

Evaluierung der qPCR

Die Real-Time-RT-PCR-Datenanalyse im Fokus der MIQE-Richtlinie

MICHAEL W. PFAFFL, IRMGARD RIEDMAIER

LEHRSTUHL FÜR PHYSIOLOGIE, WISSENSCHAFTSZENTRUM WEIHENSTEPHAN FÜR ERNÄHRUNG, LANDNUTZUNG & UMWELT, TU MÜNCHEN

Die MIQE-Richtlinie wurde 2009 von einer Gruppe internationaler Wissenschaftler ins Leben gerufen, um die Qualität, die Richtigkeit sowie die Zuverlässigkeit der gewonnenen qPCR-Ergebnisse im Labor und in der wissenschaftlichen Literatur zu steigern.

The MIQE guidelines were established 2009 by a group of international scientist to improve the quality, the accuracy, and the reliability of the generated quantitative PCR results in the lab and in the scientific literature.

■ Das langfristige Ziel ist es, die „MIQE-Richtlinie“ als den akkuraten und zuverlässigen Qualitätsstandard für das experimentelle Design, die Durchführung der quantitativen PCR (qPCR), die Datenanalyse sowie das Publizieren der PCR-Daten zu etablieren [1]. Seit ihrem Erscheinen sind die Richtlinien in molekularbiologisch forschenden Laboratorien sowie in den wissenschaftlichen Medien ein wichtiger Diskussionspunkt.

Seit dem Siegeszug der qPCR in den 1990er-Jahren sagt man der Methode in der modernen molekularbiologischen Forschung und der molekularen Diagnostik eine schnelle und einfache Durchführbarkeit und eine unkomplizierte Anwendung nach. Aufgrund

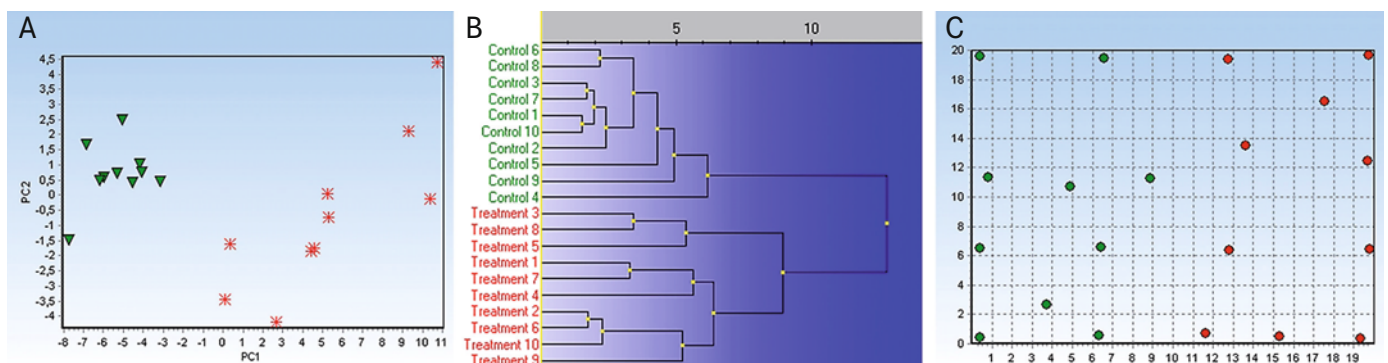
dieser Kriterien und unter dem internationalen Druck, „schnell“ zu veröffentlichen („publish or perish“), wird oft viel zu unkritisch mit den gewonnenen Daten umgegangen und in der Folge leichtfertig publiziert. In der aktuellen Literatur besteht leider ein akuter Mangel an Informationen zu den experimentellen Einzelheiten der durchgeführten Studien. In Publikationen, auch in hoch renommierten Journalen, fehlen essenzielle Informationen, die es erlauben die wissenschaftliche Qualität der dargestellten Resultate objektiv und kritisch zu bewerten. Ziel der MIQE-Richtlinie ist es, die Vollständigkeit der publizierten Methodik zu gewährleisten und die Richtigkeit der gewonnenen qPCR-

Ergebnisse sicherzustellen. Somit kann die experimentelle Transparenz gesteigert und die Wiederholbarkeit innerhalb und zwischen den Laboratorien gefördert werden [1].

Die MIQE-Guidelines stellen einen Richtliniensatz dar, gruppiert in neun Abschnitten mit 85 Einzelpunkten, der das Minimum an für die Veröffentlichung nötigen, wünschenswerten Informationen über qPCR-Experimente beschreibt. Diese Transparenz ermöglicht anderen Forschergruppen, den veröffentlichten Daten zu vertrauen und ggf. die gewonnenen Ergebnisse zu reproduzieren und zu bestätigen. Im Großen und Ganzen sollen die MIQE-Richtlinien zu einer besseren experimentellen Praxis und der Offenlegung der Methodik anregen, um eine zuverlässigere und unmissverständlichere Deutung der qPCR-Resultate zu erlauben. Ziel ist es, die Qualität der Experimente offenzulegen, um deren Ergebnisse zu evaluieren und die biologische Deutung zu ermöglichen. So soll in Zukunft vermieden werden, dass aufgrund mangelhaften Vorgehens bei qPCR-Experimenten falsche biologische Schlussfolgerungen gezogen werden.

Normalisierung

Betrachtet man die möglichen Fehlerquellen der qPCR, so sticht die fehlerhafte Analyse der PCR-Daten heraus. Die unkritische Auswahl der Referenzgene zur Normalisierung,



▲ **Abb. 1:** Expressionsdaten von zehn behandelten Tieren (Kälber, rote Farbgebung) im Vergleich mit zehn unbehandelten Kontrollen (grüne Farbgebung) mittels *principal component analysis* (PCA) (A), hierarchische Clusteranalyse (HCA) (B) und Neuronen Netzen (C), analysiert mit dem GenEx-Softwarepaket.

die fehlende Effizienzkorrektur sowie die unpassende oder falsche Statistik führen zu einer signifikanten Verzerrung der Ergebnisse und somit zu einer falschen quantitativen Aussage [2]. Die Auswahl der geeigneten Referenzgene für die Normalisierung stellt viele Wissenschaftler vor ein Problem. Man muss sich fragen: warum? Zur Verifizierung der richtigen Referenzgene wurde bereits viel publiziert, und es wurden einige Algorithmen entwickelt, welche dabei helfen, die richtigen Kandidaten zu finden; die wichtigsten Algorithmen sind hierbei geNorm [3], BestKeeper [4], Normfinder [5] oder Global Pattern Recognition [6]. Der geNorm-Algorithmus hat sich anhand der jahrelangen Erfahrung als der Goldstandard zur Normalisierung in der Expressionsanalytik herauskristallisiert. Ziel aller Analysen ist die Auswahl mehrerer stabil exprimierter und nicht regulierter Referenzgene sowie die Bestimmung der optimalen Anzahl der Gene zur Berechnung des Normalisierungsfaktors. In der Praxis werden heute immer noch Referenzgene anhand von Publikationen ausgewählt, und nicht wie empfohlen ihre Funktionalität als unregulierte Referenzgene validiert. Richtigerweise und MIQE-konform sollten in jeder durchgeführten Studie die stabilen Referenzgene bestimmt und studienspezifisch zur Normalisierung herangezogen werden. Mindestens zwei, besser drei Referenzgene werden zur Normalisierung empfohlen [1, 3]. Führt man die Normalisierung mit den validierten Referenzgenen durch und bedient sich bei der Datenanalyse frei verfügbarer Quantifizierungssoftware, kann man die technische Varianz reduzieren und die Aussagekraft des Experiments steigern [2, 3]. Normfinder, BestKeeper und geNorm sind einerseits als Freeware verfügbar oder sind teilweise in größere Softwarepakete zur relativen Quantifizierung implementiert: GenEx [7], qBase/qBaseplus [8] sowie in diversen REST-Versionen [9] oder online zur Normalisierung von qPCR-Arrays [6].

PCR-Effizienzkorrektur

Des Weiteren sticht die fehlende Effizienzkorrektur ins Auge. Klassischerweise wird die Berechnung des relativen Expressionsunterschieds über die *delta-delta-Cq*-Methode nach Livak und Schmittgen durchgeführt [10]. Die Expression eines Zielgens (Cq_{Zielgen}) wird mittels derer eines internen Referenzgens ($Cq_{\text{Referenzgen}}$) normalisiert. In einem zweiten Schritt werden die normalisierten Expressionsunterschiede ($\Delta Cq = Cq_{\text{Zielgen}} -$

$Cq_{\text{Referenzgen}}$) in der Behandlungsgruppe mit der Kontrollgruppe verglichen, und man kommt zum $\Delta\Delta Cq$ -Wert. Der relative Expressionsunterschied (R) der interessanten Zielgene berechnet sich nach der Formel: $R = 2^{-\Delta\Delta Cq}$ [10].

Die *delta-delta-Cq*-Methode setzt eine konstante Verdoppelung der DNA-Menge in jedem Real-Time-PCR-Zyklus voraus. Man legt somit eine ideale PCR-Effizienz in allen Proben ($E = 2$) zugrunde und verwendet die Normalisierung anhand eines einzelnen Referenzgens. Die wirkliche PCR-Effizienz eines einzelnen Zyklus bewegt sich im Bereich zwischen 1,7 bis 2,1. Sie ist abhängig von der Qualität des Assays und dem Grad der Inhibition in den unterschiedlichen PCR-Ansätzen [9, 11]. Deswegen ergeben geringste Schwankungen in den PCR-Effizienzen zwischen Zielgen und Referenzgen enorme Verschiebungen der Expressionsratio und führen zu verzerrten quantitativen Resultaten. Um die unterschiedlichen PCR-Effizienzen rechnerisch zu kompensieren und den MIQE-Anforderungen zu folgen, wurden Berechnungsmodelle, wie das effizienzkorrigierte relative Quantifizierungsmodell entwickelt [9, 11]. Hierbei wird die PCR-Effizienz eines jeden Gens bestimmt und im Berechnungsalgorithmus der relativen Expression (R) berücksichtigt:

$$R = \frac{(E_{\text{Zielgen}})^{\Delta Cq_{\text{Zielgen}}(\text{Kontrolle} - \text{Behandlung})}}{(E_{\text{Referenzgen}})^{\Delta Cq_{\text{Referenzgen}}(\text{Kontrolle} - \text{Behandlung})}}$$

Die effizienzkorrigierte relative Quantifizierung mittels Real-Time-qRT-PCR unter Einbeziehung mehrerer validierter Referenzgene stellt bis dato die effektivste Form der Nukleinsäure-Quantifizierung dar, die den MIQE-Richtlinien folgt [1]. Diese Methode findet auch Anwendung in den bereits genannten Softwarepaketen: Genex [7], qBase/qBaseplus [8] und REST [9].

Statistik

Der letzte Punkt in einem PCR-Experiment ist die Auswahl der richtigen Statistik zur Analyse der gewonnenen Daten. Das passende statistische Verfahren richtet sich nach dem experimentellen Design, also wie viele biologische Individuen und Wiederholungen beprobt wurden und auf welchen analytischen Ebenen technische Replikate eingebaut wurden. Im Falle der qRT-PCR sind technische Wiederholungen auf folgenden Ebenen denkbar: Extraktion, reverse Transkription (RT) und PCR. Da die Datenstruktur oft nicht gleichmäßig über die experimentellen Grup-

pen verteilt ist und die erhaltenen Cq-Werte meist nicht normalverteilt sind, stellt die Wahl der richtigen statistischen Methode große Herausforderungen an eine Analysensoftware. Der Vergleich mehrerer Behandlungsgruppen, in sich heterogene Patientengruppen, unregelmäßige Zeitreihen, unterschiedliche Gruppengrößen oder ein Vergleich über mehrere Versuchsdurchläufe erschweren die statistische Analyse. Die Wahl des richtigen und robusten statistischen Verfahrens ist oft schwer. Die Statistik im REST-Softwarepaket basiert auf einem sehr robusten und von einer Normalverteilung unabhängigen Randomisierungstest (Bootstrapping-Verfahren), bei dem beliebig viele Tausend Randomisierungen und Wiederholungen durchgeführt werden können [9]. Die REST-Software ist frei erhältlich und findet sich in vielen Versionen zu den unterschiedlichsten Anwendungsgebieten.

Multivariate Datenanalyse

Das heutige Zeitalter der Hochdurchsatz-qPCR stellt die Datenanalyse vor neue Herausforderungen. Mithilfe kommerziell erhältlicher qPCR-Arrays können 96 bzw. 384 verschiedene Transkripte quantitativ erfasst werden, welche in der Folge mit den richtigen Referenzen normalisiert werden müssen. Des Weiteren sind die Regulation bestimmter Gene und die Einteilung in Gruppen mit ähnlichem Expressionsverhalten und ähnlicher Dynamik von Interesse. Auch die Analyse von Markergenen steht im Zentrum der Expressionsanalytik und benötigt valide statistische Methoden zur Bestätigung [12]. Zur Auswertung der Expressionsprofile vieler Gene über viele biologische Proben stehen multivariate statistische Methoden zur Verfügung. Hier haben sich vor allem folgende multivariate Methoden etabliert: die *principal component analysis* (PCA), die hierarchische Clusteranalyse (HCA) und Neuronale Netzwerke (*self organizing maps*) (Abb. 1), implementiert im GenEx-Softwarepaket (<http://genex.genequantification.info>) oder als Open-Source-Software, frei verfügbare R-Programmierspakete, erhältlich über Bioconductor (www.bioconductor.org/help/search/index.html?q=qPCR).

Ausblick

Da wir uns erst am Anfang der Hochdurchsatz-qPCR-Ära befinden und wir in Zukunft mit einer riesigen Expressionsdatenflut rechnen müssen, ist es notwendig, weitere neue Algorithmen zu entwickeln, die es uns erlauben, diese Daten zuverlässig, schnell und

MIQE-konform zu analysieren. In diesem Beitrag konnten nur die wichtigsten Komplexe der qPCR-Datenanalyse aufgeführt werden.

Literatur

- [1] Bustin SA, Benes V, Garson JA et al. (2009) The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clinical Chemistry* 55:611–622
- [2] Kitchen RR, Kubista M, Tichopad A (2010) Statistical aspects of quantitative real-time PCR experiment design. *Methods* 50:231–236
- [3] Vandesompele J, De Preter K, Pattyn F et al. (2002) Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol* 3:RESEARCH0034
- [4] Pfaffl MW, Tichopad A, Prgomet C et al. (2004) Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper – Excel-based tool using pair-wise correlations. *Biotechnol Lett* 26:509–515
- [5] Andersen CA, 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
- [6] Akilesh S, Shaffer DJ, Roopenian D (2003) Customized molecular phenotyping by quantitative gene expression and pattern recognition analysis. *Genome Res* 13:1719–1727
- [7] Bergkvist A, Rusnakova V, Sindelka R et al. (2010) Gene expression profiling – clusters of possibilities. *Methods* 50:323–335
- [8] Hellemans J, Mortier G, De Paeppe A et al. (2007) qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. *Genome Biol* 8:R19
- [9] Pfaffl MW, Horgan GW, Dempfle L (2002) Relative Expression Software Tool (REST[®]) for group wise comparison and statistical analysis of relative expression results in real-time PCR. *Nucleic Acids Res* 30:e36
- [10] Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C_T}$ method. *Methods* 25:402–408
- [11] Pfaffl MW (2001) A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res* 29:e45
- [12] Riedmaier I, Becker C, Pfaffl MW et al. (2009) The use of omic technologies for biomarker development to trace functions of anabolic agents. *J Chromatogr A* 1216:8192–8199

Korrespondenzadresse:

Prof. Dr. Michael W. Pfaffl

Lehrstuhl für Physiologie

Wissenschaftszentrum Weihenstephan für

Ernährung, Landnutzung & Umwelt

Technische Universität München

Weihenstephaner Berg 3

D-85354 Freising-Weihenstephan

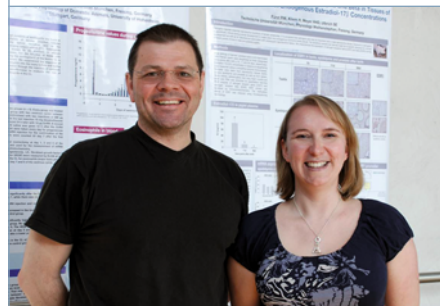
Tel.: 08161-713511

Fax: 08161-714204

Michael.Pfaffl@wzw.tum.de

<http://MIQE.gene-quantification.info>

AUTOREN



Michael W. Pfaffl

1986–1993 Studium der Agrar- und Tierwissenschaften. 1993–1996 Studium der Biotechnologie, 1997 Promotion, 2003 Habilitation (Physiologie) und Privatdozent. Seit 2011 Professor (apl) für Molekulare Physiologie am Lehrstuhl für Physiologie, TU München.

Irmgard Riedmaier

2000–2004 Studium der Biologie (Diplom) an der TU München, 2009 Promotion. Seit 2009 Wissenschaftliche Mitarbeiterin am Lehrstuhl für Physiologie, TU München.