Biomarker Insights

Libertas Academica
FREEDOM TO RESEARCH

OPEN ACCESS
Full open access to this and thousands of other papers at http://www.la-press.com.

METHODOLOGY

Standardization of Gene Expression Quantification by Absolute Real-Time qRT-PCR System Using a Single Standard for Marker and Reference Genes

Yi-Hong Zhou[1], Vinay R. Raj[2], Eric Siegel[3] and Liping Yu[4]

[1]Department of Neurological Surgery and Biological Chemistry, University of California Irvine, Irvine, CA 92697, USA. [2]Division of Medical Genetics, [3]Biostatistics, University of Arkansas for Medical Sciences, Little Rock, AR 72205, USA. [4]Ziren Research, Irvine, CA 92618, USA. Corresponding author email: yihongz@uci.edu

Abstract: In the last decade, genome-wide gene expression data has been collected from a large number of cancer specimens. In many studies utilizing either microarray-based or knowledge-based gene expression profiling, both the validation of candidate genes and the identification and inclusion of biomarkers in prognosis-modeling has employed real-time quantitative PCR on reverse transcribed mRNA (qRT-PCR) because of its inherent sensitivity and quantitative nature. In qRT-PCR data analysis, an internal reference gene is used to normalize the variation in input sample quantity. The relative quantification method used in current real-time qRT-PCR analysis fails to ensure data comparability pivotal in identification of prognostic biomarkers. By employing an absolute qRT-PCR system that uses a single standard for marker and reference genes (SSMR) to achieve absolute quantification, we showed that the normalized gene expression data is comparable and independent of variations in the quantities of sample as well as the standard used for generating standard curves. We compared two sets of normalized gene expression data with same histological diagnosis of brain tumor from two labs using relative and absolute real-time qRT-PCR. Base-10 logarithms of the gene expression ratio relative to *ACTB* were evaluated for statistical equivalence between tumors processed by two different labs. The results showed an approximate comparability for normalized gene expression quantified using a SSMR-based qRT-PCR. Incomparable results were seen for the gene expression data using relative real-time qRT-PCR, due to inequality in molar concentration of two standards for marker and reference genes. Overall results show that SSMR-based real-time qRT-PCR ensures comparability of gene expression data much needed in establishment of prognostic/predictive models for cancer patients—a process that requires large sample sizes by combining independent sets of data.

Keywords: gene expression, quantification, qRT-PCR, biomarkers

Biomarker Insights 2010:5 79–85

This article is available from http://www.la-press.com.

© the author(s), publisher and licensee Libertas Academica Ltd.

This is an open access article. Unrestricted non-commercial use is permitted provided the original work is properly cited.

Introduction

In the past decade, the traditional approach to cancer therapy has shifted more towards personalized medicine. Molecular variables such as gene expression levels are increasingly used in combination with clinical variables to model prognosis and explain variation in survival and response to therapeutic intervention.[1,2] Genome-wide gene expression profiling studies on breast,[3] lung,[4] prostate,[5] and recently brain[6] cancers have revealed panels of candidate gene expression variables. Application of these gene expression signatures in cancer should enable better quantification of the residual risk faced by patients and indicate the potential value of additional treatment. This requires establishment of prognostic and predictive models with statistically significant discriminative capability.

Although reports from the MicroArray Quality Control (MAQC) project showed inter- and intra-platform reproducibility of gene-expression measurements,[7] concerns remain on direct applicability of microarray data to identify molecular markers in prognostic modeling of patient survival outcomes. Controversial results have been reported on the comparability of data derived from different platforms, in addition to data reproducibility[8,9] and statistical complexity on model reproducibility.[10] Modeling of prognosis or patient's response to treatment requires significant sample sizes with an average of ten events per variable in a multivariate model to achieve a meaningful analysis.[10] An approach that evaluates only a select number of gene expression variables with a standardized gene expression quantification technology may provide an efficacious alternative to wide-scale microarray assessment, assuring data comparability and satisfying the statistical requirements needed to produce a meaningful analysis.[2]

Real-time quantitative reverse transcription (qRT)-PCR was developed over the last decade to sensitively and reliably quantify mRNA levels in knowledge-based gene expression studies,[11] and has been widely used to validate microarray data.[12] To normalize the variation in sample input, which is a dependent variable of RNA quantification and reverse transcription, it is in common practice to use expression of a reference gene to normalize the marker gene's expression. Both quantities rely either on PCR efficiency-based methods[13,14] or a standard curve based on serially diluted cDNA of individual genes with known quantity,[1,15] thus providing a relative quantification. As pointed out in our previous study of modeling prognosis for gliomas,[2] absolute quantification of gene expression can be achieved by "using a multi-gene (containing) DNA standard for real-time PCR, made by ligating the marker gene standard DNA and the standards for internal control genes together in one piece, in a one-to-one ratio". In this study, we evaluated the comparability of gene expression data from two sets of glioma samples of the same histology using both a relative and an absolute quantification approach. Our data demonstrates that comparable gene expression data can be achieved by a standard-curve based real-time qRT-PCR using the single standard for marker and reference genes (SSMR).

Method

Single standard for marker and reference (SSMR) genes

PCR amplicons of multiple prognostic marker genes for gliomas (*PAX6*, *PTEN*) and reference genes (*ACTB*, *RPS9*) were ligated together through multiple cloning processes. The multigene-containing DNA fragment was then amplified by PCR, purified from gel extraction, and the molar concentration of standard was calculated based on DNA concentration and molecular weight. This SSMR was subjected to a 10-fold serial dilution (1E+10 to 1E+2 molecules per 4 µl) with 10 mM Tris-HCl, pH 7.5. Four standard dilutions from 1E+6 to 1E+1 depending on the abundance of the gene were used for quantifying both marker and reference gene expression in cDNA samples.

PCR primers and real-time PCR

Using PrimerDesigner software (Ziren Research LLC, Irvine, CA), PCR primers were designed to avoid the amplification of genomic DNA and processing pseudogenes of *ACTB* and *PTEN* with >90% identity in sequence to the corresponding cDNA sequence of the gene. Real-time qRT-PCR was carried out using either a LightCycler 1.5 instrument in glass capillaries (Roche, Indianapolis, IN) or a StepOne instrument in thin-wall PCR plates (Applied Biosystems, Foster City, CA). Each reaction contained 4 µl standard/cDNA template, 1 µl 10X primer-$MgCl_2$ mix for the

gene [final 0.5 μM primer, and 2.5–4 mM $MgCl_2$], 4 μl water, and 1 μl FAST-START DNA Master SYBR Green I mix (Roche) following the standard PCR program suggested by the manufacturer.

RNA and cDNA samples

Two sets of cDNA samples of glioblastoma multiforme (GBM) from our prior modeling of glioma prognosis study[2] were used. Set I GBM cDNA collected from M.D. Anderson Cancer Center (MDACC) were synthesized from total RNA with DNase I treatment and used in studies by Sano et al 1999[16] and Zhou et al 2003[1] and described therein. Set II GBM specimens were collected at University of Arkansas for Medical Sciences (UAMS) under an IRB approved protocol. Total RNA was extracted using Trizol-based RNA isolation approach without DNase I treatment, reverse transcribed (~1 μg of RNA) using reverse transcriptase SuperScript RTII (Invitrogen) following manufacturer's protocol. Both sets of GBM cDNA samples were diluted with 10 mM Tris.HCl (pH 7.5) before using in real-time PCR.

SSMR-based qRT-PCR quantification of gene expressions in glioblastoma multiforme (GBM)

We used Ziren® Human Real-Time AqRT-PCR Standard-1001 (Ziren Research LLC, Irvine) to quantify *ABCG2*, *BMI1*, *MELK*, *MSI1*, *PROM1*, and *ACTB* and AqRT-PCR Standard-1020 to quantify *PAX6*, *PTEN*, *VEGFA*, and *ACTB* for a set of GBM cDNA from 87 patients, 66 from MDACC, and 21 from UAMS. Quantification of *CDK4*, *EGFR*, and *MMP2* was carried out in UAMS and University of California, Irvine on two sets of cDNA samples based on the same single standard containing these three genes, but not the reference gene *ACTB*. For these three genes, the data were a relative ratio to *ACTB*. In order to combine the two sets of relative quantitative data in to a single prognosis model, the data in the 2nd set was adjusted by the mean of fold difference between the 2nd vs. the 1st quantification on a set of the same cDNA samples from the 1st set.

Statistical analysis

Expression ratios for marker gene relative to reference gene (*ACTB*) were converted to their common logarithms ("logRatios"), and summarized by lab as their mean (SD) and median (quartiles). To evaluate samples from different labs for comparability, mean logRatios were evaluated for statistical equivalence between labs using a confidence-interval version of the Two One-Sided Tests (TOST) procedure of Schuirmann.[17] Briefly, two groups are considered "equivalent to within $\pm\delta$" if two one-sided statistical tests show that the difference in the two groups' means or medians simultaneously are (1) significantly more than the lower limit of equivalence at $-\delta$ and (2) significantly less than the upper limit of equivalence at $+\delta$. When both of the one-sided tests have a 5% significance level, then the entire TOST procedure has a significance level of 5% (not 10%).[18] In terms of confidence intervals, this means that the groups are considered equivalent at a 5% significance level if their difference has a 90% (not 95%) confidence interval that lies entirely inside the upper and lower equivalence limits. For these evaluations, we considered $\pm\delta = \pm 0.5$ to be reasonable limits of equivalence. When there was no detectable expression of a marker gene in a sample, we left-censored the observation at 0.8 times the minimum non-zero value of that marker gene. To accommodate the left-censoring, we estimated 90% confidence intervals via the rank-based method of Hodges and Lehmann.[19] When a gene was not equivalent between groups to within ±0.5, the difference between groups was further investigated using a single 2-sided Wilcoxon rank-sum (WRS) test. Statistical analysis was conducted using SAS 9.2 software (SAS Institute Inc., Cary, NC).

Results

SSMR-based real-time qRT-PCR provides data independent of variation in concentration of cDNA sample and the standard setting

To test the idea of absolute quantification through using SSMR, we first examined how much the quantification of gene expression varies in response to variation of the given quantity of the standard. We quantified expression levels of two genes (*PAX6* and *RPS9*) in the cDNA sample of a glioma cell line, based on two dilutions of the same SSMR, one of truthfully (T) diluted based on calculated DNA molar concentration, while the other falsely diluted (F) with half in denoted quantity. Both SSMR dilutions were

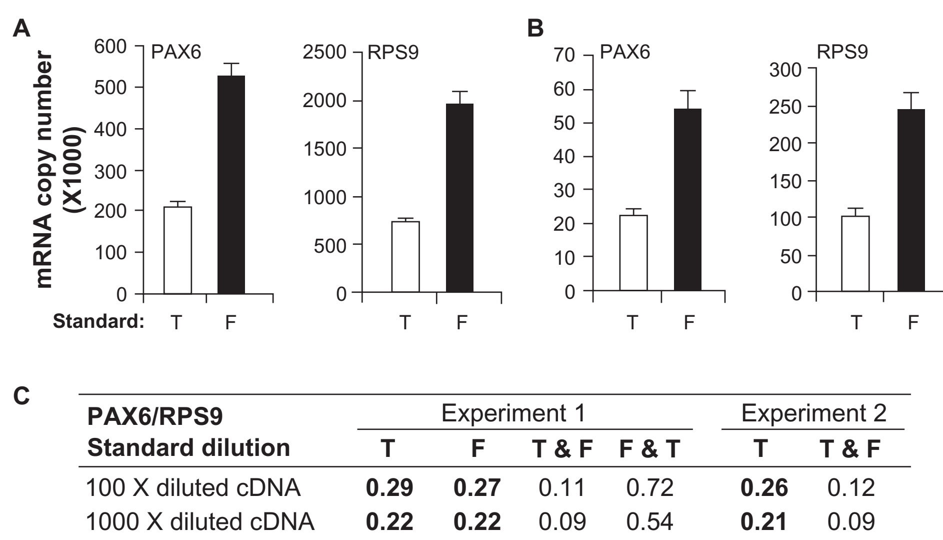
included in the same run with the same defined quantity in standard setting, and generated two standard curves to give rise to two gene expression values in the tested cDNA samples. As shown in Figure 1A and B, a 2-fold difference in gene expression levels for both marker and reference genes was seen due to a 2-fold difference in the amount of standard, like the result of taking a half-foot ruler as one foot to measure the same person's height. This result illustrated that the standard curve-based quantification of gene expression by real-time PCR relies robustly on the quantity of standard.

We then compared the ratios of *PAX6* to *RPS9*, with each mRNA copy number derived based on the same or different standard dilutes. The former is a SSMR-based quantification approach, while the latter is equivalent to using a separate standard for each gene. As shown in Figure 1C, the difference in the ratio of *PAX6*/*RPS9* in sample cDNA is less than 1.1 fold when each mRNA copy number was derived based on the SSMR. This was seen even when one standard was half the amount of the other in two quantifications, the input cDNA quantities varied by 10-folds, and the quantifications were performed at a different time. In contrast, there is a 7-fold difference in the ratio of *PAX6*/*RPS9* when each mRNA copy number was derived based on the T and F standard dilutions. Thus SSMR-based real-time qRT-PCR results in an absolute ratio of two genes that is independent of variation in concentration of cDNA sample and the standard setting.

Equivalence and difference in glioma prognosis marker gene expression data between labs

We analyzed data comparability for gene expressions quantified using real-time qRT-PCR with or without using SSMR in two sets of GBMs of comparable sample size; Set I (n = 38) from MDACC from patients operated between 1987–1997; Set II (n = 28) with 15 GBMs from MDACC and 13 GBMs from UAMS from patients operated between 2003–2006 (Fig. 2). *PAX6* and *ACTB* mRNA copy numbers were derived based on the standard curve of the SSMR for PAX6. For normalized *PAX6*, the estimated difference (90% confidence interval) in logRatios was +0.166 (–0.056 – +0.320) between Set I and set II, thus demonstrating that the two sets are statistically equivalent to within ±0.5 with respect to PAX6.

The *PTEN* mRNA copy number in the same two sets of GBMs were derived based on the standard curve of a standard for *PTEN* as reported in our previous study.[2] The normalized *PTEN* expression to *ACTB* was achieved by using the same *ACTB* mRNA copy number derived from the SSMR for PAX6, thus a relative ratio of *PTEN* to *ACTB*. The estimated difference (90% confidence interval) in *PTEN*



PAX6/RPS9	Experiment 1				Experiment 2	
Standard dilution	T	F	T & F	F & T	T	T & F
100 X diluted cDNA	**0.29**	**0.27**	0.11	0.72	**0.26**	0.12
1000 X diluted cDNA	**0.22**	**0.22**	0.09	0.54	**0.21**	0.09

Figure 1. SSMR-based qRT-PCR provides an absolute ratio for two genes. Quantification of *PAX6* and *RPS9* expression in a human glioma cell line cDNA diluted 100-fold **A**) and 1000-fold **B**) after reverse transcription from 5 µg of total RNA in 10 µl. The derived mRNA copy numbers were based on the truthfully (T) diluted and falsely (F) 2-fold further diluted SSMR, as shown by *Open* and *filled boxes*, respectively. Bar heights (error bars) represent means (SDs) from 3 independent repeats of real-time PCR. **C**) Comparison of the ratios of *PAX6*/*RPS9* with quantity of each mRNA derived based on the same or a different standard dilution.

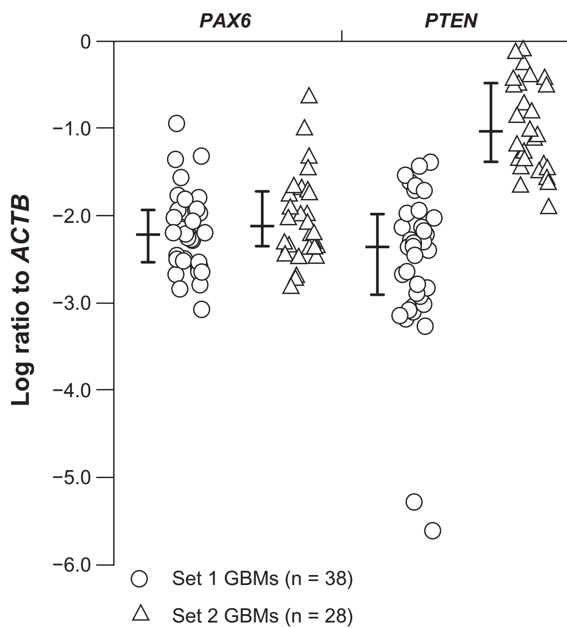


Figure 2. Comparison of normalized *PAX6* and *PTEN* expressions in two sets of GBMs with or without using SSMR in real-time qRT-PCR. Log-transformed ratio of *PAX6* to *ACTB* was an absolute ratio by using SSMR, while ratio of *PTEN* to *ACTB* was a relative ratio by using two separate gene standards. The value of each sample was plotted and the median and quartiles for each gene ratio in Set I and II GBMs are shown by horizontal and vertical bars, respectively.

logRatios was +1.484 (+1.208 – +1.720) between Set I and Set II, indicating a strong lack of equivalence. Because both ends of the confidence interval were greater than the upper equivalence limit of +0.500, *PTEN* was further evaluated for a difference between labs via single two-sided WRS test. As shown in Figure 2, the difference in median logRatios was 1.312 ($P < 0.0001$), equivalent to a 21-fold difference between labs in their *PTEN/ACTB* expression ratios, even though the same standards were used for Sets I and Set II GBMs but diluted and used by different users. Given the approximate comparability of *PAX6* expression in these two sets of GBMs, the dramatic difference in *PTEN* expression is unlikely due to changes in *PTEN* expression in the two sets of GBMs, but attributable to changes in the standard quantity, which is likely due to human error in standard dilution process, as shown in Figure 1.

We then analyzed for equivalence of gene expressions relative to *ACTB* that were included in our recent modeling of glioma prognosis that were quantified using SSMR-based real-time qRT-PCR, including the previously reported seven genes (*CDK4*, *EGFR*, *MMP2*, *VEGFA*, *PAX6*, *PTEN*, and *RPS9*) with prognosis values[2] and five stem cell marker genes (manuscript under review) that were quantified in two sets of GBMs from the two institutes' patients (MDACC and UAMS). As shown in Figure 3, 8 out of 12 genes are statistically equivalent to within ±0.500 on $\log_{10}$ scale, but 4 (*ABCG2*, *BMI1*, *MELK*, and *PROM1*) out of 12 are not. The largest institution difference (for *ABCG2*) is 0.4050 on $\log_{10}$ scale, equivalent to a 2.54-fold (2.54x) change in average expression ratio of *ABCG2* to *ACTB* between institutions ($P = 0.030$ by 2-sided WRS test). The next largest institution difference (for *BMI1*) is 0.2961, equivalent to a 1.98x change in average *BMI1*/*ACTB* ratio between institutions ($P = 0.038$ by 2-sided WRS test). The other two non-equivalent genes, *MELK* and *PROM1*, showed differences <0.2000 (fold change <1.58x) with 90% confidence intervals that contained zero, thus showing that these two genes were neither statistically equivalent nor significantly different between institutions.

Discussion

Real-time qRT-PCR technique is increasingly being used to quantify the expression levels of candidate marker genes in the identification of prognostic biomarkers. This study shows that the relative quantification method currently used in real-time qRT-PCR technique does not ensure data comparability. We introduced the method of using a SSMR in real-time qRT-PCR to achieve an absolute ratio of the marker and reference gene expression.

Gene expression quantification relies on the standard quantity for that gene. The standard curve for each gene is a linear regression curve between the log (10) value of the standard copy number and the cross points where the PCR reaches log-amplification. Given the accurate measurement of mRNA in the form of cDNA, comparing it with other samples requires normalization to make the data comparable among different samples, which is normally achieved using the ratio of the marker gene mRNA quantity to that of an appropriate internal reference gene (in practice, house-keeping genes). Quantification of the reference gene also relies on its own standard in real-time PCR. Ideally if the same copy numbers of each marker and reference gene standard are used, an absolute ratio between them in the sample can be achieved. Thus the data will be comparable to that generated separately

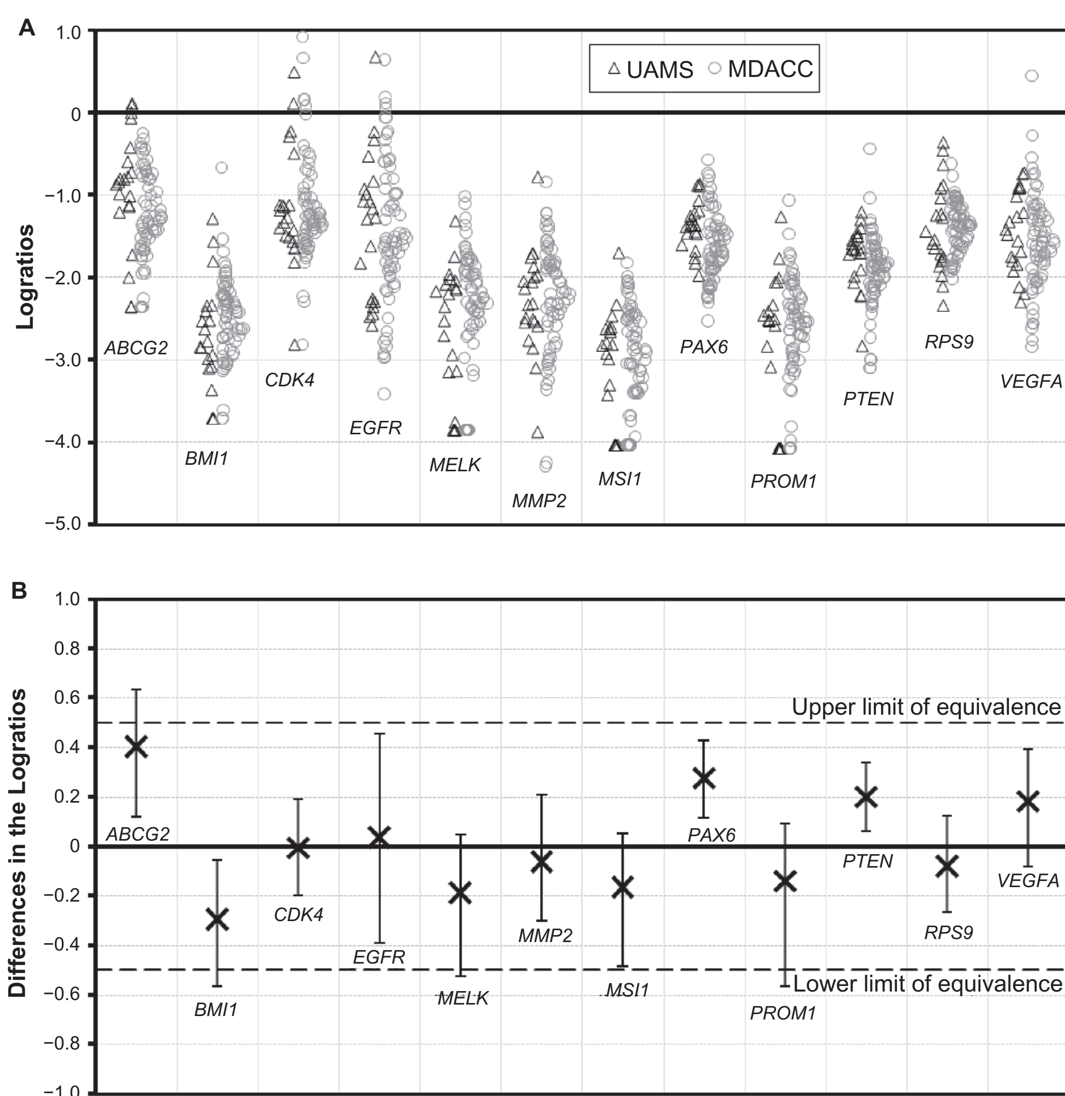


Figure 3. Equivalence test for expression of 12 prognosis marker genes relative to *ACTB* as determined by SSMR-based qRT-PCR in two sets of GBMs. **A**) Expression distributions for 87 subjects with GBM, 66 from M.D. Anderson Cancer Center (MDACC), and 21 from University of Arkansas for Medical Sciences (UAMS). **B**) Institution differences in the expression distributions shown in *A*; The X marks denote nonparametric estimates of the UAMS-minus-MDACC difference, error bars denote the 90% confidence intervals on the estimates, and horizontal dashed lines at ±0.5 $\log_{10}$ units represent the upper and lower limits of equivalence between institutions.

using the same standards for marker and reference genes. However, the variation in quantity of the two standards and human error in dilution of the standard for generating a standard curve may result in data that is not comparable, as shown by this study.

In this study, we have used a large number of GBM samples from two institutes to perform a valid statistical analysis on comparability of 12 gene expression data obtained using SSMR-based qRT-PCR. Only 4 stem cell marker genes are statistically not equivalent between two institute's GBM samples. Lack of statistical equivalence can happen when institutions are nearly equal but patient variation is large. It can also happen when patient variation is moderate but institutions are not so equal, which can be associated with changes of diagnosis standard over time. Noticeably, the MDACC GBMs were from patients operated during 1987–1997, while UAMS from 2003–2006.

The SSMR-based real-time qRT-PCR method allows independent data sets quantified at different times or by different labs comparable. We have applied it in our recent study of glioma prognosis, in which the logratios of marker gene to *ACTB* were used to establish three multivariate prognosis models: GBM, anaplastic astrocytomas (AA), and oligodendroglia tumors with patient's overall survival

time followed to maturation for GBM, 2/3 mature for AA and 1/2 mature for oligodendroglia tumors (Manuscript under review). We demonstrated that it is pivotal to obtain comparable data, which would allow us to combine data from multiple research sites and acquire a large sample size required in modeling of prognosis and identifying effect of gene expression variables. In addition to providing the proper event to variable ratio to establish statistically valid models, with standardization of gene expression quantification, the gene variables from prospective patients can be directly applied into the model based on retrospective patients.

Acknowledgments

This research was supported in part by the Seed Grant from Chao Family Comprehensive Cancer Center at the University of California, Irvine, and the Arkansas Cancer Research Center at the University of Arkansas for Medical Sciences. We would also like to acknowledge the UAMS Tissue Bank for providing tumor specimens for this study.

Abbreviations

SSMR, single standard for marker and reference genes; AqRT-PCR, Absolute quantitative reverse transcription polymerase chain reaction.

Disclosures

This manuscript has been read and approved by all authors. This paper is unique and is not under consideration by any other publication and has not been published elsewhere. The authors and peer reviewers of this paper report no conflicts of interest. The authors confirm that they have permission to reproduce any copyrighted material.

References

1. Zhou YH, et al. The expression of PAX6, PTEN, vascular endothelial growth factor, and epidermal growth factor receptor in gliomas: relationship to tumor grade and survival. *Clin Cancer Res*. 2003;9(9):3369–75.
2. Zhou YH, et al. Modeling prognosis for patients with malignant astrocytic gliomas: quantifying the expression of multiple genetic markers and clinical variables. *Neuro Oncol*. 2005;7(4):485–94.
3. Vera-Ramirez L, et al. Gene-expression profiles, tumor microenvironment, and cancer stem cells in breast cancer: Latest advances towards an integrated approach. *Cancer Treat Rev*. 2010.
4. Hou J, et al. Gene expression-based classification of non-small cell lung carcinomas and survival prediction. *PLoS One*. 2010;5(4):p. e10312.
5. Varambally S, et al. Integrative genomic and proteomic analysis of prostate cancer reveals signatures of metastatic progression. *Cancer Cell*. 2005;8(5): 393–406.
6. Verhaak RG, et al. Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell*. 2010;17(1):98–110.
7. Shi L, et al. The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. *Nat Biotechnol*. 2006;24(9):1151–61.
8. Kuo WP, et al. Analysis of matched mRNA measurements from two different microarray technologies. *Bioinformatics*. 2002;18(3):405–12.
9. Larsson O, Sandberg R. Lack of correct data format and comparability limits future integrative microarray research. *Nat Biotechnol*. 2006;24(11):1322–3.
10. katz MH. Multivariable Analysis. 1999, Cambridge: Cambridge University Press.
11. Wong ML, Medrano JF. Real-time PCR for mRNA quantitation. *Biotechniques*. 2005;39(1):75–85.
12. Canales RD, et al. Evaluation of DNA microarray results with quantitative gene expression platforms. *Nat Biotechnol*. 2006;24(9):1115–22.
13. Pfaffl MW. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res*. 2001;29(9):e45.
14. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods*. 2001;25(4):402–8.
15. Rutledge RG, Cote C. Mathematics of quantitative kinetic PCR and the application of standard curves. *Nucleic Acids Res*. 2003;31(16):e93.
16. Sano T, et al. Differential expression of MMAC/PTEN in glioblastoma multiforme: relationship to localization and prognosis. *Cancer Res*. 1999;59(8): 1820–4.
17. Schuirmann DJ. A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. *J Pharmacokinet Biopharm*. 1987;15(6):657–80.
18. Berger R, Hsu J. Bioequivalence Trials, Intersection-Union Tests and Equivalence Confidence Sets. *Statistical Science*. 1996;11:283–319.
19. Hodges JL, Lehmann EL Jr. Hodges-Lehmann Estimators. *Encyclopedia of Statistical Sciences*. 1983;3(Kotz S, Johnson NL, Read CB, editors):463–5.

Publish with Libertas Academica and every scientist working in your field can read your article

"I would like to say that this is the most author-friendly editing process I have experienced in over 150 publications. Thank you most sincerely."

"The communication between your staff and me has been terrific. Whenever progress is made with the manuscript, I receive notice. Quite honestly, I've never had such complete communication with a journal."

"LA is different, and hopefully represents a kind of scientific publication machinery that removes the hurdles from free flow of scientific thought."

Your paper will be:
- Available to your entire community free of charge
- Fairly and quickly peer reviewed
- Yours! You retain copyright

http://www.la-press.com