









An extension to: Systematic assessment of commercially available low-input miRNA library preparation kits

Fatima Heinicke ^a, Xiangfu Zhong ^{a#}, Manuela Zucknick ^b, Johannes Breidenbach ^c, Arvind Y.M. Sundaram^a, Siri T. Flåm^a, Magnus Leithaug ^a, Marianne Dalland^a, Simon Rayner ^a, Benedicte A. Lie ^{a*}, and Gregor D. Gilfillan ^{a*}

^aDepartment of Medical Genetics, Oslo University Hospital and University of Oslo, Oslo, Norway; ^bDepartment of Biostatistics, Oslo Centre for Biostatistics and Epidemiology, University of Oslo, Oslo, Norway; ^cNational Forest Inventory, Norwegian Institute for Bioeconomy Research, Ås, Norway

ABSTRACT

High-throughput sequencing has emerged as the favoured method to study microRNA (miRNA) expression, but biases introduced during library preparation have been reported. We recently compared the performance (sensitivity, reliability, titration response and differential expression) of six commercially-available kits on synthetic miRNAs and human RNA, where library preparation was performed by the vendors. We hereby supplement this study with data from two further commonly used kits (NEBNext, NEXTflex) whose manufacturers initially declined to participate. NEXTflex demonstrated the highest sensitivity, which may reflect its use of partially-randomized adapter sequences, but overall performance was lower than the QIAseq and TailorMix kits. NEBNext showed intermediate performance. We reaffirm that biases are kit specific, complicating the comparison of miRNA datasets generated using different kits.

ARTICLE HISTORY

Received 18 January 2020
Revised 18 March 2020
Accepted 22 April 2020

KEYWORDS

microRNA; miRNA; small RNA-seq; library preparation; sequencing bias; low RNA input; NGS; Next Generation Sequencing; NEBNext; NEXTflex




Introduction

Interest in miRNAs has steadily increased since their discovery in the early 1990s due to their roles in diverse biological processes [1–4] and their dysregulation associated with several diseases [5–7]. Next generation sequencing (NGS) is an attractive technology to study miRNAs because of its high sensitivity and ability to detect novel miRNAs. Several commercially-available kits are available to prepare miRNA libraries for sequencing, which entails addition of adapter sequences to the miRNAs followed by reverse transcription and cDNA synthesis. In a recent study, we compared the performance of six such kits (CATS, CleanTag, QIAseq, TailorMix, SMARTer-beta and srLp) with respect to detection rate sensitivity, reliability and ability to detect differentially expressed miRNAs [8]. However, two commonly used kits (NEBNext and NEXTflex) were not included.

Previous studies have reported differences in miRNA abundance detected by sequencing relative to the original RNA sample, which makes miRNA quantification challenging [9,10]. Sequencing library preparation, and in particular the adapter ligation steps, have been identified as the primary sources of this bias [10,11]. Most kits utilize RNA ligases to attach adapters to the miRNAs (e.g. NEBNext, QIAseq, TailorMix, CleanTag), but the efficiency of this step depends on the ligase used, the adapter

sequence and the primary and secondary structure of the miRNA [10–13]. NEXTflex reagents attempt to increase efficiency and reduce bias at this step by utilizing adapters containing stretches of random nucleotides, which increases adapter sequence diversity. Other attempts to avoid bias whilst introducing adapter sequences onto miRNAs are polyadenylation and template switching oligonucleotides (e.g. CATS) or by using single adapter circularization (e.g. SMARTer).


In this study, we aimed to complete our systematic comparison of different miRNA library preparation kits by investigating the performance of the NEBNext and NEXTflex kits (Table 1). It should be noted that although the study aimed to test low input kits handling inputs below 100 ng, the NEBNext kit is not designed for inputs below 100 ng, but was nonetheless included as it is widely used. Our previous and present studies were performed under the same conditions with one exception: While in the first study the library preparation was performed by the kit vendors themselves, for the two kits presented in the present study this step was performed at Oslo University Hospital. The results presented here were generated at the same time as those in the first study, and were sequenced on the same flow cell. This manuscript gives an overview on the results for all eight kits, with a focus on the

CONTACT Fatima Heinicke  fatima.heinicke@gmail.com; Benedicte A. Lie  b.a.lie@medisin.uio.no; Gregor D. Gilfillan  gregorg@medisin.uio.no;

 Department of Medical Genetics, Oslo University Hospital and University of Oslo, Oslo 0450, Norway

[#]Department of Biosciences and Nutrition, Karolinska Institute, Huddinge 14183 Sweden

*Contributed equally

 Supplemental data for this article can be accessed [here](#).

© 2020 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

Table 1. Small RNA library preparation methods tested in this study.

Kit name	Commercial supplier	Key features*	Max. input volume tolerated	Reported RNA input range (varies with type of input tested)	Max. number of indexes available	Method types	Approx. price per reaction (using kit with highest reaction number available)
NEXTFLEX® Small RNA-Seq Kit v3 (NEXTflex)	PerkinElmer Inc.	5-step process of 3'adapter ligation, adapter inactivation, 5'adapter ligation, reverse-transcription and PCR. 3 purification steps.	10.5 µl	Total RNA (1 ng – 2 µg), purified small RNA (from 1–10 µg total RNA), and a synthetic miRNA pool (≥100 pg)	48	Ligase based. 2-adapter procedure. Utilizes adapters with randomized 4mer ends	\$56 (48)
NEBNext® Small RNA Library Prep (E7300) (NEBNext)	New England Biolabs Inc.	Single-tube, 5-step process of 3' adapter ligation, primer annealing, 5'adapter ligation, reverse-transcription and PCR. 1–2 purification steps.	6 µl	Total RNA (100 ng–1 µg)	48	Ligase based. 2-adapter procedure	\$52 (96)

* A step is defined as a labwork period that culminates in an incubation longer than 5 minutes.

NEBNext and NEXTflex kits. For more details on the other six kits we refer to Heinicke, et al. [8].

Results

Altogether 21 samples, comprising 15 synthetic miRNA samples (five mixes processed in triplicates) and six human total RNA samples (pooled rheumatoid arthritis patients and pooled healthy controls processed in triplicates), were used to assess the performance of the different library preparation kits (Fig. 1A). To aid comparison, we present here the results

of all eight kits, with our previous results [8] displayed in faded colours in the figures. Following library preparation, the NEBNext and NEXTflex libraries were sequenced together (i.e. on the same sequencing flow cell) with the libraries from the other six library preparation kits [8]. Previously-published samples were not re-sequenced: the new data presented here was generated at the same time as those presented earlier, thus partial figures reproduced from the earlier study to aid comparison did not entail re-analysis. For NEBNext and NEXTflex, cluster density and read numbers passing filters were similar to the other kits that previously performed

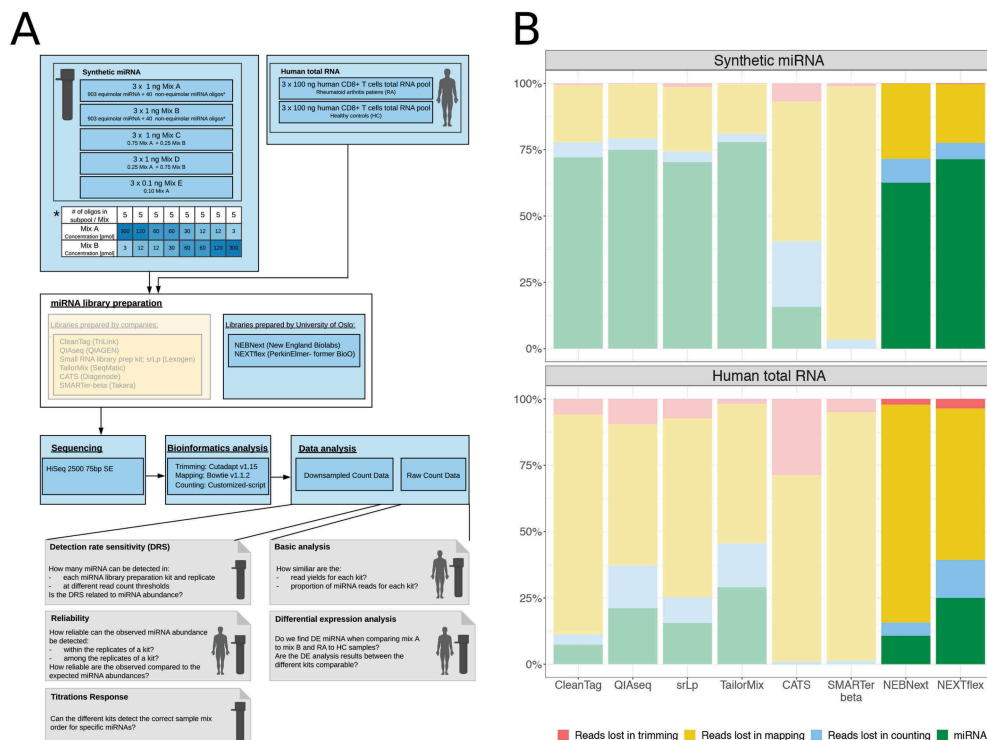


Figure 1. Experimental design and sequencing read distribution. (A) Overview of the study material, miRNA library preparation kits used, sequencing, bioinformatics and data analysis. Library preparation was performed in house in contrast to the study design presented in [8]. Grey boxes represent individual data analysis steps. (B) Percentage of reads that were removed during the bioinformatic analysis and final miRNA proportion remaining (green). Trimming refers to removal of adapter sequences, mapping to miRNA reference alignment, and counting to filtering of aligned miRNAs that did not have the same length as the reference sequence. Results presented are the mean of 15 replicates in the synthetic miRNA (upper panel) and the mean of six replicates in the human total RNA samples (lower panel). Faded colours were used to indicate previous results [8]. Images from Servier Medical Art (Servier. www.servier.com, licenced under a Creative Commons Attribution 3.0 Unported Licence) were used in (A).

well (CleanTag, QIAseq, srLp, TailorMix) (Supplementary Fig. 1 and Table 2).

Consistent with our earlier study, the greatest proportion of reads which were discarded during the bioinformatics analysis, both for NEBNext and NEXTflex, were discarded during mapping to the miRNA reference sequences (Fig. 1B, Supplementary Fig. 2 and Supplementary Table 1). Notably, the NEXTflex kit compared favourably to the best performing kits identified previously, and despite not being designed to handle sub-100 nanogram amounts, NEBNext performed adequately. To comprehensively evaluate the performance of the kits and to reduce the influence of technical aspects such as different library purification methods and raw read yields, read counts were randomly down-sampled (2.5 million reads for synthetic miRNA samples and 0.75 million reads for human total RNA samples) and, where stated, were regularized log (rlog) transformed for subsequent analysis steps.

Since the library preparation for NEBNext and NEXTflex was performed by researchers at Oslo University hospital and not by the kit vendors themselves, as in our previous study, a principal component analysis (PCA) was performed on the downscaled rlog transformed count data to assess possible bias this may have introduced (Supplementary Fig. 3). For the synthetic miRNA and human total RNA samples, the first two principal components revealed that the samples aligned by library preparation kit. These results indicate that biases caused by kit-specific factors are the dominant cause of variation observed, rather than library preparation at different locations.

To assess the detection rate sensitivity of the library preparation kits, we tested several detection thresholds in the down-sampled synthetic miRNA samples. First, we defined a miRNA to be detected in a sample when at least one read in total was registered. NEXTflex detected more miRNAs across all three replicates of the different synthetic mixes than NEBNext (Fig. 2A). Compared to our previous results, NEBNext was the kit that detected fewest miRNAs in all replicates of the different mixes. Furthermore, in mix E, where the RNA input was 10 times lower than in mixes A to D, NEBNext detected the fewest number of miRNAs across all three replicates. In contrast, NEXTflex, together with QIAseq and TailorMix missed the fewest miRNAs in one, two or all three replicates. The undetected miRNAs were generally kit specific (Supplementary Figure 4). However, some miRNAs such as EBV-1-3P and MIR-612, EBV-20-3P, MIR-548D-3P and MIR-193A-3P (miRNA annotation according to miRXPlore Reference) were undetected across several kits and replicates (Supplementary Figure 5). A comparison of the miRXPlore and miRBase v22 miRNA annotation is provided in Supplementary Table 2.

When analysing the 40 non-equimolar miRNAs, NEXTflex revealed a very high detection rate sensitivity, second only to the previously tested QIAseq kit (Supplementary Figure 5 and 6). Conversely, for NEBNext we observed the lowest detection rate sensitivity (except for the CATS and SMARTer-beta kits which were excluded from the analysis at this step already). MicroRNA detection was not solely dependent on concentration level (e.g. the kits also struggled to detect some miRNAs present at intermediate concentrations;

Supplementary Figure 6), illustrating that kit-specific biases play a role. When the random down-sampling was repeated 100 times to control for stochastic variation, the absolute number of detected miRNA changed slightly but the overall detection ratios between the different library preparation kits remained stable (Supplementary Figure 7 and 8). As before, comparing the detection rate sensitivity of the 1.0 ng synthetic miRNA samples (mix A-D) with the 0.1 ng synthetic miRNA samples (mix E) revealed no striking difference for any of the kits (Fig. 2A and Supplementary Figure 7), suggesting that kit-specific differences rather than input RNA concentration account for most of the differences observed.

Next, we examined sensitivity under more stringent detection thresholds, requiring a miRNA to be detected when at least 1, 10, 50, 100 or 200 read counts per million (CPM) were registered across all three mix replicates. With the exception of the non-equimolar miRNAs presented at the lowest concentration levels, all synthetic miRNAs should theoretically be detected at 200 CPM. However, as observed previously in Heinicke, et al. [8], the number of detected miRNAs decreased greatly with increasing CPM threshold also for the NEXTflex and NEBNext kits (Fig. 2B). Overall NEXTflex detected most miRNAs among all mixes and CPM thresholds. The number of detected miRNAs across the CPM thresholds differed significantly ($p < 0.05$) when NEXTflex was compared to NEBNext, CleanTag, srLp or mixes A, C and D of TailorMix. No statistically significant differences could be detected when NEXTflex was compared to QIAseq and mixes B and E of TailorMix (Supplementary Table 3). NEBNext detected fewer miRNAs across all mixes and CPM thresholds than NEXTflex and obtained similar results to CleanTag and srLp.

We used down-sampled and rlog transformed miRNA count data to assess reliability. The intra-rater reliability (miRNA read count concordance within the replicates of a library preparation kit) of NEBNext and NEXTflex were as strong as for the previously tested kits, although slightly weaker results were observed for mix E with NEBNext. Both kits revealed ICC values between 0.93 and 0.99 (Supplementary Table 4) and Pearson correlation coefficients above 0.91 ($p < 0.05$, Supplementary Table 5). Bland-Altman plots (data not shown) indicated no systematic differences in the measurements.

To examine inter-rater reliability (miRNA read count concordance between the library preparation kits) the first replicate of each synthetic miRNA mix, rheumatoid arthritis (RA) or healthy control sample from all six library preparation kits (NEBNext, NEXTflex, CleanTag, QIAseq, TailorMix, srLp) was chosen. Larger differences were observed between the different library preparation kits than within the replicates of a kit with regard to miRNA reads counts. Similar to our previous study, ICC values were above 0.8 for the synthetic miRNA sample mixes and above 0.95 for the RA or healthy control samples (Supplementary Table 6). The same was true for the Pearson correlation coefficients which were above 0.73 and 0.92 ($p < 0.05$) for the synthetic miRNA and human total RNA samples respectively (Supplementary Table 7). No systematic differences in the measurements were observed by Bland-Altman analysis (data not shown). The larger ICC

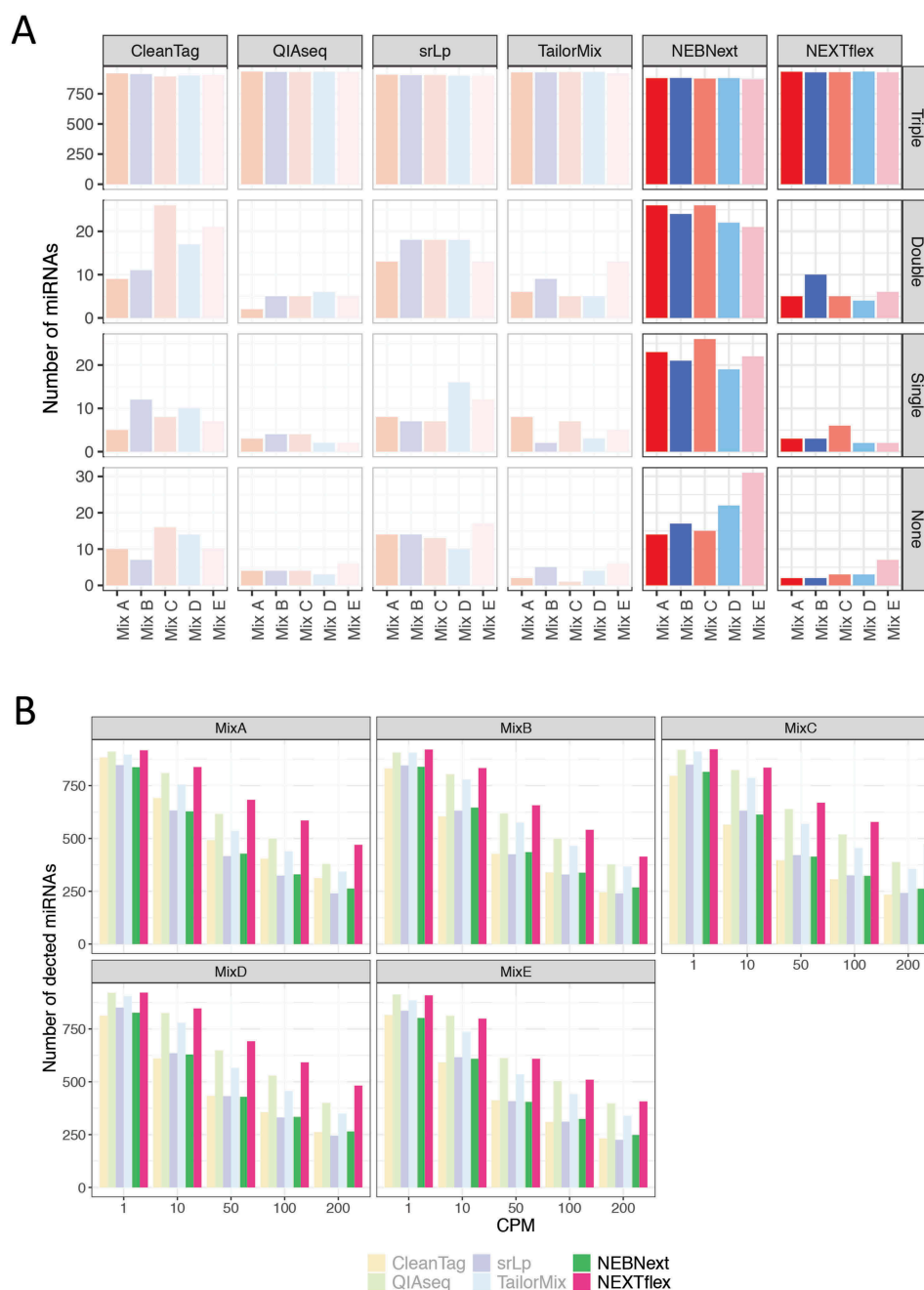


Figure 2. Detection rate sensitivity. (A) Bar charts presenting number of miRNAs detected in all replicates (Triple), in 2 out of 3 replicates (Double), in 1 out of 3 replicates (Single) or not detected in any replicate (None) across all synthetic miRNA mixes and all library preparation kits. The maximum number of detectable miRNAs is 943 (903 equimolar and 40 non-equimolar miRNA). (B) Bar charts for various read count thresholds in the synthetic miRNA samples. A miRNA is defined as detected when it is (i) expressed in all three replicates of the mix and (ii) the read counts are greater or equal to the count per million (CPM) threshold displayed on the x-axis. Faded colours were used to indicate previous results[8].

values and Pearson correlation coefficients of the human total RNA compared to the synthetic miRNA samples appear to be induced by the greater proportion of low read counts in the human total RNA samples. The applied rlog transformation stabilizes particularly the low read counts which improves the reliability for these measurements. As most biological studies are interested in the highly expressed miRNAs, we further filtered out read counts with an average expression across all human total RNA or synthetic miRNA samples of less than 20 CPM, which narrowed the differences between synthetic and human RNA samples (Supplementary Table 6 and 8).

As a further assessment of reliability, we investigated the concordance between the theoretical miRNA concentrations and the obtained read counts for the synthetic miRNA samples. For the 903 equimolar miRNAs, no significant deviation between a specific miRNA rlog read count and the median rlog read count over all equimolar miRNA was expected to be seen. The fold deviation was defined to be equimolar when its absolute value was less or equal to one. However, for the randomly chosen first replicate of mix A, only between 37.2% to 42.6% of the miRNAs were detected as equimolar. NEBNext detected the lowest number miRNAs to be

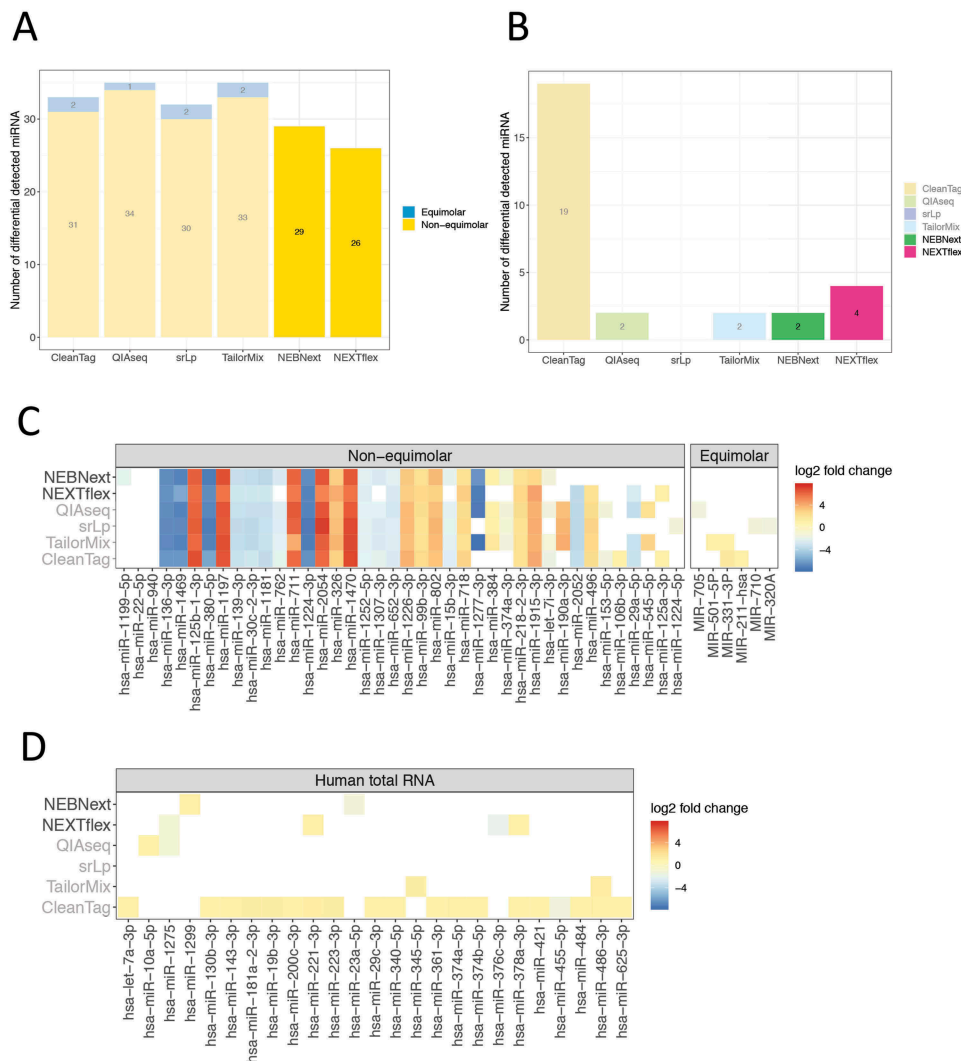


Figure 3. Differential expression analysis. Kit-specific number of differentially expressed miRNA detected for (A) synthetic miRNA samples (mix A versus mix B) and (B) human total RNA samples (RA versus healthy control). miRNA-specific log₂ fold changes across the different kits for (C) synthetic miRNA samples and (D) human total RNA samples. Faded colours or grey font were used to indicate previous results[8].

equimolar while NEXTflex detected the highest number across all tested kits (Supplementary Figure 9).

To compare the performance of the kits for quantifying miRNA levels, the read counts of the 40 non-equimolar miRNAs were correlated with the expected theoretical levels. NEXTflex showed slightly lower correlations across all samples than QIAseq, which obtained the highest correlation coefficients in our previous study (Supplementary Table 9 and Supplementary Table 8 in [8]). NEBNext was a middle-ranking kit in this correlation. However, as before, we found that none of the tested kits could accurately quantify the majority of miRNAs.

To examine kit performance in differential miRNA expression, non-down-scaled and untransformed miRNA counts, filtered to remove low-abundance miRNAs, were analysed. Between mix A and mix B of the synthetic miRNA samples, 29 out of 40 differentially expressed miRNAs were detected by NEBNext and 26 by NEXTflex (Fig. 3A). In comparison, all previously tested library preparation kits were able to detect between 32 to 35 differentially expressed miRNAs. However, of

those not all miRNAs were true positives. While only differentially expressed miRNAs were expected to be found within the pool of non-equimolar miRNA ($n = 40$), an additional one to two equimolar miRNAs were detected to be differentially expressed by the previously tested library preparation kits. This was not the case for NEBNext or NEXTflex. MiRNAs that could not be detected as differentially expressed between mix A and B were often those with the lowest concentration level differences (Fig. 3C). Quantitative reverse-transcriptase PCR assays on 16 of the 40 non-equimolar miRNAs revealed that the intended ratios for mix A and mix B were as expected (Supplementary Figure 10).

We also performed differential expression analysis between the RA patient and healthy control pools of human total RNA samples. NEBNext detected two and NEXTflex four significant differentially expressed miRNAs (Fig. 3B), but the kits did not identify the same miRNAs as differentially expressed. There was also no overlap between the differentially expressed miRNAs predicted by NEBNext and those predicted by the previously-tested miRNA library

preparation kits. For NEXTflex, three of the four miRNAs were already previously detected as differentially expressed by other kits [8]: hsa-miR-1275 was also detected by QIAseq to be down-regulated in RA patients compared to healthy controls, while hsa-miR-378a-3p and hsa-miR-221-3p were detected by CleanTag to be up-regulated in RA patients versus healthy controls (Fig. 3D).

Finally, we compared the performance of the kits in the titration response assay, which provides a measure of quantitative performance [14,15]. Downscaled and rlog transformed read counts of the 40 non-equimolar miRNAs were scored for their adherence to expected concentration orders in mixes A-D, with five miRNAs in each of the eight concentration groups Table 3. In this assay, NEBNext performed better than NEXTflex, which had an intermediate performance relative to the results reported previously [8].

Discussion

We assessed the performance of NEBNext and NEXTflex and present the results along with the six library preparation kits we tested previously [8]. Identical RNA input samples prepared at the same time point and under the same conditions were used in both studies. The prepped sample libraries from all kits were sequenced on the same flow cell and identical bioinformatics and data analysis steps were performed. However, the studies differ in the way in which the library preparation was performed: While it was performed by the kit vendors themselves in our first study [8], we performed library preparation for this additional study. Although our aim was to make the two studies as similar as possible, we cannot exclude that the different library preparation approaches and sites may have influenced the results. Furthermore, since the datasets for NEBNext and NEXTflex were generated from individual sequencing lanes, unlike for most kits in the first study which were distributed across several lanes, we cannot exclude that lane-specific effects on data quality may have influenced the conclusions in this current work.

For each performance metric examined (sensitivity, reliability, titration response and differential expression), we consistently found differences that we attribute to kit-specific biases in miRNA detection. Several studies that performed library preparation and sequencing by a single laboratory site have reported similar conclusions [10–13,16–21]. In addition to these single-site studies, Giraldez et al. performed a multi-site and multi-library preparation kit study to assess, amongst others, technical biases and reproducibility of miRNA-seq protocols [22]. Similar to our study, they distributed synthetic small RNA and RNA samples to different research laboratories for library preparation. The highest reproducibility of miRNA sequencing results was obtained for technical replicates within the same kit and laboratory, followed by replicates of the same kit across different laboratories and replicates of different kits. In agreement with these independent studies, PCA analysis of our data confirmed that the greatest source of variation could be attributed to the use of different library preparation kits.

To enable a comprehensive comparison of the performance of the different library preparation kits and to reduce the influence of technical aspects such as different library purification methods and varying raw read yield, we randomly down-sampled the miRNA read counts for the synthetic miRNA and human total RNA samples to 2.5 million and 0.75 million reads respectively. This approach of single down-sampling has previously been chosen by a number of other studies [19,23]. The absolute number of miRNA read counts will differ slightly for each random down-sampling draw, therefore we repeated the random down-sampling 100 times to assess this variation. However, we found that although there were marginal changes to read numbers, the overall miRNA expression ratios did not differ within or amongst the library preparation kits.

Jayaprakash, et al. [11] showed that small RNA profiles are dependent on the adapter sequences used during library preparation and according to their recommendation a mix of adapter sequences will enable more accurate estimation of miRNA abundance. NEXTflex is the only tested kit in our study that uses this approach by including randomized adapter termini in the procedure. Compared to the three fixed-adapter kits (NEBNext, srLp and CleanTag), the overall performance of NEXTflex with respect to detection rate sensitivity, reliability and differential expression was superior. However, QIAseq and TailorMix also used fixed adapters and performed slightly better than or equally well as NEXTflex. Even though including randomized adapter sequences during library preparation seems to improve the performance of a kit, our study suggests that additional factors influence the performance. These factors might include, for example, type of ligase or ligation temperature and ligation time. Giraldez, et al. [22] have also suggested that the concentration of polyethylene glycol used during the ligation reactions affects performance, but since buffer constituents provided by commercial vendors are kept proprietary, we were unable to examine this parameter.

With the exception of the titration response assay, NEXTflex generally displayed one of the best performances, whilst NEBNext showed average performance. In particular, the NEBNext kit displayed lower miRNA detection sensitivity than the other kits. This was especially evident for the synthetic miRNA mix E, where it displayed lower reliability. Nonetheless, the NEBNext kit did not perform markedly poorer than srLp and CleanTag, which claim to require much lower input amounts (Table 1). According to the NEBNext manual, the kit allows a minimum input of 100 ng total RNA. Mix E had the lowest miRNA content (0.1 ng in 10 ng total RNA) thus it is not surprising that NEBNext showed poorer detection sensitivity compared to the other library preparation kits. It remains possible that NEBNext would equal the detection sensitivity of the other kits if presented with recommended input amounts of over 100 ng. However, some of the miRNAs remained undetected independent of their abundance levels, which indicates that additional factors influence their detection and therefore the kit performance. This is true for all tested kits: i.e. the kits appear to have preferences for certain miRNAs. It was previously suggested that the terminal nucleotides of the miRNAs influence their

detection [9] as well as the secondary structure of the miRNA [16] and co-folding between miRNA and adapter [12], which may explain the kit-specific preferences observed.

Both the NEXTflex and NEBNext kits detected fewer differentially expressed miRNAs than the kits reported previously. Whilst this is not surprising for the NEBNext kit, which appears to be less sensitive, it was unexpected for the NEXTflex kit. However, this lower sensitivity was balanced by fewer false positive calls, which might be of advantage for studies interested in finding novel biomarkers for e.g. specific diseases or treatment responses where false positives are particularly undesirable.

In conclusion, we found considerable differences between the library preparation kits when comparing their performance. Overall, QIAseq demonstrated the best performance followed by TailorMix and NEXTflex. NEBNext, srLp and CleanTag were ranked as medium performance kits. However, when it comes to accurate quantification of miRNA, all tested kits show room for improvement.

Material and Methods

The study material was described in detail in Heinicke, et al. [8]. Briefly, synthetic miRNA and human total RNA samples were used as input into library preparation. The performances of a total of eight kits (six kits from our previous and two kits from the present publication) were compared using triplicate samples as summarized below and in Fig. 1A. Synthetic miRNA samples consisted of equimolar ($n = 962$, miRXplore Universal Reference, Miltenyi, California, United States) and non-equimolar miRNA oligonucleotides ($n = 40$, Eurofins MWG Synthesis GmbH, Bavaria, Germany) which were used to create five different mixes, A-E. Mix A and B contained the same equimolar pool of miRNAs, but differed in eight concentration ratios of the 40 non-equimolar miRNAs (Supplementary Table 1 in [8]). Mix C was a 0.75 titration of mix A and 0.25 titration of mix B while the titration ratio for mix D was vice versa. Mix E equates mix A but at a 10-fold lower concentration. *Saccharomyces cerevisiae* total RNA was added to the different mixes to obtain a more complex RNA mixture. In each mix the RNA content was 2 ng/ul and miRNAs represented approximately 10% (w/w) in mix A to D and 1% (w/w) in mix E (Supplementary

Table 2. Median and standard deviation (SD) of the raw read counts passing sequencing quality filters for each kit and sample type.

Kit	Sample Type	Median	SD
NEBNext	synthetic miRNA	15,963,032	2,098,564
	human total RNA	12,945,516	934,152
NEXTflex	synthetic miRNA	15,726,206	3,428,519
	human total RNA	9,947,511	865,005
CATS	synthetic miRNA	1,657,065	1,686,647
	human total RNA	4,368,917	610,984
srLp	synthetic miRNA	21,708,163	3,074,872
	human total RNA	9,553,164	3,234,006
QIAseq	synthetic miRNA	25,025,406	4,866,588
	human total RNA	17,161,083	1,492,933
TailorMix	synthetic miRNA	12,904,412	2,208,956
	human total RNA	11,875,567	1,275,394
SMARTer	synthetic miRNA	4,817,693	2,249,898
	human total RNA	714,966	296,656
CleanTag	synthetic miRNA	10,044,117	2,055,836
	human total RNA	19,647,913	4,898,198

Table 3. Fraction of titrating miRNAs ($n = 5$) in each of the eight concentration groups. Average log expression values for the 40 non-equimolar miRNAs were calculated across the three replicates each of mix A to D. Each miRNA was scored as titrating if the average values followed the expected trend in concentrations from high to low or vice versa across mixes A to D. Grey font indicates previous results[8].

Conc. Ratio	NEBNext	NEXTflex	CleanTag	QIAseq	srLp	TailorMix
0.01	1	0.8	1	1	1	1
0.1	1	0.6	0.8	1	1	1
0.2	0.8	1	1	1	0.8	0.8
0.5	0.8	0.6	0.8	0.6	0.4	0.6
2	0.8	0.6	0.6	0.8	0.8	0.2
5	1	1.0	0.4	1	1	0.8
10	0.6	0.8	0.6	1	1	0.6
100	0.8	0.8	0.8	1	0.8	0.8

Table 2 in [8]). The intended mix ratios were verified using RT-qPCR with 16 pre-designed TaqMan Small RNA assays (Thermo Fisher Scientific, Massachusetts, United States, Supplementary Material and Methods in [8]).

Human total RNA samples were extracted from peripheral blood CD8 + T cells from a pool of either newly diagnosed RA patients ($n = 4$) or healthy controls ($n = 4$). For all samples the RNA integrity value was above 8.5.

Library preparation for all kits except NEBNext and NEXTflex was described previously (see Supplementary Material and Methods section and Supplementary Table 2 in [8]). NEBNext and NEXTflex libraries were prepared from the 21 samples described above according to manufacturer's instructions. For the synthetic miRNA mix A to D, containing 10 ng miRNA oligonucleotides, NEBNext adapters were not diluted while NEXTflex adapters were diluted 1:2. For the synthetic miRNA mix E, containing 1 ng miRNA oligonucleotides, and the human total RNA samples the adapters were diluted 1:2 for NEBNext and 1:4 for NEXTflex. Synthetic miRNA samples mix A to D were amplified using 12 PCR cycles for NEBNext and 16 PCR cycles for NEXTflex while synthetic miRNA samples mix E and human total RNA samples were amplified using 15 PCR cycles for NEBNext and 20 PCR cycles for NEXTflex. TapeStation 2200 High Sensitivity D1000 reagents (Agilent Technologies, California, USA) were used to verify the presence of miRNA library constructs at approximately 147 bp for NEBNext and 150 bp for NEXTflex. Pippin Prep (Sage Science, Massachusetts, USA) with 3% Agarose Gel Pippin Cassettes was used to remove adapter dimers and other unwanted fragments. Per lane of the Pippin Cassette five to six samples were pooled together. Size selection was optimized to cover fragments from ca. 130bp to 160bp. Final library yields and size were measured on a Bioanalyzer 2100 using high sensitivity reagents (Agilent Technologies, Supplementary Figure 11).

Libraries were sequenced on one single-read flow cell of a HiSeq 2500 (Illumina, California, United States) with 75bp reads. Each of the 21 libraries from NEBNext and NEXTflex were sequenced independently from the previously tested library preparation kits and each on a single lane (Supplementary Figure 12). Cutadapt [24] v1.15 was used to trim the following adapter sequences from the demultiplexed fastq files: AGATCGGAAGAGCACACGTCT (NEBNext) and TGGAAATTCTCGGGTGCCAAGG (NEXTflex). For NEXTflex

we additionally clipped the first and last 4 bases of the reads to remove the random 4mers that are included in the adapters. We found 59 oligonucleotide sequences from the miRXPlore Reference to be identical to sequences in the yeast *sacCer3* genome. Those sequences were removed from the synthetic miRNA reference to avoid downstream miRNA miscounting because of the yeast fragments (Supplementary Table 3 in [8]). Trimmed reads were mapped without allowing for mismatches using bowtie [25] v.1.1.2 and counted using a customized script. For all downstream analysis, except for the basic and differential expression analysis, the mapped miRNA reads were randomly down-sampled to 2.5 million reads for the synthetic miRNA and 0.75 million reads for the human total RNA samples (seed number = 123). As the miRNA read counts in single random down-samplings may differ slightly, down-sampling was also repeated 100 times in order to generate standard deviations for the miRNA read counts. To account for the heteroscedastic behaviour of miRNA-seq data, we transformed the count data using the rlog function of DESeq2 [26] v1.20.0 where necessary.

Detection rate sensitivity was assessed by investigating which miRNAs could be detected in the synthetic miRNA samples using down-sampled read count data. The intra- and inter-rater reliability of the different kits was investigated using rlog transformed downsampled data. In an additional inter-rater reliability analysis, miRNAs with an average expression in the synthetic miRNA or human total RNA samples of less than 20 CPM were filtered out. Intra- and inter-rater reliability were assessed by calculating intra-class correlation coefficients (ICC, two-way mixed model, absolute agreement and single rater), Pearson correlations and Bland-Altman agreements.

Differential expression, using edgeR [27] v3.22.3, between mix A and B for the synthetic miRNA samples and RA patients and healthy controls was assessed using the original read count data (untransformed, not down-sampled). However, read count filtering of 3 CPM in at least two libraries for the synthetic miRNA samples and 20 CPM in at least two libraries of the human total RNA samples was applied in the differential expression analysis. A miRNA was defined as significantly differentially expressed if the absolute value of the log fold change was above 1 after adjusting for multiple testing using the method of Benjamini and Hochberg, with a false discovery rate of 0.05. For the 40 non-equimolar miRNAs of the synthetic samples, we assessed the titration response in mixes A-D using the average down-sampled rlog counts for each miRNA following the data analysis previously presented by Shippy, et al. [14]. A miRNA was scored as titrating if its average expression value followed the expected concentration trend. Further details of bioinformatic analysis are given in [8].

Sequencing fastq files and miRNA count tables have been deposited in the Gene Expression Omnibus database with accession number GSE141658.

Acknowledgments

We thank Iris Langstein and Philipp Korber for providing *S. cerevisiae* RNA. Images from Servier Medical Art (Servier. www.servier.com, licensed under a Creative Commons Attribution 3.0 Unported License) were used in Fig. 1A.

Disclosure statement

The authors declare no competing interests.

Funding

This work was supported by Helse Sør-Øst RHF grants [2015034 and 2016122]. Sequencing was performed by the Norwegian Sequencing Centre (www.sequencing.uio.no), a national technology platform hosted by Oslo University Hospital and supported by the 'Functional Genomics' and 'Infrastructure' programs of Norsk Forskningsrådet and Helse Sør-Øst RHF.

ORCID

Fatima Heinicke  <http://orcid.org/0000-0001-8253-6105>
 Xiangfu Zhong  <http://orcid.org/0000-0002-1872-1186>
 Manuela Zucknick  <http://orcid.org/0000-0003-1317-7422>
 Johannes Breidenbach  <http://orcid.org/0000-0002-3137-7236>
 Magnus Leithaug  <http://orcid.org/0000-0002-0271-0677>
 Simon Rayner  <http://orcid.org/0000-0001-8703-9140>
 Benedicte A. Lie  <http://orcid.org/0000-0002-1159-7004>
 Gregor D. Gilfillan  <http://orcid.org/0000-0001-8116-0134>

References

- Chen C-Z, Li L, Lodish HF, et al. MicroRNAs modulate hematopoietic lineage differentiation. *Science*. 2004;303(5654):83–86. .
- Sempere LF, Freemantle S, Pitha-Rowe I, et al. Expression profiling of mammalian microRNAs uncovers a subset of brain-expressed microRNAs with possible roles in murine and human neuronal differentiation. *Genome Biol*. 2004;5(3):R13. .
- Gangaraju VK, Lin H. MicroRNAs: key regulators of stem cells. *Nat Rev Mol Cell Biol*. 2009;10(2):116. .
- Xiao C, Rajewsky K. MicroRNA control in the immune system: basic principles. *Cell*. 2009;136(1):26–36. .
- Calin GA, Dumitru CD, Shimizu M, et al. Frequent deletions and down-regulation of micro-RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. *Proc Natl Acad Sci*. 2002;99(24):15524–15529. .
- Peng Y, Croce CM. The role of MicroRNAs in human cancer. *Signal Transduct Target Ther*. 2016;1:15004. .
- Furer V, Greenberg JD, Attur M, et al. The role of microRNA in rheumatoid arthritis and other autoimmune diseases. *Clin Immunol*. 2010;136(1):1–15. .
- Heinicke F, Zhong X, Zucknick M, et al. Systematic assessment of commercially available low-input miRNA library preparation kits. *RNA Biol*. 2019. .
- Linsen SE, de Wit E, Janssens G, et al. Limitations and possibilities of small RNA digital gene expression profiling. *Nat Methods*. 2009;6(7):474–476. .
- Hafner M, Renwick N, Brown M, et al. RNA-ligase-dependent biases in miRNA representation in deep-sequenced small RNA cDNA libraries. *Rna*. 2011;17(9):1697–1712. .
- Jayaprakash AD, Jabado O, Brown BD, et al. Identification and remediation of biases in the activity of RNA ligases in small-RNA deep sequencing. *Nucleic Acids Res*. 2011;39(21):e141–e41. .
- Fuchs RT, Sun Z, Zhuang F, et al. Bias in ligation-based small RNA sequencing library construction is determined by adaptor and RNA structure. *PLoS One*. 2015;10(5):e0126049. .
- Zhuang F, Fuchs RT, Sun Z, et al. Structural bias in T4 RNA ligase-mediated 3'-adapter ligation. *Nucleic Acids Res*. 2012;40(7):e54–e54. .
- Shippy R, Fulmer-Smentek S, Jensen RV, et al. Using RNA sample titrations to assess microarray platform performance and normalization techniques. *Nat Biotechnol*. 2006;24(9):1123. .

- [15] Mestdagh P, Hartmann N, Baeriswyl L, et al. Evaluation of quantitative miRNA expression platforms in the microRNA quality control (miRQC) study. *Nat Methods*. 2014;11(8):809. .
- [16] Sorefan K, Pais H, Hall AE, et al. Reducing ligation bias of small RNAs in libraries for next generation sequencing. *Silence*. 2012;3(1):4. .
- [17] Shore S, Henderson JM, Lebedev A, et al. Small RNA library preparation method for next-generation sequencing using chemical modifications to prevent adapter dimer formation. *PloS One*. 2016;11(11):e0167009. .
- [18] Wright C, Rajpurohit A, Burke EE, et al. Comprehensive assessment of multiple biases in small RNA sequencing reveals significant differences in the performance of widely used methods. *BMC Genomics*. 2019;20(1):513. .
- [19] Coenen-Stass AM, Magen I, Brooks T, et al. Evaluation of methodologies for microRNA biomarker detection by next generation sequencing. *RNA Biol*. 2018;15(8):1133–1145. .
- [20] Dard-Dascot C, Naquin D, d'Aubenton-Carafa Y, et al. Systematic comparison of small RNA library preparation protocols for next-generation sequencing. *BMC Genomics*. 2018;19(1):118. .
- [21] Barberán-Soler S, Vo JM, Hogans RE, et al. Decreasing miRNA sequencing bias using a single adapter and circularization approach. *Genome Biol*. 2018;19(1):1–9. .
- [22] Giraldez MD, Spengler RM, Etheridge A, et al. Comprehensive multi-center assessment of small RNA-seq methods for quantitative miRNA profiling. *Nature Biotechnol*. 2018. .
- [23] Sundaram AY, Hughes T, Biondi S, et al. A comparative study of ChIP-seq sequencing library preparation methods. *BMC Genomics*. 2016;17(1):816. .
- [24] Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J*. 2011;17(1):10–12. .
- [25] Langmead B, Trapnell C, Pop M, et al. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*. 2009;10(3):R25. .
- [26] Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15(12):550. .
- [27] Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010;26(1):139–140. .