

Validation of an appropriate reference gene for normalization of RT-qPCR data from rectal cancer biopsies

Alexandre Ho-Pun-Cheung^{a,b}, Caroline Bascoul-Mollevis^a, Eric Assenat^a, Frédéric Bibeau^a, Florence Boissière-Michot^a, Dominic Cellier^b, Marc Ychou^a, and Evelyne Lopez-Crapez^{a,*}

^aVal d'Aurelle Cancer Institute, Montpellier, France

^bMerck Santé, Lyon, France

* Corresponding author. Fax: +33 467 632873.

E-mail address: ecrapez@valdorel.fnclcc.fr (E. Lopez-Crapez).

Running title: Normalization of RT-qPCR data in rectal cancer

ABSTRACT

Gene expression quantification using reverse transcription-quantitative polymerase chain reaction (RT-qPCR) requires data normalization using an invariable reference gene. Here, we assessed the stability of 15 housekeeping genes in 31 tumor and normal rectal samples to validate a reliable reference gene for rectal cancer studies. Our data show that *18S* and *28S* RNA are stably expressed in all samples. Moreover, when used for normalization, *18S*, but not *28S*, greatly reduced unspecific variations of gene expression due to RNA degradation. These results demonstrate that *18S* is an appropriate reference gene for normalization of RT-qPCR data from rectal cancer samples.

Keywords:

RT-PCR

Normalization

Reference gene

Rectal cancer

Quantification

Reverse transcription-quantitative polymerase chain reaction (RT-qPCR) is the method of choice to compare changes in mRNA expression among different samples, experimental conditions or physiological states. However, gene expression levels must be normalized to obtain reliable data [1]. Generally this involves the use of one or more housekeeping genes. These genes regulate basic and ubiquitous cellular functions and their expression is assumed to be stable among individuals,

tissue types, or physiological states. Since housekeeping genes are processed as the gene(s) of interest, they are an internal reference to which target gene expression can be related to correct unspecific variations due to inaccurate amount of input RNA, RNA degradation and/or presence of reaction inhibitors. Unfortunately, commonly used reference genes are regulated and their expression may vary among individuals or following treatments [1,2]. Hence, the selected reference genes should be validated for each experimental setting. In rectal cancer, pre-therapeutic biopsies are used in gene expression studies to identify predictive markers of response to therapy [3]. However, no reference genes have been validated for this setting so far. To evaluate which genes could be appropriate for RT-qPCR normalization in rectal cancer, we assessed the stability of 15 candidate reference genes (*18S*, *28S*, *ACTB*, *ATP6*, *ATUB*, *B2M*, *GAPDH*, *HMBS*, *HPRT*, *PGK1*, *POLR2L*, *PPIE*, *PSMB6*, *RPLP0* and *TBP*) in 31 tumor and normal rectal biopsies. Candidate genes were selected on the basis of two criteria: 1) their previous use as reference genes and 2) their ability to cover a wide expression spectrum.

Forty-nine patients with rectal cancer were included in this study. For all patients, six pre-therapeutic biopsies from the primary adenocarcinoma and two from paired normal tissues were obtained by endoscopy. Biopsies were pooled according to the tissue type (tumor or normal) before RNA extraction using the RNeasy Mini Kit (Qiagen). Only samples with high RNA integrity were retained, as RNA degradation may greatly influence quantification of gene expression [4]. RNA integrity was evaluated by determining the RNA Integrity Number (RIN) using the RNA 6000 Nano LabChip kit and the Agilent 2100 bioanalyzer (Agilent Biotechnologies). RIN ranges from 1 (completely degraded RNA) to 10 (intact RNA) and the inclusion threshold for this study was set at $RIN \geq 6$. This led to the selection of 31 RNA

samples obtained from 22 paired normal/tumor rectal tissues and 9 unpaired rectal cancer samples. To ensure that equal amounts of input RNA were used for the RT reaction, RNA concentrations were accurately measured using the Hellma tray cell system combined to a SAFAS UV mc2 spectrophotometer. This system allows the analysis of extremely small volumes (0.7–4 μ l), thus avoiding dilution errors. Tissue RNA concentration ranged from 268 to 943 ng/ μ l, and the A_{260}/A_{280} ratio was between 2.0 and 2.1. To limit inter-assay variation, all samples were reverse-transcribed simultaneously using the same reagents and master mix. A 13- μ l mix containing 1 μ g total RNA, 150 ng of random hexamers (Promega), 1 μ l 10 mM dNTP Mix (Invitrogen), and 0.3 pg of a plant mRNA Spike (*A. thaliana* chlorophyll a/b-binding protein; CAB mRNA) (Stratagene) was prepared for each sample. Mixtures were heated at 65°C for 5 min, cooled on ice, and then a 7- μ l-reaction mix containing 1 μ l of SuperScriptTM III Reverse Transcriptase (200 U/ μ l) (Invitrogen), 4 μ l of 5X First-Strand Buffer (Invitrogen), 1 μ l of 0.1 M DTT (Invitrogen), and 1 μ l of SUPERase.InTM (20 U/ μ l) (Ambion) was added. First strand cDNA synthesis was carried out in an Eppendorf[®] Mastercycler[®] with a priming step at 25°C for 5 min, followed by reverse transcription at 50°C for 60 min. A final inactivation step at 70°C for 15 min completed the reaction. These samples were then used to assess the expression of the 15 candidate reference genes (see Table 1 in supplementary material). Real-time quantitative PCR (qPCR) was performed in a Rotor-GeneTM 6000 (Labgene) using the ABsoluteTM Blue QPCR SYBR[®] Green Mix (ABgene). PCR reactions were prepared in a final volume of 20 μ l. Amplification was initiated with an enzyme activation step at 95°C for 15 min, followed by 40 cycles consisting of 15 seconds of denaturation at 95 °C, 30 seconds of annealing at the temperature appropriate for each primer (see Table 2 in supplementary material), and 30 seconds

of elongation at 72°C. For each gene, a standard curve was generated using serial dilutions of HCT116 cDNA. Amplification efficiencies (E) were calculated from the slope of the standard curves according to the equation: $E=10^{[-1/\text{slope}]}$, and they ranged from 92% to 100%. To exclude between-run variations, all samples were tested in duplicate in the same analytical run. For each gene, sample relative concentrations were determined from the standard curve, and then expressed as an n-fold difference relative to a calibrator, according to the relative standard curve method. Potential inhibition of both RT and qPCR steps was assessed by following the expression of the exogenous *CAB* mRNA in each sample. In all cDNA samples, *CAB* had a mean cycle threshold (C_t) value of 21.86 ± 0.22 (95% confidence interval 21.78–21.94) at a fixed threshold of 0.06. The low standard deviation obtained suggests that there was no significant difference in the efficiency of both RT and qPCR steps among samples.

The stability of the 15 candidate reference genes in both normal and tumor samples was evaluated using descriptive statistics (Fig. 1A). For each gene, sample expression levels were presented as an n-fold difference relative to a calibrator, which was the sample with the smallest relative concentration. The most stable genes were *18S* and *28S* with less than 2.3- and 2.2-fold variation, respectively. In contrast, the expression of the other candidates fluctuated considerably, the least stable genes being *HPRT* and *PGKI* with up to 12.6- and 16-fold difference in expression among samples. Next, the relative expression levels of the candidates were compared between tumor and normal tissues (Fig. 1B). Again, *18S* and *28S* showed little variation in both normal and tumor rectal samples as previously reported in other tissue types [5-8]. Conversely, other housekeeping genes, including widely used reference genes such as *GAPDH* and *ACTB*, showed clear differences in expression. Their expression was relatively stable in normal tissues, but up-regulated and highly

variable in the paired tumor. This variability could be explained by the fact that these genes not only participate in basal cell metabolism, but are also involved in many other cellular functions. GAPDH for instance, is not only a key enzyme of glycolysis, but also a multifunctional protein implicated in various cellular processes such as DNA repair, or apoptosis [9]. These results indicate that *18S* and *28S* ribosomal RNA display stable expression in rectal biopsies. However this is not sufficient to make of them suitable reference genes for normalization as their integrity should closely match that of mRNA. This is especially true for human biopsies in which RNA is frequently degraded during sample handling. Thus, we evaluated the ability of *18S* and *28S* to correct for variations when degraded RNA was used for RT-qPCR. Different methods to degrade RNA have been described in the literature, including the use of RNase treatment [10], UV radiation [10], or thermal hydrolysis [11]. While these procedures are artificial and may differ from the natural degradation that occurs during sample handling, they allow producing RNA samples that are representative of all possible degrees of RNA degradation. In this study, intact cellular RNA (RIN = 10) from HCT116 cells was aliquoted and each aliquot was gradually degraded by hydrolysis at 70°C for different length of time (from 0 to 165 min) to obtain increasingly degraded samples with RIN values between 10 (intact RNA) and 4.7 (highly degraded RNA). We then compared the average non-normalized expression ratios of 13 target genes to the corresponding *18S*- and *28S*-normalized ratios in these artificially degraded samples (Fig. 2). The intact sample was defined as the calibrator and the transcription level of each gene was expressed as an n-fold difference relative to this calibrator. Thus, for any given gene, a ratio of 1 was expected and any variations between samples would be imputed mainly to RNA degradation. In pre-normalized data, relative gene expression levels decreased along with the RIN value

and expression ratios exhibited up to 40% error for the RIN=4.7 sample. After normalization to *18S*, the maximum variation was reduced to 5%. Conversely, with *28S* the maximum error was 25% for the RIN=4.7 sample. Moreover, *28S*-normalized ratios exhibited larger errors for the RIN=9 and RIN=8.2 samples than non-normalized data. We then modeled our data by linear regression analysis, and compared the obtained equations to the line $y = 1$, which corresponded to the ideal situation in which RNA degradation has no effect on the expression ratio. The regression lines of the non-normalized and *28S*-normalized ratios were statistically different from the line $y = 1$ ($p < 0.001$ and $p = 0.005$, respectively), while this was not the case for the average *18S*-normalized ratio ($p = 0.52$). Taken together, our results demonstrate that variations due to RNA degradation are greatly reduced when data are normalized to *18S*, whereas this is not the case when *28S* is used.

In summary, we showed that *18S* expression in rectal tissue is stable among individuals and not influenced by malignant transformation. We also demonstrated that *18S* is suitable for the normalization of unspecific variations related to mRNA degradation. We conclude that *18S* is an appropriate reference gene for RT-qPCR normalization in rectal cancer biopsies.

Acknowledgements

This work was supported by Merck Santé and ANRT (Association Nationale de la Recherche Technique).

References

- [1] S.A. Bustin, V. Benes, T. Nolan, M.W. Pfaffl, Quantitative real-time RT-PCR - a perspective, *J Mol Endocrinol.* 34 (2005) 597-601.
- [2] A. Ho-Pun-Cheung, D. Cellier, E. Lopez-Crapez, [Considerations for normalisation of RT-qPCR in oncology.], *Ann.Biol Clin (Paris).* 66 (2008) 121-129.
- [3] B.M. Ghadimi, M. Grade, M.J. Difilippantonio, S. Varma, R. Simon, C. Montagna, L. Fuzesi, C. Langer, H. Becker, T. Liersch, T. Ried, Effectiveness of gene expression profiling for response prediction of rectal adenocarcinomas to preoperative chemoradiotherapy, *J Clin Oncol.* 23 (2005) 1826-1838.
- [4] S. Imbeaud, E. Graudens, V. Boulanger, X. Barlet, P. Zaborski, E. Eveno, O. Mueller, A. Schroeder, C. Auffray, Towards standardization of RNA quality assessment using user-independent classifiers of microcapillary electrophoresis traces, *Nucleic Acids Res.* 33 (2005) e56.
- [5] 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.
- [6] 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.
- [7] C. Rubie, K. Kempf, J. Hans, T. Su, B. Tilton, T. Georg, B. Brittner, B. Ludwig, M. Schilling, Housekeeping gene variability in normal and cancerous colorectal,

pancreatic, esophageal, gastric and hepatic tissues, *Mol Cell Probes*. 19 (2005) 101-109.

- [8] S. Silvia, C. Francesca, L.I. Marco, N. Silvia, V.S. Giorgio, C.A. Raffaele, Selection of suitable reference genes for accurate normalization of gene expression profile studies in non-small cell lung cancer, *BMC Cancer*. 6:200. (2006) 200.
- [9] M.A. Sirover, New nuclear functions of the glycolytic protein, glyceraldehyde-3-phosphate dehydrogenase, in mammalian cells, *J Cell Biochem*. 95 (2005) 45-52.
- [10] S. Fleige, V. Walf, S. Huch, C. Prgomet, J. Sehm, M.W. Pfaffl, Comparison of relative mRNA quantification models and the impact of RNA integrity in quantitative real-time RT-PCR, *Biotechnol Lett*. 28 (2006) 1601-1613.
- [11] S. Mueller, Optimizing real-time quantitative PCR experiments with the Agilent 2100 bioanalyzer, Agilent Technologies - Application Note 5989-7730EN2008.

Figure legends

Fig. 1. Box plot distribution of the relative expression ratios of the 15 candidate reference genes. Normal and tumor samples were analyzed together (A) or separately (B). Grey boxes represent the 25th to 75th percentiles with the medians as black lines; whiskers mark the smallest and largest non-outlier observations, and all outliers are indicated by dots.

Fig. 2. Suitability of *18S* and *28S* for the normalization of variations due to RNA degradation. The curves represent the average expression ratios and 95% confidence interval of 13 genes (*ACTB*, *ATP6*, *ATUB*, *B2M*, *GAPDH*, *HMBS*, *HPRT*, *PGK1*, *POLR2L*, *PPIE*, *PSMB6*, *RPLP0* and *TBP*) in HCT116 samples with decreasing RIN values. For each gene, *18S*- and *28S*-normalized expression ratios were determined by dividing the non-normalized expression ratio by the expression ratio of *18S* or *28S*.