**Title:** Suitable reference genes selection for reliable normalization of gene expression in melanoma: implementation of an innovative GenExpA software

**Authors:** Dorota Hoja-Łukowicz[1*], Dawid Maciążek[2], Marcelina E. Janik[1*]

**Affiliations:**

[1] Institute of Zoology and Biomedical Research, Jagiellonian University, Gronostajowa 9, 30-387 Krakow, Poland,

[2] Smoluchowski Institute of Physics, Jagiellonian University, Prof. Lojasiewicza 11, 30-348 Krakow, Poland

**e-mail addresses:** dorota.hoja-lukowicz@uj.edu.pl, dawid.maciazek@gmail.com, marcelina.janik@uj.edu.pl

**Abstract**

Melanoma is characterized by a high mutation rate and the proper reference selection is a serious problem. The algorithms commonly used to select the best stable reference gene or pair of genes in RT-qPCR data analysis have their limitations which affects the interpretation of the results and drawing conclusions. For reliable assessment of the *B4GALT* genes expression changes in melanoma and comparing them to their expression levels in melanocytes, we implemented an innovative GenExpA software. We proposed selection of the best reference by combining the NormFinder algorithm with progressive removal of the least stable gene from the candidate genes in a given experimental model and the set of daughter models assigned to it. The reliability of references is validated by normalizing of target gene expression level through all models and is described by the coherence score. The software performs statistical analyses and generates results in the form of ready-to-use graphs and tables. GenExpA is a promising tool for basic research; it produces analyses of target gene expression in a manner independent of the experimental model and the normalizer. GenExpA and its manual is freely available at https://drive.google.com/drive/folders/1FcpBtEc_bKPvEnM5GcPZerBb_Tw_fCkt?usp=sharing or https://www.sciencemarket.pl/baza-programow-open-source#oferty. Supplemental materials are deposited under DOI number: 10.5281/zenodo.4544296.

**Keywords:** *B4GALT*; gene stability; melanoma; NormFinder; RT-qPCR

## Introduction

Achieving reliable results for normalization of the transcript level of a target gene requires an internal reference (housekeeping) gene or pair of genes with stable expression between the analysed samples and under different experimental conditions. Moreover, it is believed that the transcript levels of the reference and target genes should be expressed at roughly the same level (1). However, the transcript levels of many of the housekeeping genes typically used as

references may change significantly under physiological or pathological conditions as well as across tissue types and experimental conditions (1, 2, 3, 4, 5, 6, 7, 8). Therefore the selection and validation of the normalizer are the most critical stages in qPCR data analysis. Typically, one or more of five algorithms (geNorm (3), NormFinder (9), BestKeeper (10), the comparative ΔCt method (11) and RefFinder; http://150.216.56.64/referencegene.php) are used to select the best stable single or combination of reference genes from a panel of candidate genes. The normalizer with the lowest stability value is considered the best and is then used for calculation of target gene expression in a group of samples. We have recently shown that this commonly accepted approach seems not enough for accurate and reliable target gene transcript normalization and that it may lead to biologically incorrect conclusions (12). We demonstrated that beyond the parent group of samples of interest (experimental model) it is necessary to construct daughter groups (auxiliary models – combinations of samples without repetition), followed by selection and validation of an appropriate normalizer for each group. If we get inconsistent results for the expression level of a target gene based on a comparison of individual samples through all groups, we have to search for new references with improved stability values via progressive removal of the least stable candidate reference gene in each group, followed by re-selection and re-validation of new normalizers each time (12). However, this proposed approach of RT-qPCR data analysis has been a very time-consuming task, requiring the use of several specialist software packages. This problem led us to automate these calculations by developing the GenExpA tool (Gene Expression Analyzer). It is a comprehensive tool based on a previously described workflow for quantified as well as raw qPCR data (12). Based on the selected reference gene or pair of genes, GenExpA calculates the relative target gene expression level, performs statistical analyses and generates results in the form of graphs and tables. The innovative aspect of GenExpA is that it estimates the coherence of results for each target gene in all designed models. It determines the

robustness of normalization and works out the coherence score for the individual target gene. The advantage of this strategy is its ability to determine relative target gene expression based on the consistency of the results in all tested samples, regardless of the experimental model and the reference gene or pair of genes used. Here, based on qPCR data for B4GalT1–B4GalT7 transcripts in melanocytes and melanoma cells, we have presented step by step how to perform the analysis.

## Materials and methods

### Cell culture

Human epidermal melanocytes (adult, HEMa-LP) and human melanoma cell lines Mel202 (uveal primary), WM35 (cutaneous primary, radial/vertical growth phase), WM793 (cutaneous primary, vertical growth phase) and WM266-4 (skin metastasis) were cultured as described previously (12). All cultures were verified mycoplasma-free using the MycoAlert™ mycoplasma detection kit (Lonza).

### Reverse transcription qPCR

Total RNA extraction, the reverse-transcription reaction and real-time qPCR reactions were performed as previously described (12). RNA and cDNA concentrations as well as their quality were assessed with a NanoDrop 2000 spectrophotometer (Supplementary Data 1). Housekeeping (*GUSB*, *HPRT1*, *PGK1*, *RPS23*, *SNRPA*) and target (*B4GALT1–7*) gene-specific mRNAs were amplified with the use of TaqMan™ Gene Expression Assays, as listed in Table 1. All reactions were performed in three biological and three technical replicates. The reaction results were analysed with StepOne Software ver. 2.0. Raw Ct values were calculated using StepOne ver. 2.0, applying automatic threshold and baseline settings.

### RT-qPCR data analysis

The GenExpA tool was used for fully automated qPCR data analysis. The gene or combination of two genes with the lowest stability value were selected as the best reference by combining the model-based variance estimation with progressive removal of the least stable reference gene. For statistical analysis, the non-parametric Kolmogorov-Smirnov with subsequent post hoc Dunn's test or the ANOVA Kruskal-Wallis test were used for models comprised of two unmatched samples or three or more unmatched samples, respectively. P values less than 0.05 were taken to indicate statistical significance ($P<0.05$).

Table 1. Summary of candidate reference and target genes

| **Symbol** | **Gene name (Assay ID, TaqMan® probes; Amplicon size bp)** | **Location (GeneCards)** | **Description** |
|---|---|---|---|
| **Candidate reference genes** | | | |
| *GUSB* | Glucuronidase Beta (Hs00939627_m1; 96 bp) | 7q11.21 | Glycosaminoglycan metabolism |
| *HPRT1* | Hypoxanthine Phosphoribosyltransferase 1 (Hs99999909_m1; 100 bp) | Xq26.2-q26.3 | Generation of purine nucleotides |
| *PGK1* | Phosphoglycerate Kinase 1 (Hs00943178_g1; 73 bp) | Xq21.1 | Metabolic pathway of glycolysis |
| *RPS23* | Ribosomal Protein S23 (Hs01922548_s1; 90 bp) | 5q14.2 | poly(A) RNA binding and structural constituent of ribosome |
| *SNRPA* | Small Nuclear Ribonucleoprotein Polypeptide A (Hs00190231_m1; 123 bp) | 19q13.2 | Splicing of cellular pre-mRNAs |

| **Target genes** | | | |
|---|---|---|---|
| *B4GALT1* | Beta-1,4-Galactosyltransferase 1 (Hs00155245_m1; 70 bp) | 9p21.1 | Biosynthesis of different glycoconjugates and saccharide structures |
| *B4GALT2* | Beta-1,4-Galactosyltransferase 2 (Hs00243566_m1; 57 bp) | 1p34.1 | |
| *B4GALT3* | Beta-1,4-Galactosyltransferase 3 (Hs00937515_g1; 68 bp) | 1q23.3 | |
| *B4GALT4* | Beta-1,4-Galactosyltransferase 4 (Hs00186850_m1; 70 bp) | 3q13.32 | |
| *B4GALT5* | Beta-1,4-Galactosyltransferase 5 (Hs00941041_m1; 66 bp) | 20q13.13 | |
| *B4GALT6* | Beta-1,4-Galactosyltransferase 6 (Hs00999574_m1; 69 bp) | 18q11.1 | |
| *B4GALT7* | Beta-1,4-Galactosyltransferase 7 (Hs01011260_g1; 74 bp) | 5q35.3 | |

## Results

A flow chart of target gene expression and statistical analysis of qPCR data using the GenExpA tool is shown in Figure 1. To test the GenExpA tool, we uploaded raw qPCR data for candidate reference (*HPRT1*, *PGK1*, *RPS23*, *SNRPA*; set 1) and target (*B4GALT1*–*B4GALT7*) genes (Supplementary Table 1) as well as quantified qPCR data for candidate reference genes (Supplementary Table 2) in an experimental model of five human cell lines

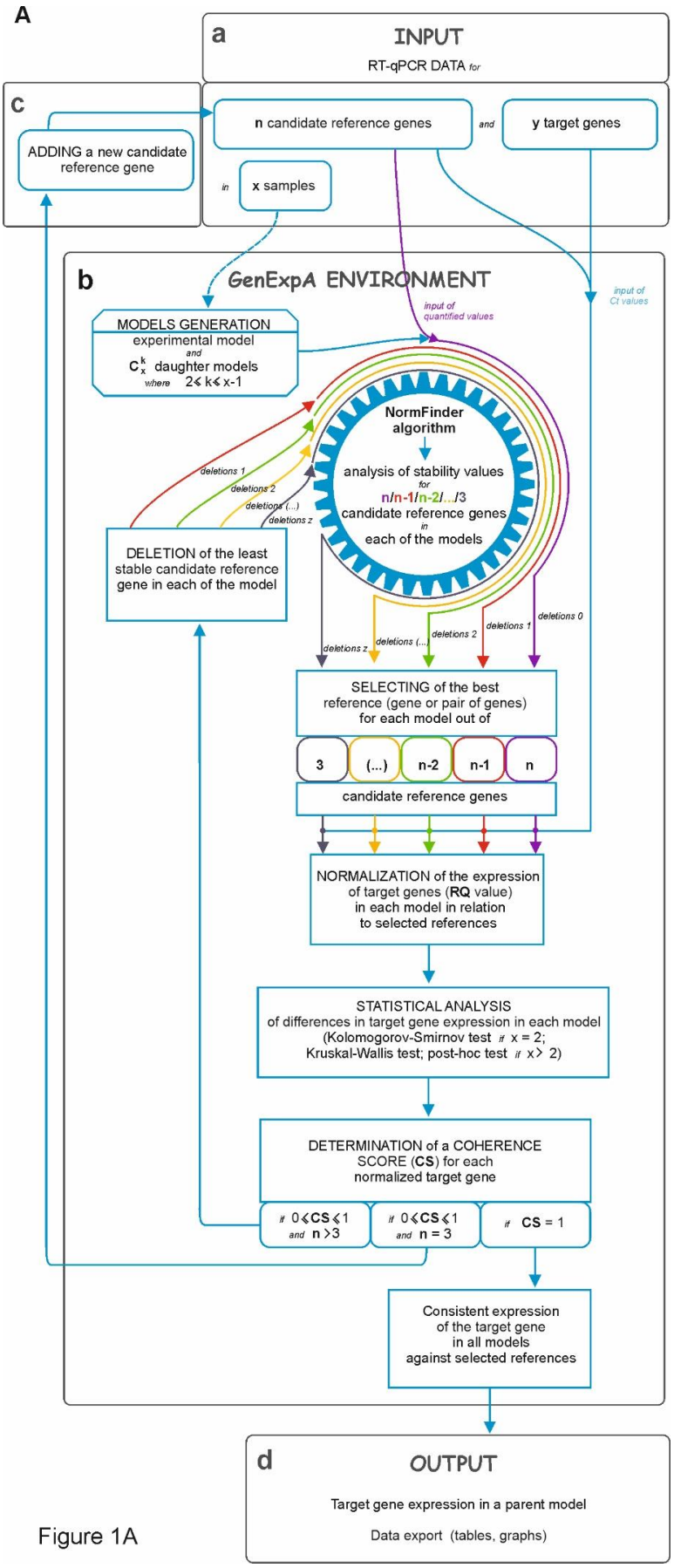
A
a
INPUT
RT-qPCR DATA for
c
ADDING a new candidate reference gene
n candidate reference genes
and
y target genes
in
x samples
b
GenExpA ENVIRONMENT
input of quantified values
input of Ct values
MODELS GENERATION
experimental model
and
$C_x^k$ daughter models
where 2⩽ k⩽ x-1
NormFinder algorithm
analysis of stability values
for
n/n-1/n-2/.../3
candidate reference genes
in
each of the models
deletions 1
deletions 2
deletions (...)
deletions z
DELETION of the least stable candidate reference gene in each of the model
deletions 0
SELECTING of the best reference (gene or pair of genes) for each model out of
3
(...)
n-2
n-1
n
candidate reference genes
NORMALIZATION of the expression of target genes (RQ value) in each model in relation to selected references
STATISTICAL ANALYSIS
of differences in target gene expression in each model
(Kolomogorov-Smirnov test if x = 2;
Kruskal-Wallis test; post-hoc test if x> 2)
DETERMINATION of a COHERENCE SCORE (CS) for each normalized target gene
if 0⩽CS⩽1 and n >3
if 0⩽CS⩽1 and n = 3
if CS = 1
Consistent expression of the target gene in all models against selected references
d
OUTPUT
Target gene expression in a parent model
Data export (tables, graphs)

Figure 1A

**B**

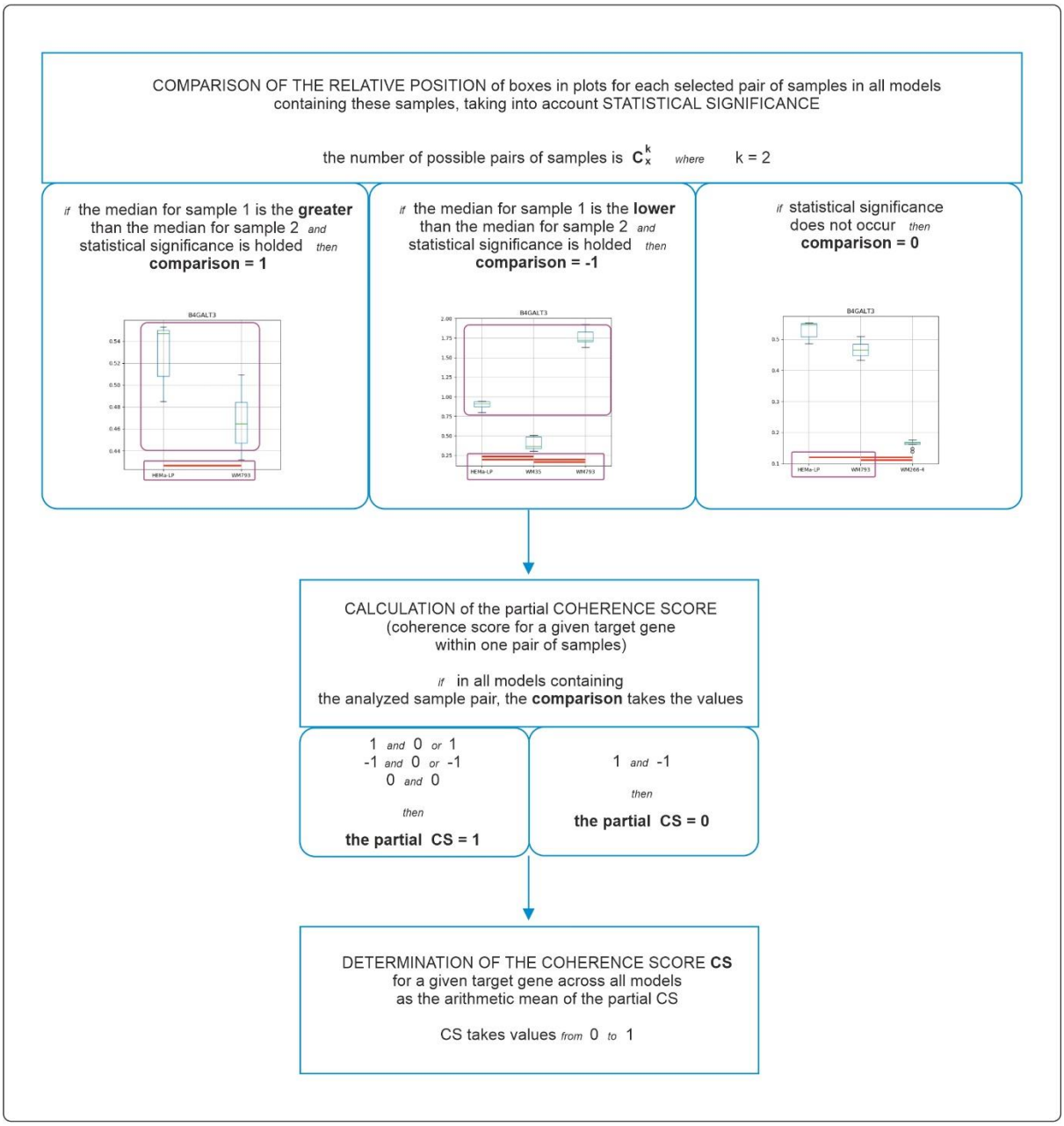


Figure 1B

**Figure 1.** Overview of the GenExpA tool. ***(A)*** (a) The user uploads a table with Ct values for 'n' candidate reference genes and 'y' target genes measured in 'x' samples. The user uploads a table with calculated quantified values for 'n' candidate reference genes. (b) In the setting mode (not shown), the user separates 'n' potential reference genes from a list of all genes, and unchecks the statistical model, wherein the non-parametric Kolmogorov-Smirnov test with subsequent post hoc Dunn's test or the ANOVA Kruskal-Wallis test_are recommended for models composed of two unmatched samples or three or more unmatched samples, respectively, in the case of a non-normal distribution. Alternatively, the pairwise t-test with Holm adjustment is recommended as a statistical model for multiple pairs of

observations in the case of a normal distribution (to determine the significance of a difference in gene expression between two or more groups) (22). In the setting mode, the user determines the P value indicating the level of statistical significance (default $P<0.05$), and single sample repetition (default 3) (not shown). The GenExpA tool automatically creates a set of possible 'z' models (groups of samples), which are combinations (without repetitions) of two or more samples. The 'z' models consist of the experimental model and its 'z-1' daughter models (auxiliary models). The user starts the automated data analysis by clicking on the 'run calculations' button; it causes the implemented NormFinder algorithm to select the best references between 'n' potential reference genes (purple line) according to the stability values of their expression in the experimental model and its auxiliary models. The best stability value is referred to the minimal combined inter- and intragroup expression variation of the gene. NormFinder works on Ct values converted to a linear scale according to the formula $E^{-\Delta Ct}$ (9). The GenExpA tool normalizes the expression of target genes (RQ value) in each model in relation to selected references, conducts a statistical analysis and calculates the coherence score (CS) separately for each target gene (details of how GenExpA determines CS are shown in ***(B)***). If the CS value is equal to 1, the target gene expression analysis is completed. In the case of inconsistent results (CS <1), the user can change the setting parameters by choosing 1 in the 'Remove repetitions' box and by marking the option 'select best remove for models' (not shown). The user starts the improved analysis by clicking on the 'run calculations' button. The GenExpA tool conducts a new analysis based on rejection of the least stable gene from a set of 'n' candidate genes in each model, followed by choosing new references from 'n-1' candidate reference genes in each model and by a reanalysis of the normalized target gene(s) (red line). This approach can be repeated serially (number in 'Remove repetitions' box should be increased by 1 each time) until the CS value is below 1 and the pool of reference genes is not less than three (from the green line to the navy blue line). (c) If the gene removal process does not yield consistency of analysis, the user can upload a new input extended with data for an additional housekeeping gene or genes and perform an improved analysis based on the removal strategy until consistent results are achieved. (d) The user can export the results and graphs of all analyses in publication-ready form. The results show coherence scores, the best reference gene or gene pair results, RQ data and statistics. ***(B)*** Flow chart of determination of the coherence score (CS) for a given normalized target gene.

(melanocytes and melanoma cells from different stages of oncogenic progression). The uploaded quantified values were calculated based on the calibration curve previously prepared for all candidate reference genes in all analysed cell lines. Working with the GenExpA interface, the potential reference genes were selected from the pool of all analysed genes

presented in the ‘Available reference genes’ window, and a panel of 26 possible models (the experimental model and 25 daughter models) was automatically designed using the ‘Generate combinations’ option. Because of the non-normal data distribution, the Kruskal-Wallis/Kolmogorov-Smirnov with subsequent post hoc Dunn’s test was chosen from the ‘Statistical model’ bar. Finally, 0 was set in the ‘Remove repetitions’ box and the ‘Single sample repetition’ and ‘Confidence’ boxes were set by default at 3 and 0.05, respectively. The ‘Select best remove for model’ option was left unselected. After clicking on the ‘Run calculation’ button, the GenExpA tool started the analysis. First, the implemented NormFinder algorithm determined the most stable reference gene or pair of genes from four candidate reference genes in each of the 26 models (Table 2, part A; remove repetition level 0). The stability value for a selected reference in the experimental model (model No. 26) was 0.349847, and the reference stability values for auxiliary models ranged from 0.069619 for reference gene pair HPRT1/RPS23 in model No. 7 to 0.494109 for reference gene pair RPS23/SNARPA in model No. 16 (Table 2, part A). Then the relative expression levels of target genes *B4GALT1–B4GALT7* were determined as relative quantification (RQ) values in each of the 26 models via normalization to the reference gene/genes assigned for these models. After clicking on the ‘Export results’ button, the GenExpA software generated tables summarizing the calculated reference genes’ stability values, the obtained RQ values and the statistical test results, as well as attributed coherence values (Supplementary Table 3). By choosing the ‘Export graph’ option, box-plots representing the medians of the obtained RQ values with statistical significance bars were generated as .png files, each representing a set of target genes in one of the analysed models (graphs for all models are compiled in Supplementary Figure 1). In this analysis, GenExpA calculated the average coherence score at level 0.94. This value resulted from unreliable/uncertain normalization of four target genes (*B4GALT3*, *B4GALT5*, *B4GALT6*, *B4GALT7*), which reached coherence scores below 1

Table 2.
Reference genes and their stability values in a panel of the experimental model and daughter models selected by NormFinder software on the base of 4 (part A) and 5 (part B) potential reference genes before (0) and after removing one (1) or two (2) the least stable genes from a given set of 4 and 5 reference genes.

| No. | model | 4 potential reference genes - part A | | 5 potential reference genes - part B | | |
|---|---|---|---|---|---|---|
| | | remove repetition level | | remove repetition level | | |
| | | 0 | 1 | 0 | 1 | 2 |
| 1 | HEMa-LP Mel202 | *HPRT1* 0,181854 | ***HPRT1 SNRPA* 0,069925** | ***HPRT1* 0,051156** | *GUSB SNRPA* 0,053992 | *HPRT1 SNRPA* 0,069925 |
| 2 | HEMa-LP WM35 | *HPRT1* 0,223074 | ***HPRT1 SNRPA* 0,147267** | ***HPRT1* 0,103310** | *HPRT1 RPS23* 0,178910 | *GUSB HPRT1* 0,219325 |
| 3 | HEMa-LP WM793 | *HPRT1* 0,279936 | ***HPRT1 SNRPA* 0,069024** | *GUSB* 0,091107 | ***GUSB SNRPA* 0,038623** | *HPRT1* 0,079924 |
| 4 | HEMa-LP WM266-4 | ***PGK1 RPS23* 0,223052** | *HPRT1 PGK1* 0,340317 | *PGK1 RPS23* 0,116316 | ***HPRT1 PGK1* 0,073448** | *HPRT1 SNRPA* 0,346148 |
| 5 | Mel202 WM35 | *HPRT1* 0,104826 | ***HPRT1 SNRPA* 0,064185** | *GUSB RPS23* 0,076568 | *HPRT1* 0,104826 | ***HPRT1 SNRPA* 0,064185** |
| 6 | Mel202 WM793 | *HPRT1 RPS23* 0,100804 | ***HPRT1 SNRPA* 0,052623** | ***HPRT1 SNRPA* 0,030737** | *GUSB SNRPA* 0,066292 | *HPRT1 SNRPA* 0,052623 |
| 7 | Mel202 WM266-4 | ***HPRT1 RPS23* 0,069619** | *HPRT1 RPS23* 0,162321 | *HPRT1 RPS23* 0,085300 | *GUSB* 0,161247 | ***RPS23* 0,012949** |
| 8 | WM35 WM793 | *SNRPA* 0,169211 | ***HPRT1 SNRPA* 0,099923** | ***GUSB RPS23* 0,071640** | *SNRPA* 0,169211 | *HPRT1 SNRPA* 0,099923 |
| 9 | WM35 WM266-4 | *HPRT1 PGK1* 0,115685 | ***HPRT1 RPS23* 0,101104** | ***RPS23* 0,095511** | *HPRT1 PGK1* 0,115685 | *HPRT1 RPS23* 0,101104 |
| 10 | WM793 WM266-4 | *HPRT1 RPS23* 0,131313 | ***HPRT1 PGK1* 0,113854** | *PGK1 SNRPA* 0,125481 | *HPRT1 PGK1* 0,082377 | ***GUSB* 0,053399** |
| 11 | HEMa-LP Mel202 WM35 | *HPRT1* 0,188414 | ***HPRT1 SNRPA* 0,113901** | ***HPRT1* 0,098380** | *GUSB HPRT1* 0,249224 | *HPRT1 RPS23* 0,318021 |
| 12 | HEMa-LP Mel202 WM793 | *HPRT1* 0,317389 | ***HPRT1 SNRPA* 0,065166** | *HPRT1* 0,238320 | ***GUSB SNRPA* 0,053718** | *HPRT1 SNRPA* 0,065166 |
| 13 | HEMa-LP Mel202 WM266-4 | ***RPS23 SNRPA* 0,240609** | *HPRT1 SNRPA* 0,337367 | *RPS23 SNRPA* 0,113205 | ***GUSB SNRPA* 0,070667** | *GUSB PGK1* 0,331792 |
| 14 | HEMa-LP WM35 WM793 | *HPRT1* 0,335229 | ***HPRT1 SNRPA* 0,113813** | *HPRT1* 0,212917 | *HPRT1* 0,172137 | ***HPRT1 SNRPA* 0,113813** |
| 15 | HEMa-LP WM35 WM266-4 | ***HPRT1 PGK1* 0,243675** | *HPRT1 SNRPA* 0,344720 | ***PGK1 RPS23* 0,147183** | *GUSB HPRT1* 0,220384 | *GUSB HPRT1* 0,248189 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 16 | HEMa-LP WM793 WM266-4 | *RPS23 SNRPA* 0,494109 | ***HPRT1 PGK1* 0,275998** | *GUSB* 0,318394 | ***GUSB SNRPA* 0,070868** | *GUSB SNRPA* 0,252192 |
| 17 | Mel202 WM35 WM793 | *HPRT1 RPS23* 0,171610 | ***HPRT1 SNRPA* 0,089572** | ***HPRT1 SNRPA* 0,068022** | *HPRT1 RPS23* 0,171610 | *HPRT1 SNRPA* 0,089572 |
| 18 | Mel202 WM35 WM266-4 | ***HPRT1 PGK1* 0,119949** | *HPRT1 RPS23* 0,177300 | *PGK1 SNRPA* 0,190394 | ***HPRT1 PGK1* 0,119950** | *HPRT1 RPS23* 0,177300 |
| 19 | Mel202 WM793 WM266-4 | ***HPRT1 RPS23* 0,108840** | *HPRT1 PGK1* 0,218254 | *PGK1 SNRPA* 0,172465 | ***GUSB SNRPA* 0,078611** | *GUSB SNRPA* 0,20656 |
| 20 | WM35 WM793 WM266-4 | ***HPRT1 RPS23* 0,168051** | *HPRT1 PGK1* 0,183833 | *PGK1 SNRPA* 0,214693 | *HPRT1 PGK1* 0,265683 | ***HPRT1 PGK1* 0,183833** |
| 21 | HEMa-LP Mel202 WM35 WM793 | *HPRT1* 0,316281 | ***HPRT1 SNRPA* 0,099055** | *HPRT1* 0,196310 | *HPRT1* 0,196654 | ***HPRT1 SNRPA* 0,099055** |
| 22 | HEMa-LP Mel202 WM35 WM266-4 | ***RPS23 SNRPA* 0,216306** | *HPRT1 SNRPA* 0,294408 | ***RPS23 SNRPA* 0,200264** | *GUSB HPRT1* 0,246682 | *HPRT1 RPS23* 0,311142 |
| 23 | HEMa-LP Mel202 WM793 WM266-4 | *RPS23 SNRPA* 0,392232 | ***HPRT1 SNRPA* 0,275124** | *HPRT1* 0,343380 | ***GUSB SNRPA* 0,070951** | *GUSB PGK1* 0,271811 |
| 24 | HEMa-LP WM35 WM793 WM266-4 | *RPS23 SNRPA* 0,422874 | ***HPRT1 SNRPA* 0,281318** | *HPRT1* 0,355958 | ***HPRT1 PGK1* 0,227386** | *HPRT1 SNRPA* 0,281318 |
| 25 | Mel202 WM35 WM793 WM266-4 | ***HPRT1 RPS23* 0,143972** | *HPRT1 PGK1* 0,220866 | ***PGK1 SNRPA* 0,210985** | *HPRT1 PGK1* 0,226756 | *HPRT1 PGK1* 0,220866 |
| 26 | HEMa-LP Mel202 WM35 WM793 WM266-4 | *RPS23 SNRPA* 0,349847 | ***HPRT1 SNRPA* 0,247033** | *HPRT1* 0,297367 | ***GUSB SNRPA* 0,196718** | *HPRT1 SNRPA* 0,247033 |

The differences in the choice of the normalizer or obtained stability values are written in **bold**.

(Supplementary Figure 2). Next, to improve the estimation of the expression of these target genes, the least stable reference gene in each model was removed by setting 1 in the 'Remove repetitions' box. Also, marking 'Select best remove for models' caused GenExpA to choose the reference from the level of removal for which the stability value was lower in a given model. This approach raised the average coherence score from 0.94 to 0.99 and gave lower stability values of the normalizers in 17 of the 26 models. Now these stability values ranged between 0.052623 for *HPRT1/SNRPA* in model No. 6 and 0.281318 for *HPRT1/SNRPA* in model No. 24 (Table 2, part A; remove repetition level 1). The stability value of the normalizer for the experimental model (model No. 26) also was improved: 0.247033. The coherence scores for

*B4GALT1, B4GALT2* and *B4GALT4* were kept at 1, and for genes *B4GALT3, B4GALT5* and *B4GALT7* reached the value 1, confirming the consistency of the analysis of the expression levels of these genes. At this point of the analysis, the coherence score for gene *B4GALT6* was below 1 (Supplementary Figure 3), suggesting that the selected references are not suitable for accurate and reliable normalization of the B4GALT6 transcript level (Supplementary Table 4 contains the tables and Supplementary Figure 4 the box-plots generated in this analysis). Executed removal of the least stable gene reduced the pool of potential reference genes to three, that is, the minimum number of genes required for NormFinder selection (9), thus ending the possibility of removing the next-weakest reference gene. To continue the analysis, we enlarged the pool of candidate reference genes by adding a new housekeeping gene, *GUSB* (inputs with Ct and quantified qPCR data are presented in Supplemental Tables 5 and 6). Then the best reference gene or pair of genes from five candidate housekeeping genes was selected (digit 0 entered in the 'Remove repetitions' box and without choosing 'Select best remove for models' box). The selected references possessed stability values ranging from 0.030737 for *HPRT1/SNRPA* (auxiliary model No. 6) to 0.355958 for *HPRT1* (auxiliary model No. 24), and 0.297367 in the experimental model (model No. 26) (Table 2, part B; remove repetition level 0). Validation of the selected references confirmed the consistency of expression analysis for genes *B4GALT1*–*B4GALT5* and *B4GALT7* but did not improve the CS value for *B4GALT6* (Supplementary Figure 5, Supplementary Table 7, Supplementary Figure 6). Removing the least stable gene from five candidate reference genes (digit 1 entered in 'Remove repetitions' box and choosing 'Select best remove for models' box) (Table 2, part B; remove repetition level 1; Supplementary Figure 7) generated inconsistent results for *B4GALT4* and *B4GALT6* (Supplementary Table 8; Supplementary Figure 8). Sequential removal of the two least stable potential reference genes (digit 2 entered in 'Remove repetitions' box and choosing 'Select best remove for models' box) allowed us to assign better stability values to 16 of the 26

models (Table 2, part B; remove repetition level 2) and to complete the analysis with the expected coherence score value of 1 for *B4GALT4* and *B4GALT6* (Supplementary Figure 9). Supplementary Table 9 contains the tables and Supplementary Figure 10 the box-plots corresponding to this analysis. At this point of the analysis, however, the coherence score dropped to 0.9 for gene *B4GALT5*.

Having the Ct values for five potential reference genes, we conducted similar selections of references from four other possible combinations of 4 candidate reference genes: *RPS23*, *SNRPA*, *HPRT1* and *GUSB* (set 2); *SNRPA*, *HPRT1*, *GUSB* and *PGK1* (set 3); *HPRT1*, *GUSB*, *PGK1* and *RPS23* (set 4); and *GUSB*, *PGK1*, *RPS23* and *HPRT1* (set 5), before and after rejection of the weakest reference genes. The results obtained from normalization of target gene expression in the 26 models were exported from the GenExpA tool as Supplementary Tables 10–17 and corresponding Supplementary Figures 11–18. A summary of all target gene expression analyses in the experimental model of interest, based on the selection of references from four (sets 1–5) and five candidate reference genes, is presented in Supplementary Figure 19. Based on it, we concluded that if, at different levels of rejection of the weakest reference gene from sets of four or five candidate reference genes, the analysis of the expression of a given target gene ended with CS = 1, the obtained results do not contradict each other (for example, compare the expression levels of *B4GALT1* in sets 1–5 and the set of five candidate reference genes – box-plots 1–13). However, starting the analysis with a higher number of potential reference genes creates an opportunity to select the reference gene with a lower stability value in the experimental model as well as auxiliary models, and therefore gives a more robust analysis of target gene expression. For example, the use of a set of five candidate reference genes (*PGK1*, *RPS23*, *SNRPA*, *HPRT1*, *GUSB*) with sequential removal of the two least stable potential reference genes gave the lowest stability value of the selected reference gene for the experimental model (0.196718) and auxiliary

models (0.012949–0.227386). On the other hand, we noticed that box-plots 13 and 91 show exactly the same results as, respectively, box-plots 4, 5, 6, 10, 12 and 82, 84, 85, 88, 90 (compare the medians of relative quantity (RQ) and statistical significance between samples). It means that in the case of *B4GALT1* and *B4GALT7* gene expression analysis, the use of four housekeeping genes from sets 2 or 5 with rejection of the most unstable housekeeping gene, or from set 3, is sufficient. The same is true for genes *B4GALT2* and *B4GALT3* and sets 3 or 5 after rejection of the most unstable housekeeping gene (compare box-plots 26 and 39 with 18, 19, 23, 25 and 31, 32, 36, 38, respectively). Therefore, adding a fifth housekeeping gene is mandatory to confirm the robustness of an analysis conducted with the above-mentioned sets of four candidate reference genes.

In the case of gene *B4GALT6*, none of the sets of four potential reference genes was designed well enough to conduct a reliable analysis. Only the application of five potential reference genes with double rejection of the weakest gene allowed us to determine a reliable normalizer in each model and to obtain a coherent analysis (box-plot 78). On the other hand, selection of the normalizer based on five potential reference genes and with 'Remove repetition' set at 2 led to an inconsistent analysis for the relative expression of gene *B4GALT5* (box-plot 65), although the same box-plots were generated with 'Remove repetition' set at 1 as well as in the case of set 3 with 'Remove repetition' set at 0 or 1 (compare box-plots 57, 58, 64 and 65), and the reference stability values for the experimental model remained the same (0.196718). It means that if a more robust analysis is needed, to improve its consistency another housekeeping gene or genes should be introduced into the pool of candidate reference genes. Another approach may be to complete the analysis with 'Remove repetition' set at 1 (box-plot 64) or 2 (box-plot 65), with indication of the CS values and stability parameters for the reference genes selected in the experimental and auxiliary models.

It is worth noting that simple application of the NormFinder algorithm (‘Remove repetition’ 0) may lead to a biological misinterpretation of the obtained results, as it strongly depends on the choice of a set of potential reference genes. For example, sets 1 and 5 showed significantly higher relative expression of gene *B4GALT6* in WM793 cells than in WM266-4 cells. In turn, set 2 and the set of five housekeeping genes showed inverse relationships that were also statistically significant (compare box-plots 66 and 74 with box-plots 68 and 76). Moreover, the relative expression level of *B4GALT6* was significantly lower in cell line WM35 than in cell line WM266-4, as shown in sets 1 and 5 (box-plots 66 and 74), or significantly higher as shown in set 4 (box-plot 72).

Confusing results are also observed for gene *B4GALT3*: in sets 1 (box-plot 27) and 5 (box-plot 35) its relative expression level was significantly lower in cell line WM35 than in cell line WM793, but in set 4 the relationship was the reverse (box-plot 33). In addition, a larger pool of candidate reference genes does not always lead to selecting the better normalizer in simple application of the NormFinder algorithm (compare the reference stability values for the experimental model and auxiliary models in set 3 and the set of five candidate reference genes; ‘Remove repetition’ 0). In contrast, our method based on gradual removal of the gene with the highest variability of expression leads to selection of a normalizer with a lower stability value (compare the reference stability values for the experimental model and auxiliary models in sets 1–5 with ‘Remove repetition’ 1 and the set of five candidate reference genes with ‘Remove repetition’ 2).

Figure 2 shows the results of GenExpA analysis for the expression levels of target genes *B4GALT1–B4GALT7* in the analysed experimental model. We obtained consistent and complete analyses for all target genes under the given stability values for the experimental and auxiliary models. These presented results demonstrate that in melanoma cells, in particular metastatic melanoma cell line WM266-4, the expression of all *B4GALT* genes decreases relative to that of

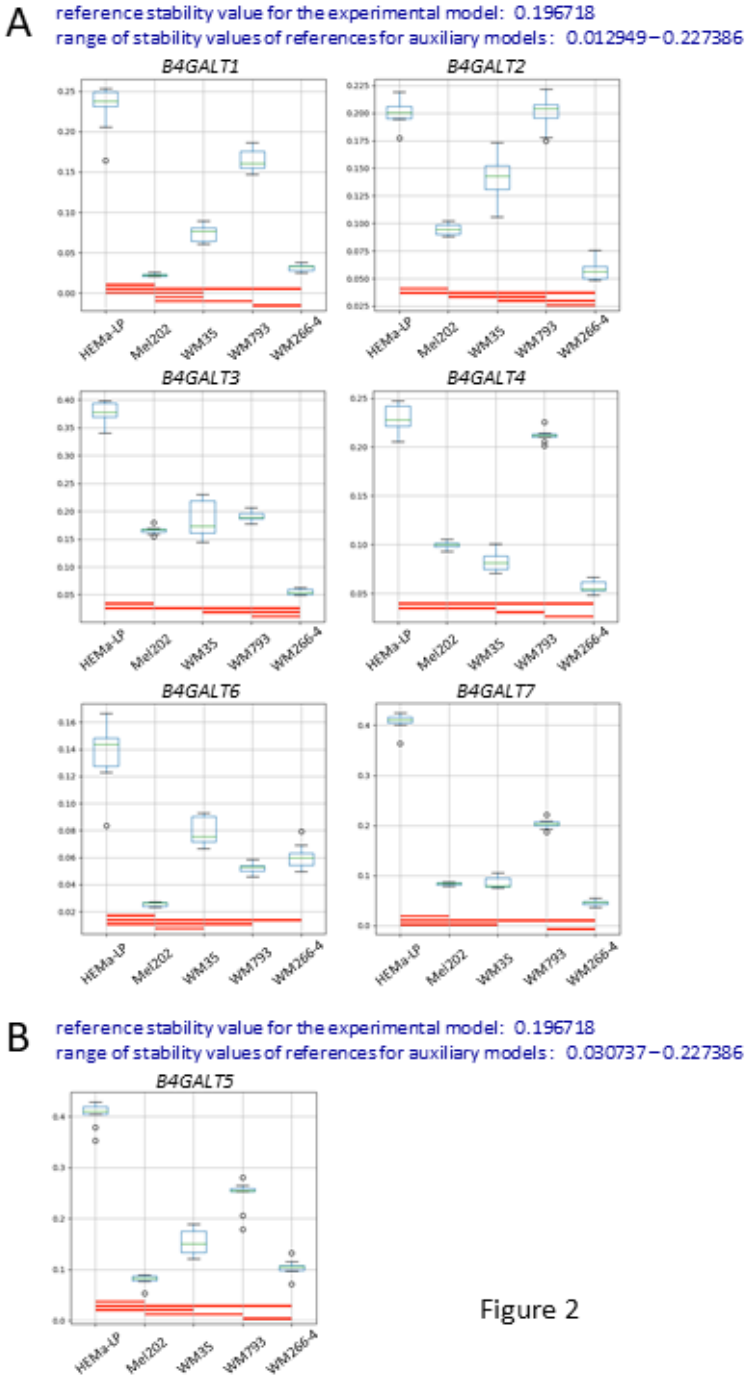


**Figure 2.** Medians of relative quantity (RQ) for target genes *B4GALT1–B4GALT7* in the experimental model normalized to a reference gene or pair of reference genes with the best stability values resulting from GenExpA analysis based on a set of five candidate reference genes and remove repetition level 2 ***(A)*** or remove repetition level 1 ***(B)*** (see Table 2, part B). The statistical analyses used the Kolmogorov-Smirnov test for models composed of two cell lines, or the Kruskal-Wallis test (ANOVA for models built of three or more cell lines). Red line represents statistical significance, P<0.05.

melanocytes. An exception is melanoma cell line WM793, in which only the expression of *B4GALT6* decreases markedly; the expression levels of the remaining *B4GALT* genes show no significant differences as compared to those for melanocytes.

**Discussion**

Obtaining reliable results from the PCR reaction is the outcome of many factors related to handling of the material, beginning with the step of RNA/DNA isolation and ending in analysis of the results. That is why the MIQE guidelines were proposed for transparent reporting of experimental details in all publication prepared with the use of qPCR technique (13). The reaction needs to be standardized in molecular analyses of all kinds of biological materials (14, 15, 16, 17). Over the past decade the MIQE guidelines have not come into common use. Dijkstra et al. (2014) made a critical analysis of qPCR result normalization in 179 publications regarding colorectal cancer. They showed that only 3% of these studies applied qPCR methodology based on the use and validation of multiple reference genes. Several statistical are used to select the best stable single gene or combination of reference genes from a panel of candidate genes. Each of these algorithms has its limitations which affect the ranking of candidate reference genes and finally the selection of the best normalizer (18, 19). Recent studies have proposed the use of at least three different methods, but the problem of how to interpret conflicting results obtained from the use of different statistical methods still puzzles researchers. They try to integrate a few algorithms to average the stability ranks; for example, the RefFinder algorithm calculates the geometric mean of ranking values obtained by four other algorithms (18). But these statistical methods differ from each other; averaging the ranks can lead to a suboptimal assessment of stable reference genes (19). It is commonly accepted that the lower the stability value of the reference, the greater the certainty of correct determination of the relative expression of the target gene. Yet it has not been clarified how low the stability value must be. In our method we used the model-dependent NormFinder algorithm, and by simulating the removal of the least stable

gene from a set of candidate genes in the experimental as well as daughter models we were able to obtain more stable normalizers for each model. This approach helped us to exclude incorrect normalizations and biological misinterpretations concerning the alteration of *ERAP1* gene expression during melanoma progression (12). We chose the NormFinder algorithm because it calculates the stability of a reference gene or pair of genes by analysing the variation of its expression within and between groups. However, genes with high overall variation influence the stability ranking of candidate reference genes. Herein we have demonstrated that simple application of NormFinder ('Remove repetition' 0) may lead to biological misinterpretation of the results, because it depends on the housekeeping gene set used to select the reference. Our method eliminates the influence of the most variable gene or genes on the stability rankings of all other candidate reference genes. In addition, breaking down the experimental model into daughter models (auxiliary models) allows us to check whether a uniform trend of target gene expression among particular cell lines is maintained. This trend decides the consistency of the analysis and is described by coherence score CS=1. Obtaining full consistency of the analysis is tantamount to its completion. If, despite the application of our strategy, a uniform trend of target gene expression among particular cell lines has not been achieved, introducing a successive housekeeping gene or genes, along with repeated rejection of the weakest gene, can lead to selection of the better reference and, in turn, biologically correct conclusions. It is important to note that once selected, a reference for a given model will not necessarily be selected once again in an experiment repeated for this model after some time, even if the pool of candidate reference genes used is exactly the same. This is because the gene expression pattern at the time point of sample collection depends on the current rate of cell mass growth and on microenvironmental conditions (20, 21).

**Conclusions**

Herein we have presented the GeneExpA tool for reference gene validation and automatic calculation of target gene expression. To each analysis, GenExpA attributes a CS value ranging from 0 to 1; a fully coherent analysis is described by a CS value of 1. We suggest starting the analysis with a set of at least four potential housekeeping genes. If the results are not satisfactory – if the CS value is below 1 – introducing another candidate housekeeping gene or genes to the set of candidate reference genes may improve them. However, even if CS is equal to 1 for a given target gene, enlarging the pool of candidate reference genes can lead to a more robust analysis. For clarity of the biological conclusions drawn, the description of the results should give stability values assigned to the experimental model and auxiliary models; this may help other researchers to compare their results and understand any differences between their outcomes and yours. GenExpA software can carry out an analysis of many genes independently at the same time.

**Acknowledgments**

Michael Jacobs line-edited the paper for submission.

**Funding**

This work was supported by the National Science Centre, Poland (grant number 2016/21/B/NZ3/00348) to D.H.-Ł; the Institute of Zoology and Biomedical Research, Jagiellonian University (grant number N18/DBS/000007).

**Conflicts of interest**

None declared.

**References**

1. Bustin,S.A. (2000) Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. J Mol Endocrinol, **25**, 169-193.

2. Goidin,D. et al. (2001) Ribosomal 18S RNA Prevails over Glyceraldehyde-3-Phosphate Dehydrogenase and β-Actin Genes as Internal Standard for Quantitative Comparison of

mRNA Levels in Invasive and Noninvasive Human Melanoma Cell Subpopulations. Anal Biochem, **295**, 17-21.

3. Vandesompele,J. et al. (2002) Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol, **3**(7), research0034.I-0034.

4. Yperman,J. et al. (2004) β-actin cannot be used as a control for gene expression in ovarian interstitial cells derived from heart valves. J Heart Valve Disease, **13**, 848-853.

5. Ryu,B. et al. (2007) Comprehensive Expression Profiling of Tumor Cell Lines Identifies Molecular Signatures of Melanoma Progression. Plos One, **2**(7), e594. doi:10.1371/journal.pone.0000594

6. Ling,D. et al. (2011) Robust RT-qPCR Data Normalization: Validation and Selection of Internal Reference Genes during Post-Experimental Data Analysis. Plos One, **6**(3), e17762. doi: 10.1371/journal.pone.0017762

7. Das,R.K. et al. (2013) Extensive Sex- and/or Hormone-Dependent Expression of Rat Housekeeping Genes. Endocrine Res, **38**(2), 105-111.

8. Zhang,Q.-L. et al. (2016) Selection of reliable reference genes for normalization of quantitative RT-PCR from different developmental stages and tissues in amphioxus. Sci Rep, 6, 37549-37559.

9. Andersen,C.L. et al. (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 Res, **64**, 5245-5250.

10. Pfaffl,M.W. et al. (2004) Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper – Excel-based tool using pair-wise correlations. Biotech Lett, **26**, 509–515.

11. Silver,N. et al. (2006) Selection of housekeeping genes for gene expression studies in human reticulocytes using real-time PCR. BMC Molecular Biology, **7**, 33-41. doi:10.1186/1471-2199-7-33

12. Janik,M.E. et al. (2018) RT-qPCR analysis of human melanoma progression-related genes – A novel workflow for selection and validation of candidate reference genes. Int J Biochem Cell Biol, **101**, 12-18.

13. Bustin,S.A. et al. (2010) MIQE précis: Practical implementation of minimum standard guidelines for fluorescence-based quantitative real-time PCR experiments. BMC Mol Biol, **11**, 74-78. doi:10.1186/1471-2199-11-74

14. Bustin,S. and Nolan,T. (2017) Talking the talk, but not walking the walk: RT-qPCR as a paradigm for the lack of reproducibility in molecular research. Eur J Clin Invest, **47**(10), 756-774. doi:10.1111/eci.12801

15. Dijkstra,J.R. et al. (2014) Critical appraisal of quantitative PCR results in colorectal cancer research: can we rely on published qPCR results? Mol Oncol, **8**(4), 813-818.

16. LaTaylor,S.C. and Mrkusich,E.M. (2014) The state of RT-quantitative PCR: firsthand observations of implementation of minimum information for the publication of quantitative real-time PCR experiments (MIQE). J Mol Microbiol Biotechnol, **24**(1), 46-52. doi:10.1159/000356189

17. Remans,T. et al. (2014) Reliable gene expression analysis by reverse transcription-quantitative PCR: reporting and minimizing the uncertainty in data accuracy. Plant Cell, **26**(10), 3829-3837. doi:10.1105/tpc.114.130641

18. Sarker,N. et al. (2018) Identification of stable reference genes for quantitative PCR in koalas. Scientific Reports, **8**, 3364-3371. doi:10.1038/s41598-018-21723-0

19. Sundaram,V.K. et al. (2019) Optimal use of statistical methods to validate reference gene stability in longitudinal studies. PLoS ONE, **14**(7), e0219440. https://doi.org/10.1371/journal.

20. Kimmerling,R.J. et al. (2018) Linking single-cell measurements of mass, growth rate, and gene expression. Genome Biol, **19**, 207-219.

21. Lemma,S. et al. (2018) Validation of Suitable Housekeeping Genes for the Normalization of mRNA Expression for Studying Tumor Acidosis. Int J Mol Sci,**19**, 2930-2944. doi:10.3390/ijms19102930

22. Hellemans,J. and Vandesompele,J. (2011) qPCR data analysis – unlocking the secret to successful results. [in:] PCR Troubleshooting and Optimization: The Essential Guide (Caister Academic Press), ISBN 978-1-904455-72-1.

**Author Contributions**

D.H-Ł and M.E.J wrote the manuscript, supervised the writing of the software, tested the software and performed experiments. D.M wrote the software and the user manual. D.H-Ł prepared figures and acquired the financial support. All authors have given approval to the final version of the manuscript.