A minimal set of internal control genes for gene expression studies in head and neck squamous cell carcinoma

Vinayak Palve 1, Manisha Pareek 1, Neeraja M Krishnan 1, Gangotri Siddappa 2, Amritha Suresh 2, Moni A Kuriakose 2, Binay Panda Corresp. 1

1 Ganit Labs, Bio-IT Centre, Institute of Bioinformatics and Applied Biotechnology, Bangalore, India
2 Head and Neck Oncology, Mazumdar Shaw Centre for Translational Research, Bangalore, India

Corresponding Author: Binay Panda
Email address: binay@ganitlabs.in

Selection of the right reference gene(s) is crucial in the analysis and interpretation of gene expression data. The aim of the present study was to discover and validate a minimal set of internal control genes in head and neck tumor studies. We analyzed data from multiple sources (in house whole-genome gene expression microarrays, previously published qPCR data and RNA-seq data from TCGA) to come up with a list of 18 genes (discovery set) that had lowest variance and high level of expression across tumors and their matched normal samples. The genes in the discovery set were ranked using four different algorithms (BestKeeper, geNorm, NormFinder, and comparative delta Ct) and a web-based comparative tool, RefFinder, for their stability and variance in expression across tissues. Finally, we validated their expression using qPCR in an additional set of tumor:matched normal samples that resulted in five genes (RPL30, RPL27, PSMC5, MTCH1 and OAZ1), out of which RPL30 and RPL27 were most stable and were abundantly expressed across the tissues. Our data suggest that RPL30 or RPL27 in combination with either PSMC5 or MTCH1 or OAZ1 can be used as a minimal set of control genes in head and neck tumor gene expression studies.
A minimal set of internal control genes for gene expression studies in head and neck squamous cell carcinoma.

Vinayak Palve¹, Manisha Pareek¹, Neeraja M Krishnan¹, Gangotri Siddappa², Amritha Suresh², Moni A Kuriakose² and Binay Panda¹*

¹Ganit Labs, Bio-IT Centre, Institute of Bioinformatics and Applied Biotechnology, Bangalore, India
²Head and Neck Oncology, Mazumdar Shaw Centre for Translational Research, Bangalore, India

* Corresponding author

Binay Panda: binay@ganitlabs.in
Abstract

Selection of the right reference gene(s) is crucial in the analysis and interpretation of gene expression data. The aim of the present study was to discover and validate a minimal set of internal control genes in head and neck tumor studies. We analyzed data from multiple sources (in house whole-genome gene expression microarrays, previously published qPCR data and RNA-seq data from TCGA) to come up with a list of 18 genes (discovery set) that had lowest variance and high level of expression across tumors and their matched normal samples. The genes in the discovery set were ranked using four different algorithms (BestKeeper, geNorm, NormFinder, and comparative delta Ct) and a web-based comparative tool, RefFinder, for their stability and variance in expression across tissues. Finally, we validated their expression using qPCR in an additional set of tumor:matched normal samples that resulted in five genes (RPL30, RPL27, PSMC5, MTCH1 and OAZ1), out of which RPL30 and RPL27 were most stable and were abundantly expressed across the tissues. Our data suggest that RPL30 or RPL27 in combination with either PSMC5 or MTCH1 or OAZ1 can be used as a minimal set of control genes in head and neck tumor gene expression studies.
Introduction

Internal control genes or housekeeping genes are important in obtaining reliable and reproducible data from gene expression studies in cancer (Eisenberg & Levanon 2013). Internal control genes should be abundantly and uniformly expressed across tumor and normal tissues and across different experimental conditions (Janssens et al. 2004). Cancers of head and neck region is the sixth most common cancer worldwide with an incidence of 550,000 cases annually (Ferlay et al. 2010). Past studies argue the lack of uniformity of expression on internal control genes based on experimental conditions (De Jonge et al. 2007; Greer et al. 2010). Although there are previous reports that describe the reference or internal control genes in head and neck squamous cell carcinoma (HNSCC) gene expression studies (Lallemant et al. 2009), use of high-throughput data like microarrays and RNA-seq from tumors and their matched normal tissues following a confirmation of expression in independent set of samples are scant in the literature. Additionally, it is crucial that the expression of internal control genes remains unaltered across temporal, spatial and experimental conditions that takes into account of the genes with a wide dynamic range of expression. Therefore, revisiting the validity of widely used genes like ACTB, TUBB and GAPDH is required along with discovery of a minimal set of internal control genes in HNSCC gene expression studies that use expression data from patient cohorts from different geography (Krishnan et al. 2015; Krishnan et al. 2016) and large consortia like TCGA (Cancer Genome Atlas 2015).

In the present study, we analyzed HNSCC gene expression data from three sources; in house microarray data (Krishnan et al. 2015; Krishnan et al. 2016) (n=21),
TCGA RNA-seq data (Cancer Genome Atlas 2015) (n=42) and qPCR data on individual genes from the previously published studies (Bär et al. 2009; Eisenberg & Levanon 2013; Lallemant et al. 2009; Martin 2016; Vandesompele et al. 2002); to come up with a set of genes (discovery set) that are stably and robustly expressed with least variance across tumor:normal pairs. We subsequently validated the expression of the discovery set in additional tumor:normal pairs (n=14) using qPCR and obtained a minimal set of two housekeeping genes by comparing and ranking their expression during validation.

Materials and Methods

Microarray data processing

The gene expression profiling using Illumina HumanHT-12 v4 expression BeadChips (Illumina, San Diego, CA), raw data collection and data processing using the R package lumi (Du et al. 2008) and individual batch-correction using ComBat (Johnson et al. 2007) is described previously (Krishnan et al. 2015; Krishnan et al. 2016). The raw signal intensities from arrays were exported from GenomeStudio for pre-processing and analyzed using R. Gene-wise expression intensities for tumor and matched control samples from GenomeStudio were transformed and normalized using VST (Variance Stabilizing Transformation) and LOESS methods, respectively, using lumi (Du et al. 2008) and top genes with least across-sample variance in expression were selected. Genes with fold change between 0.95-1.05 (tumor/normal) and with standard deviation < 0.05 for the fold changes were selected further from a larger gene list at a high level in both the microarray and the RNA-seq data (Supplementary Table 1).
**RNA-seq data processing**

RSEM-processed RNA-Seq gene expression values (TCGA pipeline, Level 3) were downloaded from the old TCGA repository ([https://genome-cancer.ucsc.edu/proj/site/hgHeatmap/](https://genome-cancer.ucsc.edu/proj/site/hgHeatmap/)). The transcripts per kilobase million (TPM) values for all the genes were extracted from the Level 3 files. Further, genes that showed zero or NA values in any of these samples were eliminated, and the log fold change values between the respective tumor and normal samples calculated by taking a log transformation of the ratios between expression values for tumor and normal. We filtered expression data that fulfilled three criteria for all samples; expressed at >= 3 TPM, tumor/normal ratio: 0.9-1.1 and standard deviation across all the tissues < 0.5. Finally genes were short-listed from a larger gene list expressed at a high level in both the microarray and the RNA-seq data (Supplementary Table 1) for the discovery set.

**Discovery set**

In addition to the microarray and RNA-seq gene expression data, the most commonly used internal control genes form the HNSCC gene expression literature (Bär et al. 2009; Eisenberg & Levanon 2013; Lallemant et al. 2009; Martin 2016; Vandesompele et al. 2002) were also considered to select the discovery set (n=18) (Supplementary Table 2). These genes were used for validation in an independent set of tumor:normal pairs (n=14) using qPCR.

**Patient samples**

Patients were accumulated voluntarily after obtaining informed consents from each patient and as per the guidelines from the institutional ethics committee of the
Mazumdar-Shaw Cancer Centre (IRB Approval number: NHH/MEC-CL/2014/197). All the tissues were frozen immediately in liquid nitrogen and stored at -80°C until further use. Only those tumors with squamous cell carcinoma (with at least 70% tumor cells and with confirmed diagnosis) and their adjacent normal tissues as confirmed by histology (Supplementary Figure 1) were included in the current study. Patients underwent staging according to AJCC criteria, and curative intent treatment as per NCCN guidelines involving surgery with or without post-operative adjuvant radiation or chemo-radiation at the Mazumdar Shaw Cancer Centre were accrued for the validation study.

RNA extraction and cDNA synthesis

The total RNA was extracted from 25mg of tissues using the RNeasy mini kit (Qiagen, cat:74104) with on-column digestion of DNA using RNase free DNase set (Qiagen, cat:79254) as per manufacturer’s instructions. The RNA quality was checked using gel electrophoresis and Agilent Bioanalyzer. Five hundred nanogram of total RNA was subjected to cDNA synthesis using Takara’s Prime Script first strand cDNA synthesis kit (cat: 6110A).

Quantitative real-time PCR (qPCR)

Quantitative PCR was carried out using KAPA Biosystem’s SYBR Fast qPCR universal master mix (cat: KK4601). The primer sequences for all the 18 discovery set reference genes and the amplification conditions are mentioned in Table 1A. The primers were either designed for this study or were chosen from the literature (Campos et al. 2009) or from the online resources (https://primerdepot.nci.nih.gov/ and
https://pga.mgh.harvard.edu/primerbank/). All amplification reactions were carried out in triplicates, using three negative controls: no template control (NTC) with nuclease free water (Ambion, cat: AM9932), no amplification control (NAC) and no primer control (NPC) in each amplification plate. We have followed the guidelines for quantitative real-time PCR experiments as suggested previously (Bustin et al. 2010).

Statistical analysis

For stability comparison of the internal candidate reference genes obtained using the discovery set, we analyzed the validation data independently using four most commonly used algorithms, Genorm (Vandesompele et al. 2002), Normfinder (Andersen et al. 2004), Bestkeeper (Pfaffl et al. 2004) and Comparative Ct (Schmittgen & Livak 2008). In addition to the comparative Ct method, we also calculated the values of mean standard deviation (Silver et al. 2006) of Ct for each gene. The results from the above tools were compared with an online tool Reffinder (Xie et al. 2012) and the results were interpreted and presented separately. Graphpad prism software version 5 was used to analyze the data and plot the graphs.

Results and Discussion

The schema for selecting genes for discovery set is depicted in Figure 1. After analyzing data from all sources, 18 genes were selected that had least variance across all the tumor and normal samples, were well annotated and had some biology known (Table 1A and Supplementary Table 2). Standard curves reflecting the linear regression ($R^2 > 0.9$) and amplification efficiency (Supplementary Figure 2, Supplementary Table 3)
were generated for the select candidate genes. Primers for all the genes showed specific amplifications as shown by the dissociation curves (Supplementary Figure 3) and by gel electrophoresis of the amplified products (Supplementary Figure 4). Data from qPCR validation of genes (Figure 2A, Supplementary Table 3) showed variable levels of expression for the 18 genes in different validation samples. As Figure 2B indicates independent results from the four different algorithms (Genorm, Normfinder, Bestkeeper and Comparative Ct) and the online tool RefFinder. As the tool RefFinder does not take amplification efficiency into account and merely sums up results from other tools, we did not rely on the results from RefFinder but used it as one of the tools along with the other four to interpret the results independently. Out of the 18 genes across all the samples, we ranked the expression of top 11 stably expressed genes namely, RPL30, RPL27, PSMC5, OAZ1, MTCH1, TSPAN21, DARS, MKRN1, RPS13, RPL5 and RPL37A. Although, there was some inconsistency between the top-ranked genes among the different algorithms/tools, for the genes RPL27 and RPL30 the data was consistent across all the algorithms As both of RPL27 and RPL30 belong to the same ribosomal family proteins, and in order to avoid any potential error due to mutual expression alterations of those genes, we considered inclusion of the next three genes PSMC5, MTCH1 and OAZ1, the expression of which were stable across all the samples. Therefore, RPL27 or RPL30 in combination with any one of the three (PSMC5 / MTCH1 / OAZ1) fulfilled all the 3 criteria for ideal internal control genes, least variance across samples, high-level of expression in both tumors and normal and at the top of the rank by the algorithms tested.
The aim of the current study was not to select a set of genes for a particular tumor stage and/or differentiation and in response to drug treatment but to select genes that are robustly expressed across multiple tumor stages and differentiation and across tumor and their matched normal samples. In HNSCC, previous attempts have been made to come up with a set of internal control genes using qPCR data (Lallemant et al. 2009; Martin 2016) or using tumors and healthy controls (Reddy et al. 2016). However, it is important that the discovery is made from a transcriptome-wide study that takes into account data from tumors and their matched control samples. This is especially important as we have data on HNSCC from large consortia, like TCGA (Cancer Genome Atlas 2015), and data from HNSCC from non-TCGA cohorts (Krishnan et al. 2015; Krishnan et al. 2016). It is not surprising that the genes in our discovery set code for structural proteins (ribosomal, cell structure and integrity) and perform essential functions like glycolysis and translation (Supplementary Table 2). In the validation study, RPL30, RPL27, PSMC5, OAZ1 and MTCH1 came as the top 5 ranked genes with stable expression across all the samples. Both RPL27 and RPL30 were previously found to be one the most stable candidate internal control genes (De Jonge et al. 2007; Lallemant et al. 2009) with stable expression in salivary samples of oral cancer patients (Martin 2016). Additionally, previous findings on non-HNSCC tumors had RPL27, RPL30 and MTCH1 in the list of top internal control genes (Popovici et al. 2009). Although, our data supports PSMC5, OAZ1 and MTCH1 were stably expressed and ranked after RPL30 and RPL27 by RefFinder, they were expressed at a lower level than RPL30 and RPL27. Therefore, we suggest using RPL30 or RPL27 in combination with
one of the three genes (*PSMC5*, *MTCH1*, *OAZ1*) to serve as a minimal set of control
genes in head and neck tumor gene expression studies.

Our study suffers from some shortcomings. First, our sample size was small. Second, the qPCR studies, although highly sensitive and widely used, could have resulted in a bias in our gene expression data. RNA input volume, cDNA synthesis efficiency, pipetting volumes and accuracy, and primer amplification efficiencies are some of the factors that may affect qPCR results (Bustin et al. 2009). Other factors that might have influenced our selection of the discovery set are inter-platform and -assay variability, temporal, spatial and experimental conditions (for example, drug treatment), and presence of transcript variants in tumor samples. A larger sample size for different experimental conditions and data from cell lines assayed by using a single kit and a single platform (for example, total transcriptome RNA-seq using a single provider's platform and assay version) may provide answers to these in the future. Third, the selection of genes in our study might have got biased as a result of the field effect (due to use of adjacent normal tissues). Use of true normal tissues may circumvent this issue. However, due to the stringency of the ethics committee rules to obtain additional normal tissues from patients without a clinical reason(s), studies with true normal tissues, outside of the adjacent normal, are least likely to take place. Fourth, in our study, like previously reported [24], a difference in the ranking of genes may be tool-specific and certain parameters (like qPCR amplification efficiency) might have influenced the ranking of the genes. A better design and a robust statistical tool to rank genes may overcome this limitation in the future.
Conclusions

Although past studies are heavily reliant on the use of ACTB, TUBB and GAPDH to normalize expression data, evidence suggests that the expression of these genes vary greatly and are not suitable internal control genes (Glare et al. 2002; Oliveira et al. 1999; Selvey et al. 2001; Zhong & Simons 1999). Additionally, expression of internal control genes may vary based on experimental conditions and tumor type. Therefore, we used a larger discovery set to validate internal control genes across HNSCC tumors and their paired normal samples to come up with a set of robust and stably expressed genes (RPL30, RPL27, PSMC5, MTCH1 and OAZ1) across tissues.

List of abbreviations

HNSCC: Head and neck squamous cell carcinoma

RPL30: 60S ribosomal protein L30;

RPL27: 60S ribosomal protein L27

NTC: No template control

NAC: No amplification control

NPC: No primer control

Acknowledgement

The authors thank Janani Hariharan in helping with the RNA-seq data analyses.
References

Andersen CL, Jensen JL, and Ørntoft TF. 2004. 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 research 64:5245-5250.

Bär M, Bär D, and Lehmann B. 2009. Selection and validation of candidate housekeeping genes for studies of human keratinocytes—review and recommendations. Journal of Investigative Dermatology 129:535-537.

Bustin SA, Beaulieu JF, Huggett J, Jaggi R, Kibenge FS, Olsvik PA, Penning LC, and Toegel S. 2010. MIQE precis: Practical implementation of minimum standard guidelines for fluorescence-based quantitative real-time PCR experiments. BMC Mol Biol 11:74.

Bustin SA, Benes V, Garson JA, Hellemans J, Huggett J, Kubista M, Mueller R, Nolan T, Pfaffl MW, and Shipley GL. 2009. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. Clinical chemistry 55:611-622.

Campos MS, Rodini CO, Pinto-Junior DS, and Nunes FD. 2009. GAPD and tubulin are suitable internal controls for qPCR analysis of oral squamous cell carcinoma cell lines. Oral Oncol 45:121-126.

Cancer Genome Atlas N. 2015. Comprehensive genomic characterization of head and neck squamous cell carcinomas. Nature 517:576-582.

De Jonge HJ, Fehrmann RS, de Bont ES, Hofstra RM, Gerbens F, Kamps WA, de Vries EG, van der Zee AG, te Meerman GJ, and ter Elst A. 2007. Evidence based selection of housekeeping genes. PLoS One 2:e898.

Du P, Kibbe WA, and Lin SM. 2008. lumi: a pipeline for processing Illumina microarray. Bioinformatics 24:1547-1548.

Eisenberg E, and Levanon EY. 2013. Human housekeeping genes, revisited. Trends in Genetics 29:569-574.

Eisenberg E, and Levanon EY. 2013. Human housekeeping genes, revisited. Trends in Genetics 29:569-574.

Ferlay J, Shin HR, Bray F, Forman D, Mathers C, and Parkin DM. 2010. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. Int J Cancer 127:2893-2917.

Glare E, Divjak M, Bailey M, and Walters E. 2002. β-Actin and GAPDH housekeeping gene expression in asthmatic airways is variable and not suitable for normalising mRNA levels. Thorax 57:765-770.

Greer S, Honeywell R, Geletu M, Arulanandam R, and Raptis L. 2010. Housekeeping genes; expression levels may change with density of cultured cells. Journal of immunological methods 355:76-79.

Janssens N, Janicot M, Perera T, and Bakker A. 2004. Housekeeping genes as internal standards in cancer research. Molecular Diagnosis 8:107-113.

Johnson WE, Li C, and Rabinovic A. 2007. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics 8:118-127.

Krishnan N, Gupta S, Palve V, Varghese L, Pattnaik S, Jain P, Khyriem C, Hariharan A, Dhas K, Nair J, Pareek M, Prasad V, Siddappa G, Suresh A, Kekatpure V, Kuriakose M, and Panda B. 2015. Integrated analysis of oral tongue squamous cell carcinoma identifies key variants and pathways linked to risk habits, HPV, clinical parameters and tumor recurrence. F1000Res 4:1215.

Krishnan NM, Dhas K, Nair J, Palve V, Bagwan J, Siddappa G, Suresh A, Kekatpure VD, Kuriakose MA, and Panda B. 2016. A Minimal DNA Methylation Signature in Oral
Tongue Squamous Cell Carcinoma Links Altered Methylation with Tumor Attributes. *Mol Cancer Res* 14:805-819.

Lallemant B, Evrard A, Combescure C, Chapuis H, Chambon G, Raynal C, Reynaud C, Sabra O, Joubert D, and Hollande F. 2009. Reference gene selection for head and neck squamous cell carcinoma gene expression studies. *BMC Molecular Biology* 10:78.

Martin JL. 2016. Validation of Reference Genes for Oral Cancer Detection Panels in a Prospective Blinded Cohort. *PLoS One* 11:e0158462.

Oliveira JG, Prados RZ, Guedes ACM, Ferreira PC, and Kroon EG. 1999. The housekeeping gene glyceraldehyde-3-phosphate dehydrogenase is inappropriate as internal control in comparative studies between skin tissue and cultured skin fibroblasts using Northern blot analysis. *Archives of Dermatological Research* 291:659-661.

Pfaffl MW, Tichopad A, Prgomet C, and Neuvians TP. 2004. 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.

Popovici V, Goldstein DR, Antonov J, Jaggi R, Delorenzi M, and Wirapati P. 2009. Selecting control genes for RT-QPCR using public microarray data. *BMC bioinformatics* 10:42.

Reddy RB, Bhat AR, James BL, Govindan SV, Mathew R, Ravindra DR, Hedne N, Illiayaraja J, Kekatpure V, Khora SS, Hicks W, Tata P, Kuriakose MA, and Suresh A. 2016. Meta-Analyses of Microarray Datasets Identifies ANO1 and FADD as Prognostic Markers of Head and Neck Cancer. *PloS one* 11:e0147409.

Schmittgen TD, and Livak KJ. 2008. Analyzing real-time PCR data by the comparative CT method. *Nature protocols* 3:1101-1108.

Selvey S, Thompson EW, Matthaei K, Lea RA, Irving MG, and Griffiths LR. 2001. β-Actin—an unsuitable internal control for RT-PCR. *Molecular and cellular probes* 15:307-311.

Silver, N., Best, S., Jiang, J., & Thein, S. L. 2006. Selection of housekeeping genes for gene expression studies in human reticulocytes using real-time PCR. *BMC molecular biology*, 7: 33.

Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, and Speleman F. 2002. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome biology* 3:research0034. 0031.

Xie F, Xiao P, Chen D, Xu L, and Zhang B. 2012. miRDeepFinder: a miRNA analysis tool for deep sequencing of plant small RNAs. *Plant molecular biology* 80:75-84.

Zhong H, and Simons JW. 1999. Direct comparison of GAPDH, β-actin, cyclophilin, and 28S rRNA as internal standards for quantifying RNA levels under hypoxia. *Biochemical and biophysical research communications* 259:523-526.
**Figure Legends:**

**Figure 1:** Schema indicating the selection criteria and validation of the internal control genes.

**Figure 2:** Expression of genes in the discovery set across tumors and normal samples (A). The Ct values were generated from qPCR experiments and plotted as box plots with median values as lines and boxes indicating 10-90 percentile values and the whiskers as the maximum and minimum values. Ranking of 18 genes using different tools (B). Y-axis for different tools represents: stability value (Genorm, Normfinder), Std.dev+-CP(Bestkeeper), $2^{\text{delCt}}$ (Comparative Ct method) and geomean of ranking (RefFinder). The columns in light grey represent stable genes across tumors and normal samples.

**Supplementary Figure and Table Legends**

**Supplementary Figure 1:** Representative histological images of the adjacent normal tissues.

**Supplementary Figure 2:** Standard curves of 18 candidate housekeeping genes in the discovery set. The qPCR was performed using 3 times dilution of the template. $R^2$ indicates the correlation coefficient and amplification efficiencies on the basis of the slopes.

**Supplementary Figure 3:** Dissociation curves for the 18 candidate housekeeping genes in the discovery set. The single peak of dissociation curve in qPCR performed for all the genes indicates the specific amplification.
Supplementary Figure 4: Electrophoresis of the PCR amplicons loaded on a 2% agarose gel.

Supplementary Table 1: A larger set of genes common between the microarray data and RNA-seq data with high expression in both tumor and normal samples.

Supplementary Table 2: Genes selected in the discovery set and their function.
Figure 1

Schema indicating the selection criteria and validation of the internal control genes.
Figure 2

Expression of genes in the discovery set across tumors and normal samples (A).

The Ct values were generated from qPCR experiments and plotted as box plots with median values as lines and boxes indicating 10-90 percentile values and the whiskers as the maximum and minimum values. Ranking of 18 genes using different tools (B). Y-axis for different tools represents; stability value (Genorm, Normfinder), Std.dev+-(CP)(Bestkeeper), $2^{-(\Delta \text{Ct})}$ (Comparative Ct method) and geomean of ranking (RefFinder). The columns in light grey represent stable genes across tumors and normal samples.
Table 1 (on next page)

Primers used in the current study (A).

The Reference column number indicates- 1: primers designed for the study; 2: primers from Hendrik de Jonge et al., 2007; 3: primers from Primerdepot (https://primerdepot.nci.nih.gov/); 4: primers from primer bank (https://pga.mgh.harvard.edu/primerbank/); 5: primers from Campos et al, 2009. Results from different tools on the internal control genes and their ranks as obtained using the validation samples (B).
| No | Gene | F/R | Sequence | Genomic Position | Exon-Intron Junction | Intron spanning | Amplicon Size (bp) | Reference |
|----|------|-----|----------|------------------|----------------------|-----------------|-------------------|-----------|
| 1  | RPL27 | F   | ACAATCACCTAATGCCCACA | chr17:43000081-43002884 | ✓ | ✓ | 146 | 1 |
| 2  | RPS29 | F   | GCATCGTGAGAGCAAGATG | chr14:49877820-49883963 | ✓ | ✓ | 213 | 2 |
| 3  | OAZ1  | F   | GGATCCCTCAATAGCCACTGC | chr19:2269621-2271410 | ✓ | ✓ | 150 | 2 |
| 4  | RPL30 | F   | ACAGCATGCGGAAAATACTAC | chr8:98041745-98042696 | ✓ | ✓ | 158 | 2 |
| 5  | LDH A | F   | TTGGTTGGGGTTGGTGCTGTT | chr11:18396916-18399514 | ✓ | ✓ | 137 | 1 |
| 6  | ACTB | F   | GCTGCCATCAGGGGAG | chr7:5529603-5530574 | ✓ | ✓ | 129 | 1 |
| 7  | RPS13 | F   | CGTCCCCACTTGTGTTAAGTT | chr11:17075584-17077438 | ✓ | ✓ | 124 | 1 |
| 8  | GAPDH | F   | GCATCTGGCCTGACTCTGA | chr12:6537873-6548138 | ✓ | ✓ | 162 | 5 |
| 9  | TUBA1B | F   | GTCGCCCTGCGCTCTTAATC | chr12:49129610-49131330 | ✓ | ✓ | 146 | 1 |
| 10 | PSMC5 | F   | TTGCACGACGAGGAGAG | chr11:63827498-63829528 | ✓ | ✓ | 124 | 1 |
| 11 | MITCH1 | F   | TTGCCCTATGGATGGAAGATG | chr6:36978607-36980780 | ✓ | ✓ | 123 | 3 |
| 12 | MKRN1 | F   | GCAGCGAGGATGATCTACGTGA | chr7:140459862-140471905 | ✓ | ✓ | 98 | 3 |
| 13 | ERGIC3 | F   | TCTCATGTCTGTTCATTTCTC | chr20:35524342-35526313 | ✓ | ✓ | 152 | 4 |
| 14 | ADRM1 | F   | CAATGCTCTCATCTTGGTC | chr20:62304519-62306588 | ✓ | ✓ | 91 | 3 |
| 15 | RPL37A | F   | TTCTGATGGCGGACTTTACC | chr21:16499965-216501378 | ✓ | ✓ | 115 | 3 |
| 16 | RPL5 | F   | TCTGATATGCTATAGGGCTGGT | chr19:92873537-92885898 | ✓ | ✓ | 136 | 4 |
| 17 | TSPAN31 | F   | GGCTATTAACCGAAGCAAGCG | chr12:57746235-57746683 | ✓ | ✓ | 117 | 4 |
| 18 | DARS | F   | AGCGCAAGAGGTTGAGGA | chr13:135979349-135986450 | ✓ | ✓ | 124 | 4 |
Table 1: Primers used in the current study (A). The Reference column number indicates- 1: primers designed for the study; 2: primers from Hendrik de Jonge et al., 2007; 3: primers from Primerdepot (https://primerdepot.nci.nih.gov/); 4: primers from primer bank (https://pga.mgh.harvard.edu/primerbank/); 5: primers from Campos et al, 2009. Results from different tools on the internal control genes and their ranks as obtained using the validation samples (B).