

Chapter 3

Selection of Reliable Reference Genes for RT-qPCR Analysis

Jan Hellemans and Jo Vandesompele

Abstract

Reference genes have become the method of choice for normalization of qPCR data. It has been demonstrated in many studies that reference gene validation is essential to ensure accurate and reliable results. This chapter describes how a pilot study can be set up to identify the best set of reference genes to be used for normalization of qPCR data. The data from such a pilot study should be analyzed with dedicated algorithms such as geNorm to rank genes according to their stability—a measure for how well they are suited for normalization. geNorm also provides insights into the optimal number of reference genes and the overall quality of the selected set of reference genes. Importantly, these results are always in function of the sample type being studied. Guidelines are provided on the interpretation of the results from geNorm pilot studies as well as for the continued monitoring of reference gene quality in subsequent studies. For screening studies including a large, unbiased set of genes (e.g., complete miRNome) an alternative normalization method can be used: global mean normalization. This chapter also describes how the data from such studies can be used to identify reference genes for subsequent validation studies on smaller sets of selected genes.

Key words qPCR, Selection of reference genes, Human candidate reference genes, Candidate reference microRNAs, geNorm

1 Introduction

Reverse transcription quantitative PCR (RT-qPCR) is currently the most popular method for the comparison of gene expression levels among biological samples. The reasons for the popularity of RT-qPCR are its sensitivity, flexibility, low cost of reagents and instruments, and the apparent ease to perform a qPCR experiment. The ease to generate qPCR data is in sharp contrast with the challenges to guarantee that the obtained results are reliable. One of the biggest challenges is proper normalization of the data.

Many different methods have been proposed to normalize qPCR data. An overview of basic normalization concepts is nicely reviewed in Huggett et al. [1]. Of all the options to normalize qPCR data, the use of reference genes (historically referred to as housekeeping genes) is undoubtedly the most popular approach. It is appealing not only for its practical simplicity but also for its

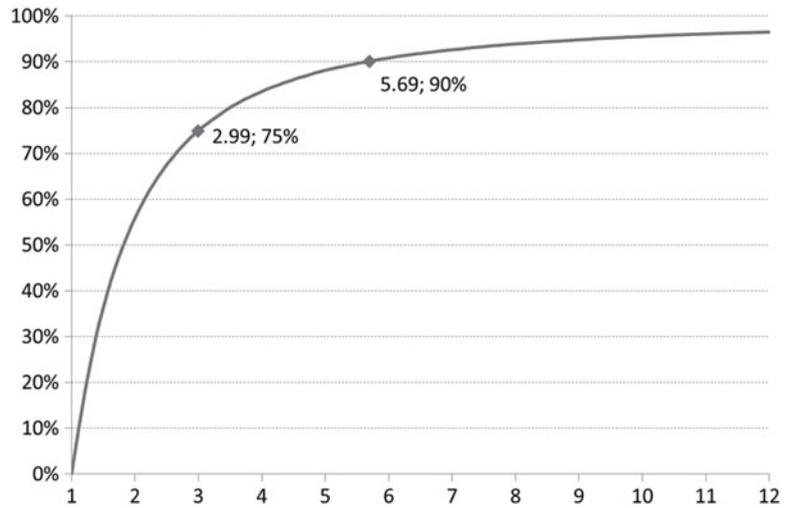


Fig. 1 Cumulative distribution of the normalization bias caused by using a single nonvalidated reference gene. Data from Vandesompele et al. [2]

potential to remove most of technical variation in cDNA concentrations between samples. However, for reference genes to meet this potential, they have to be stably expressed. Studies performed in 2002 by Vandesompele et al. and many others since then highlight the risk of blindly relying on the assumption of stable expression of reference genes [2]. The average difference in expression level of a gene of interest after normalization with any of two randomly selected nonvalidated reference genes is at least threefold in 25 % of the cases and even more than sixfold in 10 % of the cases (Fig. 1). In view of the modest degree of differential expression observed in many experiments, this bias is clearly unacceptable.

These results clearly indicate the need for proof that selected reference genes are stably expressed.

2 geNorm

A naive approach to verify stable expression is the comparison of raw Cq values for reference genes between different samples, looking for genes with minimal variation in Cq values. However, this approach is flawed because it assumes equal cDNA concentrations for all samples. Of note, cDNA inputs are often standardized by means of equal total RNA inputs. As total RNA consists mostly of ribosomal RNA and as mRNA:rRNA ratios are quite variable depending on cell type and condition, equal cDNA input amount is an unreliable measure to normalize or to validate reference genes. More robust methods that are independent of cDNA input concentrations are required to quantify the degree of variation in expression levels for reference genes. In 2002, Vandesompele et al. described the geNorm approach to identify stably expressed reference genes. This method is robust and independent of cDNA input amount (as long as the same amount is used

to measure all candidate reference genes in that sample). Since then, a number of alternative methods have been published (for overview see book chapter on “Reference gene validation software for improved normalization” [3]). The choice between any of these algorithms is less important than the fact of actually using a valid method to determine the stability of reference genes. In most cases, results will be equivalent when using different approaches.

An experiment to identify stably expressed reference genes prior to conducting the real experiment (in which the gene(s) of interest are measured) is referred to as a geNorm pilot study. Such a study typically consists of eight or more candidate reference genes evaluated for ten or more representative samples. Candidate reference genes may be selected from a list of usual suspects (examples for human (Table 1)), from commercial reference gene kits

Table 1
Human candidate reference genes

Official symbol	Official name	Ensembl gene ID	Entrez gene ID	Refseq IDs
ACTB	Actin, beta	ENSG00000075624	60	NM_001101
B2M	Beta-2-microglobulin	ENSG00000166710	567	NM_004048
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase	ENSG00000111640	2597	NM_002046, NM_001256799
GUSB	Glucuronidase, beta	ENSG00000169919	2990	NM_000181
HMBS	Hydroxymethylbilane synthase	ENSG00000256269	3145	NM_001258209, NM_000190, NM_001258208, NM_001024382
HPRT1	Hypoxanthine phosphoribosyltransferase 1	ENSG00000165704	3251	NM_000194
PPIA	Peptidylprolyl isomerase A (cyclophilin A)	ENSG00000196262	5478	NM_021130
RPL13A	Ribosomal protein L13a	ENSG00000142541	23521	NM_012423, NM_001270491, NR_073024
RPS18	Ribosomal protein S18	ENSG00000231500	6222	NM_022551
SDHA	Succinate dehydrogenase complex, subunit A, flavoprotein (Fp)	ENSG00000073578	6389	NM_004168
TBP	TATA box binding protein	ENSG00000112592	6908	NM_003194, NM_001172085
TUBB	Tubulin, beta	ENSG00000196230	203068	NM_178014
UBC	Ubiquitin C	ENSG00000150991	7316	NM_021009
YWHAZ	Tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, zeta polypeptide	ENSG00000164924	7534	NM_003406, NM_001135699, NM_145690, NM_001135700, NM_001135701, NM_001135702

(e.g., from PrimerDesign), from the literature, from experimental data in the lab (e.g., from microarray or RNA-seq), or simply from assays available in the lab. It is important to avoid selecting multiple genes from the same pathway or functional class because strong coregulation may interfere with proper analysis of expression stability. The number of candidate reference genes to be tested varies depending on sample heterogeneity and required experimental accuracy and precision. More challenging experiments with complex, heterogeneous samples and small differences in expression of the gene of interest will require a larger set of candidates to allow selection of suitable reference genes. The selection of samples needs to be representative of the actual study to be performed since different sample types may result in a different selection of reference genes. If sample subgroups exist (e.g., treated and untreated), they need to be equally represented in the geNorm pilot study to avoid selecting genes that are stably expressed in one group, but not across all samples.

Three types of information can be extracted from a geNorm analysis:

1. Ranking of candidate reference genes according to their stability.
2. The optimal number of reference genes for the analyzed sample type.
3. An assessment of the overall stability of selected reference genes (Fig. 2).

Candidate reference genes are ranked according to their M -value, a measure for relative expression stability. Unstable reference genes with high M -values are sorted to the left; the best reference genes are found on the right side. The selection of most stably expressed genes will vary from sample type to sample type. Also the optimal number of reference genes is variable and can be determined based on the V -value. The V -value is an indication for how much difference it makes when using an extra reference gene for normalization. If the added value of adding one more is limited (guideline: V below 0.15), one could as well leave it out. The average M -value for the optimal number of reference genes is an indication for their combined quality to normalize your data. Based on many sets of empirical data we suggest the following guidelines. Average M -values below 0.2 are typically seen when evaluating reference targets using genomic DNA as input (e.g., for CNV analysis), or when reference targets are very stably expressed. Average M -values between 0.2 and 0.5 are typically seen when evaluating candidate reference targets on a homogeneous set of samples (e.g., untreated cultured cells, or blood from normal individuals). Average M -values in the 0.5–1 range are expected when evaluating candidate reference targets on a heterogeneous set of samples

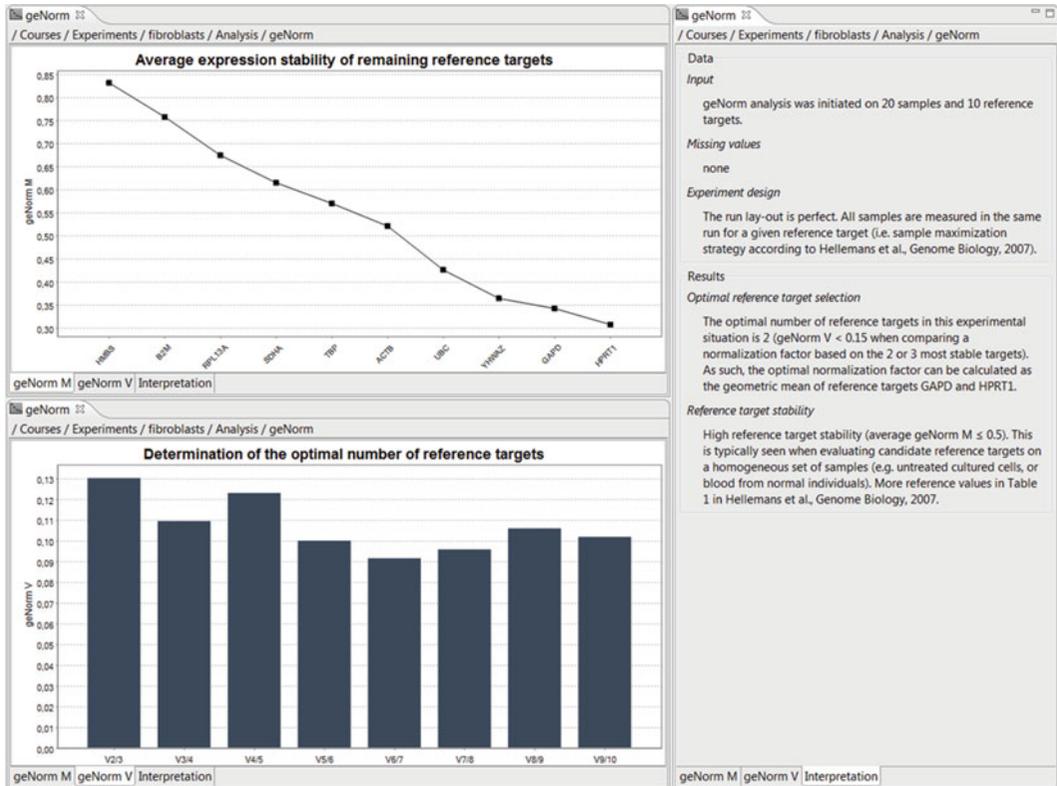
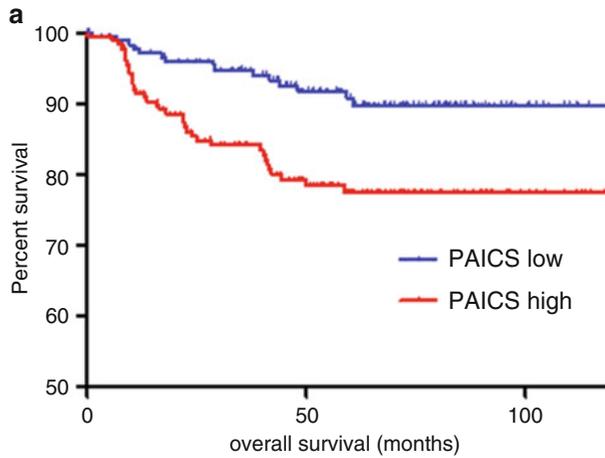


Fig. 2 Example results for a geNorm analysis on fibroblasts, generated by the qPCR data analysis software qbase+

(e.g., treated cultured cells, cancer biopsies, or samples from different tissues). Average M -values larger than 1 indicate low reference target stability. In those cases, it might prove useful to evaluate other candidate reference genes.

Reference genes selected in a geNorm pilot can be used in all subsequent experiments analyzing the same sample type in similar conditions. It is advisable to continue monitoring the stability of selected reference genes. This quality control is only possible if two or more reference genes are included in an experiment. Reference gene stability can be quantified by calculating geNorm M -values, or alternatively by inspecting the variation (expressed as CV values) on the normalized relative quantities of tested reference genes. This test is very valuable to exclude the possibility that one of your treatments or manipulations unexpectedly results in altered expression levels for one of your reference genes, and it comes at no extra cost.

Performing a geNorm pilot study does involve extra work to acquire reference gene assays, to run the pilot reactions, and to analyze and interpret the results. Because of this, many people take a pragmatic approach to the issue of reference gene selection. By consulting their colleagues, peers, or the literature they make an



b

normalization	Cox regression p -value	significant	fold change in p -value	Cox regression hazard ratio	change in hazard ratio
<i>5 reference genes</i>	0.002	yes		3.990	
AluSq	0.005	yes	2.5	2.292	-43%
HPRT1	0.006	yes	3.0	2.381	-40%
SDHA	0.085	no	42.5	1.633	-59%
HMBS	0.186	no	93.0	1.434	-64%
UBC	0.232	no	116.0	1.331	-67%

Fig. 3 (a) Kaplan–Meier survival curve for children with neuroblastoma. 366 patients were dichotomized with respect to the median PAICS mRNA expression. Normalization was done with five stably expressed reference genes (UBC, HPRT1, SDHA, HMBS, and AluSq). The survival rates of patients with PAICS expression higher than the median were significantly worse than those with PAICS expression lower than the median ($p=0.002$) (data from Vermeulen et al. [5]). **(b)** Table showing increased p -value of the Cox proportional hazards test on data from the same experiment normalized with only a single reference gene

educated guess on suitable reference genes. For easy and common sample types this may be a valid approach, as long as the necessary follow-up quality controls, as described above, are included in all experiments. For more difficult samples or demanding studies it is still advisable to perform a full-blown geNorm study to identify the optimal set of reference genes. The benefits of doing so have been demonstrated over and over again. A good set of reference genes enables the removal of most of the technical (experimentally induced) variation. This leads to the possibility to detect smaller differences in expression levels, down to 50 % in real patient samples [4], and to more significant results [5]. The example in Fig. 3 shows how a p -value of 0.002 can be obtained by the combined use of five validated reference genes (of note, clinical cancer samples often display large expression heterogeneity, requiring more reference genes than usual) (Table 1). None of the individual reference genes are as good as the combination (3 out of 5 even give rise to a nonsignificant result). Any randomly selected,

nonvalidated reference gene is likely to perform much worse. PAICS is only one of the 59 tested genes in Vermeulen et al. When doing the same analyses for the other genes, single reference gene normalization results in loss of significance in 7 % of the cases, with an overall increase in p -value of 60-fold (range 2–121,308). In addition, the hazard ratio (risk of death) changed for all tested genes, with an average impact on the magnitude by 45 %. The above results clearly indicate that multi-gene normalization allows to remove the noise in the data and to focus on the true biologically meaningful and clinically relevant differences.

3 Reference for microRNAs

When it comes to microRNA data normalization, there are no pre-defined sets of candidate reference microRNAs. Historically, qPCR users relied on small nuclear or small nucleolar RNAs for normalization (e.g., U6 or RNU44). But, similar to our arguments against ribosomal RNA for mRNA normalization, we advise against the use of these internal controls. Both rRNA and sn(o)RNA are transcribed from a different RNA polymerase and have different functions than mRNA and miRNA, respectively. Even without these arguments, it's never wise to simply rely on one or few (popular) reference genes; you really should test if the candidate reference genes are stably expressed in your experimental condition.

In 2009, we published a new method for even more accurate normalization when a large unbiased set of genes is measured [6]. We applied the method for normalization of microRNA expression profiling studies in which typically a few hundred microRNAs are measured. The method has since then been perfected [7] by attributing equal weight to each individual miRNA during normalization. The improved method is implemented in the qbase + software as “global mean normalization method”.

While the above referenced method works great, it requires many microRNAs to be measured. For follow-up studies, one typically is only interested in the validation of (part of) the differentially expressed microRNAs. To normalize that type of data, we recommend the use of multiple stably expressed microRNAs. We propose the following procedure to find such stably expressed candidate reference microRNAs:

1. Convert Cq values in non-normalized relative quantities ($RQ = 2^{\Delta Cq}$, with ΔCq being the difference in Cq between a reference sample (calibrator) and the sample of interest).
2. Normalize relative quantities using the global mean method.
3. Log transform the normalized data.
4. Calculate the standard deviation per microRNA across all tested samples.

5. Select candidate microRNAs that (a) have data for all samples, (b) have lowest standard deviation, (c) do not belong to the same miR family (use only the best miR per family; miR families can be inspected in a special miRBase file available on <ftp://mirbase.org/pub/mirbase/CURRENT/miFam.dat.gz>). We recommend selecting at least three (five or more is better) candidate reference microRNAs for use in the final experiment.
6. Verify in your final experiment that these candidate microRNAs are stably expressed (low *M*-values, guidance is offered in the geNorm report in qbase + software).

If you do not have access to large-scale microRNA profiles, consider profiling a few representative samples after which you can follow the procedure outlined higher. If that is not an option, you should setup a classic geNorm pilot experiment with sn(o)RNAs and published candidate reference microRNAs (ideally in the same type of samples). Typically, eight candidate small RNAs are measured in at least ten representative samples. The geNorm report in qbase + will help you to identify the most stably expressed genes and will suggest how many genes to use to achieve optimal normalization.

References

1. Huggett J, Dheda K, Bustin S et al (2005) Real-time RT-PCR normalisation; strategies and considerations. *Genes Immun* 6:279–284
2. Vandesompele J, De Preter K, Pattyn F et al (2002). Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol* 3:RESEARCH0034
3. Vandesompele J, Kubista M, Pfaffl MW (2009) Real-time PCR: current technology and applications. In: Logan J, Edwards K, Saunders N (eds) Caister Academic, Norfolk, England, pp 1–27
4. Hellemans J, Preobrazhenska O, Willaert A et al (2004) Loss-of-function mutations in LEMD3 result in osteopoikilosis, Buschke-Ollendorff syndrome and melorheostosis. *Nat Genet* 36: 1213–1218
5. Vermeulen J, De Preter K, Naranjo A et al (2009) Predicting outcomes for children with neuroblastoma using a multigene-expression signature: a retrospective SIOPEN/COG/GPOH study. *Lancet Oncol* 10:663–671
6. Mestdagh P, Van Vlierberghe P, De Weer A et al (2009) A novel and universal method for microRNA RT-qPCR data normalization. *Genome Biol* 10:R64
7. Barbara D, Mestdagh P, Hellemans J et al (2012) miRNA expression profiling: from reference genes to global mean normalization. *Methods Mol Biol* 822:261–272