

# Mapping human pre-rRNA processing and modification at single nucleotide resolution using long read nanopore sequencing

Received: 22 May 2025

Accepted: 13 March 2026

Cite this article as: Pastore, S., Wacheul, L., Lehmann, L. *et al.* Mapping human pre-rRNA processing and modification at single nucleotide resolution using long read nanopore sequencing. *Nat Commun* (2026). <https://doi.org/10.1038/s41467-026-71164-x>

Stefan Pastore, Ludivine Wacheul, Lioba Lehmann, Stefan Mündnich, Beat Lutz, Mark Helm, Susanne Gerber, Denis L. J. Lafontaine & Tamer Butto

We are providing an unedited version of this manuscript to give early access to its findings. Before final publication, the manuscript will undergo further editing. Please note there may be errors present which affect the content, and all legal disclaimers apply.

If this paper is publishing under a Transparent Peer Review model then Peer Review reports will publish with the final article.

# Mapping human pre-rRNA processing and modification at single nucleotide resolution using long read Nanopore sequencing

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

<sup>1</sup>Institute of Pharmaceutical and Biomedical Sciences, Johannes Gutenberg-University Mainz, Mainz 55128, Germany.

<sup>2</sup>RNA Molecular Biology, Fonds de la Recherche Scientifique (F.R.S./FNRS), Université libre de Bruxelles (ULB), Biopark campus, B-6041 Gosselies, Belgium.

<sup>3</sup>Institute of Human Genetics, University Medical Center of the Johannes Gutenberg University Mainz, Mainz 55131, Germany.

<sup>4</sup>Leibniz Institute for Resilience Research (LIR), 55122 Mainz, Germany.

<sup>5</sup>Institute of Physiological Chemistry, University Medical Center Mainz, 55128 Mainz, Germany.

<sup>6</sup>Institute for Quantitative and Computer Biosciences (IQCB), Mainz 55128, Germany.

\*Co-lead contacts

Correspondence: [buttamer@uni-mainz.de](mailto:buttamer@uni-mainz.de)

Correspondence: [denis.lafontaine@ulb.be](mailto:denis.lafontaine@ulb.be)

## Abstract

Ribosome biogenesis requires the synthesis and sequential processing of precursor rRNAs (pre-rRNAs) into mature rRNAs. Traditional methods such as northern blotting and metabolic labeling provide limited resolution. Here, we present NanoRibolyzer, a nanopore-based long-read sequencing approach that enables *ab initio* identification and quantification of rRNA precursors while simultaneously mapping RNA modifications. Using supervised and unsupervised mapping, we detect both known and previously uncharacterized pre-rRNAs and delineate cleavage events at single-nucleotide resolution. A simple cell-fractionation protocol further separates nuclear and cytoplasmic pre-rRNAs, allowing spatial deconvolution of processing pathways. By projecting each sequenced molecule in a two-dimensional space using its starting and ending coordinates, we generate an intuitive representation in which the activity of the 5'→3' and 3'→5' exoRNases can be tracked as they mature pre-rRNAs one nucleotide at a time. Targeted knockdowns of ribosome-assembly factors quantify accumulation of intermediates and reveal condition-specific processing “fingerprints” with biomarker potential. High-resolution re-analysis of known factors uncovers unexpected functions. Additionally, pseudouridine mapping shows that the primary 47S transcript is extensively modified, whereas aberrant intermediates (34S and 36S-C) are hypomodified. With its high resolution and unique discovery mode, NanoRibolyzer provides new insights into rRNA processing and modification, greatly advancing our understanding of ribosome biogenesis.

## Introduction

Ribosomes are ribonucleoprotein nanomachines responsible for protein synthesis in all living cells<sup>1,2</sup>. 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)<sup>3,4</sup>. The 47S contains sequences for three out of four rRNAs (the 18S, 5.8S, and 28S) interspersed with non-coding spacers (**Fig 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<sup>5</sup>. 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 (**Fig 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<sup>4,6,7</sup>.

More than two decades of research have led to major advances in identifying discrete processing sites and pre-rRNA intermediates, which now serve as critical markers for assessing ribosome biogenesis efficiency (**Fig 1A** and **S1**). Knowledge of these intermediates has been particularly valuable for investigating aberrant precursor production that arises during processing perturbations<sup>2</sup>. 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<sup>6,8,9,10</sup>. However, these assays require important input material (often  $\mu\text{g}$  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 poor genome annotations making it difficult to accurately map reads and distinguish between individual rDNA copies, especially with short-read sequencing technologies<sup>11,12,13,14</sup>.

Nanopore sequencing (nanopore-seq) has emerged as a promising technology to investigate ribosome biogenesis<sup>15,16,17</sup>. 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 their modifications<sup>18,19</sup>. These capabilities are particularly valuable for studying rRNA intermediates, which can vary significantly in length and abundance, and possibly modification levels. 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 integrates state-of-the-art long-read nanopore sequencing with advanced bioinformatics to achieve spatially resolved, single-nucleotide analysis of pre-rRNA intermediates. Using a streamlined nuclei isolation protocol, we systematically profile precursor and mature rRNA species in both nuclear and cytoplasmic compartments. We further introduce precursor-specific modification analysis, uncovering the spatio-temporal dynamics of rRNA modifications. Overall, NanoRibolyzer enables comprehensive detection and quantification of both known and novel processing intermediates, including cleavage events generated by endo- and exoribonucleolytic activity, as well as pseudouridine modifications.

## Results

### Nuclear isolation enabling high-resolution mapping of pre-rRNA precursors

Ribosome biogenesis begins in the nucleolus, a multiphase condensate within the nucleus, continues in the nucleoplasm, and is finalized in the cytoplasm, where mature rRNAs accumulate. In order to develop a nanopore-sequencing approach to study pre-rRNA processing, we adapted a straightforward isolation protocol involving density gradient separation using a simple benchtop centrifuge (**Fig 1B** and **S2**). This provided spatial resolution of processing events and access to purified nuclear RNA enriched in low-abundance, short-lived pre-rRNAs that are otherwise masked in total RNA by the highly abundant mature cytoplasmic rRNAs. 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') (**Fig 1B** and **S2A**).

After separation of the fractions, a quality assessment of isolated nuclei was conducted using DNA staining with DAPI, revealing debris-free and round intact nuclei<sup>20</sup> (**Fig 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) (**Fig S2D**). In contrast, the nuclear fractions exhibited higher molecular weight species, at the expected size for pre-rRNAs (**Fig S2D**). Northern blot analyses performed throughout the study consistently verified that precursor rRNA is predominantly retained within nuclear fractions, in contrast to cytoplasmic and whole-cell fractions (**Fig S3, Source Data 1**).

Since Nanopore-based RNA library preparation strategies typically capture poly(A)+ RNA<sup>21</sup>, we applied an *in-vitro* polyadenylation strategy followed by long-read cDNA sequencing (**Fig 1B**). To assess enrichment of transcripts in the nuclear fractions, we quantified two nuclear long non-coding RNAs (lncRNAs): XIST and MALAT1 (**Fig 1B, Supplementary Table S1**). 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).

Compared with whole-genome alignments, which are challenging to interpret for rRNAs due to presence of hundreds of rDNA gene copies, NanoRibolyzer aligns long rRNA reads to a single 47S reference template (45SN1; GeneID: 106631777). This approach provides a clear and unambiguous view of reads corresponding to pre-rRNAs and mature rRNAs. We generated an average of ~3 million reads per sample with an average ~75% alignment rate to the 47S template. Our bioinformatic pipeline is depicted in **Fig S4A**. For reference and to initiate precursor quantification analysis, we retrieved the positions of known processing sites and major precursors from literature<sup>8,9,10</sup> (**Fig 1A** and **S1**).

We applied NanoRibolyzer to unperturbed HEK293 cells and characterized reads mapping to mature rRNA sequences and the non-coding spacers (5' ETS, ITS1, ITS2 and 3' ETS) for the nuclear, cytoplasmic, and whole cell fractions (**Fig 1C-D**). As expected, the coverage profiles revealed that the nuclear fraction had higher coverage across the spacers (ETSs and ITSs) compared to the cytoplasmic ones, with the whole cell fraction displaying an intermediate coverage level (**Fig 1C-D**). The start and ends of these reads corresponded to known processing sites, e.g. sites 01, 1, 2, 4, etc. (see Refs<sup>8,9,10</sup>).

These observations agree with the notion that the nuclear fraction predominantly contains ribosomal assembly intermediates, in line with the fact that most steps of ribosome biogenesis occur in the nucleolus. In our initial analyses, we consistently observed that whole-cell fractions are largely redundant with cytoplasmic fractions (see **Fig 1D**, zoomed IGV). Therefore, throughout the remainder of the manuscript, and unless stated otherwise, we focus on comparisons between nuclear and cytoplasmic fractions.

We also noted that the hyper modified nucleotide 1-methyl-3- $\alpha$ -amino- $\alpha$ -carboxyl-propyl pseudouridine ( $m^1acp^3\Psi$ ) at position 1248 on the 18S rRNA was detected in whole cell and cytoplasmic fractions, but strikingly not in nuclear ones (**Fig 1C**, arrow), consistent with the fact that  $m^1acp^3\Psi$  formation is completed in the cytoplasm<sup>22</sup>.

## Mapping rRNA precursor diversity and processing pathways

Once we confirmed that nuclear sequencing reads extends across spacer regions, we employed two complementary approaches to quantify pre-rRNAs and resolve processing sites: 1) a rather classical supervised or ‘template-based’ strategy, and, 2) an unsupervised or ‘template-free’ method (**Fig S4**).

The supervised approach detects pre-rRNA intermediates and spacer fragments described in the literature, quantifying their relative abundance by maximizing reciprocal overlap (**Fig 2A** and **S4B**). These intermediates and fragments are defined based on published processing sites (**Fig 1A** and Refs<sup>1,2</sup>). Quantified data can be visualized either as heatmaps of averaged log<sub>10</sub> reads per million (**Fig 2B** and **S5A**) or histograms illustrating averaged reads per million for each condition with associated standard deviations and p-values (**Fig S5B**).

Using the supervised approach, as expected, we found that most pre-rRNA intermediates are substantially more abundant in the nucleus than in the cytoplasm (**Fig 2B** and **Fig S5A-B**; statistical analysis in **Source Data 1**, p-values on histograms). This analysis also highlights that several intermediates historically associated with ribosome biogenesis defects<sup>6</sup>, such as the 34S, 36S, or 36S-C, are not, or only marginally, produced in unperturbed cells (see, dark shade of blue in the heatmap of **Fig 2B** and **S5A**). As we will show later, when mapping rRNA modifications at the level of individual pre-rRNA precursors, these aberrant species are, strikingly, not faithfully modified. The analysis also reveals that precursors associated with early steps of large ribosomal subunit biogenesis (32S) are not detected in the cytoplasm, while precursors associated with late maturation of the small subunit (21S, 21S-C, 18S-E) and the large subunit (12S, 7S) are partially present in the cytoplasm (**Fig 2B** and **Fig S5A-B**).

Together, these results demonstrate that NanoRibolyzer provides a quantitative framework to assess the relative abundance of all major pre-rRNA precursors as well as mature rRNAs.

Having assigned long-read sequences to specific pre-rRNA precursors and mature rRNAs, we next sought to pinpoint their predominant start and end positions, thereby defining processing sites. To do so, we computed significant 5’ and 3’ boundaries from the template-based BAM files, using the 45SN1 (hg38) reference sequence. Sites occurring at least two standard deviations above the mean were classified as significant, and results were exported in TSV and BED formats for visualization (**Fig S5C**). To ensure robustness, we retained only those sites present in at least two of the three biological replicates, enabling us to identify the most reproducible processing sites and quantify their prevalence across the 47S sequence (**Source Data 1**).

Inspection of the cleavage sites revealed consistent positions within the spacer regions and at the boundaries of the mature rRNAs (**Fig S5D**; start sites in green, end sites in orange). Cleavages detected in the spacer regions, particularly those proximal to previously reported processing sites, were considered *bona fide* processing events (**Fig S5D-F**; see also **Fig 1A** and **Fig S1**). In several cases, using the literature information alone was insufficient to recognize with single nucleotide precision the RNA end points. As discussed below, this limitation will be solved by using our complementary unsupervised mapping approach (**Fig 2C-D**, **Fig 3**).

As expected, samples from the cytoplasm, where pre-ribosomes are nearly fully matured and where mature ribosomes are engaged in translation or storage, showed cleavage sites mostly restricted to the extremities of mature rRNAs (**Fig S5D**). In sharp contrast, nuclear samples displayed numerous internal cleavages within the spacer regions, reflecting ongoing pre-rRNA processing (**Fig S5E-F**).

We also found numerous cleavages within mature rRNA sequences (**Fig S5D**). These aberrant events, which generate by-products destined for degradation, have been greatly underappreciated using classical approaches such as northern blotting. Their relevance becomes clear below, where we show they can define condition-specific “fingerprints” (e.g. upon URB1 or RPL3 depletion, shown below), with potential biomarker value.

In conclusion, the supervised component of NanoRibolyzer performs very well to quantify the abundance of RNA species previously described in the literature and, with few exceptions, to map their ends with single nucleotide resolution. In essence it performs quite similarly to northern-blotting, except that, it has additionally started to reveal an under-appreciated level of cleavage within mature sequences.

Nonetheless, the supervised approach is not suitable to map the low abundant metastable (transient) species which results from the progressive exoribonucleolytic cleavage,  $5' \rightarrow 3'$  and  $3' \rightarrow 5'$ , which are known to be of paramount importance in processing, generating thousands of intermediates. To put simply, a huge diversity of species still escape characterization.

To address this limitation, we next implemented an unsupervised approach that provides a completely unbiased mapping strategy, i.e. a method that does not rely on any preconceived knowledge about pre-rRNA processing. We also developed a new display to apprehend processing intuitively.

In this approach, every single individual read sequenced is projected onto a two-dimensional matrix (**Fig 2C-D** and **Fig S4C**). The x- and y-axes extend from the transcription start site to the termination site of the primary 47S transcript, respectively. Each rRNA read is positioned on the matrix according to its 5' and 3' coordinates, enabling precise, reference-free mapping of transcript boundaries (**Fig 2C-D** and **Fig S4C**). This visualization reveals intensity “hubs” corresponding to abundant rRNA species whose ends coincide with established or putative novel processing sites (**Fig 2C-D** and **Fig S4C**; see Methods for details).

We applied the unsupervised approach to nuclear and cytoplasmic fractions of unperturbed HEK293 cells to visualize the unbiased aggregation of reads associated with the 47S transcript. In the nuclear fraction, we observed prominent accumulations of reads within the ETS and ITS regions, corresponding to processing sites and appearing as characteristic blue “pyramids” in the matrix (**Fig 2D**). For example, striking pyramids were detected between site 01 in the 5' ETS and site 2 in ITS1, defining precursors of 18S rRNA (see below), and between the 5' end of 5.8S and site 4 in ITS2, corresponding to 3'-extended 5.8S precursors destined for  $3' \rightarrow 5'$  trimming by the multi-subunit RNA exosome<sup>6,23</sup> (zoomed inset in **Fig 2D**).

Remarkably, this approach also visualizes, the activity of  $5' \rightarrow 3'$  and  $3' \rightarrow 5'$  exoRNases as they progressively trim spacer sequences one nucleotide at a time (**Fig S6A-B**). Low-abundance, metastable intermediates generated during this process, which are difficult to detect with conventional methods, align as continuous signal traces in the matrix: horizontal lines reflecting  $5' \rightarrow 3'$  exonucleolytic trimming, and vertical lines corresponding to  $3' \rightarrow 5'$  exonucleolytic activity (**Fig S6A-B**). Remarkably, the positions where the exoRNases pause can be seen as more intense dots.

In contrast, reads from the cytoplasmic fraction (shown as ‘red pyramids’, in **Fig 2D**, **Fig S6C**) mapped almost exclusively to the three mature rRNAs, 18S, 5.8S and 28S (**Fig 2D**; see inset for 5.8S). The top of each pyramid corresponds to the actual mature RNAs, whereas the interior signal to abundant degradation products. Again, such abundance of mature rRNA fragments was truly unexpected and was not characterized before. Individual matrices for all three biological replicates are shown in **Fig S6C**, underscoring the robustness of the analysis.

In conclusion, combining supervised and unsupervised mapping approaches not only exploits existing knowledge of abundant, well-characterized pre-rRNA intermediates, but also enables the discovery of previously undetected, low-abundance and metastable processing intermediates as well as degradation by-products. Our analysis reveals the progression of exoRNases on pre-rRNAs, highlighting stalling

positions. Finally, we also uncover an exceptionally rich and largely overlooked population of aberrantly cleaved species. As shown below, these aberrant products carry powerful biological information: depletion of factors acting at the same step of ribosome biogenesis yields highly similar patterns of cleavage by-products (**Fig 4E**), effectively generating condition-specific molecular fingerprints.

### 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 (**Fig 3A**). These coordinates should be considered estimates, as minor RNA species often show variability in their end points. We therefore bioinformatically selected the most predominant end points located near putative processing sites, using a threshold of at least ten reads per hub.

Analysis of the intensity matrix revealed obvious cleavages, which could all be mapped to single nucleotide resolution, many were associated with known processing sites<sup>8,9,10,24-32</sup> and there were also new sites. The detailed description of all processing sites mapped, novel and known, is provided in **Table 1** and **Fig S7**. Briefly, we mapped thirteen known processing sites to single-nucleotide resolution, revealing multiple endpoints at two sites ( $E^1$  and  $E^2$ , and  $2^1$  and  $2^2$ ). For sites already known in the literature, there was an excellent agreement between sites identified by the supervised and unsupervised approaches (**Fig S5D-F**).

In summary, we have confirmed all known processing sites and identified new sites to single nucleotide resolution, providing a powerful resource to the community for years to come.

### NanoRibolyzer enables quantification of novel pre-rRNAs

The intensity hub mapping strategy (unsupervised approach) led to the identification of novel processing sites, and we next inspected the products resulting from cleavage at these sites (**Fig 3B**).

We draw several conclusions:

- (1) the detection of precursors starting at site 01, in the 5' ETS, and ending at different positions in ITS1 (including at site 3, E, C,  $2^1$  and  $2^2$ ) indicates that cleavage in ITS1 can occur at least to some extent prior to cleavage at site A0 (**Fig 3B, Fig S7**; Hubs n° 1b and 2-4).
- (2) we confirmed that the 21S corresponds to two isoforms<sup>30</sup>: 21S<sub>s</sub> (small) and 21S<sub>L</sub> (large) (**Fig 3B, Fig S7**; Hubs n° 5a-b, and **S6B**).
- (3) we confirmed that maturation of the 3' end of 5.8S involves progressive exoribonucleolytic trimming of 12S to 7S, and ultimately to 5.8S (**Fig 3B, Fig S7**; Hubs 6b-7 and **S6B**), a process known to involve substrate handover between RNA exosome subunits<sup>6</sup>.

Having identified new cleavage sites and intermediates using the unsupervised model (**Fig 3B**), we then incorporated this information into the supervised approach to further refine our processing analysis and to formally quantify the newly detected intermediates through reciprocal overlap maximization (see above and **Fig S4B**).

This analysis revealed that the shorter pre-rRNAs are systematically more abundant than the larger pre-rRNAs from which they derive, as generally expected in a precursor-product relationship. This was true in the pathways of production of 18S, 5.8S, and, also of 28S (**Fig 3C, Fig S8A-B**).

To further illustrate how NanoRibolyzer can easily be adapted to detect and quantify new pre-rRNA species, we focused on 3' end maturation of 18S rRNA (**Fig S8C**). The biogenesis of 18S rRNA 3' end is particularly interesting because it initiates in the nucleoplasm, it ends in the cytoplasm, and it involves formation of a succession of metastable forms: the 18S-E extended respectively by 10, 24, 36, 40, and 78 nucleotides,

which are progressively trimmed to 18S-E and finally 18S (ref.<sup>26</sup>). Once we incorporated this information in our supervised approach (**Fig S8C**), it revealed the respective abundance of each intermediate across the nucleus and cytoplasm as the precursors mature across the nuclear pore complexes (**Fig S8D**). Depletion of ribosomal protein RPS26 (**Fig S9A-C**), which is known to block the final cytoplasmic steps of 18S rRNA maturation while leaving nuclear processing unaffected<sup>33-34</sup> indeed led to accumulation of the 18S-E+24, +36 and +40 in the cytoplasmic fraction but not in the nucleus (**Fig S8E-F and Fig S9**).

### NanoRibolyzer captures pre-rRNA processing perturbations

Having demonstrated the ability of NanoRibolyzer to capture pre-rRNA processing in unperturbed cells, we next aimed to evaluate how efficiently our method can detect processing perturbations. To achieve this, we selected key processing factors involved in the maturation of each spacer sequence and depleted them (**Fig 4A** and Fig S3).

For the 5' ETS, we targeted UTP18, a component of the SSU-processome, whose depletion results in accumulation of the aberrant 34S species (Ref.<sup>35</sup>). For ITS1, we selected DIMT1L and WBSCR22, whose loss results respectively in accumulation of the 21S/21S-C and the 18S-E pre-rRNAs<sup>36,37</sup>. For ITS2, we selected the endoRNase LAS1L, whose absence causes accumulation of 32S pre-rRNA<sup>38</sup>. Lastly, for the 3' ETS, we chose URB1, whose importance for 3' ETS removal has been demonstrated<sup>39</sup>. Ribosomal protein RPL3 that binds near the end of 28S rRNA but whose role in 3' ETS processing had not been established was also tested.

First, we assessed the depletion of each factor using northern blots, confirming the expected precursor accumulation (**Fig 4B** and **Fig S10-15**). For NanoRibolyzer analysis, we chose the optimal depletion time point for each factor: 48 h for UTP18, WBSCR22, and LAS1L, and 72 h for DIMT1L, URB1 and RPL3 (**Fig 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 (**Fig 4C** and **S16A-C**). The effects observed by northern blotting were all confirmed by nanopore sequencing (**Fig 4B-F** and **Fig S16**).

The depletion of UTP18 led to a substantially increased coverage upstream of site 01, corresponding to the 34S species (**Fig 4C**, purple rectangle, and **S16B**). Knocking down DIMT1L or WBSCR22 increased coverage in ITS1 downstream of the 3' end of 18S, corresponding to 21S/21S-C (in case of DIMT1L, dark brown rectangle) and 18S-E accumulation (WBSCR22, turquoise rectangle), respectively (**Fig 4C** and **S16C**). Depletion of LAS1L revealed increased reads across ITS2, in agreement with 32S accumulation (**Fig 4C** and **S16C**). Both URB1 and RPL3 depleted samples showed increased coverage across the 3' ETS (**Fig 4C** and **S16A**). For URB1, this confirms the accumulation of 3' ETS extended species in URB1-deficient cells<sup>39</sup>. The observation upon RPL3 knockdown reveals it is also indispensable for this maturation step. Each RNA species was quantified and visualized as heatmaps (**Fig 4D** and **S17A, S18B-C**) or histograms (**Fig S17B-E**) confirming the observations on the IGV traces (**Source Data 1**).

Upon instructing the supervised approach with detailed maturation steps of 18S rRNA 3' end (see above), it was possible to zoom in further. Notably, it was possible to see that all nuclear precursors of 18S-E pre-rRNA accumulates upon WBSCR22 knockdown (**Fig S18D-F**), in full agreement with ref.<sup>36</sup>.

In addition to confirming processing phenotypes previously associated with the selected factors, the nanopore-seq approach revealed several unexpected features. For example, we observed a striking and highly reproducible accumulation of fragments encompassing part of the 5' ETS starting at site 01 upon URB1 and RPL3 depletion (see green and dark yellow rectangles in **Fig 4C** and **S16B**). We also detected a remarkable "processing fingerprint" consisting of multiple consecutive cleavages within the 3' end regions of the 18S and 28S sequences, which we interpret as increased turnover (**Fig 4E** and **S16F, Source Data 1**). This is particularly interesting given that both factors are important for the same maturation step: removal of the 3' ETS (**Fig 4C** and **S16A**). Just as unexpected, we observed that DIMT1L, a small-subunit assembly

factor<sup>37</sup>, also affects 3' ETS processing and 32S maturation (**Fig 4C-D** and **Fig S16**), a phenotype that was not apparent by northern blotting.

Lastly, using the unsupervised approach, we mapped the nuclear reads onto an intensity matrix (**Fig 4F** and **S19A-F**). Upon UTP18 depletion, a unique hub corresponding to 34S was detected (purple arrow, **Fig 4F** and **S19A**). Upon DIMT1L depletion, hubs associated with 21S/21S-C and 26S were seen (brown arrows), indicating impaired processing between sites C and 2, in combination with defects in the 5' ETS (**Fig 4F** and **S19B**). Additionally, we observed notable accumulation of early precursors such as 47S and 45S, consistent with previous findings<sup>37</sup>. Depletion of WBSCR22 revealed accumulation of a hub associated with 18S-E (turquoise arrow), with shortened start coordinates, indicating some level of 5'-3' degradation of 18S precursors (**Fig 4F** and **S19C**). Knockdown of LAS1L showed reduced aggregated reads at processing site 4 and accumulation at 32S (orange arrow, **Fig S19D**). URB1 and RPL3 knockdown exhibited accumulation of reads downstream of processing site 02 (3' end of 28S) corresponding to RNAs that have retained the 3' ETS (green arrows, **Fig 4F** and **S19E-F**). Additionally, upon depletion of URB1 and RPL3, we observed an increased accumulation of products upstream of the 01 site, consistent with the coverage profiles (**S19E-F**), and an increase signal across mature rRNA regions (**S19E-F**), reflecting by-product buildup and elevated turnover (**Fig 4E** and **S17F**).

In summary, NanoRibolyzer captures previously described pre-rRNAs, newly identified pre-rRNA species, and mature rRNAs; it quantifies them, maps processing sites at single-nucleotide resolution, confirms previously assigned functions of processing factors, reveals new involvements at additional steps, and ultimately defines novel unique processing “fingerprints”.

### NanoRibolyzer uncovers spatio-temporal dynamics of RNA modification

Ribosomal RNA is abundantly decorated by covalent modifications<sup>36,40</sup>. To investigate the spatial dynamics of these modifications, we employed direct RNA sequencing (DRS), a technique that sequences native RNA molecules and detects modifications through distinct current signatures that differentiate modified from unmodified nucleotides<sup>41</sup>.

First, we selected three base modifications known to occur during late 18S biogenesis: *N*<sup>7</sup>-methylguanosine m<sup>7</sup>G, deposited by WBSCR22 at G1639 (Ref.<sup>36,37</sup>), *N*<sup>6</sup>,*N*<sup>6</sup>-dimethyladenosine (*m*<sub>2</sub>*Am*<sub>2</sub>*A*) deposited by DIMT1L at A1735 and A1736 (Ref.<sup>37</sup>) and m<sup>1</sup>acp<sup>3</sup>Ψ which is finalized by addition of the acp group by TSR3 at U1248 (Ref<sup>42</sup>) (**Fig S20A-C**). To obtain ground truth, we generated an *in vitro* transcribed sample (IVT 18S) to serve as constitutively unmodified negative control. Additionally, we used RNA extracted from cells inactivated for DIMT1L, WBSCR22, or TSR3 to serve as negative controls. We performed signal mapping refinement (conversion of sequence to raw signal) using *Remora* and analyzed the deviations around each modification site across late precursors (21S, 21S-C, 18S-E) and mature rRNA (18S) in both nuclear and cytoplasmic RNA.

For all three modifications, the signal corresponding to precursors closely resembled that observed in inactive mutants or IVT controls, indicating a lack of modification at these sites (**Fig S20D-F**). In contrast, cytoplasmic 18S showed a distinct signal shift at these positions, consistent with presence of the modification on mature rRNA. For m<sup>7</sup>G, the shift between unmodified (IVT, inactive enzyme, and precursors) and mature cytoplasmic RNA was obvious (**Fig S20D**). For *m*<sub>2</sub>*Am*<sub>2</sub>*A*, it was less obvious but visible (**Fig S20E**). For m<sup>1</sup>acp<sup>3</sup>Ψ, the situation was a bit more complex to interpret because this modification is formed in three steps: sugar 2'-O methylation by SNORA13, base methylation by EMG1, and addition of acp by TSR3. In **Fig 1C**, we discussed that m<sup>1</sup>acp<sup>3</sup>Ψ is readily detected by nanopore sequencing in the cytoplasm but not in the nucleus because the addition of the bulky acp moiety is a late event. Here the IVT control lacking all three modifications (yellow trace), and the sample extracted from cells inactivated for TSR3 lacking only the acp group (black trace) are clearly shifted from each other and from the fully modified mature cytoplasmic 18S rRNA (in red). The nuclear precursor traces (21S, 21S-C, 18S-E) superimpose nicely with the sample inactivated for TSR3, confirming earlier prediction that the addition of the acp moiety is a late event.

In conclusion, NanoRibolyzer pipeline can assign nucleotide modification status to particular pre-rRNA precursor based on raw signal analysis.

Next, we aimed to detect the abundant pseudouridine ( $\Psi$ ), a uridine isomer formed by DKC1 guided by box H/ACA snoRNAs<sup>43</sup>. For this, we used first our nuclear and cytoplasmic RNA fractions and employed SeqTagger<sup>44</sup> to multiplex up to four DRS barcodes per flow cell (two replicates per condition). Following sequencing, pseudouridine levels were quantified using *Dorado* (v7.2.0) with the “PseU basecaller”. In addition to the  $\Psi$  ratio provided by the basecaller, we calculated the U-to-C mismatch ratio, as previous studies have shown that isomerization of U to  $\Psi$  may lead to U-to-C mismatches during base calling<sup>41,45,46</sup>, a feature not integrated in the “PseU basecaller” (see **Supplementary Note S1**). To ensure selectivity, we filtered out sites with a modification ratio below 10% in each condition, and we used IVT 18S and 28S as negative controls.

We reproducibly detected all known 104  $\Psi$  sites in 18S, 5.8S, and 28S rRNAs, as previously identified by mass spectrometry<sup>40</sup> (**Fig S20G, Supplementary Data 1**). We observed that all positions are not fully modified, confirming the existence of compositionally different ribosomes with potentially specialized functions in cells. We conclude that NanoRibolyzer, coupled with “PseU basecaller”, is a robust tool for detecting ribosomal RNA pseudouridines. Using it in combination with C/U mismatch detection proved useful as it allowed to disregard hits at position not known to be modified, such as in the pre-rRNA spacers (**Fig S20G**, see e.g. green peaks in the 5' ETS).

### NanoRibolyzer captures cotranscriptional folding of the 5' external transcribed spacer

Our cDNA sequencing data revealed a consistent ~1,916-nt deletion within the 5' ETS, a feature that was absent from the corresponding DRS dataset (**Fig S21A**). When we mapped this deletion onto a secondary-structure model of the 5' ETS generated with *RNAstructure*<sup>46</sup>, we realized that it corresponds to a reverse-transcriptase bypass of a strongly predicted structural element, characterized by very extensive Watson-Crick base-pairing and a notably low  $\Delta G$ , flanked by cleavage sites A0 and 1 (**Fig S21B-C**). The extended folding of the 5' ETS, as noted by pioneers<sup>48</sup>, brings the distant A0 and 1 sites into close proximity, facilitating coordinated processing (**Fig S21C**).

### Aberrant RNA precursors are hypomodified

Lastly, to gain information about spatio-temporal conversion of U in  $\Psi$ , we followed the ‘pre-rRNA sequential extraction’ (PSE) method<sup>49</sup>, which in our hands led to the isolation of a cytoplasmic (Cp), a nucleoplasmic (Np), and a nucleolar (No) fractions, as confirmed by Western blot analysis with compartment-specific markers (**Fig 5A and S22A**).

We sequenced cDNA in triplicate and examined coverage and precursor abundance across the subcellular fractions (**Fig S22B-D**). The nucleolar fraction (No) showed strong enrichment for transcripts spanning the 5' and 3' ETS and the ITS1 and ITS2 regions (**Fig S22B**). In contrast, transcripts in the cytoplasmic fraction (Cp) were dominated by mature rRNAs. Transcripts enriched in the nucleoplasm (Np) displayed intermediate characteristics, having already undergone partial maturation (**Fig S22B**). Formal quantification revealed that the nucleolar fractions were enriched in the majority of pre-rRNA species, including 30S, 26S, 21S/21S-C, as well as 34S and 36S-C precursors, considered short-lived aberrant species<sup>6</sup> (**Fig S22C-D, Source Data 1**). Progressive processing was reflected by a gradient of precursors abundance, with the highest levels in the nucleolus, followed by the nucleoplasm and then the cytoplasm.

We next performed direct RNA sequencing (DRS), on these fractions to assess the dynamic conversion of U to  $\Psi$  from the nucleolus to the cytoplasm. Using the supervised approach revealed highly similar precursor abundance profiles between the cDNA and DRS datasets (**Fig 5B, Fig S22E-F**). Using the unsupervised approach, we identified in the DRS data, processing hubs corresponding to all key precursors

(e.g. 30S, 26S, 21S, etc., see arrowheads in Fig 5C). In addition, we could now also detect in nucleolar fraction the primary transcript, 47S, and its immediate derivative, 45S (**Fig 5C** and **Fig S22G**).

Interestingly, this led us to discover novel early intermediates, which we tentatively named 47S-01 (lacking the sequence upstream of the 01 site) and 47S-02 (without the downstream sequence of the 02 site) (**Fig 5C** and **Fig 6**). These intermediates were already visible in the processing perturbation intensity matrices, particularly upon DIMT1L depletion (**Fig 4F**). We now have further proof by direct RNA sequencing that they do exist, as these previously undocumented species presumably escaped detection due to the lower sensitivity and resolution of the methods available at the time.

We conclude that, compared with cDNA sequencing, DRS provides higher resolution, reduced background, and improved capture longer precursors (**Fig S22G-H**). We attribute these differences to the library-preparation workflow and to the use of a recently released, more processive enzyme (see **Supplementary Note S2**). Despite these distinctions, precursor abundances measured by cDNA sequencing and by DRS were remarkably similar (**Fig 5B**, **Fig S22E-F**).

Lastly, using the *PseU* basecaller, we reproducibly identified all expected  $\Psi$  sites in nucleolar, nucleoplasmic, and cytoplasmic fractions (**Fig S23-S24**; **Supplementary Data 1**). Having categorized the reads according to their respective precursor sequences (using the supervised approach), we next performed precursor-specific pseudouridine analysis to investigate the modification dynamics across the different precursors and cell compartments. A key conclusion from the precursor-specific modification mapping is that  $\Psi$  sites are already largely present on the primary 47S transcript. Another important observation is that aberrant RNAs, such as the 34S and 36S-C species, are distinctly hypomodified (**Fig 5D-E**). At this stage, we cannot determine whether these RNAs are hypomodified because they are aberrant, or whether insufficient modification contributes to their aberrant nature.

In conclusion, NanoRibolyzer can map rRNA modification at the individual precursor level.

## Discussion

Ribosome biogenesis is a sophisticated pathway involving hundreds of interconnected steps<sup>1,2</sup>. Among these, pre-rRNA processing, i.e., RNA cleavage to generate mature rRNA ends, and pre-rRNA modification serve as excellent *proxies* for monitoring the overall process (**Fig 1A**).

Traditionally, pre-rRNA processing intermediates have been analyzed using metabolic labelling, northern blotting, or primer extension. While these techniques are robust, they are limited in resolution, sensitivity, and throughput, and often require access to costly and hazardous radioactive materials. Moreover, they typically detect only a few dozen abundant precursors, despite the well-established existence of thousands of intermediates generated by progressive, one-nucleotide-at-a-time digestion by exoribonucleases, particularly during 3' end maturation of the 18S and 5.8S rRNAs<sup>50</sup>. This extraordinarily rich population of low-abundance, metastable intermediates has remained largely inaccessible, until now. NanoRibolyzer was designed to address this limitation.

RNA modification represents a second key dimension of RNA metabolism. While the diversity and distribution of rRNA modifications have been well established on mature rRNAs<sup>51</sup>, equivalent information at the level of individual processing intermediates has been lacking. NanoRibolyzer also addresses this limitation.

NanoRibolyzer is a long-read, nanopore sequencing-based pipeline for the comprehensive detection and quantification of pre-rRNA intermediates, coupled with the concomitant mapping of RNA modifications.

To gain spatio-temporal resolution of pre-rRNA processing and modification, NanoRibolyzer was applied to purified subcellular fractions. In this study, sequencing was primarily performed on nuclear and cytoplasmic enriched fractions (**Fig 1B**). A straightforward, high-throughput-compatible protocol was purposefully developed (**Fig 1B**), this was a key prerequisite for successfully establishing the method, as abundant and stable mature rRNAs would otherwise mask the detection of low-abundance, transient nuclear intermediates. Ultimately, we extended the method to include nucleolar, nucleoplasmic, and cytoplasmic fractions (**Fig 5A**). One key conclusion was that the cytoplasm, as expected, contains almost exclusively mature rRNAs, while the nucleolus and nucleoplasm are highly enriched in a diverse population of precursors retaining variable lengths of spacer sequences. In the cytoplasm, we could also monitor final trimming events of 18S rRNA precursors (**Fig S8 and S9**).

A key feature of NanoRibolyzer is that it integrates two complementary mapping strategies: (i) a supervised approach, built on prior knowledge and the large body of literature on pre-rRNA processing, and (ii) an unsupervised approach, which remains entirely agnostic of existing annotations.

The unsupervised or discovery mode was particularly powerful to appreciate the diversity and richness of the reads, which we plotted in a two-dimensional matrix based on their starting and ending coordinates. This output provides an intuitive visualization of processing efficiency under different conditions (e.g., normal vs. perturbed/disease states). In these 2D matrix displays, we could visualize exoribonucleases “in action” digesting pre-rRNAs one nucleotide at a time. RNA ends were perfectly aligned on horizontal or vertical lines in case of 5'→3' and 3'→5' digestion, respectively (**Fig S6**), and the sites of exoRNase pausing became evident.

By combining the two mapping approaches, we validated all known processing sites, often refining them to single-nucleotide resolution, and identified novel ones (**Table 1, Fig 3 and S7**). We also discovered previously uncharacterized pre-rRNA species, notably novel derivatives of the primary transcript (47S-01 and 47S-02), suggesting some plasticity at least in the sequence of the initial cleavage events (**Fig 6**).

We benchmarked our mapping strategy through high-resolution re-analysis of depletion of seven known processing factors (**Fig 4**). These factors were selected to span maturation across all four spacer regions. NanoRibolyzer not only recapitulated their known functions but also uncovered previously unappreciated roles. For example, DIMT1L, classically implicated in ITS1 processing, was also found to affect the 5' ETS

and 3' ETS. This highlights that NanoRibolyzer will become an essential tool for future ribosome-biogenesis research.

Throughout our analyses, we identified an underappreciated diversity of rRNA-derived fragments, spanning both non-coding spacers and mature rRNAs. Some of these fragments likely carry biologically relevant information, as suggested by the highly similar patterns observed upon depletion of factors acting at the same step of ribosome biogenesis. A notable example is provided by URB1 and RPL3, whose depletion impairs removal of the 3' ETS (**Fig 4**), and leads to the accumulation of fragments derived from the 5' ETS (just downstream of site O1), as well as from the 3' ends of 18S and 28S rRNAs (**Fig 4E**). Further studies will determine the extent to which such molecular “fingerprints” may serve as markers of specific perturbations or disease states.

We used both long-read cDNA sequencing (suited for processing studies and, at the time, the only protocol supporting multiplexing), and direct RNA sequencing (DRS), which allows RNA modification detection and RNA end characterization. For mapping and quantifying pre-rRNA and mature rRNA species, results from both approaches were remarkably consistent (**Fig S22**).

By design, NanoRibolyzer, like all sequencing-based approaches, entails inherent selection steps arising from cDNA library preparation and nanopore biophysics; by contrast, Northern blotting is a largely “blind” technique in which RNAs are directly resolved without prior selection, and therefore remains a valuable and complementary method.

NanoRibolyzer also maps RNA modifications, as demonstrated here with the detection of pseudouridines, known to occur at 104 sites on human rRNAs, and several base modifications (e.g.,  $m_2^6Am_2^6A$ ,  $m^7G$ ,  $m^1acp^3\psi$ ). On mature rRNAs, we confirmed all pseudouridine positions previously mapped by RNA mass spectrometry<sup>40</sup>. By analyzing pseudouridines on individual precursor rRNAs, we found that many of the modifications present on mature rRNAs were already detectable on the primary transcript (47S), indicating early modification events.

We also observed that numerous positions were only substoichiometrically modified, confirming previous reports and supporting the concept of ribosome heterogeneity<sup>52</sup>. Most notably, aberrant pre-rRNAs involved in both small and large subunit biogenesis, such as 34S, and 36S-C species (**Fig 5D-E**), were markedly hypomodified. Whether this hypomodification is a cause or consequence of the aberrant nature of these RNA species remains to be determined.

Unexpectedly, we also noted a systematic ~two kilobase deletion in our long-read cDNA data (**Fig S20**), which, upon closer inspection, corresponded to a reverse transcriptase template switch across a predicted strong secondary structure. This region has previously been proposed to spatially coordinate processing sites A0 and 1 (Ref.<sup>48</sup>).

A few recent studies have begun to use long-read sequencing (Oxford Nanopore Technologies, PacBio HiFi) to explore distinct aspects of rRNA biology, including rRNA heterogeneity (SNPs)<sup>53,54</sup>, rRNA modification<sup>17,55-57</sup>, and rRNA processing<sup>17,58</sup>, in archaea, yeasts, and human cells. Compared to protocols that rely on circularization or affinity purification of precursor ribosomes (which may introduce bias) and sequencing of associated RNAs, NanoRibolyzer is simple, robust, and easy to implement.

With the continued evolution of nanopore sequencing technology (e.g., the recent availability of multiplexing in DRS), NanoRibolyzer is poised to evolve further. In the near future, we foresee extending its use to define rRNA subtypes based on patterns of co-occurring or mutually exclusive RNA modifications, and to assess the impact of SNP variants on RNA modification.

In summary, NanoRibolyzer provides a simple, versatile, nanopore sequencing-based platform to study pre-rRNA processing and modification in human cells and, in the future, in other systems. It overcomes key limitations of classical RNA biochemical methods. By enabling single-nucleotide resolution and quantitative analysis of pre-rRNA intermediates, together with RNA modification mapping at the level of individual

species, it offers invaluable insights into ribosome biogenesis, with broad applications in basic research and in the study of disease mechanisms, including ribosomopathies and tumorigenesis, as well as in clinical diagnostics.

ARTICLE IN PRESS

## 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<sub>2</sub>.

All mutations analyses including *tsr3* <sup>-/-</sup> <sup>42</sup>, *dimt1l*-Y131G<sup>55</sup> and *wbsr22*-D82K<sup>55</sup>, were introduced in homozygous diploid in HCT116 p53 positive cells.

### siRNA inactivation experiments

Cells were reverse 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<sup>6</sup>. 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, RPL3, and RPS26 are DsiRNAs (IDT). The Silencer RNA sequences are shown in Table S2.

### Simplified nuclei isolation protocol

Detailed description of the protocol is illustrated in the Supplementary Fig S1. Briefly, samples were trypsinized and washed with cold PBS, after which they were resuspended in Nuclei Isolation Buffer (NIB - 10mM Tris-HCl (pH7.4), 10mM NaCl, 3mM MgCl<sub>2</sub>, 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 100 x g 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 100 x g 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 (20,817 x g) 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 MgCl<sub>2</sub>, 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 150 x g 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 DNase I (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.

### Pre-ribosomal sequential extraction method (PSE)

The PSE method was performed as previously described<sup>49</sup>. Briefly, cells (HEK293) were grown at at ~80% confluency and harvested on ice-cold phosphate-buffered saline. Cell were resuspended thoroughly in 0.5 ml of SN1 buffer (20 mM HEPES-NaOH [pH 7.5], 130 mM KCl, 10 mM MgCl<sub>2</sub>, 0.05% Igepal CA-630), supplemented with Cømplete protease inhibitor cocktail (Roche), followed by centrifugation 1300 x g for

3 min at 4 °C. The supernatant was collected and stored as the SN1 fraction. The pellet was washed with 0.5 ml SN1 buffer, and then resuspended in 0.3 ml of SN2 buffer (10 mM HEPES-NaOH (pH 7.5), 10 mM NaCl, 5 mM MgCl<sub>2</sub>, 0.1% Igepal CA-630, 0.5 mg/ml heparin, 600 U/ml RNasin (Promega)) supplemented with 100 U RNase-free DNase I (Qiagen), and incubated for 10 min at room temperature with gentle mixing. The lysate was centrifuged 12,300 × g for 10 min at 4 °C and the supernatant collected as the SN2 fraction. The remaining pellet was resuspended in 0.4 ml of SN3 buffer (20 mM HEPES-NaOH (pH 7.5), 200 mM NaCl, 4 mM EDTA, 0.1% Igepal CA-630, 0.04% sodium deoxycholate, 4 mM imidazole, 0.1 mg/ml heparin, 1 mM dithiothreitol (DTT), Cøplete, 600 U/ml RNasin) and incubated for 20 min at room temperature with moderate agitation. The extract was centrifuged 12,300 × g for 10 min in 4 °C and the supernatant was collected as the SN3 fraction. Total RNA was prepared from each fraction using TRIzol reagent (Life Technologies, 15596026) according to the manufacturer's protocol.

### RNA isolation

RNA isolation was performed as previously described<sup>59</sup>. 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 for 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 7,500 x g 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<sup>6</sup>. 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). The probes used for northing blotting are shown in Table S3.

### Western blot controls for SN1, SN2, and SN3 fractionation

Equivalent amounts of SN1, SN2, and SN3 fractions, each extracted from starting material equivalent to 25 µg of total protein (fractionated according to<sup>49</sup>), along with 25 µg of total protein from whole-cell lysate used as control, were separated on a 10% SDS-PAGE gel. Proteins were transferred to nitrocellulose membranes and probed in TBS containing 0.1% Tween-20 and 3% BSA. Membranes were incubated with either: 1) Anti-rabbit primary antibodies: anti-histone H3 (Proteintech, Ref 17168-1-AP), anti-nucleolin (Proteintech, Ref 10556-1-AP), and anti-α-tubulin (Proteintech, Ref 11224-1-AP), each at 1:5000 for 2 hours, followed by HRP-conjugated anti-rabbit secondary antibody (Cytiva, Ref NA934V) at 1:5000 for 2 hours, or 2) Anti-mouse primary antibodies: anti-fibrillarlin (Antibodies-online, Ref ABIN361375) or anti-UBF (Tebu Bio, Ref SC-13125), each at 1:1000 overnight, followed by HRP-conjugated anti-mouse secondary antibody (Jackson ImmunoResearch, Ref 115-035-062) at 1:5000 for 2 hours.

### 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  $\mu$ l 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.

### **In-vitro transcription of 18S and 28S rRNA.**

Synthesis of the 18S and 28S rRNA *in-vitro* transcripts was performed using HiScribe® T7 High Yield RNA Synthesis Kit following manufacturer's instruction. Briefly, 2  $\mu$ g linearized plasmid containing the mature rRNA (GenScript) were used as template material, combined with 10X Reaction Buffer, 10 mM NTP's and 2U T7 RNA Polymerase Mix. Reactions were incubated for 2h at 37°C and stopped by digestion of the template plasmid through DNase I (ThermoFisher Scientific, EN0525) following manufacturer's instructions. Purification was performed using Monarch® RNA Cleanup Kit (NEB, T2040) and product quality was assessed using capillary electrophoresis via Agilent RNA ScreenTape Analysis.

### **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.* 1 $\mu$ g of poly(A)-tailed RNA was transferred to a 1.5 ml tube and adjusted to 7.5  $\mu$ l with nuclease-free water. In a 0.2 ml PCR tube, 7.5  $\mu$ l of RNA sample were mixed with 2.5  $\mu$ l of VNP (ONT), 2.5  $\mu$ l of 10 mM dNTPs (NEB N0447), and the volume was adjusted to 11  $\mu$ l 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  $\mu$ l of 5x RT Buffer (ThermoFisher, EP0751), 1  $\mu$ l RNaseOUT (Life Technologies, 10777019), 1  $\mu$ l of Nuclease-free water, and 2  $\mu$ l Strand-Switching Primer (SSP, ONT) per sample, to a total volume of 8  $\mu$ l. 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  $\mu$ l of Maxima H Minus Reverse Transcriptase (ThermoFisher, EP0751) was added, and the total volume becomes 20  $\mu$ l. 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  $\mu$ l 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  $\mu$ l of nuclease-free water. Next, the 20  $\mu$ l of reverse-transcribed samples were prepared with 25  $\mu$ l of 2x LongAmp Taq Master Mix (NEB, N0447), 2  $\mu$ l of PR2 Primer (PR2, ONT), and 3  $\mu$ l of Nuclease-free water, to a total volume of 50  $\mu$ l. 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  $\mu$ l 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  $\mu$ l of cDNA sample with 30  $\mu$ l Nuclease-free water, 7  $\mu$ l Ultra II End-prep reaction buffer (NEB, E7546), and 3  $\mu$ l Ultra II End-prep enzyme (NEB, E7546) mix to a total volume of 60  $\mu$ l. 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  $\mu$ l of nuclease-free water. *Barcode ligation.* Barcode ligation was then performed, where 22.5  $\mu$ l of End-prepped DNA was mixed with 2.5  $\mu$ l of Native Barcode and 25  $\mu$ l of Blunt/TA Ligase Master Mix (NEB, M0367) to a total volume of 50  $\mu$ l. 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  $\mu$ l of nuclease-free

water. Lastly, the barcoded samples are pooled to a final volume of 65  $\mu\text{l}$  in a 1.5 ml Eppendorf tube. *Adapter ligation.* Adapter ligation was performed by adding 65  $\mu\text{l}$  of pooled barcoded sample, 5  $\mu\text{l}$  of Adapter Mix II (AMII, ONT), 20  $\mu\text{l}$  of 5X NEBNext Quick Ligation Reaction Buffer (NEB, B6058), and 10  $\mu\text{l}$  of Quick T4 DNA Ligase (NEB, E6056) to a total volume of 100  $\mu\text{l}$ . 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  $\mu\text{l}$  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, 1  $\mu\text{g}$  of poly(A)-tailed RNA was adjusted to final volume of 9.5  $\mu\text{l}$  with nuclease-free water. 3  $\mu\text{l}$  of NEBNext Quick Ligation Reaction Buffer (NEB B6058), 1  $\mu\text{l}$  of RT Adapter (RTA) (ONT), and 1.5  $\mu\text{l}$  of T4 DNA Ligase 2M U/ml (NEB M0202), were added to sample resulting in a total volume of 15  $\mu\text{l}$ . 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  $\mu\text{l}$  of Nuclease-free water, 2  $\mu\text{l}$  of 10 mM dNTPs (NEB N0447), 8  $\mu\text{l}$  of 5x first-strand buffer (Thermo Fisher Scientific, 18080044), and 4  $\mu\text{l}$  of 0.1 M DTT, resulting in a total volume of 23  $\mu\text{l}$ . The master mix was added to the RNA sample containing the RT adapter-ligated RNA. 2  $\mu\text{l}$  of SuperScript III reverse transcriptase (Thermo Fisher Scientific, 18080044) were added to the reaction, bringing the final volume to 40  $\mu\text{l}$ . 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  $\mu\text{l}$  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  $\mu\text{l}$  of nuclease-free water. For the adapter ligation reaction, 8.0  $\mu\text{l}$  of NEBNext Quick Ligation Reaction Buffer, 6.0  $\mu\text{l}$  of RNA Ligation Adapter (RLA), 3.0  $\mu\text{l}$  of Nuclease-free water, and 3.0  $\mu\text{l}$  of T4 DNA Ligase were mixed with 20  $\mu\text{l}$  of the eluted sample to reach a total volume of 40  $\mu\text{l}$ . The reaction was incubated for 10 minutes at room temperature. 30  $\mu\text{l}$  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  $\mu\text{l}$  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.

### 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<sup>60</sup> 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.

NanoRibolyzer is open-source and has been made freely available to the community through the Epi2ME platform for broader accessibility and use.

### Template-based quantification of rRNA precursors

The template-based algorithm associated long-reads with literature-based ribosomal intermediates<sup>8,9,10</sup>. 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 Fig 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<sup>61</sup> (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 ( $\text{length}(45\text{SN1})^2$ ) 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%.

### Determination of read cleavage sites

The determination of significant abundant cleavage sites was computed using alignment start and end sites where the absolute abundance of start and end sites was determined along the 45SN1 reference. For each template-based intermediate, the mean abundance and standard deviation (SD) of alignment start and end sites were calculated. Cleavage sites occurring two SDs above the mean in the dataset were considered significant abundant cleavage sites. For overlapping template-based intermediate cleavage site intervals, metrics were determined by calculating the mean of means and the mean of SDs. The output files were stored in TSV and BED file formats, including information about their relative abundance and the cleavage site location. The BED files were visualized using IGV.

### DRS Modification analysis

PseudoU detection was performed by using dorado version 7.2.0 with the super accuracy RNA004 model version 5.0.0 using the “pseU” flag. The resulting unaligned bam file was converted into fastq format using samtools<sup>62</sup> (<https://github.com/samtools/samtools>). The fastq file was aligned with minimap2<sup>60</sup> using the “map-ont” long read alignment flag together with the MD flag to ensure the remainder of the modification information. Subsequently, the aligned bam file was processed using pysam (<https://github.com/pysam-developers/pysam>). For each possibly modified nucleotide on a read aligned bam files contain a modification probability. A threshold of 95% modification probability was chosen to determine modifications.

For the acquisition of the raw signal analysis at specific modification sites, the data was basecalled and aligned as described above. Template-based analysis information was used to determine intermediate-specific datasets. The aligned bam file and the original pod5 together with a 45SN1 reference fasta file were processed using remora (<https://github.com/nanoporetech/remora>). Remora was used to perform the resquiggleing, which is the process of association of raw signal interval with specific nucleotides on the reference or basecalled sequence. In this case, once the intervals were associated to the reference sequence, the mean of z-normalized raw current signal covering 10 bases up- and downstream from the loci of interest were extracted. For all reads covering the locus of interest the mean of means and semi-standard deviation (deviation below and above mean) of means were determined.

### 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.

### Data availability

Nanopore sequencing data were deposited in the European Nucleotide Archive (ENA) under project accession number PRJEB82698 [ebi.ac.uk/ena/browser/view/PRJEB82698] and publicly available as of the date of publication. Owing to repository file size limitations, samples ERS24812595 and ERS24812596 were randomly subsampled before public deposition. If the entire dataset of those samples is required, please contact the authors. All data, including raw numbers for graphs are available in the Source Data file or in the Supplementary Information. Source data are provided with this paper.

### Code availability

All scripts and code used in this work have been made available on GitHub (<https://github.com/stegiopast/wf-nanoribolyzer>). The preprocessing pipeline for single samples is part of the nextflow pipeline. Code for downstream analysis is stored in Jupyter Notebooks in the github repository.

### References

- 1 Lafontaine, D. L. J. (2015). Noncoding RNAs in eukaryotic ribosome biogenesis and function. *Nature Structural & Molecular Biology*, 22(1), 11–19. <https://doi.org/10.1038/NSMB.2939>
- 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). <https://doi.org/10.15252/EMBJ.2018100278/ASSET/FE6CB780-F0D9-4514-90A3-C15074F5103F/ASSETS/GRAPHIC/EMBJ2018100278-FIG-0004-M.PNG>
- 3 Ni, C., & Buszczak, M. (2023). Ribosome biogenesis and function in development and disease. *Development (Cambridge, England)*, 150(5). <https://doi.org/10.1242/DEV.201187>
- 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. <https://doi.org/10.1038/s41392-021-00728-8>
- 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. <https://doi.org/10.4161/NUCL.23985>
- 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. <https://doi.org/10.1016/J.MOLCEL.2013.08.011>
- 7 Grummt, I. (2013). The nucleolus—guardian of cellular homeostasis and genome integrity. *Chromosoma*, 122(6), 487–497. <https://doi.org/10.1007/S00412-013-0430-0>
- 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. <https://doi.org/10.1002/1873-3468.12682>
- 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. <https://doi.org/10.1002/WRNA.1269>

- 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. <https://doi.org/10.1016/J.BIOCHI.2012.02.001>
- 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. <https://doi.org/10.1016/J.TIG.2022.02.005>
- 12 McStay, B. (2016). Nucleolar organizer regions: genomic “dark matter” requiring illumination. *Genes & Development*, 30(14), 1598–1610. <https://doi.org/10.1101/GAD.283838.116>
- 13 Agrawal, S., & Ganley, A. R. D. (2018). The conservation landscape of the human ribosomal RNA gene repeats. *PLOS ONE*, 13(12), e0207531. <https://doi.org/10.1371/JOURNAL.PONE.0207531>
- 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. <https://doi.org/10.1101/GR.157941.113>
- 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. <https://doi.org/10.1038/s41587-021-01108-x>
- 16 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>
- 17 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. <https://doi.org/10.1261/RNA.079636.123/-/DC1>
- 18 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). <https://doi.org/10.1186/S13059-020-1935-5>
- 19 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. <https://doi.org/10.1002/MCO2.316>
- 20 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. <https://doi.org/10.3390/CELLS12071051>
- 21 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. <https://doi.org/10.1038/S41592-022-01714-W>
- 22 Brand, R. C., Klotwijk, J., Planta, R. J., & Maden, B. E. (1978) Biosynthesis of a hypermodified nucleotide in *Saccharomyces carlsbergensis* 17S and HeLa-cell 18S ribosomal ribonucleic acid. *Biochem J.*, 169, 71–77
- 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. <https://doi.org/10.1038/s41594-019-0234-x>
- 24 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.

- 25 Rouquette, J., Choismel, 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. <https://doi.org/10.1038/SJ.EMBOJ.7600752>
- 26 Preti, M., O'Donohue, M. F., Montel-Lehry, N., Bortolin-Cavaillé, M. L., Choismel, 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. <https://doi.org/10.1093/NAR/GKT160>
- 27 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. <https://doi.org/10.1093/NAR/GKX253>
- 28 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. <https://doi.org/10.1101/GAD.286963.116/-/DC1>
- 29 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. <https://doi.org/10.1016/j.bcmd.2007.02.001>
- 30 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. <https://doi.org/10.1093/NAR/GKW1047>
- 31 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. <https://doi.org/10.1128/MCB.06019-11>
- 32 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. <https://doi.org/10.1083/JCB.201207131>
- 33 Larburu, N., Montellese, C., O'Donohue, M. F., Kutay, U., Gleizes, P. E., & Plisson-Chastang, C. (2016). Structure of a human pre-40S particle points to a role for RACK1 in the final steps of 18S rRNA processing. *Nucleic Acids Research*, 44(17), 8465–8478. <https://doi.org/10.1093/NAR/GKW714>
- 34 Plassart, L., Shayan, R., Montellese, C., Rinaldi, D., Larburu, N., Pichereaux, C., Froment, C., Lebaron, S., O'donohue, M. F., Kutay, U., Marcoux, J., Gleizes, P. E., & Plisson-Chastang, C. (2021). The final step of 40s ribosomal subunit maturation is controlled by a dual key lock. *ELife*, 10. <https://doi.org/10.7554/ELIFE.61254>
- 35 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. <https://doi.org/10.1074/JBC.M109.054734>
- 36 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. <https://doi.org/10.1261/RNA.047910.114>
- 37 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.

<https://doi.org/10.1091/MBC.E15-02-0073>

- 38 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. <https://doi.org/10.1128/MCB.00358-10>
- 39 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. <https://doi.org/10.1038/s41586-023-05767-5>
- 40 Taoka, M., Nobe, Y., Yamaki, Y., Sato, K., Ishikawa, H., Izumikawa, K., Yamauchi, Y., Hirota, K., Nakayama, H., Takahashi, N., & Isobe, T. (2018). Landscape of the complete RNA chemical modifications in the human 80S ribosome. *Nucleic Acids Research*, 46(18), 9289–9298. <https://doi.org/10.1093/NAR/GKY811>
- 41 Begik, O., Lucas, M. C., Prysycz, L. P., Ramirez, J. M., Medina, R., Milenkovic, I., Cruciani, S., Liu, H., Vieira, H. G. S., Sas-Chen, A., Mattick, J. S., Schwartz, S., & Novoa, E. M. (2021). Quantitative profiling of pseudouridylation dynamics in native RNAs with nanopore sequencing. *Nature Biotechnology* 2021 39:10, 39(10), 1278–1291. <https://doi.org/10.1038/s41587-021-00915-6>
- 42 Meyer, B., Wurm, J. P., Sharma, S., Immer, C., Pogoryelov, D., Kötter, P., Lafontaine, D. L. J., Wöhnert, J., & Entian, K. D. (2016). Ribosome biogenesis factor Tsr3 is the aminocarboxypropyl transferase responsible for 18S rRNA hypermodification in yeast and humans. *Nucleic Acids Research*, 44(9), 4304–4316. <https://doi.org/10.1093/NAR/GKW244>
- 43 Barozzi, C., Zacchini, F., Asghar, S., & Montanaro, L. (2022). Ribosomal RNA Pseudouridylation: Will Newly Available Methods Finally Define the Contribution of This Modification to Human Ribosome Plasticity? *Frontiers in Genetics*, 13, 920987. <https://doi.org/10.3389/FGENE.2022.920987/BIBTEX>
- 44 Prysycz, L. P., Diensthuber, G., Llovera, L., Medina, R., Delgado-Tejedor, A., Cozzuto, L., Ponomarenko, J., & Novoa, E. M. (2025). Rapid and accurate demultiplexing of direct RNA nanopore sequencing data with SeqTagger. *Genome Research*, 35(4), 956–966. <https://doi.org/10.1101/GR.279290.124/-/DC1>
- 45 Hewel, C., Wierczeiko, A., Miedema, J., Friedrich, J., Hofmann, F., Weißbach, S., Dietrich, V., Holthöfer, L., Haug, V., Mündnich, S., Schartel, L., Lehmann, L., Jensen, K. L., Diederich, S., Sys, S., Butto, T., Paul, N. W., Koch, J., Lyko, F., Kraft, F., Russo, A., Lemke, E. A., Helm, M., Linke, M., & Gerber, S. (2025). Direct RNA sequencing enables improved transcriptome assessment and tracking of RNA modifications for medical applications. *Nucleic Acids Research*, 53(22), gkaf1314. <https://doi.org/10.1093/nar/gkaf1314>
- 46 Schartel, L., Jann, C., Wierczeiko, A., Butto, T., Mündnich, S., Marchand, V., Motorin, Y., Helm, M., Gerber, S., & Lemke, E. A. (2024). Selective RNA pseudouridylation in situ by circular gRNAs in designer organelles. *Nature Communications* 2024 15:1, 15(1), 1–10. <https://doi.org/10.1038/s41467-024-53403-1>
- 47 Reuter, J. S., & Mathews, D. H. (2010). RNAstructure: software for RNA secondary structure prediction and analysis. *BMC Bioinformatics* 2010 11:1, 11(1), 129-. <https://doi.org/10.1186/1471-2105-11-129>
- 48 Michot, B., & Bachelier, J. -P. (1991). Secondary structure of the 5' external transcribed spacer of vertebrate pre-rRNA. Presence of phylogenetically conserved features. *European Journal of*

Biochemistry, 195(3), 601–609. <https://doi.org/10.1111/J.1432-1033.1991.TB15743.X>

- 49 Nieto, B., Gaspar, S. G., Moriggi, G., Pestov, D. G., Bustelo, X. R., & Dosil, M. (2020). Identification of distinct maturation steps involved in human 40S ribosomal subunit biosynthesis. *Nature Communications* 2020 11:1, 11(1), 1–17. <https://doi.org/10.1038/s41467-019-13990-w>
- 50 Schneider, C. & Bohnsack, K. E. Caught in the act-Visualizing ribonucleases during eukaryotic ribosome assembly. *Wiley interdisciplinary reviews. RNA* 14, e1766 (2023). <https://doi.org/10.1002/wrna.1766>
- 51 Milenkovic, I. & Novoa, E. M. Dynamic rRNA modifications as a source of ribosome heterogeneity. *Trends in cell biology* 35, 604-614 (2025). <https://doi.org/10.1016/j.tcb.2024.10.001>
- 52 Gay, D. M., Lund, A. H. & Jansson, M. D. Translational control through ribosome heterogeneity and functional specialization. *Trends in biochemical sciences* 47, 66-81 (2022). <https://doi.org/10.1016/j.tibs.2021.07.001>
- 53 Bailey, A. D., Talkish, J., Ding, H., Igel, H., Duran, A., Mantripragada, S., Paten, B., & Ares, M. (2022). Concerted modification of nucleotides at functional centers of the ribosome revealed by single-molecule RNA modification profiling. *ELife*, 11. <https://doi.org/10.7554/ELIFE.76562>
- 54 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. <https://doi.org/10.1016/J.XGEN.2024.100629>
- 55 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. <https://doi.org/10.1080/15476286.2023.2248752>
- 56 Delgado-Tejedor, A., Medina, R., Begik, O., Cozzuto, L., López, J., Blanco, S., Ponomarenko, J., & Novoa, E. M. (2024). Native RNA nanopore sequencing reveals antibiotic-induced loss of rRNA modifications in the A- and P-sites. *Nature Communications* 2025 15:1, 15(1), 10054-. <https://doi.org/10.1038/s41467-024-54368-x>
- 57 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). <https://doi.org/10.1016/J.MOLCEL.2024.11.014>
- 58 An, W., Yan, Y., & Ye, K. (2024). High resolution landscape of ribosomal RNA processing and surveillance. *Nucleic Acids Research*, 52(17), 10630–10644. <https://doi.org/10.1093/NAR/GKAE606>
- 59 Butto, T., Pastore, S., Müller, M., Iyer, K. V., Mündnich, S., Wierczeiko, A., Friedland, K., Helm, M., Winz, M.-L., & Gerber, S. (2024). Real-time transcriptomic profiling in distinct experimental conditions. *ELife*, 13. <https://doi.org/10.7554/ELIFE.98768.1>
- 60 Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, 34(18), 3094–3100. <https://doi.org/10.1093/BIOINFORMATICS/BTY191>
- 61 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. <https://doi.org/10.1038/NBT.1754>

62 Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., & Davies, R. M. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, 10(2). <https://doi.org/10.1093/GIGASCIENCE/GIAB008>

## Acknowledgments

This work was partly funded by Deutsche Forschungsgemeinschaft (DFG, German Research Foundation); project no. 439669440 TRR319 RMaP, TP A05/C01/C03 to M.H. and S.M, TP B07 to T.B., TP A07 to S.G. and L.L and TP C04 to S.G and S.P.. T.B and S.G. acknowledge funding from the ReALity initiative of the Johannes Gutenberg University Mainz. Research in the Lab of D.L.J.L. was supported by the Belgian Fonds de la Recherche Scientifique (F.R.S./FNRS), EOS [CD-INFLADIS, grant n°40007512], Région Wallonne (SPW EER) Win4SpinOff [RIBOGENESIS], the COST actions EPITRAN (CA16120) and TRANSLACORE (CA21154), the European Joint Programme on Rare Diseases (EJP-RD) RiboEurope and DBAGeneCure. S.G. and L.L acknowledge funding by the Boehringer Ingelheim Stiftung. S.G. acknowledges funding by the M3odel initiative from the Forschungsinitiative Rheinland-Pfalz.

## Author Contributions Statements

S.P. developed and implemented the NanoRibolyzer pipeline; S.P. and L.L. analyzed the data. T.B. L.W. and S.P. performed nuclei isolation and RNA isolation experiments. L.W. performed the knockdown experiments and northern blots. S.M. performed the in-vitro transcription of 18S and 28S rRNA. T.B. and S.P. prepared and sequenced the nanopore-seq libraries. T.B. and D.L. conceived and supervised the work, with the assistance of S.G., M.H. and B.L., which provided valuable input and feedback in various discussions. T.B., D.L. and S.P. wrote the paper, with contributions from all the authors.

## Competing Interests Statement

M.H. serves as a consultant for Moderna, Inc. All other authors declare no competing interests.

## Tables

### Table 1. Redefining the human ribosomal processing sites

Processing sites from three reviews<sup>8,9,10</sup> were compared with those identified in this study. Cleavage sites are marked by (^), and regions spanning nucleotide sites are indicated by (-). It's important to note that rRNA annotation can vary positionally due to sequence heterogeneity or incomplete rRNA annotation. In this work, all results are based on the 45SN1 sequence (GeneID:106631777; See material and methods), ensuring consistency with this specific annotations.

## Figure legends

### Fig 1. Streamlined nuclei isolation procedure

**A**, Simplified pre-rRNA processing pathway in human cells.

Three of the four mature rRNAs (18S, 5.8S and 28S), produced as a single long polycistronic precursor, 47S, by RNA polymerase I are interspersed with the 5' and 3' external (ETS) transcribed spacers and internal transcribed spacers (ITS) 1 and 2. The cleavage sites (01, A0, 1 etc.) and main processing intermediates are indicated. Some species (e.g. 34S, 36S, 36S-C) are associated with aberrant processing. **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 (cyt), whole cell (cell), and nuclear (nuc) fractions (n=3 samples each). Fold change (FC) between the conditions is shown above each comparison. One-way ANOVA followed by Tukey test for multiple comparison post hoc test, *Xist*: P(nuc-cyt)=4.94e-05; P(nuc-cell)=6.29e-05; P(cell-cyt)=0.878. *Malat1*: P(nuc-cyt)=3e-07; P(nuc-cell)=4e-07; P(cell-cyt)= 0.536. \*\*\*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. The m<sup>1</sup>acp<sup>3</sup>Ψ modification in 18S, which is visible in cytoplasm but not in the nucleus is highlighted. Positions with a mismatch frequency higher than 0.2 are colored.

**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 shown on the right of each figure 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 **Supplementary Figure S3 and S4**.

**B**, Quantification of detected pre-rRNA intermediates and mature rRNAs in the nucleus and cytoplasm (n=3 samples each).

**C**, Rationale 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 rRNAs 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 represents the exoRNase trajectories (see text for details, **Figure S3C**).

**D**, Overlaid 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).

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

**A**, Left, Intensity matrix of nuclear reads, with intensity "hubs" numbered in red. The "hubs" were identified based on intensity thresholding and closeness to *bona fide* processing sites. Right, Details of each mapped processing sites.

**B**, Illustration of the identified rRNA products in nucleus, as derived from the intensity hubs table shown in panel A. The corresponding rRNA products are shown on the left, together with the hub number (in red) and associated precursor (black). In hub 6b and 7, the (+) symbol depicts extensions 5'upstream 5.8S.

**C**, Quantification of newly detected nuclear pre-rRNAs (details in Figure S9B).

### Fig 4. Analysis of pre-rRNA processing following perturbations using NanoRibolyzer

**A**, Simplified schematics of processing with position of probes used in Northern blotting in the primary transcript (LD1844, LD2122, LD1828, and LD2612). RNA intermediates most affected upon factor knockdown highlighted with the following color-code: UTP18, purple; DIMT1L, brown; WBSCR22, cyan; LAS1L, orange; URB1, dark green; RPL3, khaki.

**B**, Defining a calibration set of reference factors to benchmark Nanoribolyzer. Factor depletion was assessed by northern blotting. Each factor was depleted in HEK293 cells in a time course (from 6- to 72-hrs). The optimal depletion time point was selected for NanoRibolyzer analysis. For UTP18, WBSCR22, and LAS1L it was 48 h. For DIMT1L, URB1, and RPL3, it was 72 h.

**C**, IGV coverage profiles of nuclear reads of control and knockdown of UTP18, DIMT1L, WBSCR22, LAS1L, URB1 and RPL3 across the 5' ETS (left), ITS1, 5.8S, and ITS2 regions (middle) and 3' ETS (left). The transcript regions affected reflect the processing perturbations and are color-coded like in panel A.

**D**, Quantification of Log<sub>2</sub>FoldChange in nuclear pre-rRNA intermediates and mature rRNAs upon knockdown of UTP18, DIMT1L, WBSCR22, LAS1L, URB1 and RPL3 knockdown samples compared to control (n=3 samples each).

**E**, Visual representation of the predominant start (green) and end (orange) sites of reads aligned to the 47S rRNA in cytoplasmic (top) and nuclear (bottom) fractions, comparing control samples with URB1 and RPL3 knockdowns. Cleavage sites are shown only if detected in at least two of the three biological replicates.

**F**, Overlaid intensity matrices of nuclear reads from control, UTP18, DIMT1L, WBSR22, LAS1L, URB1, and RPL3 knockdown samples. Representative ‘intensity hubs’ are highlighted with arrows and annotated with the corresponding RNA species where applicable.

**Fig 5. Spatio-temporal mapping of precursor-specific RNA modifications**

**A**, Schematic representation of the Pre-rRNA Sequential Extraction (PSE) method developed by the team of Prof. Dosil<sup>49</sup>. The SN1, SN2 and SN3 fractions are enriched for the cytoplasm (Cp), nucleoplasm (Np) and nucleolus (No) fractions, respectively (see analysis of marker proteins by western blots in **Fig S21A**).

**B**, Quantification of detected pre-rRNA intermediates and mature rRNAs in the nucleolus, nucleoplasm and cytoplasm (n=3 samples each) using direct RNA sequencing (RNA, left) and cDNA sequencing (right). The results are remarkably similar.

**C**, Overlaid intensity matrices of nucleolar (yellow), nucleoplasmic (blue), and cytoplasmic (red) reads. Key intensity hubs and their associated precursors are marked with arrows. The hubs detected at the top left of the 2-D display (in purple) correspond to the primary transcript, the 47S, and its derivatives: the 47S-01 and 47S-02 (both newly detected species) and the 45S. A schematic in the bottom right corner of the display illustrates the structure of the 47S and of its direct derivatives.

**D**, Heatmap depicting the level of modification each of the 42 pseudouridines detected on 18S and mapped to each individual precursor. The nucleolar signal was used to quantify  $\Psi$  in 18S-E, 21S-C, 21S, 26S, 30S, 45S, and 47S. For quantifying  $\Psi$  in 18S, it is the cytoplasmic signal that was used.

**E**, As in panel D for the 60 sites present in 28S. *in vitro* transcribed 18S and 28S were used as negative controls.

**Fig 6. Current model of human pre-rRNA processing and redefined processing sites**

Novel 47S derivatives (47S-01 and 47S-02, in purple) and hypomodified precursors (34S, and 36S-C, in red) are shown. All processing sites defined to single nucleotide resolution are indicated. See Fig S6 for more detail.

**Table 1 - Redefining the human ribosomal processing sites**

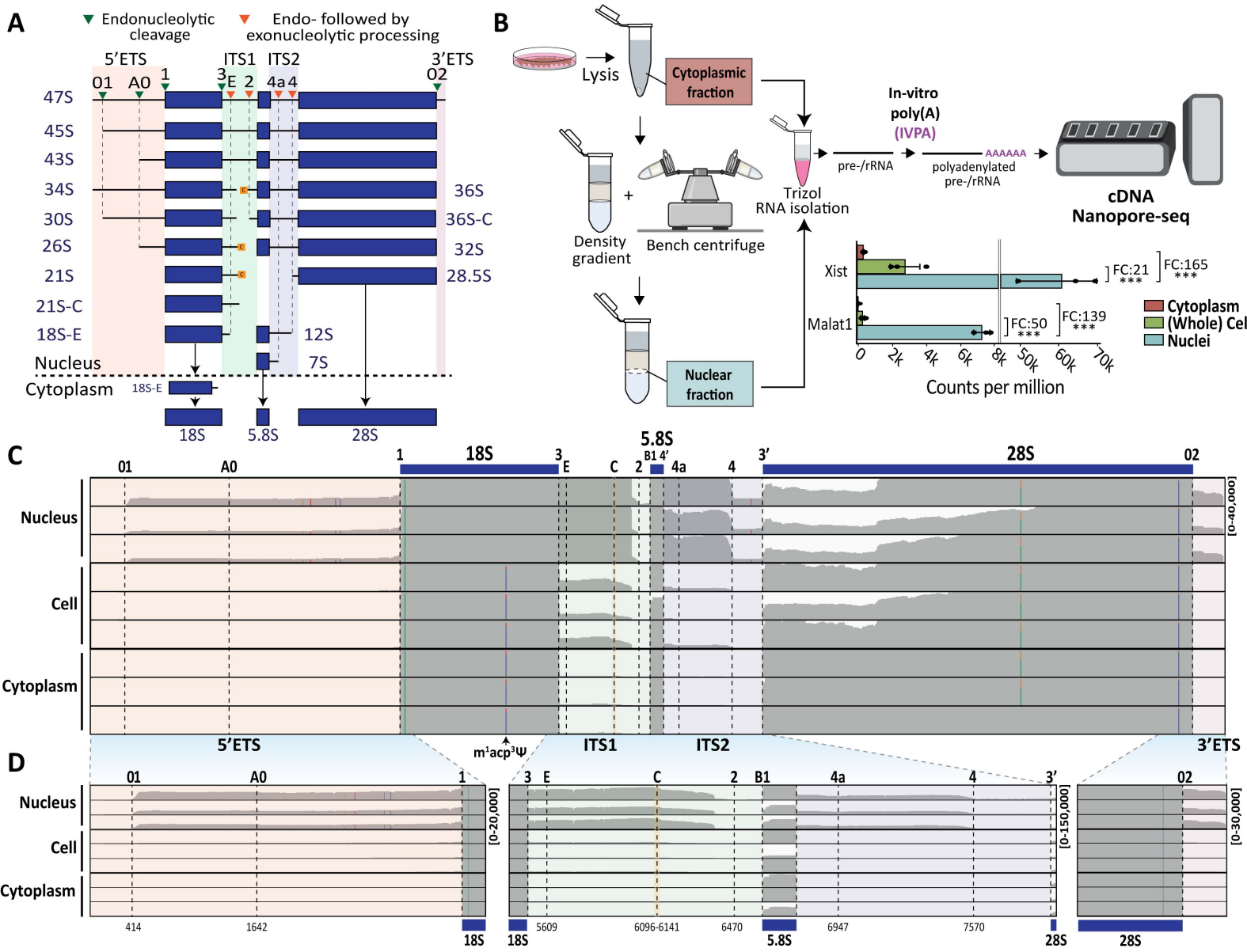
Spacer segment	Processing Site	Reported position of processing site	Proposed position of processing site (45SN1)
5' ETS	01	$\sim C^{414} - C^{416}$ $G^{420} - U^{422}$	$C^{423} \wedge U^{424}$ $G^{429} \wedge U^{430}$ $G^{443} \wedge A^{444}$
	A0	$\sim G^{1643/1642}$	$C^{1609} - U^{1627}$ $C^{1671} \wedge C^{1672}$
5' ETS (5' of 18S)	1	$U^{3653}$	$C^{3654} \wedge U^{3655}$
ITS1 (3' of 18S)	3	$A^{5527}$	$A^{5523} \wedge A^{5524}$
ITS1	E	$5605/5608$ $G^{5606} / G^{5609}$ $\sim G^{5551}$	$C^{5587} \wedge A^{5589} (E^1)$ $G^{5600} \wedge C^{5601} (+78)$ $G^{5720} - A^{5741} (E^2)$
	C	$\sim 6100-6200$ $6117-6192$ $C^{6162} / C^{6177}$	$A^{6131} \wedge A G A \wedge A^{6134}$ $A^{6147} - C^{6153}$
	2	$6477/6482(8)$ $\sim 6470 (9)$ $C^{6469} - C^{6476}$	$C^{6376} - G^{6381} (2^1)$ $A^{6458} \wedge G^{6459} (2^2)$
ITS1 (5' of 5.8S)	B1	$6617, 6623$ $6618$ $C^{6617} - A^{6622}$	$G^{6592} - G^{6602}$ $A^{6600} \wedge C^{6601}$
ITS2 (3' of 5.8S)	4'	$6779$	$C^{6755} - U^{6757}$
ITS2	4a	$6950/6980$ $\sim C^{6947}$	$C^{6861} \wedge C^{6862} (4a^1)$
	4	$\sim 7600?$ $\sim 7570?$	$C^{7559} \wedge G C G G T G T \wedge C^{7566}$
ITS2 (5' of 28S)	3'	$7935$	$A^{7921} - C^{7925}$
3'ETS (3' of 28S)	02	$12969$	$U^{12985} \wedge U U G U \wedge C^{12990}$

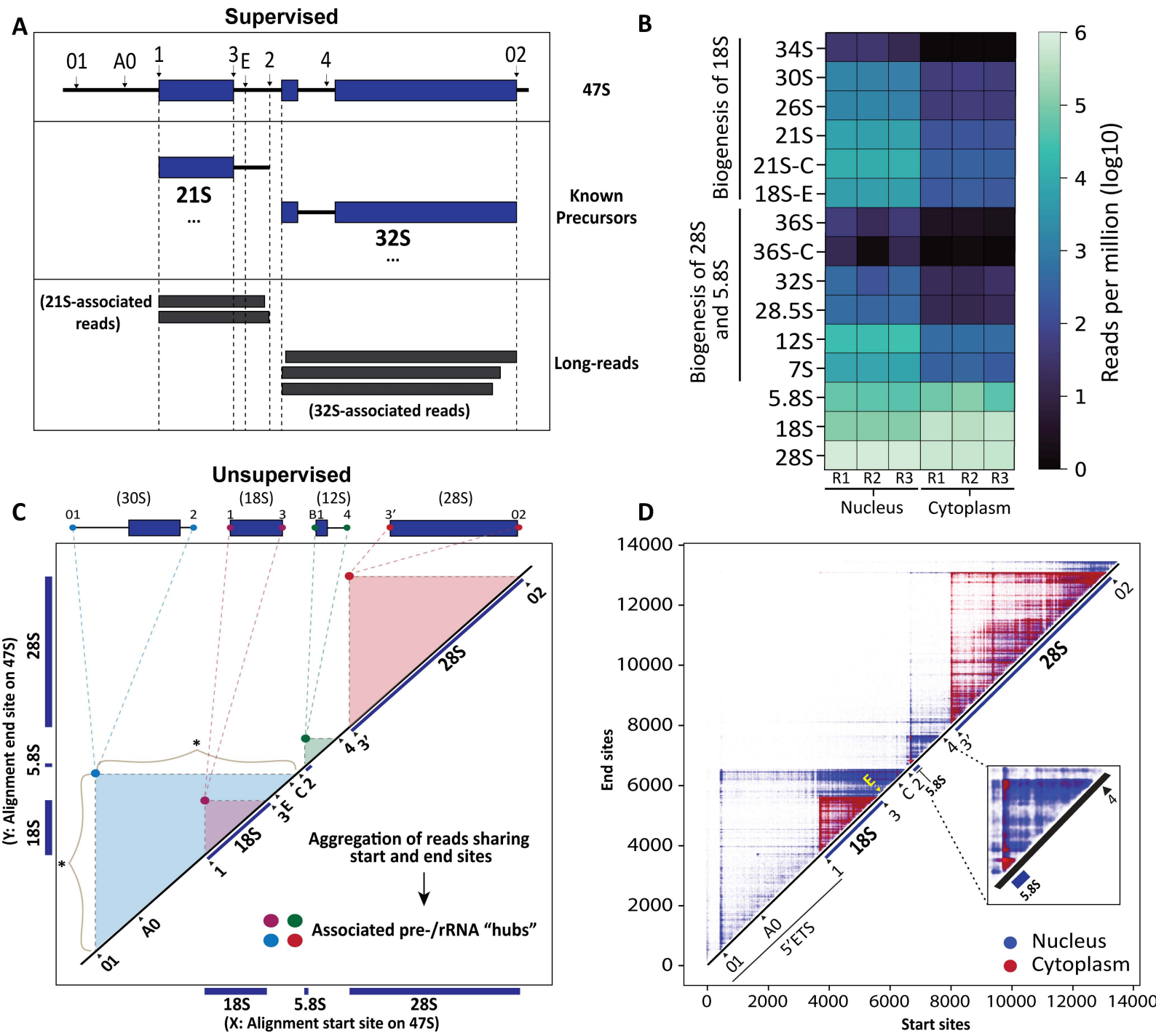
**Editorial Summary:**

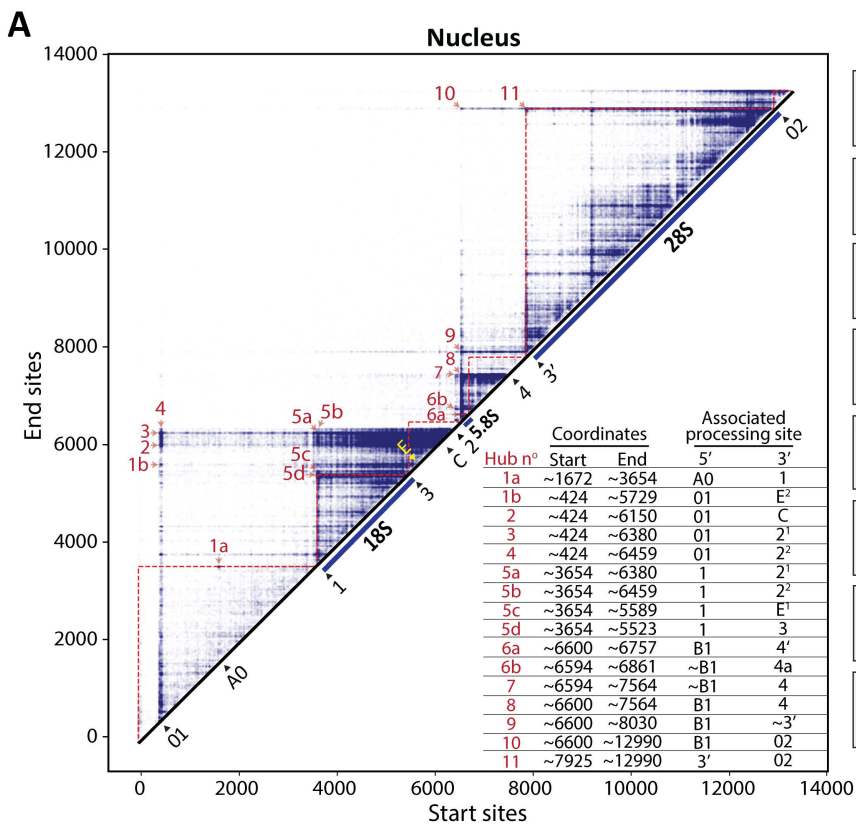
Ribosome production depends on precise processing of precursor RNAs. Here, the authors present NanoRibolyzer, a nanopore-based method that maps rRNA processing and modifications at single-nucleotide resolution, revealing new intermediates and spatial processing dynamics.

**Peer review information:** *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. A peer review file is available.

ARTICLE IN PRESS







**01**

```
CGTCCCGGGGGGCTTCTGTGATCGATGTG
      424
```

**A0**

```
ATCCCGCCGGGCGGGCCCGGCTTCCAG
      1672
```

**1**

```
CTACCTACCTACCTGGTTGATCCTGCCAGTA
      3654
```

**3**

```
CGGAAGGATCATTACGGAGCCCGGAGGGCC
      5523
```

**E<sup>1</sup>**

```
CGCACACCCACCCCCACCGCGACGCGCGG
      5589
```

**E<sup>2</sup>**

```
GAGAGAGAGAGAGAAAGAGAAAGAGGGCGT
      5729
```

**C**

```
GAGGTTAAAGACCCCTGGGGGATCGCCGTCGG
      C-region 6150
```

**2<sup>1</sup>**

```
GGGGCGGGAAACCCCGGCGCTGTGGGG
      6380
```

**2<sup>2</sup>**

```
CGGAGTCCGGTCCCGTTGTGTCTCGTCT
      6459
```

**B1**

```
TCGACCTCGTACGACTCTTAGCGGTGGATC
      6594 6600
```

**4<sup>1</sup>**

```
CTGAGCGTCTGCCGATCAATCGCCCCC
      6757
```

**4a**

```
CTAAGCGAGACCCCGGCGTCCGCCCTCC
      6861
```

**4**

```
GCGGGACCCCGGTGTCTCCCGTCTCGCG
      7564
```

**3<sup>1</sup>**

```
TCCGAGACGCGACCTCAGATCAGCGTGGCGA
      7925
```

**02**

```
TCAGCCCTCGACACAAGGTTGTCCGG
      12990
```

