Validation of suitable endogenous control genes for expression studies of miRNA in prostate cancer tissues

Short title: miRNA control genes prostate cancer

Jessica Carlsson^{1,2,5}, Gisela Helenius³, Mats Karlsson³, Zelmina Lubovac², Ove Andrén⁴, Björn Olsson² & Karin Klinga-Levan^{1, §}

Author's Affiliations:

¹Systems Biology Research Centre – Tumor Biology, University of Skövde, Skövde, Sweden

²Systems Biology Research Centre - Bioinformatics, University of Skövde, Skövde, Sweden

³Department of Laboratory Medicine, Örebro University Hospital, Örebro, Sweden

⁴Department of Urology, Örebro University Hospital, Örebro, Sweden

⁵School of Health and Medical sciences, Örebro University, Örebro, Sweden

[§]Corresponding author

Karin Klinga Levan, Department of Life Sciences, Systems Biology Research Centre – Tumor Biology, University of Skövde, SE541 28 Skövde, Sweden

Tel: +460500-448647

E-mail: Karin.klinga.levan@his.se

Key words: Endogenous control genes, microRNA, miRNA array prostate cancer, qPCR, Ct value

Abbreviations:

miRNA

microRNA

qPCR

quantitative PCR

Research article

Novelty and impact of the paper

We pinpoint to the importance of testing for optimal control genes in miRNA expression studies as we found that the stability of endogenous controls included in the arrays really behave distinctly different. The evaluation and comparison of two applications that is used in stability tests generated practically the same results.

Abstract

When performing qPCR analysis, there is a need for correction of technical variation between experiments. This correction is most commonly performed by using endogenous control genes, which are stably expressed across samples, as reference genes for normal expression in a specific tissue. In microRNA (miRNA) studies, two types of control genes are commonly used and these are small nuclear RNAs and small nucleolar RNAs. In this study, six different endogenous control genes for miRNA studies were investigated in a prostate tissue material from the Swedish Watchful Waiting cohort. The stability of the controls was investigated using two different software applications, NormFinder and BestKeeper. The results denote RNU24 as the most suitable endogenous control gene for miRNA studies in prostate tissue materials.

INTRODUCTION

MicroRNAs (miRNAs) belong to a class of small RNAs discovered in 1993 by Victor Ambros and colleagues in a genetic screen of *Caenorhabditis elegans*.¹ Later, miRNAs have been found in a diverse range of species including plants, animals and DNA viruses. MiRNAs play a key role in several biological processes including development, cell proliferation, differentiation and apoptosis.

1-4

Initially northern blots and microarrays were used for miRNA expression studies, but as qPCR protocols for quantitative miRNA expression studies were developed, these protocols have become the most widely used. Compared to other methods, such as northern blot, qPCR has some advantages because of its sensitivity and the low template requirements. In the current protocol, multiplex stem-loop primers are used in qPCR for conversion of mature miRNAs into cDNA, which allows for the simultaneous transcription of all miRNAs during a PCR.⁵

When a qPCR is performed, technical variations between the reactions have to be corrected for. One way to check for such causes of variation is to incorporate an endogenous control gene in the experiment, which can be used as a reference to normalize gene expression data sets.⁵ It is important that the control gene is carefully chosen since it has been proven that depending on which tissue is used, the expression of even the most commonly used housekeeping genes like *ACTB* (*Actin Beta*) and *GAPDH* (*Glyceraldehyde 3-phosphate dehydrogenase*) is not always stable between samples. Thus, since no universal endogenous control exists, it is recommended to test for the most suitable control gene in the tissues used.⁶

In this study, six different endogenous control genes for studies of miRNA expression were used. The endogenous control genes RNU48, RNU44, RNU43, RNU24 and RNU6B belong to a class of small non-coding RNAs, small nucleolar RNAs, while MammU6 (U6) belongs to the class small nuclear RNAs.⁶⁻⁷

The samples used for the test of stability of the control genes were formalin fixed paraffin embedded prostate cancer tissue material.

MATERIALS AND METHODS

Patient material

Patients were recruited from the population-based Swedish Watchful Waiting cohort,⁸ consisting of 1256 men with localized prostate cancer. These men had symptoms of benign prostatic hyperplasia (lower urinary tract symptoms) and were subsequently diagnosed with prostate cancer through transurethral resection (TUR-P). All men in this study were determined at the time of diagnosis to have clinical stage T1 a or T1 b, Mx, and Nx, according to the 2002 American Joint Committee on Cancer staging system TNM.⁹ The prospective follow-up time is now up to 30 years. This study includes samples from men who were diagnosed at the University Hospital in Örebro (1977–1991) and at four centers in the southeast region of Sweden: Kalmar, Norrköping, Linköping, and Jönköping (1987–1999). The study was approved by the ethical committee in the Uppsala-Örebro region. The material consisted of formalin fixed paraffin embedded (FFPE) tissues from 19 cases and the adjacent normal tissue in each individual. We collected cases randomly within each category of Gleason score (4-6, 7, 8-10) to get an equal distribution of histological differentiation.

MiRNA qPCR arrays

In this study the TaqMan[®]MicroRNA Array Set v2.0 from Applied Biosystems was used (Applied Biosystems, Foster City, CA, USA). This is a set of two cards (Card A and Card B) containing 364 TaqMan[®]MicroRNA assays plus 20 control assays per card, which enables quantification of totally 667 unique human miRNAs. Card A contains miRNAs that tend to be functionally defined, and are

broadly and/or highly expressed. The miRNAs in card B are less extensively studied and are narrowly expressed and/or expressed at low levels, and are usually not functionally defined.

Endogenous controls

The six endogenous controls investigated in this study were MammU6 (small nuclear RNA), RNU48, RNU44, RNU43, RNU24 and RNU6B (small nucleolar RNAs).⁶⁻⁷ Three of these controls were found on both cards, A and B, while the other three were only found in card B. On card A, only MammU6 was found in four replicates while the other two controls appeared just once. On card B, all six controls were found in four replicates (Table 1).

RNA extraction and cDNA preparation

Malignant and surrounding normal areas were marked by a pathologist on the paraffin blocks prior to punching out 3-4 cores ($\varnothing$ 0.6 mm) using the Tissue Micro Array equipment (Pathology devices, Westminster, USA). The cores were deparaffinized in a standard protocol using xylene and alcohols. The Recover All Total Nucleic Acid Isolation Kit (Ambion) optimized for FFPE samples was used to extract total RNA. A reverse transcription reaction using 4-10 ng of total RNA was performed using TaqMan[®] MicroRNA reverse transcription kit and Megaplex[™] RT primers, human pool v2.0 (Applied Biosystems) and the RT-product was then pre-amplified using Megaplex[™] PreAmp primers and Taqman[®] Preamp master mix (Applied Biosystems).

Quantitative PCR

The pre-amplified cDNA samples were diluted in a 0.1 x TE Buffer (pH 8.0) before use in the qPCR. The diluted pre-amplified cDNA was mixed with TaqMan[®] PCR master mix (No AmpErase UNG, Applied Biosystems) and was used in a 40 cycle qPCR reaction using TaqMan[®] MicroRNA A and B Cards v2.0. All reactions were performed on Applied Biosystems 7900 HT system, raw Ct (Cycle threshold) values were determined using the SDS software (Applied Biosystems) and

manual thresholds were applied for each miRNA. The raw Ct value represents the value where the amount of amplified cDNA reaches a defined threshold.

Calculations of stability of endogenous controls and statistical tests

The stability of the endogenous controls was evaluated using NormFinder and BestKeeper software applications.^{10 11} In NormFinder, delta Ct values were used as input and were calculated as $\text{raw Ct}_{(\text{normal})} - \text{raw Ct}_{(\text{malignant})}$. To estimate the overall expression variation and between group variations, a gene stability value is calculated.¹⁰ NormFinder returns a stability value and standard error for each endogenous control gene. BestKeeper uses raw Ct values to identify the most stable control gene based on pair-wise comparison between genes. BestKeeper determines the standard deviation and coefficient of variance (CV).¹¹

For statistical evaluations of Ct values, data were analyzed for differences among replicates by one-way ANOVA (Table 1). Student's t-test was performed for comparisons of normal and malignant tissues and a paired samples correlation was performed in the same analysis (PASW Statistics 18, SPSS Inc, Chicago, USA). In both tests the null hypotheses were assuming no differences between replicates (ANOVA), and no differences between tissue types (t-test).

RESULTS

To check for differences among replicates in the six endogenous controls, a one-way ANOVA analysis was performed. There were no significant differences ($p > 0.05$) among replicates in any of the endogenous controls (Table 1). Thus, mean Ct values were used for further evaluation of the stability of the endogenous controls included in the study by NormFinder and BestKeeper.

NormFinder returns a stability value and standard error for each endogenous control gene and BestKeeper determines the standard deviation and coefficient of variance (CV). To ease the

comparison between the results from the two software applications, standard error from NormFinder was transformed to standard deviation (Table 1).

Using both NormFinder and BestKeeper we were able to identify RNU24 and RNU44 as the two most stable control genes and with the exception of the order of MammU6 and RNU48, the concordance between the two tests was very good. The Student's t-tests revealed significant differences between malignant and normal samples in two of the endogenous controls investigated, RNU43 and RNU44 (Table 1).

To visualize the variation in expression of control genes, raw Ct values for each tissue sample was plotted in a graph (Figure 1). There are four peaks in the graph that represent deviations in expression in the least stable endogenous control RNU6B and two peaks for the second least stable control, RNU43.

DISCUSSION

Before analyzing large scale expression data, normalization is performed to minimize the variation due to technical reasons, and to increase the expression accuracy of biologically meaningful data. When analyzing data from qPCR experiments, the number of normalization methods is limited due to the small number of target genes investigated, which is not suited for the population-based normalization approaches that are commonly used in microarray studies. Therefore, normalization of qPCR data is usually dependent on using endogenous control genes as a reference for normal expression.¹²

When considering endogenous controls for miRNA expression studies, it is important that the chosen controls share similar properties with miRNAs. Properties to be considered include tissue-independent RNA stability and nucleotide size, but also whether the controls can be used in the miRNA assay design. Other small non-coding RNAs, not belonging to the miRNA class, like nuclear RNAs and nucleolar RNAs have a high abundance, are stably expressed, and are similar in

size compared to miRNAs. In addition it is unlikely that these non-coding RNAs are involved in pathways that regulate miRNAs, and therefore they fulfill the prerequisites for endogenous controls in miRNA studies.⁶

In this study, we investigated six different endogenous controls in formalin fixed paraffin embedded prostate tissue from the Swedish Watchful Waiting cohort. The six controls were present in the TaqMan[®] MicroRNA Array Set v2.0, where MammU6, RNU48, and RNU44 were present in both cards and RNU43, RNU24, and RNU6B were present only in card B. Thus, the endogenous controls were represented by different numbers of technical replicates (Table 1). We applied a one-way ANOVA on raw Ct values from the endogenous controls and this test revealed that no significant differences existed between replicates in any of the endogenous controls. Thus, we could safely use mean values for the replicates in the subsequent stability tests. Several software applications for assessing the stability of candidate control genes are available and in this study two of these software applications were used; NormFinder and BestKeeper.

In BestKeeper the value (CV) used to determine the stability value of a specific control is actually the coefficient of variation (C_v) for the same gene, multiplied with 100 ($(s/m)*100$). NormFinder calculates a stability value defined as the absolute mean value plus one SD, and it is calculated using the intergroup variation values from the control genes. GeNorm, a third software application used to calculate stability measures of endogenous controls, not used in this study, defines the stability value as the arithmetic mean of all pair wise comparisons.¹³

Both test systems used in this study, NormFinder and BestKeeper, recognized RNU24, followed by RNU44, as the two most stable controls. The third best endogenous control differed between the two software applications, as RNU48 was identified by NormFinder and MammU6 by BestKeeper. The slight difference in standard deviation between the tests may be explained by the variation in the number of input values used in the analysis. In NormFinder, delta Ct values were used (19

comparisons) whereas in BestKeeper, 38 raw Ct values were used in the calculations. According to the results from the present study we can conclude that both RNU24 and RNU44 would be suitable control genes for miRNA expression studies in prostate tissues since we could pinpoint these as the most stable control genes independently of the software application used. However, we propose that RNU24 is the best of these control genes due to the fact that the t-test, which actually verified the NormFinder results, revealed no significant difference between normal and malignant samples, which was not the case with RNU44. In addition, the slightly lower standard deviation for RNU24, compared to RNU44, further supports this choice (Table 1, Figure 1). The variation of expression of the control genes across different tissue samples illustrated in Figure 1 clearly supports the fact that RNU43 and RNU6B are not stable control genes, due to the high peaks in expression in some samples compared to the other samples and controls. Pearson correlation coefficient was calculated for all endogenous controls and a positive correlation between normal and malignant tissues was revealed in all comparisons.

Only a few studies concerning validation of endogenous controls for miRNA expression studies have been published. The fact that the controls used in these studies are not consistently the same as in our study, hamper proper comparisons with our results. Five of the control genes in our study, together with 13 other control genes, have previously been studied in different types of human leukemia samples and normal bone marrow samples using geNorm. The results were inconclusive since other control genes were considered to be stably expressed in the different tissues investigated, although RNU24 was among the most stable in all of them. The expression stability of RNU48 has previously been investigated in human breast cancer tissue as well as in prostate cancer tissue using both NormFinder and geNorm^{5, 15} and was found to be among the three most stable control genes in both types of tissue. The expression stability of RNU6B has also been investigated in two studies in human lung tumor tissue, human breast cancer tissue and canine lymphoma using

both NormFinder and geNorm. The results in these studies support our own results as the results were similar to ours, concluding that RNU6B is not a stable control gene.^{12,5,16}

So, from the results of the present study, we identified RNU24 as the most stable control gene. However, it could not be entirely verified from other studies mostly due to lack of corresponding data. The differences from our result between the studies might be due to differences in the amount of biological and technical replicates but may also be due to the stability analysis, since different software applications was used for this purpose (geNorm vs. NormFinder and BestKeeper). The result also reveals the need for a systematic evaluation of the performance of different applications in order to optimize them.

In a recently published study of miRNA profiling in prostate cancer, a miRNA gene has been used as control for normalization. This miRNA, hsa-mir130b, was the most stable miRNA in that data set, and thus more stable than the endogenous control gene RNU6B.¹⁷ As the miRNA hsa-mir 130b was included in our data set, the expression stability for this miRNA could be studied. We found that hsa-mir 130b was less stable than all of the control genes included in the present study (data not shown), suggesting that hsa-mir 130b is a poor candidate control gene.

In summary, this study investigated six endogenous candidate control genes for miRNA expression studies in prostate tissues. The controls were tested for stability across normal and malignant tissue, including statistical evaluation of the data following use of the NormFinder and BestKeeper software applications. RNU24 was proposed as the most suitable control gene for normalization of miRNA expression studies in prostate tissue. The results of this study further point at the need for evaluating endogenous control genes in every tissue investigated.

ACKNOWLEDGEMENTS

This work has been supported by the Swedish Knowledge Foundation through the Industrial PhD programme in Medical Bioinformatics at Corporate Alliances, Karolinska Institute and by Lions cancer research fund. The authors declare that they have no conflicts of interest.

REFERENCES

1. Stahlhut Espinosa CE, Slack FJ. The role of microRNAs in cancer. *Yale J Biol Med* 2006;79:131-40.
2. Ambs S, Prueitt RL, Yi M, Hudson RS, Howe TM, Petrocca F, Wallace TA, Liu CG, Volinia S, Calin GA, Yfantis HG, Stephens RM, et al. Genomic profiling of microRNA and messenger RNA reveals deregulated microRNA expression in prostate cancer. *Cancer Res* 2008;68:6162-70.
3. Shi XB, Tepper CG, White RW. MicroRNAs and prostate cancer. *J Cell Mol Med* 2008;12:1456-65.
4. Sassen S, Miska EA, Caldas C. MicroRNA: implications for cancer. *Virchows Arch* 2008;452:1-10.
5. Davoren PA, McNeill RE, Lowery AJ, Kerin MJ, Miller N. Identification of suitable endogenous control genes for microRNA gene expression analysis in human breast cancer. *BMC Mol Biol* 2008;9:76.
6. Wong L, Lee, K., Russel, I., Chen, C. Endogenous controls for Real-Time Quantitation of miRNA Using TaqMan® MicroRNA Assays: Applied Biosystems Application Note, 2007.
7. Kunkel GR, Maser RL, Calvet JP, Pederson T. U6 small nuclear RNA is transcribed by RNA polymerase III. *Proc Natl Acad Sci U S A* 1986;83:8575-9.
8. Johansson JE, Andren O, Andersson SO, Dickman PW, Holmberg L, Magnuson A, Adami HO. Natural history of early, localized prostate cancer. *JAMA* 2004;291:2713-9.
9. Sobin L, Gospodarowicz, M., Wittekind, C. TNM Classifications of Malignant Tumours, Seventh edition ed.: Blackwell Publishing Ltd, 2010.
10. Andersen CL, Jensen JL, Orntoft 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.
11. 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;26:509-15.
12. Peltier HJ, Latham GJ. Normalization of microRNA expression levels in quantitative RT-PCR assays: identification of suitable reference RNA targets in normal and cancerous human solid tissues. *RNA* 2008;14:844-52.
13. Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, Speleman F. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol* 2002;3:RESEARCH0034.
14. Mestdagh P, Van Vlierberghe P, De Weer A, Muth D, Westermann F, Speleman F, Vandesompele J. A novel and universal method for microRNA RT-qPCR data normalization. *Genome Biol* 2009;10:R64.
15. Larne O, Edsjö, A., Bjartell, A., Ceder, Y. . Development of a miRNA assay for prostate cancer detection 24th Annual Congress of the European Association of Urology, vol. 8 Stockholm, Sweden: Elsevier Science BV 2009:316-16.

16. Mortarino M, Gioia G, Gelain ME, Albonico F, Roccabianca P, Ferri E, Comazzi S. Identification of suitable endogenous controls and differentially expressed microRNAs in canine fresh-frozen and FFPE lymphoma samples. *Leuk Res* 2009.
17. Schaefer A, Jung M, Mollenkopf HJ, Wagner I, Stephan C, Jentzmik F, Miller K, Lein M, Kristiansen G, Jung K. Diagnostic and prognostic implications of microRNA profiling in prostate carcinoma. *Int J Cancer* 2010;126:1166-76.

FIGURE LEGENDS

Figure 1. The variation in expression of all control genes investigated, displayed by Ct values.

Color coding:

Blue – MammU6

Purple – RNU43

Red – RNU48

Turquoise – RNU24

Green – RNU44

Orange – RNU6B