

RESEARCH

Open Access



Conventionally used reference genes are not outstanding for normalization of gene expression in human cancer research

Jihoon Jo^{1†}, Sunkyoung Choi^{2†}, Jooseong Oh¹, Sung-Gwon Lee¹, Song Yi Choi^{3*}, Kee K. Kim^{2*} and Chungoo Park^{1*}

From International Workshop on Data and Text Mining in Biomedical Informatics
Turin, Italy. 22 - 26 October 2018

Abstract

Background: The selection of reference genes is essential for quantifying gene expression. Theoretically they should be expressed stably and not regulated by experimental or pathological conditions. However, identification and validation of reference genes for human cancer research are still being regarded as a critical point, because cancerous tissues often represent genetic instability and heterogeneity. Recent pan-cancer studies have demonstrated the importance of the appropriate selection of reference genes for use as internal controls for the normalization of gene expression; however, no stably expressed, consensus reference genes valid for a range of different human cancers have yet been identified.

Results: In the present study, we used large-scale cancer gene expression datasets from The Cancer Genome Atlas (TCGA) database, which contains 10,028 (9,364 cancerous and 664 normal) samples from 32 different cancer types, to confirm that the expression of the most commonly used reference genes is not consistent across a range of cancer types. Furthermore, we identified 38 novel candidate reference genes for the normalization of gene expression, independent of cancer type. These genes were found to be highly expressed and highly connected to relevant gene networks, and to be enriched in transcription-translation regulation processes. The expression stability of the newly identified reference genes across 29 cancerous and matched normal tissues were validated via quantitative reverse transcription PCR (RT-qPCR).

Conclusions: We reveal that most commonly used reference genes in current cancer studies cannot be appropriate to serve as representative control genes for quantifying cancer-related gene expression levels, and propose in this study three potential reference genes (*HNRNPL*, *PCBP1*, and *RER1*) to be the most stably expressed across various cancerous and normal human tissues.

Keywords: RT-qPCR, Reference genes, Human cancer

Background

To understand how genetic alterations driving tumorigenesis lead to the formation of complex cellular networks and induce biological process variation, recent

research into cancer genetics has focused on the identification of molecular differences between cancerous and normal tissues [1, 2]. Recent high-throughput transcriptionomic studies [3] have offered the opportunity to explore the molecular complexity of human cancer, and have provided evidence for classifying human cancer data into normal, benign, and malignant classes, based on their gene expression patterns. Nevertheless, the expression levels of transcriptionally identified candidate cancer genes require experimental verification via molecular methods such as quantitative reverse transcription PCR

* Correspondence: sychoi@cnu.ac.kr; kimkk@cnu.ac.kr; chungoo@jnu.ac.kr

[†]Jihoon Jo and Sunkyoung Choi contributed equally to this work.

³Department of Pathology, Chungnam National University, 282 Munhwa-Ro, Jung-Ku, Daejeon 35015, Republic of Korea

²Department of Biochemistry, Chungnam National University, 99 Daehak-Ro, Yuseong-Ku, Daejeon 34134, Republic of Korea

¹School of Biological Sciences and Technology, Chonnam National University, 77 Yongbong-Ro, Buk-Ku, Gwangju 61186, Republic of Korea



(RT-qPCR). One of the most important factors ensuring the accuracy of RT-qPCR analyses is the normalization of the identified target-gene expression level to that of a consistently expressed reference gene. To date, cancer researchers have predominantly used the *GAPDH* and *β-actin* reference genes as internal reference controls, because their mRNA expression levels are established to be high and constant in many different cells and tissues [4, 5]. However, cancerous tissues often exhibit a higher level of gene expression variability than normal tissues, due to tumor heterogeneity, genetic instability, and the fact that genetic alterations in diverse cancer types may differentially affect cellular processes at the transcriptome level. Thus, it is a challenging to determine which reference genes would best serve as internal reference controls for a range of different human cancers. Indeed, an increasing number of researches have shown the striking expression variability of known reference genes in human cancers, and subsequently recommended novel reference genes for gene expression studies in each specific human cancer type [6, 7]. These efforts with in silico analysis (e.g., geNorm, NormFinder, and Bestkeeper [8–10]) are ongoing; however, to date, no transcriptome-wide analysis for the identification of the most stably expressed consensus reference genes has been reported.

The primary objective of the present study was to conduct a screen for the most stable reference genes for the study of cancer gene expression. We exploited large-scale gene expression data from The Cancer Genome Atlas (TCGA) database, which contains 10,028 (9,364 cancerous and 664 normal) samples from 32

different cancer types. We identified novel reference genes that exhibited both a high expression and low expression-variation level across various cancerous and normal tissue types, and then demonstrated the effectiveness of these newly identified reference genes for use in RT-qPCR. Thus, the results of the present study promote a better understanding of gene expression changes in different cancer types, and will be of considerable use in facilitating the normalization of target-gene expression levels in future cancer research.

Methods

Data collection and bioinformatics analysis

The overall workflow of the present study is shown in Fig. 1. We downloaded RNA-sequence (RNA-seq) V2 data (level 3) of 34 different cancer types from the TCGA database (<http://tcga-data.nci.nih.gov/tcga/>). The TCGA RNA-seq pipeline has used two distinct measurement methods, comprising RPKM (*R*eads *P*er *K*ilobase per *M*illion mapped reads) [11] and TPM (*T*ranscripts *P*er *M*illion) [12, 13], to obtain expression levels from RNA-seq data. Given that TPM is established to produce more comparable results across various sample types than RPKM [13, 14], we used TPM-generated data for 32 of the 34 cancer types for further analyses [esophageal carcinoma (ESCA) and stomach adenocarcinoma (STAD) were excluded, since only RPKM-generated data were available for these cancer types]. Unless otherwise stated, all gene expression levels used in our analyses represent the unit of transformed (multiplied by 10⁶) normalized read counts (extracted from TCGA files with the extension “rsem.genes.normalized_results”).

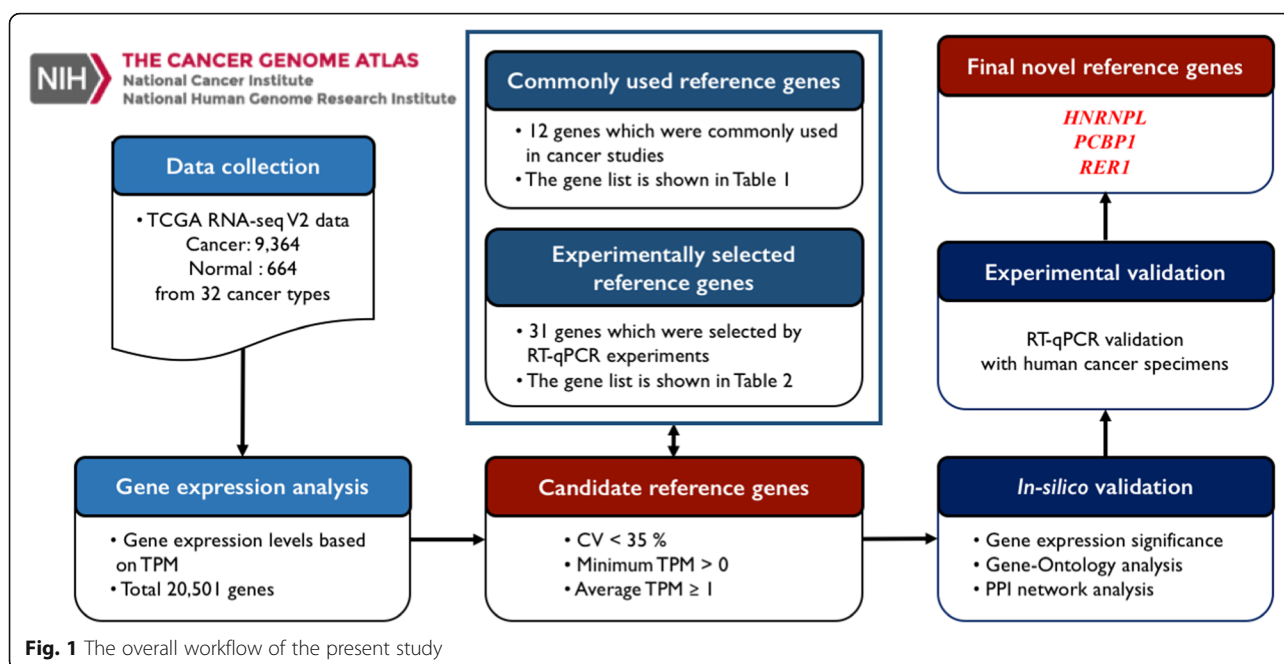


Fig. 1 The overall workflow of the present study

The human protein interaction network data were collected from the Human Protein Reference Database (HPRD release 9, <http://www.hprd.org>) [15], which includes 30,047 protein entries and 41,327 protein-protein interactions (PPIs). We extracted all binary PPIs from the HPRD, and counted the number of interactions for each protein without redundancy to estimate the size of the protein complex.

We categorized the selected reference genes according to gene ontology groups using PANTHER (<http://www.pantherdb.org/>) [16] and DAVID (<http://david.abcc.ncifcrf.gov/>) [17] tools.

Human specimens

The validity of all matched human cancerous and normal tissues was confirmed via patient clinical diagnosis. In total, 58 matched sample pairs were obtained for analysis, of which the cancerous tissue sample in each was isolated from patient breast ($n = 18$), colon ($n = 12$), thyroid ($n = 8$), lung ($n = 8$), liver ($n = 8$), kidney ($n = 2$), or cervical ($n = 2$) cancer tissues. All human tissue was trimmed to 0.5 cm² immediately after removal from the patient and stored in 5 volumes of RNAlater solution (ThermoFisher Scientific, USA) at -80°C . For the experiment, samples were used within 3 years of storage. These all utilized human specimens and data were provided by the Biobank of Chungnam University Hospital (Korea Biobank Network).

RNA preparation and RT-qPCR

Total RNA was extracted using a eCube Tissue RNA Mini Kit (PhileKorea, Korea) according to the manufacturer's instructions, and reverse-transcribed using M-MLV reverse transcriptase (Promega, USA) with random hexamers. RT-qPCR was performed with a SYBR-Green fluorescent dye (GENET BIO, Korea) and the AriaMx PCR

System (Agilent, USA). All reactions occurred under identical cycling conditions, comprising 40 cycles of amplification with denaturation (95°C , 20 s), annealing (58°C , 20 s), and elongation (72°C , 20 s). The specificity of the products generated by each primer set was confirmed by both gel electrophoresis and a melting curve analysis (Additional file 1: Table S1 and Additional file 2: Figure S1).

Results and discussion

Commonly used reference genes exhibit a high level of expression variation in both tumorous and normal tissue samples

To assess the gene expression variability within human cancerous and normal tissues, we collected gene expression data from the TCGA database, which contains 10,028 (9,364 cancerous and 664 normal) samples isolated from 32 different cancer types. We used TPM-generated data to calculate the coefficient of variation (CV, calculated as the standard deviation divided by the mean), for target gene expression levels across the analyzed samples. We initially evaluated the gene expression variability of commonly used reference genes (Table 1) [18], and found all 12 analyzed genes to exhibit a CV-value greater than 45% (Table 1). Most (23/31, 74%; Tables 2 and 3) of the experimentally selected reference genes expressed in cancer tissues were observed to exhibit a similar level of gene expression variability. We repeated this process to separately analyze cancerous and normal samples, so as to eliminate potential error caused by sample size bias (since 9,364 cancerous, but only 664 normal tissue samples were analyzed). The results of this second analysis showed the same trends in each cancer and normal group, whereby all 12 commonly used reference genes and 74% (23/31) of the experimentally selected reference genes were found to exhibit a CV value greater than 45% in both groups

Table 1 List of commonly used reference genes and their gene expression variability in 10,028 analyzed samples from TCGA database

Gene name	Description	Mean TPM value	CV(%)
ACTB	Actin Beta	4713.56	45.03
PGAM1	Phosphoglycerate Mutase 1	239.83	56.36
ALDOA	Aldolase, Fructose-Bisphosphate A	1576.45	57.97
TUBA1B	Tubulin Alpha 1b	974.69	62.63
HPRT1	Hypoxanthine Phosphoribosyltransferase 1	46.04	63.46
GAPDH	Glyceraldehyde-3-Phosphate Dehydrogenase	5445.97	67.55
B2M	Beta-2-Microglobulin	5083.41	69.57
PGK1	Phosphoglycerate Kinase 1	365.70	69.98
LDHA	Lactate Dehydrogenase A	575.04	79.40
PFKP	Phosphofructokinase, Platelet	94.20	102.74
VIM	Vimentin	1405.17	117.90
G6PD	Glucose-6-Phosphate Dehydrogenase	64.06	138.24

Table 2 List of experimentally selected reference genes

Tissues	Experimentally selected reference genes	References
Breast	PUM1, TBP, RPLP0, MRPL19, ACTB, SDHA, RPS23, HUWE1, EEF1A1, SF3A1, PPIA	[23–26]
Colon	B2M, PPIA, HPRT1, IPO8, HSP90AB1, YWHAZ, RPS13	[7, 27, 28]
Liver	HMBS, UBC, TBP, HPRT1, CTBP1	[29–33]
Lung	HPRT1, RPLP0, UBC, GAPDH, CASC3, PES1, POLR2A, YAP1, ACTB, EEF1A1, FAU, RPS9, RPS11, RPS14	[34–37]
Kidney	PPIA, RPS13, TBP	[38, 39]
Prostate	HPRT1, GAPDH, SDHA	[40, 41]
Thyroid	ACTB	[42]
HNS ^a	GAPDH, RPS18, SDHA, ALAS1	[43]

^aHNS: Head and Neck squamous cell

together (Additional file 3: Table S2). These results suggest that the reference genes most commonly used in current cancer studies may not be appropriate to serve as representative reference genes, and thus, their use may lead to erroneous quantification of cancer-related gene expression levels.

Selection of novel reference gene candidates from the TCGA database

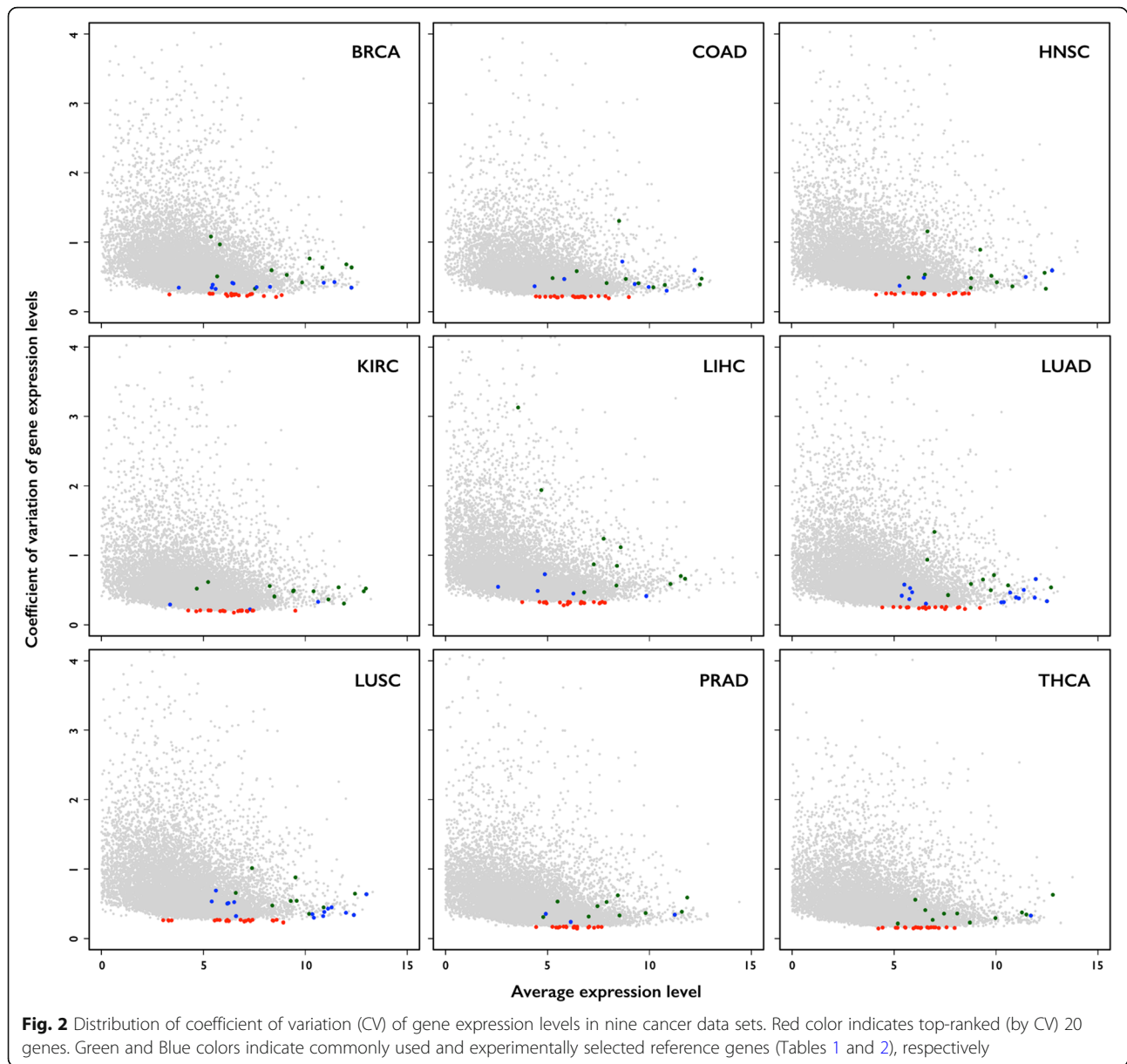
Because genetic alterations in diverse cancer types may differentially affect cellular processes at the transcriptome level, we investigated whether reference genes defined by analysis of a single type of cancerous tissue could be applied to other cancer types. Thus, we calculated and compared the CV values of > 40 samples (and their matched normal tissue samples) from nine cancer types (BRCA, COAD, HNSC, LUAD, LUSC, LIHC, PRAD, THCA, and KIRC; Additional file 4: Figure S2), that were contained within the TCGA database. Among a total set of 20 top-ranked (by CV) genes from each cancer type, no genes (1) were included in the list of commonly used reference genes, and (2) were found in more than 50% (5 out of 9) of cancer types (Fig. 2 and Additional file 5: Table S3), indicating the dependency of reference genes on cancer types.

To newly determine suitable novel genes appropriate to act as internal controls for the normalization of target gene expression in cancer research, we selected a number of genes identified (1) to exhibit unvarying expression levels across both cancerous and normal tissue samples, (2) to have a CV value < 35%, (3) a minimum TPM > 0, (4) and an average of TPM value ≥ 1 across all tissue samples. Of the 10,028 analyzed samples from the 32 different cancer types, we identified 38 candidate novel cancer-research reference genes (Fig. 3a, Additional file 1: Table S4). We subsequently evaluated whether these newly identified reference genes had the same functional characteristics as the previously

Table 3 Gene expression variability of experimentally selected reference genes in 10,028 TCGA database

Gene name	Mean TPM	STDEV	CV (%)
FAU	1392.44	538.16	38.65
CTBP1	114.52	44.26	38.65
RPS13	1571.26	675.69	43.00
UBC	1320.33	569.84	43.16
PUM1	35.09	15.17	43.22
RPS11	3969.12	1752.48	44.15
TBP	12.82	5.72	44.60
SF3A1	45.63	20.37	44.65
ACTB	4713.56	2122.56	45.03
RPS23	299.30	135.68	45.33
MRPL19	34.12	15.71	46.04
RPS14	2799.60	1316.81	47.04
IPO8	21.25	10.10	47.54
RPS9	2121.51	1065.24	50.21
PES1	66.70	34.82	52.20
RPLP0	1869.13	976.37	52.24
HUWE1	69.01	36.87	53.43
HMBS	28.07	15.41	54.91
POLR2A	60.70	33.65	55.43
HSP90AB1	1004.81	560.40	55.77
PPIA	187.35	108.62	57.98
EEF1A1	2661.35	1585.93	59.59
HPRT1	46.04	29.22	63.46
SDHA	103.74	66.11	63.72
YWHAZ	490.69	321.32	65.48
RPS18	5059.90	3336.16	65.93
GAPDH	5445.97	3678.69	67.55
B2M	5083.41	3536.78	69.57
CASC3	45.64	36.59	80.17
ALAS1	53.97	66.85	123.86
YAP1	44.89	66.00	147.04

established, commonly used reference genes. We found the average expression level of the newly identified reference genes to be significantly higher than that of the others (115.06 versus 42.93; $P < 0.0413$, using an empirical permutation test with 10,000 replications). This result is consistent with previously reported expression levels for the established reference genes [4]. Next, we determined that, as expected [4, 5, 19], the newly identified reference genes were significantly enriched in functional categories associated with transcription-translation processes, such as polyA-RNA, ribonucleoprotein, and RNA-binding (FDR < 5%, Fig. 3b). The established reference genes have been previously demonstrated to act as the ‘hubs’ of the

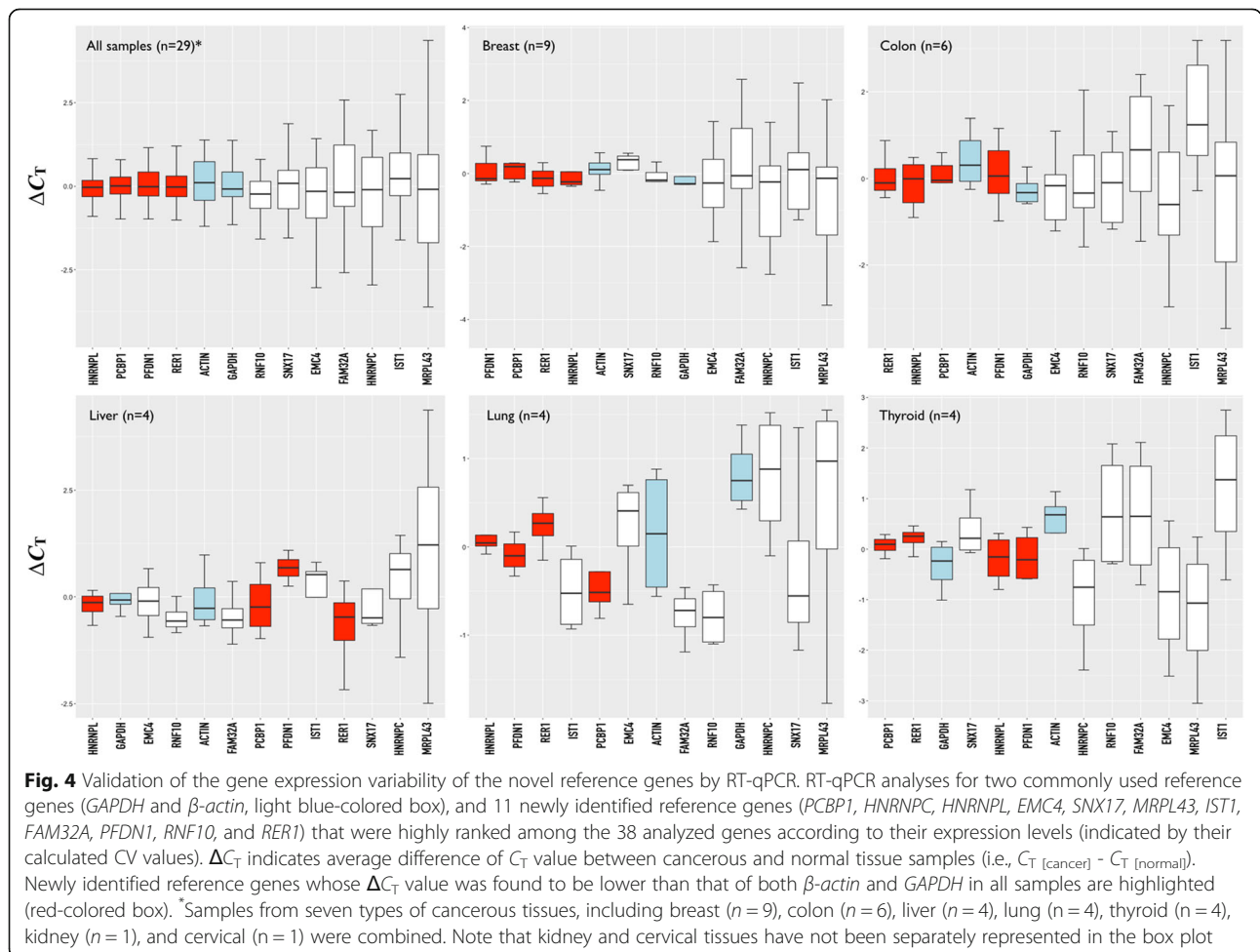
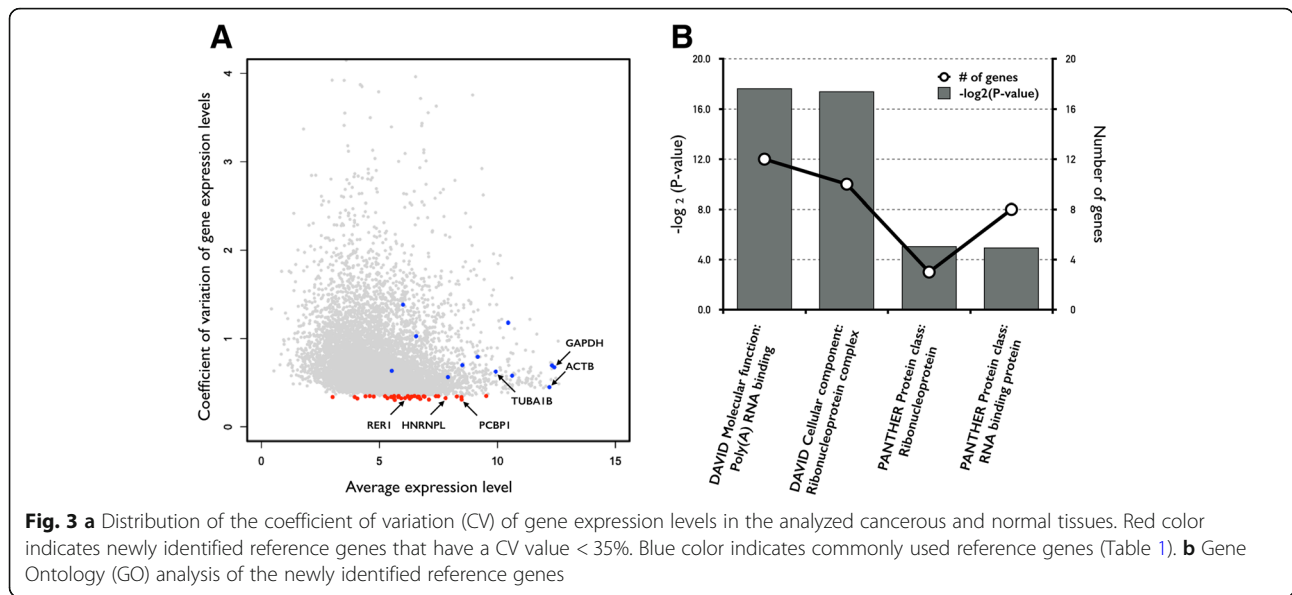


highly connected protein-protein interaction (PPI) networks [20–22]. In the present study, we observed the newly identified reference genes to be characterized by a greater number of PPI network-interaction partners than the other genes (8.42 versus 3.67; $P < 0.0185$, using an empirical permutation test with 10,000 replications), indicating their functional importance for biological systems.

RT-qPCR validation of the newly identified reference genes in human cancer tissues

We next sought to confirm the validity of the newly identified candidates as reference genes for the normalization of RT-qPCR expression data in the context of human cancer. Therefore, we compared the

RT-qPCR analysis results for two commonly used reference genes (*GAPDH* and *β-actin*) with those for the 11 most highly expressed of the newly identified reference genes (*PCBP1*, *HNRNPC*, *HNRNPL*, *EMC4*, *SNX17*, *MRPL43*, *IST1*, *FAM32A*, *PFDN1*, *RNF10*, and *RER1*) across 29 patient samples including breast, colon, liver, lung, and/or thyroid cancer types. Each human tissue was immersed in RNAlater solution immediately after extraction from the patient and stored at -80°C to minimize RNA degradation. In addition, $2\ \mu\text{g}$ of total RNA extracted from tissues was electrophoresed on 1.5% denaturing agarose gel and only 28S/18S ratio of > 2 confirmed RNA was used in the experiment. The specificity of the products generated by each primer



set was confirmed by both gel electrophoresis and a melting curve analysis (Additional file 1: Table S1 and Additional file 2: Figure S1).

Since optimal reference genes for cancer-transcriptome analysis should exhibit a low level of expression variability between cancerous and normal tissue samples, we isolated total RNA from each cancerous and normal sample from a single patient and compared their C_T values (where, C_T is the “Cycle Threshold”, defined as the number of cycles required for the fluorescence signal to exceed background level, and is inversely correlated with the amount of target nucleic acid in the sample). Of the 11 newly identified genes, *HNRNPL* ($\Delta C_T = 0.37$), *PCBP1* ($\Delta C_T = 0.42$), *PFDN1* ($\Delta C_T = 0.46$), and *RER1* ($\Delta C_T = 0.48$) were found to have a lower average C_T difference ($\Delta C_T = C_{T \text{ [cancer]}} - C_{T \text{ [normal]}}$) between cancerous and normal tissue samples than *β -actin* ($\Delta C_T = 0.58$) and/or *GAPDH* ($\Delta C_T = 0.60$), suggesting their suitability for use as consensus reference genes for gene expression studies in human cancer (Fig. 4). To ensure the reliability and robustness of these results, we reconfirmed whether these reference genes had lower ΔC_T values than *β -actin* and/or *GAPDH* in each cancer sample. *HNRNPL* was identified to have a ΔC_T value lower than that of both *β -actin* and *GAPDH* in four (breast, colon, liver, and lung) of five cancer sample types. Similarly, *PCBP1* and *RER1* had lower ΔC_T values than *β -actin* and *GAPDH* in all cancer sample types except liver cancer tissue, and *PFDN1* exhibited a lower ΔC_T value than *β -actin* and *GAPDH* in two cancer sample types (breast and lung, Fig. 4).

Conclusion

In summary, cancer is a disease characterized by complex molecular networks, in which highly heterogeneous and multifocal tumor cells cooperate with host cells within their microenvironment. Recent gene expression studies have been conducted to investigate the intricate interplay of gene expression patterns that regulate cancer invasion and metastasis at the transcriptional level; however, their accurate quantification of gene expression level is dependent upon the selection and use of reliable and appropriate reference genes for the normalization of target gene expression levels. Thus, in the present study, we performed in silico bioinformatics analyses and experimental validation to identify *HNRNPL*, *PCBP1* and *RER1* as novel candidate reference genes, whose expression is predominantly consistent, independent of cancer type, stage, and treatment status, and of patient age and gender. Although a larger sample size and more cancer types are needed for more reliable results, these novel reference genes will be invaluable for diagnosis and the prediction of patient prognosis, in a wide range of human cancers.

Additional files

Additional file 1: Table S1. Primers used for quantitative analysis of gene expression. **Table S4.** Gene expression variability of newly identified reference genes. (DOCX 28 kb)

Additional file 2: Figure S1. qPCR electrophoresis result and melting curve analysis of our reference genes. (A) Agarose gel electrophoresis showing specific reverse transcription PCR products of the expected size for each gene. (B) Melting curves generated for all genes. (TIFF 34570 kb)

Additional file 3: Table S2. Gene expression variability of commonly used and experimentally selected reference genes in each cancerous and normal group. (XLSX 24 kb)

Additional file 4: Figure S2. Nine cancer types. Nine cancer types from TCGA comprising both cancerous and matched normal data with > 40 samples. (TIFF 3075 kb)

Additional file 5: Table S3. Top 20 candidate reference genes in each cancer type. (XLSX 65 kb)

Abbreviations

BRCA: Breast invasive carcinoma; COAD: Colon adenocarcinoma; CV: Coefficient of variation; FDR: False discovery rate; HNSC: Head and neck squamous cell carcinoma; IRB: Institutional review board; KIRC: Kidney renal clear cell carcinoma; LIHC: Liver hepatocellular carcinoma; LUAD: Lung adenocarcinoma; LUSC: Lung squamous cell carcinoma; M-MLV: Moloney murine leukemia virus; PRAD: Prostate adenocarcinoma; RPKM: Reads per kilobase per million mapped reads; RT-qPCR: Quantitative reverse transcription PCR; TCGA: The cancer genome atlas; THCA: Thyroid carcinoma; TPM: Transcripts per million

Acknowledgements

The authors are grateful to valuable comments and suggestions of the reviewers.

Funding

This work was supported by research grants from the Bio-Synergy Research Project (NRF-2015M3A9C4075820) of the Ministry of Science, ICT and Future Planning through the National Research Foundation to C.P. This study was part of the project titled “Development of the methods for controlling and managing the marine ecosystem disturbing and harmful organisms (MEDHO)”, funded by the Ministry of Oceans and Fisheries of the Republic of Korea to C.P. Publication costs are funded by the project titled “Research center for fishery resource management based on the information and communication technology (ICT)”, funded by the Ministry of Oceans and Fisheries of the Republic of Korea to C.P.

Availability of data and materials

All data for this study are downloaded from TCGA (<https://cancergenome.nih.gov/>) public database and included in Tables, Supplementary tables, and Additional files.

About this supplement

This article has been published as part of *BMC Bioinformatics Volume 20 Supplement 10, 2019: Proceedings of the 12th International Workshop on Data and Text Mining in Biomedical Informatics (DTMBIO 2018)*. The full contents of the supplement are available online at <https://bmcbioinformatics.biomedcentral.com/articles/supplements/volume-20-supplement-10>.

Authors' contributions

CP and JJ designed research. CP, KKK, and SYC contributed to the research coordination. JJ, SC, JO, and SGL performed research. CP, KKK, SYC, JJ, and SC analyzed data. CP, JJ, and KKK wrote the paper. All authors read and approved the final manuscript.

Ethics approval and consent to participate

The Institutional Review Board (IRB) of Chungnam National University approved the use of human tissues in the present study (IRB number 2016-08-032). All utilized human specimens and data were provided by the Biobank of Chungnam University Hospital (Korea Biobank Network).

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Published: 29 May 2019

References

- Hornberg JJ, Bruggeman FJ, Westerhoff HV, Lankelma J. Cancer: a systems biology disease. *Biosystems*. 2006;83(2–3):81–90.
- Futreal PA, Coin L, Marshall M, Down T, Hubbard T, Wooster R, Rahman N, Stratton MR. A census of human cancer genes. *Nat Rev Cancer*. 2004;4(3):177–83.
- Cancer Genome Atlas Research N, Weinstein JN, Collisson EA, Mills GB, Shaw KR, Ozenberger BA, Ellrott K, Shmulevich I, Sander C, Stuart JM. The cancer genome atlas pan-cancer analysis project. *Nat Genet*. 2013;45(10):1113–20.
- Zhu J, He F, Hu S, Yu J. On the nature of human housekeeping genes. *Trends Genet*. 2008;24(10):481–4.
- Eisenberg E, Levanon EY. Human housekeeping genes, revisited. *Trends Genet*. 2013;29(10):569–74.
- Sharan RN, Vaiphei ST, Nongrum S, Keppen J, Ksoo M. Consensus reference gene(s) for gene expression studies in human cancers: end of the tunnel visible? *Cell Oncol (Dordr)*. 2015;38(6):419–31.
- Jacob F, Guertler R, Naim S, Nixdorf S, Fedier A, Hacker NF, Heinzemann-Schwarz V. Careful selection of reference genes is required for reliable performance of RT-qPCR in human normal and cancer cell lines. *PLoS One*. 2013;8(3):e59180.
- Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, Speleman F. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol*. 2002;3(7):RESEARCH0034.
- Andersen CL, Jensen JL, Orntoft TF. 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*. 2004;64(15):5245–50.
- Pfaffl MW, Tichopad A, Prgomet C, Neuvians TP. Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper—excel-based tool using pair-wise correlations. *Biotechnol Lett*. 2004;26(6):509–15.
- Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods*. 2008;5(7):621–8.
- Wang K, Singh D, Zeng Z, Coleman SJ, Huang Y, Savich GL, He X, Mieczkowski P, Grimm SA, Perou CM, et al. MapSplice: accurate mapping of RNA-seq reads for splice junction discovery. *Nucleic Acids Res*. 2010;38(18):e178.
- Li B, Ruotti V, Stewart RM, Thomson JA, Dewey CN. RNA-Seq gene expression estimation with read mapping uncertainty. *Bioinformatics*. 2010;26(4):493–500.
- Wagner GP, Kin K, Lynch VJ. Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theory Biosci*. 2012;131(4):281–5.
- Keshava Prasad TS, Goel R, Kandasamy K, Keerthikumar S, Kumar S, Mathivanan S, Telikicherla D, Raju R, Shafreen B, Venugopal A, et al. Human protein reference database—2009 update. *Nucleic Acids Res*. 2009;37(Database):D767–72.
- Thomas PD, Campbell MJ, Kejariwal A, Mi H, Karlak B, Daverman R, Diemer K, Muruganujan A, Narechania A. PANTHER: a library of protein families and subfamilies indexed by function. *Genome Res*. 2003;13(9):2129–41.
- Huang d W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*. 2009;4(1):44–57.
- de Jonge HJ, Fehrmann RS, de Bont ES, Hofstra RM, Gerbens F, Kamps WA, de Vries EG, van der Zee AG, te Meerman GJ, ter Elst A. Evidence based selection of housekeeping genes. *PLoS One*. 2007;2(9):e898.
- Eisenberg E, Levanon EY. Human housekeeping genes are compact. *Trends Genet*. 2003;19(7):362–5.
- Lin WH, Liu WC, Hwang MJ. Topological and organizational properties of the products of house-keeping and tissue-specific genes in protein-protein interaction networks. *BMC Syst Biol*. 2009;3:32.
- Jeong H, Mason SP, Barabasi AL, Oltvai ZN. Lethality and centrality in protein networks. *Nature*. 2001;411(6833):41–2.
- Alemu EY, Carl JW Jr, Corrada Bravo H, Hannenhalli S. Determinants of expression variability. *Nucleic Acids Res*. 2014;42(6):3503–14.
- Lyng MB, Laenkholm AV, Pallisgaard N, Ditzel HJ. Identification of genes for normalization of real-time RT-PCR data in breast carcinomas. *BMC Cancer*. 2008;8:20.
- McNeill RE, Miller N, Kerin MJ. Evaluation and validation of candidate endogenous control genes for real-time quantitative PCR studies of breast cancer. *BMC Mol Biol*. 2007;8:107.
- Gur-Dedeoglu B, Konu O, Bozkurt B, Ergul G, Seckin S, Yulug IG. Identification of endogenous reference genes for qRT-PCR analysis in Normal matched breast tumor tissues. *Oncol Res Featuring Preclinical Clin Cancer Ther*. 2009;17(8):353–65.
- Maltseva DV, Khaustova NA, Fedotov NN, Matveeva EO, Lebedev AE, Shkurnikov MU, Galatenko VV, Schumacher U, Tonevitsky AG. High-throughput identification of reference genes for research and clinical RT-qPCR analysis of breast cancer samples. *J Clin Bioinformatics*. 2013;3(1):13.
- Kheirelseid EAH, Chang KH, Newell J, Kerin MJ, Miller N. Identification of endogenous control genes for normalisation of real-time quantitative PCR data in colorectal cancer. *BMC Mol Biol*. 2010;11(1):12.
- Sørby LA, Andersen SN, Bukholm IRK, Jacobsen MB. Evaluation of suitable reference genes for normalization of real-time reverse transcription PCR analysis in colon cancer. *J Exp Clin Cancer Res*. 2010;29(1):144.
- Kim S, Kim T. Selection of optimal internal controls for gene expression profiling of liver disease. *Biotechniques*. 2003;35(3):456–458, 460.
- Gao Q, Wang XY, Fan J, Qiu SJ, Zhou J, Shi YH, Xiao YS, Xu Y, Huang XW, Sun J. Selection of reference genes for real-time PCR in human hepatocellular carcinoma tissues. *J Cancer Res Clin Oncol*. 2008;134(9):979–86.
- Fu LY, Jia HL, Dong QZ, Wu JC, Zhao Y, Zhou HJ, Ren N, Ye QH, Qin LX. Suitable reference genes for real-time PCR in human HBV-related hepatocellular carcinoma with different clinical prognoses. *BMC Cancer*. 2009;9:49.
- Liu S, Zhu P, Zhang L, Ding S, Zheng S, Wang Y, Lu F. Selection of reference genes for RT-qPCR analysis in tumor tissues from male hepatocellular carcinoma patients with hepatitis B infection and cirrhosis. *Cancer Biomark*. 2013;13(5):345–9.
- Cicinnati VR, Shen Q, Sotiropoulos GC, Radtke A, Gerken G, Beckebaum S. Validation of putative reference genes for gene expression studies in human hepatocellular carcinoma using real-time quantitative RT-PCR. *BMC Cancer*. 2008;8:350.
- Gresner P, Gromadzinska J, Wasowicz W. Reference genes for gene expression studies on non-small cell lung cancer. *Acta Biochim Pol*. 2009;56(2):307–16.
- Sharungbam GD, Schwager C, Chiblak S, Brons S, Hlatky L, Haberer T, Debus J, Abdollahi A. Identification of stable endogenous control genes for transcriptional profiling of photon, proton and carbon-ion irradiated cells. *Radiat Oncol*. 2012;7:70.
- Saviozzi S, Cordero F, Lo Iacono M, Novello S, Scagliotti GV, Calogero RA. Selection of suitable reference genes for accurate normalization of gene expression profile studies in non-small cell lung cancer. *BMC Cancer*. 2006;6:200.
- Zhan C, Zhang Y, Ma J, Wang L, Jiang W, Shi Y, Wang Q. Identification of reference genes for qRT-PCR in human lung squamous-cell carcinoma by RNA-Seq. *Acta Biochim Biophys Sin Shanghai*. 2014;46(4):330–7.
- Jung M, Ramankulov A, Roigas J, Johansson M, Ringsdorf M, Kristiansen G, Jung K. In search of suitable reference genes for gene expression studies of human renal cell carcinoma by real-time PCR. *BMC Mol Biol*. 2007;8:47.
- Dupasquier S, Delmarcelle AS, Marbaix E, Cosyns JP, Courtoy PJ, Pierreux CE. Validation of housekeeping gene and impact on normalized gene expression in clear cell renal cell carcinoma: critical reassessment of YBX3/ZONAB/CSDA expression. *BMC Mol Biol*. 2014;15:9.
- Ohi F, Jung M, Xu C, Stephan C, Rabien A, Burkhardt M, Nitsche A, Kristiansen G, Loening SA, Radonic A, et al. Gene expression studies in prostate cancer tissue: which reference gene should be selected for normalization? *J Mol Med (Berl)*. 2005;83(12):1014–24.
- Souza AF, Brum IS, Neto BS, Berger M, Branchini G. Reference gene for primary culture of prostate cancer cells. *Mol Biol Rep*. 2013;40(4):2955–62.

42. Weber R, Bertoni AP, Bessestil LW, Brasil BM, Brum LS, Furlanetto TW. Validation of reference genes for normalization gene expression in reverse transcription quantitative PCR in human normal thyroid and goiter tissue. *Biomed Res Int.* 2014;2014:198582.
43. Lallemand B, Evrard A, Combesure C, Chapuis H, Chambon G, Raynal C, Reynaud C, Sabra O, Joubert D, Hollande F, et al. Reference gene selection for head and neck squamous cell carcinoma gene expression studies. *BMC Mol Biol.* 2009;10:78.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

