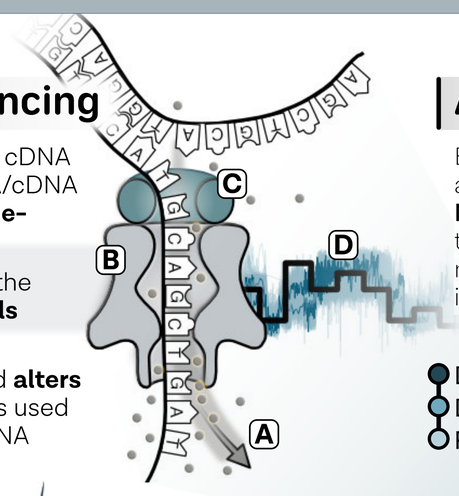
Felix Grünberger [1], Zubeir El Ahmad [1], Winfried Hausner [1], Sébastien Ferreira-Cerca [2], Dina Grohmann [1]
[1] Institute of Microbiology and Archaea Centre, Regensburg, Germany | [2] Biochemistry, Regensburg, Germany | Contact: **felix.gruenberger@ur.de**

# Expanding the transcriptomic toolbox in prokaryotes by Nanopore sequencing of RNA and cDNA molecules

**Nanopore sequencing** is a 3rd generation **single-molecule** method that allows sequencing of **full-length transcripts**.

Here, we present an experimental and bioinformatic workflow for **ONT RNA-seq** in the bacterial model organism *Escherichia coli* **(i)**, and show how we used the technolgy to analyze the stage-dependent installation of rRNA modifications **(1)** and transcription termination heterogeneity in Archaea **(2)**.

## Nanopore sequencing




- **A** An **ionic current** drives the cDNA or the RNA strand of a RNA/cDNA hybrid through a **membrane-embedded Nanopore B**.
- **C** The **motor protein unzips** the double strands and **controls** the translocation **speed**.
- **D** **Translocation** of the strand **alters** the **electric signal**, which is used to determine the RNA or DNA sequence.

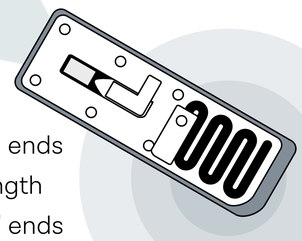
## Aim

Evaluate the applicability and advantages of **three different Nanopore RNA-seq protocols** for the simultaneous analysis of multiple transcriptomic features in **prokaryotes:**

*E. coli* K-12

- Direct RNA-seq **(DRS)**
- Direct cDNA-seq **(cDNA)**
- PCR-cDNA-seq **(PCR-cDNA)**

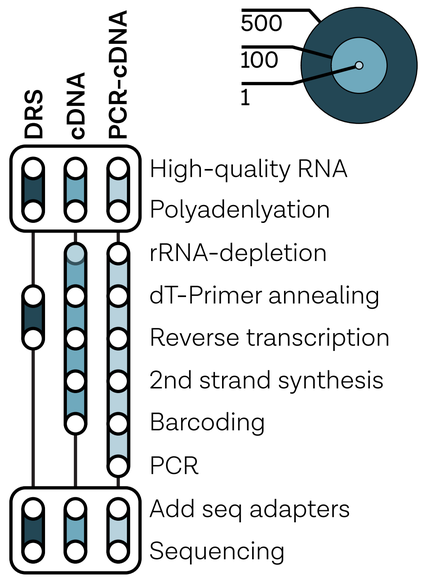
**Analysis**



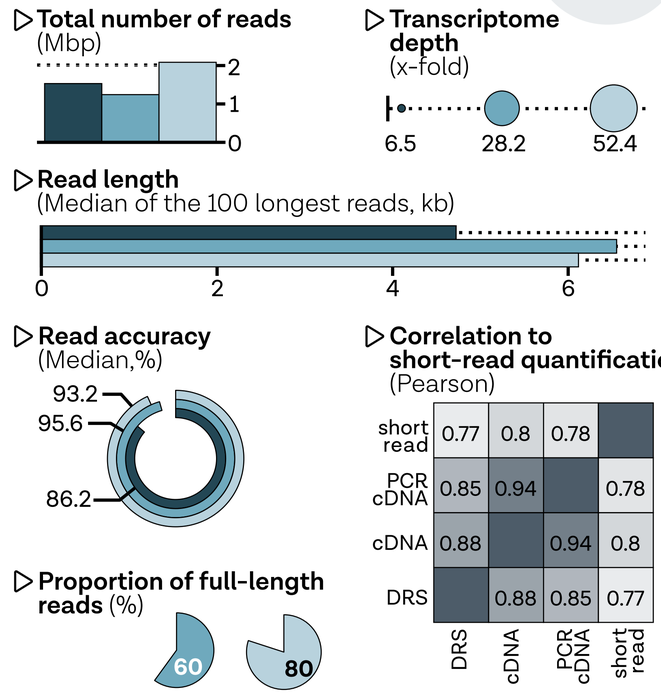
- Coverage
- Primary 5' ends
- 5' UTR length
- Primary 3' ends
- 3' UTR length
- Operon (density)

## Library prep

**Standard** ONT RNA-seq protocols (*RNA002,DCS109, PCB109*) all require **poly(A)**-tails and have different **input requirements** (ng).



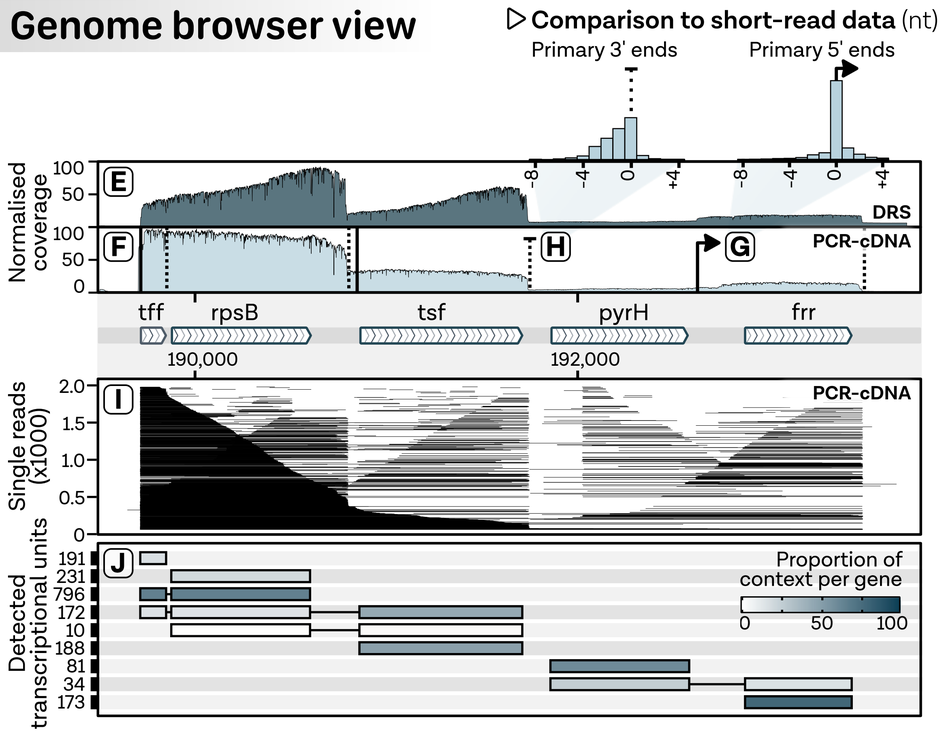

## Quality control



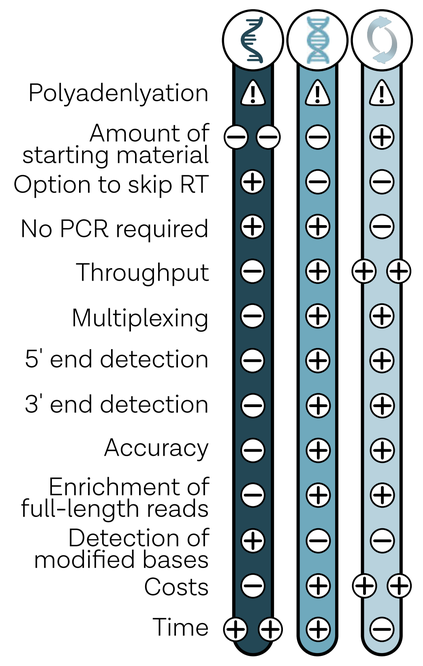

## Key features

- **E** **DRS** with 3'-5' **coverage bias**
- **F** **Full-length** read **enrichment** in the **(PCR)-cDNA** protocols improves coverage distribution over the transcript bodies
- **G** Highly **accurate 5' end detection**
- **H** Basecalling errors at poly(A) ends lead to **slightly fuzzier 3' ends**
- **I** Single-molecule resolution allows to decipher hitherto unseen **transcriptional heterogeneity**
- **J** **Complex transcriptional units** can be annotated and quantified

## Genome browser view




## Pro's and Con's



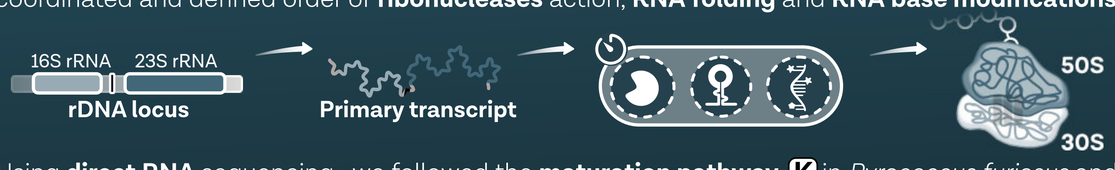

**Publication** (RNA, 2022)

**Documentation** (GitHub)

## Application 1: Stage-dependent installation of rRNA modifications

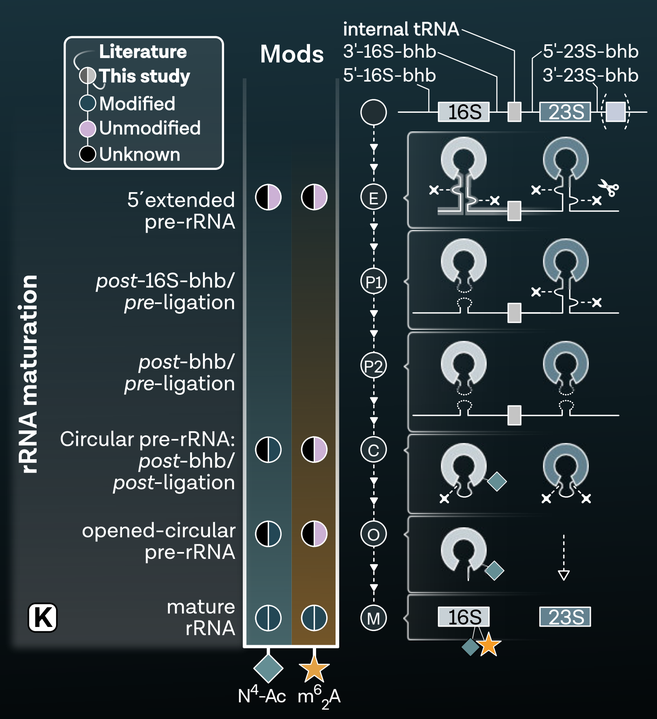
1

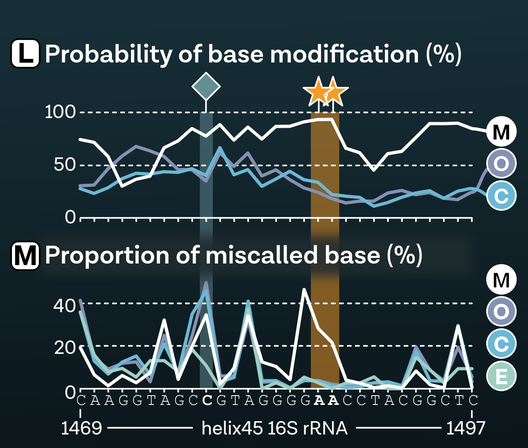
Ribosome biogenesis in Archaea starts with a polycistronic pre-rRNA, that is processed via the coordinated and defined order of **ribonucleases** action, **RNA folding** and **RNA base modifications**.




Using **direct RNA** sequencing, we followed the **maturation pathway K** in *Pyrococcus furiosus* and analyzed the **stage-dependent installation of** selected **modifications** in helix 45 of the 16S rRNA.

**RNA modifications** can lead to electric **current signals** varying from the expected theoretical distribution **L** and consequently may increase the rate of **basecalling errors M**.



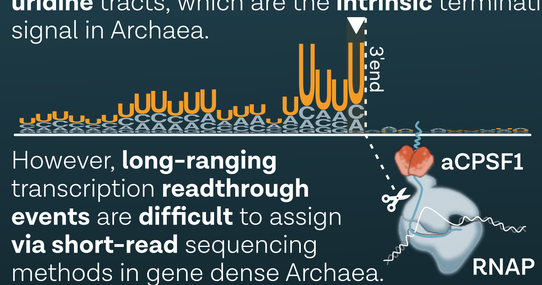




- **KsgA/Dim1**-dependent **dimethlyation** ($m^6_2A$) is completed at a **late stage** of **small ribosomal** subunit **biogenesis.**
- **Kre33/Nat10**-dependent $N^4$-cytidine **acetylation** ($N^4$-Ac) is **already installed earlier.**

## Application 2: Transcription termination heterogeneity

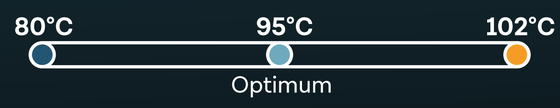
2

The efficiency of **transcription termination in Archaea** presumably depends on the cooperative action of the termination factor **aCPSF1** that specificially recognizes and cleaves after **poly-uridine** tracts, which are the **intrinsic** termination signal in Archaea.

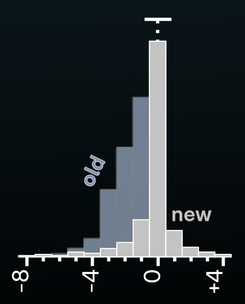



However, **long-ranging** transcription **readthrough events** are **difficult** to assign **via short-read** sequencing methods in gene dense Archaea.

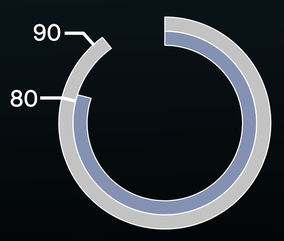
Here, we demonstrate the applicability of a **modified** and **improved** polyA-independent **PCR-cDNA** ONT protocol, to accurately **map** transcript **3' ends** in *Pyrococcus furiosus* at



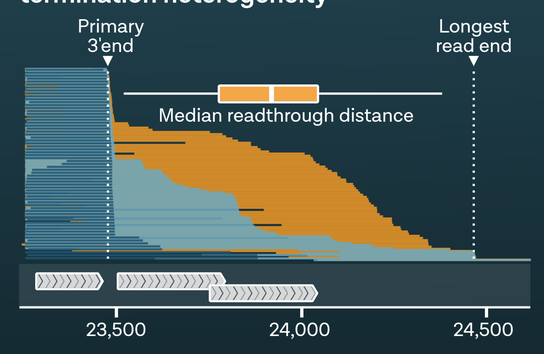

**Improved 3'end accuracy** (Comparison to short-read ends)




**Improved proportion of full-length reads (in %)**




**Single-molecule sequencing captures termination heterogeneity**




**Termination efficiency is poly(U) and temperature-dependent**