



**HAL**  
open science

## HydraPsiSeq: a method for systematic and quantitative mapping of pseudouridines in RNA

Virginie Marchand, Florian Pichot, Paul Neybecker, Lilia Ayadi, Valérie Bourguignon-Igel, Ludivine Wacheul, Denis Lafontaine, Astrid Pinzano, Mark Helm, Yuri Motorin

► **To cite this version:**

Virginie Marchand, Florian Pichot, Paul Neybecker, Lilia Ayadi, Valérie Bourguignon-Igel, et al.. HydraPsiSeq: a method for systematic and quantitative mapping of pseudouridines in RNA. *Nucleic Acids Research*, 2020, 48 (19), pp.e110. 10.1093/nar/gkaa769 . hal-02957799

**HAL Id: hal-02957799**

**<https://hal.univ-lorraine.fr/hal-02957799v1>**

Submitted on 26 Nov 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# HydraPsiSeq: a method for systematic and quantitative mapping of pseudouridines in RNA

Virginie Marchand<sup>1,\*</sup>, Florian Pichot<sup>1,2</sup>, Paul Neybecker<sup>3</sup>, Lilia Ayadi<sup>1,3</sup>,  
Valérie Bourguignon-Igel<sup>1,3</sup>, Ludivine Wacheul<sup>4</sup>, Denis L.J. Lafontaine<sup>4</sup>, Astrid Pinzano<sup>3</sup>,  
Mark Helm<sup>2</sup> and Yuri Motorin<sup>1,3,\*</sup>

<sup>1</sup>Université de Lorraine, CNRS, INSERM, IBSLor (UMS2008/US40), Epitranscriptomics and RNA Sequencing Core Facility, F54000 Nancy, France, <sup>2</sup>Institute of Pharmaceutical and Biomedical Science, Johannes Gutenberg-University Mainz, 55128 Mainz, Germany, <sup>3</sup>Université de Lorraine, CNRS, IMoPA (UMR7365), F54000 Nancy, France and <sup>4</sup>RNA Molecular Biology, ULB-Cancer Research Center (U-CRC), Center for Microscopy and Molecular Imaging (CMMI), Fonds de la Recherche Scientifique (F.R.S./FNRS), and Université Libre de Bruxelles (ULB), BioPark campus, B-6041 Gosselies, Belgium

Received May 22, 2020; Revised September 02, 2020; Editorial Decision September 03, 2020; Accepted September 06, 2020

## ABSTRACT

Developing methods for accurate detection of RNA modifications remains a major challenge in epitranscriptomics. Next-generation sequencing-based mapping approaches have recently emerged but, often, they are not quantitative and lack specificity. Pseudouridine ( $\psi$ ), produced by uridine isomerization, is one of the most abundant RNA modification.  $\psi$  mapping classically involves derivatization with soluble carbodiimide (CMCT), which is prone to variation making this approach only semi-quantitative. Here, we developed ‘HydraPsiSeq’, a novel quantitative  $\psi$  mapping technique relying on specific protection from hydrazine/aniline cleavage. HydraPsiSeq is quantitative because the obtained signal directly reflects pseudouridine level. Furthermore, normalization to natural unmodified RNA and/or to synthetic *in vitro* transcripts allows absolute measurements of modification levels. HydraPsiSeq requires minute amounts of RNA (as low as 10–50 ng), making it compatible with high-throughput profiling of diverse biological and clinical samples. Exploring the potential of HydraPsiSeq, we profiled human rRNAs, revealing strong variations in pseudouridylation levels at ~20–25 positions out of total 104 sites. We also observed the dynamics of rRNA pseudouridylation throughout chondrogenic differentiation of human bone marrow stem cells. In conclusion, HydraPsiSeq is a robust approach for the systematic mapping and accurate quantification of pseudouridines in RNAs with appli-

cations in disease, aging, development, differentiation and/or stress response.

## INTRODUCTION

Single nucleotide-resolution mapping and precise quantification of RNA modified nucleotides are key challenges in epitranscriptomics (1–4). This information is of crucial importance for analysis of biological functions of these modified RNA residues (5–8), since only precise and reliable knowledge of RNA modification profiles allows their correlation with physiological state of the cell and other factors influencing RNA maturation and functions. Quantification of RNA modification is possible by conventional and low-throughput methods such as HPLC, or combination of LC with MS analysis (9–12), but since those techniques rarely provide unambiguous mapping information, profiling of biological samples mostly relies on high-throughput approaches based on deep sequencing of second (4), and now third generation (13,14). Despite a great success in mapping different RNA modified residues in full transcriptomes, only very few available deep sequencing-based methods for RNA modification mapping combine both single-nucleotide resolution and precise quantification, since multiple enrichment steps and/or various biases during library preparation generally make quantification very complicated, or even impossible.

Deep sequencing-based methods relying on antibody capture (e.g. MeRIP, (15,16)) provide limited quantitative information on the modification level due to variable enrichment (17), but become quantitative upon inclusion of spike-in controls and parallel sequencing of both bound and unbound fractions (18). Protocols using chemical modifica-

\*To whom correspondence should be addressed. Tel: +33 3 72746629; Email: iouri.motorine@univ-lorraine.fr  
Correspondence may also be addressed to Virginie Marchand. Tel: +33 3 72746669; Email: Virginie.Marchand@univ-lorraine.fr

tion strategies without further selection of modified residues generally provide quantitative information on the modification level, in particular if the readout signal is not a full reverse transcriptase (RT)-arrest of primer extension, which is generally considered as being noisy. For instance, quantitative assessment of  $m^5C$  levels is readily achieved in bisulfite-sequencing (BS) and derived protocols, under the assumption that deamination level and  $C \rightarrow U$  conversion is close to 100%, and thus residual non-deaminated signal is essentially generated by modified  $m^5C$  (19–21). Another example of quantification for a common RNA modification is the alkaline hydrolysis-based RiboMethSeq, where protection of the phosphodiester bond between nucleotides N and N+1 correlates with 2'-O-methylation level of the first nucleotide (N). RiboMethSeq thus allows precise profiling of Nm methylation in various RNAs and under varying growth/stress conditions (22–24). Finally, several protocols, such as  $m^1A$  detection by RT misincorporation signatures (25–27), or the AlkAniline-Seq method for  $m^7G$ ,  $m^3C$  and D detection (28), provide quantitative information, but require careful calibration with artificial mixes of differentially modified RNA or corresponding synthetic oligonucleotides.

Pseudouridine, an abundant isomer of uridine, is a commonly modified nucleotide in RNA, mostly prevalent in tRNAs and rRNAs, but also present in snRNAs, snoRNAs and eukaryotic mRNAs (29–32). The broad distribution of pseudouridine residues stimulated the development of methods for their mapping in RNAs, first in low-throughput implementations using RT primer extension after chemical treatment, followed by analysis by high-resolution sequencing gels (33–37), and, more recently, by coupling with deep sequencing to increase throughput and sensitivity (31,38–40). However, despite notable differences in output data volume, both low- and high-throughput approaches described up to now were based on the use of the same chemistry, namely CMC-modification of pseudouridine residues under strongly denaturing conditions, followed by resolution of CMC-conjugates to canonical U and G residues by long incubations in alkaline bicarbonate buffer (33,36). Inevitably, due to the instability of RNA under these relatively 'harsh' conditions, the noise is high. Among other factors, this results from accelerated unspecific phosphodiester bond cleavage, thus causing numerous CMC-independent RT-stops and corresponding hits in deep sequencing. Inclusion of CMC-untreated controls (41,36,40) and/or analysis of time courses for bicarbonate CMC removal, allow to alleviate these limitations to some extent, but makes the already sophisticated experimental procedure even more complicated and relatively expensive in terms of sequencing costs. In classical RT primer extension versions, CMC- $\psi$  adducts were not quantified, since only partial linearity was observed at low modification levels, and the intensity of the signal was highly variable (33,34,42), while deep sequencing CMC-based protocols were made semi-quantitative by inclusion of synthetic spike-ins and calibrations (38,43). Other chemical cleavage/modification approaches, namely using hydrazine (33,35), acrylonitrile (44,45) and methylvinylsulfone (46) were also explored in the past, but did not become popular in practice (36).

Here, we developed a sensitive, reliable, and quantitative approach for pseudouridine mapping in RNAs based on a chemistry orthogonal to classical CMC derivatization. To achieve this goal, we explored and optimized random RNA cleavage at uridine residues by hydrazine, followed by aniline treatment for RNA chain scission at abasic sites. Protected residues (negative hits) reveal the presence of pseudouridines based on their resistance to hydrazine-dependent scission. These protection signals were compared to efficiency of neighboring cleavages at unmodified uridines to obtain a quantitative information on the pseudouridylation levels. The method termed 'HydraPsiSeq' was successfully applied for precise mapping and quantification of pseudouridine residues in yeast *Saccharomyces cerevisiae* and *Homo sapiens* rRNAs and tRNAs, as well as *S. cerevisiae* mRNAs. Application of HydraPsiSeq revealed differential rRNA pseudouridylation profiles in different human cell lines and highlighted a dynamic regulation of rRNA pseudouridylation levels during chondrogenic stem cells differentiation.

## MATERIALS AND METHODS

### Yeast strains and cultures

Yeast strains used in this study were obtained either from the lab of Denis L.J. Lafontaine or from the EU-ROSCARF collection (Germany, see Table S1 in Supporting information). Cells were grown in standard Yeast Extract/Peptone/Dextrose (YPD) to 0.6–0.7  $OD_{600}$  for exponential phase.

The box H/ACA snoRNA-associated pseudouridine synthase Cbf5 is encoded by an essential gene (47). To modulate pseudouridylation levels in RNA, we made use either of genetic depletion (pGAL-cbf5, strain YDL521-1, (47)), or expression of a catalytically-deficient allele (Cbf5-D95A allele (48,49), strain YDL2932 and its isogenic control YDL2931).

To produce rRNAs with reduced box H/ACA-mediated pseudouridylation, we depleted Cbf5 in a haploid yeast strain expressing as its sole copy of CBF5 a GAL-regulated allele (47). This allele drives the expression of CBF5 in galactose-based but not in glucose-based medium. YDL521-1 cells were grown in permissive conditions (2% galactose, 2% sucrose, 2% raffinose synthetic medium lacking histidine) to mid-log phase, washed in pre-warmed water, and resuspended in pre-warmed non-permissive conditions (2% glucose synthetic medium lacking histidine) at an  $OD_{600}$  of 0.3. Cells were maintained in log phase by continuous dilution in fresh 2% glucose synthetic medium lacking histidine and collected before transfer (time point 0), and after 24- and 48-h of transfer to glucose.

To produce rRNA deprived of box H/ACA mediated pseudouridylation, we used cells expressing as sole source of Cbf5 a catalytically deficient allele (CBF5cata, Cbf5-D95A, strain YDL2932). YDL2932 is a haploid yeast strain in which the endogenous CBF5 gene was deleted by insertion of a TRP1 cassette and rescued by expression of Cbf5-D95A from a low copy ARS/CEN HIS3 plasmid. Cells were grown to mid-log phase in YPD rich medium (1% yeast extract, 2% peptone, 2% glucose).

### RNA extraction

Total RNA from yeast cells was isolated using hot acid phenol (50,51). RNA concentration was measured on a Nanodrop One and RNA quality was assessed by capillary electrophoresis using a PicoRNA chip on Bioanalyzer 2100 (Agilent technologies, USA). Yeast PolyA+ mRNA fraction was prepared from total RNA using NEBNext Poly(A) mRNA Magnetic Isolation Module (NEB, US), according to manufacturer's instructions.

### RNA fragmentation conditions

RNA (50–300 ng) was subjected to hydrazine treatment (50% final concentration) for 30–60 min on ice. The reaction was stopped by ethanol precipitation using 0.3M NaOAc, pH5.2 and Glycoblue and incubated 30 min at  $-80^{\circ}\text{C}$ . After centrifugation, the pellet was washed twice with 80% ethanol and resuspended in 1 M aniline, pH 4.5. The reaction was incubated in dark for 15 min at  $60^{\circ}\text{C}$  and processed as described above, by ethanol precipitation.

### RNA 3'-end dephosphorylation

RNA was dephosphorylated at the 3'-end using 10 U of T4 PNK in 100 mM Tris-HCl pH6.5, 100 mM MgOAc and 5 mM  $\beta$ -mercaptoethanol and incubated for 6h at  $37^{\circ}\text{C}$ . T4 PNK was inactivated by incubation for 20 min at  $65^{\circ}\text{C}$ . RNA was extracted by phenol:chloroform and ethanol precipitated. The pellet was resuspended in RNase free water.

### Library preparation

RNA was converted to library using NEBNext Small RNA Library kit (NEB, UK) using the manufacturer's instructions. Library quality was assessed using a High Sensitivity DNA chip on a Bioanalyzer 2100. Library quantification was done using a fluorometer (Qubit 3.0 fluorometer, Invitrogen, USA).

### Sequencing

Libraries were multiplexed and subjected for high-throughput sequencing using an Illumina HiSeq 1000 instrument with 50 bp single-read runs. Libraries were loaded onto flow-cell at 10–12 pM final concentration.

### Bioinformatic pipeline

Adapter removal was performed using the Trimmomatic utility v32.0 (52), with a stringency parameter of 7, very short reads  $<8$  nt were excluded. The alignment of raw reads was conducted by Bowtie2 (53) in end-to-end mode. Mapped and sorted \*.bam file was transformed to \*.bed format. Locations of 5'-extremities of mapped unique reads were counted from \*.bed file, giving raw cleavage profile. Reads' 5'-end counts were normalized to local background in rolling window of 10 nucleotides, values for U residues were excluded to calculate the median. If RNA reference

contains multiple U-rich stretches  $>10$  in length, normalization window can be extended to 16 nt or more. Locally normalized U profiles (NormUcount, full profiles shown in figures) were used as protection U score value and further transformed to U cleavage profiles only, by omitting values for other nucleotides. Resulting U profiles were used for calculation of 'RiboMethSeq-like' scores (ScoreMEAN, A, B and PsiScore, equivalent to ScoreC in RiboMethSeq). Window of four neighboring nucleotides ( $\pm 2$  nt) was used for score calculation (54).

ScoreMEAN for each position is calculated in two steps, as follows: first, a ratio of number of cumulated 5'/3'-reads ends between preceding and following position is defined and, second, ScoreMEAN is calculated as a ratio of a drop for a given position compared to the average and variation for 4 neighboring positions ( $-2/+2$ ). PsiScore(ScoreC2 in RiboMethSeq), is calculated using the following formula:  $\text{RiboMethScore} = 1 - n_i / (0.5 \times (\text{SUM}(n_j \times W_j) / \text{SUM}(W_j) + \text{SUM}(n_k \times W_k) / \text{SUM}(W_k)))$ , where  $n_i$  is cumulated 5'/3'-end count for a given position,  $j$  varies from  $i - 2$  to  $i - 1$ ,  $k$  varies from  $i + 1$  to  $i + 2$ , Weight parameters are defined as 1.0 for  $-1/+1$  and 0.9 for  $-2/+2$  positions. ScoreA and ScoreB were previously reported (55) and are described in more details in the Supplementary Material.

Mapping of yeast mRNA reads was done to reference collection of yeast CDS (ENSEMBL release R64-1-1 *Saccharomyces cerevisiae*.R64-1-1.cdna.all.fa). Only uniquely mapped reads were taken into account for further analysis and calculation of protection scores.

### Isolation and expansion of human bone marrow mesenchymal stem cells (BMMSCs)

MSCs were isolated from human bone marrow biopsy from patient during total hip arthroplasty after informed consent. This study was approved by our local ethics committee (Registration number DC-2014-2148). Bone marrow biopsy was heparinated, diluted in Phosphate-Buffered Saline solution (PBS) and centrifuged at 1500 rpm for 5 min. The pellet was resuspended, seeded in Petri dishes at  $4 \times 10^6$  cells/dish and cultured at  $37^{\circ}\text{C}$  in a humidified atmosphere containing 5% (v/v)  $\text{CO}_2$ . The expansion phase was performed until passage P2. The expansion medium is composed of Dulbecco's modified Eagle's medium with low glucose (DMEM-LG, Gibco) supplemented with 10% fetal bovine serum (FBS, Sigma), 1 ng/ml basic fibroblast growth factor (bFGF, Miltenyi Biotec), 1% glutamine (Gibco) and 1% penicillin–streptomycin (Gibco).

### Predifferentiation of BMMSCs in monolayer and production of cartilaginous TE substitutes

To favor chondrogenic differentiation, BMMSCs were cultured at passage 3 with predifferentiation medium until confluence before seeding to biomaterials (collagen sponges). This medium was composed of Dulbecco's modified Eagle's medium with high glucose (DMEM-HG, Gibco) supplemented with 10% fetal bovine serum (FBS, Sigma), sodium pyruvate (110  $\mu\text{g}/\text{ml}$ , Gibco), bFGF (1 ng/ml, Miltenyi Biotec), 1% penicillin–streptomycin (Gibco), chon-

drogenic supplements (PAD: proline (40  $\mu\text{g/ml}$ , Sigma), L-ascorbic acid-2-phosphate (50  $\mu\text{g/ml}$ , Sigma) and dexamethasone (0.1  $\mu\text{M}$ , Sigma).

To study MSCs differentiation during cartilaginous TE substitutes production, BMMSCs were trypsinized at the end of the predifferentiation at passage 3 (trypsin-EDTA 0.05%, Gibco) and seeded into collagen sponges (Symatèse Biomatériaux, Chapanost, France) at a density of 0.5 million BMMSCs/sponge. These sponges measured 5 mm in diameter and 2 mm in thickness and were composed of 95% type I collagen and 5% type III collagen. MSCs seeded in collagen sponges were cultured at 37°C in a humidified atmosphere containing 5% CO<sub>2</sub> (v/v). Half of the sponges were cultured with a control medium (ITS) that allowed MSCs survival but not chondrogenic or other differentiation. The other half of the sponges were cultured with chondrogenic medium (TGF- $\beta$ 1). The control medium was composed of DMEM-HG supplemented with 1% ITS (ITS + premix, BD Biosciences), 1% glutamine (Gibco), sodium pyruvate (110  $\mu\text{g/ml}$ , Gibco), 1% penicillin streptomycin (Gibco), and PAD. The chondrogenic medium was prepared with control medium supplemented with TGF- $\beta$ 1 at 10 ng/ml (Miltenyi Biotech). At D7, D14, D21 and D28, analysis of cartilaginous matrix synthesis by BMMSCs inside collagen sponges was performed.

#### Analysis of cartilaginous matrix synthesis of TE substitutes

At D7, D14, D21 and D28, cartilaginous TE substitutes seeded with human MSCs were fixed ( $n = 3$  substitutes/time) with 4% paraformaldehyde during 24 hours and embedded in paraffin. 5  $\mu\text{m}$  sections were stained with Alcian blue staining to visualize proteoglycan content. To visualize collagen II content, immunohistochemical analysis was performed as described previously (56). The stained slides were observed and recorded by light microscopy (DMD108, Leica) at 4 $\times$  original magnification. The scale bars represent 500  $\mu\text{m}$ .

## RESULTS

### General overview of the HydraPsiSeq mapping technique

The quantitative HydraPsiSeq protocol for pseudouridine mapping and quantification is based on random (statistical) cleavage at all uridine residues in RNA by combination of hydrazine treatment (ring-opening of the pyrimidine base in uridines) and aniline (cleavage of the resulting RNA abasic site and release of 5'-phosphorylated RNA fragments directly competent for adapter ligation at the N+1 nucleotide in the sequence, Figure 1A, Supplementary Figure S1). The canonical RNA bases (A, C, G), as well as pseudouridines, are known to be resistant to hydrazine cleavage and thus generate only background signals (57). Decomposition of RNA abasic site (58) by consecutive  $\beta$ - and  $\delta$ -elimination generates two RNA fragments ending at the  $N - 1$  and starting at  $N + 1$  nucleotide to the U residue. A 3'-end dephosphorylation step using T4 PNK serves to remove all eventual 3'-phosphates (59) from upstream RNA fragment, and thus to ensure its compatibility with direct 3'-adapter ligation. Fragments produced by random cleavage at U residues are then selectively converted to a library

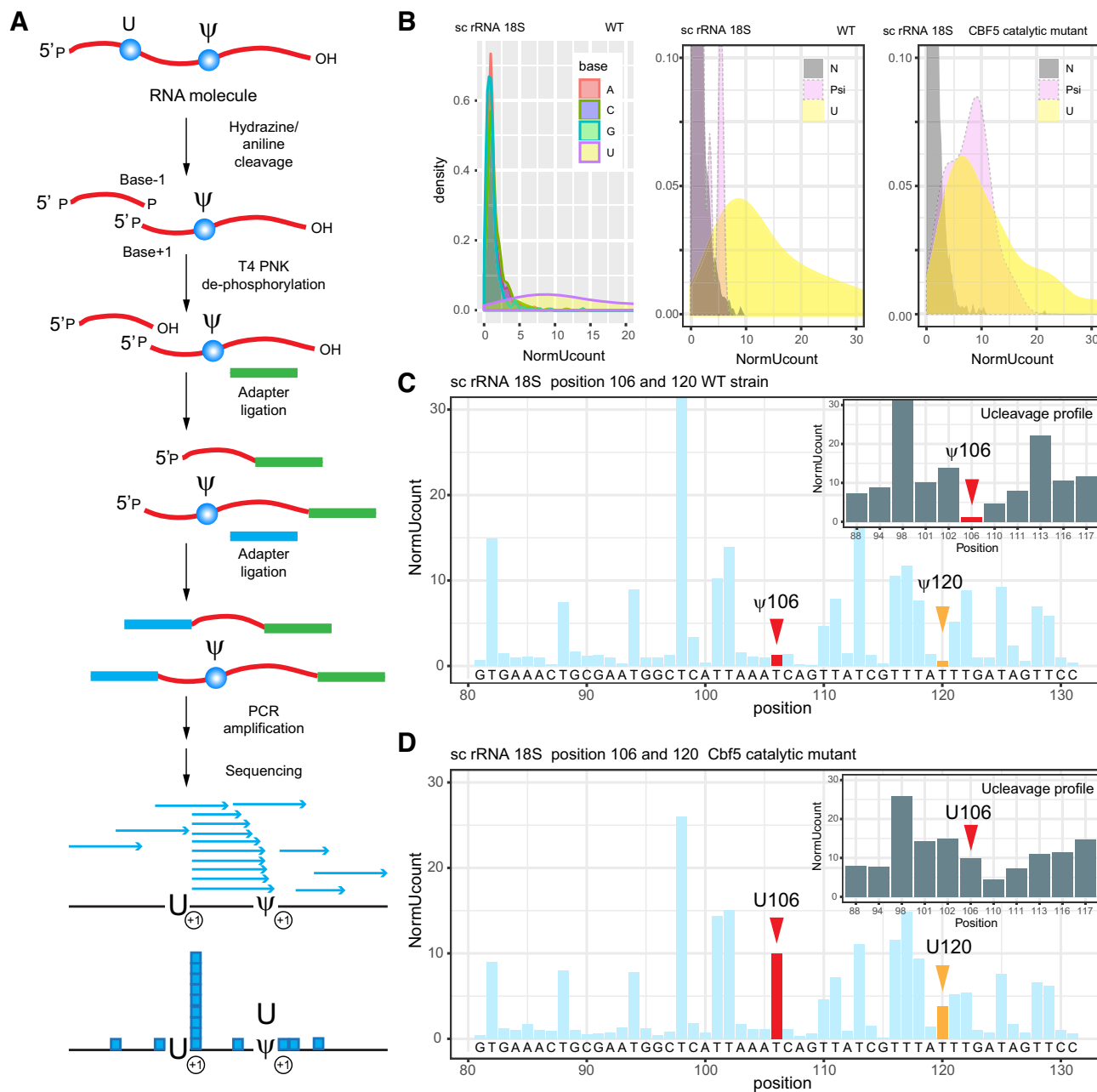
and sequenced. Mapping to the reference sequence allows counting of such reads that start at the N+1 nucleotide position relative to U residues, while  $\psi$  residues (as well as A, C and G), due to their relative resistance to hydrazine cleavage, give only background signals. These counts for A, C and G are used for local normalization of the U cleavages to background (NormUcount, see Materials and Methods), to account for local accessibility of different RNA regions. Further bioinformatic steps of analysis are similar to the well-established RiboMethSeq procedure (54), and allow calculation of scores termed ScoreMEAN, ScoreA and ScoreC for Ucleavage profiles, where  $\psi$  residues appear as protected positions (similarly to Nm residues in RiboMethSeq).

Inspection of cleavage efficiencies in yeast rRNA (used here as initial model and containing 47 known  $\psi$  sites, one each in 5S and 5.8S rRNAs, 14 in 18S rRNA and 31 in 25S rRNA (60)), demonstrated that, as expected, cleavage at A, C and G is relatively marginal (NormUcount < 5), while cleavage at U residues is very efficient (NormUcount ranging from Zero to 50–60), see Figure 1B (left panel) for 18S rRNA data, similar profiles were also obtained for 25S rRNA. Technical and biological replicates demonstrated an excellent correlation of Ucleavages in different samples (Supplementary Figure S2). Deconvoluting the U's into positions which are known to be pseudouridylated and those which are not, the pseudouridines appear to display a profile similar to that of A, C and G (i.e. 'uncleaved/protected', Figure 1B, middle panel) and well distinct from U's (the pink and yellow peaks do not overlap). The peaks of  $\psi$  cleavages in the area of 6–7 NormUcount units correspond to incompletely modified pseudouridylated sites. Further evidences that pseudouridylated positions are protected from RNA cleavage was obtained by analysis of rRNA extracted from yeast cells expressing a catalytically inactive form of Cbf5. Cbf5 is responsible for all rRNA pseudouridylations (48), except one in 5S rRNA catalyzed by Pus7 (61). In this case, positions normally pseudouridylated behaved like unmodified Us (i.e. 'cleaved/unprotected', Figure 1B, right panel, the pink and yellow peaks largely overlap).

A typical cleavage profile observed for yeast 18S rRNA at position 106 and the corresponding Uprofiles (insert) are shown in Figure 1C and D. Cleavage of pseudouridine at position 106 was very low in WT rRNA (Figure 1C), while in the absence of modification (in strain expressing Cbf5 catalytic mutant, Figure 1D), the signal was roughly equivalent to the neighboring unmodified Us in the sequence. Similar observation was done for other sites (see also position 120, Figure 1C and D).

### Optimization of the hydrazine cleavage conditions

Conditions of hydrazine cleavage were optimized with respect to incubation temperature and time to achieve a fragment size compatible with sequencing and analysis (from ~10 to 30 nt) and to maximize signal-to-noise ratio. We found that the best results were obtained for hydrazine cleavage at 0°C for 45 min as longer incubation times were detrimental to read mapping during the alignment step. This is likely because of excessive amount of short reads mapped at multiple locations (Supplementary Figure S3).



**Figure 1.** HydraPsiSeq protocol for  $\psi$  mapping and quantification. (A) General overview of the protocol. RNA is treated by hydrazine and subjected to aniline cleavage. 3'-phosphates are removed by T4 PNK treatment and adapters are ligated to 3'- and 5'-ends of RNA fragments. After sequencing and alignment to a reference sequence, 5'-ends of all fragments are counted to generate Ucleavage profiles. U residues are sensitive to hydrazine and thus efficiently cleaved, while  $\psi$  residues are resistant and provide only background signals. (B) Normalized U count for all 4 RNA nucleotides (A, C, G, U). Middle and right panels show NormUcount values for A/C/G nucleotides (N), U residues and  $\psi$  residues, in WT and strain expressing Cbf5 catalytic mutant (no Cbf5 activity, rRNA is not pseudouridylated). (C and D) Cleavage profiles for position  $\psi$ 106/ $\psi$ 120 in 18S rRNA in WT yeast strain and in the strain expressing Cbf5 catalytic mutant. Modified positions are shown in red and orange. U-only profiles are shown in inserts.

Under optimized hydrazine treatment conditions, cleavage of U residues remains rather heterogeneous, with NormUcount varying from 10–15 to >200. This may be related to particular sequence context or RNA structural effects. Even if RNA 2D and 3D structures have been described to be totally compromised in 50% (w/w) hydrazine solution (57), the use of additives, known to destabilize RNA 3D and 2D structures, such as urea/DMF or DMSO, led to accelerated hydrazine cleavage. However, these additives did not allow

to improve the U-to- $\psi$  detection specificity or make cleavage more homogeneous (data not shown).

### Pseudouridine detection and mapping

The reliability of pseudouridine detection by HydraPsiSeq was evaluated by Receiver Operating Characteristic (ROC) curves and associated Matthews correlation coefficient (MCC) values for WT yeast rRNA containing 47  $\psi$

residues altogether (60), of which 46 are installed by Cbf5 as part of H/ACA snoRNPs (47), while formation of  $\psi$ 50 in 5S rRNA is catalyzed by the stand-alone enzyme Pus7 (61). ROC curve for protection Uscore (defined as normalized protection of U residues from cleavage, see Materials and Methods) shows a reliable detection for the majority of pseudouridylation sites (Figure 2A). As anticipated, in the absence of pseudouridines in rRNA (isolated from a Cbf5 catalytic mutant), there is no distinction from U signals anymore.

Since the amount of biological material required for reliable detection and quantification is of great importance for analysis of precious biological samples, we attempted to reduce the amount of input rRNA from 300 ng, used in our initial tests, to 50 ng. Figure 2B demonstrates that even with such reduced amount of input RNA the detection of pseudouridines is highly reliable. Analysis of performance for different  $\psi$  prediction/discrimination scores (protection Uscore, ScoreMEAN, ScoreA, ScoreB and PsiScore, equivalent to ScoreC) demonstrates that Uscore and PsiScore (equivalent to ScoreC in RiboMethSeq) provide the best discrimination power between true-positive and false-positive hits (Supplementary Figure S4). Analysis of global ROC curve for all yeast rRNAs shows that 25 out of the 29 most protected U residues correspond to known  $\psi$  sites, and at the optimal MCCmax value, 35 sites are detected, with 13 false positive (FP) hits. The majority of false negative (FN) identifications correspond to partially modified positions. Measurements of PsiScore demonstrated an excellent reproducibility between technical and biological replicates and showed that molar  $\psi$  level varies from 0.8 to 1 for the majority of tested positions, only few sites were incompletely modified in yeast rRNA under normal growth conditions (complete glucose-based medium, Supplementary Figure S5). Correlation graphs (Supplementary Figures S2 and S5) also attest that hydrazine cleavage profiles were highly reproducible for all technical and biological replicates analyzed in this study. We found the average standard deviations to be of <5%, and the most variable positions in technical replicates only showed a maximum variation of 10–15%. Similar data were obtained for biological replicates as well. Thus, variations of PsiScore of >10% (0.1 in molar  $\psi$  amount per site) can be considered as significant and reproducible, opening the way for precise profiling of pseudouridine content.

Comparison of the results obtained for different sequencing libraries demonstrated that 20–25 millions of reads were sufficient for an almost perfect coverage of all *S. cerevisiae* rRNA positions, thus this number of raw reads was used in further experiments.

#### Analysis of HydraPsiSeq signals in snoRNA-deleted *S. cerevisiae* strains and in yeast cells expressing inactive Cbf5

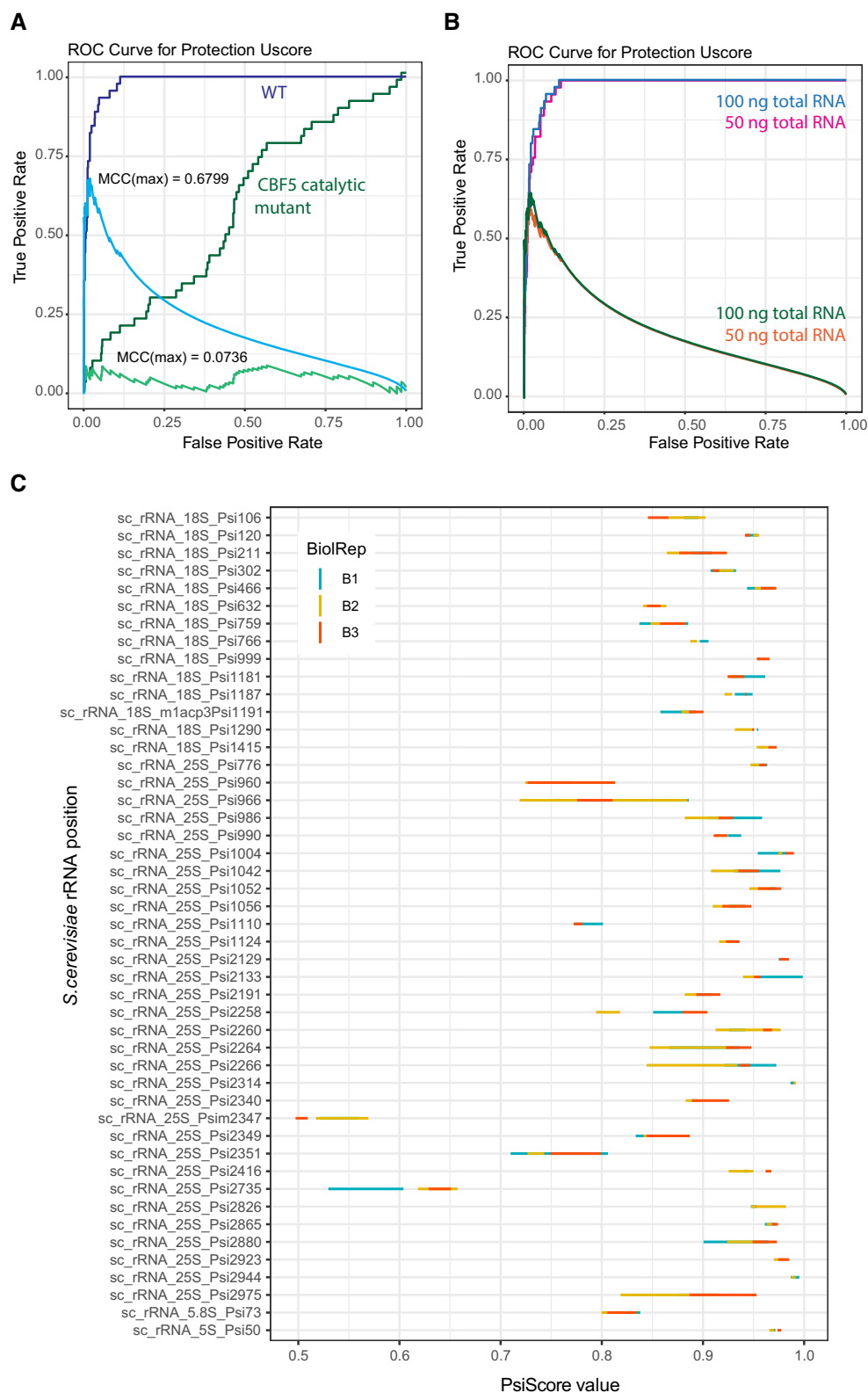
To further prove the reliability of HydraPsiSeq in  $\psi$  detection and quantification, we performed analysis of two yeast strains bearing deletions of H/ACA snoRNAs snR44 or snR81. The snoRNA snR44 guides rRNA modifications at two positions (18S- $\psi$ 106 and 25S- $\psi$ 1056), while snR81 is responsible for modification of 25S- $\psi$ 1052 (62) (Supplementary Figure S6). We also included two biological replicates

for a strain expressing an inactive Cbf5 catalytic mutant, and a time course for depletion of WT Cbf5 expressed under the control of a GAL promoter (See Supplementary Figure S7, (47)). As shown in Figure 3, the only differences in PsiScore values between WT and KO snoRNA strains were precisely those located at positions (labelled in green and blue) guided by the deleted snoRNAs, and were similar to values obtained for unmodified rRNA observed in the Cbf5 catalytic mutant. It is noteworthy that the level of the Pus7-catalyzed 5S rRNA  $\psi$ 50 was not affected upon Cbf5 inactivation or depletion (labelled in red). The Cbf5 depletion experiment highlights that for the majority of  $\psi$  sites, 48h-depletion gave roughly the same residual level of  $\psi$  as observed with the Cbf5 catalytic mutant (Supplementary Figure S7).

Analysis of unmodified rRNA extracted from the catalytically inactive Cbf5 yeast strain (Cbf5cata) also provided a calibration of PsiScore signals to ‘a zero modification level’, thus allowing precise quantification of rRNA pseudouridylation. In order to demonstrate that PsiScore can be used for precise quantification of  $\psi$  content, we performed calibration of PsiScore signals with mixes of rRNA samples extracted from WT (modified) and Cbf5cata (unmodified) yeast strains. Analysis was done in technical duplicates using 0, 5, 10, 25, 50, 75 and 100% of modified WT rRNA in the mix. Calibration curves shown in Figure 3B and Supplementary Figure S8 demonstrate a very good linear correlation between  $\psi$  content and observed PsiScore for majority of yeast rRNA sites. Precise quantification is certainly one of the most valuable features of HydraPsiSeq protocol, since, by comparison, other deep sequencing-based methods cannot afford such analysis.

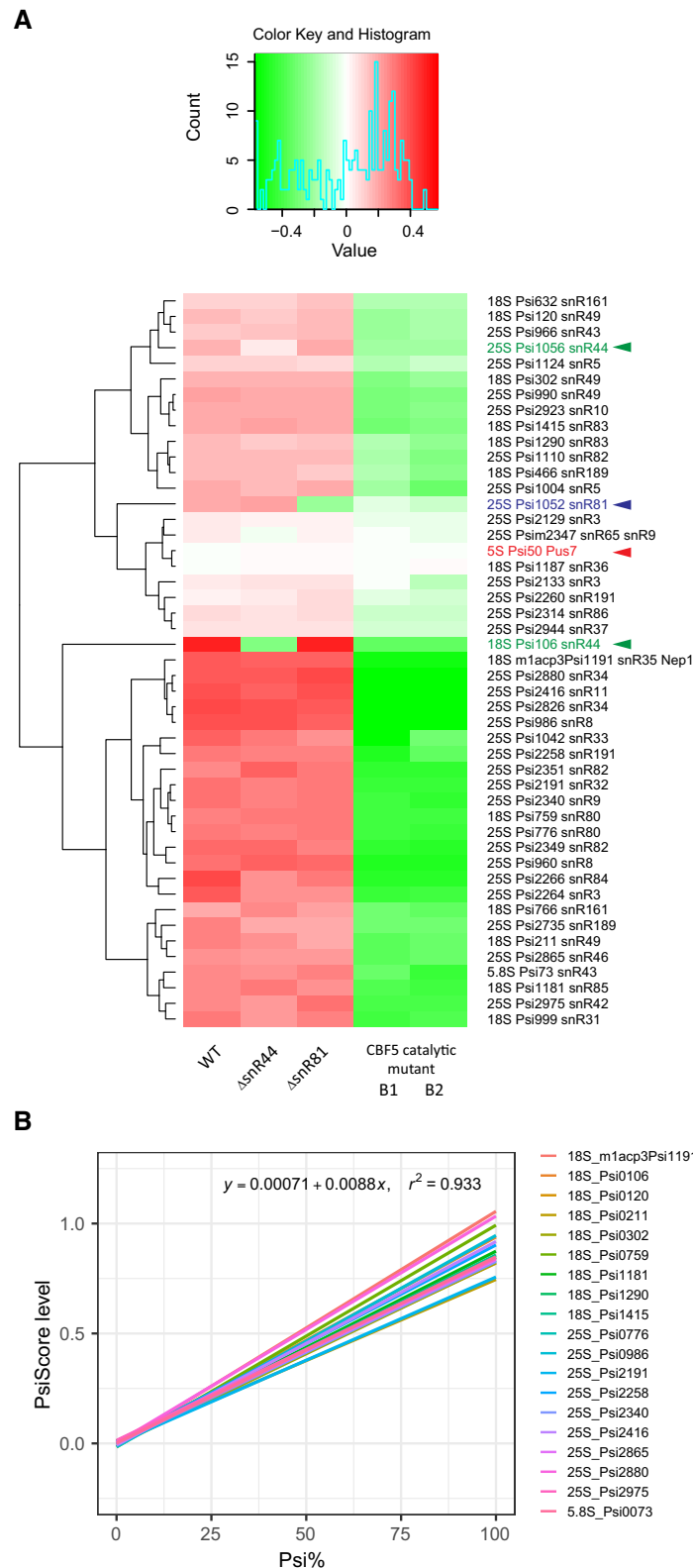
The applicability of HydraPsiSeq was also tested on yeast *S. cerevisiae* tRNAs. These molecules are highly pseudouridylated, and at least 7 tRNA: $\psi$ -synthases catalyze U to  $\psi$  conversion in tRNAs (30,63). Although hydrazine cleavage profiles in tRNAs appeared less uniform than in rRNA, pseudouridinylation sites could be clearly identified by drastic changes in protection against hydrazine/aniline cleavage in null-mutants of the corresponding Pus-enzyme (see results in cells lacking PUS1, PUS4, and PUS7 in Supplementary Figure S9). Thus, while *de novo* mapping of the pseudouridylated sites in tRNAs from WT strain was somewhat limited, combination with data from the corresponding knock-out strain allowed not only clear identification/confirmation of modified sites, but also to appreciate changes in protection and thus to quantify a modulation of individual  $\psi$  levels.

In order to extend HydraPsiSeq applications to low abundant RNAs (such as mRNAs), we performed analysis of yeast *S. cerevisiae* mRNAs from WT,  $\Delta$ PUS3 and  $\Delta$ PUS7 strains. Only 150 ng of mRNA polyA+ preparation was engaged in this analysis (this amount could even be reduced to as low as 50 ng), demonstrating low requirements for input material. With moderate sequencing depth (100 mln reads/sample) sufficient coverage for analysis was obtained for >1200 yeast mRNAs. Yeast mRNA was reported to contain ~200–250 potential  $\psi$  modification sites (31,38,64). We were able to extract NormUcount values for ~70 of them and evaluate protection level for these residues. Compilation of these data allowed to confirm moderate to high



**Figure 2.** Pseudouridine detection in WT yeast rRNA. (A) Receiver Operating Characteristics (ROC) curve for  $\psi$  detection using U protection score (dark blue). Maximal Matthews correlation coefficient (MCCmax) value is shown (light blue). The corresponding trace for unmodified rRNA from Cbf5 catalytic mutant is shown in dark green (MCC trace in light green). (B) ROC curves for  $\psi$  detection for samples with variable total RNA input. As low as 50 ng of input RNA ( $\sim 40$  ng of rRNA) is sufficient for good representativity and quantification of all rRNA  $\psi$  sites. (C) Dispersion of PsiScore values for technical and biological replicates. Analysis of *S. cerevisiae* rRNA pseudouridylation was done in three biological replicates (indicated in colors) with color bar representing dispersion of technical replicates.





**Figure 3.** (A) Differential heatmap for variations of PsiScore level for all known sites in *S. cerevisiae* rRNA. Normalization was done by the average score for a given position. SnoRNA independent  $\psi$ 50 in 5S rRNA is highlighted in red.  $\psi$ 106 and  $\psi$ 1056 guided by snR44 are highlighted in green. The identity of the *S. cerevisiae* strains used is provided at the bottom. Strain expressing Cbf5 catalytic mutant was analyzed in biological replicate. Identity of H/ACA snoRNA potentially guiding pseudouridylation is indicated together with the position. Clustering was done by hclust R function with ward.D2 algorithm. Color key for differential PsiScore is given in the insert. Visible variation of the signal for  $\psi$ 1187 is very minor upon CBF5 inactivation, but is related to drastic change of protection at the neighboring m<sup>1</sup>acp<sup>3</sup> $\psi$ 1191, and thus perturbation in the shape of Uprofile, used to calculate PsiScore. (B) – Linear correlation of PsiScore with  $\psi$  content of yeast rRNA. Calibration curve was produced using mixes of Cbf5cata and WT rRNAs, at ratio of 0, 5, 10, 25, 50, 75 and 100% of WT rRNA. HydraPsiSeq analysis was done in technical replicate, calibration curves for 19 non-clustered  $\psi$  sites are shown.

protection score (UScore 0.5–10) for ~35 sites reported by two previous studies, and out of 14 sites reported to be common between two published datasets, HydraPsiSeq confirmed high protection level at two positions (Figure 4A).

### Profiling rRNA pseudouridylation in different human cell lines

Having successfully analyzed pseudouridine sites in yeast rRNAs, tRNAs and mRNAs, we next applied HydraPsiSeq profiling to human rRNAs, containing altogether 104 reported  $\psi$  sites (65).

The HydraPsiSeq profiling was performed in biological duplicate (or triplicate) for three commonly used human cell lines, namely: fibroblasts, kidney cells (HEK293 cells), and cervix cancer cells (HeLa). As shown in Figure 4B, HydraPsiSeq profiling revealed that about half of measured pseudouridylation sites (top part of the heatmap presented in Supplementary Figure S10) displayed distinctly different PsiScore levels in the three cell lines tested.

The finding that rRNA pseudouridylation levels differ from one cell line to another, underscores the presence of differentially modified populations of ribosomes. Such differential rRNA modification was previously observed for other modifications, such as 2'-O methylation (66,67). Most affected cell type-specific pseudouridylation sites (24 sites, see Figure 4B) were located both on 18S (9 sites) and 28S rRNAs (15 sites), and mostly clustered in the region 800–1150 for 18S rRNA and were relatively evenly distributed throughout 28S rRNA domains (4 in Domain I, 1 in Domain II, 6 in Domain III, 1 in Domain V – PTC and 3 in Domain VI). Differential pseudouridylation may reflect variable efficiency of the rRNA pseudouridylation machinery (as suggested previously for 2'-O methylation (67)), and/or a novel layer of regulation at the rRNA modification level.

### Profiling rRNA pseudouridylation during human chondrogenic differentiation

To demonstrate further applications of HydraPsiSeq, we performed profiling of pseudouridine content in rRNA extracted from human bone marrow mesenchymal stem cells (BMMSCs) during their chondrogenic differentiation. Non-differentiated BMMSCs are capable to engage selective differentiation program, to become osteoblasts, chondrocytes or adipocytes and each of these programs requires specific expression of particular proteomes (68,69). Bone marrow stem cells were first extracted from clinical sample and chondrogenic differentiation was performed in type I/III collagen 3D sponges. A comparison between differentiated and non-differentiated cells was embodied in a 3D setting where the cells were either grown in the presence of TGF- $\beta$ 1, allowing the differentiation into chondrocytes, or in non-differentiating ITS medium, sustaining their growth but impeding differentiation (70,71). Chondrogenic differentiation of BMMSCs was assessed by staining with alcian blue, for detection of glycosaminoglycans, and by immunostaining for collagen II, which is a reliable chondrocyte marker (Figure 5A). In parallel, samples were collected at 4 time points (7, 14, 21 and 28 days) and subjected to HydraPsiSeq analysis.

The majority of pseudouridylation sites remained constitutively modified during the time course of differentiation, both in non-differentiating ITS medium and during differentiation in TGF- $\beta$ 1-containing medium (Supplementary Figure S11), but specific positions displayed altered pseudouridylation levels between these two states (Figure 5B). Remarkably, this subset of variable pseudouridylation sites largely overlap with the set of cell type-specific modifications (see above and Figure 5C), and some of those residues have been shown to play an important role in ribosome functions (e.g. 28S- $\psi$ 3743 (72)).

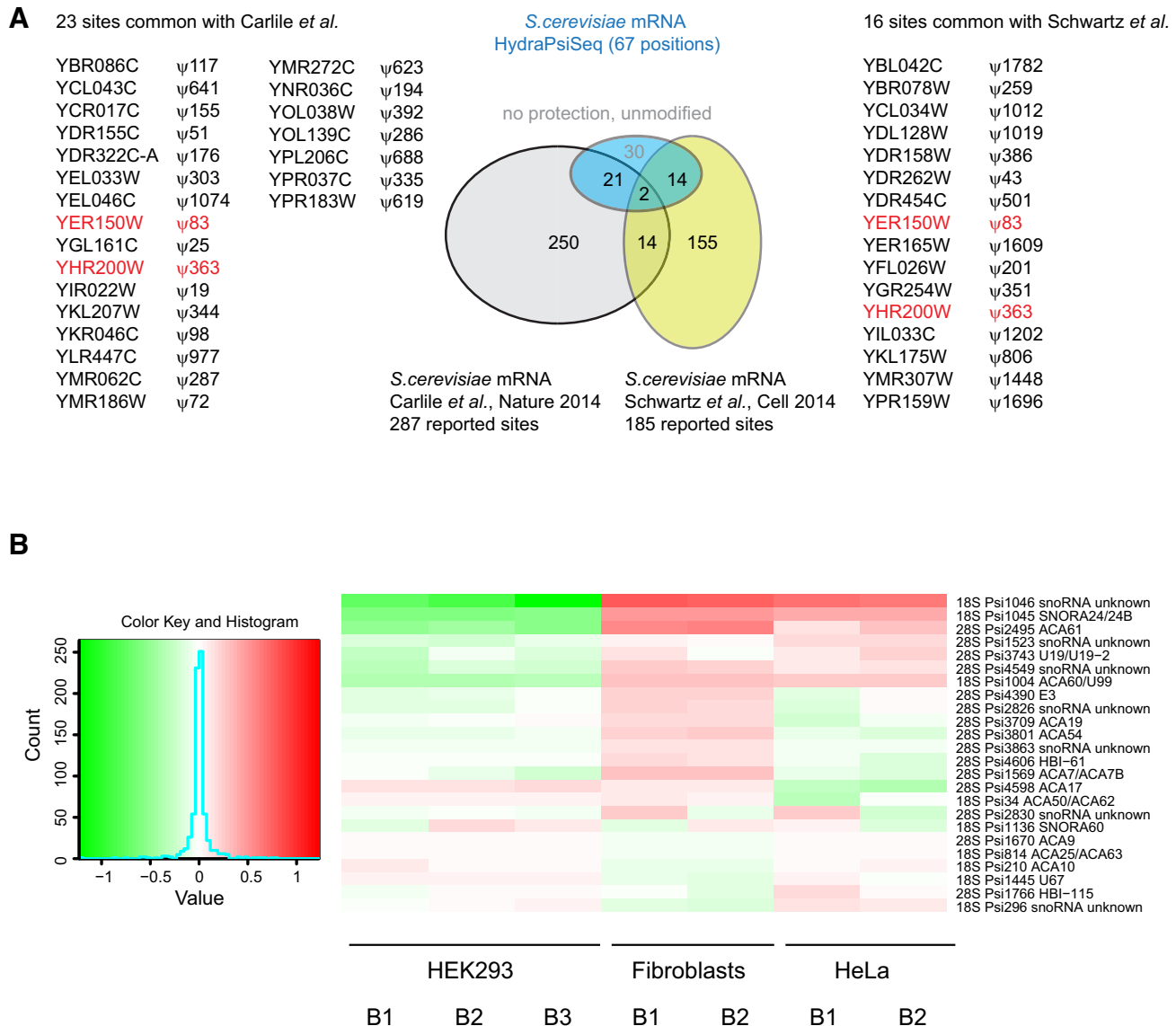
In conclusion, our analysis demonstrates that global pseudouridylation profile of human rRNAs may vary considerably between different cell lines, and throughout cell differentiation programs, supporting the emerging concept that cells produce different types of ribosomes, which may carry specialized function in order to adapt to specific translation requirements (73,74).

## DISCUSSION

### Advantages and limitations of HydraPsiSeq

A unique feature of HydraPsiSeq, by comparison to previously described high throughput methods, is that it provides for the first time quantitative information about the amounts of pseudouridylation at any particular position of an RNA molecule with nucleotide precision. HydraPsiSeq does not involve the selective enrichment of  $\psi$ -modified RNAs, and as such it is particularly efficient on abundant cellular RNAs, such as rRNAs and tRNAs. Nonetheless, we have shown that it is also efficient on lower abundant RNAs, such as yeast mRNAs. In that respect, HydraPsiSeq is similar to RiboMethSeq, which allows the systematic mapping of 2'-O methylated sites (22). By comparison, AlkAnilineSeq (28), for detection of m<sup>7</sup>G, and other base modifications, includes the selective enrichment of the modified sites (2). The absence of an enrichment step in HydraPsiSeq implies that a substantial coverage at all U positions in a given RNA sequence must be achieved (sequencing depth of 20–25 millions of reads for eukaryotic rRNA, see below), but, in return, HydraPsiSeq allows highly sensitive, reliable and precise quantification of  $\psi$  residues in RNA and, with inclusion of a calibration of the PsiScore signal to natural unmodified RNA or synthetic transcript, even to reach an absolute measurement of the modification level. Since the main advantage of this new protocol over previously published CMC-based techniques (31,38,39) resides in quantification of pseudouridines, we foresee quantitative profiling of eukaryotic rRNA as the main application, together with analysis of tRNA modification dynamics under different cell growth and physiological or pathological conditions.

Application for analysis of rare RNAs, such as mRNA or lncRNA is attainable, but mostly for validation purposes, and presumably only for the most abundant molecules, since otherwise quite substantial sequencing depth would be required for significant analysis at the transcriptome-wide level which would increase costs. This limitation of HydraPsiSeq in analysis of moderately abundant mRNAs can be overcome, providing that an appropriate enrichment step is included, e.g. through the development of a  $\psi$ -specific antibody for IP-like enrichment (17).

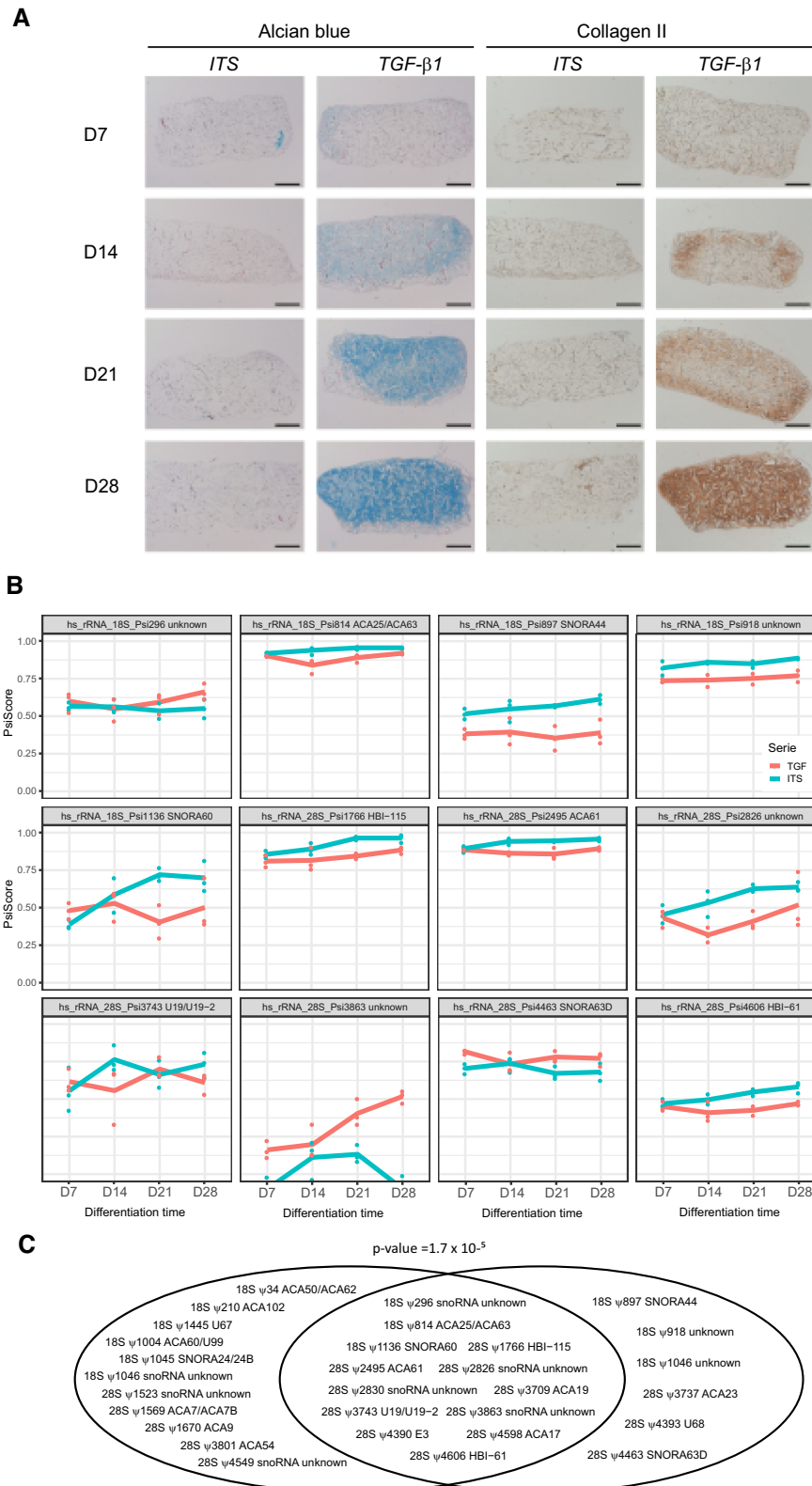


**Figure 4.** Profiling of yeast mRNA and human rRNA pseudouridylation by HydraPsiSeq. (A) Profiling of yeast *S. cerevisiae* mRNA modifications by HydraPsiSeq. Protection profiles with sufficient coverage were obtained for 67 positions previously reported to be pseudouridylated (31,38). Out of those, 23 showing moderate to high protection (UScores from 0.5–10) are common with the list from Carile *et al.* (2014) and 16 with Schwartz *et al.* (2014), two sites (indicated in red) are common for all three datasets. (B) Human cell lines (HEK293, fibroblasts and HeLa cells) were used in biological duplicate or triplicate (labeled B1/B2/B3). Profiling was done for all previously known human rRNA modifications (104 positions). Only 24 most variable positions are shown in Figure, full heatmap is given in the Supplementary Figure S10. For reference, the identity of the H/ACA snoRNA potentially guiding pseudouridylation is indicated together with the position. Clustering was done using hclust R function with ward.D2 algorithm. Color key for differential PsiScore is provided in the insert.

### Required amount of input material and sequencing depth

In our optimized protocol, HydraPsiSeq requires about 50 ng of total RNA for analysis of rRNA modifications, or equivalent amount of enriched RNA fraction for analysis of low abundant RNAs. With some, but still acceptable, loss in efficiency in the library preparation step, this amount can be even further reduced to only ~10–20 ng. This exceptional sensitivity makes HydraPsiSeq compatible with the great majority of fundamental studies on RNA modifications, and also for large-scale screening of clinical sample collections. It is noteworthy that, except for the Illumina-based

RiboMethSeq protocol, which requires 10 ng of RNA input at the minimum (22), all other reported deep-sequencing methods for analysis of RNA modifications require at least  $\mu\text{g}$  (or even mg) amounts of input RNA (75). Also of note is that, in principle, the detection of protected  $\psi$  residues by comparison to cleaved U residues (negative detection) requires much higher sequencing depth than ‘positive’ detection methods (e.g. CMC-based protocols). Our optimized protocol provides robust coverage of all U residues in human rRNA and thus reliable detection and quantification of all pseudouridines in these species, with 20–25 millions



**Figure 5.** Modulation of human rRNA pseudouridylation profile during TGF-β1 stimulated differentiation of human bone marrow stem cell into chondrocyte-like cells. (A) Accumulation of glycosaminoglycans revealed by alcian blue staining (blue, left) and immunostaining for collagen type II (brown, right). Samples were analyzed at days 7 (D7), 14 (D14), 21 (D21) and 28 (D28). The type of medium used is indicated at the top of each column. For simplicity, only one biological replicate is shown on Figure. (B) Time courses of variations of PsiScore levels at selected pseudouridylated sites in human rRNA. Only highly variable positions are shown, other are presented in Supplementary Figure S11. Differentiation was followed for three independent biological replicates for the BMMSCs issued from the same patient. (C) Human rRNA pseudouridylation sites showing variability between different cell lines and during BMMSCs differentiation. Venn diagram illustrates the overlap of two different datasets.

raw reads available for analysis (2000–3000 reads/position). This sequencing depth is a bit higher than that required in other common protocols (e.g. RiboMethSeq), but remains relatively reasonable, since no costly selection steps are necessary, and thus the overall cost of library preparation remains rather limited.

### Reproducibility and precision of pseudouridine quantification

Analysis of technical and biological replicates for yeast (Supplementary Figure S5) and human rRNA (Figure 4B) demonstrated a very low dispersion of measured protection Uscores and derived PsiScores used for quantification. Technical variability was found to be higher for partially modified rRNA positions and inversely correlated with sequencing coverage in the region used for calculation of scores. In conclusion, HydraPsiSeq now provides quantification of RNA pseudouridylation with precision at least equivalent to mass spectrometry approaches (9,10), but the use of deep sequencing allows accommodation of larger series of samples even when only minute amounts of RNA are available.

### Known false-positive/negative identifications

Random and uniform cleavage of U residues by hydrazine is the pre-requisite for successful analysis of hydrazine-resistant pseudouridines by HydraPsiSeq. Analysis of yeast and human rRNAs demonstrated that, most likely due to residual 2D and/or 3D RNA structures under hydrazine treatment at low temperature (0°C), a minor subpopulation of inaccessible U residues pertains and displays only very limited cleavage (Uscore). Thus such sites may be interpreted as false-positive hits. Another known source of false-positive hits is related to resistance of 5-modified U residues (mostly m<sup>5</sup>U, but also mo<sup>5</sup>U, etc.) to hydrazine cleavage, thus such residues are also detected as ‘protected’ U’s and can give false-positive identification. However, compared to pseudouridines, which are abundantly present in many different RNAs, these modifications are relatively rare (they are mostly found in tRNAs). Dihydrouridines (D) are most probably resistant to hydrazine themselves, but finally cleaved by the combination of alkaline conditions under hydrazine treatment and aniline scission, and thus are not revealed as false-positive protected hits in HydraPsiSeq (data not shown). Reciprocally, some rare non-U modified residues in rRNAs and tRNAs displayed an unexpectedly high sensitivity to cleavage under the conditions used in HydraPsiSeq (e.g. m<sup>7</sup>G and m<sup>3</sup>C), but this did not affect  $\psi$  detection or quantification.

### Constitutive and variable pseudouridylation sites in human rRNAs in cancer and cell differentiation programs

Profiling of different human cell lines and stem cells during chondrogenic differentiation demonstrated that the great majority of rRNA pseudouridines is constitutively modified and hardly show any variation. These represent over 80% of all known modified sites. Remarkably, however, a subset of rRNA positions display variable modification levels, and the exact profile appears to be specific for a given cell line or

cell differentiation step (Figure 5C). Similarly, only few positions vary during stem cells differentiation, and show dependence towards growth medium. The exact reasons and molecular mechanisms underlying such variability are not yet known, but one can speculate that these differences reflect adaptation of eukaryotic ribosome to translation of given mRNA species required to produce essential proteins, or protein isoforms, in particular conditions. The existence of variable pseudouridylated positions is to be interpreted in light of the observation that snoRNA H/ACA-dependent rRNA pseudouridylation (76) is strongly connected to different pathologies and development (77–79). Notably, it was reported that rRNA pseudouridylation affects the balance between cap-dependent and IRES-driven initiation of mRNA translation, and that it affects translation fidelity (49). Profiling of rRNA (and tRNA) pseudouridylation under normal and stress conditions using HydraPsiSeq will undoubtedly help addressing these and other fundamental biological questions in the future.

### SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

### ACKNOWLEDGEMENTS

We thank members of IMoPA Team RNA-RNP, Florence SCHLOTTER, Charbel ALFEGHALY and Lisa HETAL for kind gift of human cells (HeLa, HEK293 and skin fibroblasts) for RNA extraction.

*Author contributions:* V.M., M.H. and Y.M. designed HydraPsiSeq experiments, performed analysis and interpreted the data. F.P. contributed to the development of bioinformatic analysis pipeline. P.N. and A.P. provided BMMSCs differentiation model, L.A. and V.B.I. performed total RNA extraction, library preparation and deep sequencing, L.W. and D.L.J.L. supplied yeast Cbf5 depleted strains and performed depletion, V.M., D.L.J.L., M.H. and Y.M. analyzed data and wrote the paper.

### FUNDING

Y.M. is supported by Epi-ARN FRCR project from Grand Est Region; M.H. by Deutsche Forschungsgemeinschaft (DFG) [HE3397/17-1, RO 4681/6-1, HE 3397/13-2 in the SPP1784]; Research in the Lab of D.L.J.L. is supported by the Belgian Fonds de la Recherche Scientifique (F.R.S./FNRS) [‘RiboEurope’ European Joint Programme on Rare Diseases (EJPRD/JTC2019/PINT-MULTI) grant R.8015.19 and PDR grant T.0144.20]; Université Libre de Bruxelles (ULB), the Région Wallonne (SPW EER) [‘RIBOCancer’ FSO grant 1810070]; Fonds Jean Brachet, the Internationale Brachet Stiftung; Work was performed frame of the EPITRAN COST initiative [CA16120]. Funding for open access charge: lab resources.

*Conflict of interest statement.* None declared.

### REFERENCES

- Li, X., Xiong, X. and Yi, C. (2016) Epitranscriptome sequencing technologies: decoding RNA modifications. *Nat. Methods*, **14**, 23–31.

2. Helm, M. and Motorin, Y. (2017) Detecting RNA modifications in the epitranscriptome: predict and validate. *Nat. Rev. Genet.*, **18**, 275–291.
3. Grozhik, A.V. and Jaffrey, S.R. (2018) Distinguishing RNA modifications from noise in epitranscriptome maps. *Nat. Chem. Biol.*, **14**, 215–225.
4. Motorin, Y. and Helm, M. (2019) Methods for RNA modification mapping using deep sequencing: established and new emerging technologies. *Genes (Basel)*, **10**, 35.
5. Harcourt, E.M., Kietrys, A.M. and Kool, E.T. (2017) Chemical and structural effects of base modifications in messenger RNA. *Nature*, **541**, 339–346.
6. Nachtergaele, S. and He, C. (2018) Chemical modifications in the life of an mRNA transcript. *Annu. Rev. Genet.*, **52**, 349–372.
7. Zaccara, S., Ries, R.J. and Jaffrey, S.R. (2019) Reading, writing and erasing mRNA methylation. *Nat. Rev. Mol. Cell Biol.*, **20**, 608–624.
8. Linder, B. and Jaffrey, S.R. (2019) Discovering and mapping the modified nucleotides that comprise the epitranscriptome of mRNA. *Cold Spring Harb. Perspect. Biol.*, **11**, a032201.
9. Thüring, K., Schmid, K., Keller, P. and Helm, M. (2016) Analysis of RNA modifications by liquid chromatography-tandem mass spectrometry. *Methods*, **107**, 48–56.
10. Thüring, K., Schmid, K., Keller, P. and Helm, M. (2017) LC-MS analysis of methylated RNA. *Methods Mol. Biol.*, **1562**, 3–18.
11. Limbach, P.A. and Paulines, M.J. (2017) Going global: the new era of mapping modifications in RNA. *Wiley Interdiscip. Rev. RNA*, **8**, doi:10.1002/wrna.1367.
12. Jora, M., Lobue, P.A., Ross, R.L., Williams, B. and Addepalli, B. (2018) Detection of ribonucleoside modifications by liquid chromatography coupled with mass spectrometry. *Biochim. Biophys. Acta Gene Regul. Mech.*, **1862**, 280–290.
13. Novoa, E.M., Mason, C.E. and Mattick, J.S. (2017) Charting the unknown epitranscriptome. *Nat. Rev. Mol. Cell Biol.*, **18**, 339–340.
14. Zhao, L., Zhang, H., Kohnen, M.V., Prasad, K.V.S.K., Gu, L. and Reddy, A.S.N. (2019) Analysis of transcriptome and epitranscriptome in plants using PacBio Iso-Seq and Nanopore-Based direct RNA sequencing. *Front. Genet.*, **10**, 253.
15. Dominissini, D., Moshitch-Moshkovitz, S., Schwartz, S., Salmon-Divon, M., Ungar, L., Osenberg, S., Cesarkas, K., Jacob-Hirsch, J., Amariglio, N., Kupiec, M. *et al.* (2012) Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature*, **485**, 201–206.
16. Meyer, K.D., Saletore, Y., Zumbo, P., Elemento, O., Mason, C.E. and Jaffrey, S.R. (2012) Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. *Cell*, **149**, 1635–1646.
17. Slama, K., Galliot, A., Weichmann, F., Hertler, J., Feederle, R., Meister, G. and Helm, M. (2019) Determination of enrichment factors for modified RNA in MeRIP experiments. *Methods*, **156**, 102–109.
18. Molinie, B., Wang, J., Lim, K.S., Hillebrand, R., Lu, Z.-X., Van Wittenberghe, N., Howard, B.D., Daneshvar, K., Mullen, A.C., Dedon, P. *et al.* (2016) m(6)A-LAIC-seq reveals the census and complexity of the m(6)A epitranscriptome. *Nat. Methods*, **13**, 692–698.
19. Schaefer, M., Pollex, T., Hanna, K. and Lyko, F. (2009) RNA cytosine methylation analysis by bisulfite sequencing. *Nucleic Acids Res.*, **37**, e12.
20. Motorin, Y., Lyko, F. and Helm, M. (2010) 5-methylcytosine in RNA: detection, enzymatic formation and biological functions. *Nucleic Acids Res.*, **38**, 1415–1430.
21. Legrand, C., Tuorto, F., Hartmann, M., Liebers, R., Jacob, D., Helm, M. and Lyko, F. (2017) Statistically robust methylation calling for whole-transcriptome bisulfite sequencing reveals distinct methylation patterns for mouse RNAs. *Genome Res.*, **27**, 1589–1596.
22. Marchand, V., Blanloeil-Oillo, F., Helm, M. and Motorin, Y. (2016) Illumina-based RiboMethSeq approach for mapping of 2'-O-Me residues in RNA. *Nucleic Acids Res.*, **44**, e135.
23. Krogh, N., Birkedal, U. and Nielsen, H. (2017) RiboMeth-seq: Profiling of 2'-O-Me in RNA. *Methods Mol. Biol.*, **1562**, 189–209.
24. Ayadi, L., Motorin, Y. and Marchand, V. (2018) Quantification of 2'-O-Me residues in RNA using next-generation sequencing (Illumina RiboMethSeq protocol). *Methods Mol. Biol.*, **1649**, 29–48.
25. Hauenschild, R., Tserovski, L., Schmid, K., Thüring, K., Winz, M.-L., Sharma, S., Entian, K.-D., Wacheul, L., Lafontaine, D.L.J., Anderson, J. *et al.* (2015) The reverse transcription signature of N-1-methyladenosine in RNA-Seq is sequence dependent. *Nucleic Acids Res.*, **43**, 9950–9964.
26. Safran, M., Sas-Chen, A., Nir, R., Winkler, R., Nachshon, A., Bar-Yaacov, D., Erlacher, M., Rossmann, W., Stern-Ginossar, N. and Schwartz, S. (2017) The m1A landscape on cytosolic and mitochondrial mRNA at single-base resolution. *Nature*, **551**, 251–255.
27. Schwartz, S. (2018) m1A within cytoplasmic mRNAs at single nucleotide resolution: a reconciled transcriptome-wide map. *RNA*, **24**, 1427–1436.
28. Marchand, V., Ayadi, L., Ernst, F.G.M., Hertler, J., Bourguignon-Igel, V., Galvanin, A., Kotter, A., Helm, M., Lafontaine, D.L.J. and Motorin, Y. (2018) AlkAniline-Seq: Profiling of m7 G and m3 C RNA modifications at single nucleotide resolution. *Angew. Chem. Int. Ed. Engl.*, **57**, 16785–16790.
29. Cohn, W.E. (1960) Pseudouridine, a carbon-carbon linked ribonucleoside in ribonucleic acids: isolation, structure, and chemical characteristics. *J. Biol. Chem.*, **235**, 1488–1498.
30. Spenkuch, F., Motorin, Y. and Helm, M. (2014) Pseudouridine: still mysterious, but never a fake (uridine)! *RNA Biol.*, **11**, 1540–1554.
31. Carlile, T.M., Rojas-Duran, M.F., Zinshteyn, B., Shin, H., Bartoli, K.M. and Gilbert, W.V. (2014) Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells. *Nature*, **515**, 143–146.
32. Li, X., Ma, S. and Yi, C. (2016) Pseudouridine: the fifth RNA nucleotide with renewed interests. *Curr. Opin. Chem. Biol.*, **33**, 108–116.
33. Bakin, A. and Ofengand, J. (1993) Four newly located pseudouridylate residues in Escherichia coli 23S ribosomal RNA are all at the peptidyltransferase center: analysis by the application of a new sequencing technique. *Biochemistry*, **32**, 9754–9762.
34. Bakin, A., Lane, B.G. and Ofengand, J. (1994) Clustering of pseudouridine residues around the peptidyltransferase center of yeast cytoplasmic and mitochondrial ribosomes. *Biochemistry*, **33**, 13475–13483.
35. Massenet, S., Motorin, Y., Lafontaine, D.L., Hurt, E.C., Grosjean, H. and Branlant, C. (1999) Pseudouridine mapping in the Saccharomyces cerevisiae spliceosomal U small nuclear RNAs (snRNAs) reveals that pseudouridine synthase pus1p exhibits a dual substrate specificity for U2 snRNA and tRNA. *Mol. Cell Biol.*, **19**, 2142–2154.
36. Motorin, Y., Muller, S., Behm-Ansmant, I. and Branlant, C. (2007) Identification of modified residues in RNAs by reverse transcription-based methods. *Methods Enzymol.*, **425**, 21–53.
37. Behm-Ansmant, I., Helm, M. and Motorin, Y. (2011) Use of specific chemical reagents for detection of modified nucleotides in RNA. *J. Nucleic Acids*, **2011**, 408053.
38. Schwartz, S., Bernstein, D.A., Mumbach, M.R., Jovanovic, M., Herbst, R.H., León-Ricardo, B.X., Engreitz, J.M., Guttman, M., Satija, R., Lander, E.S. *et al.* (2014) Transcriptome-wide mapping reveals widespread dynamic-regulated pseudouridylation of ncRNA and mRNA. *Cell*, **159**, 148–162.
39. Li, X., Zhu, P., Ma, S., Song, J., Bai, J., Sun, F. and Yi, C. (2015) Chemical pulldown reveals dynamic pseudouridylation of the mammalian transcriptome. *Nat. Chem. Biol.*, **11**, 592–597.
40. Carlile, T.M., Rojas-Duran, M.F. and Gilbert, W.V. (2015) Pseudo-Seq: genome-wide detection of Pseudouridine modifications in RNA. *Meth. Enzymol.*, **560**, 219–245.
41. Behm-Ansmant, I., Urban, A., Ma, X., Yu, Y.-T., Motorin, Y. and Branlant, C. (2003) The Saccharomyces cerevisiae U2 snRNA:pseudouridine-synthase Pus7p is a novel multisite-multisubstrate RNA:Psi-synthase also acting on tRNAs. *RNA*, **9**, 1371–1382.
42. Massenet, S., Ansmant, I., Motorin, Y. and Branlant, C. (1999) The first determination of pseudouridine residues in 23S ribosomal RNA from hyperthermophilic Archaea Sulfolobus acidocaldarius. *FEBS Lett.*, **462**, 94–100.
43. Schwartz, S. and Motorin, Y. (2017) Next-generation sequencing technologies for detection of modified nucleotides in RNAs. *RNA Biol.*, **14**, 1124–1137.
44. Ofengand, J. (1967) The function of pseudouridylic acid in transfer ribonucleic acid. I. The specific cyanoethylation of pseudouridine, inosine, and 4-thiouridine by acrylonitrile. *J. Biol. Chem.*, **242**, 5034–5045.

45. Mengel-Jørgensen, J. and Kirpekar, F. (2002) Detection of pseudouridine and other modifications in tRNA by cyanoethylation and MALDI mass spectrometry. *Nucleic Acids Res.*, **30**, e135.
46. Emmerechts, G., Herdewijn, P. and Rozenski, J. (2005) Pseudouridine detection improvement by derivatization with methyl vinyl sulfone and capillary HPLC-mass spectrometry. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.*, **825**, 233–238.
47. Lafontaine, D.L., Bousquet-Antonelli, C., Henry, Y., Caizergues-Ferrer, M. and Tollervey, D. (1998) The box H + ACA snoRNAs carry Cbf5p, the putative rRNA pseudouridine synthase. *Genes Dev.*, **12**, 527–537.
48. Zebarjadian, Y., King, T., Fournier, M.J., Clarke, L. and Carbon, J. (1999) Point mutations in yeast CBF5 can abolish in vivo pseudouridylation of rRNA. *Mol. Cell. Biol.*, **19**, 7461–7472.
49. Jack, K., Bellodi, C., Landry, D.M., Niederer, R.O., Meskauskas, A., Musalgaonkar, S., Kopmar, N., Krasnykh, O., Dean, A.M., Thompson, S.R. *et al.* (2011) rRNA pseudouridylation defects affect ribosomal ligand binding and translational fidelity from yeast to human cells. *Mol. Cell.*, **44**, 660–666.
50. Schmitt, M.E., Brown, T.A. and Trumppower, B.L. (1990) A rapid and simple method for preparation of RNA from *Saccharomyces cerevisiae*. *Nucleic Acids Res.*, **18**, 3091–3092.
51. Collart, M.A. and Oliviero, S. (2001) Preparation of yeast RNA. *Curr. Protoc. Mol. Biol.*, doi:10.1002/0471142727.mb1312s23.
52. Bolger, A.M., Lohse, M. and Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**, 2114–2120.
53. Langmead, B. and Salzberg, S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat. Methods*, **9**, 357–359.
54. Pichot, F., Marchand, V., Ayadi, L., Bourguignon-Igel, V., Helm, M. and Motorin, Y. (2020) Holistic optimization of bioinformatic analysis pipeline for detection and quantification of 2'-O-methylations in RNA by RiboMethSeq. *Front. Genet.*, **11**, 38.
55. Birkedal, U., Christensen-Dalsgaard, M., Krogh, N., Sabarinathan, R., Gorodkin, J. and Nielsen, H. (2015) Profiling of ribose methylations in RNA by high-throughput sequencing. *Angew. Chem. Int. Ed. Engl.*, **54**, 451–455.
56. Henrionnet, C., Gillet, P., Mainard, D., Vincourt, J.-B. and Pinzano, A. (2018) Label-free relative quantification of secreted proteins as a non-invasive method for the quality control of chondrogenesis in bioengineered substitutes for cartilage repair. *J. Tissue Eng. Regen. Med.*, **12**, e1757–e1766.
57. Peattie, D.A. (1979) Direct chemical method for sequencing RNA. *Proc. Natl. Acad. Sci. U.S.A.*, **76**, 1760–1764.
58. Küpfer, P.A. and Leumann, C.J. (2007) The chemical stability of abasic RNA compared to abasic DNA. *Nucleic Acids Res.*, **35**, 58–68.
59. Wang, L.K. and Shuman, S. (2002) Mutational analysis defines the 5'-kinase and 3'-phosphatase active sites of T4 polynucleotide kinase. *Nucleic Acids Res.*, **30**, 1073–1080.
60. Taoka, M., Nobe, Y., Yamaki, Y., Yamauchi, Y., Ishikawa, H., Takahashi, N., Nakayama, H. and Isobe, T. (2016) The complete chemical structure of *Saccharomyces cerevisiae* rRNA: partial pseudouridylation of U2345 in 25S rRNA by snoRNA snR9. *Nucleic Acids Res.*, **44**, 8951–8961.
61. Decatur, W.A. and Schnare, M.N. (2008) Different mechanisms for pseudouridine formation in yeast 5S and 5.8S rRNAs. *Mol. Cell. Biol.*, **28**, 3089–3100.
62. Ma, X., Yang, C., Alexandrov, A., Grayhack, E.J., Behm-Ansmant, I. and Yu, Y.-T. (2005) Pseudouridylation of yeast U2 snRNA is catalyzed by either an RNA-guided or RNA-independent mechanism. *EMBO J.*, **24**, 2403–2413.
63. Rintala-Dempsey, A.C. and Kothe, U. (2017) Eukaryotic stand-alone pseudouridine synthases - RNA modifying enzymes and emerging regulators of gene expression? *RNA Biol.*, **14**, 1185–1196.
64. Carlile, T.M., Martinez, N.M., Schaening, C., Su, A., Bell, T.A., Zinshteyn, B. and Gilbert, W.V. (2019) mRNA structure determines modification by pseudouridine synthase 1. *Nat. Chem. Biol.*, **15**, 966–974.
65. Taoka, M., Nobe, Y., Yamaki, Y., Sato, K., Ishikawa, H., Izumikawa, K., Yamauchi, Y., Hirota, K., Nakayama, H., Takahashi, N. *et al.* (2018) Landscape of the complete RNA chemical modifications in the human 80S ribosome. *Nucleic Acids Res.*, **46**, 9289–9298.
66. Erales, J., Marchand, V., Panthu, B., Gillot, S., Belin, S., Ghayad, S.E., Garcia, M., Laforêts, F., Marcel, V., Baudin-Baillieu, A. *et al.* (2017) Evidence for rRNA 2'-O-methylation plasticity: Control of intrinsic translational capabilities of human ribosomes. *Proc. Natl. Acad. Sci. U.S.A.*, **114**, 12934–12939.
67. Sharma, S., Marchand, V., Motorin, Y. and Lafontaine, D.L.J. (2017) Identification of sites of 2'-O-methylation vulnerability in human ribosomal RNAs by systematic mapping. *Sci. Rep.*, **7**, 11490.
68. Foster, L.J., Zeemann, P.A., Li, C., Mann, M., Jensen, O.N. and Kassem, M. (2005) Differential expression profiling of membrane proteins by quantitative proteomics in a human mesenchymal stem cell line undergoing osteoblast differentiation. *Stem Cells*, **23**, 1367–1377.
69. Billing, A.M., Ben Hamidane, H., Dib, S.S., Cotton, R.J., Bhagwat, A.M., Kumar, P., Hayat, S., Youstri, N.A., Goswami, N., Suhre, K. *et al.* (2016) Comprehensive transcriptomic and proteomic characterization of human mesenchymal stem cells reveals source specific cellular markers. *Sci. Rep.*, **6**, 21507.
70. Mueller, M.B., Fischer, M., Zellner, J., Berner, A., Dienstknecht, T., Prantl, L., Kujat, R., Nerlich, M., Tuan, R.S. and Angele, P. (2010) Hypertrophy in mesenchymal stem cell chondrogenesis: effect of TGF-beta isoforms and chondrogenic conditioning. *Cells Tissues Organs (Print)*, **192**, 158–166.
71. Cicione, C., Muiños-López, E., Hermida-Gómez, T., Fuentes-Boquete, I., Díaz-Prado, S. and Blanco, F.J. (2015) Alternative protocols to induce chondrogenic differentiation: transforming growth factor-β superfamily. *Cell Tissue Bank.*, **16**, 195–207.
72. Courtes, F.C., Gu, C., Wong, N.S.C., Dedon, P.C., Yap, M.G.S. and Lee, D.-Y. (2014) 28S rRNA is inducibly pseudouridylated by the mTOR pathway translational control in CHO cell cultures. *J. Biotechnol.*, **174**, 16–21.
73. Xue, S. and Barna, M. (2012) Specialized ribosomes: a new frontier in gene regulation and organismal biology. *Nat. Rev. Mol. Cell Biol.*, **13**, 355–369.
74. Guo, H. (2018) Specialized ribosomes and the control of translation. *Biochem. Soc. Trans.*, **46**, 855–869.
75. Marchand, V., Pichot, F., Thüring, K., Ayadi, L., Freund, I., Dalpke, A., Helm, M. and Motorin, Y. (2017) Next-generation sequencing-based RiboMethSeq protocol for analysis of tRNA 2'-O-methylation. *Biomolecules*, **7**, 13.
76. Yu, Y.-T. and Meier, U.T. (2014) RNA-guided isomerization of uridine to pseudouridine-pseudouridylation. *RNA Biol.*, **11**, 1483–1494.
77. Tortoriello, G., de Celis, J.F. and Furia, M. (2010) Linking pseudouridine synthases to growth, development and cell competition. *FEBS J.*, **277**, 3249–3263.
78. McMahon, M., Contreras, A. and Ruggiero, D. (2015) Small RNAs with big implications: new insights into H/ACA snoRNA function and their role in human disease. *Wiley Interdiscip Rev RNA*, **6**, 173–189.
79. McMahon, M., Contreras, A., Holm, M., Uechi, T., Forester, C.M., Pang, X., Jackson, C., Calvert, M.E., Chen, B., Quigley, D.A. *et al.* (2019) A single H/ACA small nucleolar RNA mediates tumor suppression downstream of oncogenic RAS. *Elife*, **8**, e48847.