Correction of RT–qPCR data for genomic DNA-derived signals with ValidPrime

Henrik Laurell1,*, Jason S. Iacovoni1, Anne Abot1, David Svec2,3, Jean-José Maoret1,4, Jean-François Arnal1,5 and Mikael Kubista2,3

1Inserm/Université Paul Sabatier UMR1048, Institut des Maladies Métaboliques et Cardiovasculaires (I2MC), BP84225, 31432 Toulouse cedex 4, France, 2Laboratory of Gene Expression, Institute of Biotechnology, Academy of Sciences of the Czech Republic, Prague, 3TATAA Biocenter AB, Göteborg, Sweden, 4Plateforme GeT (Génome et Transcritpore) du Génopôle Toulouse, Toulouse, France and 5Faculté de Médecine, Université de Toulouse III and CHU de Toulouse

Received September 17, 2011; Revised November 21, 2011; Accepted December 5, 2011

ABSTRACT
Genomic DNA (gDNA) contamination is an inherent problem during RNA purification that can lead to non-specific amplification and aberrant results in reverse transcription quantitative PCR (RT–qPCR). Currently, there is no alternative to RT(−/C0) controls to evaluate the impact of the gDNA background on RT–PCR data. We propose a novel method (ValidPrime) that is more accurate than traditional RT(−/C0) controls to test qPCR assays with respect to their sensitivity toward gDNA. ValidPrime measures the gDNA contribution using an optimized gDNA-specific ValidPrime assay (VPA) and gDNA reference sample(s). The VPA, targeting a non-transcribed locus, is used to measure the gDNA contents in RT(+) samples and the gDNA reference is used to normalize for GOI-specific differences in gDNA sensitivity. We demonstrate that the RNA-derived component of the signal can be accurately estimated and deduced from the total signal. ValidPrime corrects with high precision for both exogenous (spiked) and endogenous gDNA, contributing ~60% of the total signal, whereas substantially reducing the number of required qPCR control reactions. In conclusion, ValidPrime offers a cost-efficient alternative to RT(−) controls and accurately corrects for signals derived from gDNA in RT–qPCR.

INTRODUCTION
Accurate gene expression analysis by reverse transcription (RT) quantitative PCR (qPCR) requires assays with high specificity for the target cDNA/reference gene, collectively referred to herein as the Gene-Of-Interest (GOI). It is important to have negligible signal contribution from experimental artifacts, such as primer–dimers and contaminating genomic DNA (gDNA). Traditionally, primer–dimer formation is tested using a ‘no template control’ (NTC) and gDNA contamination levels are measured with RT(−) controls [which differ from regular RT(+) reactions in that no reverse transcriptase is added]. Contamination of gDNA is an inherent problem during RNA purification due to the similar physicochemical properties of RNA and DNA. Since gDNA contamination levels are frequently not uniform between samples (1) and the sensitivity toward gDNA differs greatly between GOI assays, RT(−) controls are needed for each sample/assay pair, which substantially adds to the cost and labor in RT–qPCR profiling studies. A difference of at least five quantification cycles (Cq) between RT(+) and RT(−) reactions indicates that <3% of the total signal originates from gDNA, and is commonly used as limit to ensure accurate estimation of GOI expression. Smaller differences typically call for DNase treatment of samples.

The accuracy of gDNA background estimation, as measured with RT(−) reactions, is compromised due to the fact that GOI assays, designed to amplify target transcripts, are used even though they are not optimized for gDNA amplification. Furthermore, intrinsic characteristics of RT(−)-qPCRs that influence the result of the correction, such as amplification efficiencies, are difficult to assess. In addition, as proposed theoretically (2) and shown experimentally (3,4), a low initial number of target molecules leads to a large variability between replicates, mainly due to stochastic effects. All together, this explains the low reproducibility frequently observed in RT(−) reactions.

The qPCR assays can be either gDNA sensitive or insensitive. Whereas qPCR assays can be designed to be

*To whom correspondence should be addressed. Tel: +33 561323128; Fax: +33 561322141; Email: henrik.laurell@inserm.fr

© The Author(s) 2012. Published by Oxford University Press. This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/3.0), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
gDNA insensitive, such as those designed to target exons flanking a long intron or with primers that cross exon–exon junctions, qPCR assays for single-exon genes will readily amplify contaminating gDNA. The gDNA background signal is even further amplified in the presence of multiple genomic copies or pseudogenes. The latter are particularly troublesome since they may originate from retrotransposons without introns that are amplified even with intron-spanning assays. Thus, there exists both variation in the degree of contamination between samples and large differences between assays in terms of their sensitivity to gDNA. Therefore, general methods of controlling and correcting for gDNA contamination are essential for accurate measurements of gene expression.

As an alternative to RT(−) reactions, we have developed a procedure that determines the impact of the gDNA contamination on the measured signal much more accurately and allows validation of qPCR primers with respect to their sensitivity toward gDNA. We show in proof-of-principle experiments that efficient background correction can be performed with gDNA contamination representing ~60% of the total signal.

MATERIALS AND METHODS

Samples
All samples were from mouse (C57Bl/6J) tissues (kidney, liver, adipose tissue, uterus, peritoneal macrophages). All experimental procedures involving animals were performed in accordance with the principles and guidelines established by the National Institute of Medical Research (Inserm) and were approved by the local Animal Care and Use Committee. Prior to sampling, mice were anesthetized by intraperitoneal injection of ketamine (100 mg kg \(^{-1}\)) and xylazine (10 mg kg \(^{-1}\)). Tissues were snap-frozen in liquid nitrogen and stored at −80°C. Isolation of peritoneal macrophages has been described elsewhere (5). Macrophages were in some cases treated with 20 ng/ml LPS \textit{ex vivo} for 4 h prior to RNA extraction.

DNA extraction
C57Bl/6J mouse gDNA was extracted from whole blood using the PerfectPure DNA Blood Cell Kit, according to the recommended protocol (5′PRIME GmbH, Hamburg, Germany). Good results were also obtained with gDNA purified from mouse tails by phenol/chloroform extraction after Proteinase K digestion (6). The DNA concentration was determined spectroscopically (NanoDrop).

RNA extraction
Total RNA was extracted using a double purification protocol. Briefly, Trireagent (Sigma-Aldrich, St Louis, MI, USA) was added to the frozen tissue sample, which was homogenized in a Precellys 24 homogenizer (Bertin Technologies, France). After the extraction step, the supernatant was gently mixed with 1 Vol 70% ethanol and applied on a total RNA miniprep Genelute column, where it was washed and eluted following the instructions from the manufacturer (Sigma-Aldrich). The integrity and quality of the RNA was tested by capillary micro-electrophoresis [MultiNA (Shimadzu) or Experion (BioRad)] and spectroscopically (NanoDrop). A fraction of the RNA was DNase treated using the DNasefree kit from Ambion. To avoid inhibition of the reverse transcriptase, the volume of DNase treated RNA did not exceed 25% of the total volume during RT.

RT
Total RNA (1.0–5.0 μg) was reverse transcribed in 20–50 μl using the High Capacity cDNA Reverse Transcription Kit (Applied Biosystems) using random hexamers. The reaction mixture was incubated for 10 min at 25°C, 120 min at 37°C and finally for 5 min at 85°C, according to instructions from the manufacturer (Applied Biosystems). RT reactions were diluted 5–10-fold prior to qPCR.

Real-time qPCR

Conventional qPCR. All reactions (except when indicated) were performed in duplicate 10 μl volumes using 20 ng reverse transcribed total RNA in a StepOnePlus system (Applied Biosystems) with the SsoFast EvaGreen Supermix (BioRad) and an assay concentration of 300 nM using the cycling parameters: 95°C (20 s) followed by 40 cycles at 95°C (3 s) and 60°C (20 s). Melting curve analysis: 95°C (15 s); 60°C (60 s) and a progressive increase up to 95°C (0.5°C/min). Analysis of the data was performed with the StepOne software v.2.2.

High-throughput qPCR. The 96.96 Dynamic Arrays for the microfluidic BioMarkTM system (Fluidigm Corporation, CA, USA) (7) were used to study gene expression in 6.5 ng cDNA from mouse peritoneal macrophages or mouse uterus, as described below.

Specific target amplification. Pre-amplification of cDNA (produced from 25 to 65 ng of total RNA) was performed in the StepOnePlusycler (Applied Biosystems) at 95°C for 10 min activation step followed by 14 cycles: 95°C, (15 s), 60°C, (4 min)] in a total volume of 5 μl in the presence of all primers at a concentration of 50 nM. After pre-amplification, 20 μl Low EDTA TE Buffer [10 mM Tris pH8 (Ambion), 0.1 mM EDTA pH8 (Sigma)] was added to each sample.

Sample Mix for BioMark analysis. The pre-sample mix contained 66.7% 2X Taqman® Gene Expression Master Mix (Applied Biosystems), 6.67% 20X DNA Binding Dye Sample Loading Reagent (Fluidigm), 6.67% 20X EvaGreen® (Biotium), 20% Low EDTA TE Buffer. Sample mix was obtained by mixing 5.6 μl of the pre-sample mix with 1.9 μl of diluted cDNA.

Assay Mix. A quantity of 3.8 μl 2X Assay Loading Reagent (Fluidigm) and 1.9 μl Low EDTA TE Buffer were mixed with 1.9 μl of primers (20 μM of each forward and reverse primer).

qPCR conditions. After priming of the 96.96 Dynamic Array in the NanoFlex™ 4-Integrated Fluidic Circuits
Controller for loading and mixing under the following dynamic array was then placed again in the IFC of each assay mix were added to dedicated wells. The (IFC) Controller (Fluidigm), 5 μl of each sample and 5 μl of each assay mix were added to dedicated wells. The dynamic array was then placed again in the IFC Controller for loading and mixing under the following conditions: 50°C (2 min); 70°C (30 min) and 25°C (10 min). The loaded Dynamic Array was transferred to the BioMark™ real-time PCR instrument. After initial incubation at 50°C (2 min) and activation of the Hotstart enzyme at 95°C (10 min) cycling was performed using 95°C (15 s), and 60°C (1 min) for 35 cycles, followed by melting curve analysis (1°C/3 s).

Data analysis. Initial data analysis was performed with the Fluidigm real-time PCR analysis software v. 3.0.2 with linear derivative baseline correction and a quality correction set to 0.65.

Design of ValidPrime assays

Intergenic regions in the mouse genome with no known transcriptional activity were selected using the UCSC genome browser (http://genome.ucsc.edu/). In total, 30 assays targeting 10 different regions on 5 chromosomes were designed using PrimerBlast (NCBI). Amplification efficiencies were determined with a dilution series of gDNA (50–5000 haploid genome copies). PCR products were analyzed for purity by recording melting curves and by capillary micro-electrophoresis (MultiNA, Shimadzu), leading to the selection of five assays for limit of detection (LOD) and limit of quantification (LOQ) determination.

LOD and LOQ determination of ValidPrime assays

Five assays were selected for determination of LOD and LOQ using eight concentrations (0, 1, 2, 4, 8, 16, 32, 64 copies) in the presence of 50 ng/μl carrier yeast tRNA (Roche Molecular Biochemicals). Sequence information for the two best candidates, in terms of sensitivity and specificity, is provided in Supplementary Table S1. Except when stated otherwise, mVPA1 was used as the VPA.

GOI assay design and validation

Non-commercial GOI assays were either taken from previously published studies (5,8,9) or designed with the Primer-BLAST utility at NCBI. Sequences are reported in Supplementary Table S1. Specificity was evaluated by BLAST (mouse RefSeq database) during design and by in silico PCR (UCSC Genome Browser). Amplification efficiencies were evaluated in the BioMark system on dilutions series of both cDNA and gDNA.

Exogenous gDNA spiking experiments

Quantities ranging from 50 to 5000 haploid genome copies (corresponding to 0.15–15 ng gDNA) or water were added to 20 ng (StepOnePlus) or 6.5 ng (BioMark) cDNA. Non-spiked samples had low, but detectable gDNA levels. For the BioMark runs, the gDNA was added prior to the pre-amplification step. Genome copy number calculations were based on the NCBI m37 assembly of the C57Bl/6 mouse genome (2716965481 bp) assuming an average molecular weight of 660 g/mol/bp. The mass of a haploid mouse genome was thus estimated to be 2.98 pg.

Data analysis and statistics

CqDNA, CqRNA and %DNA were calculated using the gh-validprime software (https://code.google.com/p/gh-validprime). The GenEx software (v.5.3, www.xlmaq.se) was used for one-way ANOVA analysis and to calculate LOD. Data are presented as mean ± SD.

RESULTS

The ValidPrime method

We developed ValidPrime to estimate and correct for gDNA contribution in RT(+)-qPCR measurements in a more reliable manner than that afforded by RT(−) controls. We refer to the signal measured in an RT(+)– qPCR as CqNA (NA: Nucleic Acids) [Equation (1)], indicating contributions from RNA (CqRNA) as well as gDNA (CqDNA) as shown in Equation (2), expressed in relative quantities.

\[ C_{q_{RT+}} = C_{q_{NA}} \]
\[ 2^{-C_{q_{RNA}}} = 2^{-C_{q_{RNA}}} + 2^{-C_{q_{DNA}}} \]

 Traditionally, determination of the RNA component using RT(−) controls would be achieved using Equation (3). However, as detailed in the introduction, low reproducibility and other factors detract from the accuracy of this approach. We propose that Equation (4), derived in Supplementary Figure S1A, provides an accurate solution for CqDNA is estimated using ValidPrime, Equation (5), in which GOI refers to any transcribed ‘GOI’, including reference genes, studied in a RT–qPCR experiment. CqRNA and CqDNA refer to the signal contribution derived from RNA (cDNA) and DNA (gDNA), respectively, in a RT+ sample.

\[ C_{q_{RT+}} = -\log_2(2^{-C_{q_{RT+}}} - 2^{-C_{q_{RT-}}}) \]
\[ C_{q_{RNA}} = -\log_2(2^{-C_{q_{RNA}}} - 2^{-C_{q_{DNA}}}) \]
\[ C_{q_{DNA}} = C_{q_{Sample}} + C_{q_{GOI}} - C_{q_{VPA}} \]

For the determination of CqDNA [Equation (5)], the gDNA contamination level in a RT(+) sample (referred to as ‘Sample’) is measured with a gDNA-specific ValidPrime assay (VPA) (CqVPA). The VPA targets a non-transcribed locus present in one copy per normal haploid genome. However, since the gDNA sensitivity can be highly variable between GOI assays, the capacity of the GOI assay to amplify gDNA is compared with that of the VPA. In ValidPrime, this difference is tested on the AΔΔCt equation developed by Livak and Schmittgen (10), these calculations are distinct (Supplementary Figure S1B).

Figure 1 depicts a typical grid of qPCR data including the required controls for ValidPrime estimation of CqDNA and the subsequent correction of CqNA into CqRNA. Apart
\[ 2^{-Cq_{\text{NA}}} = 2^{-Cq_{\text{RNA}}} + 2^{-Cq_{\text{DNA}}} \quad (\text{Eq 2}) \]

<table>
<thead>
<tr>
<th></th>
<th>GOI 1</th>
<th>GOI 2</th>
<th>GOI 3</th>
<th>VPA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample 1</td>
<td>27.22</td>
<td>25.78</td>
<td>28.67</td>
<td>29.02</td>
</tr>
<tr>
<td>Sample 2</td>
<td>26.73</td>
<td>25.54</td>
<td>28.02</td>
<td>26.97</td>
</tr>
<tr>
<td>Sample 3</td>
<td>26.42</td>
<td>25.31</td>
<td>27.68</td>
<td>26.34</td>
</tr>
</tbody>
</table>

\[ Cq_{\text{DNA}} = Cq_{\text{VPA}} + Cq_{\text{gDNA}} - Cq_{\text{VPA}} \quad (\text{Eq 5}) \]

\[ Cq_{\text{NA}}^{\text{GOI}} = 29.02 + 29.62 - 28.61 = 30.03 \]

\[ Cq_{\text{RNA}} = -\log_{2} (2^{-Cq_{\text{NA}}} - 2^{-Cq_{\text{DNA}}}) \quad (\text{Eq 3}) \]

\[ Cq_{\text{NA}}^{\text{GOI}} = 27.22 \]

\[ Cq_{\text{RNA}}^{\text{GOI}} = -\log_{2} (2^{-27.22} - 2^{-30.03}) = 27.44 \]

\[ \%\text{DNA} = \frac{(2^{-Cq_{\text{DNA}}} / 2^{-Cq_{\text{NA}}}) * 100}{(2^{-30.03} / 2^{-27.22}) * 100} = 14.3 \% \quad (\text{Eq 6}) \]

**Figure 1.** ValidPrime: principles and exemplifying equations. ValidPrime uses the annotation \( Cq_{\text{NA}} \) for the signal measured in a (RT+) qPCR sample, to which both Nucleic Acids, RNA and DNA contribute, corresponding to \( Cq_{\text{RNA}} \) and \( Cq_{\text{DNA}} \) [Equation (2)]. The grid shows an example of an experimental design with 3 RT+ samples and 3 GOI assays, plus the controls required for the ValidPrime estimation of \( Cq_{\text{NA}} \) and the subsequent correction of \( Cq_{\text{NA}} \) to obtain \( Cq_{\text{RNA}} \). The term GOI is used in ValidPrime for both target transcripts and reference genes, since the calculations are independent of the gene type. The VPA column contains the data obtained with the VPA and the gDNA row contains measurements using purified gDNA as a sample. The equations under the grid illustrate the determination of \( Cq_{\text{DNA}} \), \( Cq_{\text{RNA}} \), and \( \%\text{DNA} \) for GOI 1 in sample 1 according to the color code in the grid.

from the GOI assays, that are specific for each study, the VPA has been added among the assays. In addition to samples 1–3, which correspond to any RT+ samples in qPCR study, one or several gDNA samples are added in the experimental design. The equations under the grid exemplify the calculations for GOI 1 in Sample 1. The gDNA contribution can also be expressed as a percentage of relative quantities [Equation (6)].

\[ \%\text{DNA} = \frac{(2^{-Cq_{\text{DNA}}} - 2^{-Cq_{\text{RNA}}})}{(2^{-Cq_{\text{DNA}}} - 2^{-Cq_{\text{RNA}}})} * 100 \quad (\text{6}) \]

**Assay validation**

In order to determine the accuracy of the ValidPrime method, we first designed and characterized candidate VPAs. Among 30 candidates from 10 different regions on five chromosomes, 26 amplified gDNA with efficiencies between 90 and 110%. Among the tested assays, mVPA1 (amplifying an 87-bp sequence in the qB region of chromosome 1) and mVPA5 (amplifying an 87-bp sequence in the qF region of chromosome 5) had the best characteristics in terms of sensitivity and specificity. LOD was 3.2 copies for mVPA1 (GenEx; Cut-off \( Cq \) 37; 95% CI; mean of two determinations) and 3.7 copies for mVPA5 (GenEx; Cut-off \( Cq \) 37; 95% CI) and the LOQ (SD < 45%) was 4 copies for both assays (Supplementary Figure S2). In four out of eight NTC reactions, a signal (\( Cq \) 38.1 ± 0.9) was detected with the mVPA5 assay, indicating formation of primer–dimers. However, the primer–dimer product was never observed in samples containing gDNA, as evaluated by melting curve analyses and by capillary micro-electrophoresis (MultiNA, Shimadzu).

Efficiency analysis for GOI assays was performed in the BioMark system. No amplification was observed in the NTC controls, except for Sprr2f (\( Cq \) 28.6), which was 10 cycles above the \( Cq \) measured in the sample with the lowest Sprr2f expression (\( Cq \) 18.5) and thus, far more than the proposed accepted minimal difference of five cycles between NTC and RT(+) sample (11,12). The generally low \( Cq \) values obtained with the BioMark system are explained by the 14-cycle pre-amplification step used in this protocol. The amplification efficiency was similar between assays as measured with a cDNA dilution series (95.5 ± 6.1%; mean \( R^2 \) 0.9932) and a gDNA dilution series for gDNA-sensitive assays (100.4 ± 7.7%; mean \( R^2 \) : 0.9962) (data not shown). All RNA samples used in the study had A260/A280 ratios between 1.9 and 2.0 (mean: 1.97); A260/A230 between 1.5 and 2.5 (mean: 2.13) and A260/A270 above 1.17 (mean: 1.23), where the latter tests for phenol contamination.

**Equivalence between \( Cq_{\text{DNA}} \) estimated with ValidPrime and RT(−) controls**

We next verified that the \( Cq_{\text{DNA}} \) values calculated with ValidPrime agree with those measured directly in RT(−)-qPCRs. Since a direct comparison is difficult, due to the poor reproducibility of RT(−) controls (see above), the following test was performed: RT(+) and RT(−) samples from two different tissues were spiked with 0.30 ng of gDNA (approximately 100 haploid genome copies) and measured using three gDNA-sensitive GOI assays. The data in Figure 2 are ratios of relative quantities (RQ) between either the total \( Cq_{\text{NA}} \) in RT(+) reactions or the corresponding \( Cq_{\text{DNA}} \) calculated with ValidPrime over the RQ in RT(−) reactions. As shown, tissue-dependent differences in the expression levels of the three target genes were observed [from 1.8- to 27-fold compared with RT(−) samples]. Independent of the expression level, the estimation by ValidPrime of the gDNA-derived signal levels (\( Cq_{\text{DNA}} \)) in RT(+) samples was in excellent agreement with the data from RT(−) samples, with the ratio of the relative quantities (1.20 ± 0.29) close to the theoretically expected value of 1.

**Calculation of \( Cq_{\text{RNA}} \) in RT(+) samples through the correction of signals derived from exogenously added gDNA**

Given the good correlation between ValidPrime estimation of \( Cq_{\text{DNA}} \) and RT(−) measurements, we next tested the accuracy of the calculation of the RNA-derived component \( Cq_{\text{RNA}} \) in RT(+) samples using Equation (4). In a first set of experiments, different amounts of gDNA were
added to cDNA test samples with low, but detectable, endogenous gDNA levels. All 32 GOI assays were gDNA-sensitive (Supplementary Table S1) and had gDNA amplification efficiencies similar to the VPA (i.e. passed the ValidPrime high confidence criteria detailed in Supplementary Figure S3). Both the traditional StepOnePlus microtiter plate-based qPCR (Figure 3A) and the microfluidic BioMark system (7) (Figure 3B) were used to collect raw data (CqNA) as input for ValidPrime estimations of the RNA-derived signal (CqRNA). Samples were grouped according to the level of DNA contribution. Using ValidPrime, we could accurately estimate the RNA-derived signal (CqRNA) even in samples with elevated gDNA-derived signals. However, the correction was less precise when the gDNA background exceeded 60% of the total signal. The demonstration that with ValidPrime we can identify and correct for signals derived from exogenous DNA in experimental RT–qPCR samples, using two different qPCR platforms, was first step toward a ‘proof-of-principle’. The correction is virtually independent of gene copy number since it works well both for GOI assays targeting one single locus and for genes with multiple pseudogenes (Supplementary Figure S4).

Correction of signals derived from endogenous gDNA
In order to evaluate the capacity of ValidPrime to correct for endogenous gDNA present in typical RNA preparations, a different strategy was applied. We used a gDNA-sensitive and a gDNA-insensitive assay for each GOI, with comparable amplification efficiencies. Three genes (Il1b, Serpine1 and Chi3l3) expressed in mouse macrophages were chosen as targets. Using the BioMark system, qPCR data were collected from 81 RNA preparations and the ValidPrime correction was applied. Despite identical overall gDNA content, the impact of the gDNA on the total signal obtained with the gDNA-sensitive assays differed considerably between the three genes. When the impact was limited (i.e. low %DNA), as in the case for Il1b, the effect of the ValidPrime correction was modest (Supplementary Figure S5). With increasing %DNA, as observed for Serpine1 and Chi3l3 (Figure 4A), the result of the correction becomes clearer, even in log2 scale (Figure 4B). Theoretically, given identical amplification efficiencies for the two assays and the absence of gDNA amplification, the CqRNA data in the scatter plots in Figure 4B should fall on a straight line with a slope of 1. The presence of gDNA will contribute to the signal measured with gDNA sensitive assays (x-axis) and the uncorrected CqNA data will therefore produce a slope >1. Even though the impact of the correction differs for the three genes, the CqRNA values estimated using ValidPrime restore linearity, especially for samples with a DNA contribution <60% (summarized in Figure 4C).

These data demonstrate that using ValidPrime, efficient correction of RT–qPCR data for the presence of endogenous gDNA is possible, as long as the DNA contribution to the total signal is <60%.

DISCUSSION
Since its invention in the early/mid 1990s (13,14), qPCR has undergone considerable methodological and technological advances (15). However, despite its direct impact on qPCR results, no alternative to RT(−) controls has, to our knowledge, been proposed to assess gDNA-derived contributions to the signals in RT–qPCR.
ValidPrime is a cost-efficient alternative to RT(−) controls to test for the presence of gDNA in samples. It is superior to RT(−) controls not only because of a higher accuracy, but also because fewer control reactions are required, eliminating the need for additional test reactions in the RT step. While the traditional approach for a study based on m samples and n genes requires m reverse transcription control reactions (RT−) and m × n extra qPCRs, ValidPrime only requires m + n + 1 control qPCRs and no RT(−) reactions (Table 1). As an example, in a BioMark 96.96 Dynamic Array experiment, ValidPrime reduces the number of controls by >95%.

ValidPrime is also the first method that proposes to correct for qPCR signals originating from contaminating gDNA. It is possible that the lack of accuracy and low reproducibility generally observed in RT(−) reactions has previously restrained the development of a correction-based model similar to that proposed in Equation (3). The present study includes data obtained with cDNA from five different mouse tissues analyzed with two qPCR instrument platforms, providing support for the general validity of ValidPrime.

It is important not to confuse gDNA contamination levels with the actual contribution of gDNA to the total signal, herein expressed as %DNA [Equation (6)]. Indeed, we did not observe any correlation between gDNA levels (as estimated by qPCR with the VPA) and the total signal (CqNA) measured in RT(+) qPCR reactions with GOI assays (Supplementary Figure S6). However, as evidenced from the data shown in Figure 4A and Supplementary Figure S5A, there is a clear positive correlation between %DNA and CqNA with the gDNA sensitive assay, which demonstrates the increased impact of contaminating gDNA in samples with low GOI expression levels.

The primer design strategy also strongly influences the impact of gDNA on the qPCR signal. Given the multi-exonic nature of most eukaryotic genes (16), it is conceivable that gDNA-insensitive assays can be designed for most targets in vertebrates. Regardless of the primer design strategy, the inability of a GOI assay to amplify gDNA needs to be validated experimentally. ValidPrime offers this possibility. However, for certain targets it is impossible to design transcript-specific assays. This can be due to either the presence of intronless pseudogenes or the absence of introns in single-exon genes. In order to assure a good accuracy for the ValidPrime correction, these gDNA sensitive assays should behave similarly to VPA against gDNA. In analogy with the comparative Ct method (or ΔΔCt method) (10), in which similar amplification efficiencies for the GOI and reference gene assays are presumed, estimation of CqRNA in ValidPrime assumes similar efficiencies for the GOI and gDNA assays.

When validated according to the Minimal Information for publication of Quantitative real-time PCR Experiments (MIQE) guidelines (17), gDNA-sensitive assays are in general perfectly compatible with ValidPrime. Nevertheless, when using a GOI assay for the first time with ValidPrime, and especially when Cq adjustment is requested, we recommend the inclusion of a gDNA dilution series with concentrations covering at least three log10 (e.g. 5–5000 haploid genomic copies). Consistent relation to VPA across the dilution series indicates similar amplification efficiencies of the two assays, which sanctions Cq correction with high confidence (Supplementary Figure S3). For VPAs, as well as for high confidence GOI assays, we generally observed perfectly linear amplifications from 5 to 10000 haploid genomic copies (corresponding to 0.015–30 ng) (Supplementary Figures S2 and S3). Even though it is possible that higher gDNA concentrations (i.e. >30 ng per reaction) could influence qPCR amplification efficiencies (18), such gDNA contamination levels are rarely, if ever, encountered in RT–qPCR experiments. Furthermore, we did not observe any differences in the VPA amplification between samples with purified gDNA

Figure 3. Correction of exogenous (spiked) gDNA with ValidPrime. The data are presented in linear scale as fold ratio \(2^{\Delta Cq} \), where \(Cq\) refers to the gDNA measured in both spiked (dark bars) or non-spiked controls (light bars) depending on whether or not ValidPrime correction was applied (VP−/VP+). The data are grouped based on the impact of exogenous DNA, expressed as percentage of the total signal (%DNA) in each sample. Data were collected with either 17 GOI assays on a StepOnePlus (Applied Biosystems) using mVPA1 and mVPA5 (A), or with 19 assays on a BioMark (Fluidigm) using mVPA1 (B). All assays passed the high confidence ValidPrime criteria (Supplementary Figure S3). Data are presented as the mean ± SD, with (n) designating the number of samples in each group. cDNAs were from mouse kidney or liver for the StepOnePlus studies and mouse uterus for the BioMark study.
and mixed samples, spiked with cDNA or RNA (Supplementary Table S2).

Even though we consistently observed very low variability between replicates in VPA-gDNA amplifications over a wide range of initial gDNA concentrations (Supplementary Figure S2 and Supplementary Table S2), it is advisable to use 1–10 ng gDNA (i.e. 300–3000 haploid genome copies) per qPCR, when only one gDNA concentration is included in the design. This range favors reliable and distinct gDNA amplification with the VPA and the ‘high confidence’ gDNA-sensitive GOI assays. Alternatively, an efficiency (E) based criterion can be

Figure 4. Correction of endogenous gDNA with ValidPrime. Comparison of results obtained with two assays targeting Serpine1 (left) or Chi3l3 (right) in cDNA prepared from mouse peritoneal macrophages and measured in the BioMark qPCR system. The ‘gDNA-sensitive’ assays amplify both gDNA and cDNA, while the ‘gDNA-insensitive’ assays only recognize the transcript. (A) Scatter plots showing the correlation between the %DNA [as defined in Equation (6)] and CqNA data obtained with the gDNA-sensitive assays in each of 81 independent RNA preparations (means of duplicates). The positive correlation between %DNA and Cq illustrates the increasing impact of the gDNA contamination with decreasing total signal. Mean and median values refer to %DNA levels. (B) CqNA data measured with the gDNA-insensitive assays plotted against the corresponding CqNA data (dark blue) or ValidPrime-estimated CqRNA (light blue and orange), obtained with the gDNA-sensitive assays. Samples with a DNA contribution of 60–90 % are shown in orange and those with <60% in light blue. (C) Tables summarizing the effect of ValidPrime correction and data filtering on the slope and the coefficient of determination (R²).
ValidPrime reduces the number of required control reactions in RT-qPCR

<table>
<thead>
<tr>
<th>Samples (m)</th>
<th>No. of controls</th>
<th>Assays (n)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>10</td>
</tr>
<tr>
<td>10</td>
<td>20</td>
<td>110</td>
</tr>
<tr>
<td>24</td>
<td>48</td>
<td>264</td>
</tr>
<tr>
<td>48</td>
<td>96</td>
<td>50</td>
</tr>
<tr>
<td>96</td>
<td>192</td>
<td>98</td>
</tr>
</tbody>
</table>

The roman values indicates Traditional RT− strategy: \((m \times n) + m\) and the bold values indicates ValidPrime: \((m + n + 1)\). ValidPrime replaces the need to perform RT(−) controls for all RT(+) reactions and reduces substantially the number of controls compared to a conventional set up. In an expression profiling experiment based on \(m\) samples and \(n\) assays, the RT(−) approach requires \(m\) RT(−) reactions followed by \(m \times n\) qPCR controls, whereas ValidPrime only requires \(m + n + 1\) controls. The numbers in the table are based on single measurements for both approaches. Even when \(p\) gDNA samples/concentrations are included in the experimental setup using ValidPrime, the number of control reactions \([m + (p \times n) + p]\) is still largely inferior to the RT(−) approach.

CONCLUSION

ValidPrime provides, for the first time, the opportunity to correct reliably for gDNA background in qPCR. Correction is possible for any GOI assay that consistently amplifies gDNA, given that the DNA contribution does not exceed 60% of the signal. ValidPrime is superior to traditional RT(−) controls because of its higher accuracy and the lower number of controls required, which leads to a substantial cost savings.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Tables 1 and 2, Supplementary Figures 1–7 and Supplementary References [1,10,17,22–25].

ACKNOWLEDGEMENTS

The authors are grateful for critical comments from Drs Pascal Martin, Sébastien Déjean, Coralie Fontaine and Anders Ståhlberg. Drs Geneviève Tavernier and Elodie Riant provided mouse blood and tissue samples and Hortense Bergès provided technical assistance during the isolation of mouse macrophages.
ValidPrime can be used as a reliable, cost-efficient alternative to RT(-) controls to survey gDNA background in RT–qPCR, and as a tool to determine the RNA-derived signal (\(C_{qRNA}\)) in RT(+)–qPCR reactions. To optimize its accuracy when \(C_{qRNA}\) calculation is desired, validation of GOI assays in gDNA samples is recommended, as outlined in (A). Asterisks indicates the efficiency evaluation and melting curve/electrophoresis-based analysis. This includes an evaluation of the gDNA sensitivity of GOI assays using dilution series with gDNA samples spanning at least three \(\log_{10}\) in copy number. GOI assays that do not amplify gDNA are attributed the grade A+. The amplification of gDNA by high-confidence assays should be specific and with an efficiency similar to that of the VPA (see ‘Discussion’ section and Supplementary Figure S3). For GOI assays with suboptimal, but confidently determined (17,21) efficiency, Equation (7) could be applied to adjust \(C_{qNA}\) data. To optimize specificity, there should also be consistency between the melting curves of PCR products in gDNA and cDNA samples. (B) \(C_{qRNA}\) calculation with ValidPrime-validated GOI assays. High confidence and A+ assays can be used with less gDNA samples for \(C_{qRNA}\) determination. It is recommended to confirm the absence of gDNA amplification at least once for A+ assays. Samples that do not contain sufficient gDNA to generate a signal with the VPA are attributed A*. As for gDNA insensitive A+ assays, \(C_{qRNA}\) equals \(C_{qDNA}\) (i.e. output = input) in A+ samples, since the DNA-derived signal is negligible [see Equations (2 and 4)]. For gDNA-sensitive GOI assays, \(C_{qRNA}\) is calculated by a \(C_{qDNA}\)-based correction of \(C_{qDNA}\) using Equations (4 and 5). To minimize the risk of jeopardizing the accuracy of the \(C_{qRNA}\) estimation, it is not advisable to perform correction on samples where the DNA-derived signal exceeds 60%. The calculations are facilitated using the ValidPrime software. Details on additional assay/sample grading and data output formats employed by the software are provided in Supplementary Figure S7. The \(C_{qRNA}\) output data can be used for downstream data processing, such as normalization against reference genes.

**Figure 5.** ValidPrime flowchart. ValidPrime GOI assay validation. ValidPrime can be used as a reliable, cost-efficient alternative to RT(-) controls to survey gDNA background in RT–qPCR, and as a tool to determine the RNA-derived signal (\(C_{qRNA}\)) in RT(+)–qPCR reactions. To optimize its accuracy when \(C_{qRNA}\) calculation is desired, validation of GOI assays in gDNA samples is recommended, as outlined in (A). Asterisks indicates the efficiency evaluation and melting curve/electrophoresis-based analysis. This includes an evaluation of the gDNA sensitivity of GOI assays using dilution series with gDNA samples spanning at least three \(\log_{10}\) in copy number. GOI assays that do not amplify gDNA are attributed the grade A+. The amplification of gDNA by high-confidence assays should be specific and with an efficiency similar to that of the VPA (see ‘Discussion’ section and Supplementary Figure S3). For GOI assays with suboptimal, but confidently determined (17,21) efficiency, Equation (7) could be applied to adjust \(C_{qNA}\) data. To optimize specificity, there should also be consistency between the melting curves of PCR products in gDNA and cDNA samples. (B) \(C_{qRNA}\) calculation with ValidPrime-validated GOI assays. High confidence and A+ assays can be used with less gDNA samples for \(C_{qRNA}\) determination. It is recommended to confirm the absence of gDNA amplification at least once for A+ assays. Samples that do not contain sufficient gDNA to generate a signal with the VPA are attributed A*. As for gDNA insensitive A+ assays, \(C_{qRNA}\) equals \(C_{qDNA}\) (i.e. output = input) in A+ samples, since the DNA-derived signal is negligible [see Equations (2 and 4)]. For gDNA-sensitive GOI assays, \(C_{qRNA}\) is calculated by a \(C_{qDNA}\)-based correction of \(C_{qDNA}\) using Equations (4 and 5). To minimize the risk of jeopardizing the accuracy of the \(C_{qRNA}\) estimation, it is not advisable to perform correction on samples where the DNA-derived signal exceeds 60%. The calculations are facilitated using the ValidPrime software. Details on additional assay/sample grading and data output formats employed by the software are provided in Supplementary Figure S7. The \(C_{qRNA}\) output data can be used for downstream data processing, such as normalization against reference genes.
FUNDING
Institut National de la Santé et de la Recherche Médicale.
Funding for the open access charge: Institut des Maladies Métaboliques et Cardiovasculaires, Toulouse, France.

Conflict of interest statement. M.K. and D.S. are employees of TATAA Biocenter AB, a commercial source for the ValidPrime assays.

REFERENCES