Reference genes for measuring mRNA expression

Jitesh Dundas[1] & Maurice Ling[2]

[1]Edencore Technologies, Virar (w), Thane-401303

[2]Department of Zoology, the University of Melbourne, Australia

Abstract: The aim of this review is to find answers to some of the questions surrounding reference genes and their reliability for quantitative experiments. Reference genes are assumed to be at a constant expression level, over a range of conditions such as temperature. These genes, such as GADPH and beta-actin, are used extensively for gene expression studies using techniques like quantitative PCR. There have been several studies carried out on identifying reference genes. However, a lot of evidence indicates issues to the general suitability of these genes. Recent studies had shown that different factors, including the environment and methods, play an important role in changing the expression levels of the reference genes. Thus, we conclude that there is no reference gene that can deemed suitable for all the experimental conditions. In addition, we believe that every experiment will require the scientific evaluation and selection of the best candidate gene for use as a reference gene in order to obtain reliable scientific results.

Keywords: reliability, beta-actin, reference genes, polymerase chain reactions

Introduction

Reference genes are assumed to be stably expressed and are used for many experiments [1]. Some of the examples of reference genes are *MARK3* [39], *B2M* [40], *GAPDH* [41] and *Beta-actin* [42]. These genes are known as reference genes due to their stable gene expression levels [69] and their role in aiding important regulatory functions such as protein folding and ribosome synthesis. These genes have been widely used for investigations into the patterns of diseases [70] such as cancer [26]. Such studies can help us in finding much needed information about these diseases as well as in designing better treatment strategies. Reference genes are used extensively in experimental work as they are assumed to have a minimal variation in gene expression [17], allowing reliable results from them.

Currently, there have been studies that question the reliability of these genes as reference genes for experiments [68]. Though these genes are known to present in all cells, their gene expression levels may vary depending on the cellular functions and the corresponding environmental conditions [72]. It is a well accepted fact that these reference genes are not very resistant to the

changes in their environment. They are susceptible, sometimes very easily, to any variations in temperature, stress and other attributes that may be affecting the environment [63]. Thus, the use and reliability of the experimental results involving these reference genes in the past must be validated with considerable interest. There is a need to validate these genes for the specific experiments as the behaviour of each gene may change based on the experiments, environment conditions and human accuracy [32]. Some methods, such as Bestkeeper [37], geNorm [33] and NormFinder [35], have been developed for doing the verification and selection of the reference genes. We argue that these past experiments, which had not performed the process of selecting appropriate reference genes, need to be validated as they might give different behaviour due to a change in the environmental factors such as temperature.

Contradictions in using reference genes

Recent studies have shown that the reference genes tend to fluctuate in different conditions and may give unexpected results in the involved experiments. Some of these reference genes tend to decrease in stability in the corneal neovascularization (CorNV) condition [2], with *GAPDH* [41] and *ACTB* (beta-actin) [42] being the least stably expressed. Clearly, the results from the past research work, which were using these reference genes without validation, may not be reliable due to such unexpected behaviour.

In a neuropathological study to investigate the relationship between brain-weight and mRNAs, it was found that the latter affected the expression of the genes, *Beta2M* [40] and *TATA*-binding protein (TBP) [3]. Therefore, it would be a risk to believe that these results, in the absence of reliable test cases for the validation of the known fluctuations in reference genes, did not have any effect on the involved experiments indirectly or directly. The past studies which were based on PCR (polymerase chain reaction) experiments had assumed that the reference genes were stable and independent of any experimental conditions. In a separate study [4], the reference genes were validated using qRT-PCR (quantitative real-time PCR) and the reference genes *GADPH* [41], *ALG9* [75] and *RPL13A* [73, 74]. The stability of their gene expression was tested using geNorm [33], NormFinder [35] and BestKeeper [37]. The use of quantile-normalized microarray gene expression values has shown an improvement in the selection of genes as reference genes [4].

In a study of *Daphnia pulex* specimens (that were induced with *Chaoborus midge larva*), six candidate reference genes were tested for normalization using qRT-PCR (quantitative real-time PCR) [5] and later studied. Of these, Xbp1, Tbp, CAPON and Stx16 were found to be suitably expressed. This study also stated that there was no need for verifying the expression level of reference genes. However, a study on mouse liver [39] found that Tbp was not stably expressed.

Quantitative real-time PCR (qRT-PCR) depends on the selection of appropriate and stable reference genes [6] for gene expression analysis. Thus, geNorm [33] and NormFinder [35]

algorithms were used to identify different sets of reference genes for expression data normalization in the roots, flowers, stems and leaves of flax. Differences [6] in the identification of reference genes were found due to the use of two separate algorithms, albeit without any effect on the needed analysis, in the expression of reference genes for flax (*Linum usitatissimum L*) in their various organs. This study [6] further identified *GADPH* as the most stable reference gene for such studies. Using real time PCR (polymerase chain reaction) experiments for a rat model based study [7] of three genes namely *Beta–Actin* (*ACTB*), glyceraldehyde-3-phosphate dehydrogenase (*GADPH*) and hydroxymethylbilane synthase (*HMBS*), it was found that *ACTB* and not *GAPDH* were stably expressed.

In another study [8], several genes of *Arabidopsis thaliana* were found to have more stable expression levels than traditional reference genes, when subjected to different environmental conditions. It is clear that methods such as RNA gel-blot analysis or quantitative reverse transcription-polymerase chain reaction (qRT - PCR) are dependent on stable reference genes. Thus, we believe that there is a different suitable combination of reference genes for each experimental study [8], which needs to be validated and selected, depending on the specimens and the related environmental conditions.

In a separate study [9] for observing the stability of nine reference genes in wound healing using *reverse transcription-quantitative polymerase chain reaction (RT-qPCR*) analysis, it was found that the expression levels of these reference genes were varying depending on the controlled experimental conditions. Another study found *GADPH* to be a reliable reference gene for quantitative gene expression analysis in human diploid fibroblasts (HDFs) [10]. Unicellular photosynthetic algae (dinoflagellates) from the genus *Symbiodinium* live symbiotically with coral reefs. In another study [11], cultured *Symbiodinium* species (*clade C virus*) was studied under different environmental conditions and nine housekeeping genes (HKGs) were analysed. Using five stable HKGs as reference genes for the experiment, the Hsp90 expression levels were studied for *Symbiodinium* in culture and in symbiotic relationship with coral host (*Acropora millepora*) at varying temperature levels. The results showed varying levels for Hsp90 expression levels and presented a first list of HKGs for symbiotic *dinoflagellates*. The validation of reference genes in *Symbiodinium,* in the presence of thermal and light stress, revealed a drop in the *Hsp90* expression of the reference genes [11].

In psychiatric gene expression studies [11] involving human brain tissue, the reducing intersubject variability and not experimental error was the main advantage of normalization. There has been evidence against the classification of reference genes [12] (in biological psychiatry) under the simple methods of classification. This meant that there was a need to look at a deeper and extensive approach, which included consideration of issues related to human brain studies. Using PCR (*polymerase chain reaction*) based techniques and normalization of reference genes (with geNorm [33] and NormFinder [35]), a study was undertaken to identify genes for flat oyster *Ostrea edulis*. The findings suggested using *GAPDH* [41] and *EF1*-α [35] combined as reference genes when studying expression levels in haemocytes of *Oestrea edulis*

[35]. However, *beta-actin* was found to be the least stable [13] in the study of expression levels in this case. In a separate study [68], an experiment was performed to evaluate the effect of hypoxia on expression levels of housekeeping genes. Here, the expression levels of ten reference genes were recorded for four cell lines treated with four oxygen concentrations (from 1 to 20%). The experiment revealed large variations in their gene expressions in hypoxic conditions based on the choice of cell line, oxygen concentration and the methods for statistical mean analysis. In another study [73] involving the meta-analysis of 13,629 human gene array samples, the gene expression levels [73] of the sample reference genes such as *ACTB* [42], *GAPDH* [41], *HPRT1* [76] and *B2M* [40] were used [73]. The experiments recorded values of their CV (coefficient of variation), SD (standard deviation) and MFC (the proportion of the maximum and minimum values observed within the sample set) [73] .It was found that the expression levels of these genes fluctuated dramatically [73]. The MFC value was found to range from 1.91 for *ACTB* to 15.15 for *ALDOA* [73]. Among all the genes, the CV value was found to be less than the 5% level only for *ACTB* [73], indicating the high levels of variations in the genes [73]. It is interesting to note that these genes are considered as reference genes and are assumed to be very stable for a wide range of conditions [73]. Moreover, there was no gene identified from these commonly used reference genes [73], as a deserving candidate reference gene in the top 50 recorded candidate reference [73] genes of the dataset of this experiment.

The above literature clearly shows the contradictions in the findings on reference genes, implying the need for more thorough investigation to get more reliable and accurate results [77, 78]. There is a clear dependence of the reference genes on the environmental and other aspects of the scientific study, which make their classification methodology even more specific and complex.

Current methods for identification of reference genes

There are several methods in the identification of reference genes. Internal control genes (or reference genes) can be validated using regression analysis [14].

Listed below are some of the methods used for identifying reference genes:

1) **Real Time Polymerase Chain Reaction (or RT-PCR)** has been used extensively for many years [55, 56, 57, 58, 59] for the identification of suitable reference genes. In this technique, which is a variant of polymerase chain reaction (PCR), RNA sequence is reverse transcribed into complementary DNA with the help of the reverse transcriptase enzyme. The cDNA is then amplified using traditional or real-time PCR. In a recent study [15], *RPS4, UBQ*, and *eEF1A1* genes were found to be useful in larvae studies using quantitative PCR (polymerase chain reaction). In the past, this technique has been used without any normalization of the genes, which could indicate an increased risk of

inaccurate experimental results. Thus, there is need to focus on normalisation strategies and validate the reference genes which are used in experiments [16].

2) SAGE (Sequential Analysis of Gene Expression) is a genetic sequence profiling technique[66] used for obtaining high quality, accurate and quantitative analysis of gene sequences in the given datasets. One of the main methods requires data-mining the microarray datasets for highly expressed and relatively constant transcripts [17] as done in a study involving **SAGE** [61]. SAGE is a technique used for getting a complete analysis of the gene expression patterns in the given gene datasets. This is done by isolating unique sequence tags from given mRNA sequences and then concatenating these sequences into long DNA sequences [62].This technique has the following steps [61, 62]:-

2.1. The mRNA of the observation sample is first extracted and isolated.
2.2. For each mRNA sequence, a section is removed at a pre-specified position.
2.3. All these obtained sections of mRNA sequences are then put together to form a lengthy chain like sequence.
2.4. These chains like sequences are then cloned into a vector. This is done so that the microorganisms (such as bacteria) can consume them.
2.5. These chains are then sequenced using automatic DNA sequencing technologies.
2.6. Next, the number of sequence tags are then counted using computer algorithms and software support.

In another study [17], SAGE (Serial expression of gene expression) based records were analysed and then the specificity of reference genes was evaluated using quantitative PCR (polymerase chain reaction) and the results analysed. This study suggested the usefulness of SAGE for the normalization of housekeeping genes.

3) **Microarray** data analysis is also another technique which is used to uniquely identify reference genes [39, 77, 78]. In a separate study [18], normalisation of gene expression levels was used for microarray data analysis of the canine osteoarthritic joint tissue, and then the reference genes were identified using three different algorithms. New reference genes [18] were identified using the traditionally normalised microarray data and were found to be more stable than those found using reverse transcription-quantitative polymerase chain reaction (RT-qPCR). This method proved to be very effective as atleast one gene *MRPS7,* was found to be most stable across multiple datasets. A similar method was used to identify reference genes for breast cancer studies [19] in which 3 microarray datasets of breast tumours were normalised and matched with their normal counterparts. *ACTP* (beta actin) and *SDPH* (succinate dehydrogenase) were found to be the most stable reference genes in this study involving quantitative real time polymerase chain reaction (qPCR).

4) Statistical algorithm based software such as geNorm [33] and NormFinder [35] are also used to normalize reference genes. In a separate study, while analysing the expression levels in haemocytes of *Oestrea edulis*, the reference genes for the same were identified by analysing the RNA sequences [20], using polymerase chain reaction (PCR), in pools of haemocytes that were in touch with the parasite and the haemocyte alone. The above mentioned software was used to normalise the gene expression levels in the study. This is just one of the several experiments in which the above mentioned algorithms have played an important role for improving the appropriateness of the housekeeping genes

5) In the study for ovarian tissues, twenty reference genes were identified from 52 samples [21] involving non-malignant and malignant carcinogenic cells. **One way analysis of variance (ANOVA) method** was used to study the gene changes. Next, geNorm [33] and NormFinder [35] were used for further validation.

A comparison of the existing methods is essential for a making a better choice from them for different experiment scenarios. Real-time polymerase chain reaction (RT-PCR) can quantify for much smaller samples of mRNA (even from a single cell); thus, assures higher accuracy. This quantitative method can automate processes such as finding rare targets as well as abundant targets by measuring the reaction products in each cycle for every sample. This is because during the exponential phase of amplification (this is the phase in which the samples are analysed and quantitative data is collected from the given sample).

There are some drawbacks of using real-time polymerase chain reaction (RT-PCR) when compared to other methods. It is time-consuming [64, 65] and the results are not always reliable as ethidium bromide (which is used as a staining material for detection of PCR products) has low sensitivity. Real-time polymerase chain reaction (RT-PCR) has an increased cross contamination risk of the samples under consideration here. At the same time, there had been reported instances of contamination of cDNA transcripts with genomic pseudogenes [79] or degradation of RNA [80], which may arise from RNA preparation [81]. This will significantly increase the detection error as the detection of the polymerase chain reaction (PCR) products requires the post-amplification processing of the samples. However, there had been a sample processing steps to prevent genomic pseudogene contamination [82, 83] but these will increase the number of steps in sample preparation and validation using mitochondrial DNA [84]. Moreover, the specificity of the assay is determined by the primers and which can give false-positive results. Another issue is that it is semi or even low quantitative technique, whereas the amplicon (it is a piece of DNA which is formed as the product of amplification events) can be can be visualized only after the amplification ends. Real-Time polymerase chain reaction (RT-PCR) is the most sensitive method and can discriminate closely related mRNAs. It is technically simple but it is difficult to get truly quantitative results using conventional polymerase chain reaction (PCR). Northern blotting and Ribonuclease Protection Assays (RPAs) require no amplification while in situ hybridization is qualitative in nature. Northern blotting and Ribonuclease Protection Assays (RPAs) provide good results, but consume more RNA than their counterparts. Polymerase chain reaction (PCR)

considers amplification of the DNA; thus, more sensitive. Real-time PCR gives more quantitative results and is also easier and convenient to use [65, 71].

Sequential Analysis of Gene Expression (SAGE) and Microarrays are different due to the need for prior knowledge (in the latter) of the gene sequences of the samples under consideration here [67]. Microarrays need a uniform standard for platform fabrication, assay protocols, and analysis. They have large datasets and need complex statistical algorithms for processing them with high accuracy. A significant issue, besides normalization of data and the reduction of dimensionality of data, is the sequencing requirement in SAGE. It is comparatively easy and more reliable to search sequences from SAGE databases (which store experimental data using the SAGE technique) stored at different places and then perform northern blots on them. However, Microarrays have a higher degree of random and systematic errors, due to which a comparison of experimental data from different places (experimental sources) is difficult.

What reference genes had been identified?

We present a list (Table 1) showing each of the reference genes, the related parameters and the journal source referenced here. This is in addition to an extensive list presented in an earlier study on reference genes [34].

Implications on experiments using unstable genes as reference genes

There is a growing assertion that the reference genes could be a wrong choice for experimentation if not properly validated for effective quantitative RNA analysis. A single reference gene cannot be a reliable reference gene for all experiments in all conditions [26]. Consider a scenario in which an existing gene, which is widely assumed to be a stable reference gene, is used in the traditional mode of experimentation (without any normalisation of the involved reference gene). The reference gene might have been stable for only a subset of the environmental conditions of the experiment. Could we think of the behaviour of the same reference gene if they were subjected to more extreme conditions? Based on the past studies that we reviewed in this paper, it is clear that the experimental results could be altered because of the changes in the experimental conditions. Thus, we argue based on the evidence, from the current scientific literature, that the past studies which did not perform the normalisation of the reference genes, could have issues in the accuracy of their results. This also means that the scientific results, obtained using invalid reference genes in the involved experiments, may not be acceptable as there will be certain extremes reached by the involved reference genes, which might not have been corrected. Thus, the entire experiment may have to be validated again using the corrected and normalised reference genes for a persuasive argument in favour of the obtained scientific results.

Abbreviations

1) GADPH - glyceraldehyde-3-phosphate dehydrogenase
2) PCR - polymerase chain reaction
3) MARK3 - Microtubule affinity regulating kinase 3
4) B2M - β2 microglobulin
5) CorNV - corneal neovascularisation
6) ACTB - beta-actin
7) TBP – TATA binding protein
8) qRT-PCR - quantitative real-time polymerase chain reaction
9) ALG9 - asparagine-linked glycosylation 9
10) RPL13A - ribosomal protein L13a
11) qRT-PCR - Quantitative real-time PCR
12) HMBS - hydroxymethylbilane synthase
13) qRT - PCR - quantitative reverse transcription-polymerase chain reaction
14) HKG - housekeeping genes
15) HPRT1 - hypoxanthine phosphoribosyltransferase 1
16) RT-qPCR - reverse transcription-quantitative polymerase chain reaction
17) RT-PCR - Real Time Polymerase Chain Reaction
18) SAGE - Sequential Analysis of Gene Expression
19) SDPH - succinate dehydrogenase
20) ANOVA - One way analysis of variance

21) RPA - Ribonuclease Protection Assays

Conclusions

It is clear that the use of reference genes is appropriate only if they are normalised and considered on the basis of the environmental conditions and other factors such as method of gene expression. We believe that it is wrong to consider any reference gene without validating their suitability to the undertaken experiment. Thus, the past experiments, which may have made the mistake of ignoring the need to normalise the genes, may have to validate their results in corrected conditions, so as to get accurate details and remove the possible errors due to the wrong selection of reference genes. We conclude by considering genes (normalised using techniques such as geNorm [33] and NormFinder [35]) as good reference genes for PCR based experiments. These genes if used properly, especially for phylogenetic analysis, could increase our knowledge and provide useful information about related species.

Competing interests

There are no competing interests cited by the authors of this paper.

References

1) de Kok JB, Roelofs RW, Giesendorf BA, Pennings JL, Waas ET, Feuth T, Swinkels DW, Span PN. **Normalization of gene expression measurements in tumour tissues: comparison of 13 endogenous control genes.** Laboratory Investigation 85, 154–159.2005.

2) Ren S, Zhang F, Li C, Jia C, Li S, Xi H, Zhang H, Yang L, Wang Y.. **Selection of housekeeping genes for use in quantitative reverse transcription PCR assays on the murine cornea**. Mol Vis. 2010 Jun 11; 16:1076-86.

3) Harrison PJ, Laatikainen LM, Tunbridge EM, Eastwood SL. **Human brain weight is correlated with expression of the 'housekeeping genes' beta-2-microglobulin (beta2M) and TATA-binding protein (TBP).** Neuropathol Appl Neurobiol. 2010 Jun 11.

4) 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 Jun 21; 11:47.

5) Spanier K, Leese F, Mayer C, Colbourne J, Gilbert D, Pfrender M, Tollrian R. **Predator-induced defenses in Daphnia pulex: selection and evaluation of internal reference genes for gene expression studies with real-time PCR.** BMC Mol Biol. 2010 Jun 29; 11:50.

6) Huis R, Hawkins S, Neutelings G. **Selection of reference genes for quantitative gene expression normalization in flax (Linum usitatissimum L.).** BMC Plant Biol. 2010 Apr 19;10:71.PMID: 20403198

7) Wan G, Yang K, Lim Q, Zhou L, He BP, Wong HK, Too HP. **Identification and validation of reference genes for expression studies in a rat model of neuropathic pain.** Biochem Biophys Res Commun. 2010 Aug 27.PMID: 20804730

8) Hong SM, Bahn SC, Lyu A, Jung HS, Ahn JH. **Identification and Testing of Superior Reference Genes for a Starting Pool of Transcript Normalization in Arabidopsis**. Plant Cell Physiol. 2010 Aug 26.

9) Turabelidze A, Guo S, DiPietro LA.**Importance of housekeeping gene selection for accurate reverse transcription-quantitative polymerase chain reaction in a wound healing model.** Wound Repair Regen. 2010 Aug 19.

10) Zainuddin A, Chua KH, Abdul Rahim N, Makpol S. **Effect of experimental treatment on GAPDH mRNA expression as a housekeeping gene in human diploid fibroblasts**.BMC Mol Biol. 2010 Aug 14;11:59.

11) Rosic NN, Pernice M, Rodriguez-Lanetty M, Hoegh-Guldberg O. **Validation of Housekeeping Genes for Gene Expression Studies in Symbiodinium Exposed to Thermal and Light Stress**. Mar Biotechnol (NY). 2010 Jul 29.

12) Tunbridge EM, Eastwood SL, Harrison PJ. **Changed Relative to What? Housekeeping Genes and Normalization Strategies in Human Brain Gene Expression Studies**. Biol Psychiatry. 2010 Jul 29.PMID: 20673871

13) Morga B, Arzul I, Faury N, Renault T. **Identification of genes from flat oyster Ostrea edulis as suitable housekeeping genes for quantitative real time PCR**. Fish Shellfish Immunol. 2010 Aug 7.PMID: 20696253

14) Dhar AK, Bowers RM, Licon KS, Veazey G, Read B. **Validation of reference genes for quantitative measurement of immune gene expression in shrimp.** Mol Immunol. 2009 May;46(8-9):1688-95.PMID: 19297025

15) Infante C, Matsuoka MP, Asensio E, Cañavate 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 Mar 6;9:28.PMID: 18325098

16) Huggett J, Dheda K, Bustin S, Zumla A. **Real-time RT-PCR normalisation; strategies and considerations.** Genes Immun. 2005 Jun;6(4):279-84.PMID: 15815687

17) Chari R, Lonergan KM, Pikor LA, Coe BP, Zhu CQ, Chan TH, MacAulay CE, Tsao MS, Lam S, Ng RT, Lam WL. **A sequence-based approach to identify reference genes for gene expression analysis.** BMC Med Genomics. 2010 Aug 3; 3:32.

18) 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 Jul 25; 8:62. PMID: 17651481

19) 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 tumour tissues.** Oncol Res. 2009;17(8):353-65.PMID: 19544972

20) Morga B, Arzul I, Faury N, Renault T. **Identification of genes from flat oyster Ostrea edulis as suitable housekeeping genes for quantitative real time PCR**. Fish Shellfish Immunol. 2010 Aug 7. PMID: 20696253

21) Fu J, Bian L, Zhao L, Dong Z, Gao X, Luan H, Sun Y, Song H. **Identification of genes for normalization of quantitative real-time PCR data in ovarian tissues**. Acta Biochim Biophys Sin (Shanghai). 2010 Aug;42(8):568-74.PMID: 20705598

22) Hong SM, Bahn SC, Lyu A, Jung HS, Ahn JH. **Identification and Testing of Superior Reference Genes for a Starting Pool of Transcript Normalization in Arabidopsis.** Plant Cell Physiol. 2010 Aug 26.PMID: 20798276

23) Chari R, Lonergan KM, Pikor LA, Coe BP, Zhu CQ, Chan TH, MacAulay CE, Tsao MS, Lam S, Ng RT, Lam WL. **A sequence-based approach to identify reference genes for gene expression analysis**. BMC Med Genomics. 2010 Aug 3;3:32.PMID: 20682026

24) Zainuddin A, Chua KH, Abdul Rahim N, Makpol S. **Effect of experimental treatment on GAPDH mRNA expression as a housekeeping gene in human diploid fibroblasts**. BMC Mol Biol. 2010 Aug 14;11:59.PMID: 20707929

25) Schmittgen TD and Zakrajsek BA. **Effect of experimental treatment on housekeeping gene expression: validation by real-time, quantitative RT-PCR.** J Biochem Biophys Methods. 2000 Nov 20;46(1-2):69-81.PMID: 11086195

26) Janssens N, Janicot M, Perera T, Bakker A. **Housekeeping genes as internal standards in cancer research**. Mol Diagn. 2004;8(2):107-13.PMID: 15527325

27) Cicinnati VR, Shen Q, Sotiropoulos GC, Radtke A, Gerken G, Beckebaum S. **Validation of putative reference genes for gene expression studies in human hepatocellular**

carcinoma using real-time quantitative RT-PCR.BMC Cancer. 2008 Nov 27;8:350.PMID: 19036168

28) Barsalobres-Cavallari CF, Severino FE, Maluf MP, Maia IG. **Identification of suitable internal control genes for expression studies in Coffea arabica under different experimental conditions**. BMC Mol Biol. 2009 Jan 6;10:1.PMID: 19126214

29) Narsai R, Ivanova A, Ng S, Whelan J. **Defining reference genes in Oryza sativa using organ, development, biotic and abiotic transcriptome datasets.** BMC Plant Biol. 2010 Mar 31;10:56.PMID: 20353606

30) Lee JM, Roche JR, Donaghy DJ, Thrush A, Sathish P. **Validation of reference genes for quantitative RT-PCR studies of gene expression in perennial ryegrass (Lolium perenne L.)**.BMC Mol Biol. 2010 Jan 20;11:8.PMID: 20089196

31) Artico S, Nardeli SM, Brilhante O, Grossi-de-Sa MF, Alves-Ferreira M. **Identification and evaluation of new reference genes in Gossypium hirsutum for accurate normalization of real-time quantitative RT-PCR data**. BMC Plant Biol. 2010 Mar 21;10:49.PMID: 20302670

32) Aleksandar RadoniÄ, Stefanie Thulke, Ian M Mackay, Olfert Landt, Wolfgang Siegert and Andreas Nitsche. **Guideline to reference gene selection for quantitative real-time PCR. Biochem Biophys Res Commun**. 313(4):856-62 (2004)

33) Jo Vandesompele, Katleen De Preter, Filip Pattyn, Bruce Poppe, Nadine Van Roy, Anne De Paepe and Frank Speleman**. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes**. Genome Biology 2002,3:research0034-research0034.11.doi:10.1186/gb-2002-3-7-research0034.

34) Zhu J, He F, Hu S, Yu J. **On the nature of human housekeeping genes**. Trends in Genetics, Volume 24, Issue 10, October 2008, Pages 481-484. PMID: 18786740

35) Ledet-Jensen J., Ørntoft T.: **Normalization of real-time quantitative RT-PCR data: a model based variance estimation approach to identify genes suited for normalization - applied to bladder- and colon-cancer data-sets.** Cancer Research. 2004 (64): 5245-5250.

36) Cheng C, Bhardwaj N, Gerstein M.**The relationship between the evolution of microRNA targets and the length of their UTRs**. BMC Genomics. 2009 Sep 14; 10:431.PMID: 19751524.

37) Michael W. Pfaffl, Aleš Tichopád, Christian Prgomet, Tanja P. Neuvians.**Determination of stable housekeeping genes, differentially regulated target genes and sample**

integrity: BestKeeper – Excel-based tool using pair-wise correlations. Biotechnology Letters 26: 509-515 (2004).

38) Lee S, Jo M, Lee J, Koh SS, Kim S. **Identification of novel universal housekeeping genes by statistical analysis of microarray data.** J Biochem Mol Biol. 2007 Mar 31; 40(2):226-31. PMID: 17394773.

39) 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.

40) 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 May 28; 10:240.

41) Robert D Barber, Dan W Harmer1, Robert A Coleman1 and Brian J Clark. **GAPDH AS A HOUSEKEEPING GENE: ANALYSIS OF GAPDH mRNA EXPRESSION IN A PANEL OF 72 HUMAN TISSUES**. Physiol. Genomics (March 15, 2005). doi:10.1152/physiolgenomics.00025.2005

42) Mori R, Wang Q, Danenberg KD, Pinski JK, Danenberg PV. **Both beta-actin and GAPDH are useful reference genes for normalization of quantitative RT-PCR in human FFPE tissue samples of prostate cancer.**Prostate. 2008 Oct 1; 68(14):1555-60.PMID: 18651557.

43) She X, Rohl CA, Castle JC, Kukerin AV, Johnson JM, Chen R. **Definition, conservation and epigenetic of housekeeping and tissue-enriched genes**. BMC Genomics. 2009 Jun 17;10:269.PMID: 19534766

44) Voskresenskiy AM, Sun LS. **The housekeeping gene (GA3PDH) and the long interspersed nuclear element (LINE) in the blood and organs of rats treated with cocaine.** Ann N Y Acad Sci. 2008 Aug; 1137:309-15. PMID: 18837965

45) Synnergren J, Giesler TL, Adak S, Tandon R, Noaksson K, Lindahl A, Nilsson P, Nelson D, Olsson B, Englund MC, Abbot S, Sartipy P. **Differentiating human embryonic stem cells express a unique housekeeping gene signature.** Stem Cells. 2007 Feb;25(2):473-80.PMID: 17284652

46) Paul S, Mandal SK, Bhattacharyya SS, Boujedaini N, Khuda-Bukhsh AR. **In vitro and in vivo studies demonstrate anticancer property of root extract of Polygala senega.** J Acupunct Meridian Stud. 2010 Sep;3(3):188-96.PMID: 20869020

47) Fu LY, Jia HL, Dong QZ, Wu JC, Zhao Y, Zhou HJ, Ren N, Ye QH, Qin LX. **Suitable reference genes for real-time PCR in human HBV-related hepatocellular carcinoma with different clinical prognoses.** BMC Cancer. 2009 Feb 6;9:49.PMID: 19200351

48) Khimani AH, Mhashilkar AM, Mikulskis A, O'Malley M, Liao J, GolenkoEE, Mayer P, Chada S, Killian JB, Lott ST. **Housekeeping genes in cancer: normalization of array data.**BioTechniques 38:739-745 (May 2005)

49) Glaysher S, Gabriel FG, Johnson P, Polak M, Knight LA, Parker K, Poole M, Narayanan A, Cree IA; NHS Collaborative Research Programme for Predictive Oncology. **Molecular basis of chemosensitivity of platinum pre-treated ovarian cancer to chemotherapy.**Br J Cancer. 2010 Aug 24; 103(5):656-62. Epub 2010 Aug 10.

50) 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 Nov 1; 394(1):110-6. Jul 19. PMID: 19622337.

51) Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, Speleman F.Genome Biol. **Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes**.2002 Jun 18;3(7):RESEARCH0034. PMID: 12184808.

52) Andersen CL, Jensen JL, Ørntoft 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 Aug 1;64(15):5245-50.PMID: 15289330

53) Rosic NN, Pernice M, Rodriguez-Lanetty M, Hoegh-Guldberg O. Validation of **Housekeeping Genes for Gene Expression Studies in Symbiodinium Exposed to Thermal and Light Stress.** Mar Biotechnol (NY). 2010 Jul 29. PMID: 20668900.

54) Popovici V, Goldstein DR, Antonov J, Jaggi R, Delorenzi M, Wirapati P. **Selecting control genes for RT-QPCR using public microarray data**. BMC Bioinformatics. 2009 Feb 2; 10: 42. PMID: 19187545

55) 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 Jul; 37(1):112-4, 116, and 118-9.

56) Ullmannová V, Haskovec C. **The use of housekeeping genes (HKG) as an internal control for the detection of gene expression by quantitative real-time RT-PCR.** Folia Biol (Praha). 2003;49(6):211-6.PMID: 14748434

57) Wu YY, Rees JL. **Variation in epidermal housekeeping gene expression in different pathological states.** Acta Derm Venereol. 2000 Jan-Feb;80(1):2-3.PMID: 10721821

58) Gerard CJ, Andrejka LM, Macina RA. **Mitochondrial ATP synthase 6 as an endogenous control in the quantitative RT-PCR analysis of clinical cancer samples.** PMID: 10837088. Mol Diagn. 2000 Mar; 5(1):39-46.

59) Savli H, Karadenizli A, Kolayli F, Gundes S, Ozbek U, Vahaboglu H. Expression stability of six housekeeping genes: **A proposal for resistance gene quantification studies of Pseudomonas aeruginosa by real-time quantitative RT-PCR.** J Med Microbiol. 2003 May;52(Pt 5):403-8.PMID: 12721316

60) Bustin SA. **Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays.** "Journal of Molecular Endocrinology, 25". pp. 169–193. 2000.

61) Velculescu VE, Zhang L, Vogelstein B, and Kinzler KW (1995). **Serial Analysis of Gene Expression.** Science 270, 484-487.

62) Yamamoto M, Wakatsuki T, Hada A, Ryo A. **Use of serial analysis of gene expression (SAGE) technology.** J Immunol Methods. 2001 Apr; 250(1-2):45-66. PMID: 11251221

63) L D Ke, Z Chen and W K Yung**. A reliability test of standard-based quantitative PCR: exogenous vs. endogenous standards**. Mol Cell Probes 14(2):127-35 (2000)

64) Innis MA et al. (1990). **PCR Protocols: A Guide to Methods and Applications.** Academic Press

65) Hunt M. **Real time PCR tutorial.** The Board of Trustees of the University of South Carolina. 2006. Online:- http://pathmicro.med.sc.edu/pcr/realtime-home.htm

66) Min Hu. **Serial analysis of gene expression.** Nature Protocols.2006.DOI: 10.1038/nprot.2006.269

67) Willmar D. Patino, Omar Y. Mian and Paul M. Hwang. **Serial Analysis of Gene Expression: Technical Considerations and Applications to Cardiovascular Biology.** Circ. Res. 2002;91;565-569.DOI: 10.1161/01.RES.0000036018.76903.18

68) Caradec J, Sirab N, Keumeugni C, Moutereau S, Chimingqi M, Matar C, Revaud D, Bah M, Manivet P, Conti M, Loric S. **Desperate house genes: the dramatic example of hypoxia.** Br J Cancer. 2010 Mar 16;102(6):1037-43.PMID: 20179706

69) Zhang X, Ding L, Sandford AJ. **Selection of reference genes for gene expression studies in human neutrophils by real-time PCR**.BMC Molecular Biology 2005, 6:4 doi: 10.1186/1471-2199-6-4.

70) Reverter A, Ingham A, Dalrymple BP. **Mining tissue specificity, gene connectivity and disease association to reveal a set of genes that modify the action of disease causing genes.** BioData Mining 2008, 1:8. doi:10.1186/1756-0381-1-8

71) van Guilder HD, Vrana KE, Freeman WM.**Twenty-five years of quantitative PCR for gene expression analysis.** Biotechniques 44(5):619-26.Apr 2008.PMID: 18474036

72) Greer S, Honeywell R, Geletu M, Arulanandam R, Raptis L (Feb 19 2010).**Housekeeping genes; expression levels may change with density of cultured cells.** J Immunol Methods. PMID 20171969.

73) de Jonge HJ, Fehrmann RS, de Bont ES, Hofstra RM, Gerbens F, Kamps WA, de Vries EG, van der Zee AG, te Meerman GJ, ter Elst A**. Evidence based selection of housekeeping genes.** PLoS One. 2007 Sep 19; 2(9):e898. PMID: 17878933

74) Curtis KM, Gomez LA, Rios C, Garbayo E, Raval AP, Perez-Pinzon MA, Schiller PC. **EF1alpha and RPL13a represent normalization genes suitable for RT-qPCR analysis of bone marrow derived mesenchymal stem cells**.BMC Mol Biol. 2010 Aug 17;11:61.PMID: 20716364

75) Teste MA, Duquenne M, François JM, Parrou JL. **Validation of reference genes for quantitative expression analysis by real-time RT-PCR in Saccharomyces cerevisiae.**BMC Mol Biol. 2009 Oct 30;10:99.

76) Fujimori S, Hidaka Y, Davidson BL, Palella TD, Kelley WN: **Identification of a single nucleotide change in a mutant gene for hypoxanthine-guanine phosphoribosyltransferase (HPRT Ann Arbor)**. *Hum Genet* 1988, **79**(1):39-43. doi: 10.1007/BF00291707. PMID 2896620.

77) Heng, SSJ, Chan, OYW, Keng, BMH, Ling, MHT. 2011. **Glucan biosynthesis protein G (mdoG) is a Suitable Reference Gene in *Escherichia coli* K-12.** ISRN Microbiology 2011, Article ID 469053.

78) Too, IHK, Ling, MHT. 2011. **Signal Peptidase Complex Subunit 1 (SPCS1) and Hydroxyacyl-CoA Dehydrogenase Beta Subunit (HADHB) are Suitable Reference Genes in Human Lungs.** ISRN Bioinformatics 2011, Article ID 790452.

79) Smith RD, Ogden CW, Penny MA: **Exclusive amplification of cDNA template (EXACT) RT-PCR to avoid amplifying contaminating genomic pseudogenes**. *Biotechniques* 2001, **31**(4):776-778, 780, 782.

80) Yang L, Takuno S, Waters ER, Gaut BS: **Lowly expressed genes in Arabidopsis thaliana bear the signature of possible pseudogenization by promoter degradation**. *Mol Biol Evol* 2011, **28**(3):1193-1203.

81) Pinto FL, Thapper A, Sontheim W, Lindblad P: **Analysis of current and alternative phenol based RNA extraction methodologies for cyanobacteria**. *BMC Mol Biol* 2009, **10**:79.

82) Kreuzer KA, Lass U, Landt O, Nitsche A, Laser J, Ellerbrok H, Pauli G, Huhn D, Schmidt CA: **Highly sensitive and specific fluorescence reverse transcription-PCR assay for the pseudogene-free detection of beta-actin transcripts as quantitative reference**. *Clin Chem* 1999, **45**(2):297-300.

83) Lehmann MH, Weber J, Gastmann O, Sigusch HH: **Pseudogene-free amplification of human GAPDH cDNA**. *Biotechniques* 2002, **33**(4):766, 769-770.

84) Malik AN, Shahni R, Rodriguez-de-Ledesma A, Laftah A, Cunningham P: **Mitochondrial DNA as a non-invasive biomarker: accurate quantification using real time quantitative PCR without co-amplification of pseudogenes and dilution bias**. *Biochem Biophys Res Commun* 2011, **412**(1):1-7.

Figure-1) List of Housekeeping Genes

Sr.No	Reference Genes Found	Organ/Species/Disease	Reference	Are they Reliable?
1	NDUFA1, RPL19, RAB5C, and RPS18	Human Lung samples – cancer affected.	[23]	Yes
2	glyceraldehyde-3-phosphate dehydrogenase (GAPDH)	Human diploid fibroblasts	[24]	Yes
3	beta-2 microglobulin and 18S rRNA	serum-stimulated fibroblasts	[25]	Yes
4	ATPsy6 (mATPsy6),GADPH, porphobilinogen deaminase (PBGD)	Human -number of cell lines and tumour versus matched normal tissue samples	[26]	Yes
5	HMBS,GADPH	human hepatocellular carcinoma(HCC)	[27]	Yes

6	GADPH, 14-3-3 and rpl7	Coffee Arabica	[28]	Yes
7	Xbp1, Tbp, CAPON and Stx16	Chaoborus-induced D. pulex specimens	[5]	Yes
8	LOC_Os06g11170.1	Rice	[29]	Yes
9	eEF1A (s) and YT521-B	Lolium perenne L.	[30]	Yes
10	GhUBQ14 and GhPP2A1 for cotton plants; GhACT4 and GhUBQ14 for flower development, GhACT4 and GhFBX6 for the floral organs and GhMZA and GhPTB for fruit development	Cotton, flower and fruits (plants)	[31]	Yes
11	List of Housekeeping genes	Homo Sapiens	[34]	Yes
12	ACTB	HKGs in human eESCs(embryonic stem cells)	[45]	Yes
13	AhR, p53, PCNA and β-actin	HKGs used in studying Anticancer properties in Polygela Senega organism species	[46]	Yes
14	TBP and HPRT	human HBV-related hepatocellular carcinoma	[47]	Yes
15	ACTB and 18S	human HBV-related hepatocellular carcinoma	[47]	No

16	List of Expression Data with list of several housekeeping genes	Housekeeping Genes and Cancer in humans	[48]	Yes
17	PGBD	Housekeeping gene in ovarian tumour studies	[49]	Yes
18	ACTB, ALAS1, GAPDH, HPRT1, PBGD, PUM1, RPL29 and 18S rRNA	Housekeeping gene in human serous ovarian cancer	[50]	No
19	GUSB, PPIA, and TBP	Housekeeping gene in human serous ovarian cancer.	[50]	Yes
20	S4 ribosomal protein (Rp-S4), Calmodulin (Cal), and Cytochrome oxidase subunit 1 (Cox),	Symbiodinium Exposed to Thermal and Light Stress	[53]	Yes
21	List of 50 breast cancer control genes	Breast cancer control genes in humans in ranked format.	[54] - Table -3	Yes
22	List of HK genes used for Arabidopsis studies	Arabidopsis	[6] Figure-1	Yes