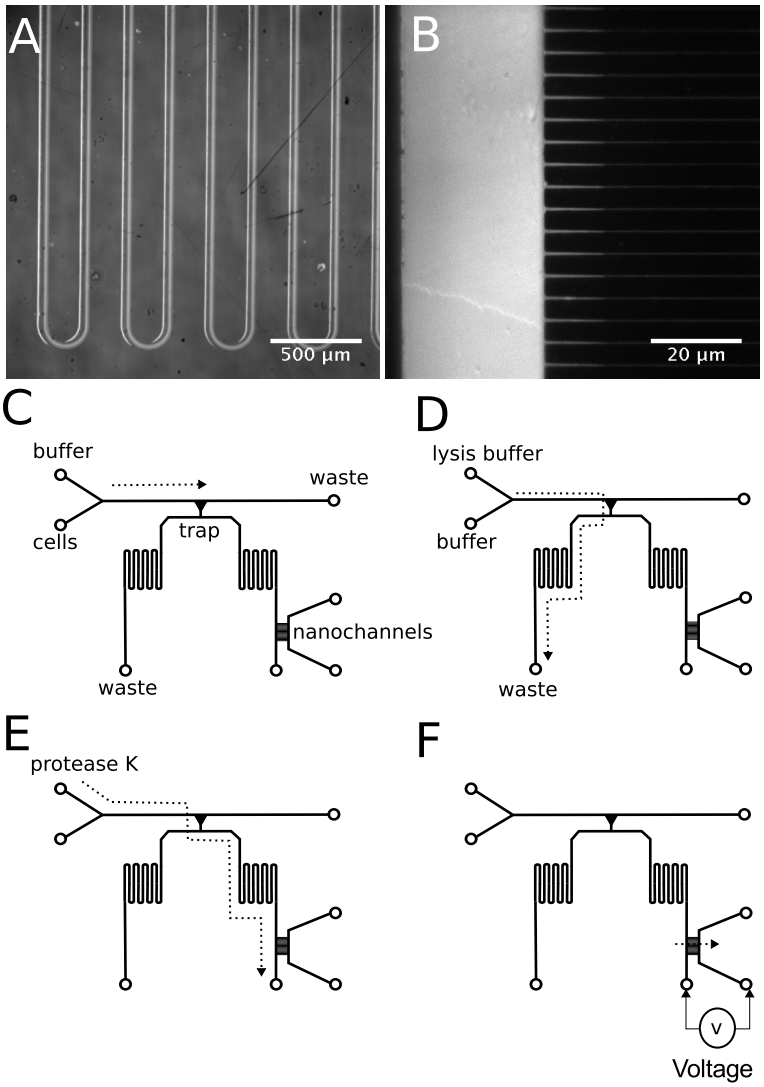


**Fig. S7.** Device design for DNA fragment sizing. (A) Bright field image of the meandering channel, and (B) of the connection between microchannel and nanochannel array on the sealed polymer device. (C–F) The four main steps of the device operation including the trapping of a cell, the lysis of the cytoplasm, the proteolysis, and the introduction of DNA fragments to the nanochannel array for sizing.

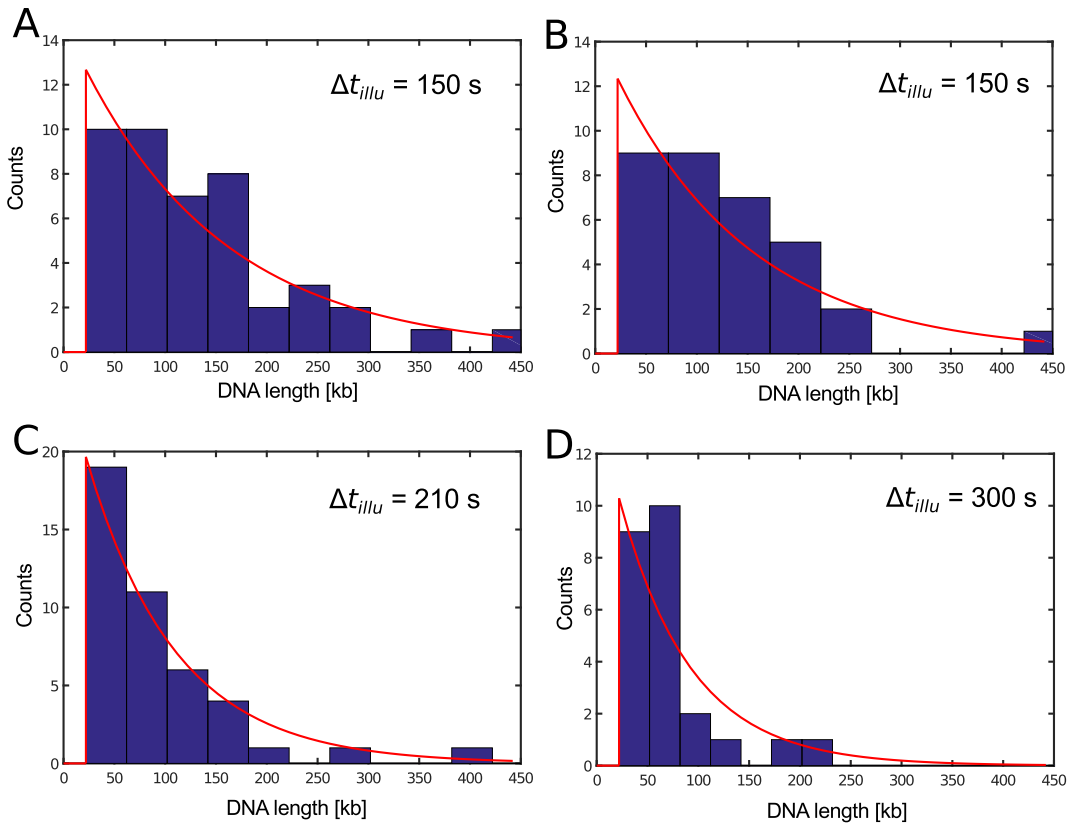


**Fig. S8.** Histograms of sampled DNA fragment lengths from four different cells after three different durations of the illumination $\Delta t_{\mathrm{illu}}$. Red curves are the expression in Eq. (1) with the estimated values for $\lambda$ from Eq. (2) and a common cut-off $K = 22\,\mathrm{kb}$. Estimated parameter values for $\lambda$ and the $P$-values for the fits are stated in **Supplementary Table S2**.

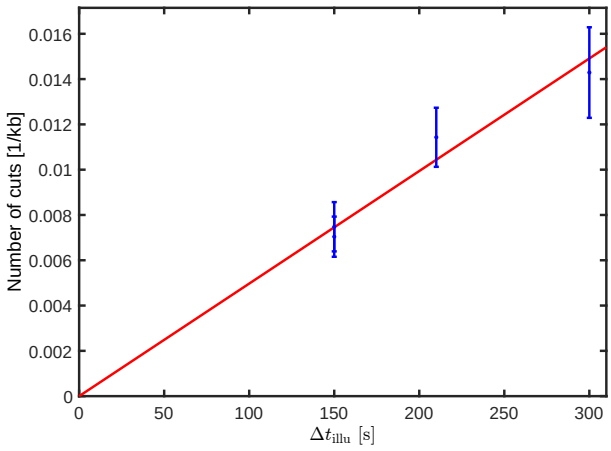


**Fig. S9.** Estimated nicking probability per kb, i.e. $\hat{p}_{\mathrm{nick}} = 1/\hat{\lambda}$. Fit to a straight line through the origin gives a slope $(5.0 \pm 0.2) \cdot 10^{-5}/(\mathrm{kb} \cdot \mathrm{s})$.

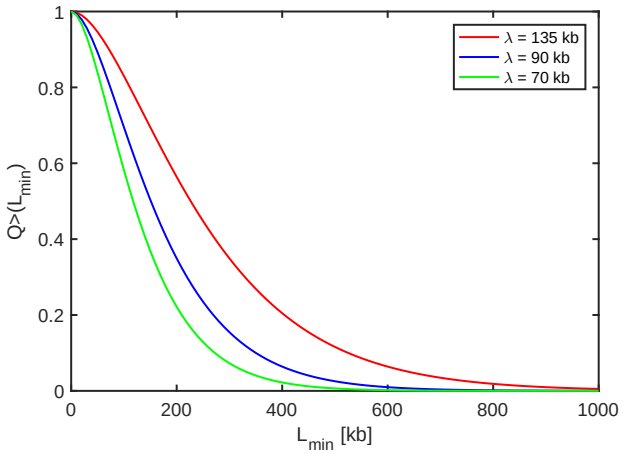


**Fig. S10.** Fraction $Q_{>}(L_{\min})$ of DNA with a length above a threshold $L_{\min}$ for three different values of the mean length $\lambda$.

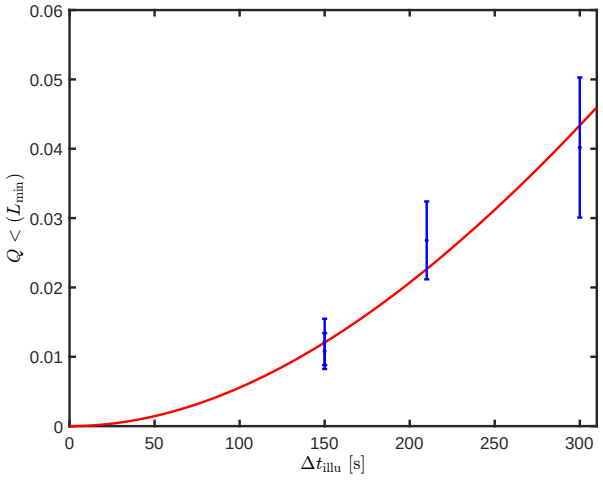


**Fig. S11.** Blue data points with error bars are the fraction of DNA with a length below the cut-off $L_{\min} = 22\,\mathrm{kb}$ versus illumination time $\Delta t_{\mathrm{illu}}$ calculated using Eq. (5) and with the nicking probabilities $p_{\mathrm{nick}}$ stated in **Supplementary Table S2**. The red curve is also calculated with Eq. (5), but with the nicking probability from the straight-line fit in **Fig. S9**, i.e., $p_{\mathrm{nick}} = 5 \cdot 10^{-5} \frac{1}{\mathrm{kb\,s}} \Delta t_{\mathrm{illu}}$.

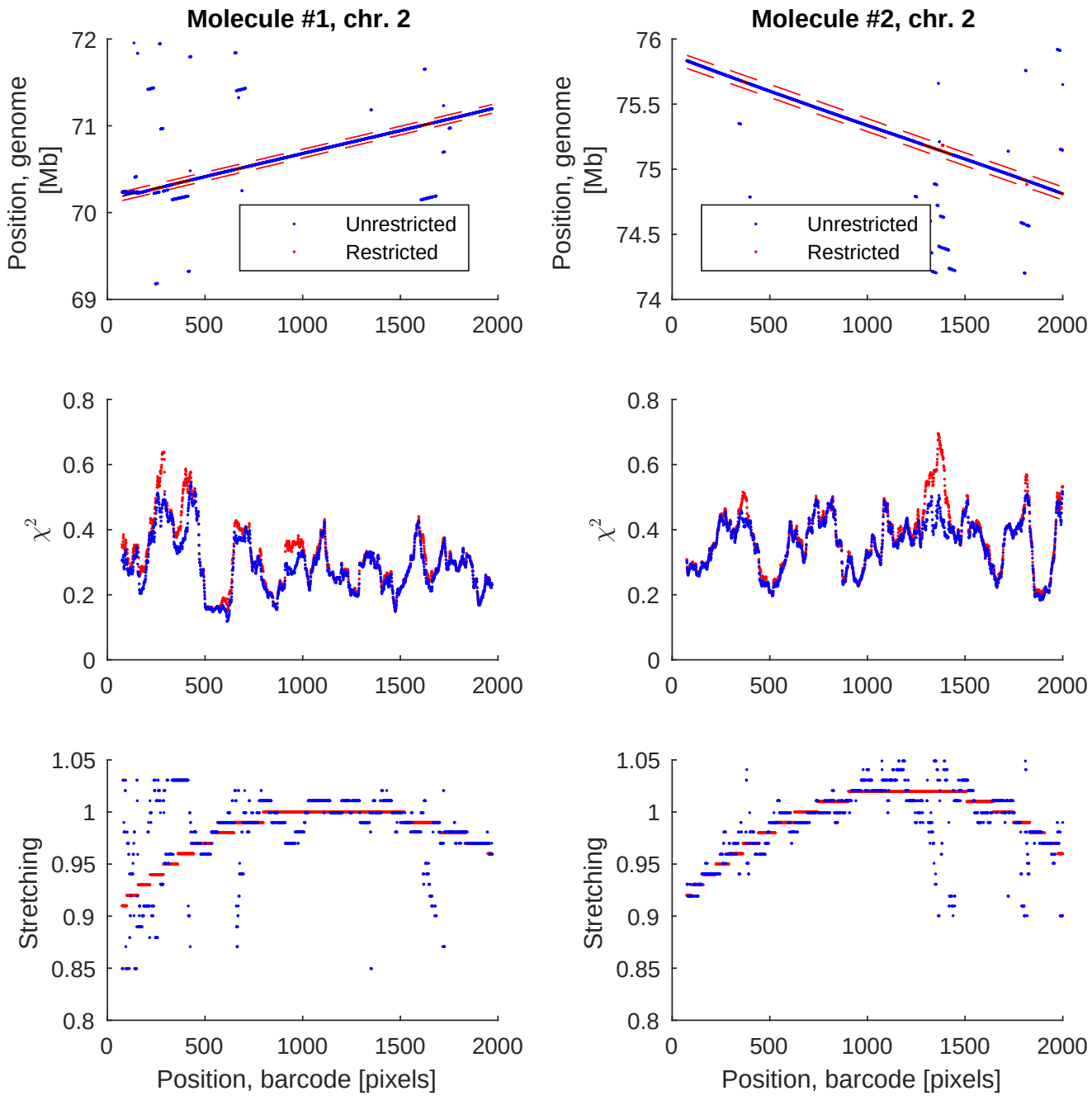


**Fig. S12.** Molecule 1 and 2. Upper panel: the positions on the reference genome (GRCh37) versus the position on the D-R map. The thin black line is a guideline that indicates the positions of the diagonal. Middle panel: the match scores, $\chi^2$, for the best match at each position. $\chi^2$ is normalized so $\chi^2$=0 denotes a perfect match, and a maximal mismatch would correspond to $\chi^2$=2. Lower panel: the degrees of stretching for the best match for each position of the window, normalized so stretching=1 means fully stretched.

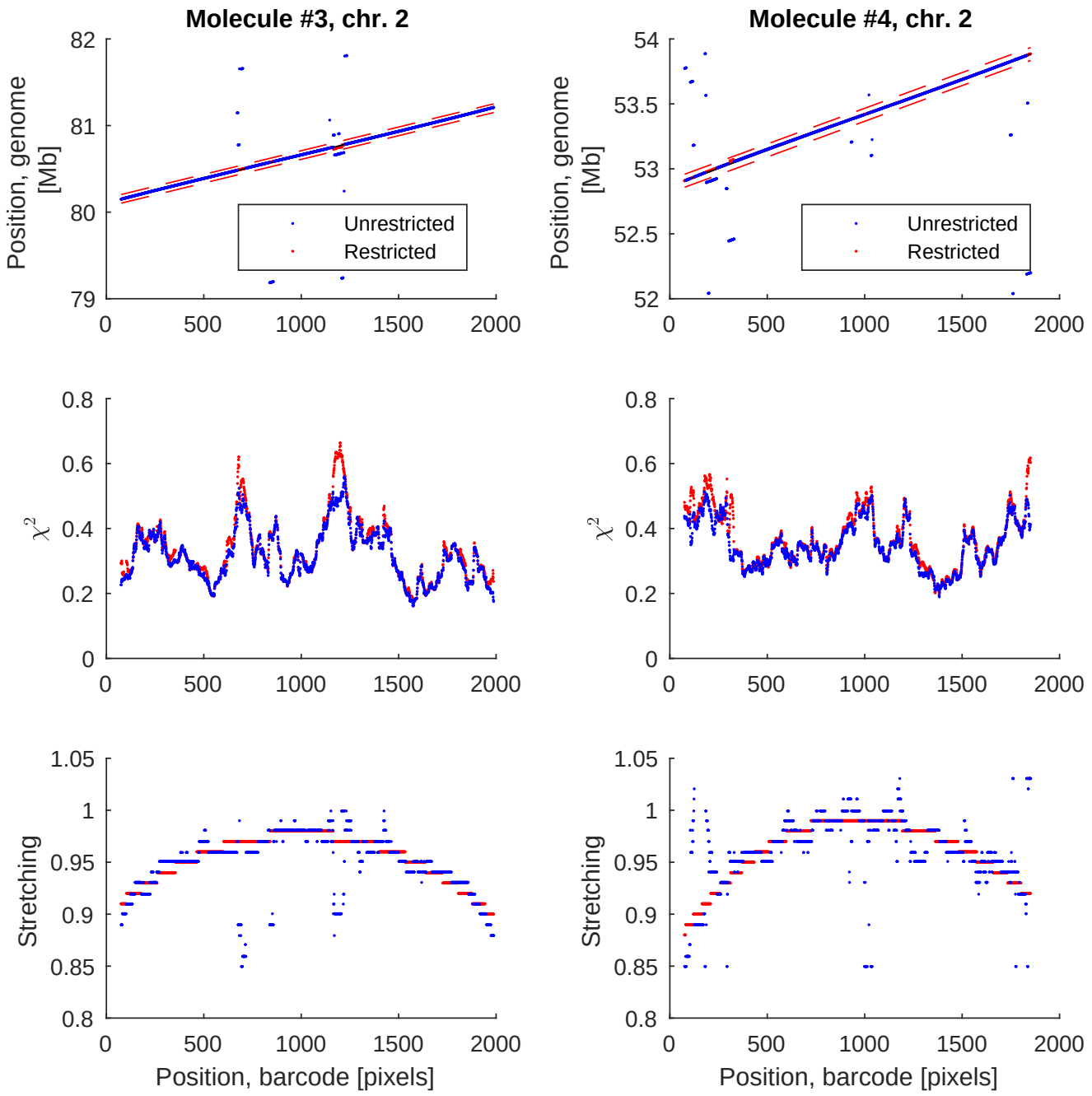


**Fig. S13.** Molecule 3 and 4. Upper panel: the positions on the reference genome (GRCh37) versus the position on the D-R map. The thin black line is a guideline that indicates the positions of the diagonal. Middle panel: the match scores, $\chi^2$, for the best match at each position. $\chi^2$ is normalized so $\chi^2$=0 denotes a perfect match, and a maximal mismatch would correspond to $\chi^2$=2. Lower panel: the degrees of stretching for the best match for each position of the window, normalized so stretching=1 means fully stretched.

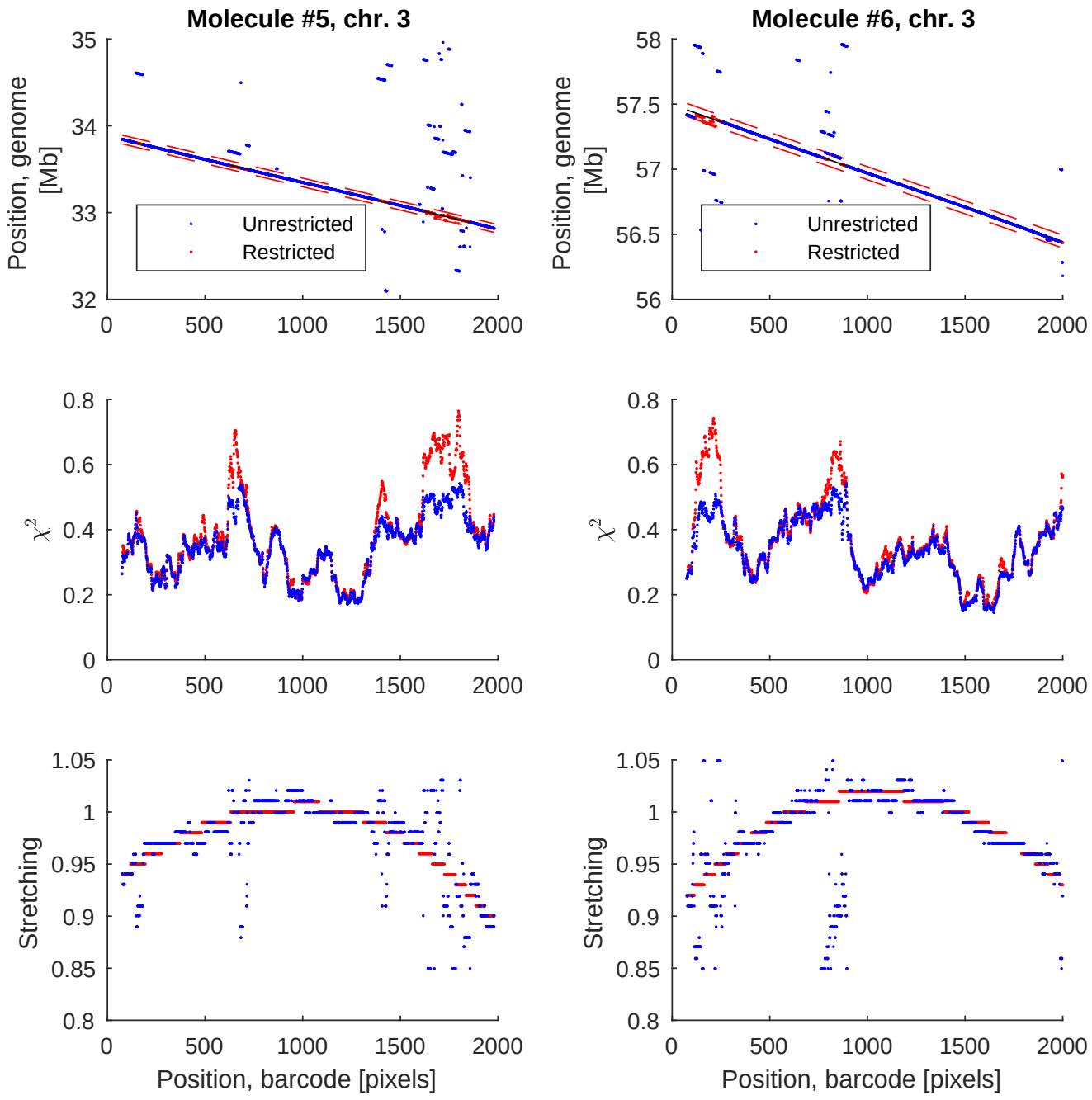


**Fig. S14.** Molecule 5 and 6. Upper panel: the positions on the reference genome (GRCh37) versus the position on the D-R map. The thin black line is a guideline that indicates the positions of the diagonal. Middle panel: the match scores, $\chi^2$, for the best match at each position. $\chi^2$ is normalized so $\chi^2$=0 denotes a perfect match, and a maximal mismatch would correspond to $\chi^2$=2. Lower panel: the degrees of stretching for the best match for each position of the window, normalized so stretching=1 means fully stretched.

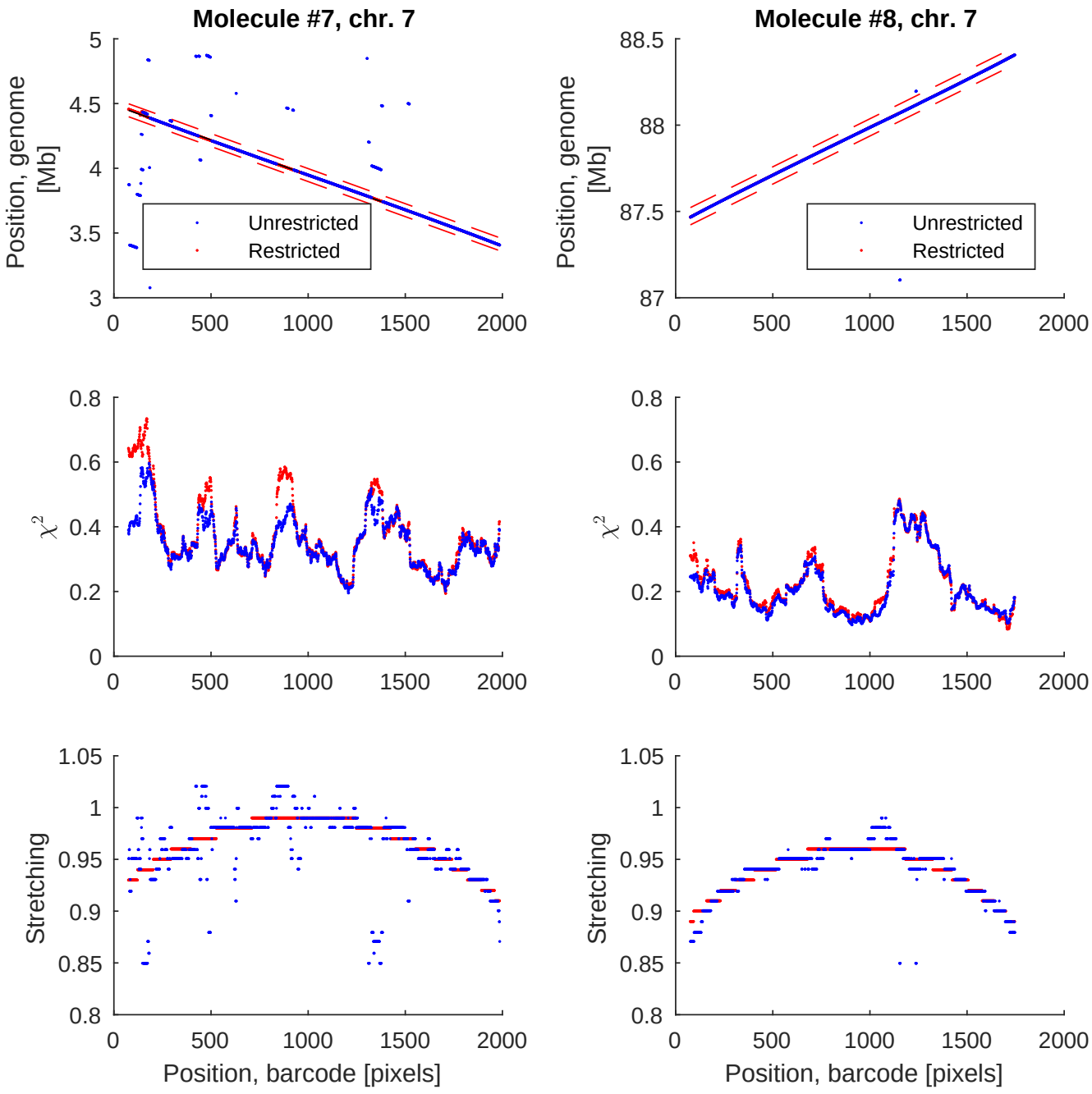


**Fig. S15.** Molecule 7 and 8. Upper panel: the positions on the reference genome (GRCh37) versus the position on the D-R map. The thin black line is a guideline that indicates the positions of the diagonal. Middle panel: the match scores, $\chi^2$, for the best match at each position. $\chi^2$ is normalized so $\chi^2$=0 denotes a perfect match, and a maximal mismatch would correspond to $\chi^2$=2. Lower panel: the degrees of stretching for the best match for each position of the window, normalized so stretching=1 means fully stretched.

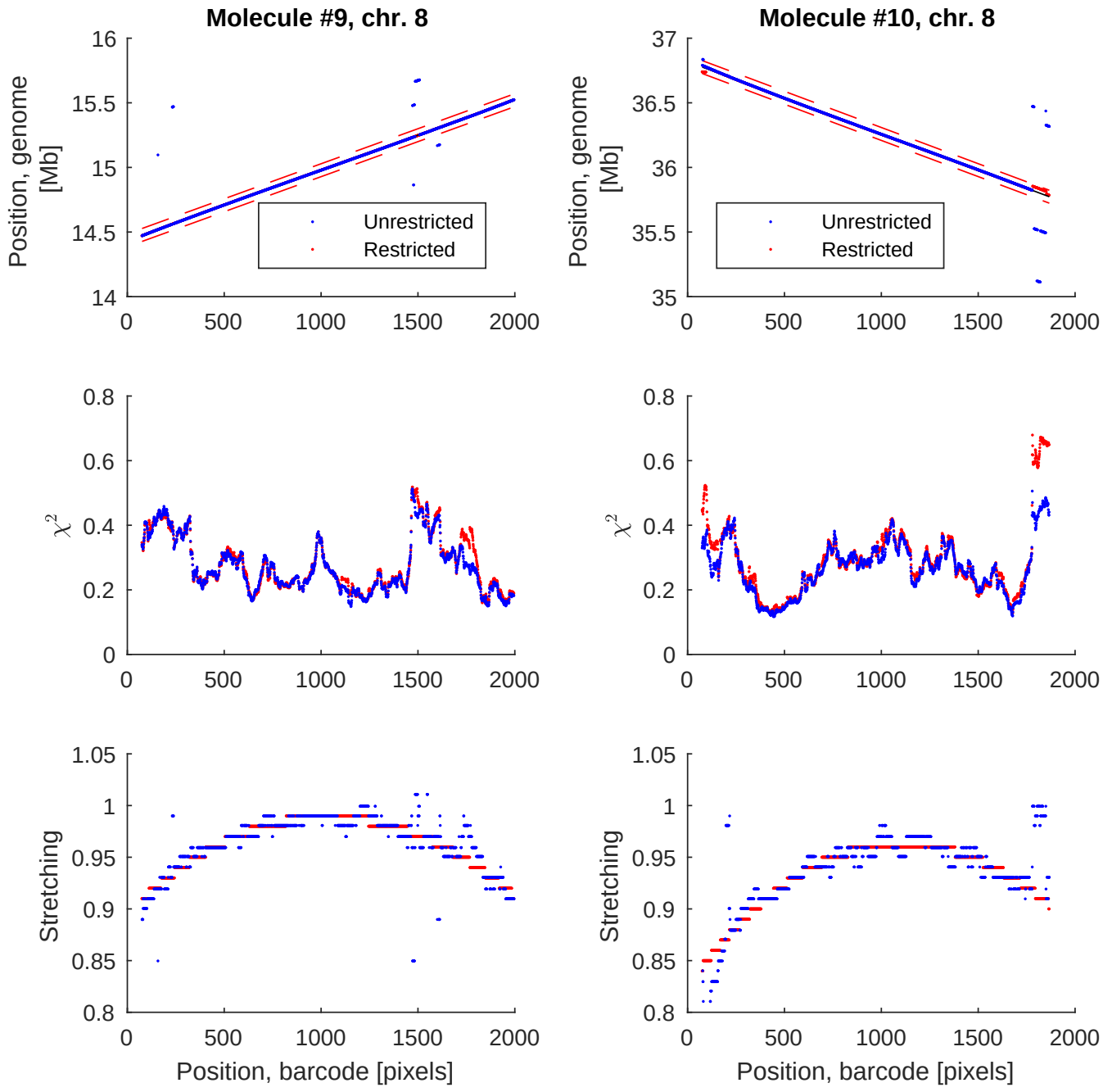


**Fig. S16.** Molecule 9 and 10. Upper panel: the positions on the reference genome (GRCh37) versus the position on the D-R map. The thin black line is a guideline that indicates the positions of the diagonal. Middle panel: the match scores, $\chi^2$, for the best match at each position. $\chi^2$ is normalized so $\chi^2$=0 denotes a perfect match, and a maximal mismatch would correspond to $\chi^2$=2. Lower panel: the degrees of stretching for the best match for each position of the window, normalized so stretching=1 means fully stretched.

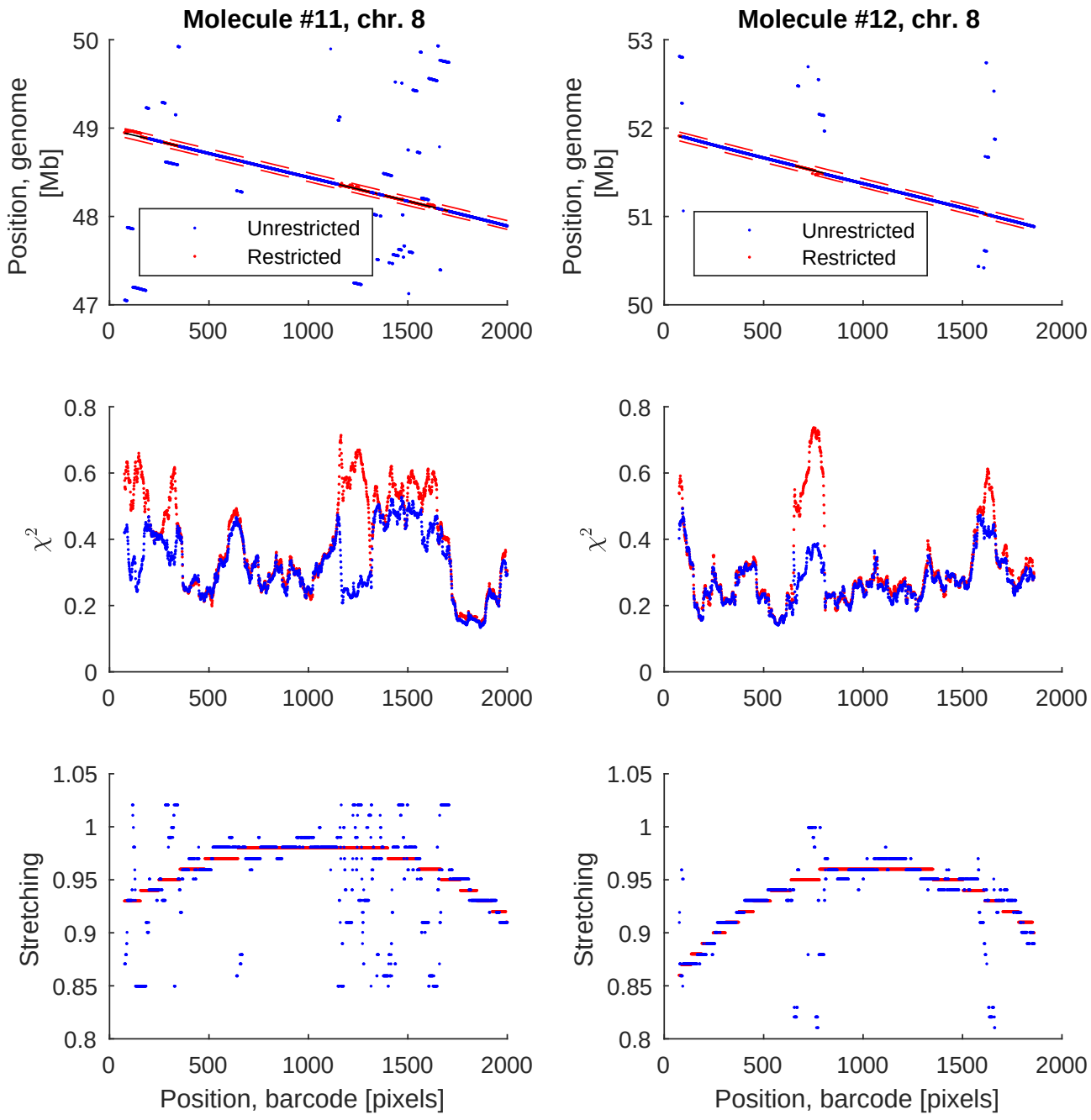


**Fig. S17.** Molecule 11 and 12. Upper panel: the positions on the reference genome (GRCh37) versus the position on the D-R map. The thin black line is a guideline that indicates the positions of the diagonal. Middle panel: the match scores, $\chi^2$, for the best match at each position. $\chi^2$ is normalized so $\chi^2$=0 denotes a perfect match, and a maximal mismatch would correspond to $\chi^2$=2. Lower panel: the degrees of stretching for the best match for each position of the window, normalized so stretching=1 means fully stretched.

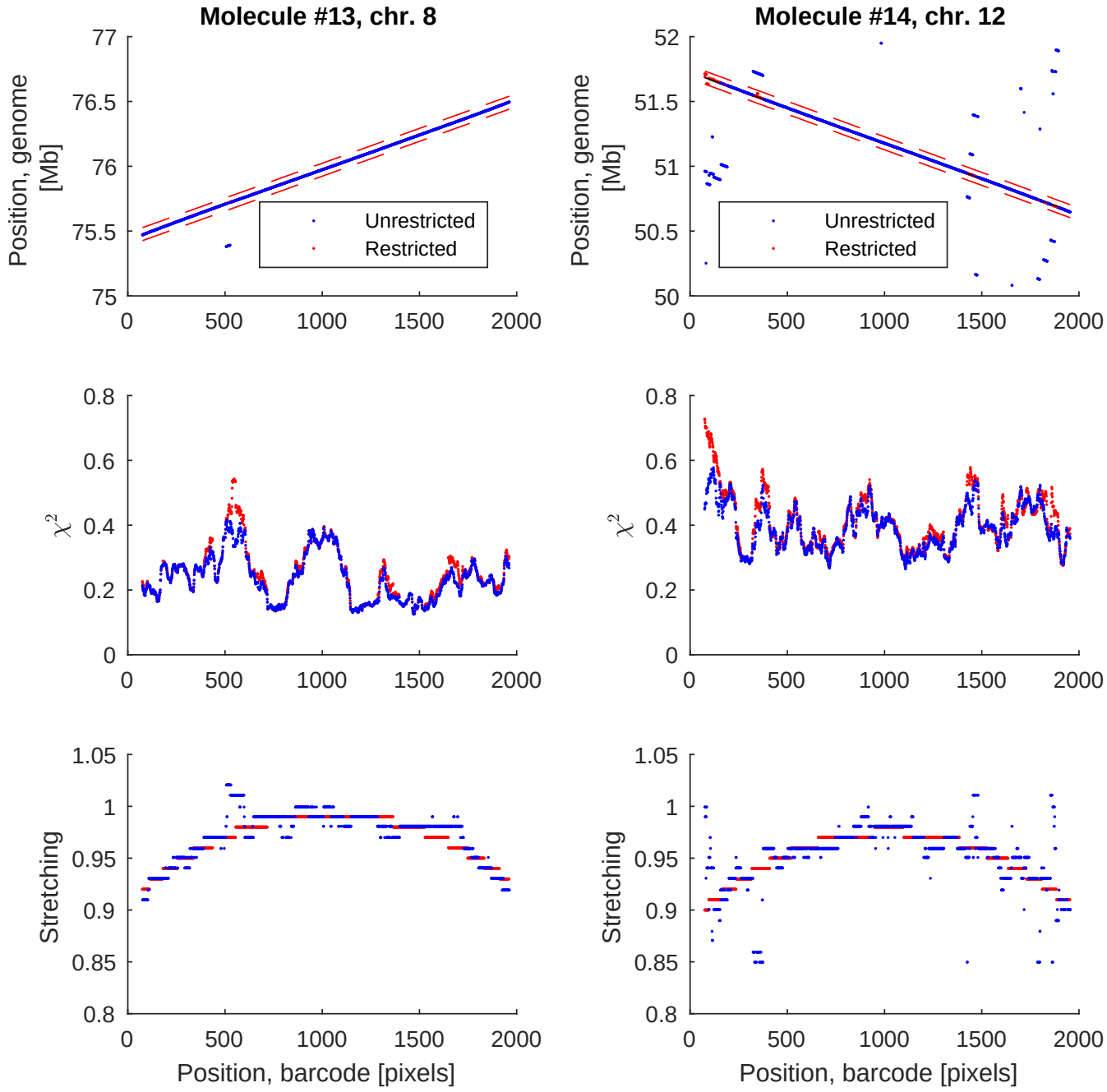


**Fig. S18.** Molecule 13 and 14. Upper panel: the positions on the reference genome (GRCh37) versus the position on the D-R map. The thin black line is a guideline that indicates the positions of the diagonal. Middle panel: the match scores, $\chi^2$, for the best match at each position. $\chi^2$ is normalized so $\chi^2$=0 denotes a perfect match, and a maximal mismatch would correspond to $\chi^2$=2. Lower panel: the degrees of stretching for the best match for each position of the window, normalized so stretching=1 means fully stretched.

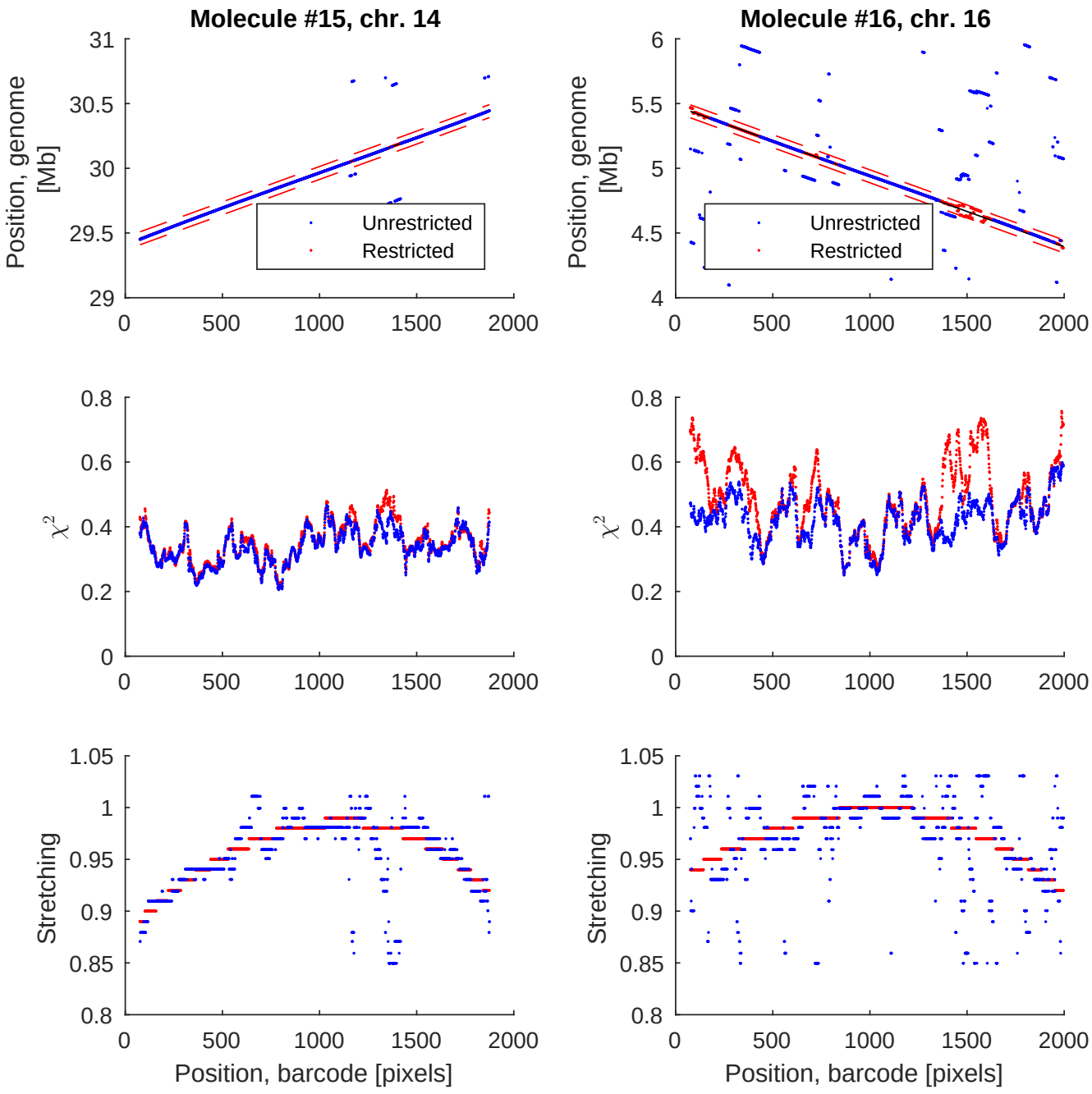


**Fig. S19.** Molecule 15 and 16. Upper panel: the positions on the reference genome (GRCh37) versus the position on the D-R map. The thin black line is a guideline that indicates the positions of the diagonal. Middle panel: the match scores, $\chi^2$, for the best match at each position. $\chi^2$ is normalized so $\chi^2$=0 denotes a perfect match, and a maximal mismatch would correspond to $\chi^2$=2. Lower panel: the degrees of stretching for the best match for each position of the window, normalized so stretching=1 means fully stretched.

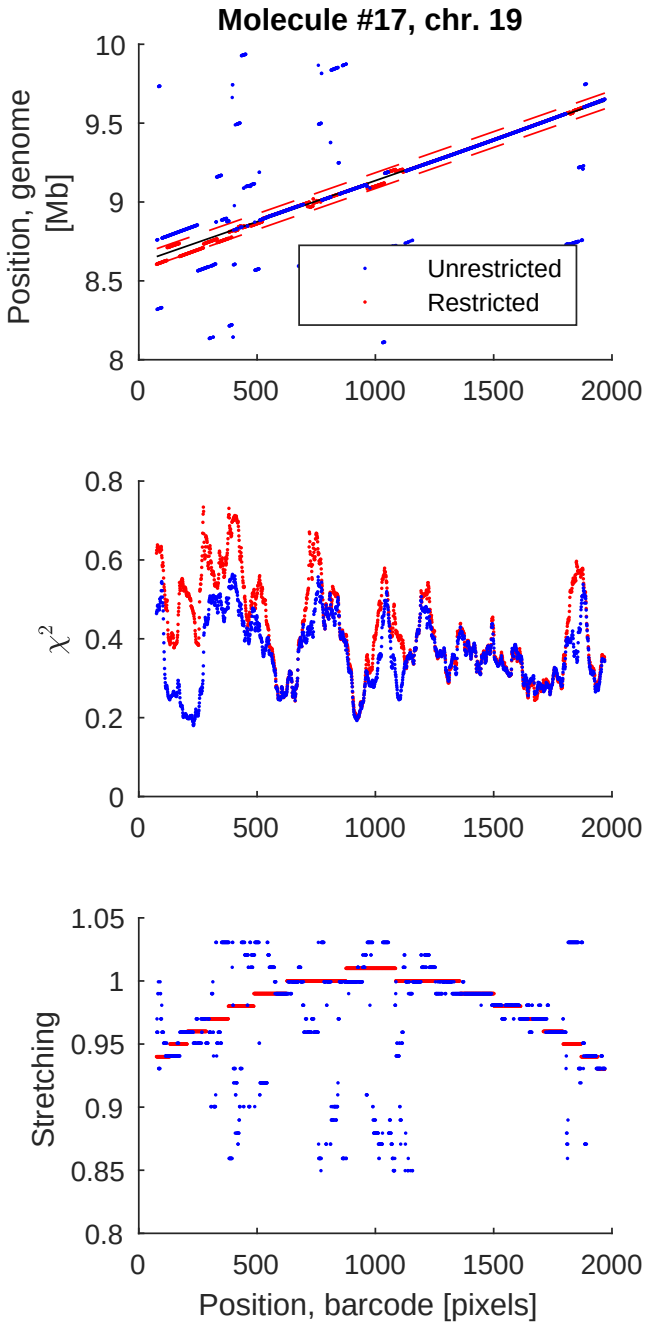


**Fig. S20.** Molecule 17. Upper panel: the positions on the reference genome (GRCh37) versus the position on the D-R map. The thin black line is a guideline that indicates the positions of the diagonal. Middle panel: the match scores, $\chi^2$, for the best match at each position. $\chi^2$ is normalized so $\chi^2$=0 denotes a perfect match, and a maximal mismatch would correspond to $\chi^2$=2. Lower panel: the degrees of stretching for the best match for each position of the window, normalized so stretching=1 means fully stretched.

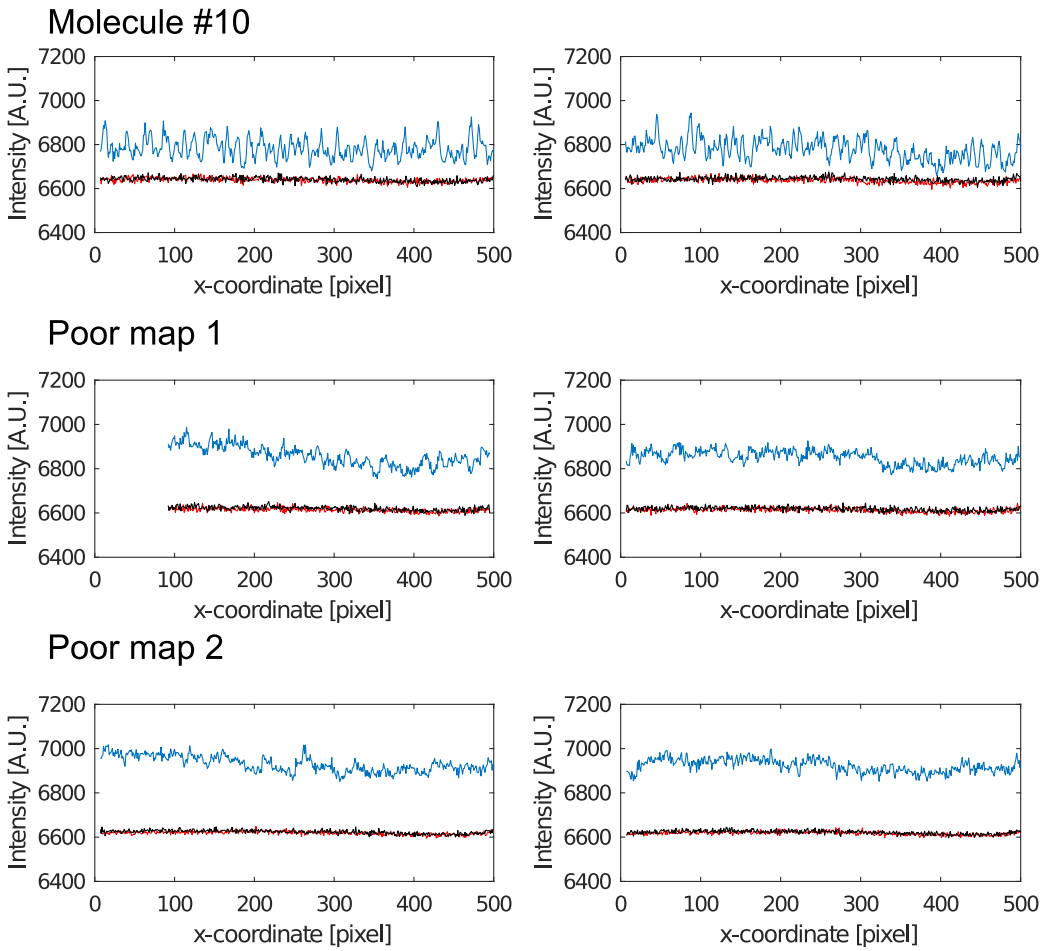


**Fig. S21.** Plots of D-R map (blue) and background (black and red) for two consecutive field-of-views for molecule #10 mapped to chromosome 8. The D-R map has high contrast with valleys near the background. Same for two other molecules that could not be mapped to the reference genome: the D-R maps have poorer contrast and higher average fluorescence.

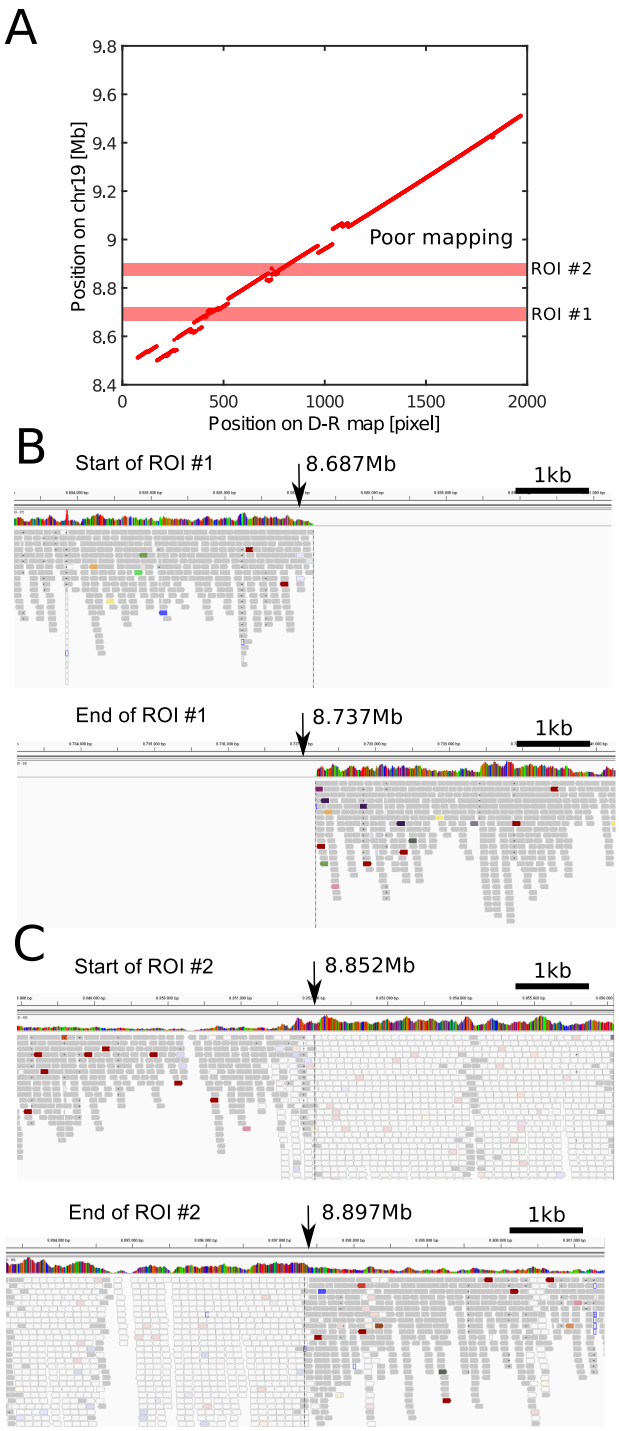


**Fig. S22.** Chromosome 19 variant. (A) D-R mapping indicates a poor match on the region 8.5-9.0 Mb on chromosome 19. (B-C) Sequencing of the LS174T cell line indicates two regions of interests on this part of chromosome 19: a deletion and a region of higher coverage but poor mapping of the reads.

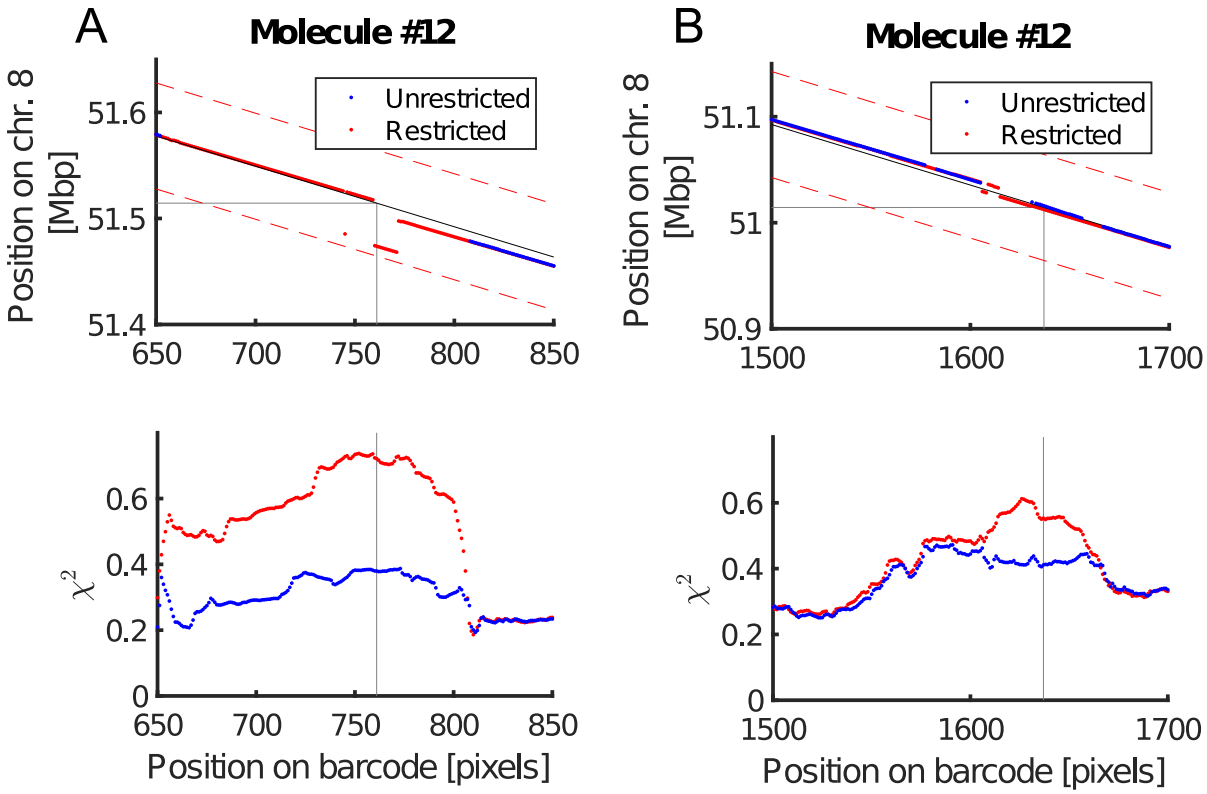


**Fig. S23.** Structural variation detection in Molecule 12 using the 'unrestricted' and 'restricted' 'Sliding Window' method described briefly above and in detail in SI Appendix to Ref. (3). Here the figures show a zoom of the data shown in Fig. S17. The thin black line shows the diagonal, and the dashed red lines marks the 50 kb search range on each side of the diagonal used in the 'restricted Sliding Window' method. The shifts in the aligned positions in the 'restricted' alignment indicate a $\sim 12$ kb deletion near position 51.514 Mb (Panel A), and a $\sim 8$ kb deletion near position 51.014 Mb (Panel B). Lower panels: the match scores, $\chi^2$, for the best match at each position. $\chi^2$ is normalized so $\chi^2$=0 denotes a perfect match, and a maximal mismatch would correspond to $\chi^2$=2.

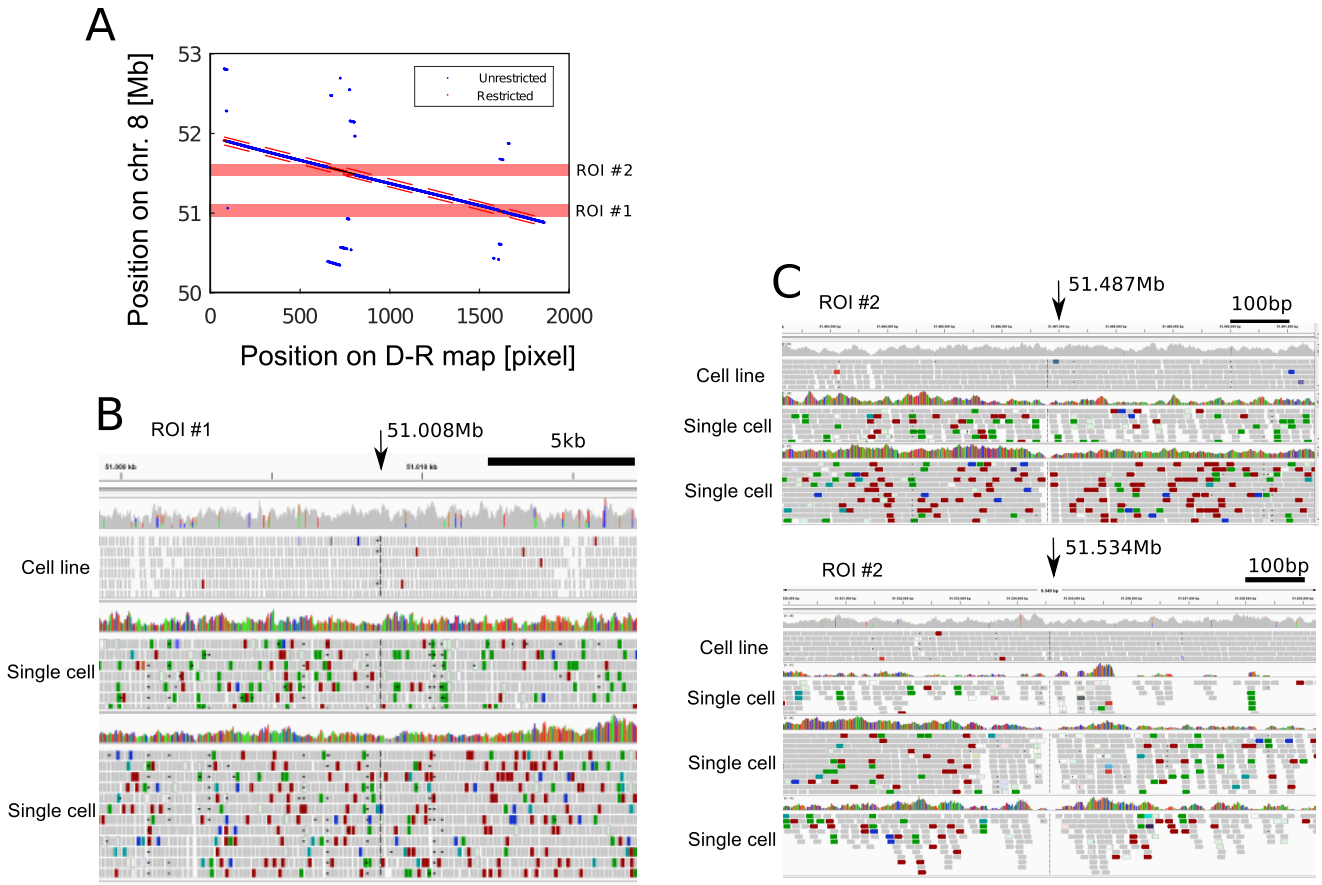


**Fig. S24.** Chromosome 8 variant detected on molecule #12. (A) Plots of the restricted and 'unrestricted' 'Sliding Window' analysis showing the two regions of interest ROI#1 and 2 where a line shift is detected on molecule 12 (see Fig. S23). (B) Integrated Genome Viewer plots of the sequencing reads for the bulk of the cell line and two single cells with good coverage in that region. Two regions of reads with wrong insert size (white reads) encompass the position 51.008 Mb (ROI#1). (C) Same for the positions 51.487 Mb and 51.534 Mb. A small deletion is visible for two of the single cells with best coverage in those regions around the ROI#2.

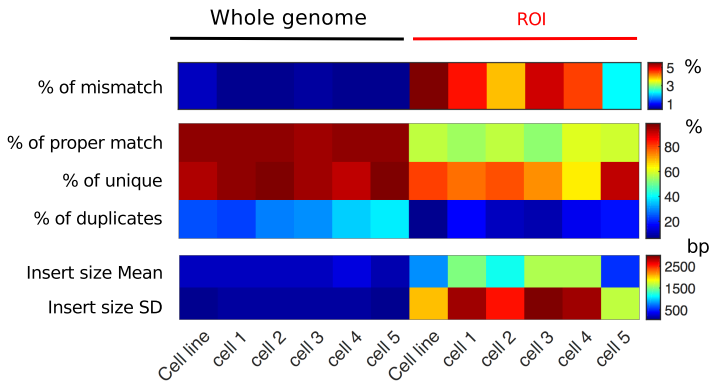


**Fig. S25.** Metrics of the mate-pair sequencing reads from the ROI on chromosome 4 compare to outside the ROI.

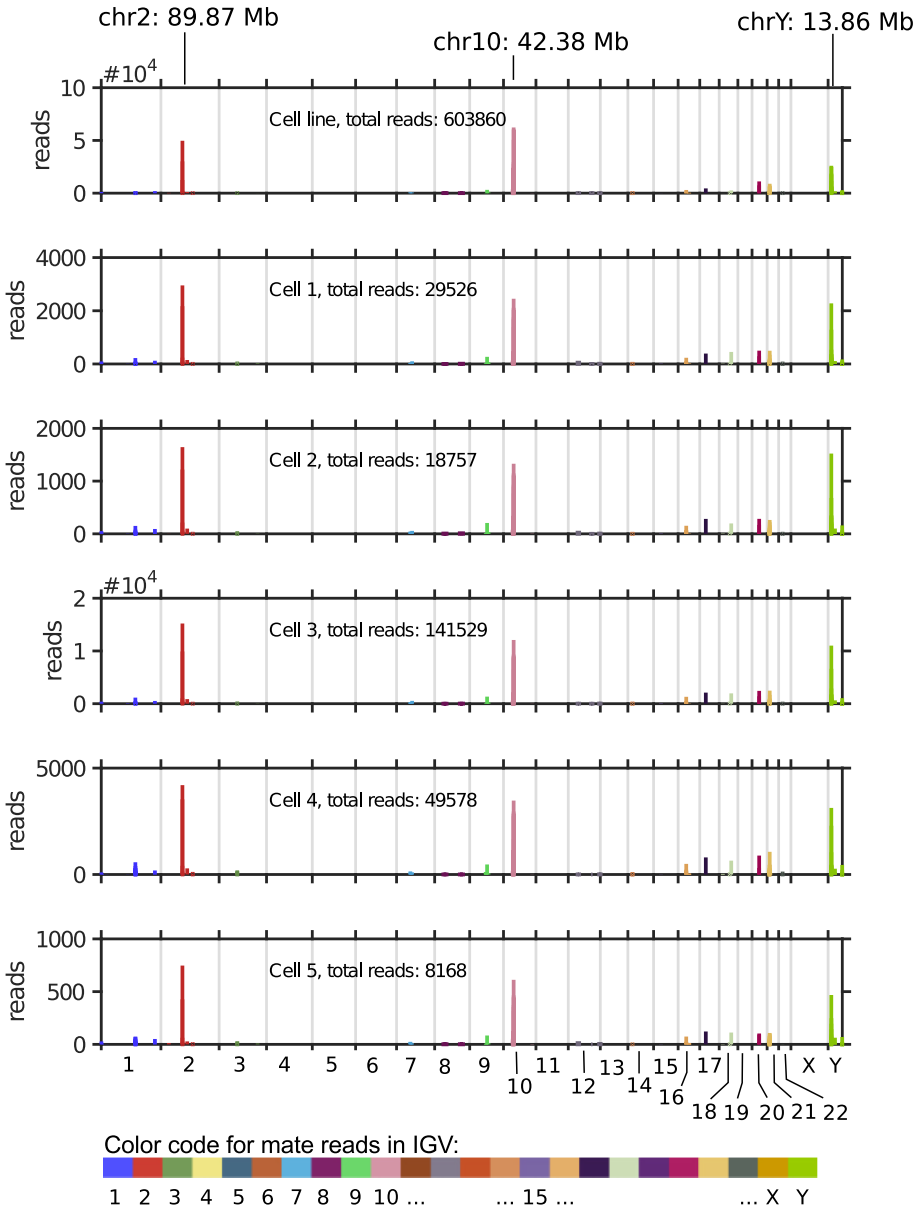


**Fig. S26.** Detailed graph showing possible candidates for a translocation from the region of interest on chromosome 4 (49.126-49.147 Mb, see **Fig. 5** main text) using DNA sequencing of the cell line (bulk) and 5 single cells. Coverage plot of the mapping location of the second read of the pair in the cases that it maps to a chromosome other than chromosome 4; reads are binned in 10 kb windows with a threshold per bin of at least 50 reads. Possible candidate positions for a translocation from chromosome 4 are 89.87 Mb on chromosome 2, 42.38 Mb on chromosome 10 and 13.86 Mb on chromosome Y.

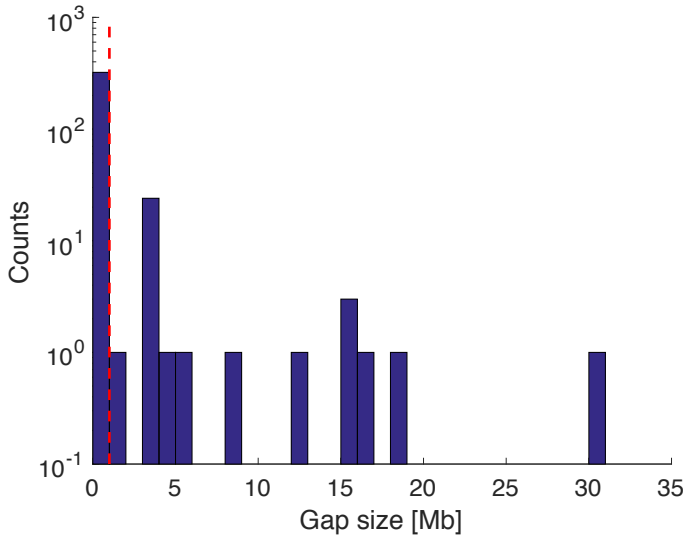


**Fig. S27.** Gap size distribution in the human reference genome GRCh37. The total number of gaps is 357. The number of gaps larger than 1 Mb (to the right of the dashed red line) is 35, and they cover 215.4 Mb.

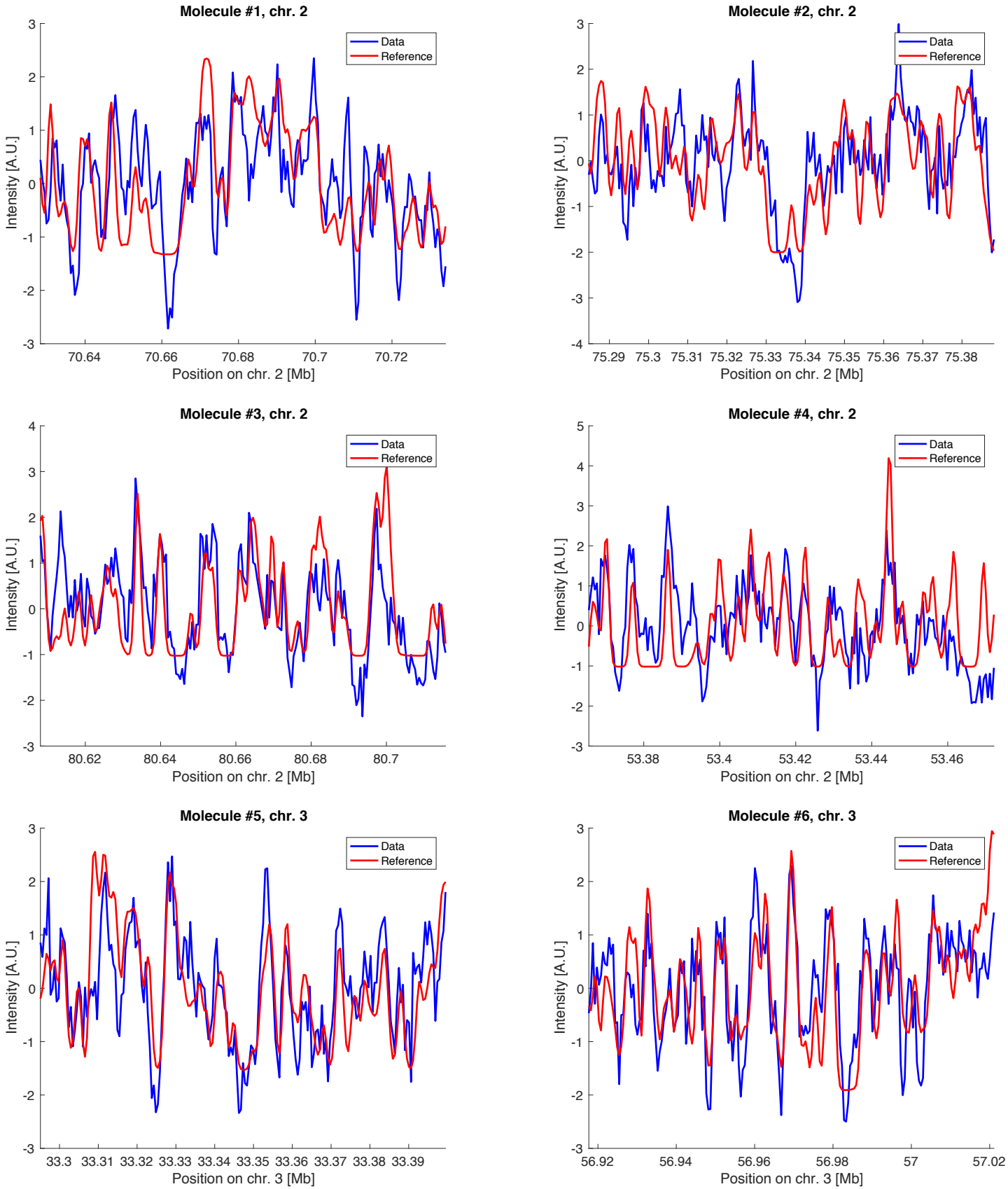


**Fig. S28.** Matches between D-R maps (pixel 900-1100) and positions on the computer-simulated whole-genome melting map for Molecules 1 to 6 (see also Fig. 3E in the main text).

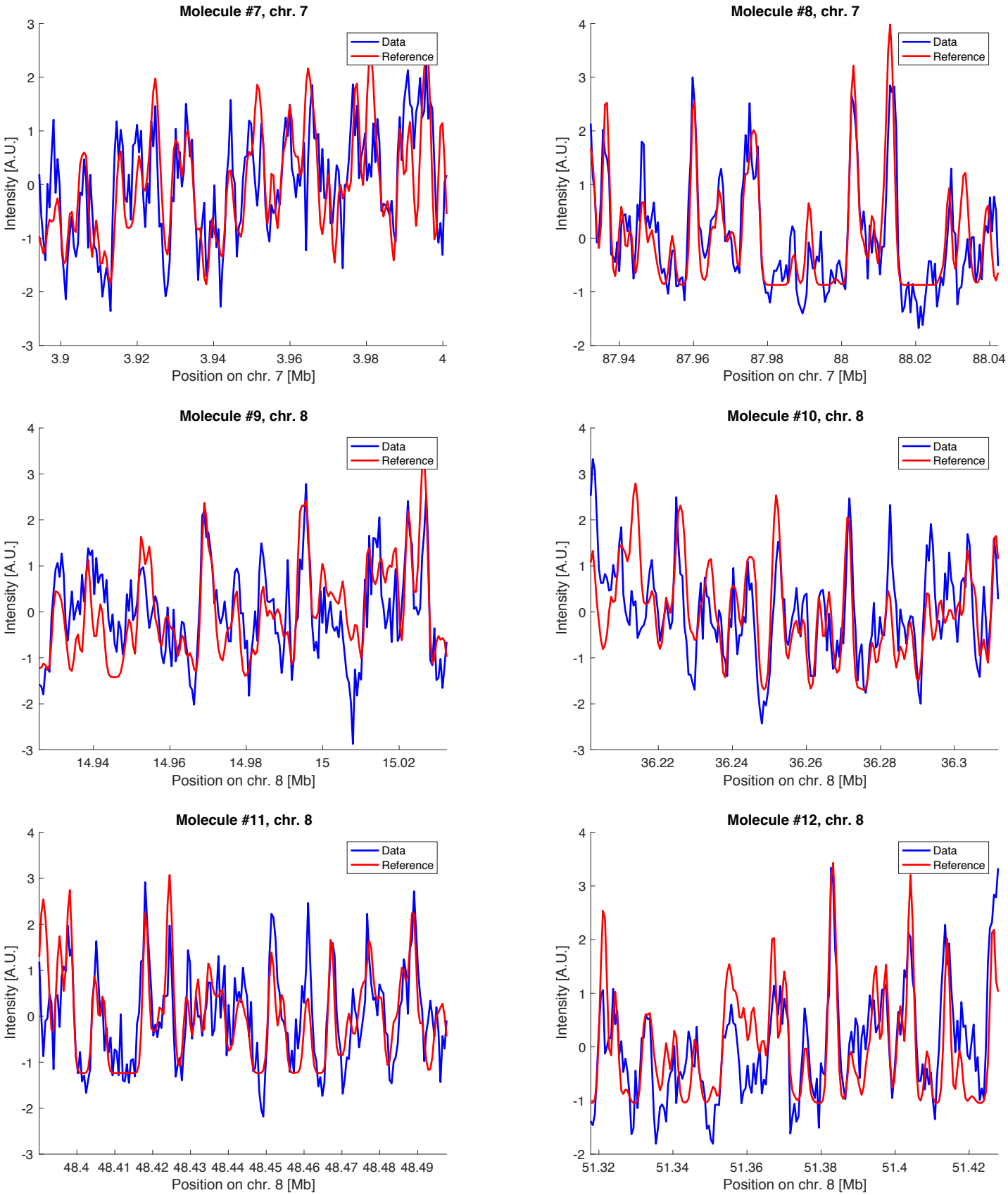


**Fig. S29.** Matches between D-R maps (pixel 900-1100) and positions on the computer-simulated whole-genome melting map for Molecules 7 to 12 (see also Fig. 3E in the main text).

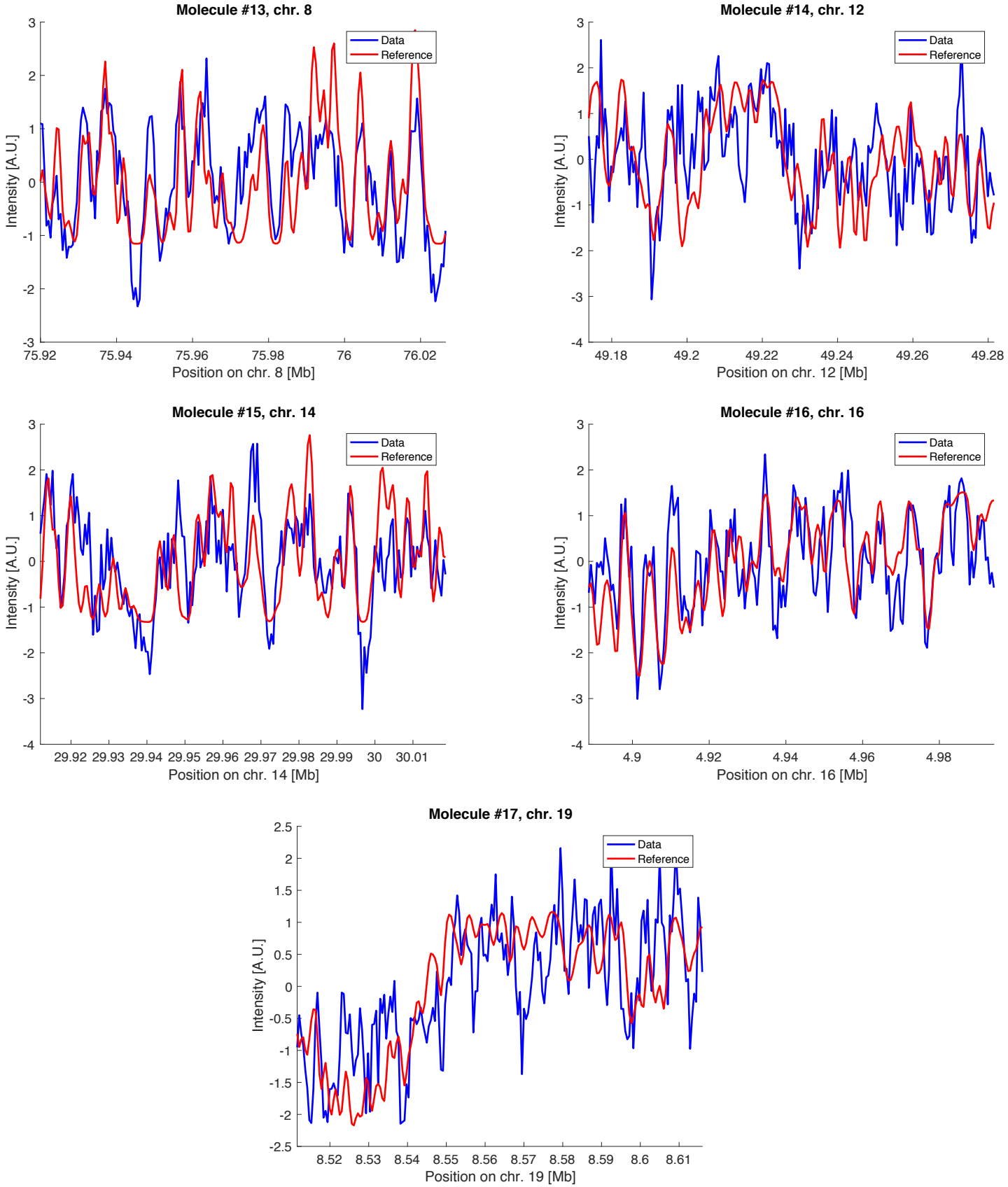


**Fig. S30.** Matches between D-R maps (pixel 900-1100) and positions on the computer-simulated whole-genome melting map for Molecules 13 to 17 (see also Fig. 3E in the main text).

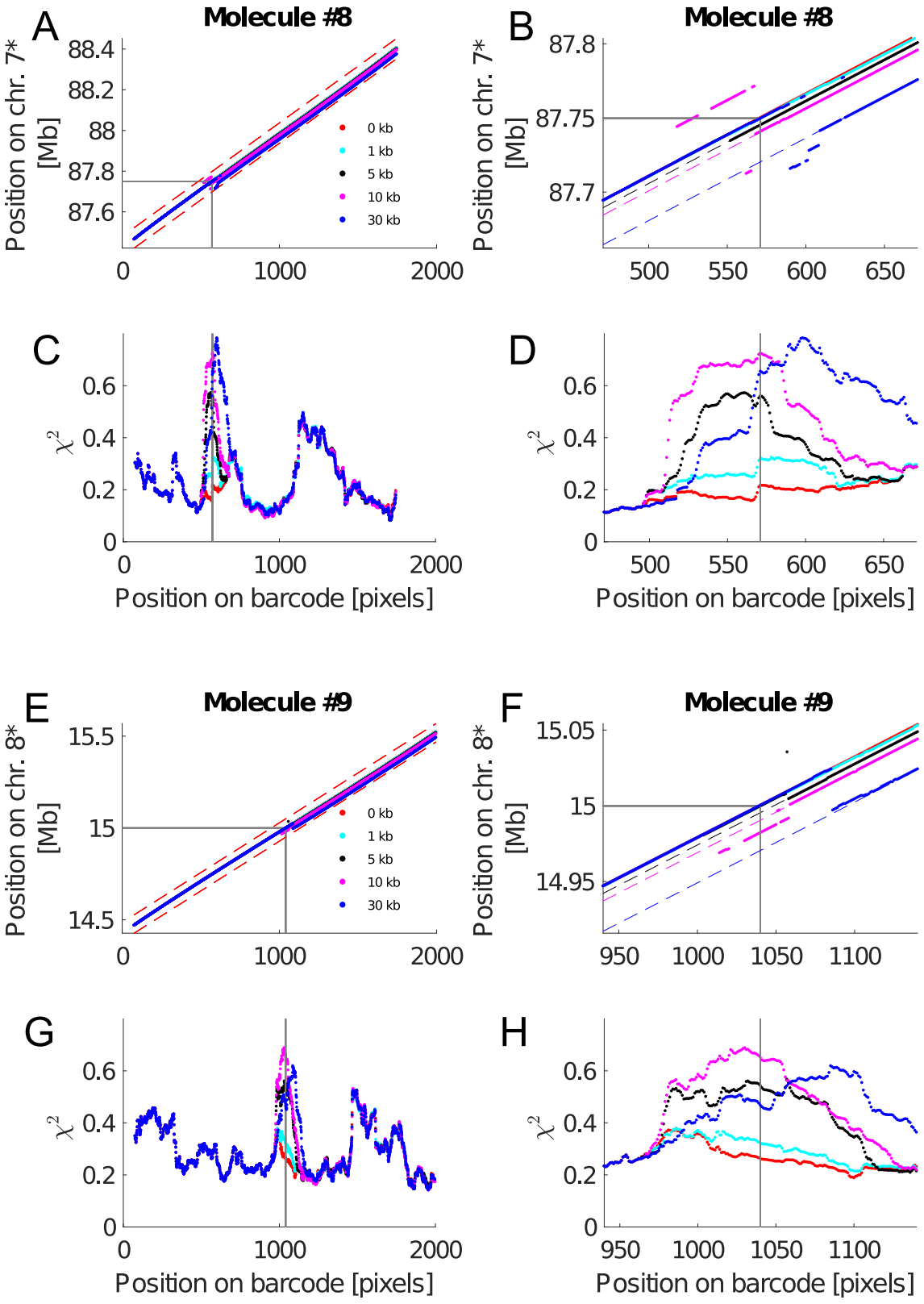


**Fig. S31.** Simulations of the detection of insertions in Molecule 8 and 9 (see Sec. 3A). Results of the restricted 'Sliding Window' analysis for the alignment of the D-R maps to the original reference genome ('0kb') and modified reference genomes with deleted sequences ('1 kb' - '30 kb'). A '*' after the chromosome number indicates that the reference genome has been altered. (A,E) Positions on the reference genomes versus the position on the D-R map. Thin, red dashed lines show the limits of the search range in the restricted 'Sliding Window' analysis. Full, gray lines mark the positions of the artificial inversions on the D-R map and on the reference genomes. (B,F) Zoom in on the data in Panels A and D. (C,G) Match scores ($\chi^2$-values normalized so $\chi^2$=0 denotes a perfect match, and a maximal mismatch would correspond to $\chi^2$=2) for the best match at each position. Vertical, gray lines mark the known position of the artificial inversion. (D,H) Zoom in on the data in Panels C and G.

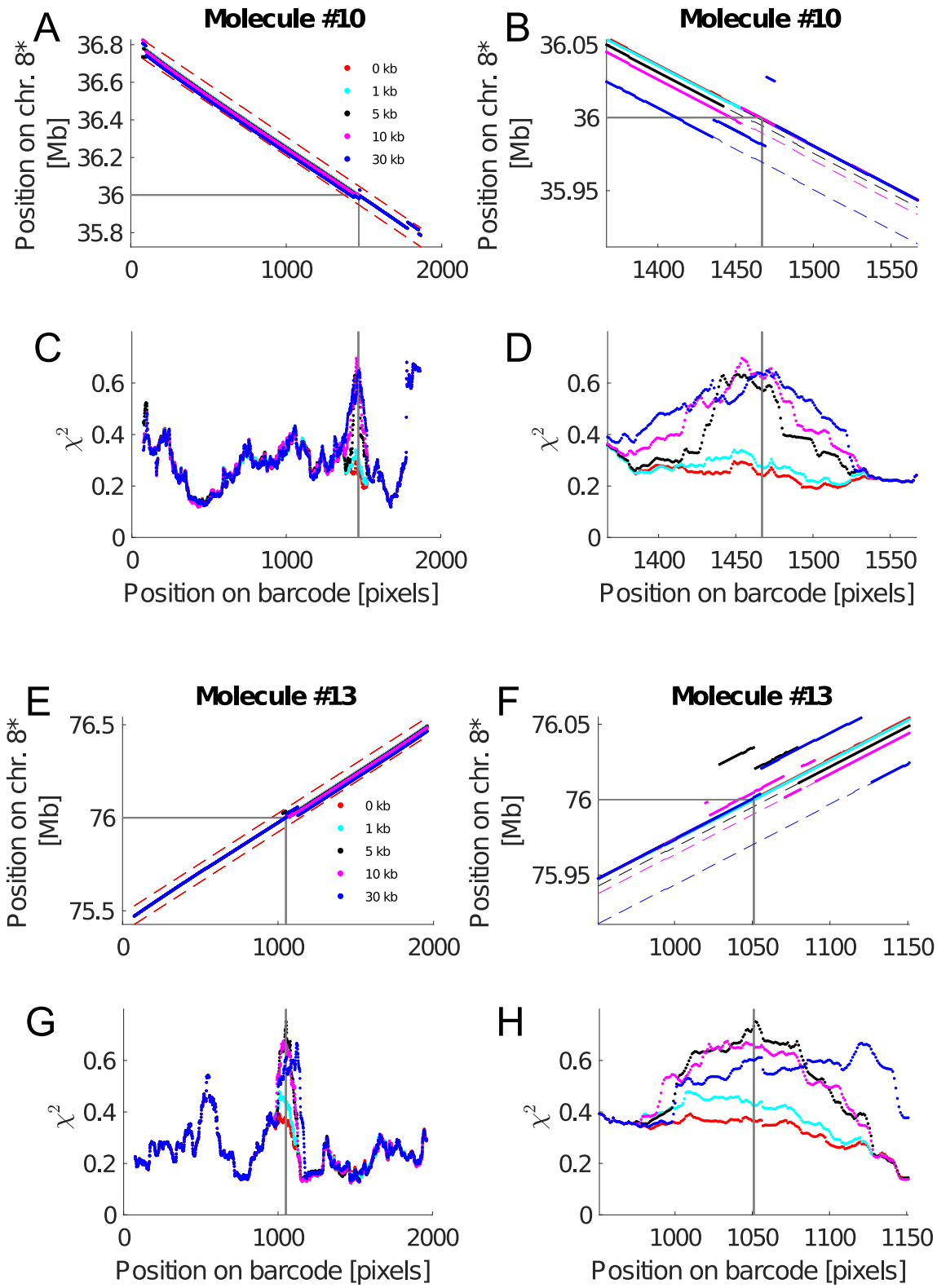


**Fig. S32.** Simulations of the detection of insertions in Molecule 10 and 13 (see Sec. 3A). Results of the restricted 'Sliding Window' analysis for the alignment of the D-R maps to the original reference genome ('0kb') and modified reference genomes with deleted sequences ('1 kb' - '30 kb'). A '*' after the chromosome number indicates that the reference genome has been altered. (A,E) Positions on the reference genomes versus the position on the D-R map. Thin, red dashed lines show the limits of the search range in the restricted 'Sliding Window' analysis. Full, gray lines mark the positions of the artificial inversions on the D-R map and on the reference genomes. (B,F) Zoom in on the data in Panels A and D. (C,G) Match scores ($\chi^2$-values normalized so $\chi^2$=0 denotes a perfect match, and a maximal mismatch would correspond to $\chi^2$=2) for the best match at each position. Vertical, gray lines mark the known position of the artificial inversion. (D,H) Zoom in on the data in Panels C and G.

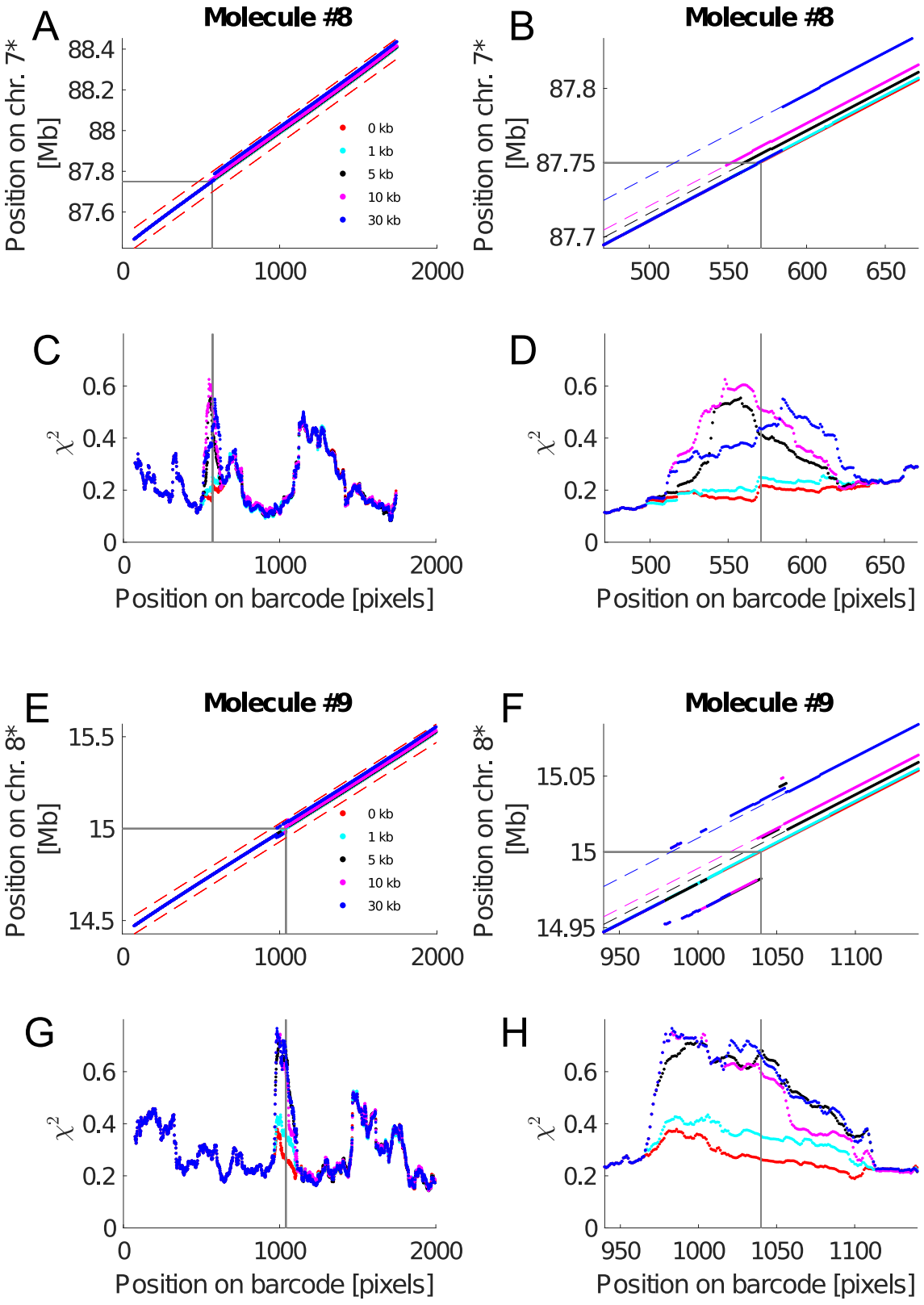


**Fig. S33.** Simulations of the detection of deletions in Molecule 8 and 9 by inserting sequences from the mouse genome in the reference genome (see Sec. 3B). Results of the restricted 'Sliding Window' analysis for the alignment of the D-R maps to the original reference genome ('0kb') and modified reference genomes with inserted sequences ('1 kb' - '30 kb'). A '*' after the chromosome number indicates that the reference genome has been altered. (A,E) Positions on the reference genomes versus the position on the D-R map. Thin, red dashed lines show the limits of the search range in the restricted 'Sliding Window' analysis. Full, gray lines mark the positions of the artificial inversions on the D-R map and on the reference genomes. (B,F) Zoom in on the data in Panels A and D. (C,G) Match scores ($\chi^2$-values normalized so $\chi^2$=0 denotes a perfect match, and a maximal mismatch would correspond to $\chi^2$=2) for the best match at each position. Vertical, gray lines mark the known position of the artificial inversion. (D,H) Zoom in on the data in Panels C and G.

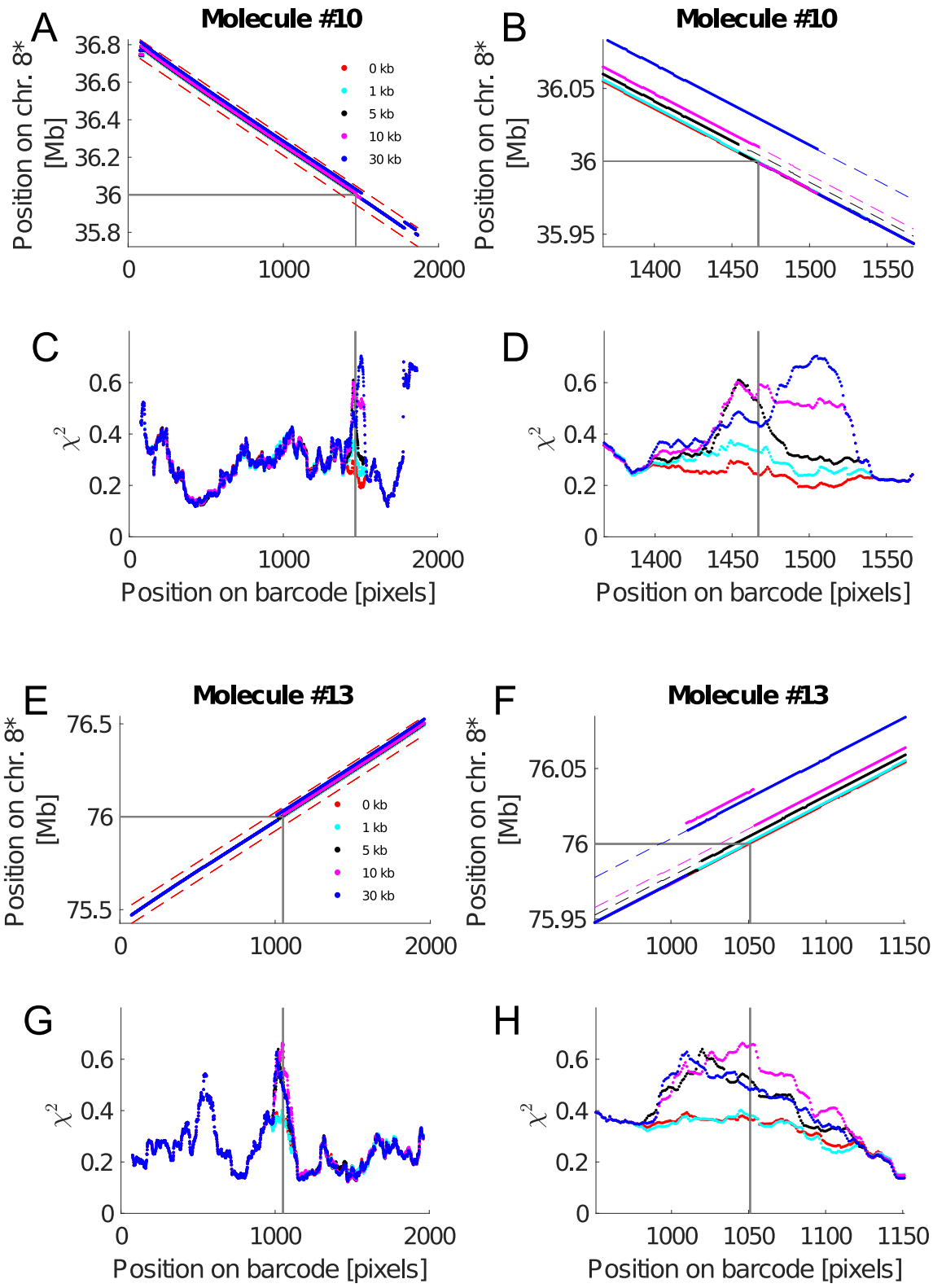


**Fig. S34.** Simulations of the detection of deletions in Molecule 10 and 13 by inserting sequences from the mouse genome in the reference genome (see Sec. 3B). Results of the restricted 'Sliding Window' analysis for the alignment of the D-R maps to the original reference genome ('0kb') and modified reference genomes with inserted sequences ('1 kb' - '30 kb'). A '*' after the chromosome number indicates that the reference genome has been altered. (A,E) Positions on the reference genomes versus the position on the D-R map. Thin, red dashed lines show the limits of the search range in the restricted 'Sliding Window' analysis. Full, gray lines mark the positions of the artificial inversions on the D-R map and on the reference genomes. (B,F) Zoom in on the data in Panels A and D. (C,G) Match scores ($\chi^2$-values normalized so $\chi^2$=0 denotes a perfect match, and a maximal mismatch would correspond to $\chi^2$=2) for the best match at each position. Vertical, gray lines mark the known position of the artificial inversion. (D,H) Zoom in on the data in Panels C and G.

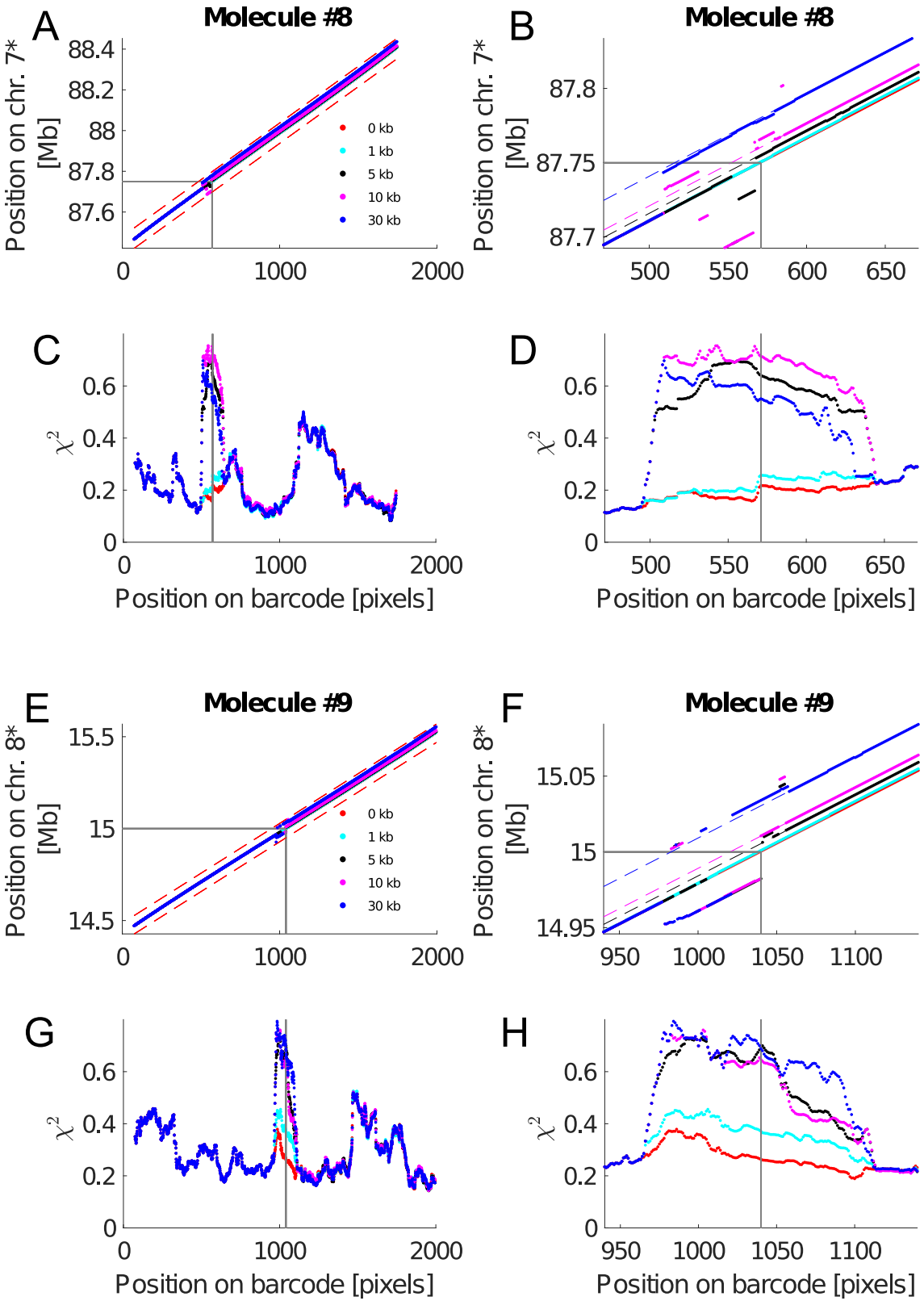


**Fig. S35.** Simulations of the detection of deletions in Molecule 8 and 9 by inserting sequences from the enterobacteria phage lambda genome in the reference genome (see Sec. 3B). Results of the restricted 'Sliding Window' analysis for the alignment of the D-R maps to the original reference genome ('0kb') and modified reference genomes with inserted sequences ('1 kb' - '30 kb'). A '*' after the chromosome number indicates that the reference genome has been altered. (A,E) Positions on the reference genomes versus the position on the D-R map. Thin, red dashed lines show the limits of the search range in the restricted 'Sliding Window' analysis. Full, gray lines mark the positions of the artificial inversions on the D-R map and on the reference genomes. (B,F) Zoom in on the data in Panels A and D. (C,G) Match scores ($\chi^2$-values normalized so $\chi^2$=0 denotes a perfect match, and a maximal mismatch would correspond to $\chi^2$=2) for the best match at each position. Vertical, gray lines mark the known position of the artificial inversion. (D,H) Zoom in on the data in Panels C and G.

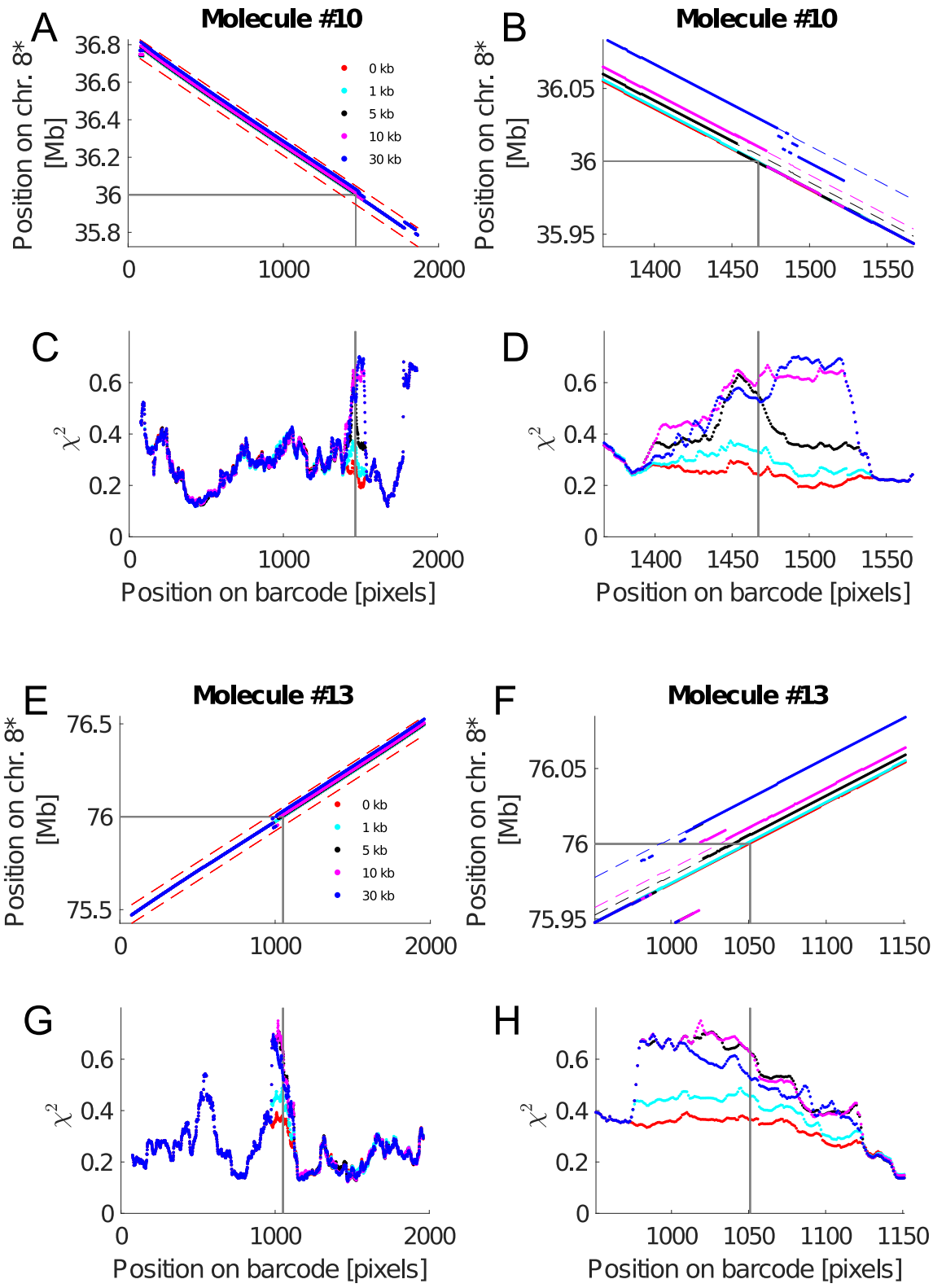


**Fig. S36.** Simulations of the detection of deletions in Molecule 10 and 13 by inserting sequences from the enterobacteria phage lambda genome in the reference genome (see Sec. 3B). Results of the restricted 'Sliding Window' analysis for the alignment of the D-R maps to the original reference genome ('0kb') and modified reference genomes with inserted sequences ('1 kb' - '30 kb'). A '*' after the chromosome number indicates that the reference genome has been altered. (A,E) Positions on the reference genomes versus the position on the D-R map. Thin, red dashed lines show the limits of the search range in the restricted 'Sliding Window' analysis. Full, gray lines mark the positions of the artificial inversions on the D-R map and on the reference genomes. (B,F) Zoom in on the data in Panels A and D. (C,G) Match scores ($\chi^2$-values normalized so $\chi^2$=0 denotes a perfect match, and a maximal mismatch would correspond to $\chi^2$=2) for the best match at each position. Vertical, gray lines mark the known position of the artificial inversion. (D,H) Zoom in on the data in Panels C and G.

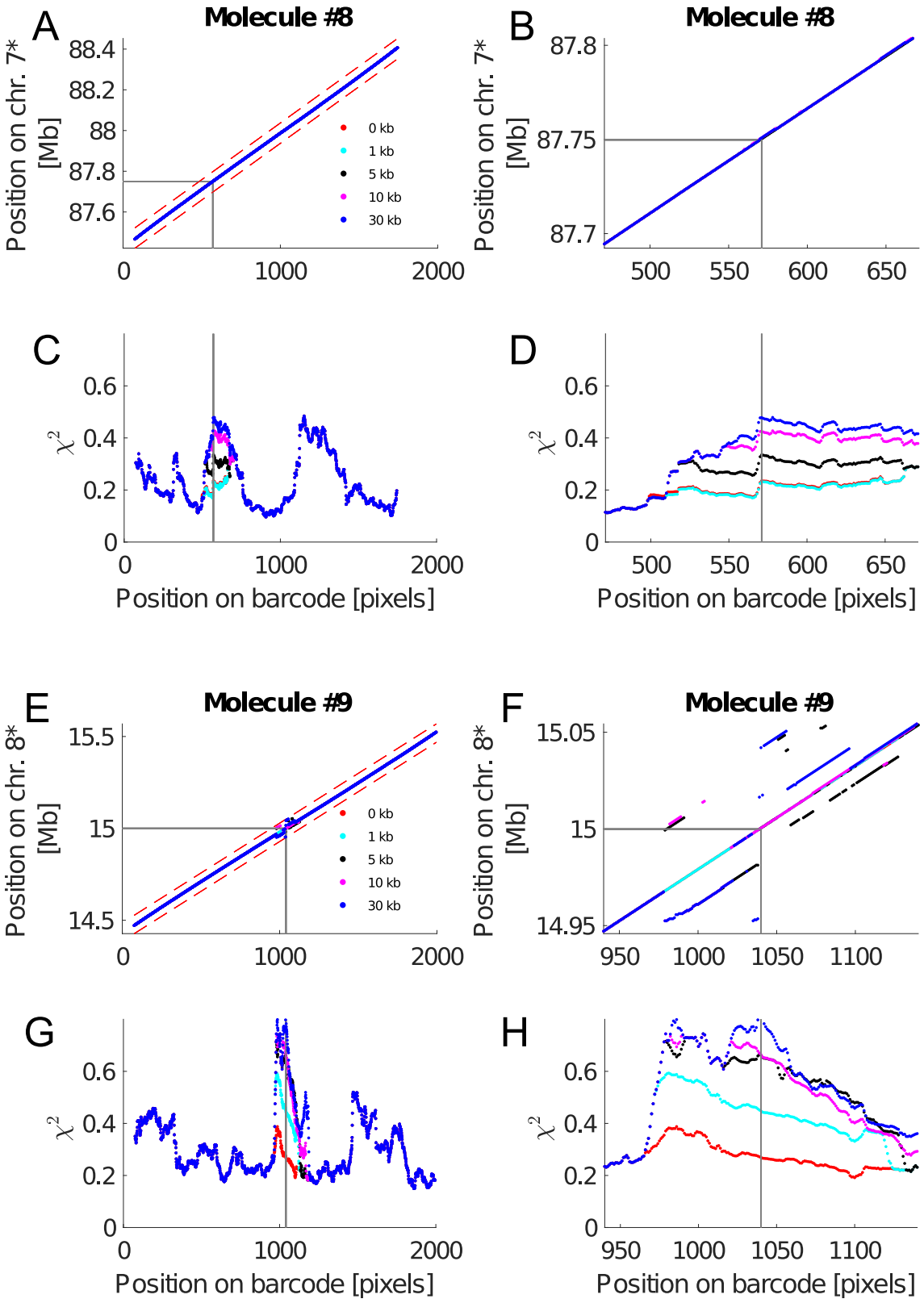


**Fig. S37.** Simulations of the detection of mismatches in Molecule 8 and 9 by replacing sequences in the reference genome with sequences of identical lengths from the mouse genome (see Sec. 3C). Results of the restricted 'Sliding Window' analysis for the alignment of the D-R maps to the original reference genome ('0kb') and modified reference genomes with replaced sequences ('1 kb' - '30 kb'). A '*' after the chromosome number indicates that the reference genome has been altered. (A,E) Positions on the reference genomes versus the position on the D-R map. Thin, red dashed lines show the limits of the search range in the restricted 'Sliding Window' analysis. Full, gray lines mark the positions of the artificial inversions on the D-R map and on the reference genomes. (B,F) Zoom in on the data in Panels A and D. (C,G) Match scores ($\chi^2$-values normalized so $\chi^2$=0 denotes a perfect match, and a maximal mismatch would correspond to $\chi^2$=2) for the best match at each position. Vertical, gray lines mark the known position of the artificial inversion. (D,H) Zoom in on the data in Panels C and G.

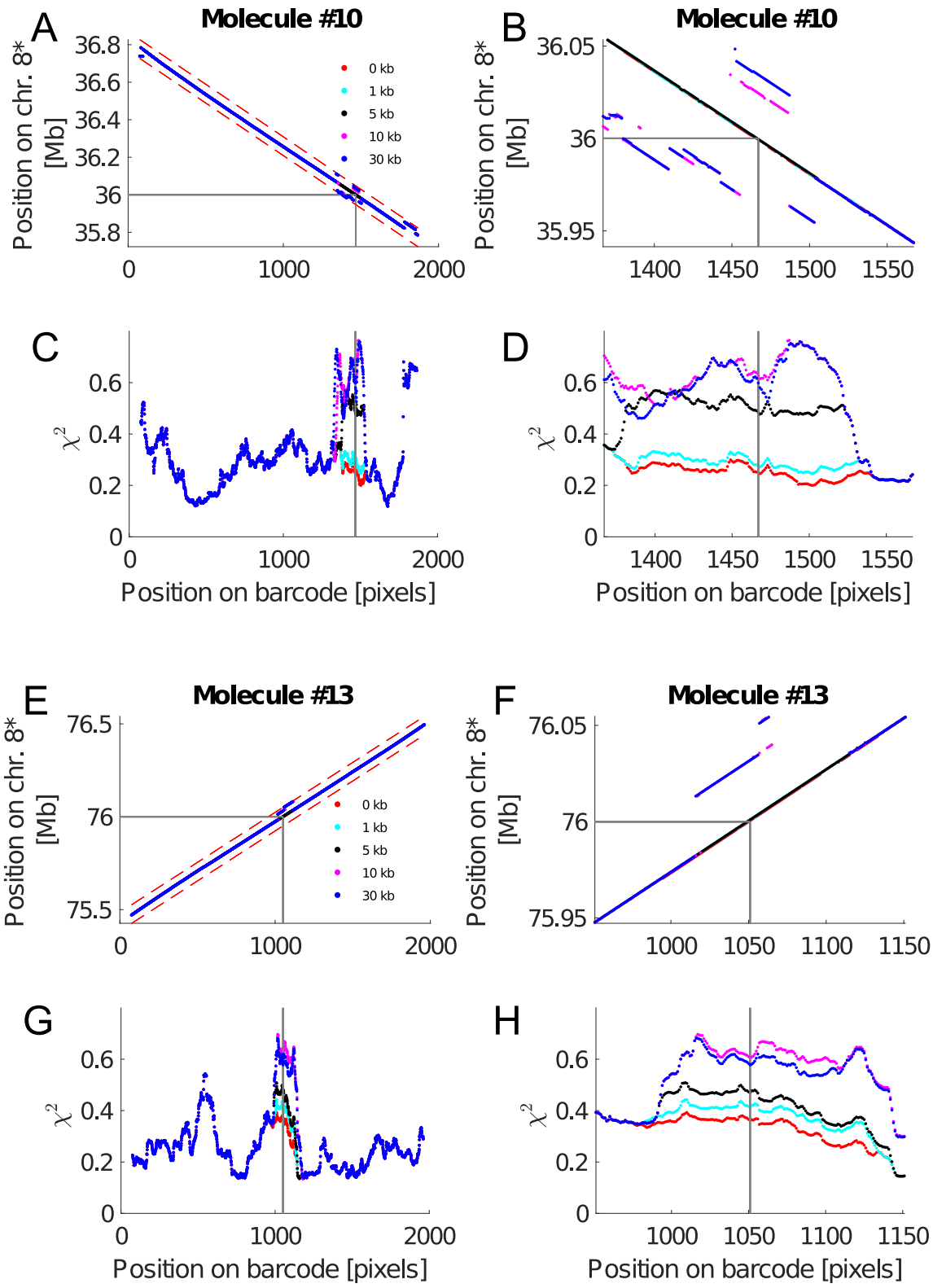


**Fig. S38.** Simulations of the detection of mismatches in Molecule 10 and 13 by replacing sequences in the reference genome with sequences of identical lengths from the mouse genome (see Sec. 3C). Results of the restricted 'Sliding Window' analysis for the alignment of the D-R maps to the original reference genome ('0kb') and modified reference genomes with replaced sequences ('1 kb' - '30 kb'). A '*' after the chromosome number indicates that the reference genome has been altered. (A,E) Positions on the reference genomes versus the position on the D-R map. Thin, red dashed lines show the limits of the search range in the restricted 'Sliding Window' analysis. Full, gray lines mark the positions of the artificial inversions on the D-R map and on the reference genomes. (B,F) Zoom in on the data in Panels A and D. (C,G) Match scores ($\chi^2$-values normalized so $\chi^2$=0 denotes a perfect match, and a maximal mismatch would correspond to $\chi^2$=2) for the best match at each position. Vertical, gray lines mark the known position of the artificial inversion. (D,H) Zoom in on the data in Panels C and G.

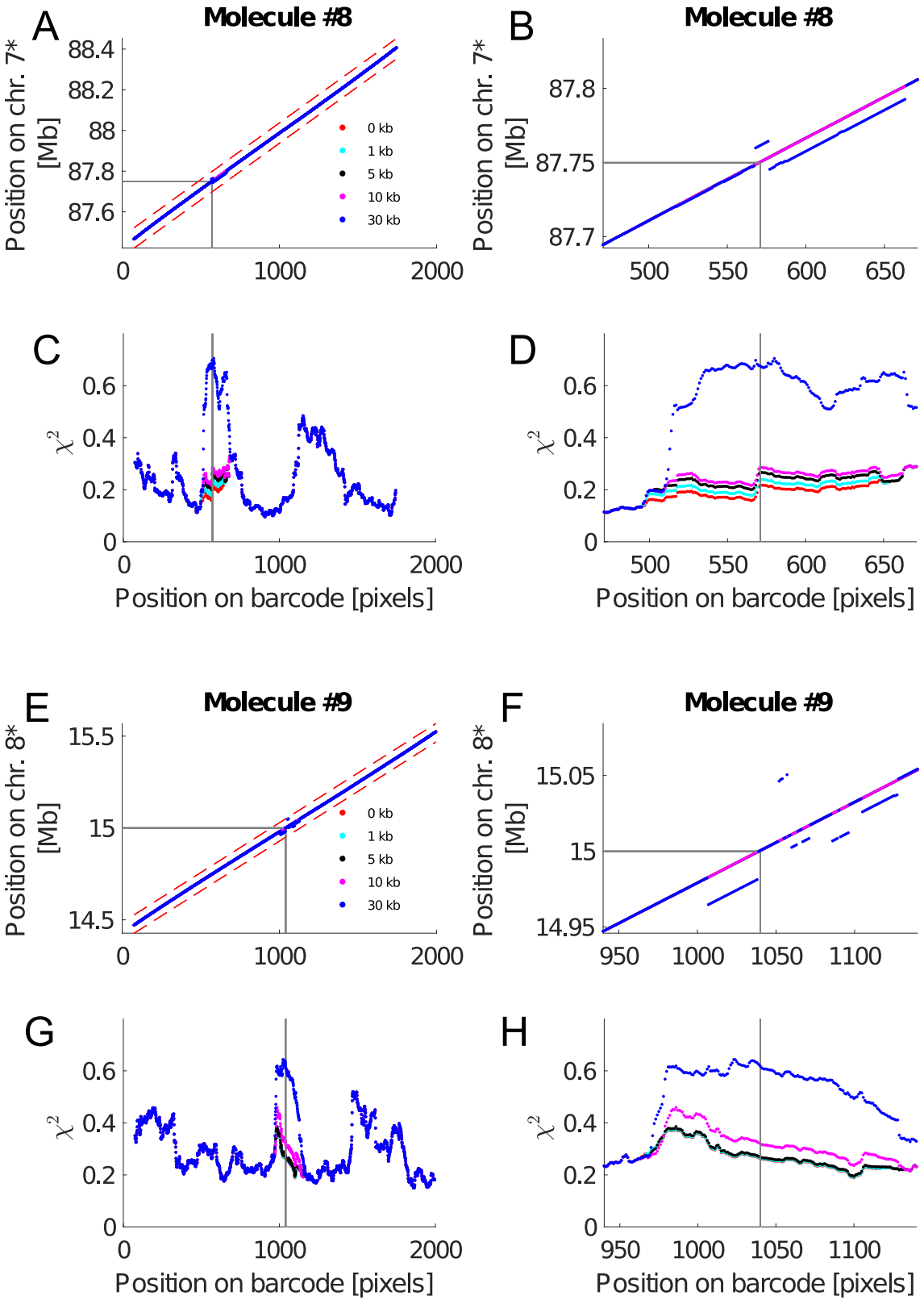


**Fig. S39.** Simulations of the detection of inversions in Molecule 8 and 9 by inverting sequences in the reference genome (see Sec. 3D). Results of the restricted 'Sliding Window' analysis for the alignment of the D-R maps to the original reference genome ('0kb') and modified reference genomes with inverted sequences ('1 kb' - '30 kb'). A '*' after the chromosome number indicates that the reference genome has been altered. (A,E) Positions on the reference genomes versus the position on the D-R map. Thin, red dashed lines show the limits of the search range in the restricted 'Sliding Window' analysis. Full, gray lines mark the positions of the artificial inversions on the D-R map and on the reference genomes. (B,F) Zoom in on the data in Panels A and D. (C,G) Match scores ($\chi^2$-values normalized so $\chi^2$=0 denotes a perfect match, and a maximal mismatch would correspond to $\chi^2$=2) for the best match at each position. Vertical, gray lines mark the known position of the artificial inversion. (D,H) Zoom in on the data in Panels C and G.

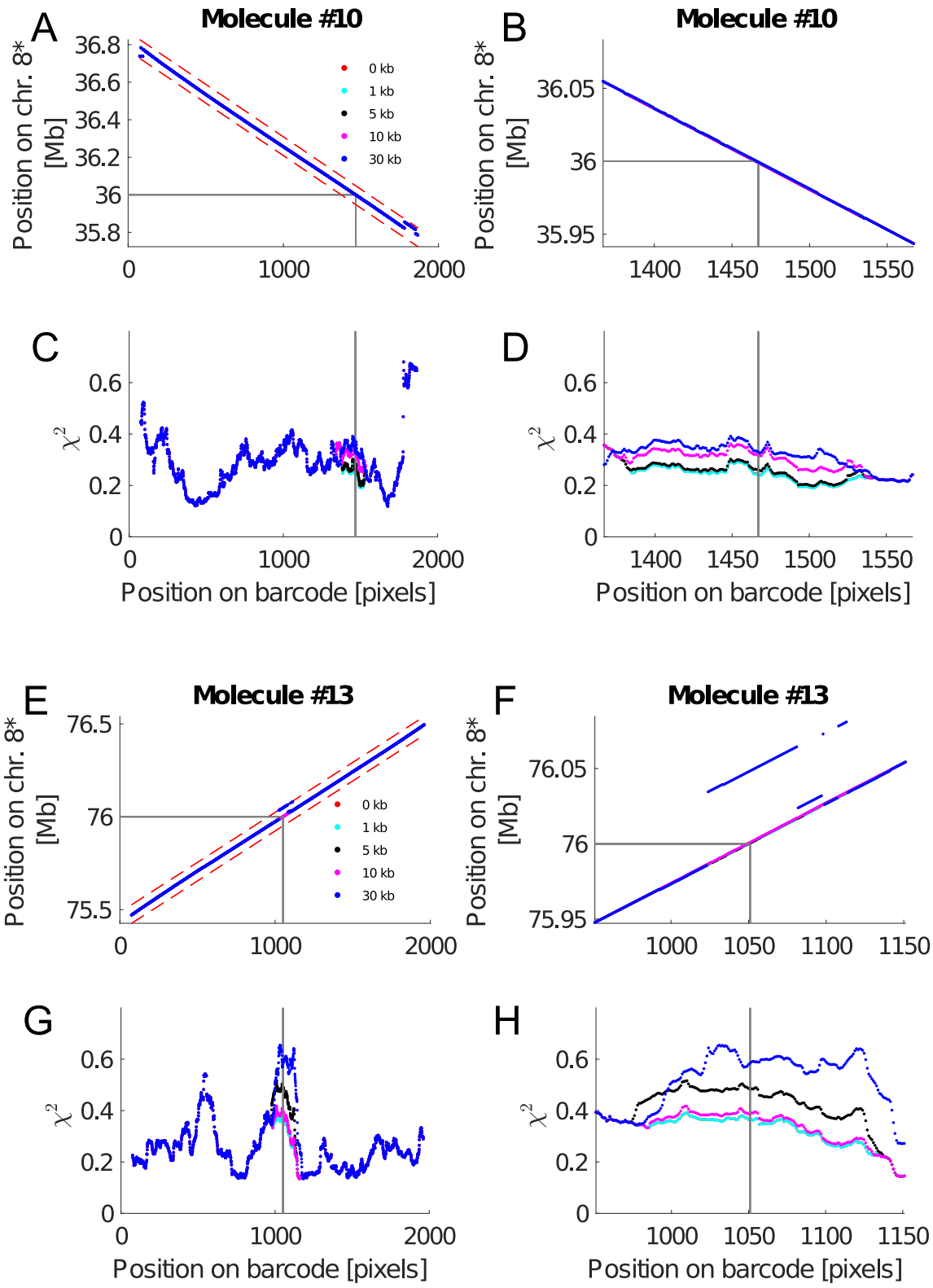


**Fig. S40.** Simulations of the detection of inversions in Molecule 10 and 13 by inverting sequences in the reference genome (see Sec. 3D). Results of the restricted 'Sliding Window' analysis for the alignment of the D-R maps to the original reference genome ('0kb') and modified reference genomes with inverted sequences ('1 kb' - '30 kb'). A '*' after the chromosome number indicates that the reference genome has been altered. (A,E) Positions on the reference genomes versus the position on the D-R map. Thin, red dashed lines show the limits of the search range in the restricted 'Sliding Window' analysis. Full, gray lines mark the positions of the artificial inversions on the D-R map and on the reference genomes. (B,F) Zoom in on the data in Panels A and D. (C,G) Match scores ($\chi^2$-values normalized so $\chi^2$=0 denotes a perfect match, and a maximal mismatch would correspond to $\chi^2$=2) for the best match at each position. Vertical, gray lines mark the known position of the artificial inversion. (D,H) Zoom in on the data in Panels C and G.

## References


1. Tegenfeldt JO, et al. (2004) The dynamics of genomic-length DNA molecules in 100-nm channels. *Proc Natl Acad Sci USA* 101(30):10979–10983.
2. Reisner W, et al. (2010) Single-molecule denaturation mapping of DNA in nanofluidic channels. *Proc Natl Acad Sci USA* 107(30):13294–13299.
3. Marie R, et al. (2013) Integrated view of genome structure and sequence of a single DNA molecule in a nanofluidic device. *Proc Natl Acad Sci USA* 110(13):4893–4898.
4. Østergaard PF, et al. (2015) Optical mapping of single-molecule human DNA in disposable, mass-produced all-polymer devices. *J Micromech Microeng* 25:105002.
5. Tostesen E, Liu F, Jenssen TK, Hovig E (2003) Speed-up of DNA melting algorithm with complete nearest neighbor properties. *Biopolymers* 70(3):364–376.
6. Liu F, et al. (2007) The human genomic melting map. *PLoS Comput Biol* 3(5):874–886.
7. Veal CD, et al. (2012) A mechanistic basis for amplification differences between samples and between genome regions. *BMC Genomics* 13(1):455.