Title

Erratum: Comprehensive multi-center assessment of small RNA-seq methods for quantitative miRNA profiling

Journal

Nature Biotechnology, 36(9)

ISSN

1087-0156

Authors

Giraldez, Maria D

Spengler, Ryan M

Etheridge, Alton

et al.

Publication Date

2018-09-01

DOI

10.1038/nbt0918-899b

Peer reviewed

nature biotechnology

Comprehensive multi-center assessment of small RNA-seq methods for quantitative miRNA profiling

Maria D Giraldez[1,17], Ryan M Spengler[1,17], Alton Etheridge[2], Paula M Godoy[3], Andrea J Barczak[3], Srimeenakshi Srinivasan[4], Peter L De Hoff[4], Kahraman Tanriverdi[5], Amanda Courtright[6], Shulin Lu[7], Joseph Khoory[7], Renee Rubio[8], David Baxter[9], Tom A P Driedonks[10], Henk P J Buermans[11], Esther N M Nolte-'t Hoen[10], Hui Jiang[12,13], Kai Wang[9], Ionita Ghiran[7], Yaoyu E Wang[8], Kendall Van Keuren-Jensen[6], Jane E Freedman[5], Prescott G Woodruff[14], Louise C Laurent[4], David J Erle[3], David J Galas[2] & Muneesh Tewari[1,12,15,16]

RNA-seq is increasingly used for quantitative profiling of small RNAs (for example, microRNAs, piRNAs and snoRNAs) in diverse sample types, including isolated cells, tissues and cell-free biofluids. The accuracy and reproducibility of the currently used small RNA-seq library preparation methods have not been systematically tested. Here we report results obtained by a consortium of nine labs that independently sequenced reference, 'ground truth' samples of synthetic small RNAs and human plasma-derived RNA. We assessed three commercially available library preparation methods that use adapters of defined sequence and six methods using adapters with degenerate bases. Both protocol- and sequence-specific biases were identified, including biases that reduced the ability of small RNA-seq to accurately measure adenosine-to-inosine editing in microRNAs. We found that these biases were mitigated by library preparation methods that incorporate adapters with degenerate bases. MicroRNA relative quantification between samples using small RNA-seq was accurate and reproducible across laboratories and methods.

RNA-seq has transformed transcriptome characterization in a wide range of biological contexts[1,2]. RNA-seq can be used to sequence long reads (long RNA-seq; for example, messenger RNAs and long non-coding RNAs) and short RNAs (small RNA-seq; for example, small non-coding RNAs such as microRNAs). These applications differ in terms of the size of the targeted RNAs and by the technical methods used and the resulting biases in the quantitative data that are produced[3]. For example, preparation of libraries for long RNA-seq, by virtue of having sufficiently long target RNA lengths, commonly utilizes primers for direct generation of cDNA from RNA. In contrast, small RNA-seq library preparation methods typically require RNA ligation or poly-A tailing steps to overcome the challenge of performing reverse transcription and subsequent PCR amplification from extremely short (for example, 16–30 nt) target RNA sequences.

Multiple approaches have been developed to overcome the challenge of uniformly and robustly generating cDNA from small RNAs for the purpose of small RNA-seq[4–9]. Protocols in use for small RNA-seq therefore vary more widely than those used for long RNA-seq, creating greater potential for variation from different library preparation protocols and different labs. In addition, small RNA-seq is increasingly used to study samples with very low RNA concentration, such as biofluids containing exosomes, other extracellular vesicles (EV)[10–16] and RNA-protein complexes[17–21]. Normalization methods[22–24] that have been developed to correct for variation in long RNA-seq data are typically not well-suited for small RNA-seq data. Although performance characteristics such as reproducibility and quantitative accuracy have been well-studied for long RNA-seq[25,26], only the reproducibility of a single library preparation protocol has been evaluated for small RNA-seq[25].

Furthermore, the performance of different small RNA-seq methods for quantifying single-nucleotide changes in RNA sequence, such as those seen with microRNA (miRNA) editing, for example, has not been systematically examined. Yet, with the rapid accumulation of small RNA-seq data (such as, the US National Institutes of Health

[1]Department of Internal Medicine, Hematology/Oncology Division, University of Michigan, Ann Arbor, Michigan, USA. [2]Pacific Northwest Research Institute, Seattle, Washington, USA. [3]Lung Biology Center, Department of Medicine, University of California San Francisco, San Francisco, California, USA. [4]Department of Obstetrics, Gynecology, and Reproductive Sciences and Sanford Consortium for Regenerative Medicine, University of California, San Diego, La Jolla, California, USA. [5]Department of Medicine, Division of Cardiovascular Medicine, University of Massachusetts Medical School, Worcester, Massachusetts, USA. [6]Neurogenomics, The Translational Genomics Research Institute (TGen), Phoenix, Arizona, USA. [7]Department of Medicine, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, Massachusetts, USA. [8]Center for Cancer Computational Biology, Dana-Farber Cancer Institute, Boston, Massachusetts, USA. [9]Institute for Systems Biology, Seattle, Washington, USA. [10]Department of Biochemistry & Cell Biology, Faculty of Veterinary Medicine, Utrecht University, Utrecht, the Netherlands. [11]Leiden Genome Technology Center, Department of Human Genetics, Leiden University Medical Center, Leiden, the Netherlands. [12]Center for Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, Michigan, USA. [13]Department of Biostatistics, University of Michigan, Ann Arbor, Michigan, USA. [14]Cardiovascular Research Institute and the Department of Medicine, Division of Pulmonary, Critical Care, Sleep, and Allergy, University of California San Francisco, San Francisco, California, USA. [15]Department of Biomedical Engineering, University of Michigan, Ann Arbor, Michigan, USA. [16]Biointerfaces Institute, University of Michigan, Ann Arbor, Michigan, USA. [17]These authors contributed equally to this work. Correspondence should be addressed to M.D.G. (giraldezjimenez@hotmail.com), D.J.G. (dgalas@pnri.org) or M.T. (mtewari@med.umich.edu).

Received 24 February 2017; accepted 1 May 2018; published online 16 July 2018; corrected online 31 July 2018; doi:10.1038/nbt.4183

(NIH) short-reads archive[27,28], EV-associated small RNA sequencing databases[29–31], The Cancer Genome Atlas (TCGA)[32], the exRNA Atlas[33], etc.), meaningful, quantitative interpretation of results, especially across studies, would benefit from a systematic examination of technical bias, its effects on accuracy and of the reproducibility of small RNA-seq.

Here we report the results of a study led by investigators from the NIH-funded Extracellular RNA Communication Consortium[34] involving nine laboratories, in which a systematic multi-protocol, multi-institution assessment was carried out to assess the accuracy, reproducibility and technical bias of small RNA-seq using standardized, synthetic reference reagents as well as biologically derived reference RNA. We also evaluated the performance of different protocols with respect to characterizing miRNA editing and identified a library preparation approach that reduces technical bias and improves the accuracy and comparability of small RNA-seq results.

RESULTS

Study design and standard reference materials for miRNA quantification

To evaluate the performance of small RNA-seq library preparation protocols across multiple laboratories, we developed standard

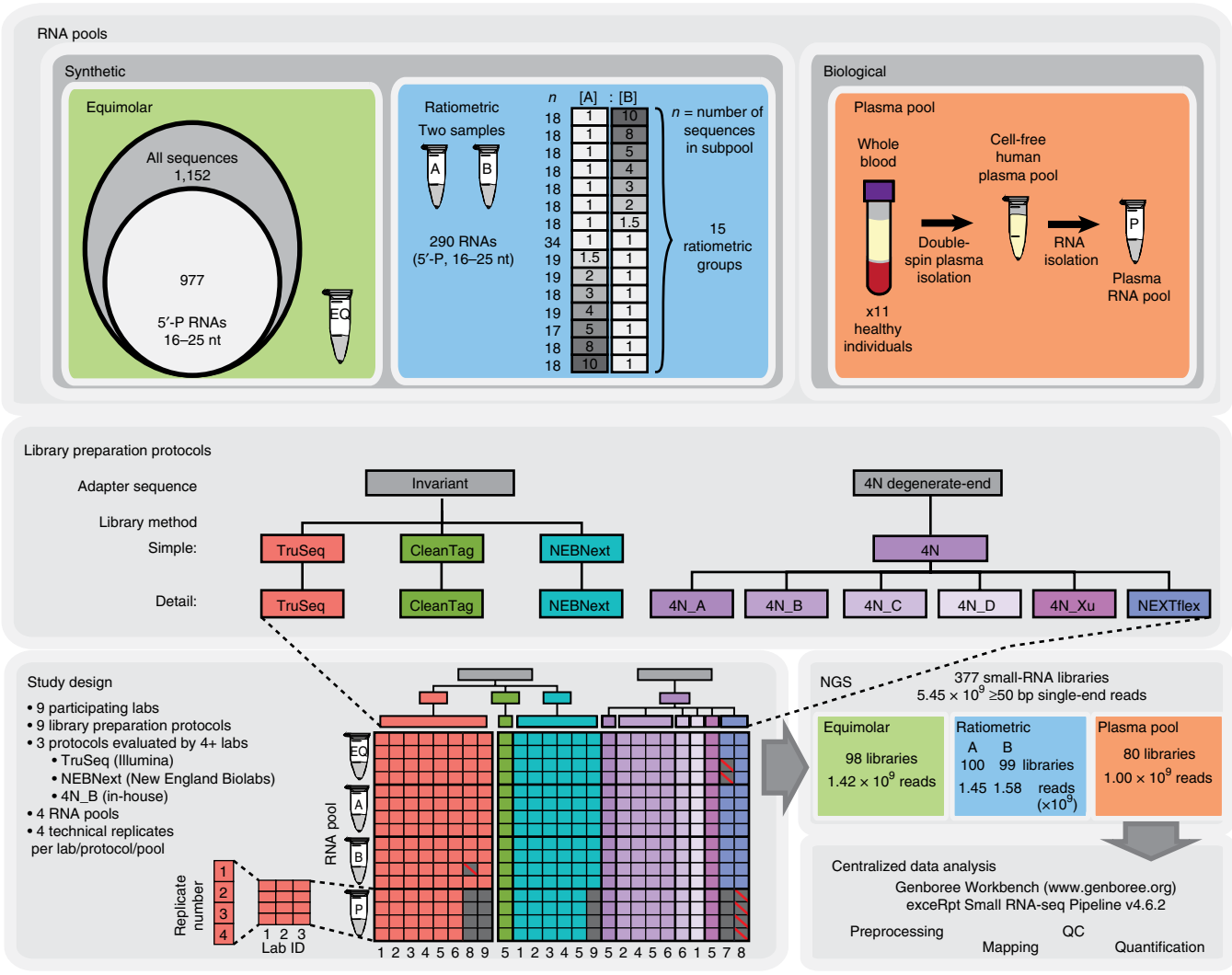


Figure 1 Overview of study design. Top, the four primary RNA pools used as common reference samples in the study are shown. Equimolar and ratiometric pools were prepared using chemically synthesized RNA oligonucleotides to establish 'ground-truth' knowledge of absolute and relative abundances, respectively. The equimolar pool consisted of 1,152 synthetic RNAs (15–90 nt) mixed at equal concentration. Downstream analyses focused on the subset of 977 16–25-nt-long RNAs with 5′-phosphorylation. The two ratiometric pools, A and B, consisted of 334 synthetic RNAs, in which subsets of RNAs were varied in relative abundance between the two pools. The RNAs were divided into 15 ratiometric subgroups. The subset of 290 16–25-nt-long RNAs were used for downstream analyses, and the number of RNAs in each ratiometric subgroup is indicated. The plasma RNA pool comprised RNA from 11 healthy males that was centrally isolated and distributed to the participating labs. Middle, nine different library preparation protocols were tested. Three commercially available kits with 'invariant' adapters and six 4N degenerate-end protocols were tested. Bottom, common reference RNA pools were distributed to nine participating labs for sequencing in quadruplicate, using a standardized common protocol (TruSeq) and at least one additional method of their choice. The breakdown of the resulting libraries is depicted in the colored grid, with the Lab IDs indicated by columns, and the replicates and pools shown in rows. Solid gray blocks indicate libraries that were not attempted. Gray blocks with a diagonal red line indicate samples in which library preparation and sequencing were attempted, but were unsuccessful. Sequencing was performed by each lab independently.

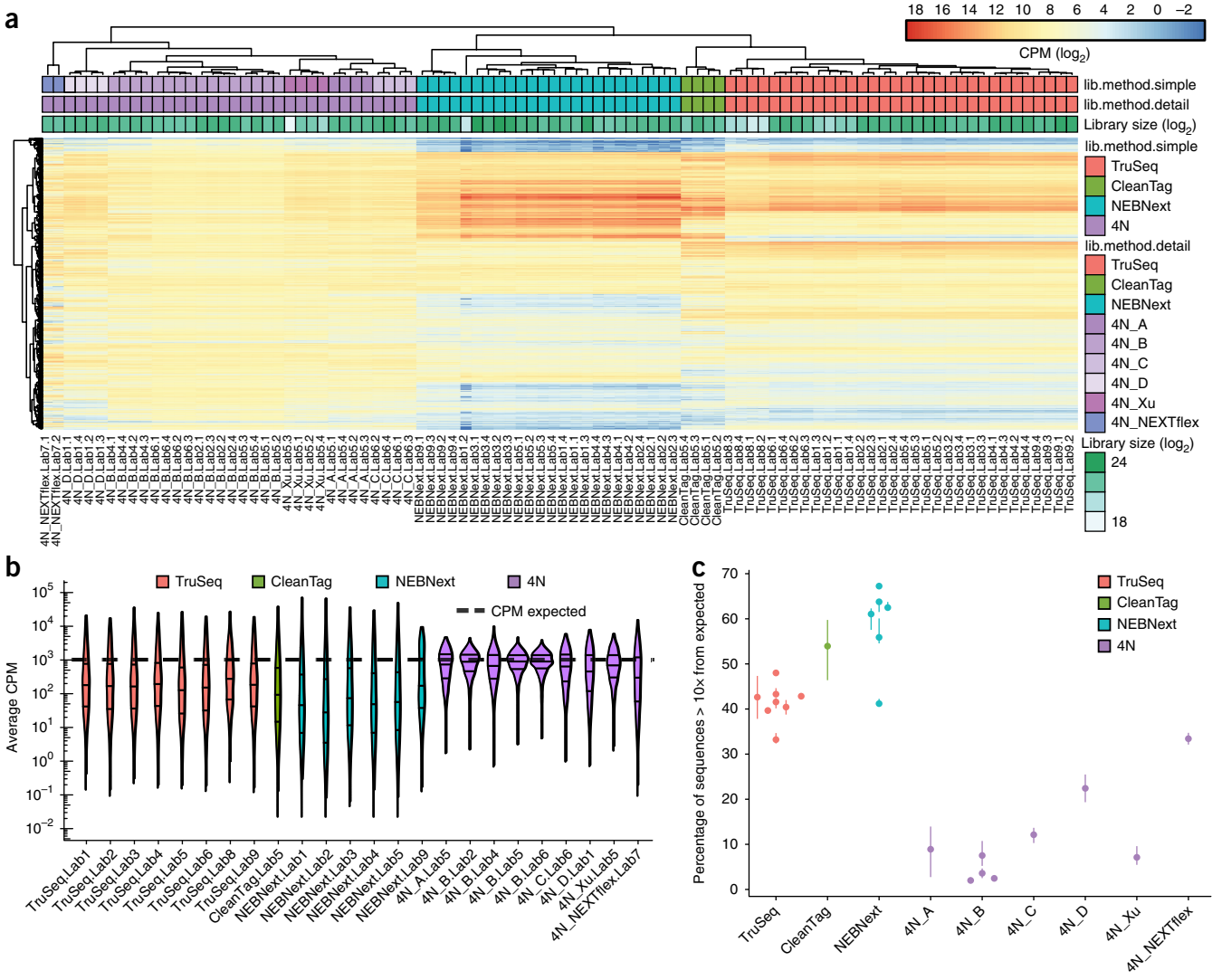


Figure 2 Equimolar pool sequencing results across multiple labs and protocols. (**a**) The heatmap shows expression levels for each synthetic RNA sequence (rows) across all replicate equimolar pool libraries (columns). Expression levels represent $\log_2$-scaled CPM calculated for 977 equimolar pool sequences, 16–25 nt in length and 5′-phosphorylated. Hierarchical clustering for rows and columns represents complete linkage clustering on Euclidean distances (the default setting for the R package, pheatmap, used for plotting). Columns are labeled at the bottom to identify replicate samples. Library size indicates the sequencing depth for each library ($\log_2$ scaled). (**b**) Violin plots summarize the mean CPM observed for each of the 16–25-nt, 5′-phosphorylated equimolar pool sequences (*y* axis, $\log_{10}$ scaled), as measured from equimolar pool libraries prepared by different institutions and using different library preparation protocols (*x* axis). The width of the violins is proportional to the density of data points at each position. The horizontal lines in each violin represent the 25th, 50th and 75th percentiles. The dashed horizontal line shows the expected CPM for each sequence in the equimolar pool (10^6 / 977 miRNAs = 1,023.5 CPM). Each violin plot and corresponding quantile lines summarize mean CPM values for $n = 977$ distinct equimolar pool sequences. The mean CPM values were calculated from $n = 4$ technical replicate libraries for each lab/library preparation method shown, with the exception of 4N_NEXTflex.Lab7 ($n = 2$). (**c**) The percentage of equimolar pool sequences sequenced at levels 10× higher (>10,235 CPM) or 10× lower (<102.35 CPM) than expected (*y* axis) are plotted for each lab. The dots and whiskers indicate the median and range of values, respectively, measured across the technical replicates for each lab. $n = 4$ technical replicate libraries per lab/library preparation method, with the exception of 4N_NEXTflex ($n = 2$).

reference samples as well as a standardized study design (**Fig. 1**). We distributed detailed instructions for library preparation and sequencing to each lab, along with four reference RNA samples (**Fig. 1** and **Supplementary Tables 1** and **2**): one equimolar pool comprising 1,152 synthetic RNA oligonucleotides, corresponding predominantly to human miRNA sequences, as well as a small set of non-miRNA oligonucleotides of varied sequence and length (15–90 nt); two synthetic small RNA pools, called ratiometric pools SynthA and SynthB, each containing the same 334 synthetic RNAs, but in which subsets of RNAs vary in relative amount between pools A and B by 15 different ratios, ranging from 10:1 to 1:10; and an RNA pool isolated from human blood plasma from 11 individuals.

The common materials were distributed to nine participating research groups (L.C. Laurent, University of California at San Diego; D.J. Erle, University of California at San Francisco; I. Ghiran/Y.E. Wang, Beth Israel Deaconess Medical Center/Dana-Farber Cancer Institute (BIDMC/DFCI); E.N.M. Nolte-'t Hoen, University of Utrecht, the Netherlands (UUTR); J.E. Freedman, University of

Massachusetts; K.Wang, Institute for System Biology (ISB); D.J. Galas, Pacific Northwest Research Institute (PNRI); K. Van Keuren-Jensen, TGen; M. Tewari, University of Michigan). Nine library preparation protocols were evaluated (Online Methods), in which at least one group prepared and sequenced quadruplicate libraries from each of the reference samples. Three of the protocols, TruSeq (Illumina), NEBNext (New England BioLabs) and CleanTag (Trilink Biotech), are commercial kits that employ adapters with invariant sequences. The remaining protocols make use of adapters with four degenerate nucleotides at the ligation ends as a strategy to reduce the bias, and we collectively refer to these as '4N' protocols. These six 4N protocols included: a commercial kit, NEXTflex (Bioo Scientific); a recently published protocol (4N_Xu)[35]; and four variants of a protocol developed by members of the consortium (protocols 4N_A, 4N_B, 4N_C and 4N_D) that we collectively refer to as 'in-house' 4N methods. The TruSeq kit served as the common reference kit for this study and was evaluated by all the groups using Illumina sequencing platforms (eight of nine groups). In addition, multiple laboratories generated libraries using the NEBNext kit (six labs) and the in-house protocol 4N_B (four labs), thereby allowing for standardized cross-lab comparisons for these two protocols, as well as to the Illumina TruSeq protocol.

In total, the nine participating groups prepared 384 libraries for miRNA quantification analysis, of which 377 (98%) were successfully sequenced and submitted for central analysis. The seven libraries that were not successfully prepared and sequenced included four plasma pool libraries (Lab8 4N_NEXTflex), two equimolar pool libraries (Lab7 NEXTflex) and one SynthB library (Lab8 TruSeq). Together, the nine participating groups collectively contributed 5.45 billion small RNA-seq reads to the analysis (**Fig. 1**). These sequencing data were centrally analyzed using the Genboree Workbench and its implementation of the Extracellular RNA Communication Consortium's exceRpt Small RNA-seq pipeline, which is specifically designed for the analysis of small RNA-seq data and uses its own alignment and quantification engine to map and quantify a range of RNAs represented in small RNA-seq data (see **Supplementary Table 3** for pipeline quality control (QC) metrics). Of the 377 samples analyzed, 364 (>96%) satisfied the minimum quality criteria (Online Methods) and were included in the analyses.

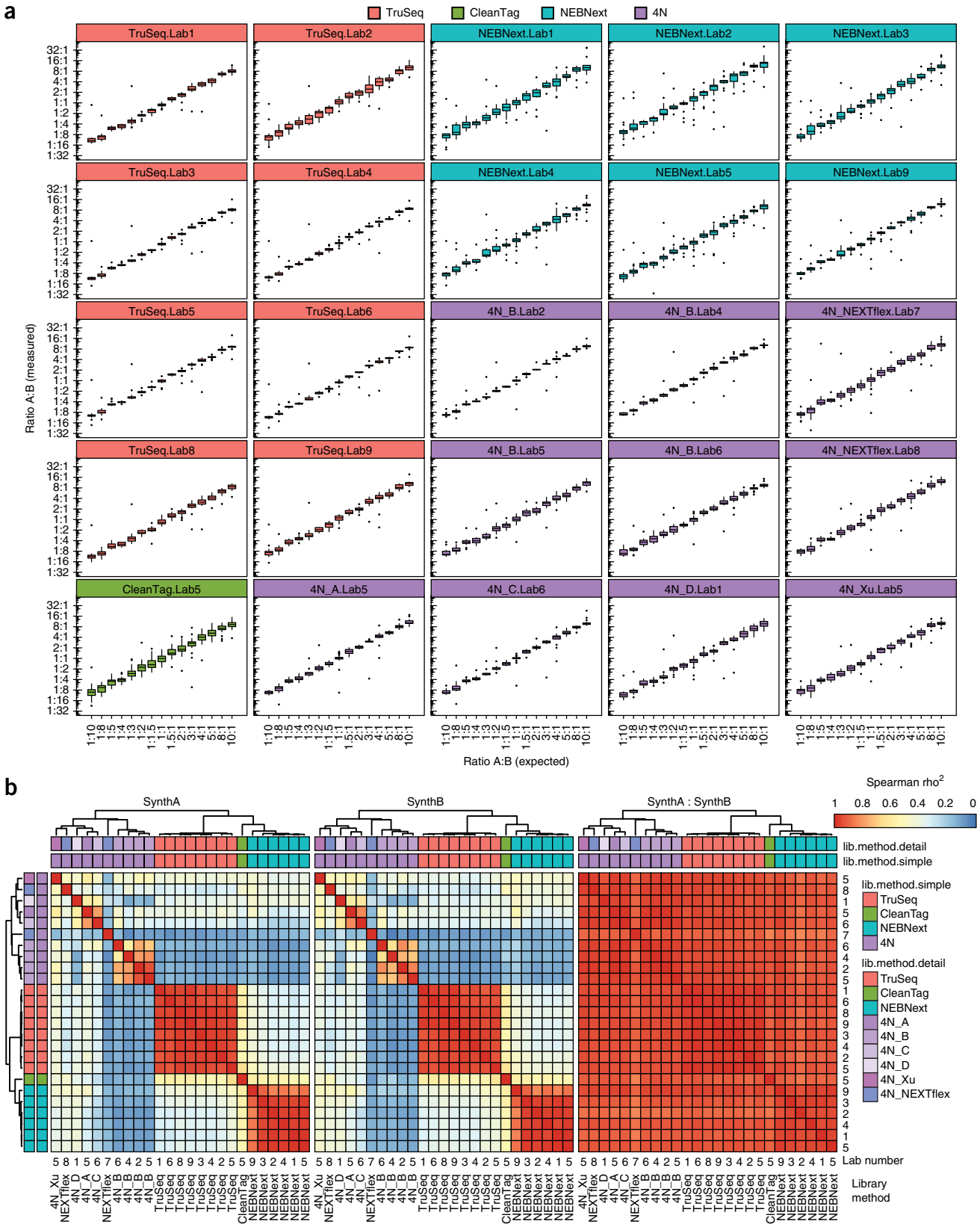
Characterization of sequence-specific bias of small RNA-seq protocols

Of the 1,152 synthetic RNAs in our equimolar pool, we focused on 977 5′-phosphorylated 16–25-nt-long RNAs, which can be captured with standard small RNA-seq protocols. The efficiency of recovery of RNA sequences varied by multiple orders of magnitude depending on the protocol, confirming that small RNA-seq protocols are associated with prominent sequence-dependent bias[4,25,36–38] (**Fig. 2a**,**b**) and that the bias is greater than that in long RNA-seq[26]. This was highly reproducible in a given protocol, both across technical replicates and laboratories using the same protocol (**Fig. 2a**). Libraries prepared by different labs clustered first into two groups, corresponding to methods with invariant (TruSeq, NEBNext and CleanTag) or degenerate (4N) adapters. In each of these two larger groups, the libraries then formed distinct clusters corresponding to the different protocols included in the study, indicating that the effect of the protocol bias is potentially greater than that of lab-to-lab variation. Consistent with this result, the ten most overrepresented and underrepresented sequences varied widely between protocols (**Supplementary Fig. 1**).

Although all of the protocols exhibited some bias, it was reduced in those using degenerate adapters (**Fig. 2b** and **Supplementary Fig. 2**). As one measure of this, we calculated the median percentage of sequences with a number of reads in counts per million (CPM) more than ten times above or below the expected value, for each protocol. This ranged from 41.6 to 61.5% for protocols using adapters with defined sequences (TruSeq, 41.6%; CleanTag, 53.9%; NEBNext, 61.5%), and from 2.8 to 22.4% for protocols using adapters with degenerate nucleotides (4N_A, 8.9%; 4N_B, 2.8%; 4N_C, 12.1%; 4N_D, 22.4%; 4N_Xu, 7.1%; 4N_NEXTflex, 17%) (**Fig. 2c**). The in-house 4N protocols showed fewer missing sequences from the equimolar pool (**Supplementary Table 4**) and when downsampling to compare the same number of total mapped reads across protocols at varying sequencing depths (**Supplementary Fig. 3**). We found that, with the in-house 4N_B protocol, even when downsampling to 10,000 total mapped reads, >90% of the miRNAs had a high probability of detection (median, 92%; range, 78–95%). In contrast, even with the best-performing invariant adapter protocol, TruSeq, <50% of miRNAs had a high probability of detection (median, 46%; range, 40–55%) at the same depth, indicating that the 4N_B protocol may require lower read depth to yield similar coverage as other library protocols.

We also assessed the reproducibility of small RNA cloning biases across labs by examining the rank-order of RNA sequence abundance. To do so, we calculated Spearman rank correlations for the equimolar synthetic pool counts between labs and protocols. As expected, the strongest correlations were found between technical replicates from the same lab and method (**Supplementary Fig. 4**). Correlations were also strong between samples generated by different labs using the same protocol (**Supplementary Table 5**). The somewhat lower correlation value observed for 4N_B (combined Rho value: 84%; top/bottom 2%: 0.66/0.95) can be attributed to the overall lower variation in read counts across miRNAs resulting from less cloning bias with this protocol. The reduced spread in the data limits the maximum absolute correlation coefficient values that can

Figure 3 Small RNA-seq accuracy and cross-protocol concordance in measuring relative expression levels between samples. (**a**) Boxplots show the observed ratio (*y* axis; $\log_2$ scale) versus expected ratio (*x* axis) for miRNAs present in each of the SynthA and SynthB synthetic RNA subpools. Observed ratios for each miRNA were calculated as mean CPM of SynthA / mean CPM of SynthB across technical replicates for each lab and library prep method. Boxes show the median + IQR; upper/lower whiskers indicate the smallest/largest observation less than or equal to 1st/3rd quartile ± 1.5 * IQR; outliers are calculated as being <1st quartile – 1.5 * IQR or >3rd quartile + 1.5 * IQR. Mean CPM ratios were calculated from $n = 4$ SynthA and $n = 4$ SynthB technical replicate libraries for each lab and library preparation method shown, with the exception of TruSeq Lab8 SynthB ($n = 3$). Those miRNAs with a mean CPM of 0 in SynthA or SynthB are not plotted. The numbers of miRNA not plotted are as follows: Truseq Labs, 1, 2, 3, 4, 5, 6 and 8: 1; Lab9: 0; CleanTag Lab5: 0; NEBNext Labs, 1, 3, 5 and 9: 0; Lab4: 1; Lab2: 3; 4N_NEXTflex Lab7: 1; other 4N: 0. The number of sequences represented in each boxplot is provided in **Supplementary Table 10**. (**b**) Heatmaps show the pairwise, squared Spearman rank correlation coefficients from sequencing the SynthA and SynthB pools. Pairwise correlation coefficients were calculated on the basis of the mean CPM across technical replicates for SynthA samples (left), SynthB samples (middle) and the ratio of SynthA: SynthB (right). The mean CPM value for each ratiometric pool sequence was calculated from $n = 4$ technical replicate libraries per lab, library preparation method and pool. Mean CPM values for $n = 290$ ratiometric pool RNAs were used for calculating each pairwise correlation coefficient. Hierarchical clustering for rows and columns is the same for all heatmaps, and is based on the average pairwise Euclidean distances calculated from the SynthA CPM and SynthB CPM correlation matrices. Column labels indicate the lab ID and library prep method; row labels indicate only lab ID, but are presented in the same order (top to bottom) as columns (left to right).



be obtained. This limitation notwithstanding, comparison across labs using different protocols showed much weaker correlations (**Supplementary Table 5**).

To dissect the source of observed bias, we evaluated the effect of several variables (5′ or 3′ terminal bases, %GC of the four 5′ or 3′end bases, overall %GC, dG [free energy], dH [enthalpy], dS

[entropy] and Tm [melting temperature]) on the number of obtained reads with different library preparation protocols (**Supplementary Figs. 5–13**). However, none of these variables substantially explained the observed bias.

Accuracy and cross-protocol concordance for relative quantification

To investigate the accuracy of relative quantification of the same small RNAs between different samples, we designed two ratiometric pools, SynthA and SynthB, each containing the same 334 synthetic RNA sequences, but varying the relative abundance of sequences between the two pools for 15 expression ratios (**Fig. 1** and **Supplementary Table 2**). All of the protocols that were tested showed close concordance between observed and expected ratios (**Fig. 3a**). We also analyzed the data using standard differential expression workflows from three commonly-used R packages (EdgeR[39,40], DESeq2 (ref. 41) and limma/voom[42]) to determine the smallest difference in abundance that could be distinguished using small RNA-seq methods. We observed that for most protocols and for the majority of miRNAs, a difference in levels of as little as 1.5-fold between the two samples could be detected (**Supplementary Fig. 14**). As shown in **Supplementary Table 6**, all of the evaluated protocols performed relatively well in detecting miRNA abundances.

We examined the rank-order of RNA sequence abundance and found that, in general, the Spearman rank correlations results obtained for the SynthA and SynthB samples were similar to those obtained for the equimolar pool: the correlation was strong when using the same protocol, but weaker across different protocols (**Fig. 3b**). In contrast, when we analyzed the concordance of the SynthA/SynthB ratios (**Fig. 3b**), we found a very strong correlation between labs not only when using the same protocol, but also across different protocols, confirming that relative quantification is resilient to variation in the protocol used (**Supplementary Table 5**).

Reproducibility of small RNA-seq protocols

To quantify intra-lab variation for each sequence, we used two metrics: the coefficient of variation (CV, 100* s.d./mean) and the quartile coefficient of dispersion (QCD, interquartile range/average of the first and third quartile). The median CV for the equimolar pool libraries ranged from 6.18% (TruSeq) to 23.92% (CleanTag) for the different library preparation methods (**Fig. 4a** and **Supplementary Table 5**). In addition, the median QCD was <0.1 for all the protocols/labs (**Fig. 4a** and **Supplementary Table 5**). We also evaluated the intra-lab variation from technical replicates of sequencing the SynthA and SynthB libraries. The calculated CV and QCD values were similar to those observed for the equimolar libraries (**Supplementary Fig. 15**).

To characterize the reproducibility of small RNA-seq libraries across laboratories, we focused on the three protocols (TruSeq, NEBNext and 4N_B) for which libraries were generated by at least three groups. In addition, of the six labs that generated libraries using the NEBNext protocol, two of the labs used somewhat modified conditions based on options provided by the manufacturer and were excluded from the analysis (Online Methods).

Using the results for the equimolar pool and treating each laboratory's results as one trial of the experiment, we calculated the CV and QCD for the mean CPM values for each RNA sequence across laboratories. The median CV across labs ranged from 30.42 (4N_B) to 35.28% (NEBNext) and the median QCD from 0.13 (4N_B) to 0.18 (TruSeq and NEBNext) (**Fig. 4b** and **Supplementary Table 5**). We confirmed that the choice of pseudo-counts for calculating CPM did not appreciably alter the CV and QCD distribution (**Supplementary Fig. 16).** In addition, repeating the inter-lab CV and QCD calculations using all combinations of $n = 3$ labs from the TruSeq, NEBNext and 4N_B equimolar pool libraries showed that results from analysis of subsets of the data were comparable to those from analysis of the full data sets (**Supplementary Fig. 17a**,**b**). We also calculated across lab variation for the SynthA and SynthB pools individually and obtained median CV and QCD values that were comparable to those described for the equimolar libraries (**Supplementary Table 5**).

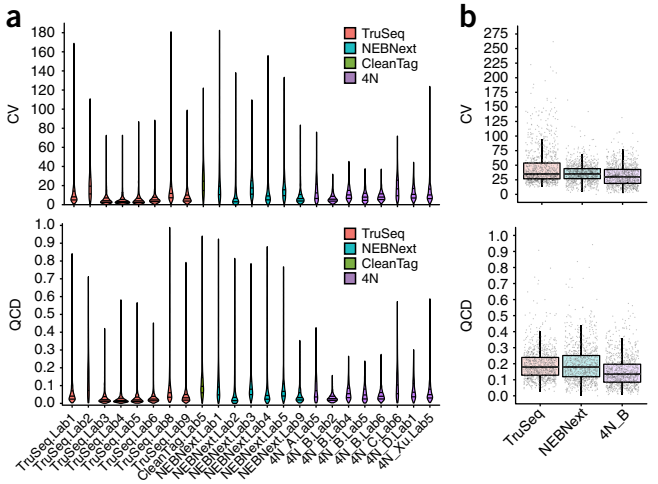


Figure 4 Reproducibility of small RNA-seq within and between labs. (**a**) Violin plots summarize the technical reproducibility of quantification for all equimolar pool sequences, as calculated from each lab and library preparation method. Reproducibility measurements, CV (top) and QCD (bottom) were calculated from CPM values. Horizontal lines in each violin indicate the 25th, 50th and 75th percentiles, calculated from the mean CPM values of $n = 977$ equimolar pool RNA sequences. Mean CPM values were calculated from $n = 4$ technical replicate libraries for each of the lab/library preparation methods shown, with the exception of TruSeq Lab1 ($n = 3$). (**b**) Boxplots summarize the sequence-specific reproducibility of quantification measured in equimolar pool libraries generated by different labs using the same protocol. CV (top) and QCD (bottom) values were calculated for each equimolar pool sequence across TruSeq ($n = 8$ labs), NEBNext ($n = 4$ labs) and 4N_B ($n = 4$ labs) library preparation protocols. The mean CPM for each sequence across technical replicates ($n = 4$ technical replicates per lab/library preparation method) was used to calculate the between-lab CV and QCD plotted here. Boxplot statistics and outliers were calculated from CV or QCD values for $n = 977$ equimolar pool sequences. The overlaid boxes indicate the median and IQR. Whiskers represent the 1st/3rd quartile ± 1.5 * IQR. Outliers are <1st quartile – 1.5 * IQR or >3rd quartile + 1.5 * IQR.

Performance of small RNA-seq protocols using biological samples

We also sought to characterize the performance of small RNA-seq protocols across labs using standard reference RNA derived from biological material to assess the reproducibility and the diversity of miRNA sequences recovered.

To perform this analysis, we shipped aliquots of RNA extracted from a pool of human blood plasma from 11 donors to the participating labs for sequencing in quadruplicate (**Fig. 1**). We focused our analysis on miRNAs because they are well-characterized and have

been extensively studied in human plasma[43]. Hierarchical clustering generally mirrored that from the synthetic pools, with technical replicates of the same protocol clustering most closely together (**Fig. 5a**), and with samples also broadly clustering according to library preparation protocol.

To evaluate the intra-lab reproducibility of plasma small RNA-seq, we calculated the CV and QCD for individual miRNA sequences across technical replicates in each lab (**Fig. 5b**). After applying minimum CPM filtering criteria as before, to focus on reliably detected miRNAs (Online Methods), we found that the median CV across the miRNAs analyzed ranged from 7.7 (TruSeq) to 24.9% (CleanTag) for different protocols. Although this degree of reproducibility seems comparable to that observed with the synthetic reference pool RNA (**Fig. 4** and **Supplementary Table 5**), it is important to note that the filtering criteria used for plasma sequencing data were different (and generally more stringent) than for the synthetic RNA sequencing data. In addition, the median QCD was ≤0.1 for all the protocols (**Supplementary Table 5**).

Unsupervised clustering of the plasma miRNA expression data revealed clear groups separated by preparation protocol (TruSeq, NEBNext and 4N_B), with results obtained from different labs using the same protocol clustering together (**Fig. 5a**). The median variability across labs measured using CV ranged from 25.7 (4N_B) to 32.9% (TruSeq) and using QCD was <0.3 for all protocols. (**Fig. 5c** and **Supplementary Table 5**). The overall reproducibility of small RNA-seq using RNA isolated from biological samples was therefore comparable to that observed using the synthetic reference RNA samples.

To assess differences between protocols in the diversity of miRNA sequences recovered from the standard reference plasma RNA, we performed an analysis of the number of miRNAs detected by each protocol in which we plotted data from different in-house 4N protocols as one group for the sake of comparison. This was done using downsampled data sets so the same total number of mature miRNA-mapping reads could be compared across protocols, at varying sequencing depths. The in-house 4N protocols recovered a larger number of miRNAs than those using defined adapter sequences (**Fig. 5d**). In addition, an indirect assessment of miRNA diversity (that is, percent of total reads accounted for by the ten most abundant miRNAs) was consistent with the conclusion that 4N protocols generate a more diverse profile of miRNAs (**Supplementary Fig. 18**).

Evaluation of small RNA-seq in miRNA A-to-I editing

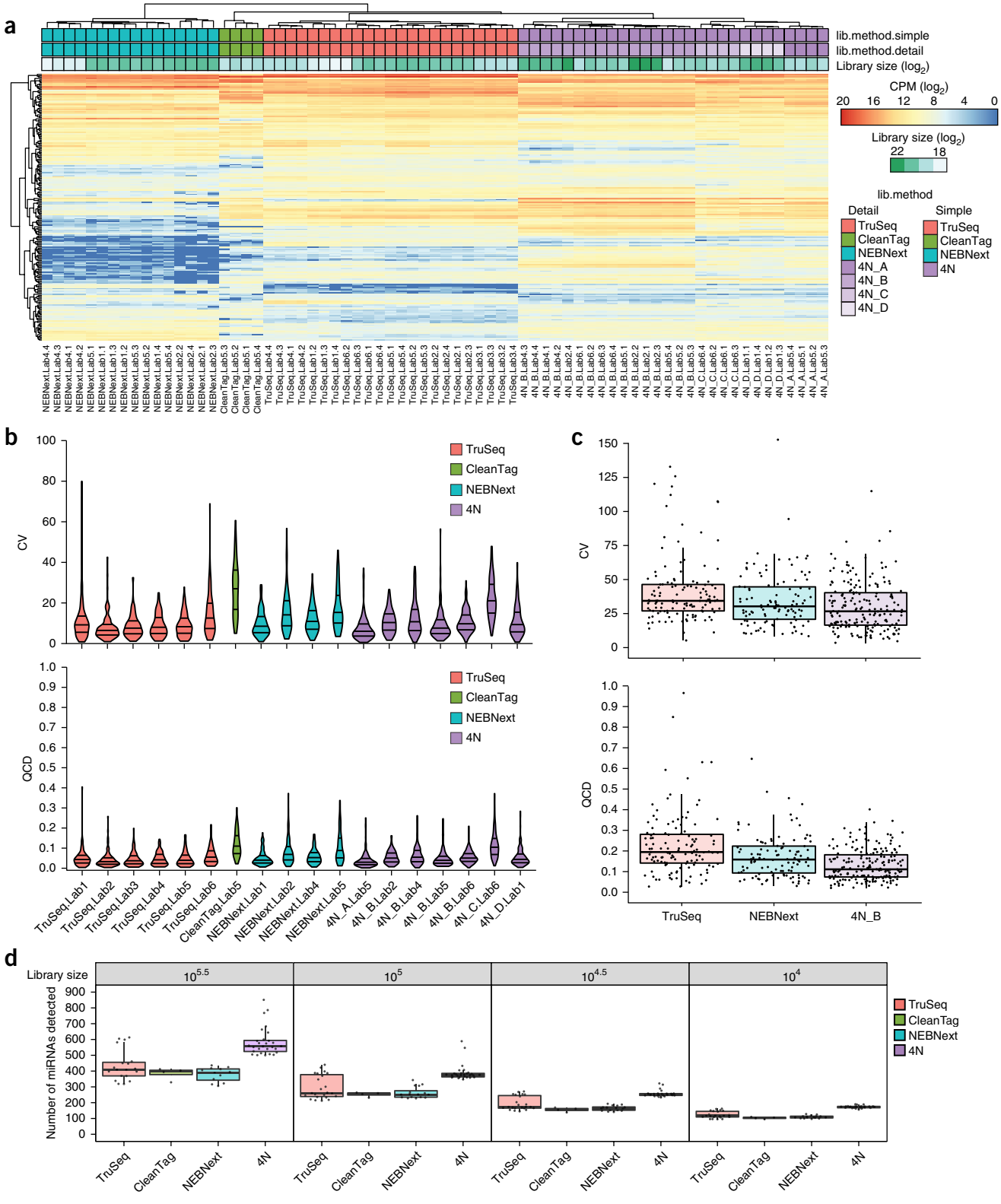
We extended our study to evaluate the performance of different protocols for quantifying sequences exhibiting adenosine to inosine (A-to-I) miRNA editing. This naturally occurring RNA modification can alter both miRNA biogenesis and regulatory functions[44,45]. We designed six synthetic RNA pools, each containing ten miRNAs that have previously been reported to undergo A-to-I editing[46,47]. Each pool combined the unedited (A) and edited (I) miRNA variants in different ratios (that is, 0, 0.1, 0.5, 5, 50 and 100% edited). Each of these mixtures was then combined with a background of 277 different, unedited human miRNAs to increase complexity in the pools (**Fig. 6a** and **Supplementary Table 7**). The six pools were sequenced by three different labs, each in triplicate, using TruSeq, NEBNext and in-house 4N_B protocols (**Fig. 6a**). The resulting 162 libraries yielded 1.42×10^9 reads aligned to editing pool sequences in total, with a median library size of 8.22×10^6 (range: 1.74×10^6 to 29.01×10^6). All 162 libraries satisfied minimum quality criteria (Online Methods and **Supplementary Table 3**).

To determine the accuracy of quantifying miRNA editing in our six synthetic pools, we compared the number of reads observed for the A and I variant oligos in each library with the expected abundance based on the known composition of the pools. Inaccurate and widely varying estimates of editing levels were apparent for many miRNAs using the NEBNext and TruSeq protocols, especially for the 1, 5 and 50% editing pools (**Fig. 6b** and **Supplementary Table 8**). In contrast, the in-house protocol, 4N_B, proved more accurate for detecting editing levels ≥1%. For example, in the 50% editing pool, where the edited and unedited forms of each miRNA are present at equivalent levels, the mean percent editing observed ranged from 19–98% and 5–95% for the TruSeq and NEBNext libraries, respectively, whereas estimates were all within 10% of the expected value (43–53%) for the 4N_B protocol.

Aside from accuracy, we calculated across-lab reproducibility (that is, precision) of the measured edited fraction in each pool for each of the evaluated protocols using CV and QCD, which are most meaningful where there are reads in both edited and unedited categories (**Supplementary Table 8**). We found that precision varied as a function of known percent editing, with greater precision being observed in the 5 and 50% edited pools compared with the 0.1 and 1% pools, as expected from the higher number of edited read counts in the former pools. Across all of the tested protocols, for the majority of miRNAs, the precision of percent editing measurements was CV <5 % in the 50% edited pool, <20% in the 5% edited pool and <25% in the 1% edited pool, and QCD < 0.3 in the 50% edited pool, <0.4 in the 5% edited pool and <0.6 in the 1% edited pool.

We evaluated the specificity and limit of detection for identifying miRNA editing by downsampling each library to 10^6 reads to allow

Figure 5 Small RNA-seq of reference plasma RNA by multiple laboratories using multiple library preparation protocols. (**a**) The heatmap shows CPM ($\log_2$ scale) for each sequence (rows) across plasma pool libraries (columns). Only mature miRNAs with a high confidence of detection are shown, requiring a minimum of 100 CPM in 90% of samples from at least one protocol (TruSeq, CleanTag, NEBNext or 4N). Hierarchical clustering for rows and columns represents complete linkage clustering on Euclidean distances. Library size indicates the sum of the mature miRNA-mapped read counts before filtering for the individual libraries ($\log_2$ scaled). (**b**) Violin plots summarize the technical reproducibility of quantification for miRNAs expressed in plasma pool libraries, as calculated from each lab and library preparation method. Reproducibility measurements, percent CV (top) and QCD (bottom), were calculated from CPM values. Horizontal lines within each violin indicate the 25, 50 and 75th percentiles, calculated from the mean CPM values of $n = 977$ equimolar pool RNA sequences. For TruSeq Lab1 $n = 3$ technical replicates; for all other lab/protocols $n = 4$. (**c**) Boxplots summarize the between-lab reproducibility for miRNAs expressed in the plasma pool libraries using TruSeq ($n = 6$ labs), NEBNext ($n = 4$ labs) and 4N_B ($n = 4$ labs) library preparation protocols. Each dot represents CV or QCD calculated across labs for a single miRNA. The between-lab CV and QCD were calculated using the mean CPMs for each sequence across technical replicates for each lab. The underlying boxes show the median and IQR. Whiskers represent the 1st/3rd quartile ± 1.5 * IQR. Outliers are <1st quartile – 1.5 * IQR or >3rd quartile + 1.5 * IQR. (**d**) Boxplots show the number of mature miRNAs detected by each protocol based on downsampling of data sets to the indicated sequencing depths. Each box summarizes number of miRNAs detected by each lab for the indicated protocol. The probability of each miRNA being detected was estimated for every sample randomly downsampled to 10^4, $10^{4.5}$, 10^5 or $10^{5.5}$ total read counts. A miRNA was only counted as detected if the probability of detection was at least 90%. Libraries with total counts less than the indicated sample size were excluded. Boxplots for 4N include only in-house 4N protocols (4N_A, 4N_B, 4N_C and 4N_D). The number of libraries summarized by each boxplot is as follows: $10^{5.5}$: TruSeq = 19; CleanTag = 4; NEBNext = 12; 4N = 28; 10^5, $10^{4.5}$ and 10^4: TruSeq = 23; CleanTag = 4; NEBNext = 16; 4N = 28. The underlying boxes show the median and IQR. Whiskers represent the 1st/3rd quartile ± 1.5 * IQR. Outliers are <1st quartile – 1.5 * IQR or >3rd quartile + 1.5 * IQR.



standardized comparisons across libraries. To calculate specificity, we first estimated the false positive frequency for each protocol by evaluating: the average percent edited reads observed in the 0% edited pool (that is, false positive edited read frequency) and the average percent unedited reads observed in the 100% edited pool (that is, false positive unedited read frequency). The overall median false positive rate was 0.1% across all protocols, all miRNAs, and both edited and unedited false positive calls (median false positive frequencies for individual protocols: TruSeq, 0.05% (edited) and 0.06% (unedited); NEBNext, 0.30% (edited) and 0.14% (unedited); 4N_B, 0.10% (edited) and 0.08%

(unedited); **Supplementary Table 8**). This corresponds to an overall specificity across all three protocols of 99.88% for calling unedited sequences and 99.91% for calling edited sequences (**Supplementary Table 8**). To calculate the limit of detection (LOD), we defined detection of editing as an observed edited count that is more than 3 s.d. above the observed edited count in the 0% edited synthetic pool. For all three protocols, the majority of miRNAs had a limit of detection at or below the 1% edited fraction, with a few miRNAs being detectable in the 0.1% edited pool (**Supplementary Table 8**). It is worth noting, however, that the LOD is expected to vary based on sequencing depth, sample complexity, relative abundance of the miRNA being studied and the pipeline used for analysis.

DISCUSSION

Our results quantitatively confirm that small RNA-seq is highly affected by sequence-related bias[4,36–38,48], which is largely protocol dependent. The observed biases were as large as 10^4-fold with some commonly used commercial library preparation protocols. This sequence-dependent bias is more severe than that previously reported for long RNA-seq[26], highlighting differences between the technologies and unique challenges involved in small RNA sequencing. In addition, this bias can be particularly vexing when working with low RNA input samples such as biofluids, preventing the reliable detection of some low-abundance small RNAs. The in-house 4N protocols that we evaluated, which employ adapters containing degenerate bases in the ligating ends, reduced the bias on the order of 100-fold and achieved better coverage at a lower sequencing depth than the widely used commercial library preparation kits with invariant adapter sequences. The magnitude of the bias observed for some sequences when using fixed adapter protocols was so high that it is unlikely to be overcome simply by increasing sequencing depth. There were, however, differences in the results of different 4N methods, suggesting that not only the use of adapters with degenerate bases, but also other factors in the protocols, such as the concentration of polyethylene glycol in ligation reactions, the time and temperature of ligations, etc., may also affect the bias. Our computational analyses of a range of sequence-related variables (for example, 5′ or 3′ terminal nucleotides, %GC of the four 5′ or 3′end nucleotides, overall %GC, dG [free energy], dH [enthalpy], dS [entropy] and Tm [melting temperature]) did not reveal strong associations, suggesting that the mechanistic basis of the bias may be complex.

Even using the best-performing 4N protocols, there is still considerable sequence-related bias, which precludes the use of read counts alone for accurate quantification of different small RNAs in a given sample. However, despite the observed biases, we found that small RNA-seq was consistently accurate for relative quantification of a given miRNA between samples, as long as the same library preparation protocol was used for the two samples being compared, which is consistent with previous observations for mRNA sequencing[26]. In this sense, all of the evaluated protocols were able to distinguish samples with as little as a 1.5-fold difference in relative abundance of most sequences examined, although the design of our ratiometric pools was such that differences smaller than 1.5-fold could not be assessed.

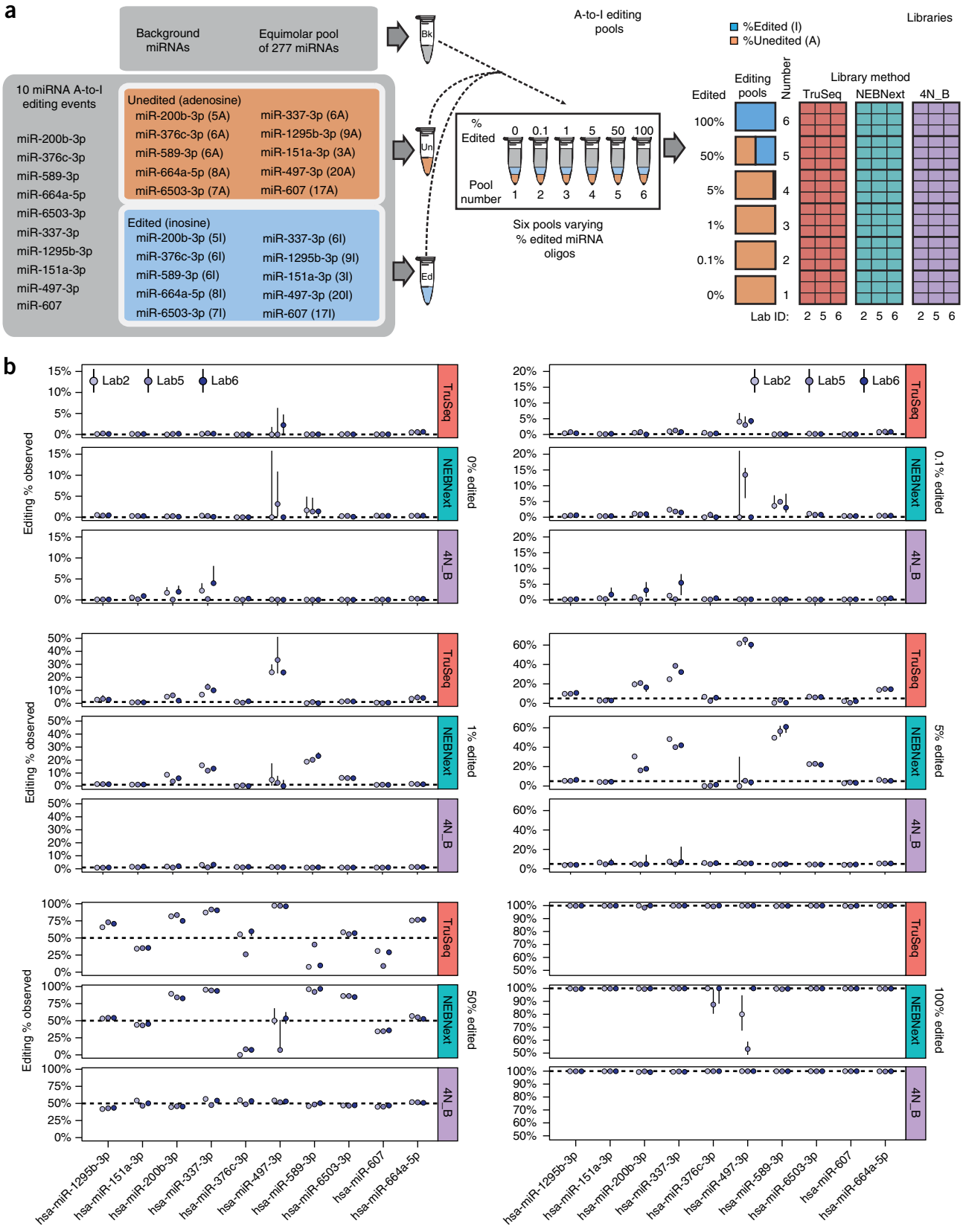
Reproducibility across laboratories is a crucial requirement for any experimental method used for research or clinical applications[49,50]. We found that for common commercial protocols as well as for our in-house 4N protocol, results were reproducible between labs with a CV ≤ 20% for most sequences. Moreover, when comparing relative quantification measurements obtained by small RNA-seq across labs, the results were highly concordant even when the centers were using different protocols.

The use of a diverse pool of synthetic RNAs allowed us uniquely to evaluate sequence-specific biases and accuracy because the 'ground truth' is known. Since biological material, with a wide range of RNA species and other macromolecules, could behave differently from the synthetic RNA pools, we also characterized the diversity of miRNAs captured in a common biological sample by each protocol. We found that the in-house 4N protocols detected a greater diversity of sequences than protocols using defined-sequence adapters. In addition, for a given protocol, the profile obtained for the biological sample was very reproducible between laboratories.

We believe that the data sets generated in this study can also serve as a valuable resource for benchmarking computational tools designed to facilitate and improve on RNA-seq analysis. This is important both for developing new software and for evaluating the suitability of using existing mRNA-seq algorithms for the analysis of small RNA-seq data sets. This could be particularly useful for benchmarking software developed to account for various technical biases found in mRNA-seq data[51–54], as our data suggest that such biases may be different in small RNA-seq data.

We also hope that our data may facilitate the development of computational approaches for normalization of data sets generated using different library preparation protocols. Although normalization algorithms are generally not intended to account for cross-platform variation, our preliminary analysis suggests that small RNA-seq protocol-specific biases largely correlate across samples. This suggests that one may be able to account for the protocol-specific differences in sequencing bias individually for each sequence, raising the possibility of cross-protocol data normalization. We performed an initial exploration of this concept using a simple approach for calculating correction factors (**Supplementary Note 1**, **Supplementary Table 9**, and **Supplementary Figs. 19** and **20**). Although this approach was able to make overall profiles from different protocols appear to be more similar to each other, its performance was not sufficient to be practically relevant at this time. We propose that synthetic RNA reference data, such as that generated here, can provide a foundation for the future development of more advanced computational approaches to enable accurate cross-protocol comparisons.

Figure 6 Library protocol performance in measuring miRNA A-to-I editing events. (**a**) A schematic depicting the experimental design for the miRNA A-to-I editing experiments. Left, ten miRNAs were synthesized with either an adenosine or inosine at a single position previously shown to be edited in human cells. The position, relative to the 5′ end of the mature miRNA, is indicated to the right of the respective miRNA IDs, along with the identity of the nucleotide. 277 other unedited human miRNAs were added at a fixed concentration to increase the background complexity of the pools. Middle, six different editing subpools were generated, using a constant amount of the background (Bk) pool and varying percentages of unedited (Un; adenosine) and edited (Ed; inosine) oligos in each pool. Right, the color-coded grid depicts the library design used in the A-to-I editing experiment. Specifically, the six editing pools were sequenced by three participating labs, using three different library preparation protocols, with each lab generating libraries in triplicate. (**b**) The observed percent editing (*y* axis) is shown for each miRNA in the six A-to-I editing pools, as measured by each of the three labs, using TruSeq, NEBNext and 4N_B protocols. The expected editing percent in each pool is both indicated to the right of each plot group and by the horizontal dotted line in each plot. The dots and whiskers represent the median and range of percent editing for each miRNA (*x* axis) as measured by the three labs. Individual miRNA percentage editing is shown for $n = 3$ technical replicate libraries for each lab and library preparation method.



We also assessed the ability of library protocols to measure miRNA A-to-I editing. Our results revealed that low bias protocols (that is, in-house 4N_B) quantify editing more accurately than protocols using defined adapter sequences (that is, TruSeq and NEBNext). It is worth noting that the accuracy of editing estimates can also be affected by low sequencing coverage. Indeed, some miRNAs had very low coverage

by at least one of the protocols, which contributed to the inaccuracy and variation in editing estimates. However, this lack of coverage is a consequence of technical biases in small RNA-seq, as 4N_B libraries all had sufficient coverage of each sequence and, at a minimum, all libraries had depth enough for ~6,000× coverage of each sequence in the pool. Thus, protocols with a higher degree of sequencing bias also have a greater potential for inaccurate estimates of editing levels, as a result of lower read coverage for some miRNAs and/or differential preferences based on a single base difference. This is relevant to miRNA editing estimates reported in the literature, given that prior studies have commonly used the more biased protocols with fixed sequence adapters[55].

METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We are grateful to J.S. Rozowsky, R. Kitchen, S.L. Subramanian, W. Thistlethwaite, M.B. Gerstein, A. Milosavljevic and N. Sakhanenko for facilitating access to the exceRpt pipeline and helpful conversations and suggestions. We are also grateful to P.A.C. 't Hoen for helpful discussions. We acknowledge funding support from the NIH Extracellular RNA Communication Common Fund grants: U01 grants HL126499 to M.T., HL126496 to D.J.G. and K.W., HL126493 to D.J.E. and P.G.W., HL126494 to L.C.L., HL126495 to J.E.F., HL126497 to I.G., and UH3 grant TR000891 to K.V.K.-J. M.D.G. acknowledges initial support from a Rio Hortega Fellowship (CM10/00084) and later from a Martin Escudero Fellowship. E.N.M.N.-`t.H. and T.A.P.D. received funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013) / ERC Grant Agreement number 337581 and from the Netherlands Organization for Scientific Research (NWO Enabling Technologies project nr 435002022). Y.E.W. received funding from the Dana-Farber Strategic Plan Initiative. K.W. received funding from DOD (W911NF-10-2-0111) and DTRA (HDTRA1-13-C-0055). Research reported in this publication was also supported by the National Cancer Institute of the NIH under Award Number P30CA046592 by the use of the following Cancer Center Shared Resource at the University of Michigan: DNA Sequencing. D.J.G. also acknowledges a special technology support award from the Pacific Northwest Research Institute to his lab. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

AUTHOR CONTRIBUTIONS

M.D.G. designed, led and coordinated the overall study. This comprised contributions throughout the entire process, including designing, preparing and distributing synthetic RNA pools, creating detailed instructions for the participating labs, performing experiments, coordinating experimental work and communication across labs, and organizing and managing all aspects of the project. In addition, she interpreted study results and was the primary writer of the manuscript. R.M.S. led the computational analyses and data management for this study, including processing data, designing and performing data analyses, identifying and applying methods to visualize data and results, and coordinating data and metadata incoming from collaborating laboratories. He also designed the composition of the ratiometric pools, interpreted results and contributed to the manuscript by preparing the figures, drafting figure legends and writing the computational methods. A.E. contributed to experimental design and preparation of synthetic RNA pools. He also performed experimental work, interpreted results and provided comments on the manuscript. He also developed and contributed the core in-house 4N protocol, variations of which were then used by multiple laboratories. M.T., D.J.G. and D.J.E. helped to design the study and interpret the results, along with contributions from the rest of the study team, including L.C.L., P.G.W., K.V.K.-J., I.G., Y.E.W., K.W., J.E.F., H.J., E.N.M.N.-`t.H. and H.P. J.B. contributed to design of statistical analyses and data interpretation. P.M.G., A.J.B., S.S., P.L.D.H., K.T., A.C., S.L., J.K., R.R., D.B. and T.A.P.D. carried out experiments. M.T. and D.J.G. supervised the overall study and did primary editing of the manuscript with substantial input from D.J.E. All of the authors contributed to reviewing, editing and/or providing comments on the manuscript.

COMPETING INTERESTS

The spouse of L.C. Laurent is an employee of Illumina, Inc., the manufacturer of the TruSeq Small RNA Library Preparation Kit. L.C. Laurent and her spouse's equity interest in Illumina, Inc. represents <<1% of the company. The other authors declare no competing financial interests.

1. Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* **5**, 621–628 (2008).
2. Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* **10**, 57–63 (2009).
3. Levin, J.Z. *et al.* Comprehensive comparative analysis of strand-specific RNA sequencing methods. *Nat. Methods* **7**, 709–715 (2010).
4. Jayaprakash, A.D., Jabado, O., Brown, B.D. & Sachidanandam, R. Identification and remediation of biases in the activity of RNA ligases in small-RNA deep sequencing. *Nucleic Acids Res.* **39**, e141 (2011).
5. Viollet, S., Fuchs, R.T., Munafo, D.B., Zhuang, F. & Robb, G.B. T4 RNA ligase 2 truncated active site mutants: improved tools for RNA analysis. *BMC Biotechnol.* **11**, 72 (2011).
6. Zhang, Z., Lee, J.E., Riemondy, K., Anderson, E.M. & Yi, R. High-efficiency RNA cloning enables accurate quantification of miRNA expression by deep sequencing. *Genome Biol.* **14**, R109 (2013).
7. Song, Y., Liu, K.J. & Wang, T.-H. Elimination of ligation dependent artifacts in T4 RNA ligase to achieve high efficiency and low bias microRNA capture. *PLoS One* **9**, e94619 (2014).
8. Baran-Gale, J. *et al.* Addressing bias in small RNA library preparation for sequencing: a new protocol recovers microRNAs that evade capture by current methods. *Front. Genet.* **6**, 352 (2015).
9. Sorefan, K. *et al.* Reducing ligation bias of small RNAs in libraries for next generation sequencing. *Silence* **3**, 4 (2012).
10. Bellingham, S.A., Coleman, B.M. & Hill, A.F. Small RNA deep sequencing reveals a distinct miRNA signature released in exosomes from prion-infected neuronal cells. *Nucleic Acids Res.* **40**, 10937–10949 (2012).
11. Nolte-'t Hoen, E. *et al.* Deep sequencing of RNA from immune cell-derived vesicles uncovers the selective incorporation of small non-coding RNA biotypes with potential regulatory functions. *Nucleic Acid Res.* **18**, 9272–9285 (2012).
12. Huang, X. *et al.* Characterization of human plasma-derived exosomal RNAs by deep sequencing. *BMC Genomics* **14**, 319 (2013).
13. Tietje, A., Maron, K.N., Wei, Y. & Feliciano, D.M. Cerebrospinal fluid extracellular vesicles undergo age dependent declines and contain known and novel non-coding RNAs. *PLoS One* **9**, e113116 (2014).
14. Lunavat, T.R. *et al.* Small RNA deep sequencing discriminates subsets of extracellular vesicles released by melanoma cells--Evidence of unique microRNA cargos. *RNA Biol.* **12**, 810–823 (2015).
15. Tosar, J.P. *et al.* Assessment of small RNA sorting into different extracellular fractions revealed by high-throughput sequencing of breast cell lines. *Nucleic Acids Res.* **43**, 5601–5616 (2015).
16. van Balkom, B.W.M., Eisele, A.S., Pegtel, D.M., Bervoets, S. & Verhaar, M.C. Quantitative and qualitative analysis of small RNAs in human endothelial cells and exosomes provides insights into localized RNA processing, degradation and sorting. *J. Extracell. Vesicles* **4**, 26760 (2015).
17. Burgos, K.L. *et al.* Identification of extracellular miRNA in human cerebrospinal fluid by next-generation sequencing. *RNA* **19**, 712–722 (2013).
18. Bahn, J.H. *et al.* The landscape of microRNA, Piwi-interacting RNA, and circular RNA in human saliva. *Clin. Chem.* **61**, 221–230 (2015).
19. Freedman, J.E. *et al.* Diverse human extracellular RNAs are widely detected in human plasma. *Nat. Commun.* **7**, 11106 (2016).
20. Wecker, T. *et al.* MicroRNA profiling in aqueous humor of individual human eyes by next-generation sequencing. *Invest. Ophthalmol. Vis. Sci.* **57**, 1706–1713 (2016).
21. Yuan, T. *et al.* Plasma extracellular RNA profiles in healthy and cancer patients. *Sci. Rep.* **6**, 19413 (2016).
22. Bullard, J.H., Purdom, E., Hansen, K.D. & Dudoit, S. Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. *BMC Bioinformatics* **11**, 94 (2010).
23. Risso, D., Ngai, J., Speed, T.P. & Dudoit, S. Normalization of RNA-seq data using factor analysis of control genes or samples. *Nat. Biotechnol.* **32**, 896–902 (2014).
24. Lin, Y. *et al.* Comparison of normalization and differential expression analyses using RNA-Seq data from 726 individual *Drosophila melanogaster*. *BMC Genomics* **17**, 28 (2016).
25. 't Hoen, P.A.C. *et al.* Reproducibility of high-throughput mRNA and small RNA sequencing across laboratories. *Nat. Biotechnol.* **31**, 1015–1022 (2013).
26. SEQC/MAQC-III Consortium. A comprehensive assessment of RNA-seq accuracy, reproducibility and information content by the Sequencing Quality Control Consortium. *Nat. Biotechnol.* **32**, 903–914 (2014).

27. Leinonen, R., Sugawara, H., Shumway, M. & International Nucleotide Sequence Database Collaboration The sequence read archive. *Nucleic Acids Res.* **39**, D19–D21 (2011).
28. Kodama, Y., Shumway, M., Leinonen, R. & International Nucleotide Sequence Database Collaboration The Sequence Read Archive: explosive growth of sequencing data. *Nucleic Acids Res.* **40**, D54–D56 (2012).
29. Kalra, H. *et al.* Vesiclepedia: a compendium for extracellular vesicles with continuous community annotation. *PLoS Biol.* **10**, e1001450 (2012).
30. Simpson, R.J., Kalra, H. & Mathivanan, S. ExoCarta as a resource for exosomal research. *J. Extracell. Vesicles* **1**, 18374 (2012).
31. Kim, D.-K. *et al.* EVpedia: an integrated database of high-throughput data for systemic analyses of extracellular vesicles. *J. Extracell. Vesicles* **2**, 20384 (2013).
32. Weinstein, J.N. *et al.* The cancer genome atlas pan-cancer analysis project. *Nat. Genet.* **45**, 1113–1120 (2013).
33. Subramanian, S.L. *et al.* Integration of extracellular RNA profiling data using metadata, biomedical ontologies and Linked Data technologies. *J. Extracell. Vesicles* **4**, 27497 (2015).
34. Ainsztein, A.M. *et al.* The NIH Extracellular RNA Communication Consortium. The NIH Extracellular RNA Communication Consortium. *J. Extracell. Vesicles* **4**, 27493 (2015).
35. Xu, P. *et al.* an improved protocol for small RNA library construction using high-definition adapters. *Methods Next Gener.* Seq. 2 2, http://dx.doi.org/10.1515/mngs-2015-0001 (2015).
36. Hansen, K.D., Brenner, S.E. & Dudoit, S. Biases in Illumina transcriptome sequencing caused by random hexamer priming. *Nucleic Acids Res.* **38**, e131 (2010).
37. Hafner, M. *et al.* RNA-ligase-dependent biases in miRNA representation in deep-sequenced small RNA cDNA libraries. *RNA* **17**, 1697–1712 (2011).
38. Fuchs, R.T., Sun, Z., Zhuang, F. & Robb, G.B. Bias in ligation-based small RNA sequencing library construction is determined by adaptor and RNA structure. *PLoS One* **10**, e0126049 (2015).
39. Robinson, M.D., McCarthy, D.J. & Smyth, G.K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
40. McCarthy, D.J., Chen, Y. & Smyth, G.K. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res.* **40**, 4288–4297 (2012).
41. Love, M.I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
42. Law, C.W., Chen, Y., Shi, W. & Smyth, G.K. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* **15**, R29 (2014).
43. Cortez, M.A. *et al.* MicroRNAs in body fluids: the mix of hormones and biomarkers. *Nat. Rev. Clin. Oncol.* **8**, 467–477 (2011).
44. Yang, W. *et al.* Modulation of microRNA processing and expression through RNA editing by ADAR deaminases. *Nat. Struct. Mol. Biol.* **13**, 13–21 (2006).
45. Kawahara, Y. *et al.* Redirection of silencing targets by adenosine-to-inosine editing of miRNAs. *Science* **315**, 1137–1140 (2007).
46. Wang, Y. *et al.* Systematic characterization of A-to-I RNA editing hotspots in microRNAs across human cancers. *Genome Res.* **27**, 1112–1125 (2017).
47. Warnefors, M., Liechti, A., Halbert, J., Valloton, D. & Kaessmann, H. Conserved microRNA editing in mammalian evolution, development and disease. *Genome Biol.* **15**, R83 (2014).
48. Linsen, S.E.V. *et al.* Limitations and possibilities of small RNA digital gene expression profiling. *Nat. Methods* **6**, 474–476 (2009).
49. Baker, M. 1,500 scientists lift the lid on reproducibility. *Nature* **533**, 452–454 (2016).
50. Goodman, S.N., Fanelli, D. & Ioannidis, J.P.A. What does research reproducibility mean? *Sci. Transl. Med.* **8**, 341ps12 (2016).
51. Risso, D., Schwartz, K., Sherlock, G. & Dudoit, S. GC-content normalization for RNA-Seq data. *BMC Bioinformatics* **12**, 480 (2011).
52. Hansen, K.D., Irizarry, R.A. & Wu, Z. Removing technical variability in RNA-seq data using conditional quantile normalization. *Biostatistics* **13**, 204–216 (2012).
53. Leek, J.T. & Storey, J.D. Capturing heterogeneity in gene expression studies by surrogate variable analysis. *PLoS Genet.* **3**, 1724–1735 (2007).
54. Li, S. *et al.* Detecting and correcting systematic variation in large-scale RNA sequencing data. *Nat. Biotechnol.* **32**, 888–895 (2014).
55. Chu, A. *et al.* Large-scale profiling of microRNAs for The Cancer Genome Atlas. *Nucleic Acids Res.* **44**, e3 (2016).

ONLINE METHODS

Reference samples. A synthetic equimolar pool containing 1,152 synthetic RNA oligonucleotides was prepared in an RNase-free environment and working on ice to minimize degradation. The pool was prepared by combining (i) the miRXplore Universal Reference from Miltenyi Biotec, which comprises 962 RNA oligonucleotides with sequences matching human and other miRNAs, and (ii) a set of 190 additional, custom-synthesized RNA oligonucleotides, to generate the pool in which each of the 1,152 RNA oligonucleotides is present at equimolar concentration. This latter set comprises miRNAs and non-miRNA sequences of varied length from 15 to 90 nt, which were synthesized, HPLC-purified and quantified spectrophotometrically by IDT. This latter set of RNA oligonucleotides is available to qualified investigators seeking to reproduce the synthetic equimolar for non-commercial purposes, by request of the corresponding authors (as long as supplies last). The resulting equimolar pool was aliquoted in prelabeled DNA-, DNase-, RNase-, and pyrogen-free screw cap tubes with low adhesion surface and stored immediately at −80 °C. Aliquots were distributed to the participant laboratories in overnight shipments with an abundant supply of dry ice. The complete list of RNA sequences comprising the equimolar pool is provided in **Supplementary Table 1**.

Two ratiometric pools, SynthA and SynthB, containing 334 synthetic RNA oligonucleotides were designed in the coordinating lab (see designing ratiometric pools section below) and synthesized by IDT. Subsets of these oligos were present in 15 different ratios between the two mixtures. These pools were also prepared, aliquoted and distributed to the participant centers following the same previously mentioned precautions to avoid RNA degradation. The complete list of sequences in the SynthA and SynthB pools, as well as their ratios, are provided in **Supplementary Table 2**.

Plasma samples from eleven healthy male donors with age ranging from 21–45 years were collected and pooled in one of the participating labs (**Supplementary Protocol 1**). The Beth Israel Deaconess Medical Center IRB approved the study protocol to consent participants and collect samples. Informed consent was obtained from all subjects, and the samples were subsequently anonymized before distributing to participating research groups. RNA was isolated from this plasma pool (**Supplementary Protocol 2**) in a single lab and aliquots of the purified RNA were mixed and distributed to the rest of the participant centers.

Library preparation and small RNA-seq of reference samples. A written guideline for library preparation and sequencing was distributed to all the participant centers. The input for library preparation was 10 femtomoles of RNA for synthetic pools and 2.1 µl of RNA for the plasma pool. Each group prepared four replicate libraries from each sample using the following small RNA library preparation protocols: Lab1 (TruSeq, NEBNext and in-house 4N_D), Lab2 (TruSeq, NEBNext and in-house 4N_B), Lab3 (TruSeq and NEBNext), Lab4 (TruSeq, NEBNext and in-house 4N_B), Lab5 (TruSeq, CleanTag, NEBNext, in-house 4N_A, in-house 4N_B and 4N_Xu), Lab6 (TruSeq, in-house 4N_B and in-house 4N_C), Lab 7* (NEXTflex), Lab 8* (TruSeq and NEXTflex) and Lab 9* (TruSeq and NEBNext). The labs marked with an asterisk did not contribute plasma libraries.

The protocols for TruSeq, CleanTag, NEBNext and NEXTflex for Illumina were performed according to the manufacturer's instructions in all labs except for NEBNext in Lab9 that performed 3′ overnight ligation. Note that some manufacturers recommended dilution of the adapters when working with low input RNA (for NEBNext, adapters were diluted, 1:2 in Lab3 and Lab9 and 1:6 in Lab1, Lab2, Lab4 and Lab5; for CleanTag 1:20 dilution of the adapters was performed). NEXTflex for Ion Torrent sequencing was performed as described in **Supplementary Protocol 3**. In-house 4N protocols A, B, C and D were performed as described in **Supplementary Protocol 4–7**. 4N_Xu protocol was performed as previously described[35]. Size selection was performed using Pippin instruments (in Lab1, Lab2, Lab4 and Lab6 for all protocols and Lab5 for in-house 4N_B only), 6% acrylamide gels (in Lab3, and Lab9 for all protocols, Lab 8 for TruSeq and Lab5 for TruSeq, NEBNext, CleanTag,4N_A and 4N_Xu) or Ampure XP beads (in Lab7 and Lab8 for NEXTflex).

Single-end libraries were sequenced using the Illumina HiSeq 2500 (Lab8 and Lab9 for all the protocols and Lab5 for TruSeq, CleanTag, 4N-Xu and 4N_A), Illumina HiSeq 4000 (Lab4 for all the protocols and Lab1 for TruSeq, 4N_D and equimolar NEBNext), Illumina NextSeq 500 (Lab2, Lab3 and Lab6 for all the protocols, Lab5 for NEBNext and in-house 4N_B and Lab1 for ratiometric and plasma NEBNext) or Ion Torrent (Lab7) platforms (see **Supplementary Table 3** which also includes information on miRNA editing libraries). All labs using Illumina sequencing performed runs specifying ≥ 50 bp single-end reads. Details on read lengths for each library are included in **Supplementary Table 3**. Each laboratory was free to choose the number of samples to pool per lane, with a target of at least 8 million reads per library. FASTQ files were uploaded to the Genboree Workbench for central data analysis.

Evaluation of miRNA editing. Ten human miRNAs previously shown in the literature to undergo adenosine-to-inosine (A-to-I) RNA editing were selected to evaluate the performance of small RNA-seq in detecting miRNA editing. To this end, we designed six pools containing different ratios of the selected synthetic edited miRNAs and their unedited counterparts (that is, 0, 0.1, 0.5, 5, 50 and 100% edited) plus 277 unrelated human miRNAs. All RNA oligonucleotides were synthesized by IDT (the complete list of sequences included in these pools is provided in **Supplementary Table 7**). The pools were prepared and aliquoted in the coordinating center and distributed to two additional labs following the same previously mentioned precautions to avoid RNA degradation. Each lab prepared three replicate libraries from 10 femtomoles of each pool using three different small RNA library preparation protocols: TruSeq, NEBNext and in-house 4N_B. The protocols for TruSeq and NEBNext were performed according to the manufacturer's instructions (note that for NEBNext, adapters were diluted 1:2). In-house 4N_B was performed as described in **Supplementary Protocol 5**. Size selection was performed using the Pippin Prep. 50 bp single-end libraries were sequenced using the Illumina NextSeq 500.

Designing ratiometric pools. 290 artificial sequences were assigned at random to 8 ratiometric groups (1, 1.5, 2, 3, 4, 5, 8 and 10×) and to either ratiometric pool SynthA or SynthB. The ratio indicates the concentration in the assigned pool relative to the other pool. For example, a sequence in the 10× pool assigned to SynthA would be present at the base concentration in SynthB and at 10× the base concentration in SynthA. To make groups of approximately equal size, we assigned 8 sequences to the 8 ratiometric groups randomly, without replacement. To ensure the total amount of oligonucleotide was approximately equal in SynthA and SynthB, an even number of sequences was assigned to each ratiometric group and were distributed equally between pools, using a similar method of equally-distributed random assignment. The random assignment was performed in Excel and the complete pool composition and ratios are shown in **Supplementary Table 2**.

Barcode splitting, FASTQ generation and data coordination. High-throughput sequencing, demultiplexing and FASTQ file generation was performed by each participating group independently. FASTQ files were uploaded to the Genboree Workbench for centralized analysis using the exceRpt small RNA analysis pipeline (http://genboree.org/java-bin/workbench.jsp).

Preprocessing, mapping and read counting. FASTQ files for the equimolar, ratiometric and plasma pools were initially processed through the exceRpt small RNA-seq Pipeline (Version 4.6.2), using the batch submission tool. For details on the exceRpt pipeline and the associated processing steps, see the Genboree Workbench documentation (http://genboree.org/theCommons/projects/exrna-tools-may2014/wiki/Small%20RNA-seq%20Pipeline). A brief description of parameters changed from the default settings or that differed between libraries is included below.

The exceRpt pipeline was used at the default settings whenever possible. The default for adapter trimming is 'auto-detect', which identifies and trims the adapter sequence for multiple library types, and all samples were initially submitted using this functionality. For 4N libraries (A, B, C, D, Xu and NEXTflex), an additional parameter was selected to indicate the degenerate sequence at the end of each adapter. The default random barcode settings were used, indicating that random 4nt sequences are present immediately 5′ and 3′ of the insert sequence. The sequence and identity of the adapter identified by the exceRpt pipeline was confirmed in the output files. Any library with a missing or incorrect adapter identified was re-submitted to the pipeline with the adapter sequence chosen manually, and a note was added to **Supplementary Table 3**.

doi:10.1038/nbt.4183

For plasma pool libraries, sequences shorter than 18 nt after adapter trimming were removed and not used for downstream analysis. For synthetic pools, the minimum length was changed to 15 nt, which corresponds to the length of the shortest sequences in the equimolar and ratiometric pools.

To quantify alignments to the full set of synthetic pool sequences, equimolar, ratiometric SynthA and ratiometric SynthB libraries were mapped to a 'Spike-In' sequence library uploaded to the Genboree Workbench. This spike-in library FASTA file contains a non-redundant set of sequences from the ratiometric and equimolar pools (**Supplementary Table 7**). Adapter-trimmed and filtered reads were mapped to the spike-in index with bowtie2 using the default Genboree Workbench alignment parameters, except that the minimum read length was reduced to 15. The number of reads aligning to each sequence was obtained from the .calibrator.mapped output files. At the time of writing, reads mapped to the spike-in sequences are removed before genomic alignment, so any endogenous alignment information from these samples was ignored. To quantify alignments to endogenous miRNAs, equimolar and plasma pool libraries were also run through the exceRpt pipeline without mapping to spike-in sequences. The default minimum read length of 18 nt was used, along with all default alignment parameters. Reads were mapped to hg19 using the STAR alignment algorithm. Multi-mapping-adjusted read counts corresponding to mature miRNAs were used for all plasma pool analyses and for the equimolar pool correction factor analyses. For all other analyses with the equimolar and ratiometric pools, the spike-in read counts from the .calibrator.mapped files were used.

Sample filtering. Unless specifically noted in the text, only libraries meeting minimum read count requirements were considered for analysis. For the synthetic pools, an average of one million reads mapping to the 'spike-in' sequences (the unique set of sequences present in the equimolar and ratiometric pools), were required across all replicate libraries. The average was taken after filtering, such that the totals were based only on 5′-phosphorylated sequences 16–25 nt in length. For the plasma pool samples, replicate libraries with fewer than 100,000 miRNA-mapping reads were removed. The entire sample was removed if more than one of the replicate libraries failed to pass the minimum count threshold.

Equimolar pool analysis. Read counts for the equimolar (and likewise for the ratiometric pools) were obtained from 'calibratormapped.counts' files included in the exceRpt pipeline output for each sample file. Sample-specific information, including the contributing lab, library preparation method and replicate number were associated with the corresponding calibrator count file, and were loaded into R for analysis. A full list of equimolar and ratiometric sequences with additional sequence information was used as a reference to merge all input files and add zero counts, where needed. Unless specifically mentioned in the text, analysis of ratiometric and equimolar libraries was limited to sequences with a 5′-phosphate modification, 16–25 nt in length. Read counts were scaled to counts per million (CPM) using the total counts from the filtered sequences.

For plots and calculations using log-transformed values, an arbitrarily small count was added to avoid taking the log of zero. To confirm that the extent of sequencing bias and reproducibility we observed was not influenced by the choice of pseudo-count for calculating CPM, we repeated our calculations in different ways with pseudo-counts ranging from 1 to 0.0001 (see **Supplementary Fig. 16**). The adjusted CPM values were calculated using the method employed by the R package, EdgeR[39,40]. This scales the user-supplied prior count (0.25; the default setting) to be proportional to the library size. The scaled prior count is calculated by multiplying the raw prior count (0.25) by the sample library size divided by the mean library size across all equimolar samples and then adding this value to the raw counts for each miRNA. Library sizes are adjusted by adding 2 × the scaled prior count value. Adjusted CPM values are finally calculated as (raw.count + adjusted.prior) * 10^6 / adjusted.library.size.

Determining overrepresented and under-represented sequences. Sequences in each equimolar pool replicate library were ranked by abundance, assigning the minimum rank value in case of ties. The ten top and bottom-ranked sequences were determined by arranging counts in descending and ascending order, respectively. TruSeq, NEBNext, CleanTag and 4N libraries were each queried for sequences consistently found in the top or bottom 10, as defined by at least 75% agreement among the libraries of at least one method.

Dissecting the source of bias. The CPM obtained for each sequence of the equimolar pool was calculated using pseudo-counts, as in the equimolar pool analysis described above, except that the library sizes were calculated from all equimolar pool sequences before filtering for length and end modifications. Sequence length, 5′and 3′ terminal bases, %GC of the four 5′or 3′end base, overall %GC, dG [free energy], dH [enthalpy], dS [entropy] and Tm [melting temperature] were calculated from the annotated sequence. UNAFold[56] (http://unafold.rna.albany.edu/) was used to obtain the dG, dH, dS and Tm values of each of the sequences comprising the equimolar pool.

Ratiometric pools analysis. Ratiometric pool counts were initially processed as described above for equimolar pools, considering only counts for 16–25 nt sequences. The ratio of SynthA:SynthB was calculated as the ratio of the mean CPM across technical replicates in SynthA / SynthB.

Ratiometric Pools: Differential Expression. Independent differential expression workflows were run for each lab and library prep method, following a standard two-group comparison between 'A' and 'B' ratiometric pools. Normalization, dispersion estimation and differential expression testing was performed using three different R packages: EdgeR[39,40], DESeq2 (ref. 41) and limma/voom[42]. For EdgeR, normalization factors were calculated using the Relative Log Expression (RLE) method, and significance was calculated (after calculating common, trended and tagwise dispersion estimates) using the default settings, based on a likelihood ratio test on the null hypothesis that ratiometric sample B – A = 0. Default settings were used for DESeq2, and significance was calculated based on a Wald Test. Significance for the limma/voom workflow was based on an empirical Bayes moderated *t* test.

Plasma pool analysis. Comparison of plasma pool libraries was limited to mature miRNAs. Read counts for mature miRNAs were taken from 'readCounts_miRNAmature_sense.txt' files provided in the exceRpt pipeline output. The read count files and associated metadata for all samples were loaded and merged in R for further analysis. Multi-mapping-adjusted read counts are calculated as part of the exceRpt pipeline and were used for all comparisons. The total number of unique reads mapping to miRNAs was taken from '.stats' files provided in the exceRpt pipeline output.

Downsampling read counts. The R package Vegan, was used to simulate random downsampling of equimolar and plasma pool count matrices. The Vegan function, drarefy, was used to estimate the probability of detection for each sequence based on random simulations of downsampling to specified levels. For the plasma pools, downsampling was performed to four different levels (10^4, $10^{4.5}$, 10^5 and $10^{5.5}$). Equimolar pools were downsampled to six different levels ($10^{6.5}$ down to 10^4 at half-log intervals). Libraries with read counts below the specified threshold were removed. A minimum probability of 0.9 was used as the threshold for detection.

Inter-protocol bias correction factors: estimation. Equimolar pool samples were processed through the exceRpt pipeline using the same input parameters as the plasma pool libraries, to obtain multi-mapping, scaled read counts for mature miRNAs that were directly comparable to the plasma pool counts. Differential expression analyses were performed using the mature miRNA read counts for the equimolar pool samples, and scaling factors were calculated for each miRNA, and were taken from the resulting $\log_2$ fold-change estimates. Key assumptions used in these calculations were: that the median mapped read level calculated for a given protocol should match the median for the 4N results given the same RNA input; that the comparisons of the results from a given protocol and the 4N protocol are performed on data processed in the same way that biological samples are processed (that is, using the exceRpt pipeline and its mapped read outputs). For details on limma and voom functionality and the parameters used, see the documentation for the limma package.

To summarize, correction factors were calculated for each pair of library preparation methods using the following workflow:

1. Filter out miRNAs with 0 counts in any library: For the subset of samples being tested, scaling factors are only calculated for miRNAs having at least one count in every sample of the two methods being tested.

2. Prepare miRNA count matrices for linear modeling: Use the R package, voom, to calculate precision weight estimates and normalize data to allow count

data to be analyzed appropriately using the limma package. Normalization is also performed between samples such that the median miRNA expression value is the same in all samples.

3. Fit miRNA-wise linear models to account for batch (lab) effect: The lmFit function from the limma package is used to fit linear models for each miRNA, estimating coefficients for each lab+library prep method. The coefficients represent the differences in expression for each miRNA between each lab+library prep method.

4. Estimate the log fold change between the two methods for each miRNA: Use the fitted model to calculate for each miRNA the average expression estimated using the method A coefficients – the average expression estimate for the method B coefficients.

The R packages limma/voom were used for read count normalization and differential expression estimates, using standard workflows suggested for RNA-seq data to account for batch (lab) effects, and then testing for the main effect of the library prep methods. For each pairwise comparison of library preparation methods, equimolar pool counts matrices were extracted and only miRNAs with ≥ 1 read count in all samples of both methods were kept for analysis. After filtering, voom was used to normalize the count data and calculate precision weight estimates that allow count data to be appropriately tested with the linear modeling schema used in the limma package. Voom was run with the default parameters, except that read counts were additionally normalized between arrays using the 'scale' method, which adjusts read counts such that the median miRNA expression value is the same in all labs. The *voom*-transformed data was supplied to the limma lmFit function, along with a design matrix indicating the coefficients to be estimated. Initially, coefficients were estimated for each lab + library prep method to model batch/lab-specific effects. The main effect of the library prep method was then calculated as the average effect of method 1 – method 2. A contrasts matrix was generated and supplied, with the fitted model, to the contrasts.fit function, followed by an empirical Bayes function to estimate the resulting statistics for each miRNA. Log_2 fold-change estimates, along with 95% confidence interval were obtained from these estimates.

Inter-protocol bias correction factors: applying corrections. The equimolar pool-derived, inter-protocol bias correction factors were applied to the corresponding plasma pool samples for testing. To apply the correction factors, count matrices for the subset of plasma pool libraries being compared were selected and were then pre-filtered and normalized in the same way as the equimolar pools in generating the correction factors, described above. MiRNAs were filtered to include only those with a correction factor estimated from the equimolar pool and at least five counts in every library in the subset of methods being compared. Correction factors were applied to the appropriate samples. For example, if correction factors were calculated as the log_2 fold-change between TruSeq and 4N samples (TruSeq - 4N), then the correction factors would be applied to the log_2-transformed TruSeq plasma pool samples by subtracting the correction factor. For the heatmaps and density plots, corrected values were added to the original count matrix of untransformed values, and unless specifically noted in the text, normalized using quantile normalization.

miRNA editing analysis. Editing libraries were trimmed of 5′ and 3′ adapters using cutadapt (version 1.9.1). Trimmed reads 16 nt and longer were aligned to editing pool sequences using bowtie2 (version 2.3.2) in local alignment mode. The first (5′) 4 nt were removed from 4N library reads during the alignment stage by adding the optional parameter '–trim5p 4'. Read counts were calculated from alignments filtered to have a minimum MAPQ of 20 and 0 mismatches to the reference sequence within the locally-aligned region. The sum totals of the filtered read counts for each library were used to calculate CPM. Down-sampling was performed using the R package, Vegan.

Life Sciences Reporting Summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this article.

Data availability and accession code availability statements. Sequencing data for all experiments can be obtained from the GEO Superseries, GSE94586. Accession numbers for the four subseries are: GSE94584 (Equimolar), GSE94585 (Ratiometric A/B), GSE94582 (Human Plasma Pool) and GSE108138 (A-to-I Editing). GEO records include raw FASTQ files and processed counts from the exceRpt pipeline.

All code, metadata and processed data files required for reproducing the figures, tables and in-text statistical summaries are freely available on GitHub (https://github.com/rspengle/CrossU01_exRNA_Manuscript2017). The repository also includes a Packrat library with a snapshot of R package versions used.

56. Markham, N.R. & Zuker, M. UNAFold: software for nucleic acid folding and hybridization. *Methods Mol. Biol.* **453**, 3–31 (2008).

doi:10.1038/nbt.4183

nature research

Corresponding author(s): Muneesh Tewari, Maria D. Giraldez and David Galas

☐ Initial submission ☐ Revised version ☒ Final submission

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Experimental design

1. Sample size

Describe how sample size was determined.

Not done. Our study was aimed at analyzing accuracy, bias and reproducibility based on ground truth composition of the samples therefore a pre-specified effect size was not used.

2. Data exclusions

Describe any data exclusions.

All the participating groups sequenced the same reference samples for the study and only those libraries not passing the QC criteria were excluded for the analysis. Reported in online Methods, sample filtering section/paragraph 1.

3. Replication

Describe whether the experimental findings were reliably reproduced.

Experiments in this paper were done in quadruplicate or triplicate and most of them included multiple centers.

4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups.

We specifically created reference samples for the study. The participating groups obtained aliquots of these reference samples for all the analyses to permit comparisons across groups. Since we were able to send the same samples to all groups, randomization was not needed. Reported in online Methods, Library preparation and small RNA-seq section/paragraph 1.

5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

There was no blinding, as it would have been difficult to truly blind the standard reference samples because they could easily be discerned during the process of working with them, based on RNA diversity and concentration. Moreover, for this type of study aimed at analyzing bias, accuracy and reproducibility of sequencing using standard reference samples, blinding is less important than it is for an experimental study assessing outcomes, for example.

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

Nature Biotechnology: doi:10.1038/nbt.4183

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
☐	☒ A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ A statement indicating how many times each experiment was replicated
☐	☒ The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
☐	☒ A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☐	☒ The test results (e.g. *P* values) given as exact values whenever possible and with confidence intervals noted
☐	☒ A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
☐	☒ Clearly defined error bars

See the web collection on statistics for biologists for further resources and guidance.

Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.	The computer code is available on GitHub at this link: https://github.com/rspengle/CrossU01_exRNA_Manuscript2017.

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* guidance for providing algorithms and software for publication provides further information on this topic.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.	We have provided the sequences contained in the synthetic pools evaluated in this study so investigators can synthesize them if desired. In addition, we are committed to providing the customs pools to any qualified investigators seeking to reproduce the synthetic reference pool for non-commercial purposes, as long as our own supply lasts.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).	Antibodies were not used in this study.

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.	Cell lines were not used in this study
b. Describe the method of cell line authentication used.	Not applicable.
c. Report whether the cell lines were tested for mycoplasma contamination.	Not applicable.
d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.	Not applicable.

Nature Biotechnology: doi:10.1038/nbt.4183

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.	This study did not involve animals.

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.	We have not used individual human samples in this study. The plasma RNA pool analyzed in this paper was obtained from eleven healthy male individuals with age ranging from 21-45 years that was collected and pooled before RNA isolation. The Beth Israel Deaconess Medical Center IRB approved the study protocol to consent participants and collect samples. The samples were subsequently anonymized before distributing to the other participating research groups. See details reported in online methods, reference samples section/paragraph 3 (page 22). The participating groups did not have access to any patient identifying information.

Nature Biotechnology: doi:10.1038/nbt.4183

Erratum: Comprehensive multi-center assessment of small RNA-seq methods for quantitative miRNA profiling

Maria D Giraldez, Ryan M Spengler, Alton Etheridge, Paula M Godoy, Andrea J Barczak, Srimeenakshi Srinivasan, Peter L De Hoff, Kahraman Tanriverdi, Amanda Courtright, Shulin Lu, Joseph Khoory, Renee Rubio, David Baxter, Tom A P Driedonks, Henk P J Buermans, Esther N M Nolte-'t Hoen, Hui Jiang, Kai Wang, Ionita Ghiran, Yaoyu E Wang, Kendall Van Keuren-Jensen, Jane E Freedman, Prescott G Woodruff, Louise C Laurent, David J Erle, David J Galas & Muneesh Tewari
Nat. Biotechnol. **doi:10.1038/nbt.4183; corrected online 31 July 2018**

In the version of this article initially published online, the text "Beth Israel Deaconess Medical Center/Dana Farber Cancer Institute (BIDMC/DFCI)" was inserted into the last sentence in the right-hand column of p.10, beginning "It is worth noting…." . In addition, on p.2, the acronym for The Cancer Genome Atlas was given as TGCA, rather than TCGA; and on p. 3, UUTR should have been defined, as University of Utrecht, the Netherlands. Finally, ref. 48 was cited in the Online Methods after "4N_Xu protocol was performed as previously described[35]"; this extra citation has been deleted. The errors have been corrected for the print, PDF and HTML versions of this article.