How to do successful gene expression analysis using real-time PCR

Jan Hellemans and Jo Vandesompele
Biogazelle, Zwijsenlaan, Belgium

experiment design
Power analysis
qPCR experiments require a sufficient number of biological samples to be included so meaningfully and statistically significant results can be obtained. All too often researchers use too few samples, preventing them from drawing strong conclusions. Power analysis may overcome the type of failure.

Sample vs gene maximization
Most qPCR experiments involve more measurements than fit in a single plate. When spreading reactions across plates one can choose to distribute genes and keep samples clustered in a plate (sample maximization), or to cluster genes and distribute samples (gene maximization). In a relative quantification study, researchers are typically interested in comparing expression levels for a given gene among samples. Since sample maximization minimizes technical variation between samples it is the preferred set-up. Whenever samples have to be spread across plates, inter-run calibrations have to be included to detect and correct for unwanted run-to-run variation.

Plate layout
Thinking about a proper plate layout, with samples and assays ordered in rows and columns enabling the use of multi-channel pipettes, streamlines plate filling and annotation and reduces the risk of pipetting mistakes. It is acceptable to have empty wells in a plate.

sample preparation
DNA extraction
DNA extraction is the start of all gene expression studies. Quality and yield are important. Depending on the assay type, RNA samples may require DNase treatment to prevent gDNA amplification.

DNA quality control
RNA quality has a profound impact on the results, in terms of the significance of differential expression, variability of reference genes and sample classification performance using a multi-gene signature. RNA quality is assessed in terms of integrity and purity. Integrity can be evaluated by means of inspection of the 18S & 28S ribosomal RNA peaks or by performing a 0.5-2.0 18S:28S ratio test. The SPUD assay can be used to measuring sample purity. Ideally, all samples are pure and fully intact. If this is not possible, it is important to compare samples of similar quality.

Pre-amplification
The amount of available RNA may be insufficient to achieve a sufficient sensitivity to screen all genes of interest. Sample pre-amplification (whole transcriptome or target specific) may offer a way out.

assay design and validation

Design
In-house-designed assays have the advantage that you may avoid gDNA co-amplification provided the gene of interest contains introns that are sufficiently large. Classic designs on the other hand may use a larger design space (all exons rather than a limited set of exon pairs) increasing the design success rate. Primers are advocated for their increased detection specificity and allow for qPCR multiplexing. A well-designed qPCR assay may however provide the same level of specificity at slower cost.

In-slice validation
The specificity of designs can be evaluated in silico with dedicated tools. Bisearch is our preferred tool because it allows for controlling the number of mismatches that would still allow qPCR amplification. To avoid obfuscatio de novo PCR runs, one has to make sure no SNPs are located at the 3’ end of the primer binding region.

Secondary structures in the primer binding region are to be avoided because they may negatively impact PCR performance.

Empirical validation
Every assay has to be validated in the lab, even after having done in silico quality assessment. This also applies for the majority of commercial assays that have not yet been lab validated.

Specificity can be evaluated by amplicon size analysis and, for SYBR assays, the interpretation of melt curves. Co-amplification of homologous sequences may be tested with these methods.

Amplification efficiencies can be assessed by measuring a serial dilution series. Good assays have efficiencies in the 92%-110% range.

relativie quality control

Efficiency correction
The delta-delta Cq quantification method assumes 100% amplification efficiency. For suboptimal assays the use of the Pfaff or dN base model, supporting variable efficiencies is recommended.

Multiple reference gene normalization
It has been clearly demonstrated that the use of single non-validated reference genes results in suboptimal or even plainly wrong results. The dN base quantification model combines the information of multiple reference genes for proper normalization.

Inter-run calibration
To detect and correct for technical run-to-run variation so called inter-run calibrations may be included in the experiment. These identical samples should yield the same result in different runs; observed technical variations can subsequently be used to correct all measurements.

Error propagation
Plotting the error bar for relative quantities involves more than simply calculating the standard deviation on the Cq values of PCR replicates. Error on the gene of interest and the reference gene(s) have to be properly combined and propagated throughout calculations to end up with the correct error on the final results.

data quality control

Amplication and melt curves
For each assay the amplification curves of all samples should run in parallel, having a similar sigmoidal shape.

For SYBR assays, the derivative melt curve can be used to detect non-specific amplification or primer dimer formation.

Replicates
Depending on the required precision, some variation in replicate Cq values can be tolerated. Objectively identified bad PCR replicates (outlier, known pipetting issue, abnormal curve) should be excluded from analysis.

Negative controls
Ideally, negative controls should not result in any amplification signal. It may be acceptable to accept some signal for negative controls as long as their Cq value is sufficiently larger than those of the other samples. Requiring a 5 cycle difference corresponds to accepting ~3% non-specific signal.

Reference gene expression stability
Even after having selected the best set of reference genes in a gelform pilot study, their stability should be verified in subsequent experiments. Both the gelform M-value and the CV on normalized relative quantities of reference genes can be used for this.

statistical analysis

Biological replicates
It is important not to drive statistical conclusions on PCR replicates. A representative, independent and sufficiently large set of biological replicates is needed (see power analysis).

Log transformed relative quantities
Many of the commonly applied statistical tests, including the t-test and ANOVA, rely on the normality of the data. Relative gene expression levels are known to not follow a normal distribution. Log transformation will make these data more symmetrical, thereby approaching normality.

Test selection
Selecting a statistical test involves more than just personal preferences. The wrong test may falsify the real impact of your study. A variety of such tests exist and may help selecting the proper test.

Multiple testing correction
The cutoff of 0.05 to call a result significant limits the chance for false positive conclusions to 5%. When testing multiple hypotheses (genes), this 5% applies to every test, resulting in a much larger overall false positive risk. This problem can be corrected for by means of multiple testing correction, typically Bonferroni, for small numbers of genes and false discovery rate control for larger screening studies.