

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

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

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 quantitative real-time PCR (qPCR) data and RNA-seq data from TCGA) to come up with a list of 18 genes (discovery set) that had the lowest variance, a 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.

Submitted 14 December 2017

Accepted 20 June 2018

Published 14 August 2018

Corresponding author

Binay Panda, binay@ganitlabs.in

Academic editor

Giuseppe Nocentini

Additional Information and Declarations can be found on page 9

DOI 10.7717/peerj.5207

© Copyright

2018 Palve et al.

Distributed under

Creative Commons CC-BY 4.0

OPEN ACCESS

Subjects Genomics, Oncology

Keywords Gene expression, Internal control genes, Head and neck squamous cell carcinoma, Housekeeping genes, Stable expression

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, 2016) and large consortia like TCGA (*The Cancer Genome Atlas Network*, 2015).

In the present study, we analyzed HNSCC gene expression data from three sources; *in house* microarray data (Krishnan *et al.*, 2015, 2016) ($n = 21$), TCGA RNA-seq data (*The Cancer Genome Atlas Network*, 2015) ($n = 42$) and quantitative real-time PCR (qPCR) data on individual genes from the previously published studies (Bär, Bär & Lehmann, 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, USA), raw data collection and data processing using the R package lumi (Du, Kibbe & Lin, 2008) and individual batch-correction using ComBat (Johnson, Li & Rabinovic, 2007) is described previously (Krishnan *et al.*, 2015, 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 variance stabilizing transformation (VST) and LOESS methods, respectively, using lumi (Du, Kibbe & Lin, 2008) and top genes with least across-sample variance in expression were selected. Genes with fold change between 0.95 and 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 (Table S1).

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/>). 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 (Table S1) for the discovery set.

Discovery set

In addition to the microarray and RNA-seq gene expression data, the most commonly used internal control genes from the HNSCC gene expression literature (Bär, Bär & Lehmann, 2009; Eisenberg & Levanon, 2013; Lallemand *et al.*, 2009; Martin, 2016; Vandesompele *et al.*, 2002) were also considered to select the discovery set ($n = 18$) (Table S2). 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 consent 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 (Fig. S1) 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 25 mg of tissues using the RNeasy mini kit (cat:74104; Qiagen, Germantown, MD, USA) with on-column digestion of DNA using RNase free DNase set (cat:79254; Qiagen, Germantown, MD, USA) as per manufacturer's instructions. The RNA quality was checked using gel electrophoresis and Agilent Bioanalyzer. A total of 500 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

Quantitative real-time 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 with nuclease free water (cat: AM9932; Ambion, Austin, TX, USA), no amplification control and no primer control in each amplification plate. We have followed the guidelines for qPCR experiments as suggested previously (Bustin *et al.*, 2010).

Table 1 Primers used in the current study (A).

A								
No.	Gene	F/R	Sequence	Genomic position	Exon-intron junction	Intron spanning	Amplicon size (bp)	Reference
1	<i>RPL27</i>	F	ACAATCACCTAATGCCACA	chr17:4300081-	✓	✓	146	This study
		R	GCCTGTCTTGATCTCTCTCAA	43002884				
2	<i>RPS29</i>	F	GCACTGCTGAGAGCAAGATG	chr14:49577820-	✓	✓	213	<i>De Jonge et al. (2007)</i>
		R	ATAGGCAGTGCCAAGGAAGA	49586363				
3	<i>OAZ1</i>	F	GGATCCTCAATAGCCACTGC	chr19:2269621-	-	✓	150	<i>De Jonge et al. (2007)</i>
		R	TACAGCAGTGGAGGGAGACC	2271410				
4	<i>RPL30</i>	F	ACAGCATGCGGAAAAATACTAC	chr8:98041745-	-	✓	158	<i>De Jonge et al. (2007)</i>
		R	AAAGGAAAAATTTTCAGGTTT	98042696				
5	<i>LDHA</i>	F	TTGTTGGGGTTGGTGTCTGTT	chr11:18396916-	-	✓	137	This study
		R	AAGGCTGCCATGTTGGAGAT	18399514				
6	<i>ACTB</i>	F	CCTTGCACATGCCGGAG	chr7:5529603-	-	✓	112	https://primerdepot.nci.nih.gov
		R	GCACAGAGCCTCGCCTT	5530574				
7	<i>RPS13</i>	F	CGTCCCCACTTGTTGAAGTT	chr11:17075584-	✓	✓	129	This study
		R	CGTACTTGTGCAACACCATGT	17077438				
8	<i>GAPDH</i>	F	GCATCCTGGGCTACACTGA	chr12:6537873-	-	✓	162	<i>Campos et al. (2009)</i>
		R	CCACCACCCTGTTGCTGTA	6538138				
9	<i>TUBA1B</i>	F	GTCGCCTTCGCCTCCTAATC	chr12:49129610-	-	✓	146	This study
		R	TCACTTGGCATCTGGCCATC	49131330				
10	<i>PSMC5</i>	F	TTGACGGACCAGAGCAGATG	chr17:63827498-	✓	✓	124	This study
		R	CTCCGGAGGTTTTGGCTCTT	63829528				
11	<i>MTCH1</i>	F	CTTGGCGTAGGTGAAGAAGC	chr6:36978607-	✓	✓	123	https://primerdepot.nci.nih.gov
		R	CATCCCCTGCTCTACGTGAA	36985885				
12	<i>MKRN1</i>	F	GCAGCAAGGGATGACTTTGT	chr7:140459862-	-	✓	98	https://primerdepot.nci.nih.gov
		R	TGTATTTATGGAGACCGCTGC	140471905				
13	<i>ERGIC3</i>	F	TCTCATGCTGCTACTGTTCCCT	chr20:35542342-	✓	✓	152	https://pga.mgh.harvard.edu/primerbank/
		R	CAATACTCAGATAGGCACAAGGC	35542833				
14	<i>ADRM1</i>	F	CAATGCTCCTCATCCTGGTC	chr20:62304519-	-	✓	91	https://primerdepot.nci.nih.gov
		R	GGAGGGTCTACGTGCTGAAG	62306228				
15	<i>RPL37A</i>	F	TTCTGATGGCGGACTTTACC	chr2:216499955-	-	✓	115	https://primerdepot.nci.nih.gov
		R	ATGAAGAGACGAGCTGTGGG	216501378				
16	<i>RPL5</i>	F	TCGTATAGCAGCATGAGCTTTC	chr1:92837537-	-	✓	136	https://primerdepot.nci.nih.gov
		R	TGTTGCAGATTACATGCGCT	92840589				
17	<i>TSPAN31</i>	F	GGCTATTAACCGAAGCAAACAGA	chr12:57746235-	-	✓	117	https://pga.mgh.harvard.edu/primerbank/
		R	GTGAGGTTGAATAAGCCACAACA	57746683				
18	<i>DARS</i>	F	AGCCGCAAGAGTCAGGAGA	chr2:135979349-	-	✓	124	https://pga.mgh.harvard.edu/primerbank/
		R	CCCGAACCAAAACTCGATCTG	135985450				

Table 1 (continued).

B

Ranking of candidate reference genes

Rank	Genorm		Normfinder 1		Comparative Ct		Bestkeeper		Reffinder comprehensive	
	Genes	Stability value (M)	Genes	Stability value (M)	Genes	$2^{-\Delta\Delta Ct}$	Genes	Pearson correlation coefficient	Genes	Geomean ranking
1	<i>RPL30</i>	0.254	<i>PSMC5</i>	0.004	<i>RPL27</i>	1.13	<i>RPL27</i>	0.278	<i>RPL30</i>	1.32
2	<i>RPL27</i>	0.254	<i>MTCH1</i>	0.005	<i>RPS13</i>	0.90	<i>RPL30</i>	0.321	<i>RPL27</i>	1.57
3	<i>PSMC5</i>	0.348	<i>TSPAN21</i>	0.006	<i>RPL30</i>	1.21	<i>RPL5</i>	0.362	<i>PSMC5</i>	3.35
4	<i>OAZ1</i>	0.365	<i>DARS</i>	0.006	<i>RPL5</i>	0.82	<i>RPS13</i>	0.391	<i>OAZ1</i>	5.03
5	<i>MKRN1</i>	0.375	<i>RPL30</i>	0.007	<i>RPL37A</i>	0.76	<i>MTCH1</i>	0.419	<i>MTCH1</i>	5.38

Note:

Results from different tools on the internal control genes and their ranks as obtained using the validation samples (B).

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, Jensen & Ørntoft, 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 Fig. 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; Table S2). Standard curves reflecting the linear regression ($R^2 > 0.9$) and amplification efficiency (Fig. S2; Table S3) were generated for the select candidate genes. Primers for all the genes showed specific amplifications as shown by the dissociation curves (Fig. S3) and by gel electrophoresis of the amplified products (Fig. S4). Data from qPCR validation of genes (Fig. 2A; Table S3) showed variable levels of expression for the 18 genes in different validation samples. As Fig. 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

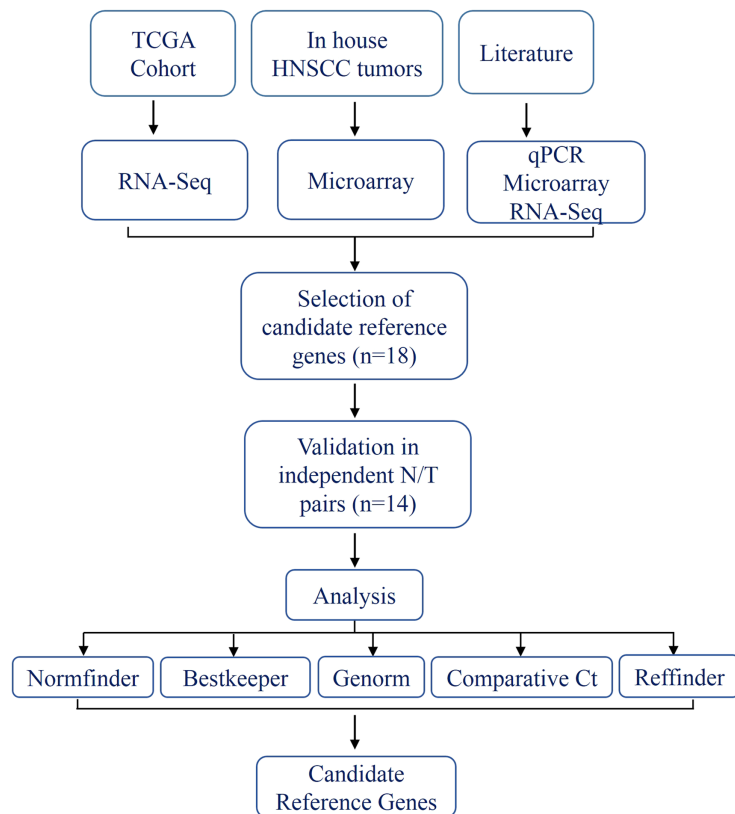


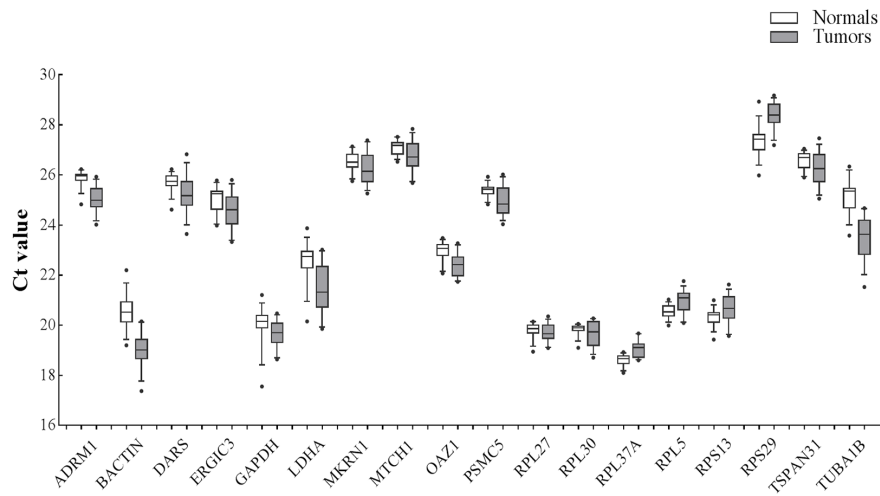
Figure 1 Schema indicating the selection criteria and validation of the internal control genes.

Full-size  DOI: 10.7717/peerj.5207/fig-1

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 three 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 (The Cancer Genome Atlas Network, 2015), and data from HNSCC from non-TCGA cohorts

A



B

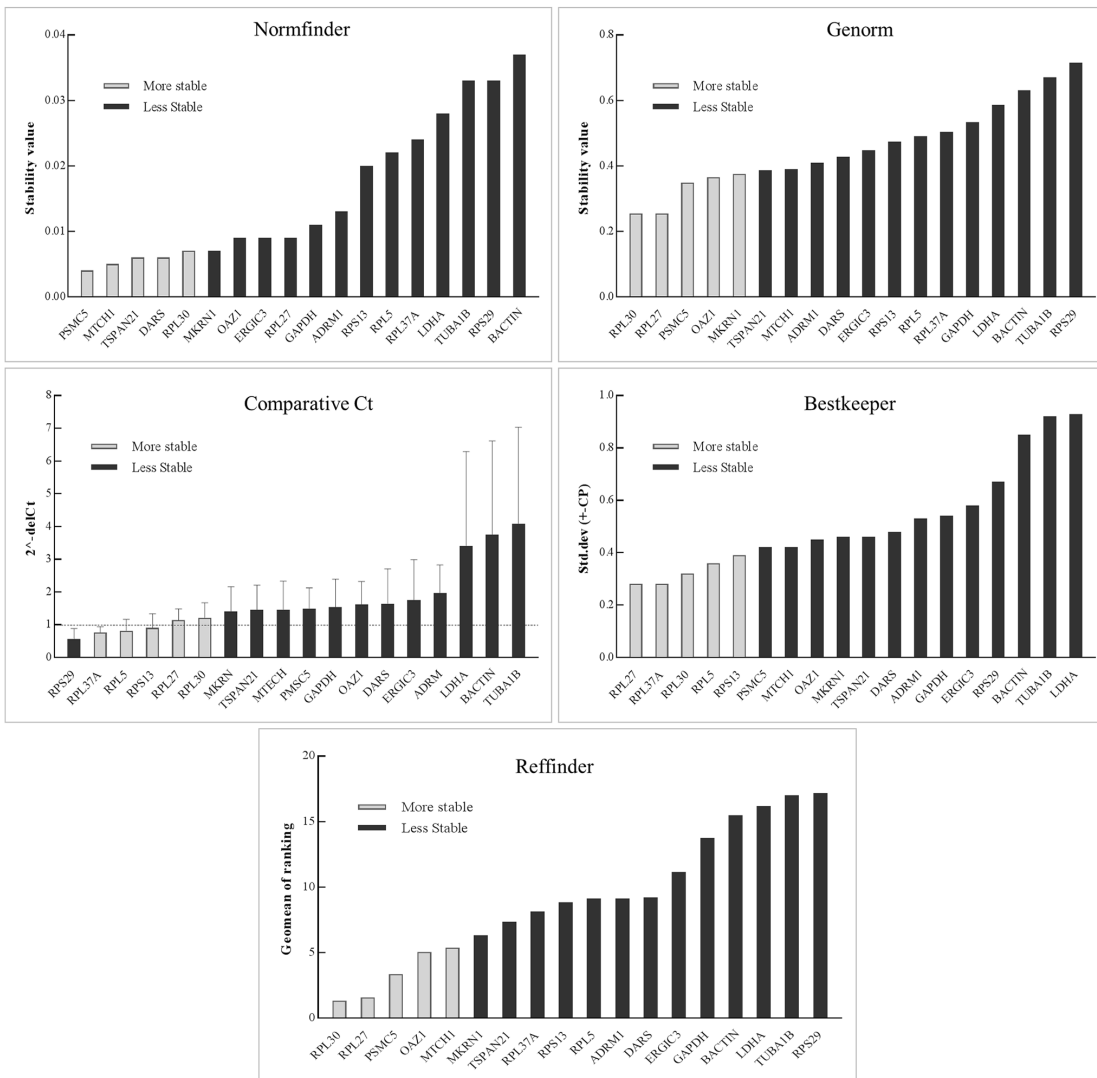



Figure 2 Expression of genes in the discovery set across tumors and normal samples. (A) The Ct values were generated from qPCR experiments and plotted as box plots with median values as lines and boxes indicating 10–90 percentile values and whiskers as the maximum and minimum values. Ranking of 18 genes using different tools (B). Y-axis for different tools represents: stability value (Genorm, Normfinder), standard deviation \pm (CP) (Bestkeeper), $2^{-(\Delta\Delta Ct)}$ (Comparative Ct method) and geomean of ranking (RefFinder). The columns in light grey represent stable genes across tumors and normal samples. [Full-size](#)  DOI: 10.7717/peerj.5207/fig-2

(*Krishnan et al., 2015, 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 ([Table S2](#)). In the validation study, *RPL30*, *RPL27*, *PSMC5*, *OAZ1*, and *MTCH1* came as the top five 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*, and *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 gotten 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 (*Popovici et al., 2009*), 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
NTC	No template control
NAC	No amplification control
NPC	No primer control.

ACKNOWLEDGEMENTS

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

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

Research presented in this article is funded by Department of Electronics and Information Technology, Government of India (Ref No: 18(4)/2010-E-Infra., 31-03-2010) and Department of IT, BT and ST, Government of Karnataka, India (Ref No: 3451-00-090-2-22). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:

Department of Electronics and Information Technology, Government of India:
18(4)/2010-E-Infra.

Department of IT, BT and ST, Government of Karnataka, India: 3451-00-090-2-22.

Competing Interests

The authors declare that they have no competing interests.

Author Contributions

- Vinayak Palve conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, approved the final draft.
- Manisha Pareek performed the experiments, approved the final draft.
- Neeraja M. Krishnan analyzed the data, authored or reviewed drafts of the paper, approved the final draft.
- Gangotri Siddappa contributed reagents/materials/analysis tools, approved the final draft.
- Amritha Suresh contributed reagents/materials/analysis tools, approved the final draft.

- Moni A. Kuriakose contributed reagents/materials/analysis tools, approved the final draft.
- Binay Panda conceived and designed the experiments, prepared figures and/or tables, authored or reviewed drafts of the paper, approved the final draft.

Human Ethics

The following information was supplied relating to ethical approvals (i.e., approving body and any reference numbers):

The Mazumdar-Shaw Cancer Centre Ethics Committee granted approval for the study (IRB Approval number: NHH/MEC-CL/2014/197).

Data Availability

The following information was supplied regarding data availability:

The raw data are provided in the [Supplemental Files](#).

Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj.5207#supplemental-information>.

REFERENCES

- Andersen CL, Jensen JL, 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**(15):5245–5250 DOI [10.1158/0008-5472.can-04-0496](https://doi.org/10.1158/0008-5472.can-04-0496).
- Bär M, Bär D, Lehmann B. 2009.** Selection and validation of candidate housekeeping genes for studies of human keratinocytes—review and recommendations. *Journal of Investigative Dermatology* **129**(3):535–537 DOI [10.1038/jid.2008.428](https://doi.org/10.1038/jid.2008.428).
- Bustin SA, Beaulieu JF, Huggett J, Jaggi R, Kibenge FS, Olsvik PA, Penning LC, Toegel S. 2010.** MIQE precis: practical implementation of minimum standard guidelines for fluorescence-based quantitative real-time PCR experiments. *BMC Molecular Biology* **11**(1):74 DOI [10.1186/1471-2199-11-74](https://doi.org/10.1186/1471-2199-11-74).
- Bustin SA, Benes V, Garson JA, Hellemans J, Huggett J, Kubista M, Mueller R, Nolan T, Pfaffl MW, Shipley GL, Vandesompele J, Wittwer CT. 2009.** The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clinical Chemistry* **55**(4):611–622 DOI [10.1373/clinchem.2008.112797](https://doi.org/10.1373/clinchem.2008.112797).
- Campos MS, Rodini CO, Pinto-Junior DS, Nunes FD. 2009.** GAPD and tubulin are suitable internal controls for qPCR analysis of oral squamous cell carcinoma cell lines. *Oral Oncology* **45**(2):121–126 DOI [10.1016/j.oraloncology.2008.03.019](https://doi.org/10.1016/j.oraloncology.2008.03.019).
- 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. 2007.** Evidence based selection of housekeeping genes. *PLOS ONE* **2**:e898 DOI [10.1371/journal.pone.0000898](https://doi.org/10.1371/journal.pone.0000898).
- Du P, Kibbe WA, Lin SM. 2008.** *lumi*: a pipeline for processing Illumina microarray. *Bioinformatics* **24**(13):1547–1548 DOI [10.1093/bioinformatics/btn224](https://doi.org/10.1093/bioinformatics/btn224).
- Eisenberg E, Levanon EY. 2013.** Human housekeeping genes, revisited. *Trends in Genetics* **29**(10):569–574 DOI [10.1016/j.tig.2013.05.010](https://doi.org/10.1016/j.tig.2013.05.010).

- Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM. 2010. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *International Journal of Cancer* 127(12):2893–2917 DOI 10.1002/ijc.25516.
- Glare E, Divjak M, Bailey MJ, Walters EH. 2002. β -Actin and GAPDH housekeeping gene expression in asthmatic airways is variable and not suitable for normalising mRNA levels. *Thorax* 57(9):765–770 DOI 10.1136/thorax.57.9.765.
- Greer S, Honeywell R, Geletu M, Arulanandam R, Raptis L. 2010. Housekeeping genes; expression levels may change with density of cultured cells. *Journal of Immunological Methods* 355(1–2):76–79 DOI 10.1016/j.jim.2010.02.006.
- Janssens N, Janicot M, Perera T, Bakker A. 2004. Housekeeping genes as internal standards in cancer research. *Molecular Diagnosis* 8(2):107–113 DOI 10.2165/00066982-200408020-00005.
- Johnson WE, Li C, Rabinovic A. 2007. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* 8(1):118–127 DOI 10.1093/biostatistics/kxj037.
- Krishnan NM, Dhas K, Nair J, Palve V, Bagwan J, Siddappa G, Suresh A, Kekatpure VD, Kuriakose MA, Panda B. 2016. A minimal DNA methylation signature in oral tongue squamous cell carcinoma links altered methylation with tumor attributes. *Molecular Cancer Research* 14(9):805–819 DOI 10.1158/1541-7786.mcr-15-0395.
- 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, 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. *F1000Research* 4:1215 DOI 10.12688/f1000research.7302.1.
- Lallemant B, Evrard A, Combescure C, Chapuis H, Chambon G, Raynal C, Reynaud C, Sabra O, Joubert D, Hollande F. 2009. Reference gene selection for head and neck squamous cell carcinoma gene expression studies. *BMC Molecular Biology* 10(1):78 DOI 10.1186/1471-2199-10-78.
- Martin JL. 2016. Validation of reference genes for oral cancer detection panels in a prospective blinded cohort. *PLOS ONE* 11(7):e0158462 DOI 10.1371/journal.pone.0158462.
- Oliveira JG, Prados RZ, Guedes ACM, Ferreira PC, 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(12):659–661 DOI 10.1007/s004030050471.
- Pfaffl MW, Tichopad A, Prgomet C, 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(6):509–515 DOI 10.1023/b:bile.0000019559.84305.47.
- Popovici V, Goldstein DR, Antonov J, Jaggi R, Delorenzi M, Wirapati P. 2009. Selecting control genes for RT-QPCR using public microarray data. *BMC Bioinformatics* 10(1):42 DOI 10.1186/1471-2105-10-42.
- 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, Suresh A. 2016. Meta-analyses of microarray datasets identifies ANO1 and FADD as prognostic markers of head and neck cancer. *PLOS ONE* 11(1):e0147409 DOI 10.1371/journal.pone.0147409.
- Schmittgen TD, Livak KJ. 2008. Analyzing real-time PCR data by the comparative CT method. *Nature Protocols* 3(6):1101–1108 DOI 10.1038/nprot.2008.73.

- Selvey S, Thompson EW, Matthaei K, Lea RA, Irving MG, Griffiths LR. 2001.** β -Actin—an unsuitable internal control for RT-PCR. *Molecular and Cellular Probes* **15**(5):307–311 DOI [10.1006/mcpr.2001.0376](https://doi.org/10.1006/mcpr.2001.0376).
- Silver N, Best S, Jiang J, Thein SL. 2006.** Selection of housekeeping genes for gene expression studies in human reticulocytes using real-time PCR. *BMC Molecular Biology* **7**:33 DOI [10.1186/1471-2199-7-33](https://doi.org/10.1186/1471-2199-7-33).
- The Cancer Genome Atlas Network. 2015.** Comprehensive genomic characterization of head and neck squamous cell carcinomas. *Nature* **517**(7536):576–582 DOI [10.1038/nature14129](https://doi.org/10.1038/nature14129).
- Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, Speleman F. 2002.** Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biology* **3**(7):research0034.1–research0034.11.
- Xie F, Xiao P, Chen D, Xu L, Zhang B. 2012.** miRDeepFinder: a miRNA analysis tool for deep sequencing of plant small RNAs. *Plant Molecular Biology* **80**(1):75–84 DOI [10.1007/s11103-012-9885-2](https://doi.org/10.1007/s11103-012-9885-2).
- Zhong H, 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**(3):523–526 DOI [10.1006/bbrc.1999.0815](https://doi.org/10.1006/bbrc.1999.0815).