TUM School of Life Sciences Weihenstephan
Division of Animal Physiology and Immunology

Technical University of Munich

Universal reference transcripts for miRNA normalization: a meta-analysis of human blood extracellular vesicle RNA-seq data sets

Alex Hildebrandt[1], Benedikt Kirchner[1], Chenna R. Galiveti[2], Esther N. M. Nolte-'t Hoen[2], Michael W. Pfaffl[1]

[1] Department of Animal Physiology and Immunology, Technical University of Munich, Freising, Germany.

[2] Department of Biochemistry and Cell Biology, Utrecht University, Utrecht, The Netherlands.

contact:
alex.hildebrandt@tum.de

DATA PROCESSING PIPELINE

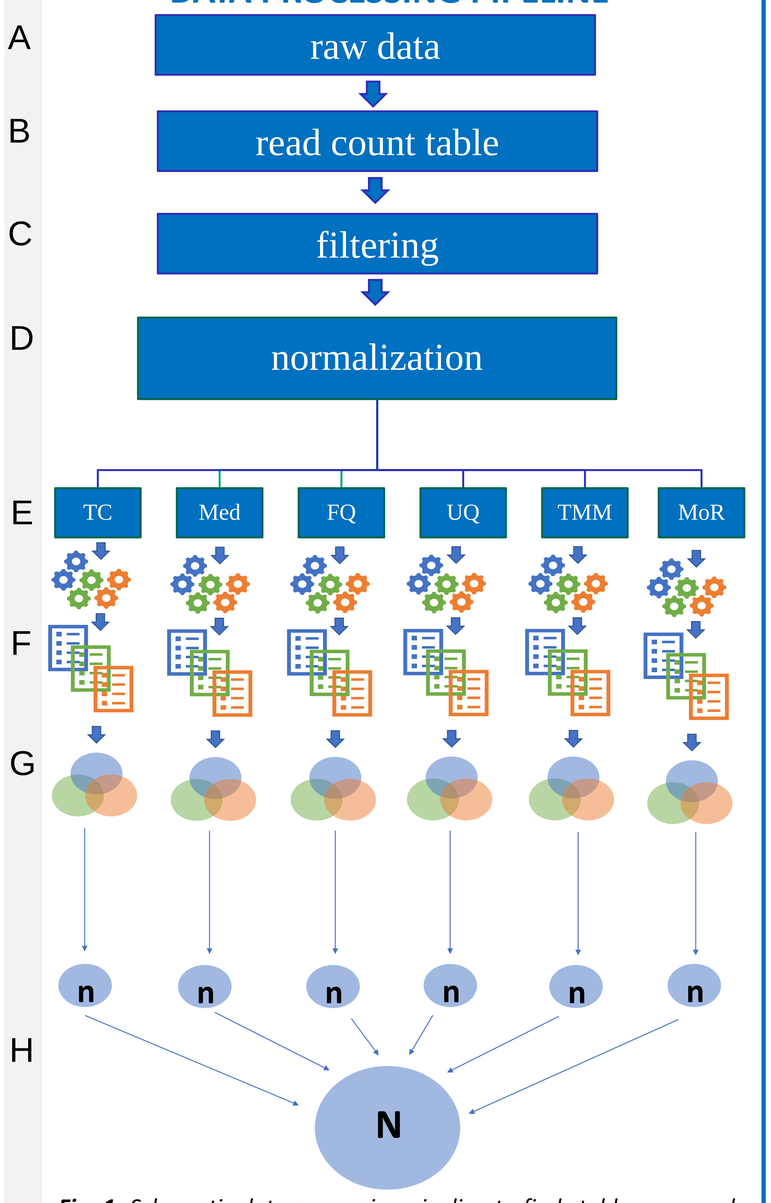


Fig. 1: *Schematic data processing pipeline to find stably expressed miRNAs. n/N is the number of overlapping miRNAs.*

INTRODUCTION

Due to their critical functions in intercellular communication, extracellular vesicles (EV) have emerged as important sources of biomarkers for pro- and diagnostic purposes. With the advent of RNA-seq as the tool of choice for unbiased biomarker screening, a major focus has been laid on miRNAs, fundamental regulators of post-transcriptional gene expression. Feasibility of RNA biomarkers in clinical practice presently still relies on validation and analysis by RT-qPCR, which in turn is depending on stably expressed reference transcripts for normalization. To assess whether a set of universal reference miRNA transcripts for normalization exists, a meta-analysis of blood derived EV samples was conducted.

METHODS

Data collection (Fig. 1 A-B)

Relevant and well annotated data sets of circulating extracellular vesicles for this meta-analysis were first downloaded from the gene expression omnibus database GEO. Second, in house data as well as data from collaborators were added. In total, 513 samples from 10 studies with 10 different physiological and pathophysiological conditions were gathered. Read count matrices were generated for each sample using a small RNA alignment and annotation pipeline [1], and available methodological and biological information was summarized for further analyses.

Filtering (Fig. 1 C)

A first filtering step was applied to the samples. Those with less than 7 % of reads mapping to miRNAs were removed, leaving 356 out of 513 samples. A second filtering step was applied to the miRNAs. miRNAs that did not appear in at least 95 % of the samples were removed as well, leaving 318 miRNAs of 2498 (found in all data sets).

Normalization (Fig.1 D)

To ensure expression estimates are comparable across miRNAs, conditions and samples, six normalization methods were applied and contrasted afterwards: Total count normalization (TC), median normalization (Med), full quantile normalization (FQ), upper quartile normalization (UQ), trimmed mean of M-value (TMM) and median of ratios normalization (MoR).

Search for candidate miRNAs (Fig. 1 E-F)

Normalized read counts were subsequently screened for candidate miRNAs using three different algorithms that each compute a specific stability metric/indicator. In addition to *bestKeeper*, which calculates the coefficient of variation (CV) across samples, we utilized *geNorm* and *normFinder*, which indicate expression stability via the M-value and stability measure rho, respectively. Subsequently, all miRNAs with a relative distance of 50 % to the best ranked miRNAs and its stability value were extracted (Fig. 2).

Overlap analysis (Fig. 1 G-H)

Candidate miRNAs of the three stability algorithms were overlapped for each normalization method to find shared miRNAs. Similarly, common miRNAs across the six normalization methods were calculated as well. Finally, miRNAs were ranked according to their stability as defined by the frequency with which they occurred in the stability algorithms and normalization methods .

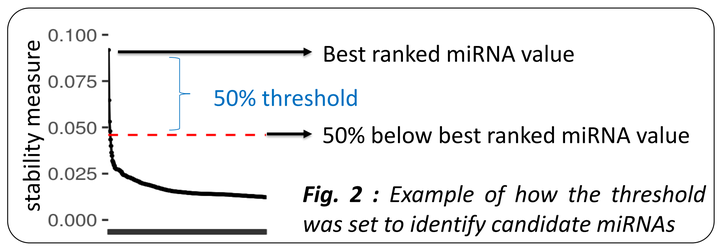


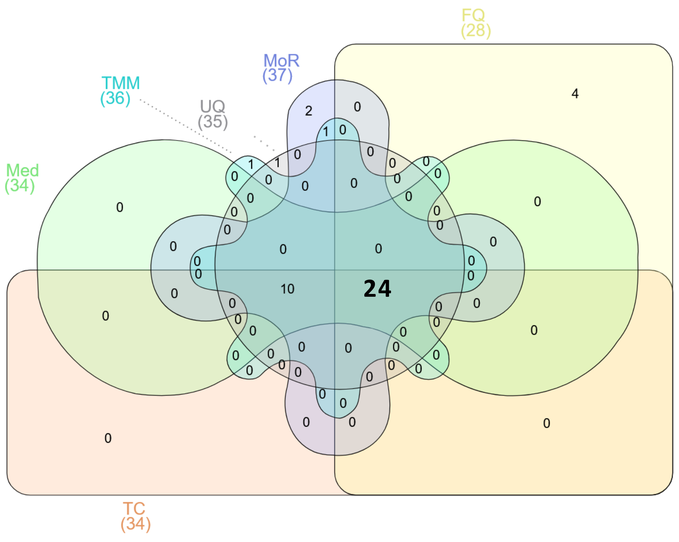
Fig. 2 : *Example of how the threshold was set to identify candidate miRNAs*

RESULTS

- Overlap analysis displayed high consensus with **24** miRNAs being shared between all normalization methods and a further 10 miRNAs common between all normalization methods except FQ, irrespective of the stability measure algorithm. Performance differences in performances for FQ were also corroborated by the highest number of miRNAs uniquely detected by only one normalization strategy although the number of normalization-specific miRNAs tended to be low in general (Fig. 3).
- Evaluation of stability measurement algorithms showed much higher variability than normalization methods with numbers of candidate miRNAs ranging from 11 (*normFinder*) to 37 (*geNorm*). Although **three** miRNAs were commonly detected by all three algorithms irrespective of the normalization strategy, no further overlap was found between *normFinder* and *bestKeeper*, and *geNorm* determined 16 miRNAs not included by the others (Fig. 4).
- Only candidate miRNAs detected by at least two stability measure algorithms were included in following analysis to counter their high variability and frequencies for normalization methods were subsequently recalculated (Table 1, left side). Comparison with the largest subset in the meta-analysis (Yuan *et al.* [2], 192 samples), processed in an identical manner, showed a very limited overlap with only one miRNA (**miR-30d-5p**) but three miRNAs not included in any candidate miRNA list of the combined data sets.

CONCLUSION

- First results suggest that circulating EVs may contain a common set of miRNAs that can be used as reference in RT-qPCR. With the availability of additional small RNA-seq data sets in the future, the robustness and validity of miRNA reference predictions by our pipeline will continue to grow.
- Based on the current data, we still recommend analyzing reference transcripts in each study individually, as current predictions are seriously hindered by the great variety of prevalent EV isolation protocols and sequencing strategies. Further standardization efforts as proposed by the MISEV guidelines [3] as well as identification of contaminants or subpopulations in EV isolates are important and necessary steps on the way to a universal set of reference miRNAs.



miRNA	frequency	miRNA	frequency
let-7i-5p	6/6	miR-652-3p	6/6
miR-101-3p	6/6	miR-9985	6/6
miR-103a-3p	6/6	let-7f-1-3p	5/6
miR-10527-5p	6/6	let-7f-2-3p	5/6
miR-10a-3p	6/6	miR-145-3p	5/6
miR-140-3p	6/6	miR-210-3p	5/6
miR-140-5p	6/6	miR-29c-5p	5/6
miR-148b-3p	6/6	miR-424-5p	5/6
miR-185-3p	6/6	miR-499a-5p	5/6
miR-19b-3p	6/6	miR-505-3p	5/6
miR-21-5p	6/6	miR-532-5p	5/6
miR-222-3p	6/6	miR-542-3p	5/6
miR-23a-3p	6/6	let-7a-3p	2/6
miR-24-3p	6/6	let-7b-3p	1/6
miR-29b-3p	6/6	miR-152-3p	1/6
miR-30d-5p	6/6	miR-22-5p	1/6
miR-331-5p	6/6	miR-221-3p	1/6
miR-340-5p	6/6	miR-24-1-5p	1/6
miR-361-3p	6/6	miR-26a-5p	1/6
miR-378a-5p	6/6	miR-30c-5p	1/6
miR-425-3p	6/6	miR-32-5p	1/6
miR-501-3p	6/6		

Fig. 3: *Overlap of candidate miRNAs by normalization method irrespective of stability measure algorithms. Candidate miRNAs are ranked by the frequency of occurrence in any of the six normalization methods.*

miRNA	frequency	miRNA	frequency	miRNA	frequency
miR-140-5p	3/3	miR-361-3p	2/3	miR-21-5p	1/3
miR-23a-3p	3/3	miR-378a-5p	2/3	miR-210-3p	1/3
miR-30d-5p	3/3	miR-425-3p	2/3	miR-22-5p	1/3
let-7f-1-3p	2/3	miR-501-3p	2/3	miR-221-3p	1/3
let-7i-5p	2/3	miR-505-3p	2/3	miR-24-1-5p	1/3
miR-103a-3p	2/3	miR-9985	2/3	miR-26a-5p	1/3
miR-10527-5p	2/3	let-7a-3p	1/3	miR-29c-5p	1/3
miR-145-3p	2/3	let-7b-3p	1/3	miR-30c-5p	1/3
miR-148b-3p	2/3	let-7f-2-3p	1/3	miR-32-5p	1/3
miR-185-3p	2/3	miR-101-3p	1/3	miR-424-5p	1/3
miR-222-3p	2/3	miR-10a-3p	1/3	miR-499a-5p	1/3
miR-24-3p	2/3	miR-140-3p	1/3	miR-532-5p	1/3
miR-29b-3p	2/3	miR-152-3p	1/3	miR-542-3p	1/3
miR-331-5p	2/3	miR-19b-3p	1/3	miR-652-3p	1/3
miR-340-5p	2/3				

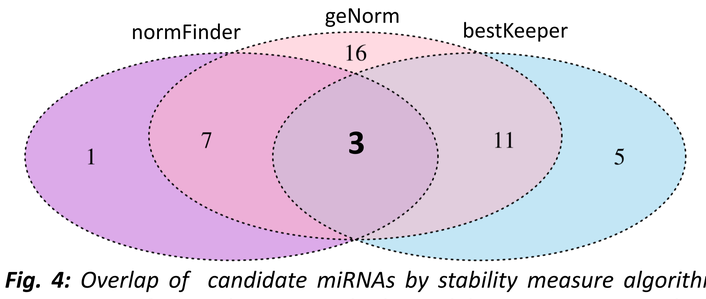


Fig. 4: *Overlap of candidate miRNAs by stability measure algorithm irrespective of normalization method. Candidate miRNAs are ordered by the frequency of occurrence in any of the three stability measurement algorithms.*

All data sets combined		Yuan *et al.* data set	
miRNA	frequency	miRNA	frequency
miR-148b-3p	6/6	miR-30d-5p	5/6
miR-425-3p	6/6	miR-146a-5p	3/6
miR-9985	6/6	miR-99a-5p	1/6
miR-140-5p	5/6	miR-99b-3p	1/6
miR-24-3p	5/6		
miR-29b-3p	5/6		
miR-30d-5p	5/6		
miR-10527-5p	4/6		
miR-23a-3p	4/6		
let-7f-1-3p	3/6		
miR-145-3p	3/6		
miR-340-5p	3/6		
miR-222-3p	2/6		
let-7i-5p	1/6		
miR-103a-3p	1/6		
miR-185-3p	1/6		
miR-361-3p	1/6		
miR-501-3p	1/6		
miR-505-3p	1/6		

Table 1: *Processing of all data sets combined (left) and the Yuan et al. data set alone (right). Frequency defines in how many normalization methods miRNAs were present during the overlap analysis.*

[1] D. Buschmann *et al.*, "Toward reliable biomarker signatures in the age of liquid biopsies - How to standardize the small RNA-Seq workflow," Nucleic Acids Res., vol. 44, no. 13, pp. 5995–6018, 2016.
[2] T. Yuan *et al.*, "Plasma extracellular RNA profiles in healthy and cancer patients," Sci. Rep., vol. 6, no. 1, pp. 1–11, 2016.
[3] C. Théry *et al.*, "Minimal information for studies of extracellular vesicles 2018 (MISEV2018): a position statement of the International Society for Extracellular Vesicles and update of the MISEV2014 guidelines," J. Extracell. Vesicles, 2018.

WE WANT YOUR DATA

Do you want to help the EV community find widely applicable reference miRNAs?

Get in contact with

alex.hildebrandt@tum.de

to support this ongoing effort by contributing your own miRNA-Seq data sets!

WE WANT YOUR DATA