

METHODOLOGY ARTICLE

Open Access



A systemic approach to screening high-throughput RT-qPCR data for a suitable set of reference circulating miRNAs

Konrad Pagacz^{1,2}, Przemyslaw Kucharski^{1,3}, Urszula Smyczynska¹, Szymon Grabia^{1,3}, Dipanjan Chowdhury⁴ and Wojciech Fendler^{1,4*}

Abstract

Background: The consensus on how to choose a reference gene for serum or plasma miRNA expression qPCR studies has not been reached and none of the potential candidates have yet been convincingly validated. We proposed a new in silico approach of finding a suitable reference for human, circulating miRNAs and identified a new set of endogenous reference miRNA based on miRNA profiling experiments from Gene Expression Omnibus. We used 3 known normalization algorithms (NormFinder, BestKeeper, GeNorm) to calculate a new normalization score. We searched for a universal set of endogenous miRNAs and validated our findings on 2 new datasets using our approach.

Results: We discovered and validated a set of 13 miRNAs (miR-222, miR-92a, miR-27a, miR-17, miR-24, miR-320a, miR-25, miR-126, miR-19b, miR-199a-3p, miR-30b, miR-30c, miR-374a) that can be used to create a reliable reference combination of 3 miRNAs. We showed that on average the mean of 3 miRNAs ($p = 0.0002$) and 2 miRNAs ($p = 0.0031$) were a better reference than single miRNA. The arithmetic means of 3 miRNAs: miR-24, miR-222 and miR-27a was shown to be the most stable combination of 3 miRNAs in validation sets.

Conclusions: No single miRNA was suitable as a universal reference in serum miRNA qPCR profiling, but it was possible to designate a set of miRNAs, which consistently contributed to most stable combinations.

Background

Molecular genetics has been a major field of study in medicine and physiology since the first successful deoxyribonucleic acid (DNA) isolation as a genetic material and conception of the correct structural model of the DNA [1, 2]. Further work, beginning with the isolation of DNA polymerase I, laid the groundwork for molecular methods of quantifying gene expression [3]. Gene expression is the most fundamental level at which genes drive the phenotype, therefore its measurement remained crucial for not only genetic studies, but also any proteomics or metabolomics research. The need for a fast and reliable way of quantifying the number of copies of a specific gene's mRNAs gave rise to real-time

quantitative polymerase chain reaction (qPCR), which since 1993 arguably became the “golden standard” of gene expression quantification and still continues to be one of the most popular techniques despite the advent of the high-throughput counterparts such as next generation sequencing or hybridization microarrays [4]. In medicine, qPCR, was at first used to detect pathogens' genetic material and ribonucleic acid (RNA) molecules, among them mRNA and miRNA [5, 6].

MiRNAs represent a group of small non-coding RNA molecules consisting of usually 18–26 nucleotides. They regulate gene expression in a sequence-specific post-transcriptional manner and their expression is often altered in diseases and pathological conditions [7, 8].

A major breakthrough in the field of miRNA studies was the observation that they are stably expressed in human serum, and plasma and as such are good candidates for biomarkers of pathological conditions [9, 10]. Such studies typically use a high-throughput method to screen

* Correspondence: Wojciech_fendler@dfci.harvard.edu

¹Department of Biostatistics and Translational Medicine, Medical University of Lodz, Lodz, Poland

⁴Dana-Farber Cancer Institute, Harvard Medical School, Boston, USA
Full list of author information is available at the end of the article



for candidate miRNAs which are subsequently validated by RT-qPCR [11]. However, qPCR results are highly dependent on parameters of the reaction and varying specificity of probes – settings unique to each experiment. This makes miRNA expression values difficult to directly compare between different qPCR experiments and a wrong choice of a reference can lead to inaccurate and biased conclusions [12–14]. To further complicate matters, qPCR measures the relative abundance of a specific miRNA in the context of a reference gene (normalizer). Therefore, qPCR accuracy relies on both technical conditions and the assumption of an unaltered and stable expression of the internal reference gene. Such a normalizer should be universally abundant in all samples of the material that is investigated and be unaffected by a variety of pathological conditions. Such normalizers have been identified at tissue level and successfully used in multiple studies on mRNA and miRNA quantification alike (ACTB, GAPDH, U6) [15–17]. In biofluids however, the ideal, a single universal reference gene does not exist, and researchers often choose the normalizer for a specific experiment making it difficult if not impossible to pool the results with other studies or perform meaningful meta-analyses. Therefore, the choice of the reference is a crucial and essential step in every qPCR analysis and should be validated on the data acquired in different conditions.

The consensus on how to choose a reference genes for serum or plasma miRNA expression qPCR studies has not been reached and none of the potential candidates have yet been convincingly validated [18, 19]. The most common protocols of normalization involve finding the most stable endogenous reference on an ad hoc, study-specific basis, focusing on normalizers efficient in specific diseases [20–24]; normalization to small-nucleolar RNAs (snoRNAs) such as RNU6B [25, 26] or, when qPCR arrays are used, normalization to mean expression of all miRNAs [27]. The latter approach may only be applied when an array of multiple miRNAs is used, making it unsuitable for validation studies of specific miRNAs or panels; the other approaches hinder the potential of comparing results between studies or rely on different RNA classes which may vary from miRNAs in terms of stability, dynamic range and amplification efficiency [28–30].

Thus, the hunt for the internal reference gene or a set of reference genes adequate for qPCR analysis of serum miRNAs continues. In this article, we proposed a new design of reference gene selection – employing four different methods of measuring expression stability, we created a framework for identification of reference miRNA sets of a variable number of elements– and tested it on all currently available datasets on Gene Expression Omnibus (GEO) platform to find the optimal set of human serum reference miRNA genes.

Results

Dataset characterisation and miRNAs filtering

We characterised the included datasets in the Additional file 1: Table S1 [31–37].

Implementation validation

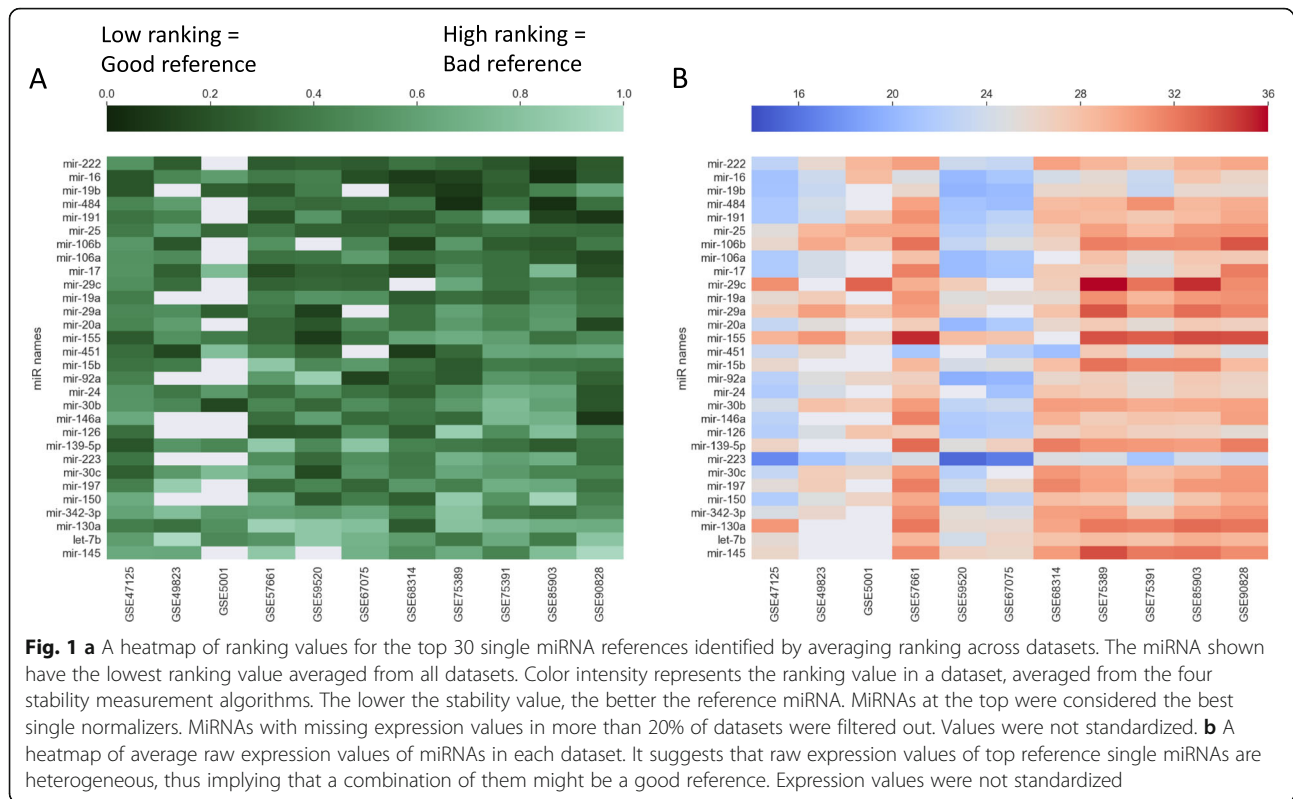
Bland-Altman analysis and Pearson's correlation ($r = 1.0000$) showed that our implementations of both versions of NormFinder stood in the excellent agreement with the original. The analysis of the raw data provided in the original GeNorm publication using MetaMirs indicated that our implementation mirrored the results obtained from the original. There was not publicly available version of BestKeeper algorithms, and any results published in the original publication, so we couldn't perform the validation for our BestKeeper implementation.

Single miRNA analysis

We found out that mean rankings, calculated from all sets, of miR-222, miR-16 and miR-19b were the lowest. We performed single miRNA analysis and aggregated the results by averaging their rankings from all datasets. This suggested that those three were the best universal, single miRNA references after selecting miRNAs that were present in more than 80% of the datasets (Fig. 1a). The heatmap of raw expression values showed great heterogeneity of expression amongst the best reference single miRNAs (Fig. 1b). We thus concluded that finding a single best normalization gene would be impossible, as not a single one miRNA achieved the lowest normalization scores in all datasets (see Additional file 1).

Comparison of rankings between single, combinations of two and three miRNAs

We compared the mean rankings of all combinations of two and three miRNAs as well as mean expression of all miRNAs using the schema shown in Fig. 2. Kruskal-Wallis testing showed $p < 0.0001$ for the comparison (Fig. 3a). In the post-hoc analysis statistically significant were the comparisons between single miRNA and two miRNAs combinations ($p = 0.0210$), three miRNAs combinations ($p < 0.0001$) and mean expression of all miRNAs ($p = 0.0025$), however the difference in the mean ranking was not significant between two and three miRNAs combinations ($p = 0.2861$). Dividing the data into datasets, it was clear that triples of miRNAs proved to be on average the best normalization factors in all datasets occupying the 1st place in rankings in all datasets (Fig. 3b). We also noted that the mean of 2 and 3 miRNAs was on average a better reference gene than its component single miRNAs (Fig. 3c, d), but in around 50% of cases at least one of the component miRNAs was a better reference than the combination. Therefore, we concluded that combinations of three miRNAs proved



to be the best normalization factors in all four algorithms.

Choice of the set of reference miRNAs

Data showed that it was impossible to find a universal single miRNA or a 2- or 3-miRNA combination, which could be reliably used in all 11 datasets as a reference gene. This was partly due to the fact that the overlap of the miRNAs' presence in the datasets was poor (see Additional file 1). However, we found out there were miRNAs that consistently created part of the top 10 reference combinations of 2 and 3 miRNAs, mainly miR-222, miR-17, miR-320a and miR-27a (see Additional file 1).

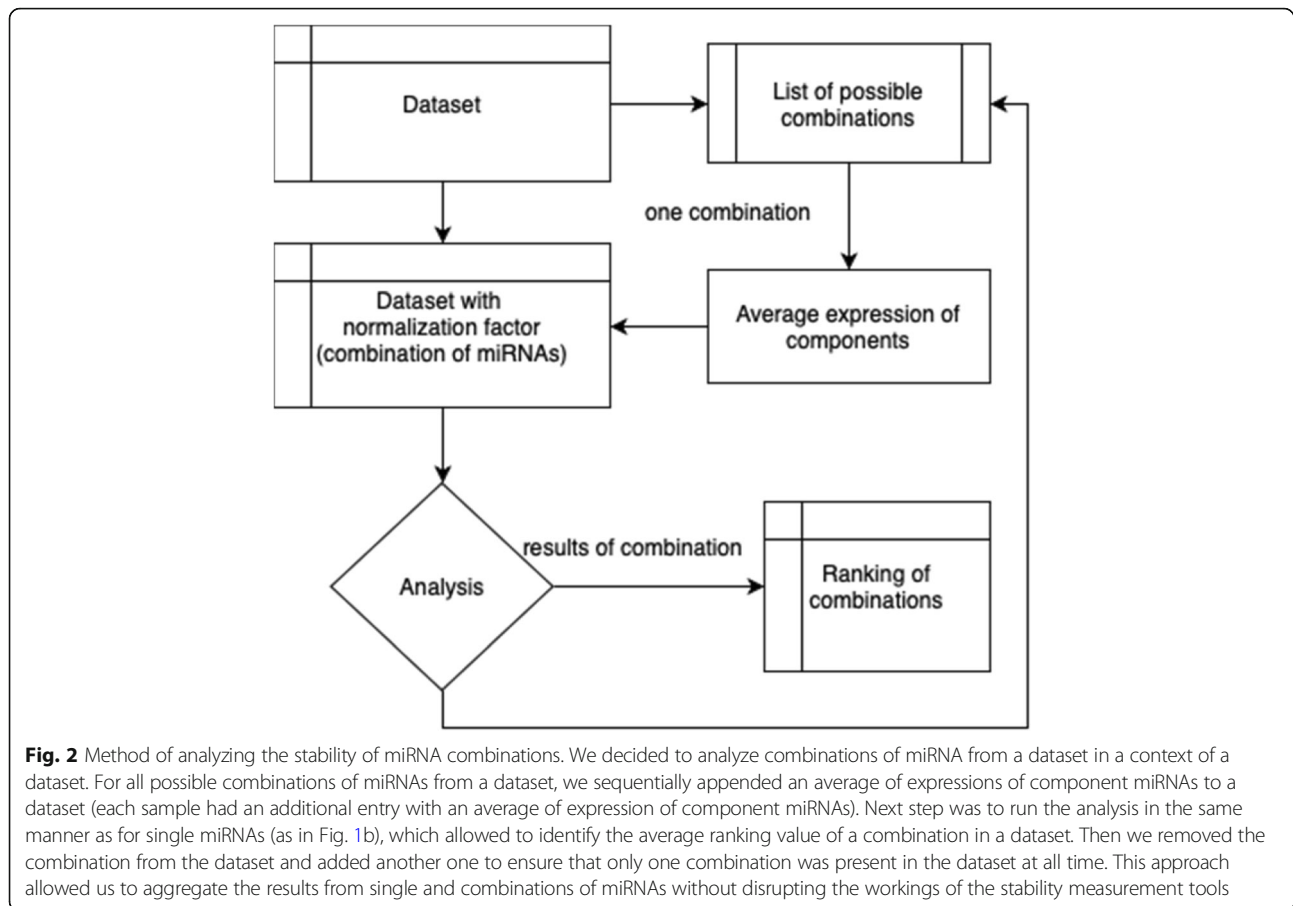
We have chosen a set of 13 reference miRNAs: miR-222, miR-92a, miR-27a, miR-17, miR-24, miR-320a, miR-25, miR-126, miR-19b, miR-199a, miR-30b, miR-30c, miR-374. According to our pipeline, we first analyzed the 11 dataset rankings of combinations of 2 miRNAs, specifically combinations that placed first in each ranking. We found out there were multiple combinations placed first in each dataset. This was possible, because our algorithm evaluated one combination at a time in the context of an original dataset. After assessing possible sets of reference miRNAs in the validation step on the dataset rankings of combinations of 3 miRNAs, we proposed a set with the lowest normalization score and with possibly minimal known dynamic range in serum. By deriving combinations of 3 miRNAs our

chosen dataset covered all first positions in the 11 dataset rankings both for 2 and 3 miRNA combinations.

Pairwise analysis of miRNAs from the 11 datasets showed the strongest affinity between: miR-374a and miR-19b, between miR-374a and miR-17, and weaker affinity between miR-25 and miR-126 (Fig. 4). miR-374a, miR-222, miR-25 and miR-126 had the highest contribution to creating the most stable combinations of 3 miRNAs (Fig. 4).

External validation of the chosen set of miRNAs

We validated the set of 13 reference miRNAs on three external qPCR datasets – two unpublished datasets from patients with head and neck tumors and one publicly available dataset from a study including patients with rheumatoid arthritis [38] – see Additional files 2, 3 and 4. Figure 5 represents the results of the external validation. Rankings of the combinations of the chosen miRNAs clustered towards lower ranking. Validation data confirmed that combinations of two and three miRNAs were a better reference than a single miRNA. We also identified that our chosen set showed low mean ranking of derived three-miRNA combinations in the overall distribution of mean ranking of combinations derived from random 13 miRNAs (Fig. 6). Average ranking of combinations derived from the chosen set was lower than 83.32, 84.76 and 97.45% of all average rankings in three validation sets, respectively. This positive control indicated that our choice



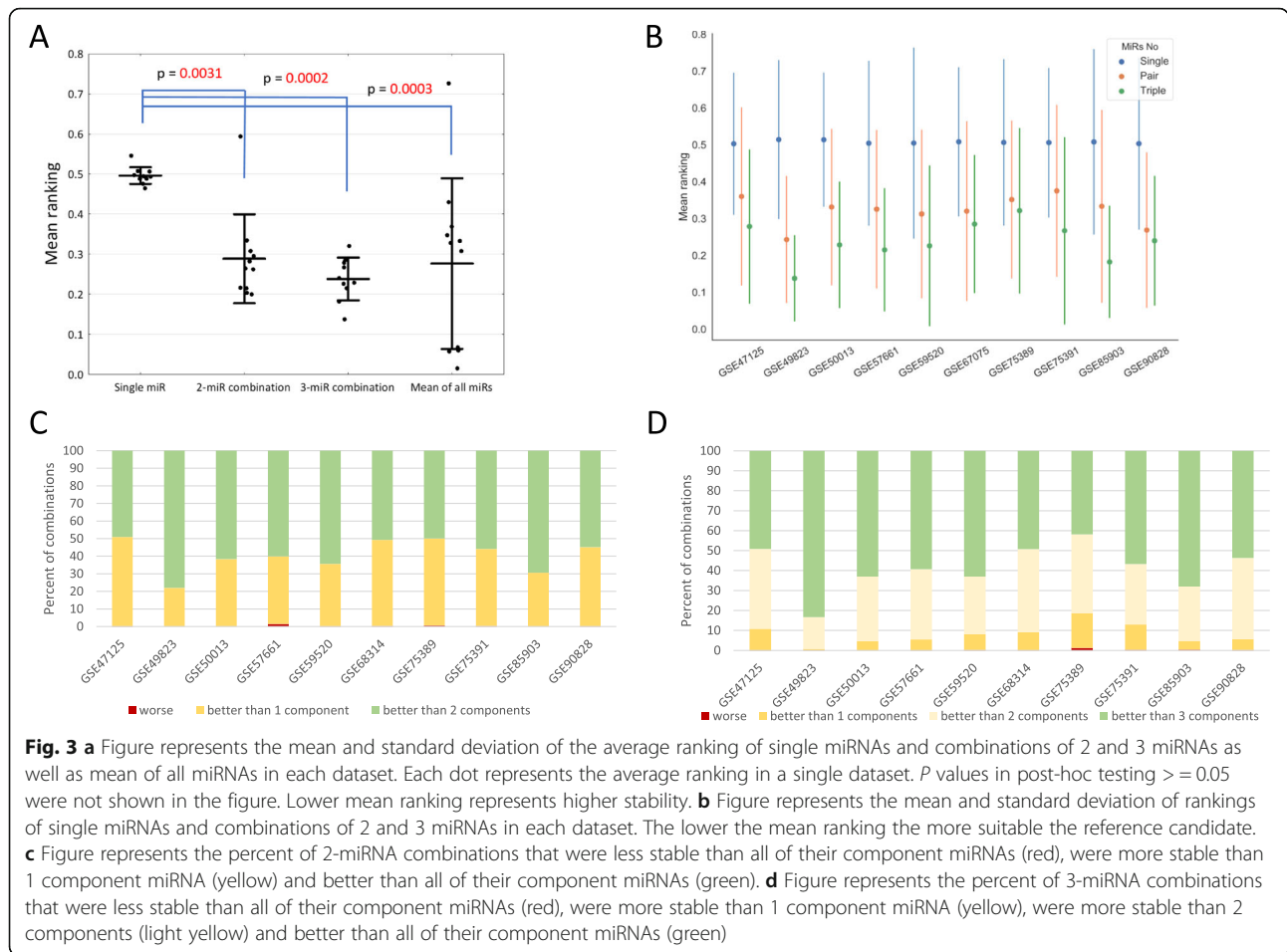
of a set created more stable references than any random 13 miRNAs, which validated our approach to selecting the set. We found out that for three external datasets the best combination of 2 chosen miRNAs placed 3rd in the combined rankings and multiple combinations of 3 chosen miRNAs placed 1st in the combined rankings (Table 1). miR-24, miR-222 and miR-27a constituted the combination with the lowest average ranking in validation analysis, among combinations of 3 miRNAs present in all two validation datasets (Additional file 1: Table S2). Detailed rankings of combinations derived from the chosen set and the best combinations in validation sets are located in the Additional files 1: Table S3 and distribution of mean rankings of combinations of 2 miRNAs in comparison with the mean of our chosen set is in the Additional file 5. As such we concluded that our normalization scheme is a valid tool for normalizing serum miRNA qPCR data and the proposed set of 13 miRNAs, emphasizing one combination of 3 miRNAs (miR-24, miR-222 and miR-27a), can be used as a viable reference for such experiments.

Discussion

Our study shows that combinations of two or preferably three miRNAs make for a better reference than single

miRNAs across a variety of clinical conditions and experimental setup. While it is difficult to pinpoint a single best combination of miRNAs that can be used in all situations, a set composed of miRNAs chosen from among: miR-222, miR-92a, miR-27a, miR-17, miR-24, miR-320a, miR-25, miR-126, miR-19b, miR-199a, miR-30b, miR-30c, miR-374a seems to be a safe, conservative choice that can be readily adopted as a standard for circulating miRNA biomarker studies.

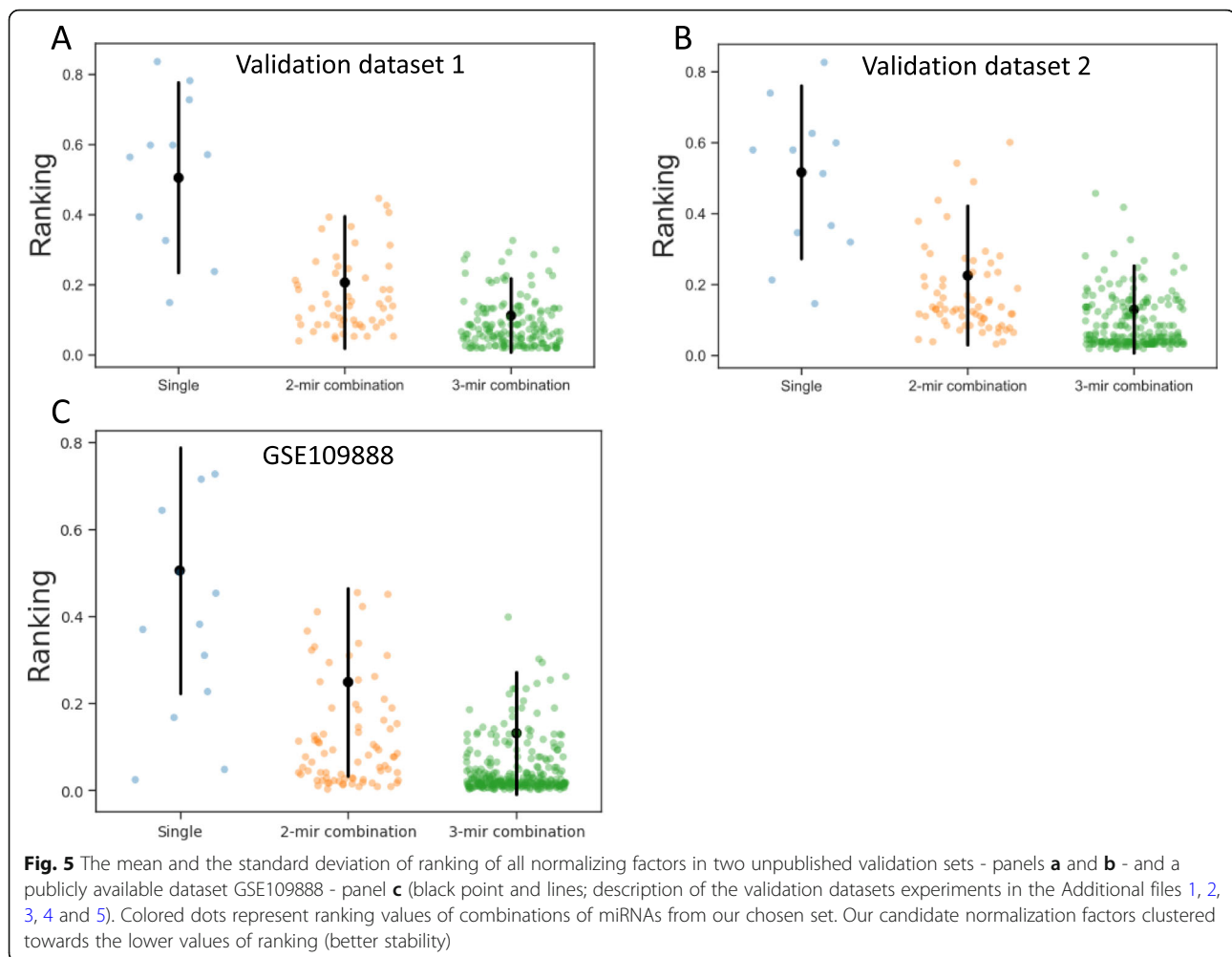
We proposed a set of miRNAs that we validated on new data to show that only 13 miRNAs were needed to be included in an analysis to acquire a stable endogenous normalization factor. We propose to normalize qPCR data to the combination of 3 miRNAs, which have the lowest normalization score, equivalent to the lowest ranking, using our algorithm pipeline and deriving combinations from the set of 13 proposed miRNAs. Our approach found a good reference in a systemic way taking into the account variety of qPCR datasets. The inclusion of datasets with different patients' conditions and treatments ensured that our results could be generalized as much as possible and the impact of different conditions of experiments on the choice was minimized.



Using spike-in reference has emerged as a trend and in fact has been used in many cases, but is not without specific drawbacks, all of which limit its applicability in biofluid studies. Spike-in methods operate on two assumptions: 1) the same amount of spike-in RNA is added to each sample; 2) synthetic spike-in transcripts behave in the same way as endogenous transcripts. It has been shown that both of those assumptions are often false and consequently disrupt the results [39, 40]. This is due to the inherent biological variability of sample storage, quality, degree of degradation and potential confounding factors. Therefore, a known-concentration spike-in may produce erroneously globally increased or decreased expression level of all evaluated miRNAs. While in experimental conditions such as cell cultures or isogenic animals, between-sample variability is largely reduced by the methodological constraints, in the clinical setting an endogenous standard is thus a far more safe point of reference as even in a degraded sample the miRNA/reference ratio should remain largely the same if both are affected by the physical, biological and chemical factors similarly. Given that our proposed references also members of the miRNA family both the investigated

ones and the reference ones should maintain their relative ratio indicative for the investigated pathological condition even across samples of varied quality. Additionally, there has been no consensus on the amount of the spike-in control added to the sample, which still leads to inter-experiment bias, while any endogenous reference potentially services more than one experiment.

The biggest obstacle to overcome in the study were long computational times. The need to calculate normalization scores for each new combination was time-consuming. Even though a single combination did not take long to calculate (time below 1 s), the sheer number of combinations going as high as 10^7 made our whole analysis take hours in the case of 2 miRNAs combination to days in the case of 3 miRNAs combinations. We explored other avenues of tackling the issue of long computation time by reducing the number of miRNAs included in the creation of combinations. We checked whether the best reference single miRNAs could be combined into the best reference combinations of 2 and 3 miRNAs. We showed that such an approach did not guarantee that the combinations would be a good reference, since some combinations created from miRNAs were worse than their component miRNAs.

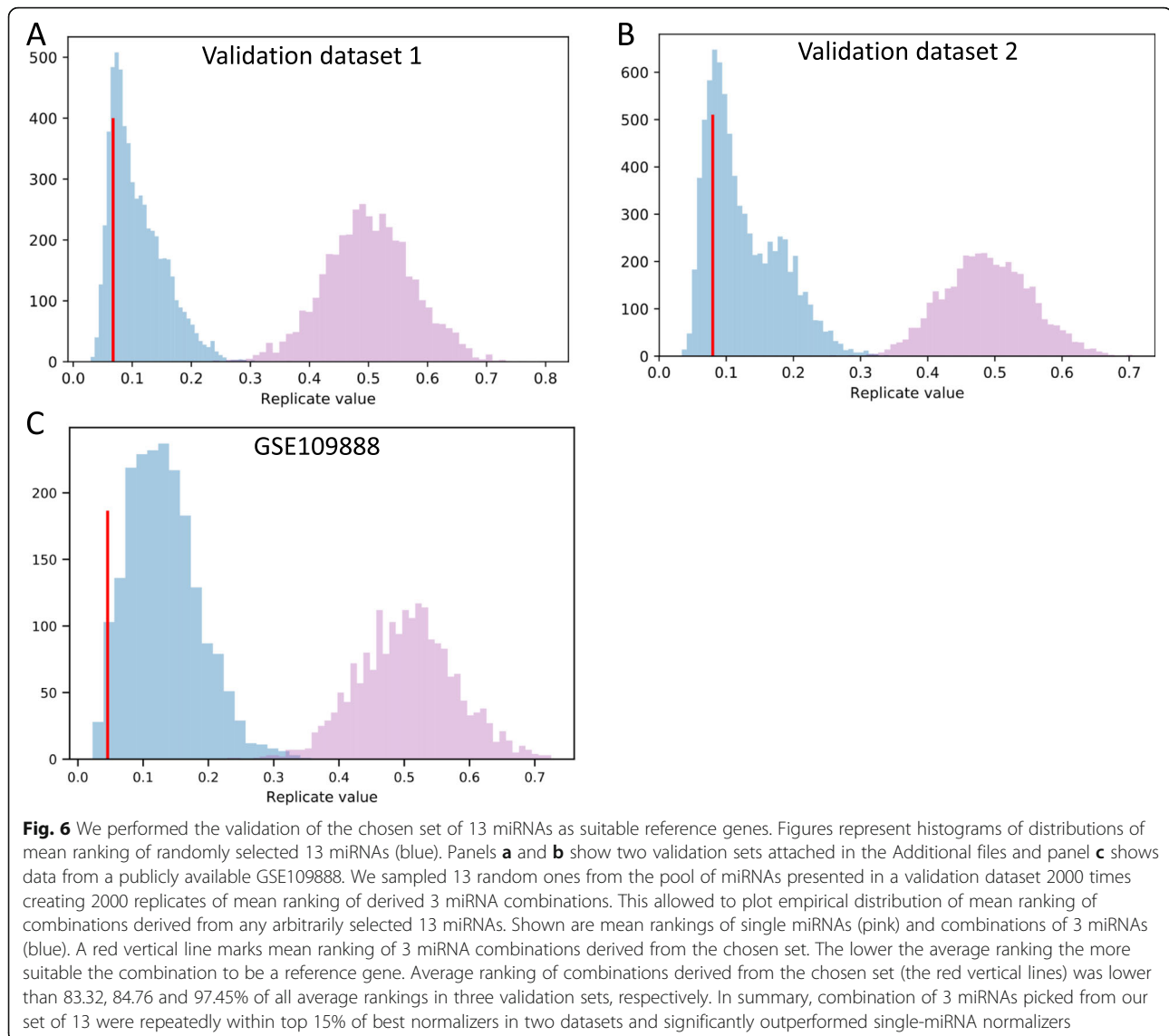


normalization scores of single miRNAs. All the studies above included a step of literature-based arbitrary pre-selection of candidate miRNAs.

The aforementioned approaches have several areas, which we improved in our work. First of all, we showed that choice of miRNA reference should not be made based on a single qPCR study, because no single good reference miRNA was reproducible in all experiments. Moreover, we proved that the mean expression of a combination of 2 or 3 miRNAs was a better reference than the expression of a single miRNA. In that regard our analysis mirrored the conclusions already made before by others [43, 44].

In order to determine the potential factors that would impair the performance of miRNAs included in our normalizer set we performed a literature search of biological significance of the chosen miRNAs. Due to large number of pathological conditions that potentially impact the levels of circulating miRNAs, we compiled a list of conditions which had been evidenced to significantly alter expression levels of the corresponding miRNAs from the proposed, reference set (Table 2). This should

allow for an informed decision about what miRNAs to include in a reference panel depending on known pathological conditions in a studied population. Moreover, we summarized the data about previous usage of aforementioned miRNAs as reference miRNAs in paragraphs below. Curiously, miR-222 has already been established as a serum reference miRNA in patients with pleural effusion and in the study of estrogen-responsive miRNAs associated with acquired protein S deficiency in pregnancy [41, 53]. Combination of 5S-rRNA and miR-92a enhanced the normalization quality compared to using only 5S-rRNA in the study of optimal small-molecular reference RNA for body fluid identification [54]. miR-27a was found to be stably expressed in rectal cancer tissue, but the downregulation of its exosomal expression has been associated with amyotrophic lateral sclerosis [55, 56]. miR-17 was found to be overexpressed in many human cancer tissues and to promote cell growth. miR-17 is a member of miR-17-92 cluster, which had been termed onco-miR-1 and its overexpression was proposed to be an early non-specific sign of cancer [57]. miR-24 was



notably a worst reference in a cardiovascular diseases' study involving 7 small non-coding RNAs (miR-16, SNOU6, 5S, miR-19b, miR-24, miR-15b, let-7i), although the authors of the study employed only one normalization algorithm – BestKeeper on top of comparative delta Ct analysis [58]. Serum expression of miR-320a was not previously considered as a reference miRNA, but was connected to several conditions including: metabolic

syndrome, epithelial ovarian cancer and inflammatory bowel disease [59, 60]. miR-25 had been previously confirmed as a suitable circulating reference gene [20, 24, 41]. Expression of circulating miR-126 was found to be associated with disease free survival in patients with squamous cell lung cancer [45]. It was also implicated in the suppression, migration and invasion of non-small-cell lung cancer cells via targeting CCR1 as well as other molecular

Table 1 Performance of the chosen miRNAs in the validation datasets

| Validation dataset | Place of the best single miRNA | Place of the best pair | Place of the best triple |
|--------------------|--------------------------------|------------------------|--------------------------|
| 1 | 6 | 3 | 1 |
| 2 | 4 | 3 | 1 |
| GSE109888 | 1 | 1 | 1 |

We validated our chosen set on two qPCR datasets. In each dataset we used our algorithms to calculate normalization rankings for all possible combinations of miRNAs present in a dataset and then noted the highest place in the rankings of the combination derived from our chosen set. Full list in the Additional file 1: Table S3. The combination with the lowest average (0.0231) ranking from the validation datasets consisted of miR-24, miR-222 and miR-27

Table 2 List of exclusions from consideration as a reference miRNA based on the known, experimentally validated connections

| microRNA | Reports on differential expression in serum or plasma due to specific clinical conditions | Reference |
|--------------|--|-------------|
| hsa-miR-222 | Gastric cancer | [45] |
| hsa-miR-92a | Acute myeloid leukemia, acute myocardial infarction, potential marker of atherosclerosis, metabolic syndrome, bladder cancer, systemic lupus erythematosus | [46–49] |
| hsa-miR-27a | Acute pulmonary embolism, colorectal cancer | [52, 61] |
| hsa-miR-17 | Elevated in many types of cancer, acute ischemic stroke, endometriosis, acute myocardial infarction | [62–64] |
| hsa-miR-24 | Coronary heart disease, type 2 diabetes mellitus, early breast cancer | [65, 66] |
| hsa-miR-320a | Esophageal adenocarcinoma, Barrett's esophagus, arrhythmogenic cardiomyopathy | [67, 68] |
| hsa-miR-25 | Osteosarcoma, breast cancer, papillary thyroid carcinoma, pancreatic cancer, gastric cancer, hepatocellular carcinoma, esophageal adenocarcinoma | [69–74] |
| hsa-miR-126 | Non-small-cell lung cancer, chronic heart failure, acute myocardial infarction, ischemic stroke, type 2 diabetes mellitus | [64, 75–78] |
| hsa-miR-19b | Knee osteoarthritis, diabetic cardiomyopathy, acute myocardial infarction, gastric cancer, prostate cancer, lung cancer | [79–84] |
| hsa-miR-199a | Glioma, hepatocellular carcinoma, acute myocardial infarction, colorectal cancer, osteosarcoma | [85–89] |
| hsa-miR-30b | Active tuberculosis | [90] |
| hsa-miR-30c | HLTV-1 infection, Duchenne muscular dystrophy, active pulmonary tuberculosis | [91–93] |
| hsa-miR-374 | Acute Graft-versus-Host disease | [94] |

We have performed a literature search in MEDLINE and PMC databases (through www.ncbi.nlm.nih.gov/pubmed portal) looking for published evidence of pathological conditions influencing levels of circulating miRNAs included in the proposed reference set. Only human and experimentally validated studies were included, and the expression of a miRNA must have differed significantly between study and control groups

functions, but it was not considered as a suitable reference gene [46, 47]. miR-19b had been found out as a decent reference miRNA in the previously mentioned evaluation of 7 potential normalizers in studies focused on cardiovascular diseases [58]. The findings of Zuberi et al. suggested that miR-199a downregulation might be a potential indicator for progression of epithelial ovarian cancer [48]. This miRNA has not been previously viewed as a potential reference miRNA. miR-30b was previously considered as a reference miRNA in self-collected cervicovaginal tissue specimens in a study that evaluated 11 reference candidate small non-coding RNAs [49]. miR-30c had been used before as a biomarker for detection of autologous blood transfusions and as an early predictor of recurrence of localized stage I non-small cell lung cancer after surgical resection [50, 51]. The search of PubMed publication database turned out 3 articles about miR-374a and none suggested specific profiles that included miR-374a nor its suitability as a reference miRNA.

Conclusions

Our work showed that using single miRNAs as references in biofluids provides a significantly worse reference for qPCR than a combination of two or three miRNAs. We also showed that the strategy of normalizing to a combination of miRNAs was more stable and predictable than normalizing to the average of expression of all miRNAs in a dataset. The proposed set of 13 miRNAs that reproducibly contributed to the selection of best normalizer combinations should be validated in further studies on patients with different clinical conditions, nonetheless we

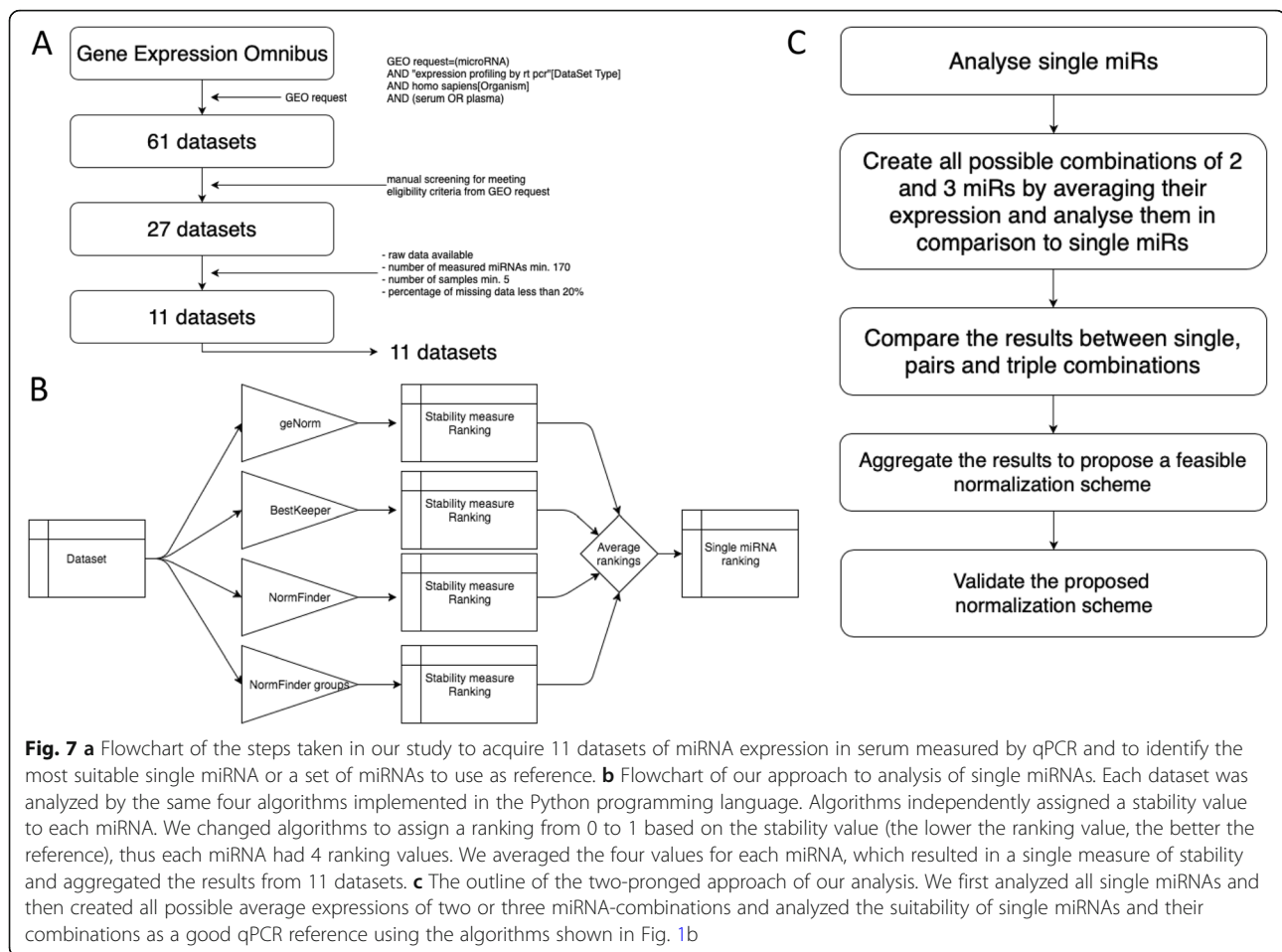
have showed that based on our in silico validation the arithmetic mean of 3 miRNAs (miR-24, miR-222 and miR-27a) was sufficiently stable to be used as a reference for serum miRNA qPCR profiling.

Methods

Data acquisition

To identify datasets pertaining to biofluid miRNA profiling with qPCR, we carried out a GEO search with a query: ((miRNA) AND “expression profiling by RT PCR”[DataSet Type]) AND *Homo sapiens* [Organism] AND (serum OR plasma), which yielded 61 hits (accession date: 30.10.2017). Afterwards, we extracted datasets, which fulfilled all of the following criteria: include miRNA high-throughput methods, raw data must be published, number of miRNAs measured at least 170 – many studies included only a limited array of miRNAs, often as a validation of several miRNAs previously identified by another measurement method, number of samples at least 5 – normalizing algorithms require it for reliable calculations, percentage of missing data lower than 20% (Fig. 7a). We manually curated each dataset removing duplicates miRNAs and chip-specific control probes. While curating a single dataset, we also removed all miRNAs, which did have missing values in that particular dataset. We treated not achieving a threshold of detection in RT-qPCR as missing values.

All the data used in this study can be accessed via the GEO platform (<https://www.ncbi.nlm.nih.gov/geo>) using accession link found in the Additional file 1: Table S1.



Algorithms implementation, normalization scoring, aggregating results from different algorithms

We implemented three algorithms: BestKeeper, NormFinder (in two versions), GeNorm in the Python programming language following formulas extracted from the appropriate original publications [12, 24, 52]. Python serves well for handling large-scale data and extensive calculations, because it provides well-optimized built-in functions and modules tailored to deal with data in tabular form (*pandas* module). BestKeeper proposes a set of housekeeping genes and then introduces the correlation coefficient as a reference. It ranks genes basing on the standard deviation of their expression across samples. Genes with lower standard deviation are considered more stable. It calculates the normalization score for each possible set of housekeeping genes. The NormFinder model-based algorithm is proposed in two variants – for datasets with multiple groups of samples and for datasets with only one group. For a generic algorithm, the inter- and intragroup variations are estimated for each candidate gene. The combined variations are the basis for stability value. The authors state that the proper measure of stability is the distribution in the

developed model, but for practical reasons it is reduced to one-dimensional value equal to the sum of mean and standard deviation of the aforementioned distribution. GeNorm ranks reference genes according to stability value M , where lower M means higher stability [12]. The basis for calculation of stability measure for a chosen gene is \log_2 -transformed ratio of expressions between the ranked gene and each of the remaining genes in the dataset. The stability measure M is then calculated as an arithmetic mean of standard deviations of pairwise expression ratios. In the original algorithm, the procedure begins with a set of candidate genes, which is a small subset of all available genes. This approach is proposed due to computational efficiency of the solution. For each candidate gene, stability value M is computed, and the gene with highest M value (i.e. worst stability) is excluded from the dataset. Then, the M -value calculations are repeated for the remaining genes, until there are only two best genes left.

Then we assigned a ranking specific for each algorithm, which varied in an interval $[0, 1]$ and was calculated as follows:

$$\text{Dataset ranking} = \frac{\text{place in the sorted array of normalization scores}-1}{\text{number of elements in the array}}$$

The final normalization score of a miRNA in a dataset was an arithmetic average of four algorithm-specific rankings (Fig. 7b). The final normalization scores of all miRNAs created a normalization ranking. The lower the ranking, the better the normalization factor.

Implementation validation

We validated the results we received from our implementation of algorithms. For NormFinder, we carried out an analysis of RT-qPCR miRNA expression (GEO ID: GSE47125) with the readily available Microsoft Excel plug-in created by the authors of the original NormFinder article and our implementation of NormFinder. We compared the results using Bland-Altman plot analysis and Pearson's correlation. There was no readily available, original, open-source and updated to work with current versions of operating systems' software implementation of GeNorm and BestKeeper. In case of GeNorm we used the raw data provided by the authors of the original manuscript and ran the analysis for the leukocytes dataset checking for the overlap of the order of the genes. The original BestKeeper article did not include any analyzed data and there is no implementation available in the open-access fashion, so we could not validate our implementation of BestKeeper with the original. In the face of the same problem, Marabita et al. decided to implement their own simplified version of the algorithm, but we decided not to use it [18].

Multiple miRNAs as a single normalization factor

BestKeeper and GeNorm do not allow scoring of multiple miRNAs as a single normalization factor, so we had to devise a way of combining multiple miRNAs. We decided to average expression values of multiple miRNAs (either two or three) by arithmetic mean, thus creating a new entry compatible with the rest of the original dataset. The length of such prepared dataset was larger than the original by one – the new combination of multiple miRNAs. Then we could input the original dataset with the added entry to normalization algorithms, which allowed for evaluation of the combination set in the context of the whole dataset.

Inclusion of all possible combinations of miRNAs into the dataset would distort the image of the measured expression, as the number of all possible combinations would always be vastly greater than that of actual miRNAs quantified. This, in turn, would significantly impact the standard deviation of the expression of miRNAs and hinder the possibility to calculate the normalization scores of NormFinder and GeNorm algorithms. Therefore, we decided to add single combinations sequentially

and by doing so minimize the effect of the addition of every new entry. Notably, expanding the dataset by adding all possible combinations would be without consequence for BestKeeper as it does not use standard deviation or its derivatives, but we wanted to create a uniform pipeline for all three algorithms. Therefore, only one combination at a time was present in a dataset during calculations of normalization scores. We analysed all possible combinations of 2 and 3 miRNAs in the aforementioned way (Fig. 7c). We did not attach more than one combination of miRNAs to a dataset simultaneously.

Combining information from different datasets – finding the normalization set

According to our workflow (Fig. 7c), we created a combined ranking for all normalization factors – single miRNA, combinations of 2 and 3 miRNAs. We algorithmically found a small set of single miRNAs, which could be further used to create combinations of pairs and triple miRNAs selected from the set. We wanted to ensure that the best miRNAs present in the highest-scoring combinations of 2 and 3 miRNAs were included amongst the chosen set of miRNAs and that the set is as small as possible without compromising the stability of combinations created from it.

We focused on the 11 dataset rankings of combinations of 2 miRNAs to find a reliable set of miRNAs and validated their performance on the 11 dataset rankings of combinations of 3 miRNAs. First step was to find the smallest miRNA set that could be used to create combinations of 2 miRNAs, which placed as the best normalization factors in the 11 dataset rankings of combinations of 2 miRNAs. The second step involved validating whether a chosen set could be used to create combinations of 3 miRNAs that also placed first in all the 11 datasets. At last we excluded sets containing miRNAs with a known labile or dynamic expression in serum or plasma.

We performed a pairwise analysis of miRNAs to better illustrate the relationships between them in the chosen set. Therefore, we counted the number of times two miRNAs occurred in all combinations of 3 miRNAs, which placed 1st in the dataset rankings. We divided each singular count by the number of combinations in a dataset containing the counted combination and summed the counts from all occurrences of a pair. This ensured that the counts were not weighted by existence of a one dataset with a multitude of miRNAs. The resulting matrix was used to create a chord diagram illustrating the “normalizing affinity” between pairs of miRNAs from the chosen set.

External validation of results

We have acquired three serum miRNAs qPCR datasets (description of the datasets in the Additional files 2, 3

and 4). We tested our chosen set of reference miRNAs, which we acquired in the previous step. We took each dataset and ran our normalization algorithms for them and then verified how our chosen miRNAs and combinations derived from them scored and compared with the normalization ranking of the dataset. Additionally, we reported the mean ranking of all combinations derived from the chosen 13. We wanted to know whether our miRNA set was truly better than any random set, therefore we sampled random 13 miRNAs 2000 times and estimated the distribution of the mean of rankings of combinations derived from a set of 13 random miRNAs. We reported the position of the mean of our chosen set in this distribution by providing the percentage of random sets surpassed by our set, which we treated as *p*-value of a Monte-Carlo testing procedure.

Statistical analysis

Initial pre-processing of the expression data and tidying was performed in Microsoft Excel 2016. We did the Kruskal-Wallis testing for comparisons of rankings between single and combinations of 2 and 3 miRNAs using statistical software package Statistica (13.1 StatSoft). We devised the Monte-Carlo analysis in the Python programming language using module ‘itertools’.

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s12864-020-6530-3>.

Additional file 1. Supplementary tables. Contains supplementary tables created by the authors.

Additional file 2. Validation dataset 1. Contains results of the qPCR profiling of circulating miRNAs in people with head and neck tumors. Samples are in columns, miRNAs are in verses.

Additional file 3. Validation dataset 2. Contains results of the qPCR profiling of circulating miRNAs in people with head and neck tumors. Samples are in columns, miRNAs are in verses.

Additional file 4. Validation dataset GSE109888. Publicly available on Gene Expression Omnibus platform under the accession number GSE109888. Contains results of the qPCR profiling of circulating miRNAs in people with rheumatoid arthritis [38]. Accession link: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE109888>.

Additional file 5. Supplementary figures. Contains additional figures not included in the main body of the manuscript: 7 additional figures containing validation of the proper implementation of the three used algorithms and figures from the Monte Carlo simulations on the external validation datasets.

Abbreviations

DNA: Deoxyribonucleic acid; GEO: Gene expression omnibus; RNA: Ribonucleic acid; rt-qPCR: real time quantitative polymerase chain reaction; snoRNA: small-nucleolar RNA

Acknowledgements

Not applicable.

Authors' contributions

KP contributed to the design of the work, analysis and interpretation of the data and prepared the first draft of the manuscript and revised it. PK

contributed to the design of the work, analysis and interpretation of the data, created new software used in this work and revised the manuscript. US contributed to the design of the work, created new software used in this work and revised the manuscript. SG created new software used in this work and revised the manuscript. DC contributed to the conception and the design of the work and revised the manuscript. WF contributed to the conception and design of the work, interpretation of the data and revised the manuscript. KP, PK, US, SG, DC, WF approved the submitted version of the manuscript and agreed both to be personally accountable for the author's own contributions and to ensure that questions related to the accuracy or integrity of any part of the work, even ones in which the author was not personally involved, are appropriately investigated, resolved, and the resolution documented in the literature.

Funding

KP, PK, SG and WF were supported by the First TEAM project funded by Smart Growth Operational Programme and the Foundation for Polish Science. We thank NVIDIA Corporation for kindly supplying us with a QUADRO P6000 graphical processing unit we used to run all experiments. The funding bodies played no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Availability of data and materials

The 11 Gene Expression Omnibus datasets used in the process of finding miRNA reference genes are available on the GEO platform (<https://www.ncbi.nlm.nih.gov/geo/>) under appropriate GSE identifiers and accession links (see Additional file 1: Table S1).

Validation datasets used in this study were attached as Additional files to this article. One of them is publicly available in the GEO database (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE109888>).

Ethics approval and consent to participate

Bioethical Committee of Medical University of Lodz approved the tissue gathering and experimental protocols of this study in the decision number RNN/290/17KE proclaimed on 05.09.2017. All participants signed an informed consent form.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Department of Biostatistics and Translational Medicine, Medical University of Lodz, Lodz, Poland. ²Postgraduate School of Molecular Medicine, Medical University of Warsaw, Warsaw, Poland. ³Institute of Applied Computer Science, Lodz University of Technology, Lodz, Poland. ⁴Dana-Farber Cancer Institute, Harvard Medical School, Boston, USA.

Received: 24 July 2019 Accepted: 22 January 2020

Published online: 31 January 2020

References

1. Avery OT, Macleod CM, McCarty M. Studies on the chemical nature of the substance inducing transformation of pneumococcal types: induction of transformation by a deoxyribonucleic acid fraction isolated from pneumococcus type III. *J Exp Med*. 1944;79(2):137–58.
2. Watson JD, Crick FHC. Molecular structure of nucleic acids: a structure for Deoxyribose nucleic acid. *Nature*. 1953;171(4356):737–8. <https://doi.org/10.1038/171737a0>.
3. Lehman IR, Bessman MJ, Simms ES, Kornberg A. Enzymatic synthesis of deoxyribonucleic acid. I. Preparation of substrates and partial purification of an enzyme from *Escherichia coli*. *J Biol Chem*. 1958;233(1):163–70.
4. Higuchi R, Fockler C, Dollinger G, Watson R. Kinetic PCR analysis: real-time monitoring of DNA amplification reactions. *Biotechnology (N Y)*. 1993;11(9):1026–30.
5. Schmittgen TD, Jiang J, Liu Q, Yang L. A high-throughput method to monitor the expression of microRNA precursors. *Nucleic Acids Res*. 2004;32(4):e43.

6. Jiang J, Lee EJ, Gusev Y, Schmittgen TD. Real-time expression profiling of microRNA precursors in human cancer cell lines. *Nucleic Acids Res.* 2005; 33(17):5394–403.
7. Lagos-Quintana M, Rauhut R, Lendeckel W, Tuschl T. Identification of novel genes coding for small expressed RNAs. *Science.* 2001;294(5543):853–8.
8. O'Connell RM, Rao DS, Chaudhuri AA, Baltimore D. Physiological and pathological roles for microRNAs in the immune system. *Nat Rev Immunol.* 2010;10:111. <https://doi.org/10.1038/nri2708>.
9. Ge Q, Zhou Y, Lu J, Bai Y, Xie X, Lu Z. MiRNA in plasma exosome is stable under different storage conditions. *Molecules.* 2014;19(2):1568–75.
10. Mitchell PS, Parkin RK, Kroh EM, Fritz BR, Wyman SK, Pogosova-Agadjanyan EL, et al. Circulating microRNAs as stable blood-based markers for cancer detection. *Proc Natl Acad Sci U S A.* 2008;105(30):10513–8.
11. van Rooij E. The art of microRNA research. *Circ Res.* 2011;108(2):219–34.
12. Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, et al. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.* 2002; 3(7):RESEARCH0034.
13. Rho H-W, Lee B-C, Choi E-S, Choi I-J, Lee Y-S, Goh S-H. Identification of valid reference genes for gene expression studies of human stomach cancer by reverse transcription-qPCR. *BMC Cancer.* 2010;10:240.
14. Wu Z-J, Tian C, Jiang Q, Li X-H, Zhuang J. Selection of suitable reference genes for qRT-PCR normalization during leaf development and hormonal stimuli in tea plant (*Camellia sinensis*). *Sci Rep.* 2016;6:19748.
15. Leal MF, Astur DC, Debieux P, Arliani GG, Silveira Franciozi CE, Loyola LC, et al. Identification of suitable reference genes for investigating gene expression in anterior cruciate ligament injury by using reverse transcription-quantitative PCR. *PLoS One.* 2015;10(7):e0133323.
16. Duan Z-Y, Cai G-Y, Li J-J, Bu R, Wang N, Yin P, et al. U6 can be used as a housekeeping gene for urinary sediment miRNA studies of IgA nephropathy. *Sci Rep.* 2018;8(1):10875.
17. Cicinnati VR, Shen Q, Sotiropoulos GC, Radtke A, Gerken G, Beckebaum S. Validation of putative reference genes for gene expression studies in human hepatocellular carcinoma using real-time quantitative RT-PCR. *BMC Cancer.* 2008;8:350.
18. Marabita F, de Candia P, Torri A, Tegnér J, Abrignani S, Rossi RL. Normalization of circulating microRNA expression data obtained by quantitative real-time RT-PCR. *Brief Bioinform.* 2016;17(2):204–12.
19. Rice J, Roberts H, Rai SN, Galandiuk S. Housekeeping genes for studies of plasma microRNA: a need for more precise standardization. *Surgery.* 2015; 158(5):1345–51.
20. Niu Y, Wu Y, Huang J, Li Q, Kang K, Qu J, et al. Identification of reference genes for circulating microRNA analysis in colorectal cancer. *Sci Rep.* 2016;6:35611.
21. Chen J, Li K, Pang Q, Yang C, Zhang H, Wu F, et al. Identification of suitable reference gene and biomarkers of serum miRNAs for osteoporosis. *Sci Rep.* 2016;6:36347.
22. Song J, Bai Z, Han W, Zhang J, Meng H, Bi J, et al. Identification of suitable reference genes for qPCR analysis of serum microRNA in gastric cancer patients. *Dig Dis Sci.* 2012;57(4):897–904.
23. Li Y, Xiang GM, Liu LL, Liu C, Liu F, Jiang DN, et al. Assessment of endogenous reference gene suitability for serum exosomal microRNA expression analysis in liver carcinoma resection studies. *Mol Med Rep.* 2015; 12(3):4683–91.
24. Drobna M, Szarzynska-Zawadzka B, Daca-Roszak P, Kosmalka M, Jaksik R, Witt M, et al. Identification of Endogenous Control miRNAs for RT-qPCR in T-Cell Acute Lymphoblastic Leukemia. *Int J Mol Sci.* 2018;19(10):2858.
25. Roth C, Rack B, Muller W, Janni W, Pantel K, Schwarzenbach H. Circulating microRNAs as blood-based markers for patients with primary and metastatic breast cancer. *Breast Cancer Res.* 2010;12(6):R90.
26. Serafin A, Foco L, Blankenburg H, Picard A, Zanigni S, Zanon A, et al. Identification of a set of endogenous reference genes for miRNA expression studies in Parkinson's disease blood samples. *BMC Res Notes.* 2014;7:715.
27. Benes V, Castoldi M. Expression profiling of microRNA using real-time quantitative PCR, how to use it and what is available. *Methods.* 2010;50(4): 244–9.
28. Reid G, Kirschner MB, van Zandwijk N. Circulating microRNAs: association with disease and potential use as biomarkers. *Crit Rev Oncol Hematol.* 2011; 80(2):193–208.
29. Benz F, Roderburg C, Vargas Cardenas D, Vuur M, Gautheron J, Koch A, et al. U6 is unsuitable for normalization of serum miRNA levels in patients with sepsis or liver fibrosis. *Exp Mol Med.* 2013;45:e42.
30. Gee HE, Buffa FM, Camps C, Ramachandran A, Leek R, Taylor M, et al. The small-nucleolar RNAs commonly used for microRNA normalisation correlate with tumour pathology and prognosis. *Br J Cancer.* 2011;104(7):1168–77.
31. Ohyashiki K, Umezumi T, Katagiri S, Kobayashi C, Azuma K, Tauchi T, et al. Downregulation of plasma miR-215 in chronic myeloid leukemia patients with successful discontinuation of imatinib. *Int J Mol Sci.* 2016;17(4):570.
32. Aherne ST, Madden SF, Hughes DJ, Pardini B, Naccarati A, Levy M, et al. Circulating miRNAs miR-34a and miR-150 associated with colorectal cancer progression. *BMC Cancer.* 2015;15(1):329.
33. Rijlaarsdam MA, van Agthoven T, Gillis AJM, Patel S, Hayashibara K, Lee KY, et al. Identification of known and novel germ cell cancer-specific (embryonic) miRNAs in serum by high-throughput profiling. *Andrology.* 2015; 3(1):85–91.
34. Shen J, Wang A, Wang Q, Gurvich I, Siegel AB, Remotti H, et al. Exploration of genome-wide circulating microRNA in hepatocellular carcinoma: MiR-483-5p as a potential biomarker. *Cancer Epidemiol Biomark Prev.* 2013; 22(12):2364–73.
35. Selth LA, Townley SL, Bert AG, Stricker PD, Sutherland PD, Horvath LG, et al. Circulating microRNAs predict biochemical recurrence in prostate cancer patients. *Br J Cancer.* 2013;109(3):641–50.
36. Kayano M, Higaki S, Satoh J, Matsumoto K, Matsubara E, Takikawa O, et al. Plasma microRNA biomarker detection for mild cognitive impairment using differential correlation analysis. *Biomark Res.* 2016;4(1):22.
37. Asano M, Umezumi T, Katagiri S, Kobayashi C, Tauchi T, Gotoh M, et al. Up-regulated exosomal miRNA-140-3p in CML patients with musculoskeletal pain associated with discontinuation of tyrosine kinase inhibitors. *Int J Hematol.* 2017;105(4):419–22.
38. Yue J, Lau TCK, Griffith JF, Xu J, Xiao F, Shi L, et al. Circulating miR-99b-5p as a novel predictor of erosion progression on high-resolution peripheral quantitative computed tomography in early rheumatoid arthritis: a prospective cohort study. *Int J Rheum Dis.* 2019;22(9):1724–33.
39. Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* 2010;11(3):R25 Available from: <http://genomebiology.biomedcentral.com/articles/10.1186/gb-2010-11-3-r25>.
40. Grün D, van Oudenaarden A. Design and analysis of single-cell sequencing experiments. *Cell.* 2015;163(4):799–810 Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0092867415013537>.
41. Tay JW, James I, Hughes QW, Tiao JY, Baker RL. Identification of reference miRNAs in plasma useful for the study of oestrogen-responsive miRNAs associated with acquired protein S deficiency in pregnancy. *BMC Res Notes.* 2017;10(1):312.
42. Mallona I, Lischewski S, Weiss J, Hause B, Egea-Cortines M. Validation of reference genes for quantitative real-time PCR during leaf and flower development in *Petunia hybrida*. *BMC Plant Biol.* 2010;10:4.
43. Thellin O, Zorzi W, Lakaye B, De Borman B, Coumans B, Hennen G, et al. Housekeeping genes as internal standards: use and limits. *J Biotechnol.* 1999;75(2–3):291–5.
44. Derveaux S, Vandesompele J, Hellemans J. How to do successful gene expression analysis using real-time PCR. *Methods.* 2010;50(4):227–30.
45. Ulivi P, Petracci E, Marisi G, Baglivo S, Chiari R, Billi M, et al. Prognostic Role of Circulating miRNAs in Early-Stage Non-Small Cell Lung Cancer. *J Clin Med [Internet].* 2019;8(2):131. Available from: <http://www.mdpi.com/2077-0383/8/2/131>
46. Alhasan I. MiR-126 modulates angiogenesis in breast Cancer by targeting VEGF-A-mRNA. *Asian Pac J Cancer Prev.* 2019;20(1):193–7.
47. Liu R, Zhang Y-S, Zhang S, Cheng Z-M, Yu J-L, Zhou S, et al. MiR-126-3p suppresses the growth, migration and invasion of NSCLC via targeting CCR1. *Eur Rev Med Pharmacol Sci.* 2019;23(2):679–89.
48. Zuberi M, Khan I, Gandhi G, Ray PC, Saxena A. The conglomeration of diagnostic, prognostic and therapeutic potential of serum miR-199a and its association with clinicopathological features in epithelial ovarian cancer. *Tumour Biol.* 2016;37(8):11259–66.
49. Babion I, Snoek BC, van de Wiel MA, Wilting SM, Steenbergen RDM. A strategy to find suitable reference genes for miRNA quantitative PCR analysis and its application to cervical specimens. *J Mol Diagn.* 2017;19(5):625–37.
50. Patnaik SK, Kannisto E, Knudsen S, Yendamuri S. Evaluation of microRNA expression profiles that may predict recurrence of localized stage I non-small cell lung cancer after surgical resection. *Cancer Res.* 2010;70(11):36–45.
51. Leuvenberger N, Schumacher YQ, Pradervand S, Sander T, Saugy M, Pottgiesser T. Circulating microRNAs as biomarkers for detection of autologous blood transfusion. *PLoS One.* 2013;8(6):e66309.

52. Pfaffl MW, Tichopad A, Prgomet C, Neuvians TP. Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper–Excel-based tool using pair-wise correlations. *Biotechnol Lett*. 2004; 26(6):509–15 Available from: <http://www.ncbi.nlm.nih.gov/pubmed/15127793>.
53. Han H-S, Jo YN, Lee JY, Choi S-Y, Jeong Y, Yun J, et al. Identification of suitable reference genes for the relative quantification of microRNAs in pleural effusion. *Oncol Lett*. 2014;8(4):1889–95.
54. Fujimoto S, Manabe S, Morimoto C, Ozeki M, Hamano Y, Tamaki K. Optimal small-molecular reference RNA for RT-qPCR-based body fluid identification. *Forensic Sci Int Genet*. 2018;37:135–42.
55. Eriksen AHM, Andersen RF, Pallisgaard N, Sorensen FB, Jakobsen A, Hansen TF. MicroRNA expression profiling to identify and validate reference genes for the relative quantification of microRNA in rectal Cancer. *PLoS One*. 2016; 11(3):e0150593.
56. Xu Q, Zhao Y, Zhou X, Luan J, Cui Y, Han J. Comparison of the extraction and determination of serum exosome and miRNA in serum and the detection of miR-27a-3p in serum exosome of ALS patients. *Intractable Rare Dis Res*. 2018;7(1):13–8.
57. Bobbili MR, Mader RM, Grillari J, Dellago H. OncomiR-17-5p: alarm signal in cancer? *Oncotarget*. 2017;8(41):71206–22.
58. Wang X, Zhang X, Yuan J, Wu J, Deng X, Peng J, et al. Evaluation of the performance of serum miRNAs as normalizers in microRNA studies focused on cardiovascular disease. *J Thorac Dis*. 2018;10(5):2599–607.
59. Beuzelin D, Kaeffer B. Exosomes and miRNA-loaded biomimetic Nanovehicles, a focus on their potentials preventing Type-2 diabetes linked to metabolic syndrome. *Front Immunol*. 2018;9:2711.
60. Pan C, Stevic I, Muller V, Ni Q, Oliveira-Ferrer L, Pantel K, et al. Exosomal microRNAs as tumor markers in epithelial ovarian cancer. *Mol Oncol*. 2018; 12(11):1935–48.
61. Wang Q, Ma J, Jiang Z, Wu F, Ping J, Ming L. Diagnostic value of circulating microRNA-27a/b in patients with acute pulmonary embolism. *Int Angiol*. 2018;37(1):19–25.
62. Wu J, Du K, Lu X. Elevated expressions of serum miR-15a, miR-16, and miR-17-5p are associated with acute ischemic stroke. *Int J Clin Exp Med* [Internet]. 2015; 8(11):21071–9. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/26885038>
63. Jia S -z, Yang Y, Lang J, Sun P, Leng J. Plasma miR-17-5p, miR-20a and miR-22 are down-regulated in women with endometriosis. *Hum Reprod* [Internet]. 2013;28(2):322–30. Available from: <https://academic.oup.com/humrep/article-lookup/doi/10.1093/humrep/des413>
64. Xue S, Liu D, Zhu W, Su Z, Zhang L, Zhou C, et al. Circulating MiR-17-5p, MiR-126-5p and MiR-145-3p Are Novel Biomarkers for Diagnosis of Acute Myocardial Infarction. *Front Physiol* [Internet]. 2019;10. Available from: <https://www.frontiersin.org/article/10.3389/fphys.2019.00123/full>
65. Sochor M, Basova P, Pesta M, Dusilkova N, Bartos J, Burda P, et al. Oncogenic MicroRNAs: miR-155, miR-19a, miR-181b, and miR-24 enable monitoring of early breast cancer in serum. *BMC Cancer* [Internet]. 2014;14(1):448. Available from: <http://bmccancer.biomedcentral.com/articles/10.1186/1471-2407-14-448>
66. Deng X, Liu Y, Luo M, Wu J, Ma R, Wan Q, et al. Circulating miRNA-24 and its target YKL-40 as potential biomarkers in patients with coronary heart disease and type 2 diabetes mellitus. *Oncotarget* [Internet]. 2017;8(38). Available from: <http://www.oncotarget.com/fulltext/18593>
67. Sommariva E, D'Alessandra Y, Farina FM, Casella M, Cattaneo F, Catto V, et al. MiR-320a as a Potential Novel Circulating Biomarker of Arrhythmogenic Cardiomyopathy. *Sci Rep* [Internet]. 2017;7(1):4802. Available from: <http://www.nature.com/articles/s41598-017-05001-z>
68. Chiam K, Wang T, Watson DI, Mayne GC, Irvine TS, Bright T, et al. Circulating Serum Exosomal miRNAs As Potential Biomarkers for Esophageal Adenocarcinoma. *J Gastrointest Surg* [Internet]. 2015;19(7):1208–15. Available from: <http://link.springer.com/10.1007/s11605-015-2829-9>
69. Fujiwara T, Uotani K, Yoshida A, Morita T, Nezu Y, Kobayashi E, et al. Clinical significance of circulating miR-25-3p as a novel diagnostic and prognostic biomarker in osteosarcoma. *Oncotarget* [Internet]. 2017;8(20). Available from: <http://www.oncotarget.com/fulltext/16498>
70. Hesari A, Azizian M, Darabi H, Nesaei A, Hosseini SA, Salarinia R, et al. Expression of circulating miR-17, miR-25, and miR-133 in breast cancer patients. *J Cell Biochem* [Internet]. 2019;120(5):7109–14. Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1002/jcb.27984>
71. Li M, Song Q, Li H, Lou Y, Wang L. Circulating miR-25-3p and miR-451a May Be Potential Biomarkers for the Diagnosis of Papillary Thyroid Carcinoma. Ray RB, editor. *PLoS One* [Internet]. 2015;10(7):e0132403. Available from: <http://dx.plos.org/10.1371/journal.pone.0132403>
72. Deng T, Yuan Y, Zhang C, Zhang C, Yao W, Wang C, et al. Identification of Circulating MiR-25 as a Potential Biomarker for Pancreatic Cancer Diagnosis. *Cell Physiol Biochem* [Internet]. 2016;39(5):1716–22. Available from: <https://www.karger.com/Article/FullText/447872>
73. ZiaSarabi P, Sorayayi S, Hesari A, Ghasemi F. Circulating microRNA-133, microRNA-17 and microRNA-25 in serum and its potential diagnostic value in gastric cancer. *J Cell Biochem* [Internet]. 2019;120(8):12376–81. Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1002/jcb.28503>
74. Mirzaei H, Sahebkar A, Mohammadi M, Yari R, Salehi H, Jafari M, et al. Circulating microRNAs in Hepatocellular Carcinoma: Potential Diagnostic and Prognostic Biomarkers. *Curr Pharm Des* [Internet]. 2016;22(34):5257–69. Available from: <http://www.eurekaselect.com/openurl/content.php?genre=article&issn=1381-6128&volume=22&issue=34&page=5257>
75. Grimolizzi F, Monaco F, Leoni F, Bracci M, Staffolani S, Bersaglieri C, et al. Exosomal miR-126 as a circulating biomarker in non-small-cell lung cancer regulating cancer progression. *Sci Rep* [Internet]. 2017;7(1):15277. Available from: <http://www.nature.com/articles/s41598-017-15475-6>
76. Khanaghaei M, Tourkianvalashani F, Hekmatimoghaddam S, Ghasemi N, Rahaie M, Khorramshahi V, et al. Circulating miR-126 and miR-499 Reflect Progression of Cardiovascular Disease; Correlations with Uric Acid and Ejection Fraction. *Heart Int* [Internet]. 2016;11(1):heartint.500022. Available from: <http://journals.sagepub.com/doi/10.5301/heartint.5000226>
77. Long G, Wang F, Li H, Yin Z, Sandip C, Lou Y, et al. Circulating miR-30a, miR-126 and let-7b as biomarker for ischemic stroke in humans. *BMC Neurol* [Internet]. 2013;13(1):178. Available from: <http://bmcneurol.biomedcentral.com/articles/10.1186/1471-2377-13-178>
78. Zhang T, Li L, Shang Q, Lv C, Wang C, Su B. Circulating miR-126 is a potential biomarker to predict the onset of type 2 diabetes mellitus in susceptible individuals. *Biochem Biophys Res Commun* [Internet]. 2015;463(1–2):60–3. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0006291X15009237>
79. Kong R, Gao J, Si Y, Zhao D. Combination of circulating miR-19b-3p, miR-122-5p and miR-486-5p expressions correlates with risk and disease severity of knee osteoarthritis. *Am J Transl Res* [Internet]. 2017;9(6):2852–64. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/28670374>
80. Copier CU, León L, Fernández M, Contador D, Calligaris SD. Circulating miR-19b and miR-181b are potential biomarkers for diabetic cardiomyopathy. *Sci Rep* [Internet]. 2017;7(1):13514. Available from: <http://www.nature.com/articles/s41598-017-13875-2>
81. Wang K-J, Zhao X, Liu Y-Z, Zeng Q-T, Mao X-B, Li S-N, et al. Circulating MiR-19b-3p, MiR-134-5p and MiR-186-5p are Promising Novel Biomarkers for Early Diagnosis of Acute Myocardial Infarction. *Cell Physiol Biochem* [Internet]. 2016; 38(3):1015–29. Available from: <https://www.karger.com/Article/FullText/443053>
82. Zhang J, Song Y, Zhang C, Zhi X, Fu H, Ma Y, et al. Circulating MiR-16-5p and MiR-19b-3p as Two Novel Potential Biomarkers to Indicate Progression of Gastric Cancer. *Theranostics* [Internet]. 2015;5(7):733–45. Available from: <http://www.thno.org/v05p0733.htm>
83. Osip'yants AI, Knyazev EN, Galatenko A V., Nyushko KM, Galatenko V V., Shkurnikov MY, et al. Changes in the Level of Circulating hsa-miR-297 and hsa-miR-19b-3p miRNA Are Associated with Generalization of Prostate Cancer. *Bull Exp Biol Med* [Internet]. 2017;162(3):379–82. Available from: <http://link.springer.com/10.1007/s10517-017-3620-6>
84. Zaporozhchenko IA, Morozkin ES, Skvortsova TE, Ponomaryova AA, Rykova EY, Cherdynitseva N V., et al. Plasma miR-19b and miR-183 as Potential Biomarkers of Lung Cancer. Dahiya R, editor. *PLoS One* [Internet]. 2016;11(10):e0165261. Available from: <https://dx.plos.org/10.1371/journal.pone.0165261>
85. Chai C, Song L-J, Yang B, Han S-Y, Li X-Q, Li M. Circulating miR-199a-3p in plasma and its potential diagnostic and prognostic value in glioma. *Eur Rev Med Pharmacol Sci* [Internet]. 2016;20(23):4885–90. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/27981547>
86. Yin J, Hou P, Wu Z, Wang T, Nie Y. Circulating miR-375 and miR-199a-3p as potential biomarkers for the diagnosis of hepatocellular carcinoma. *Tumor Biol* [Internet]. 2015;36(6):4501–7. Available from: <http://link.springer.com/10.1007/s13277-015-3092-0>
87. Xue S, Zhu W, Liu D, Su Z, Zhang L, Chang Q, et al. Circulating miR-26a-1, miR-146a and miR-199a-1 are potential candidate biomarkers for acute myocardial infarction. *Mol Med* [Internet]. 2019;25(1):18. Available from: <https://molmed.biomedcentral.com/articles/10.1186/s10020-019-0086-1>
88. Nonaka R, nishimura J, Kagawa Y, Osawa H, Hasegawa J, Murata K, et al. Circulating miR-199a-3p as a novel serum biomarker for colorectal cancer. *Oncol Rep* [Internet]. 2014;32(6):2354–8. Available from: <https://www.spandidos-publications.com/10.3892/or.2014.3515>

89. Zhou G, Lu M, Chen J, Li C, Zhang J, Chen J, et al. Identification of miR-199a-5p in serum as noninvasive biomarkers for detecting and monitoring osteosarcoma. *Tumor Biol* [Internet]. 2015;36(11):8845–52. Available from: <http://link.springer.com/10.1007/s13277-015-3421-3>
90. Xin H, Yang Y, Liu J, Li X, Li M, Feng B, et al. Association between tuberculosis and circulating microRNA hsa-let-7b and hsa-miR-30b: A pilot study in a Chinese population. *Tuberculosis* [Internet]. 2016;99:63–9. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S1472979216300257>
91. Fayyad-Kazan M, ElDirani R, Hamade E, El Majzoub R, Akl H, Bitar N, et al. Circulating miR-29c, miR-30c, miR-193a-5p and miR-885-5p: Novel potential biomarkers for HTLV-1 infection diagnosis. *Infect Genet Evol* [Internet]. 2019; 74:103938. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S1567134819301595>
92. Llano-Diez M, Ortez CI, Gay JA, Álvarez-Cabado L, Jou C, Medina J, et al. Digital PCR quantification of miR-30c and miR-181a as serum biomarkers for Duchenne muscular dystrophy. *Neuromuscul Disord* [Internet]. 2017;27(1):15–23. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0960896616302905>
93. Spinelli S V, Fernández R del V., Zoff L, Bongiovanni B, Díaz A, D'Attilio L, et al. miR-30c is specifically repressed in patients with active pulmonary tuberculosis. *Tuberculosis* [Internet]. 2017;105:73–9. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S1472979217300173>
94. Crossland RE, Norden J, Juric MK, Green K, Pearce KF, Lendrem C, et al. Expression of Serum microRNAs is Altered During Acute Graft-versus-Host Disease. *Front Immunol* [Internet]. 2017;8. Available from: <http://journal.frontiersin.org/article/10.3389/fimmu.2017.00308/full>

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://www.biomedcentral.com/submissions)

