Translational Oncology
www.transonc.com

Volume 7 Number 6 December 2014 pp. 687–693 687

The Importance of Reference Gene Analysis of Formalin-Fixed, Paraffin-Embedded Samples from Sarcoma Patients — An Often Underestimated Problem[1,2]

Ninna Aggerholm-Pedersen*, Akmal Safwat†, Steen Bærentzen‡, Marianne Nordsmark†, Ole Steen Nielsen§, Jan Alsner* and Brita S. Sørensen*

*Department of Experimental Clinical Oncology, Aarhus University Hospital, Aarhus, Denmark; †Department of Oncology, Aarhus University Hospital, Aarhus, Denmark; ‡Department of Pathology, Aarhus University Hospital, Aarhus, Denmark; §Faculty of Health, Aarhus University, Aarhus, Denmark

Abstract

Objective: Reverse transcription quantitative real-time polymerase chain reaction is efficient for quantification of gene expression, but the choice of reference genes is of paramount importance as it is essential for correct interpretation of data. This is complicated by the fact that the materials often available are routinely collected formalin-fixed, paraffin-embedded (FFPE) samples in which the mRNA is known to be highly degraded. The purpose of this study was to investigate 22 potential reference genes in sarcoma FFPE samples and to study the variation in expression level within different samples taken from the same tumor and between different histologic types. *Methods:* Twenty-nine patients treated for sarcoma were enrolled. The samples encompassed 82 (FFPE) specimens. Extraction of total RNA from 7-μm FFPE sections was performed using a fully automated, bead-base RNA isolation procedure, and 22 potential reference genes were analyzed by reverse transcription quantitative real-time polymerase chain reaction. The stability of the genes was analyzed by RealTime Statminer. The intrasamples variation and the interclass correlation coefficients were calculated. The linear regression model was used to calculate the degradation of the mRNA over time. *Results:* The quality of RNA was sufficient for analysis in 84% of the samples. Recommended reference genes differed with histologic types. However, *PPIA, SF3A1*, and *MRPL19* were stably expressed regardless of the histologic type included. The variation in Cq value for samples from the same patients was similar to the variation between patients. It was possible to compensate for the time-dependent degradation of the mRNA when normalization was made using the selected reference genes. *Conclusion: PPIA, SF3A1*, and *MRPL19* are suitable reference genes for normalization in gene expression studies of FFPE samples from sarcoma regardless of the histology.

Translational Oncology (2014) 7, 687–693

Introduction

Sarcoma is a heterogeneous group of rare tumors whose resistance to radiotherapy, chemotherapy, and molecularly targeted therapies is a major problem in clinical practice hindering major breakthroughs in outcome in the past decades [1].

New biological knowledge such as the discovery of the important role played by cancer stem cells, genetics, and epigenetic changes on resistance to treatment and eventual prognosis has shown that factors other than those traditionally measured and tested can be detrimental to treatment outcome. Gene expression analysis is increasingly important to understand tumor initiation/progression, giving insight

Address all correspondence to: Ninna Aggerholm-Pedersen, MD, MSc, PhD student, Department of Experimental Clinical Oncology, Aarhus University Hospital, Norrebrogade 44, 8000 Aarhus C, Denmark.
E-mail: aggerholm@oncology.dk

[1] Support: The work was supported by the A. P. Møller and Christine McKinney Møller Foundation.
[2] Conflicts of interest: None declared.
Received 14 August 2014; Revised 24 September 2014; Accepted 26 September 2014

1936-5233/14
http://dx.doi.org/10.1016/j.tranon.2014.09.012

into complex regulatory networks and uncovering new biological processes involved in tumorigenesis.

Reverse transcription quantitative real-time polymerase chain reaction (RT-qPCR) is a powerful technique which can detect low levels of mRNA. RT-qPCR is a fast and efficient technique for quantification of the gene expression levels [2]. However, the results are dependent on appropriate selection of normalization factors to account for errors and differences generated by the multistep process. Variables such as the differences between tissues in overall transcriptional activity need to be controlled in gene expression analysis. The use of internal control (reference) genes is the most commonly used method. Reference genes should be constitutively expressed and should have a stable expression under the experimental conditions and tissue preparation techniques. The expression of internal reference genes varies significantly under different experimental conditions. Therefore, erroneous information could be generated if normalization is based on genes that themselves are regulated or affected by the preparation of tissue samples [3,4].

The clinical material often available for analysis from sarcoma patients is routinely collected formalin-fixed, paraffin-embedded (FFPE) samples processed under different conditions. Factors such as time to fixation, fixation time, and others may contribute to variations in gene expression levels by introducing RNA degradation [5,6]. It is also well known that mRNA from FFPE tissue specimens is poorly preserved and highly degraded [7]. In rare cancer types such as sarcoma, the FFPE tissue samples are indispensable and invaluable resources for studying gene expression levels. Other studies have shown that using RNA from FFPE tissue samples is feasible [8–10].

Due to these issues, candidate reference genes must be validated in defined experimental conditions. The purpose of the study was to investigate 22 potential reference genes in FFPE tissue samples with various histology types of sarcoma and to investigate the variation within different samples taken from the same tumor.

Material and Methods

Patient Samples

Twenty-nine patients treated for sarcoma at the Sarcoma Centre of Aarhus University Hospital between 1991 and 2001 were enrolled in this study. The selected patients' characteristics are presented in Table 1. The tissue samples represented surgical specimens, and for most patients, more than one tissue sample from different locations in the tumor was analyzed. The tissue samples encompassed 82 FFPE tissue samples (between 1 and 6 per patients, with a median of 3 samples per patient) of different histologic types: pleomorphic undifferentiated sarcoma (PUS; n = 10), liposarcoma (n = 8), leiomyosarcoma (n = 2), malignant peripheral nerve sheath tumor (MPNST, n = 2), synovial sarcoma (n = 2), osteosarcoma (n = 1), Ewing (n = 1), and fibrosarcoma (n = 1). No information about the time from sampling to fixation or the fixation time was available. The tumor tissue samples were chosen based on the pathology report. The tissue sample was excluded from the analysis if more than one gene was not detectable or if the sample was 100% necrotic. One patient was excluded because of 100% normal tissue in all samples. This resulted in excluding 18 samples, and the analysis was done on 64 tissue samples from 25 patients (Table 1).

RNA Extraction and cDNA Synthesis

Extraction of total RNA from 7-μm FFPE sections from tissue samples and empty paraffin blocks was performed by the Tissue Preparation System with VERSANT Tissue Preparation Reagents. This is a fully automated, bead-based RNA isolation method

Table 1. Sample Characteristics

Patient	Histology	Time	Age	Samples	No. of Genes	% Tumor	% Necrosis
(No.)		(year of biopsy)		(No.)	not detected *		
1	Ewing	1999	24	2	0/0	90/90	0/0
2	Fibrosarcoma	1994	36	1	0	100	0
3	Leiomyosarcoma	1995	66	3	0/0/0	100/100/100	0/0/0
4	Leiomyosarcoma	1994	53	3	0/1/6	100/90/80	0/0/20
5	Liposarcoma	1999	69	3	0/0/0	30/30/0	60/60/100
6	Liposarcoma	2001	54	3	0/0/0	80/80/100	0/0/0
7	Liposarcoma	1993	80	3	0/0/0	100/20/30	0/80/70
8	Liposarcoma	1994	45	3	8/0/8	20/100/100	80/0/0
9	Liposarcoma	2000	64	3	0/0/0	0/0/0	0/0/0
10	Liposarcoma	2001	56	3	0/0/0	50/50/60	0/0/30
11	Liposarcoma	1994	77	3	0/1/0	90/80/40	10/20/60
12	Liposarcoma	1995	77	3	0/0/0	100/90/100	0/10/0
13	PUS	1993	74	3	19/0/1	0/0/0	100/80/0
14	PUS	1998	64	3	0/0/0	20/20/60	0/0/0
15	PUS	1993	73	3	0/0/0	100/100/60	0/0/0
16	PUS	1993	53	1	0	90	10
17	PUS	1994	76	4	0/0/0/0	85/70/70/70	10/0/20/10
18	PUS	1993	67	3	0/0/0	75/30/40	30/60/60
19	PUS	1993	64	6	6/0/0/0/0/0	100/50/100/95/100	0/50/0/0/0
20	PUS	1993	43	3	0/0/0	80/60/0	20/40/0
21	PUS	1993	87	2	0/1	60/40	30/50
22	PUS	1993	63	3	0/0/0	100/100/90	0/0/10
23	PUS	1993	81	3	15/6/17	10/10/10	90/90/90
23	Osteosarcoma	1994	54	2	0/0	80/90	0/0
24	MPNST	2000	51	3	0/0/8	90/100/0	10/0/0
25	MPNST	1998	37	3	0/0/0	90/80/20	0/10/80
26	Synovial sarcoma	1998	50	3	0/0/0	95/60/60	5/0/10
27	Synovial sarcoma	1991	45	3	0/0/0	100/60/70	0/40/30
29	Synovial sarcoma	1997	33	1	9	90	

* The following genes were not included in the analysis: *HMBS*, *HPRT1*, *TFRC*, and *PSMC4*.

(Siemens Healthcare Diagnostics, Tarrytown, NY) as described elsewhere [11]. Previous data with similar FFPE tissue samples have shown that although the RNA integrity number values are very low due to highly degraded RNA [7], it is still possible to obtain reproducible qPCR data. Furthermore, the RNA quantity could not be measured on a spectrophotometer because of the elution buffer from the VERSANT Tissue preparation reagents kit (Siemens) which absorbs light at the same wavelength as RNA. However, high-quality and reproducible qPCR data can be obtained using short amplicons (<100 bp) and a preamplification step [7,12]. The data from the RT-qPCR were used as an internal quality control, with the quantity and the quality of the RNA reflected in the Cq values of the stable expressed reference genes. The extraction included DNase I treatment to remove potential genomic DNA according to the manufacturer's protocol. cDNA was generated using the High Capacity cDNA Archive Kit (Applied Biosystems [ABI], Foster City, CA) according to the manufacturer's protocol. The RNA samples were used within 2 months after preparation and were stored at −70°C until use. The cDNA were analyzed immediately after preparation.

Primers and RT- qRCR

Before RT-qPCR, 10 cycles of preamplification were performed according to the manufacturer's protocol (Taqman Preamp Master Mix, ABI) with modification as described elsewhere [7]. The reference genes were selected from commonly used reference genes and from the literature with the following search terms using the PubMed database: ("PCR" and "reference" and "sarcoma") or ("genes"[Mesh] and "Sarcoma"[Mesh] and "reference") or ("aPCR"[Mesh] and "Sarcoma"[Mesh]) or ("qPCR" and "sarcoma"[Mesh]) or ("qPCR" and reference" and "genes") or ("qPCR" and "housekeeping" and "genes"). A further selection was done by excluding all non–cancer-related publications. The reference genes selected for further analysis are shown in Table 2. Assays were selected with an exon junction spanning probe and if possible purchased from ABI. For CALM2 and RPL37A, no ABI assays were suitable, and specifically designed assays were supplied by DNA Technology (Aarhus, Denmark). Assays with amplicons between 60 and 100 bp were chosen (assay numbers or primers and probes are shown in Table 2). The experiment was performed on an ABI PRISM 7900HT in 384-well plates assembled by Biomek 3000 (Beckman Coulter). The qPCR reaction consisted of 3 μl cDNA, 0.75 μl 20× TaqMan Assay, and 7.5 μl of Taqman Gene Expression Mastermix (ABI), in a total volume of 15 μl. The cDNA was not quantified as we used the results of the RT-qPCR as quality control. The experiment was initiated by one cycle at 95°C for 10 minutes, followed by 40 cycles of 95°C for 15 seconds and 60°C for 1 minute. Samples were loaded in duplicate along with negative controls. Negative controls consisted of both samples from empty paraffin blokes and no template controls. All negative controls where run on each plate. No signal was detected. Positive controls were included consisting of pooled RNA from cell lines. Cq values with a standard deviation above 0.3 between duplicated samples were dismissed, and the RT-qPCR were repeated.

Data Analysis and Statistics

The stability of the reference genes was analyzed using the RealTime Statminer (Integromics, Madison, WI), version 20. Cq values above 35 were regarded as below detection limit. Thresholds were set manually in the SDS2.1 software. The GeNorm algorithm calculates an internal control gene-stability measure M. A low value of M means a high relative stability. GeNorm includes the top-ranked genes as reference genes, and the recommended number of reference genes depends on the pairwise variation as described by Vandesompele et al. A pairwise variation under 0.15 was chosen as previously described [13]. NormFinder collects two genes as reference genes and not necessarily the gene with the best stability scores. The algorithm for NormFinder has been described by Andersen et al. [14]. Table 3 shows the M score calculated by GeNorm and the stability score calculated by NormFinder. For both algorithms, the genes were ranked, and a total rank was established based on all the different analysis entities when analyzing all tissue together, liposarcoma and MHF tissue samples together, and liposarcoma tissue samples and PUS tissue

Table 2. Summary of the Candidate Reference Genes and Function

Gene Symbol	Gene Name	Gene Function	Assay ID *	Amplicon
ACTB	Actin, βActin, β	Structural protein (cytoskeletal)	Hs01060665_g1	63
ACTR3	ARP3 actin-related protin 3 homolog	Major constituent of the ARP2/3 complex	Hs01029161_m1	72
B2M	β-2-microglobulin	Beta-chain of major histocompatibility complex class I molecules	Hs00984230_m1	80
CALM2	Calmodulin 2	Calcium-binding messenger protein	**	72
CHCHD1	Coiled-coil-helix domain containing 1	Nuclear protein	Hs00415054_m1	79
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase	Glycolysis	Hs02758991_g1	97
GUSB	Glucuronidase, β	Degradation of dermatan and keratan sulfates	Part no 4333767F	81
HMBS	Hydroxymethylbilane synthase	Heme synthesis, porphyrin metabolism	Hs00609293_g1	62
HPRT1	Hypozanthine phosphoribosyl transferase	Purine salvage	Part no 4333768T	100
IPO8	Importin 8	Nuclear protein import	Hs00183533_m1	71
MRPL19	Mitochondrial ribosomal protein L19	Component of the large ribosomal subunit	Hs00608519_m1	72
NDFIP1	Nedd4 family interacting protein 1	Integral Golgi membrane proteins	Hs00228968_m1	67
POLR2A	RNA polymerase II	Catalyzes the RNA synthesis	Hs01108291_m1	86
PPIA	Peptidylpropyl isomerase A	*cis-trans* isomerization of proline imidic peptid bonds	Part no. 4333763F	98
PSMC4	Proteasome 26S subunit	ATP-dependent degradation of ubiquitinated protien	Hs00197826_m1	83
PUM1	Pumilio homolog 1	translational regulation of mRNA	Hs00472881_m1	77
RPL37A	Ribosomal protein L37a	Protein synthesis	***	65
RPLP0	Ribosomal protein, large, P0	Structural protein of ribosomes	Hs00420895_gH	76
rRNA18s	18S ribosomal RNA	Structural RNA for the small component of ribosomes	Hs03928990_g1	61
SF3A1	Splicing factor 3a, subunit 1	pre-mRNA splicing	Hs01066327_m1	64
TBP	TATA box binding protin	RNA prolymerase II transcription factor	Hs00427621_m1	65
TFRC	Transferrin receptor	Uptake of iron	Hs00951083_m1	66

* Supplied by ABI.

** Primer (forward/reverse) GAGCGAGCTGAGTGGTTGTG/AGTCAGTTGGTCAGCCATGCT probe: TCGCGTCTCGGAAACCGGAGC, supplied by DNA Technology.

*** Primer (forward/reverse) TGTGGTTCCTGCATGAAGACA/GTGACAGCGGAAGTGGTATTGTAC probe:TGGCTGGCGGTGCCTGGA, supplied by DNA Technology.

Table 3. Ranking of the Genes by Tissue Type and Algorithm

Gene Symbol	All Tissue Samples				Liposarcoma and MFH Together				Liposarcoma				MFH				Overall
	M Value*	Rank	S Score	Rank	M Value	Rank	S Score	Rank	M Value	Rank	S Score	Rank	M Value	Rank	S Score	Rank	Rank
ACTB	0.89	7	0.39	5	0.89	11	0.22	11	0.86	13	0.49	12	**0.47**	1	0.36	7	8
ACTR3	1.41	14	0.55	14	0.99	14	0.18	8	0.86	12	0.43	10	0.76	10	0.40	11	12
B2M	1.41	15	0.68	18	1.22	16	0.22	13	1.24	17	0.81	18	1.11	19	0.61	19	17
CALM2	**0.80**	3	0.25	1	0.79	5	0.13	3	0.67	6	0.38	6	0.71	9	0.40	10	4
CHCHD1	0.83	5	0.41	7	0.80	6	0.16	5	0.63	5	0.40	8	0.72	8	0.34	4	6
GAPDH	1.61	17	0.61	16	1.50	18	0.29	17	1.58	18	0.95	19	1.07	17	0.48	14	18
GUSB	1.70	18	0.49	11	0.88	10	0.19	10	2.14	19	0.76	17	0.95	15	0.54	16	14
IPO8	**0.81**	4	0.49	12	**0.75**	4	0.11	1	0.61	4	0.25	2	0.85	12	0.50	15	7
MRPL19 *	**0.75**	1	0.34	4	0.81	7	0.14	4	**0.39**	1	0.36	4	**0.55**	3	0.29	2	1
NDFIP1	1.28	13	0.44	8	1.36	17	0.31	18	1.20	16	0.65	16	0.69	7	**0.30**	3	13
POLR2A	1.18	11	0.49	10	0.87	9	0.13	2	0.82	11	0.47	11	0.91	14	0.44	13	11
PPIA *	0.86	6	**0.29**	2	**0.63**	1	0.17	7	0.73	8	**0.40**	7	0.64	4	0.36	6	3
PUM1	1.55	16	0.60	15	0.94	12	0.22	12	0.93	14	0.57	15	0.98	16	0.57	17	15
RPL37A	1.08	9	0.49	13	0.97	13	0.28	15	**0.39**	2	**0.37**	5	0.78	11	0.38	9	10
RPLP0	1.00	8	0.46	9	0.87	8	**0.17**	6	0.79	10	0.42	9	0.88	13	0.43	12	9
rRNA18s	1.21	12	0.64	17	1.15	15	0.28	16	1.00	15	0.56	14	1.07	18	0.61	18	16
SF3A1 *	1.16	10	0.39	6	**0.63**	2	0.18	9	**0.64**	3	0.23	1	**0.47**	2	**0.27**	1	2
TBP	**0.75**	2	**0.33**	3	**0.73**	3	**0.25**	14	0.71	7	0.35	3	0.66	6	0.37	8	5

The bold and underlined values are the recommended reference genes by the two different algorithms GeNorm and NormFinder when analyzing all tissue samples, liposarcoma and PUS together, and liposarcoma and PUS separately. From these results, an overall ranking was determined, and the three best ranking genes were selected as reference genes.

* When considering the total rank of the different types analyzed, these 3 genes have the lowest score.

samples as separate groups. The three best ranked genes were used as reference genes.

ΔCq values were generated by normalizing the geometric mean of the most stable reference genes. ΔCq values were calculated in RealTime Statminer version 20.

The intrasamples variation (the standard deviation of single measurements) was calculated and reported with a 95% confidence interval (CI) for each gene. The highest and lowest ΔCq values were used, and only patients with two or more samples were included. The interclass correlation coefficients were calculated and reported with a 95% CI for each gene to determine the intra and inter variability from each sample. The assumption of the error following a normal distribution was met for all three genes analyzed. The linear regression model was used to calculate the degradation of the mRNA over time. All statistical analysis was performed by using Stata version 12.

Ethics

The study has been approved by the Ethics Committee of Denmark (no. 1-10-72-233-12) and the Danish Agency of Data Protection (no. 2012-41-0657).

Results

Of the 82 tissue samples from 29 sarcoma patients, 18 tissue samples were excluded due to more than one missing value in the qPCR analysis. This meant that the potential reference genes could be analyzed in 78% of all tissue samples (64 samples from 25 patients). The Cq value for *PSMC4, HPRT1, HMBS,* and *TFRC* were below the detection limit in 15%, 22%, 23%, and 54% of the analyzed tissue samples, respectively. These genes were therefore excluded from the analysis. Table 1 shows the different histologic types and year of biopsy, age of the patient, how many samples from each tumor were collected, and the number of genes not detected in each tissue samples. The last two columns show the percentage of tumor tissue and necrotic tissue in each sample.

To establish which reference was the most stable, the 18 detectable reference genes were analyzed using both the GeNorm and NormFinder algorithm. When all tissue samples were analyzed together, the reference genes found to be the most stable genes were *MRPL19, CALM2 IPO8, TBP* and *TBP, PPIA* for GeNorm and NormFinder, respectively. When analyzing liposarcoma and PUS tissue samples together, the most stable reference genes were *TBP, PPIA, SF3A1, IPO8* and *TBP, RPLPO* for GeNorm and NormFinder, respectively. When analyzing PUS tissue samples alone, the most stable genes were *SF3A1, ACTB, MRPL19* and *NDFIP1, SF3A1* for GeNorm and NormFinder, respectively. When analyzing liposarcoma tissue samples alone, the most stable reference genes were *SF3A1 RPL7A, MRPL19* and *PPIA, RPL37A* GeNorm and NormFinder, respectively. The ranking of the genes is shown in Table 3, and the top-ranking genes, regardless of the algorithm used and tissue included in the analysis, were *MRPL19, PPIA*, and *SF3A1*. The mean Cq value for these three reference genes was 26.9 (95% CI: 26.7 to 27.1). The least stable genes regardless of tissue type and algorithm were *B2M, GAPDH, GUSB, PUM1*, and *rRNA18s*.

The ΔCq values were calculated using the most stable genes *MRPL19, PPIA*, and *SF3A1* as reference genes for the same three genes to establish whether or not there was a difference in the ΔCq value between different tissue samples taken from the same patient (Figure 1). The measurement error or the standard deviation of single measurements (ΔCq value) was 0.61 (95% CI: 0.47 to 0.86), 0.33 (95% CI: 0.26 to 0.47), and 0.44 (95% CI: 0.34 to 0.61) for *PPIA, SF3A1,* and *MRPL19,* respectively. The intraclass correlation coefficients were calculated to find how strongly samples from the same patients resembled each other and show the proportion of total variance within each patient. The intraclass correlation coefficients were 0.57 (95% CI: 0.34 to 0.79), 0.61 (95% CI: 0.40 to 0.82), and 0.47 (95% CI: 0.21 to 0.72) for *PPIA, SF3A1*, and *MRPL19,* respectively. Supplementary Figure 1 shows the ΔCq value for each patient against sample number. Each line represents one patient, according to the investigated genes and histologic type of the patients. Supplementary Figure 2 shows the Bland-Altman plot of difference versus average, comparing the highest and lowest ΔCq value in each patient. When analyzing liposarcoma alone, the measurement error of the ΔCq value for each individual was 0.38 (95% CI: 0.25 to 0.78), 0.25 (95% CI: 0.17 to 0.52), and 0.29 (95% CI: 0.19 to 0.58) for *PPIA, SF3A1*, and *MRPL19,* respectively, and for PUS, 0.76 (95%

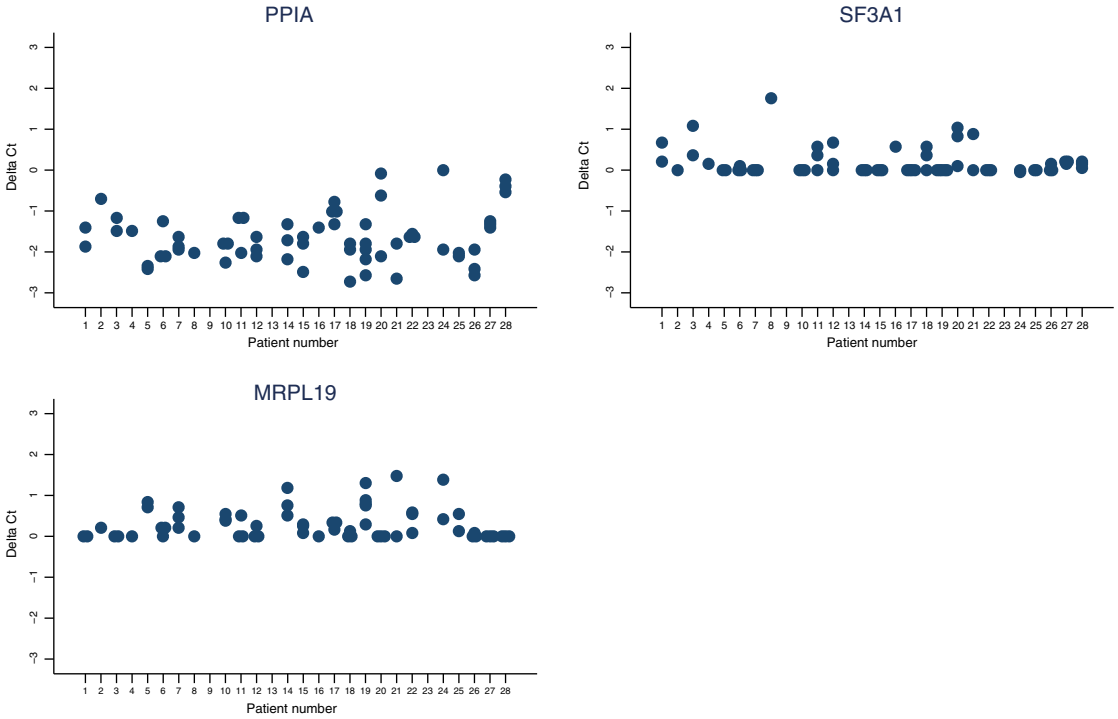

Figure 1. Cq value for the three most stably expressed genes for each patient. The Cq values were normalized to the geometric mean of the same three genes. The patient numbers are represented at the *y*-axis. The histologic type for patient number 1 was Ewing sarcoma; for patient number 2, fibrosarcoma; for patient numbers 3 to 4, leiomyosarcoma; for patient numbers 5 to 12, liposarcoma; for patient numbers 13 to 23, PUS; for patient number 24, osteosarcoma; for patient numbers 25 and 26, MPNST; and for patient numbers 27 to 29, synovial sarcoma. Four patients were not included as explained in Material and Methods.

CI: 0.52 to 1.39), 0.43 (95% CI: 0.30 to 0.79), and 0.60 (95% CI: 0.41 to 1.10) for *PPIA, SF3A1*, and *MRPL19*, respectively. Figure 2 shows the variation for all potential reference genes when all samples were normalized to the three most stable genes. The outliner values seen for each gene represent different patients.

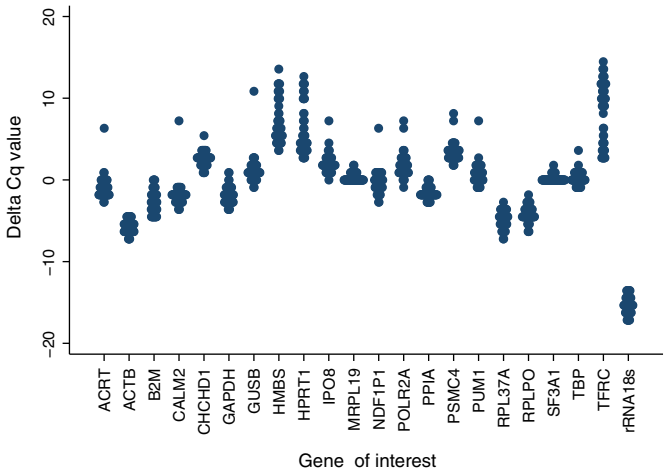

Figure 2. Median centered Cq values of the 22 studied reference genes, normalized to the three highest-ranking reference genes. All histologic types are represented in this figure. The number of tissue samples for each gene is 64. The *y*-axis represents each gene analyzed.

The raw Cq values were used to establish the degradation rate of the three most stable reference genes. For *PPIA, SF3A1*, and *MRPL19*, the degradation was −0.37 Cq/year (95% CI: −0.48 to −0.27), −0.34 (95% CI: −0.44 to −0.23), and −0.32 (95% CI: −0.43 to −0.23). There was no difference in the slope of the line $P = .84$. This gives an overall half-life of the mRNA of 2.9 years (95% CI: 2.4 to 3.6) (Figure 3a). When normalization of the three most stable genes to the geometric mean of the same three genes, the degradation was −0.03 (95% CI: −0.09 to 0.03), 0.00 (95% CI: −0.03 to 0.03), and 0.00 (95% CI: −0.04 to 0.03) for *PPIA, SF3A1*, and *MRPL19*, respectively (Figure 3b).

Discussion

Few studies have addressed the issue of identifying steadily expressed reference genes in sarcoma samples, and to the best of the authors' knowledge, this is the first article to address the problem in FFPE tissue samples from sarcoma patients. Reference gene selection is crucial because it directly influences the interpretation of RT-qPCR data.

According to the Minimum Information for Publication of Quantitative Real-time PCR experiments guidelines [15], the RNA integrity and purity should be quantified. However, it is well known that the mRNA from FFPE tissue samples is highly degraded and chemical modifications occur during fixation and storage [5,6,16]. Although we could not measure the purity and integrity of mRNA in the present study, it was still possible to quantify the expression levels of 18 reference genes in 78% of the tissue samples. This is by itself an important finding because, in the everyday clinic, the FFPE tissue samples often are the only available specimens for research. Due to the low quality of mRNA from old FFPE tissue samples, a proper choice

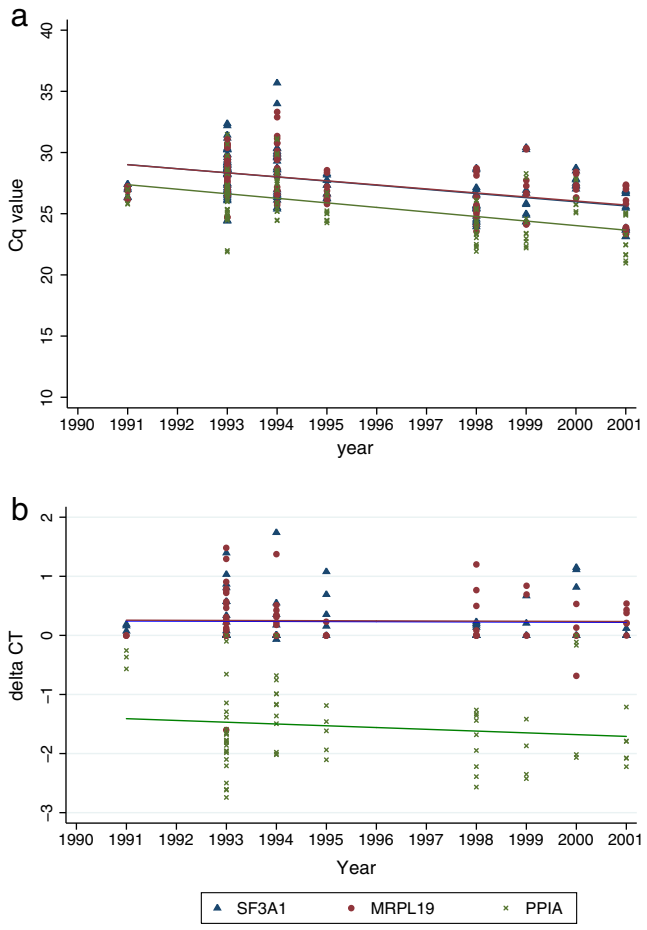

Figure 3. The top figure shows the Cq value for the most stably expressed genes, with the slope of the line representing the degradation of the RNA/year. The bottom figure shows the Cq value for the most stably expressed genes normalized to the geometric mean of the same three genes according to the age for the FFPE tissue blocks. The *y*-axis represents the year of diagnosis.

of suitable reference genes is even more important. The selection of reference genes is, however, also highly dependent on the tissues or cells under investigation [17,18], and the expression level of the reference genes can even change with different anatomical regions [19]. In this study, we used GeNorm and NormFinder to find the most stably expressed reference genes. However, the two methods apply different statistical algorithms, and it was therefore not surprising that they gave somewhat different results. Also, different reference genes are recommended for different histologic types of sarcoma (Table 3). For example, *NDFIP1* was recommended by the NormFinder algorithm when analyzing PUS, but this potential reference gene was poorly ranked when analyzing liposarcoma. Nevertheless, there was general agreement that *MRPL19, PPIA,* and *SF3A1* were within the 10 top-ranking genes regardless of the tissue under investigation or algorithms used. Vandesomple et al. have recommended at least three appropriate reference genes for calculation of a normalization factor because of the variation in the expression of reference genes [13], which were also used in this study.

The selected reference genes in this study have been shown to be stable expressed in other tissues than sarcoma. In studies investigating reference genes for RT-qPCR experiments using FFPE, *MRPL19* has also been shown to be among the top-ranked reference genes for FFPE samples from both breast [20] and lung cancer patients [7]. *PPIA* has been found to be among the most suitable reference genes for normalization in human FFPE epithelial ovarian tissue samples [21] and from fresh frozen tissue samples in colon cancer [22]. *SF3A1* has been found to be among the most stable genes in fresh frozen breast cancer specimens [23].

Because reference genes stably expressed in one cancer type may be less stably expressed in another cancer or under a different experimental condition [3,4,17,24], we investigated both the intraindividual variation and interindividual variation for all tissues together and for the liposarcoma and PUS separately. These results showed that 57%, 61%, and 47% of the variation in ΔCq values for *PPIA, SF3A1,* and *MRPL19*, respectively, were a variation between patients rather than within each patient, taking the effect of different histologic types into account. This means that the variation within the samples of each patient is not different than the variation detected between patients, which is desirable when choosing proper reference genes. When looking at the measurement error for all tissues and for PUS and liposarcoma alone, the main intraindividual variation was seen in the PUS samples.

It is important here to note that *GAPDH, B2M,* and *rRNA18s* commonly used as reference genes in other studies were, in fact, the least stable genes regardless of tissue type and analysis algorithms in these FFPE tissue samples. The less stable expression of *GAPDH* is in agreement with previously published results showing an eight-fold difference in expression levels in different tissue types [3]. Another often used reference gene, *ACTB*, has been shown to have a 22-fold difference in expression levels when comparing different tissue types [3]. In our experiments, we also found variation among different histologic types.

The total RNA is predominantly composed of rRNA and does not always reflect mRNA content [25], which is why ribosomal RNA might not be a good reference gene. RNA from FFPE samples is highly degraded, nucleic acids are cross-linked to proteins, and base modifications are introduced during the fixation process [26]. The half-life of the mRNA was found to be 2.9 years (95% CI: 2.4 to 3.6), which is within the range of previously published results [27]. When normalizing the three most stable genes to the geometric mean of the same three genes, the effect of degradation was eliminated. However, it is important to analyze the degradation of both target genes and reference genes to ensure the same degradation level.

It is of great importance that every research group investigating gene expression, protein expression, or DNA uses proper reference genes. The results will never get better than the reference genes used in the study. The lack of proper reference gene analysis before conducting gene expression analysis may lead to misleading conclusions and possible undesirable clinical consequences. However, as this study also shows, it might not be possible to exclude or include any specific set of reference genes without testing the genes in the specific experimental setting.

In conclusion, *PPIA, SF3A1*, and *MRPL19* are suitable reference genes for normalization in gene expression profiling studies of FFPE samples from primary soft tissue sarcoma. Although there may be differences in the recommended reference genes according to the histologic sarcoma subtype, it is possible to find reference genes that are stably expressed regardless of the histology of the sarcoma. We recommend that this specific set of reference genes be used and validated in independent experimental setting.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.tranon.2014.09.012.

Acknowledgments

We acknowledge Alice Baden, Charlotte Wulf-Høyer, Nanna S. L. Hansen, and Mogens Jøns Johansen for excellent technical assistance. The work was supported by the A. P. Møller and Christine McKinney Møller Foundation.

References

[1] Jain A, Sajeevan KV, Babu KG, and Lakshmaiah KC (2009). Chemotherapy in adult soft tissue sarcoma. *Indian J Cancer* **46**(4), 274–287.

[2] Bustin SA (2000). Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. *J Mol Endocrinol* **25**(2), 169–193.

[3] Warrington JA, Nair A, Mahadevappa M, and Tsyganskaya M (2000). Comparison of human adult and fetal expression and identification of 535 housekeeping/maintenance genes. *Physiol Genomics* **2**(3), 143–147.

[4] Zhong H and Simons JW (1999). Direct comparison of GAPDH, beta-actin, cyclophilin, and 28S rRNA as internal standards for quantifying RNA levels under hypoxia. *Biochem Biophys Res Commun* **259**(3), 523–526.

[5] Macabeo-Ong M, Ginzinger DG, Dekker N, McMillan A, Regezi JA, Wong DT, and Jordan RC (2002). Effect of duration of fixation on quantitative reverse transcription polymerase chain reaction analyses. *Mod Pathol* **15**(9), 979–987.

[6] Nam SK, Im J, Kwak Y, Han N, Nam KH, Seo AN, and Lee HS (2014). Effects of fixation and storage of human tissue samples on nucleic acid preservation. *Korean J Pathol* **48**(1), 36–42.

[7] Soes S, Sorensen BS, Alsner J, Overgaard J, Hager H, Hansen LL, and Kristensen LS (2013). Identification of accurate reference genes for RT-qPCR analysis of formalin-fixed paraffin-embedded tissue from primary non–small cell lung cancers and brain and lymph node metastases. *Lung Cancer* **81**(2), 180–186.

[8] Walter RF, Mairinger FD, and Wohlschlaeger J, et al (2013). FFPE tissue as a feasible source for gene expression analysis — a comparison of three reference genes and one tumor marker. *Pathol Res Pract* **209**(12), 784–789.

[9] von Ahlfen S, Missel A, Bendrat K, and Schlumpberger M (2007). Determinants of RNA quality from FFPE samples. *PLoS One* **2**(12), e1261.

[10] Ludyga N, Grunwald B, and Azimzadeh O, et al (2012). Nucleic acids from long-term preserved FFPE tissues are suitable for downstream analyses. *Virchows Arch* **460**(2), 131–140.

[11] Bohmann K, Hennig G, and Rogel U, et al (2009). RNA extraction from archival formalin-fixed paraffin-embedded tissue: a comparison of manual, semiautomated, and fully automated purification methods. *Clin Chem* **55**(9), 1719–1727.

[12] Toustrup K, Sorensen BS, Nordsmark M, Busk M, Wiuf C, Alsner J, and Overgaard J (2011). Development of a hypoxia gene expression classifier with predictive impact for hypoxic modification of radiotherapy in head and neck cancer. *Cancer Res* **71**(17), 5923–5931.

[13] 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**(7) [RESEARCH0034].

[14] Andersen CL, Jensen JL, and Orntoft TF (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**(15), 5245–5250.

[15] Bustin SA, Benes V, Garson JA, Hellemans J, Huggett J, Kubista M, Mueller R, Nolan T, Pfaffl MW, and Shipley GL (2009). The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin Chem* **55**(4), 611–622.

[16] Srinivasan M, Sedmak D, and Jewell S (2002). Effect of fixatives and tissue processing on the content and integrity of nucleic acids. *Am J Pathol* **161**(6), 1961–1971.

[17] Barber RD, Harmer DW, Coleman RA, and Clark BJ (2005). GAPDH as a housekeeping gene: analysis of GAPDH mRNA expression in a panel of 72 human tissues. *Physiol Genomics* **21**(3), 389–395.

[18] Wisnieski F, Calcagno DQ, Leal MF, dos Santos LC, Gigek Cde O, Chen ES, Pontes TB, Assumpcao PP, de Assumpcao MB, and Demachki S (2013). Reference genes for quantitative RT-PCR data in gastric tissues and cell lines. *World J Gastroenterol* **19**(41), 7121–7128.

[19] Zhou ZJ, Zhang JF, Xia P, Wang JY, Chen S, Fang XQ, and Fan SW (2014). Selection of suitable reference genes for normalization of quantitative real-time polymerase chain reaction in human cartilage endplate of the lumbar spine. *PLoS One* **9**(2), e88892.

[20] Drury S, Anderson H, and Dowsett M (2009). Selection of REFERENCE genes for normalization of qRT-PCR data derived from FFPE breast tumors. *Diagn Mol Pathol* **18**(2), 103–107.

[21] Cai J, Li T, Huang B, Cheng H, Ding H, Dong W, Xiao M, Liu L, and Wang Z (2014). The use of laser microdissection in the identification of suitable reference genes for normalization of quantitative real-time PCR in human FFPE epithelial ovarian tissue samples. *PLoS One* **9**(4), e95974.

[22] Sorby LA, Andersen SN, Bukholm IR, and Jacobsen MB (2010). Evaluation of suitable reference genes for normalization of real-time reverse transcription PCR analysis in colon cancer. *J Exp Clin Cancer Res* **29** [144,9966-29-144].

[23] Maltseva DV, Khaustova NA, Fedotov NN, Matveeva EO, Lebedev AE, Shkurnikov MU, Galatenko VV, Schumacher U, and Tonevitsky AG (2013). High-throughput identification of reference genes for research and clinical RT-qPCR analysis of breast cancer samples. *J Clin Bioinform* **3**(1) [13,9113-3-13].

[24] Rubie C, Kempf K, Hans J, Su T, Tilton B, Georg T, Brittner B, Ludwig B, and Schilling M (2005). Housekeeping gene variability in normal and cancerous colorectal, pancreatic, esophageal, gastric and hepatic tissues. *Mol Cell Probes* **19**(2), 101–109.

[25] Solanas M, Moral R, and Escrich E (2001). Unsuitability of using ribosomal RNA as loading control for northern blot analyses related to the imbalance between messenger and ribosomal RNA content in rat mammary tumors. *Anal Biochem* **288**(1), 99–102.

[26] Masuda N, Ohnishi T, Kawamoto S, Monden M, and Okubo K (1999). Analysis of chemical modification of RNA from formalin-fixed samples and optimization of molecular biology applications for such samples. *Nucleic Acids Res* **27**(22), 4436–4443.

[27] Tramm T, Sorensen BS, Overgaard J, and Alsner J (2013). Optimal reference genes for normalization of qRT-PCR data from archival formalin-fixed, paraffin-embedded breast tumors controlling for tumor cell content and decay of mRNA. *Diagn Mol Pathol* **22**(3), 181–187.