Systematic assessment of commercially available low-input miRNA library preparation kits

Fatima Heinicke[1], Xiangfu Zhong[1], Manuela Zucknick[2], Johannes Breidenbach[3], Arvind Sundaram[1], Siri T. Flåm[1], Magnus Leithaug[1], Marianne Dalland[1], Andrew Farmer[4], Jordana M. Henderson[5], Melanie A. Hussong[6], Pamela Moll[7], Loan Nguyen[8], Amanda McNulty[4], Jonathan M. Shaffer[6], Sabrina Shore[5,9], HoiChong Karen Yip[8], Jana Vitkovska[7], Simon Rayner[1], Benedicte A Lie[1]*, Gregor D. Gilfillan[1]*

[1] Department of Medical Genetics, Oslo University Hospital and University of Oslo, Oslo, 0450 Norway

[2] Department of Biostatistics, Oslo Centre for Biostatistics and Epidemiology, University of Oslo, Oslo, Norway

[3] Norwegian Institute for Bioeconomy Research, National Forest Inventory, Ås, 1431Norway

[4] Takara Bio USA, Inc., Mountain View, CA, USA

[5] TriLink Biotechnologies LLC, San Diego, CA, USA

[6] QIAGEN Sciences, Frederick, MD 21703, USA

[7] Lexogen GmbH, Vienna, Austria

[8] SeqMatic, LLC, Fremont, CA 94539, USA

[9] Current address: Singular Genomics, La Jolla, CA, USA

* contributed equally

Abstract

High-throughput sequencing is increasingly favoured to assay the presence and abundance of micro RNAs (miRNAs) in biological samples, even from low RNA amounts, and a number of commercial vendors now offer kits that allow miRNA sequencing from sub-nanogram (ng) inputs. However, although biases introduced during library preparation have been documented, the relative performance of current reagent kits has not been investigated in detail. Here, six commercial kits capable of handling <100ng total RNA input were used for library preparation, performed by kit manufactures, on synthetic miRNAs of known quantities and human biological total RNA samples. We compared the performance of miRNA detection sensitivity, reliability, titration response and the ability to detect differentially expressed miRNAs. In addition, we assessed the use of unique molecular identifiers sequence (UMI) tags in one kit. We observed differences in detection sensitivity and ability to identify differentially expressed miRNAs between the kits, but none were able to detect the full repertoire of expected miRNAs. The reliability within the replicates of all kits was good, while larger differences were observed between the kits, although none could accurately quantify the majority of miRNAs. UMI tags, at least within the input ranges tested, offered little advantage to improve data utility. In conclusion, biases in miRNA abundance are heavily influenced by the kit used for library preparation, suggesting that comparisons of datasets prepared by different procedures should be made with caution. This article is intended to assist researchers select the most appropriate kit for their experimental conditions.

Keywords: microRNA, miRNA, small RNA-seq, library preparation, sequencing bias, low RNA input, NGS, Next Generation Sequencing, UMI

Introduction

Micro RNAs (miRNAs) are ~22 nucleotide long non-coding small RNAs that regulate gene expression at a post-transcriptional level by binding to their mRNA targets to inhibit translation. First discovered in the early 1990s [1 2], miRNAs have been shown to impact biological processes such as cellular differentiation and development [3-6 7-9 10-12 13 14]. Alterations in miRNA expression have been observed in various diseases [15 16 17] and an accurate method for detecting and measuring miRNA expression is therefore crucial. In recent years, next generation sequencing (NGS) has evolved as the method of choice. The main advantages of NGS, compared to qPCR and microarray techniques, are the possibility to discover novel miRNAs and the ability to detect differences in miRNA sequences on a single base level. Furthermore, NGS enables the study of low-abundance miRNAs, which is especially useful when examining miRNAs in specific cell types or body fluids like serum and plasma. Accordingly, the latest miRNA library preparation kits allow inputs as low as 100 picograms total RNA. The library preparation process typically consists of (i) addition of adapter sequences onto the miRNA, (ii) reverse transcription and (iii) PCR amplification prior to sequencing. The kits investigated in this study used both two adapter and single-adapter circularization protocols which can broadly be divided into two classes: those employing RNA ligases (e.g. T4 RNA ligase) and those employing polyadenylation (poly-A) and template-switching oligonucleotides to attach adapter sequences to the single-stranded miRNAs.

Despite the reported advantages of NGS, the miRNA abundance detected by sequencing and that in the original sample have been shown to differ by up to four orders of magnitude [18]. In particular, the addition of adapters onto the miRNA insert has been identified as a major contributor to this bias [19 20]. For protocols utilizing T4 RNA ligase, adapter ligation is

71 influenced by the ligase used, the miRNA insert and adapter primary sequence, as well as the
72 GC content and the secondary structures of miRNA insert and adapter [19-23]. For poly-A
73 utilizing protocols, the enzyme poly (A) polymerase has also been reported to be influenced
74 by miRNA primary sequence and secondary structure [24]. Other reported possible sources of
75 bias during library preparation include the reverse transcription and PCR steps, with PCR in
76 particular able to introduce both amplification bias and duplicate reads, but results have been
77 contradictory [19-21]. A recent study recommended the use of unique molecular identifiers
78 (UMI) to mitigate the reverse-transcription and PCR biases in future experiments [25]. Previous
79 studies also reported that the incorporation of UMIs into sequence adapters resulted in
80 improved accuracy both in RNA-seq and smallRNA-seq analysis [26 27].
81
82 In this study we aimed to systematically assess the miRNA repertoire and frequency observed
83 in NGS data using six different low-input library preparation protocols (Table 1).
84 Commercial vendors marketing kits stating compatibility with total RNA amounts ≤100 ng
85 were invited to participate. The performance of the protocols was compared with regard to
86 their detection rate sensitivity, reliability and ability to identify differentially expressed
87 miRNAs. In addition, the relevance of UMIs was studied. All analyses were performed on
88 low-input well-defined synthetic miRNA and human-derived total RNA samples.
89

90 ## Results

91 ### Experimental design and miRNA read yields

92 Synthetic miRNA and biologically derived human total RNA samples (21 samples in total)
93 were distributed to participating companies for library preparation (Figure 1a). Upon return
94 of libraries, library yield and size were measured (Supplementary Figure 2). Libraries were

pooled based on their miRNA content with the aim of obtaining at least 5 million reads apiece. Based on the presence of additional small RNAs in some libraries, particularly those prepared with the CATS and SMARTer-beta reagents, which displayed a wider range of insert sizes, we did not expect all libraries to produce similar read counts at this stage. Nevertheless, a large number of reads from both CATS and SMARTer-beta kits failed to pass QC threshold filters.

The number of sequencing reads obtained ranged from 400,000 to more than 33 million reads for the individual libraries (Supplementary Table 4).

For all library preparation kits, the greatest proportion of reads were discarded during mapping, most likely as a result of not allowing for any mismatches (Figure 1b and Supplementary Figure 3). The absolute number of reads excluded differed between the kits. As anticipated, a higher fraction of miRNA reads was obtained in the synthetic miRNA samples compared to the human total RNA samples (Figure 1b), since the human total RNA samples also contain additional classes of small RNA. SMARTer-beta and CATS returned the lowest proportion of miRNA reads both in the synthetic miRNA and the human total RNA samples compared to the other library preparation kits while TailorMix, followed by QIAseq produced the highest proportions of miRNA reads (Figure 1b).

To comprehensively evaluate the sensitivity and reliability of the library preparation kits, the synthetic miRNA samples were randomly down-sampled to 2.5 million and human total RNA samples to 0.75 million miRNA reads. The libraries of SMARTer-beta and CATS did not reach these thresholds and were therefore excluded from further analysis. The results presented hereafter are based on calculations using the down-sampled data with the exception

of the differential expression and UMI analyses for which raw (not down-sampled) miRNA reads were used.

Detection rate sensitivity

Applying a relaxed detection threshold where miRNAs were defined as detected if one or more read counts were registered, the detection rate sensitivity for all kits ranged from 94.7% to 99.1%, and all miRNAs could be detected in at least one kit and replicate. QIAseq followed by TailorMix detected the highest numbers of miRNAs in all three replicates in all the mixes (Figure 2a). QIAseq and TailorMix also missed the fewest miRNAs in either one, two, or all three triplicates. When comparing the detection rate sensitivity of the 1.0ng synthetic miRNA samples (mix A-D) with the 0.1ng synthetic miRNA samples (mix E), no striking difference in the number of detected miRNAs could be observed for any of the kits (Figure 2a).

Most of the miRNAs that were undetected in QIAseq and TailorMix were neither detected by the other two kits (Supplementary Figures 4 and 5). TailorMix was the only reagent that detected each of the 903 equimolar miRNAs in at least one sample (Supplementary Figure 5). srLp followed by CleanTag showed the highest numbers of kit-specific undetectable miRNAs.

Analysis of the 40 non-equimolar miRNAs revealed that miRNAs undetected in one or more replicates belonged mostly to miRNAs present at low levels (Supplementary Figure 6), with QIAseq showing the highest detection rate, again followed by TailorMix. Notably, CleanTag and srLp failed to detect some miRNAs present at relatively high concentrations in all the replicates (mix C and D, Supplementary Figure 6). However, even though the majority of the

non-equimolar miRNAs could be detected in all replicates, the analysis indicated that factors in addition to miRNA abundance influence detection rate sensitivity.

We next compared the performance at different detection thresholds, i.e. 1, 10, 50, 100, 200 read counts per million (CPM) for synthetic miRNA samples in all mix triplicates for each kit (Figure 2b). With the exception of some of the non-equimolar miRNA oligonucleotides present at the lowest concentration, all synthetic miRNAs should in theory obtain CPM values above 200 with the library size of 2.5 million mapped miRNA reads. However, a sharp decline in detection was observed at increasing CPM thresholds. Nonetheless, QIAseq followed by TailorMix consistently detected the highest number of miRNAs across all thresholds.

Intra-rater and Inter-rater reliability

Rlog transformed miRNA count data were used for the for intra- and inter-rater reliability calculations. Intra-rater reliability calculations (the concordance between miRNA read counts within the replicates of the library preparation kit) revealed excellent reliability for the synthetic miRNA and the human total RNA samples within all tested kits with ICC values above 0.99 and 0.98, respectively (Supplementary Table 5). Similarly, very strong correlations were found when Pearson correlation coefficients were calculated ($r > 0.97$, $p < 0.05$) (Supplementary Figure 7 and 9a). Bland-Altman plots, which describe the agreement between two replicates by presenting the difference of them against the mean, also showed good agreement (Supplementary Figure 8 and 9b). For all Bland-Altman comparisons the bias was close to 0. The line of equality (not presented in our Bland-Altman plots) was always within the agreement limits, which indicates a lack of systematic error in

the measurements within the replicates. All in all, strong intra-rater reliabilities were observed within the samples prepared by each kit.

For the inter-rater reliability calculations (concordance of read counts seen between the different kits) the first replicate of each mix, RA or healthy control sample was randomly chosen. The synthetic miRNA and the human total RNA samples revealed good and excellent inter-rater reliability with ICC values above 0.83 and 0.95 respectively (Supplementary Table 6). The correlation between the different kits was above 0.76 ($p < 0.05$) for the synthetic miRNA and above 0.92 for the human total RNA samples (Supplementary Figure 10). However, differences in the correlations between the reagents were seen for the synthetic miRNA samples. The kits with the highest correlations ($r > 0.94$, $p < 0.05$) were, independent of whether mix A-E was considered, CleanTag and srLp while QIAseq showed the lowest correlation to the other kits. The Bland-Altman plots revealed no systematic error when comparing the different kits to each other (Supplementary Figures 11 and 12). The limits of agreements were smallest for CleanTag and srLp across all tested mixes in the synthetic miRNA samples indicating a high agreement between those two kits. In summary, a modest-to-good inter-rater reliability was obtained when comparing the mix-specific replicates of the four miRNA library preparation kits with each other, with QIAseq showing the greatest differences from the other reagents.

The reliability measured against the theoretical miRNA concentration was only assessed for the synthetic miRNA samples. For the 903 equimolar miRNAs, the fold deviation of the first replicate of mix A from the median count for that sample was calculated as a rlog ratio (Supplementary Figure 13). When the absolute value of the rlog fold deviation for a miRNA was less than or equal to one, the miRNA was counted as equimolar. For the four kits this

was the case for 39.8 to 42.0% of the equimolar miRNAs. The remaining miRNAs showed a bias towards over-representation (positive rlog fold change) rather than under-representation. The coefficient of variation of the rlog counts across all replicates for the equimolar miRNAs was lowest for QIAseq, followed by TailorMix, CleanTag and srLp, respectively (Supplementary Table 7).

For the 40 non-equimolar miRNAs, the correlation between the rlog counts of each library preparation kit and their theoretical concentration varied between the mixes for all kits. Overall, mix A and mix E showed greater correlations ($0.41 < r > 0.61$, $p < 0.05$) than mix B to mix D ($0.08 < r > 0.47$, $p < 0.05$) (Supplementary Table 8). QIAseq showed the highest correlation coefficients across all samples. All in all, these results suggest that on one hand the reliability between the reagents is good, but on the other hand that none of the reagents are ideally suited for accurate miRNA quantification.

Differential expression

Most miRNA profiling studies aim to identify differentially expressed (DE) miRNAs between samples of interest. When comparing mix A and mix B of the synthetic miRNA samples, ideally all 40 non-equimolar miRNAs should be detected as DE with a log2 fold change greater than or equal to one. All kits detected between 32 to 35 DE miRNAs (Figure 3a). However, some of those miRNAs (CleanTag, TailorMix and srLp=2 and QIAseq=1) were from the pool of equimolar miRNAs. Of the 40 non-equimolar miRNAs, 26 were detected to be DE by all kits, although they did not always agree on the log fold changes (Figure 3b). The non-equimolar miRNAs hsa-miR-1199-5p, hsa-miR-22-5p and hsa-miR-940, which were three of the ten miRNAs expected to show the lowest fold differences (fold change of 2) between mix A and mix B, could not be detected as DE by any of the reagents.

In order to control as best possible that the levels of miRNA in mix A and mix B were as expected, we performed quantitative reverse-transcriptase PCR assays on 16 selected non-equimolar miRNAs (Supplementary Figure 14), which confirmed the intended ratios in the starting material.

Differential expression analysis of the human total RNA samples revealed different numbers of DE miRNAs detected by the kits. CleanTag detected 19 DE miRNAs, QIAseq and TailorMix detected two DE miRNAs each, while srLp did not detect any (Figure 3c). With the exception of hsa-miR-486-3p, no overlap between the DE miRNAs was seen amongst the kits (Figure 3d).

Titration response

The titration response of the 40 non-equimolar miRNAs in mixes A - D (Figure 1a) was compared by scoring a miRNA as titrating or non-titrating based on detection in the expected concentration order in the four mixes. Since there were five miRNAs at each chosen concentration, the fraction of titrating miRNAs (0, 0.2, 0.4, 0.6, 0.8 or 1) was calculated for each reagent kit for each concentration group (Table 2). The highest fraction of titrating miRNAs was seen for QIAseq, which correctly scored all miRNA concentrations with greater than 2-fold differences in mix A through mix D.

Effectiveness of QIAseq unique molecular identifier sequence tags

QIAseq was the only kit included in this study that implements unique molecular identifiers (UMIs) during library preparation, which are claimed to enable more accurate quantification of miRNAs. For both synthetic miRNA and human total RNA samples, very strong Pearson

243 correlations were observed between the rlog transformed raw read and UMI counts
244 (Supplementary Figure 15). Comparison of the rlog sum of all UMI and ordinary read counts
245 revealed the sum of UMI counts to be negligibly smaller than the ordinary read counts for
246 both synthetic miRNA and human total RNA samples (Supplementary Figure 16a, b).
247
248 To further examine whether UMI read counts might reduce undesirable over-representation
249 of miRNAs that were favourably amplified or sequenced, we examined the abundance of the
250 ten miRNAs with the highest ordinary read counts for each sample and compared this to their
251 respective UMI counts (Supplementary Figure 16c,d). Amongst those miRNAs no
252 overestimation of the ordinary read counts was observed compared to the UMI counts.
253

254 **Discussion**

255 Several publications have revealed discrepancies between the frequencies of miRNAs present
256 in the original samples and those detected by sequencing approaches [18 20]. The adapter
257 ligation steps in the small library preparation procedure, in addition to miRNA sequence and
258 structure, have emerged as being most critical when trying to explain the discrepancy [18 19 21 43
259 44 20 22]. As an alternative to the ligase-dependent ligation step in library preparation, poly-
260 adenylation based procedures have been developed. Additional biases might be introduced
261 during reverse transcription and PCR steps, but in this case results have been contradictory
262 ([19-21], [25]). The use of UMI tags has therefore been suggested to remove this potential bias[25].
263 Here we performed a comprehensive comparison of six low input small RNA sequencing
264 reagents utilizing both ligase-depend, polyA-based and single-adapter methods, including one
265 kit that employed UMI tags. Note that we assessed here only the performance of the kits to
266 identify miRNAs; other small RNA species that may be captured were not assessed.

Sequencing yields and miRNA read proportions

Considerably different numbers of raw reads were obtained from the different kits. The kits from TailorMix and QIAseq returned the highest miRNA read counts both in the synthetic miRNA and the human total RNA samples. However, raw read outputs cannot be used to judge the performance of a method. Furthermore, since the samples from SMARTer-beta were sequenced alone in a single lane, we cannot exclude that technical issues affecting only that lane were responsible for the low raw read numbers that passed filters. The input range tested in this study was at or below the indicated range stated for the SMARTer-beta kit (100 ng -1 ug total RNA or 2 ng – 200 ng enriched small RNA); this may have resulted in the observed poor performance. Since this study was performed, the kit has been re-optimized and released with a new formulation and improved performance. Nonetheless, the low proportion of reads mapping to miRNAs from both the CATS and SMARTer-beta was clearly evident, which could be attributable to inefficient removal of other small RNA species during library preparation. However, greater numbers of reads that were not counted as miRNA (due to imperfect match in length to the database reference sequence) were noticeable for CATS, which may indicate that polyadenylation-based methods are trickier to process during data analysis, due to uncertainties on the length of the poly-A tail added. To reduce the influence of technical aspects (e.g. different library size selection and purification methods, as well as raw read yields) on the comparisons, all miRNA counts were down-sampled to the same levels. CATS and SMARTer-beta did not reach the selected thresholds and were therefore excluded from further analysis.

Detection rate sensitivity

When applying low detection rate sensitivity thresholds, most synthetic miRNAs could be detected by the four remaining kits, indicating that all of them may be suited to assess the overall miRNA repertoire. However, when applying more stringent detection thresholds ranging from 1cpm to 200cpm, greater differences in detection rates between the kits became evident, and QIAseq and TailorMix emerged as the most sensitive. It is worth noting that kit specific biases played a greater role in miRNA detection than input RNA amounts, at least within the ranges tested here (0.1-1.0 ng miRNA).

Reliability

Intra-rater reliability showed very high concordance between miRNA counts within the replicates of a miRNA library preparation kit, independent of the kit, for both synthetic miRNA and human total RNA inputs. Similar results have been reported by Giraldez, et al. [44] and Wright, et al. [25], although they refer to intra-rater reliability as reproducibility and consistency respectively. The intra-rater reliability was strong both for 0.1ng and 1.0ng synthetic miRNA samples for all kits in our study (data not shown) which is promising given current interest in using low RNA inputs derived from small biological specimens.

In concordance with the findings reported by Giraldez, et al. [44], Coenen-Stass, et al. [45] and Wright, et al. [25], inter-rater reliability (concordance of read counts seen between the different kits, also called reproducibility or consistency across replicates) was lower compared to the intra-rater reliability. In particular, QIAseq deviated from the other kits, but we stress that this does not indicate poorer performance. QIAseq employs a different 3´adapter sequence compared to the other three kits which may underlie the dissimilar preference for subset of miRNAs observed. These observations underscore the emerging conclusion that kit-specific

differences should be considered by any researchers comparing miRNA-seq datasets, as supported by another recent study [46]. Notably, the concordance between the miRNA counts measured and the expected concentration for the synthetic miRNA samples was low, and revealed that none of the library preparation kits could accurately quantify the majority of miRNAs.

Differential expression

Differential expression analysis of synthetic miRNA mix A versus mix B revealed that all kits could detect at least 31 out of 40 non-equimolar miRNA correctly as DE (fold change ≥ 2). MiRNAs hsa-miR-1199-5p, hsa-miR-22-5p and hsa-miR-940 were never detected as DE by any of the kits. These miRNAs were present at two-fold concentration differences, the lowest fold change tested, which can be challenging. In general, all reagents displayed greater problems to detect small fold-change differences, reminiscent of results seen in the recent study by Giraldez, et al. [44].

Our study offered the additional possibility to study levels of false positive DE miRNAs detected from the 903 equimolar miRNAs. Equimolar miRNAs found to be DE were characteristically detected as DE with low fold-changes and showed little agreement between the kits, consistent with their being false positive calls. Taken together, QIAseq showed slightly higher sensitivity (true positives) and slightly higher specificity (fewer false positives) than the other reagents, although the false-positive calls did fall within the expected rate set for the analysis (False discovery rate = 0.05). Reinforcing these conclusions, the titration response analysis clearly demonstrated the superior performance of the QIAseq reagents to most faithfully represent the levels of miRNAs in input material.

It nonetheless appears that the different reagents have differing preferences for particular miRNAs. The primary sequence of terminal miRNA nucleotides [18], secondary structure affecting ligation sites [47] and co-folding of the miRNA and ligated adapters [21] have all been documented as sources of bias affecting miRNA detection. Interestingly, the 3' adapter sequence in the QIAseq kit differs from the other three kits analysed. However, our attempts to explain the differences observed between the kits based on primary sequence or secondary structure analyses were inconclusive (data not shown).

Greater differences between kits were observed by examining DE miRNAs detected when comparing the RA patient pool and healthy control pool of human CD8+ T cell RNA, where the number of DE miRNA varied between none (srLp) to 19 (CleanTag). There are few preceding studies of miRNAs from blood-isolated CD8+ cells in rheumatoid arthritis, but some of the miRNAs found to be DE in this study have previously been associated with RA, e.g. miR-221-3p [48], miR-223-3p [49-51], miR-374b-5p [52] and miR-486-3p [52], however further confirmation is needed. Worryingly, in addition to the varying number of DE miRNA detected by the different kits, there was almost no concordance between the miRNAs identified. Taken together, it is advisable to interpret DE miRNA results from studies employing different library preparation methods with caution.

Re-analysis of QIAseq dataset utilising UMIs

Reverse transcription and PCR-amplification may be potential sources of bias during library preparation, and PCR can also introduce duplicate reads. QIAseq was the only kit tested to address the issue of duplicate reads by the inclusion of UMIs, however, under the employed conditions, no appreciable difference between UMI counts and the ordinary read counts were detected, mirroring the findings of Wong, et al. [46]. Fu, et al. [27] observed that higher fractions

of PCR duplicates could be observed when reducing the starting material, but when comparing the 1.0ng and 0.1ng synthetic miRNA samples, no difference in the proportion of PCR duplicates was seen. Nonetheless, it remains possible that at lower concentrations than tested here, UMIs may prove useful for the elimination of duplicates to improve dataset quality.

In conclusion, the QIAseq kit from QIAGEN consistently demonstrated performance at, or near, the top for all metrics examined. It should be mentioned that QIAGEN made an error affecting samples 1-8 in their first attempt at library preparation and were supplied with replacements. With the exception of performance in the titration response assay, the TailorMix kit from SeqMatic closely followed. Lexogen´s srLp and Trilink´s CleanTag kit also performed well, and the majority of differences we detected point to kit-specific biases. However, whilst the experiments conducted here show that sequencing is a very sensitive method for detecting miRNAs, even at low abundance, it is also clear that none of the kits performed impressively with regard to accurately reflecting the relative input levels of all miRNAs. There is clearly room for improvements in this regard for the development of further enhanced reagents or methods to accurately quantitate miRNA levels.

Material and Methods

Study material

The performance of six miRNA library preparation kits was examined using low-input material consisting of synthetic miRNA samples or human-derived total RNA samples. To maximize the possibility that each procedure was performed under optimum conditions,

388 samples were distributed to the kit vendors for library construction. Sequencing libraries were
389 returned to the Norwegian Sequencing Centre for sequencing and data analysis.

390 **Synthetic miRNA samples**

391 The synthetic miRNA samples consisted of a mixture of equimolar and non-equimolar
392 miRNAs. The miRXplore Universal Reference (Miltenyi, California, United States),
393 comprising 962 HPLC purified, 5' phosphorylated, synthetic oligonucleotides of human,
394 mouse, rat and viral miRNA origin, was used as an equimolar miRNA pool. For the non-
395 equimolar pool, 40 additional HPLC purified, 5'phosphorylated, synthetic oligonucleotides
396 representing human miRNA were purchased from Eurofins MWG Synthesis GmbH (Bavaria,
397 Germany). Altogether five different miRNA mixes were created (denoted mix A to mix E,
398 Figure 1a). Mix A and Mix B consisted of the equimolar miRNA pool supplemented with the
399 non-equimolar pool present at eight different concentration ratios between the two mixes
400 spanning a 100-fold range (Supplementary Table S1). Mix C was a titration of 0.75 mix A
401 and 0.25 mix B, while mix D was a titration of 0.25 mix A and 0.75 mix B. In the case of
402 mixes A-D, the total miRNA concentration was 30 nM, with individual equimolar miRNAs
403 present at 30 pM and other miRNAs ranging from 3 – 300 pM. Mix E consisted of the same
404 miRNAs as mix A but at a 10-fold lower concentration. Due to the low concentrations in the
405 five synthetic miRNA mixes, the samples were blended with yeast (*Saccharomyces*
406 *cerevisiae)* total RNA, which does not contain known endogenous miRNAs [28], to minimise
407 degradation and loss of material due to adhesion to plasticware, and to mimic the more
408 complex total RNA mixtures encountered under typical usage. In each mix, the final RNA
409 content was 2 ng/µl, with miRNA representing approx. 10% (w/w) of the total amount (mixes
410 A-D) or 1% (mix E). The samples were distributed in triplicates to the participating vendors.
411 To each of the triplicates in mix A to mix E, one additional specific miRNA (miR-147a, miR-
412 212-3p or miR-412-3p) was added to check that the replicates were processed independently

413 throughout library preparation and were not combined into a single sample to increase
414 reproducibility.
415
416 To verify the intended ratios of the synthetic miRNA sample starting material, quantitative
417 reverse-transcriptase PCR was performed using 16 pre-designed TaqMan® Small RNA
418 assays (Thermo Fisher Scientific, Waltham, MA USA) according to manufacturer's
419 instructions. Assay details are provided in Supplementary Material and Methods. Relative
420 abundances of miRNAs in mixes A and B were measured by absolute quantification relative
421 to a standard curve.

422 **Human-derived total RNA samples**

423 Peripheral blood CD8+ T cells were magnetically sorted from newly diagnosed rheumatoid
424 arthritis (RA) patients (n=4) and healthy controls (n=4) using the EasySep cell isolation
425 system (Stemcell technologies, Vancouver, Canada). The RNA/DNA/Protein Purification Kit
426 (Norgen Biotek, Ontario, Canada) was used to isolate total RNA. Only RNA samples with
427 RNA integrity values above 8.5 were used for downstream analysis. To ensure the desired
428 amount of total RNA input for the miRNA library preparation, the four RA patients and the
429 four healthy controls were mixed together to obtain one pooled RA and one pooled healthy
430 control sample respectively. Triplicates of these different sample types were distributed to the
431 participants.

432 **miRNA library preparation**

433 Each participant was asked to prepare miRNA libraries from the 21 samples described above
434 using their specific miRNA library preparation kit. For optimization purposes the participants
435 received a further 20 ng of synthetic miRNA (blend of Mix A and Mix B) and 200 ng total
436 human RNA. All participants were requested to use the same Illumina i7 index sequence for

437 the same sample to avoid any possible effect of these sequences on the downstream library
438 preparation and sequencing process. Detailed sample and index information can be found in
439 Supplementary Table S2.
440
441 At the time of writing, four of the six kits were commercially available in the formats used
442 for this study (CATs, QIAseq, CleanTag and TailorMix). A fifth kit, srLp, was also
443 commercially available, but with different index primer sequences. For comparison purposes
444 and to avoid possible bias arising from the use of different indexes, this participant
445 synthesised custom index primers complying with the index sequences specified in this
446 article. The SMARTer kit used in the study had not been released for purchase, but a
447 modified version is now available. It should be noted that this study is not exhaustive, since
448 two library preparation suppliers meeting the input amount inclusion criteria (PerkinElmer,
449 formerly Bioo Scientific, and NEB) declined to participate. Detailed descriptions of the
450 library preparation conditions employed by the producers of the specific reagents are
451 supplied in the Supplementary Material and Methods.
452

453 **Sequencing**

454 All libraries were sequenced at the Norwegian Sequencing Center on the same single-read
455 flow cell of a HiSeq 2500 (Illumina, San Diego, CA) with 75 bp reads generated using v4
456 clustering and SBS reagents according to the manufacturer´s instructions. To avoid
457 sequencing lane bias, the libraries of srLp, QIAseq, TailorMix, CATS and CleanTag were
458 randomly distributed over five lanes of the flow cell, equivalent to sequencing 21 libraries per
459 lane (Supplementary Table S4). Due to concerns that the SMARTer beta libraries contained a
460 large proportion of non-miRNA inserts (higher molecular weight products than expected,
461 making it challenging to obtain equivalent numbers of reads per sample), these libraries were

sequenced independently from the other participants on a single lane (Supplementary Figure 1),

Bioinformatic analysis

Read mapping and reference sequences

Primary base calling and quality scoring was performed using RTA v1.18.66.4 (Illumina), followed by demultiplexing and processing with Bcl2fastq v2.18.0.12 (Illumina).

For trimming of the 3' adapter, cutadapt v1.15 [29] with parameter –m 10 was used. Detailed information about adapter sequences is provided in the Supplementary Material and Methods.

Read mapping was performed using bowtie v1.1.2 [30] with parameters –a and --norc. No mismatch was allowed. As reference, the expected pools of synthetic miRNAs (962 synthetic equimolar miRNAs originating from the miRXplore universal reference and 40 non-equimolar miRNAs) were used for the synthetic miRNA samples, and the mature human miRNA sequences specified in miRBase[31] v21 for the human total RNA samples. We confirmed that all replicates had been processed separately by verifying the presence/absence of spiked replicate-specific miRNAs in the datasets from each sample. Further analysis revealed that 59 of the miRNA sequences included in the miRXplore Universal Reference were identical to sequences in the *Saccharomyces cerevisae* (sacCer3) genome (Supplementary Table S3). To avoid potential miscounting of yeast fragments in the downstream analysis, these miRNA were excluded and only the remaining 903 miRNA of the miRXplore Universal Reference were analysed further. Mapped reads (restricted to miRNAs matching exactly to the reference sequence and length) were counted using a custom python script (available upon request).

Read count modelling

With the exception of differential expression and UMI analysis, all further downstream analyses were performed on down-sampled mapped miRNA reads to minimise confounding factors arising from sources such as read numbers and proportions of adapter dimer reads, which can be influenced by the purification method chosen and by pipetting errors. Random down-sampling to 2.5 million reads was performed for the synthetic miRNA samples and to 0.75 million reads for the human total RNA samples. The seed number was set to 123.

In miRNA-seq count data, the average observed variance across samples increases with higher average expression of the miRNA. If this heteroscedastic behaviour of the count data is not taken into account, the results of most downstream analyses will be dominated by highly expressed and highly variable miRNAs. We therefore transformed count data, where indicated, with the rlog function of DeSeq2[32] (v1.20.0), which produces a superior homoscedastic output than log2 transformation for low- and high-expressed genes [32].

Data analysis

Detection rate sensitivity and reliability

Data and statistical analyses were performed using R v3.5.2 [33] and Python v2.7.13. Unless otherwise stated, ggplot2 [34] was used for data visualization. Synthetic miRNA and human total RNA down-sampled read count data were used in the detection rate sensitivity analysis. Upset plots were produced using the R package UpSetR[35] v1.4.0.

Rlog transformed synthetic miRNA and human total RNA count data were used for assessing the reliability of the library preparation kits, on which intra-class correlation (ICC), Pearson correlation and Bland-Altman agreements calculations were performed. For ICC, the two-way mixed effects model, absolute agreement and single rater (ICC(3,1)) were applied using

509 the R package psych[36] v1.8.4. ICC values were interpreted according to the recommendations
510 of Koo and Li [37] where ICC values above 0.9, between 0.75 and 0.9, between 0.5 and 0.75
511 and below 0.5 indicate excellent, good, moderate and poor reliability respectively. Thresholds
512 described by Chan [38] were used for the Pearson correlation where correlations above 0.8,
513 between 0.6 and 0.8, between 0.3 and 0.6 and below 0.3 are described as very strong,
514 moderately strong, fair and poor respectively. The R corrplot package[39] v0.84 was utilized for
515 correlation plots and the R BlandAltmanLeh package[40] v0.31 for Bland Altman calculations.

516 **Differential expression and titration response**

517 Original read count data of mix A and mix B were used for the differential expression
518 analysis using the R package edgeR[41] v3.22.3. For the synthetic miRNA samples a read count
519 filtering of 3 counts per million (cpm) in at least two libraries was applied to the differential
520 expression analysis while a filter of 20cpm in at least two libraries was used for the human
521 total RNA samples. miRNAs were defined as significantly differentially expressed after
522 multiple testing adjustment with the methods of Benjamini and Hochberg controlling for a
523 false discovery rate of 0.05. In addition, only those miRNA with |log2 FC| > 1 between the
524 tested conditions were kept.
525
526 The titration response of the 40 non-equimolar miRNAs of the synthetic miRNA samples was
527 examined in mixes A to D according to the analyses published by Shippy, et al. [42]. Average
528 rlog expression values for each miRNA were calculated across the three replicates of each of
529 mixes A to D. If the average expression values for each miRNA followed the expected
530 concentration trend (across the four possible concentrations seen in each mix), it was scored
531 as titrating. Any deviations from the expected trend were scored as non-titrating.

UMI analysis

QIAGEN's analysis tool Geneglobe was used for assessing the effectiveness of QIAseq's UMIs. For the synthetic miRNA samples the option "other" was chosen for mapping while "human" was chosen for the human total RNA samples during the primary data analysis. The resulting count table included UMI (after PCR duplicate removal) and raw (before PCR duplicate removal) read counts for each miRNA in the samples. Before analysing the correlation between UMI and raw read counts, the counts were rlog transformed.

Acknowledgements

We thank Iris Langstein and Philipp Korber for *S. cerevisiae* RNA. Dalia Daujotyte and Stephanie Bannister are also thanked for reviewing the results and manuscript.

Funding statement

This work was supported by the South-Eastern Regional Health Authorities under Grant 2015034 and under Grant 2016122.

Data availability statement

Raw sequencing fastq files and miRNA count tables will be made available in the Gene Expression Omnibus database.

Disclosure of interest

AF and AM are employees of Takara Bio USA Inc., JMH is an employee of, and SS a former employee, of TriLink Biotechnologies LLC. MAH and JMS are employees of QIAGEN Sciences. PM and JV are employees of Lexogen GmbH. LN and HKY are employees of SeqMatic LLC. FH, XZ, MZ, JB, AS, STF, ML, MD, SR, BAL and GDG report no conflict of interest.

References

1. Wightman B, Ha I, Ruvkun G. Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. *Cell* 1993;75(5):855-62.
2. Lee RC, Feinbaum RL, Ambros V. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* 1993;75(5):843-54.
3. Chen C-Z, Li L, Lodish HF, et al. MicroRNAs modulate hematopoietic lineage differentiation. *Science* 2004;303(5654):83-86.
4. Ramkissoon SH, Mainwaring LA, Ogasawara Y, et al. Hematopoietic-specific microRNA expression in human cells. *Leukemia Research* 2006;30(5):643-47.
5. 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 Biology* 2004;5(3):R13.
6. Giraldez AJ, Cinalli RM, Glasner ME, et al. MicroRNAs regulate brain morphogenesis in zebrafish. *Science* 2005;308(5723):833-38.
7. Bernstein E, Kim SY, Carmell MA, et al. Dicer is essential for mouse development. *Nature Genetics* 2003;35(3):215.
8. Murchison EP, Partridge JF, Tam OH, et al. Characterization of Dicer-deficient murine embryonic stem cells. *Proceedings of the National Academy of Sciences of the United States of America* 2005;102(34):12135-40.
9. Gangaraju VK, Lin H. MicroRNAs: key regulators of stem cells. *Nature Reviews Molecular Cell Biology* 2009;10(2):116.
10. Thai T-H, Calado DP, Casola S, et al. Regulation of the germinal center response by microRNA-155. *Science* 2007;316(5824):604-08.
11. Xiao C, Rajewsky K. MicroRNA control in the immune system: basic principles. *Cell* 2009;136(1):26-36.
12. O'Connell RM, Rao DS, Chaudhuri AA, et al. Physiological and pathological roles for microRNAs in the immune system. *Nature Reviews Immunology* 2010;10(2):111.
13. Chen J-F, Mandel EM, Thomson JM, et al. The role of microRNA-1 and microRNA-133 in skeletal muscle proliferation and differentiation. *Nature Genetics* 2006;38(2):228.

14. Zhao Y, Samal E, Srivastava D. Serum response factor regulates a muscle-specific microRNA that targets Hand2 during cardiogenesis. *Nature* 2005;436(7048):214.
15. 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. *Proceedings of the National Academy of Sciences* 2002;99(24):15524-29.
16. Peng Y, Croce CM. The role of MicroRNAs in human cancer. *Signal transduction and targeted therapy* 2016;1:15004.
17. Furer V, Greenberg JD, Attur M, et al. The role of microRNA in rheumatoid arthritis and other autoimmune diseases. *Clinical Immunology* 2010;136(1):1-15.
18. Linsen SE, de Wit E, Janssens G, et al. Limitations and possibilities of small RNA digital gene expression profiling. *Nature methods* 2009;6(7):474.
19. 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 Research* 2011;39(21):e141-e41.
20. 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):1697-712.
21. 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.
22. Zhuang F, Fuchs RT, Sun Z, et al. Structural bias in T4 RNA ligase-mediated 3′-adapter ligation. *Nucleic Acids Research* 2012;40(7):e54-e54.
23. 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.
24. Yehudai-Resheff S, Schuster G. Characterization of the E. coli poly (A) polymerase: nucleotide specificity, RNA-binding affinities and RNA structure dependence. *Nucleic Acids Research* 2000;28(5):1139-44.
25. 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. doi: 10.1186/s12864-019-5870-3
26. Hong J, Gresham D. Incorporation of unique molecular identifiers in TruSeq adapters improves the accuracy of quantitative sequencing. *BioTechniques* 2017;63(5):221-26.
27. Fu Y, Wu P-H, Beane T, et al. Elimination of PCR duplicates in RNA-seq and small RNA-seq using unique molecular identifiers. *BMC Genomics* 2018;19(1):531.
28. Drinnenberg IA, Weinberg DE, Xie KT, et al. RNAi in budding yeast. *Science* 2009;326(5952):544-50.
29. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet journal* 2011;17(1):pp. 10-12.
30. Langmead B, Trapnell C, Pop M, et al. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology* 2009;10(3):R25.
31. Kozomara A, Griffiths-Jones S. miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Research* 2013;42(D1):D68-D73.
32. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology* 2014;15(12):550.

635 33. R Core Team. R: A language and environment for statistical computing. 2013
636 34. Wickham H, Chang W. ggplot2: An implementation of the Grammar of Graphics.
637 *CRAN* 2008
638 35. Conway JR, Lex A, Gehlenborg N. UpSetR: an R package for the visualization of
639 intersecting sets and their properties. *Bioinformatics* 2017;33(18):2938-40.
640 36. Revelle W. Package 'psych'. *CRAN* 2015
641 37. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation
642 coefficients for reliability research. *Journal of Chiropractic Medicine*
643 2016;15(2):155-63.
644 38. Chan Y. Biostatistics 104: correlational analysis. *Singapore Med J* 2003;44(12):614-9.
645 39. Wei T, Simko V, Levy M, et al. Package 'corrplot'. *Statistician* 2017;56:316-24.
646 40. Lehnert B. R Package "BlandAltmanLeh". *CRAN* 2014
647 41. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for
648 differential expression analysis of digital gene expression data. *Bioinformatics*
649 2010;26(1):139-40.
650 42. Shippy R, Fulmer-Smentek S, Jensen RV, et al. Using RNA sample titrations to assess
651 microarray platform performance and normalization techniques. *Nature*
652 *Biotechnology* 2006;24(9):1123.
653 43. Dard-Dascot C, Naquin D, d'Aubenton-Carafa Y, et al. Systematic comparison of small
654 RNA library preparation protocols for next-generation sequencing. *BMC*
655 *Genomics* 2018;19(1):118.
656 44. Giraldez MD, Spengler RM, Etheridge A, et al. Comprehensive multi-center
657 assessment of small RNA-seq methods for quantitative miRNA profiling. *Nature*
658 *Biotechnology* 2018
659 45. Coenen-Stass AM, Magen I, Brooks T, et al. Evaluation of methodologies for
660 microRNA biomarker detection by next generation sequencing. *RNA Biology*
661 2018;15(8):1133-45.
662 46. Wong RK, MacMahon M, Woodside JV, et al. A comparison of RNA extraction and
663 sequencing protocols for detection of small RNAs in plasma. *BMC genomics*
664 2019;20(1):446.
665 47. Sorefan K, Pais H, Hall AE, et al. Reducing ligation bias of small RNAs in libraries for
666 next generation sequencing. *Silence* 2012;3(1):4.
667 48. Pandis I, Ospelt C, Karagianni N, et al. Identification of microRNA-221/222 and
668 microRNA-323-3p association with rheumatoid arthritis via predictions using
669 the human tumour necrosis factor transgenic mouse model. *Annals of the*
670 *Rheumatic Diseases* 2012;71(10):1716-23.
671 49. Shibuya H, Nakasa T, Adachi N, et al. Overexpression of microRNA-223 in
672 rheumatoid arthritis synovium controls osteoclast differentiation. *Modern*
673 *Rheumatology* 2013;23(4):674-85.
674 50. Fulci V, Scappucci G, Sebastiani GD, et al. miR-223 is overexpressed in T-
675 lymphocytes of patients affected by rheumatoid arthritis. *Human Immunology*
676 2010;71(2):206-11.
677 51. Lu MC, Yu CL, Chen HC, et al. Increased miR-223 expression in T cells from patients
678 with rheumatoid arthritis leads to decreased insulin-like growth
679 factor-1-mediated interleukin-10 production. *Clinical & Experimental*
680 *Immunology* 2014;177(3):641-51.

52. Murata K, Furu M, Yoshitomi H, et al. Comprehensive microRNA analysis identifies miR-24 and miR-125a-5p as plasma biomarkers for rheumatoid arthritis. *PloS One* 2013;8(7):e69118.

Appendices

Supplementary Figures (document: Supplementary_figures_Heinicke_etal2019.docx)

Supplementary Tables (document: Supplementary_tables_Heinicke_etal2019.xlsx)

Supplementary Material and Methods (document:

Supplementary_Material_and_Methods_Heinicke_etal2019.docx)

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

Method Name	Commercial supplier	Key points*	Max. input volume tolerated	Reported RNA input range (varies with type of input tested)	Maximum number of indexes available	Method types
CATS Small RNA-Seq Kit (CATS)	Diagenode s.a.	Single-tube, 4-step process of polynucleotide tailing, reverse transcription and PCR amplification. 1 purification step (2 if size selection required).	8 µl	0.1 – 100 ng	24	Poly-A based. 2-adapter procedure
Small RNA-Seq Library Prep Kit (srLp)	Lexogen GmbH [Cat. No. 052, 058]	4-step procedure of 3' adapter ligation, 5' adapter ligation, reverse transcription and PCR. 2 purification steps (3 if size selection required).	6 µl	0.05 – 1000 ng	96	Ligase based. 2-adapter procedure
QIAseq miRNA Library Kit (QIAseq)	QIAGEN [Cat. No. 331502 or 331505]	5-step procedure of 3' adapter ligation, 5' adapter ligation, reverse transcription and PCR. 2 purification steps (3 if size selection required).	5 µl	1 – 500 ng	96	Ligase based. UMI incorporated. 2-adapter procedure
TailorMix microRNA Sample Preparation Kit Version 3 (TailorMix)	SeqMatic LLC.	4-step procedure of 3' adapter ligation, 5' adapter ligation, reverse transcription and PCR. 2 purification steps, including a final PAGE gel excision.	6 µl	1– 1000 ng	96	Ligase based. 2-adapter procedure
SMARTer® miRNA-seq Kit (Beta version) (SMARTer)	Takara Bio USA Inc.	5-step procedure of 3' mono-adapter ligation, dephosphorylation, adapter dimer blocking, circularisation, reverse transcription and PCR. 2 purification steps, including a final PAGE gel excision.	4 µl	100 – 1000 ng total RNA or 2 – 200 ng enriched small RNA	48	Ligase based. Single adapter procedure

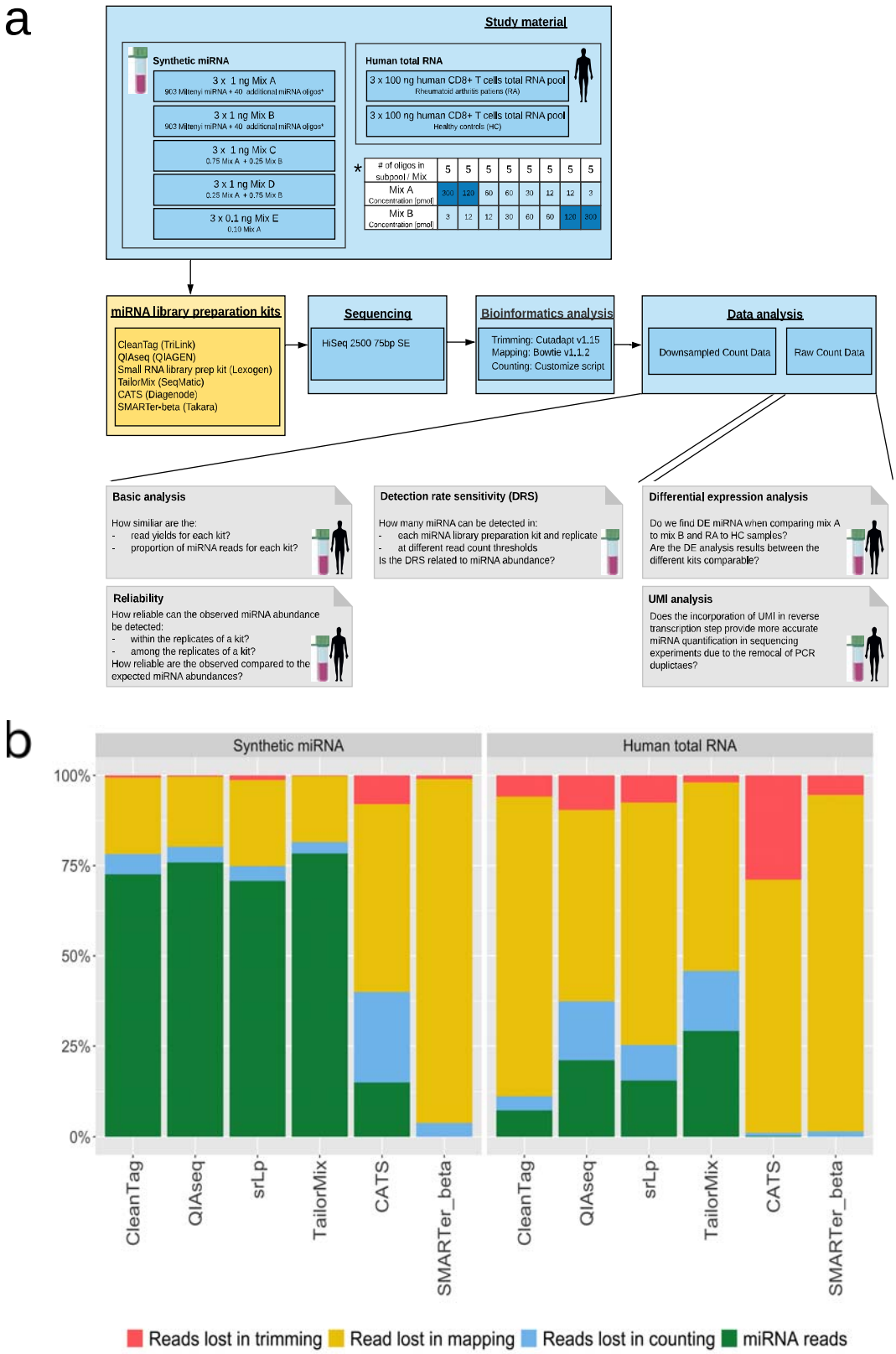
CleanTag™ Small RNA Library Prep Kit (CleanTag)	TriLink BioTechnologies, LLC.	Single-tube, 4-step procedure of 3' adapter ligation, 5' adapter ligation, reverse transcription and PCR. 1 purification step.	10µl	1 – 1000 ng	48	Ligase based. 2-adapter procedure

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

Table 2: Fraction of titrating miRNAs (n=5) in each of the eight concentration groups. Average rlog expression values for the 40 non-equimolar miRNAs were calculated across the three replicates each of mixes 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.

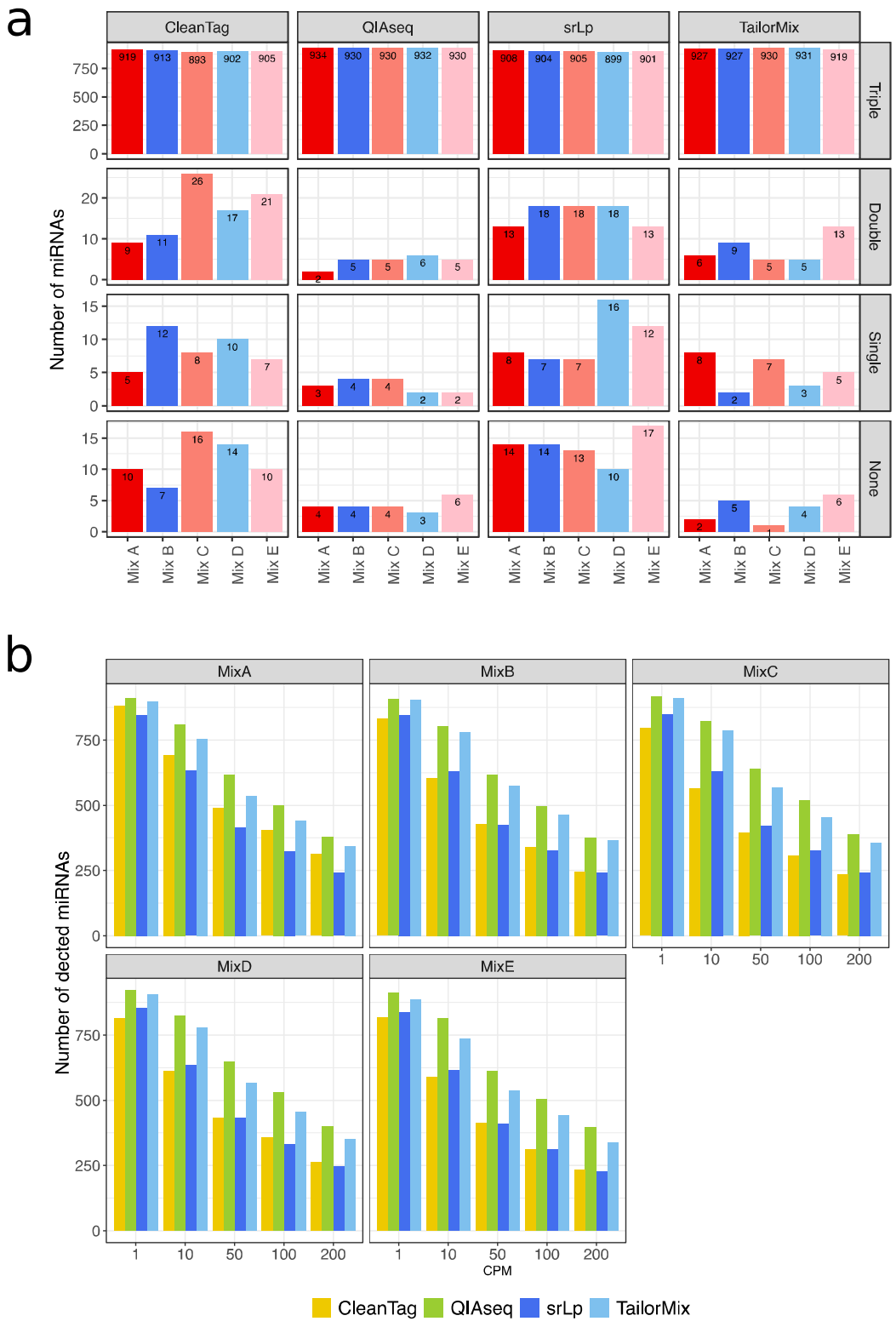
Conc. Ratio	CleanTag	QIAseq	srLp	Tailor Mix
0.01	1	1	1	1
0.1	0.8	1	1	1
0.2	1	1	0.8	0.8
0.5	0.8	0.6	0.4	0.6
2	0.6	0.8	0.8	0.2
5	0.4	1	1	0.8
10	0.6	1	1	0.6
100	0.8	1	0.8	0.8

745 **Figure 1**



746

747 **Figure 2**



748

Figure 3

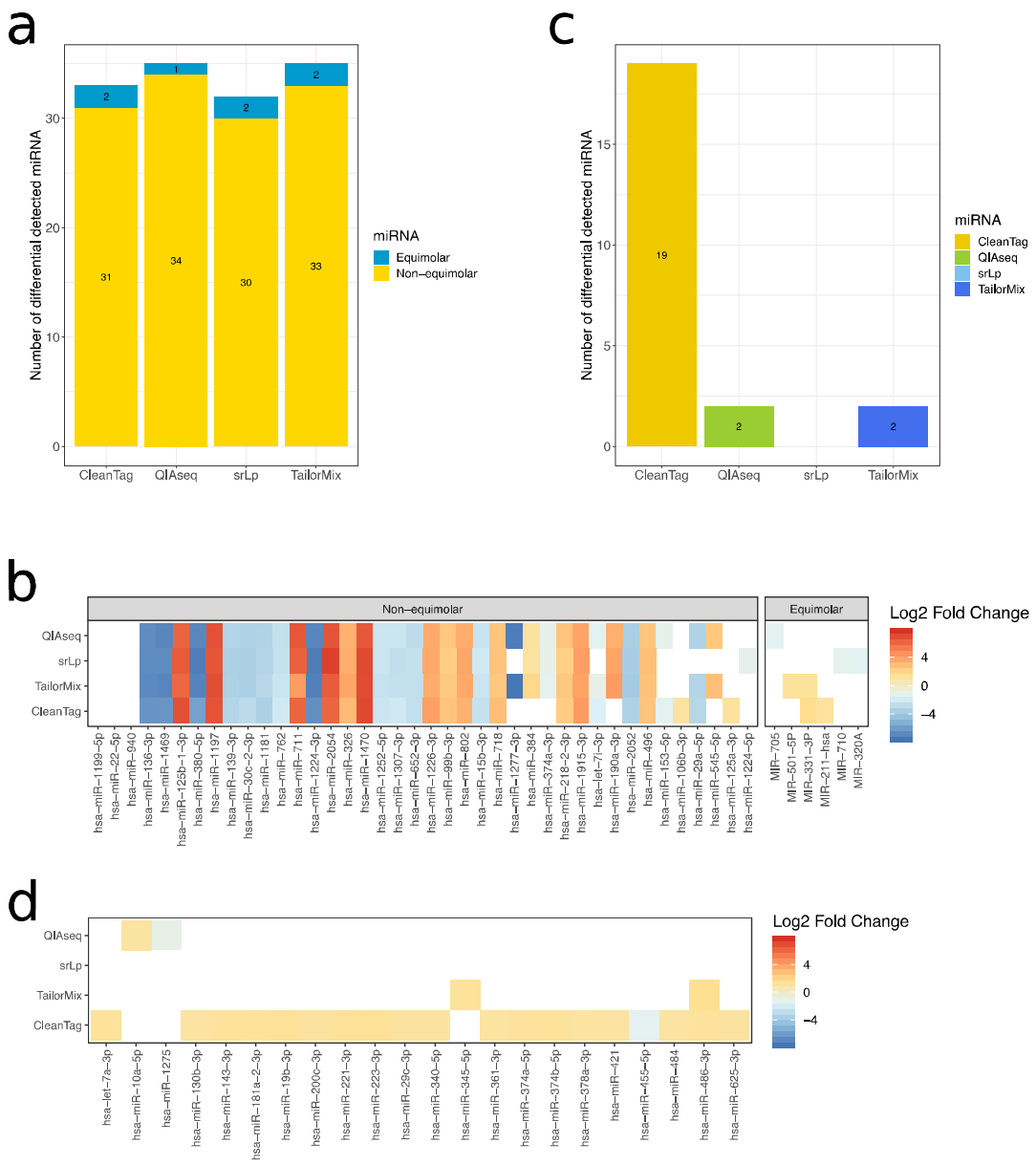


Figure	Caption
1	**Experimental design and sequencing read distribution. A: Overview of the study material, miRNA library preparation kits used, sequencing, bioinformatics and data analysis. Steps presented in blue boxes were performed in-house, while the step presented in the yellow box was executed by the indicated library preparation vendors. 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 (left) and the mean of six replicates in the human total RNA samples (right).** **Figure 1 A was created using images from Servier Medical Art (Servier. www.servier.com, licensed under a Creative Commons Attribution 3.0 Unported License).**
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. The colours of the bars represent the reagents.**
3	**Differential expression analysis. Kit-specific number of differentially expressed miRNA detected for A: synthetic miRNA samples (mix A versus mix B) and C: human total RNA samples (RA versus healthy control). miRNA-specific log2 fold changes across the different kits for B: synthetic miRNA samples and D: human total RNA samples.**

757