Choice of endogenous control for gene expression in nonsmall cell lung cancer

D.W. Liu*, S.T. Chen# and H.P. Liu*

ABSTRACT: This study attempts to identify a suitable endogenous control gene for real-time RT-PCR in nonsmall cell lung cancer (NSCLC) tissues.

Expression of seven common endogenous control genes (glyceraldehyde-3-phosphate dehydrogenase (GAPDH), v-abl Abelson murine leukaemia viral oncogene homologue 1, beta-2-microglobulin, hypoxanthin phosphoribosyltransferase 1 (HPRT1), phosphoglycerate kinase 1, peptidylprolyl isomerase A, and ribosomal protein, large, P0) in 18 heterogenous NSCLC tumour specimens, 10 normal lung tissues and six NSCLC cell lines were analysed by quantitative RT-PCR. The variances and correlation coefficients of cycle threshold (Ct) value of each control gene in three tissue groups and subgroups were compared. The difference and correlation coefficients between the Ct value for each control gene and the mean Ct value of the remaining control genes were calculated.

The GAPDH gene transcript showed the least variance and linear regression analysis demonstrated that GAPDH and HPRT had the strongest correlation in pooled tumour and normal lung tissues. Furthermore, GAPDH expression value showed stringent correlation and had the lowest difference with the mean expression value of the remaining endogenous control genes.

Among the seven common endogenous control genes, glyceraldehyde-3-phosphate dehydrogenase is the most suitable for quantitative RT-PCR reaction in nonsmall cell lung cancer tissue samples.

KEYWORDS: Endogenous control genes, nonsmall cell lung cancer, RT-PCR

AFFILIATIONS
*Division of Thoracic and Cardiovascular Surgery, and
#Division of Endocrinology and Metabolism, Chang Gung Memorial Hospital, TaoYuan, Taiwan, Republic of China.

CORRESPONDENCE
D.W. Liu
Division of Thoracic and Cardiovascular Surgery
Chang Gung Memorial Hospital
5, Fu-Hsing St.
KweiShan
TaoYuan
Taiwan
Republic of China
Fax: 11 88633285818
E-mail:
dahwelliu@adm.cgmh.org.tw

Lung cancer is the leading cause of death from cancer among males and females in the world [1]. More people die each year from lung cancer than of colon, breast and prostate cancer combined. Nearly 7,000 people in Taiwan die from lung cancer each year [2]. Despite advances in treatment between 1995–2005, long-term survival for patients with non-small cell lung cancer (NSCLC) remains poor. The overall 5-yr survival rate remains <20%, as most patients present with advanced disease [3]. To increase survival rates, lung cancer must be detected as early as possible. Real-time quantitative PCR technology has recently been applied avidly in quantifying circulating DNA for early lung cancer diagnosis [4]. Additionally, quantitative PCR has been used in the study of lung cancer to identify molecular characteristics [5], the target gene and its expression [6], and predicting prognosis [7]. However, to accurately quantify gene expression, parallel amplification of the target gene with one or more endogenous control gene is a critical issue. An ideal endogenous control gene always shows the same level of expression across all tissue samples, cells,

experimental treatments and designs. Unfortunately, such a universal endogenous control gene either does not exist [8] or it has not yet been found [9–14]. Therefore, the constant expression of an endogenous control gene chosen for the experimental setup required testing and validation before any reliable data could be obtained. A few reports have already validated one or more suitable endogenous control genes for RT-PCR in specific organs or tissues such as: Ableson gene for leukaemia [15]; protein kinase cyclic guanosine monophosphate (cGMP)-dependent type-1 (PGK1) for both T- and B-cell lymphoma [16]; human acidic ribosomal protein for pulmonary tuberculosis [17]; 18s rRNA and cyclophilin for renal biopsies [18]; ubiquitin B, glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and translationally controlled tumour protein for colon cancer [19]; heat shock 90 kD protein 1b, testis enhanced gene transcript; and adenosine tiphosphate synthase for bladder cancer [19].

Recent lung cancer studies have employed various endogenous control genes, such as GAPDH [20], β-actin [21], TATA-binding protein

[22], 18s-rRNA [23] and phenylalanine hydroxylase [24], for RT-PCR. However, no endogenous control gene has been validated in lung cancer and normal lung tissues. Therefore, this study attempted to identify an endogenous control gene suited for investigation of experimental designs covering various stages and types of NSCLC tumours, normal lung tissues and NSCLC cell lines. The gene expression levels of seven commonly used endogenous control genes were assessed by quantitative RT-PCR. The variances of each gene and correlation coefficients between control genes were analysed.

MATERIAL AND METHODS
Tumour tissues and cell lines
Primary tumour samples (n=18) and corresponding nonmalignant lung tissues (n=10) were obtained from NSCLC patients (nine males and nine females, aged 28–81 yrs, mean 63 ± 9 yrs) who had been treated with curative resectional surgery at Chung Gung Memorial Hospital (Taoyuan, Taiwan, Republic of China (ROC)) between June 2000 and March 2003. There were 14 adenocarcinomas, three squamous carcinomas, one large cell carcinoma (12 stage I, two stage II and four stage III) and 10 normal lung tissues. Surgically removed samples were immediately frozen in liquid nitrogen and stored at -80°C until used. Frozen surgical specimens were sectioned and mounted onto slides. After fixation, the cells were stained with haematoxylin and eosin stain. The slides were carefully examined for pathology by two of the authors. Areas of contaminating stroma cells in these tumoural samples were calculated under low- and high-powered microscope magnifying fields. The percentage of stromal cells was 11.95 ± 8.06 (4–35%). Six NSCLC cell lines were used: NCI-H23, NCI-H226, NCI-H460, and NCI-H1299. These were obtained from G.M. Tsai (Taipei Veteran General Hospital, Taiwan, ROC). Calu-1 was a generous gift from S.H. Liau (Chang Gung University TaoYuan, Taiwan, ROC) and NCI-A549 was obtained from Y.C. Liu (Chang Gung Memorial Hospital). All NSCLC cell lines were grown and propagated in Royal Park Memorial Institute medium-1640 supplemented with 10% foetal calf serum (Invitrogen, Carlsbad, CA, USA). The study was approved by the ethical committees at the Chang Gung Memorial Hospital and informed consent was obtained from each patient.

RNA purification and cDNA preparation
Total RNA was isolated with Trizol reagent (Invitrogen, Carsbad, CA, USA) according to the manufacturer's instruction. The amount of total RNA was quantified by spectrophotometric OD260 measurement. First-strand cDNA synthesis was carried out using the M-MLV reverse transcriptase method (Invitrogen). Both 8 μg of total RNA and an adequate amount of diethyl pyrocarbonate water in a total volume of 21 μL were incubated for 5 min at 95°C. After adding 2.74 μL of 0.3 $\mu g \cdot \mu L^{-1}$ random primers, 4 μL of 0.1 M dithiothreitol (DTT), 8 μL of $5 \times$ reverse-transcription (RT) buffer, 1 μL of RNase-out 40 $U \cdot \mu L^{-1}$ and 2 μL of Moloney murine leukaemia virus (M-MLV) reverse transcriptase (200 $U \cdot \mu L^{-1}$) in a total volume of 40 μL, the RNA mixture was incubated at 37°C overnight. The cDNA was then diluted 1:16 for application in RT-PCR.

RT-PCR
Seven endogenous control genes (v-abl Abelson murine leukaemia viral oncogene homologue 1 (ABL1; NM-007313), beta-2-microglobulin (B2M; NM-004048), hypoxanthine phosphoribosyltransferase 1 (HPRT1; NM-000194), GAPDH (NM-002046), PGK1 (NM-000291), peptidylprolyl isomerase A (PPIA; NM-021130), and ribosomal protein, large, P0 (RPLP0; NM-001002)) were measured by quantitative RT-PCR using the 5′ nuclease technology on an ABI PRISM 7700 sequence detection system (Applied Biosystems, Foster City, CA, USA). The sequences of the probes were labelled with FAM® dye-MGB (Applied Biosystems). Specific primers and probes of HPRT1, PGK1, PPIA and RPLP0 (Applied Biosystems) were Human TaqMan® (Applied Biosystems) predeveloped endogeneous controls or designed virtually by means of primer express software to select those primers and probes that met the TaqMan® PCR requirements. The sequences designed for the primers and probes are shown in table 1.

The PCR reactions were prepared in a final volume of 25 μL, with a final concentration of $1 \times$ TaqMan® Universal PCR Master Mix (Applied Biosystems) together with 400 $nm \cdot L^{-1}$ of each forward and reverse primer, and 100 $nm \cdot L^{-1}$ probe (TaqMan). Then, 62.5 ng of cDNA were added, and samples were processed under the following cycling parameters. The thermal cycling condition comprised an initial incubation at 50°C for 2 min, 95°C for 10 min, followed by 40 cycles of denaturation at 95°C for 15 s, and annealing and extension at 60°C for 1 min. Each measurement was performed in duplicate and the threshold cycle (Ct) was determined. No template was used as a control, and a standard curve of four to five serial dilutions of a cDNA mixture were performed for all runs.

Statistical analysis
All data were analysed using standard computer software. Independent samples tests were performed with Levene's test for equality of variance and paired t-test for equality of means between Ct values of tumour subgroups (adenocarcinoma and other pathological types). Correlations were determined by Pearson's test. All p-values <0.001 were considered statistically significant.

RESULTS
Variable expression of seven endogenous control genes
Differential expression levels and dispersion of individual Ct values from the mean Ct value of seven endogenous control genes in NSCLC tumours, normal lung tissues, and NSCLC cell lines were identified (fig. 1). Data are presented as mean $\pm$ SD. The lowest expression level (the highest Ct value (mean $\pm$ SD)) was seen for the HPRT1 gene (Ct=26.87 ± 0.93) and the highest expression was for the PPIA gene (Ct=20.22 ± 1.44) in NSCLC tumours. The lowest and the highest expression levels seen in the normal lung tissues, respectively, were, HPRT1 gene (Ct=26.53 ± 0.60) and B2M gene (Ct=20.16 ± 0.66). The lowest and the highest expression levels in the NSCLC cell lines, respectively, were ABL1 gene (Ct=23.51 ± 1.38) and PPIA (Ct=17.43 ± 0.66) (table 2). There was a trend that all endogenous control genes, except B2M, expressed higher in NSCLC cell lines than in lung tumours and normal lung tissues. The variances of Ct value of each endogenous control gene were compared across different

TABLE 1	The sequences designed for the primers and probes		
ECG	**Probe**	**Primer**	
		Forward	**Reverse**
GAPDH	CAAGCTTCCCGTTCTCAGCC	GAAGGTGAAGGTCGGAGTC	GAAGATGGTGATGGGATTTC
ABL1	CCATTTTTGGTTTGGGCTTCACACCATT	TGGAGATAACACTCTAAGCATAACTAAAGGT	GATGTAGTTGCTTGGGACCCA
B2M	CCTCCATGATGCTGCTTACATGTCTC	GAGTATGCCTGCCGTGTG	AATCCAAATGCGGCATC

All sequences are written in 5'–3' sequence. ECG: endogenous control genes; GAPDH: glyceraldehyde-3-phosphate dehydrogenase; ABL1: v-abl Abelson murine leukaemia viral oncogene homologue 1; B2M: beta-2-microglobulin.

sample groups (table 3). GAPDH (0.71) expression, showed the smallest variance for NSCLC tumours (n=18) followed by that of RPLP0 (0.76). The smallest variability in normal lung tissues (n=10) was HPRT1 (0.36) or ABL1 (0.36) expressions, followed by that of GAPDH (0.40). In NSCLC cell lines (n=6), RPLP0 (0.75) expression had the smallest variance, followed by that of GAPDH (0.31). However, the smallest variance observed in NSCLC tumours and normal lung tissues combined (n=28) was for the expression of GAPDH (0.59), followed by that of HPRT1 (0.69) and RPLP0 (0.71). Conversely, the B2M expression showed the largest variance in lung tumour and normal lung tissues combined (2.17, table 3). To investigate whether pathologically different tumours affect the systemic variation, the NSCLC tumours were further divided into two subgroups according to their histopathological patterns: adenocarcinoma (AD; n=14) and the remaining other pathological types (three squamous cell carcinomas, one large cell carcinoma). The expression of GAPDH had the smallest variance for both subgroups (0.67 and 0.57), and the B2M and PGK1 expression had the largest variance (1.69 and 2.57; table 3). Overall, it seemed that the expression of GAPDH had the smallest variance and B2M had the largest variance among the seven endogenous control genes. Variations of the subgroup had minimal affect on systemic variation.

Correlation between endogenous control genes
Although different expression levels were noted for the endogenous control genes in NSCLC tumour samples, normal lung tissues and NSCLC cell lines, some endogenous control genes expressed were strongly correlated (table 4). The strongest correlation was between GAPDH and HPRT1 (r=0.9328, p<0.0001) in pooled NSCLC tumours and normal lung tissues (fig. 2 and table 4) and between the Ct value of endogenous control genes in NSCLC tumours or normal lung tissues (data not shown). However, the Ct values of B2M showed weak correlation with other endogenous control genes in lung tumours, normal lung tissue, NSCLC cell lines and the pooled lung tumour and normal lung tissues (table 4).

### Linear regression and difference plot of Ct and mean Ct
To examine whether the expression of a single gene can represent the mean expression of the whole cluster of seven genes, the Ct value of each individual endogenous control gene *versus* the mean Ct of the remaining six endogenous control genes for tumour and normal lung tissue samples was plotted. Coefficients of correlation (r-value) were then calculated for

lung cancer, normal lung tissues and combined tumour and normal lung tissues (table 5). The expression of GAPDH showed a high correlation coefficient for lung cancer tissues (r=0.8382, p<0.001), normal lung tissues (r=0.9702, p<0.001), and combined lung cancer and normal lung tissues (r=0.8605, p<0.001). The GAPDH expression in NSCLC cell lines was weakly correlated with the mean Ct value of the remaining six endogenous control genes (r=0.77, p=0.073, data not shown) The mean difference between the Ct of each gene and the mean Ct of the remaining six genes was also calculated. The mean difference referred to fixed distance between the expression of each endogenous control gene and the mean expression of the remaining genes. Clearly, GAPDH expression had the lowest mean difference, *i.e.* its expression was almost equal to the mean expression of the remaining six endogenous control genes. Most importantly, the accuracy (2×SD) represents the degree of diversion from the expression of each endogenous control gene to that of the remaining genes. Again, GAPDH expression only maximally deviates ±0.47 Ct (PCR cycles) from the expression measurement of other genes combined (for 95% of the tissue samples) in pooled lung cancer and normal lung tissues (fig. 3).

DISCUSSION
The main findings of the present study are as follows: 1) GAPDH mRNA expression had the least variance in pooled lung tumour and normal lung tissue samples, adenocarcinoma, and other pathological type subgroup; 2) GAPDH mRNA expression had strong correlation with the mRNA expression of each of the six other endogenous control genes in lung tumour and in normal lung tissue samples (data not shown); 3) GAPDH mRNA expression had the strongest correlation with that of the remaining six endogenous control genes in pooled lung tumour and normal lung tissue samples; and 4) GAPDH mRNA expression had the lowest mean difference and the highest accuracy with that of the remaining six endogenous control genes in pooled lung tumour and normal lung tissue samples.

This study analysed, by quantitative RT-PCR assay, the gene expression of seven common endogenous control genes in NSCLC tumours, normal lung tissues and NSCLC cell lines. In general, the expression levels of these seven endogenous control genes were lower in tissue specimens than in NSCLC cell lines (tables 2 and 3). This phenomenon is also noted in leukaemia cell lines [16]. A possible mechanism for high endogenous control gene expression in cell lines is that serum
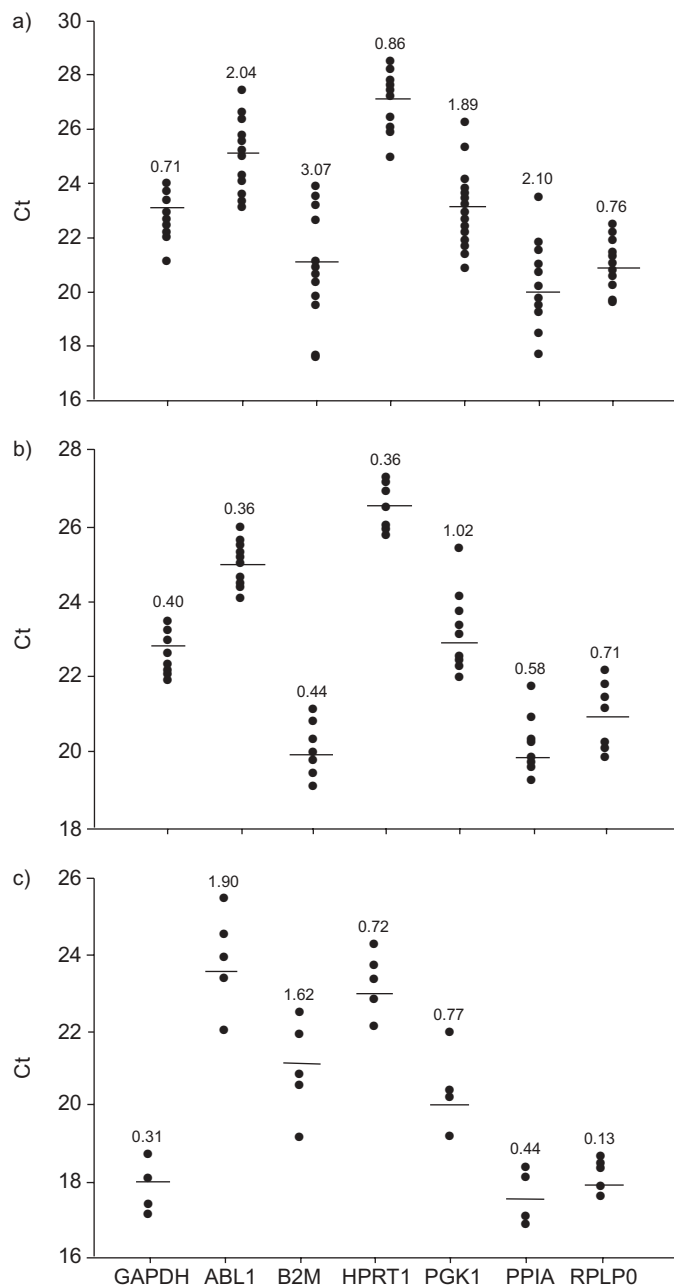
FIGURE 1. Differential mRNA expression of seven endogenous control genes in three different groups of lung tissues. a) The lung tumour tissue group. b) The normal lung tissue group. c) The nonsmall cell lung cancer cell lines. The horizontal bar represents the mean cycle threshold (Ct) value of each endogenous control gene in each specific tissue group. Numbers above the dot plots represent variance of expression of a particular gene among the samples. GAPDH: glyceraldehyde-3-phosphate dehydrogenase; ALB1: v-abl Abelson murine leukaemia viral oncogene homologue 1; B2M: beta-2-microglobulin; HPRT1: hypoxanthin phosphoribosyl-transferase 1; PGK1: phosphoglycerate kinase 1; PPIA peptidylprolyl isomerase A; RPLP0: ribosomal protein, large, P0.

stimulation influences the expression of several common endogenous control genes [25]. The mechanism of this phenomenon requires further study. To identify the appropriate endogenous control genes for RNA quantity normalisation, two criteria were adopted for selection of the most

TABLE 2	Comparison of mean cycle threshold (Ct) value across different sample groups and subgroups					
	GAPDH	ABL1	B2M	HPRT1	PGK1	PPIA	RPLP0
T	22.89	25.03	20.79	26.80	22.89	20.22	20.84
N	22.73	24.98	20.16	26.53	23.21	20.25	20.84
CL	17.93	23.51	21.21	23.37	20.34	17.43	18.21
T+N	22.83	25.01	20.57	26.75	23.01	20.23	20.84
AD	22.72	24.73	20.41	26.79	22.74	20.01	20.75
OT	23.46	26.10	22.13	27.15	23.41	20.98	21.14

Data are represented as mean Ct values of each endogenous control gene in different sample groups and subgroups. GAPDH: glyceraldehyde-3-phosphate dehydrogenase; ALB1: v-abl Abelson murine leukaemia viral oncogene homologue 1; B2M: beta-2-microglobulin; HPRT1: hypoxanthin phosphoribosyltransferase 1; PGK1: phosphoglycerate kinase 1; PPIA: peptidylprolyl isomerase A; RPLP0: ribosomal protein, large, P0; T: lung tumours (n=18), N: normal lung tissues (n=10); CL: nonsmall cell lung cancer cell lines (n=6); T+N: pooled lung tumour and normal lung tissue samples (n=28); AD: adenocarcinoma (n=14); OT: other pathological types of tumours (n=6).

suitable endogenous control gene: 1) the lowest variance in NSCLC tissue gene expression; and 2) the most strongly correlated gene expression among seven endogenous control genes.

To select the most suitable endogenous control gene in NSCLC cell lines for RT-PCR, the two selection criteria were applied. The mRNA expressions of seven endogenous control genes were evaluated in six different NSCLC cell lines. Expressions of RPLP0 (0.13) and GAPDH (0.31) had the least variability of these NSCLC cell lines. Correlation analysis demonstrated that the Ct values of the RPLP0 gene correlated strongly with most Ct values of the remaining six genes. The Ct values of the GAPDH gene correlated strongly with those of the B2M, HPRT1, PGK1 and RPLP0 genes, and poorly with those of ABL1 and PPIA genes (data not shown). Furthermore, the Ct values of the RPLP0 gene correlated best with the mean Ct values of the remaining six genes (r=0.8407). The accuracy of the RPLP0 gene was as good as that of the GAPDH gene (0.84). We concluded that the RPLP0 gene is the most suitable for endogenous control in NSCLC cell lines. This analytical result is compatible in other cell lines, as the RPLP0 gene is the most suitable for endogenous control in B- and T-cell lines [16].

In lung tumours, the normal lung tissue group and the NSCLC cell line group, the gene GAPDH expression had the second lowest variance (0.72, 0.40 and 0.31). However, in pooled lung tumour and normal lung tissue group, GAPDH expression had the lowest variance (0.59). The GAPDH was also the gene with the lowest variance both in adenocarcinoma and other pathological type subgroups (tables 2 and 3). The expression of GAPDH has exhibited marked variability among normal colon epithelium [8], human prostate carcinoma [26] and normal human cells [27]. However, analytical results showed that GAPDH gene expression had the lowest variance in pooled NSCLC tumours and normal lung tissues (table 2). Alternatively, B2M, an integral part of the histocompatibility

TABLE 3	Comparison of variance across different sample groups and subgroups						
	GAPDH	**ABL1**	**B2M**	**HPRT1**	**PGK1**	**PPIA**	**RPLP0**
T	0.71	2.04	3.07	0.86	1.89	2.10	0.76
N	0.40	0.36	0.44	0.36	1.02	0.58	0.71
CL	0.31	1.90	1.62	0.72	0.77	0.44	0.13
T+N	0.59	1.41	2.17	0.69	1.56	1.51	0.71
AD	0.67	1.34	1.69	0.96	1.33	1.53	0.90
OT	0.57	1.79	1.95	0.70	2.57	0.75	0.64

Data are represented as variances of cyclic threshold values of each endogenous control gene in different sample groups and subgroups. GAPDH: glyceraldehyde-3-phosphate dehydrogenase; ABL1: v-abl Abelson murine leukaemia viral oncogene homologue 1; B2M: beta-2-microglobulin; HPRT1: hypoxanthin phosphoribosyltransferase 1; PGK1: phosphoglycerate kinase 1; PPIA peptidylprolyl isomerase A; RPLP0: ribosomal protein, large, P0; T: lung tumours (n=18); N: normal lung tissues (n=10); CL: nonsmall cell lung cancer cell lines (n=6); T+N: pooled lung tumour and normal lung tissue samples (n=28); AD: adenocarcinoma (n=14); OT: other pathological types of tumours (n=6).

leukocyte antigen (HLA) class I complex, was a strongly affected endogenous control gene in NSCLC tumours and NSCLC cell lines. This result is compatible with the dramatic changes of B2M in different lymphoma [28, 29].

Variance represents the degree of deviation from the mean. It is possible that one endogenous control gene, which expresses the least variance of the Ct value, does not have to correlate well with the other endogenous control gene. Analysis by correlation coefficient method only has shortcomings. Correlation coefficients may be unsuited to analysing the expression stability of the tested control genes because the data range between the minimum and maximum expression levels, or any outlying value, could have a dramatic influence on the slope of the regression line and, consequently, on the value of the correlation coefficient [8]. Take GAPDH, for example: a

TABLE 4	Correlation between seven endogenous control genes in pooled tumour and normal lung tissues					
ECG	**ABL1**	**B2M**	**HPRT1**	**PGK1**	**PPIA**	**RPLP0**
GAPDH	0.7183	0.5801	0.9328	0.6595	0.6771	0.7890
ABL1		0.6174	0.6782	0.7399	0.8004	0.3594
B2M			0.5205	0.5152	0.5169	0.3183
HPRT1				0.5684	0.5798	0.6775
PGK1					0.9066	0.5768
PPIA						0.5903

Data are represented as coefficients (r-value) between each two endogenous control genes. ECG: endogenous control genes; GAPDH: glyceraldehyde-3-phosphate dehydrogenase; ALB1: v-abl Abelson murine leukaemia viral oncogene homologue 1; B2M: beta-2-microglobulin; HPRT1: hypoxanthin phosphoribosyltransferase 1; PGK1: phosphoglycerate kinase 1; PPIA peptidylprolyl isomerase A; RPLP0: ribosomal protein, large, P0.
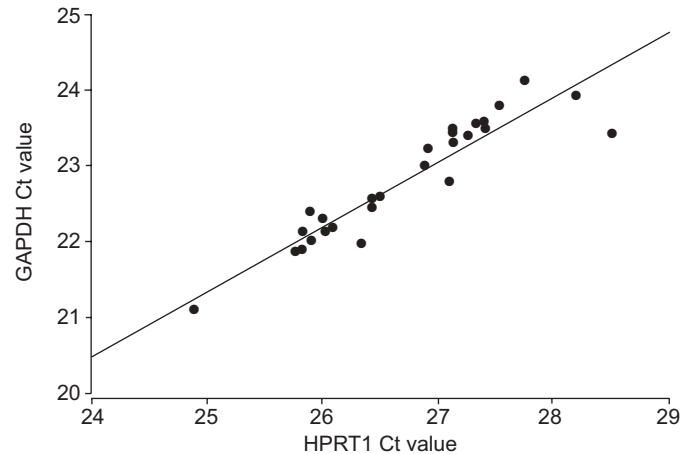

FIGURE 2. Scatter plot of the quantitative RT-PCR of endogenous control gene expression results. A correlation coefficient was calculated as r-value (r=0.9328), glyceraldehyde-3-phosphate dehydrogenase (GAPDH; y-axis) *versus* hypoxanthin phosphoribosyltransferase 1 (HRPT1; x-axis) mRNA expression value. In this correlation, 28 samples (18 lung tumours, 10 normal lung tissues) were examined. Ct: cyclic threshold.

lineal regression measurement was performed by plotting Ct values of the expression levels between each two endogenous control genes in NSCLC tumours, normal lung tissues, NSCLC cell lines and three pooled groups of tissues. The GAPDH had an obvious strong correlation with other endogenous control genes (table 4). However, GAPDH did not have a lower variance of Ct value in pooled lung tumours, normal tissues and lung cancer cell lines. This finding is due to the Ct values

TABLE 5	Coefficients of correlation and independent relationship between each individual gene and the mean cyclic threshold (Ct) of the remaining six genes				
Control genes	**Coefficients of correlation**			**MD T+N**	**AC T+N**
	T	**N**	**T+N**		
GAPDH	0.8382	09702	0.8605	0.09	0.94
ABL1	0.8116	0.8364	0.8049	2.63	1.41
B2M	0.6630	0.3579	0.6010	-2.53	2.35
HPRT1	0.7302	0.9863	0.7660	4.66	1.22
PGK1	0.8253	0.8972	0.8106	0.29	1.48
PPIA	0.8289	0.9212	0.8361	-2.80	2.40
RPLP0	0.5661	0.8672	0.6161	-2.20	1.61

Data are represented as Pearson's correlation coefficients (r-value). The mean difference (Ct-mean Ct) represents the difference in height of expression, whereas the accuracy represents $2 \times$ SD. T: lung cancer tissues; N: normal lung tissues; T+N: pooled lung cancer and normal lung tissues; MD: mean difference; AC: accuracy; GAPDH: glyceraldehyde-3-phosphate dehydrogenase; ALB1: v-abl Abelson murine leukaemia viral oncogene homologue 1; B2M: beta-2-microglobulin; HPRT1: hypoxanthin phosphoribosyltransferase 1; PGK1: phosphoglycerate kinase 1; PPIA peptidylprolyl isomerase A; RPLP0: ribosomal protein, large, P0.
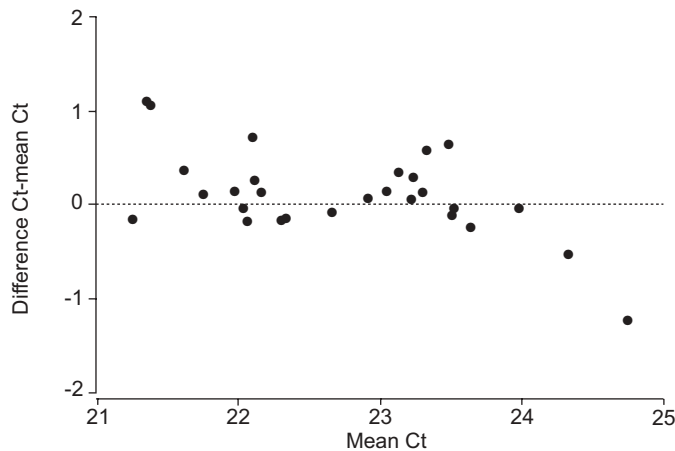
FIGURE 3. Difference plot. The plot represents the expression of glyceraldehyde-3-phosphate dehydrogenase (GAPDH; cyclic threshold (Ct) value) minus mean Ct value of the remaining six endogenous control genes (y-axis) compared with the mean expression (Ct value) of six other endogenous control genes (x-axis). The accuracy refers to $2 \times SD$.

of the lung cancer cell lines being located in the left lower end of the regression line (data not shown). The current authors, therefore, conducted both variance analysis and lineal regression measurement by plotting Ct values of the expression levels and calculating their variances between endogenous control genes in NSCLC tumours and normal lung tissues. The data showed that the Ct value of GAPDH had the lowest variance and correlated strongly with other endogenous control genes in pooled NSCLC tumour and normal lung tissues (tables 2 and 3, fig 2).

The GAPDH mRNA expression was also compared with β-actin mRNA and 18S-rRNA expressions in NSCLC. The results demonstrated that the GAPDH mRNA expression had the least variance (0.017), 18S-rRNA expression variance was next (0.031) and the β-actin mRNA expression had the greatest variance (0.4143). The GAPDH mRNA expression correlated best with 18S-rRNA expression and β-actin mRNA expression correlated poorly with 18S-rRNA expression (data not shown). Although 18S-rRNA expression had a low variance, it has limitations [16]. The 18S-rRNA gene is more stable and is less affected by degradation than most other genes; it is not expressed as mRNA, thus limiting its usage as an endogenous control. Overall, when applying the same methods (variance and correlation) to compare GAPDH mRNA, β-actin mRNA and 18S-rRNA expressions, the most suitable gene to serve as an internal RNA quality and quantity control in NSCLC specimens was GAPDH.

ANDERSEN *et al*. [19] stated that the variation, on average, of multiple genes is smaller than the variation in individual genes. In spite of this fact, the use of multiple normalisation genes, rather than a single gene, does not imply improved normalisation. It is of particular importance with regard to the clinical point of view as the choice of one single gene suitable for endogenous control is more economical and convenient than multiple genes in dealing with high throughput of heterogenous clinical samples. Therefore, to select a simple gene that could replace multiple gene measurements, the

current authors tested the correlation between the expression of each single endogenous control gene and the mean expression of the remaining six genes. The results showed that GAPDH expression is strongly correlated with the mean expression of the other six endogenous control genes, both in lung cancer and normal lung tissues (table 5). In addition, GAPDH had the lowest mean difference (Ct-mean Ct) and accuracy ($2 \times SD$). In difference plot, an accuracy of ± 0.47 PCR cycles (Ct) implies that normalisation may differ only 1.5-fold ($2^{0.48}$), using the expression of GAPDH instead of the mean expression of all genes.

In conclusion, the authors propose that GAPDH is the most suitable endogenous control among these seven genes for RNA quality and quantity control in both lung tumours and normal lung tissues. In clinical practice, when only a small amount of RNA from clinical samples is available (*e.g.* after microdissection or needle biopsies) and the gene of interest has a high expression, the high-expression GAPDH is suitable for the endogenous control gene. Alternatively, when the gene of interest has intermediate or low expression or in all situations when sufficient cells are present, a low-copy housekeeping control (best for RNA isolation efficiency, RNA quality, and RT-efficiency, HPRT1, for example) is the choice after GAPDH. Based on findings in this investigation, the expression of the genes of interest in different samples of NSCLC should be normalised by concomitant measurement of the GAPDH gene products in order to eliminate or lower systemic bias. Compared with previously published procedures for identifying suitable normalisation genes by purely mathematical modelling [19], the present approach provides more direct measures, takes into account systematic differences between sample subgroups and is less tedious.

The authors suggest that glyceraldehyde-3-phosphate dehydrogenase can be used as an endogenous control gene for real-time quantitative PCR in studies of gene expression in nonsmall cell lung cancer tumours and normal lung tissue samples.

REFERENCES

1 Jemal A, Murray T, Samuels A, Ghafoor A, Ward E, Thun MJ. Cancer statistics, 2003. *CA Cancer J Clin* 2003; 53: 5–26.

2 Taiwan Area Main Causes of Death (2004), Abstract and Analysis (Chinese version), Department of Health, Executive Yuan, Taiwan, R.O.C. website. http://www.doh.gov.tw/statistic.

3 Kuo CW, Chen YM, Chao JY, Tsai CM, Perng RP. Non-small cell lung cancer in very young and very old patients. *Chest* 2000; 117: 354–359.

4 Sozzi G, Conte D, Leon M, *et al*. Quantification of free circulating DNA as a diagnostic marker in lung cancer. *J Clin Oncol* 2003; 21: 3902–3908.

5 Nacht M, Dracheva T, Gao Y, *et al*. Molecular characteristics of non-small cell lung cancer. *Proc Natl Acad Sci USA* 2001; 98: 15203–15208.

6 Player A, Gillespie J, Fujii T, *et al*. Identification of TDE2 gene and its expression in non-small cell lung cancer. *Int J Cancer* 2003; 107: 238–243.

7 Gautschi O, Bigosch C, Huegli B, *et al.* Circulating deoxyribonucleic acid as prognostic marker in non-small-cell lung cancer patients undergoing chemotherapy. *J Clin Oncol* 2004; 22: 4157–4164.

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: research0034.1–0034.11.

9 Bustin SA. Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. *J Mol Endocrinol* 2000; 25: 169–193.

10 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.

11 Suzuki T, Higgins PJ, Crawford DR. Control selection for RNA quantitation. *Biotechniques* 2000; 29: 332–337.

12 Thellin O, Zorzi W, Lakaye B, *et al.* Housekeeping genes as internal standards: use and limits. *J Biotechnol* 1999; 75: 291–295.

13 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.

14 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: 143–147.

15 Beillard E, Pallisgaard N, van der Velden VH, *et al.* Evaluation of candidate control genes for diagnosis and residual disease detection in leukemic patients using ''real-time'' quantitative reverse-transcriptase polymerase chain reaction (RQ-PCR) - a Europe against cancer program. *Leukemia* 2003; 17: 2474–2486.

16 Lossos IS, Czerwinski DK, Wechser MA, Levy R. Optimization of quantitative real-time RT-PCR parameters for the study of lymphoid malignancies. *Leukemia* 2003; 17: 789–795.

17 Dheda K, Huggett JF, Bustin SA, Johnson MA, Rook G, Zumla A. Validation of housekeeping genes for normalizing RNA expression in real-time PCR. *Biotechniques* 2004; 37: 112–119.

18 Schmid H, Cohen CD, Henger A, Irrgang S, Schlondorff D, Kretzler M. Validation of endogenous controls for gene expression analysis in microdissected human renal biopsies. *Kidney Int* 2003; 64: 356–360.

19 Andersen CL, Jensen JL, Omntoft 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.

20 Calvo R, West J, Franklin W, *et al.* HOX and WNT7A expression in human lung cancer. *Proc Natl Acad Sci USA* 2000; 97: 12776–12781.

21 Yuan A, Yang PC, Yu CJ, *et al.* Interleukin-8 messenger ribonucleic acid expression correlates with tumor progression, tumor angiogenesis, patient survival, and timing of relapse in non-small-cell lung cancer. *Am J Respir Crit Care Med* 2000; 162: 1957–1963.

22 Shih JY, Yang SC, Hong TM, *et al.* Collapsin response mediator protein-1 and the invasion and metastasis of cancer cells. *J Natl Cancer Inst* 2001; 93: 1392–1400.

23 Bepler G, Sharma S, Cantor A, *et al.* RRM1 and PTEN as prognostic parameters for overall and disease-free survival in patients with non-small-cell lung cancer. *J Clin Oncol* 2004; 22: 1878–1885.

24 Dworakowska D, Jassem E, Jassem J, *et al.* MDM2 gene amplification: a new independent factor of adverse prognosis in non-small cell lung cancer (NSCLC). *Lung Cancer* 2004; 43: 285–295.

25 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.

26 Ripple MO, Wilding G. Alteration of glyceraldehyde-3-phosphate dehydrogenase activity and messenger RNA content by androgen in human prostate carcinoma cells. *Cancer Res* 1995; 55: 4234–4236.

27 Mansur NR, Meyer-Siegler K, Wurzer JC, Sirover MA. Cell cycle regulation of the glyceraldehyde-3-phosphate dehydrogenase/uracil DNA glycosylase gene in normal human cells. *Nucleic Acids Res* 1993; 21: 993–998.

28 Aviles A, Zepeda G, Diaz-Maqueo JC, *et al.* Beta 2 microglobulin level as an indicator of prognosis in diffuse large cell lymphoma. *Leuk Lymphoma* 1992; 7: 135–138.

29 Lossos IS, Akasaka T, Martinez-Climent JA, Siebert R, Levy R. The BCL6 gene in B-cell lymphomas with 3q27 translocations is expressed mainly from the rearranged allele irrespective of the partner gene. *Leukemia* 2003; 17: 1390–1397.