Supplementary Information

Nanopore sequencing of native Adeno-associated virus single-stranded DNA using a transposase-based rapid protocol

Marco T. Radukic[1,†], David Brandt[2,†], Markus Haak[2], Kristian M. Müller[1,*] and Jörn Kalinowski[2,*]

[1] Faculty of Technology, Bielefeld University, 33594 Bielefeld, Germany

[2] Center for Biotechnology (CeBiTec), Bielefeld University, 33594 Bielefeld, Germany

* To whom correspondence should be addressed: Jörn Kalinowski; Tel: +49 521 106 8756; Fax: +49 521 106 89041; Email: joern@cebitec.uni-bielefeld.de or Kristian Müller. Tel: +49 521 106 6323; Fax: +49 521 106 156318; Email: kristian@syntbio.net

† These authors contributed equally to this work.

Contents

1 Methods

1.1 Production of M13KO7 helper phage

E. coli ER2738 from an over-night culture were used to inoculate lysogeny broth (LB) supplemented with ampicillin (100 µg/ml) and tetracycline (25 µg/ml)=to $OD600_{10mm}$ = 0.1 and grown to $OD600_{10mm}$ = 0.5 at 37 °C in Erlenmyer flasks on an orbital shaker. Next, the culture was infected with 4×10^9 pfu/ml M13KO7 (New England Biolabs) helper phage and further incubated for 1.5 hours. Kanamycin was added to a final concentration of 70 µg/ml and, after further four hours of incubation, the culture supernatant containing the phages was separated from cells by centrifugation. Phage ssDNA was prepared from the supernatant by the QIAprep Spin M13 Kit (Qiagen) as per the manufacturer’s instructions. 6.2 µg ssDNA were obtained from a 3 ml preparation.

1.2 qPCR assay

We performed all qPCR measurements with primers and reagents according to the following tables. All dilution series for standard curves and primer dilutions were performed with sterile filtered Millipore MilliQ water in presence of 0.05% Pluronic F68. All pipetting was done with Sarstedt Biosphere low retention filter tips. Final primer concentration was

125 nM. The reaction volume was 20 µl in 96 well Sarstedt Lightcycler plates. Quantification cycles were determined with the Roche 2nd derivative Max algorithm within the LightCycler 480 software, release 1.5.0 SP4, version 1.5.0.39.

All measurements were performed in technical duplicates. Sample DNA was extracted as described in the methods section and further diluted for the assay with MilliQ water containing 0.05% Pluronic F68. Nucleic acid quantification of AAV producer plasmids for standard curve preparation was performed spectroscopically with a Nanodrop 2000c. The ratio OD260/OD280 was between 1.85 and 1.88.

Table 1: Primer sets used for qPCR measurement.

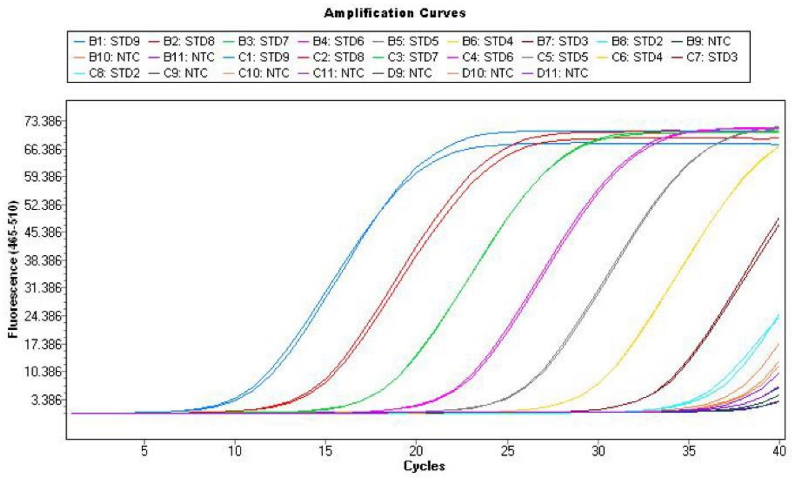
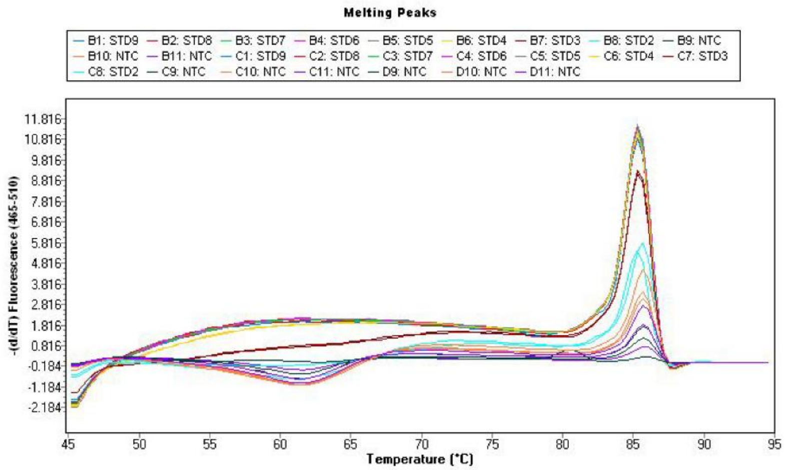
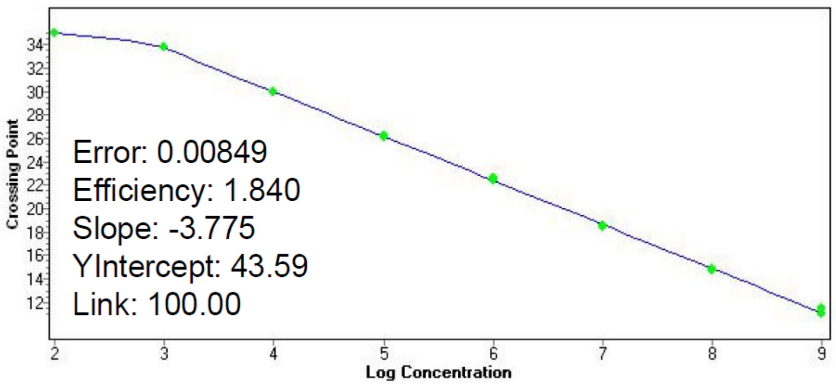
Target	**Primer sequences**	**Amplicon length**	**qPCR program**	**Primer efficiency**	**Linear range**	**Limit of detection**
CMV promoter	5‘-GGGACTTTCCTACTTGGCA 5‘-GGCGGAGTTGTTACGACA	200 bp	A	1.84	10^3 – 10^9 per reaction	<10^3, *
AAV serotype 2 Rep	5’-CGGAGAAGCAGTGGATCCA 5’-ATTTGGGACCGCGAGTTG	76 bp	B	1.82	10^3 – 10^9 per reaction	<10^3, *
Adenovirus gene E4	5’-ACTACGTCCGGCGTTCCAT 5’-GGAGTGCGCCGAGACAAC	68 bp	A	1.85	10^3 – 10^9 per reaction	<10^3, *
β lactamase (Ampicillin resistance) gene, *bla*	5’-CAACTTTATCCGCCTCCATC 5’-AAGCCATACCAAACGACGAG	138 bp	A	1.91	10^3 – 10^9 per reaction	<10^3, *

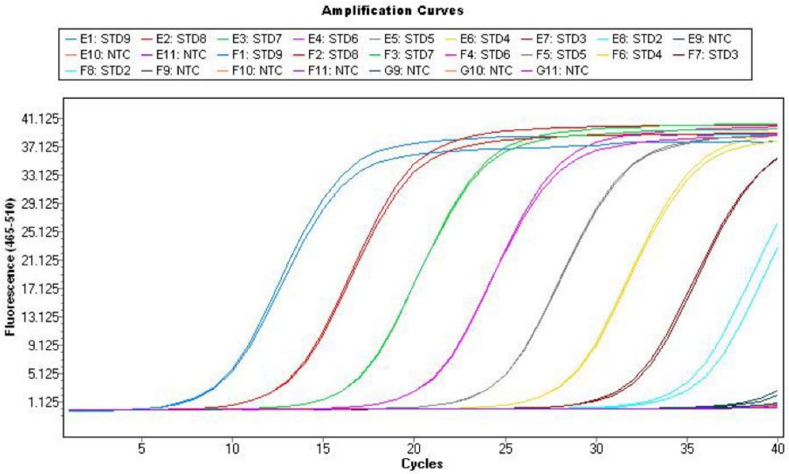
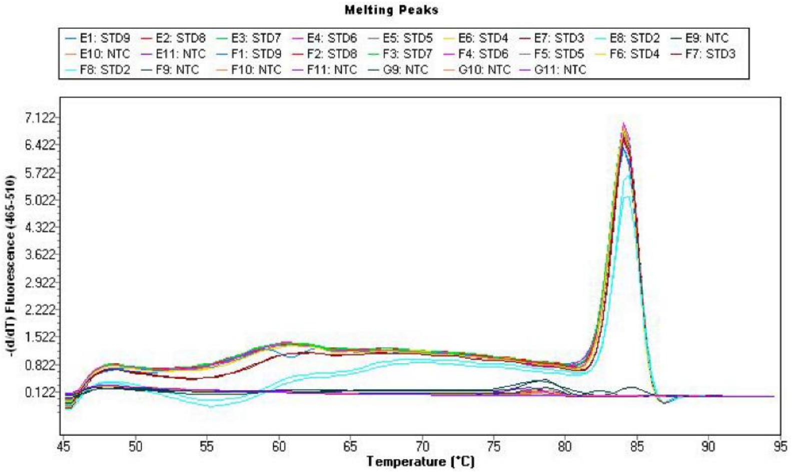
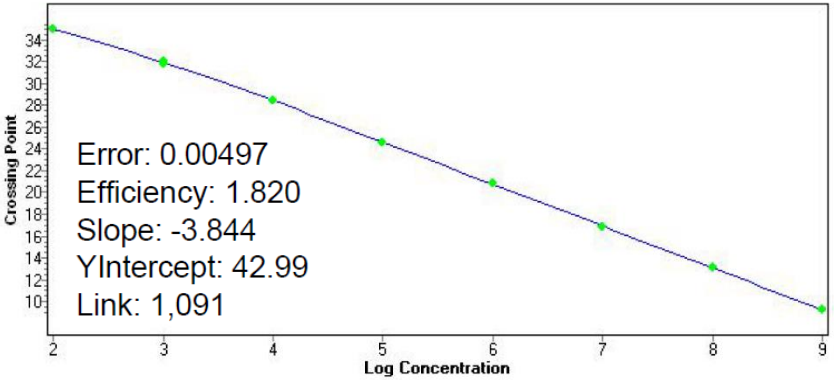
* Standard error of regression method

Table 2: qPCR programs.

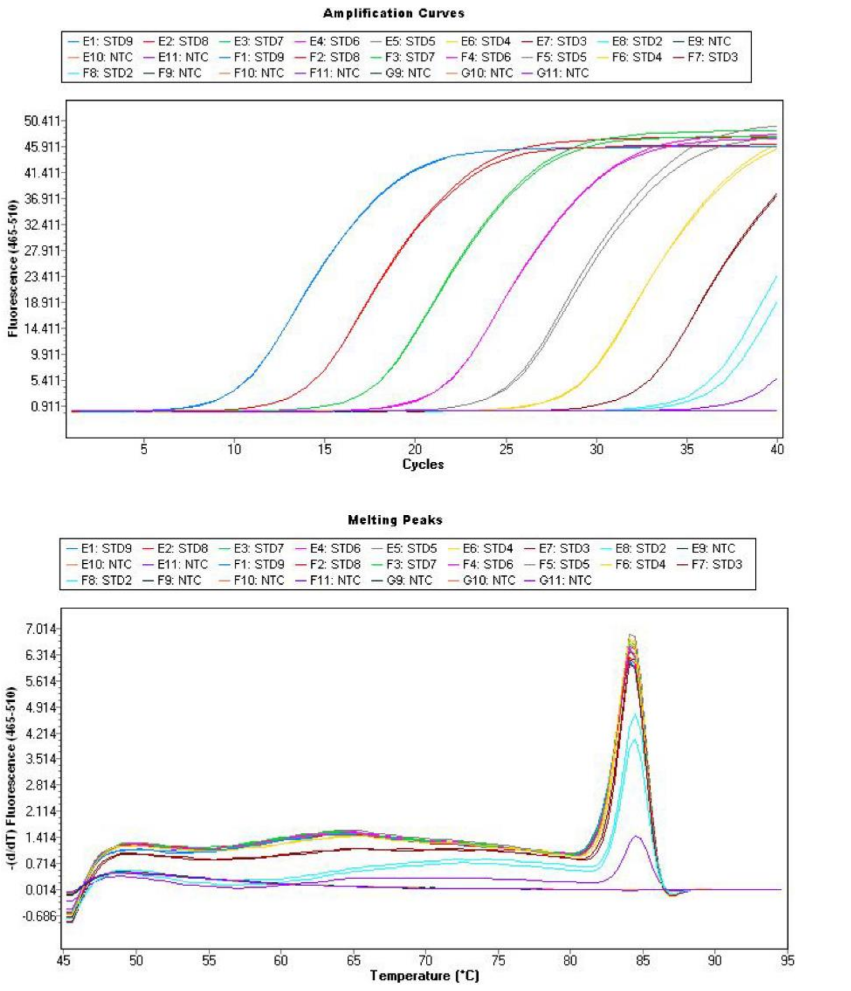
qPCR program	**Instrument, Assay**	**Master Mix**	**Program**
A	Roche LightCycler 480II SYBR-Green type assay	Promega GoTaq qPCR Master Mix	a. 95 °C, 10 min. b. 95 °C, 15 sec. c. 60 °C, 1 min. d. to b., 39x e. 45 °C to 95 °C at 0.11 °C/sec
B	Roche LightCycler 480II SYBR-Green type assay	Promega GoTaq qPCR Master Mix	a. 95 °C, 10 min. b. 95 °C, 15 sec. c. 55 °C, 15 sec. d. 60 °C, 1 min. e. to b., 39x f. 45 °C to 95 °C at 0.11 °C/sec

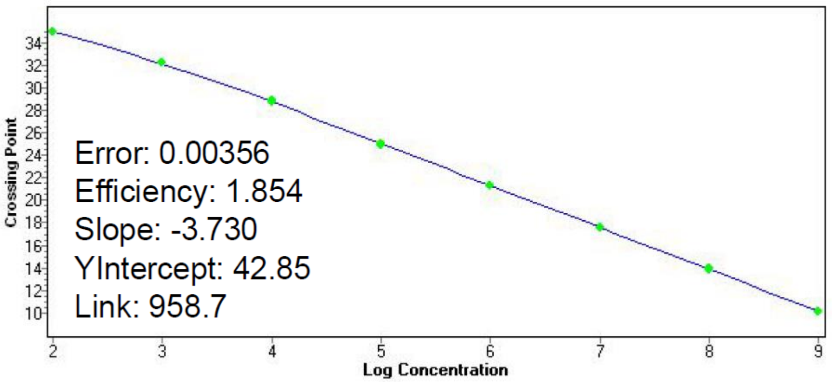
Table 3: Standard curves for qPCR primer sets

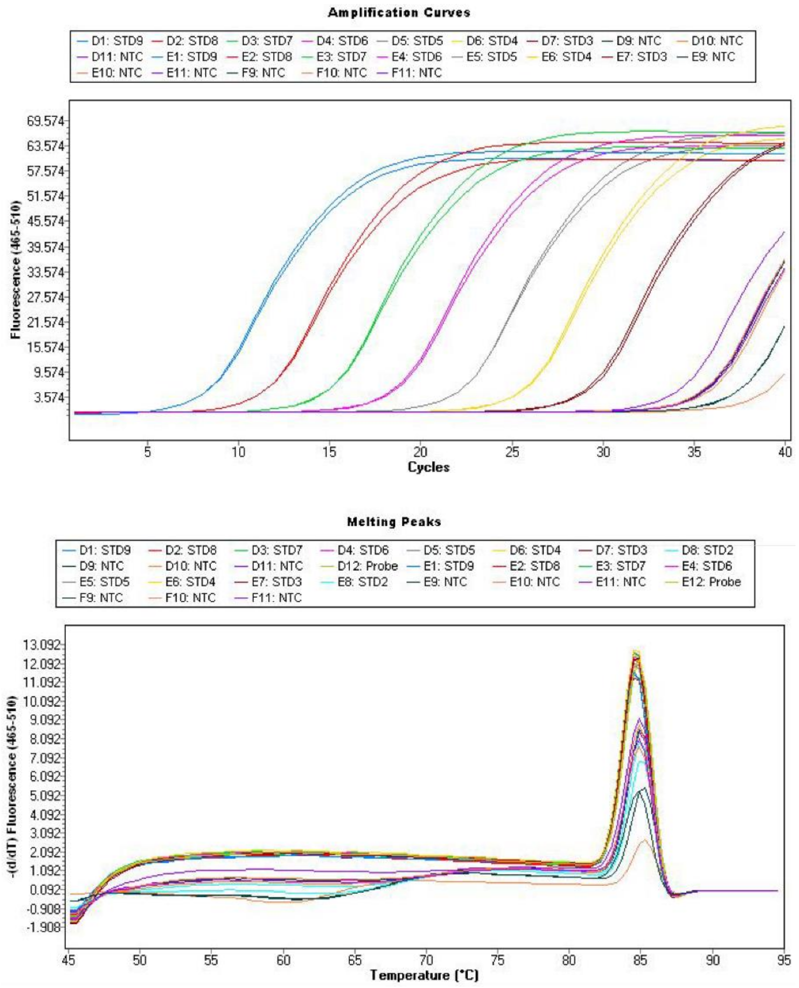
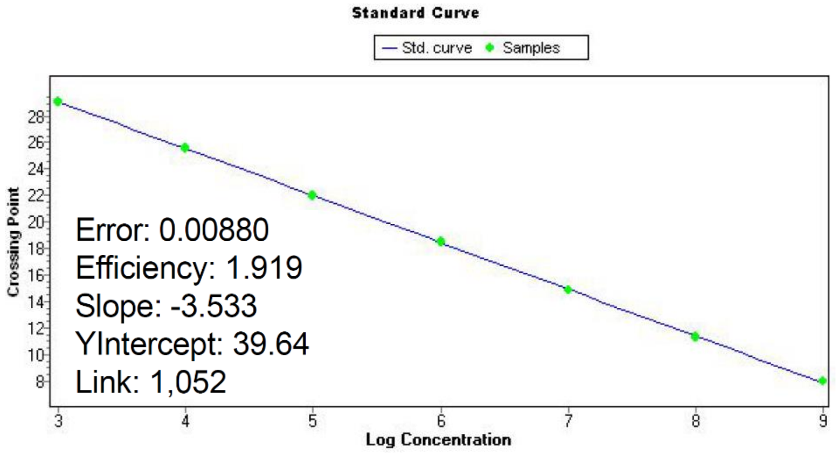
Primer set	**Fluorescence graph and melting curve**	**Standard curve**
CMV	 	

Rep	 Amplification Curves E1: STD9 E2: STD8 E3: STD7 E4: STD6 E5: STD5 E6: STD4 E7: STD3 E8: STD2 E9: NTC E10: NTC E11: NTC F1: STD9 F2: STD8 F3: STD7 F4: STD6 F5: STD5 F6: STD4 F7: STD3 F8: STD2 F9: NTC F10: NTC F11: NTC G9: NTC G10: NTC G11: NTC Fluorescence (465-510) 41.125 37.125 33.125 29.125 25.125 21.125 17.125 13.125 9.125 5.125 1.125 Cycles 5 10 15 20 25 30 35 40  Melting Peaks E1: STD9 E2: STD8 E3: STD7 E4: STD6 E5: STD5 E6: STD4 E7: STD3 E8: STD2 E9: NTC E10: NTC E11: NTC F1: STD9 F2: STD8 F3: STD7 F4: STD6 F5: STD5 F6: STD4 F7: STD3 F8: STD2 F9: NTC F10: NTC F11: NTC G9: NTC G10: NTC G11: NTC -(d/dT) Fluorescence (465-510) 7.122 6.422 5.722 5.022 4.322 3.622 2.922 2.222 1.522 0.822 0.122 Temperature (°C) 45 50 55 60 65 70 75 80 85 90 95	 Crossing Point 34 32 30 28 26 24 22 20 18 16 14 12 10 Log Concentration 2 3 4 5 6 7 8 9 Error: 0.00497 Efficiency: 1.820 Slope: -3.844 YIntercept: 42.99 Link: 1,091

E4


Amplification Curves
E1: STD9
E2: STD8
E3: STD7
E4: STD6
E5: STD5
E6: STD4
E7: STD3
E8: STD2
E9: NTC
E10: NTC
E11: NTC
F1: STD9
F2: STD8
F3: STD7
F4: STD6
F5: STD5
F6: STD4
F7: STD3
F8: STD2
F9: NTC
F10: NTC
F11: NTC
G9: NTC
G10: NTC
G11: NTC
Fluorescence (465-510)
50.411
45.911
41.411
36.911
32.411
27.911
23.411
18.911
14.411
9.911
5.411
0.911
5
10
15
20
25
30
35
40
Cycles
Melting Peaks
-(d/dT) Fluorescence (465-510)
7.014
6.314
5.614
4.914
4.214
3.514
2.814
2.114
1.414
0.714
0.014
-0.686
45
50
55
60
65
70
75
80
85
90
95
Temperature (°C)


Crossing Point
34
32
30
28
26
24
22
20
18
16
14
12
10
Error: 0.00356
Efficiency: 1.854
Slope: -3.730
YIntercept: 42.85
Link: 958.7
2
3
4
5
6
7
8
9
Log Concentration

bla		

Amplification Curves
D1: STD9 D2: STD8 D3: STD7 D4: STD6 D5: STD5 D6: STD4 D7: STD3 D9: NTC D10: NTC
D11: NTC E1: STD9 E2: STD8 E3: STD7 E4: STD6 E5: STD5 E6: STD4 E7: STD3 E9: NTC
E10: NTC E11: NTC F9: NTC F10: NTC F11: NTC
Fluorescence (465-510)
69.574 63.574 57.574 51.574 45.574 39.574 33.574 27.574 21.574 15.574 9.574 3.574
5 10 15 20 25 30 35 40
Cycles
Melting Peaks
D1: STD9 D2: STD8 D3: STD7 D4: STD6 D5: STD5 D6: STD4 D7: STD3 D8: STD2
D9: NTC D10: NTC D11: NTC D12: Probe E1: STD9 E2: STD8 E3: STD7 E4: STD6
E5: STD5 E6: STD4 E7: STD3 E8: STD2 E9: NTC E10: NTC E11: NTC E12: Probe
F9: NTC F10: NTC F11: NTC
-(d/dT) Fluorescence (465-510)
13.092 12.092 11.092 10.092 9.092 8.092 7.092 6.092 5.092 4.092 3.092 2.092 1.092 0.092 -0.908 -1.908
45 50 55 60 65 70 75 80 85 90 95
Temperature (°C)

Standard Curve
Std. curve
Samples
Crossing Point
28 26 24 22 20 18 16 14 12 10 8
3 4 5 6 7 8 9
Log Concentration
Error: 0.00880
Efficiency: 1.919
Slope: -3.533
YIntercept: 39.64
Link: 1,052

2 Script for the simulation of the transposase reaction

The following script was developed for and tested on GNU Octave 5.1.0. Sequences to be analysed can be put between the brackets in line 2. The script will output results to a graph and a text file in the user folder (under MS Windows).

```
1   #Input DNA Sequence
2   DNAseq="Put sequence of interest between brackets";
3   variants = length(DNAseq)-1;
4   r_m = ones (variants*2, 2);
5   
6   #Basic conversions
7   DNAseq( DNAseq == "G" )="S";
8   DNAseq( DNAseq == "C" )="S";
9   
10  #Calculate %GC upstream
11  for i = 1:variants
12    #length of DNA fragment
13    r_m([i], [1]) = variants-i+1;
14  
15    #%GC
16    r_m([i], [2]) = columns(strfind(substr(DNAseq, i+1), "S")) / r_m([i], [1]) * 100;
17  
18  endfor
19  
20  #Calculate %GC downstream
21  for i = variants+1:rows(r_m)
22    #lenght of DNA fragment (same as above)
23    r_m([i], [1]) = i-variants;
24  
25    #%GC
26    r_m([i],  [2])  =  columns(strfind(substr(fliplr(DNAseq),  variants-i),  "S"))  /
27  r_m([i], [1]) * 100;
28  endfor
29  
30  save results.mat r_m
31  plotmatrix(r_m)
```

3 Transposase adapter sequences

The following transposase adapter sequences were used for the realignment of untrimmed reads in order to estimate transposition sites:

Samples M13mp18 ssDNA and M13KO7 ssDNA:

5′-GCTTGGGTGTTTAACCTTCAGGGAACAAACCAAGTTACGTGTTTTCGCATTTATCGTGAAACGCTTTCGCGTTTTTCGTGCGCCGCTTCA

Sample M13mp18 phagemid dsDNA:

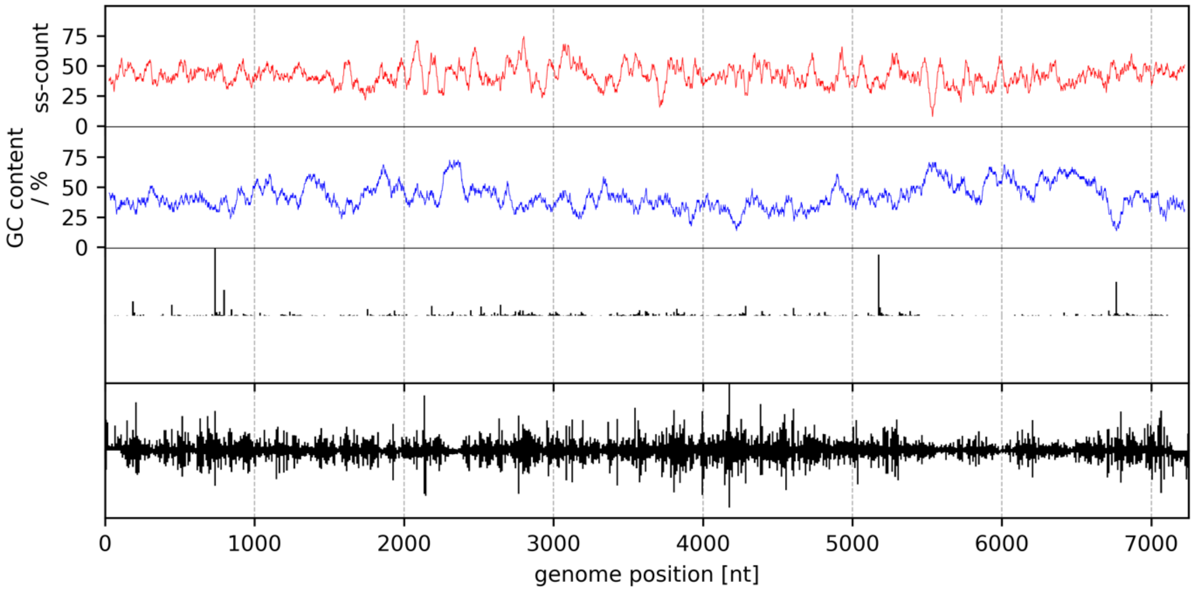
5′-GCTTGGGTGTTTAACCAACTAGGCACAGCGAGTCTTGGTTGTTTTCGCATTTATCGTGAAACGCTTTCGCGTTTTTCGTGCGCCGCTTCA

AAV sample 2 (run 2):

5’-GCTTGGGTGTTTAACCGTTTTCGCATTTATCGTGAAACGCTTTCGCGTTTTTCGTGCGCCGCTTCA

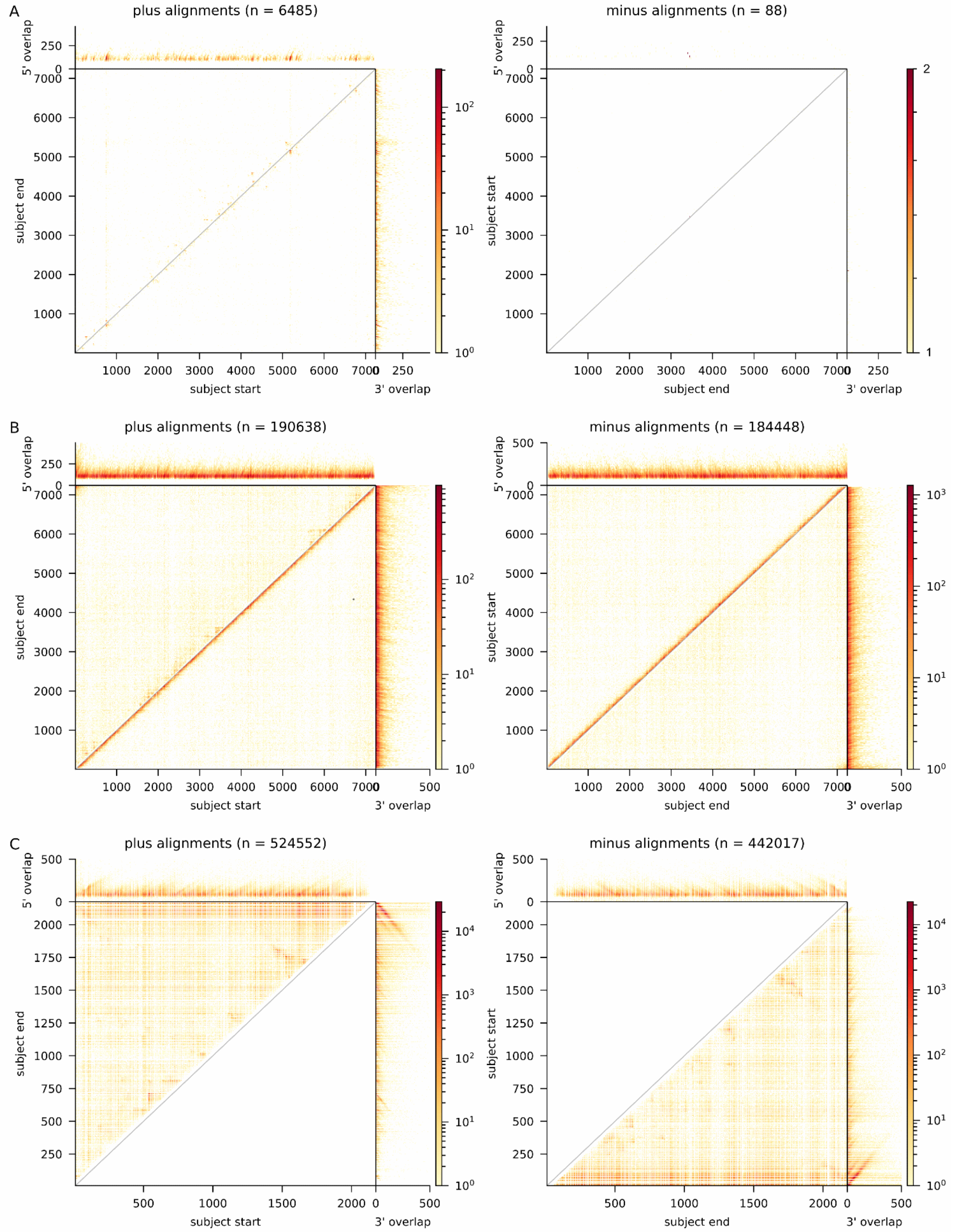
4 Supplementary Figures

Supplementary Figure S1



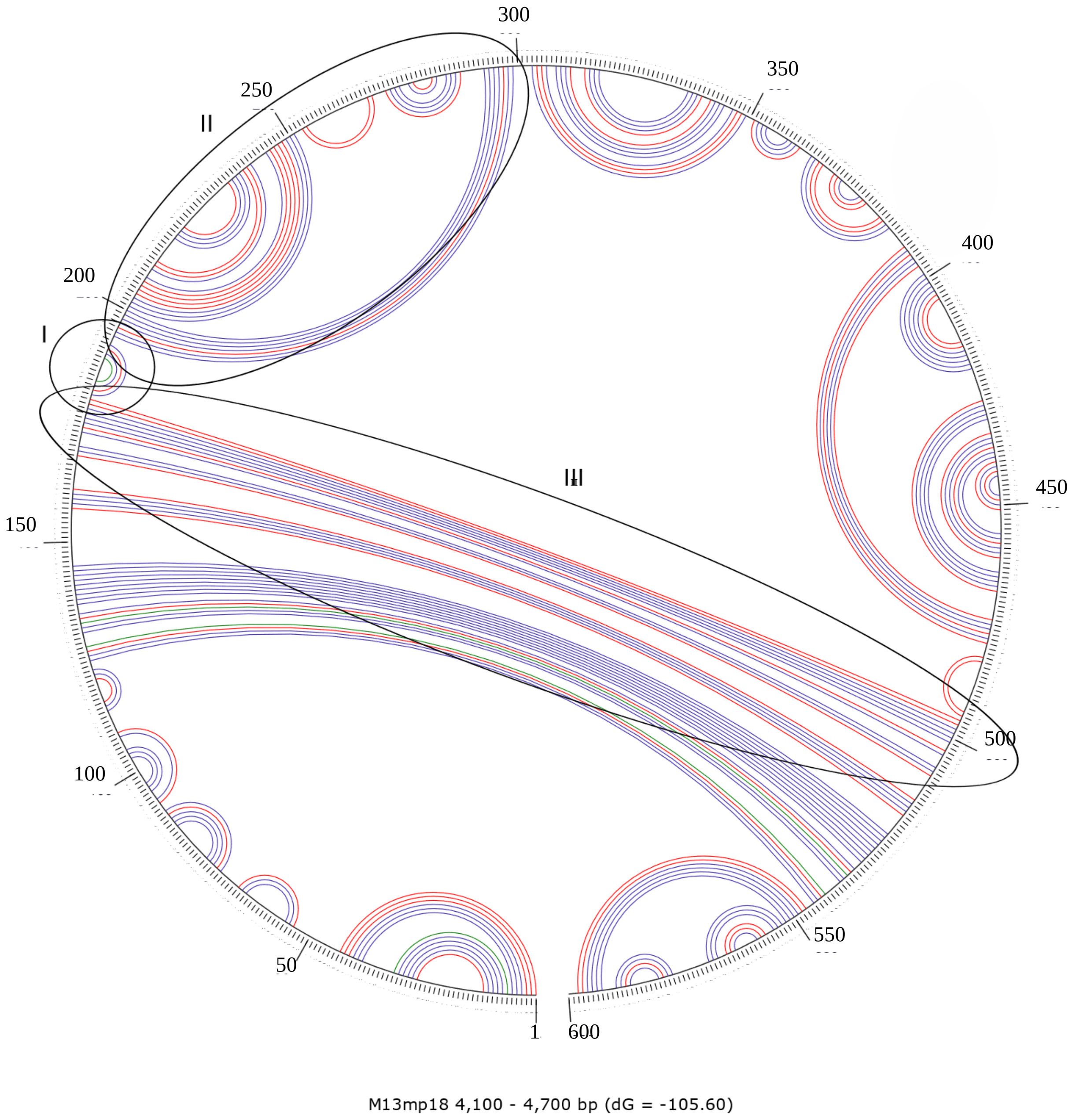
Supplementary Figure S1. Plus strand ss-count, GC content and relative transposase insertion sites of samples M13mp18 ssDNA (top) and M13mp18 dsDNA (bottom). The ss-count is based on 100 predicted DNA folding structures. Both ss-count and GC content are averaged over a moving window of 50 nt. The estimated read starts are binned in 15 nt bins and normalized to the maximal bin count.

Supplementary Figure S2



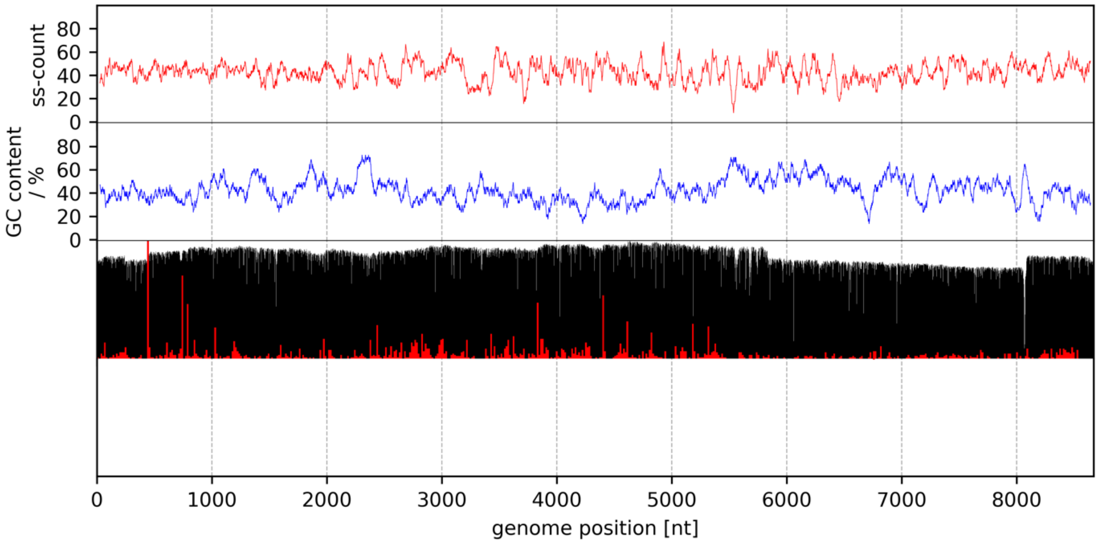
Supplementary Figure S2. Binned subject start and end positions for untrimmed reads mapped against ssM13 (A), dsM13 (B) and rAAV (C) genomes. Bins are sized 15 x 15 nt for ssM13 and dsM13 and 5 x 5 nt for rAAV. The length of unmapped stretches of reads is plotted as 5′ overhang and 3′ overhang, respectively. Read count per bin is scaled logarithmically.

Supplementary Figure S3:



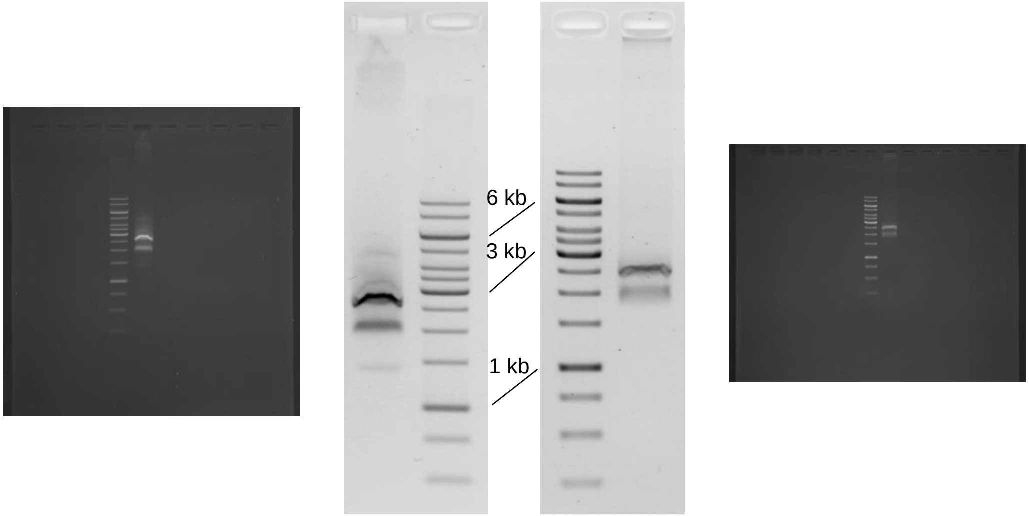
Supplementary Figure S3. Circular structure plot of predicted folding from base 4100 to 4700 of M13mp18 genomic DNA. Structure prediction carried out using mfold_util 4.7 with standard options. Possible hairpin loops serving as targets for transposase insertion are highlighted.

Supplementary Figure S4



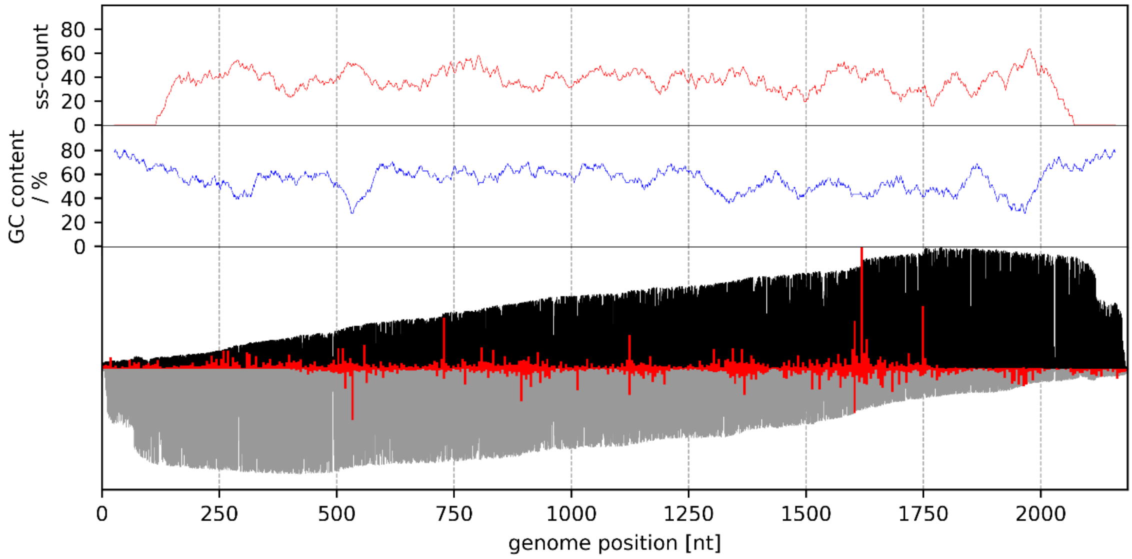
Supplementary Figure S4. Plus strand ss-count, GC content, relative coverage and relative transposase insertion sites of sample M13KO7 ssDNA. The ss-count is based on 100 predicted DNA folds. Both ss-count and GC content are averaged over a moving window of 50 nt. The estimated read starts are binned in 15 nt bins and normalized to the maximal bin count.

Supplementary Figure S5



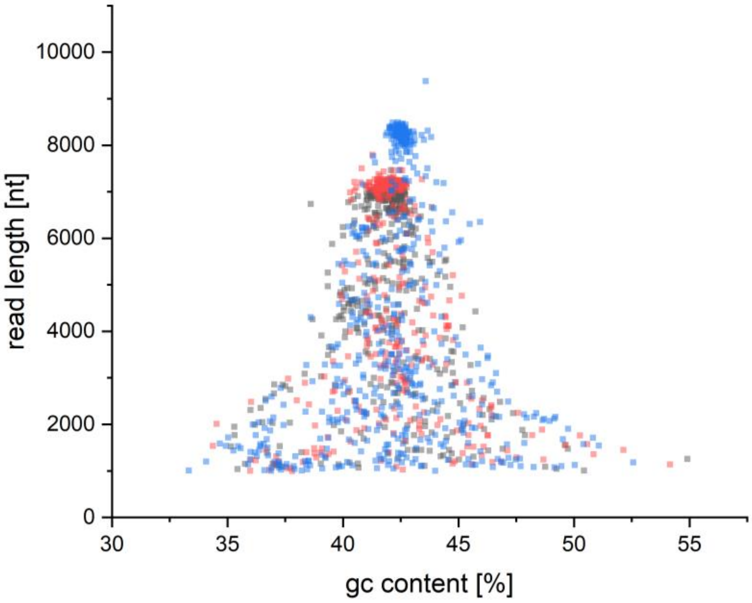
Supplementary Figure S5. Agarose gel electrophoresis of DNA prepared from rAAV sample 1 used for multiplexed sequencing (run 1). Both gels show a 5 µl sample from the same preparation before (left) and after a freeze-thaw cycle. The gel was 1% agarose in TAE buffer, run at 120 V for 50 min. Staining by SYBR Gold nucleic acid stain (Thermo). The marker was Gene Ruler 1 kb (Thermo). Gels displayed with inverted colors and spread histograms. Uncropped gels are given. The unfrozen sample shows three distinct bands. One band runs below the expected genome size of 2.2 kb. One runs right at the expected size and one runs between 2.5 and 3 kb. Upon freeze-thawing, we observe that the smallest fragment disappears, and aggregates are seen at the sample pocket, whereas the other bands are preserved. We hypothesize that the smallest fragment resembles true single-stranded genomes, while the band of the expected size resembles two at least partly hybridized genomes. Facilitated hybridization of ssDNA oligos by freeze-thaw cycles has been observed before (Elghanian et al., 1997, Science, DOI: 10.1126/science.277.5329.1078). Higher order non-covalent multimers seem also to be present. Under the investigated conditions as estimated from the band intensities, a larger part of the sample appears to be in hybridized states.

Supplementary Figure S6



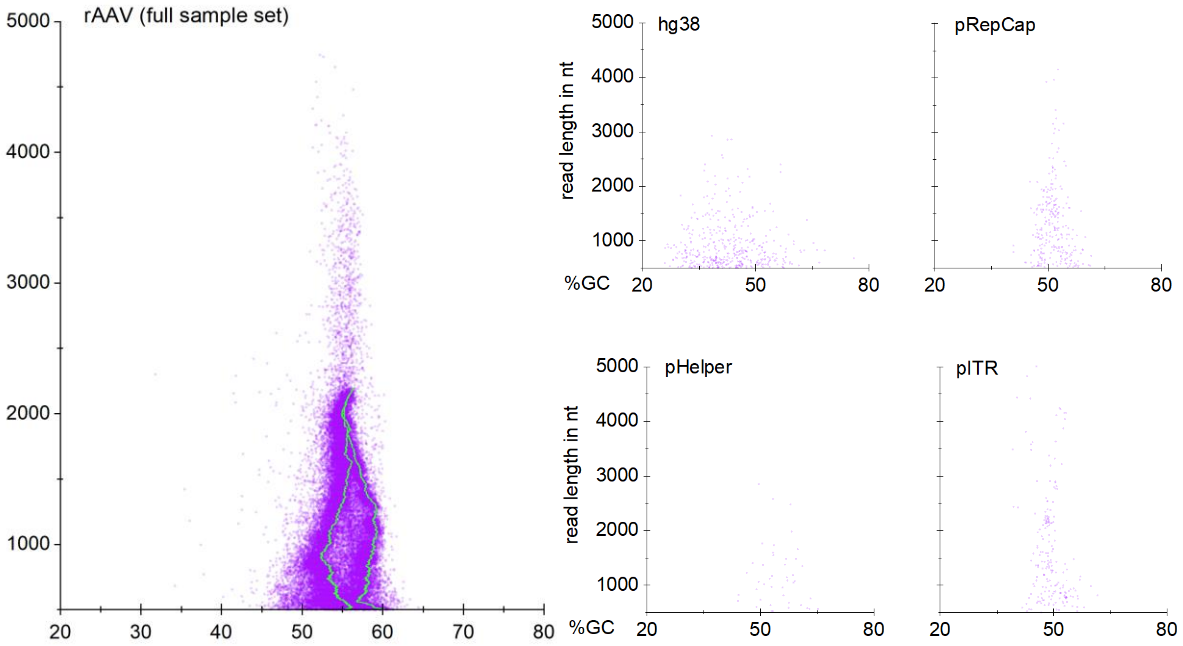
Supplementary Figure S6. Plus strand ss-count, GC content, relative coverage and relative transposition sites of sample rAAV (run 2). The ss-count is based on 100 predicted DNA folds. Both ss-count and GC content are averaged over a moving window of 50 nt. The estimated read starts are binned in 5 nt bins and normalized to the maximal bin count.

Supplementary Figure S7



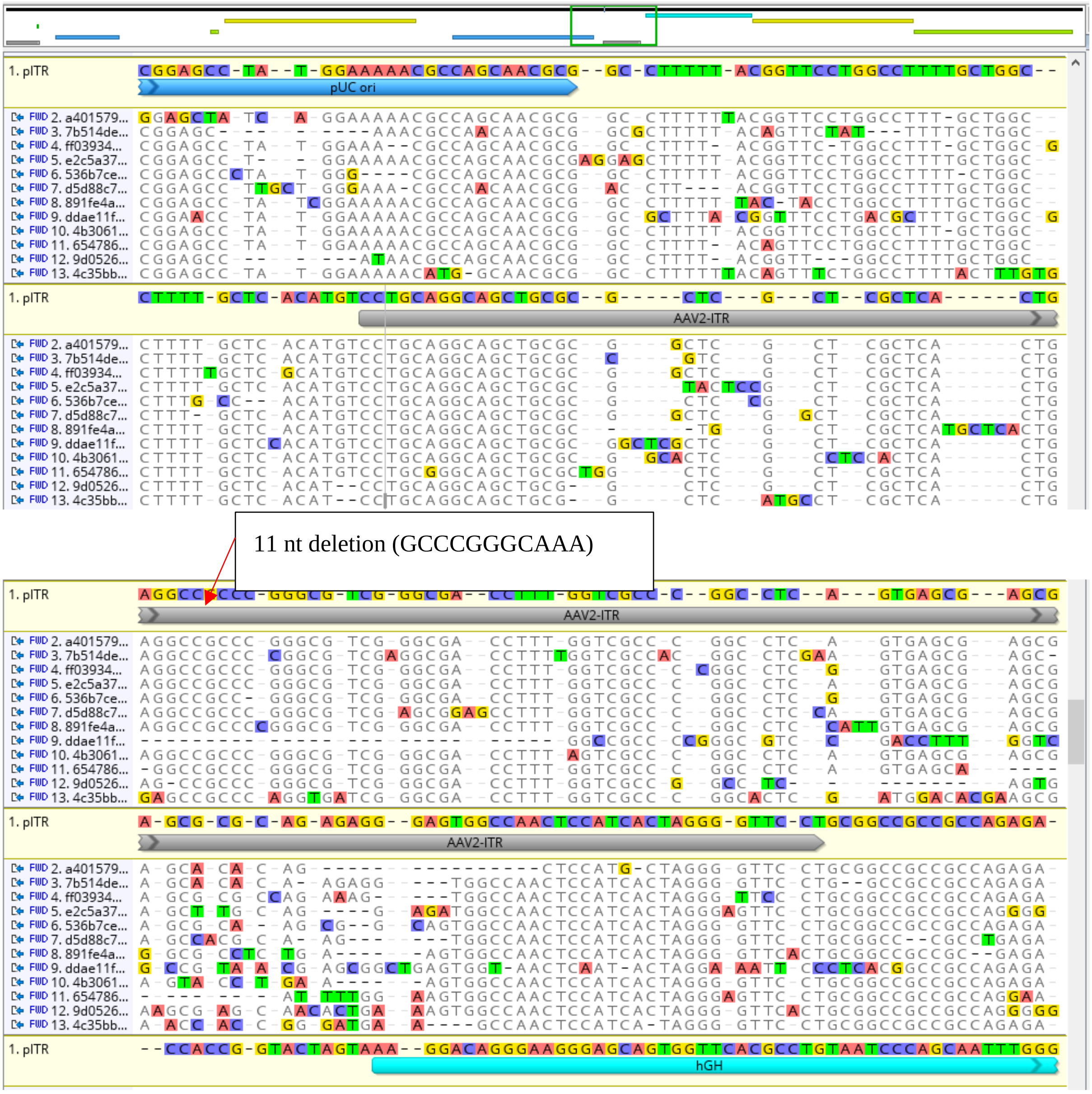
Supplementary Figure S7. GC content *versus* read length for a commercial M13mp18 dsDNA phagemid (grey, partially overlaid by red) and M13mp18 ssDNA (M13mp18, red), as well as M13 in-house propagated helper phage (blue). Most reads are of similar length and GC content, indicating that an individual molecule is only fragmented once by the transposase. Conical tailing to minor extends hints on premature sequencing breakoffs and the prevalence of double-cut genomes.

Supplementary Figure S8



Supplementary Figure S8. GC content *versus* read length for the multiplexed AAV sequencing run, grouped by BLAST assignments to the reference library. One magenta dot represents on read. Of all 36239 reads that passed the quality threshold of >500 nt, 52 (0.14%) were longer than 5000 nt, 51 of which fell into the pITR bin. This finding highlights the importance of complete Benzonase digest prior to capsid disruption. Additionally, the data indicates that oversized reads likely stem from the pITR backbone instead of genome multimers or the other producer plasmids.

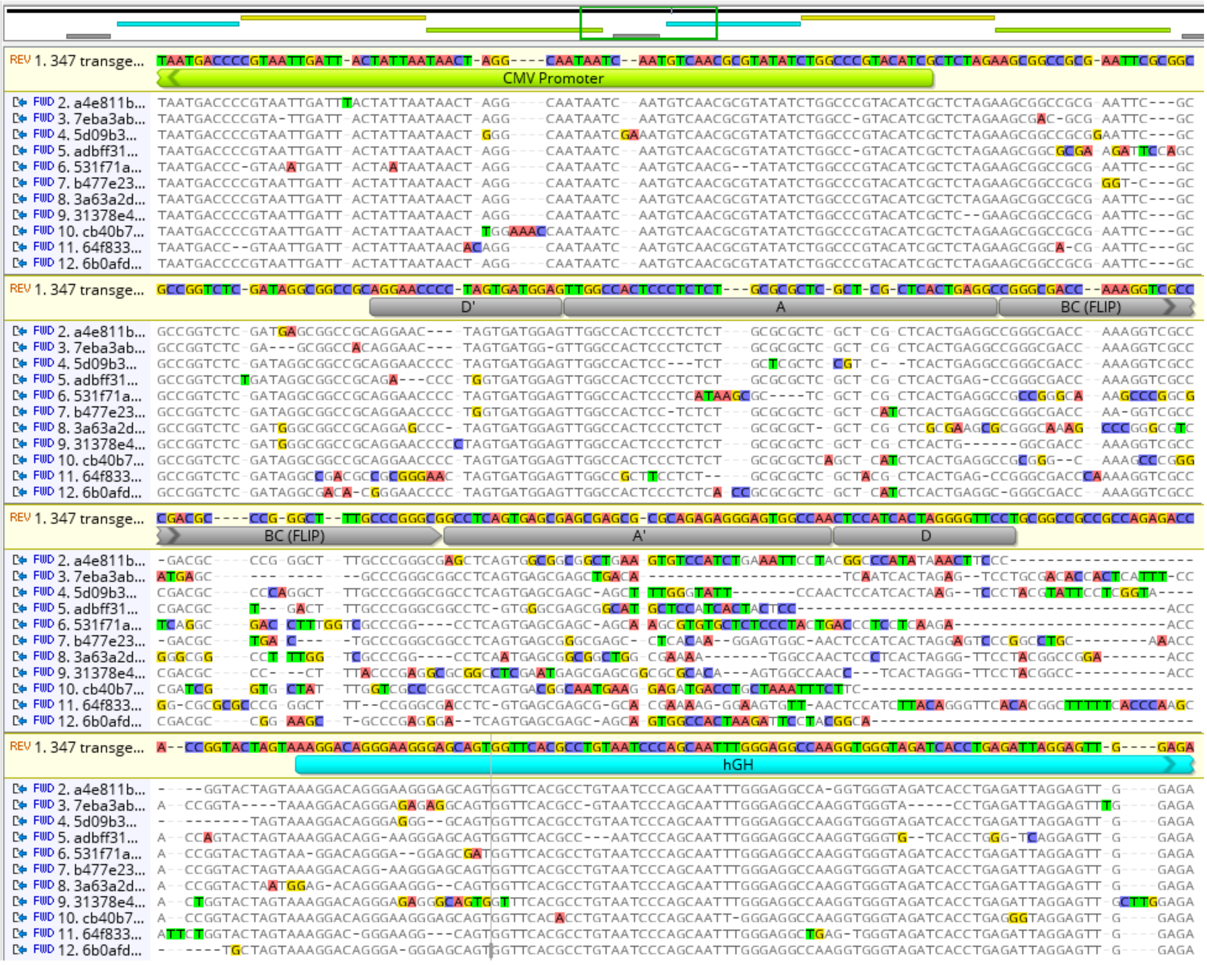
Supplementary Figure S9



11 nt deletion (GCCCGGGCAAA)

Supplementary Figure S9. 12 read examples aligned to the junction ITR between genome-backbone fusions are shown. Junction ITRs still harbour a 11 nt deletion and sequencing quality drops downstream of the ITR internal palindrome (12 read examples mapped to pITR).

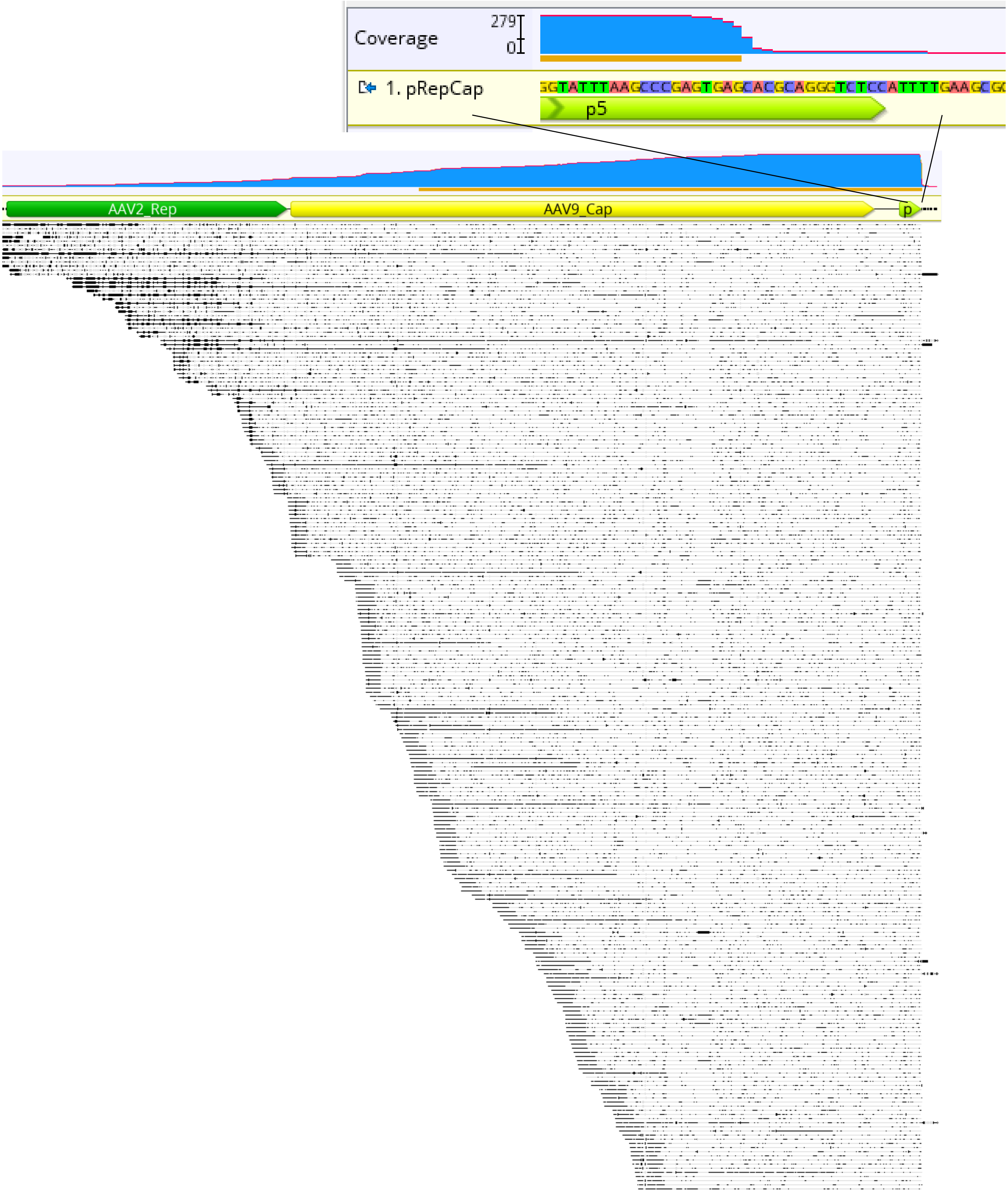
Supplementary Figure S10



```
REV 1. 347 transge...  TAATGACCCCGTAATTGATT-ACTATTAATAACT-AGG----CAATAATC--AATGTCAACGCGTATATCTGGCCCGTACATCGCTCTAGAAGCGGCCGCG-AATTCGCGGC
FWD 2. a4e811b...      TAATGACCCCGTAATTGATTTACTATTAATAACT-AGG----CAATAATC--AATGTCAACGCGTATATCTGGCCCGTACATCGCTCTAGAAGCGGCCGCG-AATTC---GC
FWD 3. 7eba3ab...      TAATGACCCCGTA-TTGATT-ACTATTAATAACT-AGG----CAATAATC--AATGTCAACGCGTATATCTGGCC-GTACATCGCTCTAGAAGCGAC-GCG-AATTC---GC
FWD 4. 5d09b3...       TAATGACCCCGTAATTGATT-ACTATTAATAACT-GGG----CAATAATCGAAATGTCAACGCGTATATCTGGCCCGTACATCGCTCTAGAAGCGGCCGCGGAATTC---GC
FWD 5. adbff31...      TAATGACCCCGTAATTGATT-ACTATTAATAACT-AGG----CAATAATC--AATGTCAACGCGTATATCTGGCC-GTACATCGCTCTAGAAGCGGCGCGA-AGATTCCAGC
FWD 6. 531f71a...      TAATGACCC-GTAAATGATT-ACTAATAATAACT-AGG----CAATAATC--AATGTCAACG--TATATCTGGCCCGTACATCGCTCTAGAAGCGGCCGCG-AATTC---GC
FWD 7. b477e23...      TAATGACCCCGTAATTGATT-ACTATTAATAACT-AGG----CAATAATC--AATGTCAACGCGTATATCTGGCCCGTACATCGCTCTAGAAGCGGCCGCG-GGT-C---GC
FWD 8. 3a63a2d...      TAATGACCCCGTAATTGATT-ACTATTAATAACT-AGG----CAATAATC--AATGTCAACGCGTATATCTGGCCCGTACATCGCTCTAGAAGCGGCCGCG-AATTC---GC
FWD 9. 31378e4...      TAATGACCCCGTAATTGATT-ACTATTAATAACT-AGG----CAATAATC--AATGTCAACGCGTATATCTGGCCCGTACATCGCTC--GAAGCGGCCGCG-AATTC---GC
FWD 10. cb40b7...      TAATGACCCCGTAATTGATT-ACTATTAATAACT-TGGAAACCAATAATC--AATGTCAACGCGTATATCTGGCCCGTACATCGCTCTAGAAGCGGCCGCG-AATTC---GC
FWD 11. 64f833...      TAATGACC--GTAATTGATT-ACTATTAATAACACAGG----CAATAATC--AATGTCAACGCGTATATCTGGCCCGTACATCGCTCTAGAAGCGGCA-CG-AATTC---GC
FWD 12. 6b0afd...      TAATGACCCCGTAATTGATT-ACTATTAATAACT-AGG----CAATAATC--AATGTCAACGCGTATATCTGGCCCGTACATCGCTCTAGAAGCGGCCGCG-AATTC---GC

REV 1. 347 transge...  GCCGGTCTC-GATAGGCGGCCGCAGGAACCCC-TAGTGATGGAGTTGGCCACTCCCTCTCT---GCGCGCTC-GCT-CG-CTCACTGAGGCCGGGCGACC--AAAGGTCGCC
FWD 2. a4e811b...      GCCGGTCTC-GATGAGCGGCCGCAGGAAC----TAGTGATGGAGTTGGCCACTCCCTCTCT---GCGCGCTC-GCT-CG-CTCACTGAGGCCGGGCGACC--AAAGGTCGCC
FWD 3. 7eba3ab...      GCCGGTCTC-GA---GCGGCCACAGGAAC----TAGTGATGG-GTTGGCCACTCCCTCTCT---GCGCGCTC-GCT-CG-CTCACTGAGGCCGGGCGACC--AAAGGTCGCC
FWD 4. 5d09b3...       GCCGGTCTC-GATAGGCGGCCGCAGGAACCCC-TAGTGATGGAGTTGGCCACTCC---TCT---GCTCGCTC-CGT-C---TCACTGAGGCCGGGCGACC--AAAGGTCGCC
FWD 5. adbff31...      GCCGGTCTCTGATAGGCGGCCGCAGA---CCC-TGGTGATGGAGTTGGCCACTCCCTCTCT---GCGCGCTC-GCT-CG-CTCACTGAG-CCGGGCGACC--AAAGGTCGCC
FWD 6. 531f71a...      GCCGGTCTC-GATAGGCGGCCGCAGGAACCCC-TAGTGATGGAGTTGGCCACTCCCTCATAAGCGC----TC-GCT-CG-CTCACTGAGGCCGCCGGGCA--AAGCCCGGCG
FWD 7. b477e23...      GCCGGTCTC-GATAGGCGGCCGCAGGAACCCC-TGGTGATGGAGTTGGCCACTCC-TCTCT---GCGCGCTC-GCT-CATCTCACTGAGGCCGGGCGACC--AA-GGTCGCC
FWD 8. 3a63a2d...      GCCGGTCTC-GATGGGCGGCCGCAGGAGCCC--TAGTGATGGAGTTGGCCACTCCCTCTCT---GCGCGCT--GCT-CG-CTCGCGAAGCGCGGGCAAAG--CCCGGGCGTC
FWD 9. 31378e4...      GCCGGTCTC-GATGGGCGGCCGCAGGAACCCCCTAGTGATGGAGTTGGCCACTCCCTCTCT---GCGCGCTC-GCT-CG-CTCACTG------GGCGACC--AAAGGTCGCC
FWD 10. cb40b7...      GCCGGTCTC-GATAGGCGGCCGCAGGAACCCC-TAGTGATGGAGTTGGCCACTCCCTCTCT---GCGCGCTCAGCT-CATCTCACTGAGGCCGCGGG--C--AAAGCCCGGG
FWD 11. 64f833...      GCCGGTCTC-GATAGGCCGACGCCGCGGGAAC-TAGTGATGGAGTTGGCCGCTTCCTCT-----GCGCGCTC-GCTACG-CTCACTGAG-CCGGGCGACCCAAAAGGTCGCC
FWD 12. 6b0afd...      GCCGGTCTC-GATAGGCGACA-CGGGAACCCC-TAGTGATGGAGTTGGCCACTCCCTCTCA-CCGCGCGCTC-GCT-CATCTCACTGAGGC-GGGCGACC--AAAGGTCGCC

REV 1. 347 transge...  CGACGC----CCG-GGCT--TTGCCCGGGCGGCCTCAGTGAGCGAGCGAGCG-CGCAGAGAGGGAGTGGCCAACTCCATCACTAGGGGTTCCTGCGGCCGCCGCCAGAGACC
FWD 2. a4e811b...      -GACGC----CCG-GGCT--TTGCCCGGGCGAGCTCAGTGGCGGCGGCTGAA-GTGTCCATCTGAAATTCCTACGGCCCATATAAACTTCCC--------------------
FWD 3. 7eba3ab...      ATGAGC----------------GCCCGGGCGGCCTCAGTGAGCGAGCTGACA---------------------TCAATCACTAGAG--TCCTGCGACACCACTCATTT-CC
FWD 4. 5d09b3...       CGACGC----CCCAGGCT--TTGCCCGGGCGGCCTCAGTGAGCGAGC-AGCT-TTGGGTATT-------CCAACTCCATCACTAAG--TCCCTACGTATTCCTCGGTA----
FWD 5. adbff31...      CGACGC----T---GACT--TTGCCCGGGCGGCCTC-GTGGGCGAGCGGCAT-GCTCCATCACTACTCC----------------------------------------ACC
FWD 6. 531f71a...      TCAGGC----GAC-CTTTGGTCGCCCGG----CCTCAGTGAGCGAGC-AGCA-AGCGTGTGCTCTCCCTACTGACCCTCCTCAAGA-----------------------ACC
FWD 7. b477e23...      -GACGC----TGA-C-----TGCCCGGGCGGCCTCAGTGAGCGGGCGAGC--CTCACAA--GGAGTGGC-AACTCCATCACTAGGAGTCCCGGCCTGC--------AAACC
FWD 8. 3a63a2d...      GGGCGG----CCT-TTGG--TCGCCCGG----CCTCAATGAGCGGCGGCTGG-CGAAAA-------TGGCCAACTCCCTCACTAGGG-TTCCTACGGCCGGA-------ACC
FWD 9. 31378e4...      CGACGC----CC----CT--TTACCCGAGGCGCGGCCTCGAATGAGCGAGCGGCGCGCACA---AGTGGCCAACC---TCACTAGGG-TTCCTACGGCC----------ACC
FWD 10. cb40b7...      CGATCG----GTG-CTAT--TTGGTCGCCCGGCCTCAGTGACGGCAATGAAG-GAGATGACCTGCTAAATTTCTTC------------------------------------
FWD 11. 64f833...      GG-CGCGCGCCCG-GGCT--TT--CCGGGCGACCTC-GTGAGCGAGCG-GCA-CGAAAAG-GGAAGTGTT-AACTCCATCTTACAGGGTTCACACGGCTTTTTCACCCAAGC
FWD 12. 6b0afd...      CGACGC----CGG-AAGC--T-GCCCGAGGGA--TCAGTGAGCGAGC-AGCA-GTGGCCACTAAGATTCCTACGGCA-----------------------------------

REV 1. 347 transge...  A--CCGGTACTAGTAAAGGACAGGGAAGGGAGCAGTGGTTCACGCCTGTAATCCCAGCAATTTGGGAGGCCAAGGTGGGTAGATCACCTGAGATTAGGAGTT-G----GAGA
FWD 2. a4e811b...      -----GGTACTAGTAAAGGACAGGGAAGGGAGCAGTGGTTCACGCCTGTAATCCCAGCAATTTGGGAGGCCA-GGTGGGTAGATCACCTGAGATTAGGAGTT-G----GAGA
FWD 3. 7eba3ab...      A--CCGGTA----TAAAGGACAGGGAGAGAGGCAGTGGTTCACGCC-GTAATCCCAGCAATTTGGGAGGCCAAGGTGGGTA-----CCTGAGATTAGGAGTTTG----GAGA
FWD 4. 5d09b3...       ----------TAGTAAAGGACAGGGAGGG--GCAGTGGTTCACGCCTGTAATCCCAGCAATTTGGGAGGCCAAGGTGGGTAGATCACCTGAGATTAGGAGTT-G----GAGA
FWD 5. adbff31...      A--CCAGTACTAGTAAAGGACAGG-AAGGGAGCAGTGGTTCACGCC---AATCCCAGCAATTTGGGAGGCCAAGGTGGGTG--TCACCTGGG-TCAGGAGTT-G----GAGA
FWD 6. 531f71a...      A--CCGGTACTAGTAA-GGACAGGGA--GGAGCGATGGTTCACGCCTGTAATCCCAGCAATTTGGGAGGCCAAGGTGGGTAGATCACCTGAGATTAGGAGTT-G----GAGA
FWD 7. b477e23...      A--CCGGTACTAGTAAAGGACAGG-AAGGGAGCAGTGGTTCACGCCTGTAATCCCAGCAATTTGGGAGGCCAAGGTGGGTAGATCACCTGAGATTAGGAGTT-G----GAGA
FWD 8. 3a63a2d...      A--CCGGTACTAATGGAG-ACAGGGAAGGG--CAGTGGTTCACGCCTGTAATCCCAGCAATTTGGGAGGCCAAGGTGGGTAGATCACCTGAGATTAGGAGTT-G----GAGA
FWD 9. 31378e4...      A--CTGGTACTAGTAAAGGACAGGGAGAGGGCAGTGGTTTCACGCCTGTAATCCCAGCAATTTGGGAGGCCAAGGTGGGTAGATCACCTGAGATTAGGAGTT-GCTTGGAGA
FWD 10. cb40b7...      A--CCGGTACTAGTAAAGGACAGGGAAGGGAGCAGTGGTTCACACCTGTAATCCCAGCAATT-GGGAGGCCAAGGTGGGTAGATCACCTGAGGGTAGGAGTT-G----GAGA
FWD 11. 64f833...      ATTCTGGTACTAGTAAAGGAC-GGGAAGG---CAGTGGTTCACGCCTGTAATCCCAGCAATTTGGGAGGCTGAG-TGGGTAGATCACCTGAGATTAGGAGTT-G----GAGA
FWD 12. 6b0afd...      -------TGCTAGTAAAGGACAGGGA-GGGAGCAGTGGTTCACGCCTGTAATCCCAGCAATTTGGGAGGCCAAGGTGGGTAGATCACCTGAGATTAGGAGTT-G----GAGA
```

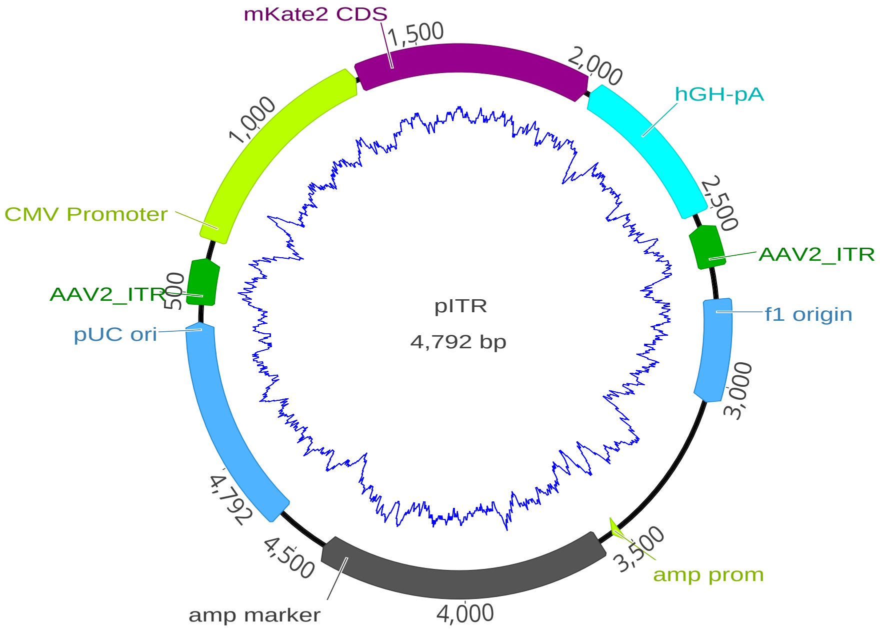
Supplementary Figure S10. Junction ITRs between genome-genome fusions have two D-sequences and sequencing quality drops downstream of the internal palindrome (BC). The 11 nt deletion present on producer plasmid level is absent in the genome-genome fusions (11 read examples mapped to *in silico* constructed genome-genome fusions).

Supplementary Figure S11



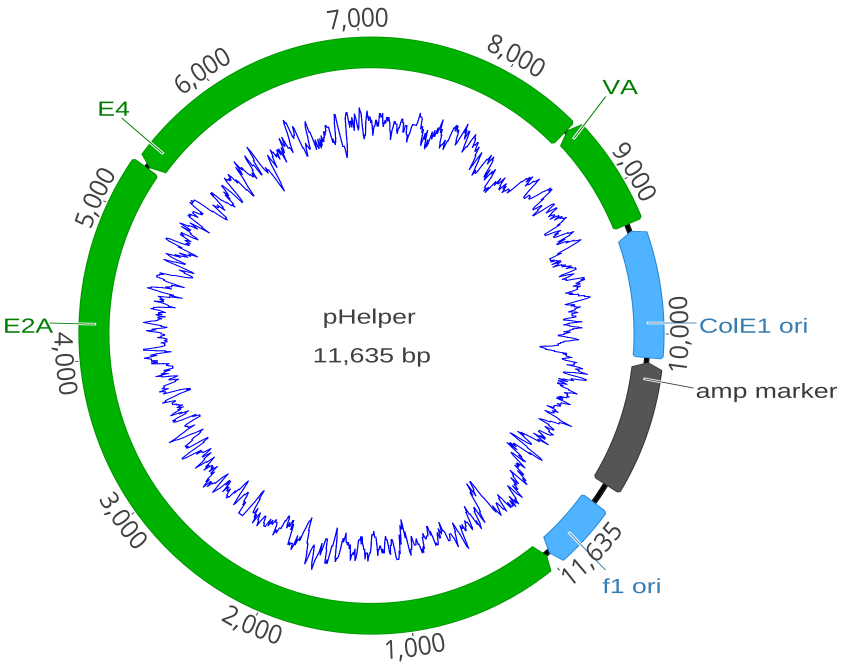
Supplementary Figure S11. p5 coverage drops sharply at position +12 of the promoter’s TATA-box.

Supplementary Figure S12



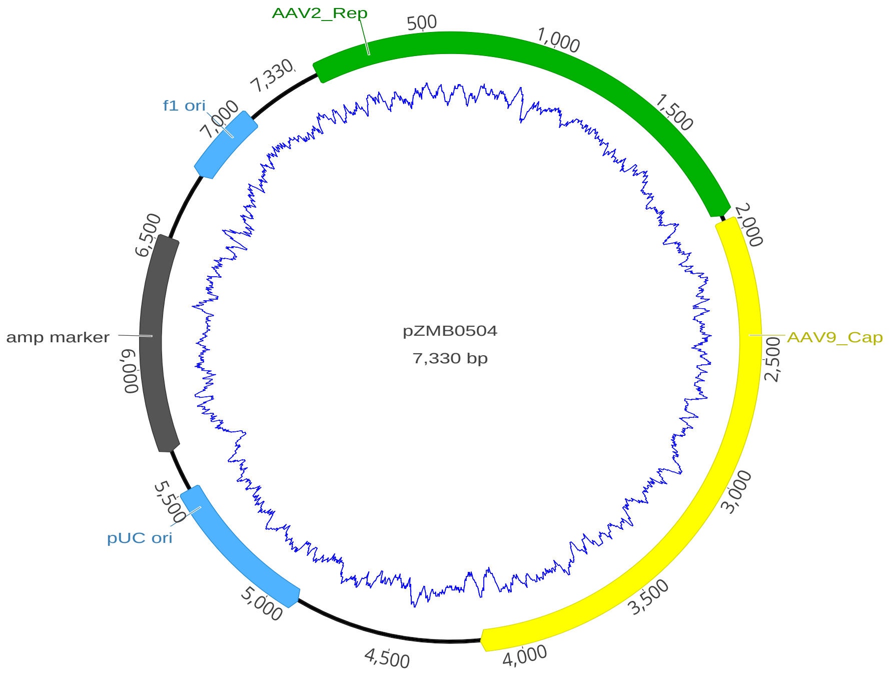
Supplementary Figure S12. Plasmid map of pITR (pUC19bb_ITR_EXS_CMV_mKate2_hGH-pA) with GC content graph. The right ITR (proximate to the hGH-pA) harbors a 11 nt deletion.

Supplementary Figure S13



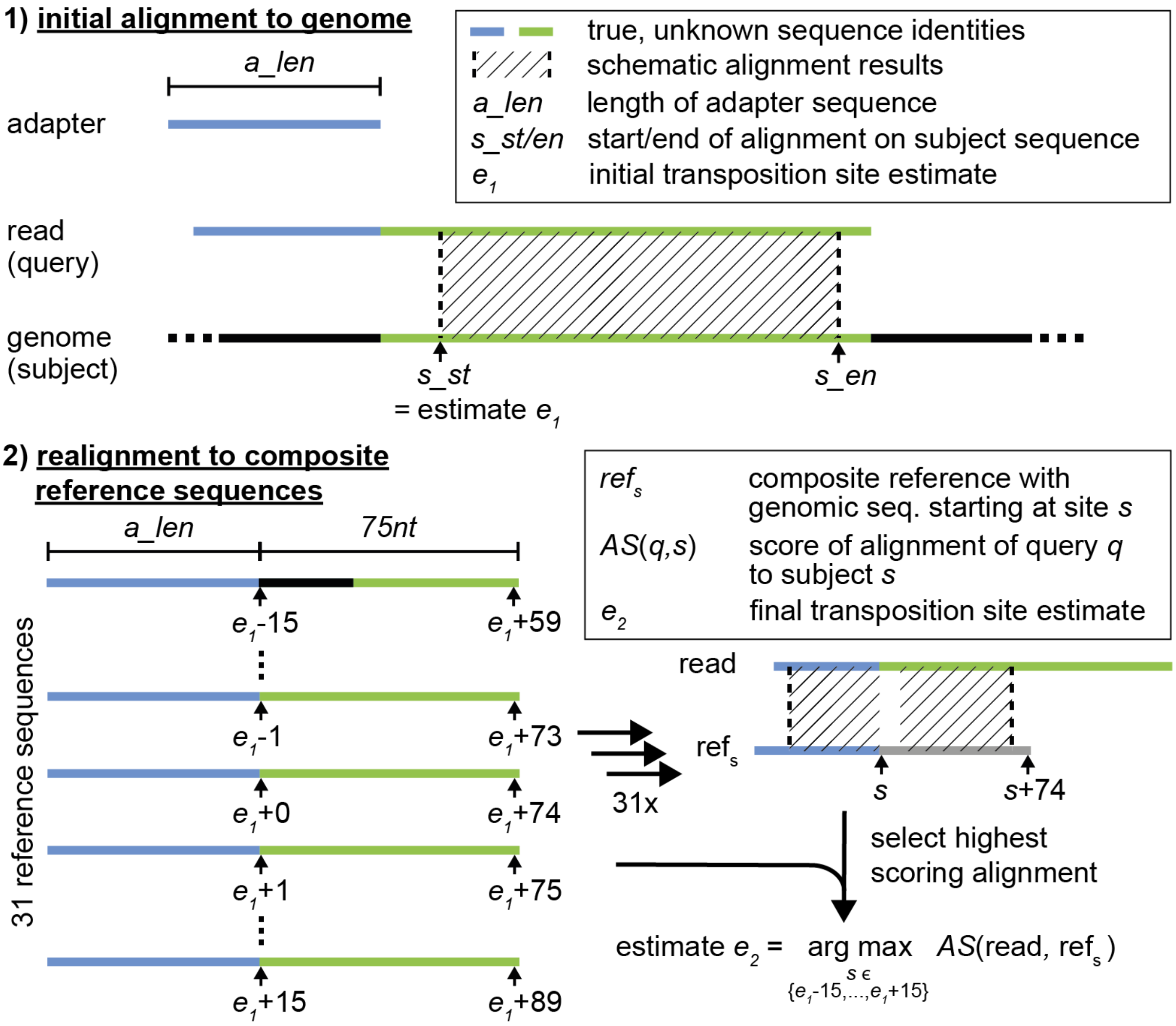
Supplementary Figure S13. Plasmid map of pHelper (Agilent Technologies) with GC content graph.

Supplementary Figure S14



Supplementary Figure S14. Plasmid map of pRepCap (pRep2Cap9) with GC content graph.

Supplementary Figure S15


1) initial alignment to genome
a_len
adapter
true, unknown sequence identities
schematic alignment results
a_len length of adapter sequence
s_st/en start/end of alignment on subject sequence
e_1 initial transposition site estimate
read (query)
genome (subject)
s_st = estimate e_1
s_en
2) realignment to composite reference sequences
a_len
75nt
31 reference sequences
e_1-15
e_1+59
e_1-1
e_1+73
e_1+0
e_1+74
e_1+1
e_1+75
e_1+15
e_1+89
ref_s composite reference with genomic seq. starting at site s
AS(q,s) score of alignment of query q to subject s
e_2 final transposition site estimate
read
ref_s
31x
s
s+74
select highest scoring alignment
estimate e_2 = arg max_{s ∈ {e_1-15,...,e_1+15}} AS(read, ref_s)

Supplementary Figure S15. Overview of the estimation of transposase insertion sites. 1) At first, each read is aligned to the genome using minimap2 and the map-ont preset, rejecting secondary alignments and alignments shorter than 100 nt. The start of the aligned genomic sequence is taken as the initial estimate e_1 of the transposase insertion site. 2) The estimate is refined by realigning the read against 31 composite reference sequences ref_s with $s \in \{e_1\text{-}15, \ldots, e_1, \ldots, e_1+15\}$, each of which is composed of the adapter sequence followed by 75 nt genomic sequence starting at site s of the genome. The final transposase insertion site estimate is set to site s of the reference sequence ref_s that produces the highest alignment score.