Pitfalls and recommendations for microRNA expression analysis using qPCR

Guidelines to microRNA qPCR with examples from the miRCURY LNA™ Universal RT microRNA PCR system

Introduction

Over the last few years, qPCR has become the most widely used method for the study of microRNAs. It is fast, extremely sensitive and offers linear detection over several orders of magnitude. Exiqon's miRCURY LNA™ Universal RT microRNA PCR system can profile microRNAs on panels in just three hours and offers a linear range of 7 orders of magnitude. The ingenious design, using LNA™ and two microRNA specific primers, allows individual microRNAs to be accurately quantified from as little as 1pg total RNA. This level of sensitivity enables microRNA profiling from difficult samples such as FFPE, LCM and body fluids including serum/plasma and urine. However, in order to get biologically relevant results, it is important to set up the qPCR experiment correctly.

Even very small changes in microRNA expression levels, e.g. in comparing different disease stages, might be biologically significant. Yet, it will require a sufficient number of samples and correct normalization to reveal the differences with statistical significance. An insufficient number of replicates may obscure discovery of small but important differences. Poor normalization can lead to incorrect conclusions regarding the magnitude of regulation and even direction of fold change when studying differential expression. Proper study design and reliable normalization is therefore critical when analyzing differences in microRNA expression.

In this document, we will go through the steps involved in setting up a qPCR study and explain how to perform the normalization and data analysis.

Table of Contents

Introduction . . . 1
Setting up the study . . . 2
Determining the number of replicates in each step of the study . . . 3
Minimizing biological variation . . . 3
Avoiding technical variation . . . 4
Controls for optimal data validation . . . 5
Setting up the correct negative controls . . . 5
A no enzyme control . . . 5
A no template control (NTC) . . . 5
A mock RT reaction . . . 5
Purpose and use of positive controls . . . 6
Quality control using SYBR® Green . . . 7
Limit of detection . . . 7
Detection of microRNAs expressed at very low levels . . . 7
Normalization and data analysis . . . 8
Interplate calibration . . . 8
Selection of good reference genes . . . 9
Presentation of results . . . 10
Concluding remarks . . . 10
References . . . 11

Setting up the study

There are several important parameters to consider when setting up a qPCR study to ensure obtaining correct data. In this section, we will discuss some of the most important parameters such as sample size, replicas and controls.

What is the ideal study sample size?

The sample size needed to get a statistically relevant answer from the study depends both on the inter-sample variance, and on the expected amplitude of differences the study is intended to show. If you are looking for very large differences in a homogeneous sample type where a low biological variance can be expected (i.e. 10-fold changes in treated versus non-treated cell culture) you may not need a large data-set to show differences that are both biologically relevant and statistically significant. However, if you are searching for small differences in a very heterogeneous sample set where a high biological variance can be expected (i.e. two-fold changes in connection to disease states, measured on biopsies taken from a tissue type with a complex cell content in a wide set of human population) you may need a very large number of samples within each group in order to obtain statistical significance for the biologically relevant differences. Performing a power analysis can be a very helpful tool to evaluate the number of samples needed, and can be performed in a simple way using the experimental design feature in the Exiqon GenEx qPCR analysis software.

A case study is used throughout the document to better demonstrate the principles in study design and normalization:

Example Study

In the example study, age dependent disease progression in a genetically modified mouse model is investigated and compared to normal mice at the same time points. Thus, we have 6 biological groups (two genotypes with each 3 time points) to compare. We designed Exiqon Pick&Mix plates with 18 microRNAs that were found to be differentially expressed in a previous study. As will be discussed in this document, selection of the proper number and types of control genes are of particular important for the study outcome. In the example study, six small RNAs were selected for candidate reference genes based on prior knowledge of these often being stably expressed across different sample types (we chose 3 microRNAs and 3 small non-coding RNAs). The study set-up is illustrated in Figure 1.

Determining the number of replicates in each step of the study

The purpose of performing replicates (whether they be biological or technical) of any test is to remove noise caused by variation, and calculate statistical confidence intervals. It is important to evaluate the necessity of including both biological and technical replicates in the study. Which replicas should be performed depends on the steps introducing most variation, i.e. biological replicates as in sample size and technical replicates as in RNA extraction, RT and PCR steps. The number of replicas also depends on the level of variation, but should never be lower than three. If only two replicates are performed, and differ highly, it will be impossible to determine which one is the outlier – three replicates mediate pinpointing of an outlier.

"The number of replicates should never be lower than three"

Minimizing biological variation

Biological variation arises mainly from the differences between individuals. The aim of a study is usually to identify differences between two biological groups. Biological variation within groups may be low if the group to be studied is very homogeneous (genetically and environmentally). This could be true for in-bred animal strains, but also for small closed populations secluded from the surrounding world (e.g. certain island, mountain and tribal populations). If the specimen of interest is cell culture, the variation may be close to negligible. On the other hand, if the population to be studied is very heterogeneous), the biological variation within groups could be very large and demand for a high number of biological replicates. Heterogeneous populations are samples collected from a general population which could include a mix of ethnicities. Hence, samples from clinical studies including microRNA biomarker development studies are typically heterogeneous.

In summary, designing a study on a very homogeneous biological population has the advantage of keeping biological variation within groups low, thereby allowing small differences to show even within a reasonably low number of biological replicates. This, of course, is at the risk of hiding important biological differences in a larger population. On the other hand, using a heterogeneous study population will increase the chance of discovering important differences within the groups – but the higher level of variation may result in the need of a prohibitively large number of biological replicates in order to reveal these differences.

Figure 1

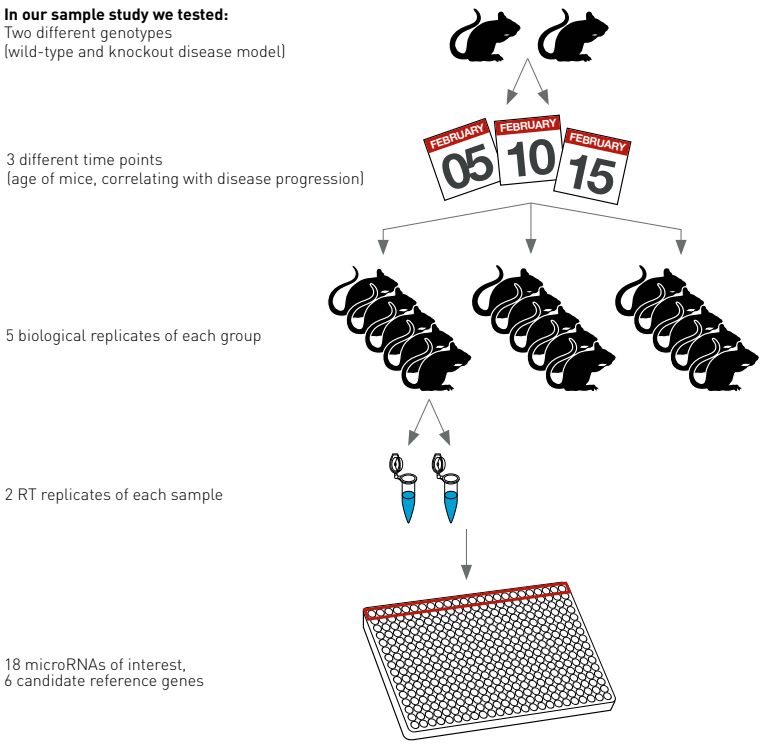


Figure 1. Set up of the example study. Two different genotypes were tested at three time points. Five biological replicates were used for each group. Two RT replicates were used for each sample and real-time PCR amplifications were performed using Exiqon Pick&Mix plates.

Avoiding technical variation

No technical manipulation of samples is performed exactly identically each time. The result of this is introduction of a certain level of variation with each step of sample manipulation. This can be countered by performing technical replicates. In general, qPCR studies comprise many different experimental steps to perform and thus many potential points for introduction of technical variance. This is also the case in microRNA qPCR. The use of a robust qPCR platform will help minimize technical variation. Figure 2 shows the high day-to-day reproducibility of the miRCURY LNA™ Universal RT microRNA PCR platform even between different days of study performance minimizing the risk of technical variance. Which types of technical replicates make the most sense to perform depends on the relative levels of variation each step introduces. For example, if the extraction procedure introduces a lot of technical variation while the PCR step is very reproducible, then making replicates at the PCR level may not give much improvement of the dataset – while replicates at the extraction level may remove a significant amount of noise and thereby reveal statistical significance to biological differences.

Sampling is a technical step which is often ignored, but may in fact introduce significant technical variation (often confused as biological variation). If the sample to be studied is a biopsy taken from a complex tissue type with many different structures and cell types (such as the kidney), three samples taken from the same individual may well be very different. On the other hand, three subsequent blood samplings from the same individual would probably be quite similar – whereas the time of day for sampling could introduce differences between samples.

Another point to consider in the design of a study is the method for RNA extraction of the samples that will be investigated. Different extraction methods may introduce different levels of variation, thus affecting the need for replicates. This becomes particularly important if you are interested in microRNA expression in body fluids such as serum/plasma, urine or cerebrospinal fluid. In most sample types, the majority of RNA comes from the larger RNA types such as mRNA, tRNA and particularly rRNA. However, if the sample is a virtually cell free body fluid, the total level of RNA may be very low – even if microRNAs are sufficiently expressed for an analysis to be feasible. As recovery of RNA is never 100%, starting out with a low level of total RNA can result in a significant loss of microRNA during extraction, resulting in lower sensitivity and higher variation in your experiments. This may be improved by adding an RNA carrier not containing microRNA, e.g. bacteriophage MS2 total RNA or yeast tRNA (see Tips box page 5). Recommendations for

Figure 2

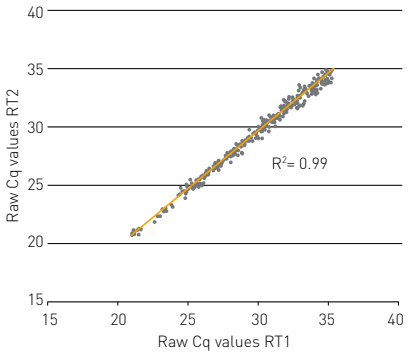


Figure 2. Day-to-day reproducibility of miRCURY LNA™ Universal RT microRNA PCR system. Different RT reactions using 40ng heart and liver total RNA were profiled on the Ready-to-Use PCR human panel I and II on different days. The correlation between raw Cq values from all microRNAs with signals below 35 Cq values is shown (total of 297 data points).

Figure 3

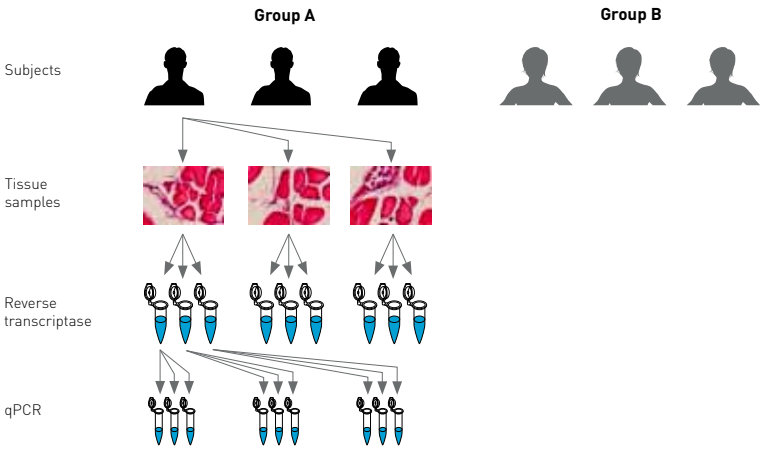


Figure 3. Example of an experimental setup. Proposed pilot study to determine the contribution to variation of each technical or biological level. This nested study uses triplicates at each level, and can be analyzed using a nested ANOVA.

working with serum/plasma can be found at www.exiqon.com/serum-plasma-guidelines

In our example study, the highest level of variation comes from individual differences between mice – therefore, we have chosen to use 5 biological replicates within each group. At the technical level we wanted to be able to remove RT and PCR outliers, so we performed duplicates at the RT level.

Our study was a relatively small validation study. Often a larger study is needed, and thus it will be desirable to keep the number of replicates to a minimum, while still performing those giving a significant contribution to the data-set. Before setting up a large, costly and time consuming study, it may therefore be helpful to perform a smaller pilot study simply with the purpose of determining levels and sources of variation, which replicates to perform, and how many biological samples needed within each group to show the expected levels of differences. This can be done through a nested design where three replicates are performed at each potential level of variation and analyze these using a nested ANOVA (Tichobad et al (2009), see Figure 3. If you are using GenEx for your data analysis, you can also find help in designing and analyzing your nested experiment here. Furthermore, good discussions and help can be found at www.qPCRforum.com

Controls for optimal data validation

In order to ensure that experiments have performed well, and enable troubleshooting if they have not, it is always a good idea to include a number of controls. How many, and which are relevant depends on the type of study and the main risks inherent to the study. The experimental controls include various types of negative and positive controls.

Setting up the correct negative controls

Negative controls are important for two purposes of equal importance: One is to reveal potential contamination in the experiment, and the other is to determine background levels for each assay. Background may arise from contamination, primer-dimer formation, or unspecific amplification and can result in an amplification curve with a high Cq in a SYBR® Green assay. The different types of negative controls reveal different potential issues.

Cq values
Different types of qPCR cyclers use different terms for their metric of quantification (Ct, Cp, TOP). We have chosen to use the standardized nomenclature quantification cycle (Cq) covering all of these terms, as suggested in the MIQE guidelines (Bustin et al, 2009).

Carrier RNA
The purpose of adding carrier RNA is to increase the total RNA amount as well as the RNA complexity in a sample as detailed below.

The increased RNA concentration will have the effect of minimizing loss of specific RNA template due to stickiness of plastic, extraction filters etc. This is particularly useful in sample types with low concentrations of total RNA, such as serum/plasma or synthetic RNA targets. The increased complexity works to increase the number of low-energy DNA duplex formation, thus competing for undesired unspecific bindings such as primer-dimer formation. Examples of carrier RNA not containing microRNA are MS2 bacteriophage total RNA and yeast tRNA.

Examples of negative controls to consider include:

1. A no enzyme control - This is an RT reaction performed without enzyme, but with RNA template. It will reveal if there is any DNA contamination in the RNA sample (either genomic or amplicon from previous experiments). At the same time, the RNA will act as carrier and help prevent primer-dimer formation. Thus, the level of primer-dimer is likely to resemble the level of primer dimer formation in a biological sample if any.

2. A no template control (NTC) - An NTC in the PCR reaction (using water or buffer as template) will reveal if the water or master mix used has been contaminated. An NTC in the RT reaction (RT on water or buffer) will equally reveal if the RT reagents have been contaminated (but it will not be possible to discriminate between contamination of RT reagents or PCR reagents). Both mentioned controls will also reveal if primer-dimer is an issue. However, it should be noted that because the primers are present in relative high concentration as the only nucleic acids in an NTC control, they may show increased tendency to form primer-dimers relative to when a template (RNA and DNA) is present (weak binding to unspecific targets will compete with low-energy primer-dimer formation).

3. A mock RT reaction - This is an RT reaction carried out only with carrier RNA. It may reveal unspecific amplification levels. If performing this control, it is important that the carrier RNA is free of any microRNA. Bacteriophage MS2 total RNA or yeast tRNA are examples of commonly used carrier RNA. This control becomes particularly important in experiments where carrier RNA has been used in the extraction, as it will reveal if the carrier used causes increased background.

An even better control for this may be to perform a mock extraction (extraction of carrier only), and use this as template in an RT reaction. This will additionally reveal any potential contamination during extraction. Table 1 shows how the mock RT of MS2 RNA was used as a negative control in the example study to verify that there is no discernible background in the microRNA assays. The small level of background seen for the small non-coding RNA could come from primer-dimer formation. However, the unspecific amplification takes off so late that it will have no effect on the specific amplification signal from our mouse samples.

Table 1

Roche - LightCycler 480 (Cp-values)	let-7a	miR-378	miR-145	RNU5g
MS2 (mock neg control 1)	no amp	no amp	no amp	35.18
MS2 (mock neg control 2)	no amp	no amp	no amp	no amp
Sample 1 (wt), 6 weeks	30.13	25.46	23.84	21.62
Sample 1 (wt), 6 weeks	28.23	25.28	23.6	21.5
Sample 2 (wt), 6 weeks	28.37	24.15	22.53	20.49
Sample 2 (wt), 6 weeks	27.19	23.84	22.13	20.13
Sample 4 (ko), 6 weeks	32.2	27.69	26.81	23.24
Sample 4 (ko), 6 weeks	31.97	27.61	26.48	22.96
Sample 5 (ko), 6 weeks	30.71	27.45	25.06	21.9
Sample 5 (ko), 6 weeks	30.07	26.62	24.77	21.6

Table 1. An example of running a mock RT negative control. MS2 bacteriophage total RNA was added as sample in the RT reaction. The PCR reaction is set to a cut-off at 40 cycles. PCR reactions with amplification after 40 cycles ("no amp") are considered blank. The data confirm that there is no discernible background in the microRNA assays. RNU5g shows little background but of no concern since the Cp values are very high compared to the specific values.

Negative controls should also be used to set the cut-off during data pre-processing. In most assays, there should be no background, but a few assays may have a slight tendency for primer-dimer formation resulting in an amplification curve late in the cycling protocol. The cut-off for what is considered a specific amplification signal and what is considered background should therefore be set based on the negative controls. For assays with no background, a high Cq cut-off (e.g. 37-40) can be accepted, whereas assays with background should use a cut-off lower than the observed background value. How low a Cq value should be set depends on the level of variation in the background. Typically, background values show higher levels of variation than specific values. The aim will be to set the cut-off at a level including specific amplification arising from low microRNA copy-numbers in the study, while eliminating amplification signal suspected to arise from unspecific amplification. Often subtracting 2-3 Cq values from the observed background will be reasonable.

Purpose and use of positive controls

Positive controls have the main purpose of ensuring the quality of the samples and reactions, and aid in troubleshooting if the results look sub-optimal. There are several options for good positive controls.

Adding a synthetic RNA template as spike-in in the RT reaction will give a nice control checking if the cDNA synthesis has worked well. Additionally, it will give an idea of the presence of inhibitors, either of the RT or PCR reaction. Not all assays are equally sensitive to different inhibitors. This means the presence of some inhibitors may not be revealed by the spike-in. Therefore, we would also recommend monitoring the efficiency of all assays, as described in the subsequent section "Quality control using SYBR® Green".

A spike-in RNA template can also be used to monitor the uniformity and efficiency of the extraction procedure. In this case, the spike-in should be added to the sample lysate with or immediately after the lysis buffer – otherwise, degradation is a significant risk. However, if the PCR reaction fails, it will not be possible to discriminate between a bad extraction and a failing RT reaction.

Note that an RNA spike-in is supplied with the miRCURY LNA™ Universal RT microRNA PCR system cDNA synthesis kit, and that the primers for amplifying the cDNA copy of the spike-in RNA are supplied with the SYBR® Green master mix kit as well as present in the pre-designed panels and for selection in the custom-designed Pick&Mix plates. Further RNA spike-in templates and detection primers are available upon request through the Exiqon customer support, allowing usage of different spike-ins to monitor extraction and cDNA synthesis.

Figure 4

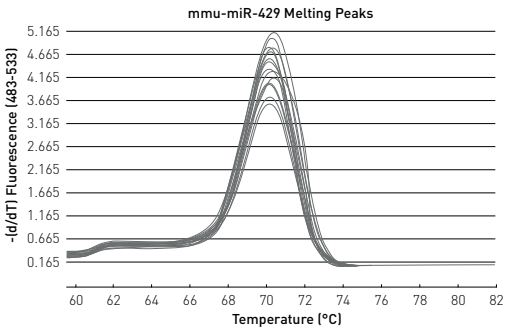


Figure 4. Melting curve analysis is useful for assay quality control. The figure shows the dissociation curves for one of the miRCURY LNA™ Universal RT microRNA PCR, mmu-miR-429 assays across different samples used in the example study. The shape of the curve and the reproducibility of T_m (peak melting temperature) can be used as quality control. The melting curves show only one major T_m peak corresponding to a well-defined T_m of the amplicon demonstrating specific amplification of mmu-miR-429.

In case of no amplification signal in the qPCR, troubleshooting may be carried out to discriminate between a true negative (the microRNA of interest is not present in the sample at a detectable level) and a malfunction of the assay. In this process, it should be kept in mind that validation may reveal different results from a screening, particularly if different methods were used. In order to discriminate between true and false negatives, the use of a positive control RNA target may prove useful. If a certain tissue or cell type has a known high expression level of the microRNA of interest, a good quality commercially available RNA sample may be used as positive control. In our example study, we know in advance that all the microRNAs of interest are well expressed in our samples. However, not all microRNA species are expressed in all cell and tissue types, and it may prove difficult to find a biological sample to act as positive control. In this case, a synthetic RNA template can be used. The target sequence for each of the LNA™ microRNA PCR primer sets is given in the assay product details at www.exiqon.com/miRNA-pcr-primer

Quality control using SYBR® Green

In addition to negative and positive controls, assay properties such as amplicon melting temperature (T_m) and assay efficiency may be used for quality control. The use of SYBR® Green detection in the miRCURY LNA™ Universal RT microRNA PCR system allows generation of dissociation curves, and the T_m of the specific amplicon is reproducible from run to run. Therefore, the T_m can be used to verify that the same amplicon is amplified from sample to sample. Note that dissociation curves should always contain a single peak. Figure 4 shows an example of the melting curves for mmu-miR-429 on one of 4 plates in the example study.

While the melting curve is a good control, it should be noted that different cyclers may have different temperature calibrations, and this can cause T_m to differ from instrument to instrument. For this reason, the expected T_m of each amplicon should be determined on the instrument used, using a good positive control. The assay efficiency can be calculated based on the amplification curve using e.g. LinRegPCR, which can be downloaded from the "downloads" menu at www.hartfaalcentrum.nl A reduced efficiency compared to the expected (as determined using a good quality positive control) may indicate the presence of inhibitors in the sample.

Again, the efficiency should be determined based on the instrument used for amplification, but a good pointer is that all Exiqon's LNA™ microRNA PCR primer sets have been validated to have an efficiency within the range of 1.80 to 2.10. Some assays may be more sensitive to inhibition than others, and thus it is quite possible to experience inhibition of some of these assays when used for microRNA detection in the same sample.

Figure 5

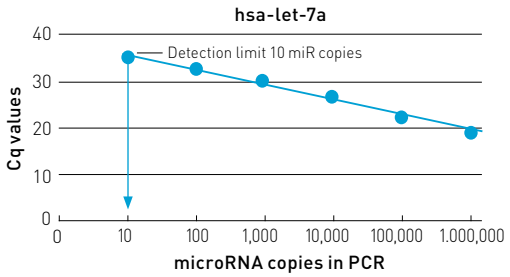


Figure 5. Exiqon's LNA™ microRNA PCR primers allow detection down to 10 microRNA copies. Data are generated from a serial dilution of hsa-let-7a synthetic microRNA (from 10^8 to 1000 copies) used for cDNA synthesis followed by real-time PCR amplification.

Limit of detection

More than 95% of the primer sets available in Exiqon's miRCURY LNA™ Universal RT microRNA PCR portfolio are sensitive down to the equivalent of 10 RNA copies in the PCR reaction – and all of them will detect a minimum of 1000 copies (see Figure 5 and tips box about relevance of limit of detection). However, if you wish to determine the limit of detection of a specific assay in the sample types of your experiment, you will need to perform a dilution series. This is also the more commonly used method for efficiency calculation (where LinRegPCR is an alternative). A dilution series is a necessity if planning a study with absolute microRNA quantification.

The limit of detection is a true indicator of sensitivity

It is a common misconception that lower Cq values from qPCR assays mean higher sensitivity. A Cq value has to be compared either to another Cq value in another sample (leading to a calculation of relative expression) or to a standard curve where known amounts of the target has been analyzed (leading to absolute quantification).

In order to find out what the true sensitivity of a qPCR assay is, it is necessary to run a dilution series of known input amounts including a negative sample where the template is not present.

The limit of detection is usually defined as the last point on that curve where the curve is still linear (e.g. any point that does not lie on the linear regression curve must be excluded).

Before performing this experiment, you should decide whether a cDNA dilution series is sufficient, or an RNA dilution series is more appropriate. cDNA dilution is certainly both easier and cheaper, since you need only making one RT reaction and then dilute on this. By diluting the RNA, you will need to perform multiple RT reactions – but you will also take into account the potential reduction in RT efficiency on dilute RNA samples, the detection limit of the RT step, and potential inhibitors of the RT reaction. All of this information will be lost in a cDNA dilution series. When diluting the RNA sample, we strongly recommend diluting in to a carrier RNA (such as bacteriophage MS2 total RNA) to avoid loss of RNA due to plastic adhesion.

Detection of microRNAs expressed at very low levels

If the microRNA(s) of interest is expressed at very low levels the microRNA may eventually fall below the limit of detection in the qPCR reaction. This may be circumvented by increasing the RNA sample input. In good quality RNA samples, we have been able to increase sample input by at least 10-fold without inhibition.

However, if there is even a low concentration of PCR inhibitors present in the sample, increasing the RNA input may prove detrimental to the assay performance. In complex samples such as serum and plasma, which contain low levels of microRNAs combined with presence of PCR inhibitors, the Cq values obtained will often be in the high end (in the range 30-35). This is perfectly acceptable as long as the signal can easily be discerned from the background. We advise trusting the obtained Cq values, rather than attempting to add more sample with the risk of inhibition. Increasing the amount of cDNA input in the PCR reaction (i.e. diluting the cDNA less) should be done with caution, as the RT buffer contains chemicals that may inhibit the PCR if present in too high concentrations. Our experience with these types of complex samples is that the cDNA should not be diluted less than 40x (compared to the recommended 100x) before addition to the qPCR reaction.

Serum/plasma samples

Read more about microRNA profiling in blood serum/plasma at www.exiqon.com/serum-plasma-guidelines

Normalization and data analysis

The purpose of most microRNA qPCR studies is to look for biological differences. However, these may be masked behind technical variation. The purpose of normalization is to filter out this technical variation, thus making the true biological differences clearer. As described above, technical variation may arise from the differences in sample collection and handling, RNA extraction procedure and resulting RNA quality.

The effect of these types of biological difference and technical variation can be reduced by normalization of qPCR data to either the global mean, or one or more endogenous control genes, in order to allow correct quantification of microRNA levels. Selection of endogenous control genes is far from trivial and advice on how to better select the controls is given below.

Interplate calibration

Some cycler types, e.g. the Roche LC480 using the 2nd derivative method for Cq calculations, give very reproducible results from run to run. In our hands, there is no need for interplate calibration when using LC480. However, many qPCR cyclers do produce technical differences in amplification signal from run to run – particularly if basing the Cq on baseline and threshold calculations. For this reason, it is often advisable to have all replicates and samples for one gene located on the same plate. However, in large studies with many samples and/or many microRNAs of interest (e.g. Exiqon's pre-designed human and rodent panels), this is not doable. Instead, this calls for inter-plate calibration, where the Cq values across plates are calibrated to a reference with identical expression in each plate. In all Exiqon's ready-to-use panels, whether the pre-designed panels or the custom-designed Pick&Mix panels, such a interplate calibrator (IPC) is present in triplicate.

Formula used by GenEx for interplate calibration:

$$Cq_{norm} = Cq_{GOI} - \frac{1}{m}\sum_{j=1}^{n}\left(Cq_{IPC_j} - \frac{1}{n}\sum_{i=1}^{n} Cq_{IPC_i}\right)$$

where GOI is the gene of interest, m is the number of interplate calibrators in run "m", and n is the total number of interplate calibrators.

Each IPC well contains both the detection primer set and the DNA template (making amplification independent of the sample). The purpose of triplicate presence is to avoid discarding an entire plate in case of pipetting error to an IPC well. In addition, we have placed the three IPC replicates in unique patterns for each of our plate types, making it possible to identify the plate used by its IPC pattern. Interplate calibration can be performed with the formula in the textbox. For users of Exiqon's GenEx qPCR analysis software, interplate calibration is integrated in the

pre-processing described in the GenEx qPCR analysis software guide www.exiqon.com/mirna-pcr-analysis

Selection of good reference genes

It is critical and far from trivial to select the optimal reference gene or combination of reference genes for normalization of quantitative microRNA analysis. If the study is a screening using full panels containing several hundred microRNAs, the best method for normalization may well be the global mean of all expressed microRNAs (Mestdagh et al. 2009). In smaller studies, one or more stably expressed endogenous controls can be used. These can be chosen from a screening study where they showed stable expression with the same behavior as the global mean (Mestdagh et al. 2009; Chang et al. 2010). If no such screening has been performed, a number of candidate reference genes should be selected from the literature – a summary can be found in Meyer et al, 2010.

Listed below are features that characterize a good endogenous control candidate for normalization of microRNA quantification:

- Expression at similar level to the microRNAs of interest in the study
- Invariant expression across all samples in the study (i.e. not regulated under the experimental conditions)
- Similar small size as the microRNAs (i.e.. similar stability, extraction and quantification efficiency)

It is important that the reference genes are empirically validated for each study. No single reference gene can be recommended for use across all types of tissue and cells. Even among the most commonly used reference genes (e.g. beta-actin and GAPDH for mRNA expression) significant variations in expression level between samples can be observed (De Kok et al, 2005).

In general, we recommend that when studying microRNA expression, the endogenous reference genes used should be stably expressed microRNAs rather than longer RNA species such as snoRNAs and snRNAs including U6. This is because microRNAs are so short that they may have very different behavior during extraction and reverse transcription compared to longer transcripts (Vandesompele et al., 2002). Depending on origin of sample, some commonly used reference gene microRNAs include (Bargaje et al, 2010; Chang et al, 2010; Liang et al, 2007; Peltier and Latham, 2008):

"Use stably expressed microRNAs as endogenous references"

hsa-miR-103	hsa-miR-423-3p	hsa-miR-191
hsa-miR-16	hsa-miR-423-5p	hsa-let-7a

Exiqon's guidelines for microRNA qPCR in serum/plasma discuss reference gene microRNA candidates specifically for use with serum and plasma samples., see www.exiqon.com/serum-plasma-guidelines

If no stably expressed microRNAs can be found, other small non-coding RNA may be used – but keep in mind that these may be more than 100 nucleotides long, in contrast to the ~18-25 nucleotides for microRNA. Genes such as U6 (107 nt) and 5S rRNA (121 nt) have often been used as reference genes. These are, however, not the optimal choices. First of all, they are both expressed within cells at much higher levels than most microRNA, while not being expressed at significant levels in cell-free body liquids. Secondly, they are both much larger than microRNA.

Figure 6

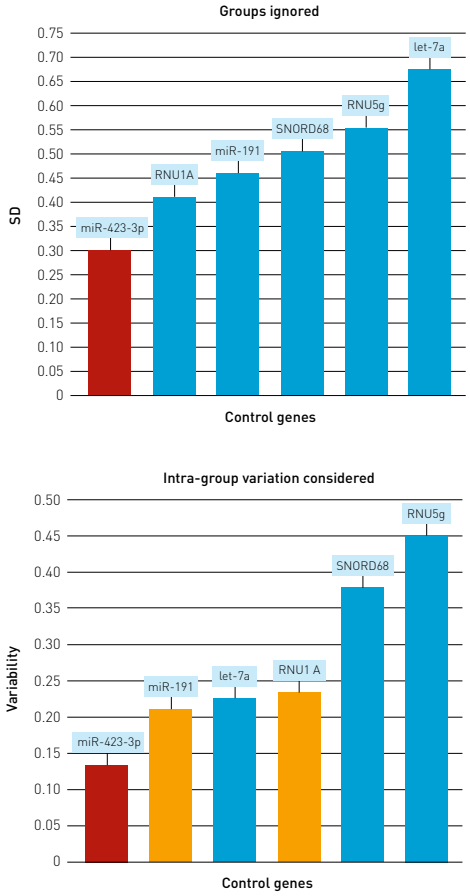


Figure 6. Example on how to select the optimal reference genes from the candidates tested. Exiqon's GenEx software was used to evaluate and identify reference genes for our example study. A NormFinder test was run ignoring groups, and with groups. miR-423-5p, miR-191 and RNU1A came out as stable in both tests; let-7a shows up as stable in one test but not in the other. We chose to use miR-423-5p, miR-191 and RNU1A as reference genes. Red marks the best single reference gene, yellow marks the optimal pair of reference genes.

Special care should be taken for the choice of reference genes if the sample is any kind of body fluids where microRNAs are found extracellularly in exosomes or in micro vesicles. These sample types are virtually cell free, and thus RNA types expressed only within cells or released as degradation products (such as U6, 5S rRNA, snoRNAs etc.) will not be appropriate. For these sample types, the recommendation of normalizing only to microRNA is emphasized even more.

Generally, it is found that evaluation of several controls is necessary in order to find the most appropriate candidates for each microRNA quantification study. If there are no previous data to base the choice on, we would recommend testing 5-6 different candidates. From these, normalization to 2-3 stably expressed genes, preferably microRNAs, would typically be sufficient.

In our example study, we chose three microRNAs and three small non-coding RNA as candidate reference genes. We analyzed their stability using Normfinder, part of the GenEx software package, and compared the results with or without taking inter-group variation into account (Figure 6). If we look at overall expression in all samples, we see that miR-423-3p, RNU1A and miR-191 all have standard deviations below 0.5. If we add intergroup variation to the analysis, taking into account the risk that candidate reference genes could vary with genotype and disease stage, we see that the same three genes are highlighted. Let-7a also shows low variability when considering groups, but is discarded due to large variation between samples.

"Normalize to 2-3 stably expressed reference genes"

Presentation of results

Once the qPCR experiment has been performed and the data pre-processed properly, getting the results becomes easy if using specific qPCR data analysis software such as Exiqon GenEx qPCR software (www.exiqon.com/qpcr-software). This qPCR software package, makes it possible to create publication ready figures. Statistics can be completed with a few clicks and does not require any prior biostatistician skills.

Looking at our example study, with just a few clicks in the Exiqon GenEx qPCR software we created a principal component analysis indicating that there is no difference between the wild-type animals at the three time points studied, and that the disease model grouped with the wt at the first time point. However, as the disease progresses over time, the samples group distinctly separate from the wild-type (Figure 7A). A heat-map confirms this pattern, and further gives a strong indication as to which microRNAs would make the strongest disease markers (Figure 7B). ANOVA test, t-test and various other tests are equally simply created using the Exiqon GenEx qPCR software package.

Figure 7

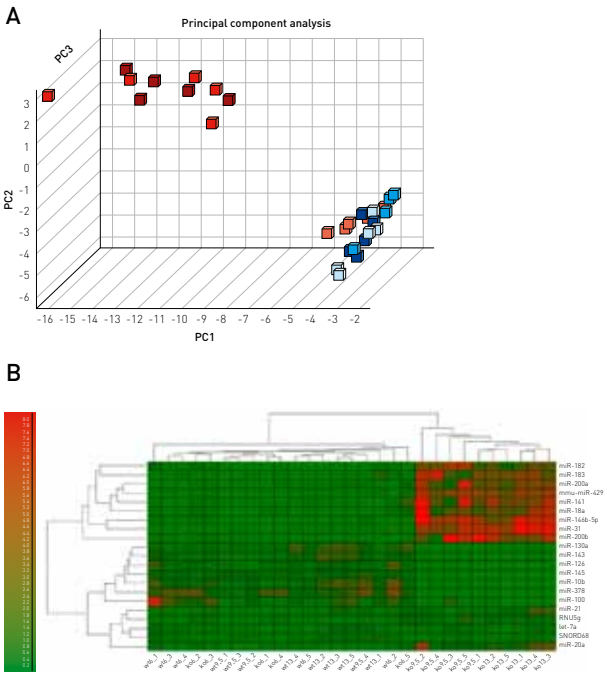


Figure 7. Presenting the qPCR study results by PCA analysis and heat map. Using the example the figure shows how results from a qPCR study easily are presented using the Exiqon GeneEx software. A prinicipal component analysis (A) shows that two groups (darker shades of red) group together, and distint from the remainding four groups (lighter red and shades of blue). A heat-map (B) confirms this grouping, and further gives and impression of which microRNAs co-regulate, and which have the higher level of regulation.

Concluding remarks

It is clear that properly setting up qPCR experiments is of critical importance for the success of the study. As described in this guide a proper qPCR experiment requires profound understanding of the qPCR technique and the purpose of the study as well as careful consideration of each step of the experiment in terms of selection of sample size, replicates and the role of normalization. As shown in the example study, qPCR is a very powerful technique when properly used. We were able to identify promising microRNA biomarker candidates which can now be studied further.

By following the guidelines presented here, you will have a good foundation for properly setting up your own microRNA qPCR study and obtaining data correctly. Choose your experimental setup wisely and you have taken the first step towards the next big microRNA break-through.

References

Bargaje, R. et al. (2010), RNA 16, 16-25
Bustin, S.A. et al.(2009), Clin. Chem., 55, 611-22
Chang, K.H. et al. (2010), BMC Cancer, 10, 173
De Kok, J.B. et al. (2005), Lab Invest., 85, 154-9
Liang, H. et al. (2007), BMC 8, 166
Mestdagh, P. et al. (2009), Genome Biol, 10, R64
Meyer, S.U. et al. (2010), Biotechnol Lett., 32, 1777-88
Peltier, H.J. and Latham, G.J. (2008), RNA, 14, 844-52
Tichopad, A. et al. (2009), Clin Chem., 55, 1816-23

Additional reading and resources:

The following publications can be downloaded from **www.exiqon.com/mirna-pcr**
- Profiling of microRNA in Blood Serum/Plasma - guidelines
- Instruction Manual for microRNA LNA™ Universal RT microRNA PCR
- RNA Purification from Blood Plasma & Serum – Human
- RNA Purification from Blood Plasma & Serum – Mouse

Experimental data for the miRCURY LNA™ Universal RT microRNA PCR system are available at **www.exiqon.com/mirna-pcr**

Data analysis software and tools are described in more details at **www.exiqon.com/mirna-pcr-analysis** including SDS template files for AB instruments and Exiqon's GenEx software access and guidelines

www.exiqon.com/e-talk displays presentations about the miRCURY LNA™ Universal RT microRNA PCR system and the advantages of using the LNA™ technology

Exiqon, LNA™, and miRCURY™ are registered trademarks of Exiqon A/S, Vedbaek, Denmark. Locked Nucleic Acids (LNA™) are covered by patents and patent applications owned by Exiqon A/S. SYBR® Green is a registered trademark of Invitrogen. Concerning miRCURY LNA™ Universal RT microRNA PCR: NOTICE TO PURCHASER: LIMITED LICENSE Purchase of this product includes an immunity from suit under patents specified in the product insert to use only the amount purchased for the purchaser's own internal research. No other patent rights are conveyed expressly, by implication, or by estoppel. Further information on purchasing licenses may be obtained by contacting the Director of Licensing, Applied Biosystems, 850 Lincoln Centre Drive, Foster City, California 94404, USA.

For life science research use only. Not for use in diagnostic procedures.

Contact information

Outside North America
Exiqon A/S · Skelstedet 16
DK-2950 Vedbaek · Denmark
Phone +45 45 650 929
Fax +45 45 661 888

North America
Exiqon Inc. · 14 F Gill Street
Woburn, MA 01801 · United States
Phone (781) 376 4150
Fax (781) 376 4152

www.exiqon.com

EXIQON
Seek Find Verify

918391 - 03/2011