Cas9-Assisted Targeting of CHromosome segments (CATCH) for targeted nanopore sequencing and optical genome mapping

Supplementary information

Guide RNA preparation:

Guide RNAs were designed as previously described (1). We designed a 117 bp template DNA containing T7 promoter sequence, 20bp target sequence, and a constant sequence that contains crRNA and tracrRNA. The online CRISPR design tool (http://crispr.mit.edu/) was used for locating a 20 bp recognition sequence (seed sequence) that was followed by the NGG protospacer-adjacent motif (PAM) and had minimal identity to other parts of the genome. One replaceable primer containing the seed sequence and the T7 promoter, and two constant primers containing constant sequences were assembled by overlap PCR to generate the 117 bp template. Two changeable primers with different seed sequences were designed in order to cut the genome at two locations.

Primer sequences:

1. gRNA1- Changeable:
   5'-TAATACGACTCACTATAggtgcggatatctcggtagtGTTTTAGAGCTAGAAATAGCAA-3'

2. gRNA6- Changeable:
   5'-TAATACGACTCACTATAgtaaatctggggatggcgctGTTTTAGAGCTAGAAATAGCAA-3'

3. gR-F- Constant:
   5'-GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTC-3'

4. gR-R- Constant:
   5'-AAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACT-3'

Overlap PCR reaction to create the template for the gRNA:

	Volume for one reaction (µl)
Phusion high-fidelity DNA polymerase buffer, 5X (NEB)	5
dNTPs (10 mM stock concentration)	0.5
gR1 or gR6 (10 µM stock concentration)	1.25

gR-F (10 µM stock concentration)	0.625
gR-R (10 µM stock concentration)	1.25
Phusion high fidelity DNA polymerase (NEB)	0.25
Ultra-pure water	16.12
Final volume	25

Conditions for overlap PCR:

Initial denaturation	98 ºC	2 min
Amplification X 35	98 ºC	30 sec
	52 ºC	1 min
	72 ºC	30 sec
Final extension	72 ºC	10 min

The PCR product was mixed with glycerol and run in a 2% agarose gel with 1X TAE buffer. The 117 bp band was cut from the gel and cleaned with Freeze and squeeze (Bio-Rad) and then with the QIAquick PCR cleanup kit (Qiagen). In order to increase the DNA concentration, a second PCR reaction was performed with two primers complementary to the ends of the 117 bp template, and the purified PCR product which was used as a template.

Primer sequences:

1. 2nd PCR Forward gR1:
   5'-TAATACGACTCACTATAggtgc-3'
2. 2nd PCR Forward gR6:
   5'-TAATACGACTCACTATAgtaaatctgg-3'
3. 2nd PCR Reverse:
   5'-AAAAGCACCGACTCGGTG-3'

2nd PCR reaction:

	Volume for one reaction (µl)
Phusion high-fidelity DNA polymerase buffer, 5X (NEB)	5
dNTPs (10 mM stock concentration)	0.5
2nd PCR Forward gR1 **or** 2nd PCR Forward gR6 (10 µM stock concentration)	1.25
2nd PCR Reverse (10 µM stock concentration)	1.25
Phusion high-fidelity DNA polymerase (NEB)	0.25
Ultra-pure water	16.75

Final volume	25

Conditions for second PCR for 35 cycles:

Initial denaturation	98 ºC	2 min
Amplification X 35	98 ºC	30 sec
	52 ºC	1 min
	72 ºC	30 sec
Final extension	72 ºC	10 min

The product was purified with QIAquick PCR cleanup kit (Qiagen). Following this step, the typical DNA concentration was 100-200 ng/µl.

For conversion of template DNA to RNA, we performed an *in vitro* transcription (IVT) reaction using HiScribe T7 RNA synthesis kit (NEB) according to manufacturer's instructions. One µg of purified PCR product was used for each reaction, and samples were incubated overnight to optimize yield. Typically, two identical reactions were performed in parallel. gRNA was purified by phenol-chloroform purification and isopropanol precipitation. Phenol (pH 4.3) and chloroform were mixed 1:1 and 200 µl were added to each IVT reaction. The reactions were centrifuged at 14,000 g at 4 °C for 10 minutes. The upper phase was gently transferred to a 1.5 ml tube, and 22 µl of 3 M NaAc and 220 µl of isopropanol were added and mixed by inverting the tube several times. The tube was incubated at -20 °C for at least 30 minutes and centrifuged for 30 minutes at 14,000 g at 4 °C. The supernatant was carefully discarded, and the pellet was washed with 1 ml of 75% ethanol followed by 2 minutes of centrifugation at 14,000 g at 4 °C. The supernatant was discarded. The pellet was briefly allowed to dry and was resuspended in 100 µl RNase-free water. gRNA was stored at -80 °C until use (typical concentration: 1.5 µg/µl).

Plug preparation

E. coli K-12 MG1655 cells were grown in LB overnight, diluted the next day 1:300 in LB, and grown until O.D. of 0.6. Typically, 4 ml of culture (~1 x 10^9 cells) were used for preparation of five plugs. Cells were centrifuged at 10,000 g for 10 minutes, and pellet was washed twice with 1 ml PBS followed by centrifugation. After washes, the pellet was resuspended with 375 µl cell suspension buffer (CHEF mammalian DNA extraction kit, Bio-Rad) and incubated at 43 °C for at least 5 minutes. An aliquot of 225 µl of 2% low melting agarose (CleanCut agarose, Bio-Rad) was melted for 10 minutes at 70 °C followed by incubation at 43 °C for 5 minutes. The agarose was added to the resuspended cells (final concentration of 0.7% agarose), and mixed gently by pipetting up and down four times while avoiding bubble formation. The mixture was immediately cast into a plug mold (~95 µl per plug) with mixing after casting five plugs. Plugs were incubated at 4 °C for 30 minutes until solidified. Up to five plugs were incubated in a 50 ml conical tube with 125 µl freshly prepared lysosome (Sigma, stock concentration 10 mg/ml in

Tris-HCl, pH 8) in 1.5 ml lysozyme buffer (10 mM Tris-HCl, pH 8, 0.1 M NaCl, 1 mM EDTA, 5% Triton x-100), at 37 °C for 30 minutes. Lysozyme solution was then discarded through a sieve (Bio-Rad), and up to five plugs were incubated twice at 50 °C with 167 µl fresh Proteinase K solution (Qiagen) in 2.5 ml lysis buffer (BioNano Genomics) with occasional shaking. The first incubation was performed for 2 hours, followed by a second overnight incubation with fresh proteinase K and lysis buffer solution. Next, the Proteinase K solution was discarded, and the plugs were incubated with 50 µl RNase (Qiagen) in 2.5 ml TE (pH 8) for 1 hour at 37 ºC with occasional shaking. Plugs were tapped to the bottom of 50-ml conical tubes and washed four times with wash buffer (10 mM Tris, pH 8, 50 mM EDTA) by adding 10 ml through the sieve for each wash, and shaking for 15 minutes on a horizontal platform mixer at 180 rpm at room temperature. Plugs were then stored in wash buffer at 4 ºC or used for Cas9 digestion.

In-gel Cas9 digestion

Prior to Cas9 digestion, plugs were washed twice with Tris-HCl (pH 8) in 50 ml conical tubes as described above, followed by a single wash with 500 µl 1X Cas9 buffer (NEB) for 30 minutes. Plugs were then placed on a parafilm strip, each plug was cut to three equal pieces, and each piece was placed in a 1.5ml tube. gRNA and Cas9 were pre-assembled prior to fragmentation reaction by incubation of Cas9 (NEB) protein with the gRNAs for 30 minutes at room temperature. Following pre-assembly, 1/3 plug (~30 µl each) was added and samples were incubated with preassembled gRNA-Cas9 at 37 ºC for 2 hours.

Cas9/gRNA assembly reaction for 1/3 plug:

	Volume (µl)
Cas9 buffer, 10X (NEB)	4
gRNA1 (300 ng/µl stock concentration)	4
gRNA6 (300 ng/µl stock concentration)	4
Cas9 (NEB, 1 µM stock concentration)	0.5
Ultra-pure water	27.5

Finally, 3 µl Proteinase K (Sigma, 20 mg/ml stock concentration) was added to each tube, and samples were incubated at 43 °C for 3 hours in order to remove excess Cas9 bound to the DNA.

Separation by PFGE

A 250 ml solution of 0.9% sea plaque low melting agarose gel (Lonza) was prepared in 0.3X TBE. DNA isolated from 15 digested plugs (~12 µg DNA) was loaded onto the gel and run on a Rotaphor instrument (Biometra) for 24 hours in 0.25X TBE buffer. The bands corresponding to ~200 kbp were cut out of the gel. The agarose was melted at 70 °C for 5 min followed by

incubation at 43 °C for 10 minutes. Next, 2 µl GELase (Epicenter) was added for digestion of the agarose and incubated for 45 minutes at 43 °C. The DNA was purified from agarose by isopropanol precipitation (as described above for purification of gRNA). A total of 200 ng DNA was recovered from the gel; 125 ng were used for nanopore library construction and the rest for optical mapping.

Nanopore sequencing

For construction of the nanopore sequencing library, a low input expansion pack kit in combination with SQK-MAP007, R9 version kit (Oxford Nanopore Technologies) were used according to manufacturer's instructions with minor modifications. Approximately 125 ng of fragmented, purified DNA was used as starting material for the library. DNA was incubated with 1X AMPure XP beads for 20 minutes, washed, and eluted with 10 mM Tris-HCl (pH 8.5) for 30 minutes at 37 °C. R9.4 flowcells were run for 48 hours on MinION sequencing device, and reads were base-called using the Metrichor software. Fast5 files that either passed ('pass reads') or failed ('fail reads') base-calling quality metrics were converted to FASTA or FASTQ files using poretools (version 0.6.0) (2) using -- type best option. Alignment of reads to the *E. coli* reference genome was performed using BWA-MEM (version 0.7.12) (3) with –x ont2D parameters and coverage was calculated using the Galaxy wrapper for BEDTools genomecov (version 2.26.0, (4, 5)). The resulting BEDGraph files were loaded onto the UCSC genome browser (https://genome.ucsc.edu/) for visual evaluation of read coverage. *De novo* assembly was performed on 2D pass reads using Canu (version 1.4, (6)) with default parameters, specifying a genome size of 200 kbp. Long reads as well as assembled contig were converted into CMAP files imitating an optical map using the Knickers software (version 1.5.5) provided by BioNano Genomics by specifying the Nt.BspQI recognition sequence (Figure S1).

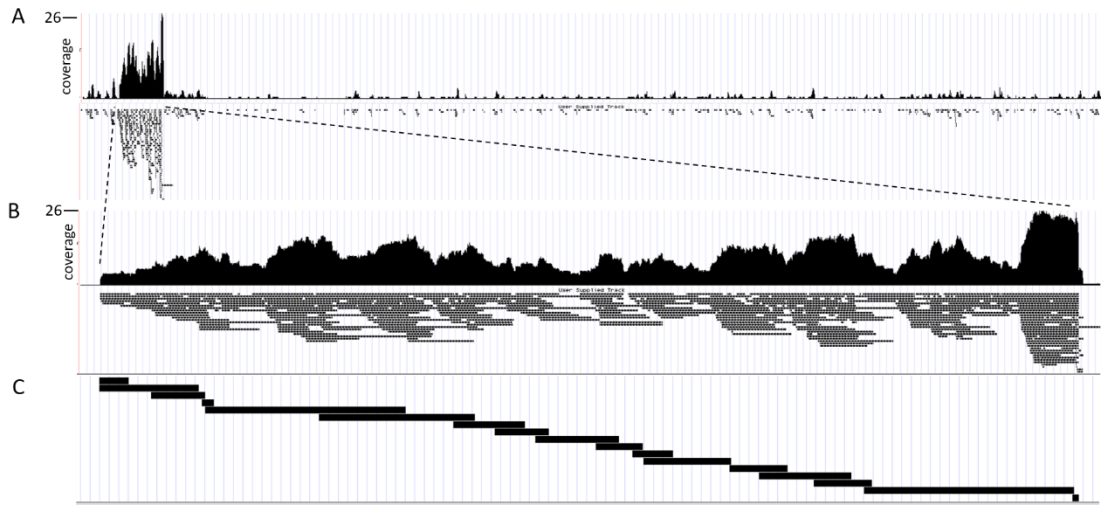

Figure S1. Nanopore sequencing coverage analysis of the 200 kbp target region. A. Coverage analysis (top) and read alignment (bottom) of the longest 10% reads to the entire *E. coli* genome. B. Coverage analysis (top) and read alignment (bottom) of the longest 10% reads to

the 200 kbp target sequence. C. Alignment of 17 reads that yielded a complete coverage of the 200 kbp target region. Results were visualized by the UCSC genome browser.

The bam2R function from the R package deep SNV was used to count the number of times each nucleotide appeared in each reference position, based on the alignment of both pass reads and all reads to the 200 kbp target reference from the *E. coli* genome. The correct ratio in each position was determined by dividing the number of reads with a correct base call compared to the reference sequence by the total number of reads aligned to the same position (Figure S2).

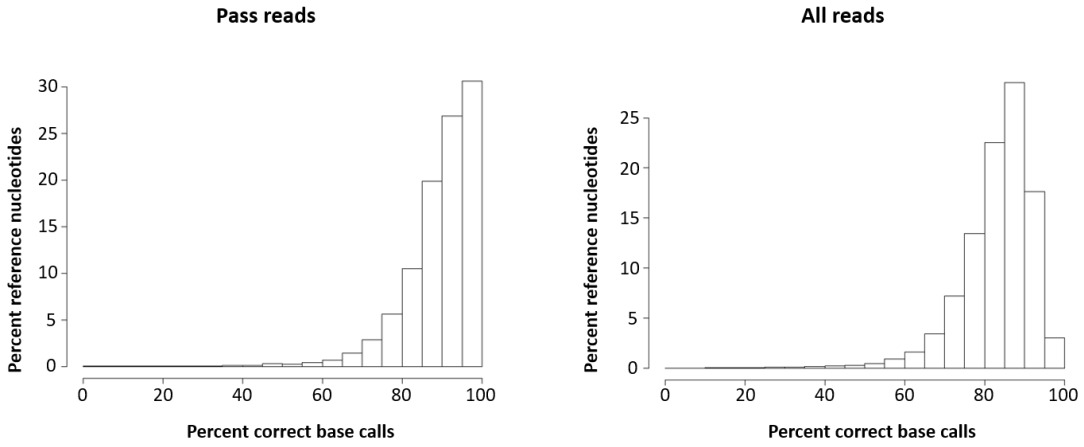

Figure S2. Histogram of the percent of reads base-called correctly compared to the reference for each position along the 200 kbp target region.

Optical mapping sample preparation, imaging, and analysis

For genetic labeling of the 200 kbp fragment, we performed nicking with the endonuclease Nt.BspQI (NEB) at 37 °C for 2 hours in a 10 µl reaction volume. DNA was added last in order to maintain its integrity and minimize fragmentation. In the following steps (labeling, ligation and staining), all the components of the reaction were mixed in separate tubes, vortexed, and then transferred to the tube containing the DNA. The reaction was gently mixed avoiding vortexing or pipetting.

Nicking reaction:

	Volume (µl)
3.1 buffer, 10X (NEB)	1
Nt.BspQI (10,000 U/ml stock concentration, NEB)	1
Digested DNA (75 ng)	8

For incorporation of the fluorescent nucleotides we incubated the nicked DNA with Taq DNA polymerase (NEB) that was supplemented with dUTP-atto 532 (Jena Bioscience) in addition to dATP, dGTP, and dCTP (Sigma) at 72 °C for 1 hour.

Labeling reaction:

	Volume (µl)
ThermoPol buffer, 10X (NEB)	1.5
dUTP-atto 532 (25 µM stock concentration)	0.36
dATP, dGTP and dCTP (25 µM stock concentration, Sigma)	0.36 each
Taq DNA polymerase (5,000 U/ml stock concentration, NEB)	1
Nicked DNA	10

Nicks were then ligated with Taq DNA Ligase at 37 °C for 30 min.

Ligation reaction:

	Volume (µl)
10X ThermoPol buffer (NEB)	0.5
NAD^+ (50 mM stock concentration, NEB)	0.4
dNTPs (1 mM stock concentration, Sigma)	0.4
Taq DNA ligase (2,000 U/ml stock concentration, NEB)	1
Nicked-labeled DNA	15

The nicked-labeled DNA was then stained with yoyo-1 (DNA stain, BioNano Genomics) and stored at 4 °C overnight:

Staining solution:

	Volume (µl)
Flow buffer, 4X (BioNano Genomics)	15
DTT, 5X (BioNano Genomics)	12
DNA stain (BioNano Genomics)	1.5
EDTA (0.5 M stock concertation)	1.5
Ultra pure water	10
Nicked-labeled-repaired DNA	20

Loading of DNA in nanochannels and imaging were performed on an Irys instrument (BioNano Genomics). Detection of imaged molecules and fluorescent labels along each molecule was performed by AutoDetect (version 2.1.4, BioNano Genomics). Alignment to the reference genome and *de novo* assembly were performed using IrysView (version 2.3, BioNano Genomics). Coverage analysis was visualized by the UCSC genome browser. Molecule quality

report (MQR) was performed with minimum molecule length of 100 kbp and $P < 1 \times 10^{-6}$. A custom BioNano Genomics assembler based on Overlap-Layout-Consensus paradigm was used for *de novo* assembly with default parameters, except that $P < 1 \times 10^{-9}$ was used for the extend merge iterations. Overall three contigs with a total length of 520 kbp were obtained, including one contig that contained the 200 kbp target sequence with confidence of 1×10^{-30}, and two other contigs that flank the target sequence (Figure S3). The flanking contigs were probably obtained due to partial digestion of molecules at one fragmentation position. The *de novo* assembly resulted in the target 200 kbp contig with five missing nicking sites. Sequencing results confirmed that these were false negative labels, possibly due to the low resolution of the optical maps. Specifically, two adjacent labels with a distance of less than 1.5 kbp between them can be considered as a single label in the assembly process.

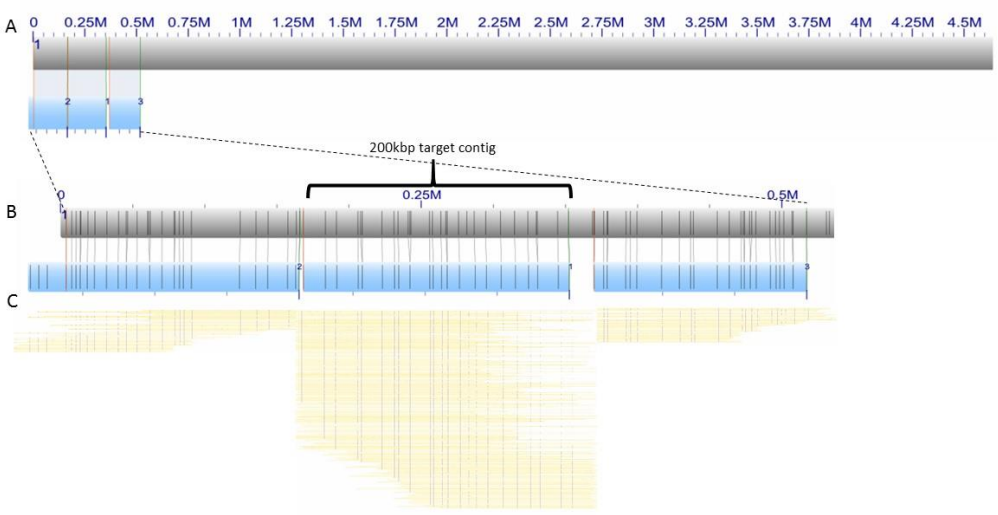

Figure S3. *De novo* assembly of optical mapps using CATCH. A. Digitized representation of the *E. coli* genome reference (gray) and three aligned contigs (blue). B. Alignment of the three contigs (blue) to the *E. coli* genome reference (gray). Gray lines on the reference and contigs indicate expected and detected Nt.BspQI nicking sites, respectively. Matches between the expected and detected nicking sites are presented with light gray lines. C. Digitized representation of individual molecules aligned to each contig (yellow). Blue dots indicate detected Nt.BspQI nicking sites.

REFERENCES

1. Jiang,W., Zhao,X., Gabrieli,T., Lou,C., Ebenstein,Y. and Zhu,T.F. (2015) Cas9-Assisted Targeting of CHromosome segments CATCH enables one-step targeted cloning of large

gene clusters. *Nat. Commun.*, **6**, 8101.

2. Loman,N.J. and Quinlan,A.R. (2014) Poretools: a toolkit for analyzing nanopore sequence data. *Bioinformatics*, **30**, 3399–3401.
3. Li,H. and Durbin,R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, **25**, 1754–1760.
4. Quinlan,A.R. and Hall,I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**, 841–842.
5. Goecks,J., Nekrutenko,A., Taylor,J. and Galaxy Team,T. (2010) Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol.*, **11**, R86.
6. Koren,S., Walenz,B.P., Berlin,K., Miller,J.R. and Phillippy,A.M. (2016) Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *bioRxiv*.