Molecular Therapy
Methods & Clinical Development
Review

Consensus guidelines for the validation of qRT-PCR assays in clinical research by the CardioRNA consortium

David de Gonzalo-Calvo,[1,2] Monica Marchese,[3] Jan Hellemans,[4] Fay Betsou,[5] Nanna Lond Skov Frisk,[6,7] Louise Torp Dalgaard,[6] Päivi Lakkisto,[8,9] Carole Foy,[10] Andreas Scherer,[11,12] María Laura Garcia Bermejo,[12,13] Yvan Devaux,[14] and On behalf of EU-CardioRNA COST Action CA17129 (https://cardiorna.eu/)[15]

[1]Translational Research in Respiratory Medicine, University Hospital Arnau de Vilanova and Santa Maria, IRBLleida, 25198 Lleida, Spain; [2]CIBER of Respiratory Diseases (CIBERES), Institute of Health Carlos III, 28029 Madrid, Spain; [3]Integrated Bed to Bench Operations (IBBO), Luxembourg Institute of Health, 1445 Strassen, Luxembourg; [4]Biogazelle, 9052 Zwijnaarde, Belgium; [5]Laboratoire National de Santé (LNS), 3555 Dudelange, Luxembourg; [6]Department of Science and Environment, Roskilde University, 4000 Roskilde, Denmark; [7]Department of Clinical Immunology, Naestved Hospital, 4700 Naestved, Denmark; [8]Minerva Foundation Institute for Medical Research, 00290 Helsinki, Finland; [9]Department of Clinical Chemistry, University of Helsinki and Helsinki University Hospital, 00029 Helsinki, Finland; [10]National Measurement Laboratory, LGC, Teddington TW11 0LY, UK; [11]Institute for Molecular Medicine Finland FIMM, Helsinki Institute for Life Science HiLIFE, University of Helsinki, 00014 Helsinki, Finland; [12]European Research Infrastructure for Translational Medicine EATRIS, 1081 HZ Amsterdam, the Netherlands; [13]Biomarkers and Therapeutic Targets Group, Ramon y Cajal Health Research Institute (IRYCIS), RedinRen, 28034 Madrid, Spain; [14]Cardiovascular Research Unit, Department of Population Health, Luxembourg Institute of Health, 1A-B rue Edison, 1445 Strassen, Luxembourg

Despite promising findings, quantitative PCR (qPCR)-based tests for RNA quantification have experienced serious limitations in their clinical application. The noticeable lack of technical standardization remains a huge obstacle in the translation of qPCR-based tests. The incorporation of qPCR-based tests into the clinic will benefit from guidelines for clinical research assay validation. This will ultimately impact the clinical management of the patient, including diagnosis, prognosis, prediction, monitoring of the therapeutic response, and evaluation of toxicity. However, clear assay validation protocols for biomarker investigation in clinical trials using molecular assays are currently lacking. Here, we will focus on the necessary steps, including sample acquisition, processing and storage, RNA purification, target selection, assay design, and experimental design, that need to be taken toward the appropriate validation of qRT-PCR assays in clinical research. These recommendations can fill the gap between research use only (RUO) and *in vitro* diagnostics (IVD). Our contribution provides a tool for basic and clinical research for the development of validated assays in the intermediate steps of biomarker research. These guidelines are based on the current understanding and consensus within the EU-CardioRNA COST Action consortium (www.cardiorna.eu). Their applicability encompasses all clinical areas.

INTRODUCTION

A literature search on biomarkers and cardiovascular diseases (CVDs) highlights the poor correlation between the efforts performed in the initial steps of the development of quantitative PCR (qPCR) assay-based biomarkers, i.e., discovery and preclinical stages and their incorporation into clinical practice. There are a number of barriers that contribute to this poor implementation. The lack of technical standardization constitutes a key limitation in the incorporation of qPCR-assay-based biomarkers into the clinic. Limitations are also linked to the absence of consensus reference values, poor harmonization of the study populations, and the barriers in collaboration between academia, physicians, and industry. For instance, despite the thousands of noncoding RNA (ncRNA)-based biomarker studies published to date, there is a paucity of potential indicators that have been successfully translated into clinical practice, mainly due to the lack of reproducibility of research findings. Kok et al.[1] nicely illustrate the situation for coronary artery disease (CAD)-associated circulating microRNA (miRNA) biomarkers based on a literature review yielding 13 miRNAs found to be up- or downregulated in more than one study, of which more than half (7 out of 13) showed a contradictory result between studies (e.g., for miR-21, two studies showed upregulation and one study showed downregulation). This lack of reproducibility has also been addressed in several publications,[2–4] with reported causes ranging from technical analytical aspects to variable patient inclusion criteria and underpowered studies to sample quality. As such, the field

https://doi.org/10.1016/j.omtm.2021.12.007.

[15]Twitter: @cost17129

Correspondence: Yvan Devaux, Cardiovascular Research Unit, Department of Population Health, Luxembourg Institute of Health, 1A-B rue Edison, 1445 Strassen, Luxembourg.
E-mail: yvan.devaux@lih.lu

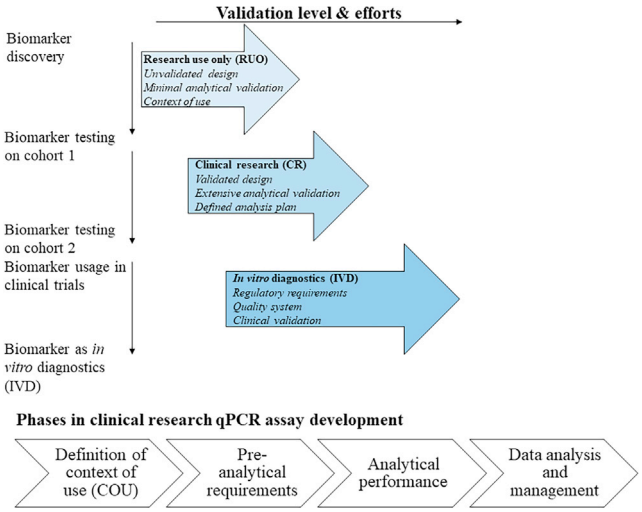

Figure 1. Workflow from research-use assays to *in vitro* diagnostic tests suitable for clinical practice

CR, clinical research; COU, context of use; IVD, *in vitro* diagnostics; LDT, laboratory-developed test, RUO: research use only.

of *in vitro* diagnostics (IVD)-grade quantitative reverse transcription PCR (qRT-PCR) assays for clinical use, initially developed in research laboratories, is still in its infancy. The incorporation of novel molecular biomarkers for clinical decision-making and patient management, i.e., diagnosis, prognosis, prediction, and monitoring of the therapeutic response or toxicity, need clear assay validation guidelines to be followed in the context of clinical research.

In this context, basic and clinical researchers often resort to the use of laboratory-developed assays with variable and undefined quality, commercial research use only (RUO) assays or, in the best-case scenarios, laboratory-developed assays validated in accordance with guidance such as minimum information for the publication of quantitative real-time PCR experiments (MIQE) guidelines.[5] The difference between such assays and certified IVD assays is significant. Laboratory-developed assays for clinical research are typically less controlled and standardized and do not need to comply with regulations, such as the European *In Vitro* Diagnostic Regulation (IVDR 2017/746). The European regulatory framework, based on the aforementioned IVDR and the Clinical Trials Regulation 2014/536, leaves a gray area relative to the status of laboratory assays that are used in the context of clinical trials. Poorly validated assays are not appropriate for large-scale clinical biomarker studies. Therefore, researchers would benefit from guidelines on the validation of what we refer to as clinical research (CR) assays, an assay type filling the gap between RUO and IVD that addresses the specific needs of researchers in the development of biomarkers. To some degree, such CR assays are similar to laboratory-developed test (LDT) assays in that they have undergone more thorough validation without reaching the status of a certified IVD assay.

Glossary

Analytical precision (or precision): closeness of two or more measurements to each other

Analytical sensitivity: the ability of a test to detect the analyte

Analytical specificity: the ability of a test to distinguish target from nontarget analytes

Analytical trueness / analytical accuracy: closeness of a measured value to the true value

Clinical research: in this article, "clinical research" encompasses clinical studies involving patients and/or healthy controls and their biomaterials, of which the objectives are related to therapeutic interventions (clinical trials), diagnostic or prognostic developments, understanding of disease mechanisms. Good Clinical Laboratory Practice (GCLP) standards typically apply to laboratory work in clinical research

Context of use: statement that describes the appropriate use of a product or test

Fit-for-purpose: a conclusion that the level of validation associated with a medical product development tool is sufficient to support its context of use

Negative predictive value: the predictive ability of a test to identify the absence of the disease in individuals with negative test results

Positive predictive value: the predictive ability of a test to identify disease in individuals with positive results

True positive rate / Sensitivity: proportion of positives that are correctly identified

True negative rate / Specificity: proportion of negatives that are correctly identified

By defining a CR level validation, researchers can more easily license out RUO assays that are affordable and easy to obtain in the early stages of biomarker research to diagnostic test manufacturers or clinical laboratory providers. This progression is visually represented in Figure 1. Here, we will focus on the necessary steps that need to be taken toward the appropriate validation of qRT-PCR workflows for CR and clinical use. Overall, the objective of this review is not to provide regulatory guidance for compliance with agency requirements. The aim is to provide supplementary practical and technical support in the specific context of qRT-PCR for which existing regulations are not always easy to apply or are unknown to researchers usually working outside of the regulated frameworks.

CONSIDERATIONS FOR BIOMARKER IDENTIFICATION, VALIDATION, AND CLINICAL USE

A biomarker is a characteristic, a measurable indicator of normal or pathologic biological processes, the responses to an exposure or intervention (including therapeutic interventions), or the risk of developing a medical condition or disease.

According to their intended use, biomarkers can be structured into several categories: susceptibility/risk, diagnostic, monitoring, prognostic, predictive, pharmacodynamics/response, and safety. Thus, with the right set of biomarkers, questions can be addressed such as how a condition will develop (prognostic), who will benefit from a treatment (predictive/stratification), will a treatment be efficacious (pharmacodynamic or surrogate) and beneficial (monitoring/response), will it be safe or toxic (safety), and how stable the health condition of a patient would be.

In general, the validation of a biomarker includes an evaluation of the analytical performance (trueness, precision, and analytical sensitivity and specificity) and the clinical performance (specificity, sensitivity, and predictive values). Analytical trueness (or analytical accuracy) refers to the closeness of a measured value to the true value, while analytical precision refers to the closeness of two or more measurements to each other and includes establishing the repeatability and reproducibility of the test. Analytical sensitivity is the ability of a test to detect the analyte (usually the minimum detectable concentration or LOD), and analytical specificity is the ability of a test to distinguish the target from nontarget analytes (in qPCR assays, the detection of a target sequence rather than other, nonspecific sequences).

Clinical performance is the ability of a test to correctly discriminate between the presence or absence of disease, and it is evaluated using measures of diagnostic accuracy.[6] The diagnostic sensitivity of a test is reflected in the true positive rate (TPR), meaning the correct identification of subjects with the disease, while the diagnostic specificity of a test is measured as the true negative rate (TNR), meaning the correct identification of subjects without disease. Positive predictive value (PPV) is the predictive ability of a test to identify disease in individuals with positive results, while negative predictive value (NPV) is the predictive ability of a test to identify the absence of disease in individuals with negative test results. Predictive values are dependent on the prevalence of disease.

The thresholds of these performance characteristics depend on the context of use (COU) and adhere to the "fit-for-purpose" (FFP) concept and must ideally be decided prior to the test. Properly defined, FFP is "a conclusion that the level (or rigor) of validation associated with a medical product development tool (assay) is sufficient to support its COU,"[7] where validation is the process of testing an assay performance, including the measures of the calibration of the instruments, the standardization of the experimental processes, the accuracy, the precision, and the reproducibility.[8]

The COU elements are laid out in the US Food and Drug Administration (FDA) and European Medicines Agency (EMA) guidelines and provide an informative and structured framework for identifying a biomarker's utility.[9,10] COU elements include (1) what aspect of the biomarker is measured and in what form, (2) what is the clinical purpose of the measurements, and (3) what is the interpretation and decision/action based on the measurements.

Biomarkers that are expected to support clinical decision-making have to be validated according to a formal qualification process, which concludes that a biomarker allows a specific interpretation and an application according to its COU in clinical product development. One example of the use of a therapeutic COU is in the allocation to specific treatment regimens. Biomarkers can influence the decision of the cessation of a patient's participation in a clinical trial, can establish a drug's proof of concept in a patient population, support clinical dose selection, and serve to enrich clinical trials for populations of interest, and can help evaluate treatment responses. This concept can be generalized to all phases of discovery research and drug development with the term FFP.

Hence, the intended use of a biomarker determines the choice of analytical methods as well as the stringency of the performance criteria of the biomarker during its validation process. An adjustment of methods with regard to new findings during the FFP validation process may be necessary and may concern the selection of preanalytical conditions. A validation phase with preliminary performance acceptance criteria, followed by larger sample sizes and off-site tests of the biomarker assay, is essential for the determination of the robustness and usability in more extended settings.

Nonclinical biomarkers (RUOs), which are the output of the FFP validation process, do not require regulatory submission. RUO biomarkers can be used as "good-enough" biomarkers internally or possibly for publication. However, full validation will be required in case the biomarker will be developed further toward clinical use, e.g., as a companion diagnostic.

For regulatory acceptance, a biomarker needs to follow and fully comply with the path of a qualification that outlines how a biomarker will be used in the clinical setting. The outcome will be a valid biomarker within its COU, one that is measured in an analytical test system with well-established performance characteristics and for which there is an established framework or body of evidence that elucidates the physiological, toxicological, pharmacological, or clinical significance of the test results.[9–11]

Controlled experimental environments, which include experimental designs with predefined acceptance criteria, preanalytical requirements, qualified equipment, trained operators, analytical performance, and data stewardship, are prerequisites for reliable and thereby meaningful measurement of biomarkers.

GUIDELINES FOR VALIDATION OF CR-GRADE qPCR ASSAYS

Many references on validation requirements exist, ranging from academic guidelines and international standards such as ICH and ISO to national regulatory requirements such as the 21 CFR Part 820 (US) and IVDR 2017/746 (EU).[8,12,13] A practical challenge for the use of many of these standards is the fact that they have not been tailored to the specifics of PCR-based tests. ISO 20395:2019 (Biotechnology — Requirements for evaluating the performance of quantification methods for nucleic acid target sequences — qPCR and digital PCR [dPCR]), which was recently published, addresses specific underpinning requirements. Inspired by these standards, we present guidelines for the practical analytical validation of CR-grade qPCR assays in clinical samples.

It is worth noting that the concept of a guideline explicitly allows researchers to deviate from them as a function of the specific needs and characteristics of their tests. A few examples where a risk analysis could alter the validation of qPCR assays include the following: (1) the specificity of an assay is typically a crucial assay property (for reference genes, this requirement may be relaxed since their measurements are required to represent the total amount of RNA, not necessarily the amount of a specific gene); (2) the limit of detection (LOD) is an important property for assays testing weakly expressed genes or for tests optimized to work with minimal input quantities (for tests where the analyte is highly abundant, defining the LOD addresses specific points in standards and regulatory requirements without significantly contributing to the quality of the assay); or (3) the capacity of an assay to accurately quantify small expression differences requires a thorough validation when used for biomarkers that rely on such small expression differences. On the other hand, less extensive validation experiments may be sufficient for tests that rely on the detection of large expression differences.

Method development for RUO

Even when aiming for a higher level of analytical assay validation, it is useful to start with RUO-level validation. Such validation is fast and easy and allows for the elimination of inferior designs before proceeding to the more extensive CR (or ultimately IVD)-level validation.

Prior to the initiation of the analytical validation, the qPCR master mix components and the thermocycler and cycling conditions must be selected, optimized, and then fixed. Frequently, these decisions are implicit by relying on standard procedures used within the laboratory, but one needs to be aware that the obtained validation results only apply to that setting. As there are considerable differences between qPCR master mixes with regard to assay performance, polymerase reliability, and optimized conditions, defining the conditions is crucial. Importantly, validated results only apply to the settings used, meaning that revalidation is required whenever the setup changes.

RUO-level analytical validation (for genes of interest and reference genes) should at least cover the following aspects: (1) correct amplification on a positive control sample (for example, Cq values below 28), (2) lack of amplification on a negative control sample, (3) evidence of RT and amplification efficiency, and (4) specificity of an amplified product (melting temperature and/or amplicon length analysis). An experimental workflow providing data for validation could be performed as follows:

1 .Samples, controls, and standards preparation
- Select and prepare a positive control
  i. In general, RNA, rather than cDNA, is preferred as a positive control. The RNA may be derived from a positive sample or generated by *in vitro* transcription from a vector or synthetic sequence.
  ii. cDNA sample. Ideally derived from an RNA sample that has undergone the same process (collection, extraction, RT) as clinical samples and that is known to express the gene of interest. Alternatively, this could also be a commercial sample available in large quantities and with sufficient and similar RNA quality to the intended test samples. Clinical biobanks, under ISO certification, may be a source of samples if no commercial controls are available.
  iii. A cloned and sequence-verified DNA fragment containing the transcript of interest.
  iv. A synthetic DNA fragment containing the transcript sequence of interest.
- Select and prepare negative controls
  i. No-template control testing for contamination and primer-dimer formation (e.g., water or carrier RNA).
  ii. Genomic DNA testing for gDNA coamplification. A RT reaction without reverse transcriptase added will also test for gDNA contamination of the RNA eluate.
  iii. "Extraction blank" to control for contamination at the extraction stage.
  iv. Optional: cDNA from other species to test for cross-species reactivity (important in preclinical studies where human genes are used in a hybrid context).
  v. Optional: cDNA from a specific RNA sample known not to contain the transcript or variant of interest (this is the ideal negative control because of its complex matrix similar to real samples).
- Create a dilution series to establish the linearity, PCR efficiency, and analytical range of the assay (Figure 2)
  i. Sample type options
    ◦ cDNA, if the gene of interest is highly expressed.
    ◦ Plasmid or long double-stranded synthetic DNA fragments.
    ◦ Short single-stranded chemically synthesized templates, e.g., 60-mer oligonucleotides of the first and last 30 nucleotides of the amplicon. This solution can be applied also for genes with very low or rare expression. Of note, for probe-based assays, the oligonucleotide (oligo) should also contain the probe binding site.
    ◦ *In-vitro*-transcribed synthetic RNA fragments representing the entire amplicon.

Considerations for sample material for qPCR dilution series
Purpose: Determine qPCR efficiency and assay linearity

cDNA based on endogenously expressed RNA

Pros:
- ✓Simple workflow: Establish a pool of representative RNA samples
- ✓Composition of samples is very close to, and representative of, actual experimental samples with regard to potential competing transcripts/inhibitors/contaminants

Cons:
- ✓Low-abundance RNAs may not allow suffient dilution to estimate PCR efficiency and assay linearity
- ✓It may be difficult to obtain enough sample material to also use as a calibrator in later assay development. Material may be challenging to standardize.

cDNA based on *in-vitro*-transcribed RNA

Pros:
- ✓*In-vitro*-transcription may yield high amounts of standardized material
- ✓Contains the same buffer system as the cDNA (the *in-vitro*-transcribed RNA is reverse transcribed using the same protocol as actual samples).
- ✓Works well with low-abundance targets

Cons:
- ✓Need to clone target amplicon or cDNA species into a suitable T7/SP6 promoter vector
- ✓Due diligence must be taken to prevent sample contamination with control material as the concentration is often very high compared with samples

Double-stranded DNA/plasmid DNA

Pros:
- ✓High amounts of standardized material
- ✓Works well with low-abundance targets

Cons:
- ✓Buffer and template/contaminant composition is less similar to experimental samples
- ✓Need to clone target amplicon or cDNA species into a vector
- ✓Due diligence must be taken to prevent sample contamination with control material as the concentration is often very high compared with samples

Single-stranded chemically synthesized DNA oligos

Pros:
- ✓High amounts of standardized material
- ✓Works well with low-abundance targets

Cons:
- ✓Buffer and template/contaminant composition less similar to experimental samples
- ✓Cost of synthesizing long oligos
- ✓Due diligence must be taken to prevent sample contamination with control material as the concentration is often very high compared with samples

Figure 2. Considerations for sample material for qPCR dilution series

- RNA from cell lines, patient samples, and/or reference materials.

The last option is applicable to one-step qRT-PCR.

ii. Six-point (or more) 10-fold dilution series, ideally spanning the 10^1 to 10^6 copies/qPCR reaction range. If the input quantities for this dilution series are well known (because of the use of qualified reference samples or because of calibration by means of dPCR), the dilution series could also be used for absolute quantification.

2 .Tests and analyses

- In total, 21–33 reactions/assay.
- Perform triplicate qPCR reactions for all samples.
- Amplicon size analysis should be performed (agarose gel or microfluidic electrophoresis). Positive control samples should yield a single sized product of correct length. For negative controls with a signal, it may support troubleshooting by differentiating between primer-dimer formation and template contamination.
- Evaluate amplification plots for correct amplification in positive control samples.
- For DNA-binding dye assays (e.g., SYBR Green):
  i. Evaluate melting curves.
  ii. Positive control samples should yield a characteristic Tm peak. Different tools enable the prediction of the Tm peak (or multiple peaks for some amplicons with nonuniform GC% distributions) as a reference value.
  iii. Negative samples should display no primer-dimer peak (typically broader peak at low Tm) nor a template Tm peak.
- Evaluate negative control samples. Ideally, no amplification is observed. In cases where the sequence of interest is very highly expressed, a low level of false positive signals (high Cq value) might be tolerated.
- Analyze the dilution series for:
  i. Linearity in a standard curve. Deviations from linearity may be observed at the ends of the dilution series, e.g., because of plateauing due to contamination or primer-dimer formation or because of inhibition due to the use of too much cDNA. The testable range is restricted to the linear range of the dilution series.
  ii. Amplification efficiency. Ideally in the 90%–110% range. Efficiencies of 80%–90% or 110%–120% are suboptimal but acceptable for some assays, e.g., when the effect size (difference) is sufficiently large and such small technical imperfections do not interfere with their separation.

CR-level validation

A CR assay consists of the entire workflow from RNA template measurement to data analysis. The standardized treatment of samples at each step is imperative to ensure biomarker performance.

Reverse transcription

During reverse transcription, the complementary DNA template is synthesized. Ideally, this reaction should generate a 1-to-1 DNA complementary to the RNA template, but this is very rarely achievable. Depending on the type and quality of RNA investigated, a reverse transcription priming method should be selected. The RT reaction may use random primers (6–9 mers), oligo dT primers, a combination of both of the aforementioned primers, or target primers, corresponding to those used in the qPCR reaction. Validation experiments should include the test of saturation by an RNA template, which is performed as multiple cDNA synthesis reactions with increasing amounts of purified RNA, including an RNA spike-in (an external control) followed by qPCR for the intended targets. Here, there should be linearity between the Cq levels of the targets and the input amount of RNA within a range corresponding to (patho)physiological levels of the RNA template. Alternatively, increasing amounts of synthetic target RNA can be added to multiple reverse transcription reactions of one specific RNA sample. Similarly, to test for reaction inhibition during cDNA synthesis, we also recommend adding a synthetic spike-in RNA molecule during reverse transcription that is different from the one used for spike-in during RNA isolation. The amount of RNA spike-in should generally be in the linear dynamic range of the assay, which would often be in the attomol range of added

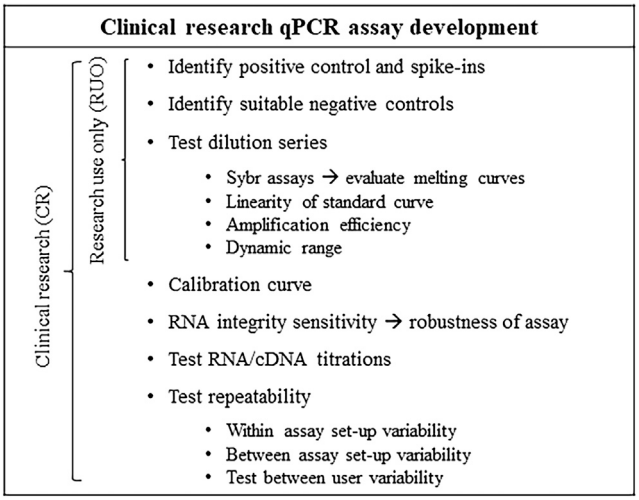

Figure 3. Clinical research qPCR assay development

spike-in. This is fundamental to correct for differences in the RT reaction efficiency.

Target selection

Different approaches lead to the identification of target sequences. Target selection is beyond the scope of these guidelines but depends on the following questions: is one interested in the gene with all its isoforms or in a subset of transcripts encoding a particular protein? Does the assay have to detect a specific allele or fusion event? Are there any regions to be avoided for assay design because of homology with mouse sequences that would interfere with analysis of murine xenografts?

Reference genes

Importantly, most qPCR tests measuring gene expression levels rely on reference genes for data normalization. Assay validation is thus not limited to the assays measuring the target of interest but must include assays for selected reference genes as well. Since the quality of the final test results strongly depends on the quality of normalization by means of the selected reference genes, it is important to ensure that multiple and stably expressed reference genes are selected. A pilot study evaluating eight candidate reference genes in 12 representative samples and analyzed by a tool such as geNorm[14] or NormFinder[15] can provide the needed assurance that the reference genes are unaffected by the experimental conditions and are stable in the population and sample type of interest. Additionally, for small RNAs determined in body fluids, a set of reference small RNAs has already been suggested.[16] Nevertheless, a critical evaluation of these reference small RNAs is fundamental for each experimental condition and study population.

Assay design

To increase the success rate of analytical assay validation, the basics of a design by tools such as Primer3[17] should be complemented with additional bioinformatic analyses testing for the specificity, transcript coverage, and absence of secondary structures and common single-nucleotide polymorphisms (SNPs) in primer or probe annealing regions. Such analyses may be integrated in assay design by programs such as primerXL,[18] performed independently with online tools such as BiSearch (for specificity)[19] and UNAfold (secondary structures),[20] or performed manually in genome browsers (transcript coverage and SNP overlap).

Despite the benefit of such *in silico* examinations, they cannot provide a full guarantee of optimal performance when put in practice in the lab. Therefore, multiple designs can be fed into downstream analytical validation, allowing for the selection of the best performing assay and increasing the likelihood of identifying at least one assay meeting all requirements.

Experimental design for CR assay validation of qPCR-based tests

We provide an example of a design that may be used for most gene expression tests (Figure 3). This design may, however, have to be modified to meet the requirements of certain specific analytical COUs, e.g., when dealing with transcript or allele-specific assays, when working in a multispecies context, or when aiming for multiplexing. The proposed number of biological replicates may also be increased, e.g., for biomarkers with a small effect size.

The researcher also needs to predefine the assay performance acceptance criteria. These might relate to parts of the method (for instance, the qPCR amplification efficiency) but should always include criteria for the full workflow starting from the matrix to the end result. The latter are typical test-related criteria such as accuracy, precision, specificity, sensitivity, LOD, and limit of quantification (LOQ). The results of this validation have to be properly documented, e.g., electronically.

Due to the lack of suitable reference samples, the accuracy of normalized gene expression levels is difficult to determine. Inspired by the mixing experiments proposed in the microarray quality control (MAQC) studies but enhanced to ensure the testability of differential expression for any gene, we propose an approach relying on known mixtures of *in-vitro*-transcribed RNAs to support the accuracy analysis of qRT-PCR tests. To make this experiment representative of the actual analysis applying reference gene normalization, it includes both transcripts for the gene(s) of interest (GOI) and for the reference gene(s).

To systematically assess the robustness of the test for RNA integrity, we propose the evaluation of a series of RNA samples with varying degrees of artificial degradation. This degradation can be achieved via heating, sonication, UV radiation, or incubation with ribonucleases.[21,22] If samples are known to systematically yield high-quality RNA, and its integrity is systematically tested, one may consider skipping this robustness analysis. Evaluation of the quality of RNA can be achieved by spectrophotometry and by microfluidic electrophoretic methods based on the RNA integrity number (RIN) or an equivalent metric. It is more difficult to evaluate the quality of ncRNA, such as

miRNAs, but this can be done either by verifying the presence of ubiquitous ncRNA molecules (e.g., RNU-24, miR-16, miR-221) or by measuring quality scores that have been described in specific COUs.[23,24]

Consensus experimental design for a standard gene expression assay is presented hereafter (for absolute and relative quantification). The recommended acceptance criteria for precision and accuracy depends on the intended purpose of the assay:

1 .Preparation 1: Establish a calibration curve
- Design and order two (or more) double-stranded synthetic DNA templates containing a promotor for *in vitro* transcription (e.g., T7 or SP6) followed by a series of concatenated amplicon sequences. To support quality control on the dilution series, each such template may also include the amplicon sequence of a control assay with a proven performance.
  i. GOI template containing amplicon sequences for the GOIs: promotor – GOI1 – GOI2 – … – GOIn – control. If a large number of assays for different GOIs are to be tested, one may consider making different GOI templates, with each containing a subset of GOI amplicon sequences.
  ii. REF template containing amplicon sequences for the reference genes: promotor – REF1 – REF2 – … – REFn – control.
- Optional: determine the copy number of the synthetic templates by means of dPCR using the control assay. If dPCR is not available, one has to rely on the quantities described by the oligo producer to estimate the template copy number.
- Create a dilution series for both GOI and REF templates spanning the 5×10^6 to 5 copies/PCR reaction range.
- Confirm the correctness of the dilution series by testing the control assay. With qPCR, the correctness of fold dilutions can be verified. If dPCR is available, the accuracy of the dilution points can also be assessed. If these deviate more than 2-fold (or any fold difference that is relevant for the intended purpose), a new dilution series should be made.

2 .Preparation 2: Establish an RNA integrity series
- Heat-treat an RNA sample for different times (75°C for 1–10 min/μg RNA). Ideally, the same representative RNA sample (same matrix and extraction method) should be used. If such RNA is already too degraded to be used as a starting point for the RNA integrity series or the available material is too limiting for these experiments, one may have to rely on commercial samples such as the MAQC RNA.
- Assess RNA integrity by means of microfluidic electrophoresis (see above). The process of artificial RNA degradation and integrity assessment may have to be iterated until proper timings have been found to generate the targeted RNA integrity range. Create cDNA for the selected RNA samples.
- For DNA analysis, an equivalent test can be set up.

3 .Preparation 3: Establish an RNA/cDNA titration series
- *In vitro* transcription of the two templates containing the GOI and REF, respectively (two separate reactions).
- Create two different mixtures of GOI and REF RNA to mimic RNA samples with different expression levels.
  i. A: mixture containing equal amounts of GOI and REF RNA.
  ii. B: mixture containing the same amount of REF RNA as sample A but only 1/1,000 of GOI RNA.
- Create RNA mixes with variable fractions of GOIs (mimicking samples with variable expression levels):
  i. C: 90% A + 10% B
  ii. D: 75% A + 25% B
  iii. E: 50% A + 50% B
  iv. F: 25% A + 75% B
  v. G: 10% A + 90% B
- Spike mock RNA with synthetic RNA mixes A–G. The mock RNA provides for a more natural, complex RNA background. Any RNA sample void of amplifiable sequences may be suitable. RNA from bacteriophage MS2 is often a good candidate mock RNA sample.
- Synthesize cDNA (by performing RT) for all spiked RNA samples and for unspiked mock RNA (sample H).
- For gDNA analyses, an equivalent test, not needing *in vitro* transcription and cDNA synthesis, can be set up.

4 .Preparation 4: Repeat extraction and cDNA synthesis
- Four representative clinical samples with sufficient material should be selected for at least 2 extractions. If material for individual samples is limiting, one may consider homogeneous pooling and mixing of samples to obtain samples with sufficient material to support repeat extraction.
- Perform repeat extraction, reflecting the different sources of variation (day of extraction, extraction kit lot number, person executing the extraction).
- Perform cDNA synthesis independently for the two sets of 4 RNA extracts.

5 .Test 1: Perform qPCR measurements to assess amplification efficiency, primer-dimer formation, robustness towards RNA degradation, qPCR repeatability, and specificity on cDNA samples.
- In total, 48–96 reactions/assay.
- Perform qPCR reactions in 3–6 replicates for all candidate assays on 16 samples:
  i. Dilution series of synthetic DNA templates (see Preparation 1).
  ii. No-template control (NTC) with carrier RNA.
  iii. 5 cDNA samples from RNA integrity series.
  iv. 3 cDNA samples from representative clinical samples.
- Assess specificity on amplicons from the 3 representative cDNA samples using size and, for assays using intercalating dyes, melt curve analysis.
- Analyze qPCR data:
  i. Determine slope, intercept, linear dynamic range, coefficient of variation (r^2), and amplification efficiency from dilution series data.
  ii. Determine qPCR repeatability for cDNA from representative clinical samples and for the different samples of the dilution series. The latter may reveal concentration-dependent

Table 1. Relative input quantities

Sample	Expected expression level
A	1.00
C	0.9001
D	0.75025
E	0.50005
F	0.25075
G	0.10009
B	0.001

repeatability, typically showing increasing variability as concentrations approach the detection limit.

iii. Assessment of the absence or degree of primer-dimer formation on the NTC sample.

iv. Assess the robustness of the assay against RNA degradation. A robust assay would show identical normalized relative quantity (NRQ) values for artificially degraded RNA as for intact, undegraded RNA. Lower NRQ values for degraded samples would reveal an impact of certain degrees of RNA degradation.

6 .Test 2: Perform qPCR measurements to assess the trueness and repeatability of the assay (NRQ values)

- In total, 48 reactions/assay.
- qPCR setup.
  i. On 12 samples: an NTC, the 7 cDNA samples (A–G) derived from the RNA/cDNA titration series (see Preparation 3), and 4 cDNA samples derived from the same repeat extraction and reverse transcription.
  ii. Using assays for both the reference genes and a set of GOIs.
  iii. Triplicate qPCR reactions (or other replicate number as will be used in later testing—more replicates may improve data quality).
  iv. Repeat the analysis of the 12 samples twice within the same run. One set containing the 4 cDNA samples of one extraction and reverse transcription round, and the other set containing those of repeat extraction and reverse transcription.
- Data analysis
  i. Calculate normalized expression levels for the GOIs by normalizing their relative quantities with the geometric mean of the relative quantities of the selected reference genes.[25]
  ii. Determine trueness by comparing the observed normalized expression level against the known relative input quantities (see Table 1).
  iii. Determine repeatability by comparing the normalized expression levels for the two repeats within the same run.

7 .Test 3: Repeat qPCR measurements to assess the between-run precision of the assay

- In total, 48 reactions/assay.
- Repeat qPCR measurements of Test 2.
  i. Prepared by a different person.
  ii. On a different day.
  iii. Measured on a different qPCR instrument.
- Determine between-run precision of the assay by comparing the NRQ values for the two repeats between the runs. The results for samples A–G only reflect the repeatability of the qPCR measurements, including pipetting errors. The results for the other 4 samples also reflect the variability of other parts of the workflow (RNA extraction and cDNA synthesis).

8 .Validation of a run.

Quality control steps to monitor relative quantification assay performance and accept or reject a run include the following: (1) cDNA prepared from a reference RNA sample should show Cq < 30; (2) reverse transcription NTC sample should give indeterminate Cq or Cq at least 5 Cq units higher than the highest Cq value of the positive test samples used in the assay; (3) NTC sample should give indeterminate Cq or Cq at least 5 Cq units higher than the highest Cq value of the positive test samples used in the assay; (4) if amplification efficiencies are remeasured, they should fall within their predefined acceptance range (typically 90%–110%); or (5) if multiple reference genes are used to normalize data, their stability measures should fall within their predefined acceptance range (typically geNorm M value < 0.5).[14,25]

Quality control steps to monitor absolute quantification assay performance and accept or reject a run include the following: (1) NTC sample gives an undetermined result or a result below the lowest point of the calibration curve, (2) the calibration curve R^2 >99%, (3) the calibration curve slope is between −2.9 and −3.8, or (4) there are at least four consecutive data points in the calibration curve (defining the "validated range" for the run).

DISCUSSION

The guidelines proposed above, which are summarized in Figure 4, are not mandatory and are proposed to assist researchers in validating their assays before implementation in CR and clinical trials. They are not intended to impose formal validation requirements but rather to describe the consensus obtained within the European CardioRNA consortium (CardioRNA COST Action CA17129) on a relevant set of validation experiments. These guidelines are focused on the analytical validation of singleplex gene expression assays. Several aspects are explicitly not covered: multiplex assays, data analysis, cross-species amplification (xenograft, infectious), preamplification, genotyping or allele-specific PCRs, or dPCR. These assays may require other approaches that are inspired by these guidelines or are completely tailor made.

We envisage reviewing these guidelines in 3 years based on the feedback received from the CR community, the practical experience gained in using these guidelines, the new consensus in the scientific community, and new technological developments and approaches.

CONCLUSIONS, REMARKS, AND PERSPECTIVES

The increase in RNA-focused research in the last decade has led to great advances in the general and specific knowledge of the

Checklist for reference procedure

Sample acquisition & storage:

- Sample species acquired uniformly
- Samples stored at –80°C

RNA purification

- Same RNA isolation method for all samples
- Spike-ins

RT-PCR assay

- Positive and negative controls
- Choice of normalizer justified
- Assay linearity and efficiency
- Sensitivity (limit of detection/quantification)
- Specificity of amplification

Dataanalysis

- Predefined data analysis plan
- Consistency
- Predefined boundaries of sample/result acceptance

Figure 4. Checklist for reference procedure

transcriptome and pathophysiological mechanisms of diseases, leading to the identification of new biomarkers that are useful in clinical practice. Despite the promising findings, qPCR-based tests for RNA quantification have experienced serious limitations in their direct clinical application. The development and commercialization of novel qRT-PCR-based tools is a laborious process, and successful assay validation requires substantial resources. Ultimately, the establishment and application of evidence-based recommendations for CR assays may reduce the time and cost of obtaining new assays from the research laboratory to clinical practice and the market.

These recommendations can fill the gap between RUO and IVD. They are the output of a collective effort of the EU-CardioRNA consortium with collaboration and endorsement by the European Research Infrastructure for Translational Medicine (EATRIS) (https://www.EATRIS.eu) and the Biobanking and BioMolecular Resources Research Infrastructure (BBMRI) (www.BBMRI.eu), with both being part of the EU-AMRI alliance (European Alliance of Medical Research Infrastructures), whose vision is to alleviate the attrition rate of biomarkers early in their development to assure accuracy and to help save costs, time, and expectancies of patients and clinicians. We are confident that application of these guidelines will result in more effective biomarkers development for many diseases but that are, above all, useful in clinical practice.

ACKNOWLEDGMENTS

This manuscript is based upon work from EU-CardioRNA COST Action CA17129 (www.cardiorna.eu) supported by COST (European Cooperation in Science and Technology). D.d.G.-C. wants particularly to acknowledge Marta Molinero and Jo Vandesompele for their technical support. D.d.G.-C. has received financial support from the Instituto de Salud Carlos III (Miguel Servet 2020: CP20/00041), which is co-funded by the European Social Fund (ESF)/"Investing in your future." This work is supported by the Instituto de Salud Carlos III (PI20/00577), which is co-funded by European Regional Development Fund (ERDF)/"A way to make Europe." CIBERES is an initiative of the Instituto de Salud Carlos III. Y.D. is funded by the EU Horizon 2020 project COVIRNA (grant agreement 101016072), the National Research Fund (grants C14/BM/8225223, C17/BM/11613033, and COVID-19/2020-1/14719577/miRCOVID), the Ministry of Higher Education and Research, and the Heart Foundation-Daniel Wagner of Luxembourg. P.L. is funded by the Finnish Cultural Foundation, The Finnish Foundation for Cardiovascular Research, The Finnish Society of Clinical Chemistry, and the Finnish Foundation for Laboratory Medicine.

AUTHOR CONTRIBUTIONS

D.d.G.-C., M.M., J.H., F.B., N.L.S.F., A.S., L.T.D., and P.L. designed the article and wrote the manuscript. C.F., M.L.G.B., and Y.D. revised the manuscript. All authors read and approved the final manuscript.

DECLARATION OF INTERESTS

D.d.G.-C. holds a patent on miRNAs as biomarkers. Y.D. holds patents related to diagnostic and therapeutic applications of RNAs. The other authors declare no competing interests.

REFERENCES

1. Kok, M.G.M., de Ronde, M.W.J., Moerland, P.D., Ruijter, J.M., Creemers, E.E., and Pinto-Sietsma, S.J. (2018). Small sample sizes in high-throughput miRNA screens: a common pitfall for the identification of miRNA biomarkers. Biomol. Detect. Quant. *15*, 1–5.
2. Faraldi, M., Gomarasca, M., Sansoni, V., Perego, S., Banfi, G., and Lombardi, G. (2019). Normalization strategies differently affect circulating miRNA profile associated with the training status. Sci. Rep. *9*, 1584.
3. Godoy, P.M., Barczak, A.J., DeHoff, P., Srinivasan, S., Etheridge, A., Galas, D., Das, S., Erle, D.J., and Laurent, L.C. (2019). Comparison of reproducibility, accuracy, sensitivity, and specificity of mirna quantification platforms. Cell Rep. *29*, 4212–4222.e5.
4. Leidner, R.S., Li, L., and Thompson, C.L. (2013). Dampening enthusiasm for circulating microRNA in breast cancer. PLoS One *8*, e57841.
5. Bustin, S.A., Benes, V., Garson, J.A., Hellemans, J., Huggett, J., Kubista, M., Mueller, R., Nolan, T., Pfaffl, M.W., Shipley, G.L., et al. (2009). The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. Clin. Chem. *55*, 611–622.
6. Šimundić, A.-M. (2009). Measures of diagnostic accuracy: basic definitions. EJIFCC *19*, 203–211.
7. Updated, L. (2021). BEST (Biomarkers, EndpointS, and Other Tools) Resource (FDA-NIH Biomarker Working Group).
8. Lee, J.W., Devanarayan, V., Barrett, Y.C., Weiner, R., Allinson, J., Fountain, S., Keller, S., Weinryb, I., Green, M., Duan, L., et al. (2006). Fit-for-purpose method development and validation for successful biomarker measurement. Pharm. Res. *23*, 312–328.
9. Agency, E.M. (2020). Qualification of Novel Methodologies for Drug Development: Guidance to Applicants. 31 (January 2012).
10. FDA (2018). Context of Use, ttps://www.fda.gov/drugs/biomarker-qualification-program/context-use.

11. FDA (2018). Biomarker Qualification: Evidentiary Framework, https://www.fda.gov/regulatory-information/search-fda-guidance-documents/biomarker-qualification-evidentiary-framework.

12. Narrandes, S., and Xu, W. (2018). Gene expression detection assay for cancer clinical use. J. Cancer *9*, 2249–2265.

13. Cummings, J., Ward, T.H., Greystoke, A., Ranson, M., and Dive, C. (2008). Biomarker method validation in anticancer drug development. Br. J. Pharmacol. *153*, 646–656.

14. Vandesompele, J., de Preter, K., Pattyn, F., Poppe, B., van Roy, N., de Paepe, A., and Speleman, F. (2002). Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol. *3*, 1.

15. Andersen, C.L., Jensen, J.L., and Ørntoft, T.F. (2004). 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. *64*, 5245–5250.

16. Dakterzada, F., Targa, A., Benítez, I.D., Romero-ElKhayat, L., de Gonzalo-Calvo, D., Torres, G., Moncusí-Moix, A., Huerto, R., Sánchez-de-la-Torre, M., Barbé, F., et al. (2020). Identification and validation of endogenous control miRNAs in plasma samples for normalization of qPCR data for Alzheimer's disease. Alzheimer's Res. Ther. *12*, 163.

17. Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M., and Rozen, S.G. (2012). Primer3–new capabilities and interfaces. Nucleic Acids Res. *40*, e115.

18. Lefever, S., Pattyn, F., De Wilde, B., Coppieters, F., De Keulenaer, S., Hellemans, J., and Vandesompele, J. (2017). High-throughput PCR assay design for targeted resequencing using primerXL. BMC Bioinformatics *18*, 400.

19. Tusnády, G.E., Simon, I., Váradi, A., and Arányi, T. (2005). BiSearch: primer-design and search tool for PCR on bisulfite-treated genomes. Nucleic Acids Res. *33*, e9.

20. Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. *31*, 3406–3415.

21. Unger, C., Kofanova, O., Sokolowska, K., Lehmann, D., and Betsou, F. (2015). Ultraviolet C radiation influences the robustness of RNA integrity measurement. Electrophoresis *36*, 2072–2081.

22. Mathay, C., Yan, W., Chuaqui, R., Skubitz, A.P.N., Jeon, J.P., Fall, N., Betsou, F., and Barnes, M. (2012). Short-term stability study of RNA at room temperature. Biopreserv. Biobank. *10*, 532–542.

23. Ammerlaan, W., Trouet, J., Sachs, M.C., Guan, P., Carithers, L., Lambert, P., Frasquilho, S., Antunes, L., Kofanova, O., Rohrer, D., et al. (2018). Small nucleolar RNA score: an assay to detect formalin-overfixed tissue. Biopreserv. Biobank. *16*, 467–476.

24. Saliminejad, K., Khorram Khorshid, H.R., Soleymani Fard, S., and Ghaffari, S.H. (2019). An overview of microRNAs: biology, functions, therapeutics, and analysis methods. J. Cell Physiol. *234*, 5451–5465.

25. Hellemans, J., Mortier, G., de Paepe, A., Speleman, F., and Vandesompele, J. (2007). qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. Genome Biol. *8*, 1.