IJC
International Journal of Cancer

# A normalization strategy for the analysis of plasma microRNA qPCR data in colorectal cancer

Sara Pizzamiglio[1], Stefano Bottelli[1], Chiara Maura Ciniselli[1], Susanna Zanutto[2,3], Claudia Bertan[4], Manuela Gariboldi[2,3], Marco Alessandro Pierotti[5] and Paolo Verderio[1]

[1] Unit of Medical Statistics, Biometry and Bioinformatics, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy
[2] Department of Experimental Oncology and Molecular Medicine, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy
[3] Molecular Genetics of Cancer, Fondazione Istituto FIRC di Oncologia Molecolare, Milan, Italy
[4] Department of Pathology, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy
[5] Scientific Directorate, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy

Dear Editor,

We read with great interest the article by Wang *et al.*,[1] recently published on the International Journal of Cancer, investigating the capability of plasma microRNA (miRNA) expression of predicting presence of colorectal cancer (CRC). The authors at first performed a discovery phase based on the generation of high-throughput data using miRNA microarrays, and then they evaluated the candidate miRNAs in the plasma samples of training and validation sets by relative quantification based on quantitative real-time PCR (qPCR). At the end they identified three miRNAs that yielded high diagnostic accuracy in discriminating CRC patients from individuals belonging to the healthy group. In both microarray and qPCR analyses miR-1228 was used as endogenous control for data normalization. However, as the authors pointed out, the identification of appropriate endogenous control miRNA(s) for data normalization is a critical step. Dealing with qPCR data, a frequent approach to cover this step is the use of invariant endogenous control(s) or reference miRNA(s) commonly identified from a pilot study with representative samples from the experimental conditions. Several articles[2–4] showed that data-driven methods for normalization of high-throughput qPCR data such as mean expression value perform best in reducing variations between arrays with respect to a single endogenous control. However, data-driven methods are preferentially used during the discovery phase and almost never applicable in the subsequent validation studies that are mainly focused on a limited number of miRNAs. To overcome this issue, we developed a comprehensive procedure that, starting from a data-driven normalization method based on high-throughput qPCR data identify a small set of miRNAs to be used as reference for data normalization in view of subsequent validation studies.

The whole procedure is represented in Figure 1. Briefly, by taking advantage of the comparative cycle threshold (Ct) method,[5] the relative expression for each considered $i$-th ($i = 1, 2, \ldots, N$) miRNA was computed as follow: $\Delta\text{Ct}\_N_i = \text{Ct}_i - \text{gm}\_N$, where gm_$N$ is the geometric mean of the N microRNAs detected in all the samples. Then a small subset of $K$ ($K < N$) miRNAs was identified according to specific selection criteria, such as low variability, consistent expression between comparison groups and low correlation with the other potentially considered reference miRNAs. Once ranked according to their stability,[6,7] these K miRNAs were forwardly combined in $S$ ($S = K$) sets, so that the first set contains the most stable miRNA, the second set contains the two most stable miRNAs and so on until the last set which contains all the K miRNAs. Within each $j$-th set ($j = 1, 2, \ldots, S$) the geometric mean of the miRNA(s) belonging to the set was computed (gm_$S_j$) and then the relative expression of each $i$-th miRNA was obtained as $\Delta\text{Ct}\_S_{ji} = \text{Ct}_{ji} - \text{gm}\_S_j$. By starting from these values the relative expression ($\log_2\text{RQ}_{ji} = -\Delta\text{Ct}\_S_{ji}$) of each of the N miRNAs was evaluated between comparison groups using a nonparametric test.[8] The set of miRNAs showing results with the highest agreement[9] with that obtained when considering the relative expression computed by using the geometric mean of the N miRNAs was identified (best subset).

We applied this procedure in a similar context of Wang *et al.*[1] by considering 80 plasma samples from individuals at high risk of developing CRC who presented (49 cases) or not (31 controls) precancerous lesions/cancer, as detected by colonscopy. The miRNA profile of each sample was analysed using the Taqman® Array Human microRNA Card A (Applied Biosystem, Foster City, CA) containing 381 mature miRNAs. We obtained $N = 69$ miRNAs detected in all samples ($\text{Ct} < 40$) and, among them, $K = 11$ miRNAs were selected and ranked according to their stability. The best subset consisted in 7 miRNAs and, when used for data normalization, allowed us to find a panel of miRNAs potentially able to identify individuals carrying precancerous lesions/cancer to be validated in subsequent studies. Notably, Kang *et al.*[10] recently published an algorithm for normalization of miRNAs in plasma. In addition, in this case the final aim is to obtain a small subset of miRNAs to be used as reference in the subsequent validation step. Interestingly, by applying this algorithm on our data, the panel of potentially relevant miRNAs resulted overlapping with those obtained with our strategy thus confirming the robustness of the latter.

In conclusion, although we agree with Wang *et al.*[1] on the need to validate in more studies the miR-1228 they used as endogenous control, we believe that a more successful normalization strategy could be to exploit the high-throughput data generated during the discovery phase. One possible method could be the procedure we have proposed for selecting a subset of reference miRNAs during the discovery phase based on qPCR high-throughput data to be transferred in the relative quantification of promising miRNAs in subsequent studies.

Letter to the Editor

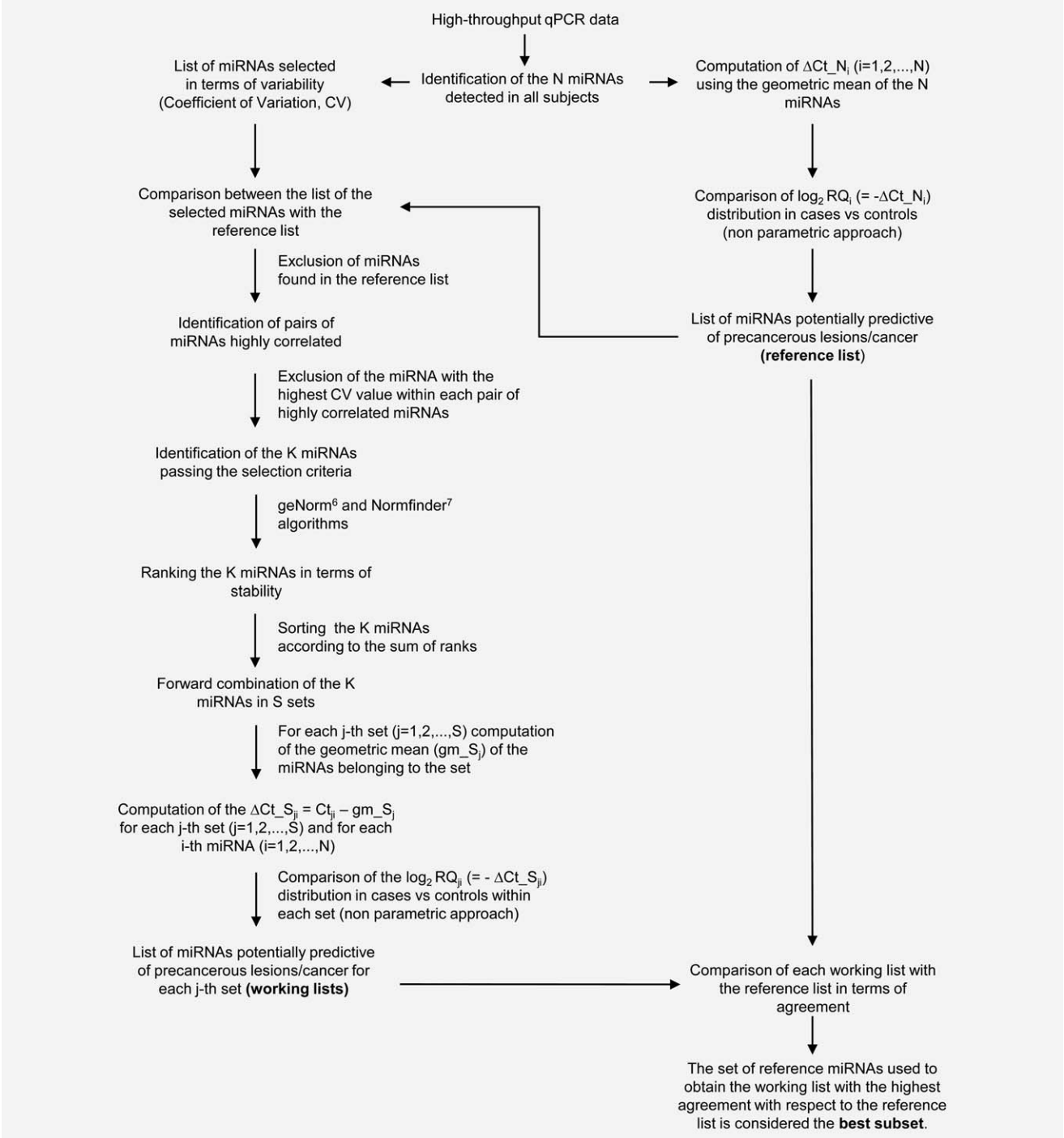


**Figure 1.** Procedure for the identification of the best subset of reference miRNAs.

Yours sincerely,
Sara Pizzamiglio,
Stefano Bottelli,
Chiara M. Ciniselli,
Susanna Zanutto,
Claudia Bertan,
Manuela Gariboldi,
Marco A. Pierotti and
Paolo Verderio

## References


1. Wang S, Xiang J, Li Z, et al. A plasma microRNA panel for early detection of colorectal cancer. *Int J Cancer*, in press. DOI: 10.1002/ijc.28136.
2. Deo A, Carlsson J, Lindlöf A. How to choose a normalization strategy for miRNA quantitative real-time (qPCR) arrays. *J Bioinform Comput Biol* 2011;9: 795–812.
3. Mestdagh P, Van Vlierberghe P, De Weer A, et al. A novel and universal method for microRNA RT-qPCR data normalization. *Genome Biol* 2009;10:R64.
4. Mar JC, Kimura Y, Schroder K, et al. Data-driven normalization strategies for high-throughput quantitativeRT-PCR. *BMC Bioinformatics* 2009;10:110.

5. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta CT) Method. *Methods* 2001;25:402–8.
6. Vandesompele J, De Preter K, Pattyn F, et al. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol* 2002;3:RESEARCH0034.
7. Andersen CL, Jensen JL, Ørntoft TF. Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. *Cancer Res* 2004;64:5245–50.
8. Hollander M, Wolfe DA. Nonparametric statistical methods, 2nd edn. New York: Wiley, 1999. 816 p.
9. Fleiss JL. Statistical methods for rates and proportions, 2nd edn. New York: Wiley, 1981. 321 p.
10. Kang K, Peng X, Luo J, et al. Identification of circulating miRNA biomarkers based on global quantitative real-time PCR profiling. *J Anim Sci Biotechnol* 2012;3:4.

**Grant sponsor:** Associazione Italiana per la Ricerca sul Cancro (AIRC); **Grant numbers:** 10529 and 12162; **Grant sponsor:** Eurocanplatform: A European Platform for Translational Cancer Research, European Union grant; **Grant number:** 260791
**DOI:** 10.1002/ijc.28530
**History:** Received 25 July 2013; Accepted 30 Sep 2013; Online 8 Oct 2013

**Correspondence to:** Paolo Verderio, PhD, Unit of Medical Statistics, Biometry and Bioinformatics, Fondazione IRCCS Istituto Nazionale dei Tumori, Via Venezian, 1, 20133 Milan, Italy, Tel.: +39 0223903201, Fax: +39 0223902095, E-mail: paolo.verderio@istitutotumori.mi.it