Research Communication

High Expression Stability of Microtubule Affinity Regulating Kinase 3 (MARK3) Makes it a Reliable Reference Gene

Ching-Yee Chia[1], Claudia WX Lim[1], Wan-Ting Leong[1] and Maurice HT Ling[2,3]

[1]*Nanyang Girls' High School, Ministry of Education, Singapore*
[2]*Department of Zoology, The University of Melbourne, Parkville, Victoria, Australia*
[3]*School of Chemical and Life Sciences, Singapore Polytechnic, Singapore*

Summary

Difference in gene expressions is characteristic of the function of different cell types and those genes with low expression variance can be used as standards for quantitative gene expression studies. Microarray technology is used to study global gene expression within a cell; hence, represents a suitable source of data to mine for genes with low expression variance. The coefficient of variation (COV) of each gene was determined and a threshold of less than 0.1 COV was used to select stably expressed genes in each data set. Our results showed that microtubule affinity-regulating kinase 3 (MARK3) has the lowest COV in eight microarray datasets. In addition, the gene expression of housekeeping genes, which is very likely to be stably expressed, tends to fluctuate highly under different conditions, marking them as being less reliable for use as reference genes. © 2010 IUBMB

IUBMB *Life*, 62(3): 200–203, 2010

INTRODUCTION

In the recent years, experiments related to gene studies (*1–6*) often require the usage of reference genes–genes that are constitutively and constantly expressed in different environmental conditions (*7–9*). Housekeeping genes, which are essential for the maintenance of the cell, are generally assumed to be stably expressed (*10*). However, a number of studies (*8, 11–13*) have shown that mRNA levels of housekeeping genes like *beta-actin* and *GAPDH* can vary in different conditions (*14*). Thus, housekeeping genes may not be appropriate standards for gene

expression studies. Previous studies had indicated that *RPS4*, *UBQ*, and *eEFIAL* were more stably expressed in the larvae of flatfish (*15*) while *TBP*, *RPL13A*, and *B2M* were more stably expressed in osteoarthritic bones (*2*). This suggests that the expressions of reference genes differ greatly in different organisms. Hence, it is unlikely for a single gene to be stably expressed in every organism.

Microarrays and real-time polymerase chain reaction (PCR) are commonly used methods to measure gene expressions (*3, 7, 16*). Microarrays allow for the comparison of thousands of gene expression simultaneously under the effects of treatments and diseased states (*17*). Hence, microarray data represent a suitable data source for isolating genes (*18*) with potentially low expression variance which can be identified as low coefficient of variation (COV). COV is the ratio of the standard deviation to the arithmetic mean and has been used as an efficient method to study gene expression variations (*19*).

In this study, we determine the different levels of gene expressions in mouse liver to find genes with low COV that can be used as standard genes for future quantitative gene expression studies. Our results suggested that housekeeping genes may not necessarily be constantly expressed.

MATERIALS AND METHODS

Microarray Data

Eight data sets of gene expression in the mouse liver under different conditions were obtained from Gene Expression Omnibus (GEO, www.ncbi.nlm.nih.gov/geo), a public microarray resource. Briefly, these data sets originated from the following studies; LDL receptor deficient mice fed on either a low fat diet or a high fat, western-style diet for 12 weeks with 12,548 genes located (GDS279); mice injected with intraperitoneal cytokine injection examined at 4 h with 12,558 genes located (GDS280); analysis of animals fed a diet containing metformin, glipizide, rosiglitazone, or soy isoflavone extract and compared to hepatic

Table 1

Number of genes with coefficient of variation (COV) under values of 0.05, 0.06, 0.07, 0.08, 0.09, and 0.1 in their respective data sets

Dataset	<0.05	<0.06	<0.07	<0.08	<0.09	<0.1
GDS279	1	10	37	86	169	310
GDS280	302	940	2060	3479	4928	6177
GDS1808	1	8	40	130	283	524
GDS2413	85	97	112	124	136	146
GDS2497	34,075	39,761	42,263	43,445	44,018	44,362
GDS2693	29,494	35,617	39,607	41,889	43,020	43,606
GDS3221	24,142	30,091	34,329	37,392	39,611	41,252

Table 2

Specific genes and their COV values

Genes	COV	Genes	COV
Mark3	<0.08	*Eif2s3x*	<0.1
4833439L19Rik	<0.09	*Epn1*	<0.1
9530068E07Rik	<0.09	*H2-K1*	<0.1
Ddb1	<0.09	*Hdgf*	<0.1
Dnajc8	<0.09	*Mark2*	<0.1
Fbp1	<0.09	*Myo1b*	<0.1
Rpl5	<0.09	*Rpl3*	<0.1
Rpl10	<0.09	*Pcid2*	<0.1
Rpl7	<0.09	*Pex14*	<0.1
Srrm1	<0.09	*Pgm2*	<0.1
Ubl5	<0.09	*Pigs*	<0.1
2410166I05Rik	<0.1	*Prdx1*	<0.1
Ahcyl1	<0.1	*Rps3*	<0.1
Ap3b1	<0.1	*Rps3a*	<0.1
Arpc5	<0.1	*Stard3*	<0.1
Bscl2	<0.1	*Sucla2*	<0.1
Cd151	<0.1	*Trpc4ap*	<0.1
Cox7c	<0.1	*Zdhhc9*	<0.1
Cstf1	<0.1	*Zwint*	<0.1
Cyb5	<0.1		

gene expression profile produced by long-term caloric restriction with 12,593 genes located (GDS1808); analysis of day 13.5 and 15.5 p38 alpha-deficient hematopoietic mice with 45,101 genes located (GDS2693); comparison of embryonic day 13.5 and 15.5; analysis of livers from NMR-1 females fed human and chimpanzee diets for 2 weeks with 45101 genes located (GDS3221); analysis of liver of C57BL/6 males maintained on a high-fat high-calorie diet supplemented with 0.04% resveratrol, a compound in red wine that extends lifespan of diverse species with 27,397 genes located (GDS2413); analysis of livers of 13 weeks following treatment with chemicals that test positive in two-year rodent cancer bioassays with 45,102 genes located (GDS2497).

Selection of Reference Gene Candidates

Arithmetic mean and standard deviation were calculated for each gene in each of the seven data sets and their COV were determined as a quotient of the standard deviation to the arithmetic mean. A threshold of less than 0.1 COV were used to select stably expressed genes in each data set.

RESULTS

The number of genes with COV ranging from <0.05 to <0.1 from each dataset were tabulated in Table 1. With GDS279 having the smallest number of gene probes compared to other data sets, it has limited the possible number of genes with COV <0.1 common in all data sets. Dataset GDS2497 has over 45102 probes, which results in the high number of genes having a COV lower than <0.1.

Despite the fact that there are close to 40,000 genes with COV <0.1 in datasets like GDS2497, GDS2693, and GDS3221, many genes are not expressed consistently under the different conditions, resulting in the small number of 39 genes with the COV of <0.1 present across the seven datasets. Nephroblastoma Overexpressed gene (*Nov*), a gene that is essential for cell growth, is expressed with a low COV of 0.052 under the conditions of being p38alpha-deficient and hematopoietic (GDS2693),

but under the conditions of being on different diets for 12 weeks (GDS279), it is expressed with a high COV of 0.561.

Analysis of the seven data set have identified 39 genes possibly the most stably expressed, with the criteria that the gene's COV is <0.1 across each of the conditions as shown in Table 2. Our results indicated that *Microtubule Affinity-Regulating Kinase 3 (MARK3)* is the most invariantly expressed gene with a COV of <0.08—the lowest COV across all conditions, followed by eight other genes with a COV of <0.09; namely *Damage-specific DNA Binding protein 1(Ddb1)*, *D-fructose-1, 6-bisphosphate 1-phosphohydrolase 1 (Dnajc8)*, *Fbp1*, *Ribosomal Protein L5 (Rpl5)*, *Ribosomal Protein L10 (Rpl10)*, *Ribosomal Protein L17 (Rpl7)*, *Serine/arginine Repetitive Matrix 1(Srrm1)* and *Ubiquitin-Like 5 (Ubl5)*.

The seven common housekeeping genes mentioned in the introduction was shown to have great fluctuations in their gene expression level under different conditions as shown in Table 3.

Glyceraldehyde-3-Phosphate Dehydrogenase (GAPDH) has a range of COV values from 0.005 in GDS2693 to 0.47 in GDS279. *Ribosomal Protein S4 (RPS4)* is found in the mice liver as *Ribosomal Protein S4x (RPS4X)*, the latter having fluctuating COV values ranging from 0.003 in GDS2693 to 0.22 in GDS279. *Ubiquitin 1 (Ubqln 1)*, which has similar functions as Ubiquitin (UBQ) in the larvae of the flatfish, has a range of COV values from 0.007 in GDS2693 to 0.43 in GDS279. For *TATA box binding protein (TBP)*, the COV values range from 0.07 in GDS280 to 0.62 in GDS279. *Ribosomal Protein L13a (RPL13A)* has fluctuating COV values of 0.013 in GDS2693 to

Table 3
Housekeeping genes and their COV values

Gene	GDS 279	GDS 280	GDS 1808	GDS 2413	GDS 2497	GDS 2693	GDS 3221
GAPDH	0.47	0.13	0.16	N.F	N.F	0.005	0.016
RPS4X	0.072	0.15	0.22	N.F	0.011	0.003	0.015
Ubqln1	0.43	0.12	0.19	0.14	0.02	0.007	0.004
TBP	0.62	0.07	0.30	0.12	N.F	N.F	N.F
RPL13A	N.F	N.F	N.F	1.75	0.04	0.013	0.12
B2M	0.13	0.074	0.11	1.09	0.01	0.025	0.012

N.F, not found in dataset.

1.75 in GDS2413. *Beta-2-microglobulin (B2M)* has COV values that range from 0.01 in GDS2497 to 1.09 in GDS2413.

Most of the housekeeping genes are also not present in certain datasets, which can mean either the genes are not present in mouse liver, or it was not found in the datasets. For example, *GAPDH* is not found in datasets GDS2413 and GDS2497. *RPS4X* is absent in GDS2413, *TBP* absent in GDS2497, GDS2693 and GDS3221, *RPL13A* absent in GDS279, GDS280 and GDS1808. Out of the six housekeeping genes mentioned above, only *Ubqln1* and *B2m* is consistently present in all seven datasets.

DISCUSSION

The use of housekeeping genes in experimentations requiring genes is based on the assumption that the expressions of these genes are stable under varying conditions (*19*). In this study, microarray data sets were used to identify genes with low expression variances for use as standards in quantitative gene expression studies.

Genes identified in previous studies (*2, 15*) were found to have either high fluctuation in mouse liver data set or the genes are not even present in several of the data sets as shown in Table 3. On the other hand, analysis of the data sets had identified nine genes with less than 10% COV across each of the given conditions, namely, *Mark3, Ddb1, Dnajc8, Fbp1, Rpl5, Rpl10, Rpl7, Srrm1*, and *Ubl5*, suggesting that these genes may be more suitable than those previously identified (*2, 15*) as standard genes for expression studies. This may also suggest that genes with low expression variance in one organism may not imply similar low variance in gene expression levels in other organisms.

Out of the nine specific genes, three genes are ribosomal proteins: *Rpl5, Rpl10*, and *Rpl7*. These ribosomal protein genes are constitutively expressed in all cell types and are essential for the biogenesis of new ribosome for the synthesis of proteins. The essentiality of ribosomal proteins to the function of a cell suggests that ribosomal proteins are more likely to be stably expressed, due to the fact that any increase or decrease of the gene level will result in an abnormal amount of ribosomal level, which may result in mutation or disabled cells as shown by Thorrez et al. (*20*).

In addition to ribosomal proteins, Srrm1 is also involved in numerous pre-mRNA processing event (*21*). These suggest that genes with their resulting protein products that are involved in translation of mRNA are more likely to be constitutively expressed at a constant level as varying availability of proteins in the translational process may result in variability in translational efficiency (*22*). Increase in translational inefficiency has been implicated in the molecular aging process (*23*), suggesting that variability in translational efficiency is not desirable and should be minimised.

Of these nine genes, at least five are known to be common housekeeping genes: *Mark3, Ddb1, Rpl5, Rpl7, Srrm1*, and *Ubl5*. Hence, although it is shown to be true that some of the housekeeping genes have a higher chance of being stably expressed (*24*), many of the common housekeeping genes like *GAPDH* were proven to have high fluctuations in the gene expression level as shown in Table 3.

It has been shown (*11*) that the gene expressions of housekeeping genes like *b-actin* and *GAPDH* can vary in different conditions; thus, lowering the likelihood that housekeeping genes have a higher chance of being stably expressed.

GAPDH is an enzyme that breaks down glucose for energy and carbon molecules (*25*). This can account for its high COV values in GDS279 and GDS1808, as the mice in the two datasets were given a special diet of high fat or restricted calories, making the expression of GAPDH in the mice liver vary to adapt to the different diets. RPS4X, a disease resistance protein that activates defences in response to bacterial pathogens, has a high COV value for certain of the datasets. GDS280, where the mice are injected with intraperitoneal cytokine injection which can enhance immunoglobulin production, can affect the gene expression level for RPS4X since immunoglobulin detects and fights against virus and bacteria and has almost the same function as PRS4X (*10*). This further suggest that while some housekeeping genes are stably expressed, others may not, and it differs for every organism and different conditions, indicating that housekeeping genes are not always reliable.

With such a high fraction of these nine genes being ribosomal proteins, and genes that are involved with the translation process, there is a very high possibility that ribosomal proteins

will tend to be the most stably expressed, even more so than housekeeping genes. Housekeeping genes may not necessarily be expressed stably in most conditions (*8, 11, 13*), so they might not be most suitable to be used for experimentations related to gene studies (*12*).

REFERENCES

1. Kreike, B., Halfwerk, H., Armstrong, N., Bult, P., Foekens, J.A., Veltkamp, S.C., Nuyten, D.S., Bartelink, H., and van de Vijver, M.J. (2009) Local recurrence after breast-conserving therapy in relation to gene expression patterns in a large series of patients. *Clin. Cancer Res.* **15**, 4181–4190.

2. Maccoux, L.J., Clements, D.N., Salway, F., and Day, P.J. (2007) Identification of new reference genes for the normalisation of canine osteoarthritic joint tissue transcripts from microarray data. *BMC Mol. Biol.* **8**, 8–62.

3. Langnaese, K., John, R., Schweizer, H., Ebmeyer, U., and Keilhoff, G. (2008) Selection of reference genes for quantitative real-time PCR in a rat asphyxial cardiac arrest model. *BMC Mol. Biol.* **9**, 53.

4. Paolacci, A.R., Tanzarella, O.A., Porceddu, E., and Ciaffi, M. (2009) Identification and validation of reference genes for quantitative RT-PCR normalization in wheat. *BMC Mol. Biol.* **10**, 11.

5. Pombo-Suarez, M., Calaza, M., Gomez-Reino, J.J., and Gonzalez, A. (2008) Reference genes for normalization of gene expression studies in human osteoarthritic articular cartilage. *BMC Mol. Biol.* **9**, 17.

6. Rhinn, H., Marchand-Leroux, C., Croci, N., Plotkine, M., Scherman, D., and Escriou, V. (2008) Housekeeping while brain's storming Validation of normalizing factors for gene expression studies in a murine model of traumatic brain injury. *BMC Mol. Biol.* **9**, 62.

7. Kidd, M., Nadler, B., Mane, S., Eick, G., Malfertheiner, M., Champaneria, M., Pfragner, R., and Modlin, I. (2007) GeneChip, geNORM and gastrointestinal tumours: novel reference genes for real-time PCR. *Physiol. Genomics* **30**, 363–370.

8. Kriegova, E., Arakelyan, A., Fillerova, R., Zatloukal, J., Mrazek, F., Navratilova, Z., Kolek, V., du Bois, R.M., and Petrek, M. (2008) PSMB2 and RPL32 are suitable denominators to normalize gene expression profiles in bronchoalveolar cells. *BMC Mol. Biol.* **9**, 69.

9. Tatsumi, K., Ohashi, K., Taminishi, S., Okano, T., Yoshioka, A., and Shima, M. (2008) Reference gene selection for real-time RT-PCR in regenerating mouse livers. *Biochem. Biophys. Res. Commun.* **374**, 106–110.

10. Gubern, C., Hurtado, O., Rodríguez, R., Morales, J.R., Romera, V.G., Moro, M.A., Lizasoain, I., Serena, J., and Mallolas, J. (2009) Validation of housekeeping genes for quantitative real-time PCR in in vivo and in vitro models of cerebral ischaemia. *BMC Mol. Biol.* **10**, 57.

11. Brunner, A.M., Yakovlev, I.A., and Strauss, S.H. (2004) Validating internal controls for quantitative plant gene expression. *BMC Plant Biol.* **4**, 14.

12. Fink, T., Lund, P., Pilgaard, L., Rasmussen J.G., Duroux, M., and Zachar, V. (2008) Instability of standard PCR reference genes in adi-

pose-derived stem cells during propagation, differentiation and hypoxic exposure. *BMC Mol. Biol.* **9**, 98.

13. Takagi, S., Ohashi, K., Utoh, R., Tatsumi, K., Shima, M., and Okano, T. (2008) Suitable reference gene for the analysis of direct hyperplasia in mice. *Biochem. Biophys. Res. Commun.* **377**, 1259–1264.

14. Strube, C., Buschbaum, S., Wolken, S., and Schnieder, T. (2008) Evaluation of reference genes for quantitative real-time PCR to investigate protein disulfide isomerase transcription pattern in the bovine lungworm *Dictyocaulus viviparus*. *Gene* **425**, 36–43.

15. Infante, C., Matsuoka, M.P., Asensio, E., Cañavate, J.P., Reith, M., and Manchado, M. (2008) Selection of housekeeping genes for gene expression studies in larvae from flatfish using real-time PCR. *BMC Mol. Biol.* **9**, 28.

16. Caelers, A., Berishvili, G., Meli, M.L., Eppler, E., and Reinecke, M. (2004) Establishment of a real-time RT-PCR for the determination of absolute amounts of IGF-I and IGF-II gene expression in liver and extrahepatic sites of tilapia. *Gen. Comp. Endocrinol.* **137**, 196–204.

17. De, R.K. and Ghosh, A. (2009) Interval based fuzzy systems for identification of important genes from microarray gene expression data: application to carcinogenic development. *J. Biomed. Inform.* **42**, 1022–1028.

18. Frericks, M. and Esser, C. (2008) A toolbox of novel murine housekeeping genes identified by meta-analysis of large scale gene expression profiles. *Biochimica et Biophysica Acta* **1779**, 830–837.

19. Gjuvsland, A.B., Plahte, E., and Omholt, S.W. (2007) Threshold-dominated regulation hides genetic variation in gene expression networks. *BMC Syst. Biol.* **1**, 57.

20. Thorrez, L., Van Deun, K., Tranchevent, L., Van Lommel, L., Engelen, K., Marchal, K., Moreau, Y., Van Mechelen, I., and Schuit, F. (2008) Using ribosomal protein genes as reference: a tale of caution. *PLoS ONE* **3**, e1854.

21. Le Hir, H., Maquat, L.E., and Moore, M.J. (2000) Pre-mRNA splicing alters mRNP composition: evidence for stable association of proteins at exon-exon junctions. *Genes Dev.* **14**, 1098–1108.

22. Meng, Z., Jackson, N.L., Choi, H., King, P.H., Emanuel, P.D., and Blume, S.W. (2008) Alterations in RNA-binding activities of IRES-regulatory proteins as a mechanism for physiological variability and pathological dysregulation of IGF-IR translational control in human breast tumor cells. *J. Cell Physiol.* **217**, 172–183.

23. Balajee, A.S., Machwe, A., May, A., Gray, M.D., Oshima, J., Martin, G.M., Nehlin, J.O., Brosh, R., Orren, D.K., and Bohr, V.A. (1999) The Werner syndrome protein is involved in RNA polymerase II transcription. *Mol. Biol. Cell* **10**, 2655–2668.

24. Coulson, D.T.R., Brockbank, S., Quinn, J.G., Murphy, S., Ravid, R., Irvine, G.B., and Johnston, J.A. (2008) Identification of valid reference genes for the normalization of RT qPCR gene expression data in human brain tissue. *BMC Mol. Biol.* **9**, 46.

25. Tisdale, E.J., Kelly, C., and Artalejo, C.R. (2004) Glyceraldehyde-3-phosphate dehydrogenase interacts with Rab2 and plays an essential role in endoplasmic reticulum to Golgi transport exclusive of its glycolytic activity. *J. Biol. Chem.* **279**, 54046–54052.