

Moving from discovery to validation in circulating microRNA research

Paolo Verderio¹, Stefano Bottelli¹, Chiara M. Ciniselli¹, Marco A. Pierotti², Susanna Zanutto^{3,4}, Manuela Gariboldi^{3,4}, Sara Pizzamiglio¹

¹ Unit of Medical Statistics, Biometry and Bioinformatics, National Cancer Institute, IRCCS Foundation, Milan - Italy

² Scientific Directorate, National Cancer Institute, IRCCS Foundation, Milan - Italy

³ Department of Experimental Oncology and Molecular Medicine, National Cancer Institute, IRCCS Foundation, Milan - Italy

⁴ Molecular Genetics of Cancer, FIRC Institute of Molecular Oncology, Milan - Italy

ABSTRACT

Background: MicroRNAs (miRNAs), small noncoding RNAs, are involved in tumorigenesis and in the development of various cancers. Quantitative real-time polymerase chain reaction (qPCR) is the most commonly used tool to investigate miRNA expression, and qPCR low-density arrays are increasingly being used as an experimental technique for both the identification of potentially relevant miRNAs and their subsequent validation. Due to the reduced number of microRNAs to be validated, this phase is generally performed on ad hoc customized cards for which a technical robustness is assumed similar to that of the high-throughput cards used during the identification phase.

Methods: With the aim of investigating the degree of reproducibility between the 2 types of cards, we analyzed plasma-circulating miRNAs evaluated in 60 subjects enrolled in a colorectal cancer screening program.

Results: Our results showed a reproducibility between the 2 methods that was not fully satisfactory, with a concordance correlation coefficient equal to 0.69 (95% confidence interval, 0.12-0.92).

Conclusions: This report highlights the need to add a technical validation step to the high-throughput-based miRNA identification workflow, after their discovery and before the validation step in an independent series.

Keywords: Biomarkers, miRNAs, Preanalytical/analytical factors, Reproducibility, Technical validation

Introduction

MicroRNAs (miRNAs) are a class of small noncoding RNAs that play an important role in tumorigenesis and in the development of various cancers (1). Quantitative real-time polymerase chain reaction (qPCR) is the most commonly used assay to investigate miRNA expression, and qPCR low-density arrays are the most widely used technique for both the identification and the subsequent validation of the potentially relevant miRNAs (2-4). Available high-throughput qPCR low-density arrays (e.g., TaqMan miRNA low-density arrays) allow the simultaneous expression profiling of several miRNAs. They

represent a suitable tool for discovery purposes where the intrinsic lack of precision (absence of replicates) and specificity (multiple tests) are balanced by the opportunity of performing large-scale screenings for selecting promising miRNAs to be further investigated. In contrast, customized low-density arrays, designed with replicates of the miRNAs identified, ideally offer the possibility to increase the level of both precision and specificity of the experiment. There are, however, some differences among the 2 methodologies, that could significantly affect the results of the validation. Although it is commonly recognized that the technical robustness of the customized arrays represents a crucial step toward the process of addressing the clinical utility of the selected miRNAs, this issue is generally disregarded by assuming an a priori satisfactory level of reproducibility between the high-throughput assay and the customized one. The 2 assays, although based on the same principles, are implemented according to specific protocols that differ for some preanalytical (i.e., number of primers included in the solutions used for the preamplification and for the reverse transcription reaction) as well as analytical steps (PCR platform setting, number of thermal cycles and number of replicates). To our knowledge, although a good level of intra-reproducibility and inter-reproducibility was reported (5-7) for both assays, no information is available about their mutual reproducibility.

Received: July 24, 2014

Accepted: December 17, 2014

Published online: February 18, 2015

Corresponding author:

Dr. Paolo Verderio

Unit of Medical Statistics, Biometry and Bioinformatics

National Cancer Institute, IRCCS Foundation

Via Venezian, 1

20133 Milan, Italy

paolo.verderio@istitutotumori.mi.it



Based on the above considerations, we examined, from a statistical point of view, the transition from the discovery to the validation phase, by evaluating the reproducibility between the high-throughput and the customized assay for the evaluation of miRNA expression levels in plasma samples.

Materials and Methods

We considered 60 plasma samples from fecal immunochemical test positive (FIT+) subjects enrolled in the colorectal cancer screening program promoted by the Milan Local Health Authority and ongoing at our institute. Among the 60 subjects, 38 individuals (cases) presented a precancerous or cancerous lesion at colonoscopy, whereas 22 individuals resulted with no lesions (controls). The miRNA profile of each sample was analyzed using the TaqMan Array Human microRNA Card A (Applied Biosystems, Foster City, CA, USA) containing 381 mature miRNAs for the identification of miRNAs differentially expressed between cases and controls. Using this high-throughput assay (Megaplex card) in the discovery phase, we identified a set of 7 potentially relevant circulating miRNAs with a significant different expression (p -value ≤ 0.05) in the 2 groups compared (cases and controls), according to the Kruskal-Wallis test (8). These miRNAs were included in an ad hoc designed Custom TaqMan Array microRNA card (Customized card) for their validation. Total RNA (including small RNAs) extracted from 400 μ L of plasma, as previously described (9), was used for both the cards.

A Megaplex card for each of the 60 samples was prepared using the standard Megaplex Pools protocol with the following modifications: 10 μ L of reverse transcription (RT) product was added to the preAmp Reaction Mix, a preamplification step of 14 cycles was performed, and no post-preamplification and PCR reaction dilution were done (Optimized Blood Plasma Protocol for Profiling Human miRNAs Using the OpenArray Real-Time PCR System; Applied Biosystems). Since we started from a small amount of total RNA, these modifications were required to increase, with an unbiased approach, the quantity of specific cDNA targets of the 381 miRNAs. qPCR was done using FAST chemistry (Applied Biosystems, Foster City, CA, USA) in an ABI PRISM 7900 HT Real-Time PCR system (Applied Biosystems, Foster City, CA, USA).

Starting from the same 60 samples, a total of 8 customized cards were designed with the selected miRNAs in duplicate, together with their corresponding primer mix. Reactions were done according to the manufacturer's protocol (Life Technologies), with the following modifications: 4 μ L of total RNA was converted into cDNA, 9 μ L of RT product was preamplified, and a preamplification step of 14 cycles was done. Post-preamplification and PCR reaction dilution, even in this case, were not required. Customized cards were processed similarly to the Megaplex ones.

Data analysis was performed using as pivotal measure the \log_2 RQ (relative quantity) values obtained with the comparative cycle threshold (Ct) method (10); accordingly, the relative expression of each of i -th ($i = 1, \dots, 7$) miRNAs considered was computed as follows: $\Delta Ct_i = Ct_i - Ct_{ref}$, where Ct_{ref} is the average of the miRNAs identified as reference. The last were identified by using an ad hoc algorithm we recently developed that allows the selection of a subset of reference miRNAs suitable for data

normalization (11, 12). By starting from these values, the relative quantity was obtained as $RQ_i = 2^{-\Delta Ct_i}$ (i.e., $\log_2 RQ_i = -\Delta Ct_i$).

To investigate the technical reproducibility between the Megaplex and the customized cards, we computed as measure of agreement the concordance correlation coefficient (CCC) and its 95% confidence interval (95% CI) (13) starting from the \log_2 RQ values of the 7 miRNAs of interest, evaluated in the same 60 samples with the 2 methods. In line with our previous experience (14, 15), the observed value of CCC was considered fully satisfactory only when the lower limit of the 95% CI was equal to or greater than 0.80. The statistical analyses were carried out with SAS software (version 9.2; SAS Institute Inc., Cary, NC, USA).

Results

We obtained a CCC value equal to 0.69 (95% CI, 0.12-0.92) between the Megaplex and customized card by comparing the expression levels of the 7 miRNAs considered, in the same 60 samples. Specifically, taking into account the lower limit of the 95% CI of CCC, the 2 methods did not reach a fully satisfactory agreement. Similar results were obtained when Ct values were used for the comparison (data not shown). In addition the wideness of this interval strongly suggested the existence of a high variability between the 2 sets of data compared. For purposes of description and with the aim of investigating the role of each miRNA considered, we report in Figure 1 the distribution of the \log_2 RQ differences between the 2 methods for each i -th miRNA considered, within each k -th ($k = 1, \dots, 60$) sample ($\delta_{ik} = \Delta Ct_{ik(\text{megaplex})} - \Delta Ct_{ik(\text{customized})}$). The most critical miRNA in terms of reproducibility is miR-D, for which the customized card provided an overestimation (with respect to the expected value of zero). In addition miR-D showed the highest variability (interquartile range = 1.814). The most reliable result in terms of mean difference ($\bar{\delta}_i$) was observed for miR-E ($\bar{\delta}_{(\text{miR-E})} = 0.026$). According to the Kruskal-Wallis test, 6 of the 7 miRNAs considered (86%) had a significantly different expression in cases vs. controls (p -value ≤ 0.05), considering the \log_2 RQ values obtained with the customized card (Tab. I).

Discussion

The development of new cancer-related circulating biomarkers is a multiphase process that begins with their discovery, followed by a validation step (16, 17).

This brief report highlights the need to add a technical validation step to the high-throughput-based miRNA identification workflow, after the discovery of the miRNAs and before the validation step in an independent series. The introduction of an additional step permits us to verify the reproducibility of the assays and to correctly select miRNAs that will have a greater chance of succeeding in the subsequent validation. These considerations can be generalized to any research scenario involving high-throughput-based discovery of putative biomarkers, irrespective of the similarity of the underlying array's principles and chemistry, to reduce costs and time and to obtain robust new biomarkers to be transferred into the clinical setting. In conclusion, the statistical approach we have presented can be viewed as a diagnostics tool to be used for the evaluation of the reproducibility

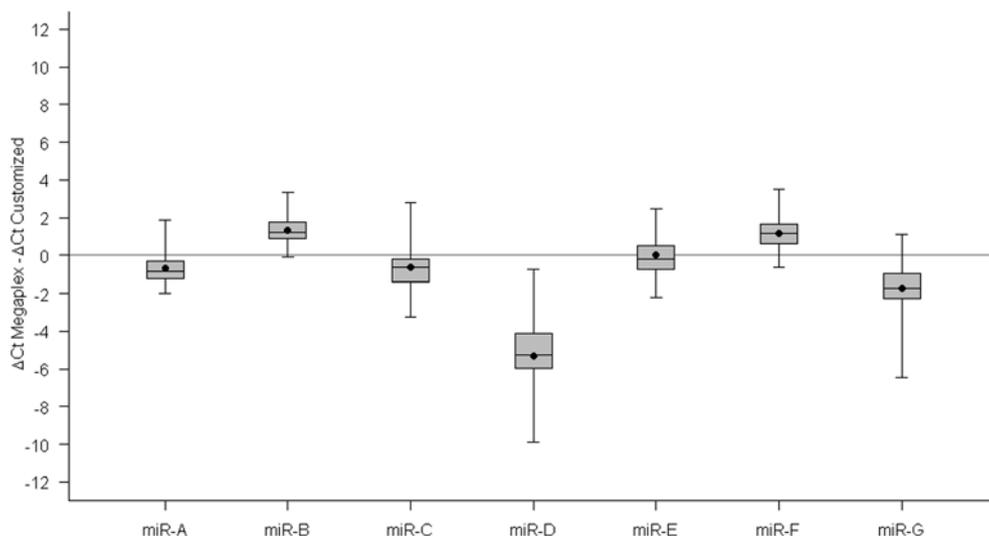


Fig. 1 - Boxplot of the distribution of the Δ Ct (Cycle threshold) differences between the Megaplex and the customized cards. Each box shows the 25th and 75th percentiles of the difference distribution; the horizontal line and the dot inside the box indicate the median and the mean, respectively. The limits of the 2 whiskers correspond to minimum and maximum values. The continuous horizontal line corresponds to the zero value.

TABLE I - Technical validation results

miRNA	Median \log_2 RQ (Cases)		Median \log_2 RQ (Controls)		Kruskal-Wallis p-value	
	Megaplex	Customized	Megaplex	Customized	Megaplex	Customized
miR-A	0.874	0.056	0.199	-0.714	0.002	0.002
miR-F	-1.569	-0.493	-0.460	0.345	0.020	0.036
miR-E	-0.522	-0.635	0.777	0.049	0.022	0.049
miR-C	-2.706	-3.145	-3.180	-3.712	0.028	0.038
miR-B	3.372	4.798	4.393	5.497	0.031	0.172
miR-G	-3.185	-4.734	-3.900	-4.997	0.043	0.028
miR-D	0.900	-4.119	0.480	-4.633	0.045	0.023

miRNA = microRNA; RQ = relative quantity.

in the technical validation step of the workflow involved in biomarker identification.

Disclosures

Experiments on human subjects: Written informed consent was obtained from all subjects upon approval of the study by the institutional review board and independent ethics committee.

Financial support: This work was supported by the Associazione Italiana per la Ricerca sul Cancro (AIRC, Grants 10529 and 12162 to M.A.P.).

Conflict of interest: The authors have no conflicts of interest to declare.

Meeting presentation: This work was presented at the WIN Symposium 2014, in Paris, France, 23-24 June 2014.

References

1. Reid JF, Sokolova V, Zoni E, et al. miRNA profiling in colorectal cancer highlights miR-1 involvement in MET-dependent proliferation. *Mol Cancer Res*. 2012;10(4):504-515.
2. Laudanski P, Charkiewicz R, Kuzmicki M, Szamatowicz J, Charkiewicz A, Niklinski J. MicroRNAs expression profiling of eutopic proliferative endometrium in women with ovarian endometriosis. *Reprod Biol Endocrinol*. 2013;11(1):78.
3. Zearo S, Kim E, Zhu Y, et al. MicroRNA-484 is more highly expressed in serum of early breast cancer patients compared to healthy volunteers. *BMC Cancer*. 2014;14(1):200.
4. Fortunato O, Boeri M, Verri C, et al. Assessment of circulating microRNAs in plasma of lung cancer patients. *Molecules*. 2014; 19(3):3038-3054.
5. Jensen SG, Lamy P, Rasmussen MH, et al. Evaluation of two commercial global miRNA expression profiling platforms for detection of less abundant miRNAs. *BMC Genomics*. 2011; 12(1):435.
6. Betts GN, Eustace A, Patiar S, et al. Prospective technical validation and assessment of intra-tumour heterogeneity of a low density array hypoxia gene profile in head and neck squamous cell carcinoma. *Eur J Cancer*. 2013;49(1):156-165.
7. Chen Y, Gelfond JA, McManus LM, Shireman PK. Reproducibility of quantitative RT-PCR array in miRNA expression profiling and comparison with microarray analysis. *BMC Genomics*. 2009;10(1):407.
8. Hollander M, Wolfe DA. The two-sample dispersion problem and other two-sample problems. In: Hollander M, Wolfe DA, eds. *Nonparametric statistical methods*. New York: Wiley; 1999: 141-188.

9. Zanutto S, Pizzamiglio S, Ghilotti M, et al. Circulating miR-378 in plasma: a reliable, haemolysis-independent biomarker for colorectal cancer. *Br J Cancer*. 2014;110(4):1001-1007.
10. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) method. *Methods*. 2001;25(4):402-408.
11. Pizzamiglio S, Bottelli S, Ciniselli CM, et al. A normalization strategy for the analysis of plasma microRNA qPCR data in colorectal cancer. *Int J Cancer*. 2014;134(8):2016-2018.
12. Verderio P, Bottelli S, Pizzamiglio S, Ciniselli CM, Gariboldi M, Pierotti MA. NqA: an R-based algorithm for the normalization and analysis of microRNA qPCR data. *Anal Biochem*. 2014;461C:7-9.
13. Marubini E, Pizzamiglio S, Verderio P. Agreement between observers: its measure on a quantitative scale. *Int J Biol Markers*. 2005;20(1):73-78.
14. Paradiso A, Volpe S, Iacobacci A, et al; Italian Network for Quality Assessment of Tumor Biomarkers. Quality control for biomarker determination in oncology: the experience of the Italian Network for Quality Assessment of Tumor Biomarkers (INQAT). *Int J Biol Markers*. 2002;17(3):201-214.
15. Terrenato I, Arena V, Pizzamiglio S, et al. External Quality Assessment (EQA) program for the preanalytical and analytical immunohistochemical determination of HER2 in breast cancer: an experience on a regional scale. *J Exp Clin Cancer Res*. 2013;32(1):58.
16. Verderio P, Mangia A, Ciniselli CM, Tagliabue P, Paradiso A. Biomarkers for early cancer detection: methodological aspects. *Breast Care (Basel)*. 2010;5(2):62-65.
17. Verderio P. Assessing the clinical relevance of oncogenic pathways in neoadjuvant breast cancer. *J Clin Oncol*. 2012;30(16):1912-1915.