Available online at www.sciencedirect.com
SCIENCE DIRECT®
Biochemical and Biophysical Research Communications 337 (2005) 306–312
BBRC
www.elsevier.com/locate/ybbrc

Selection of control genes for quantitative RT-PCR based on microarray data

Natalia Shulzhenko [a,1], Anatoly Yambartsev [b], Amador Goncalves-Primo [a], Maria Gerbase-DeLima [a], Andrey Morgun [a,*,1]

[a] Immunogenetics Division, Pediatrics Department, Universidade Federal de Sao Paulo—UNIFESP, Sao Paulo, SP, Brazil
[b] Institute of Mathematics and Statistics, Universidade de Sao Paulo—USP, Sao Paulo, SP, Brazil

Received 26 August 2005
Available online 16 September 2005

Abstract

Use of internal reference gene(s) is necessary for adequate quantification of target gene expression by RT-PCR. Herein, we elaborated a strategy of control gene selection based on microarray data and illustrated it by analyzing endomyocardial biopsies with acute cardiac rejection and infection. Using order statistics and binomial distribution we evaluated the probability of finding low-varying genes by chance. For analysis, the microarray data were divided into two sample subsets. Among the first 10% of genes with the lowest standard deviations, we found 14 genes common to both subsets. After normalization using two selected genes, high correlation was observed between expression of target genes evaluated by microarray and RT-PCR, and in independent dataset by RT-PCR ($r = 0.9$, $p < 0.001$). In conclusion, we showed a simple and reliable strategy of selection and validation of control genes for RT-PCR from microarray data that can be easily applied for different experimental designs and tissues.

Keywords: Normalization; RT-PCR; Microarray; Housekeeping gene; Internal reference genes; Heart transplant

In studies of target gene expression by RT-PCR, the use of one or more internal reference genes is necessary to control for RNA quality, reverse transcription efficiency, and overall transcriptional activity in samples. For this purpose, housekeeping genes, such as glycerol-3-phosphate dehydrogenase (*GAPDH*), β-actin, and 28S or 18S ribosomal RNAs, are commonly employed [1,2]. However, several studies have demonstrated that expression of these genes might vary as a result of tissue type, experimental conditions or pathological state [3–6]. In addition, the presence of intron-lacking pseudogenes for most housekeeping genes might interfere with the mRNA quantification unless the samples are treated with DNAse. This treatment, however, is often incomplete, leaving trace amounts of DNA and may also affect RNA, complicating the quantification of medium and low expressed target genes [7]. Ribosomal 28S or 18S RNAs, which reflect the total RNA levels, besides requiring DNAse treatment, are not representative of the mRNA fraction. Therefore, instead of the simple use of any housekeeping gene, a careful selection of internal reference genes is necessary. Indeed, different authors tried to make such selection among some housekeeping/maintenance genes [4,5,8,9]. However, microarray studies containing information on the expression of huge amount of potential control transcripts allow a more comprehensive and unbiased selection of genes appropriate for normalization.

In the present study, we elaborated a new algorithm for the selection and validation of control genes for quantitative RT-PCR based on microarray results and illustrated it by the study of gene expression in endomyocardial biopsies during acute cardiac rejection and infection.

* Corresponding author. Present address: Ghost Lab, LCMI, NIAID, NIH, 9000 Rockville Pike, Bethesda, MD 20892, USA. Fax: +1 301 496 4286.
E-mail address: morguna@niaid.nih.gov (A. Morgun).
[1] These authors contributed equally to this work.

0006-291X/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.bbrc.2005.09.048

Methods

Samples. Seventy-five endomyocardial biopsies with known cases of acute rejection ($n = 24$), non-rejection ($n = 44$), or post-transplant *Trypanosoma cruzi* infection (Chagas' disease) reactivation ($n = 7$) were obtained during routine biopsy procedures within the first six months after transplantation from 32 adult cardiac transplant recipients. The protocol was approved by the Ethics Committee of Universidade Federal de São Paulo and informed consent was obtained from all subjects under study.

Microarrays. Microarrays were performed in two independent sets of samples at two different time points and therefore constituted two subsets (first and second) in order to increase the statistical power of the analysis. Total RNA was isolated with the TRIzol reagent (Invitrogen) and amplified in two rounds [10] with aminoallyl-UTP (Ambion) incorporation during the second round of in vitro transcription and then labeled with Cy5 or Cy3 fluorophores (Amersham) and hybridized on 14K oligonucleotide microarrays. Oligos were designed and synthesized by Qiagen/Operon and printed onto the slides at NIAID Microarray Facility, NIH. All test RNAs were Cy5 labeled and competitively hybridized with the reference RNA sample that consisted of a pool of amplified RNAs from 92 non-rejection and 35 rejection biopsies. The images were scanned and expression values were obtained with the GenePix Scanner and Software (Axon Instruments). Differentially expressed genes (manuscript submitted) were revealed using BRB ArrayTools (version 3.0) developed by Dr. Richard Simon and Amy Peng.

We searched for control genes among those that passed the following spot filtering criteria: spots with a diameter between 10 and 300 μm; signal-to-background ratio >2; >50% of spot pixels are one standard deviation (*SD*) above background; present in at least 90% of arrays. Global normalization was used to median the center of the log-ratios on each array. The data discussed in this work have been deposited in NCBI's Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo/) under GEO Series Accession No. GSE2596.

Real-time RT-PCR. After reverse transcription of 1 μg of total RNA with oligo(dT) primer and SuperScript II reverse transcriptase using a common procedure described elsewhere [11], we applied the SYBR Green I real-time PCR technique using an ABI PRISM 7700 Sequence Detection System (Applied Biosystems). Primers (available upon request) were constructed using Primer Express software (Applied Biosystems) based on reference mRNA sequences. PCR was performed in a total volume of 11 μl containing 5.5 μl SYBR Green PCR Master Mix, 125 nM of sense and antisense primers, and 1 μl cDNA. We used the default cycling conditions of the ABI Prism Software (Applied Biosystems). Using a pool of 30 individual genomic DNAs, we checked for the absence of pseudogenes and contaminant genomic DNA amplifications. For quantification we used a cycle threshold (C_t) value against a standard curve constructed by amplification of the known copy number of target DNA in 10-fold serial dilutions.

Normalization index. Since we used two genes (*HK1*, hexokinase 1 and *DDOST*, dolichyl-diphosphooligosaccharide-protein glycosyltransferase) as controls, we wanted to combine their expression levels in a single value with an equal contribution by each gene independently of its level of expression. For this, expression values of one gene in each sample should be scaled in relation to the other. Therefore, we built a scatter-plot graph of *HK1* and *DDOST* expression values and estimated the linear regression coefficients (Fig. 2). Next, using these coefficients we "scaled" *DDOST* levels to the *HK1* levels as follows: $DDOST^{\#} = 0.4692\ DDOST + 2.1876$. Finally, a normalization index was calculated as a mean of $DDOST^{\#}$ and *HK1* values.

Theory and statistical analysis. The data can be represented by two tables $\{\xi_{ij}^{(1)}\}$ and $\{\xi_{ij}^{(2)}\}$, where $\xi_{ij}^{(r)}$ is the level of expression of the ith gene in the jth sample in the rth experimental table ($i = 1, \ldots, N$, $j = 1, \ldots, M_r$, and $r = 1, 2$). Note that the number of samples M_1 and M_2 can differ between the tables. For any gene i in the rth experimental table we calculate the standard deviation $SD(i,r)$ of its expression level $(\xi_{i1}^{(r)}, \xi_{i2}^{(r)}, \ldots, \xi_{iM_r}^{(r)})$:

$$(\xi_{i1}^{(1)}, \xi_{i2}^{(1)}, \ldots, \xi_{iM_1}^{(1)}) \to SD(i,1),$$
$$(\xi_{i1}^{(2)}, \xi_{i2}^{(2)}, \ldots, \xi_{iM_2}^{(2)}) \to SD(i,2).$$

We chose a value of standard deviation as a measure of variation of gene expression. For each table we ranked genes according to their standard deviations in increasing order. Denote $(i)_1$ and $(i)_2$ to be the order numbers of gene i in these two ordered subsets. We construct an underlying random model, supposing that for each gene i its orders $(i)_1$ and $(i)_2$ are independent and have a uniform distribution in the set $\{1, 2, \ldots, N\}$, where N is the total number of genes.

We chose arbitrary q% of genes with the lowest standard deviations in each group. Denote R to be a q% of total number of genes (N). Let G_1 and G_2 be the sets of these selected genes

$$\begin{aligned} G_1 &= \{\text{genes } i, \text{ such that } (i)_1 \leqslant R\}, \\ G_2 &= \{\text{genes } i, \text{ such that } (i)_2 \leqslant R\}. \end{aligned} \tag{1}$$

We are interested in the event that the order numbers of some genes are less than R in both groups simultaneously. Denote $G_{1,2}$ to be a set of overlapping genes

$$G_{1,2} = G_1 \cap G_2 = \{\text{genes } i, \text{ such that } (i)_1 \leqslant R \text{ and } (i)_2 \leqslant R\}. \tag{2}$$

Let s_1 be the maximal order number among overlapping genes in the first group and s_2 in the second group:

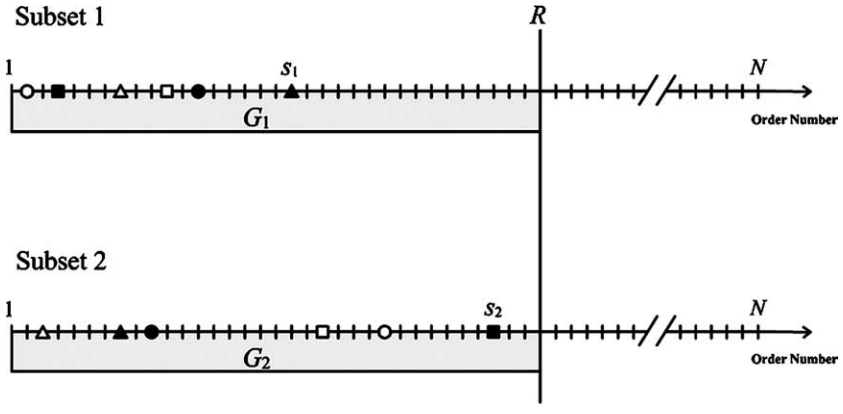

Fig. 1. Schematic representation of the model designed for selection of candidate control genes. The two scales represent order numbers of genes ranked by increasing standard deviations values in each subset. R is the arbitrary percentage (q%) of the total number of genes (N). Genes with order numbers less than or equal to R in both subsets are denoted by circles, squares or triangles. s_1 and s_2 represent highest order numbers of overlapping genes in each subset.

$$s_1 = \max\{(i)_1\text{: gene } i \text{ belongs to the set } G_{1,2}\},$$
$$s_2 = \max\{(i)_2\text{: gene } i \text{ belongs to the set } G_{1,2}\}. \quad (3)$$

Note that usually a gene with order number s_1 in the first group and a gene with order number s_2 in the second group are not the same. Let $K = |G_{1,2}|$ be the number of genes in the $G_{1,2}$ set. See Fig. 1 for a graphic illustration of denotations and model design.

Consider the event $A(s_1, s_2; K)$ where at least K ($0 \leqslant K \leqslant R$) genes appear in the $G_{1,2}$ set with the maximal order numbers s_1 and s_2 for the corresponding subsets:

$$A(s_1, s_2; K) = \{\text{exists at least } K \text{ genes } i_1, \ldots, i_K, \text{ such as } (i_l)_1 \leqslant s_1 \text{ and } (i_l)_2 \leqslant s_2 \text{ for any } l = 1, \ldots, K\}, \quad (4)$$

where s_1 and s_2 are equal to or less than R by definition (3). The probability of the event (4) gives us a P value for our observed number of genes K

$$P\{A(s_1, s_2; K)\} = \sum_{l=K}^{\min(s_1, s_2)} \binom{N}{l} \frac{s_1! s_2! (N - s_1)! (N - s_2)! (N - l)!}{(s_1 - l)! (s_2 - l)! N! N! (N - s_1 - s_2 + l)!}.$$

Note that there is an individual probability ($s_1 s_2 / N^2$) for one gene to have its order number no more than s_1 in the first group and no more than s_2 in the second. We can consider (4) as an event where at least K successes occurred in the Bernoulli model with N trials, with a success probability $s_1 s_2 / N^2$. Considering this approximation model we obtain

$$P\{A(s_1, s_2; K)\} \cong \sum_{l=K}^{N} \binom{N}{l} \left(\frac{s_1 s_2}{N^2}\right)^l \left(1 - \frac{s_1 s_2}{N^2}\right)^{N-l}. \quad (5)$$

Choosing q% sufficiently small, for example less than or equal to 10%, the probability (5) gives us a good approximation of the P value.

Note. The probability (5) can be calculated with Microsoft Excel in the following form

$$= 1 - BINOMDIST(K - 1, N, s_1 * s_2 / (N * N), TRUE), \quad (6)$$

where parameter *TRUE* means that function *BINOMDIST* is the cumulative distribution.

For other analyses we employed Spearman correlation and the Mann–Whitney test.

Results

Algorithm

Stage 1. Microarray filter. Apply spot and gene quality filters in order to select genes whose expression is reliably detected in most arrays.

Stage 2. Statistical analysis.

(1) obtain two subsets of microarrays (randomly or as the result of the two stage experimental design, e.g., training and test groups);
(2) calculate standard deviation (*SD*) of expression for each gene in each subset of experiments and rank genes in each subset based on *SD*;
(3) among the first 10% of genes with lowest *SD*, find genes overlapping in both subsets;
(4) determine the maximum order number (position) of an overlapping gene in each subset and calculate the probability of finding this number of genes by chance using Eq. (6).
(5) select genes whose expression evaluated in microarray experiments does not differ between the groups of samples under study.

Stage 3. Bioinformatics.

(1) verify gene sequence information in GenBank in order to design adequate RT-PCR (avoid genomic DNA amplification: check for the absence of processed pseudogenes, locate primers in different exons);
(2) if possible, select candidate control genes with levels of expression similar to those of target genes;

Stage 4. RT-PCR verification.

(1) perform RT-PCR for selected control gene(s);
(2) compare control genes expression levels obtained by RT-PCR between groups of the samples under study and select those without differences;
(3) calculate normalization index (see above) and use it for normalization of target gene expression.

Illustration of the algorithm

After applying the filtering criteria, as described under Methods, to the microarrays from endomyocardial biopsies, 591 genes (N) entered the next analysis step. In each subset of arrays, we ranked the genes according to their expression *SD*s. Among the first 10% (q%) of the genes with lowest *SD*s, we found 14 genes (K) common to both subsets (Table 1). Note that the same gene is usually found in different positions in each subset. For example, hexokinase 1 is the 29th gene in the first subset and the 21st in the second one. Next, we calculated the probability of finding by chance 14 genes common to both subsets of arrays in the 10% interval with positions $s_1 = 58$ and $s_2 = 45$. Using Eq. (6) in this case we obtain the following P value

$$P = 1 - BINOMDIST(14 - 1; 591; 58 * 45/(591 * 591); TRUE) = 0.0002.$$

To define for which genes quantitative RT-PCR could easily be designed and set up, we analyzed reference mRNA sequences for the 14 genes. Eight genes were rejected because of incomplete gene/mRNA sequence information or presence of pseudogenes that would greatly complicate adequate primer design.

The next criterion for gene selection was the absence of differences in expression between the groups analyzed, i.e., between rejection and non-rejection and between rejection and infection (Chagas' disease reactivation) (Table 2). Among the six remaining genes, three (*MYBPC3*, *CRIP2*, and *ACTC*) presented difference of expression in at least one comparison and were excluded from further analysis (Table 1).

Among the remaining three genes (*DDOST*, *HK1*, and *MYL2*), the first two presented expression levels similar to those of targets genes, while *MYL2* expression was much higher (many of the spots for *MYL2* were saturated). Both of the selected genes code for ubiquitously expressed molecules. *DDOST* encodes an intracellular enzyme that catalyzes N-glycosylation of newly formed peptides and *HK1* is

Table 1
List of genes overlapping between two subsets of samples in the first 10% of low-varying genes and selection of candidate control genes among them

Gene symbol	Gene description	First subset of samples		Second subset of samples		Sum of the order numbers	Pseudogene or incomplete sequence information	Difference between sample groups	Difference in the level of expression from target genes	Evaluation by RT-PCR
		Order number[a]	*SD*	Order number[a]	*SD*					
DDOST	Dolichyl-diphosphooligosaccharide-protein glycosyltransferase	8	0.356	9	0.312	17	NO	NO	NO	YES
MYBPC3	Myosin binding protein C, cardiac	23	0.377	3	0.265	26	NO	YES	n/a	n/a
EEF1G	Eukaryotic translation elongation factor 1 γ	19	0.373	12	0.324	31	YES	n/a	n/a	n/a
EEF1D	Eukaryotic translation elongation factor 1 δ	18	0.370	32	0.371	50	YES	n/a	n/a	n/a
HK1	Hexokinase 1	29	0.390	21	0.339	50	NO	NO	NO	YES
CRIP2	Cysteine-rich protein 2	16	0.369	44	0.394	60	NO	YES	n/a	n/a
MYL2	Myosin, light polypeptide 2, regulatory, cardiac, slow	**58**	0.433	2	0.253	60	NO	NO	YES	n/a
RBAF600	Retinoblastoma-associated factor 600	38	0.406	22	0.347	60	YES	n/a	n/a	n/a
ACTC	Actin, α, cardiac muscle	20	0.375	41	0.384	61	NO	YES	n/a	n/a
VPS4A	Vacuolar protein sorting 4A (yeast)	24	0.380	39	0.383	63	YES	n/a	n/a	n/a
OK/SW-cl.56	β 5-Tubulin	44	0.410	25	0.358	69	YES	n/a	n/a	n/a
SCMH1	Sex comb on midleg homolog 1 (*Drosophila*)	54	0.428	17	0.330	71	YES	n/a	n/a	n/a
DDX9	DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 9 (RNA helicase A)	55	0.430	19	0.336	74	YES	n/a	n/a	n/a
LOC51080	sec61 homolog	43	0.410	**45**	0.396	88	YES	n/a	n/a	n/a

SD, standard deviation; n/a, not applicable because the gene was eliminated in the previous step. Highest order numbers in each subset are given in bold.
[a] Position of the gene after classification of 591 genes by increasing order of *SD*.

Table 2
Median values of expression of six candidate control genes in endomyocardial biopsies evaluated by microarray and comparisons between them in the groups of samples analyzed

Gene symbol	Rejection (*R*)	Non-rejection (NR)	Infection (*I*)	*P*[a]	
				R vs. NR	*R* vs. *I*
DDOST	0.76	0.77	0.69	0.41	0.21
MYBPC3	0.86	0.54	0.97	0.31	0.01*
HK1	0.94	1.02	0.83	0.21	0.73
CRIP2	0.66	0.84	0.85	0.02*	0.33
MYL2	0.90	0.93	0.90	0.44	0.48
ACTC	1.34	0.78	0.82	0.05*	0.03*

Values are expressed as ratios between signal intensities in sample and reference RNAs.
[a] Mann–Whitney test.
* <0.05.

a glucose-phosphorylating enzyme that is involved in the first step of several metabolic pathways.

We evaluated expression of *DDOST* and *HK1* by RT-PCR in 46 biopsy samples previously analyzed by microarray. As expected, the comparison of absolute number of transcript copies in 1 μl of cDNA between different clinical groups did not show any difference for any of the two control genes, with median values of 725, 640, and 503 for *HK1* and 1769, 1415, and 1393 for *DDOST* in non-rejection, rejection, and Chagas' reactivation groups, respectively.

The normalization index was calculated (see Methods) for all samples based on linear regression between *DDOST* and *HK1* expression (Fig. 2) and used for the normalization of target gene expression. For 10 target genes selected from differentially expressed genes revealed by microarray, we observed similar differences for microarray and RT-PCR results between compared groups (fold change), $r = 0.94$, $p < 0.0001$ (Fig. 3A). Moreover, we performed RT-PCR for the same genes in the independent group of biopsy samples ($n = 30$). The differences detected in two sets of RT-PCR experiments also showed a high correlation with $r = 0.90$, $p < 0.0001$ (Fig. 3B).

Discussion

Currently, the use of real-time RT-PCR is a highly recommended approach for the validation of microarray results [12]. Therefore, it is very convenient to select internal control gene(s) for RT-PCR based on the existing microarray results. Moreover, microarrays containing a large number of genes have a much higher chance to reveal transcripts acceptable as internal controls than the evaluation of "traditional" housekeeping gene(s). In fact, genes with small variations across the arrays and not showing difference in expression between compared groups of samples are candidate control genes.

Some studies have explored microarrays as a source of potential control genes [13–16]. Each of them selects candidates for control genes among those with the lowest variation. Andersen et al. [13] proposed a model for gene ranking that incorporates intra- and intergroup variation

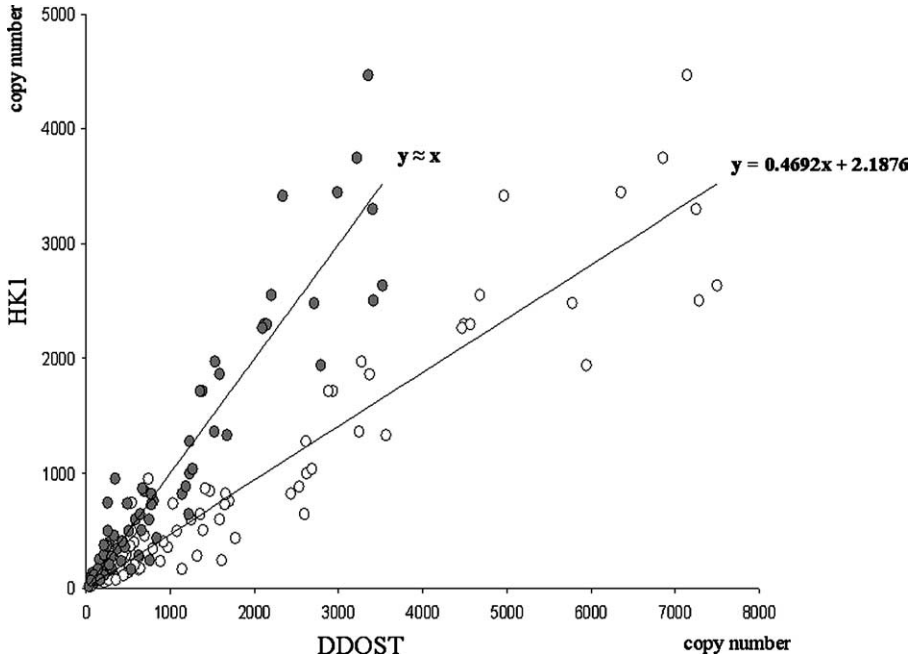

Fig. 2. Linear regression between *HK1* and *DDOST* gene expressions. Open circles represent absolute values of expression in copy number per microliter of cDNA. Filled circles represent values of expression after recalculation of *DDOST* levels using the equation: $DDOST^{\#} = 0.4692 * DDOST + 2.1876$.

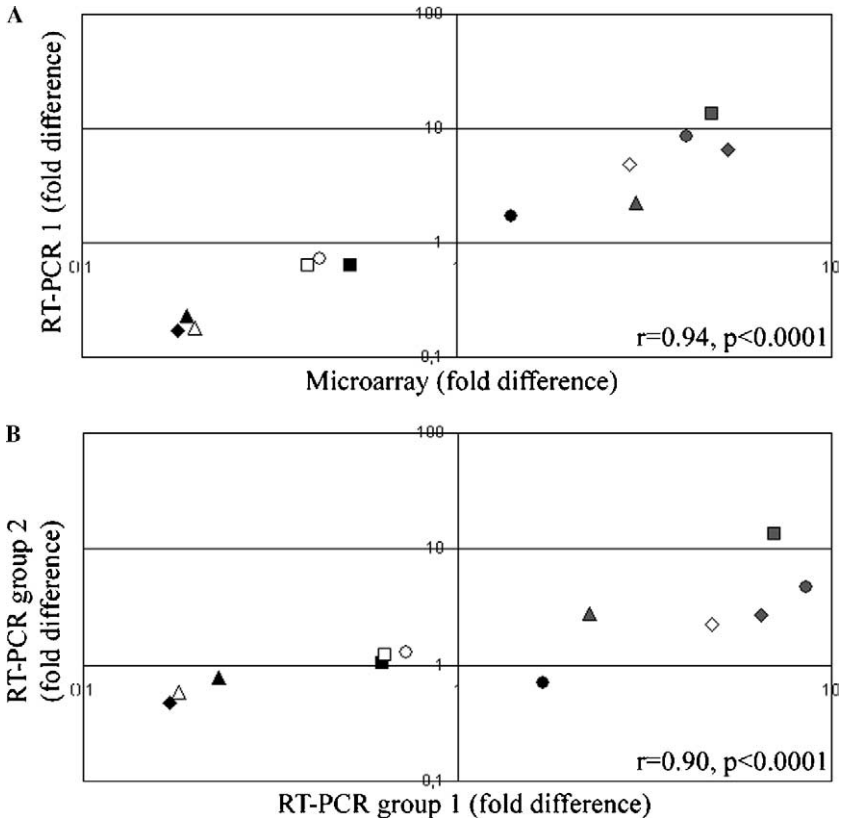

Fig. 3. Correlation between results obtained by microarray and RT-PCR (A) and between two independent groups of samples evaluated by RT-PCR (B). Each symbol represents the fold difference between expression levels in two clinical situations for one target gene.

estimates. Unfortunately, we could not analyze our data using their software (NormFinder) because it does not accept data with missing values. Since missing values are very common in microarray results, this will preclude a wide use of NormFinder, at least in its current version. Furthermore, none of the previous reports evaluated the probability of

finding candidate genes by chance. Consequently, genes found in these studies could be used for normalization of samples previously analyzed by microarray but might not be suitable for independent datasets. In fact, the probability estimate is critical to decide whether the applicability of selected genes is restricted to the same dataset from which these genes were selected (if probability > 0.05) or could be extended to new independent experiments with a similar design/question (if probability < 0.05).

The two set approach used in our study allows both obtaining genes with a low probability of finding them by chance and selection genes with uniform low expression variation. The benefit of such approach can be clearly illustrated by our data concerning the myoglobin gene, which would have been ranked as the second low-varying gene if data had not been divided into two subsets. However, this gene was not present among the overlapping genes between the two subsets in the first 10% of low-varying genes and, consequently, could not be considered as a candidate for control gene.

Although in our study only two genes satisfied all the adopted criteria, one could find a larger number of candidate control genes. The amount of control genes necessary for adequate normalization depends on several factors such as magnitude of expected differences for target genes, tissue type, RNA quality, and other practical and economical implications [6,13].

If the number of candidates satisfying the above cited criteria is greater than desired, an additional selection could be applied based on the level of variation (e.g., sum of order numbers in our case) and the probability of a difference between the groups studied (e.g., selecting genes with the highest P values for inter-group comparisons). Another important issue is how to combine gene expression if more than one gene is used for normalization. Vandesompele et al. [6] chose the geometrical mean to combine expression of several genes for normalization purposes. However, this approach leads to non-proportional contribution of different genes depending on their expression levels. In the present study, we re-scaled the expression level of one gene to the expression level of the second one before calculating the mean, thus allowing equal input of the two selected genes into the process of normalization, independently of their expression levels. Similarly, this could be applied to a larger amount of genes by re-scaling the levels of expression of all these genes to the expression level of anyone of them before mean or geometrical mean calculation.

Finally, with the objective to explore how well we can normalize gene expression using our control genes and whether their applicability could be extended to new experiments we evaluated the expression of candidate control genes and target genes by quantitative real-time RT-PCR in the group of samples previously analyzed by microarray and in a new independent group. The high correlation observed between fold differences of target gene expression obtained by microarray and RT-PCR, and in two sets of RT-PCR experiments is important evidence of consistency of our strategy. Furthermore, in our recent work on uterine cervical cancer (manuscript in preparation) we successfully applied this strategy to the analysis of gene expression in relation to the response to treatment.

In conclusion, we showed a simple and reliable strategy for the selection and validation of control genes for RT-PCR from microarray data that can be easily applied to different experimental designs and tissues. *HK1* and *DDOST* are the genes that can be used as internal controls for RT-PCR in endomyocardial biopsies when analyzing acute cardiac rejection and Chagas' disease reactivation.

Acknowledgments

We thank Anna Smirnova and Ricardo Borra for helpful discussion.

References

[1] T. Suzuki, P.J. Higgins, D.R. Crawford, Control selection for RNA quantitation, Biotechniques 29 (2000) 332–337.
[2] O. Thellin, W. Zorzi, B. Lakaye, B.B. Coumans, G. Hennen, T. Grisar, A. Igout, E. Heinen, Housekeeping genes as internal standards: use and limits, J. Biotechnol. 75 (1999) 291–295.
[3] J. Bereta, M. Bereta, Stimulation of glyceraldehyde-3-phosphate dehydrogenase mRNA levels by endogenous nitric oxide in cytokine-activated endothelium, Biochem. Biophys. Res. Commun. 217 (1995) 363–369.
[4] P.J. Gibbs, C. Cameron, L.C. Tan, S.A. Sadek, W.M. Howell, House keeping genes and gene expression analysis in transplant recipients: a note of caution, Transpl. Immunol. 12 (2003) 89–97.
[5] H.K. Hamalainen, J.C. Tubman, S. Vikman, T. Kyrola, E. Ylikoski, J.A. Warrington, R. Lahesmaa, Identification and validation of endogenous reference genes for expression profiling of T helper cell differentiation by quantitative real-time RT-PCR, Anal. Biochem. 299 (2001) 63–70.
[6] J. Vandesompele, P.K. De, F. Pattyn, B. Poppe, R.N. Van, P.A. De, F. Speleman, Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes, Genome Biol. 3 (2002), RESEARCH0034.
[7] G.J. Hurteau, S.D. Spivack, mRNA-specific reverse transcription-polymerase chain reaction from human tissue extracts, Anal. Biochem. 307 (2002) 304–315.
[8] J.B. de Kok, R.W. Roelofs, B.A. Giesenforf, J.L. Pennings, E.T. Waas, T. Feuth, D.W. Swinkels, P.N. Span, Normalization of gene expression measurements in tumor tissues: comparison of 13 endogenous control genes, Lab. Invest. 85 (2005) 154–159.
[9] M. Fischer, M. Skowron, F. Berthold, Reliable transcript quantification by real-time reverse transcriptase-polymerase chain reaction in primary neuroblastoma using normalization to averaged expression levels of the control genes HPRT1 and SDHA, J. Mol. Diagn. 7 (2005) 89–96.
[10] E. Wang, L.D. Miller, G.A. Ohnmacht, E.T. Liu, F.M. Marincola, High-fidelity mRNA amplification for gene profiling, Nat. Biotechnol. 18 (2000) 457–459.
[11] N. Shulzhenko, A. Morgun, G.F. Rampim, M. Franco, D.R. Almeida, R.V. Diniz, A.C. Carvalho, M. Gerbase-DeLima, Monitoring of intragraft and peripheral blood TIRC7 expression as a diagnostic tool for acute cardiac rejection in humans, Hum. Immunol. 62 (2001) 342–347.
[12] R.F. Chuaqui, R.F. Bonner, C.J. Best, J.W. Gillespie, M.J. Flaig, S.M. Hewitt, J.L. Phillips, D.B. Krizman, M.A. Tangrea, M. Ahram, W.M. Linehan, V. Knezevic, M.R. Emmert-Buck, Post-analysis

follow-up and validation of microarray experiments, Nat. Genet. 32 (2002) 509–514.
[13] C.L. Andersen, J.L. Jensen, T.F. Orntoft, 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. 64 (2004) 5245–5250.
[14] P. Jin, Y. Zhao, Y. Ngalame, M.C. Panelli, D. Nagorsen, V. Monsurro, K. Smith, N. Hu, H. Su, P.R. Taylor, F.M. Marincola, E. Wang, Selection and validation of endogenous reference genes using a high throughput approach, BMC Genomics 5 (2004) 55.
[15] P.D. Lee, R. Sladek, C.M. Greenwood, T.J. Hudson, Control genes and variability: absence of ubiquitous reference transcripts in diverse mammalian expression studies, Genome Res. 12 (2002) 292–297.
[16] J.A. Warrington, A. Nair, M. Mahadevappa, M. Tsyganskaya, Comparison of human adult and fetal expression and identification of 535 housekeeping/maintenance genes, Physiol. Genomics 2 (2000) 143–147.