Selection of appropriate control genes to assess expression of tumor antigens using real-time RT-PCR

Joeri L. Aerts, Monica I. Gonzales, and Suzanne L. Topalian
National Institutes of Health, Bethesda, MD, USA

BioTechniques 36:84-91 (January 2004)

*Real-time reverse transcription PCR (RT-PCR) is a sensitive and accurate method to monitor gene expression and is often used to profile the expression of putative tumor antigens in the context of immunotherapy. However, this technique consists of several steps, including cell processing, RNA extraction, RNA storage, assessment of RNA concentration, and cDNA synthesis prior to PCR. To compensate for potential variability introduced in this procedure, the expression of housekeeping genes is commonly assessed in parallel with the expression of the gene of interest. In this study, the expression of a variety of housekeeping genes in a panel of 26 different human tumor and embryonal cell lines was assessed using real-time RT-PCR. For some control genes, the variability in expression was significant between different cell lines, despite the equalization of quantities of input RNA. The greatest variability was found for GAPDH. The lowest variability was found for β-glucuronidase (*GUS*) and 18S rRNA. While real-time RT-PCR is a powerful tool for gene expression analysis, these results suggest that the choice of control genes to normalize the expression of the gene of interest is critical to the interpretation of experimental results and should be tailored to the nature of the study.*

INTRODUCTION

The expression of the gene of interest in normal and neoplastic tissues is profiled in many cancer immunology studies evaluating the value of a candidate tumor antigen. Ideally, expression should be determined at the protein level because this will ultimately determine the recognition potential of the antigen by the immune system. However, because no specific antibodies are yet available for many novel candidate tumor antigens, one frequently has to resort to gene expression as a surrogate marker of protein expression.

Recently, several fluorochrome-based assays for the quantification of PCR amplification products have been introduced. One of the most frequently employed platforms is real-time reverse transcription PCR (RT-PCR), which was used in this study. This system offers several advantages over traditional methods for direct detection of mRNA such as Northern blot analysis and RNase protection assays or PCR-based techniques such as competitive PCR. These include high sample throughput in a 96-well plate format, the requirement for relatively small quantities of mRNA, and precision. Experimental variability encountered with real-time RT-PCR can be minimized by quantitating the expression of a control gene together with the gene of interest to compensate for differences in mRNA amount and quality between different samples. Ideally, the control gene should be expressed stably and at a similar level in all tissues under investigation, and its expression should not be influenced by the experimental conditions. However, several studies have suggested that even widely used control genes such as β-actin and GAPDH are unsuitable in certain situations (1).

To address the issue of appropriate control gene selection, we analyzed the expression of 11 different control genes commonly used in real-time RT-PCR in human embryonal and neoplastic cell lines derived from a wide variety of tissues. Our results suggest that control gene selection is critical to the interpretation of tumor antigen expression studies based on real-time RT-PCR techniques.

MATERIALS AND METHODS

Cell Lines

Four melanoma cell lines (624-mel, 938-mel, 1087-mel, and 1558-mel) were generated at the Surgery Branch, National Cancer Institute (NCI) (National Institutes of Health, Bethesda, MD, USA), according to methods that have been previously described (2). The renal carcinoma cell lines UOK150, 1764RCC, 2192RCC, and 2194RCC were a gift from Dr. Ken-ichi Hanada (Surgery Branch, NCI). The breast cancers MCF-7 and SK-BR-3; the prostate cancers DU-145 and PC-3; the colon carcinomas COLO 320 and HT-29; the leukemias MOLT-4, THP-1, and K-562; the cervical carcinoma HeLa; the embryonal cell lines 293, WI-38, and HEPM; the teratocarcinomas PA-1 and NCCIT; the pancreatic cancer Capan-2; and the Ewing's sarcomas 6647 and TC-71 were all obtained from the American Type Culture Collection (Manassas, VA, USA). The cells were cultured in RPMI 1640 medium plus 10% fetal bovine serum (FBS), with 10 mM HEPES buffer, pH 7.5, 2 mM L-glutamine, 250 ng/mL amphotericin B, 50 IU/mL penicillin, 50 µg/mL streptomycin, and 50 µg/mL gentamicin sulfate.

RNA Isolation and cDNA Synthesis

To determine gene expression, total RNA was isolated from cultured cell lines using RNeasy® columns (Qiagen, Valencia, CA, USA) according to the manufacturer's instructions. Isolated RNA was eluted in 40 µL RNase-free water and subsequently used as the template for one round of reverse transcription to generate cDNA as follows. One microgram of purified total RNA was added to a Master Mix (Applied Biosystems, Foster City, CA, USA) containing a final concentration of 1× reverse transcription buffer (20 mM Tris-HCl, pH 8.4, 50 mM KCl), 5 mM $MgCl_2$, 2 mM dNTP, 2.5 µM of random hexamer primers, 0.4 U/µL RNase inhibitor, and 1.25 U/µL SUPERSCRIPT™ II reverse transcriptase (Invitrogen, Carlsbad, CA, USA) in a 20-µL reaction. Reverse transcription was performed according to the SUPERSCRIPT First-Strand Synthesis System instructions (Invitrogen).

Table 1. TaqMan Primer and Probe Sequences

Gene	Forward Primer	Reverse Primer	TaqMan Probe
β-actin	5′-GCGAGAAGATGACCCAGATC-3′	5′-CCAGTGGTACGGCCAGAGG-3′	5′-FAM-CCAGCCATGTACGTTGC-TATCCAGGC-TAMRA-3′
Neo-PAP[a]	5′-GGAGAGGGAGACGCAGGAA-3′	5′-CTTTTGTTGACGCTGGCTGTC-3′	5′-FAM-TGAAAGAGATGTCTGCAAA-CACCG-TAMRA-3′
GUS	5′-GAAAATATGTGGTTGGAGAGCTCATT-3′	5′-CCGAGTGAAGATCCCCTTTTTA-3′	5′-FAM-CCAGCACTCTCGTCGGT-GACTGACTGTTCA-TAMRA-3′

Neo-PAP, neo-poly(A)polymerase; *GUS*, β-glucuronidase.
[a]Primers and probes were designed using the Primer Express® software (version 1.0; Applied Biosystems).

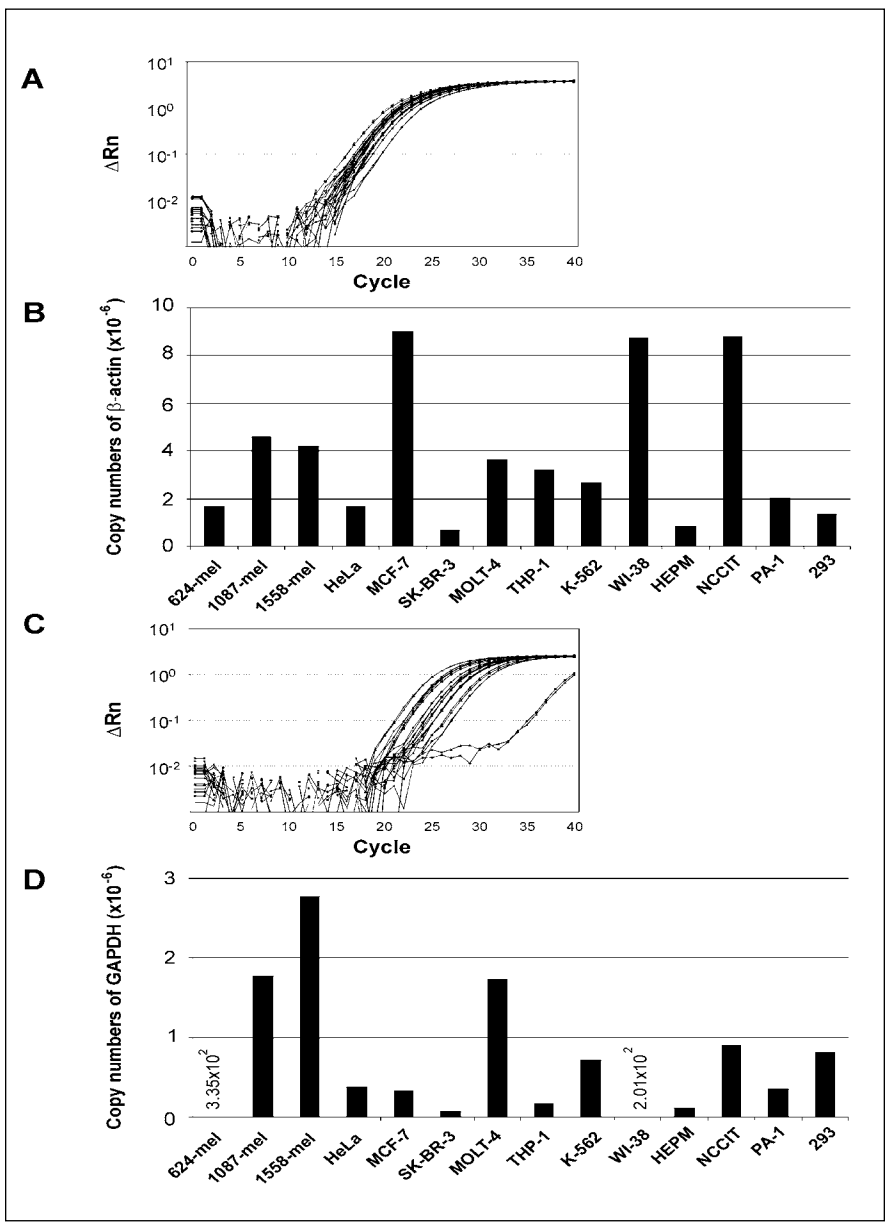

Figure 1. β-Actin and GAPDH expression in 14 human cell lines. The same RNA samples were used as templates for β-actin and GAPDH quantification. (A and C) Cycle numbers and (B and D) calculated copy numbers are shown. ΔRn, normalized fluorescence intensity.

Ten percent of the cDNA reaction was used as the template for real-time PCR.

Real-Time PCR

For the quantitative analysis of mRNA expression, we used the ABI PRISM® 7700 Sequence Detection System (Applied Biosystems). Primers and probes for β-actin (3), neo-poly(A) polymerase (*neo-PAP*), and β-glucuronidase (*GUS*) (4) were synthesized by Applied Biosystems (sequences shown in Table 1). Probes were labeled at the 5′ end with the reporter dye molecule FAM (emission wavelength, 518 nm) and at the 3′ end with the quencher dye TAMRA (emission wavelength, 582 nm). The 3′ end of the probe was additionally phosphorylated to prevent extension during PCR. For the detection of GAPDH, TaqMan® GAPDH Control Reagents (Applied Biosystems) were used, and for the detection of 18S rRNA, TaqMan Pre-Developed Assay Reagents (Applied Biosystems) were used. For PCR, 12.5 µL TaqMan Universal PCR Master Mix, 800 nM of each primer, and 200 nM of probe were combined and adjusted to a total volume of 20 µL with RNase-free water. Finally, cDNA generated from 100 ng total RNA was added to a total reaction volume of 25 µL. The conditions used for PCR were 2 min at 50°C, 10 min at 95°C, and then 40 cycles of 15 s at 95°C and 1 min at 60°C. For each experiment, a no-template reaction was included as a negative control. Each cDNA sample was tested in duplicate, and the mean values were calculated. Duplicate values varied by no more than 20% from the mean. Copy num-

ber quantification was based on the TaqMan principle, depending on the number of PCR cycles required for the threshold detection of the fluorescence signal [cycle threshold (C_t)], as previously described (5,6). Briefly, 10-fold serial dilutions of known concentrations of plasmid constructs containing the relevant sequences (β-actin, GAPDH, *neo-PAP*, and *GUS*) or of a purified PCR product (18S rRNA) were prepared so as to create standard curves for quantification of unknown samples. All plasmid dilutions were made in the presence of 20 ng/mL of *Escherichia coli* 16S and 23S rRNA (Roche Applied Science, Mannheim, Germany) as a carrier.

The TaqMan Human Endogenous Control Plate (Applied Biosystems) precoated with lyophilized primers and probes for 11 different commonly used human control genes was also used to assess gene expression in six different cancer cell lines. The plate contains an internal positive control that consists of a synthetic product plus primers and probes to amplify and quantify the template. Thus, the potential influence of inhibitors present in the tumor-derived cDNA samples could be assessed. PCR was set up according to the manufacturer's instructions. Briefly, for each tumor analyzed, 600 µL TaqMan Universal Master Mix were pipetted into a tube, and cDNA generated from 2.4 µg RNA was added. The volume was adjusted to a total of 1200 µL, and 50 µL of this mixture were added to each of the 24 wells in the plate (i.e., 11 control genes plus internal control in duplicate). The same PCR cycling conditions were used as described above. To determine gene expression using the endogenous human control plate, we used the ΔC_t method. Values obtained from a given cell line were normalized to expression in 293 cells, chosen arbitrarily to serve as a reference control, by subtracting the C_t values ($C_{t\ 293}$ - $C_{t\ cell\ line}$). Because each amplification PCR cycle leads to a doubling of the initial amount of material, 1 C_t value corresponds to a 2-fold difference and 3.3 C_t values correspond to a 10-fold difference in expression. All samples were run in duplicate, and the mean values were calculated.

RESULTS AND DISCUSSION

Variability in Expression of GAPDH and β-Actin in 14 Different Cell Lines

The selection of an appropriate control gene for real-time RT-PCR analysis is often based on historical evidence from the literature. β-Actin and GAPDH are among the most commonly used genes for this purpose. However, GAPDH expression has been shown to vary with cell cycle (7) and developmental stage (8). For β-actin, several highly homologous isoforms and pseudogenes have been identified, which introduce a risk of false positiv-

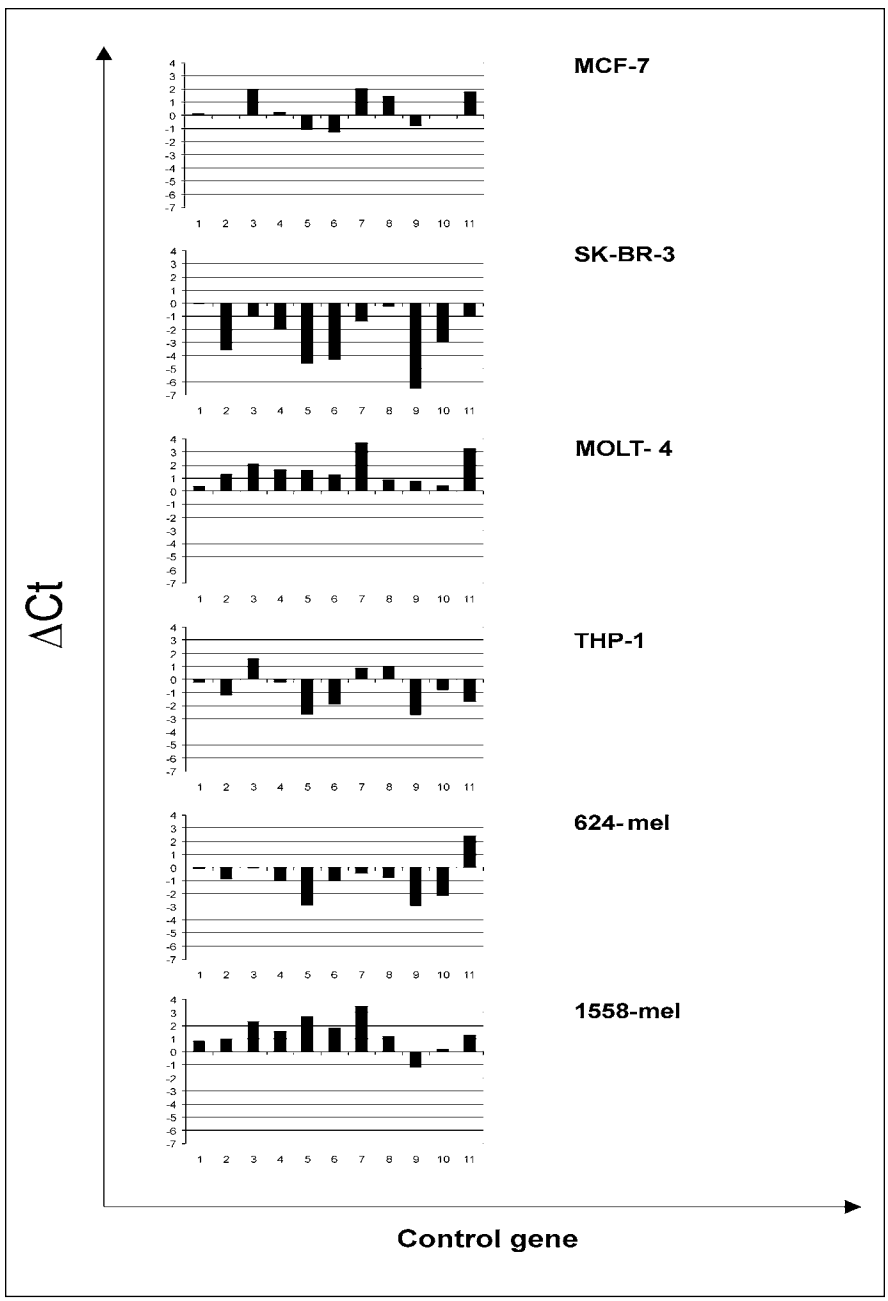

Figure 2. Relative expression of 11 different control genes in 6 cell lines. X-axis: 1, 18S rRNA; 2, acidic ribosomal protein; 3, β-actin; 4, cyclophilin; 5, GAPDH; 6, phosphoglycerokinase; 7, β2-microglobulin; 8, *GUS*; 9, hypoxanthine ribosyl transferase (*HPRT*); 10, TATA binding protein (*TBP*); 11, transferrin receptor. To normalize mRNA copy numbers, all cycle threshold (C_t) values were compared to expression in 293 cells ($\Delta C_t = C_{t\ 293} - C_{t\ cell\ line}$).

ity. To address the issue of choosing optimal control genes for real-time RT-PCR, we first evaluated the variability in expression of β-actin and GAPDH in 14 different cultured human cell lines (Figure 1). For β-actin, the fold difference in copy number between the lowest (breast cancer, SK-BR-3) and the highest (breast cancer, MCF-7) expressing cell lines was 14-fold. However, for GAPDH, the difference was 13,865-fold between the lowest (embryonal, WI-38) and the highest (melanoma, 1558-mel) expressing lines. Thus, from fixed quantities of input total RNA template derived from 14 different cell lines, real-time RT-PCR revealed a much greater consistency in the expression of β-actin, a cytoskeletal protein, than GAPDH, a glycolytic enzyme.

Selection of Control Genes With Least Variability

Based on the results derived from comparing the expression of β-actin and GAPDH in 14 cell lines, 6 of these cell lines were chosen for further analysis of control gene expression. These included cell lines showing the extremes of variability for β-actin (MCF-7, SK-BR-3, and THP-1) and GAPDH (624-mel, 1558-mel, and MOLT-4). The expression of 11 different control genes in general use for real-time RT-PCR was analyzed using the TaqMan Human Endogenous Control Plate (Figure 2). Some cell lines seemed to manifest consistent differences in the expression of all 11 control genes relative to 293 cells; SK-BR-3 expressed significantly higher levels of all control genes, while the MOLT-4 was consistently lower. However, other cell lines did not manifest such consistency. For instance, 624-mel was among the highest GAPDH expressing lines relative to 293 cells, while it was among the lowest transferrin receptor expressing lines. Therefore, it was unlikely that variable gene expression solely reflected variable amounts of mRNA contained in the six total RNA samples used as templates for real-time RT-PCR or global problems with reverse transcription. Among 11 genes analyzed, the greatest variability in expression was observed for GAPDH, which is consistent with our previous results. The range of ΔC_t values for GAPDH expression among the cell lines was 7.3 (ΔC_t = 2.7 for 1558-mel; ΔC_t = -4.6 for SK-BR-3), translating to a range of 154-fold between the lowest and highest expression relative to 293 cells. The variability in expression of 18S rRNA was shown to be the lowest (range of expression was 2.1-fold among the 6 cell lines tested), which is consistent with other studies. However, several drawbacks to using 18S rRNA as a control gene include its high abundance of expression (generally $C_t < 10$), making accurate quantitation difficult, its relative resistance to degradation compared to mRNA species, and the imperfect correlation of expression of rRNA with mRNA. The control genes β-actin, TATA box binding protein, and *GUS* also showed low variability of expression among the 6 cell lines assayed, with ranges of 10.1-, 10.1-, and 4.6-fold, respectively.

It is possible that a similarly conducted analysis of control gene expression using a different panel of cell lines

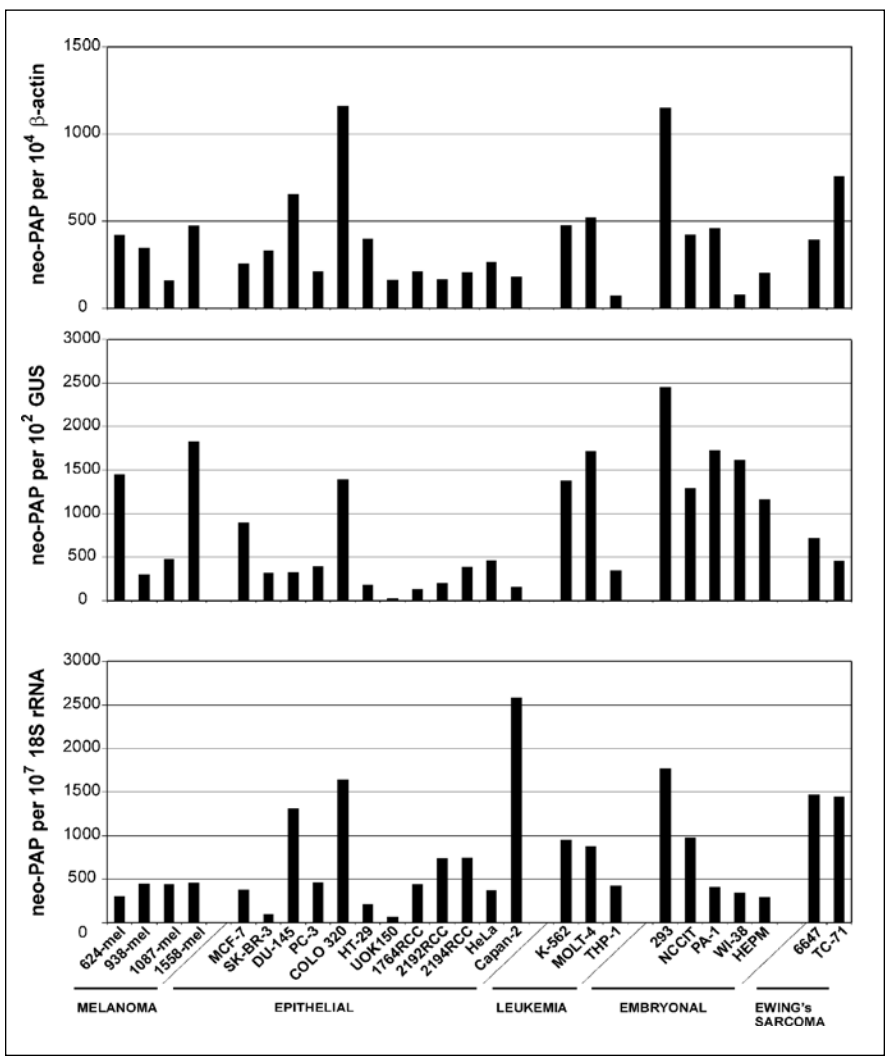

Figure 3. Expression of neo-poly(A) polymerase (*neo-PAP*) copy numbers normalized using β-actin, *GUS*, or 18S rRNA. Twenty-six cell lines were assessed, including four melanomas (624-mel, 938-mel, 1087-mel, and 1558-mel); two breast cancers (MCF-7 and SK-BR-3); two prostate cancers (DU-145 and PC-3); two colon carcinomas (COLO 320 and HT-29); four renal carcinomas (UOK150, 1764RCC, 2192RCC, and 2194RCC); one cervical carcinoma (HeLa); one pancreatic cancer (Capan-2); three leukemias (K-562, MOLT-4, and THP-1); three embryonal cell lines (293, WI-38, and HEPM); two teratocarcinomas (NCCIT and PA-1); and two Ewing's sarcomas (6647 and TC-71). The same cDNA samples were used as templates for all real-time PCRs, and *neo-PAP* with control gene reactions were always run on the same plate.

might yield somewhat different results. However, these data suggest that prior to conducting any study utilizing real-time RT-PCR, a comparison of utilization of several different control genes should be executed within the parameters of that study. The choice of a suitable control gene will depend on the scope and nature of the experiment to be performed (9).

Use of Selected Control Genes to Normalize Expression of *Neo-PAP*

When evaluating the potential clinical value of a tumor-associated protein, it is important to establish the expression profile of the gene of interest in different tumors and normal tissues, and real-time RT-PCR can be applied for this purpose. Based on the results obtained from screening 11 control genes, 3 genes demonstrating the least variability in expression among the tumor samples evaluated (18S rRNA, *GUS*, and β-actin) were used as controls to quantify the expression of *neo-PAP* in 26 different cell lines (10). *Neo-PAP* was used as a prototype of a novel tumor antigen expressed in a variety of cancers and is being evaluated as a potential therapeutic target (11). Twenty-six different cell lines were examined; as shown in Figure 3, some cell lines consistently expressed high (293 and COLO 320) or low (UOK150 and THP-1) copy numbers of *neo-PAP* relative to the other lines using any of the three control genes. However, for others, the relative expression was dependent on the control gene that was used. In the instance of the Capan-2, *neo-PAP* expression normalized with the control gene 18S rRNA was among the highest of the 26 cell lines analyzed, while normalization of expression with *GUS* or β-actin showed it to be among the poorest expressing cell lines. Thus, even with careful selection of three control genes for analysis, inconsistencies in calculated *neo-PAP* expression were observed for some cell lines.

In conclusion, the "perfect" control gene for real-time RT-PCR probably does not exist. Several recommendations have been made to increase the reliability of the normalization of gene expression, such as the concurrent use of several housekeeping genes (12) or normalization to spectrophotometrically quantified total RNA amounts (13), which may nevertheless contain varying quantities of mRNA or interfering contaminants. In our comparative study of several commonly used control genes for real-time RT-PCR analysis, 18S rRNA, *GUS*, and β-actin seemed to manifest the least variability of expression among a wide variety of human cell lines investigated. With the appropriate emphasis on optimizing the choice of control genes to the nature and design of a particular study, the real-time RT-PCR technique remains a powerful tool for gene expression analysis.

ACKNOWLEDGMENTS

The authors wish to thank Ken-ichi Hanada (Surgery Branch, NCI) for providing the renal carcinoma cell lines, Nancy Boeckx (Laboratory of Hematology, UZ Gasthuisberg) for the GUS plasmid, Yong Li (Surgery Branch, NCI) for the GAPDH plasmid, and Steven A. Rosenberg (Surgery Branch, NCI) for advice and support.

REFERENCES

1. **Sturzenbaum, S.R. and P. Kille.** 2001. Control genes in quantitative molecular biological techniques: the variability of invariance. Comp. Biochem. Physiol. B Biochem. Mol. Biol. *130*:281-289.
2. **Topalian, S.L., D. Solomon, and S.A. Rosenberg.** 1989. Tumor-specific cytolysis by lymphocytes infiltrating human melanomas. J. Immunol. *142*:3714-3725.
3. **Schoof, E., M. Girstl, W. Frobenius, M. Kirschbaum, H.G. Dorr, W. Rascher, and J. Dotsch.** 2001. Decreased gene expression of 11beta-hydroxysteroid dehydrogenase type 2 and 15-hydroxyprostaglandin dehydrogenase in human placenta of patients with preeclampsia. J. Clin. Endocrinol. Metab. *86*: 1313-1317.
4. **Gabert, J., E. Beillard, V.H. van der Velden, W. Bi, D. Grimwade, N. Pallisgaard, G. Barbany, G. Cazzaniga, et al.** 2003. Standardization and quality control studies of "real-time" quantitative reverse transcriptase polymerase chain reaction (RQ-PCR) of fusion gene transcripts for residual disease detection in leukemia—a Europe against cancer program. Leukemia [Epub ahead of print.]
5. **Heid, C.A., J. Stevens, K.J. Livak, and P.M. Williams.** 1996. Real time quantitative PCR. Genome Res. *6*:986-994.
6. **Lizee, G., J.L. Aerts, M.I. Gonzales, N. Chinnasamy, R.A. Morgan, and S.L. Topalian.** 2003. Real-time quantitative reverse transcriptase-polymerase chain reaction as a method for determining lentiviral vector titers and measuring transgene expression. Hum. Gene Ther. *14*:497-507.
7. **Mansur, N.R., K. Meyer-Siegler, J.C. Wurzer, and M.A. Sirover.** 1993. Cell cycle regulation of the glyceraldehyde-3-phosphate dehydrogenase/uracil DNA glycosylase gene in normal human cells. Nucleic Acids Res. *21*:993-998.
8. **Calvo, E.L., C. Boucher, Z. Coulombe, and J. Morisset.** 1997. Pancreatic GAPDH gene expression during ontogeny and acute pancreatitis induced by caerulein. Biochem. Biophys. Res. Commun. *235*:636-640.
9. **Thellin, O., W. Zorzi, B. Lakaye, B. De Borman, B. Coumans, G. Hennen, T. Grisar, A. Igout, et al.** 1999. Housekeeping genes as internal standards: use and limits. J. Biotechnol. *75*:291-295.
10. **Topalian, S.L., S. Kaneko, M.I. Gonzales, G.L. Bond, Y. Ward, and J.L. Manley.** 2001. Identification and functional characterization of neo-poly(A) polymerase, an RNA processing enzyme overexpressed in human tumors. Mol Cell Biol. *21*:5614-5623.
11. **Topalian, S.L., M.I. Gonzales, Y. Ward, X. Wang, and R.F. Wang.** 2002. Revelation of a cryptic major histocompatibility complex class II-restricted tumor epitope in a novel RNA-processing enzyme. Cancer Res. *62*: 5505-5509.
12. **Vandesompele, J., K. De Preter, F. Pattyn, B. Poppe, N. Van Roy, A. De Paepe, and F. Speleman.** 2002. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol. *3*:1-11.
13. **Tricarico, C., P. Pinzani, S. Bianchi, M. Paglierani, V. Distante, M. Pazzagli, S.A. Bustin, and C. Orlando.** 2002. Quantitative real-time reverse transcription polymerase chain reaction: normalization to rRNA or single housekeeping genes is inappropriate for human tissue biopsies. Anal. Biochem. *309*:293-300.

Received 22 September 2003; accepted 30 October 2003.

Address correspondence to Suzanne L. Topalian, Surgery Branch, National Cancer Institute, National Institutes of Health 10/ 2B47, Bethesda, MD 20892, USA. e-mail: suzanne_topalian@nih.gov