Real-time transcriptomic profiling in distinct experimental conditions

Tamer Butto[1,#,§], Stefan Pastore[1,2,#] , Max Müller[1], Kaushik Viswanathan Iyer[1], Marko Jörg[1], Julia Brechtel[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 four experimental setups: 1) transcriptome profiling of distinct human cellular populations, 2) identification of experimentally enriched transcripts, 3) transcriptional analysis of cells under heat shock conditions and 4) 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 1hr 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.

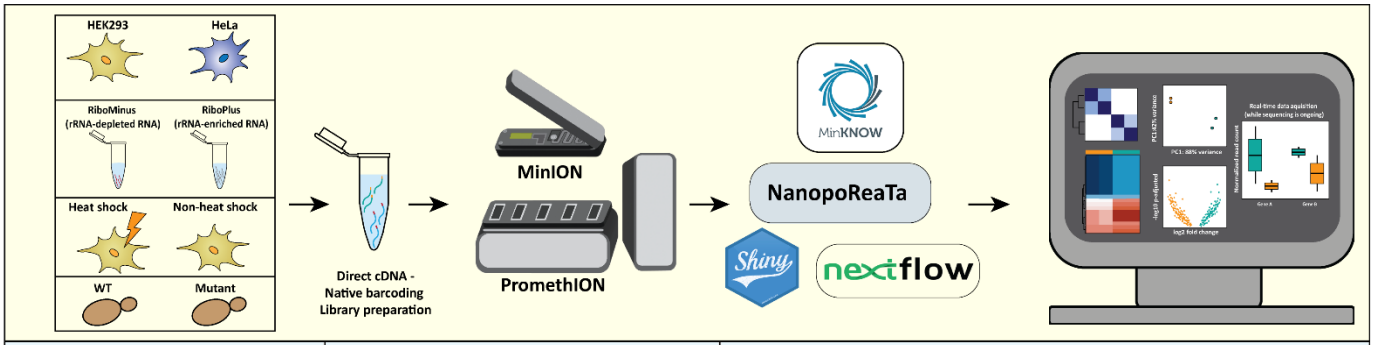
Recently, we presented NanopoReaTA, the first real-time analysis toolbox for comparative transcriptional analyses of Nanopore-seq data (Wierczeiko and Pastore et al. 2023). NanopoReaTA provides an interactive Graphic user interphase (GUI) that allows users to perform transcriptional analyses of cDNA 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 four experimental setups that include: 1) transcriptome profiling of distinct human cellular populations 2) identification of experimentally-enriched transcripts, 3) transcriptional profiling of cells exposed to heat shock and 4) 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. 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 distribution (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 transcript 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 hr) following library preparation, which included the synthesis of dscDNA library for the desired RNA sample (~2 hrs). Samples were barcoded, and adapter ligated (~1 hr). 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. To simulate distinct experimental designs, we structured three different setups. In the first, we employed 10 biological replicates per cell-type providing enhanced reliability and precision of the statistical analysis, reducing

the impact of variability and allowing for more accurate identification of significant differences. The second setup simulated a scenario with 2 replicates per condition, focusing on testing the effect of limited replication while generating high-throughput data using PromethION. The final setup compared the performance of 2 replicates per condition using a MinION, simulating an early-stage evaluation of the experimental approach. Following barcoding, the samples were loaded into a PromethION or MinION flow cells and sequenced for 24-hrs. We tracked the basic sequencing metrics using ONT's MinKnow software and activated NanopoReaTA as soon as the sequencing initiated. All supplementary figures related to this experimental setup are provided in Supplementary Document S1.

One-hour post-sequencing initiation (PSI), we gathered basic sequencing metrics from the MinKNOW software, 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 (**Supplementary Figures 2-4**). We monitored the amount of cDNA generated and loaded for sequencing, observing relatively consistent throughput between HEK293 and HeLa (**Supplementary Figures 2-4**). Detailed descriptions of the experimental setup are provided in Supplementary Material 2.

Next, we performed real-time DGE and DTE analyses to monitor the transcriptional changes between HEK293 and HeLa, 1-hr PSI. As an initial quality control, we inspected the Sample-to-sample similarity plot and principal component analysis (PCA). For the 10-replicate PromethION experimental setup, we observed a separation between the two conditions, where PC1 represents 28% of the variance while PC2 represents 5% of the variance, already 1-hr PSI (**Figure 2B-C, Supplementary Figure 5A-B**). Similar observations were noted in the 2-replicate PromethION experimental setup with distinct separation between the two conditions, where PC1 represents 68% of the variance while PC2 represents 17% of the variance (**Supplementary Figure 6A-B**). 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 2D, Supplementary Table 1**). Notably, in the 10 replicate per condition setup, we identified 23 annotated genes enriched in HEK293 and 35 genes enriched in HeLa, 1-hr PSI (**Figure 2D, Supplementary Figure 5C**). Early comparison with 2 replicate per condition setup, revealed genes such as *ACTB*, *FTL* and *S100A6* enriched in HeLa whereas mitochondrial rRNA gene such as *MT-RNR2* and *MT-RNR1* were enriched in HEK293 (**Supplementary Figure 5C-D, 6C-D**). In addition, NanopoReaTA offers an interactive utility enabling users to input specific genes and visualize both raw and normalized read counts. Using the "Gene-wise" utility, we introduced several of the DEGs as input genes and visualized their raw and normalized gene counts (**Supplementary Figures 5E and 6E**). 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 (**Supplementary Figure 8A-D, 9A-D**). Therefore, when examining the differentially expressed transcripts (DETs) in the 10 replicate setup at 1-hr PSI, we detected 24 transcripts enriched in HEK293 and 36 transcripts enriched in HeLa (**Supplementary Figure 8C-D**). In the 2 replicate per condition setup, we detected 2 transcripts enriched in HEK293 and 6 transcripts enriched in HeLa (**Supplementary Figure 9C-D**). Subsequently, as data was collected at 2 hrs, 5 hrs, 10 hrs, and 24 hrs 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 2-4**), as well as DEGs and DETs (**Supplementary Figures 5C-D, 6C-D,7A-H,8C-D,9C-D**), was observed with the advancement of sequencing. This analysis

demonstrates the ability to capture real-time changes in gene expression between distinct conditions, thus providing a valuable quality control measure.

At 24-hr PSI, we examined the sample-to-sample distance plots and PCA plots and noted enhanced separation between the conditions, particularly evident in PC2, which accounted for 88% of the variance (**Figures 2E and 2F, Supplementary Figure 5A**). Similar trend was detected 2 replicate per condition setup (**Supplementary Figure 6A**). This clustering trend persisted consistently throughout the entire sequencing process, evident in both DGE and DTE analyses (**Supplementary Figure 5A and 6A**). We identified 81 genes enriched in HEK293 and 111 genes enriched in HeLa (**Figure 2G, Supplementary Table 1**).

To assess the consistency 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. In the 10-replicate setup, we identified 56 annotated genes that were consistently detected across all time points, with an increased number of DEGs detected from 1-hr to 24-hrs 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.

Lastly, we overlapped differentially expressed genes (DEGs) across various experimental setups to assess the reproducibility of transcriptomic profiling between HEK293 and HeLa cells. Here, we identified 16 DEGs that were consistently detected in HEK293 and 51 in HeLa across setups (**Supplementary Figure 10A-B**). Notably, MinION sequencing showed fewer unique DEGs compared to PromethION due to its lower throughput but demonstrated consistent enrichment patterns within each condition. To validate that the DEGs 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 DEGs that overlapped across at least two experimental setups (47 upregulated in HEK293 and 114 upregulated in HeLa) and cross-referenced them with Harmonizome expression profiles. This analysis confirmed consistent enrichment patterns across our experimental setups (**Supplementary Figure 10C-D**), identifying 19 DEGs shared in HEK293 and 46 in HeLa (**Supplementary Figure 10D**). Notably, genes such as *RPS19* and *RPL18* were specifically enriched in HEK293 compared to HeLa, while *ACTB*, *CLU*, and *ID3* were enriched in HeLa compared to HEK293 (**Supplementary Figure 10C-D**). These findings highlight the robustness of NanopoReaTA in detecting transcriptional differences across platforms and conditions, demonstrating its utility for real-time differential gene expression analysis.

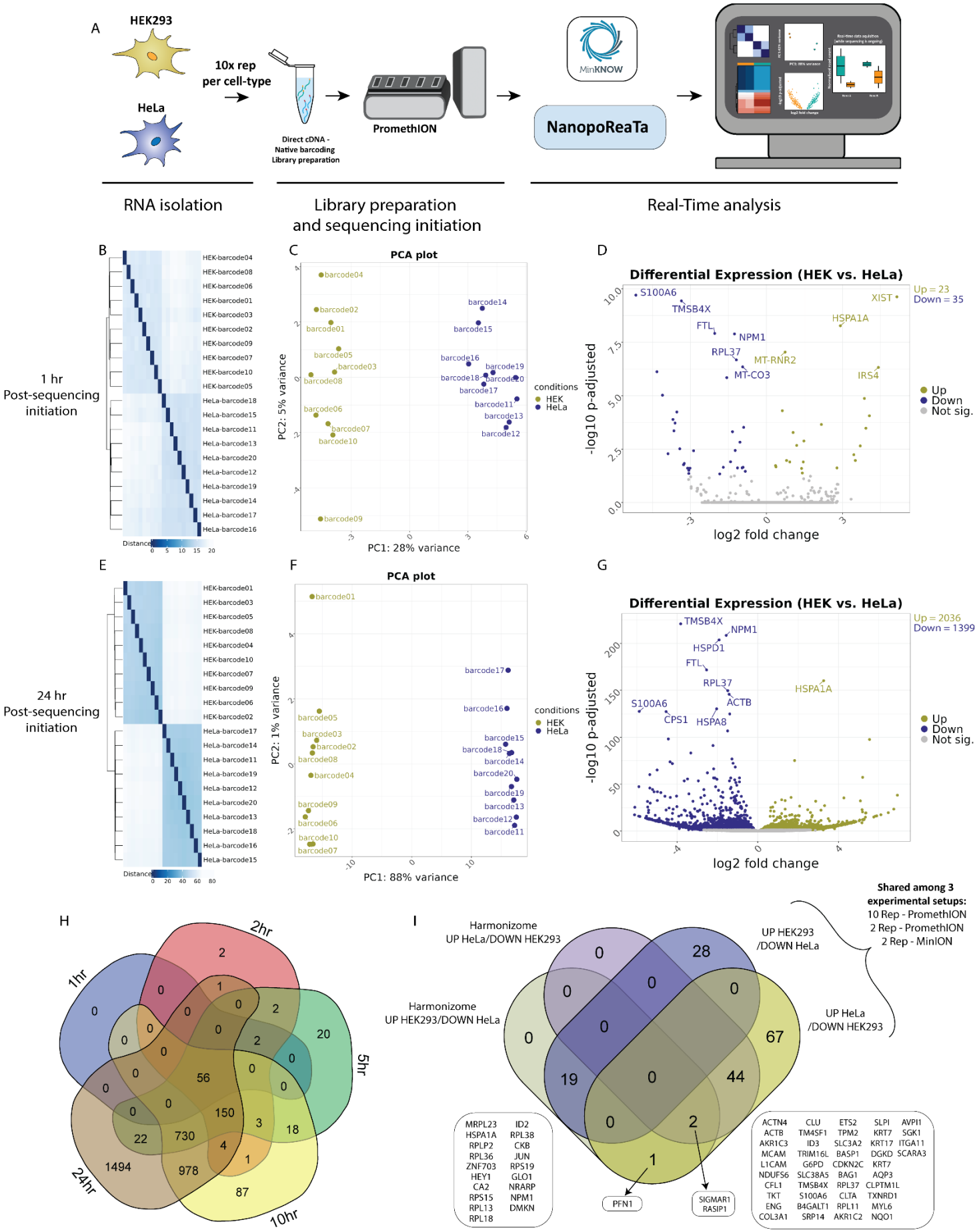


Figure 2. Real-time transcriptomic analysis between HEK293 and HeLa using NanopoReaTA. A. Experimental strategy. RNA was isolated from HEK293 and HeLa cells, the dscDNA library was prepared which included sample barcoding and adapter ligation. Samples were loaded and sequenced using a PromethION R10 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-hr post-sequencing initiation between HEK293 and HeLa samples. Selected data plots showing sample-to-sample distance plot **(B)**, Principal component analysis (PCA) **(C)** and volcano plot **(D)** 1hr-

post sequencing initiation. **E-G.** Differential gene expression 24-hrs 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-hrs 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-hrs time point), 19 genes were found to be enriched in HEK293 and depleted in HeLa, while 46 genes showed enrichment in HeLa and depletion in HEK293. One gene (PFN1) was enriched in HeLa and enriched in HEK293 Harmonizome database.

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 2A, Supplementary Figure 11A**). 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 11B**). 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 hrs, with data collected at the same intervals as mentioned earlier (**Figure 1**). All supplementary figures related to this experimental setup are provided in Supplementary Document S2.

As for the previous experiment, 1-hr 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 11C-F, Supplementary Table 2**). Interestingly, despite generating more total reads and identifying more mapped reads for TotalR and RiboP, we observed a higher number of identified gene and transcript counts in RiboM compared to the other two conditions (**Supplementary Figure 11D**). 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 12, Supplementary Figure 15, Supplementary Figure 18**).

Next, we inspected the PCA and dissimilarity plots **(Figure 3B, Supplementary Figures 13A-B, 14A-B, 16A-B, 17A-B, 19A-B, 20A-B).** We noticed clear distinctions between all three condition comparisons, evident as early as 1-hr 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-hr PSI, we identified 2 annotated genes enriched and 3 depleted in RiboM compared to TotalR (**Figure 3C**, **Supplementary Table 2**), 5 enriched in RiboP compared to TotalR (**Figure 3J**) and 7 enriched and 1 depleted in RiboP compared to RiboM (**Figure 3Q, Supplementary Figures 13C-D, 16C-D, 19C-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 13E, 16E, 19E**). When examining the DETs, we identified transcripts associated with the respective rRNA enrichment groups (**Supplementary Figures 14A-E,17A-E,20A-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-hrs 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-hr 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 12C-D, Supplementary Figure 15C-D, Supplementary Figure 18C-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.

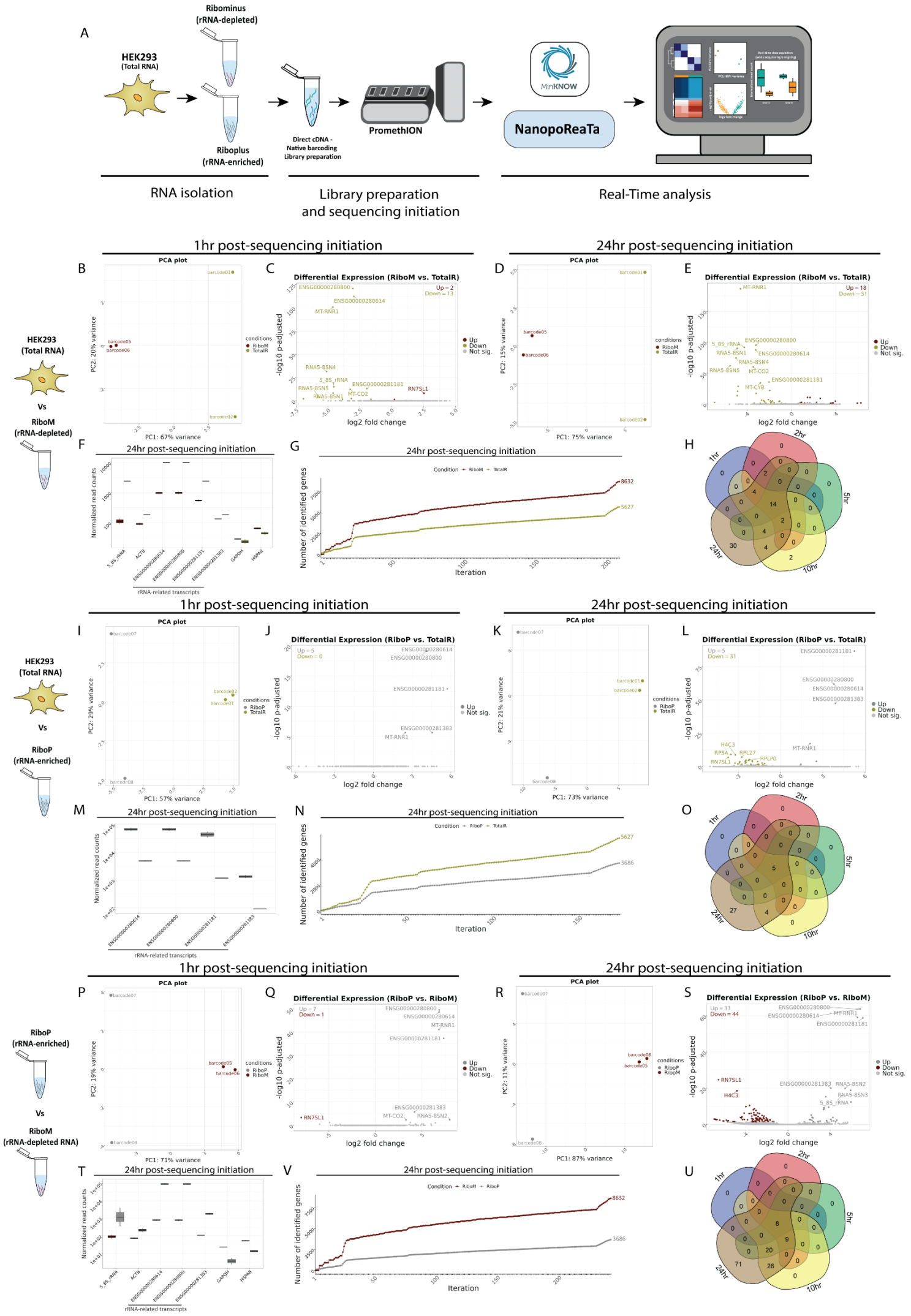

A
HEK293
(Total RNA)
Ribominus
(rRNA-depleted)
Riboplus
(rRNA-enriched)
Direct cDNA -
Native barcoding
Library preparation
PromethION
MinKNOW
NanopoReaTa
RNA isolation
Library preparation
and sequencing initiation
Real-Time analysis
1hr post-sequencing initiation
24hr post-sequencing initiation
HEK293
(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
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.** Number of identified genes with >0 aligned reads 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**).

Consequently, this resulted in a higher number of detected genes compared to TotalR as well while RiboP exhibited the fewest detected genes, as anticipated.

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 RiboM 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.

Real-time monitoring of transcriptional changes under heat shock stress

To create an experimental setup that mimics biological discovery conditions, we conducted a heat shock experiment using HEK293 cells and analyzed the resulting transcriptional changes with NanopoReaTA. We employed six biological replicates for each condition: heat shock (HS) and non-heat shock control (NHS), using a ribodepletion protocol to enhance the capture of mRNA transcripts linked to heat-shock induction (**Figure 4A, Supplementary Figure 21A**). All supplementary figures related to this experimental setup are provided in Supplementary Document S3.

First, we gathered basic sequencing metrics from the MinKNOW software, including the total reads generated per sample, along with mapped reads, gene counts, and transcript counts (**Supplementary Figure 21B-E**, **Supplementary Table 3**). Basic quality control metrics were also collected, such as the number of detected genes, gene variability, individual and combined read length distributions, and usage timings of the tools applied by NanopoReaTA (**Supplementary Figure 22 and Supplementary Table 3**).

1-hr PSI, we observed a slight separation between the two conditions in the principal component analysis (PCA), with PC1 accounting for 13% of the variance and PC2 accounting for 11% (**Figure 4B, Supplementary Figure 23A-B**). Similar trends were observed at the transcript level (**Supplementary Figure 24A-B**). We identified three annotated genes including *HSPA1A*, *DNAJB1*, and *HSP90AA1* enriched in the HS condition (**Figure 4C-D, Supplementary Figure 23C-D**). We observed similar dynamics in the identified DETs (**Supplementary Figure 24C-D**). These genes were shown to be upregulated under heat shock conditions in previous reports (Neueder et al. 2014, Yonezawa and Bono 2023, Sanchez-Briñas et al. 2023), demonstrating the efficacy of NanopoReaTA in capturing the expected DEGs early during the sequencing process.

At the 24-hr PSI, we noted a more pronounced separation between the conditions, with PC1 representing 36% of the variance and PC2 representing 23% (**Figure 4E, Supplementary Figure 23A**). This clustering trend was consistent throughout the entire sequencing process and was evident in both DGE and DTE analyses (**Supplementary Figure 23A-B, 24A-B**). In total, we identified 43 genes enriched in the HS condition and 7 genes enriched in the NHS condition (**Figure 4F-G, Supplementary Figures 23C-D**).

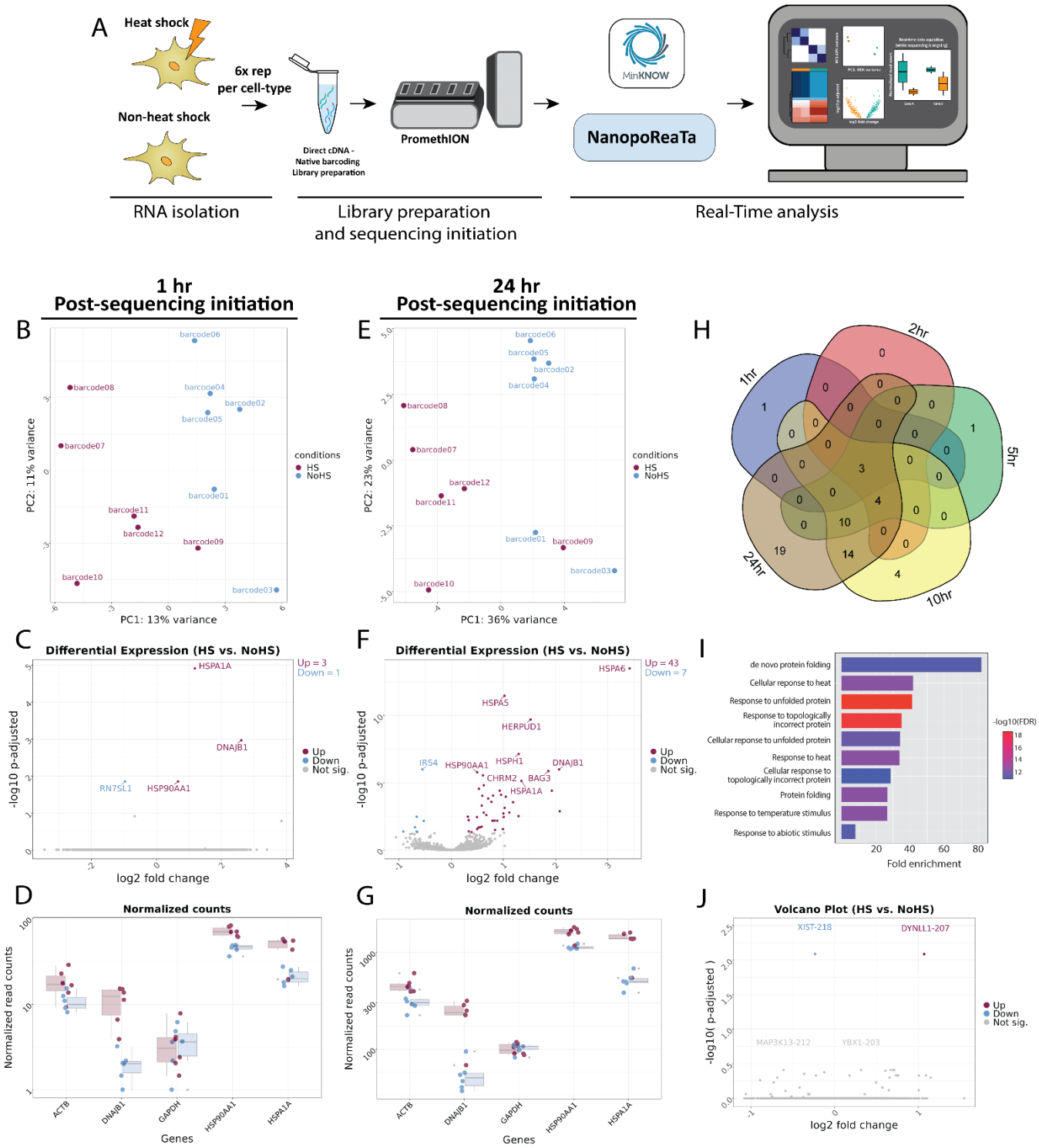


Figure 4. Real-time transcriptomic analysis between heat-shock and non-heat shock treated cells using NanopoReaTA. A. Experimental strategy. RNA was isolated from HEK293 treated with heat shock on non-heat shock treatment (n=6) and selective purification of distinct RNA populations was performed using 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-D** Differential gene expression 1-hr post-sequencing initiation between HEK293 and HeLa samples. Selected data plots showing PCA **(B)**, volcano plot **(C)** and normalized counts for selected genes **(D)** 1hr post sequencing initiation. **E-G.** Differential gene expression 24-hrs post-sequencing initiation between HEK293 and HeLa samples. Selected data plots showing PCA **(E)**, volcano plot **(F)** and normalized counts for selected genes **(G)** 24-hrs post-sequencing initiation. **H.** Five-way Venn diagram showing the differentially expressed gene overlaps between the distinct collected time points. **I.** Gene Ontology (GO) enrichment analysis of upregulated gene in HS compared to NHs conditions. **J**. Volcano plot depicting differential transcript usage between HS and NHS conditions.

We compared the total differentially expressed genes identified at each time point to evaluate the dynamic changes detected during sequencing. Notably, the three heat shock-associated genes,

HSPA1A, *DNAJB1*, and *HSP90AA1*, were consistently detected across the entire sequencing process (Figure 4H).

Next, we performed Gene Ontology (GO) term analysis using ShinyGO (Ge, Jung, and Yao 2020) to examine the biological processes associated with the HS condition. This analysis revealed terms such as “de novo protein folding,” “response to unfolded protein,” and “response to heat” among the enriched DEGs in HS, confirming our expected results (**Figure 4I**, **Supplementary Table 3**). These results demonstrate NanopoReaTA's ability to rapidly identify biologically relevant information even during the early stages of sequencing.

Lastly, we employed DEXSeq and DRIMSeq for differential transcript usage analysis to identify specific transcripts enriched in each condition. We identified DYNLL1-207 isoform as upregulated in HS compared to NHS (**Figure 4J**). *DYNLL1* is a dynein light chain involved in intracellular transport, apoptosis regulation, and cancerogenesis (Liu et al. 2024). Although there is no direct evidence linking *DYNLL1* to the heat shock response, its known roles in stress-related pathways and cellular homeostasis suggest it may have an indirect or previously unexplored involvement in the stress response.

NanopoReaTA's ability to detect differential transcript usage offers a valuable tool for isoform-specific research, especially when paired with appropriate experimental designs.

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 5A, Supplementary Figure 25A**).

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* (each condition was tested in triplicate). 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*)) (**Figure 6A, Supplementary Figure 42A**).

All *S. cerevisiae* knockout strains derived from BY4741 were prepared using homologous recombination following standard procedures and each condition was tested in (biological) triplicates. 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 the following section we will focus on describing the detected changes in differentially expressed genes/transcripts, however general sequencing overviews, as well as detailed analyses for all individual time points are presented in **Supplementary Document S4** (for yeast setup 1, containing **Supplementary Figures 25-41**) and **Supplementary Document S5** (for yeast setup 2, containing **Supplementary Figures 42-55**). For each experimental setup, we collected data from NanopoReaTA, including general sequencing overviews with the count of detected genes (**Supplementary Figures 25,42**), with individual and combined read length distribution (**Supplementary Figures 27, 30, 33, 36, 39** for yeast setup 1; **Supplementary Figures 44, 47, 50, 53** for yeast setup 2), gene expression variability (**Supplementary Figures 26,43**, and the timing of tool utilization. We conducted real-time

analyses for both DGE and DTE, documenting all associated information at each time point (**Supplementary Documents S4,S5, Supplementary Tables 4,5).**

Yeast setup 1

new1Δ-pEV(*HIS3*) vs WT-pEV(*HIS3*). For yeast setup 1, we aimed to test NanopoReaTA's 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*) versus WT-pEV(*HIS3*) at 1hr PSI, the PCA revealed separation between the samples based on their respective conditions **(Figure 5B, 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 5C, 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*) (**Supplementary Figure 4P, Supplementary Figure 28C**). These observations were consistent throughout the entire sequencing period, extending up to the 24hr mark (**Figure 5D-E, Supplementary Figure 28C-F, Supplementary Table 4**). 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.

WT-pNew1(HIS3) vs WT-pEV(HIS3). Next, 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 (**Supplementary Figure 26H, Supplementary Figure 31A**). However, the sample-to-sample distance plot did not uncover significant differences between the replicates (**Supplementary Figure 31B**). The distinct clustering observed in PCA was consistently maintained throughout the entire sequencing process until the 24-hr PSI mark (**Supplementary Figure 26H-I, Supplementary Figure 31A**). 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 (**Supplementary Figure 26J**) and transcript (**Supplementary Figure 32C-D**) identified at the 1-hr 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 (**Supplementary Figure 26K-M, Supplementary Table 4**). 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 yeast Gapdh-encoding genes *TDH1*, *TDH2*, and *TDH3* (**Supplementary Figure 26L, Supplementary Figure 31E**). 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 1h PSI. As in WT, the PCA successfully differentiated the samples according to their respective conditions (**Supplementary Figure 26N, Supplementary Figure 34A**). Nevertheless, the sample-to-sample distance plot did not reveal notable differences between the replicates (**Supplementary Figure 34B**). The distinct clustering observed in PCA remained consistent throughout the entire sequencing process, until the 24-h PSI point (**Supplementary Figure 26O,**

Supplementary Figure 34A). NEW1 emerged as the only gene and transcript enriched in pNew1(*HIS3*) as compared to pEV(*HIS3*) at 1h PSI (**Supplementary Figure 26P, Supplementary Figures 34C-D,35C-D**). This observed difference persisted throughout the entire sequencing process (**Supplementary Figure 26Q, Supplementary Figures 34C-F, 35C-D, Supplementary Table 4**). Interestingly, as sequencing was going through, the number of DEG/T increased up until the 24h PSI mark where 13 genes were enriched and 2 were depleted in *new1Δ*-pNew1(*HIS3*) compared to *new1Δ*-pEV(*HIS3*) (**Supplementary Figure 26Q-S**). This experimental configuration exemplifies a rescue experiment, highlighting NanopoReaTA's ability to promptly detect the overexpressed gene.

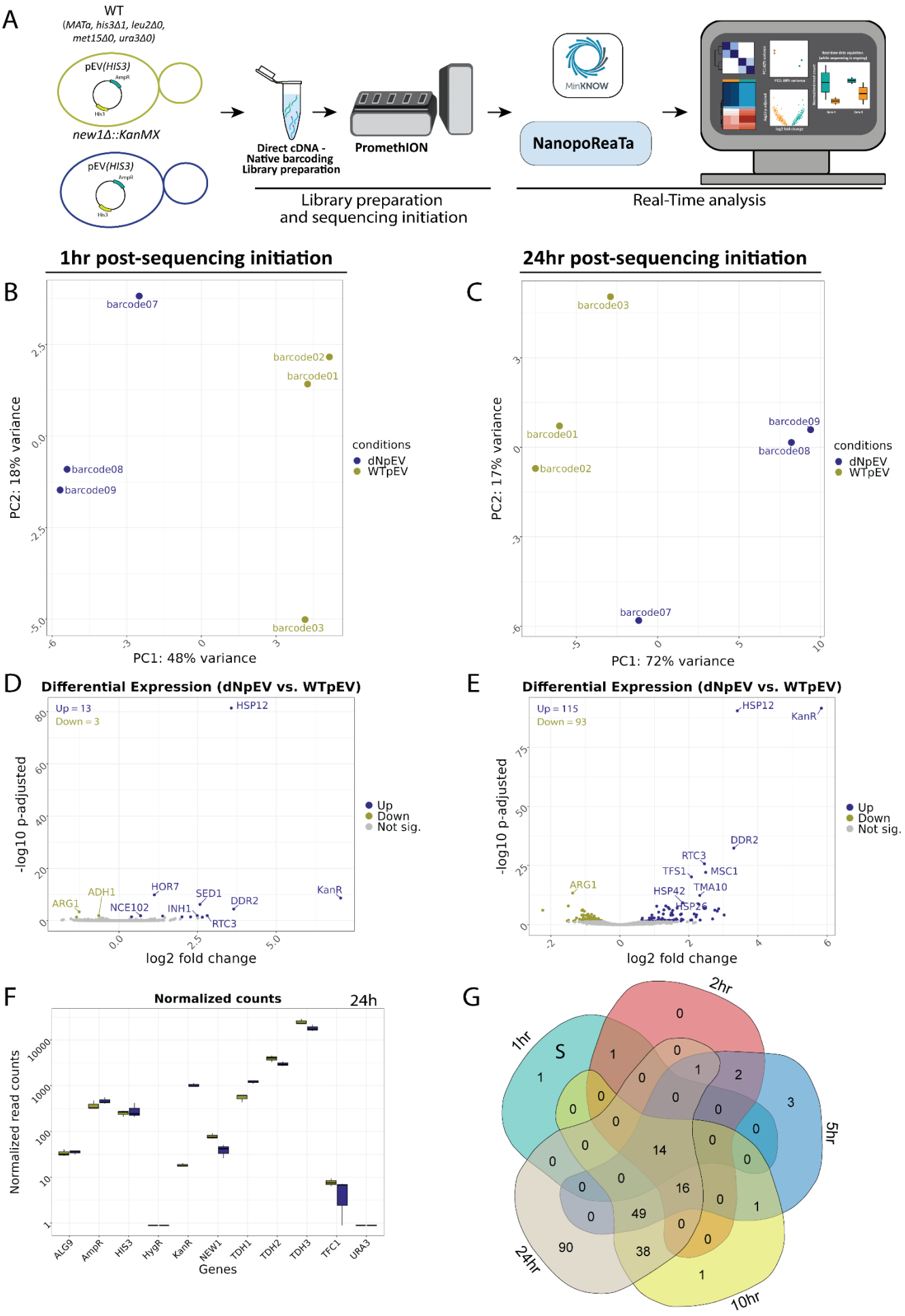

A
WT
(MATa, his3Δ1, leu2Δ0, met15Δ0, ura3Δ0)
pEV(HIS3)
new1Δ::KanMX
pEV(HIS3)
Direct cDNA - Native barcoding Library preparation
PromethION
MinKNOW
NanopoReaTa
Library preparation and sequencing initiation
Real-Time analysis
1hr post-sequencing initiation
24hr post-sequencing initiation
B
barcode07
barcode02
barcode01
barcode08
barcode09
barcode03
PC2: 18% variance
PC1: 48% variance
conditions
dNpEV
WTpEV
C
barcode03
barcode01
barcode09
barcode08
barcode02
barcode07
PC2: 17% variance
PC1: 72% variance
D
Differential Expression (dNpEV vs. WTpEV)
Up = 13
Down = 3
HSP12
ADH1
HOR7
SED1
KanR
ARG1
NCE102
INH1
DDR2
RTC3
-log10 p-adjusted
log2 fold change
Up
Down
Not sig.
E
Differential Expression (dNpEV vs. WTpEV)
Up = 115
Down = 93
HSP12
KanR
DDR2
RTC3
TFS1
MSC1
ARG1
HSP42
TMA10
HSP26
-log10 p-adjusted
log2 fold change
Up
Down
Not sig.
F
Normalized counts
24h
Normalized read counts
ALG9
AmpR
HIS3
HygR
KanR
NEW1
TDH1
TDH2
TDH3
TFC1
URA3
Genes
G
1hr
2hr
5hr
24hr
10hr
S

Figure 5. Real-time transcriptomic analysis between *new1Δ*-pEV(*HIS3*) vs WT-pEV(*HIS3*) using NanopoReaTA. A. Experimental strategy. Following library preparation (n=3), samples were loaded and sequenced using a PromethION R10 flow cell. NanopoReaTA was activated shortly after sequencing initiation and data were collected 1h, 2h, 5h, 10h, and 24h post-sequencing initiation. **B-C.** PCA plot showing the sample separation between *new1Δ*-pEV(*HIS3*) vs WT-pEV(*HIS3*) 1-h post-sequencing (B) and 24-h post-sequencing (C). **D-E.** Volcano plots showing the differentially expressed genes between *new1Δ*-pEV(*HIS3*) vs WT-pEV(*HIS3*) 1-h post-sequencing initiation (D) and 24-h post-sequencing initiation (E). **F.** Normalized gene count for selected genes 24h 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.

new1Δ-pNew1(*HIS3*) *vs WT*-pEV(*HIS3*). When comparing *new1Δ*-pNew1(*HIS3*) versus WT-pEV(*HIS3*) at 1h PSI, the PCA indicated separation between the samples based on their respective conditions (**Supplementary Figure 37A, 38A**); 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 37B, 38B**). 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 37C-F, 38C-E, Supplementary Table 4**). 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-pNew1*(HIS3) versus WT*-pNew1*(HIS3).* In the comparison of *new1Δ*-pNew1(*HIS3*) versus WT-pNew1(*HIS3*) at 1h PSI, the PCA exhibited separation between the samples based on their respective conditions (**Supplementary Figure 26T, Supplementary Figure 40A, 41A**), accompanied by clustering in the sample-to-sample distance plot (**Supplementary Figure 40B, 41B**). This clustering pattern persisted until the 24h PSI time point **(Supplementary Figure 25V, Supplementary Figure 40A, 41A**). *KanR*, *DDR2* and *HSP12* were identified as enriched genes in *new1Δ*-pNew1(*HIS3*) compared to WT-pNew1(*HIS3*) at 1h time point, appearing as both differentially expressed genes and transcripts (**Supplementary Figure 26U, Supplementary Figure 40C-F, 41C-E**). At 24h 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* (**Supplementary Figure 26W-Y, Supplementary Figure 40C-F, 41C-E, Supplementary Table 4**).

It is important to highlight that using the "gene-wise" utility, 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

rkr1Δ-pEV(*URA3*) versus WT-pEV(*HIS3*). First, we assessed 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*) (**Figure 5A, Supplementary Figures 43B-G)**. At 1h PSI, a notable separation between the conditions was evident in both the PCA and sample-to-sample distance plot (**Figure 6B, Supplementary Figure 45A-B, 46A-B**). This distinction persisted consistently until the 24-h PSI time point (**Figure 6C, Supplementary Figure 45A-B, 46A-B**). At 1h 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 6D, Supplementary Figure 45C-D, 46C-D**). At this stage, *HIS3* was not identified in the DEG analysis, but rather in the DET analysis. At the 24-h 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 6E-F, Supplementary Figure 45C-F, 46C-E**). Interestingly, 104 differentially

expressed genes were consistently identified from the 1-h mark until the final 24-h mark (**Figure 6G, Supplementary Table 5**). 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-pJlp2(*URA3*) *versus rkr1Δ*-pEV(*URA3*). Next, we tested the comparison of *rkr1Δ*-pJlp2(*URA3)* compared to *rkr1Δ*-pEV(*URA3*) for the detection of *JLP2* overexpression. One h 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*)) (**Supplementary Figure 43H, Supplementary Figure 48A, 49A**). The sample-to-sample distance plot did not uncover significant differences between the replicates (Supplementary Figure 42B, 43B). The distinct clustering observed in PCA was consistently maintained throughout the entire sequencing process (**Supplementary Figure 48A, 49A**) until the 24-h PSI mark (**Supplementary Figure 43I, Supplementary Figure 48A, 49A**). Notably, two genes/transcripts including *JLP2* and *RPL15A* were differentially expressed from 1hr PSI mark until 24hr PSI mark (**Supplementary Figure 43J-M**, **Supplementary Figure 48C-D, 49C-D, Supplementary Table 5**). 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 1h PSI, the PCA showed separation between the samples based on their respective conditions (**Supplementary Figure 43N, Supplementary Figure 51A, 52A**); Similarly, the clustering in the sample-to-sample distance plot appeared inconsistent (**Supplementary Figure 51B, 52B**). The clustering observed in PCA remained consistent throughout the entire sequencing process until the 24-h PSI time point (**Supplementary Figure 43O, Supplementary Figure 51A, 52A**). 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 (**Supplementary Figure 43P-S, Supplementary Figure 51C-F, 52C-D, Supplementary Table 5**). At the 24-h PSI mark, we observed enrichment of *JLP2* and depletion of 14 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, jlp2Δ-pEV(*URA3*) versus WT-pEV(*HIS3*). Lastly, we compared *rkr1Δ, jlp2Δ*-pEV(*URA3*) to WT-pEV(*HIS3*). At 1h PSI, noticeable separation between conditions was evident in both, PCA and sample-to-sample distance plots (**Supplementary Figure 43T, Supplementary Figure 54A-B, 55A-B**). This distinction persisted consistently until the 24-h PSI time point (**Supplementary Figure 43V, Supplementary Figure 54A-B, 55A-B**). At 1h 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*) (**Supplementary Figure 43U, Supplementary Figure 54C-D, 55C-D**). Similar to the previous observation, *HIS3* was not identified in the DEG analysis but rather in the DET analysis. By the 24-h 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*) (**Supplementary Figure 43W-Y, Supplementary Figure 54C-E, 55C-D**).

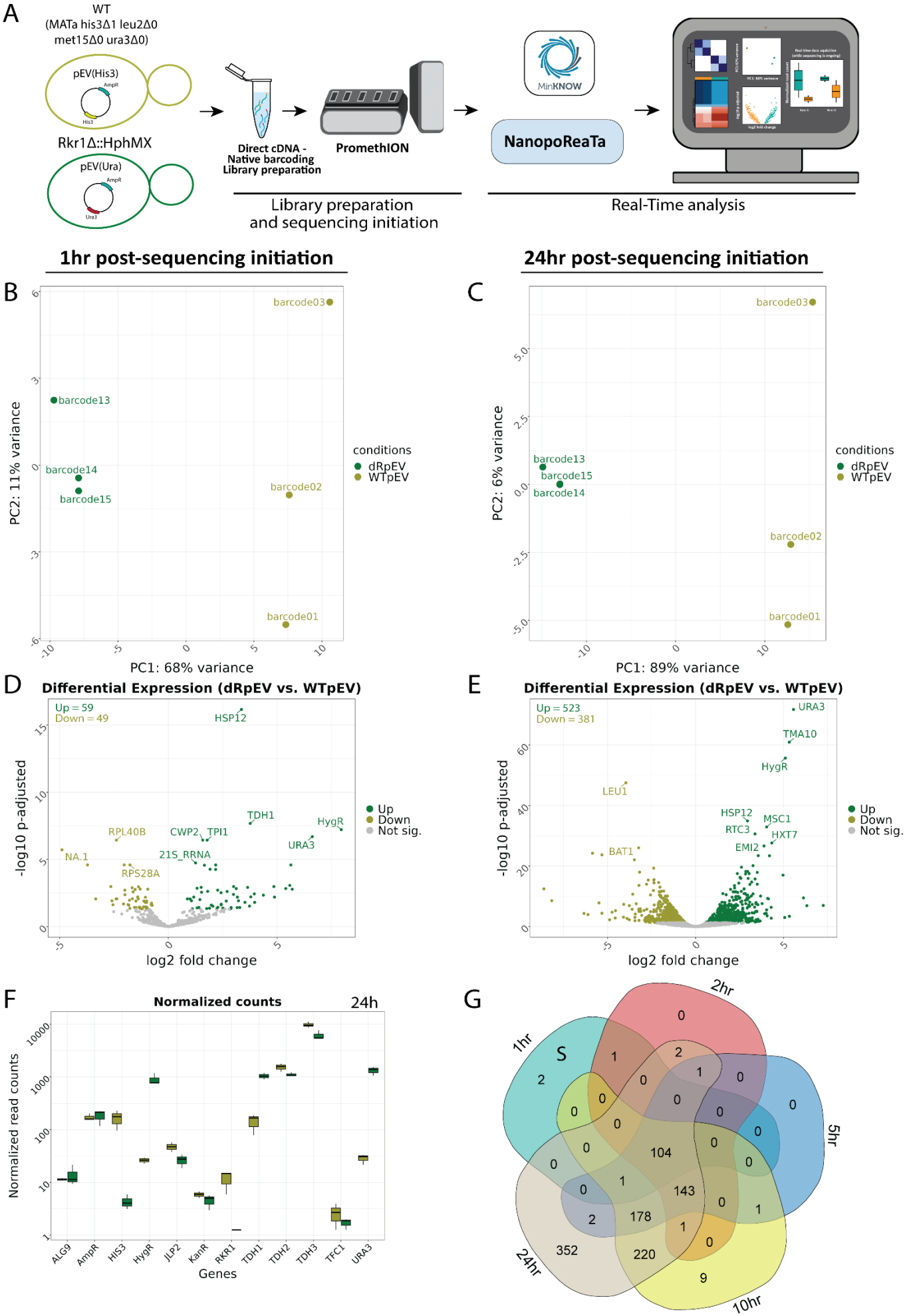

A
WT
(MATa his3Δ1 leu2Δ0
met15Δ0 ura3Δ0)
pEV(His3)
Rkr1Δ::HphMX
pEV(Ura)
Direct cDNA -
Native barcoding
Library preparation
PromethION
MinKNOW
NanopoReaTa
Library preparation
and sequencing initiation
Real-Time analysis
1hr post-sequencing initiation
24hr post-sequencing initiation
B
PC2: 11% variance
PC1: 68% variance
barcode03
barcode13
barcode14
barcode15
barcode02
barcode01
conditions
dRpEV
WTpEV
C
PC2: 6% variance
PC1: 89% variance
D
Differential Expression (dRpEV vs. WTpEV)
Up = 59
Down = 49
-log10 p-adjusted
log2 fold change
HSP12
TDH1
HygR
URA3
RPL40B
CWP2
TPI1
NA.1
21S_RRNA
RPS28A
Up
Down
Not sig.
E
Differential Expression (dRpEV vs. WTpEV)
Up = 523
Down = 381
URA3
TMA10
HygR
LEU1
HSP12
MSC1
RTC3
HXT7
EMI2
BAT1
F
Normalized counts
24h
Normalized read counts
ALG9
AmpR
HIS3
HygR
JLP2
KanR
RKR1
TDH1
TDH2
TDH3
TFC1
URA3
Genes
G
1hr
2hr
5hr
24hr
10hr
S
104
143
178
352
220
9

Figure 6. Real-time transcriptomic analysis between *rkr1Δ*-pEV(*URA3*) vs WT-pEV(*HIS3*) using NanopoReaTA. A. Experimental strategy. Following library preparation (n=3), samples were loaded and sequenced using a PromethION R10 flow cell. NanopoReaTA was activated shortly after sequencing initiation and data was collected 1h, 2h, 5h, 10h, and 24h post-sequencing initiation. **B-C.** PCA plot showing the sample separation between *rkr1Δ*-pEV(*URA3*) vs WT-pEV(*HIS3*) 1-h post-sequencing initiation (B) and 24-h post-sequencing initiation (C). **D-E.** Volcano plots showing the differentially expressed genes between *rkr1Δ*-pEV(*URA3*) vs WT-pEV(*HIS3*) 1-h post-sequencing initiation (D) and 24-h post-sequencing initiation (E). **F.** Normalized gene count for selected genes 24h 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.

A total of 101 differentially expressed genes were consistently identified from the 1-h mark until the final 24-h mark (**Supplementary Table 5)**. This experimental setup effectively detected the double knockout, as well as the expression of *HygR*, *KanR*, *URA3*, and *HIS3* in their corresponding experimental conditions.

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 principal 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).

Heat-shock experiments
HEK293 (1 Mio. cells) were plated on petri dishes (Greiner, diameter 10 cm, REF: 664160). After 48 h, plates were sealed with Parafilm and submerged in a water bath. Heat shock treatment was 42 °C for 45 min. Plates were put back at 37 °C to let the cells recover from heat shock. Non-induced cells were maintained at 37 °C.

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. RNA samples with A260/280 and A260/230 absorbance values >1.9 were taken for library preparation.

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 and RNA samples with A260/280 and A260/230 absorbance values >1.9 were taken for library preparation.

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 from 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 of the ribosomal-enriched samples (Riboplus). Qualitative analysis of the material was performed after rRNA depletion and enrichment. 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 and purified using 0.7X AMPure XP Beads and eluted in 31 µL nuclease-free water. The concentration of the pooled samples was determined. For adapter ligation, 5 µL NA, 10 µL NEBNext Quick Ligation Reaction Buffer (5X), and 5 µL Quick T4 DNA Ligase (NEB, cat # E6056) were mixed and incubated in RT for 20 min. Lastly, the library was purified with 0.7X AMPure XP Beads 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-hr 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-hr 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 hrs. 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-hr 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. M.J and J.B performed the heat-shock experiments and RNA isolation. 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

Rouillard, A. D., Gundersen, G. W., Fernandez, N. F., Wang, Z., Monteiro, C. D., McDermott, M. G., & Ma'ayan, A. (2016). The harmonizome: a collection of processed datasets gathered to serve and mine knowledge about genes and proteins. Database, 2016. https://doi.org/10.1093/DATABASE/BAW100

Uhlén, M., Fagerberg, L., Hallström, B. M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, Å., Kampf, C., Sjöstedt, E., Asplund, A., Olsson, I. M., Edlund, K., Lundberg, E., Navani, S., Szigyarto, C. A. K., Odeberg, J., Djureinovic, D., Takanen, J. O., Hober, S., … Pontén, F. (2015). Tissue-based map of the human proteome. Science, 347(6220). https://doi.org/10.1126/SCIENCE.1260419/SUPPL_FILE/1260419_UHLEN.SM.PDF

Qu, L., Hernandez, N. S., Chapman, L., Burnett, C., Gu, J., Bramlett, K., Schageman, J., Brockman, J., & Hinahon, C. (2013). Improvements to RiboMinusTM Eukaryote rRNA Depletion Probe Design and Functionality to Enable a Faster and More Complete Workflow. Journal of Biomolecular Techniques : JBT, 24(Suppl), S72. https://pmc.ncbi.nlm.nih.gov/articles/PMC3635353/

Neueder, A., Achilli, F., Moussaoui, S., & Bates, G. P. (2014). Novel Isoforms of Heat Shock Transcription Factor 1, HSF1γα and HSF1γβ, Regulate Chaperone Protein Gene Transcription. The Journal of Biological Chemistry, 289(29), 19894. https://doi.org/10.1074/JBC.M114.570739

Yonezawa, S., & Bono, H. (2023). Meta-Analysis of Heat-Stressed Transcriptomes Using the Public Gene Expression Database from Human and Mouse Samples. International Journal of Molecular Sciences, 24(17). https://doi.org/10.3390/IJMS241713444

Sanchez-Briñas, A., Duran-Ruiz, C., Astola, A., Arroyo, M. M., Raposo, F. G., Valle, A., & Bolivar, J. (2023). ZNF330/NOA36 interacts with HSPA1 and HSPA8 and modulates cell cycle and proliferation in response to heat shock in HEK293 cells. Biology Direct, 18(1), 1–14. https://doi.org/10.1186/S13062-023-00384-8/FIGURES/5

Ge, S. X., Jung, D., Jung, D., & Yao, R. (2020). ShinyGO: a graphical gene-set enrichment tool for animals and plants. Bioinformatics, 36(8), 2628–2629. https://doi.org/10.1093/BIOINFORMATICS/BTZ931

Liu, Y., Li, Z., Zhang, J., Liu, W., Guan, S., Zhan, Y., Fang, Y., Li, Y., Deng, H., & Shen, Z. (2024). DYNLL1 accelerates cell cycle via ILF2/CDK4 axis to promote hepatocellular carcinoma

development and palbociclib sensitivity. British Journal of Cancer 2024 131:2, 131(2), 243–257. https://doi.org/10.1038/s41416-024-02719-2

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

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

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