Cell Genomics, Volume *2*

Supplemental information

Direct detection of RNA modifications and structure using single-molecule nanopore sequencing

William Stephenson, Roham Razaghi, Steven Busan, Kevin M. Weeks, Winston Timp, and Peter Smibert

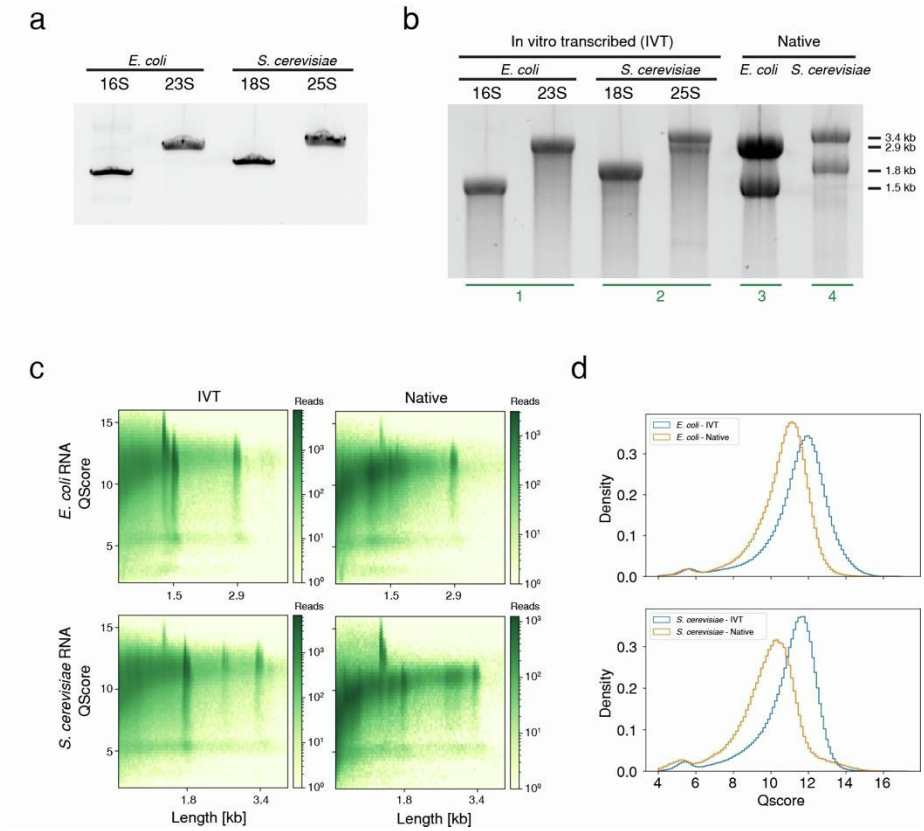

Figure S1. Generation of IVT controls and QC metrics from nanopore sequencing, related to Figure 1. a) SSU and LSU rDNA amplified from *E. coli* and *S. cerevisiae* gDNA with appended 5' T7 promoter. b) In vitro transcription (IVT) of corresponding transcription templates. Extracted total RNA (native) is shown at right. Green lines and numbering indicate the RNA used for individual MinION nanopore experiments. c) Quality score vs read length for IVT and Native samples from *E. coli* and *S. cerevisiae*. d) Quality score density distributions for native and IVT rRNA from *E. coli* (top) and *S. cerevisiae* (bottom).

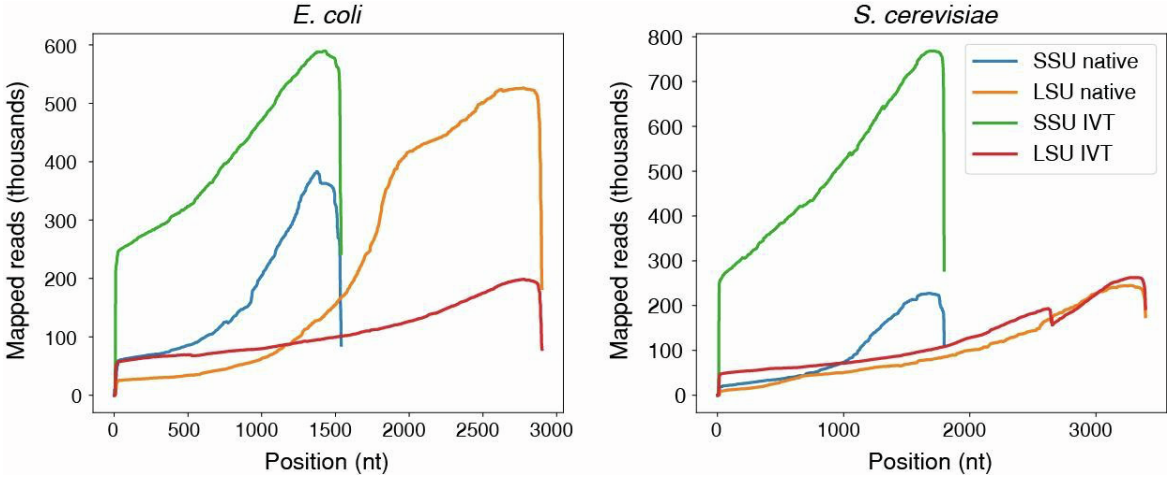

Figure S2. Mapped rRNA coverage, related to Figure 1. Mapped read coverage for SSU (16S/18S) and LSU (23S/25S) native and IVT rRNA for *E. coli* and *S. cerevisiae* after processing with *Tombo*.

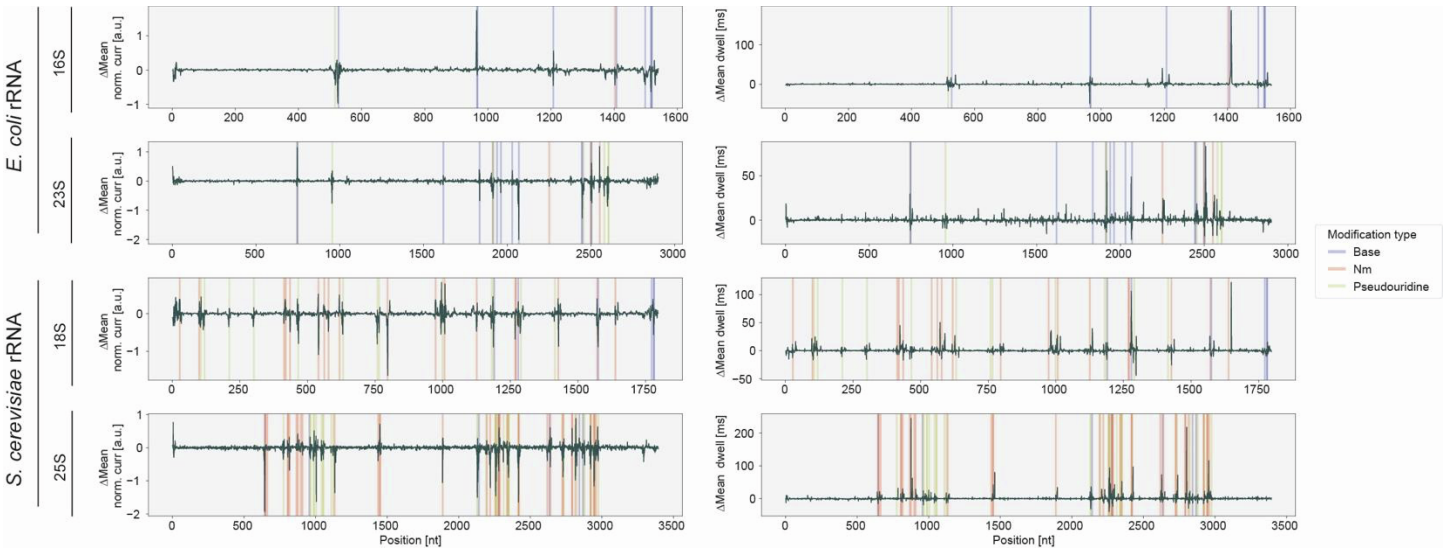

Figure S3. Mean current and dwell time changes reflect modification positions in rRNA, related to Figure 1. Mean normalized current (left column) and dwell time differences (right column) are plotted for *E. coli* (16S, 23S) and *S. cerevisiae* (18S, 25S). Known modification positions are indicated as vertical lines and reflect the modification type (non-pseudouridine base, 2'-*O*-methyl/Nm and pseudouridine)

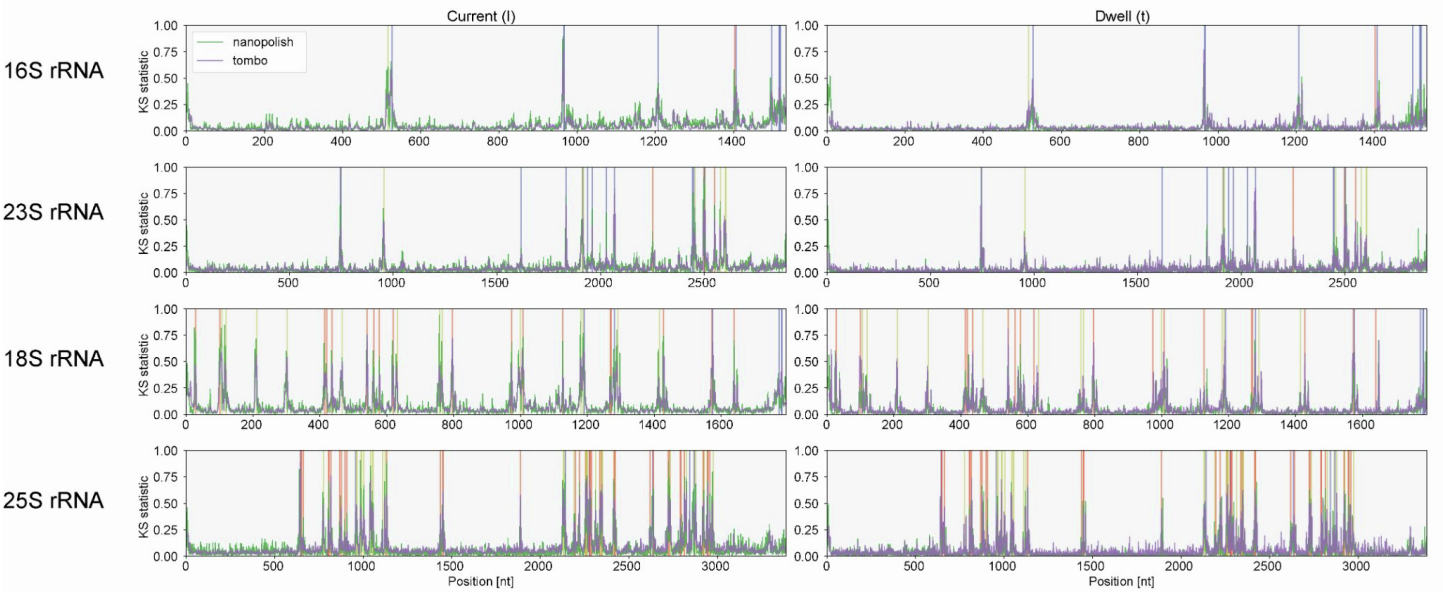

Figure S4. KS profiles of current and dwell time for rRNA, related to Figure 1 and Figure 2. KS profiles from tombo (purple) and nanopolish (green) for 16S, 23S, (*E. coli*) 18S and 25S rRNA (*S. cerevisiae*). Known modification positions are indicated as vertical lines and reflect the modification type (non-pseudouridine base, 2'-*O*-methyl/Nm and pseudouridine) Color key identical to Figure S3.

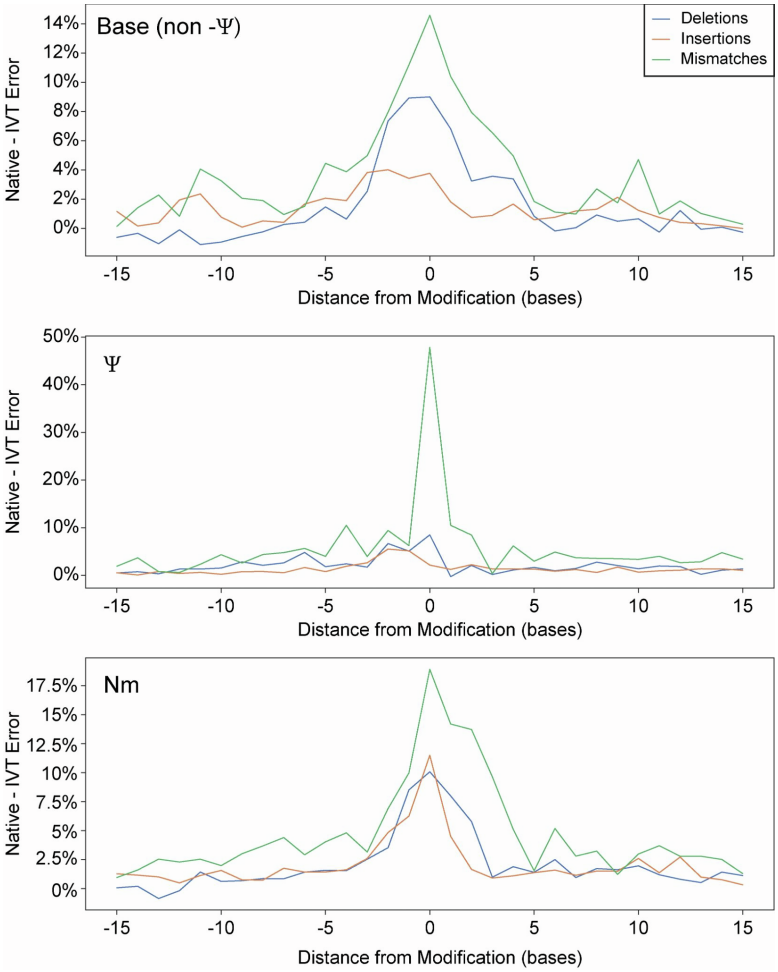

Figure S5. Proximal rRNA modification error rate analysis, related to Figure 1. Combined error rates proximal to known modification positions (non-pseudouridine base - top, pseudouridine - middle, and 2'-*O*-methyl (Nm) - bottom) in rRNA from *E. coli* (16S, 23S) and *S. cerevisiae* (18S, 25S). Error rate profiles are normalized to the IVT samples and plotted within the range -15 to +15nt surrounding the pore constriction.

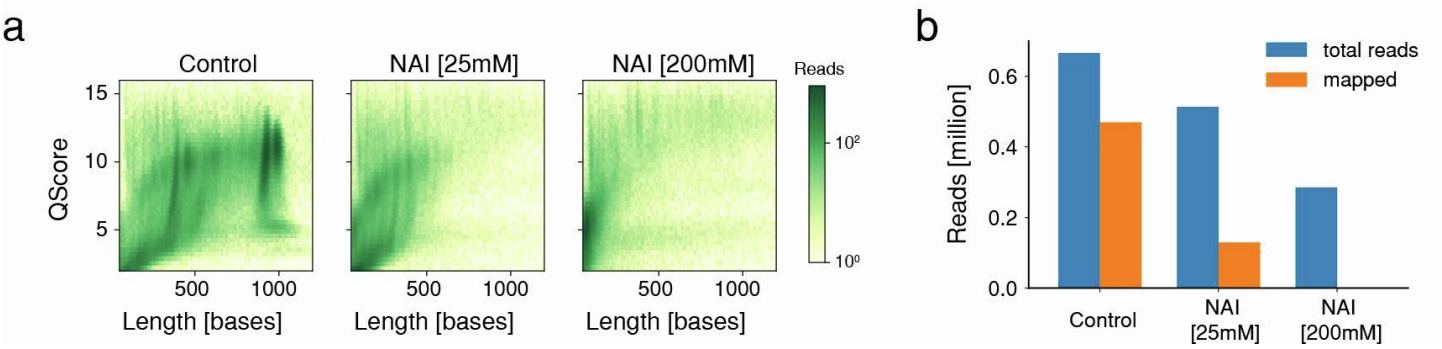

Figure S6. QC and alignment metrics of NAI modified RNA, related to STAR Methods. a) Quality score vs read length for control (unmodified) and NAI modified (25mM and 200mM final concentration) of pri-miR 17~92 cluster. b) Number of total reads and mapped reads for control (unmodified) and NAI modified reads.

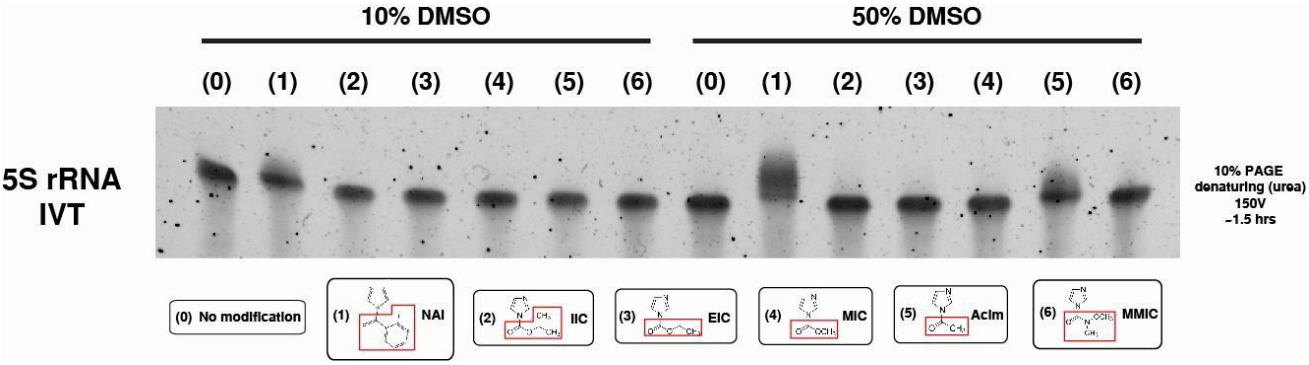

Figure S7. RNA reactivity for small-adduct RNA-reactive acylation reagents, related to Figure 3 and STAR Methods. 5 putative acylation reagents along with 1 known SHAPE reagent (NAI) were tested for reactivity against RNA in addition to a negative control (DMSO). In vitro transcribed 5S rRNA was modified in non-denaturing (10% DMSO) and denaturing (50% DMSO) conditions for 2 hours prior to resolving the RNA on a denaturing PAGE gel. Chemical modification key:

(0) DMSO vehicle only
(1) NAI (2-methylnicotinic acid imidazolide)
(2) IIC (isopropyl 1H-imidazole-1-carboxylate)
(3) EIC (ethyl imidazole-1-carboxylate)
(4) MIC (methyl imidazole-1-carboxylate)
(5) AcIm (acetylimidazole)
(6) MMIC (N-methoxy-N-methyl-1H-imidazole-1-carboxamide)

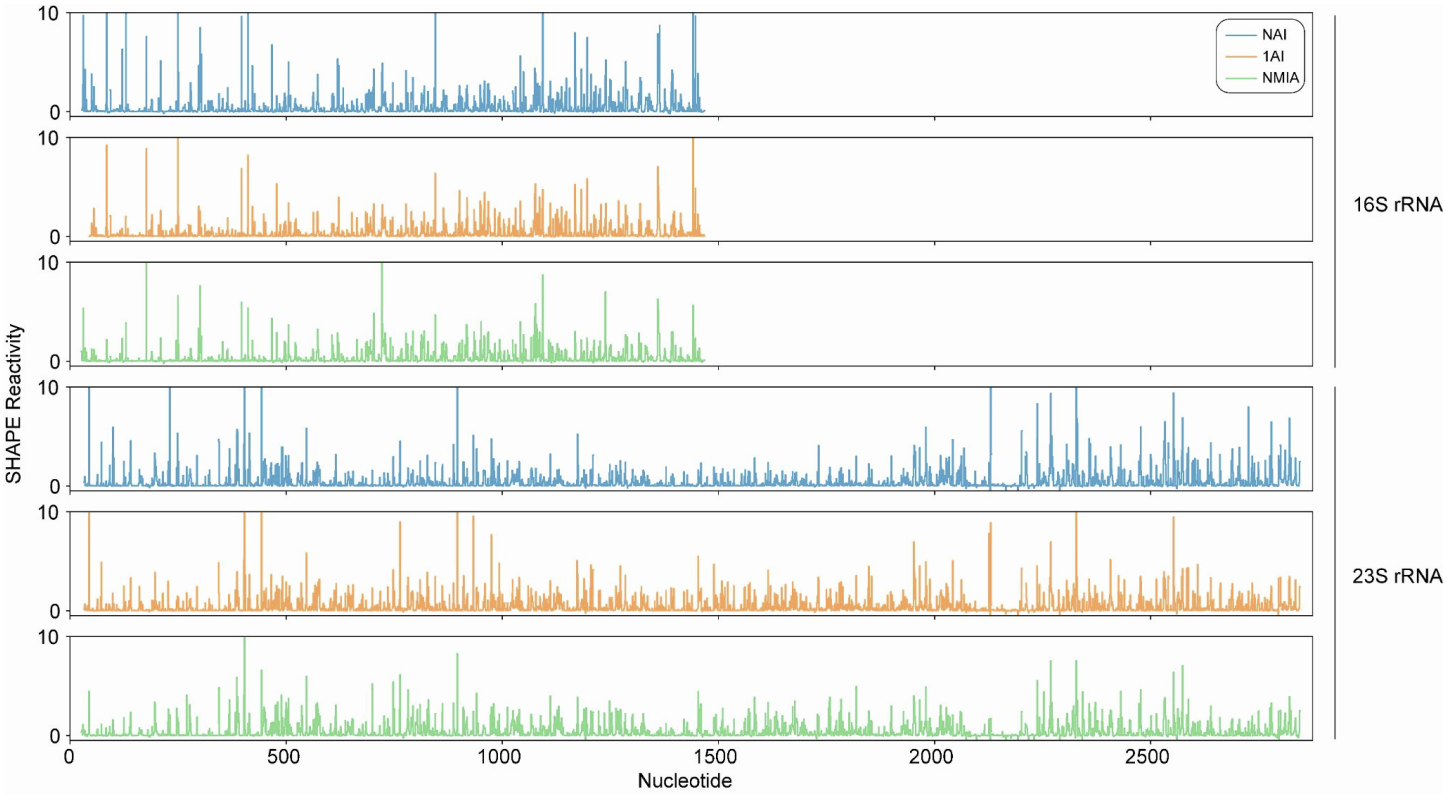

Figure S8. SHAPE-MaP reactivity profiles using NAI, AcIm, or NMIA for *E. coli* 16S and 23S rRNA, related to Figure 3.

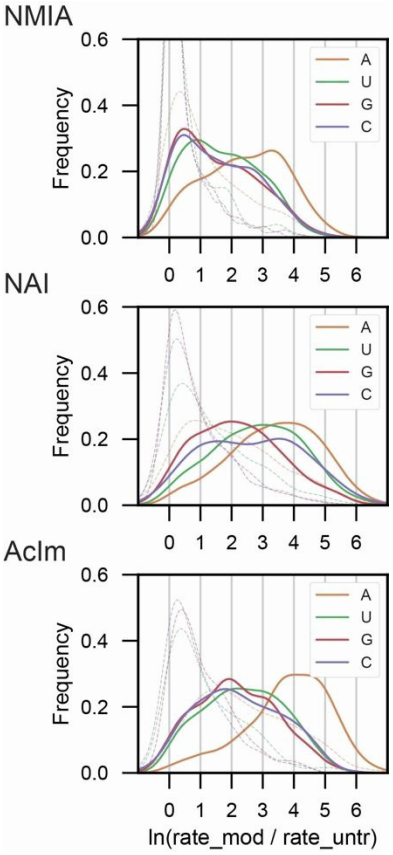

Figure S9. Per-nucleotide MaP mutation rates for SHAPE reagents, related to Figure 3. One-dimensional kernel density estimates for background-corrected mutation rates, where “rate_mod” is the SHAPE-modified rate and “rate_untr” is the unmodified control rate at a given position. Paired nucleotides are shown with a dashed line, and unpaired nucleotides with a solid line. All three reagents show the highest apparent reactivity with unpaired adenosine residues.

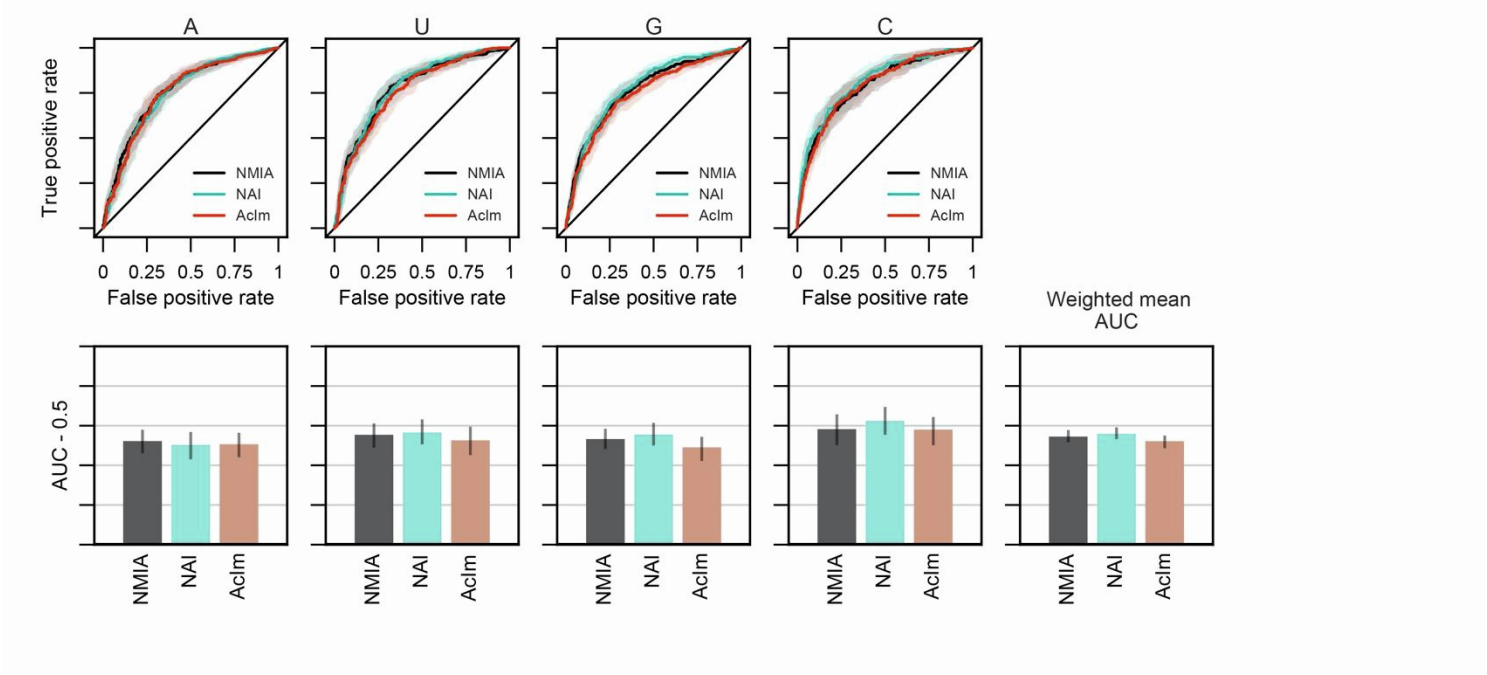

Figure S10. Receiver operator characteristic (ROC) curves for MaP reactivities as a function of base pairing status for SHAPE reagents, related to Figure 3. Analysis is based on reactivities for three reagents using 16S and 23S rRNAs. Columns are broken down by each canonical ribonucleotide. True positive rate: unpaired nucleotides with reactivity above a given threshold / total unpaired nucleotides. False positive rate: base-paired nucleotides with reactivity above a given threshold / total base-paired nucleotides. AUC: area under the ROC curve. Shaded regions are 95% bootstrap confidence intervals. Error bars indicate 95% bootstrap confidence intervals on the AUC. All three reagents show comparable agreement with base pairing status across all four ribonucleotides.

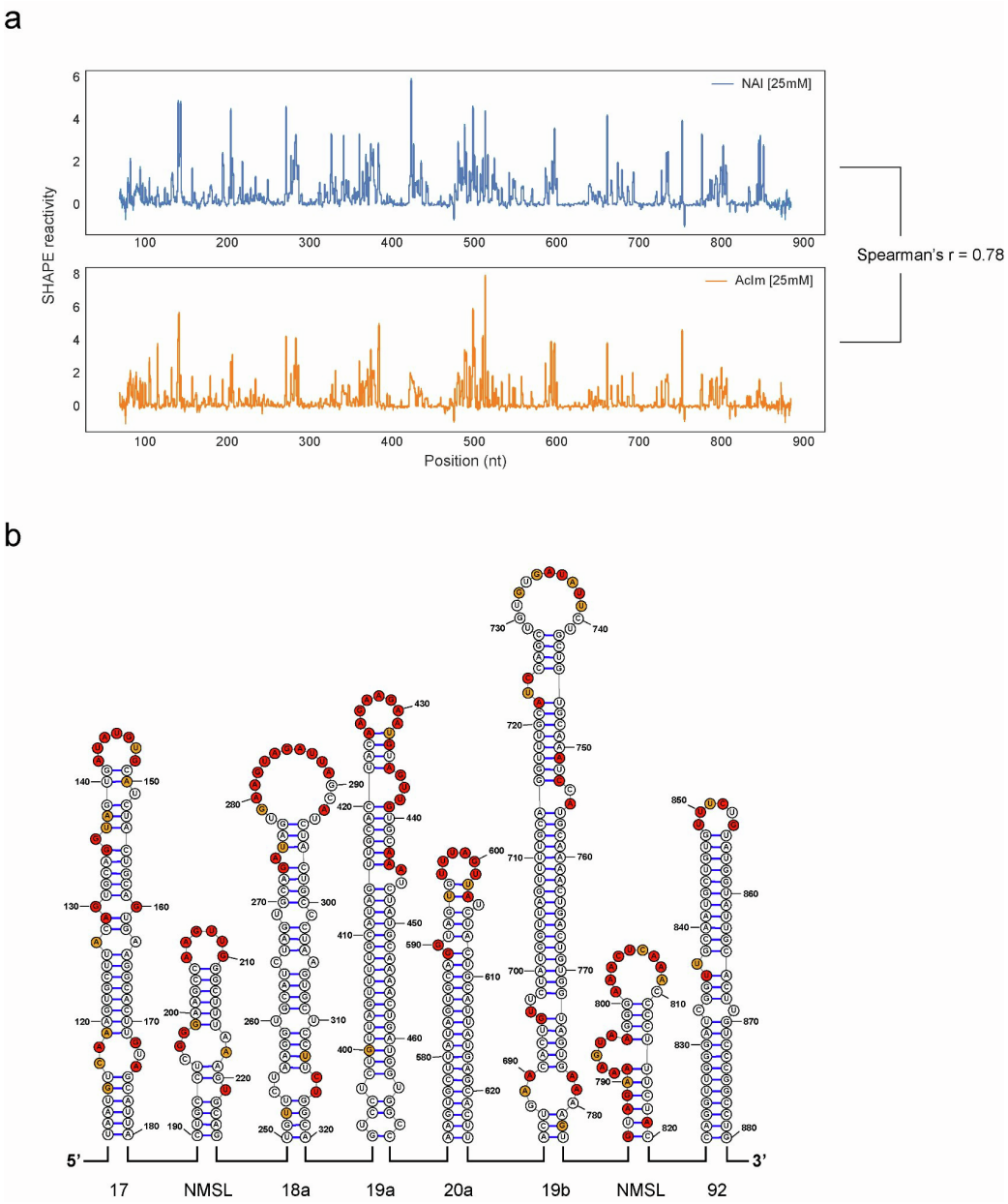

Figure S11. SHAPE-MaP profile of pri-miR 17~92, related to Figure 4. a) Normalized reactivity profiles for pri-miR 17~92 using NAI and AcIm both at 25mM final concentration. Error bars are normalized standard deviation. b) Centroid secondary structure of the pri-miR 17~92 transcript (miR hairpins shown) with the AcIm SHAPE-MaP reactivity overlaid. (Normalized SHAPE Reactivity: red > 0.85, yellow > 0.4, white < 0.4)

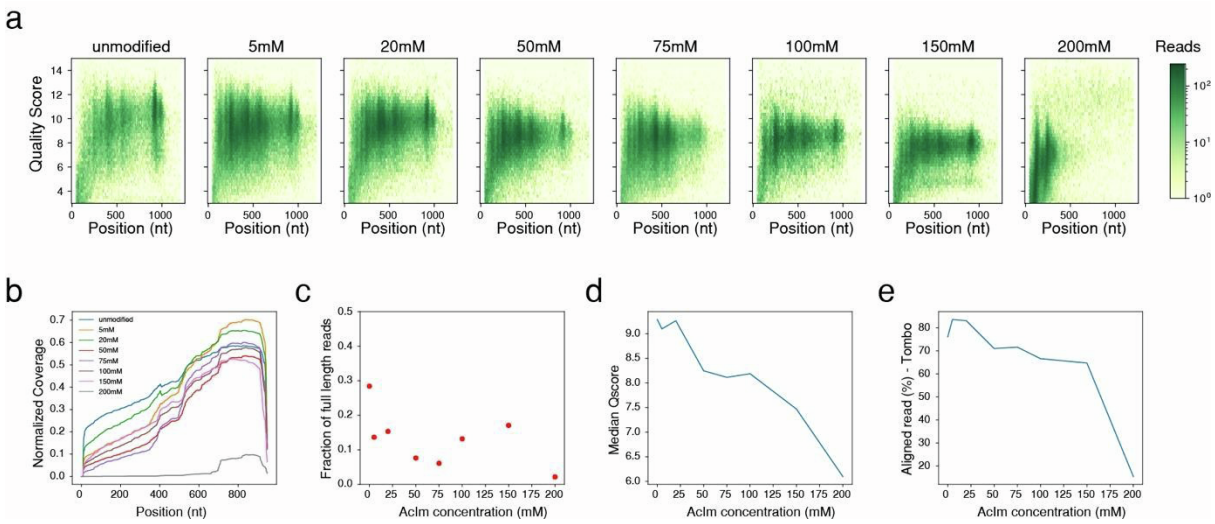

Figure S12. QC metrics direct RNA sequencing (flongle flow cell) of AcIm modified pri-miR-17~92, related to Figure 4. a) Quality score and read length heatmaps for pri-miR-17~92 modified with 0 (unmodified control), 5, 20, 50, 75, 100, 150, and 200mM final concentration of AcIm. b) Read coverage for AcIm modified pri-miR-17~92. c) Fraction of full length reads obtained for AcIm modified pri-miR-17~92. d) Median Qscore for AcIm modified pri-miR 17~92. e) Aligned read percentage after raw signal alignment with sequence using Tombo for AcIm modified pri-miR-17~92.

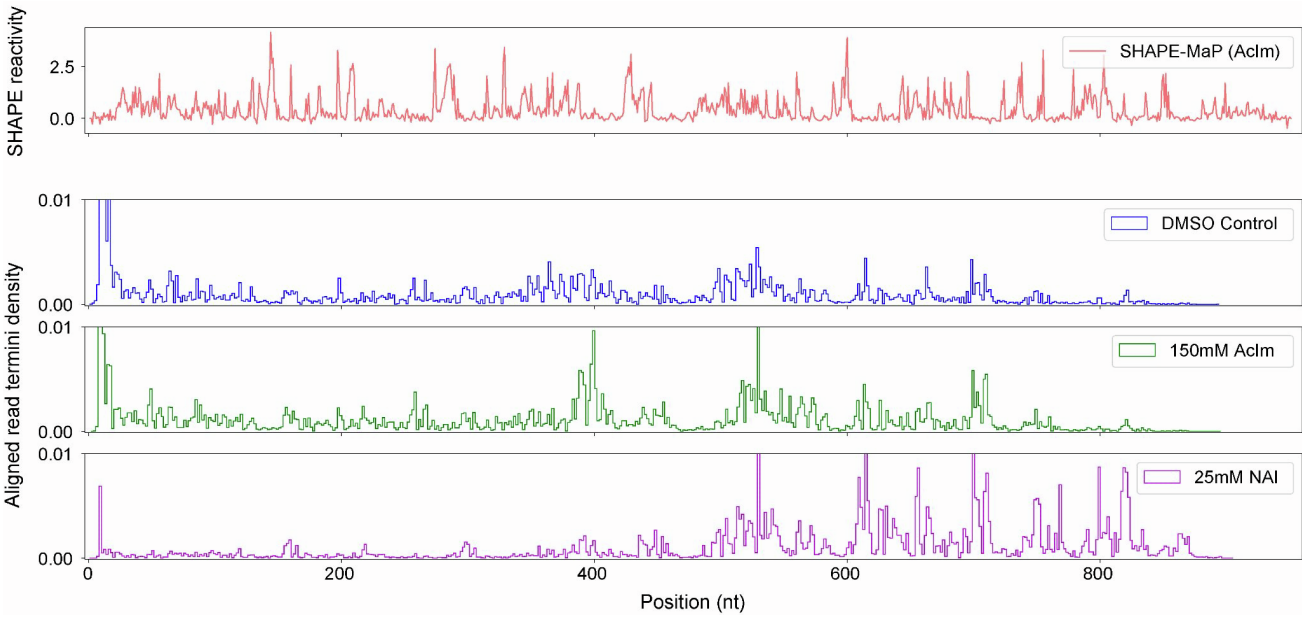

Figure S13. Nanopore aligned read termini density for unmodified and modified pri-miR-17~92, related to Figure 4 and STAR Methods. Nanopore read termini densities for unmodified (blue), AcIm (150mM) modified (green) and NAI (25mM) modified (magenta). The SHAPE-MaP reactivity is shown in red at top for comparison. Nanopore direct RNA sequencing proceeds from 3' (position 951) to 5' (position 0). The DMSO control and AcIm samples were run on Flongle flow cells. The NAI modified sample was run on GridION.

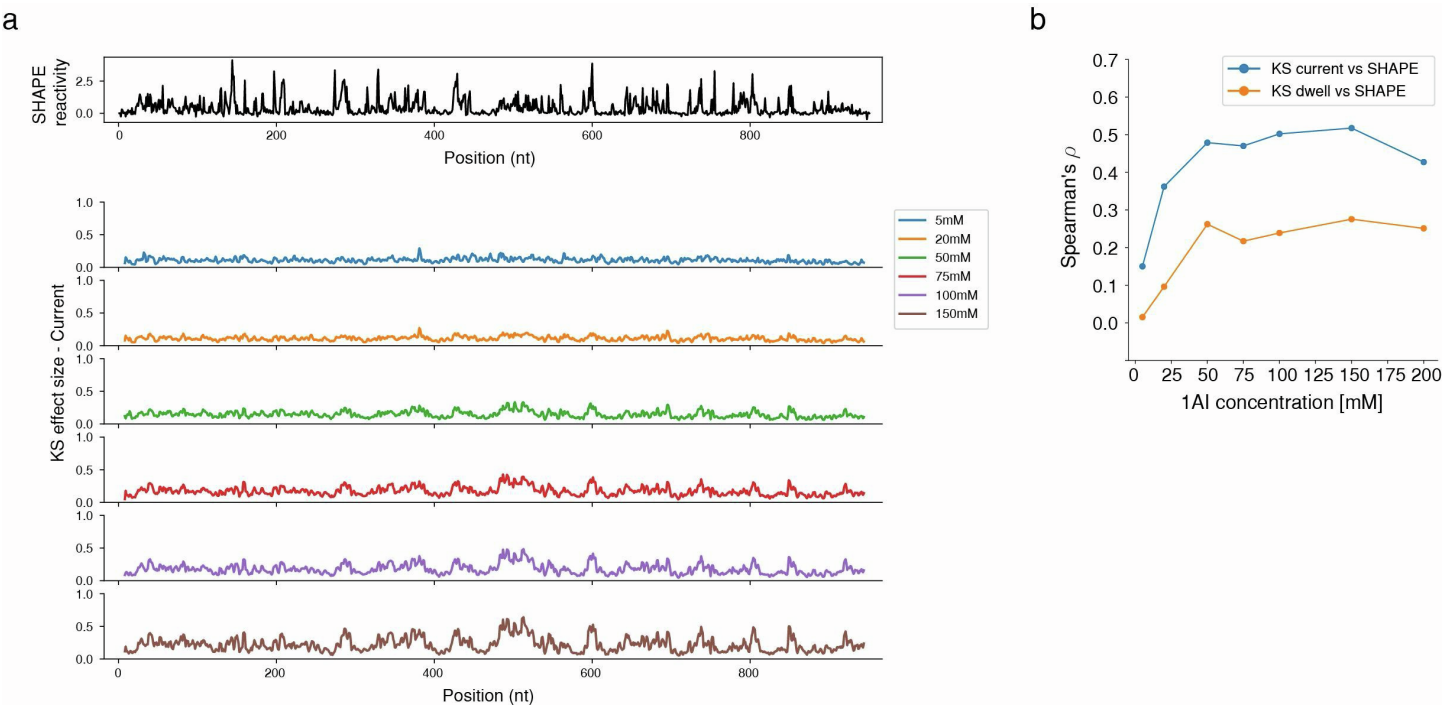

Figure S14. KS profiles of AcIm modified RNA, related to Figure 4. a) KS profiles for AcIm modified pri-miR-17~92 RNA. **b)** Spearman's rho correlation of KS profile for current and dwell time against the SHAPE-MaP for the AcIm modification concentrations.

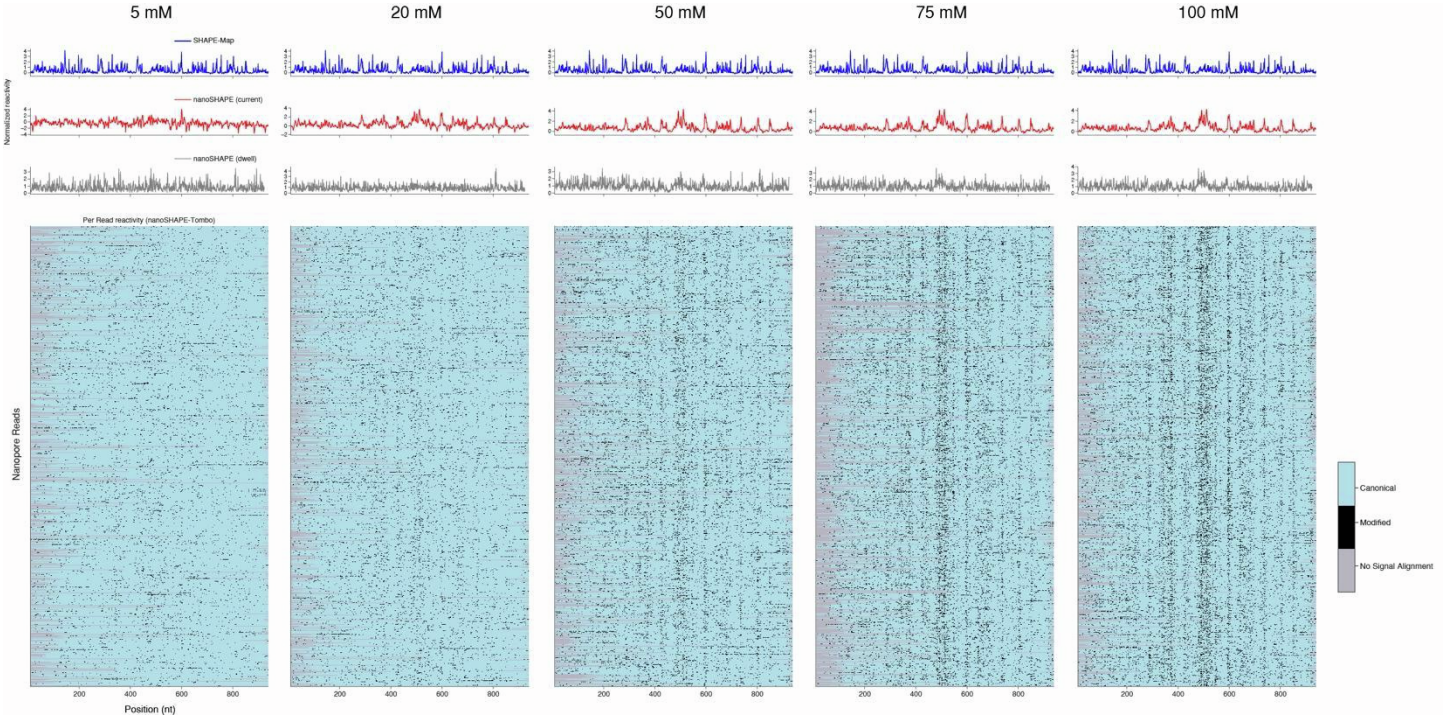

Figure S15. Per read statistical testing and normalized reactivity profiles across AcIm concentrations (5 mM, 20 mM, 50 mM, 75 mM, 100 mM), related to Figure 4.

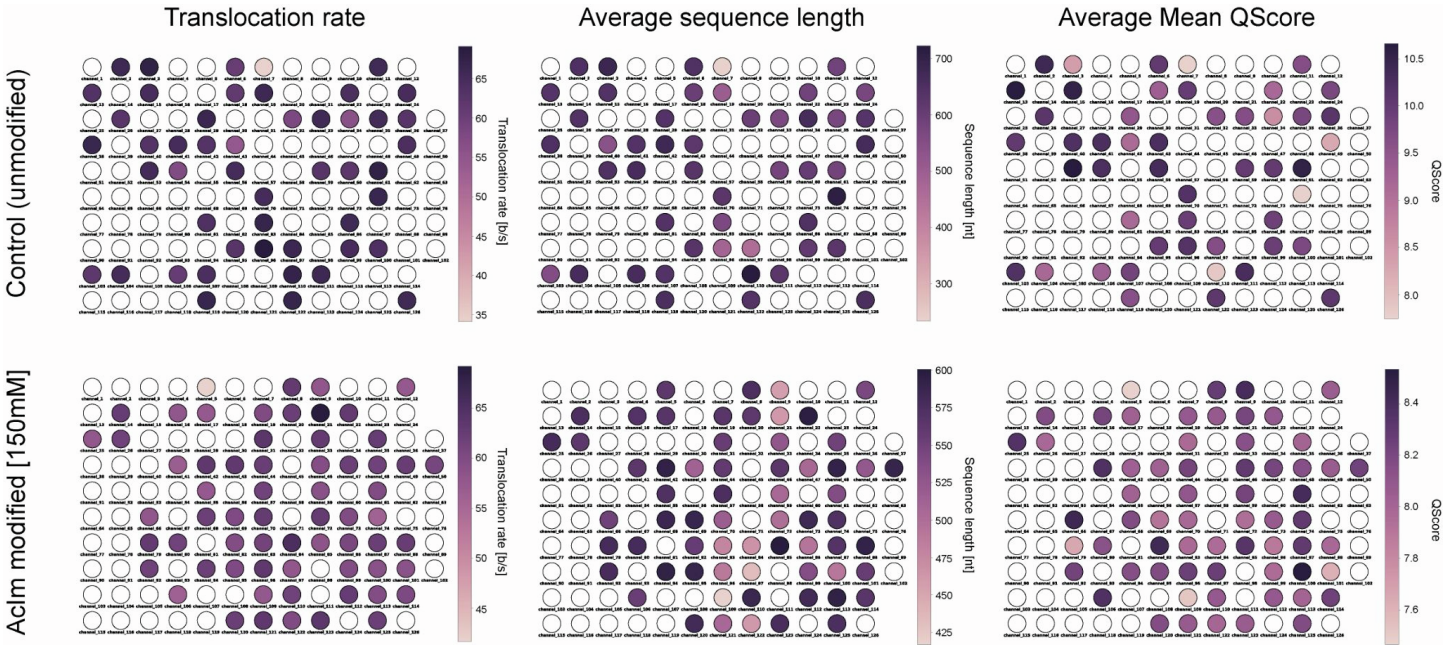

Figure S16. Flongle flow cell channel specific characteristics, related to Figure 4. Average translocation rate, average sequence length and average mean QScore across unmodified (top row) and AcIm modified (bottom row) across individual channels of two independent flongle flow cells. Inactive channels are indicated by unfilled circles.

Concentration AcIm [mM]	Reads	Full length reads (fastq)	Median Qscore	Tombo alignment [%]	Total number aligned full length reads (Tombo)
0	68693	17216	9.29	76.1	13101
5	163000	21173	9.1	83.6	17701
20	117546	16939	9.26	83.1	14076
50	97344	6720	8.25	71	4771
75	120932	6944	8.11	71.6	4972
100	52947	6389	8.19	66.6	4255
150	77992	12182	7.5	64.7	7882
200	108096	1669	6.1	15.4	257

Table S1. Flongle flow cell runs read summary, related to STAR Methods.

Primer name	Primer sequence (5'-3')
ec16f_t7p	TAATACGACTCACTATAGGGAAATTGAAGAGTTTGATCATGGCTC
ec16r	TAAGGAGGTGATCCAACCGCAGG
ec23f_t7p	TAATACGACTCACTATAGGGGGTTAAGCGACTAAGCGTACACGGT
ec23r	AAGGTTAAGCCTCACGGTTCATTAG
sc18f_t7p	TAATACGACTCACTATAGGGTATCTGGTTGATCCTGCCAGTAGTC
sc18r	TAATGATCCTTCCGCAGGTTCACCTAC
sc25f_t7p	TAATACGACTCACTATAGGGGTTTGACCTCAAATCAGGTAGGAGTA
sc25r	ACAAATCAGACAACAAAGGCTTAATCTC

Table S2. Primers for rDNA amplification from *E. coli and S. cerevisiae* gDNA, related to STAR Methods.