OPEN

Validation of Reference Genes for Normalization of Relative qRT-PCR Studies in Papillary Thyroid Carcinoma

S. Adeleh Razavi[1,2,3], Mandana Afsharpad[4], Mohammad Hossein Modarressi[5], Maryam Zarkesh[1], Parichehreh Yaghmaei[3], Shirzad Nasiri[6], S. Mohammad Tavangar[7], Hanieh Gholami[8], Afsoon Daneshafrooz[1] & Mehdi Hedayati [1]*

Quantitative reverse transcription polymerase chain reaction (qRT-PCR) in thyroid tumors require accurate data normalization, however, there are no sufficient studies addressing the suitable reference genes for gene expression analysis in malignant and normal thyroid tissue specimens. The purpose of this study was to identify valid internal control genes for normalization of relative qRT-PCR studies in human papillary thyroid carcinoma tissue samples. The expression characteristics of 12 candidate reference genes (GAPDH, ACTB, HPRT1, TBP, B2M, PPIA, 18SrRNA, HMBS, GUSB, PGK1, RPLP0, and PGM1) were assessed by qRT-PCR in 45 thyroid tissue samples (15 papillary thyroid carcinoma, 15 paired normal tissues and 15 multinodular goiters). These twelve candidate reference genes were selected by a systematic literature search. GeNorm, NormFinder, and BestKeeper statistical algorithms were applied to determine the most stable reference genes. The three algorithms were in agreement in identifying GUSB and HPRT1 as the most stably expressed genes in all thyroid tumors investigated. According to the NormFinder software, the pair of genes including 'GUSB and HPRT1' or 'GUSB and HMBS' or 'GUSB and PGM1' were the best combinations for selection of pair reference genes. The optimal number of genes required for reliable normalization of qPCR data in thyroid tissues would be three according to calculations made by GeNorm algorithm. These results suggest that GUSB and HPRT1 are promising reference genes for normalization of relative qRT-PCR studies in papillary thyroid carcinoma.

Among the methods used for gene expression studies including Northern blotting, microarrays, serial analysis of gene expression (SAGE), ribonuclease protection assays (RPAs) and quantitative reverse transcription polymerase chain reaction (qRT-PCR), the last one has been employed more than the others because of its sensitivity, specificity and low template requirements[1]. qRT-PCR gives more quantitative results and it is also easy and more convenient to use when compared to other methods. In the relative approach of qRT-PCR, it is necessary to apply a normalizing gene as an internal control to correct the differences between the compared samples in order to achieve the most accurate and reliable results[2]. Normalizing genes usually are the basic metabolism genes named housekeeping genes (HKGs) that are frequently used in studies. However, a single reference gene cannot be used for all experiments in all conditions because contrary to expectations, HKGs expression levels are affected by the

[1]Cellular and Molecular Endocrine Research Center, Research Institute for Endocrine Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran. [2]Department of Research and Development (R&D), Saeed Pathobiology & Genetics Laboratory, Tehran, Iran. [3]Department of Biology, Faculty of Basic Sciences, Science and Research Branch, Islamic Azad University, Tehran, Iran. [4]Cancer Control Research Center, Cancer Control Foundation, Iran University of Medical Sciences, Tehran, Iran. [5]Department of Medical Genetics, School of Medicine, Tehran University of Medical Sciences, Tehran, Iran. [6]Department of Surgery, Shariati Hospital, School of Medicine, Tehran University of Medical Sciences, Tehran, Iran. [7]Department of Pathology, Shariati Hospital, School of Medicine, Tehran University of Medical Sciences, Tehran, Iran. [8]Endocrine Physiology Research Center, Research Institute for Endocrine Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran. *email: hedayati47@gmail.com

Gene	M	B	r	R^2	E	%E
GAPDH	−3.15	32.48	0.991	0.982	2.08	%108
ACTB	−3.34	26.07	0.997	0.994	1.99	%99
HPRT1	−3.44	34.39	0.992	0.984	1.95	%95
TBP	−3.04	32.74	0.999	0.998	2.13	%113
B2M	−3.57	28.81	0.999	0.998	1.91	%91
PPIA	−3.29	37.03	0.998	0.996	2.01	%101
18SrRNA	−3.33	21.61	0.994	0.988	2.00	%100
HMBS	−3.53	36.20	0.998	0.997	1.92	%92
GUSB	−3.13	26.45	0.994	0.988	2.08	%108
PGK1	−3.38	36.40	0.999	0.997	1.98	%98
RPLP0	−3.34	33.93	0.977	0.955	1.99	%99
PGM1	−3.59	35.03	0.963	0.927	1.90	%90

Table 1. Standard curve results of selected candidate reference genes: M; slope, B; intercept, r; Pearson's correlation coefficient, R^2; coefficient of determination, E; amplification efficiency, %E; amplification efficiency percentage.

specimen, method, experiment, and environment conditions. Thus, every specific experiment will require selecting the best reference gene to obtain accurate and reliable results[3].

Currently, many cancer researchers investigate tumor marker genes to identify the pathological patterns of diseases and find much needed information about the predicting of the tumor's behavior. However, expression patterns of HKGs may vary in different malignant tissues, different malignant cell subtypes, or even in the same type of carcinoma. This illustrates the difference in the personalized metabolism drivers of cancer cells when different cases are compared[2].

Although the incidence rate of thyroid cancer is currently low but increasing, and it is estimated that it will be doubled and become the third most common cancer in women by 2019[4]. There are four main histopathological types of thyroid cancers including papillary thyroid carcinoma (PTC), follicular thyroid carcinoma (FTC), anaplastic thyroid carcinoma (ATC) with the origin of thyroid follicular cells and medullary thyroid carcinoma (MTC) with the origin of thyroid parafollicular cells. PTC is the most common and least aggressive histological type of thyroid cancers that its annual incidence rate is 0.29 per 100,000 persons in Iran[5].

The most important diagnostic method for thyroid cancers is fine-needle aspiration biopsy (FNAB), however, 15–30% of thyroid FNABs are cytologically indeterminate[6]. For overcoming this problem, molecular based techniques are suggested to discover new molecular markers discriminating malignancy from benignity[7]. One of the most common methods for the investigation of thyroid tumor markers is qRT-PCR; however, there are no sufficient studies introducing the suitable reference genes for gene expression analysis in malignant and normal thyroid tissue samples. Definitely, without validation of reference genes, the results of gene expression studies in thyroid cancers are less reliable due to such unexpected behavior of HKGs. Therefore it is essential to determinate the genes with the highest level of expression stability for the improvement of relative qRT-PCR in further thyroid cancer research. The purpose of this study was to identify valid internal control genes for normalization of relative qRT-PCR studies in human PTC tissue samples.

Results

Quality of RNA samples. A260 and A280 were used to assess the concentration and purity of the isolated RNA. The mean and standard deviation (mean (ng/μl) ± SD) of A260/A280 ratio in PTC, paired normal tissues (PNT), and multinodular goiters (MNG) groups were 1.85 ± 0.15, 1.80 ± 0.12 and 1.93 ± 0.09 respectively.

Amplification efficiency and specificity. To determine the amplification efficiency, serial dilutions of several cDNA templates were run for all 12 paired primers, and the results were used to depict a standard curve. The equation of the linear regression line, along with Pearson's correlation coefficient (r) and the coefficient of determination (R^2), confirmed the qPCR optimization (Table 1, Supplemental Fig. 1). The melt-curve analysis of all 12 qPCR assays exhibited a single signal peak represented the specific product (Supplemental Fig. 2).

Candidate reference genes expression. Expression levels of the 12 reference genes in thyroid tissues are shown for PTC, PNT, and MNG groups in Fig. 1. For each HKG, the Ct coefficient of variation (CtCV%) value was calculated using the following formula: CtCV% = SD/mean × 100% (Table 2). According to these data, 18SrRNA showed the higher CtCV%, which means the greater level of dispersion around the mean. HMBS had the lower value of CtCv% which means the more precise estimation.

Identification of the most stable reference genes. *NormFinder algorithm.* This algorithm merges group division, the absolute copy number of the gene, and the random expression variations (stability value) caused by biological and experimental factors. Then, it ranks the reference genes in order; the lower the stability value the reliable the reference gene[8]. The five most stable HKGs in each studied group ranked by NormFinder (from most stable to least stable) were as follows: (i) Total: GUSB, HPRT1, PGM1, TBP, RPLP0; (ii) PTC vs. PNT: HPRT1, GUSB, HMBS, GAPDH, PPIA; (iii) PTC vs. MNG: GUSB, RPLP0, HMBS, PGK1, PGM1; (iv) PNT vs.
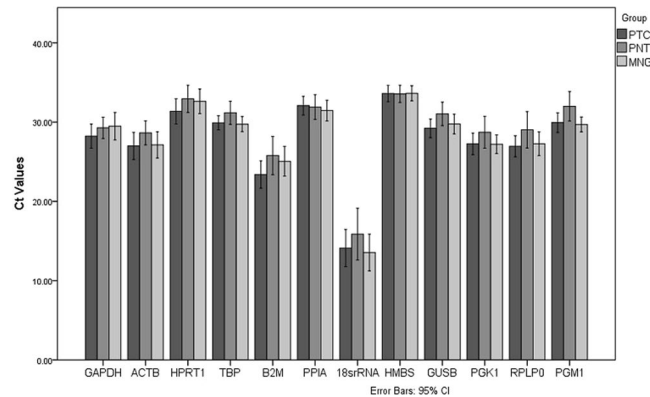
Figure 1. The mean Ct values and standard deviations of 12 candidate reference genes in papillary thyroid carcinoma (PTC), paired normal tissue (PNT) and multinodular goiter (MNG) groups.

Gene	CtCV%			
	PTC	PNT	MNG	Total
GAPDH	9.7%	8.3%	10.6%	9.6%
ACTB	11.4%	9.6%	11.1%	10.8%
HPRT1	9.2%	9.4%	8.6%	9.1%
TBP	5.4%	8.5%	5.8%	6.9%
B2M	13.3%	17.0%	13.5%	15.0%
PPIA	6.6%	8.8%	7.5%	7.6%
18SrRNA	30.0%	37.2%	30.9%	33.4%
HMBS	5.5%	5.8%	5.1%	5.4%
GUSB	7.3%	8.6%	7.6%	8.1%
PGK1	9.0%	12.6%	7.9%	10.2%
RPLP0	8.9%	14.3%	9.9%	11.6%
PGM1	7.5%	10.5%	5.7%	8.8%

Table 2. The Ct coefficient of variation (CtCV%) values of 12 candidate reference genes in papillary thyroid carcinoma (PTC), paired normal tissue (PNT), multinodular goiter (MNG), and total tissues.

MNG: GUSB, HPRT1, GAPDH, ACTB, HMBS (Fig. 2A). According to the NormFinder algorithm, the best combination of two HKGs are 'GUSB and HPRT1' or 'GUSB and HMBS' or 'GUSB and PGM1' (Table 3).

GeNorm algorithm. This algorithm calculates a stability value, M, as the average pair-wise variation of each reference gene in relation to all the other reference genes led to the elimination of the least stable gene. The process is followed by recalculation of the M values resulting in the ranking of the most stable genes. HKG with the lower M value has a higher expression stability[9]. Using GeNorm the five most stable HKGs evaluated (from most stable to least stable) were: (i) Total: GUSB, HPRT1, HMBS, TBP, GAPDH; (ii) PTC vs. PNT: HPRT1, GUSB, GAPDH, TBP, HMBS; (iii) PTC vs. MNG: GUSB, HPRT1, HMBS, RPLP0, PGM1; (iv) PNT vs. MNG: HPRT1, GUSB, GAPDH, HMBS, TBP (Fig. 2B). The GeNorm algorithm was also used to calculate the optimal number of reference genes required for a reliable normalization using a normalization factor (V_{NF} value). According to GeNorm, a combination of three reference genes was the optimal number of genes required for reliable normalization of qPCR data in thyroid tissues (Fig. 3).

BestKeeper algorithm. This algorithm computes the average Ct values, standard deviation (SD), and the coefficient of variation (CV) for each gene and SD > 1 is regarded as inappropriate for use as a stable reference gene. Then the program estimates the geometric mean of the stable control genes' Ct values and creates pair-wise correlations between this index and each gene. HKG with the higher Pearson's correlation coefficient value is considered as the most stable gene[10]. This tool has a limitation that it has been designed only for 10 genes, therefore, following the NormFinder and GeNorm analyses, the two least stable genes (B2M and 18SrRNA) were omitted then the algorithm was utilized. According to the BestKeeper, the five most stable HKGs in the groups studied (from most stable to least stable) were as follows: (i) Total: HPRT1, ACTB, GUSB, GAPDH, RPLP0; (ii) PTC vs. PNT: HPRT1, ACTB, PGM1, GUSB, GAPDH; (iii) PTC vs. MNG: GUSB, HPRT1, HMBS, RPLP0, ACTB; (iv) PNT vs. MNG: HPRT1, ACTB, GUSB, GAPDH, RPLP0 (Fig. 2C).
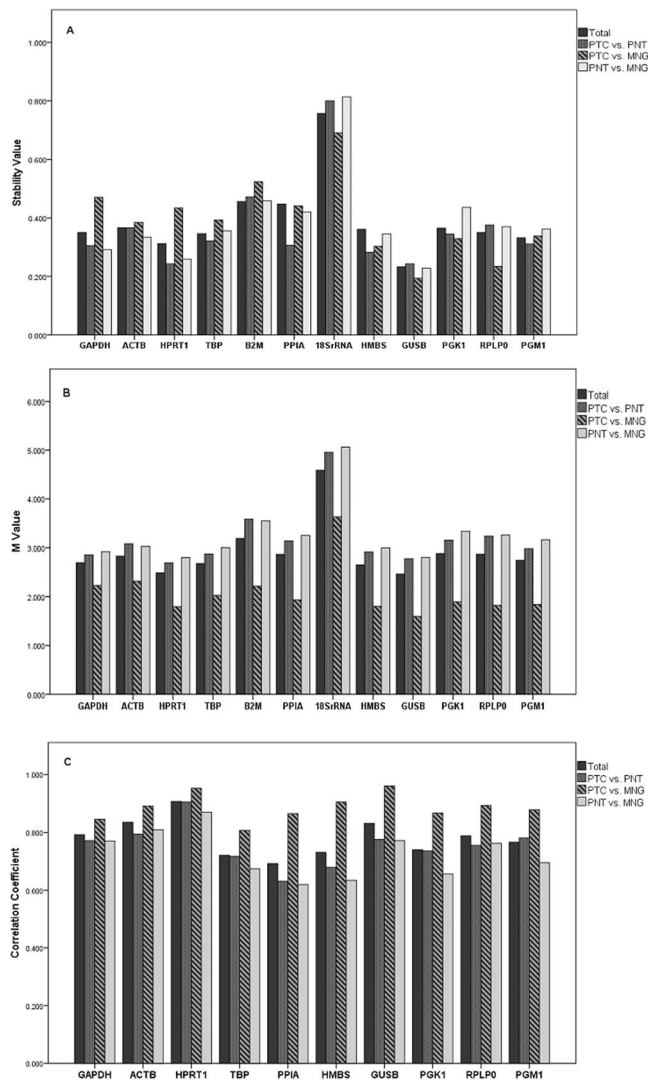
Figure 2. Candidate housekeeping gene's stability results. (**A**) The NormFinder results. The lower stability value corresponds to more stably expressed gene. (**B**) The GeNorm results. The lower M-value corresponds to more stably expressed gene. (**C**) The BestKeeper results. The higher correlation coefficient corresponds to more stably expressed gene. PTC; papillary thyroid carcinoma, PNT; paired normal tissue, MNG; multinodular goiter.

	NormFinder		GeNorm	BestKeeper
Group	**Single HKG**	**Combined HKGs**	**Single HKG**	**Single HKG**
Total	GUSB	HMBS & GUSB	GUSB	HPRT1
PTC vs. PNT	HPRT1	HPRT1 & GUSB	HPRT1	HPRT1
PTC vs. MNG	GUSB	GUSB & PGM1	GUSB	GUSB
PNT vs. MNG	GUSB	HPRT1 & GUSB	HPRT1	HPRT1

Table 3. The most stable housekeeping genes by NormFinder, GeNorm and BestKeeper. HKG; housekeeping gene, MNG; multinodular goiter, PNT; paired normal tissue, PTC; papillary thyroid carcinoma.

Discussion

In order to monitor changes occurring in the pattern of PTC gene expression for the biomarkers identification, confirming stable and reliable internal reference genes for relative qRT-PCR is highly crucial. The ideal reference gene in thyroid tissue specimens should provide constant transcriptions at any types of thyroid tissue and at any time of the cell cycle or cellular differentiation. However, it is a desire and such as that HKG may probably not exist at all, so we can only find genes with the least expressed variation.

To select the suitable internal reference genes for relative quantification analysis in PTC, we chose a panel of 11 candidate HKGs by a systematic search then added PGM1 to the panel because of its stability according to our
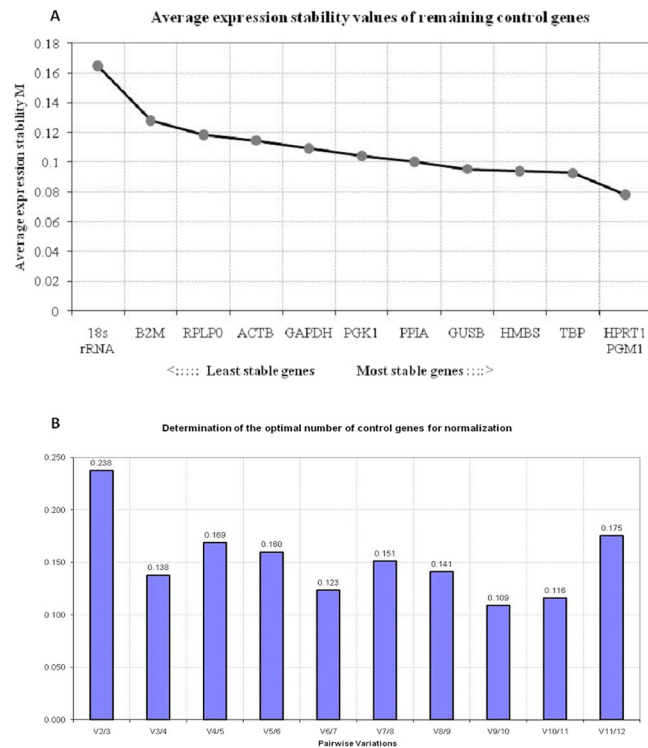
Figure 3. Selection of the most suitable reference genes using the GeNorm algorithm. (**A**) Average expression stability value (M) of housekeeping genes calculated by stepwise exclusion of the least stable genes. (**B**) Determination of the optimal number of housekeeping genes required for reliable normalization of qPCR data in thyroid tissues based on the pairwise variation values.

previous studies[11,12]. These genes were investigated in three types of thyroid lesion tissues including PTC, PNT, and MNG. The data obtained were analyzed by the three most commonly used software programs (NormFinder, GeNorm, and BestKeeper) for analyzing gene expression stability. Following the analysis, NormFinder and GeNorm selected the GUSB, and BestKeeper documented the HPRT1 as the most stable gene in the total sample group. All three algorithms recorded HPRT1 as the most stably expressed gene in PTC vs. PNT group. The three software recommended that GUSB is the most stable HKG in PTC vs. MNG. Also, NormFinder selected the GUSB, and GeNorm and BestKeeper selected the HPRT1 as the most stable gene in PNT vs. MNG group. Although, the three software programs use different mathematical models for the identification of the most stable reference gene, there were only minor variations in the selection of the single most stable reference gene by the three algorithms.

To date, only a few groups performed similar HKGs selection studies on human thyroid tissues or cells. In previous studies conducted by Weber *et al*. and Santin *et al*., ACTB was introduced as the best single reference gene among several other candidates tested[13,14]. Weber *et al*. evaluated 6 reference genes (ACTB, B2M, GAPDH, HPRT1, SDHA, and YWHAZ) in 14 thyroid specimens (7 normal thyroid tissues and 7 goitrous tissues) using the NormFinder algorithm[13]. Santin *et al*. evaluated 4 reference genes (ACTB, B2M, GAPDH, and TBP) in primary culture cells obtained from normal human thyroid tissue treated with progesterone or estradiol and chose NormFinder as an analyzer algorithm[14]. Contrary to these two studies, ACTB was ranked at the middle of the stable HKGs list in our study. The different results in studies are somewhat similar confirm that there is no universal or gold standard reference gene for all thyroid tissue samples because even highly stable expressed HKGs are prone to be affected by ethnicity and genetic status of examined tissues. On the other hand, diet, life style, and physical activity could have epigenetic effects influencing the expression of genes. In the case of cancerous tissues, due to the heterogeneity at the molecular, cellular and tissue levels, different results in similar studies may be obtained. However, pathological states of tissues and materials and methods used for an individual experiment are the other possible factors for inconsistency between our results and what the others have shown in previous studies. These different observations emphasize the need for determination of stably expressed reference genes in particular thyroid tissues and/or particular thyroid cell subtypes. Several studies on primary thyroid cancers and thyroid cancer-derived cell lines have employed GAPDH as a reliable internal reference gene because the transcription of GAPDH is stable in most experimental conditions[15–17]. However, the results of present investigation demonstrated GAPDH was not listed within the highest stable genes in the tissues and experimental setup used here. On the other hand, GAPDH has pseudogene, and it may amplify genomic DNA that lead to a false estimation of GAPDH expression[18].

In our study conditions, using the three algorithms, GUSB and HPRT1 were identified as the most stably expressed genes in paired malignant and nonmalignant thyroid tissue samples. It is noteworthy that the results of

the three software programs were close to each other. Therefore, since the rankings of the candidate HKGs were similar, our data suggest that GUSB and HPRT1 could be considered as the optimal internal reference genes for normalizing the relative gene quantitation in PTC studies. However, these findings are completely different from those found by Chantawibul *et al*. They demonstrated that GAPDH was the most and HPRT1 was the least stably expressed genes in thyroid tissues[19]. Although the stability value obtained with NormFinder was extraordinary high in Chantawibul *et al*. study, the differences may be attributed to the experimental factors and the sample selection included (13 goiters, 6 adenomas, 2 follicular carcinomas, 2 papillary carcinomas, and 2 lymphocytic thyroiditis).

At present, usage of the geometric mean of several reference genes rather than a single one is widely accepted to reduce the influence of fluctuations. Minimum Information for Publication of Quantitative Real-Time PCR Experiments (MIQE) recommends that the optimal number and the utility of HKGs must be validated for particular tissues and specific experimental conditions[20]. According to our study, GeNorm program suggested a combination of 3 most stable reference genes for gene expression studies in thyroid lesions. NormFinder software program found that the combination of GUSB and HMBS or HPRT1 or PGM1 were the best combinations.

In comparison to previous study, the results of current study provide promising evidence identifying potential HKGs for gene expression analyses in PTC. However, we stored the thyroid tissue samples in RNA later which is a commercial product and is not routinely used in investigations. Therefore, our results might not be similar to the samples were flash frozen and stored at −80 °C. For these types of stored thyroid tissues, the utility of reference genes must be validated. On the other hand, our study was limited to the most frequent type of thyroid cancer, PTC; however, there are several histological types and subtypes of thyroid cancer with different cellular origins, characteristics and prognoses including FTC and ATC (follicular thyroid cell-derived tumors), and MTC (parafollicular C cell-derived tumors). For these types of thyroid tissues, the stable reference genes must be investigated.

Conclusions

GUSB and HPRT1 evidenced as the most stable reference genes when a panel of 12 candidate reference genes were compared in PTC, PNT, and MNG specimens stored in RNA later. These genes can be considered as single reference genes for normalization of relative qRT-PCR studies in PTC. However, it is suggested to utilize the validated combination of 'GUSB and HPRT1' or 'GUSB and HMBS' or 'GUSB and PGM1' as a much more reliable normalization strategy.

Materials and Methods

Literature review for selection of candidate reference genes.   A systematic literature search was performed in Web of Science and PubMed databases using the following key words: 'reference genes' or 'housekeeping genes', and 'cancer' or 'carcinoma' or 'malignancy' or 'neoplasm' in English and in all years. The initial search visualized 123 articles. After applying the inclusion criteria (Fig. 4) 34 articles were included in the study[21–54]. According to the usage rate, 11 candidate reference genes were selected for validation in the study. The phosphoglucomutase 1 (PGM1) gene was also added to the list because of its stability that was exhibited in our previous RT-qPCR studies[11,12]. All selected reference genes are listed in Table 4.

Tissue samples.   Patients who had undergone near-total or total thyroidectomy from November 2015 to August 2016 in Shariati Hospital, Tehran Iran were initially enrolled in the study. The present study was approved by the Institutional Review Board and Ethics Committee of Research Institute for Endocrine Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran (Ethical approval number: IR.SBMU.REC.1397.110) and written informed consent was obtained from all participants.

Following the thyroid excision, part of the tumor and adjacent tissues were immediately treated in RNAlater® RNA Stabilization Reagent (Qiagen, Hilden, Germany) prior to further processing. Before archival storage at −80 °C, the tissues were incubated overnight in RNAlater at 2–8 °C, and then they were removed from the reagent and transferred to −80 °C until tested.

According to the postoperative pathological reports and the histological confirmation, a total of 45 thyroid tissue samples taken from 30 thyroid gland consisting of 15 PTC, 15 PNT belonged to the patients with PTC tumors, and 15 goiterus tissues belonged to the patients with MNG were selected. The approximate size of each sample was 0.5 cm. The pathological characteristics of the tissues used in the study are summarized in Table 5. Tumor staging was determined by the 7th edition of the American Joint Committee on Cancer Tumor-Node-Metastasis (TNM) staging system[55].

RNA extraction.   Total RNA was extracted from each tissue sample using TRIzol reagent (Invitrogen Life Technologies, Carlsbad, CA, USA), according to the manufacturer's instructions.

DNase I was used to eliminate residual genomic DNA contamination. The quantity and quality of the isolated RNA were measured using NanoDropND-1000 spectrophotometer (ThermoScientific, Waltham, MA, USA). The purity of total RNA was determined using the A260/A280 ratio. The integrity of the RNA samples was confirmed by electrophoresis on 1.0% agarose gel (Ultra-Pure™ Agarose; Invitrogen).

Complementary DNA (cDNA) synthesis.   Three μg of total RNA was reversely transcribed to cDNA using Thermo Scientific RevertAid Reverse Transcriptase (1 μL of 200 U/μL), random hexamer primers (1 μl of 100 μM), dNTPS (2 μL of 10 mM), and RiboLock RNase-inhibitor (0.5 μL of 40 U/μL), incubated for 10 min at 25 °C, followed by 60 min at 42 °C in a total volume of 20 μL. The reaction was terminated by heating the reactions at 70 °C for 10 min.
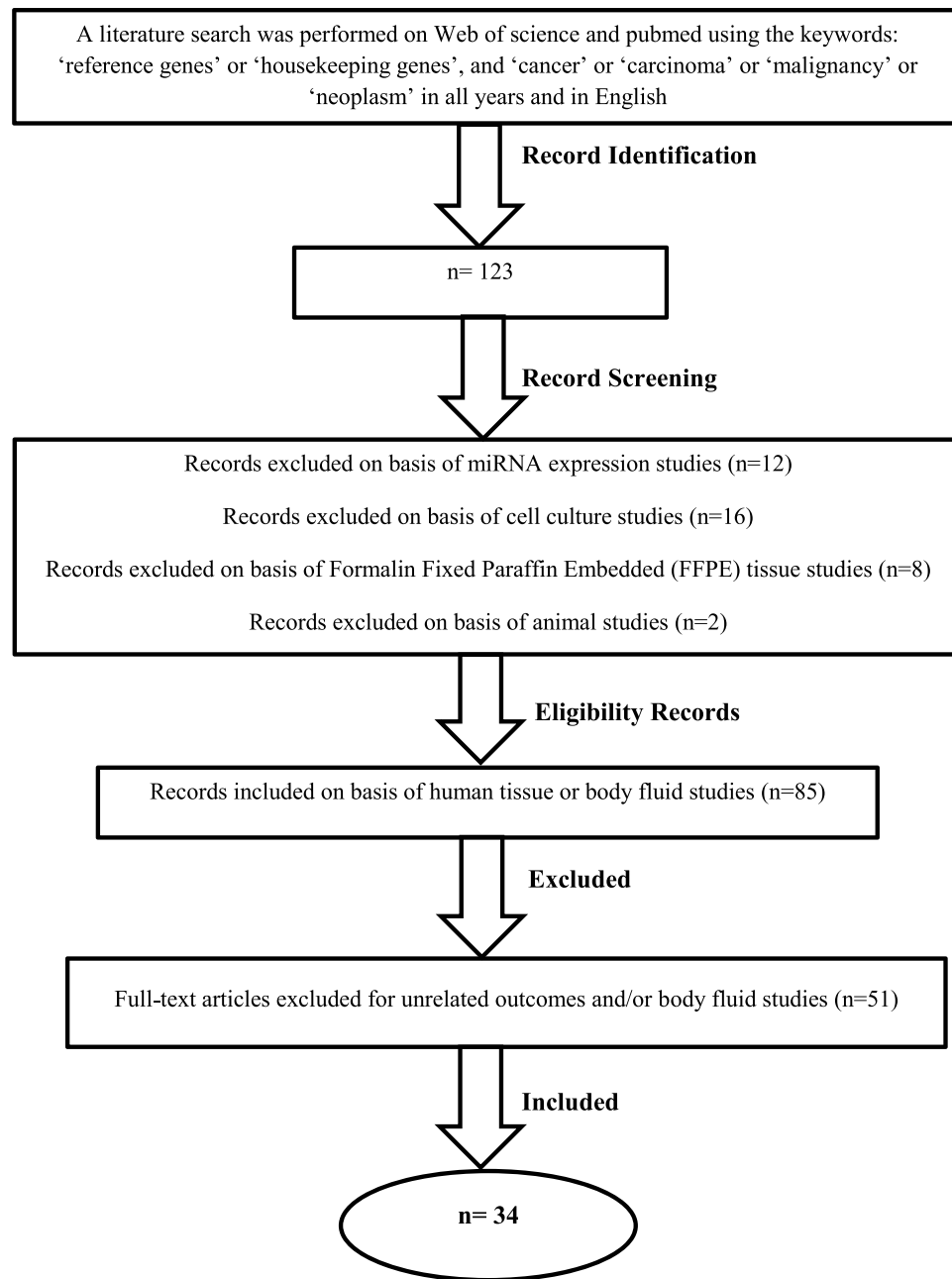
A literature search was performed on Web of science and pubmed using the keywords: 'reference genes' or 'housekeeping genes', and 'cancer' or 'carcinoma' or 'malignancy' or 'neoplasm' in all years and in English

Record Identification

n= 123

Record Screening

Records excluded on basis of miRNA expression studies (n=12)

Records excluded on basis of cell culture studies (n=16)

Records excluded on basis of Formalin Fixed Paraffin Embedded (FFPE) tissue studies (n=8)

Records excluded on basis of animal studies (n=2)

Eligibility Records

Records included on basis of human tissue or body fluid studies (n=85)

Excluded

Full-text articles excluded for unrelated outcomes and/or body fluid studies (n=51)

Included

n= 34

Figure 4. Review flow-chart for selection of included articles.

Quantitative polymerase chain reaction (qPCR). Primer sequences of 11 candidate reference genes were designed using the Primer 3 and Gene Runner software. Primer sequences of PGM1 were taken from the previously published study[11] (Table 4). qPCRs were performed using the rotor gene 6000 real-time PCR machine (Corbett, Life science, Sydney, Australia). All reactions were set up for total volumes of 10 μL which contained 1 μL cDNA, 0.5 μL of each forward and reverse primer, 3.75 μL of SYBR Green PCR Master Mix 2X (ThermoFisher, USA), and 4.25 μL of nuclease-free water. The cycling profile used for PCR reactions was as follows: 10 min at 95 °C for initial denaturation, followed by 40 cycles with 15 s at 95 °C, 45 s at 60 °C and 40 s at 72 °C. All samples were run in triplicate and no cDNA template samples were used as negative controls.

The specificity of the real-time PCR reactions was verified by the melt curves analysis. For the PCRs efficiency investigation, several cDNA samples were selected randomly and were diluted with a 10-fold serial dilution ranging from 1X to 100,000X. Slope, intercept, Pearson's correlation coefficient, coefficient of determination, and amplification efficiency were calculated by the Rotor-Gene 6000 Series Software (Table 1).

Statistical analysis. NormFinder[8], GeNorm[9], and BestKeeper[10] statistical algorithm, the three frequently used software programs, were applied to determine the most stable reference genes. For further statistical data analysis, SPSS 20.0 software (Chicago, IL, USA) was used.

	Gene symbol	Gene ID	Official full name	Number of used/34 papers	% of used	Primer sequence (5′ → 3′)	Amplicon size (bp)
1	GAPDH	2597	glyceraldehyde-3-phosphate dehydrogenase	31	91.2	F: CTGCTCCTCCTGTTCGACAGT R: CCGTTGACTCCGACCTTCAC	100
2	ACTB	60	actin beta	29	85.3	F: GATCAAGATCATTGCTCCTCCT R: TACTCCTGCTTGCTGATCCA	108
3	HPRT1	3251	hypoxanthine phosphoribosyltransferase 1	27	79.4	F: CCCTGGCGTCGTGATTAGTG R: CACCCTTTCCAAATCCTCAGC	91
4	TBP	6908	TATA-box binding protein	23	67.6	F: CACGAACCACGGCACTGATT R: TTTTCTTGCTGCCAGTCTGGAC	89
5	B2M	567	beta-2-microglobulin	22	64.7	F: TGCCTGCCGTGTGAACCATG R: TGCGGCATCTTCAAACCTCCA	97
6	PPIA	5478	peptidylprolyl isomerase A	18	52.9	F: GCTGTGAGGAGGTACTGCTTG R: CCTGAGAAACCAAGTCCTTAGTG	145
7	18SrRNA	100008588	RNA, 18S ribosomal N5	17	50	F: GGCCCTGTAATTGGAATGAGTC R: CCAAGATCCAACTACGAGCTT	146
8	HMBS	3145	hydroxymethylbilane synthase	16	47.1	F: GCACCCACACACAGCCTACT R: GTACCCACGCGAATCACTCTC	108
9	GUSB	2990	glucuronidase beta	13	38.2	F: CGCCCTGCCTATCTGTATTC R: TCCCCACAGGGAGTGTGTAG	91
10	PGK1	5230	phosphoglycerate kinase 1	11	32.4	F: AGCGGGTCGTTATGAGAGTC R: TGGTACAGCAGCCTTAATCC	86
11	RPLP0	6175	ribosomal protein lateral stalk subunit P0	11	32.4	F: CCTCGTGGAAGTGACATCGT R: CTGTCTTCCCTGGGCATCAC	76
12	PGM1	5236	phosphoglucomutase 1	—	—	F: AGCATTCCGTATTTCCAGCAG R: GCCAGTTGGGGTCTCATACAAA	120

Table 4. List of candidate reference genes used in the study. The eleven most widely used reference genes in tissue cancer studies are ranked from 1 to 11.

Parameter	PTC	PNT	MNG	Total
Tissue number	15	15	15	45
Histopathology				
multinodular goiter			15	
classic PTC	14	—		—
FVPTC	1			
Tumor size (cm)				
≤2	9	—	—	—
2–4	6			
Invasion				
Negative	10	—	—	—
Positive[a]	5			
Lymph node metastasis[b]				
Negative	7	—	—	—
Positive	8			
TNM stage[c]				
I, II	10	—	—	—
III, IV	5			

Table 5. Pathologic characteristics of the study tissues. FVPTC; follicular variant of papillary thyroid carcinoma, MNG; multinodular goiter, PNT; paired normal tissue, PTC; papillary thyroid carcinoma. [a]Includes 4 extracapsular invasion, and 1 extracapsular and lymphovascular invasions. [b]Includes N1a and/or N1b. [c]American Joint Committee on Cancer (AJCC) Tumor-Node-Metastasis (TNM) staging system.

Ethical approval. The study was approved by the Institutional Review Board and Ethics Committee of Research Institute for Endocrine Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran (Ethical approval number: IR.SBMU.REC.1397.110). All procedures performed in the study involving human participants were in accordance with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

References

1. Dundas, J. & Ling, M. Reference genes for measuring mRNA expression. *Theory in biosciences = Theorie in den Biowissenschaften* **131**, 215–223, https://doi.org/10.1007/s12064-012-0152-5 (2012).
2. Kozera, B. & Rapacz, M. Reference genes in real-time PCR. *Journal of applied genetics* **54**, 391–406, https://doi.org/10.1007/s13353-013-0173-x (2013).
3. Huggett, J., Dheda, K., Bustin, S. & Zumla, A. Real-time RT-PCR normalisation; strategies and considerations. *Genes and immunity* **6**, 279–284, https://doi.org/10.1038/sj.gene.6364190 (2005).
4. Aschebrook-Kilfoy, B. *et al*. The clinical and economic burden of a sustained increase in thyroid cancer incidence. *Cancer epidemiology, biomarkers & prevention: a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* **22**, 1252–1259, https://doi.org/10.1158/1055-9965.epi-13-0242 (2013).
5. Safavi, A., Azizi, F., Jafari, R., Chaibakhsh, S. & Safavi, A. A. Thyroid Cancer Epidemiology in Iran: a Time Trend Study. *Asian Pacific journal of cancer prevention: APJCP* **17**, 407–412 (2016).
6. Ward, L. S. & Kloos, R. T. Molecular markers in the diagnosis of thyroid nodules. *Arquivos brasileiros de endocrinologia e metabologia* **57**, 89–97 (2013).
7. Razavi, S. A., Modarressi, M. H., Yaghmaei, P., Tavangar, S. M. & Hedayati, M. Circulating levels of PTEN and KLLN in papillary thyroid carcinoma: can they be considered as novel diagnostic biomarkers? *Endocrine* **57**, 428–435, https://doi.org/10.1007/s12020-017-1368-4 (2017).
8. Andersen, C. L., Jensen, J. L. & Orntoft, T. F. 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, https://doi.org/10.1158/0008-5472.can-04-0496 (2004).
9. Vandesompele, J. *et al*. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome biology* **3**, Research0034 (2002).
10. Pfaffl, M. W., Tichopad, A., Prgomet, C. & Neuvians, T. P. 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).
11. Mobasheri, M. B., Shirkoohi, R. & Modarressi, M. H. Synaptonemal Complex Protein 3 Transcript Analysis in Breast Cancer. *Iranian journal of public health* **45**, 1618–1624 (2016).
12. Mobasheri, M. B. *et al*. Transcriptome analysis of the cancer/testis genes, DAZ1, AURKC, and TEX101, in breast tumors and six breast cancer cell lines. *Tumour biology: the journal of the International Society for Oncodevelopmental Biology and Medicine* **36**, 8201–8206, https://doi.org/10.1007/s13277-015-3546-4 (2015).
13. Weber, R. *et al*. Validation of reference genes for normalization gene expression in reverse transcription quantitative PCR in human normal thyroid and goiter tissue. *BioMed research international* **2014**, 198582, https://doi.org/10.1155/2014/198582 (2014).
14. Santin, A. P., Souza, A. F., Brum, L. S. & Furlanetto, T. W. Validation of reference genes for normalizing gene expression in real-time quantitative reverse transcription PCR in human thyroid cells in primary culture treated with progesterone and estradiol. *Molecular biotechnology* **54**, 278–282, https://doi.org/10.1007/s12033-012-9565-0 (2013).
15. Ban, Y. *et al*. Proteomic profiling of thyroid papillary carcinoma. *Journal of thyroid research* **2012**, 815079, https://doi.org/10.1155/2012/815079 (2012).
16. Endo, T. & Kobayashi, T. Concurrent overexpression of RET/PTC1 and TTF1 confers tumorigenicity to thyrocytes. *Endocrine-related cancer* **20**, 767–776, https://doi.org/10.1530/erc-13-0310 (2013).
17. Liu, Y. M. *et al*. Expression of HIF-1alpha and HIF-2alpha correlates to biological and clinical significance in papillary thyroid carcinoma. *World journal of surgical oncology* **14**, 30, https://doi.org/10.1186/s12957-016-0785-9 (2016).
18. Sun, Y., Li, Y., Luo, D. & Liao, D. J. Pseudogenes as weaknesses of ACTB (Actb) and GAPDH (Gapdh) used as reference genes in reverse transcription and polymerase chain reactions. *PloS one* **7**, e41659 (2012).
19. Chantawibul, S., Anuwong, A. & Leelawat, K. Validation of appropriate reference genes for gene expression studies in human thyroid gland using real-time RT-PCR. *Journal of the Medical Association of Thailand = Chotmaihet thangphaet* **95**(Suppl 3), S36–40 (2012).
20. Bustin, S. A. *et al*. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clinical chemistry* **55**, 611–622, https://doi.org/10.1373/clinchem.2008.112797 (2009).
21. Ayakannu, T. *et al*. Validation of endogenous control reference genes for normalizing gene expression studies in endometrial carcinoma. *Molecular human reproduction* **21**, 723–735, https://doi.org/10.1093/molehr/gav033 (2015).
22. Bjerregaard, H., Pedersen, S., Kristensen, S. R. & Marcussen, N. Reference genes for gene expression analysis by real-time reverse transcription polymerase chain reaction of renal cell carcinoma. *Diagnostic molecular pathology: the American journal of surgical pathology, part B* **20**, 212–217, https://doi.org/10.1097/PDM.0b013e318212e0a9 (2011).
23. Dupasquier, S. *et al*. Validation of housekeeping gene and impact on normalized gene expression in clear cell renal cell carcinoma: critical reassessment of YBX3/ZONAB/CSDA expression. *BMC molecular biology* **15**, 9, https://doi.org/10.1186/1471-2199-15-9 (2014).
24. Fjeldbo, C. S., Aarnes, E. K., Malinen, E., Kristensen, G. B. & Lyng, H. Identification and Validation of Reference Genes for RT-qPCR Studies of Hypoxia in Squamous Cervical Cancer Patients. *PloS one* **11**, e0156259, https://doi.org/10.1371/journal.pone.0156259 (2016).
25. Fu, L. Y. *et al*. Suitable reference genes for real-time PCR in human HBV-related hepatocellular carcinoma with different clinical prognoses. *BMC cancer* **9**, 49, https://doi.org/10.1186/1471-2407-9-49 (2009).
26. Gao, Q. *et al*. Selection of reference genes for real-time PCR in human hepatocellular carcinoma tissues. *Journal of cancer research and clinical oncology* **134**, 979–986, https://doi.org/10.1007/s00432-008-0369-3 (2008).
27. Gresner, P., Gromadzinska, J. & Wasowicz, W. Reference genes for gene expression studies on non-small cell lung cancer. *Acta biochimica Polonica* **56**, 307–316 (2009).
28. Guo, Y. *et al*. Selection of reliable reference genes for gene expression study in nasopharyngeal carcinoma. *Acta pharmacologica Sinica* **31**, 1487–1494, https://doi.org/10.1038/aps.2010.115 (2010).
29. Ho-Pun-Cheung, A. *et al*. Validation of an appropriate reference gene for normalization of reverse transcription-quantitative polymerase chain reaction data from rectal cancer biopsies. *Analytical biochemistry* **388**, 348–350, https://doi.org/10.1016/j.ab.2009.03.001 (2009).
30. Jung, M. *et al*. In search of suitable reference genes for gene expression studies of human renal cell carcinoma by real-time PCR. *BMC molecular biology* **8**, 47, https://doi.org/10.1186/1471-2199-8-47 (2007).
31. Kilic, Y., Celebiler, A. C. & Sakizli, M. Selecting housekeeping genes as references for the normalization of quantitative PCR data in breast cancer. *Clinical & translational oncology: official publication of the Federation of Spanish Oncology Societies and of the National Cancer Institute of Mexico* **16**, 184–190, https://doi.org/10.1007/s12094-013-1058-5 (2014).
32. Krasnov, G. S. *et al*. [Novel reference gene RPN1 for normalization of quantitative data in lung and kidney cancer]. *Molekuliarnaia biologiia* **45**, 238–248 (2011).
33. Krzystek-Korpacka, M., Diakowska, D., Bania, J. & Gamian, A. Expression stability of common housekeeping genes is differently affected by bowel inflammation and cancer: implications for finding suitable normalizers for inflammatory bowel disease studies. *Inflammatory bowel diseases* **20**, 1147–1156, https://doi.org/10.1097/mib.0000000000000067 (2014).
34. Lallemant, B. *et al*. Reference gene selection for head and neck squamous cell carcinoma gene expression studies. *BMC molecular biology* **10**, 78, https://doi.org/10.1186/1471-2199-10-78 (2009).

35. Li, Y. L., Ye, F., Hu, Y., Lu, W. G. & Xie, X. Identification of suitable reference genes for gene expression studies of human serous ovarian cancer by real-time polymerase chain reaction. *Analytical biochemistry* **394**, 110–116, https://doi.org/10.1016/j.ab.2009.07.022 (2009).

36. Liu, S. *et al.* Selection of reference genes for RT-qPCR analysis in tumor tissues from male hepatocellular carcinoma patients with hepatitis B infection and cirrhosis. *Cancer biomarkers: section A of Disease markers* **13**, 345–349, https://doi.org/10.3233/cbm-130365 (2013).

37. Ma, Y., Dai, H., Kong, X. & Wang, L. Impact of thawing on reference gene expression stability in renal cell carcinoma samples. *Diagnostic molecular pathology: the American journal of surgical pathology, part B* **21**, 157–163, https://doi.org/10.1097/PDM.0b013e31824d3435 (2012).

38. Mohelnikova-Duchonova, B., Oliverius, M., Honsova, E. & Soucek, P. Evaluation of reference genes and normalization strategy for quantitative real-time PCR in human pancreatic carcinoma. *Disease markers* **32**, 203–210, https://doi.org/10.3233/dma-2011-0875 (2012).

39. Nihon-Yanagi, Y. *et al.* beta-2 microglobulin is unsuitable as an internal reference gene for the analysis of gene expression in human colorectal cancer. *Biomedical reports* **1**, 193–196, https://doi.org/10.3892/br.2013.53 (2013).

40. Ohl, F. *et al.* Identification and validation of suitable endogenous reference genes for gene expression studies of human bladder cancer. *The Journal of urology* **175**, 1915–1920, https://doi.org/10.1016/s0022-5347(05)00919-5 (2006).

41. Ohl, F. *et al.* Gene expression studies in prostate cancer tissue: which reference gene should be selected for normalization? *Journal of molecular medicine (Berlin, Germany)* **83**, 1014–1024, https://doi.org/10.1007/s00109-005-0703-z (2005).

42. Rho, H. W. *et al.* Identification of valid reference genes for gene expression studies of human stomach cancer by reverse transcription-qPCR. *BMC cancer* **10**, 240, https://doi.org/10.1186/1471-2407-10-240 (2010).

43. Romani, C. *et al.* Identification of optimal reference genes for gene expression normalization in a wide cohort of endometrioid endometrial carcinoma tissues. *PloS one* **9**, e113781, https://doi.org/10.1371/journal.pone.0113781 (2014).

44. Saviozzi, S. *et al.* Selection of suitable reference genes for accurate normalization of gene expression profile studies in non-small cell lung cancer. *BMC cancer* **6**, 200, https://doi.org/10.1186/1471-2407-6-200 (2006).

45. Sharan, R. N., Vaiphei, S. T., Nongrum, S., Keppen, J. & Ksoo, M. Consensus reference gene(s) for gene expression studies in human cancers: end of the tunnel visible? *Cellular oncology (Dordrecht)* **38**, 419–431, https://doi.org/10.1007/s13402-015-0244-6 (2015).

46. Sorby, L. A., Andersen, S. N., Bukholm, I. R. & Jacobsen, M. B. Evaluation of suitable reference genes for normalization of real-time reverse transcription PCR analysis in colon cancer. *Journal of experimental & clinical cancer research: CR* **29**, 144, https://doi.org/10.1186/1756-9966-29-144 (2010).

47. Tilli, T. M., Castro Cda, S., Tuszynski, J. A. & Carels, N. A strategy to identify housekeeping genes suitable for analysis in breast cancer diseases. *BMC genomics* **17**, 639, https://doi.org/10.1186/s12864-016-2946-1 (2016).

48. Tsaur, I. *et al.* Reliable housekeeping gene combination for quantitative PCR of lymph nodes in patients with prostate cancer. *Anticancer research* **33**, 5243–5248 (2013).

49. Wang, X. *et al.* Validation of internal reference genes for relative quantitation studies of gene expression in human laryngeal cancer. *PeerJ* **4**, e2763, https://doi.org/10.7717/peerj.2763 (2016).

50. Wierzbicki, P. M. *et al.* Identification of a suitable qPCR reference gene in metastatic clear cell renal cell carcinoma. *Tumour biology: the journal of the International Society for Oncodevelopmental Biology and Medicine* **35**, 12473–12487, https://doi.org/10.1007/s13277-014-2566-9 (2014).

51. Yigin, A. K. *et al.* Selection of reliable reference genes for qRT-PCR analysis on head and neck squamous cell carcinomas. *Biomedical Research* **28**, 2014–2018 (2017).

52. Yu, S., Yang, Q., Yang, J. H., Du, Z. & Zhang, G. Identification of suitable reference genes for investigating gene expression in human gallbladder carcinoma using reverse transcription quantitative polymerase chain reaction. *Molecular medicine reports* **11**, 2967–2974, https://doi.org/10.3892/mmr.2014.3008 (2015).

53. Zhan, C. *et al.* Identification of reference genes for qRT-PCR in human lung squamous-cell carcinoma by RNA-Seq. *Acta biochimica et biophysica Sinica* **46**, 330–337, https://doi.org/10.1093/abbs/gmt153 (2014).

54. Zhao, L. M. *et al.* Optimization of reference genes for normalization of the quantitative polymerase chain reaction in tissue samples of gastric cancer. *Asian Pacific journal of cancer prevention: APJCP* **15**, 5815–5818 (2014).

55. Edge, S. B. & Compton, C. C. The American Joint Committee on Cancer: the 7th edition of the AJCC cancer staging manual and the future of TNM. *Annals of surgical oncology* **17**, 1471–1474, https://doi.org/10.1245/s10434-010-0985-4 (2010).

Acknowledgements

Author contributions

Conceived and designed the experiment: S.A.R. and H.G.; collected the tissue samples and prepared the pathology reports: S.A.R., S.N. and S.M.T.; conducted the experiments: M.Z., A.D. and H.G.; performed data analysis: M.A. and S.A.R.; interpreted the data, revised the manuscript for important intellectual content: M.H., M.H.M. and P.Y.; wrote the manuscript: S.A.R.; final approval for the manuscript submission: M.H. All the authors have reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41598-019-49247-1.

Correspondence and requests for materials should be addressed to M.H.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.