*Briefings in Bioinformatics*, 22(6), 2021, 1–14

https://doi.org/10.1093/bib/bbab219
Review

# A comparison of metabolic labeling and statistical methods to infer genome-wide dynamics of RNA turnover

Etienne Boileau, Janine Altmüller, Isabel S. Naarmann-de Vries† and Christoph Dieterich†

Corresponding author. Christoph Dieterich, Section of Bioinformatics and Systems Cardiology, Klaus Tschira Institute for Integrative Computational Cardiology, Department of Internal Medicine III, University Hospital Heidelberg Analysezentrum III, Im Neuenheimer Feld, 669 69120 Heidelberg, Germany Tel: +49 6221 56 36884; E-mail: christoph.dieterich@uni-heidelberg.de
†These authors contributed equally to this work.

## Abstract

Metabolic labeling of newly transcribed RNAs coupled with RNA-seq is being increasingly used for genome-wide analysis of RNA dynamics. Methods including standard biochemical enrichment and recent nucleotide conversion protocols each require special experimental and computational treatment. Despite their immediate relevance, these technologies have not yet been assessed and benchmarked, and no data are currently available to advance reproducible research and the development of better inference tools. Here, we present a systematic evaluation and comparison of four RNA labeling protocols: 4sU-tagging biochemical enrichment, including spike-in RNA controls, SLAM-seq, TimeLapse-seq and TUC-seq. All protocols are evaluated based on practical considerations, conversion efficiency and wet lab requirements to handle hazardous substances. We also compute decay rate estimates and confidence intervals for each protocol using two alternative statistical frameworks, pulseR and GRAND-SLAM, for over 11 600 human genes and evaluate the underlying computational workflows for their robustness and ease of use. Overall, we demonstrate a high inter-method reliability across eight use case scenarios. Our results and data will facilitate reproducible research and serve as a resource contributing to a fuller understanding of RNA biology.

**Key words:** metabolic RNA labeling; 4sU-tagging; RNA decay rate estimation; computational workflow; kinetic and statistical modeling

## Introduction

To estimate the kinetics of RNA synthesis or degradation, techniques including transcription shut-off and nascent RNA sequencing have been developed. Although used for decades, global transcriptional arrest methods interfere with the precise determination of decay rates by affecting RNA processing and stability [1] and have thus become less fashionable. Nascent RNA analysis methods include approaches such as run-on [2, 3], Pol II immunoprecipitation [4] and metabolic labeling [5–7]. One of the most recent, minimally invasive and widely applied metabolic labeling approaches, referred to as 4sU-tagging, involves labeling of newly transcribed RNAs with the thiol-labeled nucleoside analog 4-thiouridine (4sU), combined with genome-wide high-throughput sequencing [8–10]; 4sU is taken up by the cell,

**Etienne Boileau** is a postdoctoral research fellow in the Dieterich Lab in the Department of Internal Medicine III at the University Hospital Heidelberg.
**Janine Altmüller** is the Head of NGS core facility in the Cologne Center for Genomics at the University of Cologne and transitions to a new position as Head of Core Facility Genomics in the Berlin Institute of Health at Charité.
**Isabel S. Naarmann-de Vries** is a postdoctoral research fellow in the Dieterich Lab in the Department of Internal Medicine III at the University Hospital Heidelberg, and in the Department of Intensive Care Medicine at the University Hospital Aachen.
**Christoph Dieterich** is a Professor, Head of the Dieterich Lab in the Klaus Tschira Institute for Integrative Computational Cardiology at the University Hospital Heidelberg, and site coordinator of HiGHmed UK-HD.
**Submitted:** 18 February 2021; **Received (in revised form):** 18 May 2021

phosphorylated and incorporated into newly transcribed RNAs, with minimal interference to gene expression [11].

In a conventional RNA turnover experiment, samples are taken from several points over an extended time course in the presence of 4sU, with the aim of computing RNA decay rates or half-lives. Newly transcribed and pre-existing RNAs are separated either by biochemical enrichment after thiol-specific biotinylation [10], or by chemically inducing thymine-to-cytosine (T to C) nucleotide conversions, which are detected as point mutations in the sequencing data, enabling to distinguish the metabolically 'labeled reads' from pre-existing 'unlabeled' reads. Two recent studies published in Nature Methods, and a third in Angewandte Chemie International Edition provided three equivalent approaches, referred to as thiol(SH)- linked alkylation for the metabolic sequencing (SLAM-seq) [12], TimeLapse-Seq (referred to as TLS-seq in the present study) [13] and thiouridine to cytidine conversion sequencing (TUC-seq) [14], respectively, to study RNA dynamics directly from 4sU-labeled RNAs, without biochemical enrichment or affinity purification. So far, SLAM-seq has gained a larger acceptance [12, 15, 16], but all three methods have been used to infer metabolic RNA rates in a number of studies [17, 18]. Metabolic labeling using SLAM-seq, TLS-seq or TUC-seq is in principle applicable to all organisms, in cell culture as well as *in vivo* [15].

However, to discern between reads originating from newly transcribed or pre-existing RNAs remains a challenging task, due to non-negligible errors arising from the sequencing data and modest 4sU incorporation rates, especially for short periods of labeling [12, 19, 20]. Both SLAM-seq and TUC-seq have comparable conversion rates (>90%), while lower values were reported for TLS-seq (around 80%), but a direct comparison using the same starting material under the same conditions has not yet been performed nor a systematic analysis of RNA kinetics inferred from each protocol. Although different bioinformatics approaches have been used to quantify kinetic rates with [7, 9, 10] or without the profiling of nascent RNA [21], at the time of writing this manuscript, only two open-source softwares were readily available to be used with the latest nucleotide conversion methods. The SLAM-DUNK pipeline provides overall conversion rates and was used to analyze SLAM-seq from 3'-end sequencing (Quant-seq) data [12, 16, 19, 22]. The GRAND-SLAM approach is based on a statistical Bayesian framework, providing posterior distributions of the ratio of new-to-total RNA (NTR) [20], and was used recently in the context of single-cell RNA sequencing [23].

Notably, little effort has been made to compare enrichment-based and nucleotide conversion methods. Conventional, enrichment-based methods are considered to be more laborious and generally associated with higher costs, owing to the need to sequence different fractions, yet offer the advantage to employ resources only on the relevant material, especially for the estimation of extreme decay rates [24]. When 4sU-labeled RNAs only are sequenced after purification, e.g. using biotin-streptavidin, this is often referred to as 4sU-seq. This approach relies on appropriate normalization and may be affected by the efficiency and the specificity of separation [25]. We are not aware of any publicly available genome-wide sequencing data that would allow a direct comparison between 4sU-tagging methods, and there is little evidence to suggest whether any approach outperforms the others or to persuade users to change from biochemical to bioinformatic enrichment of newly transcribed RNAs.

Here, we systematically compare and assess the reliability of four methods to analyze cellular RNA dynamics based on 4sU-tagging: one biochemical enrichment protocol using biotin-streptavidin purification, and three nucleotide conversion protocols, i.e. the SLAM-seq, the TLS-seq and the TUC-seq protocol. Starting from total RNA derived from MCF-7 cells, we constructed a set of equivalent libraries for all methods and estimated decay rates using two different computational workflows. We used pulseR [25], a kinetic and statistical modelling framework based on RNA-seq read counts. To handle data arising from nucleotide conversion protocols, we developed a computational workflow, which we present here, that allows to compare estimates from all four methods. In addition, for the biochemical enrichment protocol, the use of ERCC spike-ins enabled us to compare fraction normalization strategies and provided a reference set of decay rates. For the three nucleotide conversion protocols, we used GRAND-SLAM [20] as an additional computational analysis workflow. We demonstrate that all four protocols are reliable and have a comparable conversion efficiency, but differ with regards to incubation time, time on hands and stability of reagents as well as requirements to handle hazardous substances. Decay rates estimated with two different computational methods were consistent for over 11 600 human genes, with variations in terms of identifiability and confidence intervals (CIs). We also addressed elements of tools usability and accessibility for non-specialist users. We provide guidance on use cases comparing biochemical enrichment and nucleotide conversion protocols that can be used when designing labeling experiments and constructing models. Our evaluation strategy, results and data can serve as a resource for the community to facilitate reproducible research, to identify the most suitable method for specific questions and to advance our understanding of key cellular processes.

## Results

### Evaluating four metabolic RNA labeling protocols

MCF-7 cells were exposed to the uridine analog 4sU for 1, 2, 4 or 8 h. Total RNA (including untreated cells at 0 h) was used as input for four different RNA dynamics analysis protocols (Figure 1a). In the biochemical enrichment protocol, for brevity referred to as the biotin-streptavidin (BSA) purification method, 4sU-labeled and unlabeled RNAs were separated by streptavidin purification after biotinylation with methylthiosulfonate-activated biotin XX (MTSEA-biotin XX, Methods) that has a conversion efficiency of at least 95% [10]. External RNA Controls Consortium (ERCC) spike-ins were added to all fractions: input, eluate (enriched or 4sU-labeled, newly-transcribed RNA) and supernatant (unlabeled, pre-existing RNA). Our analysis showed a gradual increase of biotinylation with labeling time (Figure 2a and Figure S1a, see Supplementary Data available online at http://bib.oxfordjournals.org/) and of RNA concentrations from the biotin-enriched fractions (Figure 2b). The residual amount of biotinylation signal in the supernatant was below 2.5% for every time point, further confirming the high efficiency of the streptavidin purification. These observations were also supported by comparing the relative enrichment of genes with high (MYC), intermediate (PDLIM5) or low (GAPDH) RNA turnover [26, 27] in the supernatant and eluate fractions (Figure 2c and Figure S1b, see Supplementary Data available online at http://bib.oxfordjournals.org/). The estimated contamination in the biotin-enriched fraction at 0 h was 1.40 ± 0.50% for MYC, 2.68 ± 0.50% for PDLIM5 and 1.65 ± 0.59% for GAPDH.

Total RNA was also used in equal amount for SLAM, TimeLapse (abbreviated as TLS) and TUC chemistry (Methods). RNA integrity of all samples was verified before sequencing (Figure S1c, see Supplementary Data available online at

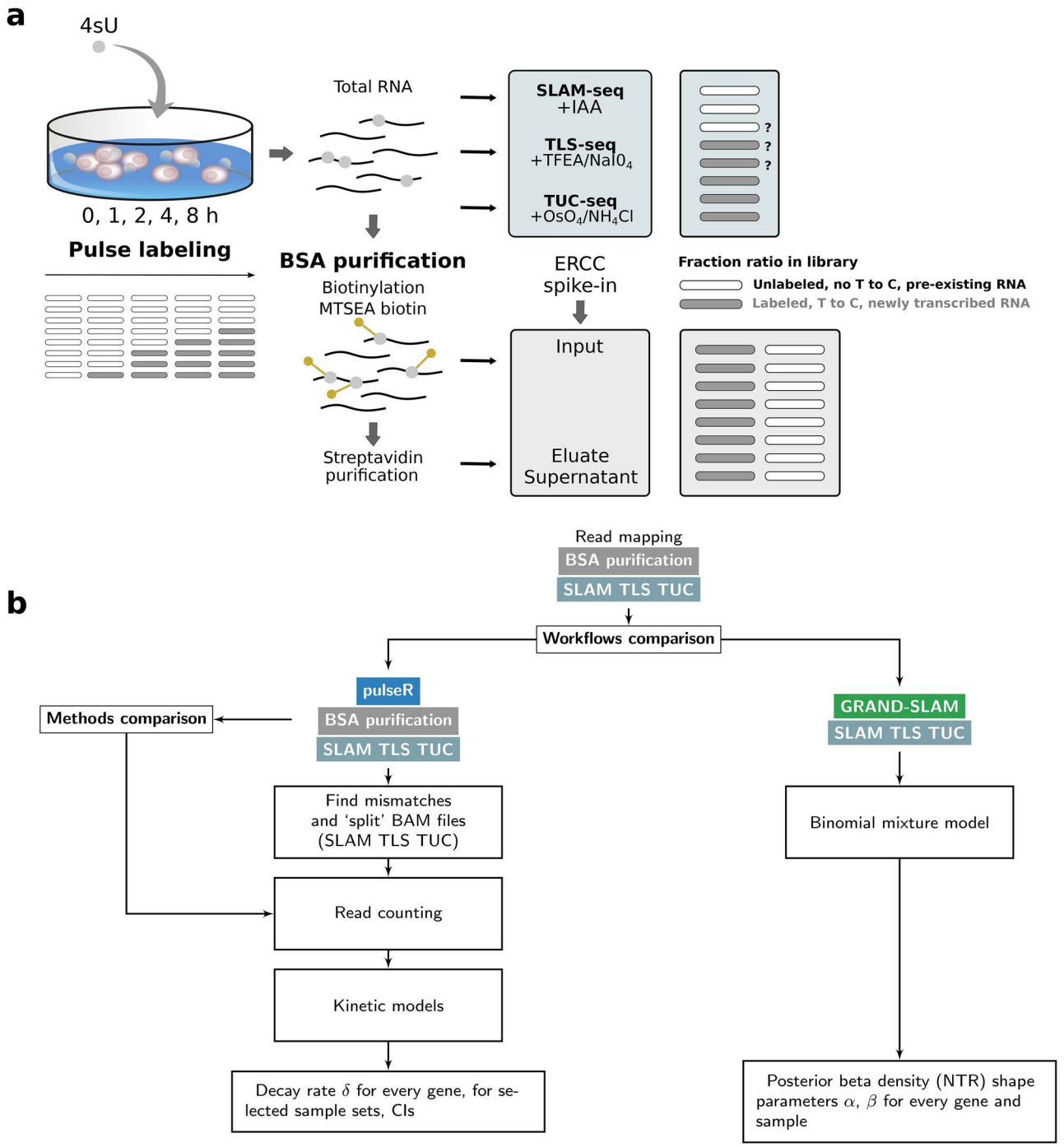


**Figure 1.** 4sU-tagging overview. (**a**) MCF-7 cells were pulse-labeled, and total RNA was used for four different labeling protocols: biochemical enrichment by BSA purification (similar to standard 4sU-seq), SLAM-seq, TimeLapse-seq (TLS-seq) and TUC-seq. SLAM-seq, TLS-seq and TUC-seq methods induce T to C substitutions, which are used to separate pre-existing and newly transcribed RNAs in the sequencing data. Theoretically, the fraction ratio is reflected by the read counts, but due to non-negligible other sources of T to C mismatches, appropriate bioinformatics tools must be employed. In the BSA purification protocol, ERCC RNA spike-in controls are added to normalize the fraction ratio in the libraries. (**b**) pulseR and GRAND-SLAM were used to estimate decay rates from all 4sU-tagging methods. pulseR is an RNA-seq count-based parameter estimation framework. GRAND-SLAM is a statistical software package to estimate NTR 'out-of-the-box'. While both methods handle nucleotide conversion labeling experiments, only pulseR provides estimates for the BSA purification protocol.

http://bib.oxfordjournals.org/). The SLAM-seq protocol employs the primary thiol-reactive compound iodoacetamide (IAA) which covalently attaches a carboxyamidomethyl group to 4sU by nucleophilic substitution [12]. The TLS-seq protocol uses a combination of 2,2,2-trifluoroethylamine (TFEA) and sodium periodate ($NaIO_4$) [13], and the TUC-seq protocol, an osmium tetroxide ($OsO_4$)-mediated transformation to convert 4sU to cytidine derivatives [14]. Instead of introducing a reverse-transcription-dependent T to C conversion as in SLAM-seq, TLS- and TUC-seq directly convert 4sU to cytidine in RNA. We

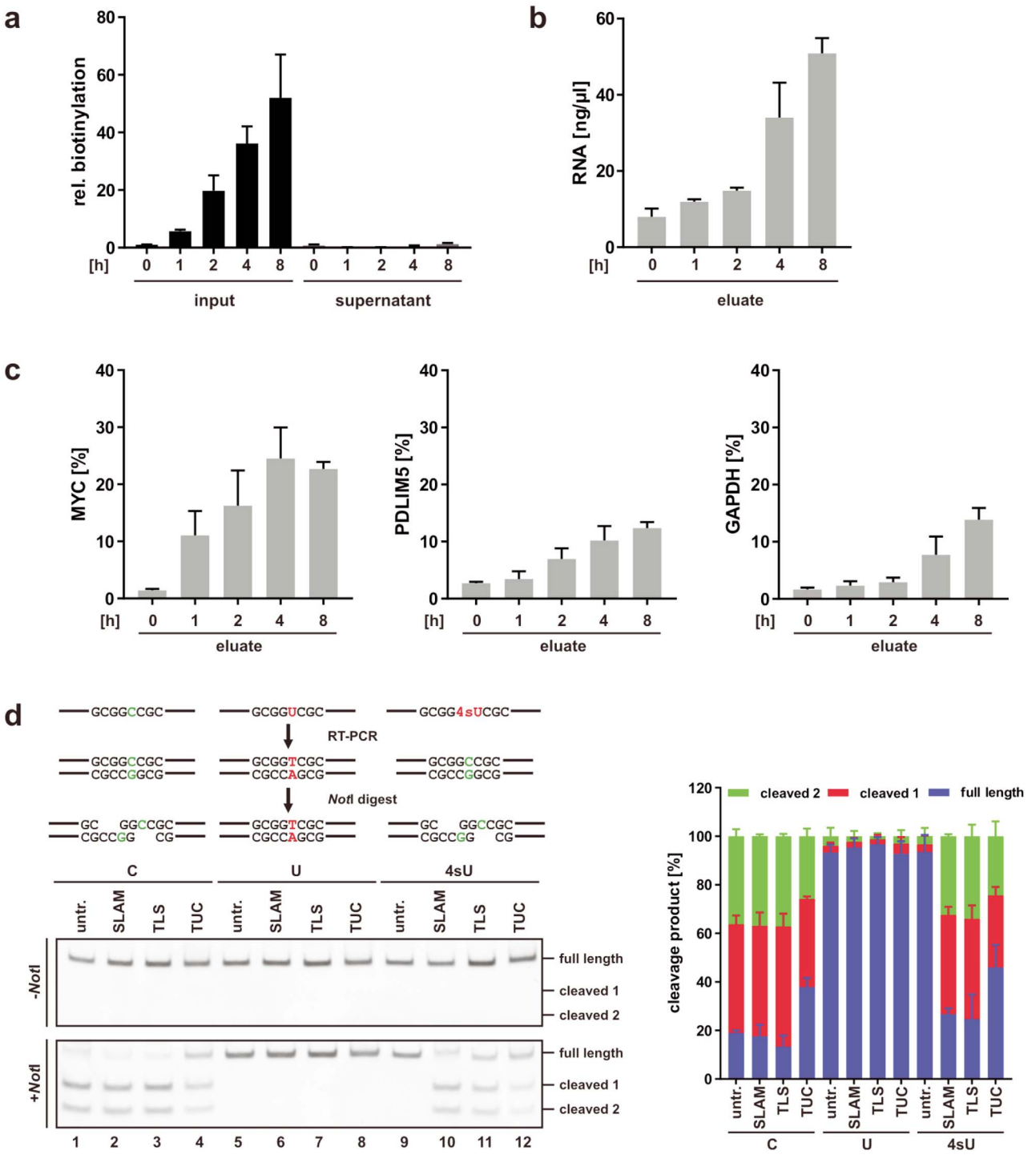


**Figure 2.** (a-c) MCF-7 cells were labeled with 4sU for up to 8 h in two replicates. RNA was either biotinylated and employed in biochemical separation or subjected to chemical conversion using SLAM-, TLS- or TUC-seq protocols. (**a**) Quantification of dot blot analysis of biotinylated RNAs (input) and supernatant of the biochemical separation with Streptavidin-HRP. (**b**) RNA concentration of biotin-enriched fractions (eluate) determined by absorption measurement. (**c**) RT-qPCR analysis of MYC, PDLIM5 and GAPDH mRNA in eluate fractions of the biochemical separation, normalized to the respective input. In a, b and c, mean and SD of two replicates are shown. (**d**) Upper left panel: schematic representation of the restriction enzyme digestion assay probing the efficiency of chemical conversion employing the positive control substrate (C), containing a *Not*I cleavage site, the negative control substrate (U) or the chemical conversion substrate (4sU). All substrates were analyzed untreated and after chemical conversion using SLAM, TLS and TUC protocols. Lower left panel: analysis of the reaction products on 10% TBE gels. Right panel: quantification of the relative fraction of full length band and the two cleavage products derived from *Not*I digestion (+*Not*I) in three independent experiments. untr. = untreated.

observed a relative increase in threshold cycle values ($C_t$ values) during cDNA library preparation for the SLAM protocol (Figure S1d, see Supplementary Data available online at http://bib.oxfordjournals.org/). This increase appeared to be independent of labeling time and was also seen in 0 h samples (treated, without 4sU) but did not affect the reliability of the method,

**Table 1.** Summary and primary chemicals used in each labeling protocol. Hazard statements and H-codes from the globally harmonized system of classification and labelling of chemicals (GHS) associated with each hazardous substance are found in Table S2, see Supplementary Data available online at http://bib.oxfordjournals.org/

| Method | Incubation | On hand | Temperature | pH | Non-hazardous substances | Hazardous substances |
|---|---|---|---|---|---|---|
| **SLAM** | 15 min | 15 min | 50°C | 8.0 | DMSO | Iodoacetamide (IAA), Sodium phosphate |
| **TLS** | 1 h | 15 min | 45°C | | | Sodium periodate ($NaIO_4$), 2,2,2-trifluoroethylamine (TFEA), EDTA |
| **TUC** | 3 h | 15 min | 50°C/25°C | 8.88 | | Osmium tetroxide ($OsO_4$), Ammonium chloride ($NH_4Cl$), Ammonium hydroxide ($NH_4OH$) |
| **BSA Biotinylation** | 30 min | 15 min | 25°C | 7.4 | HEPES, MTSEA-biotin XX | Sodium hydroxide (NaOH), EDTA |
| **BSA Streptavidin purification** | 35 min | 30 min | 25°C-65°C | 7.4 | Tris, Sodium chloride, Tween-20 | Hydrochloric acid (HCl), DTT, EDTA |

BSA = biotin-streptavidin biochemical enrichment.

as further shown below. In all three methods, 4sU-labeled and unlabeled RNA reads were separated *in silico* by the presence or absence of characteristic base substitutions (T to C), removing the need for an enrichment step. In practice, however, non-negligible other sources of T to C mismatches are present, which influence the effective labeling efficiency. Since 4sU is randomly and freely incorporated into nascent RNA, the reaction has to be highly efficient to guarantee sensitivity upon sequencing. To estimate the conversion efficiency of the three methods, we used a restriction enzyme digestion assay [13] (Methods). The conversion efficiency of the 4sU substrate in comparison to the positive control (C) and negative control (U) substrates in each of the three methods was found to be reproducible and not significantly different, SLAM 88.39 $\pm$ 3.54%, TLS 86.42 $\pm$ 13.63% and TUC 83.99 $\pm$ 13.46%, although the cleaved fraction was higher for SLAM and TLS (Figure 2d, Figure S1e, see Supplementary Data available online at http://bib.oxfordjournals.org/ and Table S1, see Supplementary Data available online at http://bib.oxfordjournals.org/). Compared with the conventional approach, these protocols require less input RNA and are less laborious and cheaper. However, in particular for the TLS- and TUC-seq chemistry, substances with considerable hazardous potential have to be handled (Table 1, Table S2, see Supplementary Data available online at http://bib.oxfordjournals.org/). For $OsO_4$ solutions, the low stability has to be considered for consecutive experiments. These considerations are further discussed below, after comparing consistency and inter-method reliability.

## Estimating RNA decay rates

Libraries were identically prepared in two replicates for every time point (0, 1, 2, 4 and 8 h) for every protocol, and in addition for every fraction in the BSA purification protocol, before being submitted to paired-end and stranded sequencing (Methods). We then applied a uniform read processing and alignment workflow to all samples (Methods). Mapping statistics are given in Supplementary File 1 (available on GitHub). The resulting mapping data were then used as input for two different computational workflows (Figure 1b, and Table 2): on the one hand, pulseR [25], a count-based parameter estimation framework using the negative binomial distribution, where RNA dynamics are described by expressions for the mean RNA abundances in different populations, and, on the other hand, GRAND-SLAM, a statistical approach used to infer the proportion of newly transcribed RNA for all nucleotide conversion methods [20]. While pulseR handles BSA purification (standard 4sU-seq or biochemical enrichment) and nucleotide conversion methods, GRAND-SLAM only works with the latter, but both ultimately rely on the assumption of a 1st order reaction with a constant rate of transcription to estimate the decay rate $\delta$ (Methods). The flexibility of pulseR enables to handle different designs, kinetic models and normalization strategies. In the standard pulseR workflow (BSA purification), normalization factors are either derived from gene counts using maximum likelihood estimation (MLE) or determined from ERCC spike-ins. For the nucleotide conversion methods, fraction normalization is not an issue. However, 'labeled' and 'unlabeled' RNAs must be separated *in silico*. This can be done by finding T to C mismatches in mapped reads, accounting for mismatches found in unlabeled RNAs and sites of SNPs that could be erroneously identified as chemically induced T to C conversions. Since GRAND-SLAM needs the alignment files as input, no pre-processing was required. However, pulseR needs RNA-seq count data as input, thus we first had to assign reads to 'labeled' and 'unlabeled' RNA populations. Reads were then counted using featureCounts, following the same approach as in the standard pulseR workflow. SNP calling, implemented in GRAND-SLAM, was done on all samples of each of the SLAM-, TLS- and TUC-seq protocols. These results were also used to filter sites in the pulseR approach (Methods). For the parameter estimation only, both pulseR and GRAND-SLAM showed a comparable usage (Table 2). While pulseR directly outputs decay rate estimates for selected time points, in GRAND-SLAM, parameters of the posterior distribution for the NTR are derived using Bayesian inference from the mapping data, without additional user intervention. To calculate decay rates for every gene, these parameters were used, weighting the contributions of each sample to the overall estimates, by maximum a posteriori (MAP) estimation (Methods). Unless otherwise stated, results are shown for all time-points.

## Assessing inter-method reliability

Decay rates $\delta$ were estimated for the BSA purification protocol using pulseR (Methods). In a first instance, ERCC spike-ins were used to determine the normalization factors in the

**Table 2.** Description, usage and comparison of computational workflows used in this study

| | pulseR | GRAND-SLAM |
|---|---|---|
| Description | Maximum likelihood estimation using RNA-seq count data [25]. R programming language. GPL-3 license. https://dieterich-lab.github.io/pulseR | Bayesian inference using mapped reads [20]. Java 1.8. License agreement for academics. https://github.com/erhard-lab/gedi/wiki/GRAND-SLAM |
| Usage[1] | Elapsed 00:40:39, Memory 3.68 GB, CPU time 97560[2] | Elapsed 00:43:11, Memory 24.84 GB, CPU time 103640[3] |
| Positives | Flexible (variable design, models, e.g. could include uridine bias, RNA processing, etc.), gene-specific and shared parameters, different read pre-processing can be used (only count data is required), can handle both chemical conversion and biochemical enrichment (incl. spike-ins), well-documented. | Black box (read counting, SNPs finding, etc.), no knowledge of specific programming language required, portable (Java), in principle more robust with limited number of samples, less affected by the choice of time points, recently extended to single-cell applications. |
| Negatives | More user involvement required (basic R knowledge), read pre-processing[4], less portable, for the chemical conversion shows more variability (more parameters to fit). | Less user control, limited documentation, difficulty to access source code, MAP estimation left to user[5], cannot be used on matched biochemical enrichment experiments (BSA or standard 4sU-seq). |

[1]The benchmark was made on the SLAM-seq samples. Jobs were run on a SMP Debian 3.16.84-1 x 86_64 GNU/Linux cluster, scheduled on the same partition, each allocated to one node and 40 CPUs. Compute nodes consisted of Intel Xeon processors E5-2650 v3, 40 cores, 256GB. Elapsed is the wall-clock time in hours:minutes:seconds, CPU time is in cpu-seconds. Memory is the maximum memory utilized per core. For GRAND-SLAM, we used the BAM file format.
[2]Only fitting is reported. Mapped reads pre-processing was done in parallel and utilized on average 22.40 GB, and ran for 04:03:55. Read counting using featureCounts utilized 678.39 MB and took 00:05:18.
[3]The index creation step is not reported.
[4]The read pre-processing and the complete pulseR workflow used in this study are available at https://github.com/dieterich-lab/ComparisonOfMetabolicLabeling.
[5]Decay rates are not readily available, they need to be computed from the posterior NTR shape parameters. We used MAP estimation, see Methods.

model. In total, 19 ERCC transcripts with the highest counts were used across replicates (Figures S2–S4, see Supplementary Data available online at http://bib.oxfordjournals.org/, and Methods). The predicted and raw abundances for all genes showed a good agreement for most replicates (Figure S5, see Supplementary Data available online at http://bib.oxfordjournals.org/). This model, referred to as 'ERCC', had the simplest description and the least number of fitted parameters, and thus provided a set of reference decay rates. In the absence of spike-ins, pulseR can also infer the normalization factors from the data. Hence, we computed decay rates for the same gene set using the same kinetic model description, but discarding the information from the ERCC spike-in read counts. We refer to this model as 'BSA'. Hence both ERCC and BSA are based on the biochemical enrichment protocol data. For the nucleotide conversion protocols, referred to as 'SLAM', 'TLS' and 'TUC', estimates were obtained with each the pulseR workflow and GRAND-SLAM. We first determined raw mismatch statistics (4sU-mediated base conversions) for each protocol and all replicates (Supplementary Files 2 and 3, available on GitHub). Mismatch rates were comparable between protocols, whether they were calculated with the pulseR workflow or with GRAND-SLAM. The rate increase with labeling time was clearly distinguishable and was also consistent along the read length, with artefactual patterns of mismatches at read ends being less pronounced at later time points (Supplementary File 2, available on GitHub). In all computational workflows, mismatches at read ends were discarded (Methods).

We sought to systematically compare the four protocols with pulseR and GRAND-SLAM using correlation as a measure of method reliability, on a common set of genes for which estimates were available for all methods (Figure 3 and Table S3, see Supplementary Data available online at http://bib.oxfordjournals.org/). The consistency between all protocols indicated a relatively high inter-method reliability (Figure 3a). A high correlation can be further appreciated from the perspective that any combination of protocol-computational method was independent from the other, associated with different experimental protocols and distinct bioinformatic treatment. These observations were further corroborated by comparing the deviation between any pair of methods (Figure 3b). In the absence of an enrichment step, as is the case for the nucleotide conversion protocols, many reads will map to pre-existing RNAs, and there will be a bias towards expressed transcripts with slower rates (shift above the diagonal when comparing nucleotide conversion versus ERCC/BSA). This was distinctly observable, particularly using only shorter or intermediate time points to estimate the decay rates (Figures S6–S8, see Supplementary Data available online at http://bib.oxfordjournals.org/), and more prominent for GRAND-SLAM, when compared with pulseR ERCC/BSA (with enrichment step). At higher rates, however, estimates from the pulseR SLAM, TLS and TUC models systematically deviated from the BSA/ERCC results, while those from GRAND-SLAM were generally more symmetric. Across the different subsets, the correlations generally remain high (Figure S9, see Supplementary Data available online at http://bib.oxfordjournals.org/). While pulseR estimates were more sensitive to the choice of time points, particularly for the nucleotide conversion protocols, MAP estimation in GRAND-SLAM was much less affected. GRAND-SLAM uses a kinetic model equivalent to the BSA model implemented in pulseR, except that it does not use the negative binomal, and where inference on the decay rates is done as if mean abundances were known. In pulseR, although all models were processed using the same workflow, there were more parameters to fit for the SLAM, TLS and TUC kinetic models, compared with the ERCC/BSA kinetic model, the former showing a bias towards faster rates for a subset of genes, particularly visible at intermediate and later time points (Figures S6–S8, see Supplementary Data available online at http://bib.oxfordjournals.org/). These fastest genes were mostly unidentifiable (see also Supplementary File 5, available on GitHub). It is not surprising if estimates obtained with these models showed a slightly lower correlation, and had higher deviations when compared with the ERCC/BSA results (i.e. correlations were marginally higher when comparing results between the pulseR ERCC/BSA kinetic model and those obtained with GRAND-SLAM, compared with those of the pulseR SLAM,

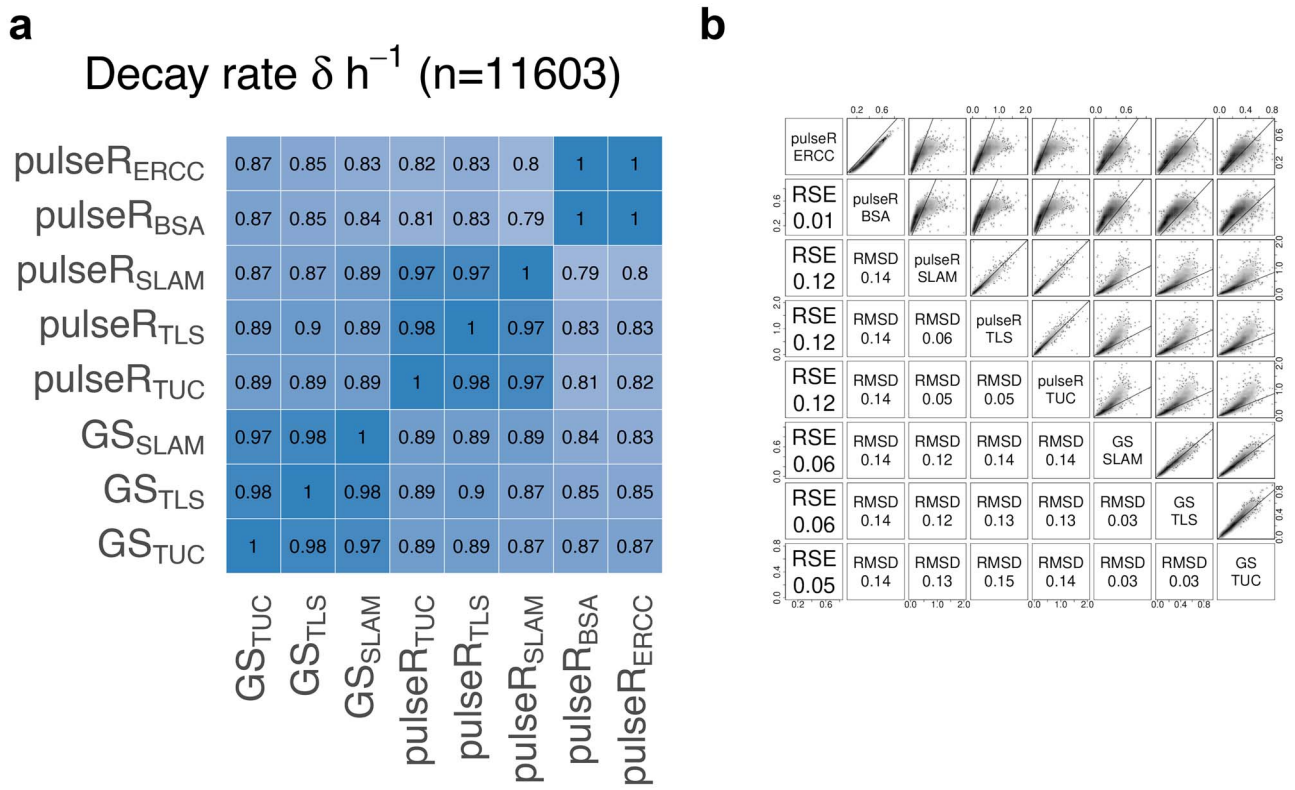


**Figure 3.** Assessing inter-method reliability. **(a)** Pearson's correlation coefficient for RNA decay rate estimates $\delta$ between any pair of methods for 11 603 common genes. **(b)** Scatter matrix of decay rate estimates $\delta$ between any pair of methods for 11 603 common genes with, on the lower diagonal elements, residual standard error from a fitted regression model using the ERCC results, and root-mean-square deviation for all other comparisons. All time points (0, 1, 2, 4 and 8 h samples) were used to estimate parameters in pulseR. The MAP estimator for $\delta$ in GRAND-SLAM was computed on the 1, 2, 4 and 8 h samples.

TLS and TUC models, irrespective of the subset of time points used).

The overlap between approximate CIs for the nucleotide conversion protocols and the ERCC results were also determined (Figures S10–S13, see Supplementary Data available online at http://bib.oxfordjournals.org/). The overlap with the pulseR SLAM, TLS and TUC results was almost entirely restricted by the width of the ERCC CIs. Since GRAND-SLAM estimated marginally slower rates than those obtained with the ERCC model, and due to its tighter intervals, the overlap was mostly restricted to the lower end of those found by the ERCC model. Across all three nucleotide conversion protocols, an average of 4264 genes had non-overlapping CIs, approximately 78% of these from the GRAND-SLAM results. However, the number of unidentifiable parameters was higher in pulseR (on average 100 genes with either one or both bounds undefined). At shorter time points (1 and 2 h), GRAND-SLAM estimates were largely non-overlapping, while at later time points (4 and 8 h), it had a more symmetric distribution with respect to the ERCC estimates (Figures S11 and S13, see Supplementary Data available online at http://bib.oxfordjournals.org/).

## Comparing computational workflows

For all shared estimates ($n$ = 11 603), RNAs with low, intermediate or high turnovers were classified consistently across all methods, and decay rates were generally identifiable (small median CI width, Figure 4a). For the other combinations of computational workflow and protocol, estimates were only available for smaller subsets of genes. By default, GRAND-SLAM excludes certain biotypes based on the annotation, and many of these genes constituted the pulseR-only intersections, which were also characterized by more variability in the estimated parameters (blue and/or grey-shaded lines, Figure 4a). Other genes, included in the pulseR analyses, had unidentifiable parameters of the posterior NTR $(\alpha, \beta)$ in GRAND-SLAM for at least one sample and thus were not included in the MAP estimation. In pulseR, on the contrary, estimates are always available, although not necessarily identifiable (based on the CI), for all genes in the input count table. Low-expression gene filtering prior to parameter estimation is, however, recommended, as is common practice in quantitative profiling of gene expression. The user has full control of how this is done. Here, we only used genes with mean total read count >50. The GRAND-SLAM-specific intersections were thus characterized by lowly expressed genes for which estimates were not available in any of the pulseR models, because of low-expression filtering (e.g. SLAM, TLS and TUC, see Figure 4b). The GRAND-SLAM estimates for these sets were associated with greater median variability (Figure 4a, SLAM GS, TLS GS and TUC GS). When considered together, all GRAND-SLAM-specific intersections showed a large read count variability (Figure 4b, $x$-axis all intersections), revealing discrepancies in read counting between the two methods for a subset of genes, despite our effort to match the GRAND-SLAM parameters. Regardless of their CIs, the GRAND-SLAM-specific estimates for these lowly expressed genes may not be reliable. There was also a statistically significant difference between decay rates estimated for all shared genes and those available in any one

method (Figure 4c GS shared versus GS only). While the median distribution for each subset differed for all comparisons, the estimates were consistent for pulseR, whether they were from the shared or the pulseR-specific gene sets (Figure 4c pulseR shared versus pulseR only). The pulseR-specific genes were also characterized by consistent gene ontology enrichment associations across protocols, which were arising from biotypes that were not included in the GRAND-SLAM analyses (Figure S14, see Supplementary Data available online at http://bib.oxfordjournals.org/).

Usage and runtime for both methods were comparable but may be affected by the chosen file format (GRAND-SLAM has its own format based on centered interval trees, but we used the BAM format for all comparisons), see Table 2. The pre-processing may also affect overall runtime in the pulseR workflow. We did not assess the influence of read mapping on estimates for each method, but it was shown previously to be negligible for GRAND-SLAM [20]. As for read counting, we were restricted in this work by the parameters modifiable in GRAND-SLAM.

### Other analyses

We also compared the pulseR workflow for the nucleotide conversion protocols using a different SNP/variant caller (Supplementary File 4, available on GitHub). The pulseR estimates were not significantly affected by the choice of variant caller, whether GRAND-SLAM or BCFtools were used to identify SNPs. Across pulseR results only, there was essentially no difference in estimates between the SLAM, the TLS and the TUC protocols. We also estimated optimal labeling times and key characteristics of RNA labeling experiments for use cases comparing biochemical enrichment and nucleotide conversion protocols, using the asymptotic theory of optimal design (Supplementary File 5, available on GitHub).

The effect of long 4sU labeling durations on cellular viability has been largely ignored in the literature, in particular in the case of SLAM-, TLS- and TUC-seq. In https://github.com/dieterich-lab/ComparisonOfMetabolicLabeling, we include a global gene expression analysis between no labeling and the different labeling durations for the nucleotide conversion protocols. We show that exposure to 4sU for up to 8 h did not result in any significant effect on gene expression. After 8 h of labeling, processes involved in cellular stress responses were represented by a small set of signature genes; however, changes in cell morphology and viability were not observable at this time point.

Biochemical enrichment protocols, such as standard 4sU-seq [28] or TT-seq [26], enable to map the whole human transient transcriptome at nucleotide resolution using ultrashort and/or progressive 4sU-tagging. Due to requirements on sequencing depth, and largely because most previous studies employed Quant-seq 3'end sequencing, these analyses were not attempted with SLAM-, TLS- or TUC-seq. As noted earlier, nucleotide conversion protocols are not currently suitable for the determination of very fast kinetics. Despite these limitations, since we used TruSeq stranded mRNA sequencing, we decided to include an RNA-seq analysis of differential exon usage using the nucleotide conversion protocols. In particular, we show that certain exons are uniquely associated to a given transcript isoform, and differentially used depending on the labeling duration, which could indicate the existence of decay rate-dependent 'isoform switching'. The supporting scripts and complete results are available at https://github.com/dieterich-lab/ComparisonOfMetabolicLabeling.

## Discussion

4sU-tagging metabolic labeling methods have not undergone comprehensive benchmarking nor have computational tools to infer decay rates from such experiments. Here, we compared four of the most recent and widely applied methods: biochemical enrichment, referred to as BSA purification (with and without ERCC spike-ins), SLAM-seq, TimeLapse-seq, referred to as TLS-seq, and TUC-seq. Our results indicate that there is no protocol that can be considered a gold standard in terms of efficiency, reproducibility and reliability, but additional considerations are also important. BSA purification has a higher cost, compared with the nucleotide conversion methods, largely due to the requirement to sequence several fractions, and because of the price of the magnetic streptavidin beads that are generally used. The nucleotide conversion protocols vary in the associated time, reaction conditions and chemicals. In general, they are less laborious than the biochemical enrichment protocol. However, integrity of the RNA should be carefully controlled after the chemical reaction. We noticed that incubation at higher temperature, e.g. for TUC-seq, can result in significant degradation. Furthermore, TLS-seq and TUC-seq require handling of substances with considerable hazardous potential. Although the required amounts are low, this should be taken into account during experimental planning. Many of the required chemicals, as $OsO_4$, are unstable in solution, and this can contribute to increase in costs for these methods.

The choice of method is also highly dependent on the experimental design. Incorporation rates may be a crucial parameter for measurements of either fast-decaying or slow-decaying RNAs, and in such cases, it may be less efficient to use nucleotide conversion protocols to estimate metabolic rates. Even after 2 h, no more than 1% of all reads actually contained a T to C conversion, thus short pulses require a large number of sequencing reads to achieve sufficient coverage for less abundant transcripts, which limits sensitivity. Longer pulses have an effect similar to an increase in sequencing depth for the labeled fraction, but contain little information to infer rates for the fastest genes. In contrast, biochemical enrichment protocols can be used to focus the sequencing capacity on the relevant molecules for a given experimental design [24]. However, as already noted, this approach requires higher quantities of starting material, it suffers from the need to normalize the fractions and estimates can be biased by the efficiency of separation or fraction contamination. Here, we showed a high efficiency of the streptavidin purification and negligible contamination of the biotin-enriched fraction. As recently advocated, contamination should be assessed at every experiment [21].

We also aimed to test broadly applicable computational methods: pulseR [25], a framework applicable to any type of metabolic labeling experiments, and GRAND-SLAM [20], which has been used to infer NTR ratios in the nucleotide conversion protocols. In pulseR, the kinetic equations describing the RNA populations must be specified. They directly influence parameter optimization and the required number of time points. In the nucleotide conversion models, the presence of background fractions, considered as nuisance parameters, adversely affected the CIs, which were calculated using the profile likelihood method [25]. For the less experienced user, the flexibility of this approach might weigh down some of its advantages, and one might consider using GRAND-SLAM, a portable statistical inference framework 'out-of-the-box'. GRAND-SLAM estimates are less affected by the choice of time points, and the software provides a detailed output for quality

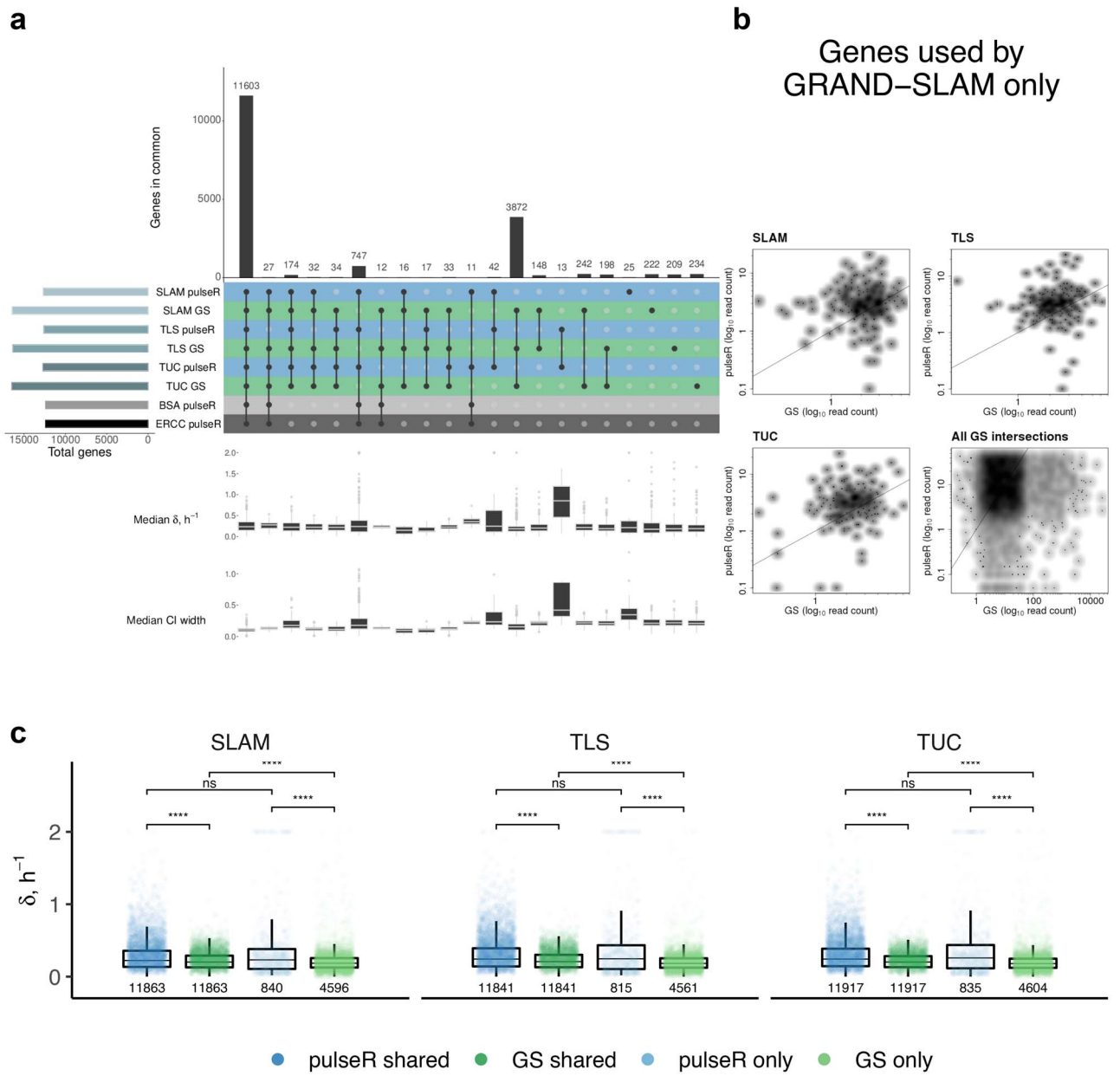


**Figure 4.** Comparing computational workflows. (**a**) UpSet plot showing the overlap between estimates across methods, median decay rate estimates $\delta$ and median CI width for each intersection. Intersections with less than 10 estimates are not shown. (**b**) Scatter plot of read counts between GRAND-SLAM and those obtained with featureCounts (pulseR workflow) for GRAND-SLAM-specific estimates. Decay rates for these genes are not available in pulseR due to low-expression gene filtering. For SLAM, TLS and TUC, total mean read count is shown. For all intersections (lower right), the median of the total mean count across protocols is shown. Gene sets are obtained by selecting appropriate intersections in a. (**c**) Box plot comparisons of decay rate estimates $\delta$ between pulseR and GRAND-SLAM for the nucleotide conversion protocols. Comparisons are made using a two-sided Mann–Whitney test. The shared genes for SLAM (resp. TLS, TUC) are those where estimates are available for both pulseR and GRAND-SLAM for the SLAM protocol (resp. TLS, TUC protocol), while the genes found only by one or the other method are genes for which estimates are available only for that method using the SLAM protocol (resp. TLS, TUC protocol). For all figures, all time points (0, 1, 2, 4 and 8 h samples) were used to estimate parameters in pulseR. The MAP estimator for $\delta$ in GRAND-SLAM was computed on the 1, 2, 4 and 8 h samples.

control, without the need to run additional analyses. Lower and upper bounds of the posterior distribution of the NTR are given as output for each gene and sample, but these are not readily usable, as one is generally interested in a single estimate per gene. Here, we computed decay rates using MAP estimation on subsets of samples, and estimated approximate CIs using the profile likelihood method, without nuisance parameters, explaining the tighter intervals obtained for GRAND-SLAM.

In summary, we recommend that this kind of analysis is updated regularly, as it is a fast-developing field. Standardized methods for benchmarking metabolic labeling protocols and extensive public data, as we present here, can facilitate the development of computational tools, serve as a resource for the community to promote reproducible research and help to advance our understanding of RNA biology.

## Methods

### Tissue culture

MCF-7 cells (ACC-115) were obtained from the Leibniz Institute DSMZ German Collection of Microorganisms and Cell Cultures. Cells were routinely tested for mycoplasma contamination with Venor GeM Classic (Minerva Biolabs). MCF-7 cells were cultured at 37°C and 5% $CO_2$ and maintained in DMEM (Thermo Fisher Scientific) supplemented with 10% fetal calf serum (Merck), 1×MEM non-essential amino acids (Thermo Fisher Scientific) and 1×Penicillin/Streptomycin (Thermo Fisher Scientific). MCF-7 cells were seeded 48 h prior to the experiment at a cell density of $0.3\times10^5$ cells per $cm^2$. Cells were labeled with 4-thiouridine (4sU) (Sigma-Aldrich) at a final concentration of 200 μM for 1, 2, 4 or 8 h. Cells were scraped in DPBS and the pellet re-suspended in Trizol (Thermo Fisher Scientific).

### Isolation of total RNA

Total RNA was isolated using the Trizol method. Briefly, the cell pellet was re-suspended in 750 μl Trizol and incubated 5 min at room temperature before addition of 200 μl chloroform. Samples were centrifuged (20 min, 10.000 g, room temperature) and the aqueous phase re-extracted with one volume chloroform: isoamylalkohol (24:1) (5 min, 10.000 g, room temperature). The RNA in the aqueous phase was precipitated with one volume isopropanol (30 min, 20.8000 g, 4°C), washed twice with 1 ml 80% ethanol in DEPC-$H_2O$ and dissolved in 25 μl DEPC-$H_2O$ (10 min, 55°C, shaking).

### Biotinylation of RNA

Total RNA was biotinylated using MTSEA biotin-XX (Biotium) as described by Duffy *et al.* [10]. Briefly, 160 μg total RNA was incubated in 10 mM HEPES pH 7.5, 1 mM EDTA and 5 μg MTSEA biotin-XX (freshly dissolved in DMF) in a total volume of 250 μl. Reactions were incubated 30 min in the dark at room temperature. Biotinylated RNA was recovered by extraction with one volume phenol:chloroform:isoamylalkohol (24:24:1) (PCI) and separated by centrifugation (5 min, 20.800 g, room temperature) using Phase-Lock-tubes. RNA was precipitated by addition of 350 μl isopropanol, 25 μl 5 M sodium chloride and 1 μl glycogen (Roche Diagnostics, 20 μg $\mu l^{-1}$) (30 min, 20.800 g, 4 °C). RNA was washed twice with 500 μl 80% ethanol in DEPC-$H_2O$ and dissolved in 160 μl DEPC-$H_2O$ (10 min, 55 °C, shaking).

### Dot blot-based detection of biotinylation

Equal volumes of input and supernatant fractions from Streptavidin purification were applied to nylon membrane (Hybond-N, GE Healthcare) using a dot blot device (Carl Roth). RNA was crosslinked twice at 254 nm using the 'Optimal Crosslink' mode of the Spectroline Select XLE-1000 crosslinker. The membrane was blocked 20 min with PBS + 10% SDS and incubated 2 h with Streptavidin-HRP (Thermo Fisher Scientific, 1:5000 in PBS + 10% SDS). Prior to detection with SuperSignal West Pico (Thermo Fisher Scientific), the membrane was washed each three times 10 min with PBS + 10% SDS, PBS + 1% SDS and PBS + 0,1% SDS. Images were acquired with the LAS4000 system (GE Healthcare). Dot blot signals were quantified using ImageQuant TL 8.1 (GE Healthcare) and normalized to a background control. Mean and standard deviation from duplicate measurements of two biological replicates are shown in Figure 2a.

### Streptavidin purification

For purification of biotinylated RNAs, the method described by Schwanhäusser *et al.* was adapted [8]; 25 μg biotinylated total RNA was adjusted to 100 μl with DEPC-$H_2O$ and filled up with Streptavidin binding buffer (Strep-BB) (20 mM Tris, pH 7.4, 0.5 M sodium chloride, 1 mM EDTA) to 200 μl. RNA was denatured 10 min at 65°C and subsequently placed on ice; 100 μl magnetic streptavidin beads (New England Biolabs) were washed once with 200 μl Strep-BB and resuspended in 100 μl Strep-BB. RNA and beads were incubated 15 min at room temperature on a rotating wheel. Beads were washed three times with 500 μl Strep washing buffer (100 mM Tris pH 7.4, 1 M sodium chloride, 10 mM EDTA, 0.1% Tween 20) prewarmed to 55°C. RNA was eluted three times with 100 μl freshly prepared 100 mM DTT and elution fractions pooled for further analysis. RNA was recovered from total RNA, supernatant and eluate by PCI extraction using Phase-Lock-tubes and isopropanol precipitation. The amount of recovered RNA was determined by Nanodrop measurement.

### ERCC spike-ins

Two hundred nanogram RNA from the streptavidin purification were spiked with 4 μl of a 1:1000 dilution of the universal ERCC RNA Control Spike-in Mix 1 (Thermo Fisher Scientific). ERCC spike-ins were used in the PulseR model to normalize the fractions. Since input and supernatant fractions were diluted each in 1:5, to mitigate the resulting amplification bias, we adjusted ERCC read counts in input and supernatant fractions as follows: for matched samples, we scaled the counts using the fold change difference in intercept between fits, when compared with the eluate fraction (the difference in mean read count for eluate and supernatant/input when the ERCC concentration is 0). We assumed non-significant two-way interactions. We only used transcripts with mean read count >50 in the input fraction. For sample 1A (no eluate fraction), we used the scaling factor average of A samples. Based on the adjusted counts, and due to the low dilution ratio (1:1000), a total of 19 ERCC transcripts were used to determine the normalization factors in the model.

### SLAM-seq sample preparation

Alkylation of 4sU-labeled RNA using iodoacetamide was performed as described by Herzog *et al.* [12]. Briefly, 10 μg 4sU-labeled total RNA was incubated in 50% DMSO, 50 mM sodium phosphate buffer pH 8.0 and 10 mM iodoacetamide (Sigma-Aldrich) for 15 min at 50°C. Reaction was quenched by addition of excess DTT and RNA was recovered by PCI (24:24:1) extraction using Phase-Lock-tubes and isopropanol precipitation.

### TimeLapse-seq sample preparation

TimeLapse-seq (TLS-seq) conversions were carried out as described by Schofield *et al.* [13]; 10 μg 4sU-labeled total RNA was added to a mixture of 600 mM TFEA, 1 mM EDTA and 100 mM sodium acetate pH 5.3. After addition of 10 mM $NaIO_4$, the reaction was incubated for 1 h at 45°C. RNA was recovered by PCI (24:24:1) extraction using Phase-Lock-tubes and isopropanol precipitation.

### TUC-seq sample preparation

RNA conversion for TUC-seq analysis was carried out as described by Riml *et al.* [14]; 10 μg 4sU-labeled total RNA was incubated with 180 mM $NH_4Cl$ and 0.45 mM $OsO_4$ for 3 h

at 25°C. RNA was recovered by PCI (24:24:1) extraction using Phase-Lock-tubes and isopropanol precipitation.

### Agarose gel analysis

RNA integrity was routinely analyzed on 1% agarose gels in TAE stained with SYBR Green II RNA Gel Stain (Thermo Fisher Scientific). 500 ng RNA was boiled in loading buffer (10 min, 65°C) and chilled on ice prior to loading. Images were acquired with the LAS4000 system (GE Healthcare).

### Reverse transcription

Equal volumes of RNA recovered from streptavidin purification or 500 ng RNA derived from chemical conversion were reverse transcribed using the Maxima H Minus First Strand cDNA Synthesis Kit (Thermo Fisher Scientific) with Random Primers according to the manufacturers protocol. Briefly, RNA was mixed in a total volume of 15 μl with 1 μl Random Primer and 1 μl dNTP solution and denatured (5 min, 65°C). Reaction was completed by addition of 4 μl 5 ×RT buffer and 1 μl Maxima enzyme and incubated 10 min at room temperature followed by 30 min at 50°C, and denaturation (5 min, 85°C).

### qPCR analysis

Reverse transcription reactions were diluted 1:20 and used for qPCR analysis on a StepOnePlus instrument (Thermo Fisher Scientific) with Power SYBR Green PCR Master Mix (Thermo Fisher Scientific) and primers listed in Table S4, see Supplementary Data available online at http://bib.oxfordjournals.org/. qPCR data obtained from the Streptavidin purification were normalized to the respective input (0–8 h) using the comparative $C_t$ method [29]. For the analysis of SLAM, TLS and TUC samples, the $\Delta C_t$ values relative to the untreated controls are shown.

### Restriction enzyme digestion assay

The assay was performed mainly as described by Schofield *et al.* [13] using identical oligonucleotides. Briefly, the synthetic DNA template was *in vitro* transcribed using the T7 Megascript kit (Thermo Fisher Scientific). For the positive control substrate (C) and the negative control substrate (U) standard nucleotides were used, for the chemical conversion substrate (4sU), UTP was replaced by 4sUTP (Jena Biosciences). Reactions were incubated for 16 h at 37°C. The template was digested with Turbo DNase (15 min, 37°C). Resulting *in vitro* transcripts (IVT) were purified by phenol-chloroform extraction followed by RNA Clean and Concentrator kit (Zymo Research). Integrity of the IVTs was analyzed by gel electrophoresis on 6% TBE-urea gels (Thermo Fisher Scientific); 120 ng of the respective IVTs were treated according to SLAM, TLS and TUC procedures as described above; 50 ng of purified products were reverse transcribed with SuperScript IV (Thermo Fisher Scientific) according to the manufacturers protocol; 2 μl of the reverse transcription was PCR amplified using Taq polymerase (Promega); 5 μl of the PCR product was either digested with *Not*I-HF (New England Biolabs) or mock digested for 1.5 h at 37°C. The products were separated on 10% TBE gels (Thermo Fisher Scientific) and stained with GelRed (Biotium). Quantification of bands from three independent replicates was performed using ImageJ to calculate the fractions of the full length band and the two cleavage products. Assuming that the C substrate resembles 100% conversion efficiency and the U substrate shows 0% conversion efficiency, the conversion efficiency of the 4sU substrate was calculated using the cleaved fraction.

### Library preparation and sequencing

All libraries were prepared using the Illumina TruSeq mRNA stranded sample preparation kit according to the standard protocol. Library preparation started with 200 ng total RNA plus 4 μl ERCC spike-in (1:1000 dilution) or 1 μg total RNA. One sample was removed from the lot due to low RNA concentration (eluate A 1 h). After poly-A selection (using poly-T oligo-attached magnetic beads), mRNA was purified and fragmented. RNA fragments underwent reverse transcription using random primers followed by second strand cDNA synthesis with DNA Polymerase I and RNase H. After end repair and A-tailing, UDI adapters were ligated. The products were then purified and amplified (14 PCR cycles) to create the final cDNA libraries. After library validation and quantification (Agilent 4200 tape station), equimolar amounts of library were pooled. The pool was quantified by using the Peqlab KAPA Library Quantification Kit and the Applied Biosystems 7900HT Sequence Detection System and sequenced on an Illumina NovaSeq 6000 sequencer with a PE100 protocol.

### Read processing and alignment

Quality clipping and adapter removal were performed with flexbar v3.5.0 [30], with –max-uncalled 1 –post-trim-length 248 –min-read-length 20 –qtrim-format sanger –qtrim TAIL qtrim-threshold 10. Reads aligning to a custom bowtie2 v2.3.5 [31] ribosomal index were discarded. All remaining reads were aligned to the human genome (GRCh38.p7) using STAR v2.6.0c [32], with –outFilterMultimapNmax 20 –outFilterScoreMin 1 –outFilterMatchNminOverLread 0.7 –outFilterMismatchNmax 999 –outFilterMismatchNoverLmax 0.05 –alignIntronMin 20 –alignIntronMax 1000000 –alignMatesGapMax 1000000 –alignSJoverhangMin 15 –alignSJDBoverhangMin 10 –alignSoftClipAtReferenceEnds No –outSAMattributes NH HI AS nM NM MD jM jI XS. Reads were then sorted with samtools v1.7 [33], and duplicates were marked and removed using Picard Tools v2.5.0 [34].

### pulseR workflow

We used MLE to obtain parameter values, as implemented in the R package pulseR v1.0.3 [25]. For a given gene, the read count follows a negative binomial distribution $X \sim NB(m(\boldsymbol{\theta}, t), k)$, where $m(\boldsymbol{\theta}, t)$ is the mean read count, which depends on the time of labeling $t$, the decay rate $\delta$, the expression level $\mu$ in the steady-state and sample normalization, hence $\boldsymbol{\theta} = (\mu, \delta, ...)^T$. The parameter $k$ is the overdispersion of the negative binomial distribution and is a shared parameter of the model. The *NB* probability distribution function is

$$P(X = x) = \frac{\Gamma(x + k)}{x!\,\Gamma(k)} \left(\frac{m}{m + k}\right)^x \left(\frac{m + k}{k}\right)^{-k}. \quad (1)$$

The logarithm of the likelihood function depends on the points $X_i$ and the model parameters $\boldsymbol{\theta}$

$$\log \mathcal{L}(\boldsymbol{\theta}, X) = \sum_i \log P(\boldsymbol{\theta}, X_i). \quad (2)$$

The maximum likelihood estimator is

$$\hat{\boldsymbol{\theta}} = \text{argmax}_\theta \log \mathcal{L}(\theta, X). \quad (3)$$

The abundance $m$ or mean read count of an RNA with synthesis rate $\sigma$ and decay rate $\delta$ is modeled using 1st order kinetics

$$\frac{dm}{dt} = \sigma - \delta m. \tag{4}$$

The expression level $\mu$ is generally derived from the total fraction, which ensures identifiability of this parameter. In the conventional approach (BSA purification), where labeled and unlabeled molecules are separated, we have

$$m_{\text{total}} = \mu \tag{5}$$

$$m_{\text{unlabeled}} = x_u \mu e^{-\delta t} \tag{6}$$

$$m_{\text{labeled}} = x_l \mu (1 - e^{-\delta t}). \tag{7}$$

In this work, the normalization coefficients $x_u$ and $x_l$ were either inferred from the data using all fractions or estimated from the ERCC spike-in read counts. To estimate the normalization coefficients $x_u$ and $x_l$, two different approaches were used. For the BSA purification without ERCC spike-ins (referred to as the BSA model), normalization factors were inferred from the data using all fractions during the fitting procedure. For the BSA purification with ERCC spike-ins (referred to as the ERCC model), the normalization factors were determined without fitting. This can done since spike-ins were added in equal amounts to all fractions after purification. This approach does not allow to recover factors to account for differences in the efficiency of separation, but for simplicity, we assumed no cross-contamination between fractions. In the model fitting, we assumed shared normalization coefficients for samples originating from the same time point and fraction. We only used genes with mean read count >50 in the total samples. Counts were obtained using featureCounts v1.5.1 [35], with -t exon -s 2 -p -B -C -M –fraction –ignoreDup. These parameters were determined to match as closely as possible the parameters used in GRAND-SLAM.

In the chemical nucleotide conversion approach (SLAM-, TLS- or TUC-seq), samples were sequenced without prior separation, and labeled and unlabeled RNAs were differentiated on the basis of T to C mismatches of reads mapped to the genome. We use the term ‘labeled’ and ‘unlabeled’ to avoid introducing additional terminology, but technically, a read can be ‘new’ without being labeled (e.g. few Ts in genomic locus, and/or no incorporation took place), and an ‘old’ read or pre-existing RNA can be erroneously tagged as labeled.

We first identified all T to C mismatches and kept those at positions with a base quality score greater than 20 that were at least 5 nt from the read ends. SNP calling was done with GRAND-SLAM on all samples of each of the SLAM-, TLS- and TUC-seq protocols. We also used BCFtools to call SNPs and reported these results in Supplementary File 4 (available on GitHub). Once the SNPs were filtered, read pairs with at least one T to C mismatch were classified as ‘labeled’, and the others were classified as ‘unlabeled’, and reads were counted using featureCounts as described above.

The accuracy of this bioinformatic separation depends on the sequencing errors and the 4sU incorporation rates. Even after a 8 h pulse, not all molecules are labeled, i.e. not all reads show a T to C conversion and thus the labeled fraction never reaches the total level $\mu$. Besides, at 0 h, some T to C mismatches are incorrectly identified as genuine conversions. Assuming that the ratio of labeled to unlabeled fractions is preserved, we have

$$m_{\text{unlabeled}} = \mu_1 + \mu_3 e^{-\delta t} \tag{8}$$

$$m_{\text{labeled}} = \mu_2 + \mu_3 (1 - e^{-\delta t}), \tag{9}$$

where $\mu_1$ is the background unlabeled fraction, which never extincts, even after long labeling times, $\mu_2$ is the background labeled fraction (the average T to C mismatch rate in unlabeled RNA) and $\mu_3$ is the difference between the maximum level of the labeled fraction and $\mu_2$, hence $\boldsymbol{\theta} = (\mu_1, \mu_2, \mu_3, \delta, ...)^T$. We impose the constraint $\mu_1 > \mu_2 + \mu_3$, i.e. the unlabeled fraction is always greater than the maximum labeled fraction. The constraint $\mu_3 > 0$ (maximum labeled fraction is greater than the background error) is automatically satisfied in the implementation. Since the model includes two more parameters, we need at least three different time points and always include the 0 h time point to have information on $\mu = \mu_1 + \mu_2 + \mu_3$. As described above, the model was fitted to the read counts from the sequenced samples for genes with mean read count >50. For the comparisons, we used the following subsets of time pionts: all time points (0, 1, 2, 4, 8 h), early time points (0, 1, 2 h), intermediate time points (0, 2, 4 h) and late time points (0, 4, 8 h).

The 95% CIs for $\delta$ were computed using a likelihood ratio test [25]. An approximate $(1 - \alpha)\%$ CI for $\boldsymbol{\theta}_0$ is the set of $\boldsymbol{\theta}$ satisfying

$$\left\{ \boldsymbol{\theta} : \log \mathcal{L}\left(\hat{\boldsymbol{\theta}}\right) - \log \mathcal{L}\left(\boldsymbol{\theta}\right) \leq \frac{1}{2} \mathcal{X}^2_{1,1-\alpha} \right\}, \tag{10}$$

where $\hat{\boldsymbol{\theta}}$ is the maximum likelihood estimate of the model parameters, ignoring normalisation factors when fitted, and where the 95% quantile of the chi-squared distribution with 1 degree of freedom is $\mathcal{X}^2_{1,1-\alpha} \approx 3.84$, for $\alpha = 0.05$. The profile-likelihood method reduces $\log \mathcal{L}$ to a function of $\delta$ by treating the other components of $\theta$ as nuisance parameters, and maximizing the likelihood over them.

## GRAND-SLAM

Globally refined analysis of newly transcribed RNA and decay rates using SLAM-seq (GRAND-SLAM) is a statistical framework to infer posterior distributions of new and old RNA from chemical nucleotide conversion metabolic labeling experiments [20]. We used GRAND-SLAM v2.0.5f with -strandness Antisense -overlap Unique -trim5p 5 -trim3p 5 on all samples of each of the SLAM-, TLS- and TUC-seq protocols. The regression model for the T to C mismatch rate in unlabeled RNA was infered from the 0 h samples in each protocol.

To obtain estimates of decay rates from GRAND-SLAM for matched subsets of time points, we use the approximate posterior beta densities for proportion parameters measured at time points 1, 2, 4 and 8 h (and similarly for early, intermediate and late time points, as described above). Solving Eq. (4), and setting the initial abundance to zero for newly synthesized RNA (labeled) and to the steady state for pre-existing RNA (unlabeled), we can derive an expression for the ratio of new to total RNA

$$NTR(\delta) = \frac{m_{\text{labeled}}}{m_{\text{labeled}} + m_{\text{unlabeled}}} = 1 - e^{-\delta t}. \tag{11}$$

We used Eq. (11) to estimate the decay rate $\delta$ from the output of GRAND-SLAM. Assuming that the proportion of new and old RNA for gene $g$ is a beta distributed random variable with shape parameters $\alpha$ and $\beta$, and probability density function $b_g(NTR;\alpha,\beta)$, then the distribution of the transformed random variable is

$$b_g(NTR(\delta);\alpha,\beta)\frac{dNTR}{d\delta} = \frac{t}{B(\alpha,\beta)}(1-e^{-\delta t})^{\alpha-1}e^{-\beta\delta t}. \quad (12)$$

The output of GRAND-SLAM provides these approximate beta posterior densities $(\alpha_i, \beta_i)$ measured at times $t_i$. We can then find the MAP estimator for the decay rate $\delta$ by maximising

$$l(\delta) = \sum_i (\alpha_i - 1)\log(1-e^{-\delta t_i}) - \beta_i \delta t_i \quad (13)$$

for any subset of time points, where $i \subseteq \{1,2,4,8\}$. We estimated the 95% CIs for $\delta$ as described above using the profile-likelihood method; however, in this case, there was only one parameter.

### Statistical and comparative analyses

All correlations were reported using the Pearson's correlation coefficient. To determine the overlap between the different genes sets, we adapted and used functions from the R package UpSetR v1.4.0 [36]. To compare selected genes sets, significance was measured using Mann–Whitney U-test (two-sided), at a threshold of * $\leq$ 0.05 (**** $\leq$ 0.0001). GO enrichment to assign functional annotation to selected genes sets was performed with topGO v2.34.0 [37]. The universe of genes consisted of all annotated genes with median (across protocols) mean count (across samples) >50, using the featureCount tables, which was the most extensive. The overlap O between CIs was determined as follows:

$$O_{min} = \max(\text{lower bound}_{ERCC}, \text{lower bound}_j) \quad (14)$$

$$O_{max} = \min\left(\text{upper bound}_{ERCC}, \text{upper bound}_j\right), \quad (15)$$

where $j$ = SLAM, TLS, TUC. With this definition, there is no overlap, e.g. if the maximum lower bound is greater than the minimum upper bound.

**Key Points**

- 4sU-tagging metabolic labeling methods include biochemical enrichment after thiol-specific biotinylation, and recent approaches such as SLAM-seq, TimeLapse-seq or TUC-seq that rely on bioinformatic enrichment of newly transcribed RNAs.
- To infer genome-wide kinetics of RNA abundance remains a challenging task, and there is currently no general solution allowing to incorporate high-throughput sequencing data from both biochemical and bioinformatic enrichment protocols.
- We propose an open-source computational workflow to compare rate estimates from all protocols, providing decay rates and confidence intervals using two different bioinformatics tools, including our proposed workflow, through a detailed practical comparison.
- We benchmark for the first time all these RNA labeling protocols, demonstrating a comparable conversion efficiency and a high inter-method reliability.
- We provide data and resources to advance reproducible research.

## Supplementary data

Supplementary data are available online at *Briefings in Bioinformatics*.

## Data and software availability

Sequencing data for MCF-7 cells for all four protocols used in this study are available at NCBI's Sequence Read Archive through the BioProject accession number PRJNA726397. The pulseR source code is available at https://github.com/dieterich-lab/pulseR under the GPL-3 license. All scripts used in the analysis presented in this study, data used for the figures as well as results generated with pulseR and GRAND-SLAM are available at https://github.com/dieterich-lab/ComparisonOfMetabolicLabeling.

## Authors' contribution

C.D. supervised the project. C.D. and I.S.N.dV. conceived and designed the experiments. J.A. performed library preparation and sequencing. I.S.N.dV. performed the experiments and contributed figures. C.D. processed data and contributed figures. E.B. processed data, developed and implemented the computational workflow, led the analyses and contributed figures. E.B. wrote the paper. All authors contributed research, analysis and writing.

## Acknowledgements

We thank Florian Erhard for sharing with us the GRAND-SLAM software. The pulseR software was developed by Alexey Uvarovskii while employed at the Dieterich Lab.

## Funding

Klaus Tschira Stiftung GmbH (grant 00.219.2013 to E.B. and C.D.).

## References

1. Tani H, Akimitsu N. Genome-wide technology for determining rna stability in mammalian cells. *RNA Biol* 2012;**9**(10):1233–8.
2. Core LJ, Waterfall JJ, Lis JT. Nascent rna sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* 2008;**322**(5909):1845–8.
3. Kwak H, Fuda NJ, Core LJ, *et al*. Precise maps of rna polymerase reveal how promoters direct initiation and pausing. *Science* 2013;**339**(6122):950–3.
4. Churchman LS, Weissman JS. Nascent transcript sequencing visualizes transcription at nucleotide resolution. *Nature* Jan 2011;**469**(7330):368–73.

5. Cleary MD, Meiering CD, Jan E, *et al*. Biosynthetic labeling of RNA with uracil phosphoribosyltransferase allows cell-specific microarray analysis of mRNA synthesis and decay. *Nat Biotechnol* January 2005;**23**(2):232–7.
6. Kenzelmann M, Maertens S, Hergenhahn M, *et al*. Microarray analysis of newly synthesized rna in cells and animals. *Proc Natl Acad Sci* 2007;**104**(15):6164–9.
7. Dolken L, Ruzsics Z, Radle B, *et al*. High-resolution gene expression profiling for simultaneous kinetic parameter analysis of RNA synthesis and decay. *RNA* July 2008; **14**(9):1959–72.
8. Schwanhäusser B, Busse D, Li N, *et al*. Global quantification of mammalian gene expression control. *Nature* May 2011; **473**(7347):337–42.
9. Rabani M, Levin JZ, Fan L, *et al*. Metabolic labeling of rna uncovers principles of rna production and degradation dynamics in mammalian cells. *Nat Biotechnol* May 2011; **29**(5):436–42.
10. Duffy EE, Rutenberg-Schoenberg M, Stark CD, *et al*. Tracking distinct rna populations using efficient and reversible covalent chemistry. *Mol Cell* 2015;**59**(5):858–66.
11. Friedel CC, Dölken L. Metabolic tagging and purification of nascent RNA: implications for transcriptomics. *Mol Biosyst* 2009;**5**(11):1271.
12. Herzog VA, Reichholf B, Neumann T, *et al*. Thiol-linked alkylation of rna to assess expression dynamics. *Nat Methods* Dec 2017;**14**(12):1198–204.
13. Schofield JA, Duffy EE, Kiefer L, *et al*. Timelapse-seq: adding a temporal dimension to rna sequencing through nucleoside recoding. *Nat Methods* Mar 2018;**15**(3):221–5.
14. Riml C, Amort T, Rieder D, *et al*. Osmium-mediated transformation of 4-thiouridine to cytidine as key to study rna dynamics by sequencing. *Angew Chem Int Ed* 2017; **56**(43):13479–83.
15. Matsushima W, Herzog VA, Neumann T, *et al*. Sequencing cell-type-specific transcriptomes with slam-itseq. *Nat Protoc* Aug 2019;**14**(8):2261–78.
16. Thiecke MJ, Wutz G, Muhar M, *et al*. Cohesin-dependent and -independent mechanisms mediate chromosomal contacts between promoters and enhancers. *Cell Rep* 2020; **32**(3):107929.
17. Luo Y, Schofield JA, Simon MD, *et al*. Global profiling of cellular substrates of human dcp2. *Biochemistry* Nov 2020;**59**(43):4176–88.
18. Gasser C, Delazer I, Neuner E, *et al*. Thioguanosine conversion enables mrna-lifetime evaluation by rna sequencing using double metabolic labeling (tuc-seq dual). *Angew Chem Int Ed* 2020;**59**(17):6881–6.
19. Neumann T, Herzog VA, Muhar M, *et al*. Quantification of experimentally induced nucleotide conversions in high-throughput sequencing datasets. *BMC Bioinformatics* May 2019;**20**(1):258.
20. Jürges C, Dölken L, Erhard F. Dissecting newly transcribed and old RNA using GRAND-SLAM. *Bioinformatics* 06 2018;**34**(13):i218–26.
21. Furlan M, Galeota E, Gaudio ND, *et al*. Genome-wide dynamics of rna synthesis, processing, and degradation without rna metabolic labeling. *Genome Res* October 2020;**30**(10): 1492–507.
22. Bryan AF, Wang J, Howard GC, *et al*. WDR5 is a conserved regulator of protein synthesis gene expression. *Nucleic Acids Res* 01 2020;**48**(6):2924–41.
23. Erhard F, Baptista MAP, Krammer T, *et al*. scslam-seq reveals core features of transcription dynamics in single cells. *Nature* Jul 2019;**571**(7765):419–23.
24. Alexey Uvarovskii IS, Vries N-d, Dieterich C. On the optimal design of metabolic rna labeling experiments. *PLoS Comput Biol* 08 2019;**15**(8):1–22.
25. Uvarovskii A, Dieterich C. pulseR: Versatile computational analysis of RNA turnover from metabolic labeling experiments. *Bioinformatics* 06 2017;**33**(20):3305–7.
26. Schwalb B, Michel M, Zacher B, *et al*. Tt-seq maps the human transient transcriptome. *Science* 2016;**352**(6290): 1225–8.
27. Hendriks G-J, Jung LA, Larsson AJM, *et al*. Nasc-seq monitors rna synthesis in single cells. *Nat Commun* Jul 2019;**10**(1): 3138.
28. Windhager L, Bonfert T, Burger K, *et al*. Ultrashort and progressive 4su-tagging reveals key characteristics of rna processing at nucleotide resolution. *Genome Res* October 2012;**22**(10):2031–42.
29. Schmittgen TD, Livak KJ. Analyzing real-time pcr data by the comparative c(t) method. *Nat Protoc* 2008;**3**(6): 1101–8.
30. Roehr JT, Dieterich C, Reinert K. Flexbar 3.0 - SIMD and multicore parallelization. *Bioinformatics* 05 2017;**33**(18): 2941–2.
31. Langmead B, Salzberg SL. Fast gapped-read alignment with bowtie 2. *Nat Methods* 2012;**9**(4):357–9.
32. Dobin A, Davis CA, Schlesinger F, *et al*. Star: ultrafast universal rna-seq aligner. *Bioinformatics* 2013;**29**(1): 15–21.
33. Li H, Handsaker B, Wysoker A, *et al*. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 06 2009;**25**(16):2078–9.
34. Picard toolkit. http://broadinstitute.github.io/picard/, 2019.
35. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 11 2013;**30**(7):923–30.
36. Conway JR, Lex A, Gehlenborg N. UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics* 06 2017;**33**(18):2938–40.
37. Alexa A, Rahnenfuhrer J. *topGO: Enrichment Analysis for Gene Ontology*. R package version 2.28.0, 2018.