

Supplementary information

Robust direct digital-to-biological data storage in living cells

In the format provided by the authors and unedited

Supplementary Information

Robust direct digital-to-biological data storage in living cells

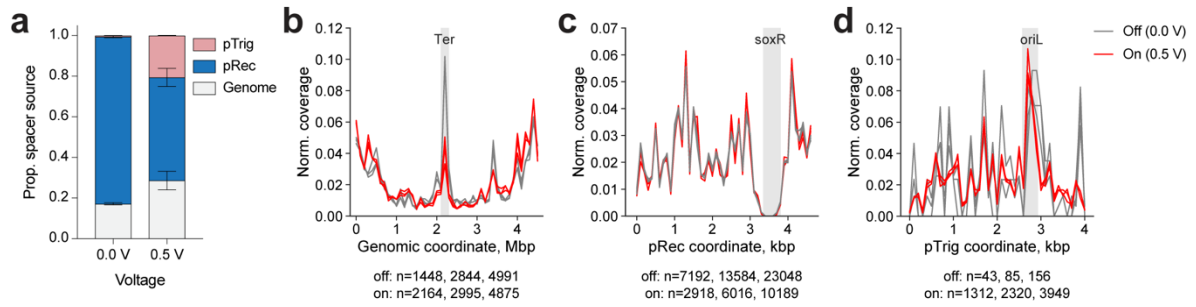
Sung Sun Yim, Ross M. McBee, Alan M. Song, Yiming Huang, Ravi U. Sheth, Harris
H. Wang*

*correspondence to: hw2429@columbia.edu

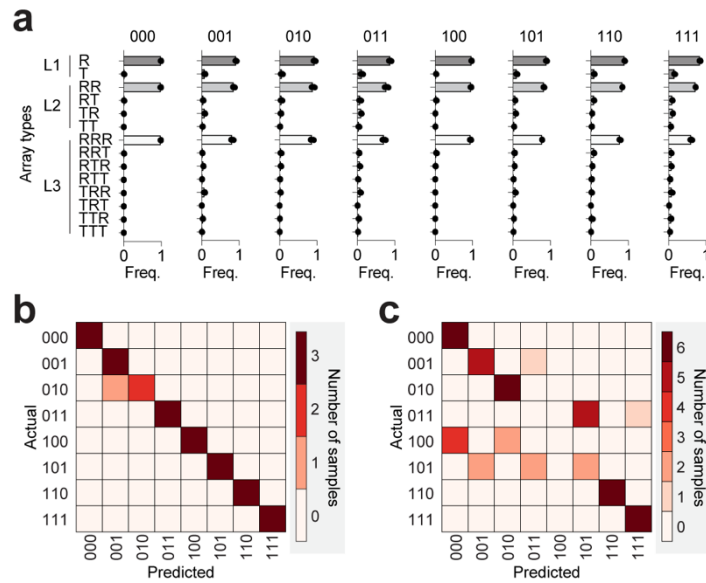
This PDF file includes:

Supplementary Figures 1 to 6

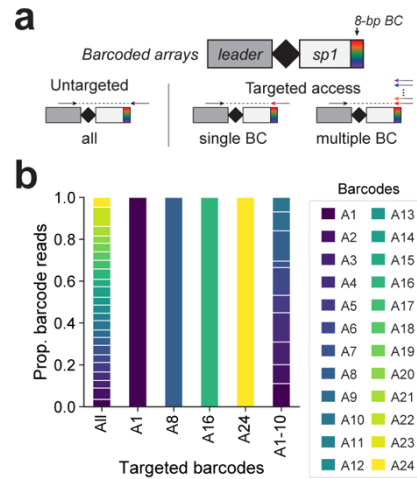
Supplementary Tables 1 to 3



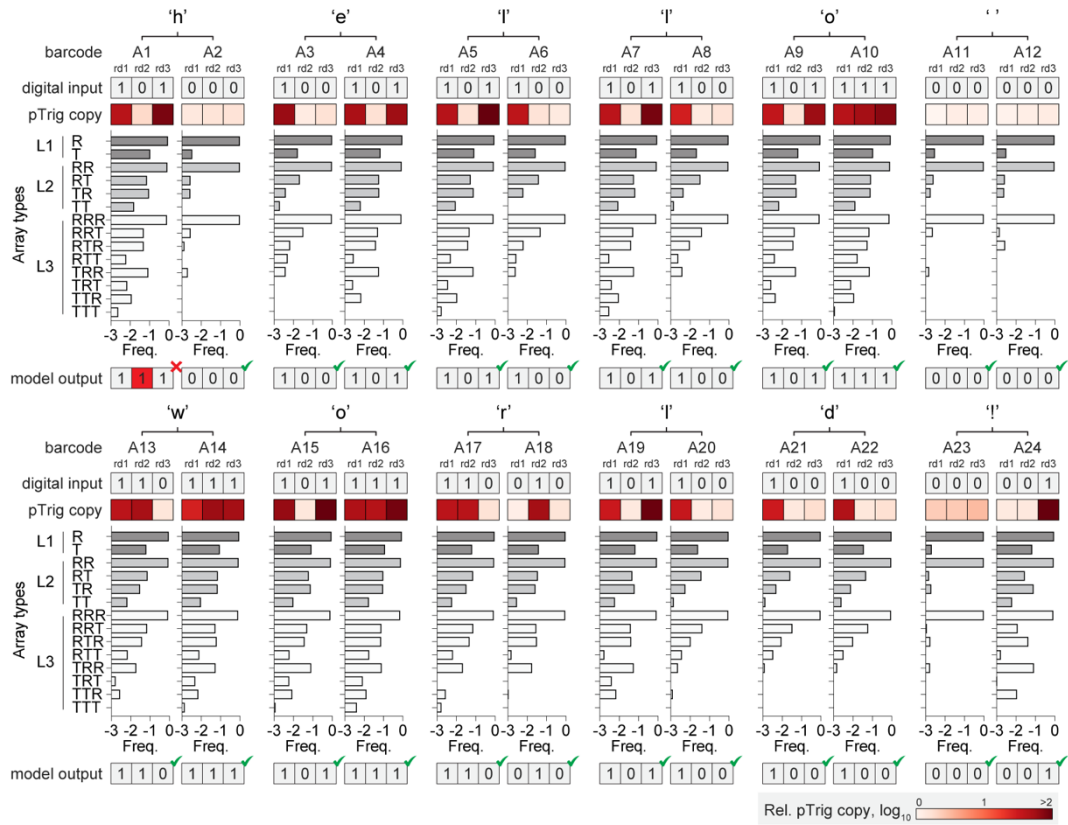
Supplementary Figure 1. Mapping spacers to their source location. (a) Proportion of source of newly incorporated spacers without (off, 0.0 V) and with (on, 0.5 V) electrical signal for 14 hours. (b) Genome-derived, (c) pRec-derived, and (d) pTrig-derived spacers were mapped to their protospacer location. Coverage was summed across 100 kb bins for genome or 100 bp bins for plasmids, and plotted. The three biological replicates without electrical stimulation (off, 0.0 V) are displayed in grey lines, and the three biological replicates with electrical stimulation (on, 0.5 V) are displayed in red lines. Genomic coordinate 0 is indexed as the midpoint of the *oriC* sequence. Spacers from the *soxR* gene region in pRec is excluded in the analysis as the region is present also in the genome. All measurements are based on three biological replicates. The number of spacers used in the analysis is displayed for each biological replicate.



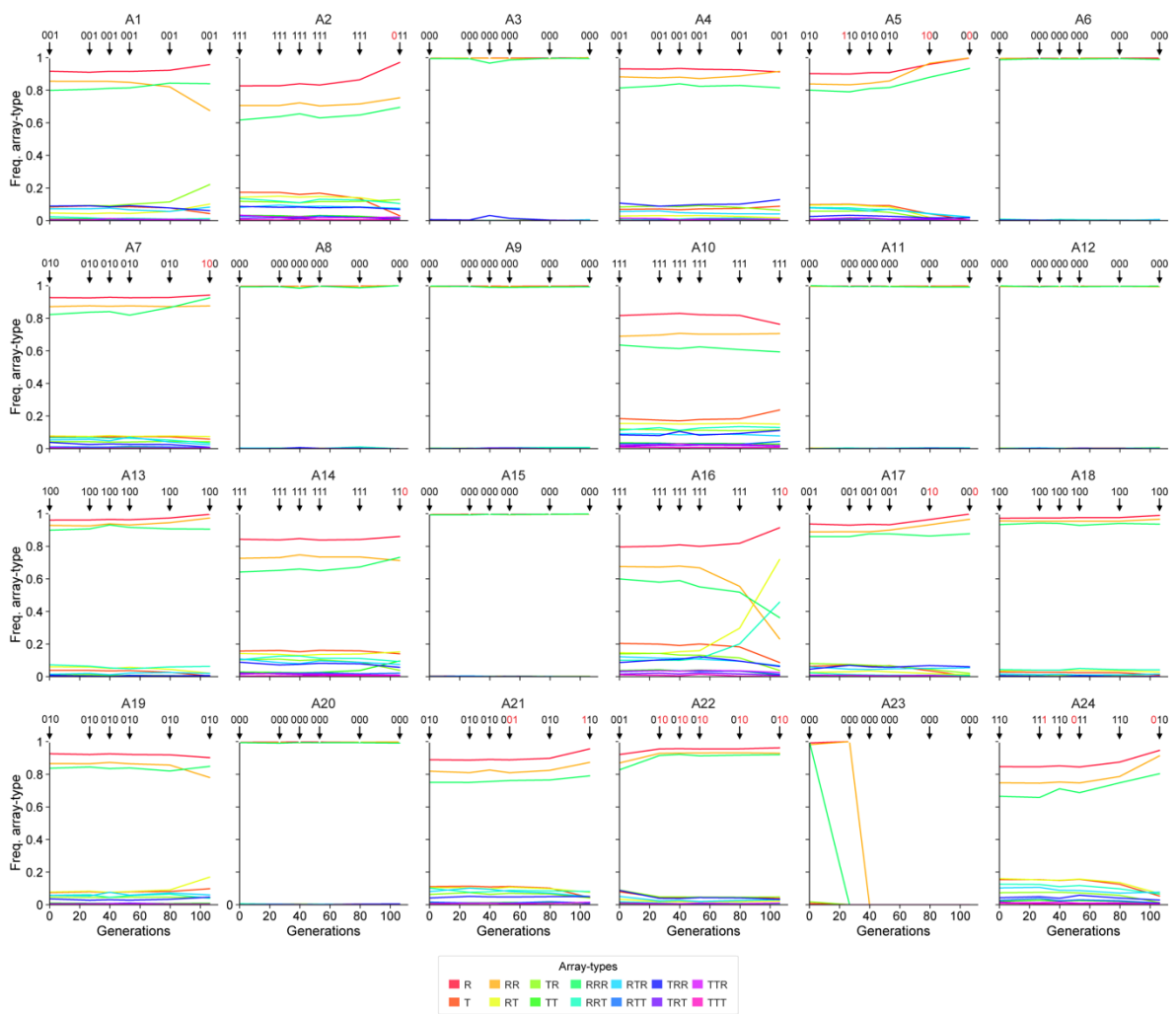
Supplementary Figure 2. Classification using Euclidean distance between observed and reference datasets. For comparison, Euclidean distance between observed (new dataset) and reference array-type frequencies was calculated, and the 3-bit profile with minimum distance to the observed data was selected as the predicted 3-bit profile as previously described in Sheth et al., *Science* **358**, 1457-1461 (2017). **(a)** The reference array-type frequencies (derived from the first three independent experiments described in **Figure 2d**, averaged) used for the classification. The frequencies values are same as in **Figure 2d** but plotted in linear scale. **(b)** Classification of the training datasets (the first three independent datasets that we extracted reference array-type frequencies). For the classification of each sample, average 163,541 of total sequencing reads with 67,620 of expanded arrays (or 23,827 of L2/L3 arrays) with uniquely mapping spacers were used. Bead-based size enrichment was performed to enrich for expanded arrays and deplete unexpanded arrays. **(c)** Classification of the test datasets (6 additional independent datasets). For classification of each sample, an average of 172,788 total sequencing reads with 89,928 reads of expanded arrays (or 38,295 of L2/L3 arrays) that uniquely map spacers were used. Bead-based size enrichment was performed to enrich for expanded arrays and deplete unexpanded arrays.



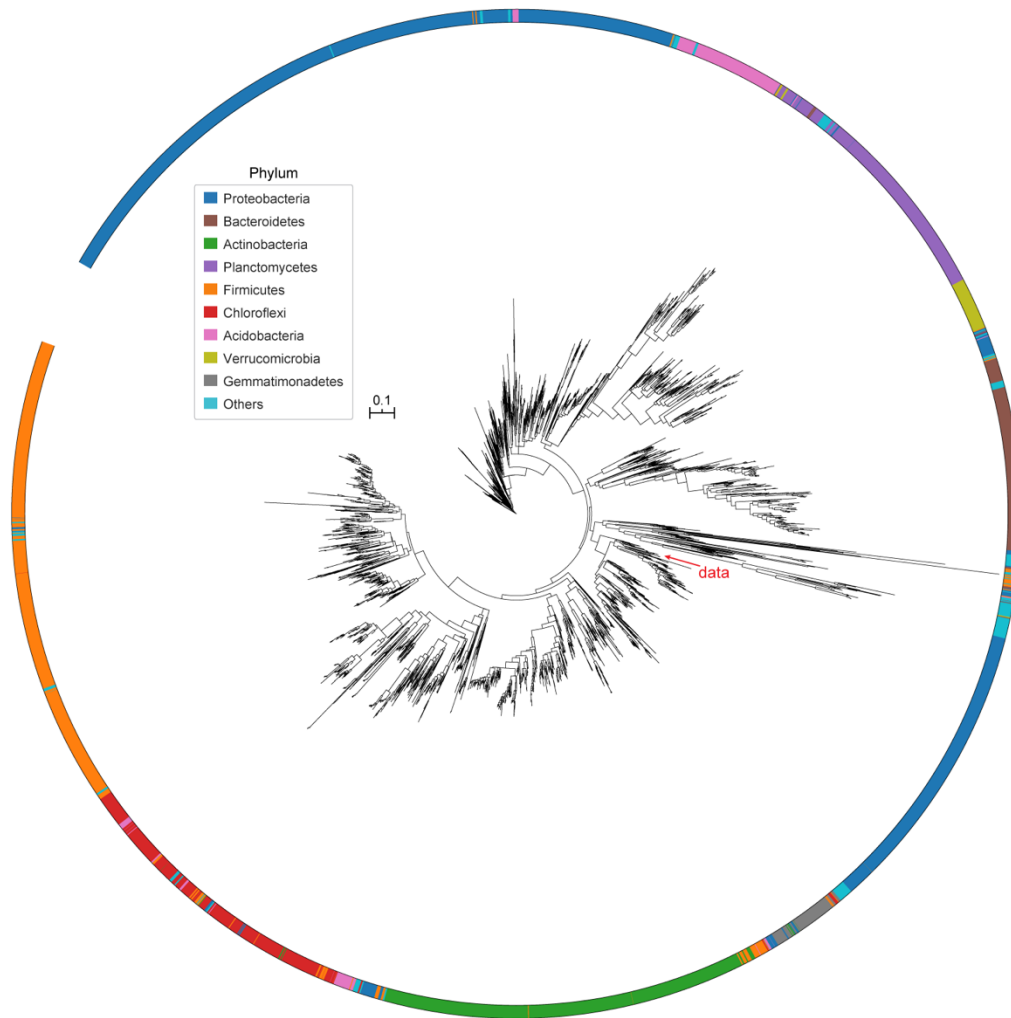
Supplementary Figure 3. Targeted access to specific barcoded cell populations in a mixed pool. (a) Schematic diagram of targeted amplification of barcoded CRISPR arrays in a mixed pool. **(b)** Targeted access to single or multiple barcoded cell populations in a mixed pool of 24 barcoded cell populations. For the targeted access to multiple (A1-10) barcoded cell populations, a single PCR reaction was performed using multiple primers simultaneously. All measurements are based on a single biological replicate.



Supplementary Figure 4. 'hello world!' encoded in barcoded *E. coli* populations by DEC 6-bit encoding. DEC 6-bit encoding table is expected to have higher error rates for data reconstruction as described in **Extended Data Fig. 7**. When the same 'hello world!' text message was encoded using DEC 6-bit encoding table, 1 bit out of 72 bits for the text message was misclassified, resulting in 'xello world!' returned from the classifier instead of 'hello world!'. For classification of each barcoded cell population, an average of 247,906 total sequencing reads with 147,073 reads of expanded arrays (or 102,475 of L2/L3 arrays) that uniquely map spacers were used. Bead-based size enrichment was performed to enrich for expanded arrays and deplete unexpanded arrays. Frequencies of array-types are in log₁₀ scale. All measurements are based on a single experimental study.



Supplementary Figure 5. Array-type fluctuations over time in replicating cells. The array-type frequencies for each array lengths in each barcoded cell population over the course of the passage experiment in **Extended Data Fig. 10** are plotted. Samples were collected at the time points indicated by arrows (day 0, 4, 6, 8, 12, and 16). For each sampling time point, the measured array-type frequencies were used for data retrieval and the classification results are displayed (black: correct, red: wrong). For classification of each sample, an average of 82,860 total sequencing reads with 40,502 reads of expanded arrays (or 17,139 of L2/L3 arrays) that uniquely map spacers were used. Bead-based size enrichment was performed to enrich for expanded arrays and deplete unexpanded arrays. All measurements are based on three biological replicates.



Supplementary Figure 6. Phylogenetic diversity of a mixed microbial community of data-encoded *E. coli* cells and natural soil microbiota. Maximum-likelihood phylogenetic tree of all 4083 OTUs detected in the mixed community based on multiple sequence alignment of V4 region of 16S rRNA sequences using MAFFT (distance scale of 0.1).

Supplementary Table 1. Barcodes for CRISPR arrays

Array#	Array_sequence	BC
A1	TGTTGTTTGTGTGATACTATAAAGTTGGTAGATTGTGACTGGCTTAAAAAATCATTAATTAATAATAGGTTATGTTTAGAGTGTTC ^{DR} CCCCG CGCCAGCGGGGATAAAACC ^{Spacer1} GAGCACAAATATCATCGCTCAAACCACTTACGGTAAGGCGATCTAATCTAGCGCGACGTC	TAAG GGGA
A2	TGTTGTTTGTGTGATACTATAAAGTTGGTAGATTGTGACTGGCTTAAAAAATCATTAATTAATAATAGGTTATGTTTAGAGTGTTC ^{DR} CCCCG CGCCAGCGGGGATAAAACC ^{Spacer1} GAGCACAAATATCATCGCTCAAACCACTTACGGCGTACTAGTCTAATCTAGCGCGACGTC	CGTA CTAG
A3	TGTTGTTTGTGTGATACTATAAAGTTGGTAGATTGTGACTGGCTTAAAAAATCATTAATTAATAATAGGTTATGTTTAGAGTGTTC ^{DR} CCCCG CGCCAGCGGGGATAAAACC ^{Spacer1} GAGCACAAATATCATCGCTCAAACCACTTACGGAGGCAGAACTAATCTAGCGCGACGTC	AGGC AGAA
A4	TGTTGTTTGTGTGATACTATAAAGTTGGTAGATTGTGACTGGCTTAAAAAATCATTAATTAATAATAGGTTATGTTTAGAGTGTTC ^{DR} CCCCG CGCCAGCGGGGATAAAACC ^{Spacer1} GAGCACAAATATCATCGCTCAAACCACTTACGGTCTGAGCTAATCTAGCGCGACGTC	TCCT GAGC
A5	TGTTGTTTGTGTGATACTATAAAGTTGGTAGATTGTGACTGGCTTAAAAAATCATTAATTAATAATAGGTTATGTTTAGAGTGTTC ^{DR} CCCCG CGCCAGCGGGGATAAAACC ^{Spacer1} GAGCACAAATATCATCGCTCAAACCACTTACGGGACTCCTAATCTAGCGCGACGTC	GGAC TCCT
A6	TGTTGTTTGTGTGATACTATAAAGTTGGTAGATTGTGACTGGCTTAAAAAATCATTAATTAATAATAGGTTATGTTTAGAGTGTTC ^{DR} CCCCG CGCCAGCGGGGATAAAACC ^{Spacer1} GAGCACAAATATCATCGCTCAAACCACTTACGGTAGGCATGTCTAATCTAGCGCGACGTC	TAGG CATG
A7	TGTTGTTTGTGTGATACTATAAAGTTGGTAGATTGTGACTGGCTTAAAAAATCATTAATTAATAATAGGTTATGTTTAGAGTGTTC ^{DR} CCCCG CGCCAGCGGGGATAAAACC ^{Spacer1} GAGCACAAATATCATCGCTCAAACCACTTACGGCTCTACTAATCTAGCGCGACGTC	CTCT CTAC
A8	TGTTGTTTGTGTGATACTATAAAGTTGGTAGATTGTGACTGGCTTAAAAAATCATTAATTAATAATAGGTTATGTTTAGAGTGTTC ^{DR} CCCCG CGCCAGCGGGGATAAAACC ^{Spacer1} GAGCACAAATATCATCGCTCAAACCACTTACGGCGAGGCTGTCTAATCTAGCGCGACGTC	CGAG GCTG
A9	TGTTGTTTGTGTGATACTATAAAGTTGGTAGATTGTGACTGGCTTAAAAAATCATTAATTAATAATAGGTTATGTTTAGAGTGTTC ^{DR} CCCCG CGCCAGCGGGGATAAAACC ^{Spacer1} GAGCACAAATATCATCGCTCAAACCACTTACGGAAGAGGCACTAATCTAGCGCGACGTC	AAGA GGCA
A10	TGTTGTTTGTGTGATACTATAAAGTTGGTAGATTGTGACTGGCTTAAAAAATCATTAATTAATAATAGGTTATGTTTAGAGTGTTC ^{DR} CCCCG CGCCAGCGGGGATAAAACC ^{Spacer1} GAGCACAAATATCATCGCTCAAACCACTTACGGTAGAGGACTAATCTAGCGCGACGTC	GTAG AGGA
A11	TGTTGTTTGTGTGATACTATAAAGTTGGTAGATTGTGACTGGCTTAAAAAATCATTAATTAATAATAGGTTATGTTTAGAGTGTTC ^{DR} CCCCG CGCCAGCGGGGATAAAACC ^{Spacer1} GAGCACAAATATCATCGCTCAAACCACTTACGGGCTCATGACTAATCTAGCGCGACGTC	GCTC ATGA
A12	TGTTGTTTGTGTGATACTATAAAGTTGGTAGATTGTGACTGGCTTAAAAAATCATTAATTAATAATAGGTTATGTTTAGAGTGTTC ^{DR} CCCCG CGCCAGCGGGGATAAAACC ^{Spacer1} GAGCACAAATATCATCGCTCAAACCACTTACGGATCTCAGTCTAATCTAGCGCGACGTC	ATCT CAGG
A13	TGTTGTTTGTGTGATACTATAAAGTTGGTAGATTGTGACTGGCTTAAAAAATCATTAATTAATAATAGGTTATGTTTAGAGTGTTC ^{DR} CCCCG CGCCAGCGGGGATAAAACC ^{Spacer1} GAGCACAAATATCATCGCTCAAACCACTTACGGACTCGCTAATCTAGCGCGACGTC	ACTC GCTA
A14	TGTTGTTTGTGTGATACTATAAAGTTGGTAGATTGTGACTGGCTTAAAAAATCATTAATTAATAATAGGTTATGTTTAGAGTGTTC ^{DR} CCCCG CGCCAGCGGGGATAAAACC ^{Spacer1} GAGCACAAATATCATCGCTCAAACCACTTACGGGAGACTACTAATCTAGCGCGACGTC	GGAG CTAC
A15	TGTTGTTTGTGTGATACTATAAAGTTGGTAGATTGTGACTGGCTTAAAAAATCATTAATTAATAATAGGTTATGTTTAGAGTGTTC ^{DR} CCCCG CGCCAGCGGGGATAAAACC ^{Spacer1} GAGCACAAATATCATCGCTCAAACCACTTACGGCGTAGTACTAATCTAGCGCGACGTC	GCGT AGTA
A16	TGTTGTTTGTGTGATACTATAAAGTTGGTAGATTGTGACTGGCTTAAAAAATCATTAATTAATAATAGGTTATGTTTAGAGTGTTC ^{DR} CCCCG CGCCAGCGGGGATAAAACC ^{Spacer1} GAGCACAAATATCATCGCTCAAACCACTTACGGCGGAGCCTAATCTAGCGCGACGTC	CGGA GCCT
A17	TGTTGTTTGTGTGATACTATAAAGTTGGTAGATTGTGACTGGCTTAAAAAATCATTAATTAATAATAGGTTATGTTTAGAGTGTTC ^{DR} CCCCG CGCCAGCGGGGATAAAACC ^{Spacer1} GAGCACAAATATCATCGCTCAAACCACTTACGGTACGCTGCTAATCTAGCGCGACGTC	TACG CTGC
A18	TGTTGTTTGTGTGATACTATAAAGTTGGTAGATTGTGACTGGCTTAAAAAATCATTAATTAATAATAGGTTATGTTTAGAGTGTTC ^{DR} CCCCG CGCCAGCGGGGATAAAACC ^{Spacer1} GAGCACAAATATCATCGCTCAAACCACTTACGGATCGCAGTCTAATCTAGCGCGACGTC	ATGC GCAG
A19	TGTTGTTTGTGTGATACTATAAAGTTGGTAGATTGTGACTGGCTTAAAAAATCATTAATTAATAATAGGTTATGTTTAGAGTGTTC ^{DR} CCCCG CGCCAGCGGGGATAAAACC ^{Spacer1} GAGCACAAATATCATCGCTCAAACCACTTACGGTAGCGCTCTAATCTAGCGCGACGTC	TAGC GCTC
A20	TGTTGTTTGTGTGATACTATAAAGTTGGTAGATTGTGACTGGCTTAAAAAATCATTAATTAATAATAGGTTATGTTTAGAGTGTTC ^{DR} CCCCG CGCCAGCGGGGATAAAACC ^{Spacer1} GAGCACAAATATCATCGCTCAAACCACTTACGGACTGAGCGTCTAATCTAGCGCGACGTC	ACTG AGCG
A21	TGTTGTTTGTGTGATACTATAAAGTTGGTAGATTGTGACTGGCTTAAAAAATCATTAATTAATAATAGGTTATGTTTAGAGTGTTC ^{DR} CCCCG CGCCAGCGGGGATAAAACC ^{Spacer1} GAGCACAAATATCATCGCTCAAACCACTTACGGCCTAAGACTAATCTAGCGCGACGTC	CCTA AGAC
A22	TGTTGTTTGTGTGATACTATAAAGTTGGTAGATTGTGACTGGCTTAAAAAATCATTAATTAATAATAGGTTATGTTTAGAGTGTTC ^{DR} CCCCG CGCCAGCGGGGATAAAACC ^{Spacer1} GAGCACAAATATCATCGCTCAAACCACTTACGGCGATCAGTCTAATCTAGCGCGACGTC	CGAT CAGT
A23	TGTTGTTTGTGTGATACTATAAAGTTGGTAGATTGTGACTGGCTTAAAAAATCATTAATTAATAATAGGTTATGTTTAGAGTGTTC ^{DR} CCCCG CGCCAGCGGGGATAAAACC ^{Spacer1} GAGCACAAATATCATCGCTCAAACCACTTACGGTGCAGCTAATCTAGCGCGACGTC	TGCA GCTA
A24	TGTTGTTTGTGTGATACTATAAAGTTGGTAGATTGTGACTGGCTTAAAAAATCATTAATTAATAATAGGTTATGTTTAGAGTGTTC ^{DR} CCCCG CGCCAGCGGGGATAAAACC ^{Spacer1} GAGCACAAATATCATCGCTCAAACCACTTACGGTCGACGCTCTAATCTAGCGCGACGTC	TCGA CGTC

Colors: Leader-DR-Spacer1-BC-Pad

Supplementary Table 2. Simple parity check error correction

Actual input: 00000[0] / Error corrected output: 00000[0] (100%)

Possible 1 st 3-bit call	Possible 2 nd 3-bit call	Prediction probability	# of 1 in the 1 st 5 bits	Checksum	Checksum correct?	Can be recovered properly?
000	00[0]	100%	0	0	T	T

Actual input: 00001[1] / Error corrected output: 00001[1] (100%)

Possible 1 st 3-bit call	Possible 2 nd 3-bit call	Prediction probability	# of 1 in the first 5 bits	Checksum	Checksum correct?	Can be recovered properly?
000	01[1]	83.33%	1	1	T	T
000	00[1]	16.67%	0	0	F	T

Actual input: 00010[1] / Error corrected output: 00010[1] (100%)

Possible 1 st 3-bit call	Possible 2 nd 3-bit call	Prediction probability	# of 1 in the first 5 bits	Checksum	Checksum correct?	Can be recovered properly?
000	10[1]	83.33%	1	1	T	T
000	11[1]	16.67%	2	0	F	T

Actual input: 00011[0] / Error corrected output: 00011[0] (100%)

Possible 1 st 3-bit call	Possible 2 nd 3-bit call	Prediction probability	# of 1 in the first 5 bits	Expected checksum	Checksum correct?	Can be recovered properly?
000	11[0]	83.33%	2	0	T	T
000	01[0]	16.67%	1	1	F	T

Actual input: 00100[1] / Error corrected output: 00100[1] (100%)

Possible 1 st 3-bit call	Possible 2 nd 3-bit call	Prediction probability	# of 1 in the first 5 bits	Expected checksum	Checksum correct?	Can be recovered properly?
001	00[1]	100%	1	1	T	T

Actual input: 00101[0] / Error corrected output: 00101[0] (100%)

Possible 1 st 3-bit call	Possible 2 nd 3-bit call	Prediction probability	# of 1 in the first 5 bits	Expected checksum	Checksum correct?	Can be recovered properly?
001	01[0]	100%	2	0	T	T

Actual input: 00110[0] / Error corrected output: 00110[0] (100%)

Possible 1 st 3-bit call	Possible 2 nd 3-bit call	Prediction probability	# of 1 in the first 5 bits	Expected checksum	Checksum correct?	Can be recovered properly?
001	10[0]	100%	2	0	T	T

Actual input: 00111[1] / Error corrected output: 00111[1] (100%)

Possible 1 st 3-bit call	Possible 2 nd 3-bit call	Prediction probability	# of 1 in the first 5 bits	Expected checksum	Checksum correct?	Can be recovered properly?
001	11[1]	100%	3	1	T	T

Actual input: 01000[1] / Error corrected output: 01000[1] (100%)

Possible 1 st 3-bit call	Possible 2 nd 3-bit call	Prediction probability	# of 1 in the first 5 bits	Expected checksum	Checksum correct?	Can be recovered properly?
010	00[1]	100%	1	1	T	T

Actual input: 01001[0] / Error corrected output: 01001[0] (100%)

Possible 1 st 3-bit call	Possible 2 nd 3-bit call	Prediction probability	# of 1 in the first 5 bits	Expected checksum	Checksum correct?	Can be recovered properly?
010	01[0]	100%	2	0	T	T

Actual input: 01010[0] / Error corrected output: 01010[0] (100%)

Possible 1 st 3-bit call	Possible 2 nd 3-bit call	Prediction probability	# of 1 in the first 5 bits	Expected checksum	Checksum correct?	Can be recovered properly?
010	10[0]	100%	2	0	T	T

Actual input: 01011[1] / Error corrected output: 01011[1] (100%)

Possible 1 st 3-bit call	Possible 2 nd 3-bit call	Prediction probability	# of 1 in the first 5 bits	Expected checksum	Checksum correct?	Can be recovered properly?
010	11[1]	100%	3	1	T	T

Actual input: 01100[0] / Error corrected output: 01100[0] (100%)

Possible 1 st 3-bit call	Possible 2 nd 3-bit call	Prediction probability	# of 1 in the first 5 bits	Expected checksum	Checksum correct?	Can be recovered properly?
011	00[0]	83.33%	2	0	T	T
001	00[0]	16.67%	1	1	F	T

Actual input: 01101[1] / Error corrected output: 01101[1] (97.22%), 00100[1] (2.78%)

Possible 1 st 3-bit call	Possible 2 nd 3-bit call	Prediction probability	# of 1 in the first 5 bits	Expected checksum	Checksum correct?	Can be recovered properly?
011	01[1]	69.44%	3	1	T	T
011	00[1]	13.89%	2	0	F	T
001	01[1]	13.89%	2	0	F	T
001	00[1]	2.78%	1	1	T	F

Actual input: 01110[1] / Error corrected output: 01110[1] (97.22%), 00111[1] (2.78%)

Possible 1 st 3-bit call	Possible 2 nd 3-bit call	Prediction probability	# of 1 in the first 5 bits	Expected checksum	Checksum correct?	Can be recovered properly?
011	10[1]	69.44%	3	1	T	T
011	11[1]	13.89%	4	0	F	T
001	10[1]	13.89%	2	0	F	T
001	11[1]	2.78%	3	1	T	F

Actual input: 01111[0] / Error corrected output: 01111[0] (97.22%), 00101[0] (2.78%)

Possible 1 st 3-bit call	Possible 2 nd 3-bit call	Prediction probability	# of 1 in the first 5 bits	Expected checksum	Checksum correct?	Can be recovered properly?
011	11[0]	69.44%	4	0	T	T
011	01[0]	13.89%	3	1	F	T
001	11[0]	13.89%	3	1	F	T
001	01[0]	2.78%	2	0	T	F

Actual input: 10000[1] / Error corrected output: 10000[1] (100%)

Possible 1 st 3-bit call	Possible 2 nd 3-bit call	Prediction probability	# of 1 in the first 5 bits	Expected checksum	Checksum correct?	Can be recovered properly?
100	00[1]	100%	1	1	T	T

Actual input: 10001[0] / Error corrected output: 10001[0] (100%)

Possible 1 st 3-bit call	Possible 2 nd 3-bit call	Prediction probability	# of 1 in the first 5 bits	Expected checksum	Checksum correct?	Can be recovered properly?
100	01[0]	100%	2	0	T	T

Actual input: 10010[0] / Error corrected output: 10010[0] (100%)

Possible 1 st 3-bit call	Possible 2 nd 3-bit call	Prediction probability	# of 1 in the first 5 bits	Expected checksum	Checksum correct?	Can be recovered properly?
100	10[0]	100%	2	0	T	T

Actual input: 10011[1] / Error corrected output: 10011[1] (100%)

Possible 1 st 3-bit call	Possible 2 nd 3-bit call	Prediction probability	# of 1 in the first 5 bits	Expected checksum	Checksum correct?	Can be recovered properly?
100	11[1]	100%	3	1	T	T

Actual input: 10100[0] / Error corrected output: 10100[0] (100%)

Possible 1 st 3-bit call	Possible 2 nd 3-bit call	Prediction probability	# of 1 in the first 5 bits	Expected checksum	Checksum correct?	Can be recovered properly?
101	00[0]	83.33%	2	0	T	T
111	00[0]	16.67%	3	1	F	T

Actual input: 10101[1] / Error corrected output: 10101[1] (97.22%), 11100[1] (2.78%)

Possible 1 st 3-bit call	Possible 2 nd 3-bit call	Prediction probability	# of 1 in the first 5 bits	Expected checksum	Checksum correct?	Can be recovered properly?
101	01[1]	69.44%	3	1	T	T
101	00[1]	13.89%	2	0	F	T
111	01[1]	13.89%	4	0	F	T
111	00[1]	2.78%	3	1	T	F

Actual input: 10110[1] / Error corrected output: 10110[1] (97.22%), 11111[1] (2.78%)

Possible 1 st 3-bit call	Possible 2 nd 3-bit call	Prediction probability	# of 1 in the first 5 bits	Expected checksum	Checksum correct?	Can be recovered properly?
101	10[1]	69.44%	3	1	T	T
101	11[1]	13.89%	4	0	F	T
111	10[1]	13.89%	4	0	F	T
111	11[1]	2.78%	5	1	T	F

Actual input: 10111[0] / Error corrected output: 10111[0] (97.22%), 11101[0] (2.78%)

Possible 1 st 3-bit call	Possible 2 nd 3-bit call	Prediction probability	# of 1 in the first 5 bits	Expected checksum	Checksum correct?	Can be recovered properly?
101	11[0]	69.44%	4	0	T	T
101	01[0]	13.89%	3	1	F	T
111	11[0]	13.89%	5	1	F	T
111	01[0]	2.78%	4	0	T	F

Actual input: 11000[0] / Error corrected output: 11000[0] (100%)

Possible 1 st 3-bit call	Possible 2 nd 3-bit call	Prediction probability	# of 1 in the first 5 bits	Expected checksum	Checksum correct?	Can be recovered properly?
110	00[0]	83.33%	2	0	T	T
010	00[0]	16.67%	1	1	F	T

Actual input: 11001[1] / Error corrected output: 11001[1] (97.22%), 01000[1] (2.78%)

Possible 1 st 3-bit call	Possible 2 nd 3-bit call	Prediction probability	# of 1 in the first 5 bits	Expected checksum	Checksum correct?	Can be recovered properly?
110	01[1]	69.44%	3	1	T	T
110	00[1]	13.89%	2	0	F	T
010	01[1]	13.89%	2	0	F	T
010	00[1]	2.78%	1	1	T	F

Actual input: 11010[1] / Error corrected output: 11010[1] (97.22%), 01011[1] (2.78%)

Possible 1 st 3-bit call	Possible 2 nd 3-bit call	Prediction probability	# of 1 in the first 5 bits	Expected checksum	Checksum correct?	Can be recovered properly?
110	10[1]	69.44%	3	1	T	T
110	11[1]	13.89%	4	0	F	T
010	10[1]	13.89%	2	0	F	T
010	11[1]	2.78%	3	1	T	F

Actual input: 11011[0] / Error corrected output: 11011[0] (97.22%), 01001[0] (2.78%)

Possible 1 st 3-bit call	Possible 2 nd 3-bit call	Prediction probability	# of 1 in the first 5 bits	Expected checksum	Checksum correct?	Can be recovered properly?
110	11[0]	69.44%	4	0	T	T
110	01[0]	13.89%	3	1	F	T
010	11[0]	13.89%	3	1	F	T
010	01[0]	2.78%	2	0	T	F

Actual input: 11100[1] / Error corrected output: 11100[1] (100%)

Possible 1 st 3-bit call	Possible 2 nd 3-bit call	Prediction probability	# of 1 in the first 5 bits	Expected checksum	Checksum correct?	Can be recovered properly?
111	00[1]	100%	3	1	T	T

Actual input: 11101[0] / Error corrected output: 11101[0] (100%)

Possible 1 st 3-bit call	Possible 2 nd 3-bit call	Prediction probability	# of 1 in the first 5 bits	Expected checksum	Checksum correct?	Can be recovered properly?
111	01[0]	100%	4	0	T	T

Actual input: 11110[0] / Error corrected output: 11110[0] (100%)

Possible 1 st 3-bit call	Possible 2 nd 3-bit call	Prediction probability	# of 1 in the first 5 bits	Expected checksum	Checksum correct?	Can be recovered properly?
111	10[0]	100%	4	0	T	T

Actual input: 11111[1] / Error corrected output: 11111[1] (100%)

Possible 1 st 3-bit call	Possible 2 nd 3-bit call	Prediction probability	# of 1 in the first 5 bits	Expected checksum	Checksum correct?	Can be recovered properly?
111	11[1]	100%	5	1	T	T

Supplementary Table 3. Primers used in this study

Seq_primer	Sequence
CB501	AATGATACGGCGACCACCGAGATCTACACTAGATCGCctggcttaaaaaatcattaatta ataataggttatgttaga
CB502	AATGATACGGCGACCACCGAGATCTACACCTCTCTATctggcttaaaaaatcattaatta ataataggttatgttaga
CB503	AATGATACGGCGACCACCGAGATCTACACTATCCTCTctggcttaaaaaatcattaatta ataataggttatgttaga
CB504	AATGATACGGCGACCACCGAGATCTACACAGAGTAGActggcttaaaaaatcattaatta ataataggttatgttaga
CB505	AATGATACGGCGACCACCGAGATCTACACGTAAGGAGctggcttaaaaaatcattaatt aataataggttatgttaga
CB506	AATGATACGGCGACCACCGAGATCTACACACTGCATActggcttaaaaaatcattaatta ataataggttatgttaga
CB507	AATGATACGGCGACCACCGAGATCTACACAAGGAGTctggcttaaaaaatcattaatta ataataggttatgttaga
CB508	AATGATACGGCGACCACCGAGATCTACACCTAAGCCTctggcttaaaaaatcattaatta ataataggttatgttaga
CB509	AATGATACGGCGACCACCGAGATCTACACCGTCTAATctggcttaaaaaatcattaatta ataataggttatgttaga
CB510	AATGATACGGCGACCACCGAGATCTACACTCTCTCCGctggcttaaaaaatcattaatta ataataggttatgttaga
CB511	AATGATACGGCGACCACCGAGATCTACACTCGACTAGctggcttaaaaaatcattaatta ataataggttatgttaga
CB512	AATGATACGGCGACCACCGAGATCTACACTTCTAGCTctggcttaaaaaatcattaatta ataataggttatgttaga
CB513	AATGATACGGCGACCACCGAGATCTACACCCTAGAGTctggcttaaaaaatcattaatta ataataggttatgttaga
CB514	AATGATACGGCGACCACCGAGATCTACACGCGTAAGActggcttaaaaaatcattaatt aataataggttatgttaga
CB515	AATGATACGGCGACCACCGAGATCTACACCTATTAAGctggcttaaaaaatcattaatta ataataggttatgttaga
CB516	AATGATACGGCGACCACCGAGATCTACACAAGGCTATctggcttaaaaaatcattaatta ataataggttatgttaga
CB701	CAAGCAGAAGACGGCATAACGAGATTCGCCTTAaggtttgagcgatgatatttgct
CB702	CAAGCAGAAGACGGCATAACGAGATCTAGTACGggtttgagcgatgatatttgct
CB703	CAAGCAGAAGACGGCATAACGAGATTTCTGCCTggtttgagcgatgatatttgct
CB704	CAAGCAGAAGACGGCATAACGAGATGCTCAGGAggtttgagcgatgatatttgct
CB705	CAAGCAGAAGACGGCATAACGAGATAGGAGTCCggtttgagcgatgatatttgct

CB706	CAAGCAGAAGACGGCATAACGAGAT	CATGCCTAgggttgagcgatgatatttgct
CB707	CAAGCAGAAGACGGCATAACGAGAT	GTAGAGAGgggttgagcgatgatatttgct
CB708	CAAGCAGAAGACGGCATAACGAGAT	CCTCTCTGgggttgagcgatgatatttgct
CB709	CAAGCAGAAGACGGCATAACGAGAT	AGCGTAGCgggttgagcgatgatatttgct
CB710	CAAGCAGAAGACGGCATAACGAGAT	CAGCCTCGgggttgagcgatgatatttgct
CB711	CAAGCAGAAGACGGCATAACGAGAT	TGCCTCTTgggttgagcgatgatatttgct
CB712	CAAGCAGAAGACGGCATAACGAGAT	TCCTCTACgggttgagcgatgatatttgct
CB713	CAAGCAGAAGACGGCATAACGAGAT	TCATGAGCgggttgagcgatgatatttgct
CB714	CAAGCAGAAGACGGCATAACGAGAT	CCTGAGATgggttgagcgatgatatttgct
CB715	CAAGCAGAAGACGGCATAACGAGAT	TAGCGAGTgggttgagcgatgatatttgct
CB716	CAAGCAGAAGACGGCATAACGAGAT	GTAGCTCCgggttgagcgatgatatttgct
CB717	CAAGCAGAAGACGGCATAACGAGAT	TACTACGCgggttgagcgatgatatttgct
CB718	CAAGCAGAAGACGGCATAACGAGAT	AGGCTCCGgggttgagcgatgatatttgct
CB719	CAAGCAGAAGACGGCATAACGAGAT	GCAGCGTAgggttgagcgatgatatttgct
CB720	CAAGCAGAAGACGGCATAACGAGAT	CTGCGCATgggttgagcgatgatatttgct
CB721	CAAGCAGAAGACGGCATAACGAGAT	GAGCGCTAgggttgagcgatgatatttgct
CB722	CAAGCAGAAGACGGCATAACGAGAT	CGCTCAGTgggttgagcgatgatatttgct
CB723	CAAGCAGAAGACGGCATAACGAGAT	GTCTTAGGgggttgagcgatgatatttgct
CB724	CAAGCAGAAGACGGCATAACGAGAT	ACTGATCGgggttgagcgatgatatttgct
CBR1	CTGGCTTAAAAAATCATTAAATAATAGGTTATGTTTAGAGTGTTCCCCGC	GCCAG
CBI1	CGGGGATAAACCGAGCACAAATATCATCGCTCAAACC	

CB5XX: Illumina P5 adaptor sequence-BC (8 bp)-primer binding site

CB7XX: Illumina P7 adaptor sequence-BC (8 bp)-primer binding site

CBR1: Custom read1 sequencing primer (spiked into well 12)

CBI1: Custom index1 sequencing primer (spiked into well 13)

Barcoded array Sequence

seq primer

BC_Array_CB7	CAAGCAGAAGACGGCATAACGAGAT	GACGTCGCGCTAGATTAGA
BC_Array_CBI	AACCGAGCACAAATATCATCGCTCAAACCACTTACGG	

BC_Array_CB7: Illumina P7 adaptor sequence-primer binding site

BC_Array_CBI: Custom index1 sequencing primer (spiked into well 13)

Random access Sequence

seq primer

RA701	CAAGCAGAAGACGGCATAACGAGAT	CGCTAGATTAGATCGCCTTA
RA702	CAAGCAGAAGACGGCATAACGAGAT	CGCTAGATTAGACTAGTACG
RA703	CAAGCAGAAGACGGCATAACGAGAT	CGCTAGATTAGATTCTGCCT

RA704	CAAGCAGAAGACGGCATAACGAGAT	CGCTAGATTAGAGCTCAGGA
RA705	CAAGCAGAAGACGGCATAACGAGAT	CGCTAGATTAGAAGGAGTCC
RA706	CAAGCAGAAGACGGCATAACGAGAT	CGCTAGATTAGACATGCCTA
RA707	CAAGCAGAAGACGGCATAACGAGAT	CGCTAGATTAGAGTAGAGAG
RA708	CAAGCAGAAGACGGCATAACGAGAT	CGCTAGATTAGACAGCCTCG
RA709	CAAGCAGAAGACGGCATAACGAGAT	CGCTAGATTAGATGCCTCTT
RA710	CAAGCAGAAGACGGCATAACGAGAT	CGCTAGATTAGATCCTCTAC
RA711	CAAGCAGAAGACGGCATAACGAGAT	CGCTAGATTAGATCATGAGC
RA712	CAAGCAGAAGACGGCATAACGAGAT	CGCTAGATTAGACCTGAGAT
RA713	CAAGCAGAAGACGGCATAACGAGAT	CGCTAGATTAGATAGCGAGT
RA714	CAAGCAGAAGACGGCATAACGAGAT	CGCTAGATTAGAGTAGCTCC
RA715	CAAGCAGAAGACGGCATAACGAGAT	CGCTAGATTAGATACTACGC
RA716	CAAGCAGAAGACGGCATAACGAGAT	CGCTAGATTAGAAGGCTCCG
RA717	CAAGCAGAAGACGGCATAACGAGAT	CGCTAGATTAGAGCAGCGTA
RA718	CAAGCAGAAGACGGCATAACGAGAT	CGCTAGATTAGACTGCGCAT
RA719	CAAGCAGAAGACGGCATAACGAGAT	CGCTAGATTAGAGAGCGCTA
RA720	CAAGCAGAAGACGGCATAACGAGAT	CGCTAGATTAGACGCTCAGT
RA721	CAAGCAGAAGACGGCATAACGAGAT	CGCTAGATTAGAGTCTTAGG
RA722	CAAGCAGAAGACGGCATAACGAGAT	CGCTAGATTAGAACTGATCG
RA723	CAAGCAGAAGACGGCATAACGAGAT	CGCTAGATTAGATAGCTGCA
RA724	CAAGCAGAAGACGGCATAACGAGAT	CGCTAGATTAGAGACGTCTGA

qPCR_primer	Sequence
genome-F	GCGAGCGATCCAGAAGATCT
genome-R	GGGTAAAGGATGCCACAGACA
pTrig-F	CGCTCTATGATCCAGTTCGATTT
pTrig-R	TCCGTATGCCATGCGTTTAT