

Identification of Reference Genes across Physiological States for qRT-PCR through Microarray Meta-Analysis

Wei-Chung Cheng¹, Cheng-Wei Chang¹, Chang-Ray Chen¹, Min-Lung Tsai², Wun-Yi Shu³, Chia-Yang Li¹, Ian C. Hsu^{1*}

1 Department of Biomedical Engineering and Environmental Sciences, National Tsing Hua University, Hsinchu, Taiwan, **2** Institute of Athletics, National Taiwan Sport University, Taichung, Taiwan, **3** Institute of Statistics, National Tsing Hua University, Hsinchu, Taiwan

Abstract

Background: The accuracy of quantitative real-time PCR (qRT-PCR) is highly dependent on reliable reference gene(s). Some housekeeping genes which are commonly used for normalization are widely recognized as inappropriate in many experimental conditions. This study aimed to identify reference genes for clinical studies through microarray meta-analysis of human clinical samples.

Methodology/Principal Findings: After uniform data preprocessing and data quality control, 4,804 Affymetrix HU-133A arrays performed by clinical samples were classified into four physiological states with 13 organ/tissue types. We identified a list of reference genes for each organ/tissue types which exhibited stable expression across physiological states. Furthermore, 102 genes identified as reference gene candidates in multiple organ/tissue types were selected for further analysis. These genes have been frequently identified as housekeeping genes in previous studies, and approximately 71% of them fall into Gene Expression (GO:0010467) category in Gene Ontology.

Conclusions/Significance: Based on microarray meta-analysis of human clinical sample arrays, we identified sets of reference gene candidates for various organ/tissue types and then examined the functions of these genes. Additionally, we found that many of the reference genes are functionally related to transcription, RNA processing and translation. According to our results, researchers could select single or multiple reference gene(s) for normalization of qRT-PCR in clinical studies.

Citation: Cheng W-C, Chang C-W, Chen C-R, Tsai M-L, Shu W-Y, et al. (2011) Identification of Reference Genes across Physiological States for qRT-PCR through Microarray Meta-Analysis. PLoS ONE 6(2): e17347. doi:10.1371/journal.pone.0017347

Editor: Joaquín Dopazo, Centro de Investigación Príncipe Felipe, Spain

Received: September 15, 2010; **Accepted:** January 31, 2011; **Published:** February 24, 2011

Copyright: © 2011 Cheng et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This study was funded by the National Science Council (NSC), Taiwan grant number 97-2623-7-007-015-NU (<http://web1.nsc.gov.tw/mp.aspx?mp=7>). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: ichsu@mx.nthu.edu.tw

Introduction

Reference genes (RGs) are widely used to normalize the expression level for removing potential artifacts caused by sample preparation level and detection as well as to provide an accurate comparison of gene expression among different samples. Traditional reference genes (tRGs) are housekeeping genes (HKGs), such as ACTB, GAPDH, and HPRT, and usually serve as internal controls in Northern blot, RNase protection assays, conventional RT-PCR assays, and quantitative real-time PCR (qRT-PCR). The assumption is that these genes are defined as maintaining basic cellular functions [1] and are expressed at a constant level across samples, physiological states, and treatments. However, numerous studies have already shown that tRGs are regulated and their expression levels are varied under certain experimental conditions [2,3,4,5].

qRT-PCR is often considered as the golden standard for quantitative gene expression analysis. However, the use of inappropriate RGs can result in incorrect findings if the expression levels of the chosen RGs are influenced by the experimental conditions [3,6]. Researchers should make sure that the chosen RGs are suitable for the experiment they conducted. Thus,

identification of RGs and their validation within specific biological conditions under investigation are critical issues.

Previous research identified RGs by selecting them from a list of tRGs for specified biological conditions according to the results of qRT-PCR [7,8,9,10,11,12]. Microarray screening is an alternative approach and has the potential to identify novel RGs whose expression levels are more stable than that of tRGs. Moreover, the increasing amount of microarray data is an excellent source for the identification of genes with the most stable expression [13,14,15,16]. Most research using microarray analysis identified RGs for specific biological conditions, for example, evolution [17], differentiation [18], development [19], treatment [20], cancer [13,14,21,22,23,24,25,26], other diseases [27,28,29,30] or comparing different physiological stages of a single organ [21,23,25]. A number of studies have identified RGs with relatively stable expressions across tissue types [31] and among metadata which pooled multitudes of arrays ignoring cell types and experimental conditions [32,33]. However, no results have been reached for a consistent set of RGs. Many researchers assume that no RG is universally stable in its expression in all situations [14,22,23,28,34]. The ideal set of RGs depends on the biological conditions and should be selected and evaluated for each series of experiments.

This study aimed to identify RGs for clinical studies by meta-analysis of human clinical samples. These RGs had to demonstrate a stable expression across various physiological states in individual tissue/organ type. After the removal of poor quality arrays, 4,804 Affymetrix U133A arrays performed on human clinical samples were selected from the M²DB, a microarray meta-analysis database [35]. These arrays were classified into 4 physiological states and 13 organ/tissue types. Genes showing stable expressions within and between physiological stages for a single tissue were identified as RGs for that particular tissue. Our results recommended a number of sets of RGs for various organ/tissue types. Additionally, we have found that the genes that are frequently identified as RGs for multiple organ/tissue types are highly related to the functional category, Gene Expression (GO: 0010467). These genes are frequently classified as HKGs in previous studies. Besides, our results suggest that RGs identified in this study are candidates as control genes for qRT-PCR in clinical studies.

Analysis

Microarray data collection, quality control, and pre-processing

Expression data were collected from the M²DB, which compiles more than 10,000 well-annotated, published, human clinical Affymetrix GeneChip arrays. We excluded poor quality arrays (8% of the total), that did not match the criteria of the 95 percentile of PMVO [36], according to the QC metrics of the M²DB. Then, according to the annotation of the M²DB, samples related to the same organ/tissue type and the same physiological state were classified into a single group. An organ/tissue type was included into this study if it has at least two groups, which contained at least 10 HG-U133A arrays, in the organ/tissue type. In summary, this study included 4,804 HG-U133A arrays classified into 13 organ/tissue types and 4 physiological states (Normal, Abnormality, Disease, and Cancer or Tumor). Table 1 gives the summary of the number of arrays classified in each organ/tissue and physiological state. The data uniformly processed by the GC Robust Multi-array Average (GCRMA) algorithm [37]

were downloaded from the M²DB. Intensities (without log transformation) of the probe sets with the same Entrez GeneID were averaged to represent the expression of the corresponding gene.

Selection of Reference Gene Candidates

The definition of an RG in this study is that a gene stably expressed for each organ across different physiological states. RGs for each organ/tissue type were identified using the following criteria:

1. $\bar{I}_i > 100$ and FP > 80%.
2. $\frac{\sigma_i}{\bar{I}_i} < 0.3$
3. $Max(\bar{I}_i/\bar{I}_j) < 1.2$

Where \bar{I}_i and \bar{I}_j denote the mean intensity of the gene in arrays of *i*th and *j*th physiological states respectively. σ_i is the standard deviation of intensity in *i*th physiological states. Max() is the maximum ratio of mean intensity. For a gene, FP is fraction Present which is the fraction of arrays called present in a single organ/tissue type [38]. The first criterion identified genes that are truly expressed in a tissue. For each gene, the expression values were averaged for each physiological state. A gene was retained if the average expression level exceeded the selected threshold value 100 and FP was larger than 80%. Filtering data by FP increases the correlation between Affymetrix GeneChip and qRT-PCR expression measurements [39]. Genes with their expression values satisfy these two thresholds are most likely to be truly expressed. The second criterion used the coefficient of variation, standard deviation divided by mean intensity, to verify whether the genes exhibited stable expressions in a physiological state. The third criterion used fold change of expression to filter out genes that differentially expressed across physiological states in a single organ/tissue type. The fold change refers to the ratio of mean intensity of physiological states and represents the expression differences between physiological states. Table 2 shows the number of genes which are stably expressed within individual physiological state (the first and second criteria), stably expressed

Table 1. Summary of arrays classified into 4 physiological states and 13 organ/tissue types.

Organ/Tissue Types	Physiological States				Total
	Normal	Cancer or Tumor	Disease	Abnormality	
blood	252	403	514	137	1,306
lung	44	92	66	128	330
bone marrow	39	559	19	0	617
brain	229	139	202	0	570
uterus	18	15	13	0	46
breast	10	1,229	0	3	1,242
kidney	9	21	0	10	40
bladder	12	64	0	0	76
lymph node	12	33	0	0	45
prostate	15	59	0	0	74
testis	17	102	0	0	119
muscle	75	0	109	44	228
heart	30	0	81	0	111
Total	762	2,716	1,004	322	4,804

doi:10.1371/journal.pone.0017347.t001

Table 2. Summary of the number of genes passed different criteria in 13 organ/tissue types.

Organ/Tissue types	Stable Within Physiological States*				Stable Between Physiological States [§]	RGs [¶]
	Normal	Cancer or Tumor	Disease	Abnormality		
blood	133	203	479	238	162	11
lung	211	117	768	581	195	16
bone marrow	186	66	382	-	301	21
brain	184	60	491	-	657	15
uterus	761	1,454	1,548	-	1,271	276
breast	352	108	-	2,263	378	17
kidney	201	1,041	-	2,542	421	31
bladder	362	200	-	-	1,385	89
lymph node	2,212	495	-	-	1,030	150
prostate	734	106	-	-	1,989	65
testis	238	173	-	-	713	13
muscle	327	-	198	478	1,103	93
heart	742	-	794	-	2,406	250

*The criterion is that the CV of the intensity of the gene in the physiological state is smaller than 30%. CV, coefficient of variation, is equal to standard deviation divided by mean.

§The criterion is that the maximum of the fold change of mean intensity between physiological states is smaller than 1.2.

¶The number of genes stably expressed within and between physiological states.

doi:10.1371/journal.pone.0017347.t002

across physiological states (the first and third criteria), and qualified as RGs (all three criteria) for each organ/tissue type. For example, by apply the first two criteria, the counts of genes stably expressed within the four physiological states in blood are 133, 203, 479, and 238, respectively. By applying the first and third criteria, there were 162 genes stably expressed across physiological states in blood. Finally, 11 genes passed all three criteria were identified as RGs for blood. Data S1 gives the complete lists of RGs for respective organ/tissue types.

Frequent Reference Genes

The genes which were identified as RGs for at least three organ/tissue types are denoted as frequent reference genes (fRGs). Table 3 displays a list of 102 fRGs and the corresponding numbers of organ/tissue types for which the RGs were identified. Some tRGs, such as ACTB, B2M, UBC, RPL13A and RPLP0, are also on this list. Gene ontology was used to analyze the gene function of fRGs. A set of GO terms (14 terms) was chosen to give a broad overview of gene function. Figure S1 generated by QuickGO [40] is a graphical view of the term lineage of these 14 terms in Gene Ontology. Figure 1 shows the percentage of fRGs counts in these 14 terms. Approximately 61%, 15%, and 7% of fRGs belong to Translation (GO: 0006412), RNA Processing (GO: 0006396), and Transcription (GO: 0006350) respectively. Moreover, these three terms are children of Gene Expression (GO: 0010467) (Figure S1). Approximately 71% of the fRGs fall into this functional category. These are basic cellular functions referring to HKGs. When compared with 8 lists of HKGs identified by microarray or EST analysis in 7 previous studies [16,41,42,43,44,45,46], fRGs were frequently classified as HKGs in these lists. Furthermore, the percentages of these HKGs lists falling into Gene Expression (GO: 0010467) range from 22.4 to 35.1 (Table 4). These percentages are much lower than that of fRGs. In addition, these 14 terms cover 84% of fRGs. The other 16% of fRGs do not belong to these 14 GO terms, and half of these genes do not refer to any GO terms.

Expression profiles of tRGs and fRGs

Six tRGs and six fRGs were selected to examine the expression profiles. The 6 housekeeping genes (ACTB, B2M, GAPDH, PKG1, RPLP0, and PPIA) have been commonly used as reference genes for qRT-PCR in numerous studies. In this study, the 6 fRGs (HUWE1, TPT1, EEF1A1, LRRC40, RPS20, RPL37A, and RPL41) are the most frequently identified RGs in various organ/tissue types (Table 3). Three of the housekeeping genes, ACTB, B2M, and RPLP0, are also identified as fRGs. Although the other three housekeeping gene are not fRGs, they are still identified as RGs for one or two organs/tissue types. Figure 2 depicts the intensity profile of the 12 genes (6 tRGs and 6 fRGs) in various physiological states of 13 organ/tissue types. The RGs exhibit consistent expressions in the corresponding organ/tissue type. The 6 fRGs exhibit more stable expression than the 6 tRGs do both within and between organ/tissue types.

Discussion

We examined the variability of gene expression within and between various physiological states in 13 organ/tissue types. Lists of RGs were identified for the corresponding organ/tissue types. Clinical research usually focused on various physiological states for a single organ/tissue type (such as cancer classification [47,48,49]). The relative expression level of an ideal RG for clinical studies should not be significantly influenced by physiological states. Previous studies, which used microarray screening to identify RGs, mostly focused on a specific physiological state in an organ/tissue type. Some research identified universal RGs by pooling all of microarray data from public repositories ignoring organ/tissue types and physiological states [31,32,33]. Different from them, our study broadly searched RGs in various physiological stages of 13 organ/tissue types. To achieve this goal, we classified samples into four physiological states according to information found in the M²DB. Then, we applied several criteria to identify expressed genes with consistent expression within and between physiological

Table 3. fRGs and the corresponding numbers of organ/tissue types for which fRGs were identified.

Num. of organ/ tissue types	Gene Symbol
3	SEPT2, ATG4B, B2M, BTF3, DAZAP2, DDB1, DDX17, EIF4G2, ENSA, EWSR1, FNTA, HDAC3, HDLBP, HMGB1, HMG2, HNRNPA1, MORF4L1, MTCH1, PI4KB, PUM1, RPL10, RPL11, RPL17, RPL19, RPL22, RPL4, RPL5, RPS12, RPS15, RPS28, RPS3, RPS7, TBC1D9B, TCEB3
4	ACTB, CCDC72, EEF1G, EEF2, FTHP1, GDI1, GTF2F1, GTPBP6, RPL12, RPL24, RPL27A, RPL30, RPL37, RPL38, RPL39, RPL7, RPL7A, RPLP0, RPLP1, RPS16, RPS2, RPS24, RPS25, RPS27A, RPS3A, RPS4X, RPSA, SKP1, SNRPB2, SRP14, USP34
5	ACTG1, EEF1D, EIF1, MYL12B, OAZ1, RPL13A, RPL15, RPL21, RPL27, RPL31, RPL32, RPS13, RPS14, RPS15A, TOX4, UBA52
6	PNN, RPL34, RPL9, RPS10, RPS11, RPS17, RPS18, RPS23, RPS27, RPS29, UBB, UBC
7	NACAP1, RPL23A
8	EEF1A1, LRRC40, RPS20
9	RPL41
10	RPL37A, TPT1
11	HUWE1

doi:10.1371/journal.pone.0017347.t003

states as RGs. Genes satisfied these selection criteria indeed exhibited stable expression and results indicated that the tRGs are not always the best choice for reference of qRT-PCR (Figure 2). Although numbers of genes in our RG list had been reported as

RGs for some experimental conditions in previous studies, our results specified which gene could be RG in particular organ/tissue types. For example, ACTB, the most frequently used tRG, is also in our fRGs list, but we suggested that ACTB can only be

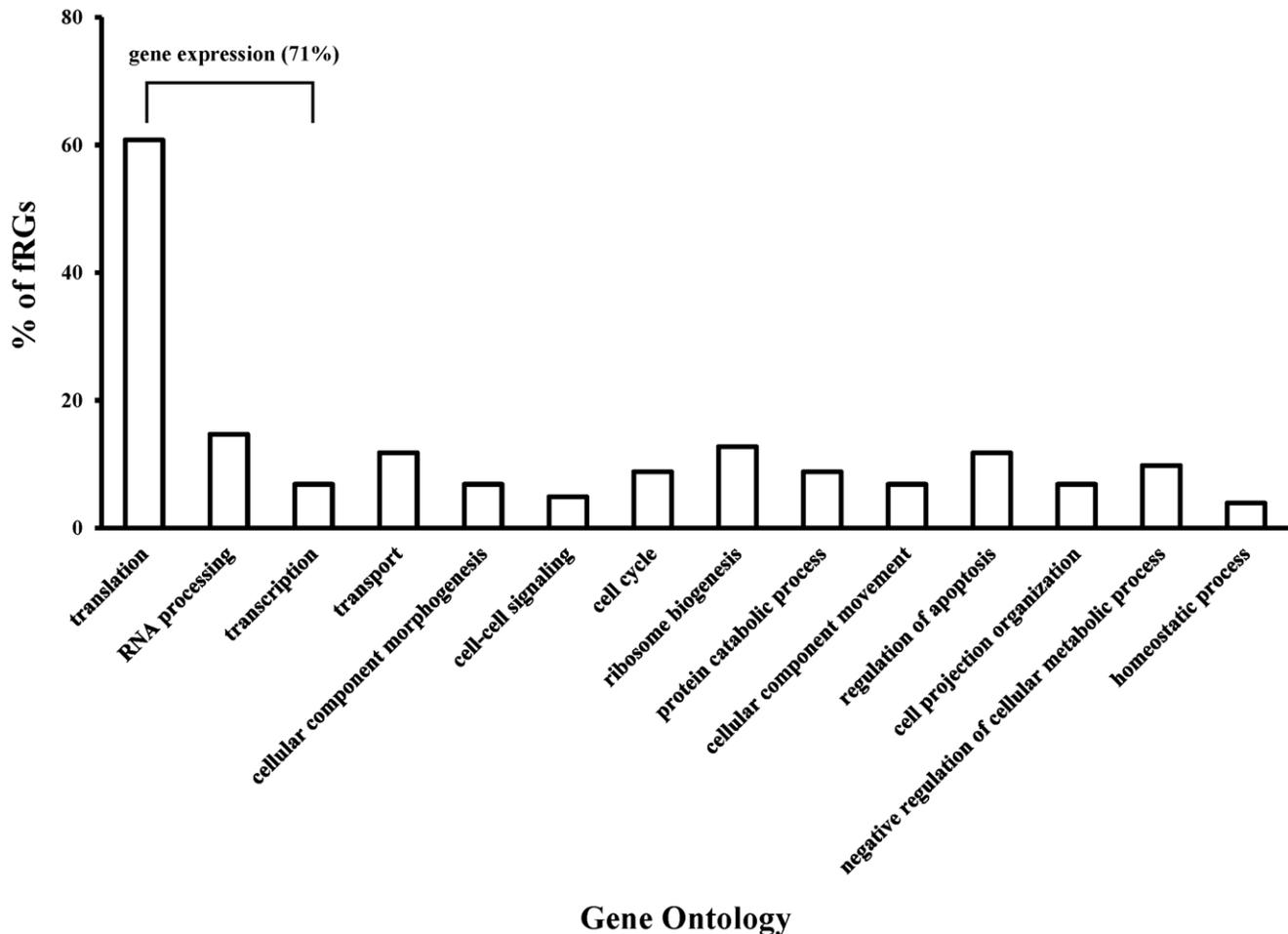


Figure 1. Gene Ontology Functional analysis of fRGs. The percentage of fRGs counted in 14 GO terms which give a broad overview of gene function. Gene expression is the parent term of transcription, translation, and RNA processing in Gene Ontology and contains 71% of fRGs.
doi:10.1371/journal.pone.0017347.g001

Table 4. Comparison of fRGs with HKG lists of previous studies.

References	Tech.	% of overlap [*]	% in Gene Expression (GO: 0010467) [§]
Warrington et al. 2000 [16]	Microarray	59.8	31.9
Hsiao et al. 2001 [46]	Microarray	58.8	35.1
Eisenberg et al. 2003 [45]	Microarray	43.1	27.2
Tu et al. 2006 [44]	Microarray	75.5	22.4
Zhu et al. 2008 [42]	EST	92.2	24.5
Zhu et al. 2008 [42]	Microarray	85.3	26.1
Dezso et al. 2008 [43]	Microarray	81.4	24.3
She et al. 2009 [41]	Microarray	68.6	29.1
fRGs	Microarray	-	70.6

^{*}The percentage of fRGs falls into HKG lists.

[§]The percentage of genes in these lists falls into Gene Expression (GO: 0010467) category in Gene Ontology.
doi:10.1371/journal.pone.0017347.t004

served as RG in three organ/tissue types out of the total thirteen organ/tissue types which we investigated. Furthermore, unlike some previous studies, our results indicated that there is no universal RG for all experimental conditions listed in our study. As the result, it also shows that choosing randomly any HKGs for normalization is risky and may lead to erroneous results.

With rapidly accumulating metadata, microarray meta-analysis is becoming more important in microarray research. One major concern is that as more datasets are included into analysis, the more variance could contribute to the result. Ramasamy et al. had suggested several key issues for microarray meta-analysis [50]. Using pre-processed data based on different algorithms will introduce variations into meta-analysis and the resulting data are unlikely to be directly comparable. As Ramasamy et al. point out, even for studies conducted using the same microarray platform; the raw data should be uniformly pre-processed and normalized using the same algorithm to remove systematic biases for all tested datasets. Several studies have suggested considering data quality within the context of microarray meta-analysis [50,51,52]. Poor quality data must be identified and eliminated during data processing [50,53]. In this study, we adopted single platform for analysis to avoid the variance of combining different platforms, and then uniformly pre-processed all arrays to eliminate the technical variance of data transformation and removed poor quality arrays to alleviate laboratory-to-laboratory variance [35]. Moreover, we used the 12 tRGs and fRGs in Figure 2 to evaluate the effect of QC (Figure S2). The CV of intensity for these genes with QC was lower compared to those without QC. This result suggests that including poor quality arrays could lead to increase expression variation. The advantageous effect of excluding poor quality data is apparent when processing muscle tissues. More than 40% of muscle sample arrays were identified as poor quality arrays (8% of total arrays are poor quality). This result shows that expression variations of RGs are greatly reduced when poor quality muscle arrays were excluded.

Most genes included in lists of fRGs were commonly referred to as HKGs in previous studies (Table 4). To a certain extent, this result is in line with the original concept of using HKGs as RGs for normalization. However, contrary to commonly held assumptions, no HKGs were consistently expressed across all tissues in our study. Moreover, no genes maintained a stable expression level under all conditions (various organ/tissue types and physiological states) (Table 4). In fact, this observation has been mentioned in previous studies [14,22,23,28,34] which presumed there is no universal RGs

for all experimental conditions. Furthermore, approximately 71% of fRGs' were related to the function of Gene Expression (including Transcription, Translation, and RNA Processing). The percentage is much higher than those of HKGs lists by previous studies (Table 4). Consequently, fRGs are highly related to HKGs and maintained at relatively stable level. This result indicates that the genes in the Gene Expression (GO: 0010467) category are more likely to be stably expressed across physiological states and organ/tissue types. This may imply these genes play more important roles than general HKGs. Besides, we found that half of the fRGs were ribosomal protein genes. A meta-analysis study conducted by de Jonge et al. revealed 15 reference genes with the most constant expression, and 13 out of 15 genes were ribosomal proteins [32]. In contrast, Thorrez et al. demonstrated that ribosomal protein genes exhibited important tissue-dependent variations in mRNA expression [54]. Thorrez's results were based on the study of 70 microarrays, representing 22 tissues. The authors cautioned against using ribosomal protein genes as a reference [54]. Our study, which preserved more sample conditions, resolves the contradictory conclusions by these two studies. Our results depicts that some ribosomal protein genes maintained relative stability of expression across organ/tissue types, however, some ribosomal proteins exhibited significant tissue-dependent expression (for example, RPLP0 in Figure 2). The RGs identified in this study expressed stably across physiological states in a single organ/tissue type. Thus, a number of ribosomal protein genes tallied with the criterion could be identified as RGs. For example, in this study, more than half of RGs for breast are ribosomal protein genes, which is consistent with the results of a meta-analysis to identify RGs for breast cancer [26]. However, if the experiment is conducted by various organ/tissue types, it required further verification to use ribosomal protein genes as reference.

UBB, UBC, and UBA52 in the list of fRGs are known as functions related to protein ubiquitination, as well as numerous essential cellular functions. They have been identified as RGs in breast cancer [26]. UBC is a tRG and has also been identified as an RG in colon cancer [14]. TPT1 was initially described as a growth-related protein, and it was recently shown being involved in calcium homeostasis [55]. This implies the expression stability of TPT1 could influence the calcium stability in cells. It could be the reason that TPT1 was identified as RG in previous studies [14,29] and for 10 organ/tissue types in this study. RPL41 and EEFA1 in the list of fRGs have also been recognized as RG for liver [23] and myocardium [29] respectively. GAPDH, the most common tRG, was identified as RG only for heart and muscle in

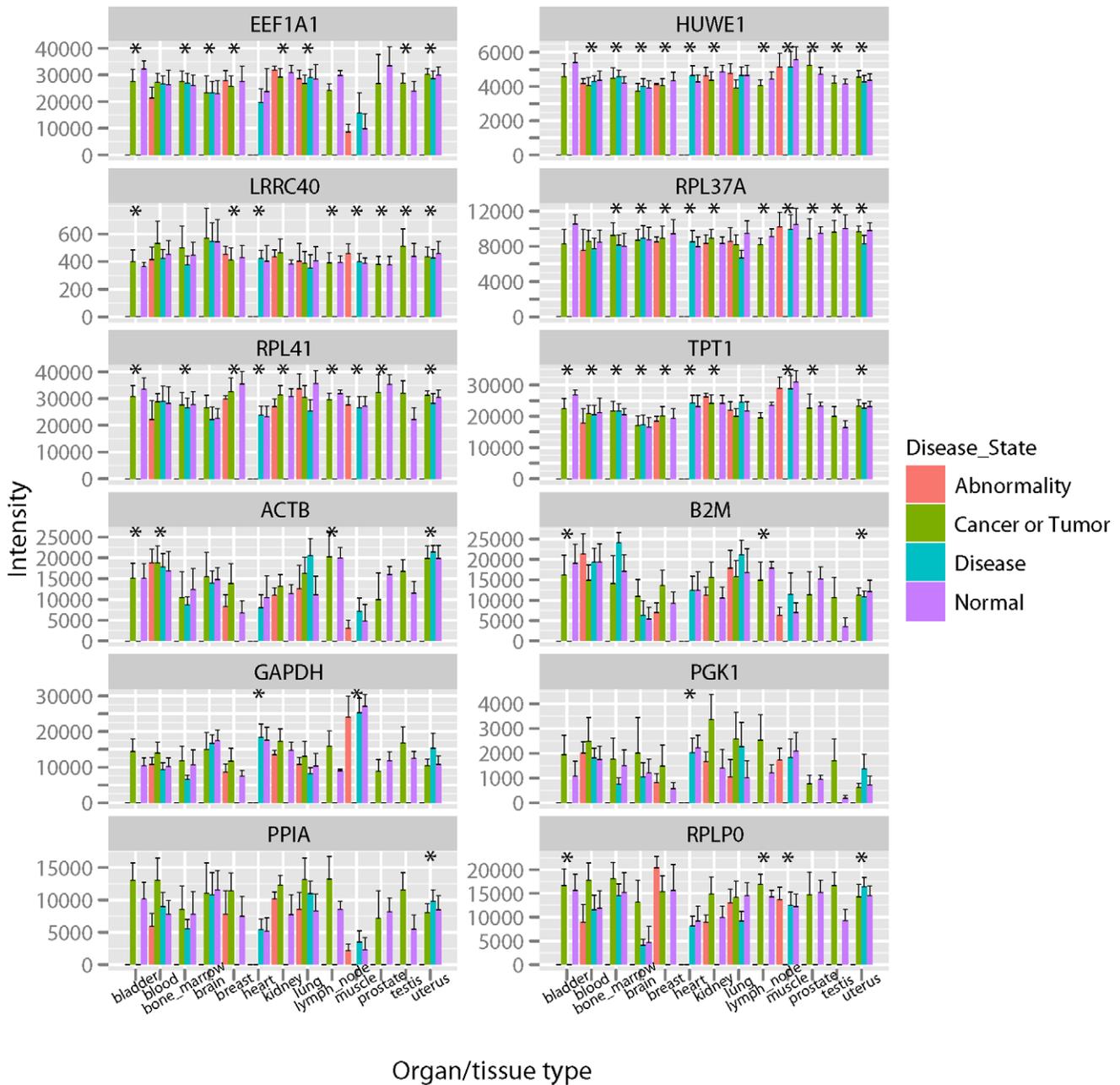


Figure 2. Expression profiles of 6 tRGs and 6 fRGs for 4 physiological states in 13 organ/tissue types. The upper and lower halves of the figure are 6 fRGs and 6 tRGs respectively. The error bar is the standard deviation of intensity. * denotes the gene identified as RG in the organ/tissue type. doi:10.1371/journal.pone.0017347.g002

this study, but this is partially consistent with the previous study which identified GAPDH as a RG for myocardium [29]. HUWE1, which is related to histone ubiquitination [56] and protein polyubiquitination [57], was the top-ranked RG in our result. Although HUWE1 was not the most stable gene in individual organ/tissue type, it was the gene most frequently identified as RG in this study, and suggested to be a novel RG candidate for clinical studies.

Geometric averaging of multiple RGs rather than using single RG for normalization of qRT-PCR is an alternative strategy [58]. We have supplied lists of RG candidates for researchers to confirm their qRT-PCR results under particular experimental conditions.

Choosing several RG candidates from our RG lists to perform qRT-PCR could help researchers to confirm one or multiple RGs for use as references.

For some organ/tissue types, there were only dozens of samples for identifying RGs, despite the thousands of arrays included in this study (Table 1). This might underestimate the variance of expression among individuals or physiological conditions and might lead to increased false positive rate. For example, 276 RG candidates were identified for the uterus (Table 2). There is a limitation of accuracy in identifying RG upon small number of samples. However, our RG list can be good candidates for researchers to identify the true RG by qRT-PCR but not choosing

HKGs randomly as reference. Researchers can exclude unsuitable RGs which had been shown variable expression in our results. Using the same example, the most used tRG, GAPDH, is not included in the 276 RG candidates for the uterus. Thus, researchers could choose several candidate genes in our list for further validation by qRT-PCR but GAPDH. In the future, with rapidly accumulated microarray metadata, we could gather more clinical arrays and subdivide them by detailed physiological states and organ/tissue types. Accordingly, the more accurate RGs could be identified for clinical studies.

In summary, this study performed microarray meta-analysis to compile lists of RG candidate for 13 organ/tissue types. We provided lists of RG candidates for researchers to select single or multiple genes as references for the normalization of qRT-PCR in clinical studies. We also found that fRGs were recognized as HKGs in previously studies and about 71% of fRGs were functional annotated to Gene Expression (GO:0010467). The percentage is also much higher than that of HKG lists. To our best knowledge, this is the first study considering different physiological states as well as identifying RGs for various organ/tissue types. In our results, the tRGs are not the best choice for reference of qRT-PCR in most conditions, and the RGs identified in this study are more reliable than tRGs for normalization in qRT-PCR for clinical studies.

References

- Butte AJ, Dzau VJ, Glueck SB (2001) Further defining housekeeping, or "maintenance," genes Focus on "A compendium of gene expression in normal human tissues". *Physiol Genomics* 7: 95–96.
- Wu YY, Rees JL (2000) Variation in epidermal housekeeping gene expression in different pathological states. *Acta Derm Venereol* 80: 2–3.
- Tricarico C, Pinzani P, Bianchi S, Paglierani M, Distante V, et al. (2002) Quantitative real-time reverse transcription polymerase chain reaction: normalization to rRNA or single housekeeping genes is inappropriate for human tissue biopsies. *Anal Biochem* 309: 293–300.
- Beillard E, Pallisgaard N, van der Velden VH, Bi W, Dee R, et al. (2003) Evaluation of candidate control genes for diagnosis and residual disease detection in leukemic patients using 'real-time' quantitative reverse-transcriptase polymerase chain reaction (RQ-PCR) - a Europe against cancer program. *Leukemia* 17: 2474–2486.
- Rubie C, Kempf K, Hans J, Su T, Tilton B, et al. (2005) Housekeeping gene variability in normal and cancerous colorectal, pancreatic, esophageal, gastric and hepatic tissues. *Mol Cell Probes* 19: 101–109.
- Bas A, Forsberg G, Hammarstrom S, Hammarstrom ML (2004) Utility of the housekeeping genes 18S rRNA, beta-actin and glyceraldehyde-3-phosphate-dehydrogenase for normalization in real-time quantitative reverse transcriptase-polymerase chain reaction analysis of gene expression in human T lymphocytes. *Scand J Immunol* 59: 566–573.
- Erkens T, Van Poucke M, Vandesompele J, Goossens K, Van Zeveren A, et al. (2006) Development of a new set of reference genes for normalization of real-time RT-PCR data of porcine backfat and longissimus dorsi muscle, and evaluation with PPAR α . *BMC Biotechnol* 6: 41.
- Cicinatti VR, Shen Q, Sotiropoulos GC, Radtke A, Gerken G, et al. (2008) Validation of putative reference genes for gene expression studies in human hepatocellular carcinoma using real-time quantitative RT-PCR. *BMC Cancer* 8: 350.
- Fu LY, Jia HL, Dong QZ, Wu JC, Zhao Y, et al. (2009) Suitable reference genes for real-time PCR in human HBV-related hepatocellular carcinoma with different clinical prognoses. *BMC Cancer* 9: 49.
- Lyng MB, Laenkholm AV, Pallisgaard N, Ditzel HJ (2008) Identification of genes for normalization of real-time RT-PCR data in breast carcinomas. *BMC Cancer* 8: 20.
- Coulson DT, Brockbank S, Quinn JG, Murphy S, Ravid R, et al. (2008) Identification of valid reference genes for the normalization of RT qPCR gene expression data in human brain tissue. *BMC Mol Biol* 9: 46.
- Exposito-Rodriguez M, Borges AA, Borges-Perez A, Perez JA (2008) Selection of internal control genes for quantitative real-time RT-PCR studies during tomato development process. *BMC Plant Biol* 8: 131.
- Saviozzi S, Cordero F, Lo Iacono M, Novello S, Scagliotti GV, et al. (2006) Selection of suitable reference genes for accurate normalization of gene expression profile studies in non-small cell lung cancer. *BMC Cancer* 6: 200.
- Andersen CL, Jensen JL, Orntoft 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 Res* 64: 5245–5250.
- Schmid H, Cohen CD, Henger A, Irrgang S, Schlondorff D, et al. (2003) Validation of endogenous controls for gene expression analysis in microdissected human renal biopsies. *Kidney Int* 64: 356–360.
- Warrington JA, Nair A, Mahadevappa M, Tsyganskaya M (2000) Comparison of human adult and fetal expression and identification of 535 housekeeping/maintenance genes. *Physiol Genomics* 2: 143–147.
- Fedrigo O, Warner LR, Pfefferle AD, Babbitt CC, Cruz-Gordillo P, et al. (2010) A pipeline to determine RT-QPCR control genes for evolutionary studies: application to primate gene expression across multiple tissues. *PLoS One* 5.
- Hamalainen HK, Tubman JC, Vikman S, Kyrölä T, Ylikoski E, et al. (2001) Identification and validation of endogenous reference genes for expression profiling of T helper cell differentiation by quantitative real-time RT-PCR. *Anal Biochem* 299: 63–70.
- Narsai R, Inanova A, Ng S, Whelan J (2010) Defining reference genes in *Oryza sativa* using organ, development, biotic and abiotic transcriptome datasets. *BMC Plant Biol* 10: 56.
- Zhou L, Lim QE, Wan G, Too HP (2010) Normalization with genes encoding ribosomal proteins but not GAPDH provides an accurate quantification of gene expressions in neuronal differentiation of PC12 cells. *BMC Genomics* 11: 75.
- Kidd M, Nadler B, Mane S, Eick G, Malfertheiner M, et al. (2007) GeneChip, geNorm, and gastrointestinal tumors: novel reference genes for real-time PCR. *Physiol Genomics* 30: 363–370.
- Su IJ, Chang CW, Wu YC, Chen KC, Lin CJ, et al. (2007) Selection of DDX5 as a novel internal control for Q-RT-PCR from microarray data using a block bootstrap re-sampling scheme. *BMC Genomics* 8: 140.
- Waxman S, Wurbach E (2007) De-regulation of common housekeeping genes in hepatocellular carcinoma. *BMC Genomics* 8: 243.
- Nguewa PA, Agorreta J, Blanco D, Lozano MD, Gomez-Roman J, et al. (2008) Identification of importin 8 (IPO8) as the most accurate reference gene for the clinicopathological analysis of lung specimens. *BMC Mol Biol* 9: 103.
- Gur-Dedeoglu B, Konu O, Bozkurt B, Ergul G, Seckin S, et al. (2009) Identification of endogenous reference genes for qRT-PCR analysis in normal matched breast tumor tissues. *Oncol Res* 17: 353–365.
- Popovici V, Goldstein DR, Antonov J, Jaggi R, Delorenzi M, et al. (2009) Selecting control genes for RT-QPCR using public microarray data. *BMC Bioinformatics* 10: 42.
- Shulzhenko N, Yambartsev A, Goncalves-Primo A, Gerbase-DeLima M, Morgun A (2005) Selection of control genes for quantitative RT-PCR based on microarray data. *Biochem Biophys Res Commun* 337: 306–312.
- Maccoux IJ, Clements DN, Salway F, Day PJ (2007) Identification of new reference genes for the normalisation of canine osteoarthritic joint tissue transcripts from microarray data. *BMC Mol Biol* 8: 62.
- Pilbrow AP, Ellmers IJ, Black MA, Moravec CS, Sweet WE, et al. (2008) Genomic selection of reference genes for real-time PCR in human myocardium. *BMC Med Genomics* 1: 64.
- Folkersen L, Kurtovic S, Razuvaev A, Agardh HE, Gabrielsen A, et al. (2009) Endogenous control genes in complex vascular tissue samples. *BMC Genomics* 10: 516.

Supporting Information

Data S1 The complete lists of RGs for the 13 organ/tissue types. For each gene, the CV and mean intensity of various physiological states are also included in this file.
(XLS)

Figure S1 The lineage of 14 GO terms.
(TIF)

Figure S2 The CV of intensity of 12 genes in 13 organ/tissue types with/without QC filtering.
(TIF)

Acknowledgments

We are grateful to the National Center for High-performance Computing for computer time and facilities.

Author Contributions

Conceived and designed the experiments: WC MT. Analyzed the data: WC WS CRC CWC. Wrote the manuscript: WC IH. Edited the paper: IH CL CRC CWC.

31. Lee S, Jo M, Lee J, Koh SS, Kim S (2007) Identification of novel universal housekeeping genes by statistical analysis of microarray data. *J Biochem Mol Biol* 40: 226–231.
32. de Jonge HJ, Fehrmann RS, de Bont ES, Hofstra RM, Gerbens F, et al. (2007) Evidence based selection of housekeeping genes. *PLoS One* 2: e898.
33. Kwon MJ, Oh E, Lee S, Roh MR, Kim SE, et al. (2009) Identification of novel reference genes using multiplatform expression data and their validation for quantitative gene expression analysis. *PLoS One* 4: e6162.
34. Lee PD, Sladek R, Greenwood CM, Hudson TJ (2002) Control genes and variability: absence of ubiquitous reference transcripts in diverse mammalian expression studies. *Genome Res* 12: 292–297.
35. Cheng WC, Tsai ML, Chang CW, Huang CL, Chen CR, et al. (2010) Microarray meta-analysis database (M2DB): a uniformly pre-processed, quality controlled, and manually curated human clinical microarray database. *BMC Bioinformatics* 11: 421.
36. Asare AL, Gao Z, Carey VJ, Wang R, Seyfert-Margolis V (2009) Power enhancement via multivariate outlier testing with gene expression arrays. *Bioinformatics* 25: 48–53.
37. Wu ZJ, Irizarry RA, Gentleman R, Martinez-Murillo F, Spencer F (2004) A model-based background adjustment for oligonucleotide expression arrays. *Journal of the American Statistical Association* 99: 909–917.
38. McClintick JN, Edenberg HJ (2006) Effects of filtering by Present call on analysis of microarray experiments. *BMC Bioinformatics* 7: 49.
39. Mieczkowski J, Tyburczy ME, Dabrowski M, Pokarowski P (2010) Probe set filtering increases correlation between Affymetrix GeneChip and qRT-PCR expression measurements. *BMC Bioinformatics* 11: 104.
40. Binns D, Dimmer E, Huntley R, Barrell D, O'Donovan C, et al. (2009) QuickGO: a web-based tool for Gene Ontology searching. *Bioinformatics* 25: 3045–3046.
41. She X, Rohl CA, Castle JC, Kulkarni AV, Johnson JM, et al. (2009) Definition, conservation and epigenetics of housekeeping and tissue-enriched genes. *BMC Genomics* 10: 269.
42. Zhu J, He F, Song S, Wang J, Yu J (2008) How many human genes can be defined as housekeeping with current expression data? *BMC Genomics* 9: 172.
43. Dezso Z, Nikolsky Y, Sviridov E, Shi W, Serebriyskaya T, et al. (2008) A comprehensive functional analysis of tissue specificity of human gene expression. *BMC Biol* 6: 49.
44. Tu Z, Wang L, Xu M, Zhou X, Chen T, et al. (2006) Further understanding human disease genes by comparing with housekeeping genes and other genes. *BMC Genomics* 7: 31.
45. Eisenberg E, Levanon EY (2003) Human housekeeping genes are compact. *Trends Genet* 19: 362–365.
46. Hsiao LL, Dangond F, Yoshida T, Hong R, Jensen RV, et al. (2001) A compendium of gene expression in normal human tissues. *Physiol Genomics* 7: 97–104.
47. Wan Y-W, Sabbagh E, Raese R, Qian Y, Luo D, et al. (2010) Hybrid Models Identified a 12-Genes Signature for Lung Cancer Prognosis and Chemoresponse Prediction. *PLoS One* 5: e12222.
48. Espinosa E, Sánchez-Navarro I, Gámez-Pozo A, Marin AP, Hardisson D, et al. (2009) Comparison of Prognostic Gene Profiles Using qRT-PCR in Paraffin Samples: A Retrospective Study in Patients with Early Breast Cancer. *PLoS One* 4: e5911.
49. Rizzi F, Belloni L, Crafa P, Lazzaretti M, Remondini D, et al. (2008) A Novel Gene Signature for Molecular Diagnosis of Human Prostate Cancer by RT-qPCR. *PLoS One* 3: e3617.
50. Ramasamy A, Mondry A, Holmes CC, Altman DG (2008) Key Issues in Conducting a Meta-Analysis of Gene Expression Microarray Datasets. *PLoS Med* 5: e184.
51. Owzar K, Barry WT, Jung SH, Sohn I, George SL (2008) Statistical challenges in preprocessing in microarray experiments in cancer. *Clin Cancer Res* 14: 5959–5966.
52. Cahan P, Rovegno F, Mooney D, Newman JC, St Laurent G, 3rd, et al. (2007) Meta-analysis of microarray results: challenges, opportunities, and recommendations for standardization. *Gene* 401: 12–18.
53. Larsson O, Sandberg R (2006) Lack of correct data format and comparability limits future integrative microarray research. *Nat Biotechnol* 24: 1322–1323.
54. Thorrez L, Van Deun K, Tranchevent LC, Van Lommel L, Engelen K, et al. (2008) Using ribosomal protein genes as reference: a tale of caution. *PLoS One* 3: e1854.
55. Arcuri F, Papa S, Meini A, Carducci A, Romagnoli R, et al. (2005) The translationally controlled tumor protein is a novel calcium binding protein of the human placenta and regulates calcium handling in trophoblast cells. *Biol Reprod* 73: 745–751.
56. Liu Z, Oughtred R, Wing SS (2005) Characterization of E3Histone, a novel testis ubiquitin protein ligase which ubiquitinates histones. *Mol Cell Biol* 25: 2819–2831.
57. Zhong Q, Gao W, Du F, Wang X (2005) Mule/ARF-BP1, a BH3-only E3 ubiquitin ligase, catalyzes the polyubiquitination of Mcl-1 and regulates apoptosis. *Cell* 121: 1085–1095.
58. Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, et al. (2002) Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol* 3: RESEARCH0034.