

Direct RNA sequencing for long and short transcripts, poly(A) tail measurement, and modified base detection

Unique native RNA sequencing chemistry from Oxford Nanopore enables simultaneous analysis of isoform expression, RNA modifications, and poly(A) tails at a single-molecule resolution for long and short RNAs

Contact details: sample.tech@nanoporetech.com
More information at www.nanoporetech.com

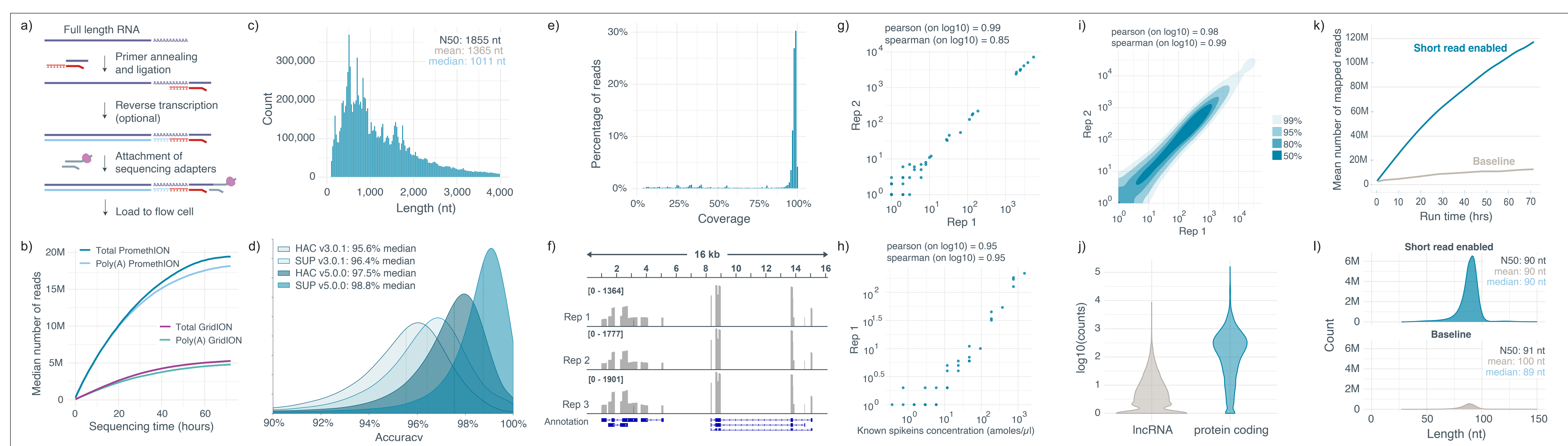


Fig. 1 Direct RNA nanopore sequencing a) workflow b) output c) read lengths d) accuracy e) spike-in coverage f) spike-in transcripts on IGV g) spike-in gene count replicates and h) expected gene counts i) human gene counts j) by biotype k) short RNA read output l) short RNA read lengths.

Flexible transcriptome profiling with the latest native nanopore chemistry (SQK-RNA004)

The direct RNA sequencing workflow, using the Direct RNA Sequencing Kit (SQK-RNA004) and corresponding RNA-specific 'RP4' flow cells, has been updated to new run conditions, resulting in an approximate increase in output of 20%, and Induro® Reverse Transcriptase (NEB) to capture longer transcripts. Native RNA molecules are sequenced directly, without the need for cDNA conversion and PCR steps, enabling the detection of modified RNA bases. A primer ligated to the 3' end of RNA transcripts enables poly(A) tail information to be retained. A recommended reverse transcription step helps scaffold the molecule as the native strand passes through the nanopore (Fig. 1a). We used Universal Human Reference RNA (UHRR) (Agilent) spiked with a set of control transcripts of known lengths and concentrations (SIRV-Mix 3) (Lexogen), prepared libraries in triplicate, and sequenced on GridION™ and PromethION™ devices. The kit recommends 300 ng poly(A)+ RNA (or 1 µg total RNA) as input and delivers approx. 5 M reads from one MinION™ and approx. 20 M reads from one PromethION™ Flow Cell in 72 h (Fig. 1b). Reads are basecalled using Dorado during sequencing with the MinKNOW™ software, or post-run using standalone Dorado. Dorado can estimate poly(A) tail lengths from raw nanopore signal. The median length for aligned reads was >1kb (Fig. 1c). Improvements in basecalling models provide a marked increase in accuracy, reaching median accuracies of 97.5% with high (HAC) and 98.8% with super accuracy (SUP) basecalling modes (Fig. 1d). We used the spike-ins to assess the method's ability to recover reads spanning full-length transcripts and quantify expression. The majority of reads covered transcripts from 5' to 3' end (Fig. 1e) with clear delineation of exon/intron boundaries (Fig. 1f). Count correlation between samples (Fig. 1g) and relative to known concentration (Fig. 1h) was high. Human gene counts are highly correlated between samples and have a high dynamic range (Fig. 1i). The method quantifies both protein and long non-coding transcripts (Fig. 2j). We used a 90 nt *in-vitro* transcript to demonstrate a new feature that enables the sequencing of reads as short as 50 nt at very high depth (>100 M) (Fig. 1k-l).

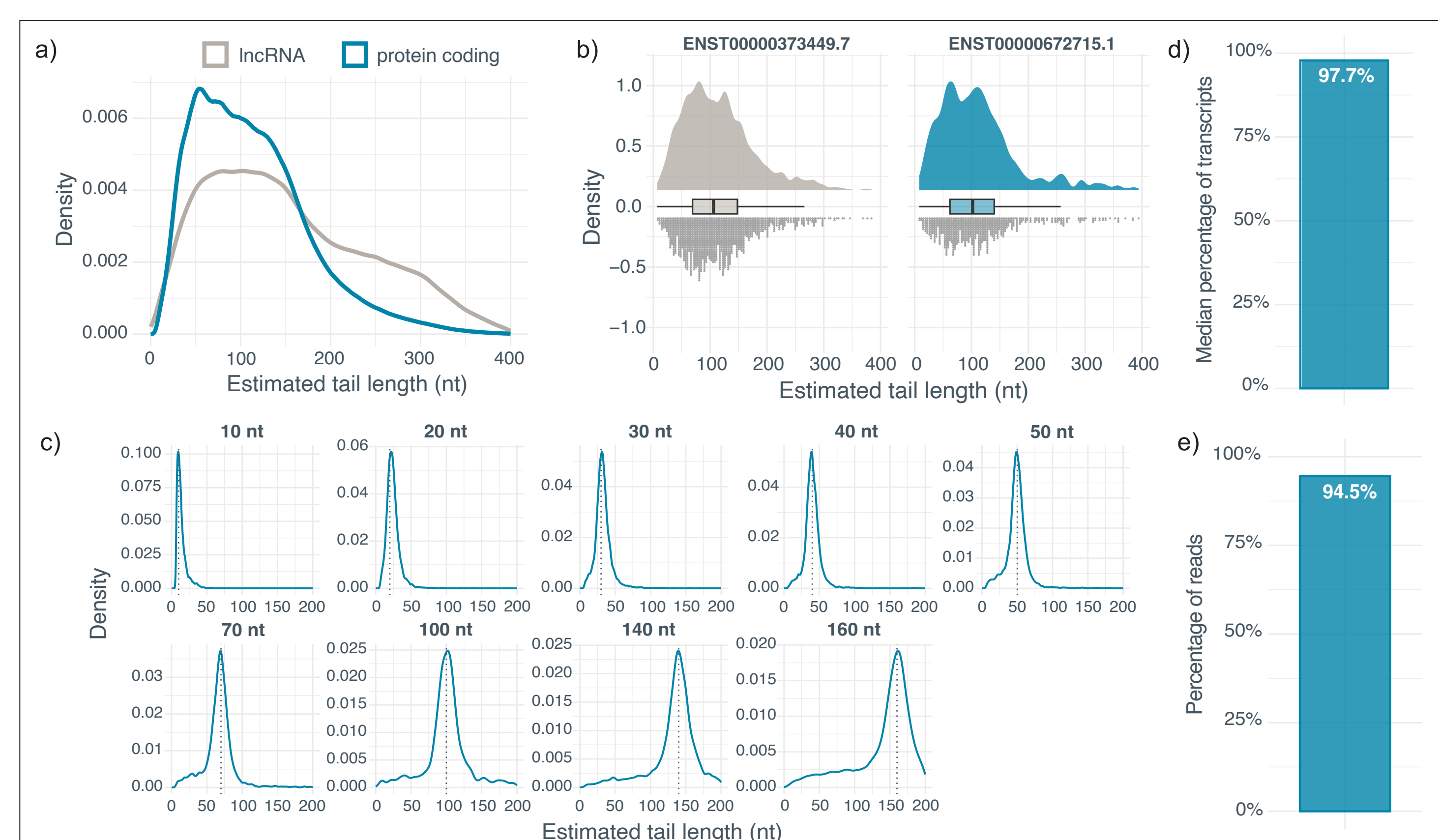


Fig. 2 Poly(A) tail lengths for transcript a) biotypes b) isoforms c) standards of known tail length d) percent of human reads with a tail e) percent of tail standard reads with a tail.

Simultaneous poly(A) tail length estimation

Library preparation with SQK-RNA004 retains the full poly(A) tail length sequence. Poly(A) tail length information can be easily estimated for each read during basecalling with the Dorado basecaller. A range of poly(A) tail lengths were captured in reads assigned to human protein coding and long non-coding transcripts (Fig. 2a). Poly(A) tail lengths for transcripts of the same gene can be compared (Fig. 2b). In-house tail length standards ranging from 10–160 nt in length were correctly estimated (Fig. 2c). A poly(A) tail is detected by Dorado in >94% of all human and tail standard reads (Fig. 2d-e).

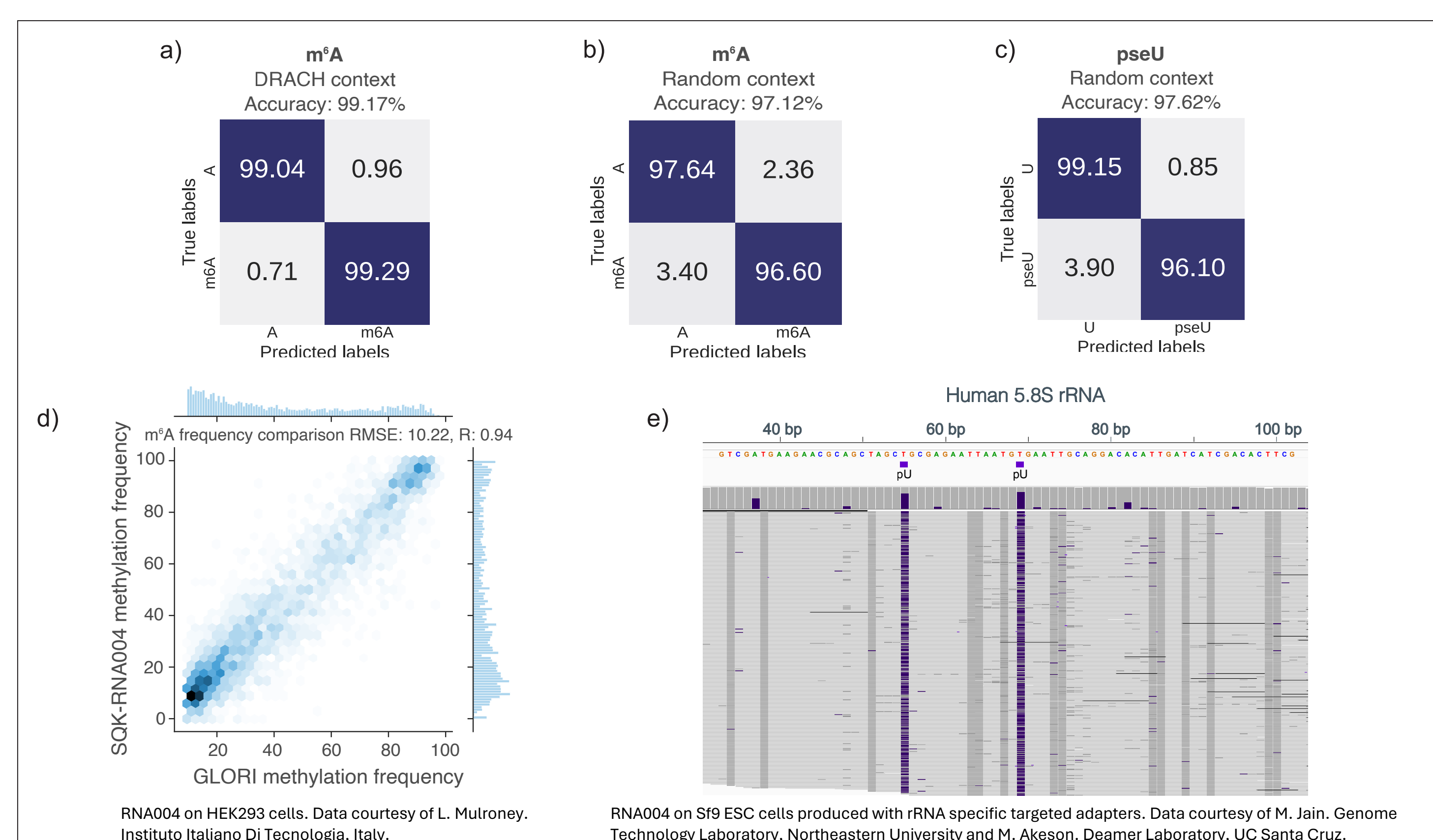


Fig. 3 RNA modification detection accuracies for a) m⁶A in DRACH context b) m⁶A in all contexts c) pseU in all contexts d) m⁶A compared to GLORI e) pseU on IGV.

Easy detection of m⁶A and pseU

SQK-RNA004, together with improved basecalling models, enable detection of RNA modification sites at single-molecule, single-nucleotide resolution directly from raw signal during basecalling. Improved, DRACH-motif context (Fig. 3a) and all-context m⁶A calling (Fig. 3b) and new all-context pseU (Fig. 3c) models are available in the production basecaller Dorado. Direct detection of m⁶A with nanopore sequencing has a high correlation with an orthogonal method GLORI (Liu et al.) (Fig. 3d). Modified bases can be easily visualised using IGV (Fig. 3e).