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

Claus Lindbjerg Andersen,[1] Jens Ledet Jensen,[1,2] and Torben Falck Ørntoft[1]

[1]*Molecular Diagnostic Laboratory, Department of Clinical Biochemistry, Aarhus University Hospital, Aarhus N, Denmark, and* [2]*Department of Theoretical Statistics, Department of Mathematical Sciences, Aarhus University, Aarhus C, Denmark*

ABSTRACT

Accurate normalization is an absolute prerequisite for correct measurement of gene expression. For quantitative real-time reverse transcription-PCR (RT-PCR), the most commonly used normalization strategy involves standardization to a single constitutively expressed control gene. However, in recent years, it has become clear that no single gene is constitutively expressed in all cell types and under all experimental conditions, implying that the expression stability of the intended control gene has to be verified before each experiment. We outline a novel, innovative, and robust strategy to identify stably expressed genes among a set of candidate normalization genes. The strategy is rooted in a mathematical model of gene expression that enables estimation not only of the overall variation of the candidate normalization genes but also of the variation between sample subgroups of the sample set. Notably, the strategy provides a direct measure for the estimated expression variation, enabling the user to evaluate the systematic error introduced when using the gene. In a side-by-side comparison with a previously published strategy, our model-based approach performed in a more robust manner and showed less sensitivity toward coregulation of the candidate normalization genes. We used the model-based strategy to identify genes suited to normalize quantitative RT-PCR data from colon cancer and bladder cancer. These genes are *UBC*, *GAPD*, and *TPT1* for the colon and *HSPCB*, *TEGT*, and *ATP5B* for the bladder. The presented strategy can be applied to evaluate the suitability of any normalization gene candidate in any kind of experimental design and should allow more reliable normalization of RT-PCR data.

INTRODUCTION

When performing a quantitative reverse transcription-PCR (RT-PCR) analysis, several parameters need to be controlled to obtain reliable quantitative expression measures. These include variations in initial sample amount, RNA recovery, RNA integrity, efficiency of cDNA synthesis, and differences in the overall transcriptional activity of the tissues or cells analyzed.

To date, the most frequently used approach for normalization of the above-mentioned parameters is an internal control gene. For a gene to be valid as a control gene, its expression should not vary in the tissues or cells under investigation. The ideal internal control gene is universally valid, with a constant expression level across all thinkable tissue samples, cells, experimental treatments, and designs. Unfortunately, literature shows that as of yet, no such housekeeping gene has been found (1–7), and a recent study touching on the subject even stated that such a gene does not exist (8). It might be true that a universally valid housekeeping gene does not exist, but luckily for most experimental designs, this is not necessary either. Most experimental designs are restricted to a few tissue types or a few different histological

stages of the same tissue, *e.g.*, normal and tumor, untreated and treated, and so forth, and it is likely that one or more genes are constitutively expressed across such restricted designs. However, the task of identifying these genes is not trivial. It is composed of two steps: first, to identify which genes are likely candidates; and second, to verify the stability of these candidates.

In this study, we aimed at identifying normalization genes suited for investigations of experimental designs covering various stages of bladder or colon cancer. We screened high-density oligonucleotide array-based expression profiles to identify possible bladder and colon normalization gene candidates. The next task was to evaluate the expression stability of these candidates in our RT-PCR experimental designs. According to the generally accepted criterion, a suited normalization gene is a gene that shows no expression variation (or at least, only a limited extent of variation) across the investigated sample set. From our point of view, this criterion is too simple. In practice, all genes will show some variation, making it important to remember that in most RT-PCR experiments, the sample set consists of two or more sample groups, *e.g.*, normal and tumor, and that even limited systematic intergroup variation of the candidate could lead to an incorrect interpretation of the results. Thus, we believe that in addition to "overall expression variation," the candidate should also be evaluated as to whether it shows systematic variation across the sample subgroups. Evaluating expression stability of a candidate normalization gene represents a circular problem (*i.e.*, how can the expression stability of a candidate be evaluated if no reliable measure is available to normalize the candidate?). We overcome this circular problem by developing a novel evaluation strategy, called the "model-based approach to estimation of expression variation." In this article, this approach will be outlined, applied, and compared with another expression variance estimation procedure reported recently.

MATERIALS AND METHODS

GeneChip Expression Profiling Data

Expression profiling data from 99 bladder (normal and T_a, T_1, and T_{2-4}) and 161 colon (normal and Dukes' A, B, C, and D) samples were generated previously, in-house, using Affymetrix HuGeneFL GeneChip arrays [70 bladder and 54 colon specimens (9, 10)] and HG-U133A GeneChip arrays (29 bladder and 107 colon specimens).[3] All GeneChip raw data were normalized using the RMA normalization approach in the Bioconductor Affy package (11, 12).

Selection of Normalization Gene Candidates

Fundamentally, a normalization gene candidate has to fulfill two criteria: (*a*) it has to be stably expressed in the tissue of interest; and (*b*) it has to have an expression level above background. To obtain approximately the same number of candidate genes for the two tissues, we took those genes with a coefficient of variation of <0.20 and an average signal intensity of >1800 for colon, whereas we took genes with a coefficient of variation of <0.40 and an average signal intensity of >350 for bladder.

[3] T. Ørntoft, unpublished data.

To minimize the likelihood of selecting false positive genes, we used the fact that for each tissue, we had expression profiling data available from two independent sample sets, each analyzed with its own GeneChip array format, *i.e.*, with different probe and probe set designs, making it possible to preferentially choose the candidates among the genes fulfilling the inclusion criteria in both sample sets. Attention was paid to avoid genes whose GeneChip array probe sets were sensitive to more than a single transcript, *e.g.*, probe sets sensitive to multiple splice variants.

Primer Design

Primers were designed using the primer analysis software Primer Express v2.0 (Applied Biosystems). Designs were based on publicly available sequences (Table 1). Whenever possible, the forward and reverse primers were placed in exons separated by an intron at least 500-bp long. To limit sensitivity toward pseudogenes, care was taken to place at least one of the primers in a region where the gene of interest was distinct from these. To avoid problems caused by sequence variability between individual patients, primers were designed to sequences free of known polymorphisms. In the cases in which the gene of interest had multiple splice variants, primers were designed specifically to the variant detected by the GeneChip array. For each gene, a minimum of two forward and two reverse primers were designed. All combinations of these were tested in real-time RT-PCR. The primer pairs in Table 2 all produced a single product and amplified the target transcript with equal efficiency over a 1000-fold range of input material.

Patient Material

We acquired 28 tumor biopsy samples (10 T_a, 8 T_1, and 10 T_{2-4} samples) from bladder and 10 normal and 30 tumor samples (10 Dukes' A, 10 Dukes' B, and 10 Dukes' C samples) from colon after the amount of tissue necessary for routine pathology examination had been removed. We froze the tumor samples immediately after surgery and stored them at $-80°C$ in a guanidinium thiocyanate-containing solution. Informed consent was obtained in all cases, and protocols were approved by the scientific ethical committee of Aarhus County.

RNA Purification and cDNA Preparation

We isolated total RNA from crude tumor biopsy samples using a Polytron homogenizer and the RNAzol B RNA isolation method (WAK-Chemie Medical GmbH). RNA integrity was confirmed by gel electrophoresis or Bioanalyzer (Agilent Technologies). First-strand cDNA synthesis was carried out using the SuperScript II System (Life Technologies, Inc.). One μg of total RNA and 2 μl of 250 μM random nonamers in a total volume of 12 μl were incubated for 10 min at 70°C and chilled on ice. After adding 4 μl of 1st Strand Buffer, 1 μl of DTT (0.1 M), 2 μl of deoxynucleotide triphosphate mix (10 mM), and 1 μl of SuperScript reverse transcriptase II (200 units/μl), the reaction was incubated for 1 h at 42°C and, finally, for 5 min at 95°C. The cDNA was diluted 1:20 for use in real-time PCR.

Table 1 *Candidate normalization genes evaluated in this study*

Symbol	Gene name	Accession no.[a]	Locus link	Function	Location[b]	Pseudo-genes[c]	Tissue[d]
FLOT2	Flotillin 2	NM_004475.1	2319	May act as a scaffolding protein within caveolar membranes. May be involved in epidermal cell adhesion and epidermal structure and function.	17q11.2	No	B
ATP5B	ATP synthase, H+ transporting, mitochondrial F1 complex, β polypeptide	NM_001686.1	506	Produces ATP from ADP in the presence of a proton gradient across the membrane. The β chain is the catalytic subunit.	12q13.3	Yes	B
HSPCB	Heat shock 90-kDa protein 1, β	ENST00000265413	3326	Molecular chaperone. Has ATPase activity (by similarity).	6p21.1	Yes	B
S100A6	S100 calcium-binding protein A6 (calcyclin)	ENST00000292162	6277	May function in stimulation of Ca^{2+}-dependent insulin release, stimulation of prolactin secretion, and exocytosis.	1q21.3	Yes	B
TEGT	Testis enhanced gene transcript (BAX inhibitor 1)	ENST00000267115	7009	Suppressor of apoptosis	12q13.12	No	B
CFL1	Cofilin 1 (non-muscle)	ENST00000308162	1072	Controls reversibly actin polymerization and depolymerization. It is the major component of intranuclear and cytoplasmic actin rods.	11q13.1	Yes	B and C
FLJ20030	Hypothetical protein FLJ20030	ENST00000218328	54789	Unknown	Xp11.22	No	B and C
TPT1	Tumor protein, translationally controlled 1	ENST00000255477	7178	Unknown	13q14.13	Yes	B and C
UBB[e]	Ubiquitin B	NM_018955.1	7314	Protein degradation	17p11.2	Yes	B and C
UBC[e]	Ubiquitin C	ENST00000280590	7316	Protein degradation	12q24.31	Yes	B and C
RPS13[f]	Ribosomal protein S13	NM_001017.2	6207	Encodes a ribosomal protein that is a component of the 40S subunit	11p15.1	Yes	B and C
RPS23[f]	Ribosomal protein S23	NM_001025.1	6228	Encodes a ribosomal protein that is a component of the 40S subunit	5q14.2	Yes	B and C
GAPD	Glyceraldehyde-3-phosphate dehydrogenase	NM_002046	2597	Catalyzes the reversible oxidative phosphorylation of glyceraldehyde-3-phosphate	12p13.31	Yes	B and C
ACTB	Actin, β	NM_001101.2	60	Cytoskeletal structural protein involved in various types of cell motility	7p22.1	Yes	B and C
CLTC	Clathrin, heavy polypeptide (Hc)	NM_004859.1	1213	Clathrin is the major protein of the polyhedral coat of coated pits and vesicles.	17q23.2	Yes	C
NACA	Nascent-polypeptide-associated complex alpha polypeptide	ENST00000228312	4666	The nascent-polypeptide-associated complex binds newly synthesized polypeptide chains as they emerge from the ribosome and prevents them from being incorrectly translocated to the endoplasmic reticulum	12q13.3	Yes	C
SUI1	Putative translation initiation factor	ENST0000031083	10209	May act as a negative growth regulator	17q21.2	Yes	C
TUBA6	Tubulin alpha 6	ENST00000301072	84790	Cytoskeletal structural protein	12q13.12	Yes	C

[a] Primer design based on this sequence. The database sources are the NCBI Reference Sequence database (http://www.ncbi.nlm.nih.gov/RefSeq/) and the Ensembl database (http://www.ensembl.org/).
[b] Ensembl cytogenetic band.
[c] Determined by BLAT search of Human June 2002 Freeze (Ref. 14).
[d] Identified from bladder (B) and colon (C).
[e] Genes belonging to the same functional class.
[f] Genes belonging to the same functional class.

Table 2 *Primer sequences for candidate normalization genes*

Symbol	Forward primer	Reverse primer	Amplicon size	Intron spanning
FLOT2	TGCCGTGGTGCAGAGAGA	GGTGTCTGCCATGAACTTCACA	115	Yes
ATP5B	TCACCCAGGCTGGTTCAGA	AGTGGCCAGGGTAGGCTGAT	80	Yes
HSPCB	AAGAGAGCAAGGCAAAGTTTGAG	TGGTCACAATGCAGCAAGGT	120	Yes
S100A6	ACAAGCACACCCTGAGCAAGA	CCATCAGCCTTGCAATTTCA	99	Yes
TEGT	TGCTGGATTTGCATTCCTTACA	ACGGCGCCTGGCATAGA	151	Yes
CFL1	GAAGGAGGATCTGGTGTTTATCTTCT	GCTGGCATAAATCATTTTGCTCTT	73	Yes
FLJ20030	GGCTTTGGACTTGCAGAATGTT	ATAATGTAGTGTGTTACTAGTTGTCCTTTTCTC	144	Yes
TPT1	GATCGCGGACGGGTTGT	TTCAGCGGAGGCATTTCC	100	Yes
UBB	GGGCGGTTGGCTTTGTT	GACCTGTTAGCGGATACCAGGAT	91	Yes
UBC	GATTTGGGTCGCGGTTCTT	TGCCTTGACATTCTCGATGGT	134	Yes
RPS13	CGAAAGCATCTTGAGAGGAACA	TCGAGCCAAACGGTGAACT	87	Yes
RPS23	TGGAGGTGCTTCTCATGCAA	AATGGCAGAATTTGGCTGTTTG	76	Yes
GAPD	TCTCCTCTGACTTCAACAGCGAC	CCCTGTTGCTGTAGCCAAATTC	126	Yes
ACTB	GCCCTGAGGCACTCTTCCA	CGGATGTCCACGTCACACTTC	100	Yes
CLTC	TCGCTACCTGGTACGTCGAAA	GCCTTTACAGTTACTGACACTTCTTCAG	150	Yes
NACA	GCAAAACAGAGTCGGAGTGAAAA	GTAACTCCTGTAACCTGCCGAAGA	77	Yes
SUI1	CAGGGTGACCAACGCAAGA	CACAAGCACTTAAAACCCATGAAC	96	Yes
TUBA6	GCAGACCCCTTCAAGTTCTAGTCA	GTAGAGCTCCCAGCAGGCATT	95	Yes

Real-Time RT-PCR

Real-time PCR analysis was performed using the primers shown in Table 2. Real-time RT-PCR was performed on ABI PRISM 7000 Sequence Detection System using the SYBR Green PCR Master Mix (Applied Biosystems). The PCR reaction consisted of 12.5 μl of SYBR Green PCR Master Mix, 300 nM of forward and reverse primers, and 2.0 μl of 1:20-diluted template cDNA in a total volume of 25 μl. Cycling was performed using the default conditions of the ABI Prism 7000 SDS Software 1.0: 2 min at 50°C, 10 min at 95°C, followed by 40 rounds of 15 s at 95°C and 1 min at 60°C. To verify that the used primer pair produced only a single product, a dissociation protocol was added after thermocycling, determining dissociation of the PCR products from 65°C to 95°C. The assay included a no-template control, a standard curve of four serial dilution points (in steps of 10-fold) of a cDNA mixture, and each of the test cDNAs.

The Model-Based Approach to Estimation of Expression Variation

The Model. Let y_{igj} be the log-transformed measured gene expression for gene i in the j'th sample of group g. We have a total of k genes and G groups, and the number of samples in group g is n_g.

A natural model is to write y_{igj} as a sum of three terms.

$$y_{igj} = \alpha_{ig} + \beta_{gj} + \varepsilon_{igj}, \qquad \text{(Eq. A)}$$

The first term, α_{ig}, represents the general expression level for candidate gene i within group g. The second term, β_{gj}, represents the amount of mRNA in the sample j. The last term, ϵ_{igj}, is the random variation caused by biological and experimental factors, with mean zero and variance σ_{ig}^2.

Our interest is to estimate the intragroup variation σ_{ig}^2 as well as the intergroup variation measured by the variation in α_{ig}, $g = 1, \ldots, G$.

Estimation of the Intragroup Variation. To estimate the intragroup variances, we form the usual residuals for an additive model, square these, and take the sum over the samples belonging to the group. Dividing the sum by $(n_g - 1)(1 - 2/k)$, we get the term s_{ig}^2, and the estimate of σ_{ig}^2 is

$$\hat{\sigma}_{ig}^2 = s_{ig}^2 - \frac{1}{k(k-1)} \sum_{v=1}^{k} s_{vg}^2 \qquad \text{(Eq. B)}$$

Most importantly, $\hat{\sigma}_{ig}^2$ is an unbiased estimate of σ_{ig}^2. For details on the derivation and properties of $\hat{\sigma}_{ig}^2$, see the supplementary data.

Estimation of the Intergroup Variation. We next turn to the problem of taking into account the intergroup variation. Let z_{ig} be the average of y_{igj} over all of the samples in group g. If θ_g is the average of β_{gj} over the samples in group g, the mean of z_{ig} is $\alpha_{ig} + \theta_g$, and the variance is σ_{ig}^2/n_g. With respect to intergroup variation, it is the variation of $z_{ig} - \theta_g$, $g = 1, \ldots, G$ that defines the suitability of gene i as a control gene.

Stability Value. Having estimated both the intra- and intergroup variation, we combine the two into a stability value, which intuitively adds the two

sources of variation and thus represents a practical measure of the systematic error that will be introduced when using the investigated gene.

Let α_i be the mean of α_{ig}. For a future experiment, it is the distribution of $z_{ig} - \theta_g - \alpha_i$ that defines the stability of gene i. However, we find a distribution to be an impractical stability measure and therefore reduce it to a one-dimensional value by defining the stability value ρ_{ig} to be the absolute value of the mean $+ 1$ SD. We use a Bayesian argument to find the distribution of $z_{ig} - \theta_g - \alpha_i$ in a future experiment (see supplementary data for details). Finally, to implement our stability measure, we need to estimate $z_{ig} - \theta_g - \alpha_i$. To this end, we make the assumption that the average of α_{ig} over the genes $i = 1, \ldots k$ is independent of the group g. Then $z_{ig} - \theta_g - \alpha_i$ is naturally estimated by $d_{ig} = z_{ig} - \bar{z}_{i\cdot} - \bar{z}_{\cdot g} + \bar{z}_{\cdot\cdot}$, where a bar indicates an average, and the stability value becomes

$$\rho_{ig} = \frac{\hat{\gamma}^2 |d_{ig}|}{\hat{\gamma}^2 + \hat{\sigma}_{ig}^2/n_g} + \sqrt{\hat{\sigma}_{ig}^2/n_g + \frac{\hat{\gamma}^2 \hat{\sigma}_{ig}^2/n_g}{\hat{\gamma}^2 + \hat{\sigma}_{ig}^2/n_g}} \qquad \text{(Eq. C)}$$

where γ^2 is the variance of α_{ig}. We estimate γ^2 by

$$\frac{1}{(k-1)(G-1)} \sum_{i=1}^{k} \sum_{g=1}^{G} d_{ig}^2 - \frac{1}{kG} \sum_{i=1}^{k} \sum_{g=1}^{G} \hat{\sigma}_{ig}^2/n_g \qquad \text{(Eq. D)}$$

if this is positive; otherwise, we set it to 0.

To get a single value for gene i, we let ρ_i be the average of ρ_{ig}, $g = 1, \ldots, G$. For details on the derivation and for a stability measure based on an average of several control genes, see the supplementary data.

Requirements. The validity of the approach is related to the number of samples and candidates analyzed, *i.e.*, the more samples and candidates, the better the estimates. The sample set should minimally contain 8 samples/group, and the number of candidates should be at least 3 for technical reasons, but 5–10 are recommended. It is a further requirement that the candidates are chosen from a set of genes with no prior expectation of expression difference between groups. This requirement is used to say that the average expression level is approximately the same in the different groups. Thus, instead of assuming the individual candidate genes to show no systematic intergroup variation, we assume the average of the candidate genes to show no systematic variation.

RESULTS AND DISCUSSION

Selection of Normalization Gene Candidates (from Array-Based Expression Profiles). The large and constantly growing amount of array-based expression profiling data represents an excellent source for identification of genes with minimal expression variation. Using the criteria specified in "Materials and Methods," we screened 99 bladder and 161 colon sample expression profiles for such genes. The genes meeting the criteria were, for the most part,

novel with respect to the normalization issue but also included a few known normalization genes, *e.g., UBC* and *ACTB*. Interestingly, the most commonly reported normalization gene, *GAPD*, did not meet the criteria in either of the tissues. However, for reasons of comparison, we decided to include *GAPD* in the list of candidates anyway.

Fourteen gene candidates for bladder and 13 gene candidates for colon were selected for additional studies (see Table 1 for full gene name, accession number, function, chromosomal location, and existence of pseudogenes; see Table 2 for primer sequences and indication of whether the primer pair is intron-spanning).

Special attention was paid to select genes whose proteins belong to different functional classes, which theoretically should reduce the chance that the genes might be coregulated. However, for the purpose of testing the robustness of our expression variance estimation approach, in two instances we included genes belonging to the same class (Table 1).

Evaluation of Candidates in Real-Time RT-PCR Experiment Designs. The expression level of the bladder and colon normalization gene candidates was determined in 28 and 40 samples, respectively. The raw expression values are available as a text file (see the supplementary data).

A Model-Based Approach for Estimation of Expression Variation. In our pursuit of suitable normalization genes, we first turned to previously reported approaches for evaluating normalization gene candidates (8, 13). However, we soon realized that none of these sufficiently evaluated the critical points outlined in the "Introduction," especially with respect to sample subgroups. Thus, we decided to develop a novel approach.

This approach entails application of a mathematical model to describe the expression values measured by RT-PCR, separate analysis of the sample subgroups, estimation of both the intra- and the intergroup expression variation, and calculation of a candidate gene "stability value." It is important to note that during the development of the model-based approach we made several choices with regard to the statistical methods used and that other choices could have been made that would have given a similar result.

A description, based on real data, of the relationship between the stability value and the intra- and intergroup expression variations as well as a study demonstrating the reproducibility of the strategy is available as supplementary data.

Ranking of the Candidate Normalization Genes. We have written a Visual Basic application for Microsoft Excel, termed Norm-Finder, which automatically calculates the stability value for all candidate normalization genes tested on a sample set containing any number of samples organized in any given number of groups. Norm-Finder is freely available from the authors on request. We used this application to rank all our colon and bladder normalization gene candidates according to their stability value (Table 3).

The top three ranked candidates were *UBC, GAPD*, and *TPT1* for colon and *HSPCB, TEGT*, and *ATP5B* for bladder (*i.e.*, of the tested candidates, these will introduce the least systematic error when used as normalization genes). Interestingly, the ranking order indicates that none of the known normalization genes among the candidates are suited for normalizing bladder cancer quantitative RT-PCR data. *GAPD* did not conform to the candidate selection criteria for either of the tissues; however, for colon, it is among the top three ranked candidates. Although seemingly surprising, this finding probably just reflects differences in the platforms used for quantification (quantitative RT-PCR and GeneChip arrays). We note in passing that *ACTB*, a gene commonly used for normalization, was ranked poorly in both tissues.

We next investigated the reproducibility of the variance estimations used to form the rankings. The expression levels of a number of the

Table 3 *Candidate genes ranked according to their expression stability—estimated using either the model-based or the pairwise comparison approach*[a]

Colon		Bladder	
Model-based	Pairwise	Model-based	Pairwise
UBC	RPS23 and TPT1	HSPCB	CFL1 and UBC
GAPD		TEGT	
TPT1	RPS13	ATP5B	ATP5B
UBB	SUI1	UBC	HSPCB
TUBA6	UBC	RPS23	GAPD
RPS13	GAPD	RPS13	TEGT
NACA	TUBA6	CFL1	RPS23
CFL1	UBB	FLJ20030	RPS13
SUI1	NACA	TPT1	TPT1
ACTB	CFL1	UBB	FLJ20030
CLTC	CLTC	FLOT2	FLOT2
RPS23	ACTB	GAPD	UBB
FLJ20030A	FLJ20030A	S100A6	ACTB
		ACTB	S100A6

[a] The candidates are listed with decreasing expression stability from top to bottom.

candidates were measured again in 12 T_a and 14 T_{2-4} bladder tumors (the raw data, assay design, and detailed description of this experiment are available as supplementary data). Comparison of the estimated inter- and intragroup variations with the estimations from the original experiment revealed a tight concordance, indicating a fine reproducibility (see Supplementary Fig. 1).

We then ranked the candidates using another recently published strategy to identify stably expressed genes; in the remainder of this article, this strategy is called the pairwise comparison approach (8). A comparison of the rankings produced by this and the model-based approach revealed them as different. The two rankings of the colon candidates showed no resemblance whatsoever, whereas the rankings of the bladder candidates showed some similarity (Table 3). These discrepancies are of course caused by the differences between the approaches. The model-based approach top ranks the candidates with minimal estimated intra- and intergroup variation, in contrast to the pairwise comparison approach, which tends to select those genes with the highest degree of similarity of the expression profile across the sample set (see "Robustness of the Model-Based Approach"). The latter approach implies that the candidates with minimal expression variation do not necessarily become top ranked. This we will demonstrate in the following using the T_a and T_{2-4} groups of our bladder data. Using these two groups, the intergroup variation corresponds to the difference between the average expression levels of the T_a and the T_{2-4} group. Furthermore, the estimated intragroup variations can be used to calculate a confidence interval for the difference. This is shown in Fig. 1.

Fig. 1 clearly demonstrates the different characteristics of the candidates top ranked by the two approaches. The model-based approach selects as the two best genes the candidates with minimal combined inter- and intragroup expression variation, whereas the pairwise comparison approach selects two genes with a low intragroup variation and roughly the same nonvanishing intergroup variation. To demonstrate the effect of using the different genes for normalization, we have drawn up two target genes with log difference between the group of −0.5 and 0.5. Using the candidates top ranked by the model-based approach for normalization will result in both these targets being identified as differentially expressed and with a similar-sized fold change, although in the opposite direction. Using the candidates top ranked by the pairwise approach, the target gene with log difference of −0.5 will not be detected as significantly differentially expressed, and the size of the fold change for the target gene with log difference of 0.5 will be erroneously overestimated.

The Model-Based *versus* the Pairwise Comparison Approach. We believe the model-based approach, with its account of sample groups and its direct estimation of expression variation, provides a
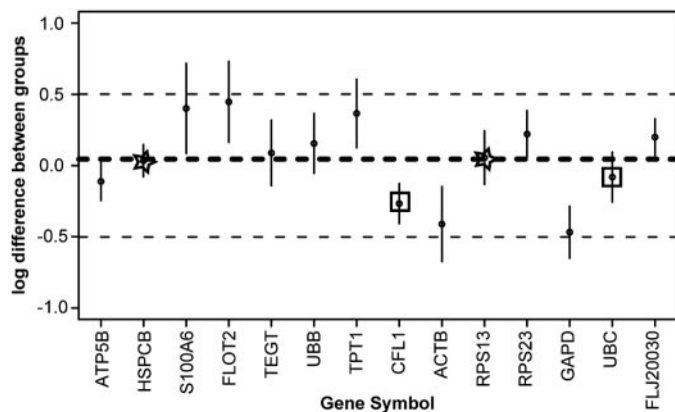
Fig. 1. Candidates top ranked by the model-based and pairwise comparison approaches. Using the T_a and T_{2-4} groups of bladder data, the intergroup variation is plotted (*small black circles*) as the expression difference between the two groups. The intragroup variation is indicated by *vertical bars* that give a confidence interval for the difference. Because the average of the intergroup variations is almost 0 (the *thick dashed line*), an intergroup variation of >0 implies that the gene shows systematically higher expression in T_a tumors than in T_{2-4} tumors and the opposite if the value is <0. The top two ranked candidate genes of the model-based approach are indicated by *stars*, and the top two ranked candidate genes of the pairwise comparison approach are indicated by *squares*. The *two thin dashed lines* represent hypothetical target genes with a log difference of 0.5 and -0.5.

more precise and robust measure of gene expression stability than does the pairwise comparison approach. The latter basically ranks the genes according to the similarity of their expression profiles. Imagine a situation in which the sample set consists of two sample subgroups, and all of the candidates but one show some difference between the groups; in this case, the optimal candidate with no difference between the groups would be excluded early on in the pairwise comparison approach, whereas in the model-based approach, it would present with the smallest stability value of all candidates. Furthermore, ranking genes according to the similarity of their expression profiles, as is done in the pairwise comparison approach, is problematic if there are co-regulated genes among the candidates. These will inadvertently have a tendency to show very similar expression profiles and thus, independent of their expression stability, to be top ranked. For a more detailed description of the differences between the model-based and the pairwise comparison approaches, see the supplementary data.

Robustness of the Model-Based Approach. To enable assessment of the robustness of the model-based approach toward correlation among the candidate genes, we deliberately sought to include a few coregulated pairs of genes in our list of candidates. This was done in two instances by including genes encoding proteins belonging to the same functional class (*UBB* and *UBC*; *RPS13* and *RPS23*).

A correlation analysis of our data (bladder and colon) revealed the expression of *RPS13* and *RPS23* to be correlated, whereas the expression of *UBB* and *UBC* was not correlated. A little to our surprise, the analysis also revealed the expression of the candidates *CFL1* and *ACTB* to be correlated. The protein encoded by *CFL1* acts as regulator of actin polymerization and depolymerization, thus it is not completely surprising that the expressions of the two genes correlate.

The above-mentioned analysis indicates that at present, it is very difficult, if not impossible, to foresee expression correlation of the candidates, implying that robustness toward expression correlation is of crucial importance. We will demonstrate this below using our colon data. To assess the robustness of the candidate rankings toward correlation among the candidate genes, we compared the results obtained using all candidate genes and the results obtained when excluding *RPS23* (Fig. 2). Fig. 2 clearly demonstrates that the model-based approach, in contrast to the pairwise comparison approach, was

not influenced by the exclusion of *RPS23*. Interestingly, the top three ranked candidates of the pairwise comparison approach after *RPS23* exclusion include *CFL1* and *ACTB* (the other candidate pair with correlated expression), emphasizing the tendency of the approach to top rank candidates with correlated expression rather than minimal variation.

Taken together, these results demonstrate that candidate coregulation does not significantly affect the model-based approach and equally clearly demonstrate that sensitivity to coregulation is a major weakness of the pairwise comparison approach.

Normalization Factor (NF). In situations where no optimal normalization gene has been found, it may be prudent to normalize the data using a NF based on multiple normalization genes rather than a single gene. The rationale is simple; the variation in the average of multiple genes is smaller than the variation in individual genes. Despite this fact, the use of a NF does not necessarily imply improved normalization. Only if the NF is generated from a cautiously selected set of genes can an improvement be expected. Furthermore, compared with a single normalization gene, a NF has the practical implication that multiple normalization genes have to be measured each time instead of only one. This may be impractical, particularly when only few target genes need to be studied, or when limited amounts of RNA are available. Thus, the number of genes to include in the NF must be carefully considered. Intuitively, the number of genes will be a trade-off between practical considerations and minimizing the variation in the NF.

Usually, the use of NFs is only considered in situations in which the available normalization gene candidates show significant expression variation. This variation will typically involve both intra- and intergroup variation. In particular, the intergroup variation is critical for the NF. If the genes are not cautiously selected, the intergroup variation will not be eliminated by the NF but transferred to it. We will demonstrate this using the data displayed in Fig. 1. Imagine a NF
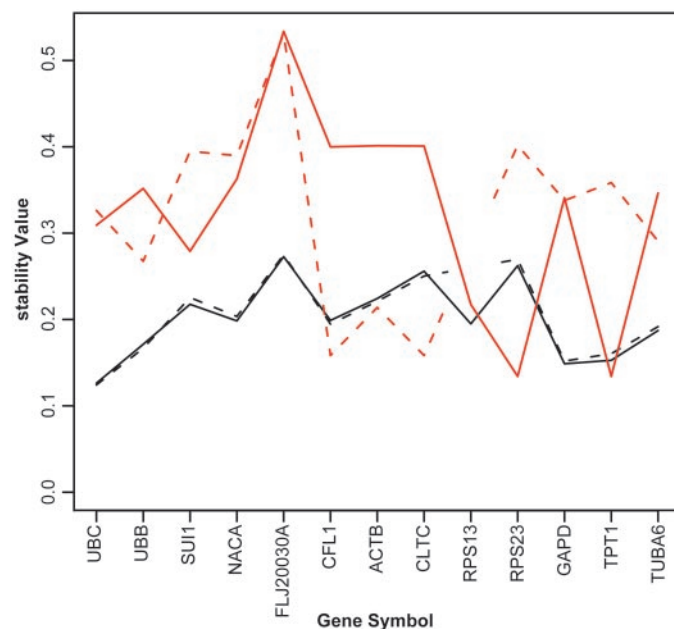

Fig. 2. Robustness of the model-based and the pairwise comparison approaches. To assess the robustness of the candidate rankings toward correlation among the candidate genes, we compared the results obtained using all candidate genes (*solid lines*) and the results obtained when excluding *RPS23* (*dashed lines*). The *black curves* represent the stability values of the model-based approach, and the *red curves* represent the values used in the sequential step of the pairwise comparison approach (precisely, these represent the average pairwise SD for each gene at the time of its exclusion in the sequential procedure). The top ranked genes are those with the smallest values for each method.

generated on the basis of the three genes *S100A6, FLOT2,* and *TPT1,* all of which show intergroup variation oriented in the same direction (*i.e.*, slightly higher expression in the T_a tumors than the T_{2-4} tumors). This NF represents no improvement compared with the individual genes because their intergroup variation is transferred to it. Thus, even though the overall variation of the NF is reduced, its use would still cause the differential expression of the target gene 0.5 log difference between the groups to be missed. If, instead, the NF was based on genes with opposite-directed intergroup variation, such as *S100A6, TPT1,* and *ACTB,* it would represent an improvement, and both target genes would correctly be identified as differentially expressed.

It is clear that a strict gene selection procedure is necessary to generate functional NFs and that knowledge of potential intergroup expression variation is a fundamental prerequisite.

The model-based approach outlined above provides all of the measures needed for qualified selection of NF genes. First of all, the estimated intergroup expression variations allow for selection of the suited genes. Secondly, the intragroup variance estimates provide a natural way of identifying the number of genes to include. The optimal number of genes is reached when addition of a further gene leads to a negligible reduction in the average of the gene variance estimates.

In conclusion, we have described a procedure that directly and robustly evaluates gene expression stability. The expression variation of each evaluated candidate is directly estimated. This enables the user to put the candidates in context because the estimated variation directly indicates the introduced error associated with using them. Compared with previously published procedures for identifying suitable normalization genes, the present approach provides more direct measures, takes into account systematic differences between sample subgroups, and is less affected by correlated expression of the candidate genes. We used the approach to identify genes suited to normalize quantitative RT-PCR data from colon cancer and bladder cancer.

These genes are *UBC, GAPD,* and *TPT1* for colon and *HSPCB, TEGT,* and *ATP5B* for bladder.

The variance estimation approach outlined in this article can be applied to evaluate the suitability of any normalization gene candidate in any kind of experimental design and should allow more reliable normalization of RT-PCR data.

REFERENCES

1. Bustin SA. Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. J Mol Endocrinol 2000;25:169–93.
2. Bustin SA. Quantification of mRNA using real-time reverse transcription PCR (RT-PCR): trends and problems. J Mol Endocrinol 2002;29:23–39.
3. Schmittgen TD, Zakrajsek BA. Effect of experimental treatment on housekeeping gene expression: validation by real-time, quantitative RT-PCR. J Biochem Biophys Methods 2000;46:69–81.
4. Suzuki T, Higgins PJ, Crawford DR. Control selection for RNA quantitation. Biotechniques 2000;29:332–7.
5. Thellin O, Zorzi W, Lakaye B, et al. Housekeeping genes as internal standards: use and limits. J Biotechnol 1999;75:291–5.
6. Tricarico C, Pinzani P, Bianchi S, et al. Quantitative real-time reverse transcription polymerase chain reaction: normalization to rRNA or single housekeeping genes is inappropriate for human tissue biopsies. Anal Biochem 2002;309:293–300.
7. Warrington JA, Nair A, Mahadevappa M, Tsyganskaya M. Comparison of human adult and fetal expression and identification of 535 housekeeping/maintenance genes. Physiol Genomics 2000;2:143–7.
8. Vandesompele J, De Preter K, Pattyn F, et al. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol 2002;3:RESEARCH0034.
9. Dyrskjot L, Thykjaer T, Kruhoffer M, et al. Identifying distinct classes of bladder carcinoma using microarrays. Nat Genet 2003;33:90–6.
10. Frederiksen CM, Knudsen S, Laurberg S, Orntoft TF. Classification of Dukes' B and C colorectal cancers using expression arrays. J Cancer Res Clin Oncol 2003;129:263–71.
11. Bolstad BM, Irizarry RA, Astrand M, Speed TP. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics 2003;19:185–93.
12. Irizarry RA, Bolstad BM, Collin F, et al. Summaries of Affymetrix GeneChip probe level data. Nucleic Acids Res 2003;31:e15.
13. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(−Delta Delta C(T)) method. Methods 2001;25:402–8.
14. Kent WJ. BLAT: the BLAST-like alignment tool. Genome Res 2002;12:656–64.