

# DNA and RNA References for qRT-PCR Assays in Exfoliated Cervical Cells

Martin Steinau, Mangalathu S. Rajeevan, and  
Elizabeth R. Unger

From the Human Papillomavirus Program, Centers for Disease  
Control and Prevention, Atlanta, Georgia

**The noncritical use of housekeeping genes, RNA mass, or cell number for normalization in quantitative reverse transcriptase-polymerase chain reaction (qRT-PCR) assays has come under scrutiny in recent years, highlighting the need to evaluate references in the immediate context of the relevant samples and experimental design. The purpose of this study was to select appropriate references for normalizing qRT-PCR assays of gene expression in exfoliated cervical cells. We used total nucleic acid extracts from 30 samples, representing the full spectrum of pre-invasive cervical neoplasia. We determined the DNA content by quantitative PCR for the single-copy gene  $\beta$ -globin and total RNA content using quantitative image analysis of ribosomal bands. In addition, qRT-PCR for 13 candidate housekeeping genes was performed. We used two analysis methods, geNorm and NormFinder, to identify the best combination of reference genes and then correlated housekeeping gene expression with DNA content and gel representation of ribosomal RNA. *ACTB* was the most stable single gene. The addition of *PGK1* and *RPLP0* increased the robustness in qRT-PCR applications not stratified by disease. These genes also showed the highest correlation with DNA contents in the same samples. If special attention to intraepithelial lesions is appropriate, *RPL4* and *PGK1* are recommended as the best combination of two genes. (J Mol Diagn 2006, 8:113–118; DOI: 10.2353/jmoldx.2006.050088)**

Quantitative measures of gene expression between samples require some form of normalization to a reference that provides a common basis for the comparison, essentially controlling for amount of starting material on the basis of cell number (ie, DNA content) or amount of RNA or specific transcript. The reference should be invariable and should especially not be affected by variables included in the experimental design to avoid introducing systematic errors. Most methods of quantitative reverse transcriptase-polymerase chain reaction (qRT-PCR) use one or more so called “housekeeping genes” whose expression is considered to be stable. A number of re-

cent papers have indicated the inappropriateness of any single reference gene for normalization and highlighted the need for validating the status of references for each type of specimen analyzed.<sup>1–5</sup> A housekeeping gene may be quite stable in one sample or tissue source and vary considerably in another. In fact, the specific conditions within an experiment or observational study may also have an effect on any kind of reference.<sup>6</sup>

The purpose of this study was to select appropriate references for normalizing qRT-PCR assays of gene expression in exfoliated cervical cells. We have been investigating the gene expression of exfoliated cervical cells using cDNA microarrays to identify differential expression associated with cervical disease.<sup>7</sup> The terminal differentiation of exfoliated cells, partial RNA degradation,<sup>8</sup> and presence of neoplastic lesions must be accommodated in determining references for gene expression. No data are available on the stability of common reference genes in this type of sample. We examined total RNA content, DNA content, and the expression of 13 candidate housekeeping genes as references for qRT-PCR of exfoliated cells.

Total RNA mass determined either by UV photospectrometry or by densitometric assessment has been widely used as a reference to normalize the amount of RNA template. Its accuracy and suitability as a reference for highly sensitive real-time PCR technology is nonetheless questionable because mRNA makes up less than 1% of the total RNA, and the approach assumes that mRNA is a constant proportion of RNA in all cells. In addition, differences in the efficiency of the enzymatic reactions of qRT-PCR are not accounted for (ie, presence of inhibitors will not be detected).

Specimen cell number is seen as a universal reference for samples. However, the impracticality of quantitative assessment of cell number in solid tissues or other complex samples has limited its use.<sup>1,9</sup> Because the cellular DNA content is generally constant, DNA quantitation using qRT-PCR could be used to determine cell number. DNA was explored for normalization of microarray gene expression studies in bacteria, but no data are available for human

---

Supported in part by the Early Detection Research Network Interagency Agreement of the National Cancer Institute (grant Y1-CN-0101-01).

Accepted for publication September 13, 2005.

The findings and conclusions in this report are those of the authors and do not necessarily represent the views of the funding agency.

Address reprint requests to Elizabeth R. Unger, Centers for Disease Control and Prevention, 1600 Clifton Rd., MS G41, Atlanta, GA 30333. E-mail: eunger@cdc.gov.

gene expression profiling by PCR.<sup>10</sup> Potential limitations of this approach include variation in the DNA-to-RNA ratio due to differentiation or disease and alterations in cellular DNA content due to polyploidy or aneuploidy. As with total RNA mass quantitation methods, sample-specific differences in tissue-borne inhibitors potentially affecting the enzymatic downstream processes are not measured.

Gene expression variation in cervical exfoliated cells might be indicative of neoplasia, and its detection holds promise as early cancer detection assays.<sup>7</sup> To empirically evaluate references for qRT-PCR in this specimen, we used total nucleic acid extracts (TNAs; both DNA and RNA) from 30 samples, representing the full spectrum of preinvasive cervical neoplasia. We determined the number of cells by measuring the DNA content by PCR for the single-copy gene  $\beta$ -globin quantitative and RNA content by using quantitative image analysis of ribosomal bands in denaturing agarose gels. In addition, we performed qRT-PCR for 13 genes selected as housekeeping genes in other studies. We used the Microsoft Excel add-ins geNorm<sup>1</sup> and Norm-Finder<sup>5</sup> to identify the best combination of housekeeping genes as references and correlated their expression with DNA content (ie, cell number) and gel representation of ribosomal RNA in the same samples.

## Materials and Methods

### Sample Collection and TNA Extraction

Samples were selected from an ongoing study of cervical neoplasia in women enrolled at the time of colposcopy. Specimen collection and processing was performed as previously described.<sup>11</sup> Briefly, endo- and ectocervical cells were collected in PreservCyt (Cytoc Corp, Marlborough, MA) and extracted with the MasterPure TNA isolation kit (Epicenter, Madison, WI). The TNAs, including both DNA and RNA, were stored at  $-70^{\circ}\text{C}$  until use.

Cervical disease status was determined based on the summary results of cytology, colposcopy, and biopsy examination. The 30 specimens used in this study were selected to represent women without abnormalities ( $n = 8$ ) and three grades of cervical intraepithelial neoplasia (CIN): CIN1 ( $n = 9$ ), CIN2 ( $n = 7$ ), and CIN3 ( $n = 6$ ). These samples also represented the spectrum of human papillomavirus (HPV) infection (HPV16 positive, other types of HPV, and no HPV detected), age (18 to 58 years), and ethnicity (black, white, Hispanic, and other) in the study.

### Quantification of Total RNA

The quality of the isolated TNAs was evaluated with spectrophotometry and ethidium bromide-stained denaturing agarose electrophoresis. Quantitation of the total RNA extracted from each sample was assessed by densitometric measurement (FluorChem Digital Imaging System; Alpha Innotech, Inc., San Leandro, CA) of the ribosomal bands visualized on ethidium bromide-stained denaturing agarose gels in comparison with a standard 28S and 18S control

marker. The standard was highly purified total RNA prepared from cultured Caski cells, quantitated using UV spectrophotometry.

### cDNA Synthesis

In 0.2-ml thin-wall PCR tubes (Robbins Scientific Corp., Sunnyvale, CA), equal volumes (2.5  $\mu\text{l}$ ) of each sample were treated with 5 U of DNase I (GenHunter Corp., Nashville, TN) in a 10- $\mu\text{L}$  reaction with 1 $\times$  RT buffer (Invitrogen Corp., Carlsbad, CA) for 30 minutes at  $37^{\circ}\text{C}$ . We removed 1  $\mu\text{l}$  to be tested for residual DNA (no-RT control). A master mix of primers and exogenous plant gene spike chlorophyll A-B binding protein (CAB) (Stratagene, La Jolla, CA) was prepared. Aliquots were added to each sample (final concentrations per sample were 300 ng of random primer, 50 ng of oligo-T<sub>12-18</sub> [Invitrogen], and 0.1 pg of CAB mRNA), and the mixtures were heated at  $65^{\circ}\text{C}$  for 5 minutes and then transferred to ice. After 1 minute on ice, the reaction mix was added (2  $\mu\text{l}$  of RT buffer, 2  $\mu\text{l}$  of dithiothreitol [100 mmol/L], 2  $\mu\text{l}$  of dNTPs [10 mmol/L], 1  $\mu\text{l}$  of dH<sub>2</sub>O, and 1  $\mu\text{l}$  of Superscript III reverse transcriptase [Invitrogen]) to the final reaction volume of 20  $\mu\text{l}$ . Samples were incubated at  $25^{\circ}\text{C}$  for 5 minutes,  $50^{\circ}\text{C}$  for 50 minutes, and  $70^{\circ}\text{C}$  for 15 minutes. All incubations were performed in a thermocycler. The product was diluted 1:5 in diethylpyrocarbonate-treated water and stored at  $-20^{\circ}\text{C}$  in 25- $\mu\text{l}$  aliquots until further processing.

### Quantitative Real-Time RT-PCR

We developed quantitative SYBR green PCR assays for the 13 endogenous "housekeeping" genes shown in Table 1, along with assays for spiked CAB and  $\beta$ -globin DNA. We generated gene-specific primer sequences with the Primer Select application of Lasergene software (DNA Star, Madison, WI) (Table 2). We tested the specificity of the amplification conditions for each primer pair by melting curve analysis and by verifying the size of the amplicon on gel electrophoresis. We used cDNA from universal human reference RNA (Stratagene) in a four-step, 10-fold dilution series to calculate the PCR efficiency for each assay over a 1000-fold range of dilution (Table 2).

PCR reactions contained 12.5  $\mu\text{l}$  of 2 $\times$  SYBR Green I Master Mix buffer (Applied Biosystems, Foster City, CA), 6.5  $\mu\text{l}$  of diethylpyrocarbonate-treated water, 2  $\mu\text{l}$  of each forward and reverse primer (20  $\mu\text{mol/L}$ ), and 2  $\mu\text{l}$  of the diluted cDNA template. Amplification was performed in an ABI Prism 7900HT (Applied Biosystems) with the following cycling conditions: 1 cycle of 10 minutes at  $95^{\circ}\text{C}$  and 45 cycles of 15 seconds at  $95^{\circ}\text{C}$ , 15 seconds at  $60^{\circ}\text{C}$ , and 45 seconds at  $72^{\circ}\text{C}$ . All reactions were run in duplicate with a no-template control for each run. A no-RT control was amplified from each sample to control for remaining DNA contaminations. Threshold values ( $C_t$ ) were acquired at  $\Delta R_n = 0.1$  and exported as tab-delimited text files.

**Table 1.** Genes Amplified

Gene name	Symbol	GenBank accession no.
$\beta$ -Globin region on chromosome 11	<i>HBB</i>	NG_000007
Phosphoglycerate kinase 1	<i>PGK1</i>	NM_000291
Ribosomal protein L4	<i>RPL4</i>	NM_000968
Large ribosomal protein P0	<i>RPLP0</i>	NM_001002
$\beta$ -Actin	<i>ATCB</i>	NM_001101
Eukaryotic elongation factor 1 $\alpha$ 1	<i>EEF1A1</i>	NM_001402
Eukaryotic translation elongation factor 1 $\gamma$	<i>EEF1G</i>	NM_001404
Glyceraldehyde-3-phosphate dehydrogenase	<i>GAPD</i>	NM_002046
TATA box binding protein	<i>TBP</i>	MN_003194
$\beta$ -2-Microglobulin	<i>B2M</i>	NM_004048
Succinate dehydrogenase complex A	<i>SDHA</i>	NM_004168
Muscleblind-like 2	<i>MBNL2</i>	NM_144778
<i>Arabidopsis thaliana</i> chlorophyll A-B binding protein 2	<i>CAB</i>	NM_115346
28S Ribosomal RNA	<i>28S</i>	M11167
18S Ribosomal RNA	<i>18S</i>	X03205

### DNA Quantification (Cell Number)

We determined the DNA content in the same TNA samples using quantitative PCR for the single-copy  $\beta$ -globin gene (*HBB*). Because DNA content is context independent and with few exceptions is the same in every cell, its  $C_t$  values directly correlate with the number of cells that were processed from the original specimens. One microliter of TNAs was amplified with *HBB* specific primers (Table 2), applying the SYBR green PCR assay and data acquisition as described above.

### Analysis

We required the coefficient variation (CV) of the  $C_t$  values for the spiked CAB among all samples to be less than 2% to control the experimental variability during cDNA synthesis. Duplicate  $C_t$  values for each candidate reference gene were averaged for each of the 30 samples, and the CVs had to be below 1% for a sample to be included in the analysis. The PCR efficiencies were calculated as  $e = 10^{-1/\text{slope}}$  for each primer pair, with the slope determined by a linear regression model over  $\log_{10}$ -transformed  $C_t$  values of the template dilution series described above.

To evaluate the reference genes that are most suitable for RT-PCR normalization, we applied two previously published Microsoft Excel-based applications: 1) geNorm<sup>1</sup> calculates a gene stability measure as the SD of the  $\log_2$ -transformed expression ratios of each housekeeping gene with all others tested throughout the samples. 2) Norm-Finder<sup>5</sup> uses a model-based approach to estimate expression stability based on intra- and intergroup variations for candidate housekeeping genes. We applied a disease model comprised of three groups with different degrees of abnormality: no disease (CIN0;  $n = 8$ ), mild dysplasia (CIN1;  $n = 9$ ), and moderate to severe dysplasia (CIN2/CIN3;  $n = 13$ ). Quantitative relationships between RNA transcripts and DNA content in the samples were analyzed via Pearson correlation in Microsoft Excel.

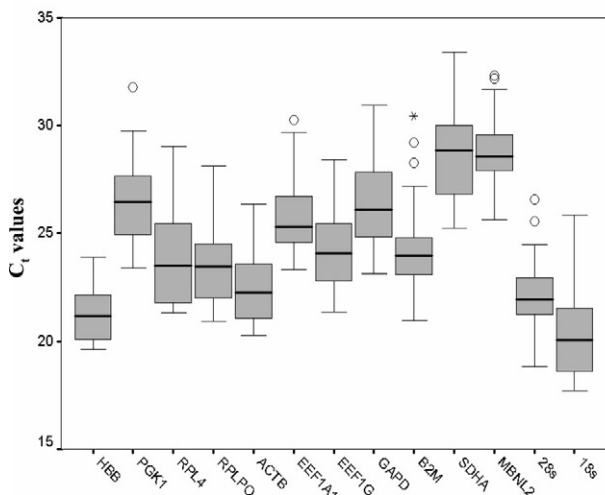
### Results

RNA from all samples in this study displayed distinct intact 28S and 18S ribosomal bands after separation by gel electrophoresis. The total RNA mass as determined by densitometric measures was between 194 and 713 ng/ $\mu$ l. Aliquots collected after DNase and before RT treat-

**Table 2.** Primer Sequences, Amplicon Length, and PCR Efficiency for All SYBR Green I Real-Time PCR Assays\*

Gene	Forward primer sequences 5' $\rightarrow$ 3'	Reverse primer sequences 5' $\rightarrow$ 3'	Amplicon length (bp)	PCR efficiency (%)
<i>HBB</i>	CAGGTACGGCTGTCATCACTTAGA	CATGGTGTCTGTTTGAGGTTGCTA	183	95.0
<i>PGK1</i>	ATTAGCCGAGCCAGCCAAAATAG	TCATCAAAAACCCACCAGCCTTCT	152	98.5
<i>RPL4</i>	GCTCTGGCCAGGGTGCTTTTG	ATGGCGTATCGTTTTGGGTTGT	154	100
<i>RPLP0</i>	GCTGTGTCGCCGTGCTGGTG	TGGTGCCCTGGAGATTTTAGTGG	130	100
<i>ATCB</i>	ATGTGGCCGAGGACTTTGATT	AGTGGGGTGGCTTTTAGGATG	107	95.5
<i>EEF1A1</i>	TGCGGTGGGTGTCATCAAA	AAGAGTGGGGTGGCAGGTATTG	123	100
<i>EEF1G</i>	TCTGGTGGGCGAACGAGTGACA	GGAAAGGCCCTGGCGGAAAGAAG	102	97.5
<i>GAPD</i>	TCCACCACCCTGTGCTGTA	ACCACAGTCCATGCCATCAC	451	95.0
<i>TBP</i>	GCCTCCCCCACCCTTCTTT	GCCACACCCTGCAACTCAACATCC	106	95.5
<i>B2M</i>	CACCCCCACTGAAAAGATG	ATATTAAGGCAAGCAAGCAGAA	167	98.0
<i>SDHA</i>	GGCGGCATTCCCACTACTACA	GCACATGCCCGACCAAGACAA	173	99.5
<i>MBNL2</i>	GTCACGTCCCCGGGCTCAACTGC	ATGGTGCTGTCTGCGGGGTGTG	137	99.5
<i>CAB</i>	CTCAGGAATGGGCAGCACTACC	CAGAATCCTACAAACGCCAACAGC	273	98.0
<i>28S</i>	TGTCGGCTCTTCTATCATTGT	ACCCAGCTCACGTTCCCTATTA	81	100
<i>18S</i>	GGCGCCCCCTCGATGCTCTTAG,	GCTCGGGCCTGCTTTGAACACTCT	89	100

\*Annealing temperature for all primer pairs was 60°C with the exception of *HBB*, which was primed at 65°C



**Figure 1.** Distributions of raw  $C_t$  values for each HK gene visualized as boxplots. Boxplots showing the distributions of raw  $C_t$  values (arithmetic means of duplicates) of the 30 samples for each of the genes tested. **Gray boxes** indicate the interquartile range with the median as the **black center bar**. The first and fourth quartile are shown as the **extended vertical bars**, and **stars** and **open circles** mark outliers and suspected outliers, respectively.

ment (no-RT controls) did not amplify with *HBB* primers above the threshold within 35 cycles. The inter-run variation was measured as the correlation coefficient of two independent PCR amplifications and was 0.997.

We found the expression level of the *TBP* transcript to be close to the background of cervical exfoliated cells. This transcript could not be reliably amplified in all sample cDNAs, even with additional template (1:2 dilution instead of 1:5). Results from the remaining 12 genes and DNA content were of sufficient quality, and the raw  $C_t$  values were distributed over comparable ranges for most genes (Figure 1).

The geNorm-calculated average gene-stability measure (M), ranked *ACTB* as the most stable gene among the 12 references tested (Table 3). Using the stepwise inclusion strategy, the combination of the three genes *PGK*, *ACTB*, and *RPLP0* showed the lowest pairwise vari-

**Table 3.** Ranking of Housekeeping Gene by Two Algorithms, GeNorm and Norm-Finder, with a Disease (CIN) Model

Rank	GeNorm	Norm-Finder
1	<b><i>ACTB</i></b>	<i>ACTB</i>
2	<b><i>RPLP0</i></b>	<i>EEF1A1</i>
3	<b><i>PGK1</i></b>	<i>18S</i>
4	<i>18S</i>	<b><i>PGK1</i></b>
5	<i>SDHA</i>	<i>GAPD</i>
6	<i>RPL4</i>	<i>EEF1G</i>
7	<i>GAPD</i>	<b><i>RPL4</i></b>
8	<i>EEF1G</i>	<i>RPLP0</i>
9	<i>EEF1A1</i>	<i>SDHA</i>
10	<i>HBB</i> (DNA)	<i>28S</i>
11	<i>28S</i>	<i>HBB</i> (DNA)
12	<i>B2M</i>	<i>B2M</i>
13	<i>MBNL2</i>	<i>MBNL2</i>

The genes in bold type indicate those that were recommended to use in combination (geometric means of  $C_t$ ) for normalization.

ation ( $V = 0.0472$ ), which was not further reduced by the addition of ribosomal *18S* as a fourth gene.

Defining three disease groups (CIN0, CIN1, and CIN2/CIN3) as a model, NormFinder identified *ACTB* as the best single gene with the a stability value (low variation) of 0.244 (Table 3). *RPL4* and *PGK* were recommended as the best combination of two genes with a stability value of 0.181.

Table 4 presents the correlation coefficients between DNA content and the  $C_t$  values of the 30 samples for each housekeeping gene transcript. There was a wide range in these correlations, but the transcripts identified as most stable with either analysis also showed the highest correlation with DNA content. *MBNL2* appeared to be completely independent from  $\beta$ -globin values. The geometric means of the three most stable genes by GeNorm (*ACTB*, *PGK1*, and *RPLP0*) had a correlation coefficient of 0.86 with  $\beta$ -globin DNA (Figure 2A). No consistent relationship was observed with RNA amounts measured by gel densitometry (Figure 2B).

## Discussion

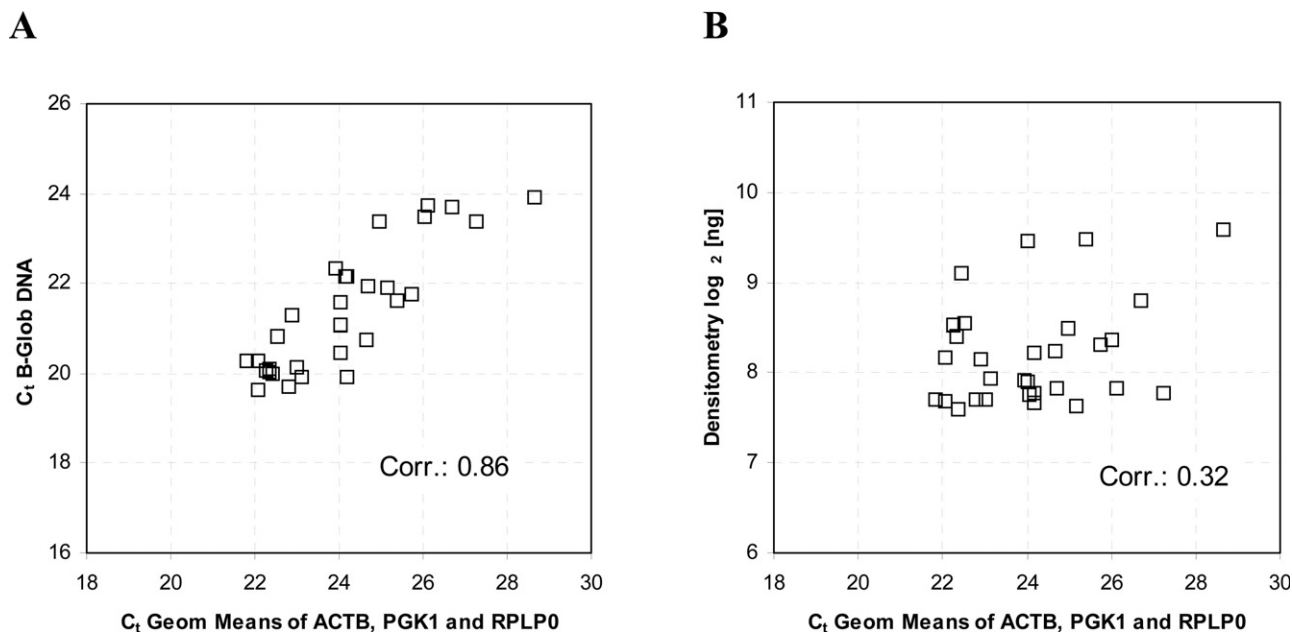
This study is the first analysis of appropriate references for exfoliated cervical cells. We measured three commonly used benchmarks—total RNA mass by densitometric detection of the *28S* and *18S* ribosomal band, DNA content by quantitative PCR amplification of genomic  $\beta$ -globin, and mRNA levels of 12 internal genes by qRT-PCR—in cervical exfoliated cells and evaluated their relationship with each other. Both GeNorm and Norm-Finder identified *ACTB* as the single most stable reference gene in this sample. Results for combinations of reference genes differed depending on the analytical approach.

GeNorm determines an internal stability measure for each gene calculated strictly from ratios. This approach seems suitable to identify a qRT-PCR reference without presumptions and resulted in the selection of *ACTB*, *RPLP0*, and *PGK1*. The model-based approach of Norm-Finder is advantageous if subpopulations with differential gene expression exist. This is especially significant if differences between these groups are being interrogated in the experiment. Accordingly, for analysis of genes

**Table 4.** Correlation Coefficients of  $C_t$  Values from Individual Candidate Reference Genes to DNA Contents throughout the Samples

Gene transcript	Correlation with DNA content
<i>PGK1</i>	0.852
<i>ACTB</i>	0.821
<i>RPLP0</i>	0.816
<i>SDHA</i>	0.795
<i>18S</i>	0.787
<i>GAPD</i>	0.734
<i>B2M</i>	0.695
<i>RPL4</i>	0.686
<i>EEF1G</i>	0.686
<i>EEF1A1</i>	0.681
<i>28S</i>	0.655
<i>MBNL2</i>	0.272





**Figure 2.** Correlation of different references. **A:** Correlation of the three most stable genes by GeNorm (geometric means of  $C_t$  values) and DNA content ( $C_t$  of B-globin). **B:** No correlation was apparent among the three most stable genes (geometric means of  $C_t$  values) and total RNA mass estimated by densitometric assessment of the 18S and 28S ribosomal bands.

indicative of the degree of CIN, a normalization factor generated from the Norm-Finder selection (*RPL4* and *PGK1*) might be preferable, because the pre-existing differences from varying CIN grades were used in the model.

The choice of references for qRT-PCR normalization is crucial for accurate comparisons of gene expression. However, the selection remains problematic for several reasons. Because of variations in the source of the sample and the biological variability encountered in the study, references must be evaluated empirically. To avoid a circular argument over the sample concentration, we made no attempt to equalize template concentration before each assay and used strictly equal input volumes from all samples. Although the problems of sample heterogeneity (contributing to variation in RNA content per cell or DNA unit) and RNA preservation (affecting observed relation between rRNA and mRNA) may well be greater for exfoliated cells than other kinds of samples, we believe our findings are relevant to other systems and indicate the importance of empirical validation of reference genes, regardless of the sample and experimental question.

Interestingly, the ranking of reference genes by GeNorm was paralleled by correlation with the DNA content individually (Table 3), and the geometric means of the three top genes (*ACTB*, *RPLP0*, and *PGK1*) showed a correlation coefficient of 0.86 with  $\beta$ -globin (Figure 2A). This relationship might indicate a certain consistency of these gene expression levels in populations of exfoliated cells.

DNA content representing the cell number rationally appears as the most robust standard, and its good correlation with a number of housekeeping genes indicates a relative robust representation of mRNA levels. Its sta-

bility was nonetheless low when the CIN model was applied through Norm-Finder. An explanation could be given by haplotype variations that frequently occur in neoplastic lesions.<sup>12</sup> Furthermore, the gene stability was calculated in relation to the other transcript levels in the study, and differences of global RNA expression relative to DNA might actually be a variable factor in different disease states. However, DNA could provide a valuable absolute reference in some types of experiments. Use of an additional external RNA control, like the plant *CAB* transcript we used here, is strongly advisable to monitor sample-specific differences that occur during the cDNA synthesis.

Ribosomal 18S RNA was moderately stable in the qRT-PCR assays and ranked fourth by geNorm and third by Norm-Finder, whereas 28S rRNA was ranked low by both algorithms (Table 3). Correlation of 28S to DNA content was also insignificant compared with 18S. These results seem to underline the general supposition that 18S expression can be relatively stable and had in fact been suggested for normalization in other cell types and conditions.<sup>9,13,14</sup> Fluctuation of rRNA levels especially of 28S are nonetheless apparent in exfoliated cells, influence total RNA levels accordingly, and might therefore not adequately represent mRNA activities. The finding that densitometric estimation of total RNA mass by the intensity of rRNA bands was completely independent from all PCR quantifications—DNA and RNA—measured in the sample could originate from different degrees of partial degradation that influences UV absorbance of ribosomal bands but not the abundance of short PCR-amplified target templates. The relatively low precision of photometric technologies might further distort the true quantity. According to our results, the use of rRNA can generally

not be recommended as a standard for cervical exfoliated cells.

No standard is likely to account perfectly for all aspects of the complexity and dynamic of the transcriptome, and mRNA quantification must therefore be seen as relative to a subjective reference. To normalize for sample-specific differences and accumulative experimental errors alike, we suggest the use of *ACTB* ( $\beta$ -actin) as an acceptable standard for qRT-PCR studies in cervical exfoliated cells. The geometric means of *RPL4* and *PGK1* are recommended if special attention to intraepithelial lesions is appropriate.

## Acknowledgments

We thank Dr. Brian Gurbaxani for helpful discussions.

## References

1. Vandesompele J, Preter KD, Pattyn F, Poppe NVR, Paepe AD, Speleman F: Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol* 2002, 3:0034.1–0034.11
2. Janssens N, Janicot M, Perera T, Bakker A: Housekeeping genes as internal standards in cancer research. *Mol Diagn* 2004, 8:107–113
3. Pfaffl MW, Tichopad A, Prgomet C, Neuvians TP: Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: bestKeeper—Excel-based tool using pair-wise correlations. *Biotechnol Lett* 2004, 26:509–515
4. de Kok JB, Roelfs RW, Giesendorf BA, Pennings JL, Waas ET, Feuth T, Swinkels DW, Span PN: Normalization of gene expression measurements in tumor tissues: comparison of 13 endogenous control genes. *Lab Invest* 2004, 85:154–159
5. Andersen CL, Jensen JL, Orntoft TF: Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. *Cancer Res* 2004, 64:5245–5250
6. Huggett J, Dheda K, Bustin S, Zumla A: Real-time RT-PCR normalisation: strategies and considerations. *Genes Immun* 2005, 6:279–284
7. Steinau M, Lee DR, Rajeevan MS, Vernon SD, Ruffin MT, Unger ER: Gene expression profile of cervical tissue compared to exfoliated cells: impact on biomarker discovery. *BMC Genomics* 2005, 6:64
8. Habis AH, Vernon SD, Lee DR, Verma M, Unger ER: Molecular quality of exfoliated cervical cells: implications for molecular epidemiology and biomarker discovery. *Cancer Epidemiol Biomarkers Prev* 2004, 13:492–496
9. Bas A, Forsberg G, Hammerstrom S, Hammerstrom ML: Utility of the housekeeping genes 18S rRNA, b-actin and glyceraldehyde-3-phosphate-dehydrogenase for normalization in real-time quantitative reverse transcriptase-polymerase chain reaction analysis of gene expression in human T lymphocytes. *Scand J Immunol* 2004, 59:566–573
10. Talaat AM, Howard ST, Hale W, Lyons R, Garner H, Johnston SA: Genomic DNA standards for gene expression profiling in *Mycobacterium tuberculosis*. *Nucleic Acids Res* 2002, 30:e104
11. Rajeevan MS, Swan DC, Nisenbaum RL, Lee DR, Vernon SD, Ruffin MT, Horowitz IR, Flowers LC, Kmak D, Tadros T, Birdsong G, Husain M, Srivastava S, Unger ER: Epidemiologic and viral factors associated with cervical neoplasia in HPV-16-positive women. *Int J Cancer* 2005, 115:114–120
12. Shirata NK, Longatto Filho A, Roteli-Martins C, Espoladore LM, Pittoli JE, Syrjanen K: Applicability of liquid-based cytology to the assessment of DNA content in cervical lesions using static cytometry. *Anal Quant Cytol Histol* 2003, 25:210–214
13. Goldin D, Mamessier A, Staquet MJ, Schmitt D, Berthier-Vergnes O: Ribosomal 18S RNA prevails over glyceraldehyde-3-phosphate dehydrogenase and beta-actin genes as internal standard for quantitative comparison of mRNA levels in invasive and noninvasive human melanoma cell subpopulations. *Anal Biochem* 2001, 295:17–21
14. Morse DL, Carroll D, Weberg L, Borgstrom MC, Ranger-Moor J, Gillies RL: Determining suitable internal standards for mRNA quantification of increasing cancer progression in human breast cells by real-time reverse transcriptase polymerase chain reaction. *Anal Biochem* 2005, 342:69–77