Real-time transcriptomic profiling in distinct experimental conditions

Tamer Butto[1,#,§], Stefan Pastore[1,2,#] , Max Müller[1], Kaushik Viswanathan Iyer[1], Stefan Mündnich[1] , Anna Wierczeiko[2], Kristina Friedland[1], Mark Helm[1], Marie-Luise Winz[1] and Susanne Gerber[2,§]

[1]Institute of Pharmaceutical and Biomedical Sciences, Johannes Gutenberg-University Mainz, 55128 Mainz, Germany

[2]Institute of Human Genetics, University Medical Center of the Johannes Gutenberg University Mainz, 55131 Mainz, Germany,

[#] Authors contributed equally

[§]Joint senior authors

Abstract

Nanopore technology offers real-time sequencing opportunities, providing rapid access to sequenced data and allowing researchers to manage the sequencing process efficiently, resulting in cost-effective strategies. Here, we present focused case studies demonstrating the versatility of real-time transcriptomics analysis in rapid quality control for long-read RNA-seq. We illustrate its utility through three experimental setups: 1) transcriptome profiling of distinct human cellular populations, 2) identification of experimentally enriched transcripts, and 3) identification of experimentally manipulated genes (knockout and overexpression) in several yeast strains. We show how to perform multiple layers of quality control as soon as sequencing has started, addressing both the quality of the experimental and sequencing traits. Real-time quality control measures assess sample/condition variability and determine the number of identified genes per sample/condition. Furthermore, real-time differential gene/transcript expression analysis can be conducted at various time points post-sequencing initiation (PSI), revealing dynamic changes in gene/transcript expression between two conditions. Using real-time analysis, which occurs in parallel to the sequencing run, we identified differentially expressed genes/transcripts as early as 1-hour PSI. These changes were consistently observed throughout the entire sequencing process. We discuss the new possibilities offered by real-time data analysis, which have the potential to serve as a valuable tool for rapid and cost-effective quality checks in specific experimental settings and can be potentially integrated into clinical applications in the future.

Introduction

The field of transcriptomics aims to explore, monitor, and quantify the complete set of transcripts, including coding (e.g., mRNA), non-coding, and small RNAs, within a given cell at a given condition (Wang et al. 2009). The investigation of the transcriptome is crucial for understanding the functional elements of the genome and their role within a cell or tissue, as well as their role during development or disease manifestation (Casamassimi et al. 2017). Over the past decade, transcriptomics has witnessed significant technological advancements, especially with the rise of Next Generation Sequencing (NGS) and the extensive use of RNA sequencing (RNA-seq) (Mutz et al, 2013; Satam et al. 2023). Techniques such as RNA-seq became the primary methodology to investigate the transcriptome using high-accuracy, short-read data (Mutz et al, 2013; Satam et al. 2023, Butto et al. 2023). Additionally, several well-established bioinformatic pipelines for RNA-seq have demonstrated reliability in analyzing transcriptome data. These pipelines typically involve quality control, read alignment to a reference genome, quantification of gene expression levels, and downstream analysis of differential gene expression. Notable tools such as minimap2 (Li 2018), HISAT2 (Kim et al. 2019) or STAR (Dobin et al. 2012) are commonly employed for read alignment, while featureCounts (Liao et al. 2013) or HTSeq (Anders et al. 2014) are utilized for quantifying expression. The widely used DESeq2 (Love et al. 2014) and edgeR (Robinson et al. 2009) packages offer robust statistical methods for identifying differentially expressed genes. The reliability of such tools is evidenced by their widespread adoption in the scientific

community, which allows the extraction of meaningful insights from RNA-seq data and contributes to our understanding of gene expression dynamics in various biological contexts (Conesa et al. 2016; Ji and Sadreyev 2018; Corchete et al. 2020).

While RNA-seq coupled with NGS has revolutionized transcriptome analysis, there are still improvements to be made, mainly depending on the requirements of the experimental design. The costs associated with NGS RNA-seq experiments can be a considerable factor, particularly when dealing with a large number of samples or the experimental approach requires high sequencing depth (Conesa et al. 2016; Ji and Sadreyev 2018). For instance, higher read-depth often yields more comprehensive information (e.g., splicing/isoform detection analyses) (Zhang et al. 2017; Hardwick et al. 2019). However, this comes at the expense of higher costs. Secondly, the library preparation process for NGS RNA-seq poses inherent challenges since it involves fragmentation of the reverse transcribed cDNA and introducing potential PCR bias during library amplification (Ozsolak and Milos 2011). These steps may introduce a limitation in accurately representing the investigated transcriptome, as certain sequences might be preferentially amplified over others, ultimately resulting in the loss of valuable information. Lastly, repetitive sequences pose a significant obstacle in their analysis, especially when employing short-read sequencing technologies (*Ozsolak and Milos, 2011)*. For instance, the precise alignment of short reads to repeat regions/elements remains problematic due to the intrinsic nature of such reads *(Ozsolak and Milos, 2011)*. Thus, it is essential to consider alternative sequencing strategies to address these obstacles.

One noteworthy alternative is long-read sequencing, such as Nanopore sequencing (Nanopore-seq). This technology, developed by Oxford Nanopore Technologies (ONT), has emerged as an innovative method for sequencing native long-read nucleic acids, including genomic DNA, cDNA, and RNA (Lu, Giordano, and Ning 2016, Wang et al. 2021, Zheng et al. 2023). The library preparation procedure involves straightforward steps, integrating a specific adapter at the end of the nucleic acid. This facilitates the efficient "reading" of intact nucleic acids, even ultra-long fragments. (Kono and Arakawa 2019, Amarasinghe et al. 2020). Integrating long-read sequencing with transcriptomics allows for the capture of entire transcripts, providing distinct advantages in detecting various RNA isoforms, repetitive sequences, and long mRNA transcripts (Amarasinghe et al. 2020; Wang et al. 2021). In addition, one key advantage of Nanopore-seq lies in its capability of real-time sequencing. This feature provides the opportunity to gain rapid access to the sequenced data, enabling researchers to either manage the sequencing process or stop it once the desired results are achieved (Wang et al. 2021). The latter allows for washing and reusing consumables, thus significantly lowering the sequencing costs. Moreover, adaptive sampling offers opportunities to enrich or deplete specific genes or transcripts during runtime. (Wang et al. 2024). A few studies have reported real-time analysis tools coupled with Nanopore-seq, primarily focusing on genomic or metagenomic DNA applications. For instance, real-time analysis platforms like EPI2ME by ONT (https://labs.epi2me.io/) and minoTour (Munro et al. 2022) provide continuous access to real-time metrics and analysis, streamlining the sequencing process. Algorithmic tools such as BOSS-RUNS (Weilguny et al. 2023), RawHash (Fırtına et al. 2023), and BoardION (Bruno et al. 2021) introduce dynamic decision strategies, hash-based similarity searches for efficient real-time analysis, and interactive web applications for ONT sequencing runs. Additional real-time detection tools, such as Metagenomic (Sanderson et al. 2018) and NanoRTax (Rodríguez-Pérez et al. 2022), provide immediate analytical pathways, concentrating on assessing metagenomic composition and viral detection tools. This diverse array of tools collectively addresses various aspects of Nanopore sequencing, spanning real-time analysis, algorithmic enhancements, metagenomic exploration, and current signal mapping. However, the combination of real-time analysis alongside comprehensive transcriptomic analysis has not been extensively explored.

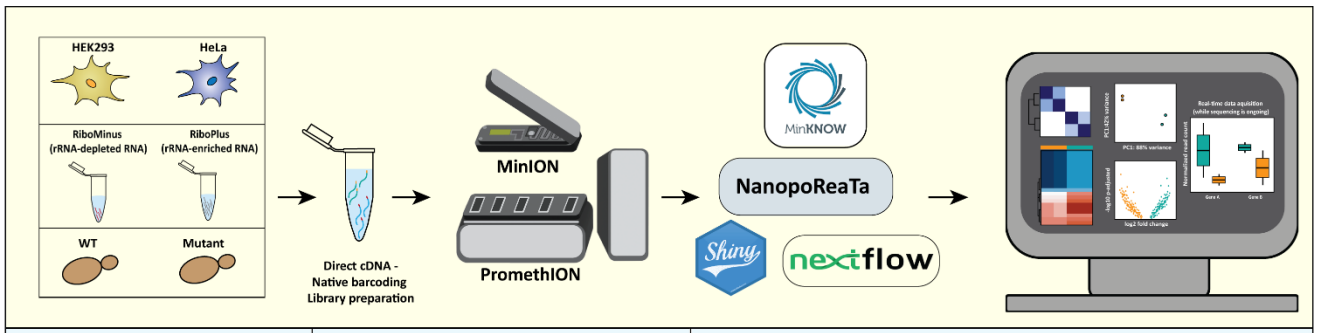
Recently, we presented NanopoReaTA, the first real-time analysis toolbox for comparative transcriptional analyses of Nanopore-seq data (Wierczeiko and Pastore et al. 2023). NanopoReaTA provides an interactive Graphic unit interphase (GUI) that allows users to perform transcriptional analyses of cDNA and/or direct RNA libraries. The new possibilities offered by real-time analysis are precious for fast and cost-effective quality control. In addition, they have the potential to significantly impact clinical applications where speed and efficiency are crucial, e.g., in diagnostics. Here, we present streamlined case studies, demonstrating the utility of real-time analysis using NanopoReaTA in various rapid quality control layers.

Results

Experimental design

We designed three experimental setups that include: 1) transcriptome profiling of distinct human cellular populations 2) Identification of experimentally-enriched transcripts and 3) Identification of an experimentally manipulated gene (KO and overexpression) in yeast strains (Fig 1). The latter demonstrates that real-time analysis using NanopoReaTA can also be applied to non-mammalian samples, provided the genome annotation files are available. We have designed a streamlined pipeline (experimental and bioinformatic) to monitor the detection speed of the transcriptional changes occurring between distinct conditions. RNA was isolated from all samples using Trizol, and library preparation was performed using polyA enrichment and first and second cDNA synthesis strategies (See material and methods). All samples were barcoded using ONT native barcoding. We simultaneously performed a pairwise comparison between two distinct conditions (Supplementary Material 1 - "Step-by-step use of NanopoReaTA"). According to the capabilities of the computational device (256 GB RAM) and the size of the reference genome (Human ~40 GB RAM, Yeast ~8 GB RAM) in use, multiple instances of NanopoReaTA were used in parallel. We set up 5 data collection time points from the sequencing initiation, including 1hr, 2hr, 5hr,10hr, and 24hr post-sequencing initiation (PSI). While sequencing, we exported several analyzed datasets, including general sample overviews such as read length (per sample and condition), gene expression variability (per sample and condition), changes in gene composition (per sample and condition), and processing time.

Additionally, we performed analyses of real-time differential gene/transcript expression (DGE/DTE) and differential transcript usage (DTU) between the two conditions, providing valuable quality control for the experimental setup. Analyses of DGE and DTE were performed by DESeq2 (Love et al. 2014), which is integrated into NanopoReaTA's pipeline. For DTU, we integrated analysis tools such as DEXSeq (Anders et al. 2012) and DRIMSeq (Nowicka and Robinson 2016). This feature offers insights into specific gene isoforms differentially expressed between distinct conditions. All output tables and figures produced by NanopoReaTA were systematically gathered and arranged to track the real-time detection of transcriptional changes during sequencing.



	Sample preparation and RNA isolation	Library preparation and sequencing initiation	Real-Time analysis and visualization
Preparation time:	~1hr	~3hr	Data collection: 1h , 2h, 5hr, 10hr, 24hr post-sequencing initiation

Figure 1. Streamlined pipeline for the rapid experimental setup and utilization of NanopoReaTA. Sample preparation involved Trizol RNA isolation (~1 hour) following library preparation, which included the synthesis of dscDNA library for the desired RNA sample (~2 hours). Samples were barcoded, and adapter ligated (~1 hour). Sequencing was performed using a PromethION R10 flow-cell or MinION R9 flow cell (for HEK293 and HeLa), and real-time data analysis and visualization occurred alongside ongoing sequencing. For this study, figures were exported at 1hr, 2hr, 5hr, 10hr, and 24hr psi, providing insights into the dynamic transcriptional changes of long-read RNA-seq between distinct conditions.

Efficient segregation of distinct cellular populations using NanopoReaTA's rapid transcriptome profiling

To demonstrate the rapidity and precision of real-time analysis in detecting transcriptional changes, we chose two distinct cell populations with unique transcriptomes and monitored alterations while sequencing was in progress. HEK293 (Human Embryonic Kidney) and HeLa (cancer) cells were selected due to their simplicity, widespread use, and ease of manipulation. For each cell type, we utilized 2 biological replicates. Following barcoding, the samples were loaded into a PromethION flow cell and sequenced for 24 hours. We tracked the basic sequencing metrics using ONT's MinKnow software and activated NanopoReaTA as soon as the sequencing initiated. Furthermore, in this experimental setup, we sequenced HEK293 and HeLa samples using a MinION R9 flow cell. This allowed us to demonstrate the reproducibility of our real-time analyses for detecting transcriptional changes between samples and highlight variations in data analysis when comparing low-throughput (MinION) and high-throughput (PromethION) data collection.

One-hour PSI, we gathered basic sequencing metrics from the MinKnow device, including total reads generated per sample, along with mapped reads, gene counts, and transcript counts (Supplementary Figure 1, Supplementary Table 1). At this stage, we gathered basic quality control information, including the number of detected genes, gene variability, individual and combined read length distribution, and the usage timings of the tools applied by NanopoReaTA for quality control (Supplementary Figure 2). Interestingly, despite loading similar cDNA amounts for each sample, we observed a reduction in the overall sequencing throughput and detected gene number for HEK293 compared to HeLa (Supplementary Figure 2). We tested such observations in the MinION experimental setup, where all cDNA generated during library preparation was loaded. Here, we observed a relatively similar throughput between HEK293 and HeLa (Supplementary Figure 3). Detailed description of potential reasons for these discrepancies is provided in Supplementary Material (Supplementary Material 2).

Next, we performed real-time DGE and DTE analyses to monitor the transcriptional changes between HEK293 and HeLa, 1-hour PSI. As an initial quality control, we inspected the Sample-to-sample similarity plot and principal component analysis (PCA) (Figure 1B, 1C). Such analyses reduce the dimensionality of the data allowing to monitor the clustering of similar data sets (same condition) while separating distinct data sets. Here, we observed distinct separation between the two conditions, where PC1 represents 68% of the variance while PC2 represents 17% of the variance, already 1-hour PSI. Similar observations were noted in the MinION sequencing setup (Supplementary Figure 4A, 4B). Real-time measurements like these offer valuable quality control insights into experimental replicates' and conditions' quality, influencing the decision to continue sequencing based on their clustering.

In the next phase, we inspected the differentially expressed genes (DEGs) presented in the output volcano plots and the top 20 DEGs (based on fold-change and adjusted p-value) (Figure 1D, Supplementary Table 1). Notably, we identified 5 annotated genes enriched in HEK293 and 3 genes enriched in HeLa, 1-hour PSI (Figure 1D, Supplementary Figure 4C). HEK293 enriched transcripts included 5.8S rRNA and mitochondrial rRNA transcripts whereas HeLa enriched transcripts represented 18S rRNA (Figure 4D, Supplementary Figure 4C, 4D). We speculate that these early differences are detected due to the elevated abundance of rRNA transcripts within a cell. NanopoReaTA offers an interactive utility enabling users to input specific genes and visualize both raw and normalized read counts (Supplementary Figure 4E). Using the "Gene-wise" utility, we introduced several of the DEGs as input genes and visualized their raw and normalized gene counts (Supplementary Figure 4E). This application proves valuable for monitoring specific genes of interest such as cell-specific marker genes. Following that, we conducted a DTE analysis using Salmon based read counts (Patro et al. 2017) in contrast to the DGE analysis using featureCounts (Liao et al. 2013) based read counts. It is acknowledged that employing diverse analysis tools may yield varying numbers and specific differentially expressed genes or transcripts. Thus, we intended to include several established tools, ensuring that significant results are consistently identified across different methods. DTE analyses offered by NanopoReaTA generate similar visual representations to DGE, including PCA, Sample-to-sample distance plots, volcano plots, and heatmaps (refer to Supplementary Figure 6A-D). Therefore, when examining the differentially expressed transcripts (DETs) at 1-hour PSI, we detected 3 transcripts enriched in HEK293 and 6 transcripts enriched in HeLa (Supplementary Figure 6C). Differentially expressed genes associated with these transcripts were also detected in DGE. These observations are further evident in the heatmaps generated for the top 20 differentially expressed genes (see

Supplementary Figure 4D) and transcripts (Supplementary Figure 6D), as generated, and exported from NanopoReaTA. Subsequently, as data was collected at 2 hours, 5 hours, 10 hours, and 24 hours PSI, our objective was to compare the entire dataset thus provide dynamic real-time view of the RNA sequencing run. As expected, a noticeable increase in the number of identified genes (Supplementary Figure 2C, 2J-K), as well as DEGs and DETs (Supplementary Figure 4C-D, Supplementary Figure 6C-D), was observed with the advancement of sequencing. This analysis demonstrates is the first evidence in the course of the analysis for the existence of differences in gene expression between different conditions and represents a valuable and dynamic quality control.

At 24-hour PSI, we examined the sample-to-sample distance plots and PCA plots (Figure 2E, 2F). Additionally, we collected basic quality control data, including the number of detected genes, gene variability, individual and combined read length distribution, and the usage timings of tools applied by NanopoReaTA, (see Supplementary Figure 2H-N). At this stage, we noted enhanced separation between the conditions, particularly evident in PC2, which accounted for 92% of the variance. This clustering trend persisted consistently throughout the entire sequencing process, evident in both DGE and DTE analyses (Supplementary Figure 4A, Supplementary Figure 6A). At this stage, we identified 81 genes enriched in HEK293 and 111 genes enriched in HeLa (Figure 2G, Supplementary Table 1). To assess the reproducibility of the results provided by NanopoReaTA, we cross-referenced the total differentially expressed genes identified at each time point, providing insights into the dynamic changes detected throughout sequencing. Remarkably, we identified 8 annotated genes (5 enriched in HEK293 and 3 enriched in HeLa) that were consistently detected across all time points, from 1-hour to 24-hours PSI (Figure 2H). Two additional differentially expressed genes (DEGs) were identified from the 2-hour time point, followed by 23 DEGs from the 5-hour time point and an additional 17 DEGs at 10 hours PSI (Figure 2H). These observations highlight the dynamic detection of DEGs during the ongoing sequencing process, emphasizing that distinctions in the most abundant transcripts likely emerge early after sequencing initiation. Subsequently, we cross-verified the identified DEGs between the PromethION and MinION sequencing. Here, we identified a total of 88 DEGs, which accounts for 70% of the DEGs identified in MinION sequencing and 45% of the DEGs identified in PromethION sequencing (25 enriched in HEK293/depleted in HeLa and 63 enriched in HeLa/depleted in HEK293) (Supplementary Table 1). These results are expected, as PromethION sequencing, with its deeper sequencing depth, generally yields more reliable results compared to MinION sequencing due to improved coverage, lower error rates, better detection of low-abundance transcripts, and enhanced statistical power. Nevertheless, we demonstrate that a notable portion of DEGs showed overlapping patterns between the sequencing platforms, aligning consistently with the conditions' enrichment/depletion profiles. (Supplementary Figure 5I).

NanopoReaTA offers a valuable utility for exploring differential transcript usage using DEXSeq (Anders et al. 2012) and DRIMSeq (Nowicka and Robinson 2016). DEXSeq provides an overview of isoforms differentially expressed through a volcano plot (Figure 6E), while DRIMSeq offers transcript-specific abundance presented as boxplots for individual selections. In our analysis, isoforms like RPL13A-201, RPS-201, and Eno1-201 showed high enrichment in HEK293, while GAPDH-211 was highly enriched in HeLa (Supplementary Figure 6E-I, Supplementary Table 1). Visualizing and exporting the expression of these specific DTUs can provide valuable biological insights into isoform-specific enriched transcripts.

Lastly, to validate that the DEGs and DETs correspond to each condition, we utilized the Harmonizome database (Rouillard et al. 2016) which contains a collection of datasets that compares the differential gene expression across different cell lines ("HPA Cell Line Gene Expression Profiles", Uhlén et al. 2015). We selected all the DEGs at the 24-hour time point (PromethION setup) and cross checked the enrichment of these genes within HEK293 and HeLa samples. Here, we observed a total of 71 genes (30 enriched in HEK293/depleted in HeLa and 41 enriched in HeLa/depleted in HEK293) corresponding to our real-time observations of NanopoReaTA (Figure 2I, Supplementary Figure 4F,4G). Genes such as RPS19, CA2, RPL18, and RPS4X enriched in HEK293/depleted in HeLa, as well as genes like ActB, Clu, ID3, and S100A6 enriched in HeLa/depleted in HEK293, were a few examples of the DEGs observed in the MinION experimental setup. These observations further support the overlap with the Harmonizome database. Taken all together, we show that the real-time transcriptional profiling between distinct cell populations is accurate and rapid, already revealing results at 1-hour PSI.

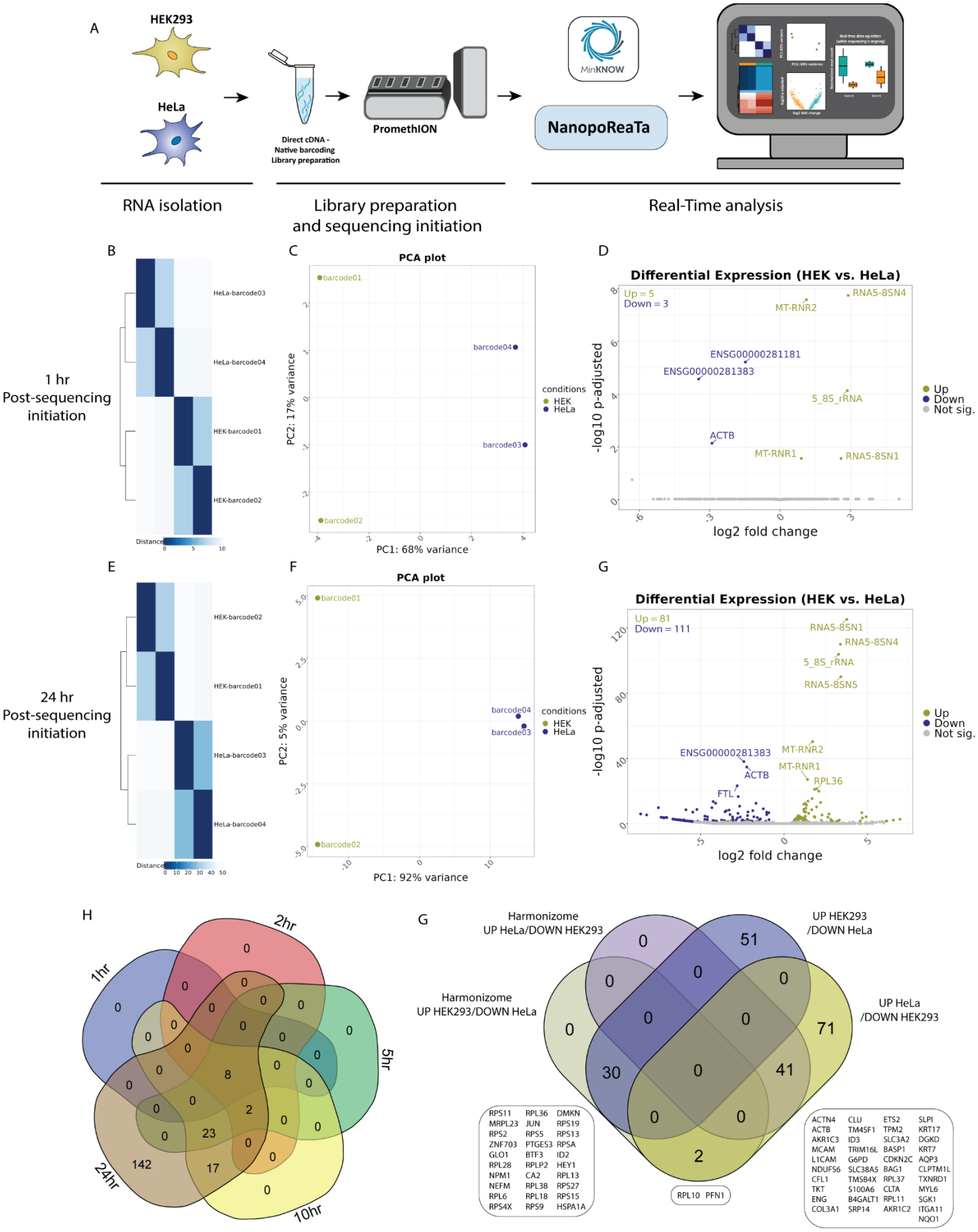


Figure 2. Real-time transcriptomic analysis between HEK293 and HeLa using NanopoReaTA. A. Experimental strategy. RNA was isolated from HEK293 and HeLa cells, the dscDNA library was prepared which included sample barcoding and adapter ligation. Samples were loaded and sequenced using a PromethION R10 flow cell. NanopoReaTA was activated shortly after sequencing initiation and data was collected 1hr, 2hr, 5hr, 10hr, and 24hr post-sequencing initiation. **B-D** Differential gene expression 1-hour post-sequencing initiation between HEK293 and HeLa samples. Selected data plots showing sample-to-sample distance plot **(B)**, pPCA **(C)** and volcano plot **(D)** 1hr- post sequencing initiation. **E-G.** Differential gene expression 24-hours post-

sequencing initiation between HEK293 and HeLa samples. Selected data plots showing sample-to-sample distance plot **(E)**, Principal component analysis (PCA) **(F)** and volcano plot **(G)** 24-hours post-sequencing initiation. **H.** Five-way Venn diagram showing the differentially expressed gene overlaps between the distinct collected time points. **I.** Validation of identified differentially expressed genes through Harmonizome database (Rouillard et al. 2016). For this analysis, we utilized "HPA Cell Line Gene Expression Profiles" (Uhlén et al. 2015). Among the DEGs (identified at the 24-hours time point), 35 genes were found to be enriched in HEK293 and depleted in HeLa, while 41 genes showed enrichment in HeLa and depletion in HEK293.

Real-time analysis provides rapid identification of experimentally enriched transcripts.

Next, we aimed to assess the rapid detection capabilities of NanopoReaTA for experimental-enriched transcripts. To achieve this, we performed ribosomal depletion using RiboMinus rRNA depletion (Thermo Scientific, K2561) on the previously tested HEK293 samples (Figure 1A, Supplementary Figure 7A). Different fractions of enriched transcripts, including ribosomal-depleted transcripts (RiboMinus/RiboM) and the depleted rRNA (RiboPlus/RiboP), were collected, along with total RNA (TotalR) from HEK293 as a control (Supplementary Figure 7B). We tested three comparisons including: totalR vs RiboM, totalR vs RiboP, and RiboM vs RiboP. Two replicates per condition were barcoded, and the samples were sequenced on a PromethION flow cell for 24 hours, with data collected at the same intervals as mentioned earlier (Figure 1A).

As for the previous experiment, 1-hour PSI, we gathered basic sequencing metrics from condition comparisons, including total reads generated per sample, along with mapped reads, gene counts, and transcript counts provided by NanopoReaTA (Supplementary Figure 7C-F, Supplementary Table 2). Interestingly, despite generating more total reads and identify more mapped reads for TotalR and RiboP, we observed a higher ratio in the number of identified gene and transcript counts in RiboM compared to the other two conditions (Supplementary Figure 7). These findings are intriguing, especially considering the lower amount of cDNA generated and loaded for RiboM compared to TotalR and RiboP. We discuss potential reasons for these observations in the Supplementary Material (Supplementary Material 2). Additionally, like in the previous section, we gathered basic quality control information, including the number of detected genes, gene variability, individual and combined read length distribution, and the usage timings of the tools applied by NanopoReaTA for quality control (Supplementary Figure 8, Supplementary Figure 11, Supplementary Figure 14).

Next, we inspected the PCA and dissimilarity plots (Figure 2B, Supplementary Figures 9A-B, 10A-B, 12A-B, 13A-B, 15A-B, 16A-B). We noticed clear distinctions between all three condition comparisons, evident as early as 1-hour PSI. PC1 represented 67% of the variance between RiboM vs TotalR (Figure 3B, 57% of the variance between RiboP vs TotalR (Figure 3I) and 72% of the variance between RiboP vs RiboM (Figure 3P). On the other hand, PC2 represented 20%, 29% and 19% of the variance in the respective comparison. In terms of detected DEGs 1-hour PSI, we identified 2 annotated genes enriched and 3 depleted in RiboM compared to TotalR (Figure 3C, Supplementary Table 2), 5 enriched and 0 depleted in RiboP compared to TotalR (Figure 3J) and 7 enriched and 1 depleted in RiboP compared to RiboM (Figure 3Q, Supplementary Figures 9C-D, 12C-D, 15C-D). Notably, the enriched annotated genes in RiboP and totalR compared to RiboM align with rRNA-related transcripts, which are predominantly enriched in RiboP (Figure 3F, 3M, 3T, Supplementary Figures 9E, 12E, 15E). When examining the DETs, we identified transcripts associated with the respective rRNA enrichment groups (Supplementary Figures 10A-E,13A-E,16A-E).

Similarly, we collected all the metrics corresponding to the sample/condition variability, annotated genes, and differentially expressed genes/transcripts for 2hr, 5hr, 10hr and 24hr PSI. At the 24-hours PSI time point, we observed a further separation of samples in the PCA according to their conditions (Figure 3D, 3K, 3R). Additionally, to test further the transcripts enrichment procedure, we monitored the number of identified genes in the last collection time points between the different conditions. Interestingly, by the conclusion of the 24-hour period PSI, we identified an average of 5627 genes in TotalR, 8632 genes in RiboM, and 3686 genes in RiboP (Figure 3G, 3N, 3V, Supplementary Figure 8C-D, Supplementary Figure 11C-D, Supplementary Figure 14C-D). While these results align with our expectations—given that RiboM is strongly depleted from ribosomal RNA and RiboP consists primarily of rRNA transcripts—the depletion of rRNA in the RiboM samples may have facilitated a more efficient enrichment of non-rRNA transcripts during the double-strand cDNA synthesis procedure. Consequently, this resulted in a higher number of detected genes compared to TotalR as well while RiboP exhibited the fewest detected genes, as anticipated.

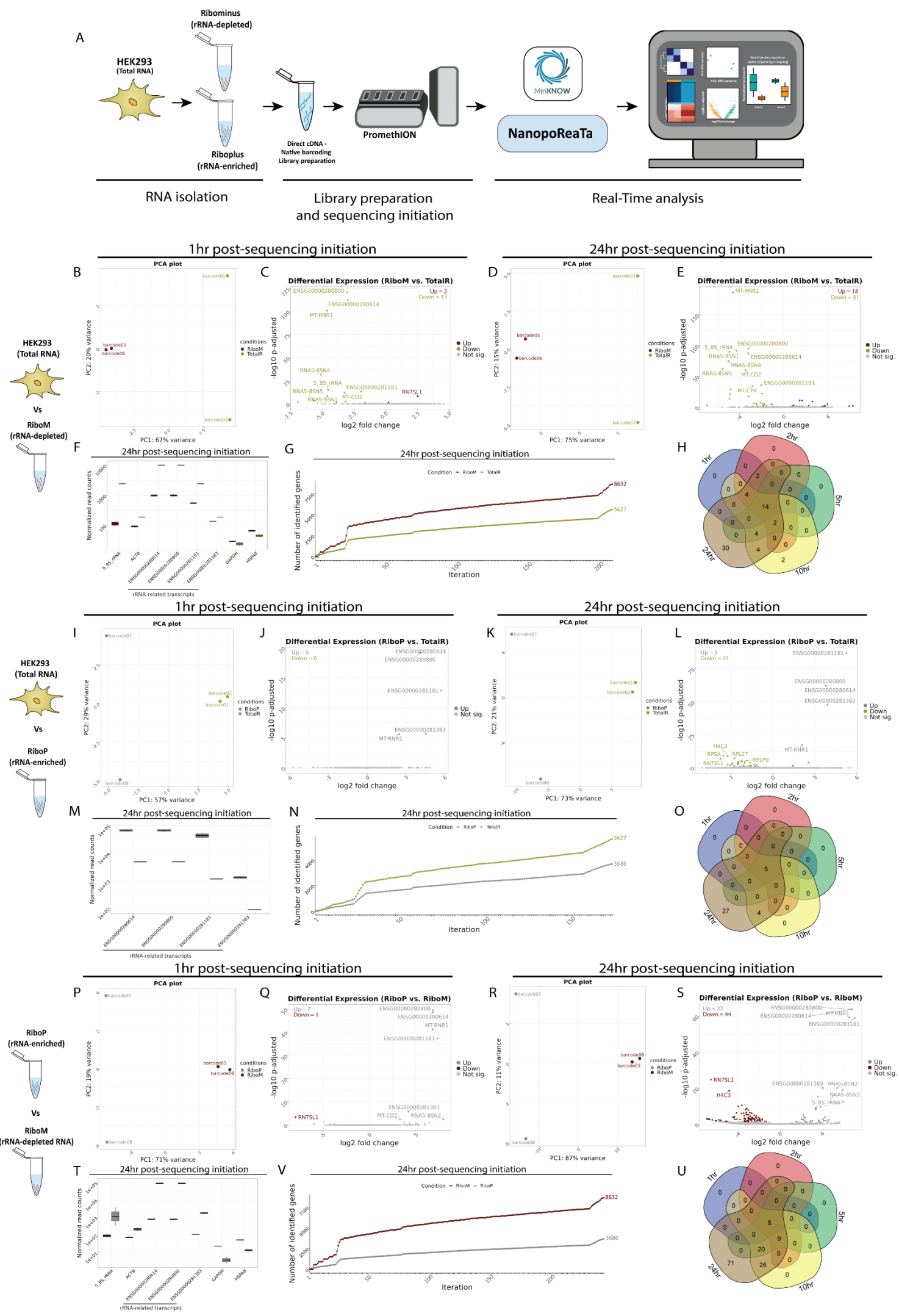

A
Ribominus
(rRNA-depleted)
HEK293
(Total RNA)
Riboplus
(rRNA-enriched)
Direct cDNA -
Native barcoding
Library preparation
PromethION
MinKNOW
NanopoReaTa
RNA isolation
Library preparation
and sequencing initiation
Real-Time analysis
1hr post-sequencing initiation
24hr post-sequencing initiation
HEK293
(Total RNA)
Vs
RiboM
(rRNA-depleted)
B
PCA plot
PC2: 20% variance
PC1: 67% variance
C
Differential Expression (RiboM vs. TotalR)
Up = 2
Down = 13
-log10 p-adjusted
log2 fold change
D
PCA plot
PC2: 15% variance
PC1: 75% variance
E
Differential Expression (RiboM vs. TotalR)
Up = 18
Down = 31
-log10 p-adjusted
log2 fold change
conditions
RiboM
TotalR
Up
Down
Not sig.
F
24hr post-sequencing initiation
Normalized read counts
rRNA-related transcripts
G
24hr post-sequencing initiation
Condition RiboM TotalR
Number of identified genes
Iteration
8632
5627
H
1hr
2hr
5hr
24hr
10hr
HEK293
(Total RNA)
Vs
RiboP
(rRNA-enriched)
I
PCA plot
PC2: 29% variance
PC1: 57% variance
J
Differential Expression (RiboP vs. TotalR)
Up = 5
Down = 0
-log10 p-adjusted
log2 fold change
K
PCA plot
PC2: 21% variance
PC1: 73% variance
L
Differential Expression (RiboP vs. TotalR)
Up = 5
Down = 31
-log10 p-adjusted
log2 fold change
M
24hr post-sequencing initiation
Normalized read counts
rRNA-related transcripts
N
24hr post-sequencing initiation
Condition RiboP TotalR
Number of identified genes
Iteration
5627
3686
O
1hr
2hr
5hr
24hr
10hr
RiboP
(rRNA-enriched)
Vs
RiboM
(rRNA-depleted RNA)
P
PCA plot
PC2: 19% variance
PC1: 71% variance
Q
Differential Expression (RiboP vs. RiboM)
Up = 7
Down = 1
-log10 p-adjusted
log2 fold change
R
PCA plot
PC2: 11% variance
PC1: 87% variance
S
Differential Expression (RiboP vs. RiboM)
Up = 33
Down = 44
-log10 p-adjusted
log2 fold change
conditions
RiboP
RiboM
Up
Down
Not sig.
T
24hr post-sequencing initiation
Normalized read counts
rRNA-related transcripts
V
24hr post-sequencing initiation
Condition RiboM RiboP
Number of identified genes
Iteration
8632
3686
U
1hr
2hr
5hr
24hr
10hr

Figure 3. Real-time transcriptomic analysis between rRNA-depleted and rRNA-enriched transcripts using NanopoReaTA. A. Experimental strategy. RNA was isolated from HEK293 and selective purification of distinct RNA populations was performed using the Ribominus™ Eukaryote kit. The dscDNA library was prepared, which included sample barcoding and adapter ligation. Samples were loaded and sequenced using a PromethION R10 flow cell. NanopoReaTA was activated shortly after sequencing initiation and data was collected 1hr, 2hr, 5hr, 10hr, and 24hr post-sequencing initiation. **B-H.** Differential gene expression between RiboM and TotalR from HEK293. Selected data plots showing PCA and volcano plots 1hr (**B-C**) and 24hr (**D-E**) post sequencing initiation. **F.** Normalized gene counts for selected genes 24hr. Normalized gene counts are visualized for selected genes per condition using boxplots. The median-of-ratios normalization method from DESeq2 was used for normalization. **G.** Gene expression variability. The number of identified genes (> 0 reads counted) is plotted after each iteration per condition. **H.** Five-way Venn diagram showing the differentially expressed gene overlaps between the distinct collected time points. **I-O.** Differential gene expression between RiboP and TotalR. Similar analyses to B-H were conducted for comparisons at 1hr (**I-J**) and 24hr (**K-L**) as well as normalized gene counts (**M**), gene expression variability (**N**) and Venn diagram (**O**). **P-U.** Differential gene expression compared between RiboM and RiboP. Similar analyses to B-H were conducted for comparisons at 1hr (**P-Q**) and 24hr (**R-S**) as well as normalized gene counts (**T**), gene expression variability (**V**) and Venn diagram (**U**).

Lastly, we overlapped the total differentially expressed genes identified at each time point to test the reproducibility of the changes detected throughout sequencing. Here, we identified 14 annotated genes in RiboM vs TotalR (Figure 3H) 5 annotated genes in RiboP vs TotalR (Figure 3O) and 8 annotated genes in RiboP vs RiboR that were consistently detected across all time points (Figure 3U). Interestingly, we noted an enrichment of mitochondrial rRNA in RiboP samples, which had previously been reported as depleted within the RiboMinus™ Eukaryote Kit, thereby reinforcing the robustness of our experimental design (Qu et al., 2013). Overall, these findings highlight further the swift detection capabilities among transcript-enriched samples, serving as a valuable quality control measure for the rapid identification of ribosomal-depleted or polyA enrichment strategies.

NanopoReaTA offers rapid quality control assessments for experimental manipulated samples.

In our final aim, we sought to highlight the flexibility of NanopoReaTA in an experimental manipulation setup and its applicability beyond human cell culture. To achieve this, we employed *S. cerevisiae* strains harboring gene knockouts or strains transformed with plasmids containing the deleted gene for overexpression. Two distinct experimental setups were designed to assess the reproducibility and detection capabilities of NanopoReaTA. In the first experimental setup (Yeast setup 1), we utilized *new1Δ::KanMX* yeast strains, where the *NEW1* gene (coding sequence only) was replaced with the KanMX cassette which contains the Kanamycin resistance gene (*KanR*). We used the wild type (WT) strain (BY4741, *MATa, his3Δ1, leu2Δ0, met15Δ0, ura3Δ0*) for comparison with the KO strain. These strains were transformed with either an empty vector with the *HIS3* selection marker (pEV(*HIS3*)) or an overexpression vector built on the same backbone, for C-terminally FLAG-tagged New1 with the same *HIS3* selection marker (pNew1(*HIS3*)) (Figure 4A, Supplementary Figure 17A). All *S. cerevisiae* knockout strains derived from BY4741 were prepared using homologous recombination following standard procedures. Furthermore, we employed customized yeast genome annotation files that included *KanR*, Hygromycin resistance gene (*HygR*) and Ampicillin resistance gene (*AmpR*; contained to allow propagation of the shuttle vectors in *E. coli*) transcripts, ensuring the detection of foreign transcripts specific to their corresponding experimental setup, thus adding an extra layer of quality control (see “Material and Methods”). For each experimental setup, we collected data from NanopoReaTA, including general sequencing overview with individual and combined read length distribution, the count of detected genes, gene expression variability, and the timing of tool utilization (Supplementary Figures 18, 21,24,27,30). Similar to the previous section, we conducted real-time analyses for both DGE and DTE, documenting all associated information at each time point (Supplementary Figures 19-20, 22-23, 25-26, 28-29, 31-32, Supplementary Table 3). For the following section we will focus on describing the detected changes in differentially expressed genes/transcripts, however general sequencing overviews for each individual time points are presented in supplementary figures.

Yeast setup 1

WT-pNew1(HIS3) vs WT-pEV(HIS3). For yeast setup 1, we assessed NanopoReaTA's capability to identify the expressed content of the transformed plasmids by contrasting the WT strain transformed with pNew1(*HIS3*) against the WT strain transformed with pEV(*HIS3*). In the comparison of WT pNew1(*HIS3*) versus WT-pEV(*HIS3*) at 1hr PSI, the PCA effectively distinguished the samples based on their respective conditions (Figure 4B, Supplementary Figure 19A). However, the sample-to-sample

distance plot did not uncover significant differences between the replicates (Supplementary Figure 19B). The distinct clustering observed in PCA was consistently maintained throughout the entire sequencing process (Supplementary Figure 19A) until the 24-hour PSI mark point (Figure 9B, Supplementary Figure 19A). The sample-to-sample distance plot indicated greater similarities between the samples, which is anticipated given the comparison involves similar WT strains harboring either pNew1(*HIS3*) or pEV(*HIS3*). Notably, *NEW1* was the sole differentially expressed gene (Figure 9D) and transcript (Supplementary Figure 20C-D) identified at the 1-hour PSI mark in the WT strain supplemented with pNew1(His3), in comparison to the WT strain *transformed* with pEV(*HIS3*). This difference was maintained throughout the whole sequencing (Figures 4E-G). Using the "Gene-wise" feature in NanopoReaTA, we tracked, in real-time, the normalized read counts of various genes of interest, such as *NEW1*, *KanR*, *AmpR*, and *HIS3*, along with housekeeping genes such as *ALG9 and TFC1* (Teste et al., 2009), and the commonly used versions of Gapdh-encoding yeast genes *TDH1*, *TDH2*, and *TDH3* (Figure 4F, Supplementary Figure 19E). This setup demonstrates an overexpression experiment, showcasing NanopoReaTA's capability to swiftly detect the overexpressed gene.

New1Δ-pNew1*(HIS3) vs new1Δ*-pEV*(HIS3).* Next, we compared *new1Δ* strains harboring either pNew1(*HIS3*) or pEV(*HIS3*) at 1hr PSI. As in WT, the PCA successfully differentiated the samples according to their respective conditions (Figure 4H Supplementary Figure 22A). Nevertheless, the sample-to-sample distance plot did not reveal notable differences between the replicates (Supplementary Figure 22B). The distinct clustering observed in PCA remained consistent throughout the entire sequencing process, until the 24-hour PSI point (Figure 4I, Supplementary Figure 22A). New1 emerged as the only gene and transcript enriched in pNew1(*HIS3*) as compared to pEV(*HIS3*) at the 1hr PSI (Figure 4J, Supplementary Figures 22C-D,23C-D). This observed difference persisted throughout the entire sequencing process (Fig 4M, Supplementary Figures 22C-D, 22F, 23C-D). Interestingly, as sequencing was going through, the number of DEG/T increased up until the 24hr PSI mark where 13 genes were enriched and 2 were depleted in *new1Δ*-pNew1(*HIS3*) compared to *new1Δ*-pEV(*HIS3*) (Figure 4K-L). This experimental configuration exemplifies a rescue experiment, highlighting NanopoReaTA's ability to promptly detect the overexpressed gene.

new1Δ-pNew1(*HIS3*) *vs WT*-pEV(*HIS3*)*.* When comparing *new1Δ*-pNew1(*HIS3*) versus WT-pEV(*HIS3*) at 1hr PSI, the PCA indicated separation between the samples based on their respective conditions (Supplementary Figure 25A, 26A); however, the clustering in the sample-to-sample distance plot appeared inconsistent, likely due to the rescue of New1 in the *new1Δ* mutant strain (Supplementary Figure 25B, 26B). Remarkably, we consistently detected both *KanR* and *NEW1* as enriched in *new1Δ*-pNew1(*HIS3*) compared to WT-pEV(*HIS3*) throughout the entire sequencing process, and both as differentially expressed genes and transcripts (Supplementary Figure 25C-F, 26C-E). These observations highlight NanopoReaTA's rapid detection capabilities in an experimental setup where the deleted gene is rescued by the overexpression plasmid compared to WT with an empty vector.

new1Δ-pEV(*HIS3*) vs WT-pEV(*HIS3*)*.* Next, we aimed to test NanopoReaTA's detection capabilities in foreign gene detection since the coding sequence of the *NEW1* gene was replaced via homologous recombination with the KanMX cassette which contains the *KanR* antibiotic resistance gene. In the comparison of *new1Δ*-pEV(*HIS3*) versuss WT-pEV(*HIS3*) at 1hr PSI, the PCA revealed separation between the samples based on their respective conditions (Figure 4N, Supplementary Figure 28A, 29A), as well as clustering in the sample-to-sample distance plot (Supplementary Figure 28B, 29B). This clustering pattern persisted until the 24hr PSI time point (Fig 4O, Supplementary Figure 28A, 29A). Notably, in every comparison group, we noted one sample that exhibited a slight separation from the condition cluster in the PCA, in contrast to the other two replicates. However, these differences did not raise significant concerns. Nonetheless, the real-time PCA clustering feature of NanopoReaTA could prove valuable when assessing biological replicates. Upon conducting differential gene and transcript expression analysis at 1hr PSI, we detected 13 genes enriched and 3 genes depleted in *new1Δ*-pEV(*HIS3*) compared to *WT*-pEV(*HIS3*) (Figure 4P, Supplementary Figure 28C). Notably, *KanR* emerged as one of the enriched genes with the highest log2 fold change, alongside the highly significant *HSP12*, both in DGE and DTE analyses.

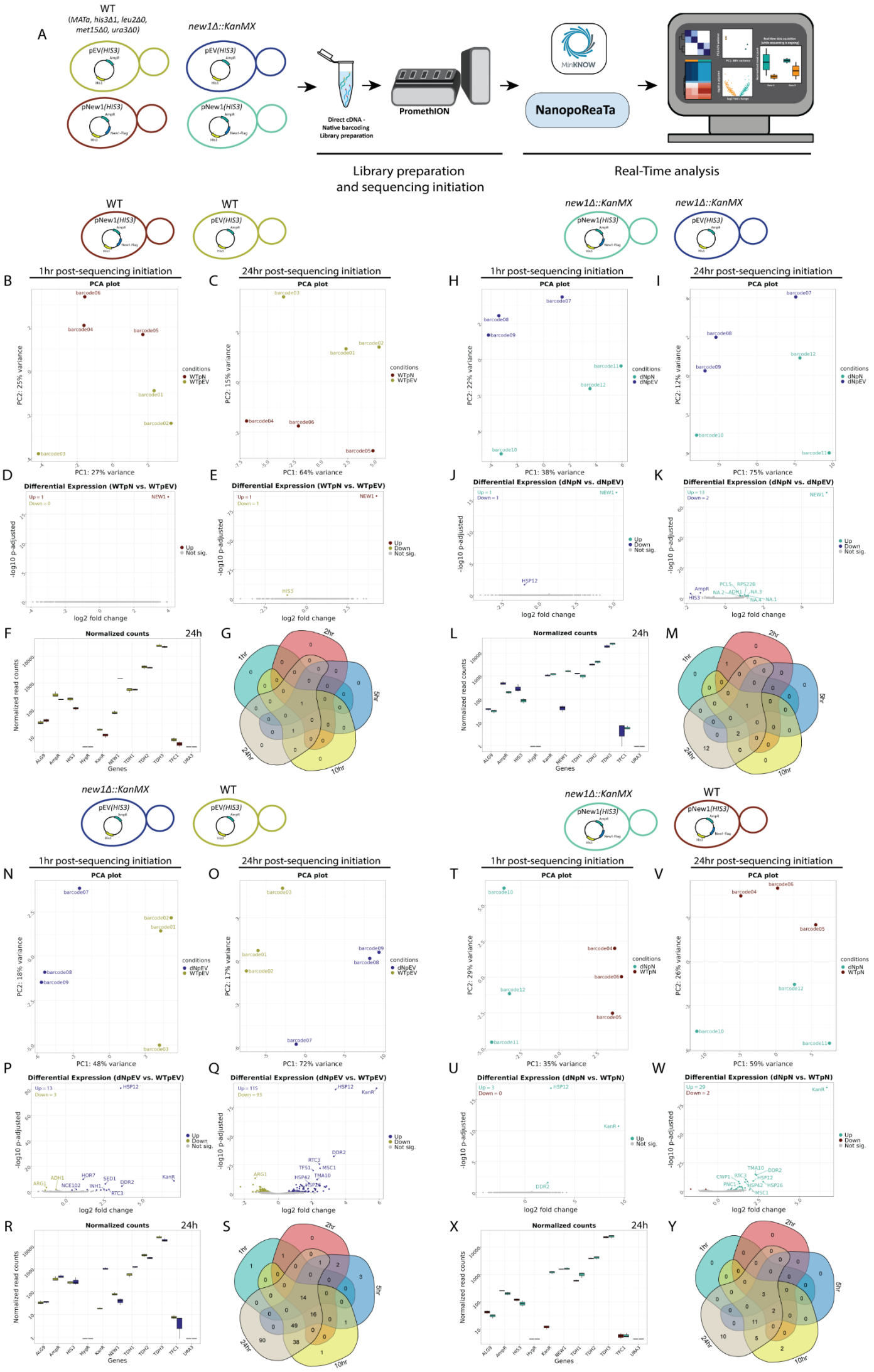

A
WT
(MATa, his3Δ1, leu2Δ0, met15Δ0, ura3Δ0)
new1Δ::KanMX
pEV(HIS3)
pNew1(HIS3)
Direct cDNA - Native barcoding Library preparation
PromethION
MinKNOW
NanopoReaTa
Library preparation and sequencing initiation
Real-Time analysis
1hr post-sequencing initiation
24hr post-sequencing initiation
PCA plot
PC1: 27% variance
PC2: 25% variance
PC1: 64% variance
PC2: 15% variance
PC1: 38% variance
PC2: 22% variance
PC1: 75% variance
PC2: 12% variance
conditions
WTpN
WTpEV
dNpN
dNpEV
Differential Expression (WTpN vs. WTpEV)
Differential Expression (dNpN vs. dNpEV)
log2 fold change
-log10 p-adjusted
Up
Down
Not sig.
Normalized counts
Normalized read counts
Genes
1hr
2hr
5hr
10hr
24hr
PC1: 48% variance
PC2: 18% variance
PC1: 72% variance
PC2: 17% variance
PC1: 35% variance
PC2: 29% variance
PC1: 59% variance
PC2: 26% variance
Differential Expression (dNpEV vs. WTpEV)
Differential Expression (dNpN vs. WTpN)

Figure 4. Real-time transcriptomic analysis between experimental manipulated yeast setup 1 samples using NanopoReaTA. A. Experimental strategy of yeast setup 1. RNA was isolated from selected yeast strains and dscDNA library was prepared which included samples barcoding and adapter ligation. Samples were loaded and sequenced using a PromethION R10 flow cell. NanopoReaTA was activated shortly after sequencing initiation and data was collected 1hr, 2hr, 5hr, 10hr, and 24hr post-sequencing initiation. **B-G.** Differential gene expression in WT-pNew1(*HIS3*) versus WT-pEV(*HIS3*). Selected data plots showing PCA (**B-C**) and volcano plots (**D-E**) 1hr and 24hr post sequencing initiation. **F.** Normalized gene count for selected genes 24hr post sequencing initiation. Normalized gene counts are visualized for selected genes per condition using boxplots. The median-of-ratios normalization method from DESeq2 was used for normalization. **G.** Five-way Venn diagram showing the differentially expressed gene overlaps between the distinct collected time points. **H-M.** Differential gene expression in *new1Δ*-pNew1(*HIS3*) versus *new1Δ*-pEV(*HIS3*). Similar analyses to B-G were conducted for PCA (**H-I**) and volcano plots (**J-K**) as well as normalized gene counts (**L**) and Venn diagram (**M**). **N-S.** Differential gene expression in *new1Δ*-pEV(*HIS3*) versus WT-pEV(*HIS3*). Similar analyses to B-G were conducted for PCA (**N-O**) and volcano plots (**P-Q**) as well as normalized gene counts (**R**) and Venn diagram (**S**). **T-Y.** Differential gene expression in *new1Δ*-pNew1(*HIS3*) versus WT-pNew1(*HIS3*). Similar analyses to B-G were conducted for PCA (**T-V**) and volcano plots (**U-W**) as well as normalized gene counts (**X**) and Venn diagram (**Y**).

These observations were consistent throughout the entire sequencing period, extending up to the 24hr mark (Figure 4Q-S, Supplementary Figure 28C-F). These results demonstrate the possibility of detection of foreign transcripts incorporated instead of knockout gene. A previous study conducted RNA-seq between *new1Δ* and WT (Kasari et al., 2019); therefore, we overlapped the identified differentially expressed genes (DEGs) to examine the commonality between the detected DEGs. Within the overlap, four upregulated genes and 11 downregulated genes, including *HSP12* and *NEW1*, were found to be common between Kasari et al. (2019) and our investigation (Supplementary Figure 28G), despite variations of growth conditions, as well as exact yeast genotypes between this study and Kasari et al. (2019). As demonstrated, NanopoReaTA can swiftly identify an experimental knockout, and in instances where the gene is replaced with a foreign gene, it can also detect this alteration effectively given that the foreign gene is incorporated into the genome annotation files.

new1Δ-pNew1*(HIS3) versus WT*-pNew1*(HIS3).* In the comparison of *new1Δ*-pNew1(*HIS3*) versus WT-pNew1(*HIS3*) at 1hr PSI, the PCA exhibited separation between the samples based on their respective conditions (Figure 4T, Supplementary Figure 31A, 32A), accompanied by clustering in the sample-to-sample distance plot (Supplementary Figure 31B, 32B). This clustering pattern persisted until the 24hr PSI time point (Figure 4V, Supplementary Figure 31A, 32A). *KanR*, *DDR2* and *HSP12* were identified as enriched genes in *new1Δ*-pNew1(*HIS3*) compared to WT-pNew1(*HIS3*) at 1hr time point, appearing as both differentially expressed genes and transcripts (Figure 4U, Supplementary Figure 31C-F, 32C-E). At 24hr PSI, only 29 genes were enriched and 2 were depleted in *new1Δ*-pNew1(*HIS3*) compared to WT-pNew1(*HIS3*), implying a potential transcriptional overcompensation facilitated by the transformed plasmid encoding *NEW1* (Figure 4W-X, Supplementary Figure 31C-F, 32C-E).

It is important to highlight that using the "gene-wise" utility (Figure 4F, 4L, 4R, and 4X), we identified discrepancies in the reads associated with each condition. For instance, in the *new1Δ* condition, where the *NEW1* gene (coding sequence only) has been replaced with the *KanMX* cassette containing the *KanR* gene, some *NEW1* transcripts still aligned and were quantified. Additionally, we observed the presence of condition-specific transcripts (e.g., *KanR*, expected only in *new1Δ* mutants) in WT conditions, though in low quantities. A detailed discussion of these discrepancies is provided in the Supplementary Material (Supplementary Material 3). These findings highlight the ability of NanopoReaTA to offer valuable quality control insights that could reveal experimental flaws, such as contaminations, which could then be rapidly addressed and rectified.

Yeast setup2

In the second experimental setup (setup 2), we employed the *rkr1Δ::HphMX* strain, where the coding sequence of the *RKR1* gene was replaced with the *HphMX cassette encoding the* Hygromycin B resistance gene *HygR*. Additionally, we used the double KO strain *jlp2Δ::KanMX, rkr1Δ::HphMX*, where the coding sequence of the *JLP2* gene was replaced with KanMX cassette containing *KanR*. For this setup, these strains were transformed with either an empty vector with the *URA3* selection marker (pEV(*URA3*)) or an overexpression vector for C-terminally HA-tagged Jlp2 with the *URA3* selection marker (pJlp2(*URA3*)).

rkr1Δ-pJlp2(*URA3*) *versus rkr1Δ*-pEV(*URA3*). First, we tested the comparison of *rkr1Δ*-pJlp2(*URA3)* compared to *rkr1Δ*-pEV(*URA3*) for the detection of *JLP2* overexpression. One hour after the initiation of sequencing, PCA successfully differentiated the samples based on their respective conditions, with a slight separation observed in one replicate (barcode 16- *rkr1Δ*-pJlp2(*URA3*)) (Figure 5B, Supplementary Figure 35A, 36A). The sample-to-sample distance plot did not uncover significant differences between the replicates (Supplementary Figure 35B, 36B). The distinct clustering observed in PCA was consistently maintained throughout the entire sequencing process (Supplementary Figure 35A, 36A) until the 24-hour PSI mark point (Figure 5C, Supplementary Figure 35A, 36A). Notably, two genes/transcripts including *JLP2* and *RPL15A* were differentially expressed (Figure 5D, Supplementary Figure 35C-D, 36C-D) from 1hr PSI mark until 24hr PSI mark (Figure 5E-G, Supplementary Table 3). Thus, NanopoReaTA was able to detect the overexpressed gene from the plasmid.

rkr1Δ, jlp2Δ-pJlp2(*URA3*) versus *rkr1Δ, jlp2Δ*-pEV(*URA3*). Similarly, we tested *rkr1Δ, jlp2Δ-pJlp2(URA3*) compared to *rkr1Δ, jlp2Δ-pEV(URA3)* for the detection of *JLP2* overexpression. When comparing *rkr1Δ, jlp2Δ*-pJlp2(*URA3*) versus *rkr1Δ, jlp2Δ*-pEV(*URA3*) at 1hr PSI, the PCA showed separation between the samples based on their respective conditions (Figure 5H, Supplementary Figure 38A, 39A); Similarly, the clustering in the sample-to-sample distance plot appeared inconsistent (Supplementary Figure 38B, 39B). The clustering observed in PCA remained consistent throughout the entire sequencing process until the 24-hour PSI time point (Figure 5I, Supplementary Figure 38A, 39A). Remarkably, we detected only *Jlp2* as enriched in *rkr1Δ, jlp2Δ*-pJlp2(*URA3*) compared to *rkr1Δjlp2Δ*-pEV(*URA3*) throughout the entire sequencing process, in DEG and DET analyses (Figure 5J-M, Supplementary Figure 38C-F, 38C-E, Supplementary Table 4). At the 24-hour PSI mark, we observed enrichment of *JLP2* and depletion of 11 genes in *rkr1Δ, jlp2Δ*-pJlp2(*URA3*) compared to *rkr1Δ, jlp2Δ*-pEV(*URA3*). These observations highlight the swift and accurate detection capabilities of NanopoReaTA in an experimental setup where similar strains are compared, and only one is transformed with an overexpression vector, thereby illustrating the rescue by the expressed gene.

rkr1Δ-pEV(*URA3*) versus WT-pEV(*HIS3*). Next, we aimed to assess an additional layer of quality control by comparing the KO strains with the WT strain, utilizing WT-pEV(*HIS3*), and detecting the distinct selection genes present in each condition. Initially, we conducted a comparison between *rkr1Δ*-pEV(*URA3*) and WT-pEV(*HIS3*). At 1hr PSI, a notable separation between the conditions was evident in both the PCA and sample-to-sample distance plot (Figure 5N, Supplementary Figure 41A-B, 42A-B). This distinction persisted consistently until the 24-hour PSI time point (Figure 5O, Supplementary Figure 41A-B, 42A-B). At 1hr PSI, 59 genes were enriched and 49 were detected as depleted in *rkr1Δ*-pEV(*URA3*) as compared to WT-pEV(*HIS3*). As expected, *URA3* and *HygR* were detected as enriched in *rkr1Δ*-pEV(*URA3*) (Figure 5P, Supplementary Figure 41C-D, 42C-D). At this stage, *HIS3* was not identified in the DEG analysis, but rather in the DET analysis. At the 24-hour PSI mark, 523 genes were enriched (including *URA3* and *HygR*) and 381 were depleted (including *HIS3*) in *rkr1Δ*-pEV (*URA3*) as compared to WT-pEV(*HIS3*) (Figure 5Q-R, Supplementary Figure 41C-F, 42C-E). Interestingly, 104 differentially expressed genes were consistently identified from the 1-hour mark until the final 24-hour mark. This experimental configuration demonstrates the detection of the knockout of an individual gene, the detection of *HygR*, *URA3* and *HIS3* selection genes as well as a large number of additional DEGs/DETs that could validate the mechanistic function of the mutant investigated, as well as differences between yeast grown in different culturing conditions.

rkr1Δ, jlp2Δ-pEV(*URA3*) versus WT-pEV(*HIS3*). Lastly, we compared *rkr1Δ, jlp2Δ*-pEV(*URA3*) to WT-pEV(*HIS3*). At 1hr PSI, noticeable separation between conditions was evident in both, PCA and sample-to-sample distance plots (Figure 5T, Supplementary Figure 44A-B, 45A-B). This distinction persisted consistently until the 24-hour PSI time point (Figure 5V, Supplementary Figure 44A-B, 45A-B). At 1hr PSI, 69 genes were enriched, and 34 were depleted in *rkr1Δ*-pEV(*URA3*) compared to WT-pEV(*HIS3*). As expected, *URA3* and *HygR* were detected as enriched in *rkr1Δ*-pEV(*URA3*) (Figure 5U, Supplementary Figure 41C-D, 42C-D). Similar to the previous observation, *HIS3* was not identified in the DEG analysis but rather in the DET analysis. By the 24-hour PSI mark, 612 genes were enriched (including *URA3* and *HygR*), and 424 were depleted (including *HIS3*) in *rkr1Δ*-pEV(*URA3*) compared to WT-pEV(*HIS3*) (Figure 5W-Y, Supplementary Figure 44C-F, 45C-E). A total of 101 differentially expressed genes were consistently identified from the 1-hour mark until the final 24-hour mark. This experimental setup effectively detected the double knockout, as well as the expression of *HygR*, *KanR*, *URA3*, and *HIS3* in their corresponding experimental conditions.

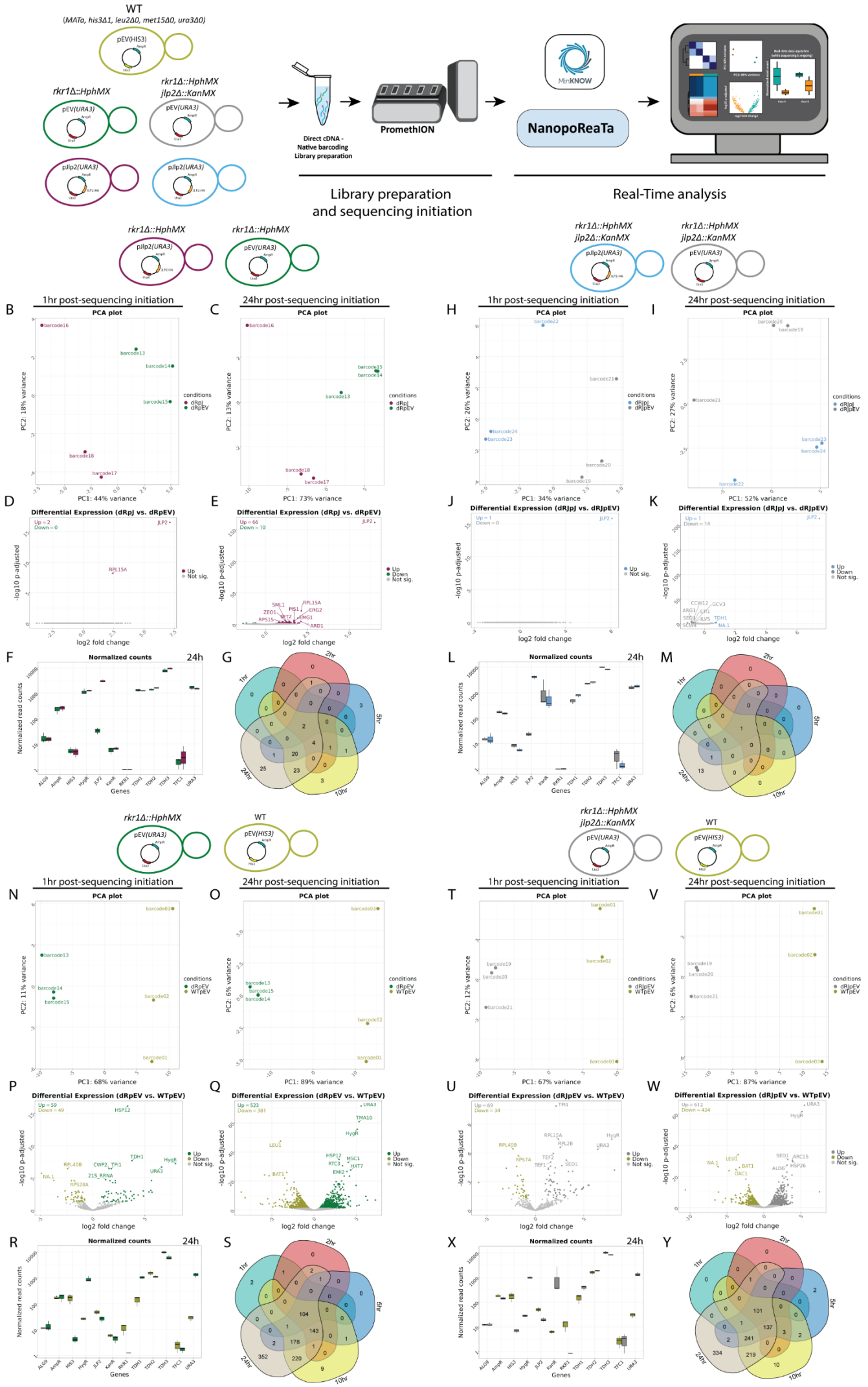

WT
(MATa, his3Δ1, leu2Δ0, met15Δ0, ura3Δ0)
pEV(HIS3)
rkr1Δ::HphMX
pEV(URA3)
rkr1Δ::HphMX
jlp2Δ::KanMX
pEV(URA3)
pJlp2(URA3)
pJlp2(URA3)
Direct cDNA - Native barcoding Library preparation
PromethION
MinKNOW
NanopoReaTa
Library preparation and sequencing initiation
Real-Time analysis
B
1hr post-sequencing initiation
PCA plot
PC1: 44% variance
PC2: 18% variance
C
24hr post-sequencing initiation
PCA plot
PC1: 73% variance
PC2: 13% variance
conditions
dRpJ
dRpEV
D
Differential Expression (dRpJ vs. dRpEV)
Up = 2
Down = 0
log2 fold change
-log10 p-adjusted
Up
Not sig.
E
Differential Expression (dRpJ vs. dRpEV)
Up = 66
Down = 10
F
Normalized counts
24h
Normalized read counts
Genes
G
1hr
2hr
5hr
10hr
24hr
H
1hr post-sequencing initiation
PCA plot
PC1: 34% variance
PC2: 26% variance
I
24hr post-sequencing initiation
PCA plot
PC1: 52% variance
PC2: 27% variance
conditions
dRjpJ
dRjpEV
J
Differential Expression (dRjpJ vs. dRjpEV)
Up = 1
Down = 0
K
Differential Expression (dRjpJ vs. dRjpEV)
Up = 1
Down = 14
L
Normalized counts
24h
M
N
1hr post-sequencing initiation
PCA plot
PC1: 68% variance
PC2: 11% variance
conditions
dRpEV
WTpEV
O
24hr post-sequencing initiation
PCA plot
PC1: 89% variance
PC2: 6% variance
P
Differential Expression (dRpEV vs. WTpEV)
Up = 59
Down = 49
Q
Differential Expression (dRpEV vs. WTpEV)
Up = 523
Down = 381
R
Normalized counts
24h
S
T
1hr post-sequencing initiation
PCA plot
PC1: 67% variance
PC2: 12% variance
V
24hr post-sequencing initiation
PCA plot
PC1: 87% variance
PC2: 6% variance
U
Differential Expression (dRjpEV vs. WTpEV)
Up = 69
Down = 34
W
Differential Expression (dRjpEV vs. WTpEV)
Up = 612
Down = 424
X
Normalized counts
24h
Y

Figure 5. Real-time transcriptomic analysis between experimental manipulated yeast setup 2 samples using NanopoReaTA. A. Experimental strategy of yeast setup 2. RNA was isolated from selected yeast strains and the dscDNA library was prepared, which included sample barcoding and adapter ligation. Samples were loaded and sequenced using a PromethION R10 flow cell. NanopoReaTA was activated shortly after sequencing initiation and data was collected 1hr, 2hr, 5hr, 10hr, and 24hr post-sequencing initiation. **B-G.** Differential gene expression in *rkr1Δ*-pJlp2(*URA3*) versus *rkr1Δ*-pEV(*URA3*). Selected data plots showing PCA (**B-C**) and volcano plots (**D-E**) 1hr and 24hr post sequencing initiation. **F.** Normalized gene counts for selected genes 24hr post-sequencing initiation. Normalized gene counts are visualized for selected genes per condition using boxplots. The median-of-ratios normalization method from DESeq2 was used for normalization. **G.** Five-way Venn diagram showing the differentially expressed gene overlaps between the distinct collected time points. **H-M.** Differential gene expression in *rkr1Δ, jlp2Δ*-pJlp2(*URA3*) versus *rkr1Δ, jlp2Δ*-pEV(*URA3*). Similar analyses to B-G were conducted for PCA (**H-I**) and volcano plots (**J-K**) as well as normalized gene counts (**L**) and Venn diagram (**M**). **N-S.** Differential gene expression in *rkr1Δ*-pEV(*URA3*) versus WT-pEV(*HIS3*). Similar analyses to B-G were conducted for PCA (**N-O**) and volcano plots (**P-Q**) as well as normalized gene counts (**R**) and Venn diagram (**S**). **T-Y.** Differential gene expression in in *rkr1Δ, jlp2Δ*-pEV(*URA3*) versus WT-pEV(*HIS3*). Similar analyses to B-G were conducted for PCA (**T-V**) and volcano plots (**U-W**) as well as normalized gene counts (**X**) and Venn diagram (**Y**).

Similar to yeast experimental setup 1, we observed unexpected findings in the reads associated with each condition using the "gene-wise" utility. For example, JL*P2* expression was detected in *rkr1Δjlp2Δ*-pEV(*URA3*) in low quantities, and *HIS3* was observed in strains, where the selection plasmid should not contain the *HIS3* selection marker. A comprehensive discussion of these observations is presented in the Supplementary Material (Supplementary Material 3). Nonetheless, NanopoReaTA can rapidly detect experiment-specific transcripts associated with the experimental condition. This application can be utilized to quickly identify knockout, knockdown or overexpression experiments and to quantify foreign transcripts that are not naturally present in the species' genome.

DISCUSSION

We presented a proof-of-concept application use of NanopoReaTA demonstrating its rapid detection capabilities of pairwise transcriptomic changes and for the first time, real-time dynamics of long read RNA-seq throughout the sequencing process. NanopoReaTA can be used as a multi-species transcriptomic detection tool revealing its broad utility. The tool requires well annotated genomes including genome sequence (FASTA files), annotated transcripts (FASTA files), gene annotation (GTF files), and gene coordinates throughout the genome (BED files). Additionally, NanopoReaTA works in combination with MinION/GridION flow cells, however, due to their reduced throughput compared to PromethION flow cells, achieving statistically meaningful results (e.g. larger number of DEGs) may be limited or take longer.

The straightforward utilization of NanopoReaTA, coupled with an intuitive graphical user interface (GUI), facilitates its smooth integration into daily experimental setups for quality checks in transcriptomic data analysis. The tool swiftly identifies transcriptomic differences between distinct cell types, compartment-enriched transcripts, or genetically manipulated cells, even within the first hour post-sequencing initiation. It is highly probable that these early detected changes represent the most significant transcripts, present or highly expressed in one condition versus absent or lowly expressed in another condition. These noteworthy early alterations persist throughout the entire sequencing process until its completion. As sequencing progresses and more reads are acquired, there is an increase in the number of detected genes, as well as genes and transcripts detected as differentially expressed (DEGs & DETs). It is important to note that these DEGs and DETs may undergo changes over the sequencing process as the data is normalized to the total read counts within the compared conditions (Evans et al. 2017). We incorporated into NanopoReaTA both differential gene/transcript expression analyses, performed by DESeq2 (Love et al. 2014), and quantification of genes and transcripts was performed by featureCounts (Liao et al. 2013) and Salmon (Patro et al. 2017), respectively. It is acknowledged that utilizing different analysis tools may lead to detection of varying numbers and tool-specific differentially expressed genes or transcripts (Thawng et al. 2023). Therefore, by offering both analyses, our intention is to provide orthogonal methodologies, ensuring that the most significant outcomes are consistently identified across different methods. Moreover, given the capability of capturing complete transcripts with long-read sequencing, we integrated a "differential transcript usage" application performed by DEXseq (Anders et al. 2012) and DRIMSeq (Nowicka and Robinson 2016). These applications are dedicated to the analysis and quantification of different isoforms per selected gene. This utility proves beneficial in

uncovering or determining the predominant isoform used between two conditions and utilizing it more frequently could unveil novel biological insights.

NanopoReaTA provides multi-layer quality control of several distinct experimental setups. On the first layer, NanopoReaTA can provide information regarding the number of genes identified, both per sample and per condition, as well as the changes in gene composition detected in each iteration compared to the previous one. When no additional genes are detected, the “Gene expression variability” lines reach a plateau, and the sequencing can be practically terminated (depending on the desired read depth). Such quality control provides a cost-efficient strategy when coupled with the Nanopore-seq washable flow cell that can be reused for separate experimental setups. Moreover, this analysis provides relevant biological insights into the number of genes expressed under specific conditions, a factor that may vary across different cell types or distinct experimental conditions. Another level of quality control can be applied when comparing distinct cell types or strains, where several cell-type/strain specific gene markers can be examined. Using the “Gene-wise” utility, these marker genes could be monitored in real-time providing quality control for the cell-type/strain specific purity as compared to distinct cell type. Combined with this, a third layer of quality control is featured while performing differential gene/transcript expression with the visualization of the principle component analysis (PCA). Such analysis could reveal rapidly the transcriptional differences between distinct cell types by monitoring the increased PC variance throughout sequencing. Ideally, similar samples (e.g. technical/biological replicates) would cluster together whereas distinct samples (e.g. distinct conditions/cell-types) will cluster separately. Similarly, such analyses could also reveal inter-sample variability between similar biological replicates, providing information about their transcriptional states (similarity or dissimilarities) and thus the reliability of the results. Lastly, NanopoReaTA could analyze foreign expressed genes using modified genome annotations that had incorporated gene sequences which are not naturally present in the species’ genome. This was demonstrated in the yeast strains experimental setup with the detection and quantification of foreign genes such as *KanR*, *AmpR* and *HygR*, providing confirmation of the incorporated mutation or transformation efficiency of the foreign vectors. Such a utility could have a major value when a gene of interest is replaced with a foreign gene (e.g., an antibiotic resistance gene) or when introducing foreign vectors harboring specific selection genes. In practice, NanopoReaTA could also be used to detect fusion-protein transcripts as well as monitor transcription efficiency from specific promoters by quantifying the expressed transcripts. On top of these multi-layered quality control detection capabilities, NanopoReaTA performs long-read RNA-seq data analyses in parallel to ongoing sequencing, providing valuable preliminary results of the experimental setup. In case all the QC criteria are fulfilled, the sequencing can be maintained until reaching the desired sequencing depth.

NanopoReaTA's usefulness in academic settings extends to reducing sequencing costs and enhancing sample quality checks prior to sequencing. However, the potential impact of real-time analysis tools in clinical settings is possibly even more far-reaching. For instance, Gorzynski et al. (2022) have introduced an efficient framework for whole genome sequencing, setting a world record in the sequencing and analysis of whole genomes. Not only is this approach technically impressive, but it also enables rapid genetic diagnosis, ultimately improving clinical diagnoses and reducing associated costs (Gorzynski et al. 2022). Additional possibilities may include employing rapid transcriptomic analyses to identify pathogen-specific transcripts or detection of disease-associated transcripts or transcripts isoforms (e.g. detection of aberrant BRCA transcripts). The integration of real-time analysis tools like NanopoReaTA could revolutionize clinical applications as a diagnostic tool, especially when considering transcriptomic data. In conclusion, NanopoReaTA stands out as a valuable tool applicable in both academic and clinical settings, offering cost-effective quality checks for specific experimental conditions while simultaneously providing valuable data through the execution of long-read RNA-seq.

Methods

Cell culture

For HEK293 and HeLa transcriptional comparison, cells were cultured and maintained in an incubator at 37°C and 5% CO_2. HEK293 and HeLa cells were cultured in Dulbecco’s modified Eagle medium (DMEM) supplemented with 10% FBS, 1% penicillin-streptomycin, and 1% L-glutamine. Once the cells

were confluent, the medium was removed and cells were washed once with 1 mL DPBS. The cells were resuspended with 0.5 mL Trizol and collected for Trizol RNA isolation.

Yeast strain growth

S. cerevisiae knockout strains derived from BY4741 were prepared by homologous recombination using standard procedures. Genotypes and culture media used for the respective strains are given in Table 1. For preparation of *S. cerevisiae* RNA, 3 ml of the respective media (Table 1) were inoculated with a single colony of the respective strain and grown overnight in an orbital shaker (30°C, 220 rpm). 25 ml of the same media were then inoculated with the respective overnight culture to an OD_{600} of 0.2 and cultured at 30°C, 220 rpm until an OD_{600} of 0.8 – 1.0 (log-phase). Cells were harvested by centrifugation at 4 °C, pellets were washed twice with Milli Q water, resuspended in Trizol and snap-frozen in liquid nitrogen.

Table 1: List of yeast strains used in this study.

Strain name	written as	Genotype	Cultured in	notes
BY4741		*MATa, his3Δ1, leu2Δ0, met15Δ0, ura3Δ0*	Not cultured	Parental strain
WT pEV(*HIS3*)	WTpEV	BY4741, pEV(*HIS3*)	HIS(-) media	From this study
WT-pNew1(*HIS3*)	WTpN	BY4741, pNew1-Flag(*HIS3*)	HIS(-) media	From this study
new1Δ-pEV(*HIS3*)	dNpEV	BY471, *new1Δ::KanMX*, pEV(*HIS3*)	HIS(-) media	From this study
new1Δ-pNew1(*HIS3*)	dNpN	BY4741, *new1Δ::KanMX*, pNew1-Flag(*HIS3*)	HIS(-) media	From this study
rkr1Δ-pEV(*URA3*)	dRpEV	BY4741, *rkr1Δ::HphMX*, pEV(*URA3*)	URA(-) media	From this study
rkr1Δ-pJlp2(*URA3*)	dRpJ	BY4741, *rkr1Δ::HphMX*, pJLP2-HA(*URA3*)	URA(-) media	From this study
rkr1Δ, *jlp2Δ*-pEV(*URA3*)	dRJpEV	*jlp2Δ::KanMX* ,*rkr1Δ::HphMX*, pEV(*URA3*)	URA(-) media	From this study
rkr1Δ, *jlp2Δ*-pJlp2(*URA3*)	dRJpJ	*jlp2Δ::KanMX, rkr1Δ::HphMX*, pJLP2-HA(*URA3*)	URA(-) media	From this study

Plasmids and media used:
pEV(*HIS3*) – Empty vector with *HIS3* selection marker.
pNew1(*HIS3*) – Overexpression vector for C-terminally FLAG-tagged New1 with *HIS3* selection marker.
pEV(*URA3*) - Empty vector with *URA3* selection marker.
pJlp2(*URA3*) - Overexpression vector for C-terminally HA-tagged Jlp2 with *URA3* selection marker.
HIS(-) media: 20 g/L glucose, 6.9 g/L Yeast Nitrogen Base without amino acids (Formedium), 1.4 g/l yeast synthetic complete drop-out medium supplements (Formedium), 76 mg/L of each: L-Tryptophan (Roth), L-Leucine (Roth), and Uracil (Formedium).
URA(-) media: 20 g/L glucose, 6.9 g/L Yeast Nitrogen Base without amino acids, 770 mg/l CSM, Single Drop-Out -Ura (Formedium).

RNA isolation
HEK 293 and HeLa. For RNA isolation, following 5 min incubation in RT, 100 µL of chloroform was added. Samples were vortexed and incubated 2 minutes at RT. Sample were centrifuged at 13000xg, 4 °C for 10 min and the upper aqueous phase was transferred into a new tube. Next, 250 µL of isopropanol was added and incubated for 15 min at RT for RNA precipitation. The samples were then centrifuged at 13000 – 15000xg at 4 °C for 30 min and supernatant was discarded. The RNA pellet was washed with cold 75% cold EtOH (stored at -20 °C) and centrifuged again at 13000 – 15000xg at 4 °C for 30 min. Supernatant was discarded and the pellet was air dried. The RNA pellet was resuspended with nuclease-free water and concentration was measured using nanodrop.

Yeast. Cells in Trizol were thawed on ice and disrupted by bead-beating with zirconia/glass beads (0.5 mm) and vortexing 10 times in cycles of 30 s vortexing at 3,000 rpm, intermittent with at least 30 s chilling on ice. Following this, 150 µL of Chloroform/Isoamyl alcohol (24:1 V/V) were added per 750 µl of Trizol, and vortexed. After centrifugation (10 min, 14,000 rpm, 4°C) and an optional second extraction

of the aqueous phase with Chloroform/Isoamyl alcohol (24:1 V/V) and water-saturated phenol (pH 4,5-5), the aqueous phase was mixed with sodium acetate pH 5.2 (to at least 0.15 M) and RNA was precipitated by addition of 2-propanol and centrifugation (20-30 min, 14,000 rpm, 4°C). The pellet was washed twice with ice-cold 75% ethanol, briefly dried and dissolved in Milli Q water. RNA concentrations were measured by Nanodrop (A_{260}).

Selective purification ribosomal-depleted (RiboMinus) and ribosomal-enriched (RiboPlus) transcripts

Selective purification of distinct RNA populations was performed using the Ribominus™ Eukaryote kit for RNA-seq (#Ambion, A10837-08) according to the manufacturer's "standard protocol" instructions, with slight modifications for specific rRNA isolation. For the procedure, 5 µg of RNA in 5 µl nuclease-free water was subjected to hybridization with 100 µl of Hybridization Buffer and 10 µl of RiboMinus™ Probe (15 pmol/µL) at 70–75ºC for 5 minutes, followed by an additional 30-minute incubation at 37ºC. The RiboMinus™ Magnetic beads were prepared according to the manufacturer's instructions. The RNA/probe mixture was then combined with RiboMinus™ Magnetic beads and incubated at 37ºC for 15 minutes. Subsequently, magnetic separation was employed to pellet the rRNA-probe complex, and the supernatant, containing ribo-depleted RNA, was collected. The remaining beads underwent a similar process for rRNA isolation using nuclease-free water, and the resulting RiboMinus™ RNA was added to the previous supernatant. To isolate the RNA form the RiboMinus™ supernatant, the sample underwent ethanol precipitation according to the manufacturer's instructions. The pooled bead samples (containing the rRNA) were further processed with Trizol RNA isolation to complete the purification. 1 µg of Total RNA from HEK293 and RiboPlus and 150ng of Ribominus were assessed on 1% TBE agarose gel stained with ethidium bromide.

Direct cDNA-native barcoding Nanopore library preparation and sequencing.

Double-stranded cDNA synthesis was carried out using the Maxima H Minus Double-Stranded cDNA Synthesis Kit (Thermo Scientific, K2561) following the manufacturer's protocol. Initially, 2-3 µg of RNA was combined with 1 µl of oligo(dT)18 (100 pmol) and 1 µl of 10 mM dNTPs, reaching a final volume of 11 µl with RNase-free water. After incubating at 65°C for 5 minutes and snap-cooling on ice, a master mix consisting of 4 µl 5x RT Buffer, 1 µl RNaseOUT, 3 µl Nuclease-free water, and 1 µl Maxima H Minus Reverse Transcriptase per sample was prepared. Incubation for 30 min at 50°C followed, and the reaction was terminated by heating at 85°C for 5 minutes. For the second strand synthesis, a master mix with 17.5 µl nuclease-free water, 10 µl of 5X second strand reaction buffer, and 2.5 µl of second strand enzyme mix per sample was supplemented to the 20 µl first strand cDNA synthesis reaction. Samples were incubated at 16°C for 60 min. Subsequently, 10 µL (100 u) of RNase I was added, and purification using AMPure XP beads-based (Agencourt, A63881) method was performed with a bead-to-sample ratio of 0.8X, eluting in 21 µl of nuclease-free water. Concentrations of second-strand cDNA samples were determined using Qubit Fluorometric Quantitation (1 µl). Following this, end-prepping was conducted with NEBNext Ultra II End Repair / dA-tailing Module (NEB, cat # E7546). A mixture of 20 µl dscDNA sample, 22 µl nuclease-free water, 5.5 µl Ultra II End-prep reaction buffer, and 2.5 µl Ultra II End-prep enzyme mix was incubated at 20°C for 15 minutes and 65°C for 10 minutes. Cleanup with 1× AMPure XP Beads was performed, and elution was carried out in 10 µl nuclease-free water. Barcoding was achieved using Native Barcoding Expansion 1-12 (EXP-NBD104, ONT) by supplementing each sample with 2.5 µl Native Barcode and 10 µl Blunt/TA Ligase Master Mix, reaching a final volume of 22.5 µl. After incubation at RT for 20 min, 2 µl of EDTA was added to each sample to stop the reaction. Barcoded samples were pooled, purified using 0.7X AMPure XP Beads, and eluted in 31 µl nuclease-free water. Concentration determination (1 µl) and adapter ligation using 5 µL NA, 10 µL NEBNext Quick Ligation Reaction Buffer (5X), and 5 µL Quick T4 DNA Ligase (NEB, cat # E6056) were performed. Pooled library purification with 0.7X AMPure XP Beads resulted in a final elution volume of 33 µl EB. Concentration of the pooled barcoded library was determined using Qubit (1 µl). Finally, the library was mixed with sequencing buffer and loading beads before loading onto a primed R10.4.1 PromethION flow cell or R9.4.1 MinION flow cell.

Nanopore-seq and NanopoReaTA data collection.
Reads were basecalled using Guppy basecaller version 3.6.1 in high-accuracy (hac) mode for PromethION sequencing and super-accuracy for MinION sequencing. For a detailed overview of NanopoReaTA's requirements, pipelines, and additional tools, please refer to Wierczeiko et al. 2023 or visit https://github.com/AnWiercze/NanopoReaTA. In this study, upon sequencing initiation, NanopoReaTA was activated following the guidelines outlined in the "Step-by-Step Use of NanopoReaTA." Two PromethION and one MinION flow cells were employed for this investigation. The cell culture samples, comprising a total of 8 barcodes (barcodes 1-8), were loaded onto the first flow cell, and data were continuously collected over a 24-hour period. For the HEK293 and HeLa experimental setup, the samples were loaded onto a MinION flow cell, and data collection took place over a 72-hour sequencing period. The yeast samples were loaded onto a separate PromethION flow cell. Yeast setup 1 (barcodes 1-12) were initially loaded, and data collection extended for 24 hours. Following this period, sequencing was halted, and the PromethION flow cell was washed using the Flow Cell Wash Kit (EXP-WSH003, ONT). Subsequently, Yeast setup 2 (barcodes 1-3, 13-24) was loaded, and data were collected over another 24-hour period. For all experimental setups (cell culture and yeast), data points were collected at 1hr, 2hr, 5hr, 10hr, and 24hr post-sequencing initiation (PSI). DEG overlap between the distinct time points, database or sequencing devices was performed using Venn diagram web-tool (https://bioinformatics.psb.ugent.be/webtools/Venn/). The collected data included general overview metrics, including the number of detected genes, gene variability, individual and combined read length distribution, as well as the usage timings of tools applied by NanopoReaTA. Additionally, detailed information on differential gene and transcript expressions, including PCA, volcano plots, sample-to-sample distance plots, and heatmaps, were organized in the Supplementary Figures.

Data access

All raw and processed sequencing data generated in this study have been submitted to the NCBI BioProject database (https://www.ncbi.nlm.nih.gov/bioproject/) under accession number PRJNA1090486.

Competing interest statement

The authors declare no conflict of interest.

Acknowledgments

This work was funded by Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) [project number 439669440 TRR319 RMaP TP B05 to M.L.W., TP A05/C01/C03 to M.H.; Project number 255344185 SPP1784, Startup Funding to M.L.W.]. T.B and S.G. acknowledge funding from the Emergent AI Center funded by the Carl-Zeiss-Stiftung. M.L.W and S.G. acknowledge funding from the Forschungsinitiative Rheinland-Pfalz and the ReALity initiative of the Johannes Gutenberg University Mainz. S.G. acknowledges funding by SFB 1551 Project No. 464588647 of the Deutsche Forschungsgemeinschaft (DFG).

Author contributions. T.B. and S.G. conceived and supervised the project. T.B. and S.P. designed the experimental strategy and performed the data collection using NanopoReaTA. S.P and A.W. implemented GUI updates in NanopoReaTA. M.M. and K.V.I. generated the yeast mutant strains, performed the overexpression experiments, and performed the RNA isolation from yeast samples. S.M. performed the RNA isolation from HEK293 and HeLa. T.B performed the selective purification of ribosomal-depleted and ribosomal-enriched transcripts from HEK293 and prepared the direct cDNA libraries. T.B. and S.P. wrote the manuscript. S.G., M.H., A.W., K.F. and M.W. edited the manuscript and provided valuable input and feedback in various discussions. All authors read and approved the final manuscript.

References

Wang Z, Gerstein M, Snyder M. 2009. RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics*, *10*1, 57–63. https://doi.org/10.1038/nrg2484

Casamassimi A, Federico A, Rienzo M, Esposito S, Ciccodicola A. 2017. Transcriptome Profiling in Human Diseases: New Advances and Perspectives. *International Journal of Molecular Sciences*, 188, 1652. https://doi.org/10.3390/ijms18081652

Mutz K, Heilkenbrinker A, Lönne M, Walter J, Stahl F. 2013. Transcriptome analysis using next-generation sequencing. *Current Opinion in Biotechnology*, *24*1, 22–30. https://doi.org/10.1016/j.copbio.2012.09.004

Satam H, Joshi K, Mangrolia U, Waghoo S, Zaidi G, Rawool S, Thakare RP, Banday S, Mishra A, Das G, Malonia SK 2023. Next-Generation Sequencing Technology: Current trends and advancements. *Biology*, *12*7, 997. https://doi.org/10.3390/biology12070997

Butto T, Mungikar K, Baumann P, Winter J, Lutz B, Gerber S. (2023). Nuclei on the rise: When Nuclei-Based Methods meet Next-Generation Sequencing. *Cells*, *12*(7), 1051. https://doi.org/10.3390/cells12071051

Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, *34*18, 3094–3100. https://doi.org/10.1093/bioinformatics/bty191

Kim D, Paggi JM, Park C, Bennett C, Salzberg SL 2019. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology*, *37*8, 907–915. https://doi.org/10.1038/s41587-019-0201-4

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras, T. 2012. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, *29*1, 15–21. https://doi.org/10.1093/bioinformatics/bts635

Liao Y, Smyth GK, Shi W. 2013. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, *30*7, 923–930. https://doi.org/10.1093/bioinformatics/btt656

Anders S, Pyl PT, Huber W. 2014. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics*, *31*2, 166–169. https://doi.org/10.1093/bioinformatics/btu638

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*12. https://doi.org/10.1186/s13059-014-0550-8

Robinson MD, McCarthy DJ, Smyth GK. 2009. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, *26*1, 139–140. https://doi.org/10.1093/bioinformatics/btp616

Conesa A, Madrigal P, Tarazona S, Gomez-Cabrero D, Cervera A, McPherson A, Szcześniak MW, Gaffney DJ, Elo LL, Zhang X, Mortazavi A. 2016. A survey of best practices for RNA-seq data analysis. *Genome Biology*, *17*(1). https://doi.org/10.1186/s13059-016-0881-8

Ji F, Sadreyev RI. 2018. RNA-SEq: Basic Bioinformatics Analysis. *Current Protocols in Molecular Biology*, *124*(1). https://doi.org/10.1002/cpmb.68

Corchete LA, Rojas EA, Alonso-López D, De Las Rivas J, Gutiérrez NC, Burguillo F. 2020. Systematic comparison and assessment of RNA-seq procedures for gene expression quantitative analysis. *Scientific Reports*, *10*(1). https://doi.org/10.1038/s41598-020-76881-x

Zhang C, Zhang, B, Lin L, Zhao S. 2017. Evaluation and comparison of computational tools for RNA-seq isoform quantification. *BMC Genomics*, *18*(1). https://doi.org/10.1186/s12864-017-4002-1

Hardwick SA, Joglekar A, Flicek P, Frankish A, Tilgner H. 2019. Getting the entire message: progress in isoform sequencing. *Frontiers in Genetics*, *10*. https://doi.org/10.3389/fgene.2019.00709

Ozsolak F, Milos PM. 2010. RNA sequencing: advances, challenges and opportunities. *Nature Reviews Genetics*, *12*(2), 87–98. https://doi.org/10.1038/nrg2934

Lu H, Giordano F, Ning Z. 2016. Oxford Nanopore MinION Sequencing and Genome Assembly. *Genomics, Proteomics Bioinformatics*, *14*(5), 265–279. https://doi.org/10.1016/j.gpb.2016.05.004

Wang Y, Zhao Y, Bollas A, Wang Y, Au KF. 2021. Nanopore sequencing technology, bioinformatics and applications. *Nature Biotechnology*, *39*(11), 1348–1365. https://doi.org/10.1038/s41587-021-01108-x

Zheng P, Zhou C, Ding Y, Liu B, Lu L, Zhu F, Duan S. 2023. Nanopore sequencing technology and its applications. *MedComm*, *4*(4). https://doi.org/10.1002/mco2.316

Kono N, Arakawa K. 2019. Nanopore sequencing: Review of potential applications in functional genomics. *Development, Growth Differentiation*, *61*(5), 316–326. https://doi.org/10.1111/dgd.12608

Amarasinghe SL, Su S, Dong X, Zappia L, Ritchie ME, Gouil Q. 2020. Opportunities and challenges in long-read sequencing data analysis. *Genome Biology*, *21*(1). https://doi.org/10.1186/s13059-020-1935-5

Munro R, Santos R, Payne A, Forey T, Osei S, Holmes N, Loose M. 2022. minoTour, real-time monitoring and analysis for nanopore sequencers. *Bioinformatics*, *38*(4), 1133–1135. https://doi.org/10.1093/bioinformatics/btab780

Weilguny L, De Maio N, Munro R, Manser C, Birney E, Loose M, Goldman N. 2023. Dynamic, adaptive sampling during nanopore sequencing using Bayesian experimental design. *Nature Biotechnology*, *41*(7), 1018–1025. https://doi.org/10.1038/s41587-022-01580-z

Fırtına C, Ghiasi NM, Lindegger J, Singh G, Cavlak MB, Mao H, Mutlu O. 2023. RawHash: enabling fast and accurate real-time analysis of raw nanopore signals for large genomes. *Bioinformatics*, *39*(Supplement_1), i297–i307. https://doi.org/10.1093/bioinformatics/btad272

Bruno A, Aury J, Engelen S. 2021. BoardION: real-time monitoring of Oxford Nanopore sequencing instruments. *BMC Bioinformatics*, *22*(1). https://doi.org/10.1186/s12859-021-04161-0

Sanderson, ND, Street, T, Foster, D, Swann, J, Atkins BL, Brent A, McNally M, Oakley S, Taylor A, Peto TEA, *et al.* 2018. Real-time analysis of nanopore-based metagenomic sequencing from infected orthopaedic devices. *BMC Genomics*, *19*(1). https://doi.org/10.1186/s12864-018-5094-y

Rodríguez-Pérez H, Ciuffreda L, Flores C. 2022. NanoRTax, a real-time pipeline for taxonomic and diversity analysis of nanopore 16S rRNA amplicon sequencing data. *Computational and Structural Biotechnology Journal*, *20*, 5350–5354. https://doi.org/10.1016/j.csbj.2022.09.024

Wierczeiko A, Pastore S, Mündnich S, Busch AM, Dietrich V, Helm M, Butto T, Gerber S. 2023. NanopoReaTA: a user-friendly tool for nanopore-seq real-time transcriptional analysis. *Bioinformatics*, *39*(8). https://doi.org/10.1093/bioinformatics/btad492

Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. 2017. Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods*, *14*(4), 417–419. https://doi.org/10.1038/nmeth.4197

Anders S, Reyes A, Huber W. 2012. Detecting differential usage of exons from RNA-seq data. *Genome Research*, *22*(10), 2008–2017. https://doi.org/10.1101/gr.133744.111

Nowicka M, Robinson MD. 2016. DRIMSeq: a Dirichlet-multinomial framework for multivariate count outcomes in genomics. *F1000Research*, *5*, 1356. https://doi.org/10.12688/f1000research.8900.2

Evans C, Hardin J, Stoebel DM. 2017. Selecting between-sample RNA-Seq normalization methods from the perspective of their assumptions. *Briefings in Bioinformatics*, *19*(5), 776–792. https://doi.org/10.1093/bib/bbx008

Thawng CN, Smith GB. 2023. Transcriptome software results show significant variation among different commercial pipelines. *BMC Genomics*, *24*(1). https://doi.org/10.1186/s12864-023-09683-w

Kasari V, Pochopien AA, Margus T, Murina V, Turnbull KJ, Zhou Y, Nissan T, Graf M, Nováček J, Atkinson GC, *et al.* 2019. A role for the Saccharomyces cerevisiae ABCF protein New1 in translation termination/recycling. *Nucleic Acids Research*, *47*(16), 8807–8820. https://doi.org/10.1093/nar/gkz600

Teste M, Duquenne M, François J, Parrou J. 2009. Validation of reference genes for quantitative expression analysis by real-time RT-PCR in Saccharomyces cerevisiae. *BMC Molecular Biology*, *10*(1). https://doi.org/10.1186/1471-2199-10-99

Gorzynski JE, Goenka SD, Shafin K, Jensen TD, Fisk DG, Grove ME, Spiteri E, Pesout T, Monlong J, Baid G, *et al.* 2022. Ultrarapid nanopore genome sequencing in a critical care setting. *The New England Journal of Medicine*, *386*(7), 700–702. https://doi.org/10.1056/nejmc2112090