Chapter 6

IDENTIFICATION OF REFERENCE GENES BY META-MICROARRAY ANALYSES

Issac HK Too[1], Sean SJ Heng[2], Oliver YW Chan[2], Bryan MH Keng[2], Ching-Yee Chia[3], Claudia WX Lim[3], Wan-Ting Leong[3], Qinghao Chu[2], Ernest JG Ang[2], Yongjie Lin[2] and Maurice HT Ling[4,5]*

[1]Department of Biological Sciences, National University of Singapore, Singapore
[2]Raffles Institution, Singapore
[3]Nanyang Girls' High School, Singapore
[4]Department of Zoology, The University of Melbourne, Australia
[5]School of Chemical and Biomedical Engineering,
Nanyang Technological University, Singapore

Citation: Too, IHK, Heng, SSJ, Chan, OYW, Keng, BMH, Chia, CY, Lim, CWX, Leong, WT, Chu, QH, Ang, EJG, Lin, YJ, Ling, MHT. 2014. Identification of Reference Genes by Meta-Microarray Analyses. James V. Rogers (ed), Microarrays: Principles, Applications and Technologies. Nova Science Publishers, Inc. Pages 81-99.

ABSTRACT

The expression levels of reference genes used in gene expression studies are assumed to not change under most circumstances. However, a number of studies have demonstrated that genes theoretically assumed to be stably expressed were found to vary under experimental conditions. In addition, previous studies have also reported that stably expressed genes in an organ, may not be stably expressed in other organs or in a different organism, suggesting the need to identify reference genes for each organ and each organism. Due to its ability to analyze the expression of thousands of genes in an experiment, microarrays present a suitable resource for the analysis and identification of reference genes. We present four cases on practical applications of microarrays whereby multiple published microarray data sets were examined to identify suitable reference genes using coefficient of variation (CV) and NormFinder. Our results suggest that microtubule affinity-regulating kinase 3 (MARK3) is a suitable reference gene for mouse liver, 40S ribosomal protein S29 (Rps29) is a suitable reference gene for mouse testes and pancreas, signal peptidase complex subunit 1 (SPCS1) and hydroxyacyl-CoA dehydrogenase beta subunit (HADHB) are suitable reference genes for human lungs, and glucan biosynthesis protein G (mdoG) is a suitable reference gene for *Escherichia coli*. Further analysis suggests that the identified reference genes are involved in fundamental biochemical processes. This supports the theoretical basis and previous studies that housekeeping genes, on the whole, are generally stably expressed. However, our results also suggest that certain housekeeping genes that are stably expressed in one tissue or

* Corresponding author: mauriceling@acm.org.

one organism may not be stably expressed in different tissues or organisms, supporting the need to identify reference genes for each tissue and organism.

ACKNOWLEDGMENTS

The authors will like to thank the following colleagues for their comments on this manuscript: JSH Oon (University of Queensland, Australia), QMA Xie (University of Queensland, Australia), PCK Au (CSIRO, Australia), CC Goh (University of Virginia, USA), and YZ Koh (University of Portsmouth, UK). SSH, OYC, BMK, CC, CWL, WL, QC, EJA, and YL contributed to this work as part of the Science Mentorship Programme under Gifted Education Branch, Ministry of Education, Singapore.

REFERENCES

[1] Chey, S; Claus, C; Liebert, UG. Validation and application of normalization factors for gene expression studies in rubella virus-infected cell lines with quantitative real-time PCR. *J Cell Biochem*, 2010, 110(1), 118-128.

[2] Bustin, SA; Nolan, T. Pitfalls of quantitative real-time reverse-transcription polymerase chain reaction. *J Biomolecular Tech*, 2004, 15(3), 155-166.

[3] Maccoux, LJ; Clements, DN; Salway, F; Day, PJ. Identification of new reference genes for the normalisation of canine osteoarthritic joint tissue transcripts from microarray data. *BMC Mol Biol*, 2007, 8, 62.

[4] Airoldi, EM; Huttenhower, C; Gresham, D; Lu, C; Caudy, AA; Dunham, MJ, et al. Predicting cellular growth from gene expression signatures. *PLoS Comput Biol*, 2009, 5(1), e1000257.

[5] Gubern, C; Hurtado, O; Rodriguez, R; Morales, JR; Romera, VG; Moro, MA; et al. Validation of housekeeping genes for quantitative real-time PCR in in-vivo and in-vitro models of cerebral ischaemia. *BMC Mol Biol*, 2009, 10:57.

[6] Czechowski, T; Stitt, M; Altmann, T; Udvardi, MK; Scheible, WR. Genome-wide identification and testing of superior reference genes for transcript normalization in Arabidopsis. *Plant Physiol*, 2005, 139(1), 5-17.

[7] Dundas, JB; Ling, MHT. Reference genes for measuring mRNA expression. *Theory in Biosci*, 2012, 131, 215-223.

[8] Jain, M; Nijhawan, A; Tyagi, AK. Validation of housekeeping genes as internal control for studying gene expression in rice by quantitative real-time PCR. *Biochem Biophys Res Commun*, 2006, 345, 646-651.

[9] Nicot, N; Hausman, JF; Hoffmann, L. Housekeeping gene selection for real-time RT-PCR normalization in potato during biotic and abiotic stress. *J Exp Botany*, 2005, 56, 2907-2914.

[10] Gibson, UE; Heid, CA; Williams; PM. A novel method for real time quantitative RT-PCR. *Genome Res*, 1996, 6, 995-1001.

[11] Sturzenbaum, SR; Kille, P. Control genes in quantitative molecular biological techniques: the variability of invariance. *Comp Biochem Physiol B Biochem Mol Biol* 2001, 130, 281-289.

[12] Takle, GW; Toth, IK; Brurberg, MB. Evaluation of reference genes for real-time RT-PCR expression studies in the plant pathogen Pectobacterium atrosepticum. *BMC Plant Biol*, 2007, 7, 50.

[13] Noriega, NC; Kohama, SG; Urbanski, HF. Microarray analysis of relative gene expression stability for selection of internal reference genes in the rhesus macaque brain. *BMC Mol Biol*, 2010, 11, 47.

[14] Remans, T; Smeets, K; Opdenakker, K; Mathijsen, D; Vangronsveld, J; Cuypers, A. Normalisation of real-time RT-PCR gene expression measurements in Arabidopsis thaliana exposed to increased metal concentrations. *Planta*, 2008, 227, 1343-1349.

[15] Glare, EM; Divjak, M; Bailey, MJ; Walters, EH. beta-Actin and GAPDH housekeeping gene expression in asthmatic airways is variable and not suitable for normalizing mRNA levels. *Thorax*, 2002, 57(9), 765-770.

[16] Gutierrez, L; Mauriat, M; Guénin, S; Pelloux, J; Lefebvre, JF; Louvet, R; et al. The lack of a systematic validation of reference genes: a serious pitfall undervalued in reverse transcription polymerase chain reaction (RT-PCR) analysis in plants. *Plant Biotechnol*, 2008, 6, 609-618.

[17] Gur-Dedeoglu, B; Konu, O; Bozkurt, B; Ergul, G; Seckin, S; Yulug, IG. Identification of endogenous reference genes for qRT-PCR analysis in normal matched breast tumor tissues. *Oncol Res*, 2009, 17(8), 353-365.

[18] Kosir, R; Acimovic, J; Golicnik, M; Perse, M; Majdic, G; Fink, M; et al. Determination of reference genes for circadian studies in different tissues and mouse strains. *BMC Mol Biol,* 2010, 11, 60.

[19] Shen, Y; Li, Y; Ye, F; Wang, F; Lu, W; Xie, X. Identification of suitable reference genes for measurement of gene expression in human cervical tissues. *Anal Biochem*, 2010, 405(2), 224-229.

[20] Rho, HW; Lee, BC; Choi, ES; Choi, IJ; Lee, YS; Goh, SH. Identification of valid reference genes for gene expression studies of human stomach cancer by reverse transcription-qPCR. *BMC Cancer*, 2010, 10, 240.

[21] Brattelid, T; Winer, LH; Levy, FO; Liestol, K; Sejersted, OM; Andersson, KB. Reference gene alternatives to Gapdh in rodent and human heart failure gene expression studies. *BMC Mol Biol*, 2010, 11, 22.

[22] Lallemant, B; Evrard, A; Combescure, C; Chapuis, H; Chambon, G; Raynal, C; et al. Reference gene selection for head and neck squamous cell carcinoma gene expression studies. *BMC Mol Biol*, 2009, 10, 78.

[23] Li, YL; Ye, F; Hu, Y; Lu, WG; Xie, X. Identification of suitable reference genes for gene expression studies of human serous ovarian cancer by real-time polymerase chain reaction. *Anal Biochem*, 2009, 394(1), 110-116.

[24] Infante, C; Matsuoka, MP; Asensio, E; Canavate, JP; Reith, M; Manchado, M. Selection of housekeeping genes for gene expression studies in larvae from flatfish using real-time PCR. *BMC Mol Biol*, 2008, 9, 28.

[25] Vandesompele, J; De Preter, K; Pattyn, F; Poppe, B; Van Roy, N; De Paepe, A; et al. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol*, 2002, 3(7), RESEARCH0034.

[26] Pfaffl, MW; Tichopad, A; Prgomet, C; Neuvians, TP. Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper--Excel-based tool using pair-wise correlations. *Biotechnol Lett*, 2004, 26(6), 509-515.

[27] Andersen, CL; Jensen, JL; Orntoft, TF. Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. *Cancer Res*, 2004, 64(15), 5245-5250.

[28] He, JQ; Sandford, AJ; Wang, IM; Stepaniants, S; Knight, DA; Kicic, A; et al. Selection of housekeeping genes for real-time PCR in atopic human bronchial epithelial cells. *Eur Respir J*, 2008, 32(3), 755-762.

[29] Ren, S; Zhang, F; Li, C; Jia, C; Li, S; Xi, H; et al. Selection of housekeeping genes for use in quantitative reverse transcription PCR assays on the murine cornea. *Mol Vis*, 2010, 16, 1076-1086.

[30] De, RK; Ghosh, A. Interval based fuzzy systems for identification of important genes from microarray gene expression data: Application to carcinogenic development. *J Biomed Inform*, 2009, 42(6),1022-1028.

[31] Galiveti, CR; Rozhdestvensky, TS; Brosius, J; Lehrach, H; Konthur, Z. Application of housekeeping npcRNAs for quantitative expression analysis of human transcriptome by real-time PCR. *RNA*, 2010, 16(2), 450-461.

[32] Chia, CY; Lim, CW; Leong, WT; Ling, MH. High expression stability of microtubule affinity regulating kinase 3 (MARK3) makes it a reliable reference gene. *IUBMB Life*, 2010, 62(3), 200-203.

[33] Thorrez, L; Van Deun, K; Tranchevent, LC; Van Lommel, L; Engelen, K; Marchal, K; et al. Using ribosomal protein genes as reference: a tale of caution. *PLoS One*, 2008, 3(3), e1854.

[34] Le Hir, H; Maquat, LE; Moore, MJ. Pre-mRNA splicing alters mRNP composition: evidence for stable association of proteins at exon-exon junctions. *Genes Dev*, 2000, 14, 1098-1108.

[35] Meng, Z; Jackson, NL; Choi, H; King, PH; Emanuel, PD; Blume, SW. 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*, 2008, 217, 172-183.

[36] Coulson, DTR; Brockbank, S; Quinn, JG; Murphy, S; Ravid, R; Irvine, GB; et al. Identification of valid reference genes for the normalization of RT qPCR gene expression data in human brain tissue. *BMC Mol Biol*, 2008, 9, 46.

[37] Chu, QH; Lin, YJ; Ang, EJG; Ling, MHT. Identification of transcriptional invariant genes in mouse endocrine glands from microarray data. *Proceedings of the 16th Youth Science Conference*, 2010, Singapore.

[38] Too, IH; Ling, MH. Signal peptidase complex subunit 1 (SPCS1) and hydroxyacyl-CoA dehydrogenase beta subunit (HADHB) are suitable reference genes in human lungs. *ISRN Bioinformatics*, 2012, 2012, Article ID 790452.

[39] Cheadle, C; Vawter, MP; Freed, WJ; Becker, KG. Analysis of microarray data using Z score transformation. *J Mol Diagn*, 2003, 5(2), 73-81.

[40] Nguewa, PA; Agorreta, J; Blanco, D; Lozano, MD; Gomez-Roman, J; Sanchez, BA; et al. Identification of importin 8 (IPO8) as the most accurate reference gene for the clinicopathological analysis of lung specimens. *BMC Mol Biol*, 2008, 9, 103.

[41] Liu, DW; Chen, ST; Liu, HP. Choice of endogenous control for gene expression in nonsmall cell lung cancer. *Eur Respir J*, 2005, 26, 1002-1008.

[42] Saviozzi, S; Cordero, F; Lo Iacono, M; Novello, S; Scagliotti, GV; Calogero, RA. Selection of suitable reference genes for accurate normalization of gene expression profile studies in non-small cell lung cancer. *BMC Cancer*, 2006, 6, 200.

[43] Kalies, KU; Hartmann, E. Membrane topology of the 12- and the 25-kDa subunits of the mammalian signal peptidase complex. *J Biol Chem*, 1996, 271(7), 3925-3929.

[44] Swanton, E; Bulleid, NJ. Protein folding and translocation across the endoplasmic reticulum membrane. *Mol Membr Biol*, 2003, 20(2), 99-104.

[45] Aoyama, T; Wakui, K; Orii, KE; Hashimoto, T; Fukushim, Y. Fluorescence in situ hybridization mapping of the alpha and beta subunits (HADHA and HADHB) of human mitochondrial fatty acid beta-oxidation multienzyme complex to 2p23 and their evolution. *Cytogenet Cell Genet*, 1997, 79(3-4), 221-224.

[46] Fu, Z; Attar-Bashi, NM; Sinclair, AJ. 1-14C-linoleic acid distribution in various tissue lipids of guinea pigs following an oral dose. *Lipids*, 2001, 36(3), 255-260.

[47] Heng, SSJ; Chan, OYW; Keng, BMH; Ling, MHT. Glucan biosynthesis protein G (mdoG) is a suitable reference gene in Escherichia coli K-12. *ISRN Microbiol*, 2011, 2011, Article ID 469053.

[48] Loubens, I; Debarbieux, L; Bohin, A; Lacroix, JM; Bohin, JP. Homology between a genetic locus (mdoA) involved in the osmoregulated biosynthesis of periplasmic glucans in Escherichia coli and a genetic locus (hrpM) controlling pathogenicity of Pseudomonas syringae. *Mol Microbiol*, 1993, 10(2), 329-340.

[49] Bohin, JP. Osmoregulated periplasmic glucans in Proteobacteria. *FEMS Microbiol Lett*, 2000, 186(1), 11-19.

[50] Page, F; Altabe, S; Hugouvieux-Cotte-Pattat, N; Lacriox, JM; Robert-Baudouy, J; Bohin, JP. Osmoregulated periplasmic glucan synthesis is required for Erwinia chrysanthemi pathogenicity. *J Bacteriol*, 2001, 183(10), 3134-3141.

[51] Meldgaard, M; Fenger, C; Lambertsen, KL; Pedersen, MD; Ladeby, R; Finsen, B. Validation of two reference genes for mRNA level studies of murine disease models in neurobiology. *J Neurosci Methods*, 2006, 156(1-2), 101-110.

[52] Keng, BMH; Chan, OYW; Hen, SSJ; Ling, MHT. Transcriptome analysis of Spermophilus lateralis and Spermophilus tridecemlineatus liver does not suggest the presence of Spermophilus-liver-specific reference genes. *ISRN Bioinformatics*, 2013, 2013, Article ID 361321.

[53] Sauer, U; Preininger, C; Hany-Schmatzberger, R. Quick and simple: quality control of microarray data. *Bioinformatics*, 2005, 21, 1572-1578.

[54] Silberberg, G; Baruch, K; Navon, R. Detection of stable reference genes for real-time PCR analysis in schizophrenia and bipolar disorder. *Anal Biochem*, 2009, 391, 91-97.

[55] Ling, MHT; Ban, Y; Wen, H; Wang, SM; Ge, SX. Conserved expression of natural antisense transcripts in mammals. *BMC Genomics*, 2013, 14(1), 243.