Available online at www.sciencedirect.com
SCIENCE DIRECT®
Analytical Biochemistry 335 (2004) 1–9
ANALYTICAL BIOCHEMISTRY
www.elsevier.com/locate/yabio

Equivalence test in quantitative reverse transcription polymerase chain reaction: confirmation of reference genes suitable for normalization

Florian Haller[a,*], Bettina Kulle[b], Stefanie Schwager[a], Bastian Gunawan[a], Anja von Heydebreck[c,1], Holger Sültmann[d], László Füzesi[a]

[a] Department of Pathology, Georg-August University of Göttingen, D-37075 Göttingen, Germany
[b] Department of Genetic Epidemiology, Georg-August University of Göttingen, D-37075 Göttingen, Germany
[c] Max Planck Institute for Molecular Genetics, D-14195 Berlin, Germany
[d] Department of Molecular Genome Analysis, German Cancer Research Center, D-69120 Heidelberg, Germany

Received 7 April 2004
Available online 8 October 2004

Abstract

In quantitative reverse transcription–polymerase chain reaction (qRT–PCR), normalization using reference genes is a common useful approach, but the validation of suitable reference genes remains a crucial problem. Use of unconfirmed reference genes may lead to misinterpretation of the expression of target genes. The aim of this study was to adapt an adequate statistical approach to identify and validate reference genes suitable for normalization in qRT–PCR assays. We introduce the equivalence test for the identification of stably expressed reference genes. To evaluate the advantages of this test, the expression of five genes widely used as reference genes (*18S, B2M, HPRT1, LMNB1*, and *SDHA*), and of two target genes (*TP53* and *MMP2*), was determined with qRT–PCR in different tissues (clear cell renal cell carcinoma, colon carcinoma, and gastrointestinal stromal tumors). We demonstrate that a stable expression of a reference gene in one tumor type does not predict a stable expression in another tumor type. In addition, we found that even within one tumor type, the expression of a reference gene was not stable for different biological groupwise comparisons. These observations confirm that there is no universal reference gene and underline the importance of specific validation of potential reference genes for any experimental condition.

Keywords: Normalization; Quantitative RT–PCR; Reference gene; Housekeeping gene; Equivalence test; Gene expression; Clear cell renal cell carcinoma; Colon carcinoma; Gastrointestinal stromal tumor

Quantitative reverse transcription–polymerase chain reaction (qRT–PCR)[2] is a highly specific and sensitive technique for the quantification of gene expression on the mRNA level [1–3]. The advantage of speed, throughput, and excellent reproducibility, together with a wide variety of instruments and reagents, makes this technique a powerful new tool in experimental research and clinical diagnostics. Two major strategies exist for interpretation of the data. The expression of a target gene can either be related to total RNA input [4–6] or be declared as a ratio to the expression level of reference or so-called housekeeping genes [7–10]. In the latter, assuming a stable expression of the reference gene in all cells, this ratio would represent the relative

* Corresponding author. Fax: +49 551 398 627.
E-mail address: florian.haller@stud.uni-goettingen.de (F. Haller).
[1] Present address: Department of Bio- and Chemoinformatics, Merck KGaA, D-64293 Darmstadt, Germany.
[2] *Abbreviations used:* qRT–PCR, quantitative reverse transcription–polymerase chain reaction; ccRCC, clear cell renal cell carcinoma; GIST, gastrointestinal stromal tumor; RT, reverse transcription.

0003-2697/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.ab.2004.08.024

expression level of the target gene. Unfortunately, a universal reference gene that is expressed at a stable level unaffected by all biological conditions in all tissues simply does not exist [6,11,12]. For nearly every gene traditionally used for normalization, differential expression dependent on at least one biological condition has been described [13–20], although not every gene is affected by a certain experimental condition in the same manner [15–17,19,21–24]. Nonetheless, normalization to a reference gene is the only option when dealing with small amounts of total RNA (e.g., using biopsies or laser-assisted cell picking), when the precise quantification of total RNA is not available, or when mRNA is used.

The aim of this study was to adapt an adequate statistical approach to identify and validate reference genes suitable for normalization in qRT–PCR assays. This statistical approach should be easy to handle and transferable to different kinds of experimental conditions. We hereby introduce the equivalence test and evaluate its usefulness by determining the expression of five genes widely used as reference genes (*18S, B2M, HPRT1, LMNB1*, and *SDHA*) by qRT–PCR in two different experimental setups. We compared tissue from a solid tumor with its matching normal tissue (clear cell renal cell carcinoma [ccRCC] vs. normal renal parenchyma, colon carcinoma vs. normal colon mucosa) and analyzed a series of 35 gastrointestinal stromal tumors (GISTs). Furthermore, we demonstrate the effect of interpreting expression of the target genes, *TP53* and *MMP2*, after normalization to the five different reference genes.

Materials and methods

Surgical specimen

Tissue samples from 10 cases of ccRCC and paired normal renal parenchyma from the same patient (Table 1), five cases of colon carcinoma, and five cases of normal colon mucosa from different patients (Table 1), and a series of 35 cases of GIST (Table 2) were cut into pieces of approximately 100–200 mg, snap frozen in liquid nitrogen immediately after surgery, and stored at −80 °C. Staging of ccRCC and colon carcinoma was performed according to the current TNM classification [25]. Malignant potential for GIST was estimated as proposed by Miettinen et al. [26]. All patients were treated surgically, and none had received chemotherapy or radiation prior to surgery.

RNA isolation

After homogenization with an Ultra Turrax T25 (IKA–Werke GmbH, Staufen, Germany), total RNA was isolated with TRIzol Reagent (Invitrogen Life Technologies, Karlsruhe, Germany) according to the manufacturer's manual, using approximately 50 mg frozen tissue per milliliter TRIzol. Total RNA concentration was quantified with the RNA 6000 nano LabChip using the Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA). Only high-quality total RNA was used, as confirmed by high peaks for *18S* and *28S* ribosomal RNA.

Reverse transcription

First-strand cDNA was generated from 5 μg of total RNA per sample using the Superscript II RNase H^- Reverse Transcriptase (Invitrogen Life Technologies) according to the manufacturer's manual, including 125 ng of random hexamer primers (Invitrogen Life Technologies) and 40 U of RNasIn Ribonuclease Inhibitor (Promega, Mannheim, Germany). RT product was aliquoted in equal volumes and stored at −20 °C.

Quantitative PCR

Gene-specific primers for *18S, LMNB1*, and *TP53* (Table 3) were designed on different exons with a 60 °C melting temperature and a length of 18–24 bp for PCR products with a length of 70–150 bp, using Primer3 software [27] (Table 3). Primers for *MMP2* were kindly provided by P. Thelen (Department of Urology, Georg-August University of Göttingen), primers for *B2M* were adapted from [28], and primers for *HPRT1* and *SDHA* were adapted from [29]. PCR was run in 20-μl reactions in triplicate on an iCycler (Bio-Rad Laboratories GmbH, Munich, Germany) using the Eurogentec qPCR Core Kit for Sybr Green I (Eurogentec, Seraing, Belgium) and gene-specific primers in a final concentration of 300 nM. The temperature profile consisted of (i) an initial step of 95 °C for 10′ for *Taq* activation, (ii) 40 cycles of 95 °C for 15″and 60 °C for 1′, and (iii) a final melt curve analysis with a temperature ramp from 60 to 95 °C with a heating rate of 3 °C/min. PCR efficiencies were calculated with a relative standard curve derived from a cDNA mixture (a two-fold dilution series with seven measuring points in triplicate) and gave regression coefficients greater than 0.98 and efficiencies greater than 95%. Gene-specific amplification was confirmed by a single peak in melt curve analysis and a single band in high-resolution agarose gel electrophoresis (SeaKem LE agarose, BMA, Rockland, ME, USA). No template controls (no cDNA in PCR) or genomic controls (no enzyme in RT) were run for each gene to detect unspecific or genomic amplification and primer dimerization, but no peaks in melt curve analysis and no band in high-resolution agarose gel electrophoresis (except *18S*, for which the proportion of genomic amplification was <0.01) were observed. Relative expression levels were calculated from the relative standard curve as described in [10] and logarithmized to

Table 1
Main clinico-pathological data in 10 cases of ccRCC, 5 cases of colon carcinoma, and 5 cases of normal colon mucosa

Case	Tissue	Age (years)	Gender	T	N	M	G
1	ccRCC	66	Female	1	X	X	2
2	ccRCC	48	Female	1a	0	X	1
3	ccRCC	69	Male	1a	X	X	2
4	ccRCC	47	Male	1b	0	X	2
5	ccRCC	85	Male	3a	X	X	2
6	ccRCC	84	Male	3b	X	X	3
7	ccRCC	57	Female	3b	X	X	2
8	ccRCC	45	Male	3b	1	1 (OSS)	2
9	ccRCC	84	Male	3b	X	1 (ADR)	2
10	ccRCC	75	Male	3b	X	X	3
11	Colon carcinoma	78	Female	3	2	X	2
12	Colon carcinoma	49	Male	3	2	X	2
13	Colon carcinoma	56	Female	3	2	X	2
14	Colon carcinoma	63	Male	3	0	X	2
15	Colon carcinoma	70	Male	3	3	X	3
16	Colon mucosa	69	Female	—	—	—	—
17	Colon mucosa	49	Male	—	—	—	—
18	Colon mucosa	56	Female	—	—	—	—
19	Colon mucosa	63	Male	—	—	—	—
20	Colon mucosa	70	Male	—	—	—	—

Table 2
Main clinico-pathological data in 35 cases of GIST

Case	Age (years)	Gender	Site	Size (cm)	Mitoses/50 high-power fields	Estimated malignancy (probably) [26]
1	69	Female	Stomach	1.7	$\leqslant 5$	Benign
2	73	Male	Stomach	3.5	$\leqslant 5$	Benign
3	80	Female	Stomach	4.0	$\leqslant 5$	Benign
4	78	Male	Stomach	4.0	$\leqslant 5$	Benign
5	77	Female	Stomach	4.0	$\leqslant 5$	Benign
6	52	Male	Stomach	4.0	$\leqslant 5$	Benign
7	45	Female	Stomach	4.5	$\leqslant 5$	Benign
8	82	Female	Stomach	4.5	$\leqslant 5$	Benign
9	72	Female	Stomach	4.5	>5	Malignant
10	77	Female	Stomach	5.0	>5	Malignant
11	64	Female	Stomach	5.5	$\leqslant 5$	Low malignant
12	67	Male	Stomach	6.0	$\leqslant 5$	Low malignant
13	47	Male	Stomach	6.0	$\leqslant 5$	Low malignant
14	64	Male	Stomach	6.0	>5	Malignant
15	64	Male	Stomach	6.0	>5	Malignant
16	66	Male	Stomach	6.3	>5	Malignant
17	65	Male	Stomach	6.5	$\leqslant 5$	Low malignant
18	53	Male	Stomach	7.0	>5	Malignant
19	59	Female	Stomach	7.5	$\leqslant 5$	Low malignant
20	80	Male	Stomach	8.0	>5	Malignant
21	66	Male	Stomach	10.0	$\leqslant 5$	Low malignant
22	39	Female	Stomach	10.0	$\leqslant 5$	Low malignant
23	70	Female	Stomach	10.0	>5	Malignant
24	78	Female	Stomach	13.0	>5	Malignant
25	63	Male	Stomach	13.0	>5	Malignant
26	49	Female	Stomach	17.0	>5	Malignant
27	61	Female	Stomach	30.0	>5	Malignant
28	54	Female	Small intestine	7.5	>5	Malignant
29	54	Male	Small intestine	8.0	$\leqslant 5$	Malignant
30	62	Female	Small intestine	8.0	>5	Malignant
31	82	Male	Small intestine	9.0	$\leqslant 5$	Malignant
32	61	Male	Small intestine	10.0	>5	Malignant
33	57	Male	Small intestine	15.0	$\leqslant 5$	Malignant
34	56	Male	Colorectal	6.5	>5	Malignant
35	66	Male	Colorectal	7.0	>5	Malignant

Table 3
Primer sequences

Gene	Forward primer	Reverse primer	Amplicon length (bp)	Amplicon melting temperature (°C)	Source
18S	ACATCCAAGGAAGGCAGCAG	TCGTCACTACCTCCCCGG	62	85.5	[a]
B2M	ACCCCCACTGAAAAAGATGA	ATCTTCAAACCTCCATGATG	114	84.0	[28]
HPRT1	TGACACTGGCAAAACAATGCA	GGTCCTTTTCACCAGCAAGCT	94	83.5	[29]
LMNB1	CTGGAAATGTTTGCATCGAAGA	GCCTCCCATTGGTTGATCC	89	80.5	[a]
SDHA	TGGGAACAAGAGGGCATCTG	CCACCACTGCATCAAATTCATG	86	81.0	[29]
TP53	GGGACAGCCAAGTCTGTGA	AATCAACCCACAGCTGCAC	97	87.0	[a]
MMP2	CCAAGTGGTCCGTGTGAAGT	CATGGTGAACAGGGCTTCAT	193	88.0	[b]

[a] Primers were designed using Primer3 software.
[b] Primers for MMP2 were kindly provided by P. Thelen (Department of Urology, Georg-August University of Göttingen).

obtain approximately normally distributed data. No significant influence of gene expression associated with RNA isolation batch, cDNA synthesis batch, or quantitative PCR batch, or with failing of the RT in individual cDNA samples, was detected.

Statistics

Descriptive statistics and *t* test for dependent and independent samples and graphs were performed with Statistica 6.0 (StatSoft, Hamburg, Germany). The *t* tests were used to detect significant groupwise differences in the logarithmized relative expression of each reference gene. The equivalence test was adapted from [30]. We used a constant level $\alpha = 0.05$ for rejection of the null hypothesis.

Equivalence test

The difference (δ) of the expected logarithmized expression levels μ_1 and μ_2 of a reference gene in two groups, 1 and 2, is given as

$$\delta := \mu_1 - \mu_2. \tag{1}$$

The classical two-sided *t* test for independent samples detects a significant difference δ with the null hypothesis formulated as H_0: $\delta = 0$ with a controllable significance level of α. This controls the mistake of incorrectly rejecting the null hypothesis. However, the absence of a significant difference does not imply that the null hypothesis is true. The type 2 error is uncontrollable. Therefore, for an adequate confirmation of a reference gene, a test for approval of equality is needed. Obviously, exact equality ($\delta = 0$) cannot be proved, but it can be tested whether the absolute value of the true difference δ is bounded by a small positive number ε. This leads to the formulation of the equivalence hypothesis:

$$H_0 : \delta \notin [-\varepsilon; \varepsilon] \text{ vs. } H_1 : \delta \in [-\varepsilon; \varepsilon]. \tag{2}$$

A reference gene can be called equivalently expressed on level α if the confidence interval $CI(\delta)$ for the difference δ of the expected logarithmized expression values μ_1 and μ_2 is part of a determined deviation area, for example,

$$[-\varepsilon; \varepsilon] = [\log_2(0.5); \log_2(2)] = [-1; 1], \tag{3}$$

when using a fold change of 2. To evaluate the confidence interval $CI(\delta)$, the distribution of the standard *t* statistic is used with confidence probability $1–2\alpha$. Now the lower δ_L border and the upper δ_U border of the symmetrical confidence interval can be calculated as

$$CI(\delta) := [\delta_L; \delta_U], \tag{4}$$

where

$$\delta_{L,U} := \bar{X}_1 - \bar{X}_2 \mp \frac{s}{\sqrt{\frac{N_1 \times N_2}{N_1 + N_2}}} \times t_{N_1+N_2-2;1-\alpha} \tag{5}$$

with a standard deviation *s* calculated by

$$s = \sqrt{\frac{(N_1 - 1) \times S_1^2 + (N_2 - 1) \times S_2^2}{N_1 + N_2 - 2}}. \tag{6}$$

For $i = 1, 2$, $\bar{X}_i$ is the mean of the logarithmized expression values X_{ij} in group $i, j = 1, \ldots, N_i$ subject j in group i, N_i is the number of individuals in group i, $t_{N_1+N_2} - 2; 1 - \alpha$ is the $1 - \alpha$ quantile of the *t* distribution with $N_1 + N_2 - 2$ degrees of freedom, and s_i is the standard deviation of the expression values in group i. The confidence interval and the values explained above can be calculated by standard statistic software. If

$$[\delta_L; \delta_U] \subset [-\varepsilon, \varepsilon], \tag{7}$$

the confidence interval is part of the determined deviation area, and H_0 (inequivalence) can be rejected on level α; therefore, H_1 (equivalence) can be accepted on level α, and the reference gene is equivalently expressed. The confidence interval $CI(\delta)$ must include 0.

This equivalence test can be used for independent samples. For dependent samples, such as tumor and paired normal tissue of one person, a slightly modified version should be used. For this situation, we introduce $d_j = X_{1j} - X_{2j}$, the difference of the logarithmized expression values of subject j. In case of dependent samples, the difference δ is redefined as $\delta := \mu_d$, where μ_d is the expected difference of the logarithmized expression levels. The same formulation for the hypothesis as in the equivalence test version for independent samples (Eq. (2)) leads to a confidence interval calculated as

$$\mathrm{CI}(\delta) := [\delta_{\mathrm{L}}; \delta_{\mathrm{U}}], \tag{8}$$

where

$$\delta_{\mathrm{L,U}} := \bar{d} \mp \frac{s}{\sqrt{N}} \times t_{N-1;1-\alpha} \tag{9}$$

with the same notation as above and the standard deviation s calculated as

$$s = \sqrt{\frac{1}{N-1}\sum_{j=1}^{N}(d_j - \bar{d})}, \tag{10}$$

where d is the mean value of d_j for subject j. The equivalence test is now performed in the same way as for independent samples: If $[\delta_{\mathrm{L}};\ \delta_{\mathrm{U}}] \subset [-\varepsilon,\ \varepsilon]$, the confidence interval is part of the determined deviation area, and H_0 (inequivalence) can be rejected on level α; therefore, H_1 (equivalence) can be accepted on level α, and the reference gene is equivalently expressed. The confidence interval $\mathrm{CI}(\delta)$ must include 0.

Parametric versions of the equivalence tests have been used. They have a higher power to detect equivalence in normally distributed data, obtained from quantitative qRT–PCR, than do nonparametric versions. However, for the case of not normally distributed data, nonparametric versions of the equivalence test also exist [30].

In qRT–PCR data analysis, there is no definite minimal cutoff for a significant fold change in differentially expressed target genes. Depending on the distribution of the data with respect to biological and methodical variance, even a fold change (f) of less than 2 may be statistically and biologically significant. Using the equivalence test, a determination area must be defined for the decision of whether the confidence interval for the distribution of the expression of a tested reference gene is part of that determination area or not. To identify a useful maximal fold change for equivalent expression applicable to the equivalence test, we decided to evaluate fold changes of 2 and 3.

Results

Comparison of ccRCC with normal renal parenchyma

There was no significant difference in the expression of any of the five reference genes detectable with the t test for dependent samples. Nevertheless, equivalent expression as confirmed by the equivalence test for dependent samples was identified only for *HPRT1* using a fold change of 3 for the cutoff (Fig. 1A). The other four genes were not confirmed to be equivalently expressed. *B2M* and *LMNB1* showed a tendency to be higher expressed in ccRCC (fold change ccRCC/renal parenchyma values of 1.3 and 1.7, respectively), whereas *18S* and *SDHA* appeared to be higher expressed in normal renal parenchyma (fold change ccRCC/renal parenchyma values of 0.6 and 0.8, respectively), but these tendencies were not significant. After normalization to the confirmed reference gene *HPRT1*, the expression of *TP53* was not significantly different between ccRCC and renal parenchyma. On the other hand, a significant difference ($P = 0.02$) in the expression of *TP53* was detectable after normalization to the unconfirmed

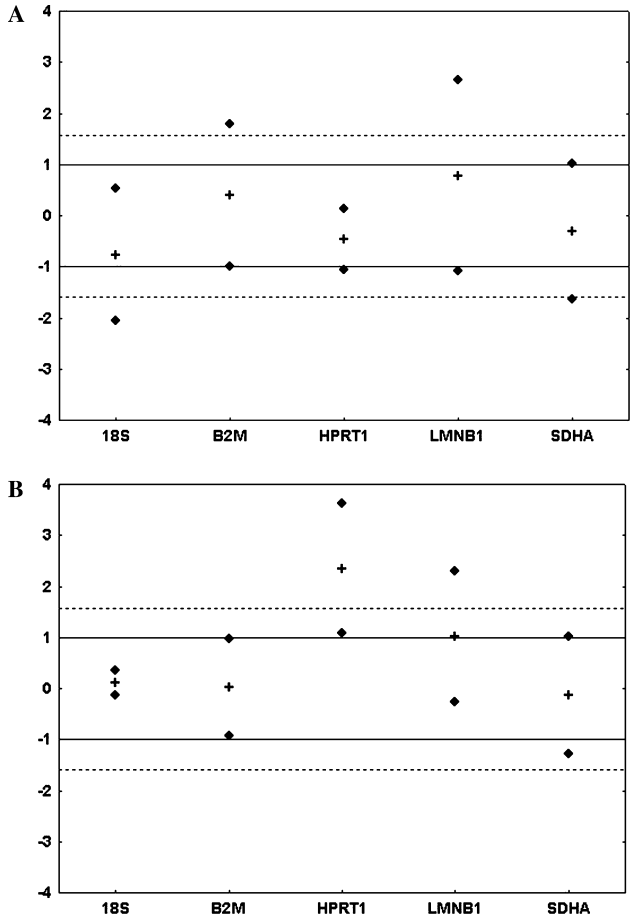

Fig. 1. Equivalence test for five reference genes in 10 ccRCCs and paired normal renal parenchyma (version for dependent samples) (A) and five colon carcinomas and normal colon mucosa (version for independent samples) (B). Shown are differences of the means (+) and matching symmetrical confidence intervals (♦) for the logarithmized relative expression of five reference genes. The deviation area [−1; 1] for a fold change of $f \leqslant 2$ is plotted as continuous lines, whereas the deviation area [−1.58; 1.58] for a fold change of $f \leqslant 3$ is plotted as dotted lines. If the symmetrical confidence interval is part of the deviation area, the reference gene is confirmed to be expressed equivalently. (A) The upper border of the symmetrical confidence interval (upper diamond) $\geqslant 1$ (1.58) indicates unequivalent higher expression in ccRCC (*B2M* and *LMNB1*), whereas the lower border of the symmetrical confidence interval (lower diamond) $\leqslant -1$ (−1.58) indicates higher expression in normal renal parenchyma (*18S* and *SDHA*). (B) The upper border of the symmetrical confidence interval (upper diamond) $\geqslant 1$ (1.58) indicates higher expression in colon carcinoma (*HPRT1* and *LMNB1*), whereas the lower border of the symmetrical confidence interval (lower diamond) $\leqslant -1$ (−1.58) indicates higher expression in normal colon mucosa (*SDHA* when using a cutoff of 2).

reference gene *LMNB1*, where *TP53* was found to be lower expressed in tissues from ccRCC compared with normal renal parenchyma (fold change ccRCC/renal parenchyma value of 0.6). There was a tendency for a higher expression of *TP53* in ccRCC after normalization to the unconfirmed reference genes *18S* and *SDHA* as well as a tendency for a higher expression of *TP53* in normal renal parenchyma after normalization to the unconfirmed reference gene *B2M*, but these tendencies were not significant.

Comparison of colon carcinoma with normal colon mucosa

The *t* test for independent samples detected a significant difference in the expression between colon carcinoma and normal colon mucosa only in the case of *HPRT1* ($P = 0.01$), which was higher expressed in colon carcinoma (fold change colon carcinoma/colon mucosa value of 5.1). On the other hand, equivalent expression confirmed by the equivalence test for independent samples was present only for *18S* and *B2M* when using a fold change of 2, whereas it was present for *SDHA* as well when using a fold change of 3 (Fig. 1B). For *LMNB1*, the expression was not equivalent, although there was no significant difference. *LMNB1* showed a tendency to be higher expressed in colon carcinoma (fold change colon carcinoma/colon mucosa value of 2.0). *HPRT1* could not be confirmed to be equivalently expressed, and the confidence interval for equivalent expression did not include 0 due to the significant differential expression of *HPRT1*. For the expression of *TP53*, no significant difference could be detected after normalization to any of the reference genes. However, *TP53* showed a tendency to be higher expressed in colon carcinoma when it was normalized to the confirmed reference genes *18S* and *B2M* (fold change colon carcinoma/colon mucosa values of 1.9 and 2.1, respectively). On the other hand, when normalized to *HPRT1*, the expression of *TP53* showed a tendency to be higher expressed in colon mucosa (fold change colon carcinoma/colon mucosa value of 0.5).

Gastrointestinal stromal tumors

The *t* test for independent samples detected no gender-specific expression for the five reference genes. Using a fold change of 2, the equivalence test for independent samples confirmed *18S, HPRT1*, and *LMNB1* to be equivalently expressed between GISTs from female and male patients (Fig. 2A). Using a fold change of 3, *B2M* and *SDHA* could also be confirmed to be equivalently expressed, but they tended to be higher expressed in GISTs from female patients. The normalized expression of *MMP2* was not significantly different between GISTs from female and male patients, without any differences between normalization to confirmed reference genes and unconfirmed reference genes. In the comparison between benign/low-malignant and malignant GISTs according to Miettinen et al. [26], the *t* test for independent samples detected a significant difference in the expression of *LMNB1* ($P = 0.04$), with higher expression levels in probably malignant GISTs (fold change benign/malignant value of 0.4). *B2M, HPRT1*, and *SDHA* were not significantly differentially expressed but showed a tendency to be higher expressed in probably malignant GISTs. Using a fold change of 2 as a cutoff, the equivalence test for independent samples confirmed only *18S* to be equivalently expressed (Fig. 2B). When a fold change of 3 was allowed, *B2M* and

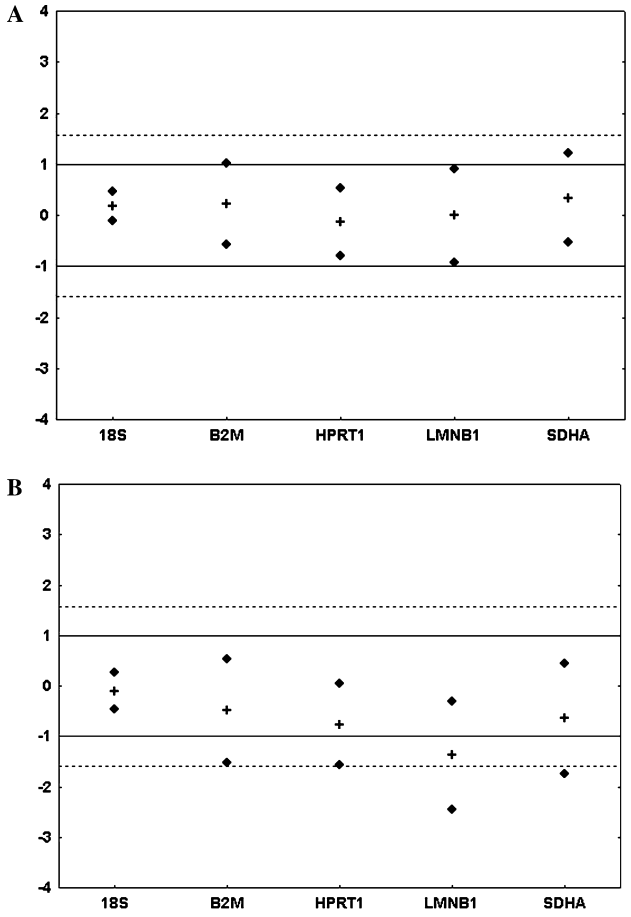

Fig. 2. Equivalence test for five reference genes in 35 GISTs in association with gender (A) and estimated malignancy (B). (A) The upper border of the symmetrical confidence interval (upper diamond) $\geqslant 1$ (1.58) indicates higher expression in female patients (*B2M* and *SDHA* when using a cutoff of 2), whereas the lower border of the symmetrical confidence interval (lower diamond) $\leqslant -1$ (-1.58) indicates higher expression in male patients. (B) The upper border of the symmetrical confidence interval (upper diamond) $\geqslant 1$ (1.58) indicates higher expression in probably benign, uncertain, or low-malignant potential GISTs, whereas the lower border of the symmetrical confidence interval (lower diamond) $\leqslant -1$ (-1.58) indicates higher expression in probably malignant GISTs (*B2M* and *HPRT1* only when using a cutoff of 2 but also *LMNB1* and *SDHA* when using a cutoff of 3).

HPRT1 were confirmed to be equivalently expressed as well. The normalized expression of the target gene *MMP2* was not significantly different among probably benign, uncertain or low-malignant, and probably malignant GISTs, without any differences between normalization to confirmed reference genes or unconfirmed reference genes.

Discussion

In this study, we have determined the expression of five commonly used reference genes [17,28,29,31,32] by qRT–PCR. We evaluated how a variable expression of a reference gene itself influenced the normalized expression of a target gene. Comparing ccRCC with normal renal parenchyma, the higher expression of the reference gene *LMNB1* in ccRCC led to the misinterpretation of higher expression of the target gene *TP53* in normal renal parenchyma, whereas there was no significant differential expression of *TP53* between ccRCC and normal renal parenchyma after normalization to the confirmed reference gene *HPRT1*. Comparing colon carcinoma with normal colon mucosa, a tendency of the target gene *TP53* for a higher expression in colon carcinoma was masked by a higher expression of the unconfirmed reference gene *HPRT1* in colon carcinoma itself. These results demonstrate how use of unconfirmed reference genes results in misinterpretation of the expression of the target genes.

In the case of *LMNB1* in ccRCC, the equivalence test, but not the *t* test, detected the variable expression of the reference gene associated with the examined tissue. This observation is due to a mathematical fact widely unregarded [30]. The classical two-sided *t* test is constructed to detect a significant difference. This test is appropriate for the identification of differentially expressed target genes. However, for the validation of stably expressed reference genes, this test is not appropriate. The conclusion that a reference gene that is not significantly differentially expressed is equally expressed is not necessarily correct. The risk of getting false negatives (type 2 error, i.e., acceptance of not differentially expressed genes as suitable reference genes even though they are not equivalently expressed) is not controllable. A comparably simple approach to solve this problem is the equivalence test [30]. Based on the inclusion of a confidence interval, the null hypothesis and the alternative are exchanged (compared with the *t* test). Therefore, the user can control the type 1 error for incorrect rejection of the null hypothesis—"different distribution," which is the adequate approach for the validation of a reference gene.

Depending on the character of the examined data, the appropriate version of the equivalence test must be used. In this article, we have introduced two different parametric versions for dependent and independent samples that were applicable to our approximately normally distributed quantitative RT–PCR data. However, nonparametric versions also exist but may be of less power and may further carry a higher risk of confirming reference genes with spurious outliers.

The decision of the equivalence test depends on the cutoff for the inclusion of the confidence interval, which is hard to justify with a mathematical approach. Instead, the cutoff must be adjusted carefully to the distribution of each specific experiment by the investigator himself or herself. A too stringent cutoff would exclude stably expressed reference genes (e.g., due to the variance of the data), whereas a too loose cutoff would include unsuitable genes. The variance of the measured expression of a reference gene is composed of methodical (e.g., RNA quantification and quality, efficiency of cDNA synthesis, pipetting error) and true biological variance. The different parts can hardly be distinguished quantitatively and may vary between different experimental setups or laboratories. In fact, the major advantage of using stably expressed reference genes with minimal biological variance for normalization of target genes is to reduce the variance of methodical errors/differences between different samples (e.g., RNA quantification and quality, efficiency of cDNA synthesis) that affect both the target and the reference gene on the same quantitative scale.

The evaluation of fold changes of 2 and 3 implicated that a cutoff of a fold change of 2 might yet be too stringent. Comparing ccRCC with normal renal parenchyma using a cutoff of 2, none of the five tested reference genes could be confirmed. According to our observations, a threefold change appeared to be a useful cutoff. In the experiment where no biological variance was expected (gender-associated comparison of GISTs), all five tested reference genes appeared to be equivalently expressed. On the other hand, no reference gene that was significantly differentially expressed, as detected by the *t* test, was false-positively stated as equivalently expressed. However, from a mathematical point of view, we cannot confirm a threefold change to be the definite cutoff for equivalent expression of suitable reference genes in qRT–PCR. Instead, we propose that the fold change of a significant differential expression of a target gene after normalization to a confirmed reference gene should be at least the fold change for the inclusion of the confidence interval for equivalent expression of the used reference gene. If several possible reference genes are confirmed by the equivalence test, the one with a minimum of variability in the confidence interval should be chosen. As proposed by Vandesompele et al. [29], normalization to the geometric average of multiple reference genes has the advantage of being less susceptible to differential expression of single reference genes in individual samples. However, this approach cannot identify a unidirectional affection of several reference

genes. A unidirectional affection in four of five tested reference genes could be shown for GISTs, where *B2M, HPRT1, LMNB1*, and *SDHA* all were higher expressed in probably malignant tumors. In such situations, a single validation of each reference gene with the equivalence test should be done in advance.

We were able to demonstrate that a stable expression of a reference gene in one tumor type does not predict a stable expression in another tumor type. According to our results, *HPRT1*, but not *18S*, was a suitable reference gene for the comparison of ccRCC with renal parenchyma, whereas it was the other way around for the comparison of colon carcinoma with colon mucosa. In addition, we found that even within one tumor type, the expression of a reference gene was not stable for different biological groupwise comparisons. *LMNB1* was a suitable reference gene for gender-associated comparisons in GISTs but not for comparisons regarding estimated malignant potential. These observations confirm that there is no universal reference gene and underline the importance of specific validation of potential reference genes for any experimental condition.

References

[1] R. Higuchi, C. Fockler, G. Dollinger, R. Watson, Kinetic PCR analysis: real-time monitoring of DNA amplification reactions, Biotechnology 11 (1993) 1026–1030.

[2] C.A. Heid, J. Stevens, K.J. Livak, P.M. Williams, Real time quantitative PCR, Genome Res. 6 (1996) 986–994.

[3] S.A. Bustin, Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays, J. Mol. Endocrinol. 25 (2000) 169–193.

[4] P. Bhatia, W.R. Taylor, A.H. Greenberg, J.A. Wright, Comparison of glyceraldehyde-3-phosphate dehydrogenase and 28S-ribosomal RNA gene expression as RNA loading controls for Northern blot analysis of cell lines of varying malignant potential, Anal. Biochem. 216 (1994) 223–226.

[5] H. Zhong, J.W. Simons, Direct comparison of GAPDH, beta-actin, cyclophilin, and 28S rRNA as internal standards for quantifying RNA levels under hypoxia, Biochem. Biophys. Res. Commun. 259 (1999) 523–526.

[6] S.A. Bustin, Quantification of mRNA using real-time reverse transcription PCR (RT–PCR): trends and problems, J. Mol. Endocrinol. 29 (2002) 23–39.

[7] L. Fink, W. Seeger, L. Ermert, J. Hanze, U. Stahl, F. Grimminger, W. Kummer, R.M. Bohle, Real-time quantitative RT–PCR after laser-assisted cell picking, Nat. Med. 4 (1998) 1329–1333.

[8] W.H. Karge, III, E.J. Schaefer, J.M. Ordovas, Quantification of mRNA by polymerase chain reaction (PCR) using an internal standard and a nonradioactive detection method, Meth. Mol. Biol. 110 (1998) 43–61.

[9] T. Suzuki, P.J. Higgins, D.R. Crawford, Control selection for RNA quantitation, Biotechniques 29 (2000) 332–337.

[10] M.W. Pfaffl, A new mathematical model for relative quantification in real-time RT–PCR, Nucleic Acids Res. 29 (2001) e45.

[11] G. Haberhausen, J. Pinsl, C.C. Kuhn, C. Markert-Hahn, Comparative study of different standardization concepts in quantitative competitive reverse transcription–PCR assays, J. Clin. Microbiol. 36 (1998) 628–633.

[12] O. Thellin, W. Zorzi, B. Lakaye, B. De Borman, B. Coumans, G. Hennen, T. Grisar, A. Igout, E. Heinen, Housekeeping genes as internal standards: use and limits, J. Biotechnol. 75 (1999) 291–295.

[13] C.J. Gerard, L.M. Andrejka, R.A. Macina, Mitochondrial ATP synthase 6 as an endogenous control in the quantitative RT–PCR analysis of clinical cancer samples, Mol. Diagn. 5 (2000) 39–46.

[14] S. Selvey, E.W. Thompson, K. Matthaei, R.A. Lea, M.G. Irving, L.R. Griffiths, Beta-actin: an unsuitable internal control for RT–PCR, Mol. Cell Probes 15 (2001) 307–311.

[15] D. Goidin, A. Mamessier, M.J. Staquet, D. Schmitt, O. Berthier-Vergnes, Ribosomal 18S RNA prevails over glyceraldehyde-3-phosphate dehydrogenase and beta-actin genes as internal standard for quantitative comparison of mRNA levels in invasive and noninvasive human melanoma cell subpopulations, Anal. Biochem. 295 (2001) 17–21.

[16] H.K. Hamalainen, J.C. Tubman, S. Vikman, T. Kyrola, E. Ylikoski, J.A. Warrington, R. Lahesmaa, Identification and validation of endogenous reference genes for expression profiling of T helper cell differentiation by quantitative real-time RT–PCR, Anal. Biochem. 299 (2001) 63–70.

[17] C. Blanquicett, M.R. Johnson, M. Heslin, R.B. Diasio, Housekeeping gene variability in normal and carcinomatous colorectal and liver tissues: applications in pharmacogenomic gene expression studies, Anal. Biochem. 303 (2002) 209–214.

[18] E.M. Glare, M. Divjak, M.J. Bailey, E.H. Walters, Beta-Actin and GAPDH housekeeping gene expression in asthmatic airways is variable and not suitable for normalising mRNA levels, Thorax 57 (2002) 765–770.

[19] B.K. Steele, C. Meyers, M.A. Ozbun, Variable expression of some ''housekeeping'' genes during human keratinocyte differentiation, Anal. Biochem. 307 (2002) 341–347.

[20] C. Tricarico, P. Pinzani, S. Bianchi, M. Paglierani, V. Distante, M. Pazzagli, S.A. Bustin, C. Orlando, Quantitative real-time reverse transcription polymerase chain reaction: normalization to rRNA or single housekeeping genes is inappropriate for human tissue biopsies, Anal. Biochem. 309 (2002) 293–300.

[21] T.D. Schmittgen, B.A. Zakrajsek, Effect of experimental treatment on housekeeping gene expression: validation by real-time, quantitative RT–PCR, J. Biochem. Biophys. Meth. 46 (2000) 69–81.

[22] K. Gorzelniak, J. Janke, S. Engeli, A.M. Sharma, Validation of endogenous controls for gene expression studies in human adipocytes and preadipocytes, Horm. Metab. Res. 33 (2001) 625–627.

[23] J. Lupberger, K.A. Kreuzer, G. Baskaynak, U.R. Peters, P. le Coutre, C.A. Schmidt, Quantitative analysis of beta-actin, beta-2-microglobulin, and porphobilinogen deaminase mRNA and their comparison as control transcripts for RT–PCR, Mol. Cell Probes 16 (2002) 25–30.

[24] R.M. Murphy, K.K. Watt, D. Cameron-Smith, C.J. Gibbons, R.J. Snow, Effects of creatine supplementation on housekeeping genes in human skeletal muscle using real-time RT–PCR, Physiol. Genom. 12 (2003) 163–174.

[25] H.L. Sobin, C. Wittekind, TNM Classification of Malignant Tumours, sixth ed., John Wiley, New York, 2003.

[26] M. Miettinen, W. El Rifai, H.L. Sobin, J. Lasota, Evaluation of malignancy and prognosis of gastrointestinal stromal tumors: a review, Hum. Pathol. 33 (2002) 478–483.

[27] S. Rozen, H. Skaletsky, Primer3 on the WWW for general users and for biologist programmers, Meth. Mol. Biol. 132 (2000) 365–386.

[28] W.J. Kim, Y. Kakehi, H. Kinoshita, S. Arao, M. Fukumoto, O. Yoshida, Expression patterns of multidrug-resistance (MDR1), multidrug resistance-associated protein (MRP), glutathione-*S*-transferase-pi (GST-pi), and DNA topoisomerase II (Topo II)

genes in renal cell carcinomas and normal kidney, J. Urol. 156 (1996) 506–511.

[29] J. Vandesompele, K. De Preter, F. Pattyn, B. Poppe, N. Van Roy, A. De Paepe, F. Speleman, Accurate normalization of real-time quantitative RT–PCR data by geometric averaging of multiple internal control genes, Genome Biol. 3 (2002) 34.

[30] S. Wellek, Testing Statistical Hypothesis of Equivalence, Chapman & Hall, London, 2003.

[31] N. Tsuji, C. Kamagata, M. Furuya, D. Kobayashi, A. Yagihashi, T. Morita, S. Horita, N. Watanabe, Selection of an internal control gene for quantitation of mRNA in colonic tissues, Anticancer Res. 22 (2002) 4173–4178.

[32] N. Koon, R. Schneider-Stock, M. Sarlomo-Rikala, J. Lasota, M. Smolkin, G. Petroni, A. Zaika, C. Boltze, F. Meyer, L. Andersson, S. Knuutila, M. Miettinen, W. El Rifai, Molecular targets for tumour progression in gastrointestinal stromal tumours, Gut 53 (2004) 235–240.