Quantitative nucleotide resolution profiling of RNA cytidine acetylation by ac4C-seq

Quantitative nucleotide resolution profiling of RNA cytidine acetylation by ac4C-seq

Supuni Thalalla Gamage*, Aldema Sas-Chen*, Schraga Schwartz,[#] Jordan L. Meier[#]

[1]Chemical Biology Laboratory, National Cancer Institute, NIH, Frederick MD 21702-1201, USA

[2]Department of Molecular Genetics, Weizmann Institute of Science, Rehovot 7610001, Israel

[*]These authors have contributed equally to this manuscript

[#]Corresponding authors: jordan.meier@nih.gov, schwartz@weizmann.ac.il

EDITORIAL SUMMARY The N4-acetylcytidine (ac4C) RNA modification is conserved in all domains of life. Here, the authors provide a detailed protocol for whole transcriptome quantitative nucleotide resolution mapping of this modified nucleobase in different organisms.

PROPOSED TWEET A new protocol describes a sequencing-based approach for quantitative nucleotide resolution mapping of N4-acetylcytidine (ac4C) RNA.
Twitter handles @Dunee_Ga, @AldemaSasChen, @SchragaSchwartz, @doc_jlmeier
ALTERNATIVE: **A new sequencing-based protocol for whole transcriptome quantitative nucleotide resolution mapping of N4-acetylcytidine (ac4C) RNA.**

PROPOSED TEASER Mapping cytidine acetylation in RNA

Related links

Key reference using this protocol:

Sinclair, W. et al. ACS Chem. Biol. 12, 2922–2926 (2017): https://doi.org/10.1021/acschembio.7b00734

Thomas, J. et al. J. Am. Chem. Soc. 140, 12667–12670 (2018): https://doi.org/10.1021/jacs.8b06636

Sas-Chen, A. et al. Nature 583, 638–643 (2020). https://doi.org/10.1038/s41586-020-2418-2

Abstract

A prerequisite to defining the transcriptome-wide functions of RNA modifications is the ability to accurately determine their location. Here we present N4-acetylcytidine sequencing (ac4C-seq), a protocol for the quantitative single nucleotide resolution mapping of cytidine acetylation in RNA. This method exploits the kinetically facile chemical reaction of ac4C with sodium cyanoborohydride under acidic conditions to form a reduced nucleobase. RNA is then fragmented, ligated to an adapter at its 3' end, and reverse transcribed to introduce a non-cognate nucleotide at reduced ac4C sites. Following adapter ligation, library preparation, and high-throughput sequencing, a bioinformatic pipeline enables identification of ac4C positions based on the presence of C>T misincorporations in reduced samples but not in controls. Unlike antibody-based approaches, ac4C-seq identifies specific ac4C residues and reports on their level of modification. The ac4C-seq library preparation protocol can be completed in ~4 d for transcriptome-wide sequencing.

Introduction

N4-acetylcytidine (ac4C) is a modified RNA nucleobase that was first discovered in 1960's during early efforts to characterize eukaryotic transfer RNAs (tRNAs)[1]. Subsequent studies found that this RNA modification is conserved in all domains of life[2,3], raising the provocative question of its function. In many bacteria, ac4C occurs at the wobble base of initiator $tRNA^{Met}$ and encourages the selective utilization of this tRNA, increasing translational fidelity[4]. In archaea, ac4C's rigidified structure was hypothesized to increase the stability of RNA to thermal denaturation as early as the 1990's[3], a supposition borne out by recent studies[5]. In eukaryotes, where ac4C is found in $tRNA^{Leu}$, and $tRNA^{Ser}$, the combined disruption of ac4C and another tRNA modification (N7-methylguanosine; m7G) destabilizes $tRNA^{Ser}$ in mutant yeast strains, inhibiting growth[6,7]. Building on these seminal studies, a breakthrough came in late 2014 and early 2015, when multiple groups concurrently reported that eukaryotic ac4C formation is catalyzed by an orphan GCN5-related N-acetyltransferase enzyme known as Nat10 (Kre33 in yeast)[8,9,10]. To direct its activity towards specific targets, Nat10 interacts with an additional adaptor protein – Thumpd1 in humans and Tan1 in yeast – which is required for modification of $tRNA^{Leu}$, and $tRNA^{Ser}$ [6,10]. Alternatively, specialized short nucleolar RNAs (snoRNAs) guide Nat10's acetylation of eukaryotic ribosomal RNA (rRNA)[11]. The unique biochemistry of Nat10/Kre33, together with the fact that cytidine acetyltransferases genes are highly conserved[12], essential in eukaryotes[13,14], and associated with several diseases including premature aging and cancer[15,16], has catalyzed the development of new methods to study ac4C.

Here we describe a detailed protocol for ac4C-seq, a high-throughput sequencing method that enables the quantitative base resolution determination of the location, abundance, and dynamics of cytidine acetylation in RNA.

Development of the protocol

Cytidine acetylation does not disrupt canonical base pairing, and is therefore undetectable in conventional RNA sequencing protocols[17]. However, ac4C has two unique chemical properties which together enable its sequencing-based detection. First, N4-acetylation withdraws electron density from cytidine's pyridine ring, increasing the susceptibility of the modified nucleobase to reduction by hydride donors[18]. Addition of two equivalents of sodium borohydride to ac4C results in formation of a reduced nucleobase, tetrahydro-N4-acetylcytidine (here termed 'reduced ac4C'). Reduced ac4C can be misread as a 'U' rather than a 'C' during reverse transcription, causing C>T mutations at sites of ac4C to be observed upon cDNA sequencing[19] (Figure 1). One limitation of this approach is that in addition to ac4C, several other modified nucleobases (m7G, dihydrouridine, N3-methylcytidine) are also susceptible to reduction by hydride donors, meaning the presence of a hydride-dependent misincorporation alone is insufficient to specify a site of cytidine acetylation. To address this, a second distinct chemical feature of ac4C, its hydrolytic lability, can be exploited to chemically deacetylate RNA for control experiments[20]. Together, these reactions establish a unique chemical signature for ac4C that consists of: 1) increased C>T conversion upon reduction by hydride donors, which is 2) abrogated by pre-treating RNA with mild alkali. Crucial elements of this approach that have been optimized for sensitive and quantitative ac4C detection include the reduction chemistry, deacetylation reaction, genetic controls, reverse transcriptase, library preparation, and analytical pipeline, all of which are described in detail under 'Experimental Design' (Figure 2 and 3).

Applications of the method

In our initial studies we have applied ac4C-seq to profile cytidine acetylation within human cell lines, fungal model organisms, and archaeal extremophiles[5]. In eukaryotes, ac4C-seq enabled the *de novo* annotation of all four previously known sites of ac4C in rRNA and tRNA, without requiring a priori knowledge of their location. The quantitative nature of the method also validated these rRNA and tRNA residues as the major targets of Nat10/Kre33-catalyzed cytidine acetylation in eukaryotes, suggesting additional ac4C sites are either absent or present at very low occupancies under physiological conditions. In addition to informing on endogenous biology, ac4C-seq can be used to study the biochemistry of Nat10 RNA acetyltransferases. As one

example, ac4C-seq has been used to monitor the ability of co-overexpressed human Nat10 and Thumpd1 to acetylate hundreds of messenger RNA (mRNA) transcripts, providing new insights into the sequence and structural determinants required for productive enzyme-substrate interaction[5]. The quantitative nature of ac4C-seq has also been powerfully employed to study RNA acetylation dynamics within archaeal hyperthermophiles. Here, application of ac4C-seq to archaea of the order Thermococcales led to the discovery of a widespread, temperature-inducible program of cytidine acetylation in rRNA, tRNA, and mRNA. This provides a model biological system for studying the physical and phenotypic consequences of cytidine acetylation, aspects of which may be conserved across evolution.

With regards to future applications, it is important to note that in eukaryotes transcriptome-wide methods may not be necessary to study the dominant sites ac4C sites found in rRNA and tRNA, although we anticipate analysis of the dynamics of these sites will benefit from quantitative sequencing. Instead, in these systems ac4C-seq may find its most powerful application in discovering and validating ac4C sites in viral genomes (evidence for which has been recently reported)[21], exploring whether ac4C occurs within mRNA at low-stoichiometries in settings marked by Nat10 overexpression such as cancer, and studying how ac4C is targeted by protein and snoRNA adapters[22,23,24]. In addition, it may be possible to integrate ac4C-seq with other hydride-based sequencing methods to simultaneously profile additional modified nucleobases including m7G[25,26], further expanding the scope of this method. Overall, these applications highlight the potential for ac4C-seq to facilitate the high resolution study of cytidine acetylation, enabling interrogation of its functional role in fundamental biology and disease.

Comparison with other methods

LC-MS

Liquid chromatography coupled to mass spectrometry (LC-MS) has been powerfully applied to study ac4C by several groups, most notably Suzuki and coworkers who used LC-MS analysis of partially digested tRNA and rRNA fragments as a primary assay to identify ac4C sites and the genes responsible for ~~subjected to~~ cytidine acetylation in bacteria, yeast, and human cell lines[8,9,27]. In a related method, Sharma et al. used a mung bean nuclease protection assay coupled to high performance liquid chromatography with UV detection (UV-HPLC) to identify ac4C within helix 34 of eukaryotic rRNA, as well as to identify the yeast snoRNAs responsible for guiding rRNA acetylation[10,11]. While these methods provide validated methods for discovery, they each require significant technical expertise and suffer from sensitivity limitations due to

their lack of signal amplification. This latter limitation may hinder the detection and quantification of ac4C in low abundance species (i.e. mRNAs) or within RNAs that are only partially modified.

Affinity-based methods

Antibody-based methods have been used to detect ac4C-containing RNAs via dot blot and Northern blotting[5,20], as well as to enrich and identify candidate ac4C-containing mRNAs in both human and viral transcriptomes via immunoprecipitation and sequencing[21,28]. These methods have the advantage of signal amplification, in the form of chemiluminescent detection for the immunoblotting detection protocols, and enrichment for immunoprecipitation-based detection. A major limitation of antibody-based methods is they cannot detect ac4C at nucleotide resolution, nor quantify its abundance at putative modification sites. Another limitation of RNA modification immunoprecipitation-sequencing methods is their susceptibility to false positives, which can be caused by even the slight degree of antibody promiscuity when attempting to enrich a low abundance modification from an enormous transcriptomic background[29,30,31].

Single-molecule sequencing

A third and emerging method for the analysis of ac4C is the use of single-molecule sequencing technologies[32]. In one such approach, RNA molecules are pushed by a motor protein through a membrane-bound pore of an Oxford Nanopore Technologies sequencer. This causes a change in the transduction of electric current through the nanopore in a manner that is sequence dependent, and also can reflect the presence of minor nucleobases such as ac4C which produce basecalling errors at modified sites. Very recently, Grunberger and coworkers applied this technology to the ribosomal RNA of archaeal hyperthermophiles and were able to profile 34 sites of cytidine acetylation, many of which have been validated by ac4C-seq[5,33]. Key advantages of single-molecule platforms is their potential utility in interrogating dense epitranscriptomic landscapes by detecting multiple modifications simultaneously across the entire length of a transcript and the avoidance of biases introduced by reverse transcription and PCR amplification during conventional RNA-seq workflows. Current limitations are their relatively low sensitivity, specificity and throughput and challenging analysis of small RNAs. The continued development of these platforms will likely facilitate additional creative applications in the study of ac4C and other RNA modifications.

Experimental design

Preparation of synthetic ac4C spike-in RNA

A valuable element of ac4C-seq is the use of synthetic RNAs spike-ins that contain ac4C at a defined position as a control for assessing modified RNA detection[5]. Since synthetic ac4C RNAs cannot be produced by standard phosphoramidite RNA synthesis, spike-ins are prepared via standard in vitro transcription. These reactions utilize a T7 RNA Polymerase template which encodes a single 'C' residue in an AUG-rich sequence and substitute N4-acetylcytidine triphosphate in the place of cytidine triphosphate[19]. Preparation of four synthetic RNA spike-ins which contain a single ac4C site in a defined sequence context (ACA, GCA, UCA, ACU) as well as four corresponding cytidine-containing RNAs allows the preparation of a spike-in mixture in which ac4C is present at a defined stoichiometry (10%, 40%, 60%, 100%) within four different sequences, allowing the sensitivity of ac4C detection to be assessed. Since ac4C is a hydrolytically labile RNA modification, synthetic spike-ins should be routinely monitored for degradation via cyanoborohydride-dependent misincorporation assay (see '*ac4C reduction and mock treatment' section*). Hydrolysis of ac4C within the spike-in would be expected to reduce C>T misincorporations at the known ac4C site. In practice, we have found aqueous solutions of the synthetic ac4C spike-in RNA mixture are stable for months when stored at -80 ºC.

RNA isolation and deacetylation control reaction

RNA to be analyzed by ac4C-seq is prepared from biological samples by standard Trizol extraction and split into three equal fractions, with an aliquot of synthetic ac4C spike-in RNA mixture added to each. One of these fractions is subjected to chemical deacetylation by treatment with mild alkali, and is referred to as the 'chemical deacetylation control' throughout this protocol. This fraction provides an important specificity control for ac4C detection, as hydrolysis of the N4-acetyl group decreases cyanoborohydride-dependent C>T misincorporations at ac4C sites, while leaving other hydride-reducible modifications unaffected. Our group has evaluated several reagents for chemical deacetylation of ac4C[20], and favors alkali hydrolysis due to the lack of competing side reactions and known compatibility with next-generation sequencing library preparation. One consideration with chemical deacetylation is that even short exposure to alkali conditions can cause fragmentation of RNA, thus it is important not to prolong the deacetylation step beyond the recommended duration. Fragmentation in general, and in the deacetylation step will affect the final size of the cDNA library, and hence will affect calculation of final dilutions prior to Illumina sequencing.

ac4C reduction and mock treatment

After addition of spike-ins, RNA is treated with sodium cyanoborohydride in the presence of strong acid to cause the reduction of ac4C-modified sites. This treatment is applied to two fractions (native RNA and the chemical deacetylation control), and are referred to below as the 'treated' or 'reduced' samples. The remaining (third) RNA sample is subjected to a mock treatment with only strong acid but no reducing agent, and serves as the non-treated input control. One consideration in choosing an appropriate reduction chemistry is the relative rate of ac4C reduction as compared to ac4C hydrolysis, as the latter reaction erases the signal for cytidine acetylation. As with the chemical deacetylation reaction, our group has evaluated several reactions for the chemical reduction of ac4C, and favors our recently reported use of sodium cyanoborohydride in acidic (pH 1) conditions. In model reactions we have observed that the reduction of ac4C under these conditions proceeds approximately 10-fold faster than acid-catalyzed hydrolysis reaction[5], providing it with excellent signal to noise when integrated with next-generation sequencing workflows.

Choice of reverse transcriptase

Since ac4C-seq detection of cytidine acetylation sites depends on analysis of chemically induced C>T misincorporation sites, the choice of RT is essential[34]. Different polymerases exhibit different abilities to read through the reduced nucleobase. For locus-specific detection of ac4C sites in rRNA, our group has used the genetically engineered Moloney Murine Leukemia Virus SuperScript III (SS-III), HIV reverse transcriptase, and Type Group II Intron Reverse Transcriptase (TGIRT-III), all of which demonstrate a mixture of read-through and stop activity when transcribing ac4C sites[19]. For next-generation sequencing analysis of ac4C we have primarily used TGIRT-III, based on its excellent compatibility with library preparation and misincorporation-based RNA modification detection[5,29], but have also obtained highly comparable results using SS-III.

Genetic controls

In addition to spike-ins, which provide a control for monitoring quantitative ac4C reduction, and the deacetylation fraction, which provides evidence that a putative modified site shares the known chemical properties of ac4C (alkaline lability), the use of organismal models in which cytidine acetyltransferase enzymes have been inactivated provides a powerful genetic control for ac4C detection. Almost all known eukaryotic organisms contain a single RNA acetyltransferase gene, which in many cases has been found to be essential. However, in humans it has been shown that genetic targeting of a site that allows only a minor splice isoform

to be produced can result in the generation of hypomorphic cell lines in which cytidine acetyltransferases are present at very low levels[21,28]. Alternatively, in yeast it has been shown that mutation of the acetyltransferase domain is not lethal and can be used to produce RNA that lacks acetylation[10]. In archaea and bacteria knockout of cytidine acetyltransferase enzymes is tolerated and results in growth defects only under specific conditions[5,27,35]. In addition to loss of function genetic controls, gain-of-function genetic controls can be used as well. For example, overexpression of Nat10 and Thumpd1 in human cells causes supraphysiological acetylation of hundreds of mRNA sites, which can be attributed directly to the activity of the ectopically overexpressed proteins[5].

Library preparation and sequencing

The treated RNA samples are used for high-throughput sequencing library construction using the ac4c-seq protocol detailed in **steps 22-66** of the **PROCEDURE** section below. In short, RNA is first fragmented to produce ~200 bp long fragments, followed by ligation of an RNA adapter, used to prime reverse transcription. Reverse transcription is then performed using the TGIRT-III enzyme, which results in C>T misincorporation at the position of a reduced-ac4C. A second adapter is then ligated to the 3′ end of the single-stranded cDNA product. Importantly, both the RNA and cDNA adapters used are identical across all libraries, minimizing biases. Barcoded primers are then used to amplify the sequencing library via PCR. Libraries are subsequently sequenced on Illumina NextSeq 500 or NovaSeq 6000 platforms generating short paired-end reads, ranging from 25 to 55 bp from each end. While sequencing in single-end mode is possible, we advice using paired-end mode, which improves the accuracy of sequence alignment and allows better estimation of sample complexity. Of note, fragmentation is not needed when profiling short molecules, such as tRNAs. Although conducting the washing steps with 0.8 volumes of ethanol partially preserves short fragments, further adaptations can be made to the protocol to facilitate analysis of these transcripts.

Data analysis

Detection of ac^{4}C sites is based on identification of a specific C->T misincorporation signature in reduced samples but not in the control samples, as detailed in **steps 67-74** of the **PROCEDURE** section below. In short, raw reads generated via Illumina sequencing are mapped to a custom genome using the STAR aligner[36], which allows for local alignment (e.g. by activating the “soft clipping” option), followed by detection of single-nucleotide variants using the JACUSA software in pileup mode[37]. The JACUSA output is then used to calculate the

misincorporation rate of each position by dividing the number of non-cytidine containing reads by the total number of reads covering that position. Higher misincorporations indicate a higher stoichiometry of ac4C at the given position. Binomial tests of significance can further be conducted to identify putative ac4C sites, whose misincorporation rate in the reduced sample is significantly higher than that of the control samples (chemical and/or genetic controls). Occurrence of the putative ac4C site within a CCG motif, where the middle cytidine is modified, further strengthens confidence in assigned sites, as the CCG tri-nucleotide sequence motif was found to be necessary for targeting by Nat10 and its homologs in eukaryotes and archaea[5]. Additionally, when relevant, putative sites should be assessed post hoc to ensure they do not originate from misalignaments and misinterpretations related to complexity and redundancy of the genome (e.g., that they do not stem from alignment to repetitive elements, nuclear mitochondrial DNA (NUMTs), paralogs, SNPs etc.)[34].

Advantages and limitations

The advantage of ac4C-seq relative to related methods are its quantitative nature and nucleotide resolution, which enable monitoring ac4C dynamics and *de novo* discovery of novel ac4C sites. A limitation that ac4C-seq shares in common with all chemical reaction-based modification sequencing protocols is that its detection limit at any given site is dependent on the stoichiometry of ac4C, the efficiency of the ac4C reduction, and the sequencing depth of that site[34]. Misincorporations caused by quantitative reduction of ac4C using the protocol described here have been shown to scale linearly with LC-MS measurements of ac4C stoichiometry in a single-stranded spike-in RNAs (see '*Preparation of synthetic ac4C spike-in RNA*' section) and twenty-seven sites of ac4C in eukaryotic and archaeal rRNA[5]. In addition, ac4C-seq has allowed detection of cytidine acetylation at >200 archaeal sites spread across tRNA, mRNAs, and ncRNAs, and within an additional 230 sites that can be induced in eukaryotic mRNA upon Nat10/Thumpd1 or Kre33/Tan1 overexpression[5]. These observations indicate that cyanoborohydride-mediated reduction of ac4C occurs efficiently across a broad range of RNA contexts to produce a signal that can be robustly detected, and suggesting the sensitivity of ac4C detection by ac4C-seq is likely to primarily depend on stoichiometry and sequencing depth. Deeper sequencing depth of low stoichiometry sites may be achieved by targeted amplicon sequencing or, theoretically, through pre-enrichment of ac4C-containing RNAs using commercially available anti-ac4C antibodies[20], although the latter approach has not yet been integrated with ac4C-seq.

One limitation of ac4C-seq is that it relies exclusively on detection of C>T misincorporations to identify and quantify ac4C sites, which has the potential to cause an underestimation of modification abundance within RNAs where the reduced ac4C nucleobase causes RT stop events. Thus, the quantitative nature of ac4C-seq may be further improved by combined quantification of both misincorporation and RT stop events[38]. While we have successfully applied this protocol to the study of ac4C in archaeal and eukaryotic tRNA, optimized quantitative analysis of tRNA acetylation may further benefit from pre-treatment of RNA with promiscuous demethylase enzymes, which have previously been shown to increase tRNA sequencing depth due to
removal of methylated modifications that block reverse transcription at those sites[39,40], and by replacing the fragmentation step with simple size selection, in order to limit degradation of these small RNAs.

Finally, it should be noted that the individual chemical treatments used in ac4C-seq can exhibit cross-reactivity with other modified nucleobases. For instance, the alkali conditions used to chemically deacetylate ac4C can also cause Dimroth rearrangements at N1-methyladenosine (m1A) sites, and cyanoborohydride in strong acid was found to cause a C>T misincorporation at a 5-formylcytidine (5fC) site in mitochondrial tRNA[5]. While ac4C can be distinguished from these nucleobases through additional filters (cyanoborohydride reactivity, occurrence at a 5'-CCG-3' consensus sequence, sensitivity to alkali treatment), the development of more specific chemical or enzymatic methods may further simplify such analyses.

Materials

Biological materials

- Cells and tissues from any organism can be used to isolate RNA for ac4C-seq. Specifically, ac4c-seq has been successfully conducted on RNA from HeLa (https://scicrunch.org/resolver/CVCL_0030) and HEK-293T (https://scicrunch.org/resolver/CVCL_0063) human cells. We have also applied ac4C-seq to RNA isolated from S.cerevisiae and from archaea of the following strains: T. kodakarensis strain TS559, P. furiosus strain COM1, Thermococcus sp. AM4, Methanocaldococcus jannaschii and Saccharolobus solfataricus. Experiments shown in the ANTICIPATED RESULTS section were carried out on HeLa and TS559 cells, as indicated in the figure legend.

! CAUTION Use aseptic techniques and practices to avoid contamination. Human cell lines used should be regularly checked to ensure that they are not infected with mycoplasma.

Reagents

- TRIzol™ reagent (Invitrogen, cat. no. 15-596-018) ! CAUTION TRIzol reagent is a serious health hazard, use personal protective equipment when handling.
- Chloroform (Sigma-Aldrich, cat. no. C2432) !CAUTION Chloroform is a health hazard. Wear personal protective equipment when handling.
- Isopropanol (Acros organics, cat. no. 32727-0010) !CAUTION Isopropanol is highly flammable; keep away from all sources of ignition.
- Ethanol (Fisher Scientific, cat. no. BP2818) !CAUTION Ethanol is highly flammable; keep away from all sources of ignition.
- Nuclease-free ultra-pure water (Invitrogen, cat. no. 10977)
- Sodium Cyanoborohydride ($NaCNBH_3$, Sigma-Aldrich, cat. no. 156159) !CAUTION This is a flammable solid and a health hazard. Keep away from sources of ignition and any possible contact with water. Use personal protective equipment when handling, and work in a chemical hood
- Sodium bicarbonate (Sigma-Aldrich, cat. No. S6297)
- Sodium acetate (3 M; pH 5.5; Invitrogen, cat. No. AM9740)
- Hydrochloric acid solution, BioReagent, suitable for cell culture (HCl, 1N; Sigma, cat. no. H9892)
- UltraPure™ 1 M Tris-HCl Buffer, pH 8.0 (Invitrogen, cat. no. 15567027)
- RNA Fragmentation Reagents (Ambion, cat. no. AM8740). Includes Fragmentation Reagent and Stop Solution.
- Silane beads - Dynabeads™ MyOne™ Silane (Thermo Fisher, cat. no. 37002D)
- RLT buffer (RNeasy Lysis Buffer) (Qiagen, cat. no. 79216)
- T4 PNK (includes 10X PNK buffer containing 700 mM Tris-HCl pH 7.6, 100 mM $MgCl_2$, 50 mM DTT; New England Biolabs, cat. no. M0201L)
- FastAP Thermosensitive Alkaline Phosphatase (includes 10X Fast AP buffer containing 100 mM Tris-HCl pH 8, 50 mM $MgCl_2$, 1 M KCl, 0.2% (vol/vol) Triton X-100, 1 mg/mL BSA; Thermo Fisher Scientific, cat. no. EF0654)
- DMSO (Sigma-Aldrich, cat. no. D8418) !CAUTION DMSO is flammable liquid; keep away from all sources of ignition.

- RNase Inhibitor, Murine (New England Biolabs, cat. no. M0314)
- Turbo DNAse (Life technologies, cat. no. AM2238)
- T4 RNA Ligase 1 (ssRNA Ligase) high conc. (includes 100 mM ATP, 50% (w/v) PEG 800, 10X T4 RNA ligase reaction buffer; New England Biolabs, cat. no. M0437M)
- TGIRT™-III reverse transcription enzyme (INGEX)
- Potassium chloride (Sigma-Aldrich, cat. no. P9333)
- 50 mM Magnesium chloride solution (New England Biolabs, cat. no. B0510A)
- dNTP Mix (10 mM Solution, Jena Bioscience, cat. no. NU-1006L).
- ExoSap-it PCR Product Cleanup Reagent (Affymetrix, cat. no. 78201)
- DTT (0.1 M, Invitrogen, cat. no. Y00147)
- Ethylenediaminetetraacetic acid disodium salt solution for molecular biology (EDTA, 0.5 M in H_2O; Sigma-Aldrich, cat. no. 7889).
- Sodium hydroxide solution BioUltra, for molecular biology, (NaOH, 10 M in H_2O; Sigma, cat. no. 72068)
- 2X KAPA HiFi HotStart ReadyMix PCR Kit (KAPA Biosystems)
- 2% (w/v) E-Gel™ EX Agarose Gel (Invitrogen, cat. no. G402002)
- 100bp DNA Ladder (New England Biolabs, cat. no. N0467S)
- Ampure Beads - Ampure XP Agencourt (60mL, Beckman Coulter, cat. no. A63881)
- Qubit™ RNA HS Assay Kits (Invitrogen, cat. no. Q32852),
- Qubit™ dsDNA HS Assay Kit (Invitrogen, cat. no. Q32854)
- High Sensitivity RNA ScreenTape Analysis (Agilent, cat. no. 5067-5579)
- High Sensitivity DNA ScreenTape Analysis (Agilent, cat. no. 5067-5584)
- RNA adaptor for first ligation: 5'- /5Phos/rArG rArUrC rGrGrA rArGrA rGrCrA rCrArC rGrUrC/3ddC/ -3' (Integrated DNA Technologies, custom order, standard desalting. Dissolved in H_2O to 10 µM and stored at -20 °C for up to 1 year)
- DNA adaptor for second ligation: 5'- /5Phos/AG ATC GGA AGA GCG TCG TGT AG/3ddC/ -3' (Integrated DNA Technologies, custom order. Dissolved in H_2O to 100 µM and stored at -20 °C for up to 3 years)
- Primer for reverse transcription: 5'- AGA CGT GTG CTC TTC CG - 3' (Integrated DNA Technologies, custom order. Dissolved in H_2O to 10 µM and stored at -20 °C for up to 3 years)
- Primers for PCR (Integrated DNA Technologies, custom order. Dissolved in H_2O to 100 µM and stored at -20 °C for up to 3 years. Working dilution to be used at step 49 is 25 µM). Barcoded primers as published before[41]. Sequences of the Read 1 primer

(2P_universal) and three possible barcoded Read 2 primers (2P_504, 2P_375, and 2p_630) are provided in Table 1.

Name	Sequence (5'-3')
Primer 2P_universal	AATGATACGGCGACCACCGAGATCTACACTCTTTCC CTACACGACGCTCTTCCGATCT
2P_504	CAAGCAGAAGACGGCATACGAGATCCTGGTAGGTGACTGGAGTTCAGA CGTGTGCTCTTCCGATCT
2P_375	CAAGCAGAAGACGGCATACGAGATTAAGCATGGTGACTGGAGTTCAGA CGTGTGCTCTTCCGATCT
2P_630	CAAGCAGAAGACGGCATACGAGATAGATGTGCGTGACTGGAGTTCAGA CGTGTGCTCTTCCGATCT

Table 1. PCR primers to be used in step 49.

Equipment

- Microcentrifuge tubes (1.7 ml; Sorenson Bioscience, cat. no. 11700)
- PCR Tubes (0.2 mL, Flat Cap; Corning, cat. no. 3745), or PCR strips (0.2 mL, flat cap; Gunster MB-P08-B)
- Refrigerated microcentrifuge (Eppendorf, model no. 5424)
- Vortex mixer
- Veriti™ 96-Well Thermal Cycler (Applied Biosystems, cat. no. 4375786)
- Agarose gel electrophoresis equipment
- E-Gel™ Safe Imager™ E-Gel Real-Time Transilluminator (Invitrogen, cat. no. G6500)
- NanoDrop 2000 Spectrophotometer (Thermo Scientific)
- DynaMag™-96 Side Magnet (Thermo Fisher, cat. no. 12331D)
- Ice bucket/cold block
- Sequencing system (Illumina, NextSeq 500 or NovaSeq 6000)
- Desiccator for storing $NaCNBH_3$ stock
- Qubit® 3.0 Fluorometer (Invitrogen, Q33216)
- 4200 TapeStation system (Agilent, G2991A)

Software and data

- STAR (version 2.5.3a) for mapping short reads to to call modification sites[36]
- Samtools (v.0.1.19) for converting, sorting and indexing the SAM/BAM mapping result files[42] (http://www.htslib.org/)
- JACUSA for base calling from BAM files[37]
- R[43] (https://www.r-project.org)
- R package lme4[44]

Reagent setup

1 M $NaCNBH_3$

! CAUTION $NaCNBH_3$ is a flammable solid, which releases flammable and toxic gases upon reaction with water and/or HCl. Thus all handling should be carried out in a chemical hood. Refer to the reagent's safety data sheet for information of disposal. In a 1.7 ml tube, dissolve 10 mg of $NaCNBH_3$ in 159.1 µl of H_2O. The solution should be made up fresh each time. Keep the tube open to avoid accumulation of gases, which are released upon addition of H_2O. Leftover Sodium Cyanoborohydride dissolved in water, should be quenched with equal volume of 1 M HCl prior to disposal.

100 mM sodium bicarbonate pH 9.5

Dissolve 8.401 mg of sodium bicarbonate in 800 µl of nuclease-free H_2O. Adjust pH to 9.5 with 5 M NaOH. Bring final volume to 1 mL with H_2O. This solution can be stored at room temperature up to a year.

FNK buffer

Mix 300 µL of 10X PNK buffer (NEB), 300 µL of 10X Fast AP buffer (Thermo Fisher Scientific) and 600 µL water. This buffer can be stored at -20 °C.

5X TGIRT-III buffer

Mix reagents to prepare the 5X TGIRT-III buffer with final composition 250 mM Tris-HCl (pH 8.3), 375 mM KCl, and 15 mM $MgCl_2$. This buffer can be stored at room temperature or -20 °C.

Procedure

Total RNA isolation ● Timing 3 h

1. Isolate total RNA from mammalian, yeast, bacterial, or archaeal cells (using at least 1×10^6 cells) using TRIzol reagent according to manufacturer protocol[45,46]. Pellet the cells by centrifugation, and discard supernatant before proceeding to the TRIzol treatment. Alternatively, hot acid phenol can be used to isolate RNA from yeast[47].

   A. Mammalian cells

      i. Cells can be grown in monolayer in culture dishes or flasks to 80% confluency (*e.g.* 8-10 $\times$ 10^6 cells in 10 cm^2 dish). Remove growth media, add 2-4 mL of cold PBS and harvest by scraping with a cell scraper. Pellet cells by centrifugation (400 $\times$ g, 4 °C, 4 min). Alternatively, TRIzol can be directly added the culture dish after removing the media to lyse the cells.

      ii. Cells can be grown in suspension to get 5-10 $\times$ 10^6 cells. Remove media and pellet cells by centrifugation (400 $\times$ g, 4 °C, 4 min).

   B. Bacteria, yeast, and archeal cells

      i. Cells can be grown in suspension to reach OD600 of ~0.6-0.8 for bacteria/archea or OD600 of 1 for yeast. Pellet cells by centrifugation (5000 $\times$ g, 4 °C, 15 min)

2. Next, add TRIzol reagent (approximately 1 mL TRIzol per 5-10 $\times$ 10^6 mammalian or yeast cells, or 1 $\times$ 10^7 bacterial or archeal cells grown in suspension) directly to the cell pellet and ensure complete lysis by vortexing the suspension.
3. After cell lysis, incubate the samples for 5 min at room temperature.
4. Add 0.2 mL chloroform per 1 mL of TRIzol used and vortex for 15 s.
5. Incubate at room temperature for 3 min, and centrifuge at 12,000 $\times$ g for 15 min, at 4 °C.
6. Transfer the colorless upper aqueous layer to a new tube.

▲ **CRITICAL STEP** Aqueous phase contains the RNA. Avoid transferring the bottom or the interface layer, which contain DNA.

7. To precipitate the RNA, add an equal volume of isopropyl alcohol to the aqueous layer. Mix by vortexing for 15 s, incubate at room temperature for 10 min, and centrifuge (12,000 $\times$ g, 4 °C, 10 min).
8. Discard supernatant and vacuum or air dry the pellet for up to 10 min. Take care to avoid over-drying RNA pellets.

9. Resuspend the pellet in 100 µL of nuclease-free water and quantify using the NanoDrop 2000 Spectrophotometer.

10. Check RNA size and quality by Agilent 4200 TapeStation system, by Agilent 2100 bioanalyzer or by running an aliquot on a 1% (w/v) agarose gel electrophoresis with 1× TBE buffer (stained with ethidium bromide)

? **TROUBLESHOOTING**

■ **PAUSE POINT** Isolated RNA can be stored at −80 °C for up to 1 year.

Chemical treatment of RNA ● **Timing 6 h**

▲ **CRITICAL** RNA from **step 10** is used for reductive treatment with $NaCNBH_3$ as well as for 2 controls. We recommend starting with 1 µg of RNA per reaction; however, this amount can be as low as 0.2 µg. Synthetic ac4C spike-in RNA mixture can be added to RNA from **step 10** before proceeding to reactions.

▲ **CRITICAL** Set up three parallel reactions; A) $NaCNBH_3$ treated (+$NaCNBH_3$), B) chemical deacetylation followed by $NaCNBH_3$ treatment (+alkali +$NaCNBH_3$), and C) mock-treated control (-$NaCNBH_3$). RNA from **step 10** can be directly used for reactions A and C. For reaction B, RNA from **step 10** should be pre-treated with alkali as described in **steps 11-16**.

11. *Chemical deacetylation*: Incubate 1 µg of RNA in 100 µL of 100 mM sodium bicarbonate pH 9.5 at 60 °C for 1 h.

▲ **CRITICAL STEP** Alkali treatment with sodium bicarbonate promotes substantial conversion (>80%) of ac4C to cytidine while also causing partial degradation of the RNA backbone, as assessed by Bioanalyzer/TapeStation (**Fig. 4**).

12. Adjust the volume of the RNA reaction to 200 µL with nuclease-free H_2O.

13. Add 0.1 volume of 3 M sodium acetate (pH 5.5) and 2.5 volumes of 100% Ethanol to the sample. Mix thoroughly and keep at −20 °C for 1 h. Centrifuge at 18,000 × g at 4 °C for 30 min.

14. Carefully remove the supernatant and wash the pellet with 500 µL of ice-cold 70% (v/v) ethanol. Centrifuge for an additional 10 min at 18,000 × g at 4 °C.

15. Discard supernatant without disturbing the pellet and vacuum or air dry the pellet for up to 10 min. Take care to avoid over-drying RNA pellets.

16. Resuspend the pellet in nuclease-free water such that the RNA concentration will be greater than 0.2 µg/µL and quantify RNA using the NanoDrop 2000 Spectrophotometer.

17. Set up three 1.7 mL Eppendorf tubes per sample.

- For *+NaCNBH$_3$ RNA*: mix 1 µg of RNA (from **step 10**) in 80 µL nuclease-free H_2O and 10 µL of 1 M $NaCNBH_3$
- For *+alkali +NaCNBH$_3$ RNA*: mix 1 µg of alkali-treated RNA (from **step 16**) in 80 µL nuclease-free H_2O and 10 µL of 1 M $NaCNBH_3$
- For *-NaCNBH$_3$ RNA*: use 1 µg of RNA (from **step 10**) in 90 µL nuclease-free H_2O.

▲ **CRITICAL STEP** Prepare a fresh solution of $NaCNBH_3$ immediately before using. For the untreated control $NaCNBH_3$ is replaced with 10 µL of nuclease-free water.

18. Add 10 µL of 1 M HCl to all three tubes, vortex briefly, spin down, and start the timer.

▲ **CRITICAL STEP** HCl will initiate the reducing reactions and hydrogen gas bubbles will be produced as the reaction progresses.

? **TROUBLESHOOTING**

19. Incubate the samples for 20 min at room temperature.

20. Add 30 uL of 1 M Tris-HCl (pH 8.0) to each sample to quench the reaction by neutralizing the pH.

21. Adjust the reaction volume to 200 µL by adding 30 µL of nuclease-free H_2O and purify RNA by ethanol precipitation as described in **steps 13-16**.

■ **PAUSE POINT** Treated RNA can be stored at −80 °C for up to 1 year.

Preparation of cDNA libraries for high-throughput sequencing ● Timing 2 days

▲ **CRITICAL** The procedures described below entail repeated steps of sample washing, meant to clean the sample from the reagents of the current reaction and resuspend it in an appropriate volume for the next step. Cleanup includes binding the reaction mix to pre-cleaned Silane beads, washing the sample on the beads, air-drying the RNA (while avoiding over-drying the pellets) and eluting it from the beads to a new tube. Although the order of these stages is constant, volumes and elution buffers may vary. Thus, we describe the "cleanup" process in detail in **steps 26-32** below, and only mention variations of it in following wash steps. Pre-cleaning of beads (**steps 26-27**), when applicable, can be conducted in parallel to the reaction it is meant to follow, and thus save time for the user. Varying the concentration of ethanol at the binding stage (**step 29**) is utilized to optimize retention of RNA/cDNA while disposing of unwanted oligonucleotide used as reagents in the reaction, as higher concentration of ethanol will enable retention of smaller fragments. Thus, when applicable, it is advised to use either 0.8 volumes of 100% ethanol for the retentions of short fragments below ~200 nt (e.g.,

tRNAs) or 0.6 volumes of ethanol for longer fragments (e.g., rRNA and mRNA). Specific volumes are indicated when relevant.

RNA fragmentation ● Timing 30 min

22. In a 0.2 mL tube, take 50-200 ng of purified RNA from **step 21** and adjust the reaction volume to 18 µL by adding nuclease-free H_2O.
23. Add 2 µL of RNA Fragmentation Reagent.
24. Heat at 70 °C for 2 min in a thermocycler.

▲ **CRITICAL STEP** Do not exceed the duration of incubation at 70 °C as this may result in excessive fragmentation of the RNA.

25. Place sample on ice/cold block and add 2 µL Fragmentation Stop Solution.
26. To pre-clean Silane Beads, take 20 µL of beads into a new tube and place on magnet. Remove supernatant and add 50 µL RLT buffer.
27. Remove supernatant and resuspend beads in 3x the sample volume (66 µL) RLT buffer.
28. To wash the samples, add the 66 µL beads to the fragmentation reaction tube from **step 25**.
29. Add 125 µL 100% ethanol, mix gently by pipetting up and down 10 times and incubate for 1 minute. Place on magnet and remove supernatant.
30. Add 100 µL of 75% (v/v) ethanol, wash by transferring from side to side of the magnet, remove supernatant.
31. Repeat 75% (v/v) ethanol wash in **step 30.** Air dry beads for 1 minute.
32. Resuspend sample in 16.5 µL of nuclease-free H_2O, mix well, place on magnet and elute by transferring 15.5 µL of the sample into a new tube.

DNAse treatment and dephosphorylation ● Timing 45 min

33. To the eluted sample, add the components as outlined in the table below and incubate at 37 °C for 30 min.

Component	Amount (µL)
Sample from **step 32**	15.5

FNK buffer	6
RNAse inhibitor	0.5 (20 U)
Turbo DNAse	1 (2 U)
FastAP	3 (3 U)
T4 PNK	4 (40 U)

34. Conduct cleanup as in **steps 26-32** by pre-washing 20 µL beads, except resuspend beads in 90 µL RLT. To elute, add 7 µL nuclease-free H_2O to beads, mix well, place on magnet and elute by transferring 6 µL of the sample into a new tube.

First adapter ligation to the RNA ● Timing 2 h

35. Add 1 µL of 10 µM RNA adapter 5' - /5Phos/rArG rArUrC rGrGrA rArGrA rGrCrA rCrArC rGrUrC/3ddC/ - 3' to the eluted RNA.
36. Incubate at 70 °C for 2 minutes and place directly on ice or a cold block.
37. Combine the following components as outlined in the table below; incubate at 23 °C for 90 minutes.

Component	Amount (µL)
Sample from **step 36**	7
10X T4 RNA ligase buffer (NEB)	2
DMSO (100%)	1.8

ATP (100 mM)	0.2
PEG 8000 (50%, w/v)	8
RNAse inhibitor	0.3 (12 U)
T4 RNA Ligase 1, high concentration	1.2 (36 U)

38. Conduct cleanup as in **steps 26-32** by pre-washing 15 µL silane beads, except resuspend the beads in 60 µL RLT and add either 64 µL or 48 µL of 100% ethanol (for short or long RNA, respectively). To elute, add 13.5 µL of nuclease-free H_2O to beads, mix well, place on magnet and elute by transferring 12.5 µL of the sample into a new tube.

First strand cDNA synthesis ● Timing 2 h

39. Add 1 µL of the primer 5´- AGA CGT GTG CTC TTC CG - 3´ (10 µM stock, 10 pmoles final concentration). Mix well by pipetting up and down 10 times.
40. Heat at 70 °C for 2 minutes in a thermocycler and transfer directly to a cold block.
41. On ice, combine the following components as outlined in the table below. Incubate at 50 °C for 60 minutes and allow the thermocycler to cool down to 4 °C before transferring to ice/cold block.

Component	Amount (µL)
Sample from **step 40**	13.5

5X TGIRT-III Buffer	4.4
100 mM DTT	1
10 mM dNTP Mix	1
RNAse inhibitor	0.5 (20 U)
TGIRT-III enzyme	0.9 (180 U)

▲ **CRITICAL STEP** It is recommended to pre-incubate TGIRT-III enzyme with the reaction mixture for 20-30 min prior to adding dNTPs to increase the reverse transcription efficiency.

42. Add 3 µL ExoSap-it and incubate at 37 °C in a thermocycler for 12 minutes. This step will digest the oligodT and primers from the sample.
43. To hydrolyze RNA, add 1 µL of 0.5 M EDTA, 2.5 µL 1 M NaOH and incubate at 70 °C for 12 minutes in a thermocycler. Add 2.5 µL 1 M HCl to neutralize NaOH.
44. Conduct cleanup as in **steps 26-32** by pre-washing 12 µL of beads, except resuspend the beads in 90 µL RLT and add either 96 µL or 72 µL of 100% ethanol (for short or long RNA, respectively). To elute, add 6.5 µL of nuclease-free H_2O to beads, mix well, place on magnet and elute by transferring 5.5 µL of the sample into a new tube.

■ **PAUSE POINT** Treated cDNA can be stored at −80 °C for the next day.

Second adapter ligation to the cDNA ● ***Timing 15 min + overnight incubation + 30 min***

45. To the 5.5 µL of cDNA, add 0.5 µL of a 100 µM stock of the DNA adaptor 5'- /5Phos/AG ATC GGA AGA GCG TCG TGT AG/3ddC/ -3'.
46. Incubate at 75 °C for 3 minutes and place on ice or cold block.
47. On ice, combine the following components as outlined in the table below. Incubate at 23 °C overnight.

Component	Amount (µL)
Sample from **step 46**	6
10X T4 RNA ligase buffer (NEB)	2
DMSO (100%)	0.8
ATP (100mM)	0.2
PEG 8000 (50%, w/v)	10
T4 RNA Ligase 1, high concentration	1.6 (48 U)

48. Conduct cleanup as in **steps 26-32** by pre-washing 5 µL of beads, except resuspend the beads in 60 µL RLT and add either 64 µL or 48 µL of 100% ethanol(for short or long RNA, respectively). To elute, add 23 µL of nuclease-free H_2O to beads, mix well, place on magnet and elute by transferring 11.5 µL into each of two tubes.

▲ **CRITICAL STEP** separation of the eluted cDNA sample into two aliquots in 0.2 mL tubes allows the user to conduct the following PCR enrichment step on half of the material, keeping the other half as backup for potential troubleshooting. The backup tube should be stored at -20 °C until needed.

■ **PAUSE POINT** After this step, the protocol can be paused at any point and cDNA kept at −20 °C till further use.

PCR enrichment ● ***Timing 30 min***

49. To the tube with 11.5 µL cDNA intended for PCR, add the following components as outlined in the table below.

Component	Amount (µL)
One aliquot from **step 48**	11.5
25 µM Read 1 PCR primer	0.5
25 µM Read 2 PCR primer (a distinct barcoded primer for each sample)	0.5
2X KAPA HiFi HotStart ReadyMix PCR Kit (KAPA Biosystems)	12.5

50. Set up and run the PCR program as follows:

Stage	Cycle number	Denature	Anneal	Extend
1	1	95 °C, 2 min		
2	2-5	98 °C, 20 s	65 °C, 15 s	72 °C, 15 s
3	6-10	98 °C, 15 s		72 °C, 15 s
4	11			72 °C, 2 min

▲ **CRITICAL STEP** Although the number of cycles in stage 3 of the PCR reaction can vary from 5-10, it is recommended to keep it to a minimum, in order to preserve the complexity of the resulting cDNA library.

? **TROUBLESHOOTING**

Quality control of cDNA libraries and preparation for high-throughput sequencing ● ***Timing 1-1.5 h***

51. Take 5 µL of the cDNA library into a new tube, add 15 µL nuclease-free H_2O. Keep the remaining 20 µL on ice until used in **steps 54-62**.
52. In a new tube, mix 3 µL of a 100 bp ladder and 17 µL nuclease-free H_2O.
53. run the cDNA library samples alongside the ladder sample on an agarose gel (Invitrogen E-Gel™ EX Agarose Gels, 2%, G402002 can be used) to check the average size of the libraries, which should appear as 200-300 bp smears (**Fig. 4c**).

? **TROUBLESHOOTING**

54. Ampure *cleanup*: To the remaining 20 µL of cDNA library add 20 µL nuclease-free H_2O.
55. Add 40 µL Ampure beads, mix well by pipetting 10 times, and let stand at room temperature for 2 minutes.
56. Place on magnet until all beads have collected and remove supernatant.
57. Add 100 µL 75% (vol/vol) ethanol and move back and forth on the magnet 15-20 times, until the beads are no longer clumpy.
58. Remove supernatant carefully and repeat **step 57.**
59. Remove supernatant and dry at room temperature on magnet until beads appear dry (2-3 minutes).
60. Resuspend sample in 40 µL of nuclease-free H_2O, mix well, place on magnet and transfer supernatant to a new tube.
61. Repeat **steps 55-59.**
62. Resuspend sample in 13 µL of nuclease-free H_2O, mix well, place on magnet and transfer supernatant to a new tube. This is the final cleaned cDNA library to be sequenced.
63. Measure library concentration by a Qubit fluorometer, using the Qubit® dsDNA HS Assay Kit. Store the library at -20 °C (expected concentration should be >2 ng/µl).
64. OPTIONAL: Perform additional quality control of the libraries using an Agilent 2100 Bioanalyzer or High Sensitivity D1000 ScreenTape on the Tapesaation system to check the size of the libraries (**Fig. 4d**).
65. Calculate final dilution of each library for Illumina sequencing: Given the concentration and average length of each resulting library (i.e., sample), calculate the concentration in

nM as follows and add H_2O to dilute to the desired concentration according to the Illumina sequencing protocol used.

$$concentration\ in\ nM = \frac{(concentration\ in\ ng/\mu L)}{\left(650\frac{g}{mol} * average\ library\ length\ in\ bp\right)} * 10\text{^}6$$

▲ **CRITICAL STEP** Final dilution of the cDNA library is based on the average size of the library (estimated from **step 53** or **step 64**) and its concentration (taken from **step 63**). The amount of material taken from each library depends on experimental needs and the sequencing protocol used. Thus, usually only an aliquot of the final library should be diluted, while the rest can be stored at -20 °C.

Illumina sequencing ● Timing 12 h

66. Sequence the generated libraries with the Illumina NextSeq 500 or NovaSeq 6000 platforms using paired-end reads, ranging from 25 to 55 bp from each end, according to manufacturer's instructions.

Data Analysis ● Timing 1+ days

CRITICAL Misincorporation-based mapping of modifications is subject to diverse sources of artifacts, ranging from ones introduced during the reverse transcription step, or in the library preparation, or ones due to the analysis including mismapping of reads, misannotation of genomic regions, or misinterpretation of sources of misincorporations (e.g. due to presence of a SNP). Thus, it is advised to incorporate computational considerations of misincorporation analysis, the rationale and implementation of which are described in Sas-Chen and Schwartz 2019[34]. These include performing local alignment, filtering out sites showing positional or directional biases, and filtering out known SNPs, nuclear mitochondrial DNA (NUMTs) and sites with multiple annotations (e.g., annotated as part of a tRNA and a mRNA).

Read mapping

67. Map paired-end reads to the reference genome relevant for the examined species, using the STAR aligner (version 2.5.3a)[36], and output a BAM file sorted by coordinates via Samtools[42].

Detection of single-nucleotide variants

68. Use the bam files from **step 67** as input to the JACUSA software to detect single nucleotide variant positions[37]. JACUSA should be applied in pileup mode and minimal coverage can be limited to reduce computational load (e.g., to detect only positions with coverage >5), to output a tabular format summarizing the abundance of each nucleotide at each position.

Calculation of misincorporation rate and statistical testing

69. Based on the output of JACUSA, calculate the misincorporation rate of each position by dividing the number of non-cytidine containing reads by the total number of reads.
70. A putative ac4C site is defined by having a statistically significant higher misincorporation rate in the $NaCNBH_3$-treated sample compared to the control samples (mock-treated, chemically deacetylated and/or a genetic control), based on the results of binomial testing. Perform testing as described in Options A and B, depending on how many replicates per condition are available.
    A. Multiple replicates per condition

    If multiple replicates for each condition are available, calculate the P-value of a generalized linear mixed-effects model (GLMM), by using the “glmer” function from the “lme4” package in R[44]. For example in an experiment with two treatment and two control samples with 5 reads in each, the number of cytidine and non-cytidines in each sample is as follows:

sample	cytidine	Non-cytidine
Treatment_1	4	1
Treatment_2	3	2
Control_1	1	4
Control_2	2	3

In R, create a dataframe (myData) in which each row represents a single site from the experiment, as in the table below, and conduct the GLMM test as follows:

GLMM.test = glmer(isCytidine ~ condition + (1|sample), data=myData, family="binomial")

To obtain the P-value apply: **GLMM.pvalue** = coef(summary(GLMM.test))[,'Pr(>|z|)'][2]

condition	isCytidine	sample
treatment	0	Treatment_1
Treatment	1	Treatment_1
Treatment	1	Treatment_1
Treatment	1	Treatment_1
treatment	1	Treatment_1
Treatment	0	Treatment_2
Treatment	0	Treatment_2
Treatment	1	Treatment_2
Treatment	1	Treatment_2
Treatment	1	Treatment_2

Control	0	Control_1
Control	0	Control_1
Control	0	Control_1
Control	0	Control_1
Control	1	Control_1
Control	0	Control_2
Control	0	Control_2
Control	0	Control_2
Control	1	Control_2
Control	1	Control_2

B. Single replicate per condition

If a single replicate is available for each condition, calculate the P-value of a $\chi 2$ test, by using the "chisq.test" function from the "stats" package in R[43],. For example, in an experiment that has only one replicate per condition (in treatment sample 20 cytidines and 80 non-cytidines, and in the control sample 95 cytidines and 5 non-cytidines), create a matrix (myMatrix) as in the table below, and conduct the chisq.test as follows:

CHISQ.pvalue = chisq.test(myMatrix)$p.value

20	80
95	5

▲ **CRITICAL STEP** In order to reduce computational load, statistical tests (detailed in **step 70**) can be calculated on a subset of sites fulfilling conditions specified in **step 71**. Note that the thresholds specified in **step 71** are dependent on the experimental design, namely type of controls and type of organism.

71. For each position, the custom code available at https://github.com/SchwartzLab/ac4c-seq registers the following parameters, and conducts statistical tests only for positions fulfilling all five conditions:
    - The base which cytidine is most frequently converted to is a thymidine.
    - Number of reads displaying a C>T conversion is greater than MINC2T (MINC2T=3 in the original study; can be defined by user when running the code).
    - Misincorporation rate in the $NaCNBH_3$-treated sample is greater than MINmisTreat (MINmisTreat = 2-3% in the original study; can be defined by user when running the code).
    - Misincorporation rate in control sample(s) is less than MAXmisCont (MAXmisCont = 1-5% in the original study; can be defined by user when running the code). Note that deacetylated controls and/or partial knock-down of the acetyltransferase may result in higher misincorporation rate compared to mock, and thus this threshold might need to be relaxed.
    - The difference in misincorporation rate of the $NaCNBH_3$-treated versus control sample is greater than MINdiff (MINdiff = 2% in the original study; can be defined by user when running the code).

Additional filtering

72. When applicable, flag sites overlapping with known single-nucleotide polymorphisms (SNPs), rRNA and tRNA repeats, repetitive regions (such as Alu repeats in primates), mitochondrial pseudogenes that have “transferred” into the nuclear DNA (aka NUMTs: nuclear mitochondrial DNA) and paralog genes derived from a common ancestor gene through duplication events.

Motif search

73. For each putative ac4C site, the custom code extracts the 10 bases upstream and downstream of the site and stores the 21-base sequence in a column named “surrSeq” within the table saved as “ac4c_significantSites.txt”. Extract relevant sequences from this column for motif analysis. Note: sequences are extracted based on the user-provided .fasta file. Make sure to use a relevant reference of the genome or transcriptome, occurding to experimental design (e.g., in case of assessing sites in the coding sequence of intron-containing mRNA, make sure to use a file containing sequences of the mature mRNA).
74. Use the 21 base long sequences as input for multiple sequence alignment via the WebLogo software (https://weblogo.berkeley.edu/logo.cgi).

▲ **CRITICAL STEP** Experimental results indicate that a CCG motif is necessary for catalysis by the acetyltransferase[5]. Thus, the presence of a CCG motif in the putative ac4C sites can serve as a quality control of the validity of the sites. The code stores the 3-base motif under the column “motif” within the table saved as “ac4c_significantSites.txt”.

Timing

Steps 1-21 (day 1-2), Isolation of total RNA and Chemical treatment of RNA: 9 h

Steps 22-65 (days 2-3), Preparation of cDNA libraries for high-throughput sequencing: 2 days

Step 66 (day 4), Illumina sequencing: 12 h

Steps 67-74 (day 5), data analysis: minimum 1 day

Troubleshooting

Troubleshooting advice can be found in Table 2.

Table 2. Troubleshooting table.

Step	Problem	Possible reason	Solution
10	Poor RNA quality and low yield	RNA degradation	Use RNase-free materials and equipment when isolating total RNA.
18	No bubbles are formed after addition of HCl to the $NaCNBH_3$-treated samples	Inactive $NaCNBH_3$, perhaps due to exposure to humidity in the air. (if the powder absorbs water from the air it is less active and its color turns from bright white to pale white)	Prepare a fresh aliquot of $NaCNBH_3$. Make sure the stock powder is properly desiccated.

50 and 53	No library is seen in the E-gel	Number of PCR cycles in **step 50** was insufficient (NOTE: this may indicate low complexity of the library, and thus should be taken into account in downstream analysis).	Place the 20 ul of sample from **step 51** back to the thermocycler and conduct 2-5 additional PCR cycles (stage 3 of **step 50**) followed by a final extension (stage 4 of **step 50**). Re-run the library on an E-gel. If the library now shows the correct 200-300 bp smear, re-do the PCR reaction on the 11.5 µL left from **step 48**, using the newly optimized number of cycles. Continue with the PCR product to **step 51**.

		No library was generated.	If no product is seen after increasing the number of PCR cycles, verify the quality of RNA and materials used and re-do library preparation (**steps 22-53**).
53	The library contains adaptor dimers (~140-bp peak) or excess primers (70–80 bp)	May result from insufficient removal of adaptors and primers from prior steps.	Conduct Ampure cleanup as in **step 54-62**, and verify that unwanted bands are removed.
		May indicate low complexity of the generated library, and thus may affect downstream quality of sequencing output.	Presence of adaptor sequences can be detected in the .fastq files generated by the Illumina platform.

Anticipated Results

RNA isolation throughout the protocol

After RNA isolation using TRIzol (**step 10**), >100 µg of RNA is recovered, and RNA size distribution should resemble that shown in **Fig. 4a**, indicating that rRNA, mRNA, and short RNAs were purified with no notable degradation. Ethanol precipitation of RNA after chemical treatments (**step 21**) results in recovery of >70% RNA.

ac4C-Seq library construction

Distribution of RNA fragments size between reduced, deacetylated, and mock-treated samples varies, with deacetylated samples usually exhibiting smaller RNA fragments (**Fig. 4b**). A library prepared using ~200 ng of total RNA results in >25 ng of amplified cDNA at concentration of >2 ng/ul (**step 63**). The size of the libraries is 200-300 bp, with the deacetylated control samples usually resulting in smaller cDNA fragments compared to reduced and mock treated samples (**Fig. 4c, d**).

Read mapping and misincorporation calculations

The mapping results in BAM files, sorted and indexed via Samtools, which can be used for visual inspection and quality control of the data via the IGV software[48]. While ac4C sites in reduced samples should exhibit a mixture of nucleotides (predominantly cytidines and thymidines), control samples will show negligible amounts of non-cytidine incorporations (**Fig. 5a**). Note that deacetylated controls and partial knock-down of the acetyltransferase may result in some non-cytidine containing reads, but their percentage should be lower compared to the reduced sample. Furthermore, misincorporations can be seen for additional modifications, however these signatures are not expected to vary between experimental conditions. Known sites can be used as positive controls during visual inspection of the data via the IGV software, and when generating the catalog of statistically significant ac4C sites (**Fig. 5b, c**).

ac4C sites and motif identification

As reduction by hydride donors leads to C>T conversions at ac4C sites, it is expected that putative ac4C sites that passed statistical testing will be enriched in C>T conversions compared to other conversion types. In contrast control samples are expected to show no enrichment for a specific conversion type. This can be assessed by plotting the number of sites displaying each conversion type in reduced and control samples, as in **Fig. 5d**.

Furthermore, we previously identified that ac4C occurs within a CCG motif, which is necessary for modification of the middle cytidine. Thus, individual putative ac4C sites, as well as the signature identified via WebLogo (**step 74**) should contain a CCG motif (**Fig. 5e**).

Few examples of putative ac4C sites identified using ac4C-seq[5] are listed in Table 3.

species	RNA type	Number of identified ac4C sites
Yeast (*S. cerevisiae*)	18S rRNA	2 (h34 and h45)

	tRNA	2 (tRNA-Ser and tRNA-Leu)
Mammalian (HeLa)	18S rRNA	2 (h34 and h45)
	tRNA	2 (tRNA-Ser and tRNA-Leu)
Archea (T. kodakarensis grown at 85 °C)	rRNA	173
	tRNA	77
	mRNA	119
	Rnase P	19
	SRP	16

Table 3: Identified putative ac4C sites

Author Contributions

A.S., S.G., J.M., and S.S. developed the protocol. A.S. and S.G. performed the experiments. A.S. designed the bioinformatic pipeline and analyzed the data. S.G. A.S., SS, and J.M. wrote the manuscript.

Acknowledgments

We thank members of the Schwartz and Meier laboratories for many helpful comments. S. Schwartz is funded by the Israel Science Foundation (543165), the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement no. 714023), and the Estate of Emile Mimran. S. Schwartz is the incumbent of the Robert Edward and Roselyn Rich Manson Career Development Chair in Perpetuity. J.L.M. is supported by the Intramural Research Program of the National Institutes of Health (NIH), the National Cancer Institute, The Center for Cancer Research (ZIA BC011488-06).

Competing Interests

The authors declare no competing interests.

~~**Data availability**~~

Data and Code availability

The custom code used for the "Calculation of misincorporation level and statistical testing" section is available at: https://github.com/SchwartzLab/ac4c-seq. Results depicted in **Figure 5** are based on ac^{4}C-seq data previously deposited in the Gene Expression Omnibus (GEO) under accession number GSE135826 as part of the publication by Sas-Chen et al 2020[5].

References

1. Zachau, H. G., Dütting, D. & Feldmann, H. Nucleotide Sequences of two Serine-Specific Transfer Ribonucleic Acids. *Angew. Chem. Int. Ed.* **5,** 422–422 (1966).
2. Ohashi, Z. *et al.* Characterization of C+ located in the first position of the anticodon of Escherichia coli tRNAMet as N4-acetylcytidine. *Biochim. Biophys. Acta.* **262,** 209–213 (1972).
3. Bruenger, E. *et al.* 5S rRNA modification in the hyperthermophilic archaea Sulfolobus solfataricus and Pyrodictium occultum. *FASEB J.* **7,** 196–200 (1993).
4. Stern, L. & Schulman, L.H. The Role of the Minor Base N4-Acetylcytidine in the Function of the Escherichia Coli Noninitiator Methionine Transfer RNA. *J. Biol. Chem.* **253**, 6132–39 (1978).
5. Sas-Chen, A. *et al.* Dynamic RNA acetylation revealed by quantitative cross-evolutionary mapping. *Nature.* **583,** 638–643 (2020).
6. Johansson, M. J. & Bystrom, A. S. The Saccharomyces cerevisiae TAN1 gene is required for N4-acetylcytidine formation in tRNA. *RNA.* **10,** 712–719 (2004).
7. Kotelawala, L., Grayhack, E. J. & Phizicky, E. M. Identification of Yeast tRNA Um(44) 2'-O-Methyltransferase (Trm44) and Demonstration of a Trm44 Role in Sustaining Levels of Specific tRNA(Ser) Species. *RNA.* **14**, 158–69 (2008).
8. Ito, S. *et al.* A Single Acetylation of 18 S rRNA Is Essential for Biogenesis of the Small Ribosomal Subunit inSaccharomyces cerevisiae. *J. Biol. Chem.* **289,** 26201–26212 (2014).
9. Ito, S. *et al.* Human NAT10 Is an ATP-dependent RNA Acetyltransferase Responsible forN4-Acetylcytidine Formation in 18 S Ribosomal RNA (rRNA). *J. Biol. Chem.* **289,** 35724–35730 (2014).
10. Sharma, S. *et al.* Yeast Kre33 and Human NAT10 Are Conserved 18S rRNA Cytosine Acetyltransferases That Modify tRNAs Assisted by the Adaptor Tan1/THUMPD1. *Nucleic Acids Res.* **43**, 2242–58 (2015).
11. Sharma, S. *et al.* Specialized box C/D snoRNPs act as antisense guides to target RNA base acetylation. *PLOS Genet.* **13,** (2017).
12. Thomas, J. M., Bryson, K. M. & Meier, J. L. Nucleotide resolution sequencing of N4-acetylcytidine in RNA. *Methods in Enzymology Chemical and Synthetic Biology Approaches To Understand Cellular Functions - Part A* 31–51 (2019). doi:10.1016/bs.mie.2019.02.022

13. Wang, T. *et al.* Identification and characterization of essential genes in the human genome. *Science.* **350,** 1096–1101 (2015).
14. Dempster, J. M. *et al.* Extracting Biological Insights from the Project Achilles Genome-Scale CRISPR Screens in Cancer Cell Lines. Preprint at https://doi.org/10.1101/720243 (2019).
15. Larrieu, D., Britton, S., Demir, M., Rodriguez, R. & Jackson, S. P. Chemical Inhibition of NAT10 Corrects Defects of Laminopathic Cells. *Science.* **344,** 527–532 (2014).
16. Tschida, B. R. *et al.* Sleeping BeautyInsertional Mutagenesis in Mice Identifies Drivers of Steatosis-Associated Hepatic Tumors. *Cancer Res.* **77,** 6576–6588 (2017).
17. Ryvkin, P. *et al.* HAMR: high-throughput annotation of modified ribonucleotides. *RNA.* **19,** 1684–1692 (2013).
18. Cerutti, P., & Miller, N. Selective Reduction of Yeast Transfer Ribonucleic Acid with Sodium Borohydride. *J. Mol. Biol*. **67**, 90260-4 (1967).
19. Thomas, J. M. *et al.* A Chemical Signature for Cytidine Acetylation in RNA. *J. Am. Chem. Soc.* **140,** 12667–12670 (2018).
20. Sinclair, W. R. *et al.* Profiling Cytidine Acetylation with Specific Affinity and Reactivity. *ACS Chem. Biol.* **12,**2922–2926 (2017).
21. Tsai, K. *et al.* Acetylation of Cytidine Residues Boosts HIV-1 Gene Expression by Increasing Viral RNA Stability. *Cell Host Microbe.* **28** (2020).
22. Balmus, G. et al. Targeting of NAT10 Enhances Healthspan in a Mouse Model of Human Accelerated Aging Syndrome. *Nat. Commun.* **9**, 1700 (2018).
23. Li, Q. *et al.* NAT10 is upregulated in hepatocellular carcinoma and enhances mutant p53 activity. *BMC Cancer.* **17** (2017).
24. Liu, X. *et al.* Deacetylation of NAT10 by Sirt1 promotes the transition from rRNA biogenesis to autophagy upon energy stress. *Nucleic Acids Res.* **46,** 9601–9616 (2018).
25. Lin, S. *et al.* Mettl1/Wdr4-Mediated mG tRNA Methylome Is Required for Normal mRNA Translation and Embryonic Stem Cell Self-Renewal and Differentiation. *Mol. Cell* **71**, 244–55.e5 (2018).
26. Marchand, V. *et al.* AlkAniline-Seq: Profiling of M G and M C RNA Modifications at Single Nucleotide Resolution. *Angew. Chem.* **57**, 16785–90 (2018).
27. Ikeuchi, Y., Kitahara, K. & Suzuki, T. The RNA acetyltransferase driven by ATP hydrolysis synthesizes N4-acetylcytidine of tRNA anticodon. *EMBO J.* **27,** 2194–2203 (2008).

28. Arango, D. *et al.* Acetylation of Cytidine in mRNA Promotes Translation Efficiency. *Cell* **175** (2018).
29. Safra, M. *et al.* The m1A landscape on cytosolic and mitochondrial mRNA at single-base resolution. *Nature* **551,** 251–255 (2017).
30. Grozhik, A. V. *et al.* Antibody cross-reactivity accounts for widespread appearance of m1A in 5'UTRs. *Nat. Commun.* **10,** 5126 (2019).
31. Helm, M., Lyko, F. & Motorin, Y. Limited antibody specificity compromises epitranscriptomic analyses. *Nat. Commun.* **10,** 5669 (2019).
32. Liu, H. *et al.* Accurate detection of m6A RNA modifications in native RNA sequences. *Nat. Commun.* **10,** 4079 (2019).
33. Grünberger, F. *et al.* Exploring prokaryotic transcription, operon structures, rRNA maturation and modifications using Nanopore-based native RNA sequencing. Preprint at https://doi.org/10.1101/2019.12.18.880849 (2019)
34. Sas-Chen, A. & Schwartz, S. Misincorporation signatures for detecting modifications in mRNA: Not as simple as it sounds. *Methods* **156,** 53–59 (2019).
35. Orita, I. *et al.* Random mutagenesis of a hyperthermophilic archaeon identified tRNA modifications associated with cellular hyperthermotolerance. *Nucleic Acids Res.* **47,** 1964–1976 (2019).
36. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *J. Bioinform.* **29**, 15–21 (2013).
37. Piechotta, M., Wyler, E., Ohler, U. *et al.* JACUSA: site-specific identification of RNA editing events from replicate sequencing data. *BMC Bioinformatics* **18,** 7 (2017).
38. Sexton, A. N., Wang, P. Y., Rutenberg-Schoenberg, M. & Simon, M. D. Interpreting Reverse Transcriptase Termination and Mutation Events for Greater Insight into the Chemical Probing of RNA. *Biochemistry* **56,** 4713–4721 (2017).
39. Cozen, A. E. *et al.* ARM-Seq: AlkB-Facilitated RNA Methylation Sequencing Reveals a Complex Landscape of Modified tRNA Fragments. *Nat. Methods* **12**, 879–84 (2015).
40. Zheng, G. *et al.* Efficient and quantitative high-throughput tRNA sequencing. *Nat. Methods* **12,** 835–837 (2015).
41. Quinodoz, S. A. *et al.* Higher-Order Inter-chromosomal Hubs Shape 3D Genome Organization in the Nucleus. *Cell*, ***174*(3),** 744–757 (2018).
42. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *J. Bioinform.* **25,** 2078–2079 (2009).

43. R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.
44. Bates, D., Mächler, M., Bolker, B., Walker, S. Fitting Linear Mixed-Effects Models Using lme4. J. Stat. Softw. **67**, 1–48 (2015).
45. Chomczynski, P. A reagent for the single-step simultaneous isolation of RNA, DNA and proteins from cell and tissue samples. *Biotechniques*. **15(3)**,532-4(1993)
46. Hummon, A. B., Lim S. R., Difilippantonio, M. J., & Ried, T. Isolation and solubilization of proteins after TRIzol® extraction of RNA and DNA from patient material following prolonged storage. *Biotechniques* **42**, 467-472 (2007)
47. Collart, M. A. & Oliviero, S. Preparation of yeast RNA. *Curr. Protoc. Mol. Biol.* **Chapter 13**, Unit13 12, (2001).
48. Robinson, J. *et al.* Integrative genomics viewer. *Nat. Biotechnol*. **29,** 24–26 (2011)

Figure legends

Figure 1. Overview of nucleobase reaction chemistry underlying N4-acetylcytidine sequencing (ac4C-seq). Endogenous N4-acetylcytidine (ac4C) forms base pairs identical to cytidine, rendering it silent in reverse transcription (RT) and complementary DNA (cDNA) sequencing experiments (top left). Treatment of ac4C-containing with $NaCNBH_3$ under acidic conditions results in formation of a reduced nucleobase, which is read as a 'U' during RT and results in misincorporations which can be detected by cDNA sequencing (bottom right). This chemistry is blocked when ac4C is chemically deacetylated by alkali hydrolysis (top right), which provides a control for specific detection of the acetylated nucleobase.

Figure 2. Experimental workflow for ac4c-seq protocol. RNA is spiked with a synthetic ac4C RNA spike-in and split into three samples. One experimental sample is treated with $NaCNBH_3$ under acidic conditions (reduction, positive signal), while two control samples are subjected to acidic conditions without reducing agent (mock-treated, control #1) and deacetylation followed by $NaCNBH_3$ treatment (deacetylated + reduction, control #2). A 3' adapter is ligated onto fragmented RNA, which is then reverse transcribed, resulting in misincorporation of 'A' in cDNA at positions of reduced ac4C. A 3′ adaptor is ligated to facilitate cDNA library construction, with

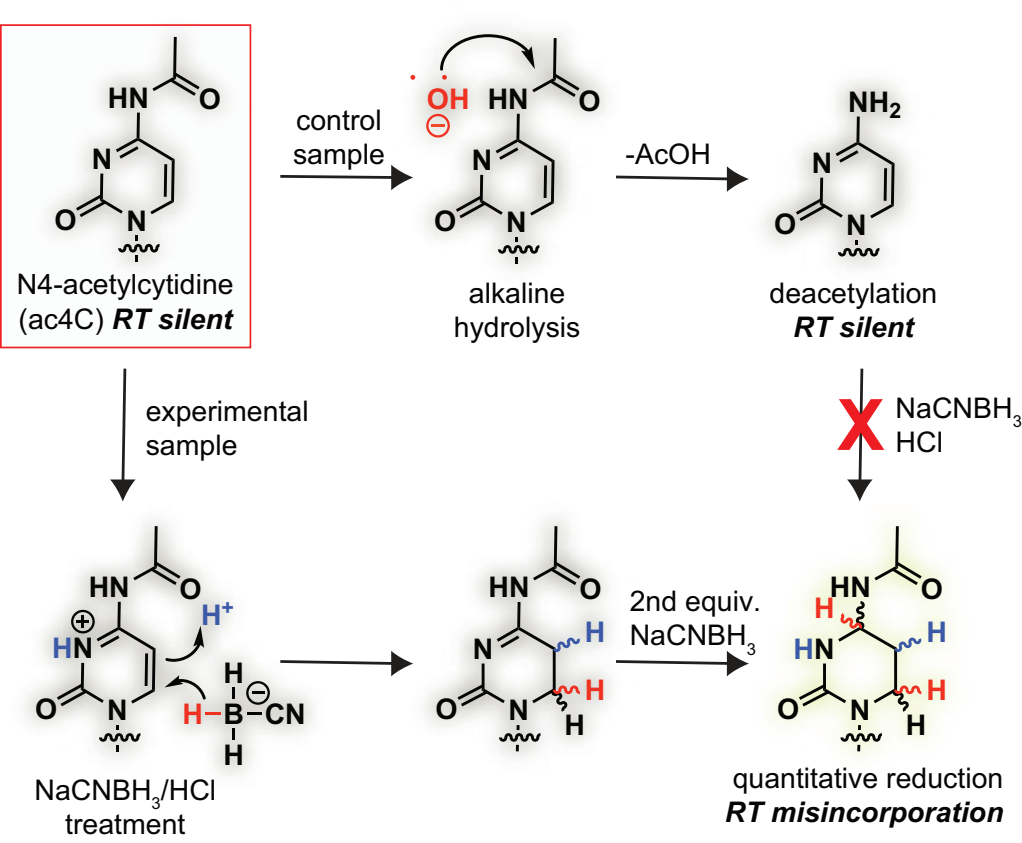
subsequent sequencing and bioinformatic analysis being used to identify ac4C-modified sites. Figure is adapted with permission from Sas-Chen, A. et al. Dynamic RNA Acetylation Revealed by Quantitative Cross-Evolutionary Mapping. Nature, **583**, 638–643 (2020). Copyright Springer Nature, 2020[5].

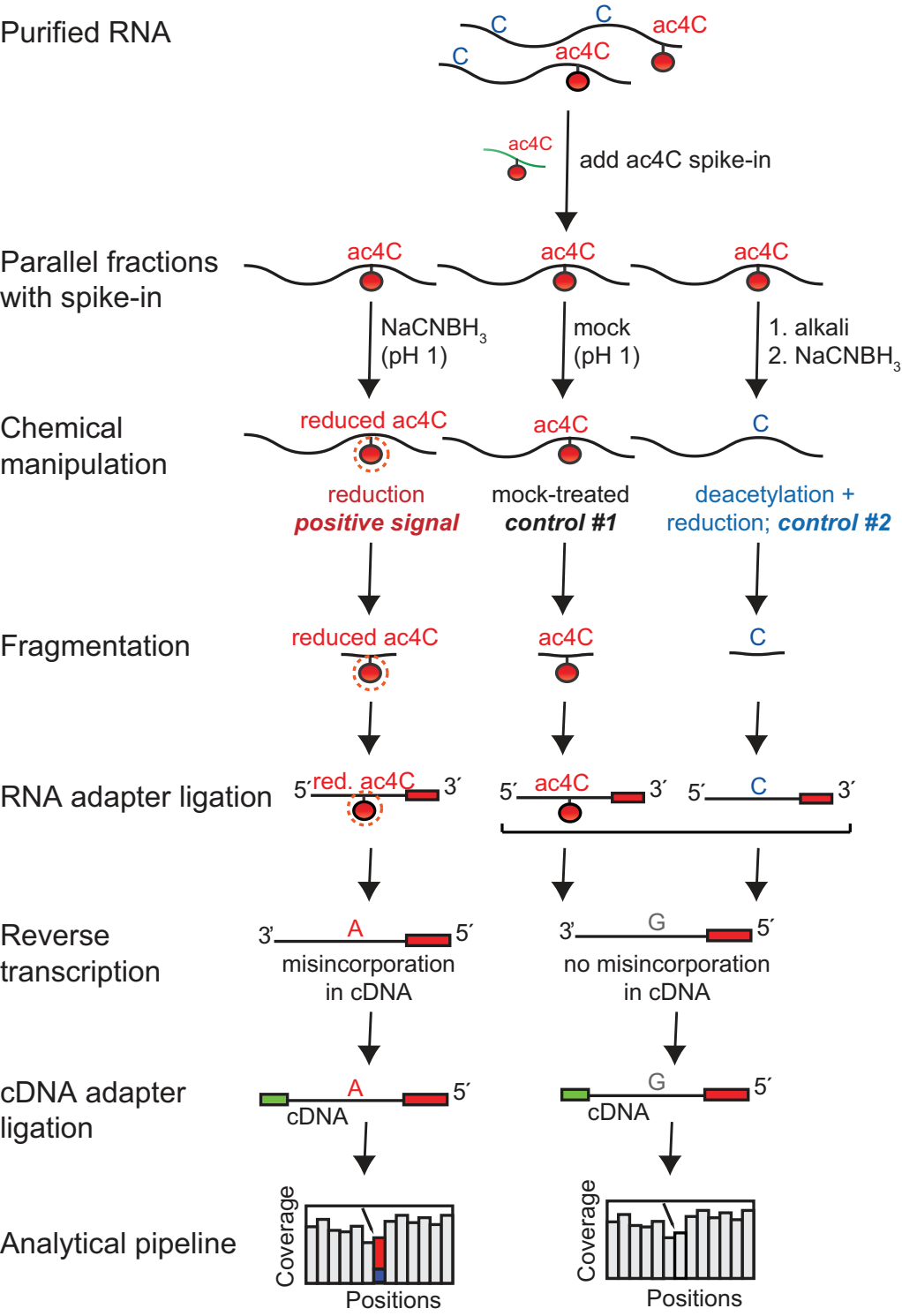
Figure 3. Flowchart illustrating steps involved in ac4C-seq. (**a**) First RNA is isolated, followed by chemical treatment to prepare experimental and control samples. These RNA samples are used to prepare cDNA libraries which undergo next-generation sequencing and are subsequently analyzed using computational pipelines. (**b**) Data analysis includes mapping of sequenced reads to the appropriate genome, calculating misincorporation rate in treatment and control samples, and conducting statistical tests to identify ac4C sites. Following additional filtering, motif analysis is applied to strengthen confidence in newly identified candidate ac4C-modified sites.

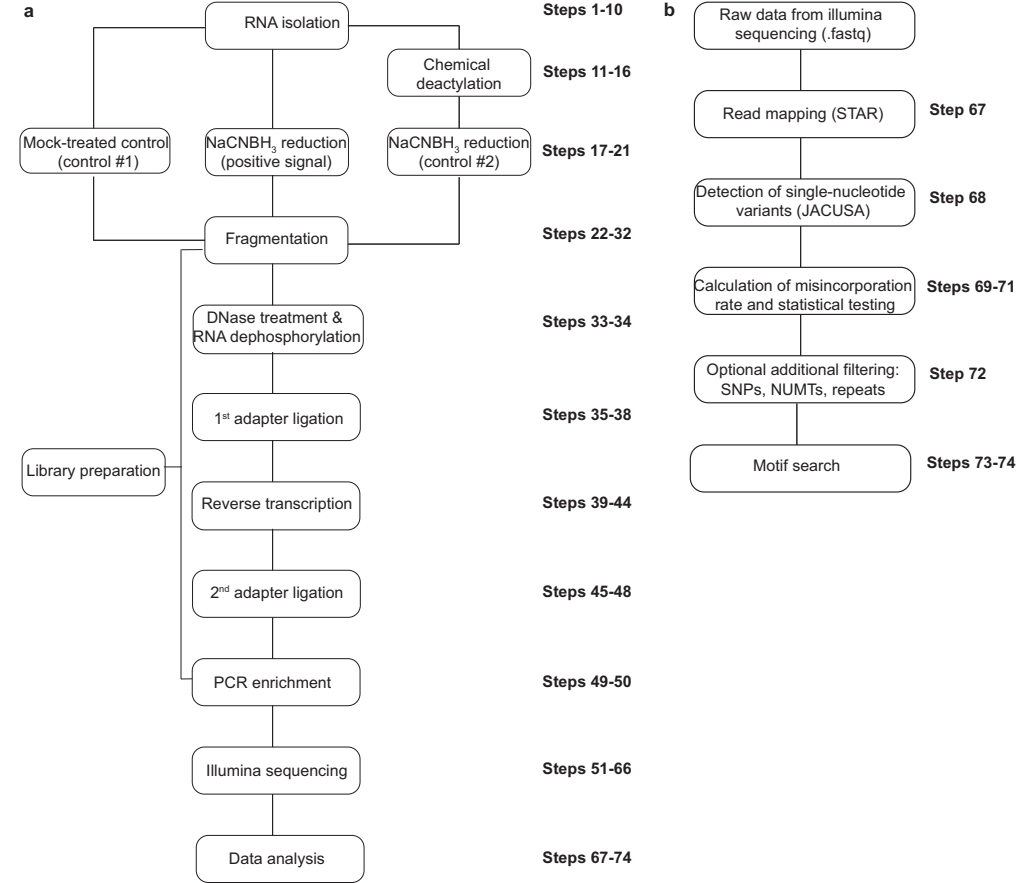
Figure 4. Quality control of RNA and cDNA libraries. (**a-b**) typical tapestation generated trace of RNA from human HeLa cells after isolation (a, **step 10**) or after treatment with $NaCNBH_3$, alkali deacetylation followed by $NaCNBH_3$ or mock (b, **step 21**). (**c-d**) typical cDNA libraries generated from RNA of T. kodakarensis, analyzed using E-gel (c, **step 53**) or tapestation (d, **step 64**). Dashed lines indicate the mean fragment size of each library. Bands in the dashed rectangle represent leftover primers from the PCR reaction (**step 50**), which are eliminated upon the final washes on Ampure beads (**step 61**). R-reduced, D-deacetylated, M-mock, FU-fluorescence units.

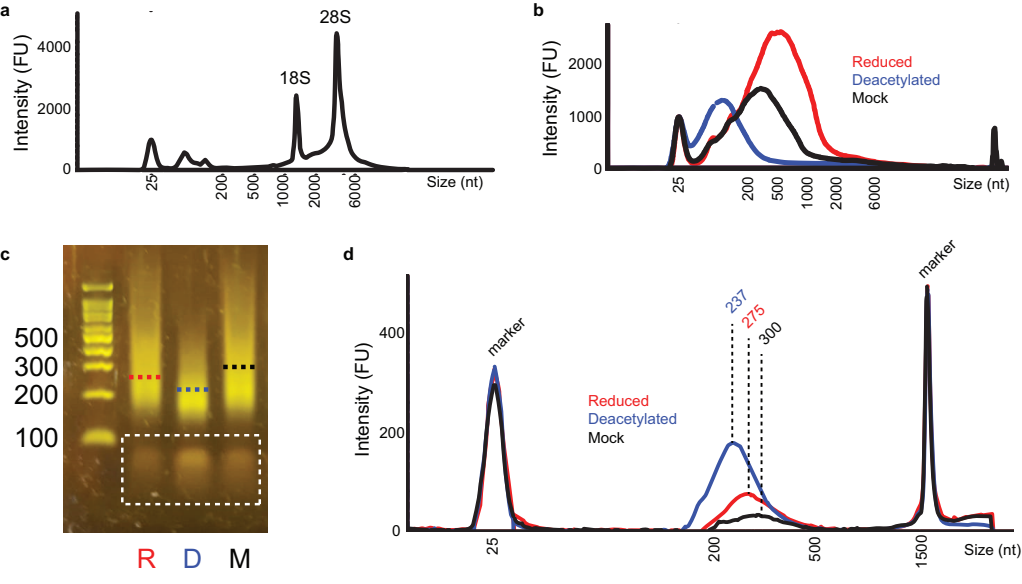
Figure 5. Anticipated results of ac4c-seq. (**a**) Coverage tracks of regions within human rRNA 18S (left) and mitochondrial tRNA Methionine (right), as depicted in the IGV browser. Positions with misincorporation rate >1% are shown in color (blue, cytidine; red, thymidine; green, adenosine; grey, misincorporation <1%). While both the ac4C in helix 45 of 18S and the f5C in mt-tRNA-Met are dependent on reduction with $NaCNBH_3$(display a misincorporation upon reduction), only ac4C is sensitive to alkali deacetylation (showing reduced misincorporation in “reduced” versus “deacetylated” conditions). m6,6A in 18S induces a misincorporation in a $NaCNBH_3$ and deacetylation independent manner and is thus similar in all conditions. mt-tRNA-Met, mitochondrial tRNA Methionine; ac4C, N4-acetylcytidine; f5C, 5-formylcytidine; m6,6A,

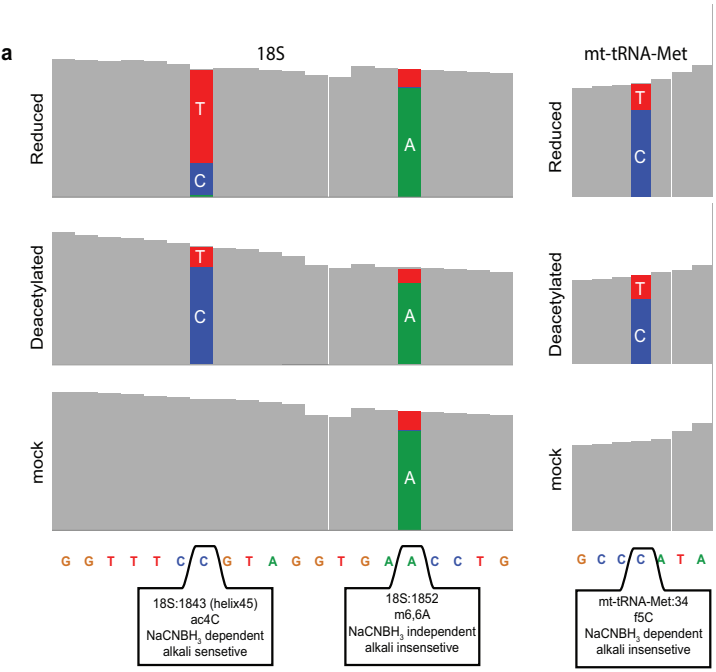
N6,N6-dimethyladenosine. (**b**) Misincorporation rates of known sites in 18S in wild-type and NAT10-depleted HeLa cells (bars, mean of 3 biological samples; error bars, standard deviation). (**c**) Statistical significance plotted against the difference in misincorporation rates between $NaCNBH_3$ and mock-treated total RNA from HeLa cells. Vertical dashed line, 5%; horizontal dashed line, P = 0.05 (χ2 test). n = 3 biological samples. (**d**) Frequency of the 12 possible misincorporation patterns (y axis) found across all statistically significant sites in T. kodakarensis, showing an enrichment of C>T conversions. (**e**) Sequence motif surrounding the ac4C sites identified in human, yeast, and archaea. Figure is adapted with permission from Sas-Chen, A. et al. Dynamic RNA Acetylation Revealed by Quantitative Cross-Evolutionary Mapping. Nature, **583**, 638–643 (2020). Copyright Springer Nature, 2020[5].

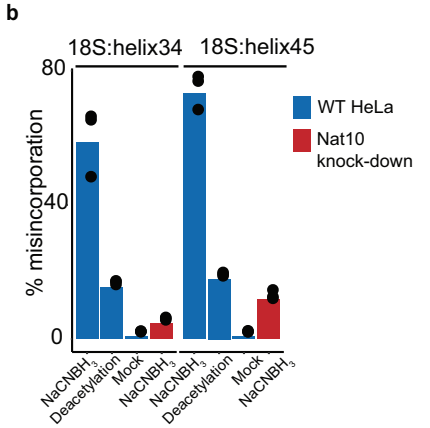

N4-acetylcytidine
(ac4C) RT silent
control
sample
alkaline
hydrolysis
-AcOH
deacetylation
RT silent
experimental
sample
NaCNBH3
HCl
NaCNBH3/HCl
treatment
2nd equiv.
NaCNBH3
quantitative reduction
RT misincorporation

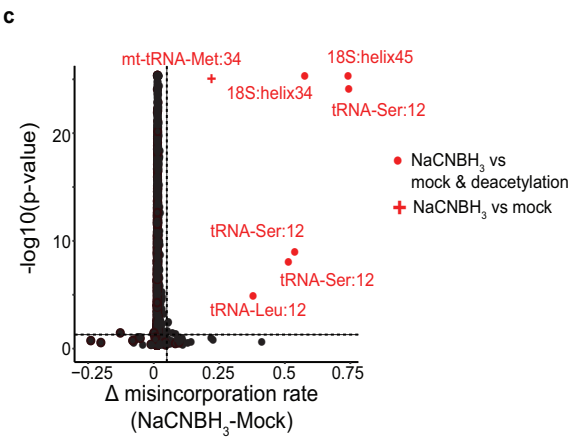

Purified RNA
C
C
ac4C
C
ac4C
ac4C
add ac4C spike-in
Parallel fractions with spike-in
ac4C
ac4C
ac4C
NaCNBH3 (pH 1)
mock (pH 1)
1. alkali
2. NaCNBH3
Chemical manipulation
reduced ac4C
ac4C
C
reduction
positive signal
mock-treated
control #1
deacetylation + reduction; control #2
Fragmentation
reduced ac4C
ac4C
C
RNA adapter ligation
5′ red. ac4C 3′
5′ ac4C 3′
5′ C 3′
Reverse transcription
3′ A 5′
misincorporation in cDNA
3′ G 5′
no misincorporation in cDNA
cDNA adapter ligation
A 5′
cDNA
G 5′
cDNA
Analytical pipeline
Coverage
Positions
Coverage
Positions

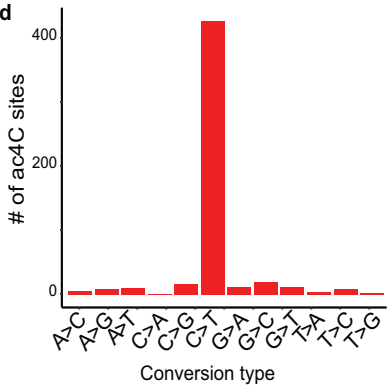

a
RNA isolation
Steps 1-10
Chemical deactylation
Steps 11-16
Mock-treated control (control #1)
NaCNBH3 reduction (positive signal)
NaCNBH3 reduction (control #2)
Steps 17-21
Fragmentation
Steps 22-32
DNase treatment & RNA dephosphorylation
Steps 33-34
1st adapter ligation
Steps 35-38
Library preparation
Reverse transcription
Steps 39-44
2nd adapter ligation
Steps 45-48
PCR enrichment
Steps 49-50
Illumina sequencing
Steps 51-66
Data analysis
Steps 67-74
b
Raw data from illumina sequencing (.fastq)
Read mapping (STAR)
Step 67
Detection of single-nucleotide variants (JACUSA)
Step 68
Calculation of misincorporation rate and statistical testing
Steps 69-71
Optional additional filtering: SNPs, NUMTs, repeats
Step 72
Motif search
Steps 73-74

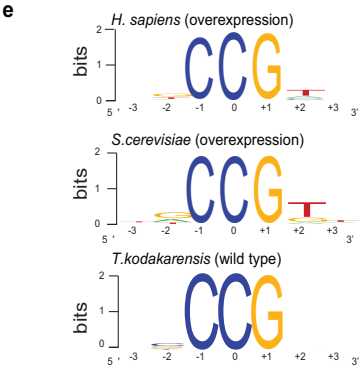

a
Intensity (FU)
4000
2000
0
18S
28S
25
200
500
1000
2000
6000
Size (nt)
b
Intensity (FU)
2000
1000
0
Reduced
Deacetylated
Mock
25
200
500
1000
2000
6000
Size (nt)
c
500
300
200
100
R
D
M
d
Intensity (FU)
400
200
0
marker
237
275
300
marker
Reduced
Deacetylated
Mock
25
200
500
1500
Size (nt)


a
18S
mt-tRNA-Met
Reduced
Deacetylated
mock
T
C
A
G G T T T C C G T A G G T G A A C C T G
G C C C A T A
18S:1843 (helix45)
ac4C
NaCNBH3 dependent
alkali sensetive
18S:1852
m6,6A
NaCNBH3 independent
alkali insensetive
mt-tRNA-Met:34
f5C
NaCNBH3 dependent
alkali insensetive


b
18S:helix34
18S:helix45
% misincorporation
80
40
0
WT HeLa
Nat10
knock-down
NaCNBH3
Deacetylation
Mock
NaCNBH3
NaCNBH3
Deacetylation
Mock
NaCNBH3


c
mt-tRNA-Met:34
18S:helix34
18S:helix45
tRNA-Ser:12
-log10(p-value)
20
10
0
tRNA-Ser:12
tRNA-Ser:12
tRNA-Leu:12
NaCNBH3 vs
mock & deacetylation
NaCNBH3 vs mock
−0.25
0
0.25
0.5
0.75
Δ misincorporation rate
(NaCNBH3-Mock)


d
# of ac4C sites
400
200
0
A>C
A>G
A>T
C>A
C>G
C>T
G>A
G>C
G>T
T>A
T>C
T>G
Conversion type


e
H. sapiens (overexpression)
S.cerevisiae (overexpression)
T.kodakarensis (wild type)
bits
CCG
5'
3'