Supplementary information for:

Multi-color Super-resolution DNA Imaging for Genetic Analysis

Murat Baday[1], Aaron Cravens[2], Alex Hastie[5], HyeongJun Kim[2], Deren E. Kudeki[3], Pui-Yan Kwok[6], Ming Xiao[5,7], Paul R. Selvin[1,2,4,7]

[1]Center for Biophysics and Computational Biology, and [2]Physics Dept, [3]Computer Science, and [4]Center for Physics of the Living Cell, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, United States
[5]BioNano Genomics Inc., San Diego, CA 92121, United States
[6]Institute of Human Genetics, University of California, San Francisco, San Francisco, California, United States
[7]Present address: Drexel University, Philadelphia, PA, United States

[8]To whom correspondence may be addressed:
P.R.S: Telephone: (217) 244-3371
Fax: (217) 244-7559
selvin@illinois.edu
M.X: Telephone: (267)499-2021
Fax: (215) 966-6001
ming.xiao@drexel.edu

SUPPLEMENTARY TEXT

Preparation of the glass coverslips and DNA mounting: Glass surface were functionalized with two different polyelectrolytes according to earlier studies.[1,2] 22mm × 30 mm coverslips were sonicated first in acetone for 30 min then in 1M KOH for 30 min. Between sonication steps, coverslips were rinsed with MilliQ water. After final drying with Nitrogen, coverslips were kept under Argon plasma cleaner for 5 minutes. Polymers Poly(acrylic acid) (PAcr) and Poly(allylamine) (PAll) (Sigma-Aldrich, St Louis, MO) PAII and PAcr were dissolved at 2 mg/mL in desterilized water and filtered through 0.22 nm filter. Cleaned coverslips were incubated in positive (PAII), negative (PAcr) and positive (PAII) polyelectrolytes consecutively about 30 minutes on shaker 150 rpm at 35°C. Then, coated coverslips were kept in high purity water at room temperature.

Coverslips were placed on glass slide (1"x3") and pipet 4 μl DNA on the edge. DNA concentration is 10 pM and has oxygen scavenger system made of T50 buffer (10 mM Tris, pH 8.0, 50 mM NaCl) with PCD, PCA[3] and Trolox[4] (Sigma-Aldrich, St Louis, MO). The DNA was stretched on the coverslip surface by capillary action which causes strong flow. Coverslips were sealed with nail polish to prevent drying solution. DNA backbone is imaged with YOYO intercalating dye at 300:1 base-pair to dye ratio.

Microscopy: Short DNA and Lambda DNA samples(samples in Figures 1a, 2a, 3a) were imaged by using total internal reflection fluorescence microscope (TIRF) which is built on an Olympus

IX-71 microscope (Olympus America Inc) with PlanApo objective (Olympus 100x1.45 NA, oil). While imaging test-DNA in Figure 2a for combined SHREC and SHRImP method, the Cy3 labels were excited with 532 nm laser (Crysta Laser ,Reno, Nv) and Cy5 is excited by 633 nm laser (Crysta/Melles Griot, Helium-Neon). DualView Apparatus (MSMI-DV-CC, Optical Insights, AZ) was used to split emission into two-color channels. Lambda DNA backbone stained with YOYO-1 was excited using 488 nm blue laser (Melles Griot, Argon-Ion). All laser lines were set to have TIRF and entered the microscopy from same port. The emitted photons were collected through triple band filter cube which has emission filter z488/532/635m and polychroic mirror z488/532/633rpc (both from Chroma Technology). The image was recorded by a back-illuminated, frame-transfer charge-coupled device (CCD) detector IxonEM (Andor Technology, Belfast, North Ireland). All DNA samples were imaged in 0.5 s frame rate. Nikon Eclipse Ti inverted microscope system was used to image the BAC DNA. Eclipse Ti has a built-in multi-color laser TIRF set up in which alternating lasers are controlled by AOTF. By using custom-made macro code, images were taken with 488, 532 and 641 nm lines sequentially, and then stage moved in 50 micron intervals spanning several regions in the same sample. Same filter sets and CCD camera is used to image BAC DNA as well.

DNA sample preparation: While preparing SHRImP test DNA, two PCR primers, one labeled at the 5' end with cy3 and the other one phosphorylated at the 5' end were used. The DNA was amplified, creating a product with one 5' end labeled with cy3 and the other end phosphorylated. Single- stranded DNA molecules were then produced with the 5' end labeled with Cy3 by using lambda exonuclease enzyme to remove one strand of the PCR product. Two oligos, both with Cy3 at their 5' end, were hybridized at different locations on the single strand, resulting in a DNA labeled with fluorophores with either 94 bp or 266 bp apart. Then, by filling in the gaps between oligos using a polymerase enzyme, a double stranded DNA with cy3 fluorophores at specific locations was afforded. The SHREC-SHRImP DNA sample was prepared in a similar way to the SHRImP sample explained in the paper, replacing one Cy3 labeled primer with Cy5 labeled primer

We obtained BAC clone in LB slabs from the BACPAC Resource Center at the Children's Hospital Oakland Research Institute (http://bacpac.chori.org/) from the BAC library CHORI-501. BAC DNA sample (cho501-1H2, GenBank Al662781.4) used in the study were prepared using Qiagen's Large-Construct Kit. The DNA sample was quantified using Nanodrop 1000 (Thermal Fisher Scientific) and their quality assessed using pulsed-field gel electrophoresis. One milligram BAC DNA was linearized with 2 U of NotI and nicked with 0.5 U nicking endonuclease Nt.BbVcI and Nt.BsmI (New England BioLabs, NEB) at 37 ^{o}C for 2 hours in NEB Buffer 3. The resultant DNA fragments were labeled with 25 nM Cy3 acyclo-dUTP and Cy5 acyclo-dCTP (Perkin Elmer) and Vent (exo-) (NEB) for 1 hour at 72 ^{o}C. The backbone of above fluorescently tagged DNA (5 ng/uL) was stained with YOYO-1 (3 nM; Invitrogen). Lambda DNA sample was nicked with Nt.BbVcI and labeled with 25 nM Cy3 acyclo-dUTP.

Image analysis: Cy3-Cy3-Cy3 labeled DNA samples (figure 1a) were excited by 532 nm laser. Double and photobleaching steps were searched to find SHRImP spots. After locating SHRImP qualified spots, photobleaching frame number was entered manually. Distances between dyes are calculated with IDL code written according to Gordon *et al.*[5] Cy5-Cy3-Cy3 DNA samples (figure 2a) are imaged with addition of DualView apparatus to separate Cy3 and Cy5 emissions.

Each spot in green channel were searched for double photobleaching steps. Distances between Cy3-Cy3 in this sample were calculated with the SHRImP code. Before and after running each experiment, nanoholes as fiduciary markers were used to get mapping function between two channels for SHREC calculations. Nanoholes were 100 nm in diameter and 1.5 micron apart from each other in both x and y direction. Mapping functions in MATLAB were used to register and map channels[6,7]. When analyzing SHREC data, mapping functions with mapping error less than 5 nm were used to calculate distances between Cy3-Cy5 dyes.

Experiments with lambda DNA were done by using 488 nm and 532 nm excitations. Backbone was labeled by YOYO-1 stain and restriction sites were labeled with Tamra fluorophore. After overlapping YOYO-1 labeled backbone and Tamra labeled restriction sites channels in ImageJ, spots on lambda DNA were picked by Point Picker Plug-in. For each lambda DNA, distances between dyes were calculated and searched for double photobleaching spots to identify qualified SHRImP spots. Location of each SHRImP spot were recognized according the other non-SHRImP dyes on the same DNA and its distance to DNA ends. Non-SHRImP and SHRImP distance were all calculated using custom made code in IDL.

BAC DNA with YOYO-1, Tamra and Cy5 channels were overlaid by using ImageJ. Later on, point picker (imageJ plug-in) was used to pick spots in Tamra and Cy5 channel which are on DNA backbone. By using IDL custom code, each picked spots were checked individually and filtered with certain FIONA error. Picked spots were checked to see whether they are SHRImP spot or not by looking at count trace over time. Distance between each spot were calculated and results were printed in an excel sheet. Data in each column in an excel sheet represented consecutive distances between neighboring spots for every DNA fragment. Search algorithm (which is developed in Mathlab) was run to match the consecutive distance list of each DNA fragment on the actual DNA reference list. If the DNA fragment were located on full BAC DNA with the search algorithm, new positions of restriction sites were assigned to their corresponding locations on full BCA DNA. If DNA is overstretched or under-stretched and there is any non-specific label on the list, restriction sites on the DNA fragment were excluded from all output list. Final output list from several experiments was plotted in a histogram in Origin Graphing Software.

Image processing for distance calculations on BAC DNA

The IDL code retrieves coordinates of spots from text file. These spots were previously picked by the point picker which is an ImageJ plug-in. After displaying DNA backbone image, retrieved spots are printed on their locations on the backbone image. Then, locations for beginning and end of stretched DNA are asked to be clicked on the image display. By using DNA end points, slopes and origins of each DNA segment are calculated. Program searches through a line on a DNA and listing corresponding spots to each DNA line. Later, 2D-Gaussian-fits are performed for all spots for all frames to get their coordinates over the time. Certain number of frames is averaged to remove the fluctuations effect in dyes' emissions. If the point is non-SHRImP spot, first 12 frames of the spot are averaged. If the point is SHRImP point, position is averaged from beginning frame till before photobleaching occurs. Averaging excludes off-values which may be due to bad photon collection in any single frame and bad fit results. After listing all the spots for each DNA segment, calculation of distances between each neighboring spots are made by using their averaged positions. While going through each spot on a DNA, some spots can be excluded by looking at its raw image's quality and trace over the time. During the spot check, SHRImP spots are detected by looking at their photobleaching behavior. If the spot is a SHRImP spot which has gone through double photobleaching, SHRImP calculation is made according to previous study[5] and the measured SHRImP distance is added to the DNA segment's distance list. Distances of each DNA segment is printed in a column of an excel sheet.

Search algorithm,

The best match for the measured data is found by moving the measured span across the template by small increments and checking how closely the points line up. The distance between two measured points is used to determine a range in which any template points will be considered possible matches with the second measured point. If a template point is within this range, the template point is added as a found point and the potential position of the next point after is updated in accordance to this new starting point, and the process is repeated to find the next point. This allows for a cascade of found points when the correct location is found as each successive point will be in the correct range. If no point is found within range of the current length, the next length is added and a new range is produced, and this continues until a point is found that falls within range or all the lengths have been used up. This allows for the program to skip over inaccurate data from erroneous measurements. If there are multiple points in the range the program chooses the one closest to the measured point they are being matched to. If matching points are found, the algorithm then goes back to check if the first measured point lines up with a template point using the distance between the first matching point and the first measured point to determine the range that a match can be in.

This scanning method is repeated multiple times, moving the starting point farther down the template until the entire length of the template has been tested. While the algorithm is traversing the template it remembers the best match it has found so far. A match is considered better than another if more points line up with the template. If two different matches have the same number of points that line up, the match where the points are closest to the template is chosen.

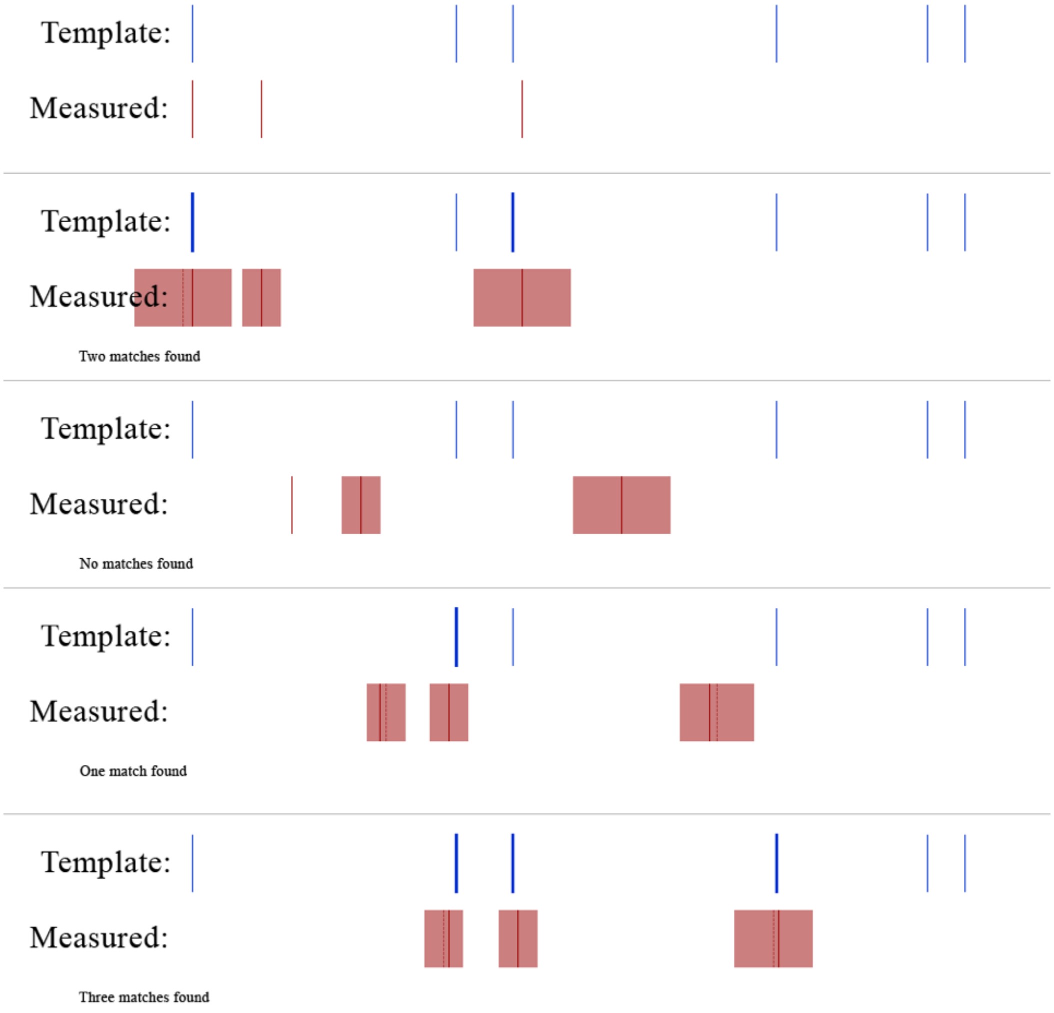

Figure S1. An illustration of the algorithm
Row 1: The template and the starting position of the measured data before a scan is run.
Row 2: The first scan. The distance between measured points one and two determines the range around point two that a match can be in. No match is found, so the distance between measured points one and three determines the range around point three that a match can be in. A template point is within the range, so it is recorded as a found point and shown as thick here. Point three is the last point, so the first point needs to be checked. The distance between points one and three is again used to determine the range that a match can be found in, but the center of this range is based on the template position of the first found point (point three) minus the distance between measured points three and one. This center is shown slightly to the left of measured point one as a lighter dotted line. A template point falls within this range and it is saved as the start of the found points. This point is also shown in bold.
Row 3: A later scan where the measured data has been shifted over to the right, having already gone through a number of previous scans. Once again the distance between measured points one and two are used to determine the range around point two to look for template points. No match is found, so the distance between measured points one and three determines the range around

point three that a match can be in. Once again, there is no template point that falls within that range. The program has hit the end of the measured points without finding any match, so it does not bother checking the starting point.

Row 4: Another later scan farther to the right determines the range around measured point two that matches can be found in based on the distance between measured points one and two. This time a match is found and recorded. It is shown in bold in the figure. The distance between measured points two and three is used to find the range for a match for measured point three, but the center of this range is determined by adding the distance between measured points two and three to the template point that was found to match measured point two. This means that measured point three is not in the center of its own range of potential matches. No matches are found for measured point three. Since the algorithm has reached the end of the measured points and has found at least one matching point, it goes back to check if there is a match for measured point one. The distance between measured point one and the first point with a match, measured point two, is used to determine the range that a match can be in, and the center of that range is determined by subtracting that distance from the location of the template point that matches measured point two. No point is found within this range.

Row 5: Once again the scan takes place even farther to the right. The range for matching points for measured point two is based on the distance between measured points one and two, and centered around measured point two. A template point is found within this range and recorded. It is shown as bold in the figure. The range for measured point three is determined by the distance between measured points two and three, and the center of that range is determined by adding that distance to the position of the template point that matches measured point two. A template point is found in this range, recorded, and shown as bold in the figure. Having reached the end of the measured data, and having recorded at least one match, the algorithm goes back to check if measured point one matches a template point. The range that is searched is determined by the distance between the first measured point and the first measured point that matches a template point, in this case measured point two. This distance is subtracted from the position of the template point that matches measured point two to determine the center of this range. A template is found in this range and recorded as the starting point of the matching points. This point is shown in bold. This is the best match found so far, with all three points lining up with the template and very little offset from any of their matching points. This sequence of matching points is saved as the best match found so far along with other relevant data.

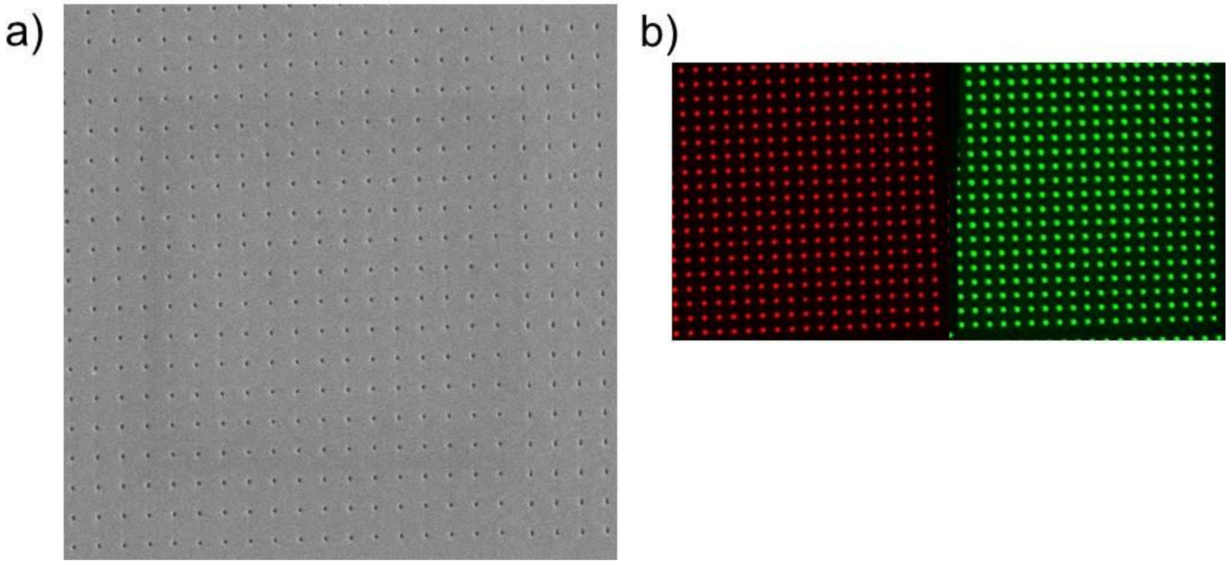

Figure S2. Nanoholes as fiduciary markers are made by using Focus Ion Beam (FIB). They are 100 nm in diameter and 1.5 μm apart from each other. First, thermal evaporator is used to deposit ~15 nm Cr and ~140 nm Ag (silver) on coverslips. Then, nanoholes image map which is drawn by CorelDRAW are patterned on coverslip by FIB. a) SEM image of nanoholes. b) TIRF image of nanoholes in green and red channels is taken by using Dualview apparatus to image SHREC sample in figure 2 of the paper. The white light was shed from the top, and only nanoholes are the region where the light can pass. And because of its broad wavelength and Dual-view optics, the transmitted light is separated according to wavelength.

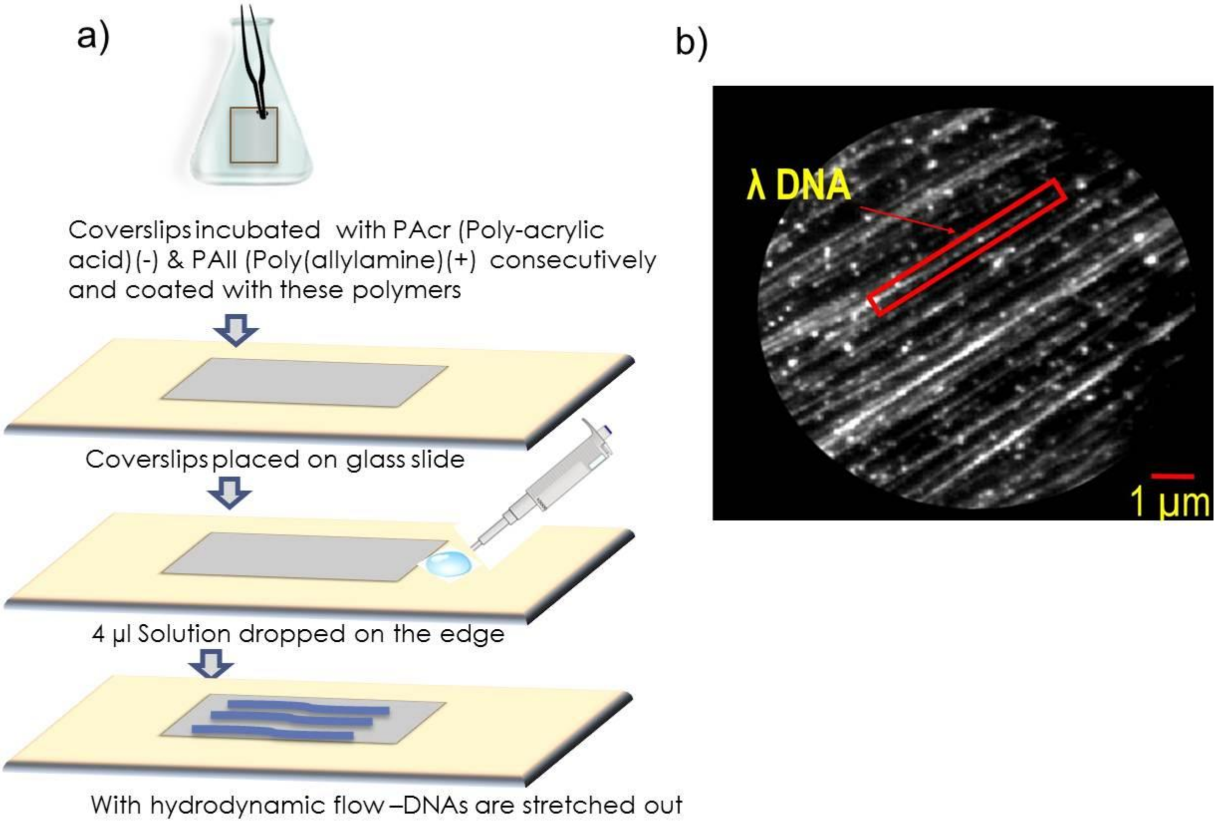
a)
Coverslips incubated with PAcr (Poly-acrylic acid) (-) & PAll (Poly(allylamine) (+) consecutively and coated with these polymers
Coverslips placed on glass slide
4 µl Solution dropped on the edge
With hydrodynamic flow –DNAs are stretched out
b)
λ DNA
1 µm

Figure S3. a) Illustration of stretching method b) Stretched lambda DNA image with YoYo on PAcr-PAII coated coverslip.

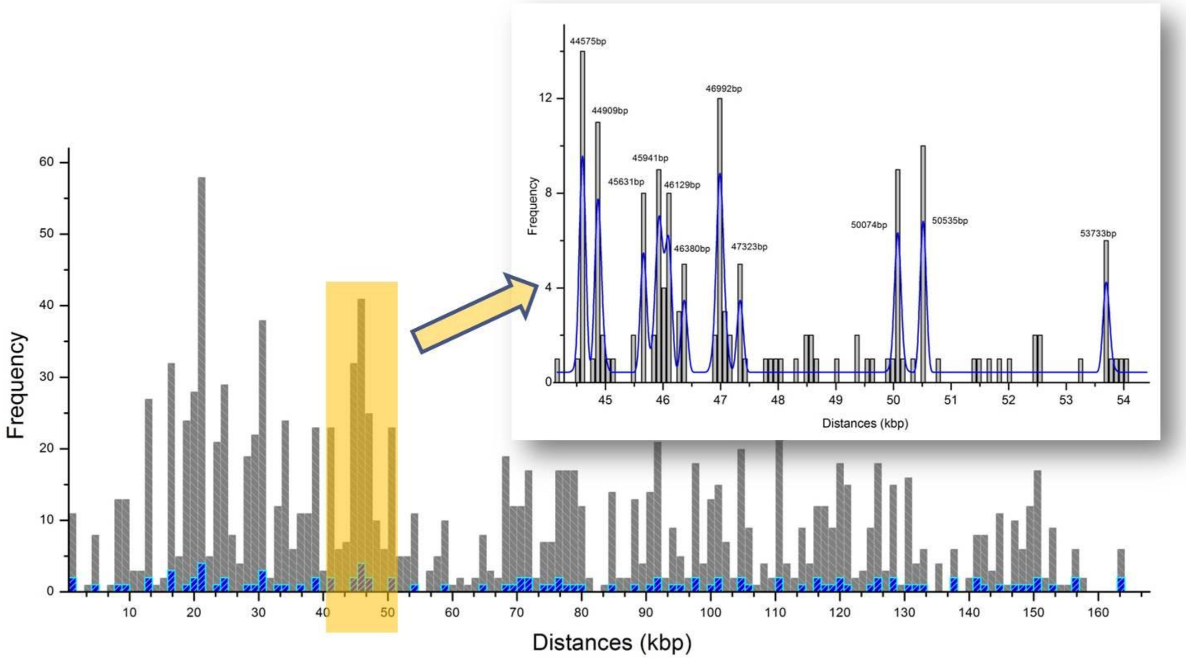

Figure S4. SHRImP/SHREC combined analysis of DNA mapping. Histogram ranges from 0-165 kbp (0-60 μm) which is the full BAC DNA length. Histogram's bin size is 200 nm. Main image has columns in two different colors. Dark gray is the histogram of experimental distances of restriction sites. Blue columns are the histogram of actual restriction list that includes both green and red labeled sites. Zoomed graph is the sub-region of main graph from 44 kbp to 55 kbp with binning size of 100bp (30 nm). Peaks are fit with multi-gaussian function to get mean values. All expected 11 restriction sites in that region is resolved as shown in zoomed histogram.

Expected	Experiment	Expected	Experiment	Expected	Experiment	Expected	Experiment
680	697	36474	36465	76895	76897	119948	119988
1092		38516	38496	78033	78043	120642	120007
4934	4933	39052	39082	78597	78573	125064	130683
7857	7858	40734	40744	79929	79897	126378	126378
9696	9733	41154	41130	84447		126438	131577
12953	12942	44579	44586	88710	88686	127661	127657
13111	13151	44892	44880	90791	90729	127928	127876
15921	15964	45652	45670	91563		130765	
16188	16266	45925	45946	91611		131581	
16323	20497	46059	46067	94113	94137	132842	
18648	18646	46324	46316	95268	95281	137529	
19782	19815	47009	47019	97195	97182	137545	
20482		47367	47372	98169	98131	141034	140969
20874	20884	50060	50088	101018		141204	141189
21185	21203	50507	50507	101476	101502	144114	144072
21275	21216	53723	53720	100239	100242	142420	
21622	21602	58383	58400	104866	104850	147481	
23780	23761	65194	65186	105036	105005	147759	147792
24826	24801	68517	68510	105996	105891	149787	149793
25096	25087	68893	68931	110148	110174	150168	150178
27917	27838	70246	70375	111001	126426	150546	150569
29187	29184	70454	70516	113998	113999	152863	152884
30200	30188	71221	71139	115926		156393	
30687	30665	71352	71338	115966		156591	156405
30804	30806	74253	74281	117839	117880	163235	163221
32533	32567	75632	75679	118725	118776	163642	163537
33832	33820	76443	76479	119837	119794		

Table S1. Restriction sites of Nb.BsmI and Nb.BbvcI enzymes on BAC DNA are located with 2D-SHRImP method. The table also includes expected distances of each site. Nb.BsmI and Nb.BbvcI are shown with green and red fonts respectively.

REFERENCES

(1) Xiao, M.; Phong, A.; Ha, C.; Chan, T.-F.; Cai, D.; Leung, L.; Wan, E.; Kistler, A. L.; DeRisi, J. L.; Selvin, P. R.; Kwok, P.-Y. *Nucl. Acids Res.* **2007**, *35*, e16–e16.

(2) Kartalov, E. P.; Unger, M. A.; Quake, S. R. *BioTechniques* **2003**, *34*, 505–510.

(3) Aitken, C. E.; Marshall, R. A.; Puglisi, J. D. *Biophys J* **2008**, *94*, 1826–1835.

(4) Rasnik, I.; McKinney, S. A.; Ha, T. *Nat Meth* **2006**, *3*, 891–893.

(5) Gordon, M. P.; Ha, T.; Selvin, P. R. *Proceedings of the National Academy of Sciences of the United States of America* **2004**, *101*, 6462 –6465.

(6) Goshtasby, A. *Image Vision Comput.* **1988**, *6*, 255–261.

(7) Goshtasby, A. *Pattern Recognition* **1986**, *19*, 459–466.