

TECHNICAL PAPER



Comprehensive identification of diverse ribosomal RNA modifications by targeted nanopore direct RNA sequencing and JACUSA2

Isabel S. Naarmann-de Vries^{a,b+}, Christiane Zorbas^{c+}, Amina Lemsara^a, Michael Piechotta^a, Felix G. M. Ernst^c, Ludvine Wacheul^c, Denis L. J. Lafontaine^{c§}, and Christoph Dieterich^{a,b§}

^aSection of Bioinformatics and Systems Cardiology, University Hospital Heidelberg, Heidelberg, Germany; ^bGerman Center for Cardiovascular Research (DZHK), Partner Site Heidelberg/Mannheim, Heidelberg, Germany; ^cRNA Molecular Biology, Université libre de Bruxelles (ULB), Fonds de la Recherche Scientifique (F.R.S./FNRS), Gosselies, Belgium

ABSTRACT

Ribosomal RNAs are decorated by numerous post-transcriptional modifications whose exact roles in ribosome biogenesis, function, and human pathophysiology remain largely unknown. Here, we report a targeted direct rRNA sequencing approach involving a substrate selection step and demonstrate its suitability to identify differential modification sites in combination with the JACUSA2 software. We compared JACUSA2 to other tools designed for RNA modification detection and show that JACUSA2 outperforms other software with regard to detection of base modifications such as methylation, acetylation and aminocarboxypropylation. To illustrate its widespread usability, we applied our method to a collection of CRISPR-Cas9 engineered colon carcinoma cells lacking specific enzymatic activities responsible for particular rRNA modifications and systematically compared them to isogenic wild-type RNAs. Besides the numerous 2'-O methylated riboses and pseudouridylated residues, our approach was suitable to reliably identify differential base methylation and acetylation events. Importantly, our method does not require any prior knowledge of modification sites or the need to train complex models. We further report for the first time detection of human rRNA modifications by direct RNA-sequencing on Flongle flow cells, the smallest-scale nanopore flow cell available to date. The use of these smaller flow cells reduces RNA input requirements, making our workflow suitable for the analysis of samples with limited availability and clinical work.

ARTICLE HISTORY

Revised 10 August 2023
Accepted 11 August 2023

KEYWORDS

RNA modification; nanopore; direct RNA-seq; ribosomal RNA; RNA methylation; cytidine acetylation

Introduction

Ribosomes are evolutionary-conserved sophisticated nanomachines responsible for protein production in all three kingdoms of life. They are composed of a small and a large subunit, termed 40S and 60S in eukaryotes. Mammalian ribosomes are composed of four ribosomal RNAs (rRNAs), the 5S, 5.8S, 18S, and 28S rRNAs and of 80 ribosomal proteins [1]. The biogenesis of ribosomes is a highly complex process that requires several hundreds of assembly factors [2–5], which are either proteins or small RNAs (small nucleolar RNAs, snoRNAs).

An important step of ribosome biogenesis is the modification of rRNAs [6,7]. Some modifications are conserved throughout kingdoms, others are not. In eukaryotes, the most abundant rRNA modifications are the 2'-O-methylation of ribose (sugar methylation) and the isomerization of uridine to pseudouridine (Ψ) [8]. In a recent comprehensive study conducted on human lymphoblast TK6 cells, 42 pseudouridines and 42 2'-O-ribose methylations were described on the 18S rRNA, whereas the 28S rRNA was shown to contain 61 pseudouridines and 68 2'-O-ribose methylations [8]. Besides these abundant modifications, rRNAs carry additional base modifications, including methylation,

acetylation, and aminocarboxypropylation [8]. Most activities responsible for rRNA modifications have been identified to date [1], however their biological role is often only partially understood. In some cases, it is the physical presence of the modification enzyme in cells rather than that of the modification which is important [9]. 2'-O-methylation and pseudouridylation are catalysed by snoRNA-guided enzyme complexes [10] and are thought to stabilize secondary and tertiary structures of modified rRNAs [11,12]. Furthermore, some of them at least are required to promote efficient and proper translation [12–14].

The importance of correct ribosome assembly for cell homeostasis is evidenced by an emerging class of syndromes designated as ribosomopathies. Ribosomopathies are ribosome biogenesis dysfunction diseases caused by ribosomal protein or ribosome assembly factor mutations [15]. Although all ribosomopathies affect ribosome biogenesis to some extent, they give rise to different syndromes [15]. For example, mutations in the snoRNA-associated pseudouridine synthetase *DKC1* (dyskerin) cause X-linked dyskeratosis congenita, which is associated with bone marrow failure, skin and mucosa alterations, and increased cancer susceptibility [16]. Mutations in *DKC1* also lead to defective

CONTACT Christoph Dieterich ✉ christoph.dieterich@uni-heidelberg.de  Section of Bioinformatics and Systems Cardiology, Klaus Tschira Institute for Integrative Computational Cardiology, Department of Internal Medicine III, University Hospital Heidelberg, Im Neuenheimer Feld 669, Heidelberg 69120, Germany

⁺These authors contributed equally to this work.

[§]These authors contributed equally to this work.

 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/15476286.2023.2248752>.

© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

telomerase maintenance [17], and it is understood that, in this case, both telomerase dysfunction and aberrant RNA modification contribute to the disease. The observed decrease in pseudouridine levels is accompanied by an impairment in the translation of specific cellular mRNAs, in particular those harbouring internal ribosome entry sites in their 5' UTR [12]. Furthermore, translational fidelity is strongly reduced by *DKC1* depletion [18,19].

In the past, the analysis of mutations in ribosomal proteins and ribosome assembly factors was made possible by the use of diverse perturbation techniques. However, it was less feasible to systematically elucidate the effect of mutations in rRNA (due to the high complexity and repetitive nature of genomic rDNA loci [20–22]), or to directly assess dynamic changes in distinct rRNA modifications in clinical or biological samples for which the amount of material is limited. Mass spectrometry- and HPLC-based methods usually require a high amount of input material (μg range) to allow for the accurate detection of ribosomal modifications [8]. Other deep-sequencing-based methods such as RiboMethSeq [23] or HydraPsiSeq [24] are restricted to specific modification types: 2'-O-ribose methylation or pseudouridine detection, respectively. Consequently, the relevance of mutations or aberrant modifications in rRNAs is not well understood.

With the advent of long-read sequencing techniques, especially the direct RNA sequencing method (direct RNA-seq) introduced by Oxford Nanopore Technologies (ONT), it has become possible to directly sequence full-length RNA molecules [25–27]. Using direct RNA-seq, RNA modifications can now also be analysed directly as exemplified recently for m⁶A [25,26,28–31] and pseudouridine [31,32]. Nanopore direct RNA-seq was previously used to analyse rRNA sequences derived from bacteria [29,30], yeast [30,32,33], and human cells [30,31]. An increasing number of computational tools are becoming available to call RNA modifications in direct RNA-seq data [30–36] with different underlying concepts. Each of these methods displays specific advantages and inherent limitations. For example, nanoRMS uses changes in the electric current intensity and trace profiles to analyse pseudouridine modifications [32]. Furthermore, the software can be adapted to identify differential 2'-O-ribose methylation sites; however, it is not intended for detection of the heterogeneous class of other base modifications. Another recent work based on signalAlign profiles known rRNA modifications at the single read level, but requires extensive prior knowledge based on a previously reported modification map [33]. Thus, this method is not suitable to identify novel modifications. We have recently introduced the JACUSA2 algorithm for the analysis of RNA modifications [31], which uses basecalling errors (Mismatch, Deletion, Insertion) in pairwise comparisons (call-2 mode), and handles replicate samples. Previously, we have demonstrated the feasibility of JACUSA2 in the analysis of m⁶A modifications on mRNA and the detection of modified uridine residues (mainly pseudouridine) on rRNA [31].

In this work, we show that JACUSA2 affords the *de novo* identification of differential modification sites with no need for training data or prior knowledge of the position and chemical nature of the modification. First, we established nanopore targeted direct RNA-seq of human 18S rRNA with

a specific adapter (rather than by performing bulk polyadenylation and non-discriminative sequencing of total RNA). In a benchmark of JACUSA2 against other published tools, we find that JACUSA2 Mis is especially useful to identify base modifications as methylation and acetylation. Moreover, JACUSA2 is faster and computationally less demanding than most other tools. To validate the detection of these RNA modifications, we employed a collection of CRISPR-Cas9-engineered human colon carcinoma cells (HCT116) lacking specific modifications on the 18S rRNA. To detect modification signatures in nanopore direct rRNA-seq data, each analysed modification was systematically compared in mutant (MUT)/knock out (KO) and isogenic wild type (WT) control cells. Using this approach, we validated, in addition to the abundant 2'-O-ribose methylated and pseudouridylated residues, several 18S rRNA base modifications (including the METTL5-catalysed m⁶A₁₈₃₂ [37], the DIMT1L-catalysed m⁶A₁₈₅₀ m⁶A₁₈₅₁ [9,38,39], the WBSCR22-catalysed m⁷G₁₆₃₉ [9,40,41], and the NAT10/SNORD13-catalysed ac⁴C₁₈₄₂). Employing JACUSA2 call-2, we show that all analysed rRNA modifications are detectable as differential sites in nanopore direct rRNA-seq data. Importantly, we provide experimental and computational evidence that our approach can estimate relative levels of modification based on calibration curves. Lastly, to expand the repertoire of biological samples that are accessible to nanopore direct rRNA-seq, the targeted rRNA-seq method was transferred to the small-scale Flongle flow cells. We demonstrate that downscaled nanopore direct rRNA-seq on these small devices equipped with only 126 pores allows for detection of rRNA modifications as efficiently as with MinION flow cell sequencing.

In summary, we show that nanopore direct rRNA-seq in combination with JACUSA2 is a fast and simple approach to identify changes in rRNA modification pattern, irrespective of the chemical nature of the analysed modifications or prior knowledge on the modification sites. Furthermore, our miniature Flongle-based sequencing approach makes the method amenable to precious low-input samples of biological and clinical relevance.

Materials and methods

Generation of HCT116 mutant cell lines

All human cell lines were generated in p53-positive diploid HCT116 cells (ATCC, #CCL-247) by genome editing. The recipient cell line was diagnosed by ATCC by short tandem repeat (STR) analysis prior to use. The HCT116 METTL5^{−/−} cell line has been described previously [37]. Here, the exon encoding the catalytic domain of the protein was precisely excised from the genome on both alleles by CRISPR-Cas9 genome editing.

To generate the HCT116 DIMT1L^{Y131G/Y131G}, WBSCR22^{D82K/D82K}, and SNORD13 KO cell lines, the selected point mutations or deletion of SNORD13 were introduced by CRISPR-Cas9 genome editing on both alleles, as follows: an *in vitro* reconstituted Cas9 RNP complex (final concentration 4 μM) consisting of specific crRNA guides (see Table S2),

a universal tracrRNA (IDT, #1072532), and the *Streptococcus pyogenes* Cas9 (IDT, #1081058), was electroporated in cells freshly resuspended in nucleofector solution V (Lonza, VCA-1003) together with a single-strand donor DNA (ssDNA, final concentration 4 μ M) in case of the point mutations. Cells were electroporated with an enhancer (IDT, #1075915, final concentration 4 μ M) in a nucleofector device (Lonza, Nucleofector 2; programme D-032). Cells were incubated for 24 h to allow them to recover and then detached and cloned by serial dilution. Individual clones were selected and diagnosed by PCR amplification of the modified area (for SNORD13 KO clones, clone #2 was used), followed by differential restriction digest for DIMT1L^{Y131G/Y131G} clones (gain of a *Bst*NI restriction site, clone #10 was used) and WBSCR22^{D82K/D82K} clones (loss of an *Eco*RV restriction site, clone #1 was used) as well as by DNA sequencing of the modified area.

Cell culture

HCT116 cells were cultured in McCoy's 5A medium (Lonza, BE12-168F) supplemented with 10% foetal bovine serum (Sigma, F7524), 100 U/ml penicillin, and 100 μ g/ml streptomycin (Lonza, DE17-602E) in a New Brunswick Galaxy 170 R incubator at 37°C and under 5% CO₂.

Isolation of total RNA

Total RNA from HCT116 cells was extracted in TriReagent solution (Thermo Fisher) according to the manufacturer's instructions.

Isolation of genomic DNA

Genomic DNA was isolated from five Mio HeLa cells using the NucleoSpin tissue kit (Macherey-Nagel) according to the manufacturer's protocol.

Generation of templates for *in vitro* transcription

The complete 18S rRNA sequence was amplified from genomic DNA by touchdown PCR with Q5 DNA polymerase (New England Biolabs) using a forward primer that introduces the T7 promoter sequence for *in vitro* transcription (IVT). The following protocol was used for touchdown PCR: 30 sec initial denaturation at 98°C, 20 cycles of touchdown (10 sec, 98°C; 20 sec, 72°C to 62°C (Δ Tm -0.5° C); 5 min, 72°C), followed by 15 cycles standard PCR at 62°C annealing temperature and final elongation (5 min, 72°C). The primer sequences are listed in Table S1.

In vitro transcription

The 18S IVT was generated using the T7 Megascript kit (Thermo Fisher Scientific) according to the manufacturer's protocol. RNA integrity was analysed on a 1% agarose gel. The IVT product was purified using RNA Clean and Concentrator kit (Zymo Research).

Polyadenylation of 18S IVT

One-microgram 18S IVT was polyadenylated with an E-PAP based Poly(A) Tailing Kit (Thermo Fisher Scientific) according to the manufacturer's instructions and purified using RNA Clean and Concentrator kit (Zymo Research).

Generation of ONT direct RNA-seq libraries for sequencing on FLO-MIN106D (R9.4.1) flow cells

Direct RNA-seq libraries were generated using the SQK-RNA002 kit (Oxford Nanopore Technologies) following the sequence-specific protocol. Universal oligo A and sequence-specific oligo B (Table S1) were annealed at a concentration of 1.4 μ M each in 10 mM Tris, pH 7.5, 50 mM NaCl (2 min, 95°C; 0.1°C/sec to 22°C). Briefly, 500 ng total RNA in a volume of 9 μ l was ligated to 1 μ l custom adapter using 1.5 μ l T4 DNA ligase (New England Biolabs) in NEB next Quick ligation buffer (3 μ l, New England Biolabs) in the presence of 0.5 μ l RNA CS (Oxford Nanopore Technologies) for 10 min at room temperature. Reverse transcription to stabilize the RNA strand was performed using Superscript IV reverse transcriptase (Thermo Fisher Scientific for 50 min at 50°C, followed by enzyme inactivation (10 min, 70°C)). Reactions were cleaned up using Agencourt RNAClean XP beads (Beckman Coulter). The RMX RNA adapter was ligated as described above, followed by Agencourt RNAClean XP purification and elution in 21 μ l elution buffer. The concentration of the library was determined using Qubit DNA HS assay (Thermo Fisher Scientific). Libraries were sequenced on a GridION X5 device equipped with MinION R9.4.1 flow cells for 48 h and basecalled with Guppy 5.0.11 in fast-basecalling mode.

Generation of ONT direct RNA-seq libraries for sequencing on Flongle flow cells

Libraries for sequencing on Flongle flow cells were prepared as above with some modifications. Libraries were prepared with 200 ng total RNA as input. After the first cleanup, RNA was eluted in 10 μ l H₂O and the following steps carried out in a smaller volume: 10 μ l RNA were ligated to 2 μ l RMX with 1 μ l T4 DNA Ligase in a total volume of 20 μ l and purified with an equal volume of Agencourt RNAClean XP beads. Libraries were eluted in 9 μ l ELB. Flongle flow cells were loaded by a community protocol to allow loading similar to the FLO-MIN106D flow cells (<https://community.nanoporetech.com/posts/a-very-gentle-relatively>). Flongle flow cells were primed with 117 μ l FLB +3 μ l FLT. Eight-microlitre library was diluted with 7 μ l H₂O and loaded with 15 μ l RRB. Flongle libraries were sequenced on a GridION X5 device equipped with Flongle adapters and Flongle flow cells for 24 h and basecalled with Guppy 5.0.11 (Figure 5: 5.1.13) in fast-basecalling mode.

Preprocessing of direct RNA-seq data

Reads were mapped using minimap2 version 2.17. BAM files were filtered to exclude secondary and poor alignments. Plus,

the MD tag was added to allow variant calling using JACUSA2 software without the need for the reference transcriptome.

Detection of modifications with JACUSA2

The JACUSA2 software [31] calculates individual scores for Mismatch, Deletion, and Insertion events. We employ the JACUSA2 call-2 run mode throughout this manuscript, which compares replicate samples from two conditions. If not otherwise stated, we use 1,000 reads for our analyses throughout the manuscript. We tested the following feature sets derived from JACUSA2 to take into account the clustering of rRNA modifications, as well as the inherent characteristic of the current nanopore pore protein to cover 5-mers: a) Mismatch score of the analysed site (M), b) Mismatch, Insertion, and Deletion scores of the analysed site (MDI), c) Mismatch score of the 5-mer context (modified site in position 3), Insertion, and Deletion score of the analysed site (M_{ConDI}) and d) Mismatch, Insertion, and Deletion score of the 5-mer context ($(MDI)_{\text{Con}}$). We either use these JACUSA2 scores and feature sets directly or employ them as input to the Local Outlier Factor (LOF) [42] method. To generate JACUSA2 score plots, our software performed WT vs. IVT comparisons on either one or three replicate samples. In case of the genetic model systems, where we can make an assumption on the number of differential sites, we applied the LOF method which predicts outliers in an unsupervised manner by measuring the density deviation of each point with respect to its neighbours. We predict positions with the highest LOF score as modified. We only compute LOF scores for positions where the JACUSA2 score is above the median of the score distribution. We compute the LOF score for all positions using the set of pairwise comparisons as features. The proportion of outliers to be captured for the analysed cases was set to 0.1–0.2% (contamination value 0.001–0.002), depending on the expected number of differential sites. We used the ‘LocalOutlierFactor’ function from the scikit-learn python package with the default neighbourhood size 20 to compute LOF scores and the matplotlib package for visualization. We labelled the identified outliers as modified site, neighbours (–2 to +2) or non-modified positions *post hoc* based on the analysis by [8].

rRNA benchmark

The rRNA benchmark has been implemented as a snakemake (v7.25.3) workflow to measure the performance of different software solutions in identifying RNA modifications. All employed tools perform pairwise comparisons where WT and KO/IVT conditions are contrasted. Performance has been measured against a set of known 18S rRNA modifications. The influence of basecalling, and the number of replicates has been investigated by running tools on data sets processed with different combinations of the aforementioned factors. Calculations have been carried out on the same node in a slurm cluster (see Supplementary Text for details). The area under the precision recall curve has been used to compare the performance of the tools (Figure S2, Figure S3 and

Supplementary Text). The running times of the tools were determined in triplicates (Table S4, Supplementary Text).

Downsampling analysis

To evaluate the effect of read coverage on the analysis, BAM files were downsampled to different amounts of reads (0.3k, 0.5k, 1k, 5k, 10k). We employ various seed values for the downsampling procedure. The generated down-samplings were subjected to the JACUSA2 call-2 analysis. To compare results across the different levels of read coverage, we calculated the distance between modification sites and the median in terms of two basic scores: the JACUSA2 Mismatch scores and the score assigned to each site by the LOF method. To avoid bias caused by the different scales of LOF scores across analyses, the normalized distance was considered so that the difference between the score of the modification site and the median is divided by the maximum LOF value.

Mixing analysis

To evaluate the ability to detect rRNA modifications with low stoichiometry, *in silico* samples with different averages of modification rates (0%, 0.5%, 5%, 10%, 25%, 50%, 75%, 100%) were designed by combining WT and KO/MUT samples of 1,000 reads. Then, differential analysis of the generated mixtures and the MUT/IVT samples was performed using JACUSA2 call-2. For the experimental mixing analysis, WT and KO/MUT RNA were mixed with the indicated ratios and libraries were prepared for sequencing on Flongle flow cells as described above. The JACUSA2 Mismatch score was compared across different mixture ratios.

Preprocessing, down-sampling, and mixing were performed using Samtools version 1.9. A Snakemake pipeline for the analysis workflow was developed and is available on Zenodo (<https://doi.org/10.5281/zenodo.8268171>).

Northern blot

The loss of SNORD13 in HCT116 SNORD13 KO was analysed by Northern blotting [43] using a probe described in Table S2. Ethidium bromide staining was used to control loading.

Primer extension assay

Primer extension assays were used to validate the loss of specific 18S rRNA modifications in CRISPR-Cas9-engineered HCT116 cell lines as described previously [38,44,45] using 2 µg total RNA and primers listed in Table S3.

Misincorporation assay

Acetylation of 18S rRNA C_{1842} was analysed by NaCNBH_3 reduction, followed by TGIRT-III reverse transcription and Sanger sequencing of the PCR product based on the method described [45–47]. Briefly, 200 ng total RNA was reduced with 100 mM NaCNBH_3 in 100 mM hydrochloric acids for 20 min at room temperature in a total volume of 100 µl. Reactions were

quenched by the addition of 30 μ l 1 M Tris pH 8.0 and purified with RNA Clean & Concentrator-5 kits (Zymo Research). 200 pg reduced RNA or non-treated control (not shown) was reverse transcribed with a primer targeting the H45 of the 18S rRNA (18S H45 rev) and TGIRT-III reverse transcriptase. RNA was mixed in a total volume of 17 μ l with 4 μ l 1 μ M RT primer and 4 μ l 5 $\times$ TGIRT buffer (2.25 M NaCl, 25 mM MgCl₂, 100 mM Tris pH 7.5) and incubated 3 min at 75°C, followed by 3 min on ice. Then, 1 μ l 100 mM DTT and 0.5 μ l TGIRT-III (Ingex) were added. Reactions were incubated for 20 min at room temperature. After addition of 1 μ l dNTP mix with reduced dGTP (10 mM dATP, dTTP, and dCTP and 5 mM dGTP), reactions were incubated 60 min at 57°C. PCR reactions were composed of 2 μ l cDNA, 2.5 μ l 10 μ M 18S H45 forward and reverse primer, respectively, 10 μ l 5 $\times$ HF buffer, 1 μ l 10 mM dNTPs, 1 μ l Phusion Hot Start Flex DNA Polymerase (New England Biolabs) and 31 μ l water. Reactions were cycled with the following programme: 30 sec initial denaturation at 98°C; 35 cycles (10 sec, 98°C; 15 sec, 67.4°C; 15 sec, 72°C) and final elongation (5 min, 72°C). PCR products were analysed on 2% agarose gels stained with GelRed (Biotium) and cleaned up with NucleoSpin Gel and PCR Clean-up Kit (Macherey-Nagel). PCR products were analysed by Sanger sequencing (LGC Genomics) with the 18S H45 forward primer. Primer sequences used in the Misincorporation assay are listed in Table S3. Chromatograms were analysed with Chromas 2.6.6 (Technelysium Pty Ltd).

Results

Targeted direct ribosomal RNA-seq in total human RNA samples

The direct analysis of rRNA sequence variants and of rRNA modifications has only become possible recently, thanks to the advent of the direct RNA-sequencing (direct RNA-seq) platform developed by Oxford Nanopore Technologies (ONT). Here, we established a protocol with minimal pre-processing to enable the analysis of low-input samples (Figure 1A,B). To prevent the laborious purification of individual rRNAs, which often suffers from material loss and introduction of biases, or additional experimental steps, such as *in vitro* polyadenylation, we established a custom adapter (analogous to [48]) for the selective sequencing of human 18S rRNA that captures the 3' end of the mature rRNA for direct rRNA-sequencing (Figure 1A). Input for our analysis was BAM files (basecalling with Guppy, alignment with minimap2) sampled to equal read numbers (1,000 or 5,000 reads). These were then subjected to a Snakemake workflow for pairwise JACUSA2 call-2 analysis. JACUSA2 calculates scores for different basecalling errors: Mismatch, Deletion, and Insertion, which can be combined to feature sets. Identification of significantly different sites may be formulated as an outlier detection problem with local outlier factorization (LOF) (Figure 1B). The human 18S rRNA harbours 91 modification sites, with pseudouridines (psU) and 2'-O-ribose methylation (Nm) being the largest classes (Figure 1C).

First, we compared the performance of two sequence-specific adapters with different lengths to the standard oligo(dT) adapter (RTA) employing an *in vitro* transcribed

(IVT) 18S rRNA. A ten nucleotides long adapter was sufficient to efficiently capture the 18S rRNA (Figure S1A,B). Furthermore, this analysis revealed sequencing of incomplete (missing 3' end) and reverse strand RNA molecules with the standard RTA. This problem was not observed with the sequence-specific adapter (Figure S1D-E). This adapter was then used for direct rRNA-seq of rRNA from human HCT116 cells and compared to the sequencing of a full-length 18S IVT (Figure 1D). The cellular rRNA displayed a higher mismatch rate than the IVT (coloured lines, Figure 1D), which is indicative of the presence of rRNA modifications detectable *via* basecalling errors with the JACUSA2 algorithm [31].

To elucidate whether JACUSA2 is indeed suitable to identify well-defined rRNA modifications [8], we analysed the 18S rRNA from HCT116 wild-type cells (WT) and the corresponding IVT using JACUSA2 call-2 (WT vs. IVT) and calculated Mismatch (Mis), Deletion and Insertion scores. Furthermore, we considered the 5-mer context (as the R9.4.1 nanopore protein covers five nucleotides at a time) to test whether it contributes to the modification signal. This results in four feature sets that were derived from JACUSA2 as described in the Methods section.

The performance of these JACUSA2 feature sets was benchmarked against other tools designed for the identification of RNA modifications by direct RNA-seq, namely xPore [34], Nanocompore [36], EpiNano [28,32,35] and Eligos2 [30]. We noticed that JACUSA2, also on replicate samples, and EpiNano have the shortest run times (Table S4). All tools were run on fast and high-accuracy (HAC) basecalled data. The AUC for precision and recall was calculated for all modifications and additionally stratified according to modification type (pseudouridine, 2'-O-ribose methylation and other modifications) for one and if possible three replicates (Figure S2). For all modifications, JACUSA2 Mis, EpiNano linear and EpiNano delta had the best performance, with EpiNano being slightly better on HAC basecalled data (Figure S2, panel 1). Interestingly, the performance of many tools was strongly dependent on the modification type. Here, EpiNano was best in detecting 2'-O-ribose methylation sites (Figure S2, panel 3), whereas JACUSA2 Mis and JACUSA2 MDI were superior in detection of diverse 'other' modifications (Figure S2, panel 4). We further stratified the 2'-O ribose methylation sites according to the nucleobase (Figure S3). This reveals that in case of Cm the 5-mer context and the use of replicates are important for modification calling (Figure S3, panel 2). However, as the Mis score is the main determinant for all other modification types, we decided to use JACUSA2 Mis on fast basecalled data throughout this study.

The JACUSA2 Mis scores for the WT – IVT comparison were further stratified for every modification type (Figure 1E), revealing the capability to detect ac⁴C, m²A, m¹acp³psU, m⁶A and m⁷G in nanopore direct RNA-seq data. JACUSA2 Mis scores on HAC basecalled data (Figure S4) and for the other JACUSA2 feature sets (Figure S5) are provided in the Supplementary Material.

In conclusion, Nanopore direct rRNA-seq coupled to JACUSA2 analysis is suitable for the detection of diverse RNA modifications in human rRNAs and may be especially useful for detecting base modifications as methylation and acetylation.

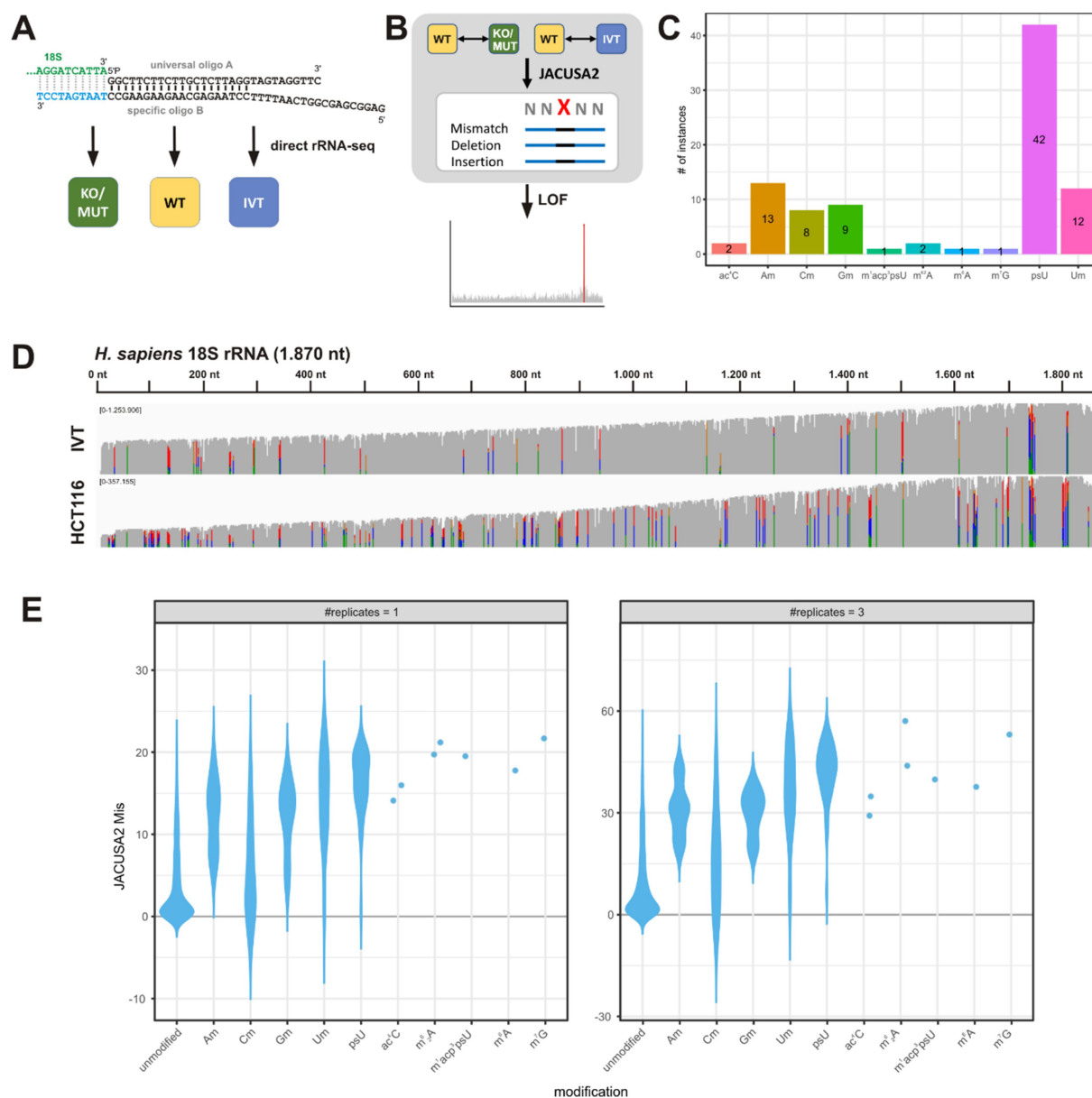


Figure 1. Setup of targeted direct rRNA sequencing. (A) a custom adapter is employed for sequencing of the human 18S rRNA (direct rRNA-seq) in samples representing different modification status: wild type (WT), knock out/mutant (KO/MUT) or *in vitro* transcribed rRNA (IVT). The custom adapters consist of an universal oligo A annealed to a sequence specific oligo B. The sequence-specific part of oligo B (light blue) has a length of 10 nts and anneals with the 3' end of the 18S rRNA (green). (B) Overview of the analysis workflow. BAM files from the direct rRNA-sequencing (A) are compared pairwise employing JACUSA2 call-2, which calculates scores for different error profiles (Mismatch, Deletion, Insertion). These scores can be combined to feature sets, taking only the target site, or as well the 5-mer context into account. Significant outliers are identified by Local outlier Factorization (LOF). (C) Abundance of modification types on the 18S rRNA. (D) Coverage of nanopore direct rRNA-seq of 18S IVT and 18S rRNA from HCT116 WT cells sequenced on MinION R9.4.1 flow cells. Allele frequency threshold = 0.2. Mismatches are indicated by IGV default colours. (E) Violin plot summarizing the JACUSA2 call-2 analysis of the 18S rRNA from HCT116 WT cells and 18S IVT. Shown is the JACUSA Mis score for all modification types on fast basecalled data as indicated. Left panel: analysis of a single replicate, right panel: analysis of three replicates.

Nanopore direct rRNA-seq enables the detection of site-specific RNA modifications

Modification sites on rRNAs, including human rRNAs, have been characterized extensively by a range of techniques, including classical RNA biochemistry (such as primer extension), HPLC, mass spectrometry, short-read deep sequencing-based methods (see Introduction), and more recently, by CryoEM. For most of these modifications, the responsible enzymes and, when relevant, the antisense snoRNA guides have been identified and characterized. We made use of this

knowledge to analyse specific rRNA base methylations of the 18S rRNA in human cells lacking individual modifications by targeted nanopore direct RNA-seq as a proof-of-concept.

For this, we generated human colon carcinoma-derived cell lines (HCT116) genetically engineered by CRISPR-Cas9 genome editing to harbour either a knock-out (KO) or a catalytic-dead variant (MUT) of selected methyltransferases (see Material and Methods for details). The cell lines and affected modifications are listed in Table 1. The METTL5 KO cell line was described previously, and it lacks the m⁶A modification at

position 1832 of the human 18S rRNA [37]. In addition, we generated cell lines that express catalytic dead variants of DIMT1L and WBSCR22 (Figure 2). DIMT1L catalyses the double m⁶A modification at positions 1850 and 1851 of the 18S rRNA [9,38,39], whereas WBSCR22 catalyses the m⁷G modification at position 1639 of the 18S rRNA

[9,40,41]. The conserved D82 amino acid in the catalytic pocket of WBSCR22 (Figure 2A) was mutated to a lysine, resulting in loss of a diagnostic *EcoRV* site as monitored by differential restriction digest (Figure 2B). As expected, in cells expressing the WBSCR22 D82K variant, the methylation of G₁₆₃₉ was no longer detected in a primer extension assay

Table 1. Genetically engineered HCT116 cell lines analysed by direct rRNA-seq.

Cell line	Missing modification	Reference modification enzyme
HCT116 WT	-	
HCT116 WBSCR22 ^{D82K/D82K}	18S m ⁷ G ₁₆₃₉	[9,40,41]
HCT116 METTL5 ^{-/-}	18S m ⁶ A ₁₈₃₂	[37,49]
HCT116 DIMT1L ^{Y131G/Y131G}	18S m ⁶ A ₁₈₅₀ m ⁶ A ₁₈₅₁	[9,38,39]
HCT116 SNORD13 ^{-/-}	18S ac ⁴ C ₁₈₄₂	[43,50,51]

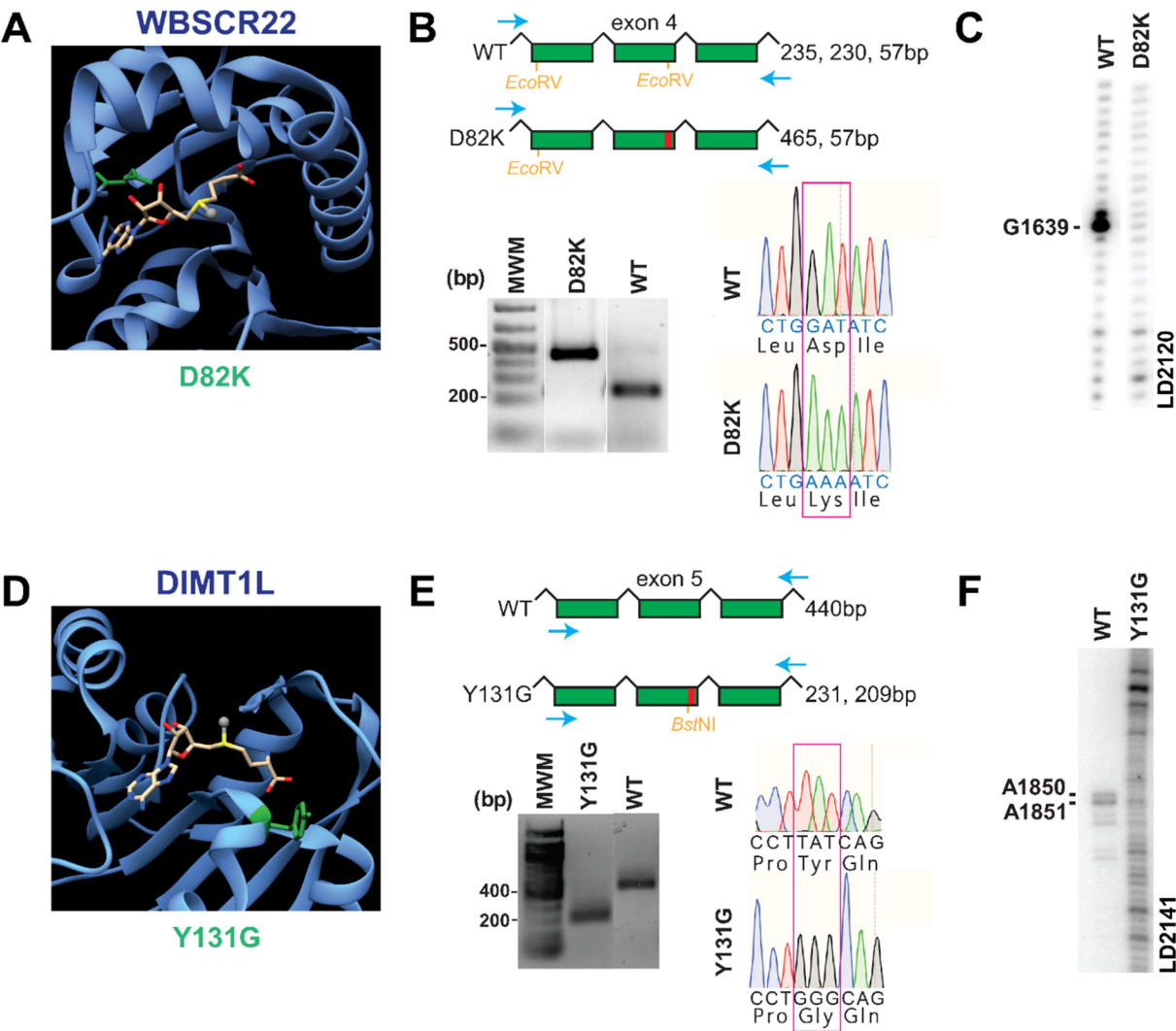


Figure 2. Generation and validation of human cell lines expressing a catalytically inactive form of the 18S rRNA methyltransferases WBSCR22 or DIMT1L. A mutation encoding a single amino-acid substitution in the catalytic pocket of the methyltransferase (Y131G for DIMT1L, D82K for WBSCR22), was introduced on both alleles of HCT116 cells by CRISPR-Cas9 genome editing. (A) 3-D model based on the yeast homolog of WBSCR22, BUD23 (PDB 4QTU). The mutated residue is highlighted in green in the catalytic pocket of the enzyme. The methyl donor cofactor, S-adenosyl methionine (SAM), is depicted in stick representation with the methyl group to be transferred as a grey sphere. (B) Diagnostic evaluation by differential restriction of PCR fragments, and by DNA sequencing. The PCR products were amplified from genomic DNA extracted from the mutant cell line and the isogenic control (WT). The primers used (blue arrows) and the size of the expected fragments upon *EcoRV* digestion are shown. Red box, catalytic domain. Restriction digests were analysed by agarose electrophoresis followed by ethidium bromide staining. DNA sequencing profiles at the mutation site are shown. (C) Loss of RNA modification in the mutant cell line was confirmed by primer extension assay performed with oligo LD2120 on total RNA cleaved at m⁷G following treatment with NaBH₄ and aniline as in [52]. (D) 3-D model based on human DIMT1L (PDB 1ZQ9), coloured elements as in A. (E) Diagnostic evaluation of differential restriction of PCR fragments after *BstNI* digestion as in B. (F) For DIMT1L, the primer extension was performed on total RNA with oligonucleotide LD2141 as in [9]. Loss of modification was further confirmed by HPLC analysis (data not shown).

(Figure 2C) following NaBH_4 -aniline treatment [44]. In case of DIMT1L, the Y131 residue, located in the catalytic pocket of the enzyme, was substituted by a glycine (Figure 2D). The introduction of the mutation resulted in the gain of a *Bst*NI restriction site (Figure 2E). A primer extension assay revealed that the DIMT1L Y131G variant led to loss of the $\text{A}_{1850/1851}$ double dimethylation (Figure 2F), a result which was further confirmed by HPLC analysis (data not shown). Note that the double m^6A modification is rather ‘bulky’ and is naturally causing a reverse transcription drop-off. For both WBSR22 and DIMT1L, the bi-allelic knock-in was confirmed by DNA sequencing of the region of interest (Figure 2B,E).

These cell lines were employed for pairwise comparisons with the isogenic WT control (Figure 3). We detected characteristic base calling errors in the WT control in all analysed cases as higher Mismatch frequencies in IGV snapshots (Figure 3A–C) in comparison to WT samples. However, different modification types affected base calling differently. Strikingly, $\text{m}^7\text{G}_{1639}$ affected basecalling not only at the actual modified position but also strongly at the neighbouring residues (–3 to +3), providing a powerful signature (Figure 3A). On the other hand, the m^6A and m^5A modifications resulted mainly in basecalling errors at the target site (Figure 3B,C).

The pairwise comparison of WT and KO/MUT cell lines with JACUSA2 call-2 and 1,000 reads per condition revealed

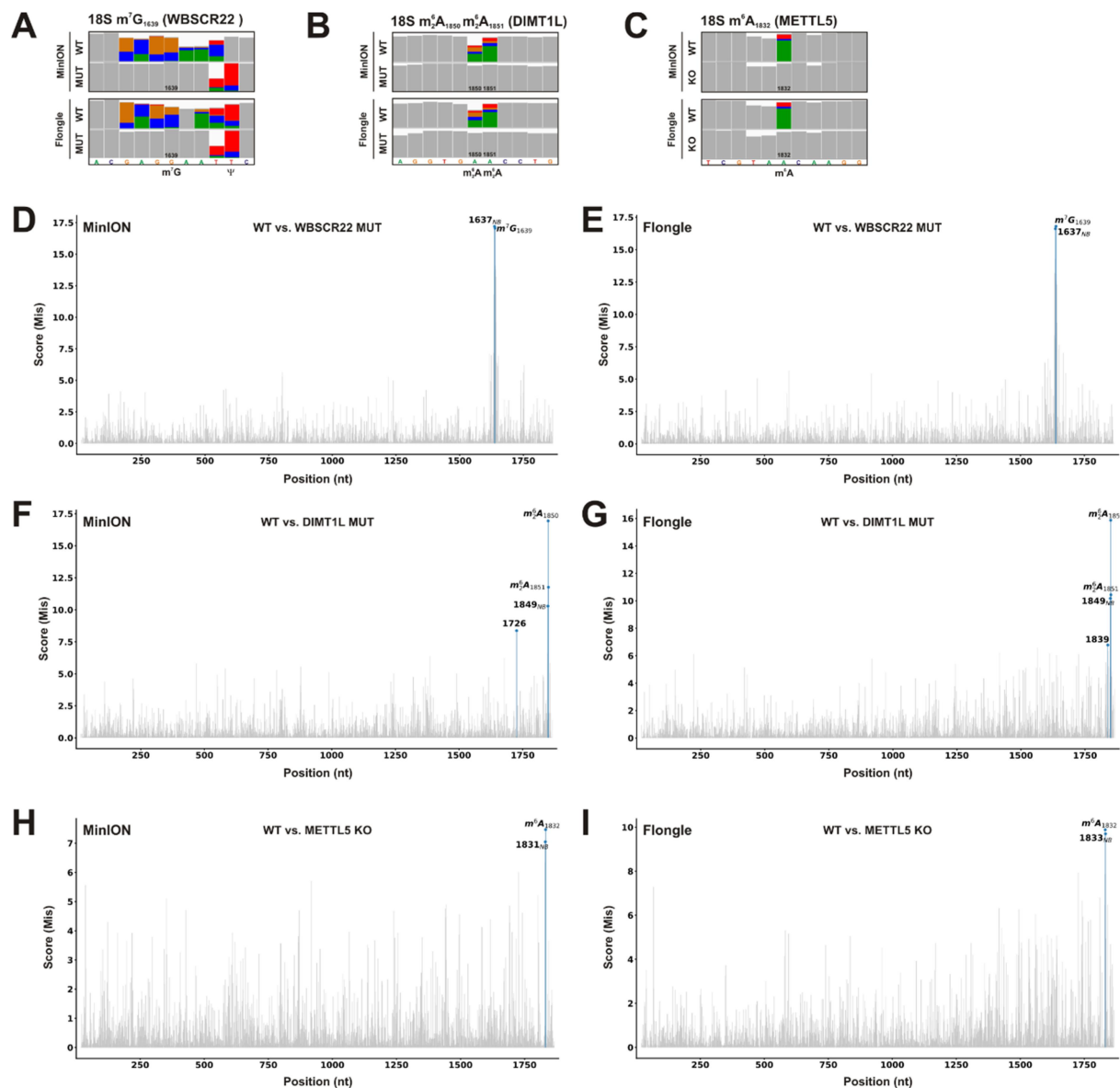


Figure 3. Detection of site-specific modifications in genetically engineered HCT116 cells by nanopore direct rRNA-seq and JACUSA2 call-2. HCT116 cells as listed in Table 1 were subjected to nanopore direct rRNA-seq, either on a MinION or Flongle flow cell as indicated. (A–C) IGV snapshots of the regions of interest from MinION or Flongle sequencing as indicated. The target site as well as other described modifications are annotated. Allele frequency threshold = 0.2. (D,F,H) Barplots of the pairwise comparisons of MinION derived data by JACUSA2 call-2 considering the Mismatch score. Significant outliers detected by Local outlier Factorization (contamination value = 0.001 for WBSR22 and METTL5, 0.002 for DIMT1L) are labelled in blue. Outliers in the 5-mer context of the target site are marked by ‘NB’. (E, G, I) Barplots of the pairwise comparisons of Flongle derived data, as in D,F,H. (D,E) analysis of 18S $\text{m}^7\text{G}_{1639}$ employing HCT116 WBSR22^{D82K/D82K}. (F,G) analysis of 18S $\text{m}^6\text{A}_{1850}$ $\text{m}^5\text{A}_{1851}$ employing HCT116 DIMT1L^{Y131G/Y131G}. (H,I) analysis of 18S $\text{m}^6\text{A}_{1832}$ employing HCT116 METTL5^{–/–}.

high JACUSA2 Mis scores for the target sites, but not for non-related positions (Figure 3D-I). Here, we considered modifications as outliers (i.e. unusual JACUSA2 scores within a set of positions) employing the previously introduced Local Outlier Factorization (LOF) [42] to rank positions by their degree of outlieriness (see Methods). The LOF approach has two hyperparameters: neighbourhood size and contamination. The contamination value determines the proportion of points with the highest LOF scores to be called as outliers. As we expect to find only one or two outliers, we applied here strict contamination values of 0.001 (WBSCR22, METTL5) and 0.002 (DIMIT1L), respectively, which corresponds to 0.1% and 0.2% of sites. Strikingly, for all analysed modifications, the target site as well as one to two adjacent sites were identified as significant outliers (blue) (Figure 3D-I).

Here, we used the JACUSA2 Mis score to call modification sites. However, we would like to note that depending on the data and modification type, also the Deletion and Insertion score as well as the 5-mer context may improve modification calling (Table S5).

Current input requirements may preclude the use of nanopore direct RNA-seq for samples with limited availability. Furthermore, multiplexing of samples with barcoding is currently not officially supported for direct RNA-seq by ONT, increasing the cost for direct RNA-seq experiments. To overcome these problems, we aimed to transfer the targeted direct rRNA-seq approach described above from the standard MinION flow cells (512 channels with 4 pores each) to the recently introduced smaller Flongle flow cells (126 pores), which require less RNA input. The overall results of the JACUSA2 analysis were highly comparable for MinION and Flongle flow cell-derived data (Figure 3 compares panels D, F, and H with E, G, and I). Also, the IGV snapshots for the analysed modification sites were remarkably comparable between MinION and Flongle data (Figure 3A-C, compare upper and lower panels).

In summary, all analysed base methylation sites on the 18S rRNA were detected with JACUSA2 in the MinION flow cell data as well as in the Flongle flow cell data. In addition, the generated mismatch profile was highly similar between MinION and Flongle sequencing. Overall, the Flongle-based approach enables the analysis of low input samples such as patient-derived material making it amenable to clinical biology.

The optimal coverage is determined by the modification type

To exclude the impact of coverage on our analysis, we used 1,000 reads throughout this work to analyse differential RNA modifications (Figure 3). To elucidate whether a higher number of reads would be beneficial for the analysis, or if an even lower number of reads may be useful, we sampled different read numbers from our MinION datasets. As expected, the difference between the JACUSA2 score of the target site and the median JACUSA2 score increases, when more reads are considered for analysis (Figure S6A-C, left panels). To evaluate the robustness of outlier identification, we calculated the

normalized distance of the target site LOF scores to the median LOF score (Figure S6A-C, right panels). Surprisingly, we noticed a small decrease in the normalized LOF score distance for WBSCR22 (Figure S6A, right panel), which was, on the other hand, accompanied by a decrease in the standard deviation at higher read numbers. Robustness in the detection of the METTL5-catalysed m⁶A modification was increased by higher read numbers, as indicated by the increased normalized LOF score distance and decreased standard deviation at 5,000 and 10,000 reads, respectively (Figure S6C, right panel). For the DIMIT1L target sites, the identification is mostly independent of the number of analysed reads (Figure S6B). In summary, 1,000 reads are sufficient to detect the analysed modifications. In most cases, more reads are not beneficial and less reads are sufficient.

Nanopore direct rRNA-seq is suitable to estimate modification levels

As described above, in a clear cut situation, when a cell line harbouring a knock-out (KO) or catalytically dead variant (MUT) of an enzyme is compared to a wild-type cell line (Figure 3, Figure S6), 1,000 reads were well suited for the analysis of differential RNA modifications. However, in physiological or pathological contexts, more subtle changes in modification levels are often expected. We were therefore interested to learn to what extent rRNA modifications could also be analysed at substoichiometric levels with our approach. We approached this question both experimentally and *in silico* by mixing either RNA or sequencing reads from WT and KO/MUT at various ratios. We first sampled *in silico* a total of 1,000 reads from all samples. For the reference 'Mix' sample, different ratios of WT and KO/MUT reads were bioinformatically mixed as indicated (Figure 4A-D). Five replicate samples from all mixing ratios were analysed by pairwise JACUSA2 comparison as outlined above. Interestingly, m⁷G₁₆₃₉ and m⁶A_{1850/1851} were consistently detected with only 5–10% modified reads, whereas m⁶A₁₈₃₂ had a detection threshold of around 25%. Importantly, an increase in the JACUSA2 score with increasing modification frequency was detected for all analysed modification sites (Figure 4A-D), indicating that nanopore direct rRNA-seq can also be used for estimation of modification levels. For the m⁷G and m⁶A₁₈₅₀ modifications, the JACUSA2 score seems to approach saturation (Figure 4A,C).

To validate these findings, experimental mixing analyses were performed with WBSCR22 MUT and METTL5 KO RNAs (Figure 4A,B). Importantly, the results of the experimental mixing analyses were comparable to the *in silico* mixing analyses. For both analysed modification types, the JACUSA2 score increases with increasing modification levels and approached saturation at higher modification levels (Figure 4A,B). In conclusion, we show that nanopore direct rRNA-seq in combination with JACUSA2 analysis can be used to estimate modification levels providing an appropriate calibration curve is established that may also be generated *in silico*.

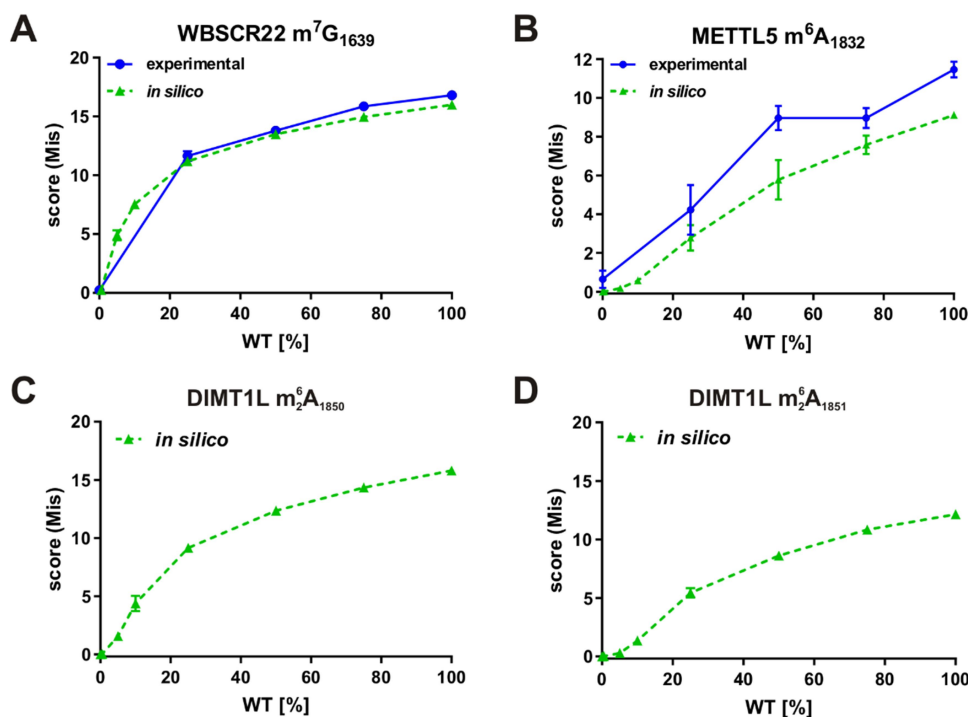


Figure 4. Analysis of the influence of modification levels on the JACUS2A score determined by *in silico* or experimental mixing analysis. For the *in silico* mixing analysis 1,000 reads were downsampled from the MinION sequencing data shown in Figure 3. The ‘Mix’ sample was composed of modified (WT) and unmodified (KO/MUT) reads as indicated that were derived from the downsampled data with 5 different seeds. For the experimental mixing analysis, RNA from WT and WBSCR22 MUT (A) or METTL5 KO (B) cell lines was mixed in the indicated amounts and subjected to direct rRNA-seq on Flongle flow cells. As for the *in silico* mixing, they were analysed with 1,000 sampled reads. JACUS2A call-2 analysis considering the Mismatch score. (A) Analysis of 18S m⁷G₁₆₃₉ in HCT116 WT and WBSCR22 MUT cells. (B) Analysis of 18S m⁶A₁₈₃₂ in HCT116 WT and METTL5 KO cells. (C,D) analysis of 18S m²A₁₈₅₀ or m²A₁₈₅₁, respectively, employing HCT116 WT and DIMT1L MUT cells.

Beyond methylation – detection of acetylcytidine

In human cells, NAT10, which is essential for pre-18S rRNA processing, works together with SNORD13 (U13) to install the acetylation of 18S at position C₁₈₄₂ [43,50,51], which is one of the two ac⁴C modifications found on the 18S rRNA. We generated an HCT116 SNORD13 KO cell line by CRISPR-Cas9 genome editing (Figure S7A,B). The loss of C₁₈₄₂ acetylation in this cell line was validated by a primer extension assay (Figure S7C) and direct rRNA-seq – JACUS2A analysis that identified ac⁴C₁₈₄₂ as a *bona fide* outlier (Figure 5A,B).

As the identification of small changes in rRNA modification levels is presumably of biological and clinical interest, we compared our approach to an orthogonal method, namely a C-to-T misincorporation assay based on borohydride reduction [45–47]. To simulate a scenario of small changes in acetylation, we analysed WT/SNORD13 KO mixtures with only 10%, 20%, and 30% of KO RNA (Figure 5C,D). Nanopore direct rRNA-seq revealed a gradual increase of JACUS2A scores in pairwise comparison to the WT RNA (Figure 5C). Interestingly, JACUS2A is particularly sensitive to relatively small changes in C₁₈₄₂ acetylation. On the other hand, the C-to-T misincorporation assay could not discriminate between 100% WT and 90% WT/10% SNORD13 KO, revealing an inherent limitation of this approach (Figure 5D).

We therefore concluded that nanopore direct rRNA-seq can detect even small changes in modification levels, whereas the misincorporation assay is only suitable to detect larger differences.

In summary, we established the nanopore-based targeted direct RNA-sequencing of human ribosomal RNAs and analysed widespread rRNA modifications employing JACUS2A. We show that direct rRNA-seq on the nanopore can be scaled down to the Flongle device, enabling the analysis of clinical or biological samples with limited availability.

Discussion

Modifications of human rRNAs have been characterized for many years, and until recently were considered as relatively constitutive. Emerging evidence however supports the hypothesis of ribosomal heterogeneity [53], which includes the production of differentially modified ribosomes that may contribute to the aetiology and progression of several diseases including cancer, developmental and cardiovascular disorders.

Thus, there is a need to develop techniques for the in-depth characterization of the entire rRNA modification repertoire in a quantitative fashion. Ideally, such techniques should be amenable to low input material and high throughput analysis.

Consequently, the sequence specific and quantitative analyses of various RNA modification types with moderate amounts of input material are progressing fast with the advent of the direct RNA-sequencing method. However, current analyses have focused mostly on highly specific types of RNA modifications as m⁶A and pseudouridine [28,30,31,34–36], or did not address dynamic changes in human samples with clinical relevance [32,33].

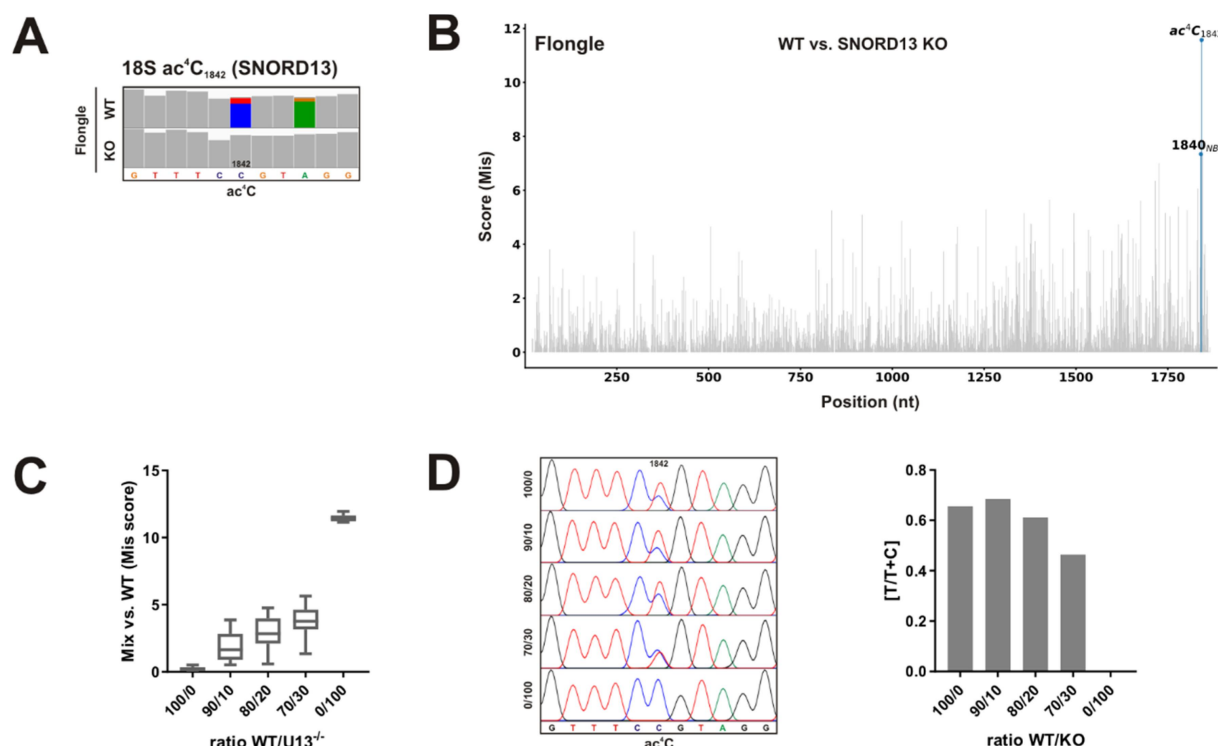


Figure 5. Analysis of 18S ac⁴C₁₈₄₂ by nanopore direct rRNA-seq and an orthogonal method employing HCT116 SNORD13 KO cells. (A) IGV snapshots of the region of interest from HCT116 WT and SNORD13 KO cells as indicated. The target site is annotated. Allele frequency threshold = 0.18. (B) Barplots of the pairwise JACUSA2 call-2 analysis. Significant outliers detected by LOF (contamination value = 0.001) are labelled in blue. JACUSA2 Mismatch score was considered. (C) JACUSA2 Mis scores of pairwise comparisons of different experimental RNA mixes against the HCT116 WT RNA. Reads were downsampled to 1,000 reads with different seeds ($n = 15$). (D) 18S helix 45 ac⁴C misincorporation assay based on NaCNBH₃ reduction and TGIRT-III reverse transcription for RNA mixtures as indicated. Left panel: Sanger sequencing traces. Right panel: quantification of the C-to-T misincorporation.

Here, we established the targeted direct RNA-seq of human ribosomal RNAs with a selection step involving the use of custom adapters (direct rRNA-seq). Our targeted approach ensures that only properly processed rRNAs with defined 3' ends are sequenced [54–56], and no unfaithfully processed RNAs (Figure S1). We benchmarked our JACUSA2-based approach against other tools designed to detect RNA modifications in direct RNA-seq data and demonstrate that JACUSA2 may be considered especially for the detection of pseudouridines and the diverse class of 'other' modifications (Figure 1E, Figure S2), whereas the detection of 2'-O-ribose methylation sites by different tools differs substantially between the different nucleobases (Figure S3). In future work, there will be a need to develop more sophisticated models that also consider, for example, the context. Based on the comparison of HCT116 WT and IVT samples, we identified almost all known modification sites on the basis of positive JACUSA2 Mis scores (Figure 1E). This validates the general ability of JACUSA2 to identify differential rRNA modifications of all types (pseudouridine, 2'-O methylation of ribose, m¹acp³psU, ac⁴C, m⁷G, m⁶A, m₂A on the 18S rRNA). However, the main application of JACUSA2 and the comprehensive workflow presented here is the comparative analysis of biological or clinical samples that differ in a limited number of modification sites. In line with previous findings [31,32], we show that Mismatches are the main determinant for the identification of modification sites (Figure 1E). In specific cases, for example Cm modifications, also the Deletion and Insertion scores as well as the 5-mer context may

be considered (Figure S5). The rationale of taking the 5-mer context into account is based on the fact that the current nanopore protein (R9.4.1) covers five nucleotides at a time. This implies that in case of an RNA modification, also the basecalling at neighbouring sites may be altered.

It is important to emphasize that our prediction of modification sites with JACUSA2 does not rely on any prior knowledge, as we only use modification maps (such as those established experimentally by mass spectrometry by Taoka *et al.* [8]) for *post hoc* evaluation of our results. As for every nanopore error-profile-based approach, the chemical nature of *de novo* identified RNA modification sites needs to be determined by other methods.

Employing a collection of genetically engineered HCT116 cell lines lacking individual modifications (Figure 2), we demonstrated that the recently introduced JACUSA2 software [31] detects all targeted 18S rRNA base methylations in sequencing data produced with the standard MinION flow cells, namely m⁷G₁₆₃₉, m⁶A₁₈₃₂, and the double dimethylation m₂A₁₈₅₀ m₂A₁₈₅₁. These base modifications cause distinct basecalling errors that are reflected by the Mismatch scores calculated by JACUSA2 (Figure 3). Importantly, we could detect rRNA modifications just as efficiently with the small-scale Flongle flow cells (Figure 3), enabling the analysis of precious material available only in limited amounts. This is of particular interest as barcoding of direct RNA-seq libraries is currently not officially supported by Oxford Nanopore Technologies. Sequencing on the Flongle flow cells yields data with comparable quality. All base methylations were

identified as significant outliers in this analysis based on 1,000 reads employing the LOF method. The LOF method determines sites that are significantly different from their neighbours as outliers, thus we recommend it for outlier detection when the expected number of differential sites is known or expected to be small. In case the number of differentially modified sites cannot be estimated *a priori*, the contamination value (which defines how many sites are identified as outliers), can be fitted automatically by `sklearn.neighbors.LocalOutlierFactor`, as described [42]. We do not recommend the LOF method for cases with a high number of differential sites as the WT vs. IVT comparison. Here, we recommend to identify candidate sites based on the JACUSA2 score (see Figure 1E). By downsampling analyses of different read numbers from the MinION data, we show that, remarkably, as little as 300 to 500 reads were sufficient to identify the analysed modification sites (Figure S6). With exception of the METTL5-catalysed m⁶A modification, more reads did not improve results.

We show that the relative modification levels of selected base modifications can be estimated from the JACUSA2 scores based on a calibration curve generated experimentally or *in silico* (Figure 4). In most analysed cases, the JACUSA2 score approaches saturation at higher modification levels (Figure 4A–D); thus, an estimation of modification levels should always be based on an appropriate calibration curve. Only in case of METTL5-catalysed m⁶A₁₈₃₂, we observed some offset between experimental and *in silico* mixing data (Figure 4B). We speculate that this may be caused by slight variations of the 18S m⁶A₁₈₃₂ level in the different biological replicates of HCT116 WT samples used.

Besides the different base methylation sites, the human 18S rRNA harbours in addition two acetylated cytidines. To further expand the repertoire of modifications that may be analysed by nanopore direct rRNA-seq and JACUSA2, we made use of an HCT116 SNORD13 KO cell line. SNORD13 works together with NAT10 to install 18S ac⁴C₁₈₄₂ [43,50,51]. Importantly, ac⁴C₁₈₄₂ was identified as a *bona fide* outlier in the comparison of SNORD13 KO cells to HCT116 WT cells (Figure 5A,B). Nanopore direct rRNA-seq only requires small amounts of input material and is suitable to identify changes in different modification types at a time. However, we were as well interested in the sensitivity of the method to identify also small changes in modifications that may occur in samples with biological or clinical relevance. To this end, we compared JACUSA2 to an orthogonal validation method based on borohydride reduction of ac⁴C and subsequent reverse transcription misincorporation (Figure 5D). Interestingly, mixture experiments with only small amounts of KO RNA, revealed that direct rRNA-seq – JACUSA2 is particularly sensitive to small changes in RNA modification (Figure 5).

Although we did not cover the complete repertoire of rRNA modifications by dedicated KO cell lines at the time, our data suggest that our workflow captures many if not all rRNA modifications through basecalling error (Mismatch, Insertion, Deletion) analysis.

In this work, we applied JACUSA2 call-2 in pairwise comparisons to identify differences in rRNA modification. A number of other computational pipelines are already available in the literature to analyse rRNA modifications in

nanopore direct RNA-seq data and were compared here to JACUSA2 (Figure S2), including Nanocompore [36], EpiNano [35], xPore [34], and Eligos2 [30], which are either based on the detection of basecalling errors, such as in the case of JACUSA2, or which infer modification pattern from changes in the current signal traces. A recently introduced approach, based on signalAlign, estimates modification probabilities on a single read level, but is limited to and only looks at previously described modification sites [33]. Furthermore, some methods are until now only established for specific modification types as m⁶A or pseudouridine [28,32]. The major challenge in the analysis of RNA modifications compared to DNA modifications is the diversity of modification types. Additionally, the expression of different rRNA variants may complicate the analysis.

Importantly, JACUSA2 does not depend on prior knowledge or on training of complex models to identify modification sites. However, in this work, we took advantage of the accurate mapping of rRNA modification sites by mass spectrometry for *post hoc* validation of our findings. In contrast to some other workflows, JACUSA2 supports the handling of replicate samples as well as pairwise comparison [31], enabling either the comparison to an unmodified reference sequence or the identification of differences between biological and clinical samples. JACUSA2 can be used with both Guppy basecalling modes (fast vs. HAC), but of course all analysed samples should be basecalled identically. For the detection of methylation and acetylation, we recommend the fast basecalling mode (Figure S2). We show here that JACUSA2 is not limited to specific modification types but can potentially be used to map and quantify all modification types that cause basecalling errors (Mismatch, Deletion, Insertion). Thus, JACUSA2 could be applied also to the detection of new modification sites, whose chemical nature could be later determined by other methods. It is noteworthy to mention that when we performed comparative analysis we realized that JACUSA2 was also less demanding on computing time than many other algorithms available to date (Table S4).

In conclusion, we established a targeted nanopore direct RNA-seq strategy for human rRNA and detection of modifications by JACUSA2. The down scaling to Flongle flow cells enables the study of samples with limited availability. Future work will focus on the identification of differential modification sites in samples of biological interest and clinical relevance and to understand the biological consequences of altered modification patterns.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the Klaus Tschira Foundation (grant 00.013.2021 to CD), TRR 319 - RMaP (DFG INST 247/1042-1 to CD). Research in the lab of DLJL is supported by the Belgian Fonds de la Recherche Scientifique (F.R.S./FNRS), the Université libre de Bruxelles (ULB), the European Joint Programme on Rare Diseases (RIBOEurope and DBAGeneCure), and the Region Wallonne (SPW EER) (RIBOgenesis).

Data availability statement

Sequencing data generated for this study are available at NCBF's Sequence Read Archive through the BioProject accession number PRJNA781102. A Snakemake pipeline for the analysis workflow was developed and is available on Zenodo (<https://doi.org/10.5281/zenodo.8268171>).

ORCID

Christoph Dieterich  <http://orcid.org/0000-0001-9468-6311>

References

- Bohnsack KE, Bohnsack MT. Uncovering the assembly pathway of human ribosomes and its emerging links to disease. *EMBO J*. 2019;38(13):e100278. doi: [10.15252/embj.2018100278](https://doi.org/10.15252/embj.2018100278)
- Badertscher L, Wild T, Montellese C, et al. Genome-wide RNAi Screening Identifies protein modules required for 40S subunit synthesis in human cells. *Cell Rep*. 2015;13(12):2879–2891. doi: [10.1016/j.celrep.2015.11.061](https://doi.org/10.1016/j.celrep.2015.11.061)
- Wild T, Horvath P, Wyler E, et al. A protein inventory of human ribosome biogenesis reveals an essential function of exportin 5 in 60S subunit export. *PLoS Biol*. 2010;8(10):e1000522. doi: [10.1371/journal.pbio.1000522](https://doi.org/10.1371/journal.pbio.1000522)
- Farley-Barnes KI, McCann KL, Ogawa LM, et al. Diverse regulators of human ribosome biogenesis discovered by changes in nucleolar number. *Cell Rep*. 2018;22(7):1923–1934. doi: [10.1016/j.celrep.2018.01.056](https://doi.org/10.1016/j.celrep.2018.01.056)
- Tafforeau L, Zorbas C, Langhendries JL, et al. The complexity of human ribosome biogenesis revealed by systematic nucleolar screening of pre-rRNA processing factors. *Mol Cell*. 2013;51(4):539–551. doi: [10.1016/j.molcel.2013.08.011](https://doi.org/10.1016/j.molcel.2013.08.011)
- Sharma S, Lafontaine DLJ. 'View from a bridge': a new perspective on eukaryotic rRNA base modification. *Trends Biochem Sci*. 2015;40(10):560–575. doi: [10.1016/j.tibs.2015.07.008](https://doi.org/10.1016/j.tibs.2015.07.008)
- Sloan KE, Warda AS, Sharma S, et al. Tuning the ribosome: The influence of rRNA modification on eukaryotic ribosome biogenesis and function. *RNA Biology*. 2017;14(9):1138–1152. doi: [10.1080/15476286.2016.1259781](https://doi.org/10.1080/15476286.2016.1259781)
- Taoka M, Nobe Y, Yamaki Y, et al. Landscape of the complete RNA chemical modifications in the human 80S ribosome. *Nucleic Acids Res*. 2018;46(18):9289–9298. doi: [10.1093/nar/gky811](https://doi.org/10.1093/nar/gky811)
- Zorbas C, Nicolas E, Wacheul L, et al. The human 18S rRNA base methyltransferases DIMT1L and WBSCR22-TRMT112 but not rRNA modification are required for ribosome biogenesis. *MBoC*. 2015;26(11):2080–2095. doi: [10.1091/mbc.E15-02-0073](https://doi.org/10.1091/mbc.E15-02-0073)
- Watkins NJ, Bohnsack MT. The box C/D and H/ACA snoRnps: key players in the modification, processing and the dynamic folding of ribosomal RNA. *Wiley Interdiscip Rev RNA*. 2012;3(3):397–414. doi: [10.1002/wrna.117](https://doi.org/10.1002/wrna.117)
- Helm M. Post-transcriptional nucleotide modification and alternative folding of RNA. *Nucleic Acids Res*. 2006;34(2):721–733. doi: [10.1093/nar/gkj471](https://doi.org/10.1093/nar/gkj471)
- Penzo M, Montanaro L. Turning Uridines around: role of rRNA pseudouridylation in ribosome biogenesis and ribosomal function. *Biomolecules*. 2018;8(2):38. doi: [10.3390/biom8020038](https://doi.org/10.3390/biom8020038)
- Liang XH, Liu Q, Fournier MJ. Loss of rRNA modifications in the decoding center of the ribosome impairs translation and strongly delays pre-rRNA processing. *RNA*. 2009;15(9):1716–1728. doi: [10.1261/rna.1724409](https://doi.org/10.1261/rna.1724409)
- Monaco PL, Marcel V, Diaz JJ, et al. 2'-O-Methylation of ribosomal RNA: towards an epitranscriptomic control of translation? *Biomolecules*. 2018;8(4):doi: [10.3390/biom8040106](https://doi.org/10.3390/biom8040106)
- Kampen KR, Sulima SO, Vereecke S, et al. Hallmarks of ribomopathies. *Nucleic Acids Res*. 2020;48(3):1013–1028. doi: [10.1093/nar/gkz637](https://doi.org/10.1093/nar/gkz637)
- Mason PJ, Bessler M. The genetics of dyskeratosis congenita. *Cancer Genet*. 2011;204(12):635–645. doi: [10.1016/j.cancergen.2011.11.002](https://doi.org/10.1016/j.cancergen.2011.11.002)
- Mitchell JR, Wood E, Collins K. A telomerase component is defective in the human disease dyskeratosis congenita. *Nature*. 1999;402(6761):551–555. doi: [10.1038/990141](https://doi.org/10.1038/990141)
- Penzo M, Rocchi L, Brugiere S, et al. Human ribosomes from cells with reduced dyskerin levels are intrinsically altered in translation. *FASEB J*. 2015;29(8):3472–3482. doi: [10.1096/fj.15-270991](https://doi.org/10.1096/fj.15-270991)
- Jack K, Bellodi C, Landry DM, et al. rRNA pseudouridylation defects affect ribosomal ligand binding and translational fidelity from yeast to human cells. *Mol Cell*. 2011;44(4):660–666. doi: [10.1016/j.molcel.2011.09.017](https://doi.org/10.1016/j.molcel.2011.09.017)
- Stults DM, Killen MW, Pierce HH, et al. Genomic architecture and inheritance of human ribosomal RNA gene clusters. *Genome Res*. 2008;18(1):13–18. doi: [10.1101/gr.6858507](https://doi.org/10.1101/gr.6858507)
- Parks MM, Kurylo CM, Dass RA, et al. Variant ribosomal RNA alleles are conserved and exhibit tissue-specific expression. *Sci Adv*. 2018;4(2):eaao0665. doi: [10.1126/sciadv.aao0665](https://doi.org/10.1126/sciadv.aao0665)
- Rothschild D, Susanto TT, Spence JP, et al. (2023) A comprehensive rRNA variation atlas in health and disease. *bioRxiv*. 2023.2001.2030.526360.
- Birkedal U, Christensen-Dalsgaard M, Krogh N, et al. Profiling of ribose methylations in RNA by high-throughput sequencing. *Angew Chem Int Ed Engl*. 2015;54:451–455. doi: [10.1002/anie.201408362](https://doi.org/10.1002/anie.201408362)
- Marchand V, Pichot F, Neybecker P, et al. HydraPsiSeq: a method for systematic and quantitative mapping of pseudouridines in RNA. *Nucleic Acids Res*. 2020;48(19):e110. doi: [10.1093/nar/gkaa769](https://doi.org/10.1093/nar/gkaa769)
- Garalde DR, Snell EA, Jachimowicz D, et al. Highly parallel direct RNA sequencing on an array of nanopores. *Nat Methods*. 2018;15(3):201–206. doi: [10.1038/nmeth.4577](https://doi.org/10.1038/nmeth.4577)
- Workman RE, Tang AD, Tang PS, et al. Nanopore native RNA sequencing of a human poly(A) transcriptome. *Nat Methods*. 2019;16(12):1297–1305. doi: [10.1038/s41592-019-0617-2](https://doi.org/10.1038/s41592-019-0617-2)
- Jenjaroenpun P, Wongsurawat T, Pereira R, et al. Complete genomic and transcriptional landscape analysis using third-generation sequencing: a case study of *Saccharomyces cerevisiae* CEN.PK113-7D. *Nucleic Acids Res*. 2018;46(7):e38. doi: [10.1093/nar/gky014](https://doi.org/10.1093/nar/gky014)
- Liu H, Begik O, Lucas MC, et al. Accurate detection of m(6)A RNA modifications in native RNA sequences. *Nat Commun*. 2019;10(1):4079. doi: [10.1038/s41467-019-11713-9](https://doi.org/10.1038/s41467-019-11713-9)
- Smith AM, Jain M, Mulrone L, et al. Reading canonical and modified nucleobases in 16S ribosomal RNA using nanopore native RNA sequencing. *PLoS One*. 2019;14(5):e0216709. doi: [10.1371/journal.pone.0216709](https://doi.org/10.1371/journal.pone.0216709)
- Jenjaroenpun P, Wongsurawat T, Wadley TD, et al. Decoding the epitranscriptional landscape from native RNA sequences. *Nucleic Acids Res*. 2021;49(2):e7. doi: [10.1093/nar/gkaa620](https://doi.org/10.1093/nar/gkaa620)
- Piechotta M, Naarmann-de Vries IS, Wang Q, et al. RNA modification mapping with JACUSA2. *Genome Biol*. 2022;23(1):115. doi: [10.1186/s13059-022-02676-0](https://doi.org/10.1186/s13059-022-02676-0)
- Begik O, Lucas MC, Pryszcz LP, et al. Quantitative profiling of pseudouridylation dynamics in native RNAs with nanopore sequencing. *Nat Biotechnol*. 2021;39(10):1278–1291. doi: [10.1038/s41587-021-00915-6](https://doi.org/10.1038/s41587-021-00915-6)
- Bailey AD, Talkish J, Ding H, et al. Concerted modification of nucleotides at functional centers of the ribosome revealed by single-molecule RNA modification profiling. *Elife*. 2022;11:doi: [10.7554/eLife.76562](https://doi.org/10.7554/eLife.76562)
- Pratanwanich PN, Yao F, Chen Y, et al. Identification of differential RNA modifications from nanopore direct RNA sequencing with xPore. *Nat Biotechnol*. 2021;39(11):1394–1402. doi: [10.1038/s41587-021-00949-w](https://doi.org/10.1038/s41587-021-00949-w)
- Liu H, Begik O, Novoa EM. EpiNano: Detection of m(6)A RNA modifications using Oxford Nanopore direct RNA sequencing. *Methods Mol Biol*. 2021;2298:31–52.
- Leger A, Amaral PP, Pandolfini L, et al. RNA modifications detection by comparative Nanopore direct RNA sequencing. *Nat Commun*. 2021;12(1):7198. doi: [10.1038/s41467-021-27393-3](https://doi.org/10.1038/s41467-021-27393-3)

- [37] van Tran N, Ernst FGM, Hawley BR, et al. The human 18S rRNA m6A methyltransferase METTL5 is stabilized by TRMT112. *Nucleic Acids Res.* **2019**;47(15):7719–7733. doi: [10.1093/nar/gkz619](https://doi.org/10.1093/nar/gkz619)
- [38] Lafontaine D, Vandenhoute J, Tollervey D. The 18S rRNA dimethylase Dim1p is required for pre-ribosomal RNA processing in yeast. *Genes Dev.* **1995**;9(20):2470–2481. doi: [10.1101/gad.9.20.2470](https://doi.org/10.1101/gad.9.20.2470)
- [39] Lafontaine DL, Preiss T, Tollervey D. Yeast 18S rRNA dimethylase Dim1p: a quality control mechanism in ribosome synthesis? *Mol Cell Biol.* **1998**;18(4):2360–2370. doi: [10.1128/MCB.18.4.2360](https://doi.org/10.1128/MCB.18.4.2360)
- [40] White J, Li Z, Sardana R, et al. Bud23 methylates G1575 of 18S rRNA and is required for efficient nuclear export of pre-40S subunits. *Mol Cell Biol.* **2008**;28(10):3151–3161. doi: [10.1128/MCB.01674-07](https://doi.org/10.1128/MCB.01674-07)
- [41] Haag S, Kretschmer J, Bohnsack MT. WBSR22/Merm1 is required for late nuclear pre-ribosomal RNA processing and mediates N7-methylation of G1639 in human 18S rRNA. *RNA.* **2015**;21(2):180–187. doi: [10.1261/rna.047910.114](https://doi.org/10.1261/rna.047910.114)
- [42] Breunig MM, Kriegl H-P, Ng RT, et al. LOF: identifying density-based local outliers. In: *Proceedings of the 2000 ACM SIGMOD international conference on Management of data*; Dallas, Texas, USA. Association for Computing Machinery; **2000**. p. 93–104.
- [43] Sharma S, Langhendries JL, Watzinger P, et al. Yeast Kre33 and human NAT10 are conserved 18S rRNA cytosine acetyltransferases that modify tRNAs assisted by the adaptor Tan1/THUMP1. *Nucleic Acids Res.* **2015**;43(4):2242–2258. doi: [10.1093/nar/gkv075](https://doi.org/10.1093/nar/gkv075)
- [44] Letoquart J, Huvelle E, Wacheul L, et al. Structural and functional studies of Bud23–Trm112 reveal 18S rRNA N 7 -G1575 methylation occurs on late 40S precursor ribosomes. *Proc Natl Acad Sci U S A.* **2014**;111(51):E5518–5526. doi: [10.1073/pnas.1413089111](https://doi.org/10.1073/pnas.1413089111)
- [45] Thomas JM, Briney CA, Nance KD, et al. A chemical signature for cytidine acetylation in RNA. *J Am Chem Soc.* **2018**;140(40):12667–12670. doi: [10.1021/jacs.8b06636](https://doi.org/10.1021/jacs.8b06636)
- [46] Sas-Chen A, Thomas JM, Matzov D, et al. Dynamic RNA acetylation revealed by quantitative cross-evolutionary mapping. *Nature.* **2020**;583(7817):638–643. doi: [10.1038/s41586-020-2418-2](https://doi.org/10.1038/s41586-020-2418-2)
- [47] Thomas JM, Bryson KM, Meier JL. Nucleotide resolution sequencing of N4-acetylcytidine in RNA. *Methods Enzymol.* **2019**;621:31–51.
- [48] Jain M, Olsen HE, Akeson M, et al. Adaptation of human ribosomal RNA for Nanopore sequencing of Canonical and modified nucleotides. *Methods Mol Biol.* **2021**;2298:53–74.
- [49] Leismann J, Spagnuolo M, Pradhan M, et al. The 18S ribosomal RNA m 6 a methyltransferase Mettl5 is required for normal walking behavior in Drosophila. *EMBO Rep.* **2020**;21(7):e49443. doi: [10.15252/embr.201949443](https://doi.org/10.15252/embr.201949443)
- [50] Ito S, Horikawa S, Suzuki T, et al. Human NAT10 is an ATP-dependent RNA acetyltransferase responsible for N4-acetylcytidine formation in 18 S ribosomal RNA (rRNA). *J Biol Chem.* **2014**;289(52):35724–35730. doi: [10.1074/jbc.C114.602698](https://doi.org/10.1074/jbc.C114.602698)
- [51] Bortolin-Cavaille ML, Quillien A, Thalalla Gamage S, et al. Probing small ribosomal subunit RNA helix 45 acetylation across eukaryotic evolution. *Nucleic Acids Research.* **2022**;50(11):6284–6299. doi: [10.1093/nar/gkac404](https://doi.org/10.1093/nar/gkac404)
- [52] Figaro S, Wacheul L, Schillewaert S, et al. Trm112 is required for Bud23-mediated methylation of the 18S rRNA at position G1575. *Molecular And Cellular Biology.* **2012**;32(12):2254–2267. doi: [10.1128/MCB.06623-11](https://doi.org/10.1128/MCB.06623-11)
- [53] Gay DM, Lund AH, Jansson MD. Translational control through ribosome heterogeneity and functional specialization. *Trends Biochem Sci.* **2022**;47(1):66–81. doi: [10.1016/j.tibs.2021.07.001](https://doi.org/10.1016/j.tibs.2021.07.001)
- [54] Tomecki R, Sikorski PJ, Zakrzewska-Placzek M. Comparison of preribosomal RNA processing pathways in yeast, plant and human cells - focus on coordinated action of endo- and exoribonucleases. *FEBS Lett.* **2017**;591(13):1801–1850. doi: [10.1002/1873-3468.12682](https://doi.org/10.1002/1873-3468.12682)
- [55] Preti M, O'Donohue MF, Montel-Lehry N, et al. Gradual processing of the ITS1 from the nucleolus to the cytoplasm during synthesis of the human 18S rRNA. *Nucleic Acids Res.* **2013**;41(8):4709–4723. doi: [10.1093/nar/gkt160](https://doi.org/10.1093/nar/gkt160)
- [56] Sloan KE, Mattijssen S, Lebaron S, et al. Both endonucleolytic and exonucleolytic cleavage mediate ITS1 removal during human ribosomal RNA processing. *J Cell Bio.* **2013**;200(5):577–588. doi: [10.1083/jcb.201207131](https://doi.org/10.1083/jcb.201207131)