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, Jens Ledet Jensen and Torben Falck Ørntoft

Cancer Res 2004;64:5245-5250.

Updated version
Access the most recent version of this article at:
http://cancerres.aacrjournals.org/content/64/15/5245

Supplementary Material
Access the most recent supplemental material at:
http://cancerres.aacrjournals.org/content/suppl/2004/08/24/64.15.5245.DC1.html

Cited Articles
This article cites by 14 articles, 6 of which you can access for free at:
http://cancerres.aacrjournals.org/content/64/15/5245.full.html#ref-list-1

Citing articles
This article has been cited by 100 HighWire-hosted articles. Access the articles at:
http://cancerres.aacrjournals.org/content/64/15/5245.full.html#related-urls

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.
Normalizing 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 Ørntoft1

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

ABSTRACT

Accurate normalization is an absolute prerequisite for correct measurements 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 variation estimation procedure reported recently.

MATERIALS AND METHODS

GeneChip Expression Profiling Data

Expression profiling data from 99 bladder (normal and T2–4, 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 to 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 T8, 8 Tp, and 10 Tp- 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**

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Gene name</th>
<th>Accession no.</th>
<th>Locus link</th>
<th>Function</th>
<th>Location</th>
<th>Pseudogenes</th>
<th>Tissue</th>
</tr>
</thead>
<tbody>
<tr>
<td>FLOT2</td>
<td>Flotillin 2</td>
<td>NM_004475.1</td>
<td>2319</td>
<td>May act as a scaffolding protein within caveolar membranes. May be involved in epidermal cell adhesion and epidermal structure and function.</td>
<td>17q11.2</td>
<td>No B</td>
<td></td>
</tr>
<tr>
<td>ATP5B</td>
<td>ATP synthase, H+ transporting, mitochondrial F1 complex, β-polypeptide</td>
<td>NM_001686.1</td>
<td>506</td>
<td>Produces ATP from ADP in the presence of a proton gradient across the membrane. The β chain is the catalytic subunit.</td>
<td>17q13.3</td>
<td>Yes B</td>
<td></td>
</tr>
<tr>
<td>HSPCB</td>
<td>Heat shock 90-kDa protein 1, β</td>
<td>EST0000265413</td>
<td>3326</td>
<td>Molecular chaperone. Has ATPase activity (by similarity).</td>
<td>6p21.1</td>
<td>Yes B</td>
<td></td>
</tr>
<tr>
<td>S100A6</td>
<td>S100 calcium-binding protein A6 (calcyclin)</td>
<td>EST0000292162</td>
<td>6277</td>
<td>May function in stimulation of Ca²⁺-dependent insulin release, stimulation of prolactin secretion, and exocytosis.</td>
<td>1q21.3</td>
<td>Yes B</td>
<td></td>
</tr>
<tr>
<td>TEGT</td>
<td>Testis enhanced gene transcript (BAX inhibitor 1)</td>
<td>EST0000267115</td>
<td>5009</td>
<td>Suppressor of apoptosis</td>
<td>12q13.12</td>
<td>No B</td>
<td></td>
</tr>
<tr>
<td>CFL1</td>
<td>Collin 1 (non-muscle)</td>
<td>EST0000308162</td>
<td>1072</td>
<td>Controls reversibly actin polymerization and depolymerization. It is the major component of intranuclear and cytoplasmic actin rods.</td>
<td>11q13.1</td>
<td>Yes B and C</td>
<td></td>
</tr>
<tr>
<td>FLJ20030</td>
<td>Hypothetical protein FLJ20030</td>
<td>EST0000021832</td>
<td>54789</td>
<td>Unknown</td>
<td>Xp11.22</td>
<td>No B and C</td>
<td></td>
</tr>
<tr>
<td>TPT1</td>
<td>Tumor protein, translationally controlled 1</td>
<td>EST00000255477</td>
<td>7187</td>
<td>Unknown</td>
<td>13q14.13</td>
<td>Yes B and C</td>
<td></td>
</tr>
<tr>
<td>UBB</td>
<td>Ubiquitin B</td>
<td>NM_018955.1</td>
<td>7314</td>
<td>Protein degradation</td>
<td>17p11.2</td>
<td>Yes B and C</td>
<td></td>
</tr>
<tr>
<td>UBC</td>
<td>Ubiquitin C</td>
<td>EST0000260590</td>
<td>7316</td>
<td>Protein degradation</td>
<td>12q24.31</td>
<td>Yes B and C</td>
<td></td>
</tr>
<tr>
<td>RPS1F</td>
<td>Ribosomal protein S13</td>
<td>NM_001017.2</td>
<td>6207</td>
<td>Encodes a ribosomal protein that is a component of the 43S subunit</td>
<td>11p15.1</td>
<td>Yes B and C</td>
<td></td>
</tr>
<tr>
<td>RPS2/F</td>
<td>Ribosomal protein S23</td>
<td>NM_001025.1</td>
<td>6228</td>
<td>Encodes a ribosomal protein that is a component of the 40S subunit</td>
<td>5p14.2</td>
<td>Yes B and C</td>
<td></td>
</tr>
<tr>
<td>GAPD</td>
<td>Glyceraldehyde-3-phosphate dehydrogenase</td>
<td>NM_002046</td>
<td>2597</td>
<td>Catalyzes the reversible oxidative phosphorylation of glyceraldehyde-3-phosphate</td>
<td>12p13.31</td>
<td>Yes B and C</td>
<td></td>
</tr>
<tr>
<td>ACTB</td>
<td>Actin, β</td>
<td>NM_001101.2</td>
<td>60</td>
<td>Cytoskeletal structural protein involved in various types of cell motility</td>
<td>7p22.1</td>
<td>Yes B and C</td>
<td></td>
</tr>
<tr>
<td>CLTC</td>
<td>Clathrin, heavy polypeptide (Hc)</td>
<td>NM_004859.1</td>
<td>1213</td>
<td>Clathrin is the major protein of the polyhedral coat of coated pits and vesicles.</td>
<td>17q23.2</td>
<td>Yes C</td>
<td></td>
</tr>
<tr>
<td>NACA</td>
<td>Nascent-polypeptide-associated complex alpha polypeptide</td>
<td>EST00000228312</td>
<td>4666</td>
<td>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</td>
<td>12q13.3</td>
<td>Yes C</td>
<td></td>
</tr>
<tr>
<td>ST11</td>
<td>Putative translation initiation factor</td>
<td>EST0000031083</td>
<td>10209</td>
<td>May act as a negative growth regulator</td>
<td>17q21.2</td>
<td>Yes C</td>
<td></td>
</tr>
<tr>
<td>TUBA6</td>
<td>Tubulin alpha 6</td>
<td>EST0000301072</td>
<td>84790</td>
<td>Cytoskeletal structural protein</td>
<td>12q13.12</td>
<td>Yes C</td>
<td></td>
</tr>
</tbody>
</table>

* 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/).

* Ensembl cytogenetic band.

* Determined by BLAT search of Human June 2002 Freeze (Ref. 14).

* Identified from bladder (B) and colon (C).

* Genes belonging to the same functional class.

* Genes belonging to the same functional class.
Table 2. Primer sequences for candidate normalization genes

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Forward primer</th>
<th>Reverse primer</th>
<th>Amplicon size</th>
<th>Intron spanning</th>
</tr>
</thead>
<tbody>
<tr>
<td>FLOT2</td>
<td>TGCCCAGGTTGACAGGAAGA</td>
<td>GGTGTCGCGCCCATGAACTTCAACA</td>
<td>115</td>
<td>Yes</td>
</tr>
<tr>
<td>ATPSB</td>
<td>TCACCCAGGCTGTTGACAGGA</td>
<td>AGTGGCCAAAGTGAGCTGAT</td>
<td>80</td>
<td>Yes</td>
</tr>
<tr>
<td>HSPCB</td>
<td>AAGAGGAGCAAGGGGACAACTTGGGAG</td>
<td>TGGTCAAGTACTGATCGGAACTT</td>
<td>120</td>
<td>Yes</td>
</tr>
<tr>
<td>S100A6</td>
<td>ACAAGGCAACCTGGACAA</td>
<td>CCATCAGCCTGTCATTTCC</td>
<td>99</td>
<td>Yes</td>
</tr>
<tr>
<td>TEG7</td>
<td>TGCCTGCTGGCATTTCCCTACA</td>
<td>ACGCGGCTCTGGCATATA</td>
<td>151</td>
<td>Yes</td>
</tr>
<tr>
<td>CFI1</td>
<td>GAAGGAGGATGCCCGGTGTTTACCTT</td>
<td>GCTGCCATACAATTTGGTGCTCT</td>
<td>73</td>
<td>Yes</td>
</tr>
<tr>
<td>FLJ20290</td>
<td>GCCCTTGAGCTCCAGAATTGTT</td>
<td>ATAAATGATTTGTCCTAGTGCATTC</td>
<td>144</td>
<td>Yes</td>
</tr>
<tr>
<td>TPI1</td>
<td>GATCAGCGGAGCTTTGTGT</td>
<td>TTTCCACGGGACTTTCTTCTT</td>
<td>100</td>
<td>Yes</td>
</tr>
<tr>
<td>UBB</td>
<td>GGGCCGTTGGCTTTGTT</td>
<td>GAACCTGTACCCGATACCAAGAAT</td>
<td>91</td>
<td>Yes</td>
</tr>
<tr>
<td>UBC</td>
<td>GATGTGGCTGAGCCGTCCTT</td>
<td>TGCGCTGACCTTCCGTAGT</td>
<td>134</td>
<td>Yes</td>
</tr>
<tr>
<td>RPS13</td>
<td>CGGAGAATGCTGACTGACGAGAAC</td>
<td>TGGAGCCAAACGCTGATCTC</td>
<td>87</td>
<td>Yes</td>
</tr>
<tr>
<td>RPS23</td>
<td>TGGAGGTTCTGCTCTCATTGAA</td>
<td>AATTCAGGCTTTTGGTTGTT</td>
<td>182</td>
<td>Yes</td>
</tr>
<tr>
<td>GAPD</td>
<td>TCTCTCCTGACTCAACAGGCA</td>
<td>CCGCTTGGTCTGAGCAAAATT</td>
<td>126</td>
<td>Yes</td>
</tr>
<tr>
<td>ACTB</td>
<td>GCCCTGAGGCTCCACCTCCTCA</td>
<td>GACCTGTACCTTGGACACCTTTC</td>
<td>100</td>
<td>Yes</td>
</tr>
<tr>
<td>CLTC</td>
<td>TGCTAAGCTCGCTCAGTGGC</td>
<td>GCTTTACAGTTACTGACACTTCTTC</td>
<td>150</td>
<td>Yes</td>
</tr>
<tr>
<td>NACA</td>
<td>GCAAAAGCAAGTCCGGAGATGAAAAA</td>
<td>GTAATCCCTGTAACCTGCCGAAAGA</td>
<td>77</td>
<td>Yes</td>
</tr>
<tr>
<td>SUI</td>
<td>CAGGTTGACAAAGCGAAAGA</td>
<td>CACAAGGACTTAAAACACCTTGAAC</td>
<td>96</td>
<td>Yes</td>
</tr>
<tr>
<td>TUBA5</td>
<td>GCAGACCCCTCTAGTTCATCTCAGTCA</td>
<td>GTAGAGCCCTACCAAGGGCAT</td>
<td>95</td>
<td>Yes</td>
</tr>
</tbody>
</table>

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 performed after thermocycling, determining dissociation of the PCR products. The assay included a no-template control, a standard curve 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. The last term, \( y_ig \), is the random variation caused by biological variation. The second term, \( \theta_ig \), is the average of \( y_igj \), where \( j = 1, \ldots, k \) is a group of the gene. The variation in the fold change is \( \theta_ig \), which is naturally the absolute value of the mean of array-based expression profiling data and the number of 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, 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 Variance. 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 NormFinder, 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. NormFinder 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 candidates were measured again in 12 T4 and 14 T2–4 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 T4 and T2–4 groups of our bladder data. Using these two groups, the intergroup variation corresponds to the difference between the average expression levels of the T4 and the T2–4 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
expression of RPS13 in T\textsubscript{A} tumors than in T\textsubscript{2} intergroup variation of and two thin dashed lines. The squares ranked candidate genes of the pairwise comparison 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. 1. Candidates top ranked by the model-based and pairwise comparison approaches. Using the T\textsubscript{A} and T\textsubscript{2} 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\textsubscript{A} tumors than in T\textsubscript{2} 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.](image1)

![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.](image2)
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 $T_3$ 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