Redefining human pre-rRNA processing at single nucleotide resolution using long read Nanopore sequencing

Stefan Pastore[1], Ludivine Wacheul [2], Lioba Lehmann [3], Stefan Mündnich[1], Beat Lutz [4,5], Mark Helm [1], Susanne Gerber [3], Denis L.J. Lafontaine [2,6**],Tamer Butto [1,6,*]

[1]Institute of Pharmaceutical and Biomedical Sciences, Johannes Gutenberg-University Mainz, Mainz 55128, Germany.
[2]RNA Molecular Biology, Fonds de la Recherche Scientifique (F.R.S./FNRS), Université libre de Bruxelles (ULB), Biopark campus, B-6041 Gosselies, Belgium.
[3]Institute of Human Genetics, University Medical Center of the Johannes Gutenberg University Mainz, Mainz 55131, Germany.
[4]Leibniz Institute for Resilience Research (LIR), 55122 Mainz, Germany.
[5]Institute of Physiological Chemistry, University Medical Center Mainz, 55128 Mainz, Germany.

[6]Co-lead contacts
*Correspondence: buttamer@uni-mainz.de
**Correspondence: denis.lafontaine@ulb.be

Abstract

Ribosome biogenesis involves the synthesis of precursor ribosomal RNAs (pre-rRNAs) and their processing (cleavage) into mature rRNAs. Traditional techniques like northern blotting and metabolic labeling, while robust, have limited resolution and throughput. Here, we present NanoRibolyzer, a long-read nanopore sequencing-based method that enables *ab initio* characterization and quantification of rRNA precursors. By combining supervised and unsupervised mapping approaches, it identifies both known and novel rRNA species and precisely characterizes cleavage sites at single-nucleotide resolution. Additionally, a simplified cell fractionation procedure allows for spatial resolution of nuclear and cytoplasmic pre-rRNAs. Targeted knockdown experiments further quantify specific intermediate accumulations, defining condition-specific 'fingerprints' with potential biomarker value. With its unique mapping strategy, NanoRibolyzer advances our understanding of pre-rRNA processing providing high-resolution insights into ribosome biogenesis.

Keywords

NanoRibolyzer, ribosome biogenesis, rRNA intermediates, pre-rRNA processing, fingerprinting, Nanopore-sequencing, nucleolus

Introduction

Ribosomes are ribonucleoprotein nanomachines responsible for protein synthesis in all living cells[1,2]. Ribosome biogenesis is a complex process involving the synthesis, processing, and modification of precursor ribosomal RNAs (pre-rRNAs), as well as RNA folding and packaging into functional ribosomal subunits. In eukaryotes, this pathway is initiated in the nucleolus, where a large ribosomal RNA precursor (pre-rRNA), the 47S, is synthesized by RNA polymerase I (Pol I)[3,4]. The 47S contains sequences for three out of four rRNAs (the 18S, 5.8S, and 28S) interspersed with non-coding spacers (**Figures 1A and S1**). The fourth rRNA, 5S, is produced independently by Pol III, in the nucleoplasm. Following transcription, pre-rRNAs undergo a series of maturation steps, including processing (cleavage), modification, and packaging with ribosomal proteins, to release the mature rRNAs and produce the ribosomal subunits, which are ultimately exported to the cytoplasm where they engage in translation[5]. Throughout this multistep process, the nascent transcripts undergo extensive processing by endonucleases performing precise cleavages within the external and internal transcribed spacers (ETS and ITS, respectively) often followed by exonucleases that progressively trim pre-rRNAs, ultimately releasing the mature rRNAs (**Figures 1A and S1**). The progressive trimming of pre-rRNAs by exonucleases results in the production of transient, metastable species that likely remain largely uncharacterized. Additionally, these processes contribute to the generation of poorly defined RNA ends, further highlighting the complexity and incomplete understanding of pre-rRNA processing. Disruptions occurring at any stage of the pathway may activate regulatory cascades, including surveillance leading to the accumulation of distinctive intermediates, which can significantly impact ribosome function, cellular protein synthesis, and overall cellular homeostasis[4,6].

More than two decades of research have seen significant progress in identifying major discrete processing sites and pre-rRNA intermediates, which serve as critical markers for studying the efficacy of ribosome biogenesis (**Figures 1A and S1**). Knowledge of these intermediates has been particularly valuable for investigating aberrant precursor production that arises during processing perturbations[2]. Conventional approaches for analyzing rRNA processing intermediates, such as northern blotting, metabolic labelling, or primer extension, allow for the identification of these accumulated precursors and cleavage sites, respectively[6,8,9,10]. However, these assays require important input material (often µg range), have limited resolution and throughput. Additionally, studying pre-rRNA processing by sequencing has remained challenging due to the highly repetitive nature of rDNA arrays and their poor genome annotations, which makes it difficult to accurately map reads and distinguish between individual rDNA copies, especially with short-read sequencing technologies[11,12,13,14].

Nanopore sequencing (nanopore-seq) has emerged as a promising technology to investigate ribosome biogenesis[15,16]. The key advantage of nanopore-seq lies in its ability to sequence long reads, such as cDNAs, as well as native RNA molecules via direct RNA sequencing (DRS), allowing for the investigation of entire transcripts, including processed ones[17,18]. Nanopore-based RNA library preparation strategies typically capture poly(A)+ RNA[19]. Though polyadenylation is typically associated with mRNA maturation, it also plays a critical role in rRNA surveillance and turnover, particularly for aberrant rRNA transcripts[20,21,22,23]. Disruptions of ribosome biogenesis can lead to the accumulation of aberrant, misprocessed, or truncated transcripts, which are typically degraded by surveillance mechanisms involving polyadenylation. In contrast, faithfully produced rRNA species are expected to be non-polyadenylated, highlighting the distinction between faithfully processed (non-polyadenylated) and misprocessed (polyadenylated) species. To this date, there are no tools that exploit long-read sequencing for the analysis of human ribosomal RNA precursors.

Here, we present NanoRibolyzer, a method that combines state-of-the-art long-read nanopore sequencing with advanced bioinformatics for the spatially resolved analysis of pre-rRNA intermediates at single-nucleotide resolution. By employing a streamlined nuclei isolation procedure, we systematically analyze precursor and mature rRNA populations in both the nucleus and cytoplasm. This approach enables the detection of known and novel processing intermediates, redefining all pre-rRNA processing sites with single-nucleotide resolution.

Results

A simplified nuclei isolation procedure to characterize pre-rRNA precursors

Ribosome biogenesis initiates in the nucleolus, a multiphase biomolecular condensate that resides within the nucleus. To define spatially pre-rRNA processing, we first focused on isolating highly purified nuclei fractions, from which we extracted RNA for sequencing. We developed a straightforward isolation protocol involving density gradient separation using a simple benchtop centrifuge (**Figures 1B and S2A-B**). We applied the protocol to HEK293 cells to produce highly purified nuclear and cytoplasmic RNA fractions, and as a control, whole cell total RNA ('whole cell') (**Figures 1B and S2A**).

After separation, a quality assessment of isolated nuclei was conducted using DAPI staining, revealing debris-free and round intact nuclei[24] (**Figure S2C**). RNA was extracted and electropherograms produced using a TapeStation. As expected, the whole cell and cytoplasmic fractions displayed the abundant mature 18S and 28S rRNAs (all analysis performed in triplicate throughout this work, R1-R3) (**Figure S2D**). In contrast, the nuclear fractions exhibited higher molecular weight species, at the expected size for pre-rRNAs (**Figure S2D**).

Since Nanopore-based RNA library preparation strategies typically capture poly(A)+ RNA[19], we applied an *in-vitro* polyadenylation strategy followed by long-read cDNA sequencing (**Figure 1B**). To assess enrichment of transcripts in the nuclear fractions, we quantified two nuclear long non-coding RNAs (lncRNAs): XIST and MALAT1 (**Figure 1B**). As expected, comparative analysis of the abundance of these transcripts across different cellular compartments revealed several fold-change enrichments in the nuclear fraction relative to the cytoplasm (up to 165-fold for XIST).

Having established an efficient nucleo-cytoplasmic purification protocol, we developed NanoRibolyzer, a method relying on long-read nanopore sequencing in combination with a bioinformatic pipeline specifically designed for the detection and quantification of rRNA intermediates (**Figure S3A**).

NanoRibolyzer aligns long-reads to a single 47S template (equivalent to 45SN1; GeneID:106631777), in contrast to the multiple templates found in genome annotations, offering a clearer representation of

reads associated with pre-rRNA compared to whole-genome alignment. We generated an average of ~3 million reads per sample with an average ~75% alignment rate to the 47S template. For reference and to initiate precursor quantification analysis, we retrieved the positions of known processing sites and major precursors from literature[8,9,10] (**Figure 1A and S1**).

We applied NanoRibolyzer to unperturbed HEK293 cells and characterized reads mapping to rRNA precursors and non-coding spacers (5' ETS, ITS1, ITS2 and 3' ETS) for the nuclear, cytoplasmic, and whole cell fractions (**Figure 1C-D**).

As expected, the coverage profiles revealed that the nuclear fraction had higher coverage across the ETSs and ITSs compared to the cytoplasmic ones, with the whole cell fraction displaying an intermediate coverage level (**Figure 1C-D**). The start and ends of these reads were closely aligned with annotated processing sites, such as 01,2,4, etc. (see Refs[8,9,10]).

These observations agree with the notion that the nuclear fraction predominantly contains ribosomal intermediates, reflecting that most steps of ribosome biogenesis occur in the nucleolus. In these initial analyses, we often observed that whole cell fractions were largely redundant with that of cytoplasmic fractions (see **Fig 1D**, zoomed IGV), therefore in the remainder of the manuscript we will only compare the nuclear and cytoplasmic fractions.

A novel mapping strategy to characterize rRNA precursors

After confirming that the nuclear fraction coverage profiles encompass the spacer regions, we employed two complementary strategies to quantify pre-rRNA and identify processing sites: a classical supervised (template-based) and an original unsupervised (template-free) approach.

The supervised (template-based) approach considers known discrete pre-rRNA intermediates as well as the spacer regions, quantifying their relative abundance through reciprocal overlap maximization (**Figures 2A and S3B**). These intermediates were derived from the literature using the positions of processing sites on the 47S, allowing us to generate the major precursors depicted in **Figure 1A** (Refs[1,2]). The quantified data can be displayed using an intuitive heatmap, showing averaged log10 reads per million, or a histogram, depicting the averaged reads per million for each condition (**Figure S4A-B**).

Using the supervised approach, as expected, we observed that most precursors are more abundant in the nuclear fraction than in the cytoplasmic one (**Figure 2B and S4A-B**). This also illustrates that some precursors, which were historically associated with ribosome biogenesis perturbations, are indeed not produced or only marginally in unperturbed cells (e.g. 34S, 36S, 36S-C). Thus, NanoRibolyzer offers an efficient and qualitative approach to determining the relative levels of all major pre-rRNA precursors and mature rRNAs.

The original unsupervised (template-free) approach provides an unbiased mapping method. The rationale for implementing it was to study processing steps that involve progressive exonuclease trimming of pre-rRNAs generating transient, metastable species that are often poorly characterized as they produce ill-defined ends (**Figure 1A** and **S1**). In this case, reads are plotted in a 2-D graph, forming a matrix (**Figures 2C and S3C**). The x-axis and y-axis span from the transcription start site to the termination site of the primary transcript. Each rRNA read is plotted onto the matrix based on its starting and ending positions (x and y coordinates, respectively), enabling precise mapping and analysis of transcript boundaries (**Figures 2C and S3C**). Such displays offer an intuitive visualization of intensity "hubs" representing abundant rRNA products whose ends are at, or close to, well-established or putative novel processing sites (**Figures 2C and S3C**, see methods for further detail).

We applied the unsupervised approach to the nuclear and cytoplasmic fractions of unperturbed HEK293 cells to identify the unbiased aggregation of reads associated with the 47S template. In the nuclear fraction (in blue in **Figure 2D**), we observed an accumulation of reads associated with the ETS and ITS regions, correlating with processing sites (**Figures 2D and S4C**). This approach also enabled the detection of previously unobserved low-abundance and metastable processing intermediates, as well as by-products, appearing as "processing smears" in the intensity matrix (**Figure 2C**, see asterisks). In contrast, reads in the cytoplasmic fraction (in red in **Figure 2D**) were nearly exclusively associated to the three mature rRNAs, the 18S, 5.8S and 28S (**Figures 2D**). The individual matrices for the replicates (R1, R2, and R3) are shown in Fig **S4C**, attesting to the robustness of the analysis.

In conclusion, by combining supervised and unsupervised mapping approaches, we not only leverage existing knowledge of abundant and long-lived pre-rRNA precursors—previously mapped using

conventional techniques—but also create new opportunities for identifying previously undetected, low-abundance, metastable processing intermediates and by-products.

Redefining pre-rRNA processing sites to single nucleotide resolution

Next, we followed an agnostic approach to remap the processing sites by analyzing the intensity matrix obtained by unsupervised mapping. To achieve this, we extracted approximate coordinates from detectable intensity "hubs" in the nuclear fraction and annotated the associated 5' and 3' processing sites (**Figure 3A**). These coordinates are considered estimates due to variability among minor cleavage points within a hub, thus we bioinformatically selected the most predominant ones in proximity to putative processing sites.

Analysis of the intensity matrix revealed obvious processing sites associated with several known cleavages, allowing to map them to single nucleotide resolution and to show, for some of them, they may in fact correspond to two cleavages. We will now review them briefly, one by one:

(1) For **site 01**, the start site of the associated intensity hub maps to position C^{424}^T (where the proposed cleavage site is marked by ^), which confirms previous findings[25] (**Figure 3A; Hubs nº 1b, 2, 3, and 4**).

(2) For **site A0**, we identified a hub with a start site at ^{1672}C^G, located 20 nucleotides downstream of the proposed processing site[26] (**Figure 3A; Hub nº 1a)**.

(3) For **site 1**, marking the 5' end of 18S, we detected multiple hubs containing the start site ^{3654}C^T (**Figure 3A; Hubs nº 5a, 5b, 5c, and 5d)**

(4) For **site 3**, corresponding to the 3' end of 18S, we confirmed site ^{5523}A^A **(Figure 3A; Hub nº 5d).**

(5) For **site E**, interestingly, we identified two distinct end points: one we refer to as "**b**", located 11 nucleotides downstream of the previously reported site E (Refs [27,28]), with multiple cleavages observed at coordinate C^A^{5589}, and the other we refer to as "**E^2**", located in an AG-rich region, with reproducible end cleavages around G^A^{5729} (**Figures 3A and S5A-B ; Hubs nº 1b and 5C)**.

(6) For **site C**, we identified an end point located at site C^G^{6150}, 8 nucleotides downstream of the previously described conserved "region C" (Ref.[27]) (**Figure 3A; Hub nº 2**).

(7) For **site 2**, our analysis also revealed two distinct intensity hubs: site "**2^1**", located at G^T^{6380}, and site "**2^2**", positioned near A^G^{6459}, in agreement with the former observations [30,31] (**Figures 3A and S5A-B; Figure 3A; Hubs nº 3,4, 5a, and 5b**). These sites could also be observed in the coverage profiles upstream the putative site 2 (**Figures 1C-D**). These findings reveal two potential precursors associated with 21S, as previously suggested[29].

(8) For site **B1**, we detected the intensity hub precisely at the expected 5' end of 5.8S, at ^{5600}A^CGA, defining the short form of 5.8S, the $5.8S_S$ (**Figure 3A; Hub nº 6a, 8, 9, and 10)**. Interestingly, we identified an intensity hub six nucleotides upstream of the 5.8S start site at position ^{6594}C^C (**Figure 3A; annotated as ~B1 in hubs 6b and 7**). This site most likely corresponds to the previously described 5' extended version of 5.8S (both coexist in cells), defining the long form of 5.8S, the $5.8S_L$ ([8,32,33]).

(9) For **site 4'**, corresponding to the end site of 5.8S, we confirmed the end point located at site T^T^{6757}.

(10) For the still poorly characterized **site 4a** in ITS2(Refs [8,9,10]), we reveal a very clear end point intensity hub located at AGA^C^{6861} (**Figure 3A; Hub nº 6b**).

(11) For the uncharacterized **site 4**, we identified a distinct intensity hub at the end coordinate T^{7564}GT, defining its cleavage site (**Figure 3A; Hubs nº 7-8)**.

(12) At **site 3'**, marking the start of 28S, we confirmed the start point located at site ^{7925}A^C (**Figure 3A; Hub nº 11)**.

(13) Lastly, for **site 02**, marking the end of 28S, we confirmed the end point located at site ^{12990}C^C **(Figure 3A; Hub nº 11)**.

In conclusion, NanoRibolyzer's unsupervised mapping approach enables the detection and characterization of both known and novel processing sites. We mapped thirteen known processing sites at single-nucleotide resolution, revealing multiple endpoints at two sites (E^1 and E^2, and 2^1 and 2^2).

Having shown the intensity hubs identify processing sites, we next inspected previously undetected precursors (**Figure 3B**).

For instance, several intensity hubs correspond to species starting at site 01 and ending at different sites, such as site 3 (18S 3' end), E, C, 2^1 and 2^2 (**Figure 3B**). The detection of such precursors illustrates that cleavage in ITS1 can occur prior to cleavage at site A0, at least to some extent, in unperturbed cells.

We also identified species ending at a known site but starting at unknown continuous positions ("smear"), indicating they are subject exonucleolytic digestion (**Figures S5A-B**).

Additionally, we refined further the 21S annotation, categorizing the precursor into two distinct forms: $21S_S$ (small) and $21S_L$ (large), as previously suggested[29] (**Figure 3B; Hubs nº 5a-b, and S5A-B**).

In the case of the 5.8S maturation, we identified abundant 3' extended precursors corresponding to the progressive handover between 12S and 7S exoribonucleolytic processing, involving the RNA exosome (**Figures 3B; Hubs nº 6b-7, and S5A-B**).

Overall, we showcase how rRNA products, including known and novel metastable precursors, can be detected and analyzed at high resolution and sensitivity.

NanoRibolyzer captures pre-rRNA intermediate changes following processing perturbations

So far, we have characterized pre-rRNA processing in unperturbed cells. We next aimed to evaluate how efficiently the method can detect processing perturbations. To achieve this, we selected key processing factors involved in the maturation of each spacer sequence and depleted them (**Figure 4A**).

For the 5' ETS, we targeted UTP18, a component of the SSU-processome, whose depletion results in the accumulation of the aberrant 34S species (Ref. [34]). For ITS1, we selected WBSCR22, whose depletion results in accumulation the 18S-E pre-rRNA[35,36]. For ITS2, we selected LAS1L, whose absence causes accumulation of 32S pre-rRNA[36,37]. Lastly, for the 3′ ETS, we chose URB1, recently shown to be important for 3' ETS removal[38].

First, we assessed the depletion of each factor using northern blots, confirming the expected precursor accumulation (**Figure 4B**). For NanoRibolyzer analysis, we chose the optimal depletion time point for each factor (48 h for UTP18, WBSCR22, and LAS1L, and 72 h for URB1) (**Figure 4B**).

RNA from nuclear and cytoplasmic fractions was extracted in triplicate and analyzed using cDNA Nanopore-seq, as described above. Coverage profiles revealed distinct differences in processing factor knockdowns compared to controls (**Figures 4C and S6A-C**).

Upon UTP18 depletion, increased coverage was observed upstream of the 01 site, corresponding to the 34S species. WBSCR22 KD samples showed increased coverage between the 3' end of 18S and site E, corresponding to 18S-E precursors (**Figures 4C and S6B**). Depletion of LAS1L revealed increased reads across ITS2, in agreement with 32S accumulation (**Figures 4C** and **S6A-B**). URB1 depleted samples showed increased coverage across the 3′ ETS, in agreement with the accumulation of species that retained this extension in URB1 defective cells (**Figures 4C and S6A-B**). For URB1 depletion, it is interesting to note on the IGV and the matrix display that there are abundant fragments in the 5' ETS (notably starting at 01), and, more globally, that many fragments accumulate throughout the mature sequences, indicating increased turn over.

Each RNA species was quantified using the supervised approach and visualized as a heat map (**Fig 4D**) or histograms (**Fig S7**). The key perturbations associated with each tested factor could easily be visualized through accumulation of the expected diagnostic precursor RNA species (34S for UTP18, 18S-E for WBSCR22, 32S for LAS1L, and 3' ETS extended RNAs for URB1). These results highlight NanoRibolyzer's ability to detect precursor accumulation and processing perturbations.

Lastly, we analyzed the sequencing data using an unsupervised approach, mapping nuclear reads onto the intensity matrix (**Figures 4E and S8**). For UTP18 depleted samples, we identified a unique hub corresponding to 34S coordinates (red arrow, **Figures 4E and S8A**). Analysis of the WBSCR22 KD sample revealed accumulation of a hub associated with 18S-E (green arrow), with shortened start coordinates, indicating some level of 5'-3' degradation of 18S precursors (**Figures 4E and S8B**). LAS1L KD showed reduced aggregated reads at processing site 4 and accumulation at 32S, **Figures 4E and S8C**). URB1 KD exhibited accumulation of reads downstream of processing site 02 (3' end of 28S) corresponding to 3' ETS containing RNAs **(**green arrow, **Figures 4E and S8D)**. Additionally, we observed an increased accumulation upstream of the 01 site as observed in the coverage profiles (green arrow, **Figures 4E and S8D**) and across the mature rRNA sequences.

In conclusion, NanoRibolyzer identifies and quantifies pre-rRNAs and mature rRNAs, defining processing 'fingerprints'. We observed the expected accumulation of precursors upon depletion of representative processing factors and gained novel insights, such as a possible coupling of maturation at both ends of the primary transcript (URB1).

Discussion

Ribosome biogenesis is a complex pathway involving hundreds of interconnected steps[1,2]. Among these, pre-rRNA processing (RNA cleavage) to generate mature rRNA ends serves as an excellent proxy for the overall process (**Figure 1A**). Traditionally, processing intermediates have been analyzed using metabolic labelling, northern blotting or primer extension. While these techniques are robust, they are somewhat limited in resolution, sensitivity, and throughput and often require access to costly and hazardous radioactive materials. This is particularly problematic for detecting low-abundance, metastable intermediates generated by progressive exoribonucleolytic trimming, such as the 3' end maturation of the 18S and 5.8S rRNAs[7,8,9,10](Figure 1A).

We have introduced NanoRibolyzer, an experimental and bioinformatic platform designed for the analysis and quantification of pre-rRNA intermediates using long-read Nanopore-sequencing. Coupled with a streamlined nuclei isolation protocol, we successfully detected both precursor and mature forms of nuclear and cytoplasmic rRNAs (**Figure 1B**). Utilizing long-read cDNA sequencing, which supports multiplexing for increased throughput and cost efficiency, we applied two complementary mapping approaches: a supervised (template-based) strategy for known precursors, and an unsupervised (template-free) approach, providing an agnostic characterization of reads across the entire 47S template (**Figure 2**).

The supervised approach enabled us to obtain quantitative measurements of rRNA precursors and mature rRNAs, which is particularly valuable for studying processing perturbations, such as the inactivation ribosome processing factors (**Figure 4**).

Using the unsupervised approach, we detected pre-rRNAs associated with perturbations, defining processing fingerprints and identified cleavage sites. Cross-referencing with the literature confirmed that these sites aligned with known processing sites, including 01(Ref.[25]), A0 (Ref.[26]), E (Refs [28,29]), C (Ref.[27]) and 2 (Refs[29,30,31]) and others (**Figure 3**). While these sites exhibited sequence similarity at the nucleotide level, they did not always match the expected annotated positions. This highlights the importance of accurately documenting processing sites with nucleotide resolution using a standard reference sequence. We have summarized our findings on cleavage sites definition in **Figure 5**.

Additionally, we identified novel putative processing sites and alternative precursors, providing new insights into human ribosome biogenesis. For instance, we confirmed two potential precursors associated with site 2: one corresponding to the previously described 21S precursor (site 2^1, Ref. [30,31]) and a shorter 21S variant (site 2^2, Ref. [29]), which had been observed in a previous study[28]. We also predicted the putative location of site 4, which had not been fully characterized. In the 2-D matrix display, we observed what we refer to as processing "smears" within both spacers and mature rRNA. We believe these correspond to degradation resulting from aborted subunit production, surveillance, and/or stress-induced activation of damage pathways[4,39,40]. These early observations suggest that NanoRibolyzer could also serve as a powerful tool for exploring rRNA surveillance and turnover.
A few recent studies have applied Nanopore-seq to investigate distinct aspects of rRNA biology, including rRNA modification[41,42,43], processing in archaea[44] and yeasts[45] (although the protocol used was far more complex than NanoRiolbyzer, involving the selection of precursor ribosomes by affinity purification and circularization of the isolated RNAs prior to sequencing) and more recently, to investigate rRNA heterogeneity in human cells[46,47]. Building on these developments, NanoRibolyzer

offers a straightforward and efficient approach to studying ribosome biogenesis, which can easily be adapted for use in other organisms. Combining this method with investigations of rRNA modifications and ribosomal heterogeneity will provide deeper insights into ribosome biogenesis, function and diversity across different biological models.

In conclusion, NanoRibolyzer was conceived as a tool to help decode ribosome biogenesis, initially focusing on pre-rRNA processing. By providing single nucleotide resolution description of pre-rRNA processing sites and quantifying the amounts of pre-rRNA intermediates in the pathway, this technology should be greatly beneficial to fundamental research on ribosome biogenesis and to deciphering the molecular basis of ribosome biogenesis dysfunction diseases (ribosomopathies). Additionally, the method is expected to have broad applications in clinical diagnostics.

Material and methods

Cell lines and culture

HEK293 cells (ATCC CRL-1573) were cultured in DMEM supplemented with 10% FBS, and 1% L-glutamine and maintained in an incubator at 37°C and 5% CO_2.

siRNA inactivation experiments

Cells were revere transfected with silencers (10 nM, except URB1: 15 nM) in a time course (H6, H12, H24, H48, H72) to identify the best condition for Nanoribolyzer analysis[6]. All DsiRNAs silencers were used at 10 nM final (except for URB1, 15 nM). Silencers against UTP18, WBSCR22, DIMT1L, and LAS1L are Silencer Select (Ambion). Silencers against URB1 are DsiRNAs (IDT).

Simplified nuclei isolation protocol

Detailed description of the protocol is illustrated in the Supplementary Fig S1. Briefly, samples were trypsined and washed with cold PBS, after which they were resuspended in Nuclei Isolation Buffer (NIB - 10mM Tris-HCl (pH7.4),10mM NaCl, 3mM MgCl2, 0.1% Igepal, 0.1% Tween-20,1% BSA, 0.15mM Spermine, 0.15mM Spermidine, 0.2U/ul RNase inhibitor) and homogenized using a loose pestle with ten strokes. The homogenate was then incubated on ice for 15 minutes and subsequently centrifuged at 800 rpm for 5 minutes at 4 °C. The resulting soluble fraction (cytoplasmic) was transferred to a new tube, and Trizol was added to the sample at room temperature (RT) while nuclei isolation continued. The remaining pellet was subjected to two washes with 500 µl of ice-cold NIB buffer and centrifuged at 800rpm for 5 minutes at 4 °C. After the second wash, the supernatant was removed, and the pellet was resuspended with 300 µl of ice-cold NIB buffer. For nuclei isolation an equal volume (300 µl) of 50% Optiprep solution (Stemcell technologies, 07820) was added to the homogenate sample and thoroughly resuspended by pipetting, resulting in a 25% sample/optiprep mix. Then, 600 µl of 40% Optiprep solution, followed by 30% Optiprep solution, were layered in a 2 ml Eppendorf tube. The 25% sample/optiprep mix was layered on top of the 30% solution forming three visible layers. The tube was centrifuged at top speed (14,000rpm) in a bench centrifuge for 20 minutes at 4 °C. After centrifugation, the nuclei (~600 µl) were carefully collected from the 40%-30% phase and transferred into a 1.5 ml Lo-bind tube. 600 µl of Nuclei Wash Buffer (NWB - 10mM Tris-HCl (pH7.4),10mM NaCl, 3mM MgCl2, 0.1% Tween-20,1% BSA, 0.2U/ul RNase inhibitor) was added to the nuclei and thoroughly resuspended by pipetting. The sample was then centrifuged at 1,100rpm for 10 minutes at 4 °C, and the supernatant was removed. The nuclei were washed again with 500 µl of ice-cold NWB buffer and centrifuged similarly. The supernatant was removed, leaving approximately 20 µl of solution. Quality of the nuclei was assessed using DAPI staining and visualized under the microscope using a UV filter to identify DAPI-positive nuclei. High-quality nuclei were characterized by debris-free, round or oval-shaped DAPI-stained nuclei. Once the quality of isolated nuclei was confirmed, remaining nuclei samples were treated with DNaseI (NEB). The DNase I master mix was supplemented with 0.15 mM spermidine, 0.15 mM spermine, and RNAse inhibitor (U/ul) per sample. The samples were then incubated at 37 °C for 20 minutes. Next, the sample volume was brought to 500 µl using nuclease-free water, and 500 µl of TRIzol reagent (Life Technologies, 15596026) was added to the nuclei samples to proceed with RNA isolation.

RNA isolation

Nuclear and cytoplasmic (as well as whole cell) fractions (~1 mL) were incubated in RT for at least 5 Min. 200µl Chlorophorm was added to nuclear and cytoplasmic fractions. Samples were Vortexed for 15 sec, incubated in RT fort 3 min and centrifuged for 15 min full speed (FS) at 4 °C. The upper aqueous phase (~550 µl) was transferred into a fresh Eppendorf tube and 500 µl isopropanol was added, thoroughly resuspended and incubated in RT for 15 min. Sample were centrifuge 10 min at full speed at 4 °C. The supernatant was discarded pellet was washed with 75% EtOH in nuclease free water followed by a 5 min centrifugation at 7500g at 4°C. The supernatant was discarded, and pellet was air dried for 5 – 10 min. RNA was eluted with 30 µl of RNAse free water and mix in Hula mixer for 10 min in RT. RNA concentration was measured with qubit and RNA integrity was assessed via Agilent RNA ScreenTape analysis.

Northern Blotting

Total RNA was extracted using TRIzol reagent, separated on agarose denaturing gel, and analyzed by northern blotting, as previously described[6]. The depletion optimization was performed on HEK293 cells where each depletion factor displayed the expected phenotype according to the literature. On the basis of this analysis, the following depletion time points were selected for NanoRibolyzer: UTP18 (48h), DIMT1L (72h), WBSCR22 (48h), LAS1L (48h), and URB1 (72h) (see Fig 3b).

In-vitro poly adenylation using poly(A) tailing of RNA

For in-vitro polyadenylation, “In-vitro poly adenylation using poly(A) tailing of RNA” kit was used (NEB#M0276), according to the manufacturer instructions. Briefly, 1 µg of RNA was taken in 15 µl nuclease free water and supplemented with 2 µl of 10X E. coli Poly(A) Polymerase Reaction Buffer, 2 µl ATP (10mM) and 1 µl E. coli Poly(A) Polymerase (to a total volume 20 µl). Samples were incubated at 37°C for 30 minutes. The polyadenylated RNA samples were purified using RNAClean XP beads (Beckman Coulter, A63987), according to the manufacturer instructions. In the last elution step, the sample was resuspended in 10 µl nuclease-free water and incubated at 37°C for 5 minutes in Hula mixer. The sample was placed on the magnet and once the solution was clear, the elute was transferred into a clean 1.5 ml Eppendorf tube.

Direct cDNA-native barcoding library preparation

Direct cDNA coupled with native barcoding libraries were prepared using Direct cDNA Sequencing Kit (SQK-DCS109), Native Barcoding Expansion 1-24 (EXP-NBD104, EXP-NBD114), following the manufacturer's protocol. *Reverse transcription and strand-switching.* 1µg of poly(A)-tailed RNA was transferred to a 1.5 ml tube and adjusted to 7.5 µl with nuclease-free water. In a 0.2 ml PCR tube, 7.5 µl of RNA sample were mixed with 2.5 µl of VNP (ONT), 2.5 µl of 10 mM dNTPs (NEB N0447), and the volume was adjusted to 11 µl with nuclease-free water. The samples were incubated for 10 minutes at room temperature and then snap-cooled on a pre-chilled freezer block for 1 minute. Next, a master mix was prepared, containing 4 µl of 5x RT Buffer (ThermoFisher, EP0751**)**, 1 µl RNaseOUT (Life Technologies, 10777019), 1 µl of Nuclease-free water, and 2 µl Strand-Switching Primer (SSP, ONT) per sample, to a total volume of 8 µl. The strand-switching buffer was added to the snap-cooled, annealed mRNA, and the samples were incubated at 42°C for 2 minutes in the thermal cycler. Subsequently, 1 µl of Maxima H Minus Reverse Transcriptase (ThermoFisher, EP0751) was added, and the total volume becomes 20 µl. The samples were incubated following a specific thermal protocol: 42°C for 90 mins, 85°C for 5 mins, and then holding at 4°C. After the reverse transcription, RNA degradation and second strand synthesis were performed. 1 µl of RNase Cocktail Enzyme Mix (ThermoFisher, AM2286) was added to the reverse transcription reaction and incubated for 10 minutes at 37°C. The samples were then subjected to the AMPure XP beads-based (Beckman Coulter A63881) purification method using 0.85x ratio of beads:sample and ultimately cDNA hybrid was eluted in 20 µl of nuclease-free water. Next, the 20 µl of reverse-transcribed samples were prepared with 25 µl of 2x LongAmp Taq Master Mix (NEB, N0447), 2 µl of PR2 Primer (PR2, ONT), and 3 µl of Nuclease-free water, to a total volume of 50 µl. The samples were incubated at specific temperatures in the thermocycler. Afterwards, the sample were then subjected to the AMPure XP beads-based purification method using 0.8x ratio of beads:sample and ultimately cDNA/RNA hybrid was eluted in 21 µl of nuclease-free water. The eluted

sample was quantified using a Qubit fluorometer. *End-prep.* Subsequently, end repair and dA-tailing were performed by mixing 20 µl of cDNA sample with 30 µl Nuclease-free water, 7 µl Ultra II End-prep reaction buffer (NEB, E7546), and 3 µl Ultra II End-prep enzyme (NEB, E7546) mix to a total volume of 60 µl. The samples were incubated at 20°C for 5 minutes and then at 65°C for 5 minutes. Next, the samples were subjected to AMPure XP beads-based purification using 1x ratio of beads:sample, and the cDNA was eluted with 22.5 µl of nuclease-free water. *Barcode ligation.* Barcode ligation was then performed, where 22.5 µl of End-prepped DNA was mixed with 2.5 µl of Native Barcode and 25 µl of Blunt/TA Ligase Master Mix (NEB, M0367) to a total volume of 50 µl. The reaction was incubated for 15 minutes at room temperature. The samples were then subjected to AMPure XP beads-based purification using 1x ratio of beads:sample, and the cDNA was eluted with 26 µl of nuclease-free water. Lastly, the barcoded samples are pooled to a final volume of 65 µl in a 1.5 ml Eppendorf tube. *Adapter ligation.* Adapter ligation was performed by adding 65 µl of pooled barcoded sample, 5 µl of Adapter Mix II (AMII, ONT), 20 µl of 5X NEBNext Quick Ligation Reaction Buffer (NEB, B6058), and 10 µl of Quick T4 DNA Ligase (NEB, E6056) to a total volume of 100 µl. The final libraries were incubated for 10 minutes at room temperature and then subjected to AMPure XP beads-based purification, with the cDNA being eluted with 26 µl of nuclease-free water. The sample was loaded and sequenced onto a primed PromethION flow cell as per the manufacturer instruction.

Direct RNA library preparation

Direct RNA libraries were prepared using the SQK-RNA004 kit (ONT) following the manufacturer's protocol. Briefly, 1µg of poly(A)-tailed RNA was adjusted to final volume of 9.5 µl with nuclease-free water. 3 µl of NEBNext Quick Ligation Reaction Buffer (NEB B6058), 1 µl of RT Adapter (RTA) (ONT), and 1.5 µl of T4 DNA Ligase 2M U/ml (NEB M0202), were added to sample resulting in a total volume of 15 µl. The reaction is mixed by pipetting and incubated for 10 minutes at room temperature. Next, the reverse transcription master mix was prepared by mixing 9 µl of Nuclease-free water, 2 µl of 10 mM dNTPs (NEB N0447), 8 µl of 5x first-strand buffer (Thermo Fisher Scientific, 18080044), and 4 µl of 0.1 M DTT, resulting in a total volume of 23 µl. The master mix was added to the RNA sample containing the RT adapter-ligated RNA. 2 µl of SuperScript III reverse transcriptase (Thermo Fisher Scientific, 18080044) were added to the reaction, bringing the final volume to 40 µl. The reaction was incubated at 50°C for 50 minutes, followed by 70°C for 10 minutes, and then brought to 4°C. Agencourt RNAClean XP beads (Beckman Coulter, A63987) were resuspended and 72 µl of the resuspended beads were added to the reaction. The sample was mixed by pipetting and incubated on a Hula mixer for 5 minutes at room temperature. Subsequently, the sample was subjected to two washes with 70% ethanol, and the RNA:DNA hybrids were eluted with 20 µl of nuclease-free water. For the adapter ligation reaction, 8.0 µl of NEBNext Quick Ligation Reaction Buffer, 6.0 µl of RNA Ligation Adapter (RLA), 3.0 µl of Nuclease-free water, and 3.0 µl of T4 DNA Ligase were mixed with 20 µl of the eluted sample to reach a total volume of 40 µl. The reaction was incubated for 10 minutes at room temperature. 30 µl of resuspended RNAClean XP beads were added to the adapter ligation reaction, mixed by pipetting, and incubated on a Hula mixer for 5 minutes at room temperature. The sample was then subjected to two washes with the Wash Buffer (WSB, ONT) using a magnetic rack. Following the washes, the beads were pelleted on the magnet, and the supernatant was pipetted off. The pellet was resuspended in 33 µl of Elution Buffer (EB, ONT) and were incubated at 37°C for 10 minutes in a Hula mixer. Incubation at 37°C allows the release of long fragments from the beads. The eluate was then cleared by pelleting the beads on a magnet, and the eluate was retained and transferred to a clean to 1.5 ml tube. The sample was loaded and sequenced onto primed PromethION flow cell as per the manufacturer instruction.

Quantification and statistical analysis

Implementation of NanoRibolyzer pipeline

NanoRibolyzer was implemented as a Nextflow-based workflow using Docker containers and could be installed as a plugin within Oxford Nanopore Technologies' (ONT) Epi2Me platform (https://github.com/stegiopast/wf-nanoribolyzer). Pod5 output files were basecalled using the dorado basecaller (https://github.com/nanoporetech/dorado), trimmed with Porechop

(https://github.com/rrwick/Porechop) to remove adapter sequences and aligned to the 45SN1 (equivalent to 47S) (GeneID:106631777; NW_021160023.1:480347-493697) using minimap2[48] with the map-ont flag. The read IDs of the aligned 45SN1 were used to filter the original pod5 file. The resulting unaligned BAM files were used to collect metadata on reads at a single nucleotide resolution and the rebasecalled reads were realigned to the 45SN1 reference. The final BAM files were processed using both template-based and template-free approaches.

Template-based quantification of rRNA precursors

The template-based algorithm associated long-reads with literature-based ribosomal intermediates[8,9,10]. In this approach, the pairwise minimal reciprocal overlap (MRO) between a query read and all possible intermediates was determined. The MRO was defined by calculating the minimal relative overlap of the query over the intermediate and vice versa. Once the minimal overlap for each query-intermediate pair was established, the pair with the maximal overlap was used to associate the read with the corresponding intermediate (See Figure S3). Read clusters were then stored in a tab-separated values (TSV) table, which included read IDs, absolute and relative read counts, and the start and end sites of all reads associated with each intermediate. Additionally, bed files for each intermediate were generated to facilitate visualization in the Integrative Genomics Viewer[49] (IGV). The 45SN1 reference FASTA from the NanoRibolyzer references repository was used for all analyses.

Template-free rRNA precursors

The template-free algorithm was based on the construction of a 2-dimensional (length(45SN1)^2) intensity matrix in which reads were embedded using the alignment start and end sites as coordinates. The number of reads sharing start and end site coordinates on the matrix led to the formation of intensity “hubs”. The resulting intensity matrix was stored in CSV format, which included the start site, end site, number of reads, and ID list for each intensity hub. For visualization of the matrices, absolute read counts of intensity hubs were min-max normalized applying an additional contrast enhancement of 2%.

Statistics

All statistical analyses are described in the respective figure legends. Each legend provides detailed information about the statistical metrics (such as the mean and standard deviation), sample sizes, statistical tests used and any relevant adjustments applied.

References

1 Lafontaine, D. L. J. (2015). Noncoding RNAs in eukaryotic ribosome biogenesis and function. Nature Structural & Molecular Biology, 22(1), 11–19.

2 Bohnsack, K. E., & Bohnsack, M. T. (2019). Uncovering the assembly pathway of human ribosomes and its emerging links to disease. The EMBO Journal, 38(13).

3 Ni, C., & Buszczak, M. (2023). Ribosome biogenesis and function in development and disease. Development (Cambridge, England), 150(5).

4 Kang, J., Brajanovski, N., Chan, K. T., Xuan, J., Pearson, R. B., & Sanij, E. (2021). Ribosomal proteins and human diseases: molecular mechanisms and targeted therapy. Signal Transduction and Targeted Therapy 2021 6:1, 6(1), 1–22.

5 Popov, A., Smirnov, E., Kováčik, L., Raška, O., Hagen, G., Stixová, L., & Raška, I. (2013). Duration of the first steps of the human rRNA processing. Nucleus, 4(2), 134–141.

6 Tafforeau, L., Zorbas, C., Langhendries, J. L., Mullineux, S. T., Stamatopoulou, V., Mullier, R., Wacheul, L., & Lafontaine, D. L. J. (2013). The complexity of human ribosome biogenesis revealed by systematic nucleolar screening of Pre-rRNA processing factors. Molecular Cell, 51(4), 539–551.

7 Grummt, I. (2013). The nucleolus—guardian of cellular homeostasis and genome integrity. Chromosoma, 122(6), 487–497.

8 Tomecki, R., Sikorski, P. J., & Zakrzewska-Placzek, M. (2017). Comparison of preribosomal RNA processing pathways in yeast, plant and human cells – focus on coordinated action of endo- and exoribonucleases. FEBS Letters, 591(13), 1801–1850.

9 Henras, A. K., Plisson-Chastang, C., O'Donohue, M. F., Chakraborty, A., & Gleizes, P. E. (2015). An overview of pre-ribosomal RNA processing in eukaryotes. Wiley Interdisciplinary Reviews: RNA, 6(2), 225–242.

10 Mullineux, S. T., & Lafontaine, D. L. J. (2012). Mapping the cleavage sites on mammalian pre-rRNAs: where do we stand? Biochimie, 94(7), 1521–1532.

11 Hall, A. N., Morton, E., & Queitsch, C. (2022). First discovered, long out of sight, finally visible: ribosomal DNA. Trends in Genetics : TIG, 38(6), 587–597.

12 McStay, B. (2016). Nucleolar organizer regions: genomic "dark matter" requiring illumination. Genes & Development, 30(14), 1598–1610.

13 Agrawal, S., & Ganley, A. R. D. (2018). The conservation landscape of the human ribosomal RNA gene repeats. PLOS ONE, 13(12), e0207531.

14 Floutsakou, I., Agrawal, S., Nguyen, T. T., Seoighe, C., Ganley, A. R. D., & McStay, B. (2013). The shared genomic architecture of human nucleolar organizer regions. Genome Research, 23(12), 2003–2012.

15 Wang, Y., Zhao, Y., Bollas, A., Wang, Y., & Au, K. F. (2021). Nanopore sequencing technology, bioinformatics and applications. Nature Biotechnology 2021 39:11, 39(11), 1348–1365.

16 Kono, N., & Arakawa, K. (2019). Nanopore sequencing: Review of potential applications in functional genomics. Development, Growth & Differentiation, 61(5), 316–326.

17 Amarasinghe, S. L., Su, S., Dong, X., Zappia, L., Ritchie, M. E., & Gouil, Q. (2020). Opportunities and challenges in long-read sequencing data analysis. Genome Biology, 21(1).

18 Zheng, P., Zhou, C., Ding, Y., Liu, B., Lu, L., Zhu, F., & Duan, S. (2023). Nanopore sequencing technology and its applications. MedComm, 4(4), e316.

19 Begik, O., Diensthuber, G., Liu, H., Delgado-Tejedor, A., Kontur, C., Niazi, A. M., Valen, E., Giraldez, A. J., Beaudoin, J. D., Mattick, J. S., & Novoa, E. M. (2023). Nano3P-seq: transcriptome-wide analysis of gene expression and tail dynamics using end-capture nanopore cDNA sequencing. Nature Methods, 20(1), 75–85.

20 LaCava, J., Houseley, J., Saveanu, C., Petfalski, E., Thompson, E., Jacquier, A., & Tollervey, D. (2005). RNA degradation by the exosome is promoted by a nuclear polyadenylation complex. Cell, 121(5), 713–724.

21 Shcherbik, N., Wang, M., Lapik, Y. R., Srivastava, L., & Pestov, D. G. (2010). Polyadenylation and degradation of incomplete RNA polymerase I transcripts in mammalian cells. EMBO Reports, 11, 106–111.

22 Jia, H., Wang, X., Liu, F., Guenther, U. P., Srinivasan, S., Anderson, J. T., & Jankowsky, E. (2011). The RNA Helicase Mtr4p Modulates Polyadenylation in the TRAMP Complex. Cell, 145(6), 890–901

23 Pirouz, M., Munafò, M., Ebrahimi, A. G., Choe, J., & Gregory, R. I. (2019). Exonuclease requirements for mammalian ribosomal RNA biogenesis and surveillance. Nature Structural & Molecular Biology 2019 26:6, 26(6), 490–500.

24 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 2023, Vol. 12, Page 1051, 12(7), 1051.

25 Kass, S., Craig, N., & Sollner Webbi, B. (1987). Primary Processing of Mammalian rRNA Involves Two Adjacent Cleavages and Is Not Species Specific. MOLECULAR AND CELLULAR BIOLOGY, 7(8), 2891–2898.

26 Rouquette, J., Choesmel, V., & Gleizes, P. E. (2005). Nuclear export and cytoplasmic processing of precursors to the 40S ribosomal subunits in mammalian cells. The EMBO Journal, 24(16), 2862–2872.

27 Preti, M., O'Donohue, M. F., Montel-Lehry, N., Bortolin-Cavaillé, M. L., Choesmel, V., & Gleizes, P. E. (2013). Gradual processing of the ITS1 from the nucleolus to the cytoplasm during synthesis of the human 18S rRNA. Nucleic Acids Research, 41(8), 4709.

28 Montellese, C., Montel-Lehry, N., Henras, A. K., Kutay, U., Gleizes, P. E., & O'Donohue, M. F. (2017). Poly(A)-specific ribonuclease is a nuclear ribosome biogenesis factor involved in human 18S rRNA maturation. Nucleic Acids Research, 45(11), 6822–6836.

29 Ishikawa, H., Yoshikawa, H., Izumikawa, K., Miura, Y., Taoka, M., Nobe, Y., Yamauchi, Y., Nakayama, H., Simpson, R. J., Isobe, T., & Takahashi, N. (2017). Poly(A)-specific ribonuclease regulates the processing of small-subunit rRNAs in human cells. Nucleic Acids Research, 45(6), 3437–3447.

30 Schillewaert, S., Wacheul, L., Lhomme, F., & Lafontaine, D. L. J. (2012). The Evolutionarily Conserved Protein LAS1 Is Required for Pre-rRNA Processing at Both Ends of ITS2. Molecular and Cellular Biology, 32(2), 430.

31 Sloan, K. E., Mattijssen, S., Lebaron, S., Tollervey, D., Pruijn, G. J. M., & Watkins, N. J. (2013). Both endonucleolytic and exonucleolytic cleavage mediate ITS1 removal during human ribosomal RNA processing. Journal of Cell Biology, 200(5), 577–588.

32 Hölzel, M., Orban, M., Hochstatter, J., Rohrmoser, M., Harasim, T., Malamoussi, A., Kremmer, E., Längst, G., & Eick, D. (2010). Defects in 18 S or 28 S rRNA processing activate the p53 pathway. The Journal of Biological Chemistry, 285(9), 6364–6370.

33 Haag, S., Kretschmer, J., & Bohnsack, M. T. (2015). WBSCR22/Merm1 is required for late nuclear pre-ribosomal RNA processing and mediates N7-methylation of G1639 in human 18S rRNA. RNA (New York, N.Y.), 21(2), 180–187.

34 Zorbas, C., Nicolas, E., Wacheul, L., Huvelle, E., Heurgué-Hamard, V., & Lafontaine, D. L. J. (2015). The human 18S rRNA base methyltransferases DIMT1L and WBSCR22-TRMT112 but not rRNA modification are required for ribosome biogenesis. Molecular Biology of the Cell, 26(11), 2080–2095.

35 Castle, C. D., Cassimere, E. K., Lee, J., & Denicourt, C. (2010). Las1L is a nucleolar protein required for cell proliferation and ribosome biogenesis. Molecular and Cellular Biology, 30(18), 4404–4414.

36 Shan, L., Xu, G., Yao, R. W., Luan, P. F., Huang, Y., Zhang, P. H., Pan, Y. H., Zhang, L., Gao, X., Li, Y., Cao, S. M., Gao, S. X., Yang, Z. H., Li, S., Yang, L. Z., Wang, Y., Wong, C. C. L., Yu, L., Li, J., … Chen, L. L. (2023). Nucleolar URB1 ensures 3′ ETS rRNA removal to prevent exosome surveillance. Nature 2023 615:7952, 615(7952), 526–534.

37 Goldfarb, K. C., & Cech, T. R. (2017). Targeted CRISPR disruption reveals a role for RNase MRP RNA in human preribosomal RNA processing. Genes and Development, 31(1), 59–71.

38 Idol, R. A., Robledo, S., Du, H.-Y., Crimmins, D. L., Wilson, D. B., Ladenson, J. H., Bessler, M., & Mason, P. J. (2007). Cells depleted for RPS19, a protein associated with Diamond Blackfan Anemia, show defects in 18S ribosomal RNA synthesis and small ribosomal subunit production.

39 King, K. L., Jewell, C. M., Bortner, C. D., & Cidlowski, J. A. (2000). 28S ribosome degradation in lymphoid cell apoptosis: evidence for caspase and Bcl-2-dependent and -independent pathways. Cell Death & Differentiation 2000 7:10, 7(10), 994–1001.

40 Parker, M. D., & Karbstein, K. (2023). Quality control ensures fidelity in ribosome assembly and cellular health. Journal of Cell Biology, 222(4).

41 Smith, A. M., Jain, M., Mulroney, L., Garalde, D. R., & Akeson, M. (2019). Reading canonical and modified nucleobases in 16S ribosomal RNA using nanopore native RNA sequencing. PLOS ONE, 14(5), e0216709.

42 Jain, M., Olsen, H. E., Akeson, M., & Abu-Shumays, R. (2021). Adaptation of Human Ribosomal RNA for Nanopore Sequencing of Canonical and Modified Nucleotides. Methods in Molecular Biology (Clifton, N.J.), 2298, 53–74.

43 Naarmann-de Vries, I. S., Zorbas, C., Lemsara, A., Piechotta, M., Ernst, F. G. M., Wacheul, L., Lafontaine, D. L. J., & Dieterich, C. (2023). Comprehensive identification of diverse ribosomal RNA modifications by targeted nanopore direct RNA sequencing and JACUSA2. RNA Biology, 20(1), 652–665.

44 Grünberger, F., Jüttner, M., Knüppel, R., Ferreira-Cerca, S., & Grohmann, D. (2023). Nanopore-based RNA sequencing deciphers the formation, processing, and modification steps of rRNA intermediates in archaea. RNA, 29(8), 1255–1274.

45 An, W., Yan, Y., & Ye, K. (2024). High resolution landscape of ribosomal RNA processing and surveillance. Nucleic Acids Research, 52(17), 10630–10644.

46 Rothschild, D., Susanto, T. T., Sui, X., Spence, J. P., Rangan, R., Genuth, N. R., Sinnott-Armstrong, N., Wang, X., Pritchard, J. K., & Barna, M. (2024). Diversity of ribosomes at the level of rRNA variation associated with human health and disease. Cell Genomics, 4(9), 100629.

47 Milenkovic, I., Cruciani, S., Llovera, L., Lucas, M. C., Medina, R., Pauli, C., Heid, D., Muley, T., Schneider, M. A., Klotz, L. v., Allgäuer, M., Lattuca, R., Lafontaine, D. L. J., Müller-Tidow, C., & Novoa, E. M. (2024). Epitranscriptomic rRNA fingerprinting reveals tissue-of-origin and tumor-specific signatures. Molecular Cell, 0(0).

48 Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics, 34(18), 3094–3100.

49 Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., & Mesirov, J. P. (2011). Integrative genomics viewer. Nature Biotechnology, 29(1), 24–26.

Main Figures

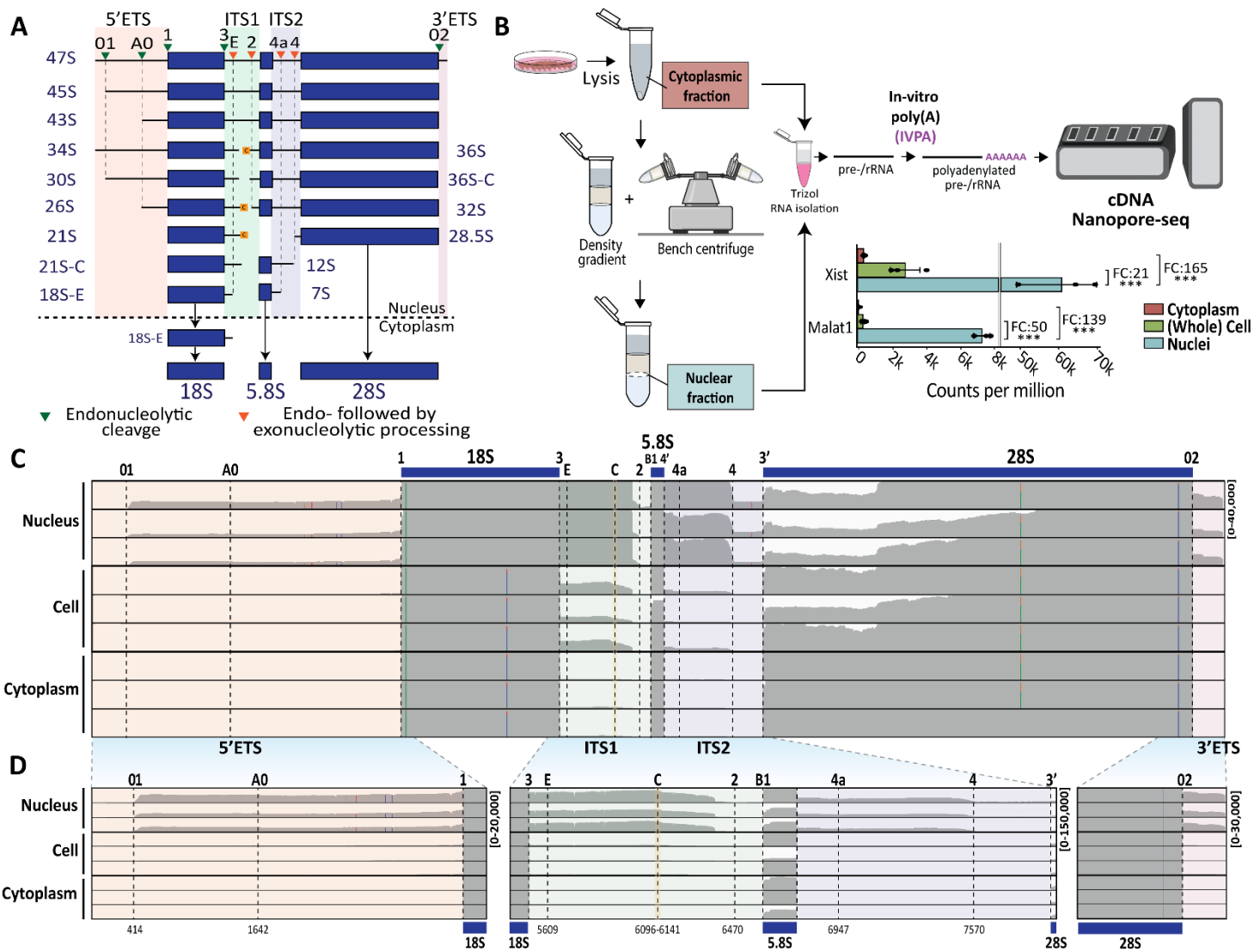


Fig 1: Streamlined nuclei isolation procedure utilized for isolating pre-rRNA.

A, Simplified pre-rRNA processing pathway in human cells. The 5' external transcribed spacer (5' ETS) contains two primary processing sites (01 and A0). Internal transcribed spacer 1 (ITS1) has three main processing sites (3, C, and 2), while internal transcribed spacer 2 (ITS2) has at least two sites (4a and 4). The 3' external transcribed spacer (3' ETS) is also shown. The boundaries for the mature rRNA components are indicated as follows: 18S rRNA (sites 1 and 3), 5.8S rRNA (sites B1 and 4'), and 28S rRNA (sites 3' and 02). The precursors for 18S rRNA biogenesis include, 30S, 26S, 21S, 21S-C, and 18S-E, while the precursors for 28S and 5.8S rRNA include, 32S, 28.5S, 12S, and 7S[8,9,10]. The 36S, 36S-C, and 34S RNAs are associated with ribosome biogenesis perturbations.

B, Nuclei isolation procedure for isolation of nuclear and cytoplasmic RNA followed by cDNA Nanopore-seq. (See Material and methods and supplementary Fig 1 for details). Normalized read count per million of *Xist and Malat1* transcripts in cytoplasmic, whole cell, and nuclear fractions (n=3). Fold change (FC) between the conditions is shown above each comparison. One-way ANOVA followed by Tukey test for multiple comparison post hoc test, ***P < 0.001.

C, IGV coverage profiles of representative nuclear, whole cell and cytoplasmic samples across 47S. Data range was normalized to 40,000 across all samples to visualize the coverage profiles within the selected regions.

D, Zoom-in IGV coverage profiles across 5' ETS and 18S (left), ITS1, 5.8S and ITS2 (middle) and 28S and 3' ETS (right) of representative nuclear, whole cell and cytoplasmic samples across 45SN1. Data range is shown on the right of each figure and was normalized across all samples to visualize the coverage profiles within the selected regions.

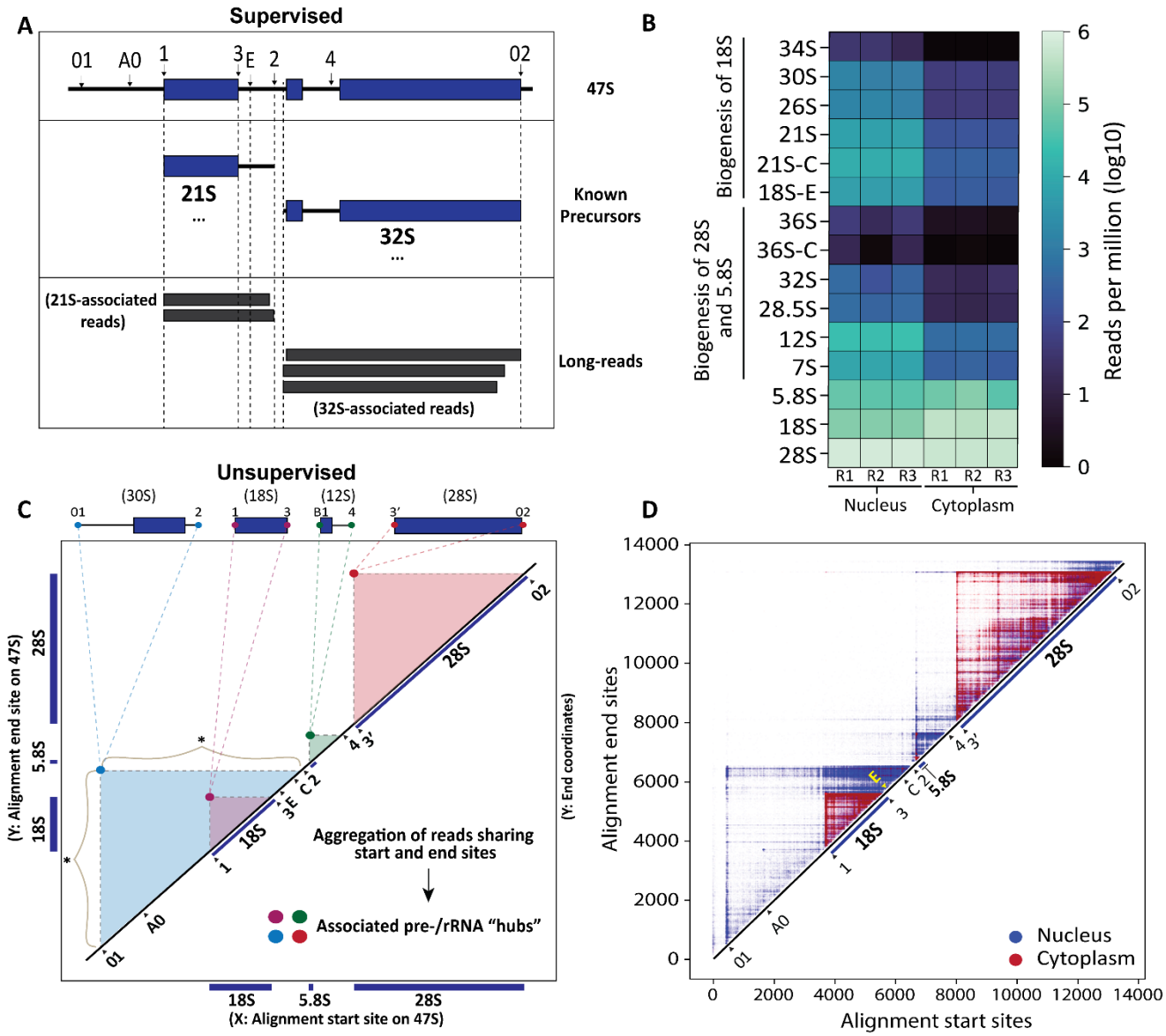


Fig 2. Quantification of rRNA precursors and processing sites using NanoRibolyzer.

A, Simplified overview of supervised (template-based) approach using minimal reciprocal overlap (MRO). Query reads are compared to literature-based intermediates (Fig 1a), and each read is assigned to the intermediate with the highest overlap based on alignment start and end positions. In the example shown, the query reads are closely associated with 21S and 32S precursors. After processing all the reads, the data is presented as a relative quantification score in reads per million, allowing for clear visualization and comparison. See more details in **Figure S3B**.

B, Quantification of detected pre-rRNA intermediates and mature rRNAs in the Nucleus and Cytoplasm (n=3).

C, Simplified illustration of the unsupervised (template-free) approach. A 2-D matrix representing the RNA45SN1 template is constructed, with each rRNA read plotted by its start (x-axis) and end (y-axis) positions. This approach maps transcript boundaries and highlights intensity "hubs," which indicate abundant rRNA products near mature rRNA or putative processing sites. In the example shown, the intensity hubs for 30S, 18S, 12S and 28S correspond to reads clustered at the start and end of the respective pre-/rRNAs. The * symbol denotes 'processing smears,' where a stable start or end site is accompanied by multiple exonucleolytic events at the opposite end of the rRNA product. See more details in **Figure S3C.**

D, Overlayed intensity matrices of nucleus (blue) and cytoplasm (red), highlighting contrasting read distributions: ETS and ITS-associated reads dominate in the nucleus (in blue), while the cytoplasm predominantly contains mature rRNA reads (in red).

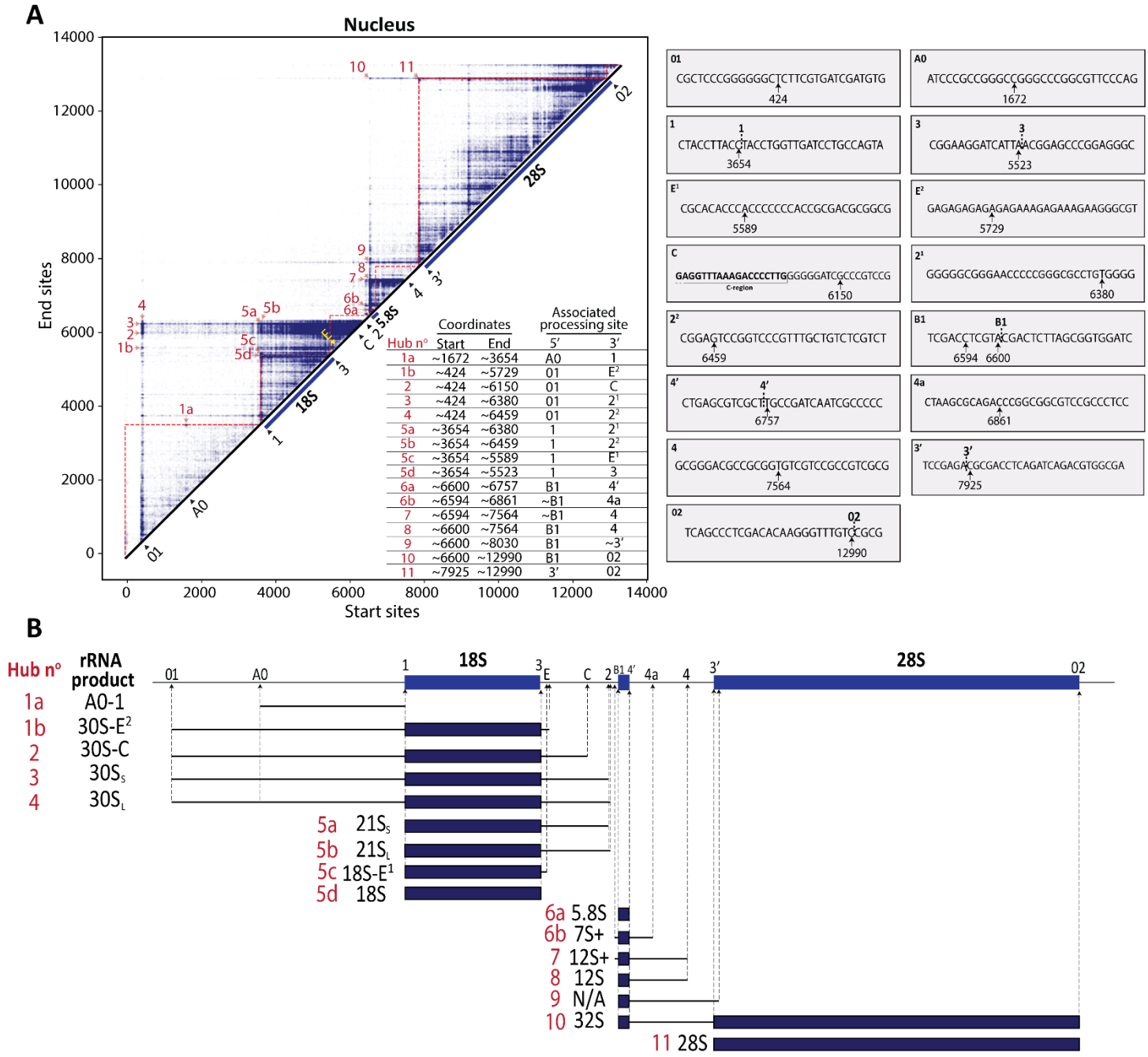


Figure 3. Identification of processing sites and novel precursors with nucleotide resolution.

A, Left, Intensity matrix of nuclear conditions, with highlighted intensity "hubs" numbered in red. The "hubs" were identified based on intensity and closeness to *bona fide* processing sites. Right, The insets provide the of each mapped processing sites. Note that "hubs" exhibiting vertical conservation at the 5' end indicates 3'-5' exonuclease activity, while horizontal conservation at the 3' end suggests 5'-3' exonuclease activity (**See Figure S5A-C**).

B, Illustration of the identified rRNA products in nucleus, as derived from the intensity hubs table shown in 3A. The corresponding rRNA products are shown on the left, alongside the site number (red) and associated precursor (black). In hubs number 6b and 7, the + symbolizes the 5' extension upstream 5.8S sequence.

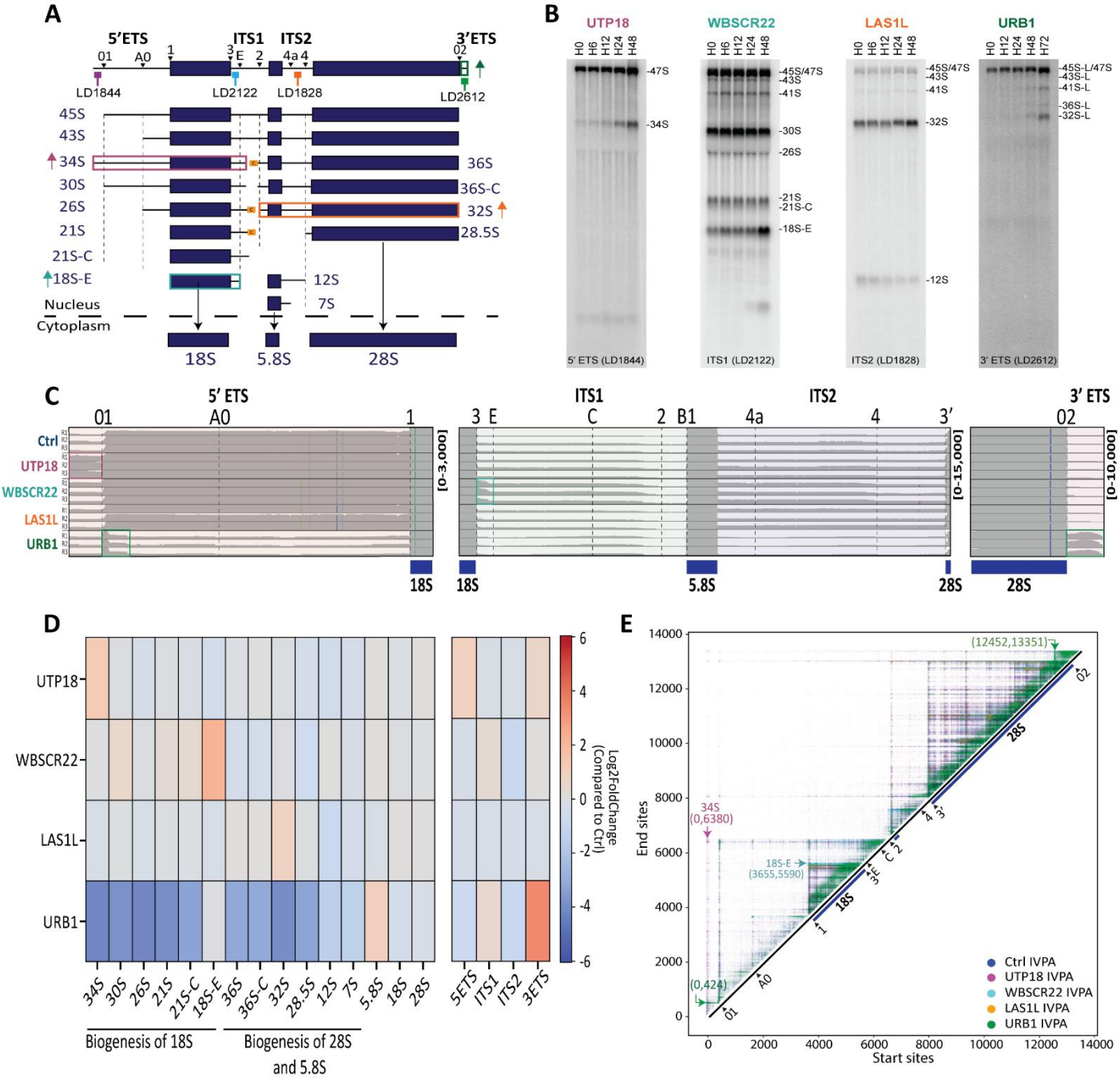


Figure 4. Quantification of pre-rRNAs following processing perturbations.

A, Major effects of knockdown of key factors involved in maturation of spacer regions. Knockdowns of UTP18 (purple, within 5' ETS), WBSCR22 (turquoise, within ITS1), LAS1L (orange, within ITS2), URB1 (green, within 3' ETS) causes the accumulation of 34S, 18S-E, 32S and 3' ETS ribosomal intermediates/spacer region, respectively. Location of probes used in panel b are indicated.

B, Assessment of intermediate accumulation following processing perturbation using northern blotting. Each target was depleted in HEK293 cells in a time course (from 6 to up to 72 h) to identify the best time point for NanoRibolyzer analysis. Based on this analysis, the following depletion time points were selected: UTP18 (48 h), WBSCR22 (48 h), LAS1L (48 h) and URB1 (72 h).

C, IGV coverage profiles of nuclear reads of control and knockdown of UTP18, WBSCR22, LAS1L and URB1 across the 5' ETS (left), ITS1, 5.8S, and ITS2 regions (middle) and 3' ETS (left). The regions affected by the processing perturbation are highlighted in the corresponding colors shown in panel a.

D, Log2FoldChange of quantified pre-rRNA intermediates and mature rRNAs in the nucleus of UTP18, WBSCR22, LAS1L and URB1 knockdown samples compared to control (n=3 each).

E, Overlayed Intensity matrix of nucleus of control, UTP18, WBSCR22, LAS1L and URB1 knockdown samples. Selected intensity hubs are indicated with arrows, displaying their coordinates (start, end) along with the associated precursor, where applicable.

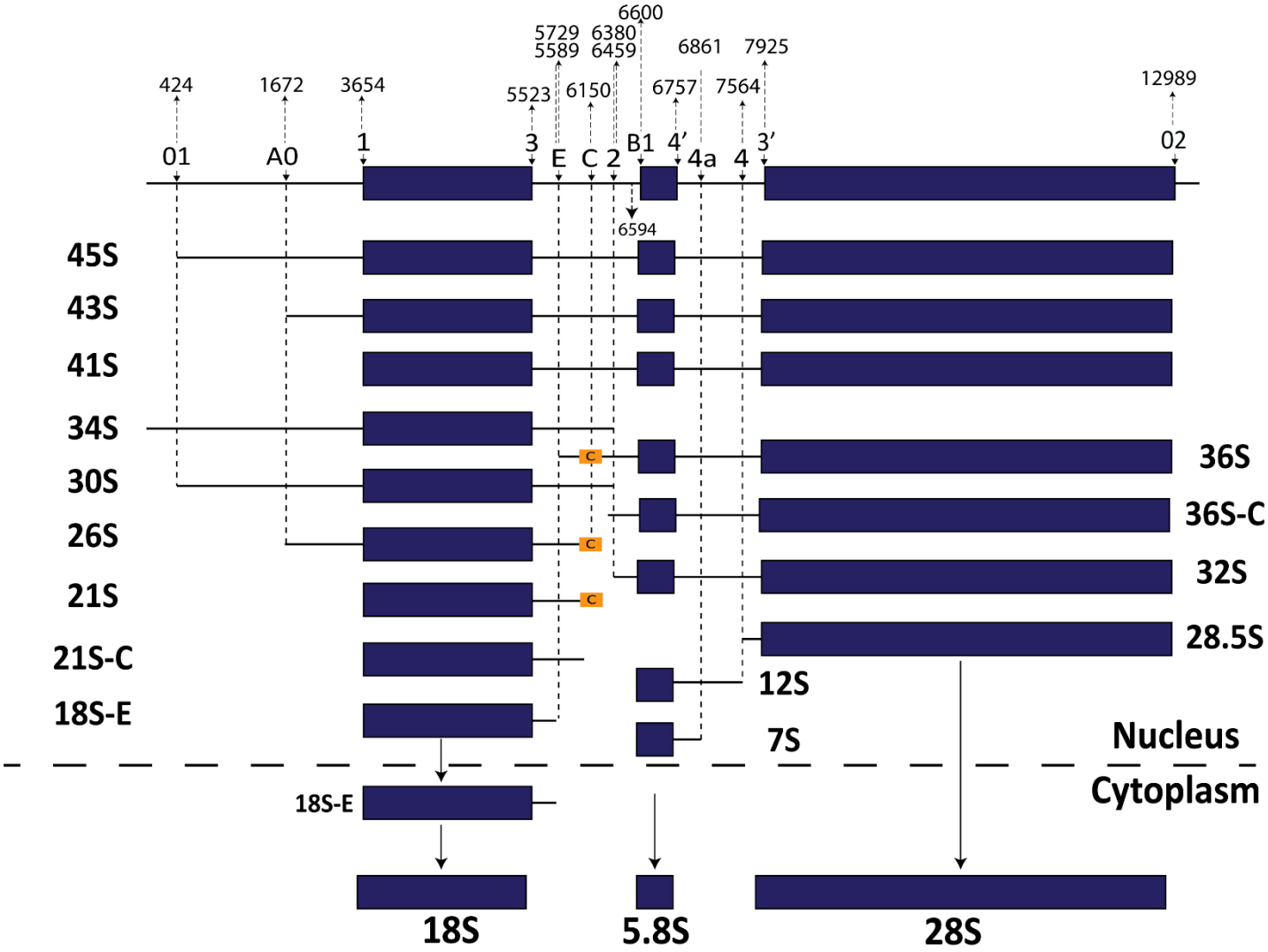
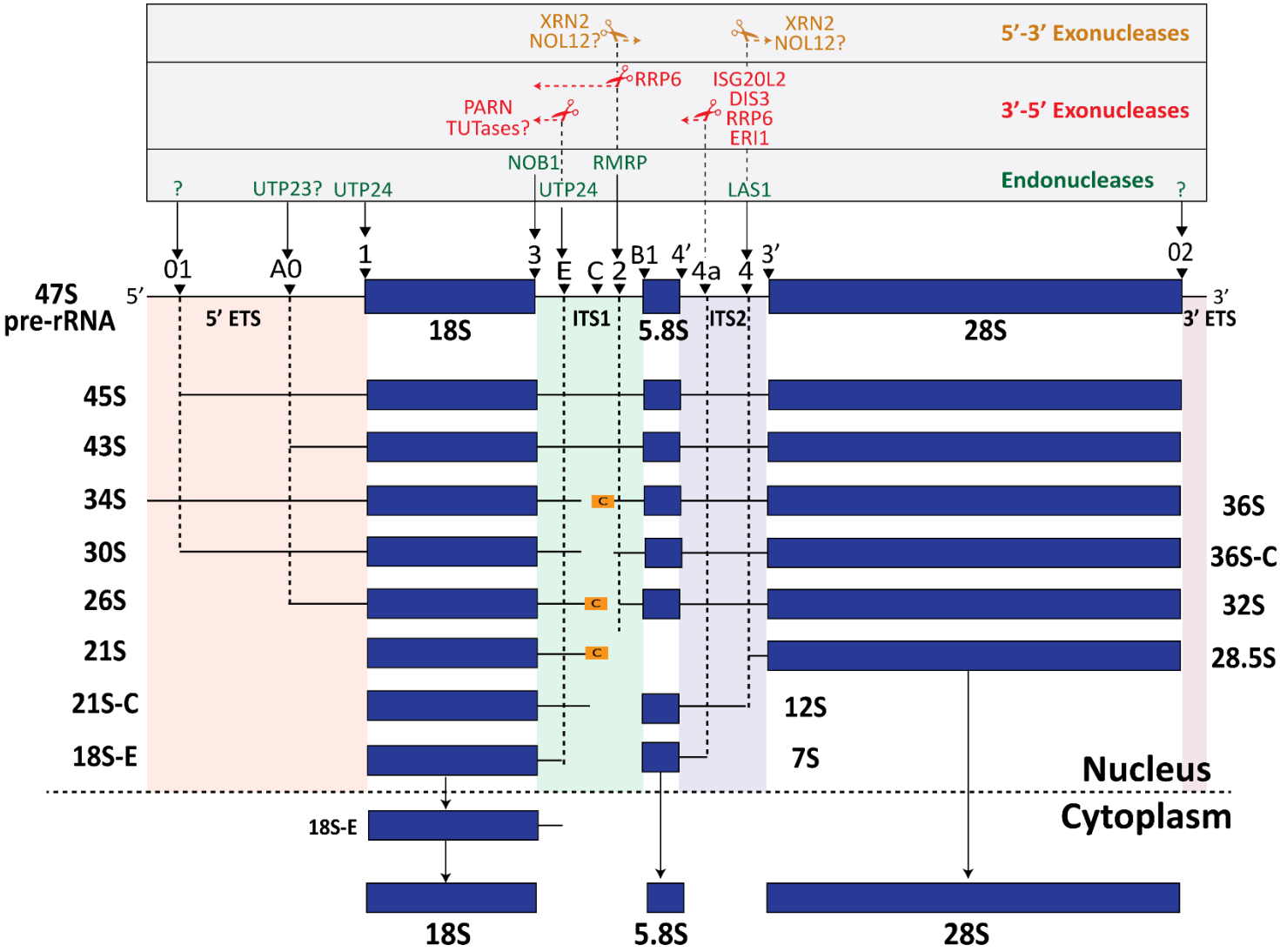


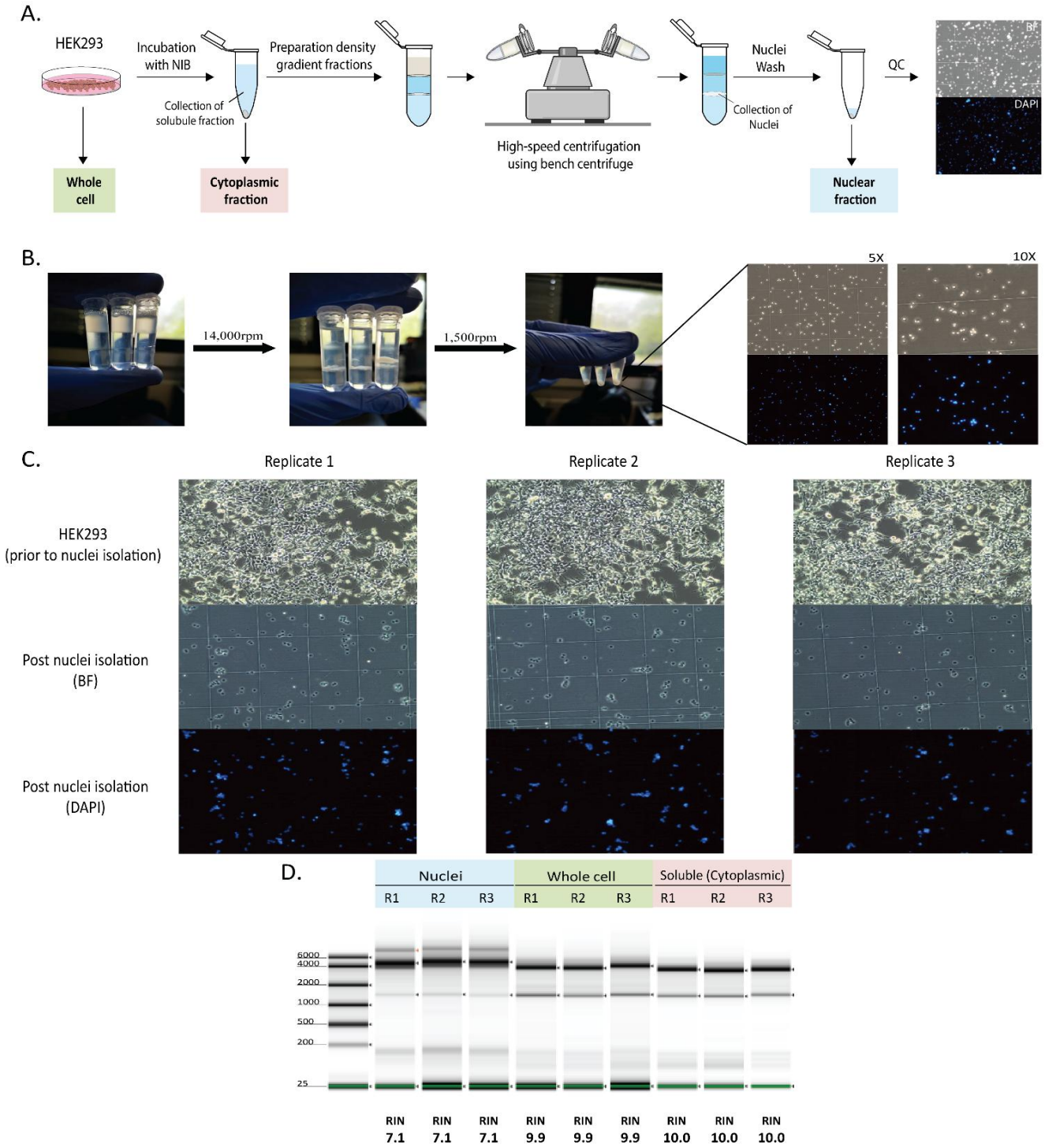
Figure 5. Summary of redefined processing sites in the 47S pre-RNA

Supplementary Figures



Supplementary Figure S1. Overview of human pre-rRNA processing pathway

Schematic overview of processing sites in the human 47S pre-rRNA, summarizing over two decades of research on rRNA processing within the external transcribed spacers (5′ ETS and 3′ ETS) and internal transcribed spacers (ITS1 and ITS2), shown below the 47S template. Enzymes involved in specific processing steps are indicated: endonucleases (green), 3′-5′ exonucleases (red), 5′-3′ exonucleases (orange), and unknown enzymes (?).
The boundaries for the mature rRNA components are indicated as follows: 18S rRNA (sites 1 and 3), 5.8S rRNA (sites B1 and 4'), and 28S rRNA (sites 3' and 02).
Early precursors include 47S, 45S and 43S. The precursors for 18S rRNA biogenesis include, 30S, 26S, 21S, 21S-C, and 18S-E, while the precursors for 28S and 5.8S rRNA include, 32S, 28.5S, 12S, and 7S [8,9,10]. The 36S, 36S-C, and 34S RNAs were associated with dysfunctional subunit biogenesis.



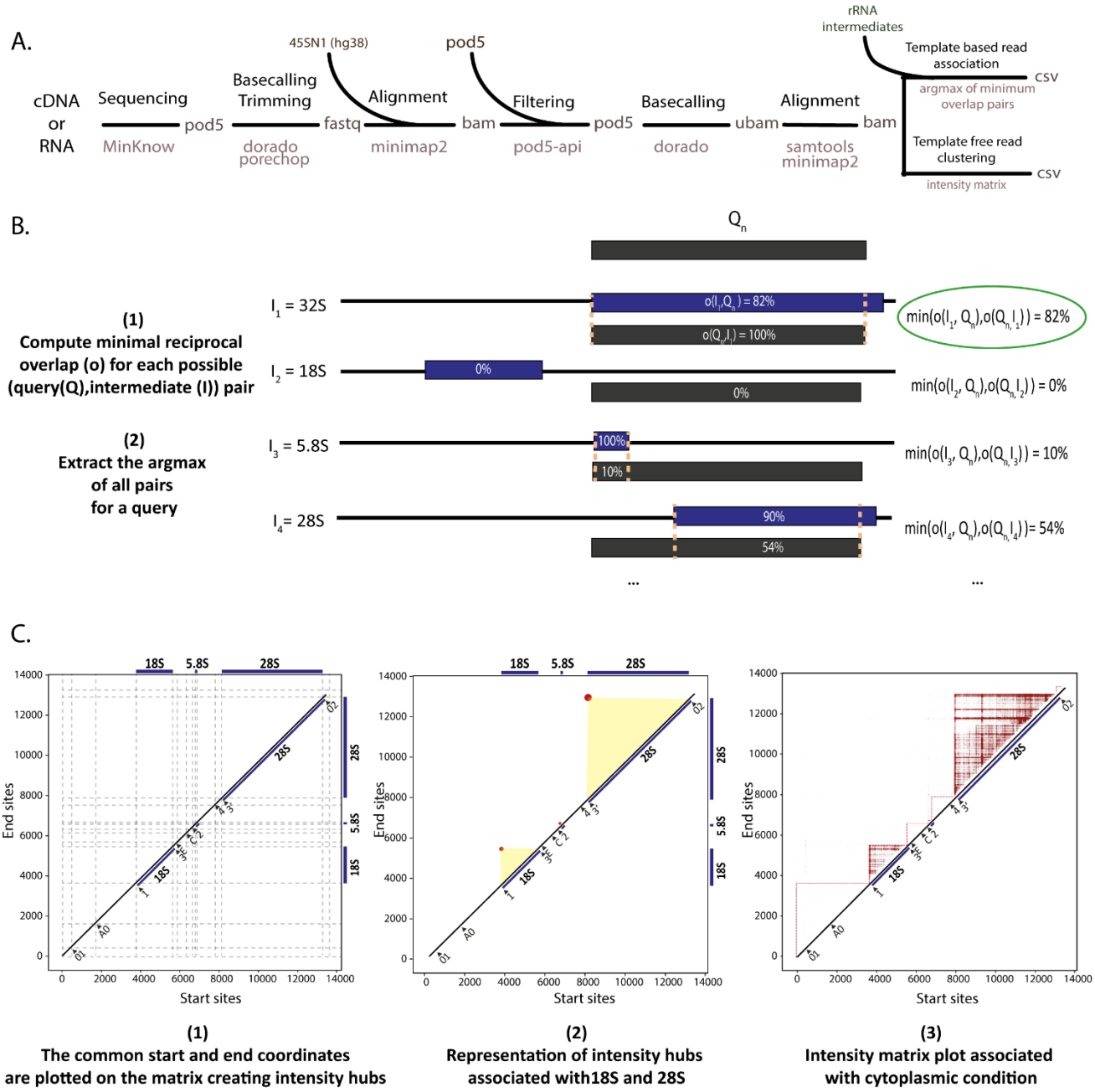
Supplementary Figure S2. Simplified nuclei isolation procedure and quality control steps

A, Detailed schematic diagram of the simplified nuclei isolation procedure. See materials and methods for details. Cells are incubated in nuclei isolation buffer (NIB) for 15 min. Following low speed centrifugation, the soluble fraction (containing the cytoplasmic fraction) is collected and stored for Trizol RNA isolation (light red). Density gradient solutions are prepared from 60% optiprep solution forming three layers where the pellet containing nuclei are placed at the top. Following high-speed centrifugation using a bench centrifuge, the nuclei are collected, washed and assessed under a microscope. Nuclei are then stored in Trizol for RNA isolation (light blue). The whole cell condition is collected at the first step and stored for Trizol RNA isolation as control (light green).

B, Illustration of the 2 ml tubes with density gradient and sample at the top of the tube. Following high-speed centrifugation (~14krpm), the lower interphase (containing the nuclei) is collected and washed using low-speed centrifugation (~1.5krpm). A sample from pelleted nuclei is taken, incubated with DAPI, and analyzed under bright field microscopy with UV filter to visualize the quality of the nuclei.

C, Representative microscope image of HEK293 cells (top) prior nuclei isolation, bright field post nuclei isolation (middle) and DAPI stained nuclei post nuclei isolation (bottom).

D, Tape-station analysis of RNA profiles of cytoplasmic (light red), whole cell (light green) and nuclear (blue) samples (n=3). RIN, RNA integrity number.

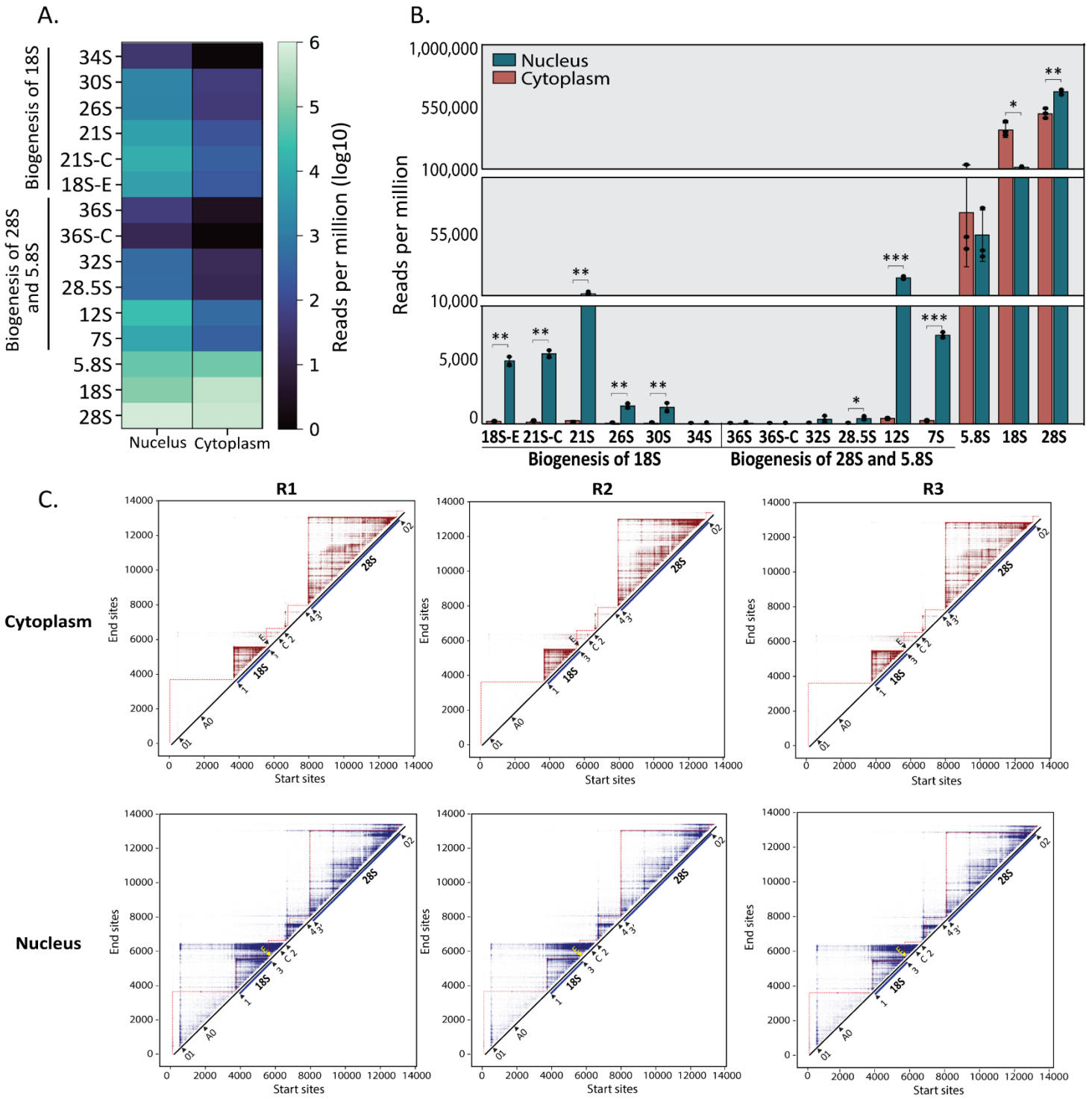


Supplementary Figure S3. Overview of the NanoRibolyzer bioinformatic pipeline

A, NanoRibolyzer pipeline. Reads of input pod5 become basecalled and trimmed. Alignment to 45SN1 of hg38 is used to extract reads aligning to ribosomal RNA. Extracted reads become rebasecalled to perform polyA-estimation and modification detection. Rebasecalled reads are realigned to 45SN1 of hg38. Template-free and template-based read association analysis is performed on realigned reads.

B, Template-based approach. Minimal reciprocal overlap (MRO) between query read and all literature-based intermediates is computed based on alignment start and end sites of the query. The query read is associated to the intermediate with the maximal query/intermediate MRO. Formal: I={Intermediates}, Q={Query reads}, A={Associated query reads}, $Qn \in Q$, A_n=argmax(min(overlap(I_i,Q_n),overlap(Q_n,I_i)) $\forall$ $I_i \in I$).

(C, Template-free approach. A 2-D matrix at the length of the RNA45SN1 template is constructed. Augmentation of reads by start and end site coordinate according to the alignment reveals intensity hubs. Intensity hubs are min-max normalized with a contrast enhancement of 2%.

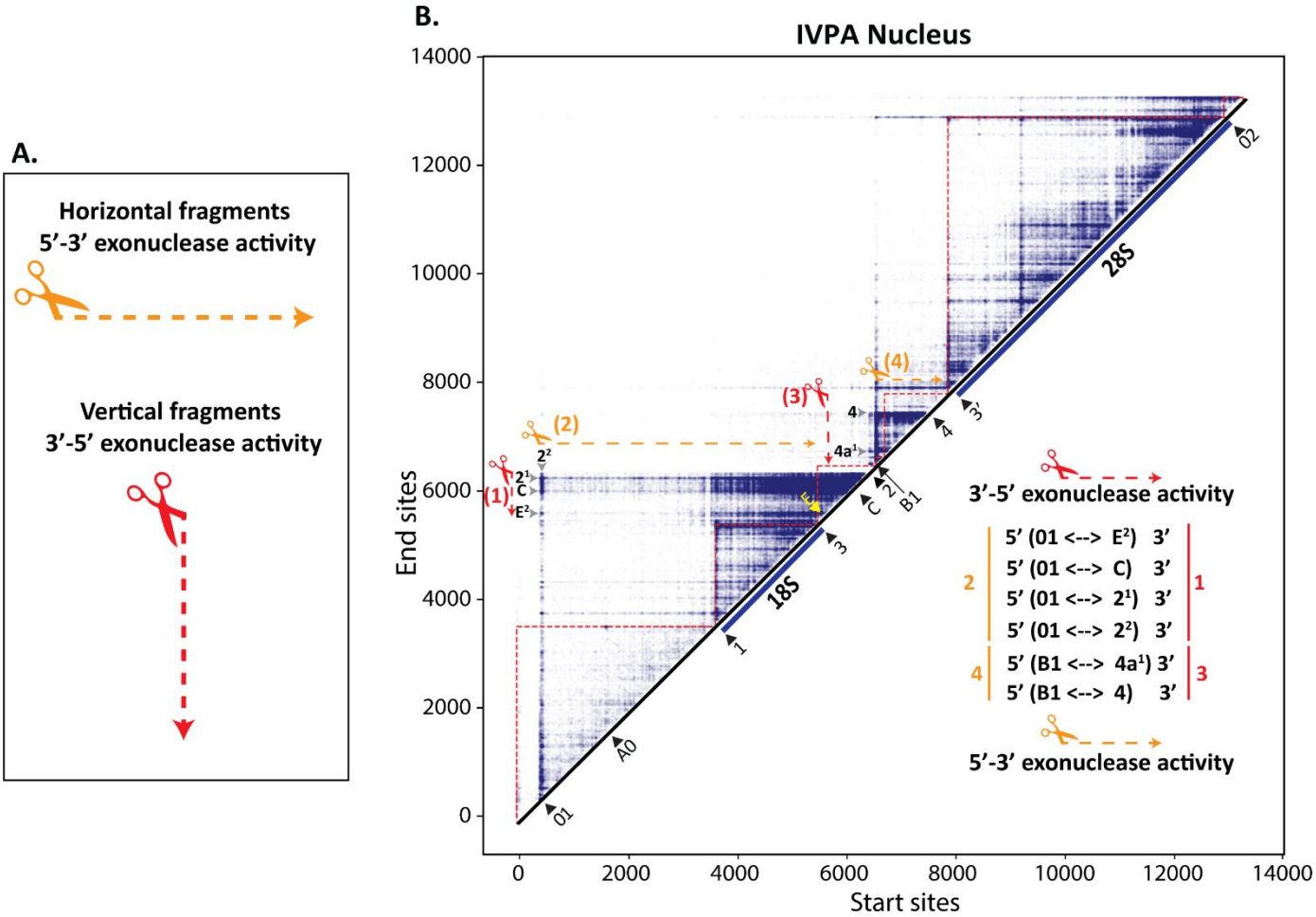


Supplementary Figure S4. Pre-rRNA intermediate visualization and quantification in nucleus, cell and cytoplasm

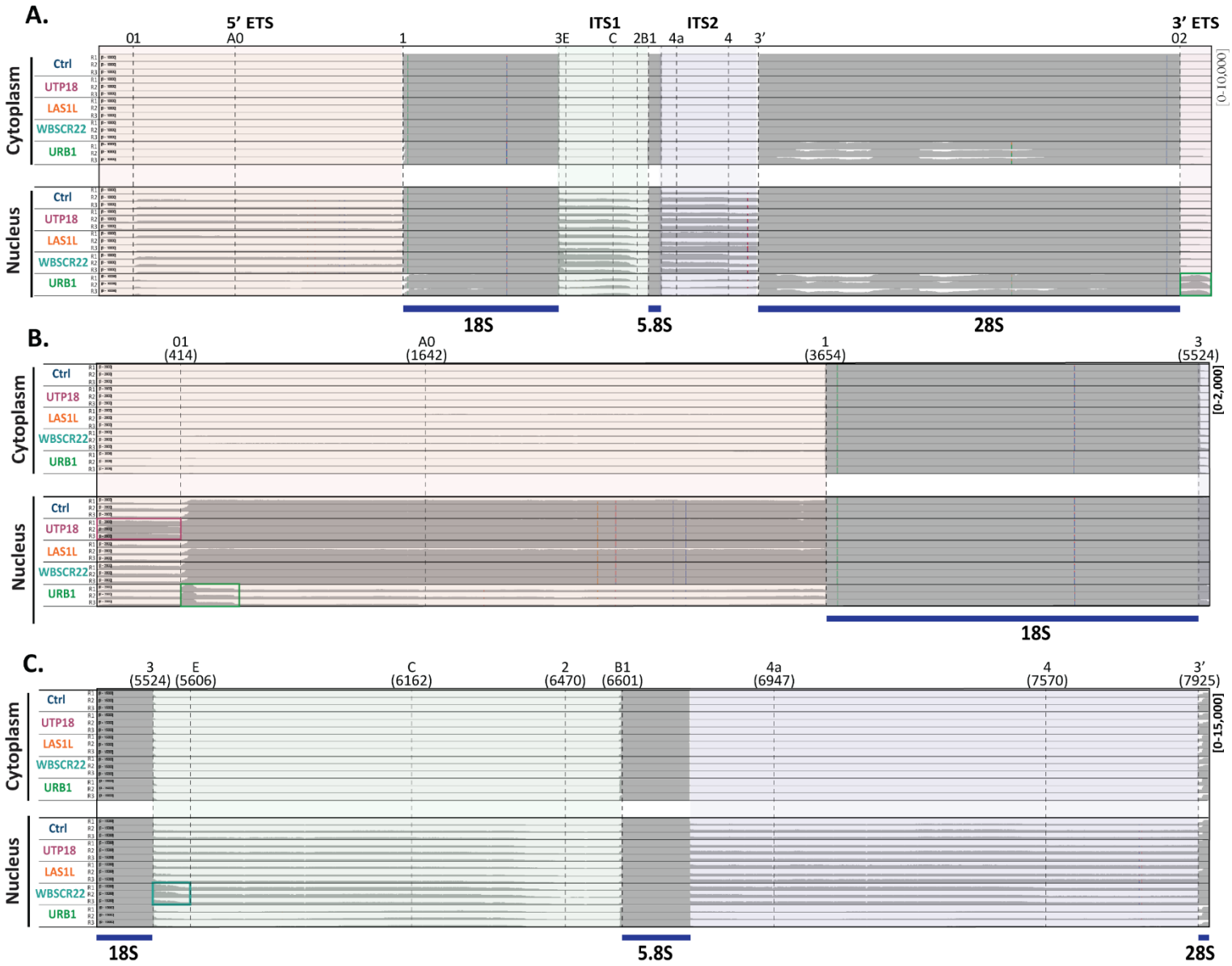
A, Heatmap illustrating pre-/rRNA abundance in nuclear and cytoplasmic fractions, represented as log10 reads per million.

B, Quantitative comparison of pre-rRNA intermediates and mature rRNA between nuclear and cytoplasmic fractions (n=3 each). Histograms display mean reads per million ± SD. Statistical significance was determined using a t-test between nucleus and cytoplasm conditions, with *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

C, Intensity matrices of cytoplasm and nuclei conditions (n=3).

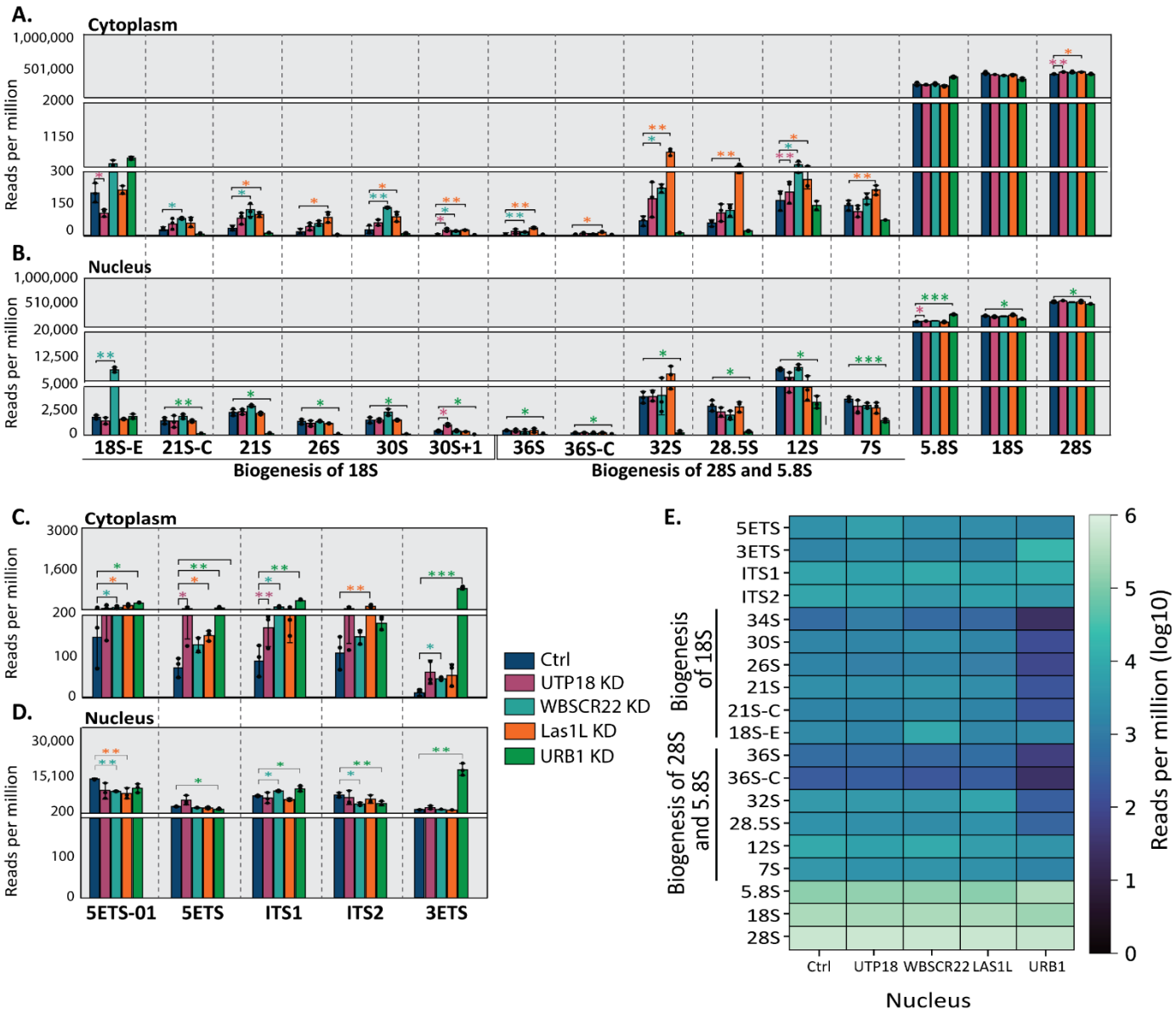


Supplementary Figure S5. Exonucleolytic cleavage identification using the intensity matrix,
A, The intensity matrix provides a clear visualization of exonucleolytic cleavage patterns: Horizontal fragmentation indicates a retained end processing site, with altered start sites, consistent with 5′-3′ exonuclease activity (orange). Vertical fragmentation indicates a retained start processing site, with altered end sites, consistent with 3′-5′ exonuclease activity (red).
B, Retained end processing sites are marked in black within the intensity matrix in the nuclear fraction. The 5’ and 3’ processing sites are indicated on the right with the directionality of exonuclease acitvity colored approripately.



Supplementary Figure S6. IGV snapshots of nucleus/cytoplasm in knockdown samples and Ctrl.

A-C, Coverage profiles of control and knockdown conditions for UTP18, WBSCR22, LAS1L and URB1 in nucleus/cytoplasm (n=3). The profiles are shown across the entire 47S (A), 5' ETS and 18S (B), and ITS1, 5.8S, and ITS2 (C). Data ranges were normalized across all samples to visualize the coverage profiles within the selected regions. Altered regions in the selected knockdowns are highlighted in purple (UTP18 KD), turquoise (WBSCR22 KD) and green (URB1 KD).

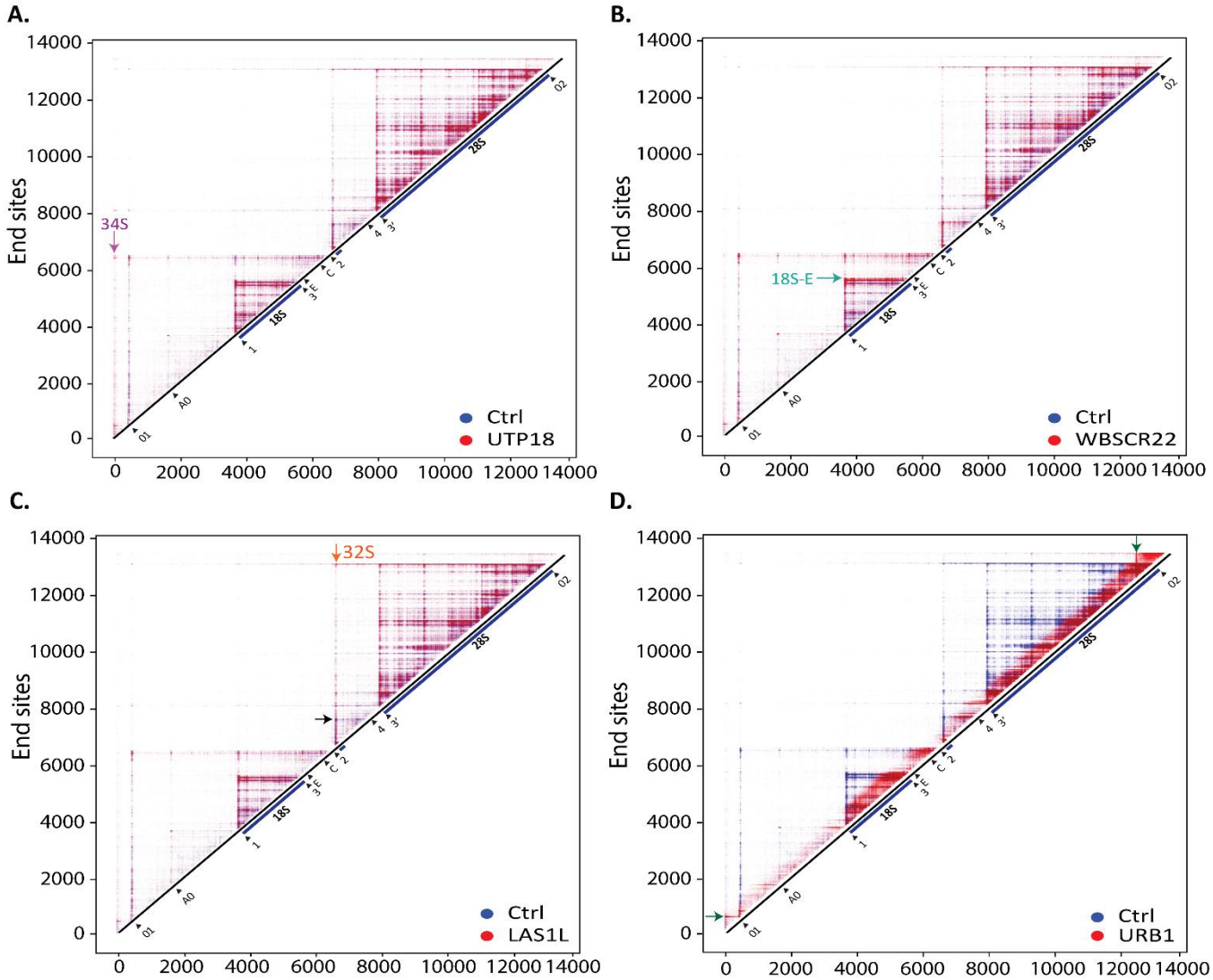


Supplementary Figure S7. Pre-rRNA intermediate quantification and visualization of cleavages sites following processing perturbations

A-B, Quantification of pre-rRNA intermediates and mature rRNAs in the cytoplasm (A) and Nucleus (B) of control and knockdown samples of UTP18, WBSCR22, LAS1L and URB1 (n=3 each).

C-D, Quantification of external and internal transcribed spacers in cytoplasm (C) and nucleus (D) across control and knockdown samples of UTP18, WBSCR22, LAS1L and URB1. Histograms display mean reads per million ± SD. Statistical significance was determined using a t-test between control and KD condition, with *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

E, Heatmap illustrating pre-/rRNA abundance in nuclear fraction of control and UTP18, WBSCR22, LAS1L, URB1 knockdowns samples, represented as log10 reads per million.



Supplementary Figure S8. Pairwise overlay of intensity matrices of processing perturbation samples

A-D,Overalyed matrices are shown in nuclear fraction nucleus between control (blue) and designated KD conditions (red) (**A**-UTP18; **B**-WBSCR22; **C** - LAS1L; D - URB1). Selected intensity “hubs” associated with the effected precursors are indicated with arrows colored according to KD condition shown in Figure 4B. The black arrow in panel C marks processing site 4, the endonucleolytic site of LAS1L.