

achtet und stört nicht bei der Analyse von klinischen Proben. Das Detektionslimit der typenspezifischen Sonden beträgt 0,1 bis 0,8 ng PCR-Produkt, die Ergebnisse sind semiquantitativ und erlauben das Abschätzen der Menge an PCR-Produkt. Der gleichzeitige Nachweis von bis zu zehn HPV-Typen in derselben Reaktion ist ohne Einschränkungen möglich.

### Validierung

Multiplex-HPV-Genotyping wurde in einem verblindeten Vergleich mit ihrem Goldstandard, dem GP5<sup>+</sup>/6<sup>+</sup>-RLB, validiert. Die Analyse von 94 klinischen Proben – vorselektiert, um eine große Vielfalt an HPV-Typen zu repräsentieren – zeigte eine gute Übereinstimmung beider Methoden (Kappa-Wert = 0,92), deutete aber auch auf eine höhere Sensitivität von MPG hin. Von allen Typisierungreaktionen waren 107 übereinstimmend positiv, 1944 mit beiden Methoden negativ und 17 diskordant. Unter diesen Reaktionen wurden 16 zusätzliche Infektionen mit MPG nachgewiesen.

### Ausblick

Aufgrund der Gegebenheiten des Lumineszenz-Arrays erlaubt MPG die gleichzeitige Identifizierung von bis zu 100 verschiedenen HPV-Typen in einer Reaktion. Weitere Typen können leicht in den Test integriert werden und damit etwa den Nachweis aller 45 bekannten genitalen HPV-Typen erlauben. Die Hochdurchsatz-Methode ist schnell und zeitsparend: Die Analyse von PCR-Produkten aus mindestens 500 klinischen Proben kann von einer Person an einem Tag durchgeführt werden. MPG ist daher bestens geeignet für große epidemiologische HPV-Studien und für die Überwachung von Impfprogrammen.

### Literatur

- [1] Parkin DM, Bray F, Ferlay J, Pisani P. Estimating the world cancer burden. *Globocan 2000*. *Int J Cancer* 2001;94:153-6.
- [2] Munoz N, F. X. Bosch, S. de Sanjose, R. Herrero, X. Castellsague, K. V. Shah, P. J. Snijders, and C. J. Meijer. Epidemiologic classification of human papillomavirus types associated with cervical cancer. *N Engl J Med* 2003;348:518-27.
- [3] Walboomers JM, Jacobs MV, Manos MM, Bosch FX, Kummer JA, Shah KV, et al. Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. *J Pathol* 1999;189:12-9.
- [4] van den Brule AJ, Pol R, Franssen-Daalmeijer N, Schouls LM, Meijer CJ, Snijders PJ. GP5<sup>+</sup>/6<sup>+</sup> PCR followed by reverse line blot analysis enables rapid and high-throughput identification of human papillomavirus genotypes. *J Clin Microbiol* 2002;40:779-87.
- [5] Schmitt M, Bravo IG, Snijders PJ, Gissmann L, Pawlita M, Waterboer T. Bead-based multiplex genotyping of human papillomaviruses. *J Clin Microbiol* 2006;44:504-12.

### Korrespondenzadresse

Dipl.-Biol. Markus Schmitt  
Deutsches Krebsforschungszentrum (DKFZ),  
Im Neuenheimer Feld 242  
D-69120 Heidelberg  
Tel.: +49-(0)6221-424937  
Fax: +49-(0)6221-424932  
eMail: markus.schmitt@dkfz.de.

## Bioinformatik

# Neue Entwicklungen in der post-qPCR-Datenanalyse

PD Dr. Michael W. Pfaffl, Technische Universität München, Freising-Weihenstephan

Die Entwicklung der Polymeraseketten-Reaktion (PCR) in den achtziger Jahren gehört ohne Zweifel zu den größten Errungenschaften in der Molekularbiologie. Mittels klassischer PCR lassen sich hochsensitiv DNA-Fragmente qualitativ und semi-quantitativ nachweisen. Um spezifische mRNA zu quantifizieren, wird vor der PCR die Reverse Transkription (RT) eingesetzt. Die Anwendung der RT-PCR zur Quantifizierung spezifischer mRNA ist heute zum Routinewerkzeug in der Expressionsanalytik geworden. Die gewonnenen Ergebnisse sind von überproportionalem Nutzen in der molekularbiologischen Forschung und molekularen Diagnostik, in der vergleichenden Expressionsanalytik sowie zur Aufklärung von Genfunktionen. Der Nachweis kann qualitativ in klassischen Thermocyclern oder in „real-time“ quantitativ mittels Echtzeit-PCR (qPCR) erfolgen. Die Ergebnisse sind direkt verfügbar, so daß der Einsatz der qPCR eine deutliche Zeitersparnis gegenüber der klassischen PCR-Methode mit sich bringt. Da die Zunahme der Fluoreszenz und die Menge an neusynthetisierten PCR-Produkten über einen weiten Bereich proportional zueinander sind, kann aus den gewonnenen Fluoreszenzdaten die eingesetzte Ausgangsmenge der DNA oder RNA bestimmt werden. Voraussetzung für einen zuverlässigen quantitativen Nachweis ist eine funktionierende Analytik und Datenauswertung, die exakte Quantifizierungsergebnisse bei ausreichender Genauigkeit und hoher Wiederholbarkeit liefert<sup>1</sup>.

Zwei generelle Quantifizierungsstrategien stehen in der qPCR zur Verfügung: Die „absolute“ Quantifizierung wird anhand einer gegebenen Kalibrierkurve durchgeführt, basierend auf einer Verdünnungsreihe von qPCR-Produkten, Plasmid-DNA, *in vitro* transkribierter RNA, synthetisierter DNA oder von RNA-Oligomeren<sup>2</sup>. Bei der relativen Quantifizierung wird dagegen die Genexpression eines Zielgens auf ein oder mehrere Referenz-Gen(e) bezogen. Man nennt diesen Vorgang auch Normalisierung der Expressionsergebnisse<sup>3</sup>. Die relative Quantifizierung läßt sich weiter optimieren, indem die unterschiedlichen qPCR-Effizienzen der untersuchten Gene miteinbezogen werden. Die Effizienz-korrigierte relative Quantifizierung mittels real-time-qRT-PCR stellt bis dato die eleganteste Form der Quantifizierung dar<sup>4</sup>.

In den vergangenen Jahren konzentrierte sich die Entwicklung vor allem auf neue Hardware und Detektionssysteme. Die post-qPCR-Datenanalyse der gewonnenen Ergebnisse stand eher im Hintergrund. Auf der Basis neuentwickelter qPCR-Applikationen und der stark wachsenden Zahl an qPCR-Datensätzen ist jetzt ein Umdenken in der Datenanalyse gefordert. Die qPCR-Datenanalyse muß transparenter und reproduzierbarer gemacht werden. Einige ausgeklügelte Algorithmen sind schon entwickelt und in existierende Software-Applikationen implementiert worden<sup>4,6</sup>, aber es fehlt noch an zuverlässigen post-qPCR-Applikationen, die leichtverständliche Ergebnisse liefern. künftig sollte das Augenmerk mehr auf die

post-qPCR-Datenanalyse, die qPCR-Bioinformatik, gelegt werden.

### Nutzen einer verbesserten qPCR-Datenanalyse

Aus den generierten Fluoreszenz-Emissionen (=Rohdaten) sollten sinnvolle qPCR-Kurven generiert und nachfolgend zuverlässige Ct-Werte sowie qPCR-Effizienzen berechnet werden. Ausgehend von Ct-Werten unterschiedlicher Gene kann dann in der Folge eine Normalisierung durch ein oder mehrere Referenz-Gene und die Verrechnung der Ct-Datenpunkte stattfinden<sup>5,7</sup>. Auch die Analyse von qPCR-Ausreißern stellt ein Problem dar und wurde bis dato sehr stiefmütterlich behandelt<sup>8</sup>. Zeigt der Kurvenverlauf die Integrität der qPCR? Ist der gefundene Ct-Wert zuverlässig oder nicht? Liegt der Ct-Wert im akzeptablen und quantifizierbaren Bereich? Nicht zu vergessen ist die statistische Auswertung der qPCR-Daten. Dies birgt einige Hindernisse, da Expressionsdaten zum Beispiel sehr selten normalverteilt sind<sup>4</sup>. Die Wahl der richtigen statistischen Methode ist eine Herausforderung. Der Vergleich mehrerer Gruppen, einer Zeitreihe, unterschiedliche Gruppengrößen oder ein Vergleich über mehrere Versuchsdurchläufe erschweren die Analyse und die darauf aufbauende Statistik. Die Analyse und statistische Auswertung komplexer Zusammenhänge muß vereinfacht und zuverlässig

Idaho Technologies

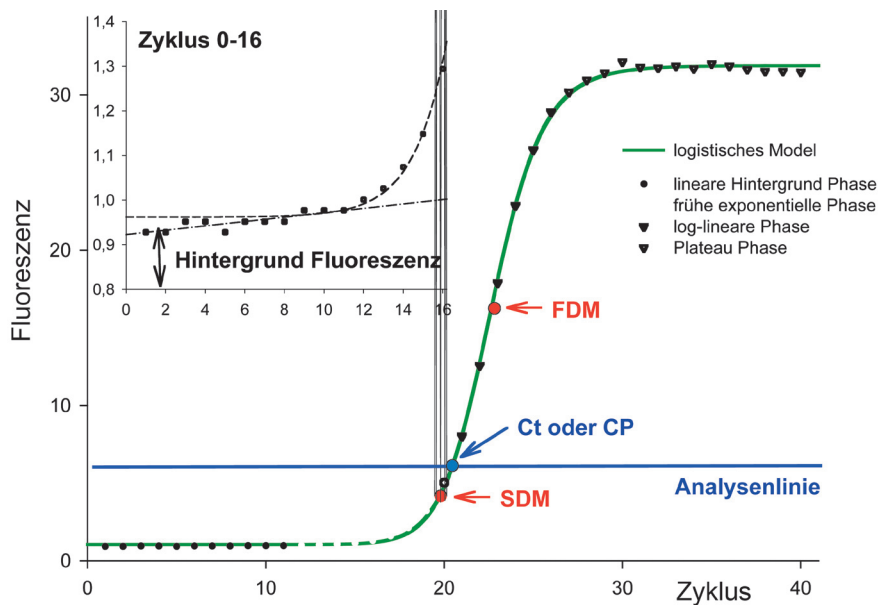


Abb. 1: Kurvenverlauf und Phasen der real-time PCR. Ct Wert Bestimmung anhand der Analyselinie oder mittels eines 4-parametrisch logistischen Modells am FDM (First Derivative Maximum) oder SDM (Second Derivative Maximum).

dargestellt werden. Letztendlich müssen die generierten qPCR-Daten richtig archiviert werden, um sie in größeren Studien miteinander vergleichen zu können. Im Zentrum des Interesses der post-qPCR-Datenanalyse stehen Standardisierung, Transparenz und die Qualitätskontrolle der qPCR.

### Cycle threshold und Hintergrundfluoreszenz

Als neue Meßgröße und indirektes Maß für die eingesetzte Startmenge werden in der qPCR der Ct-Wert (Cycle threshold) oder Crossing Point (CP) bestimmt<sup>3,5</sup>. Er entspricht der Anzahl der PCR-Zyklen die nötig sind, um ein konstant definiertes Fluoreszenzniveau (= Threshold-Fluoreszenz) zu erreichen (Abb. 1). Per Definition befindet sich am Ct in allen Reaktionsgefäßen die gleiche Menge an neusynthetisierter DNA. Die gesamte absolute oder relative Quantifizierung hängt von diesem einem Ct-Wert ab. Deshalb sollte bei der Ct-Berechnung besonderes Augenmerk auf die Art der Berechnung gelegt werden. Alle Geräte generieren zunächst Fluoreszenzrohdaten. Bereits diese Rohdaten können aufgrund der verwendeten Farbstoffe, der Geräteeigenschaften, des Kapillarmaterials, der Meßtechnik und der Datenmittelung auf der Meßebeane sehr differieren.

Die Berechnung des Ct-Wertes kann direkt aus den Fluoreszenz-Rohdaten oder aus korrigierten Kurven abgeleitet werden – einerseits über eine Hintergrund-Fluoreszenzkorrektur und/oder durch die Normalisierung eines weiteren internen Fluoreszenzstandards, etwa ROX. Des weiteren verfügt die geräteeigene Software über die Option einer Kurvenglättung zur besseren optischen Präsentation der generierten Rohdaten. Mit dieser Glätt-

funktion und der Möglichkeit, vorhandene Hintergrundfluoreszenzen herauszulöschen, werden die Kurven manipuliert und auch die Lage des Ct-Wertes verändert.

Selbst die richtige Korrektur der Hintergrundfluoreszenz ist aufgrund des nicht konstanten Verlaufes ein Problem (Abb. 1). In der Anfangsphase der qPCR kann es einen linearen oder unregelmäßigen, einen ansteigenden oder abnehmenden Fluoreszenzverlauf geben. Hier die richtige Hintergrundkorrektur über eine Vielzahl von biologischen Proben reproduzierbar vorzunehmen, kann sehr schwierig werden.

Zudem kann auch die Höhe des Fluoreszenzniveaus durch den Wissenschaftler oder durch die Auswertesoftware beeinflusst werden. In den einzelnen Softwareversionen der real-time-Hardware bieten sich meist mehrere Optionen, die Analyselinie zu legen (Abb. 1). Einerseits kann die Höhe der Analyselinie abhängig von der Varianz der Grundfluoreszenz (plus 10x Std.Dev.) sein; über mehrere Experimente hinweg kann die Analyselinie als konstant definiert werden; oder die Auswertesoftware wählt sich eine, oft sehr hoch liegende Analyselinie, die den geringsten Fehler bei der Erstellung einer Kalibrierkurve aufweist. Die große Schwierigkeit der Auswertung unterschiedlicher qPCR-Experimente liegt allerdings in der Konstanz und in der Vergleichbarkeit der generierten Ct-Werte.

Soll der Ct-Wert völlig unabhängig vom Kurvenverlauf bestimmt werden, sind etablierte sigmoidale- oder logistische-Algorithmen angezeigt. Diese berechnen den Ct ohne Analyselinie und Hintergrundkorrektur immer konstant und ohne Beeinflussung von außen<sup>6,9</sup>. Nicht nur einzelne, arbiträr ausgesuchte Zyklen werden zur Ct-Berechnung herangezogen, sondern die gesamte

Fluoreszenzkurve, mit all ihren Fluoreszenz-Rohdaten. Der Fluoreszenzverlauf wird an ein vier-parametrisches-sigmoidales oder -logistisches Modell mit sehr hoher Korrelation ( $r = 0.99$ ) angepaßt (Abb. 1). Mittels der ersten (First Derivative Maximum = FDM) oder der zweiten Ableitung (SDM) können Ct-Wert und Effizienz in der exponentiellen Phase berechnet werden<sup>3,9</sup>. Für jede Probe werden der theoretische Ct-Wert sowie die kurvenspezifische qPCR-Effizienz erhalten. Beide Werte lassen sich in Softwareapplikationen zur quantitativen Analyse nutzen.

### Relative Quantifizierung

Als am häufigsten verwendete Methode der post-qPCR Analyse soll die relative oder vergleichende Genexpressionsanalytik vorgestellt werden. Bei der relativen Quantifizierung wird die mRNA-Expression des Zielgens mit der eines Referenzgens normalisiert<sup>4,5</sup>. Zudem wird die relative Expression des zu untersuchenden Gens in der behandelten experimentellen Probe auf ein nicht behandeltes Kontrollprobenmaterial bezogen. Die Vorteile der Normalisierung mit einem internen Standard liegen in der Reduzierung der Varianz der Expressionsergebnisse, da Gewebe- und Matrixeffekte, unterschiedliche RNA Extraktionseffizienzen, die Integrität der extrahierten RNA sowie Fehler bei der reversen Transkription innerhalb einer experimentellen Probe gleichermaßen beide Gene betreffen<sup>1</sup>. In den folgenden Berechnungen des Expressionsverhältnisses heben sich diese individuellen Probeneffekte gewissermaßen wieder auf. Sehr beliebte Referenzgenkandidaten sind 18S-, 28S-rRNA, Ubiquitin-, Histon-H3, GAPDH- und beta-Aktin-mRNA.

$$\Delta \text{CP} = \text{CP Zielgen} - \text{CP Referenzgen}$$

$$\Delta \Delta \text{CP} = \Delta \text{CP Behandlung} - \Delta \text{CP Kontrolle}$$

$$- \Delta \Delta \text{CP}$$

$$\text{Ratio} = 2$$

Meist wird die Berechnung des Expressionsunterschiedes über die sogenannte  $\Delta \Delta \text{CP}$ -Methode bestimmt<sup>5</sup>. Der relative Expressionsunterschied berechnet sich nach Formel 2- $\Delta \Delta \text{CP}$ . Die Expression des Zielgens wird mittels internem Referenzgen normalisiert, und in einem zweiten Schritt werden die normalisierten Expressionsunterschiede ( $\Delta \text{CP}$ ) in der Behandlung und der Kontrolle verglichen<sup>5</sup>.

Dieses Berechnungsschema setzt eine Verdopplung der DNA-Menge in jedem Zyklus voraus. Es wird von einer optimalen real-time-PCR-Effizienz in allen Proben ausgegangen ( $E = 2$ ), was natürlich nicht der Praxis entspricht. Die „wahre“ qPCR-Effizienz dürfte sich unter optimierten Reaktionsbedingungen im Bereich von 1,7 bis 1,9 bewegen<sup>6,8</sup>. Geringste Schwankungen in den Effizienzen von Zielgen zu Referenzgen können zu enormen Unterschie-

den in den Expressionsverhältnissen führen und somit die Ergebnisse verzerren. Um den unterschiedlichen Effizienzen in den Proben Rechnung zu tragen, sind Berechnungsmodelle entwickelt worden, wie das effizienzkorrigierte relative Quantifizierungsmodell<sup>4</sup>.

### Relative Expression – Software-Tool

REST faßt die Expressions-Gruppen als CP-Mittelwerte zusammen und berechnet daraus die mittleren Expressionsratios (R) der Gruppen – normalisiert über ein bestimmtes HKG – sowie die Varianzen der einzelnen Expressionen<sup>4</sup>. Die Ergebnisse werden statistisch getestet und die Expressionsunterschiede werden in einem Ausgabefenster graphisch mit ihren Signifikanzniveaus ausgegeben. Die Statistik basiert auf einem sehr robusten und von einer Normalverteilung unabhängigen Randomisierungstest, bei dem beliebig viele tausend Randomisierungen und Wiederholungen durchgeführt werden können. Die

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

REST-Software ist frei erhältlich und findet sich in vielen Versionen zu den unterschiedlichsten Anwendungsgebieten.

### GeNorm

Ein weiteres Problem bei der relativen Quantifizierung stellt sich bei der Auswahl der geeigneten Referenzgene<sup>1</sup>. Die oft übliche Praxis, die Daten gegen ein einziges Referenzgen abzugleichen, ist unbefriedigend. Die aktive Auswahl kann die stabilsten Genexpressionen

für die Normalisierung identifizieren und so individuelle Expressionsunterschiede in den Zielgen ausgleichen. Die Normalisierung gegen mehrere Referenzgene ist deshalb für eine korrekte Behandlung von Expressionsdaten in der Regel unumgänglich und ein entschei-

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

dender Schritte bei der Erstellung valider qRT-PCR-Expressionsprofile<sup>10</sup>.

### Welches Gen ist in meiner Applikation das „beste“ Referenzgen?

Welches Gen ist unter individuellen Versuchsbedingungen nicht oder nur minimal reguliert? Zur Analyse und validen Beantwortung dieser Fragestellung wurde eine bereits frei erhältliche, Excel-basierte Software namens GeNorm entwickelt<sup>10</sup>. Ziel ist die Identifikation stabiler, exprimierter Referenzgene und der optimalen Anzahl der Gene zur Berechnung eines Index oder Normalisierungsfaktors (NF). Für die Normalisierung mit GeNorm werden zunächst die Expressionswerte mehrerer Gene in einem Panel ermittelt. Für jedes Gen werden zunächst die Ct-Werte ermittelt, in relative Expressionswerte umgerechnet und diese dann in einem paarweisen Vergleich analysiert. Nach ähnlichen Algorithmen arbeiten auch vergleichbare Software-Tools<sup>10,11</sup>. Die Erfahrung zeigt, daß in der Regel mindestens drei Referenzgene zur Normalisierung herangezogen werden sollten<sup>7,10,11</sup>.

Da wir uns erst am Beginn der post-qPCR-Daten-Ära befinden, ist es notwendig, weitere neue qPCR-Datenanalysemethoden und -Al-

gorithmen sowie statistische Modelle zu entwickeln, die die Erzeugung, Evaluierung, Verrechnung, Statistik und die Archivierung der qPCR-Daten erleichtern und standardisieren. In diesem Artikel konnten nur einige Themen der post-qPCR-Datenanalyse angeschnitten werden. Die beschriebenen Themenkomplexe können im Detail auf der korrespondierenden Internetseite nachgeschlagen, und die beschriebenen Software-Tools können kostenfrei heruntergeladen werden, siehe <http://bioinformatics.gene-quantification.info>.

### Literatur

- [1] Bustin, S.A., *J Mol Endocrinol* 25 (2000) 169-193
- [2] Pfaffl, M.W., and Hageleit, M., *Biotechnology Letters*, 23 (2001) 275-282
- [3] Rasmussen, R., (2001) *Quantification on the LightCycler*. In: Meuer, S., Wittwer, C. and Nakagawara, K. eds. *Rapid Cycle Real-time PCR, Methods and Applications* Springer Press, Heidelberg; ISBN 3-540-66736-9, 21-34
- [4] Pfaffl, M.W., Horgan, G.W., and Dempfle, L., *Nucleic Acids Res.* 30(9): (2002) e36
- [5] Livak, K.J., and Schmittgen, T.D., *Method. Methods.*, 25(4): (2001) 402-408
- [6] Rutledge RG., *Nucleic Acids Res.* 2004, 32 (2004): e178
- [7] Pfaffl M.W., Tichopad A., Prgommet C., Neuvians T. P. *Biotechnology Letters*, 26(6) (2004) 509-515.
- [8] Bar T, Stahlberg A, Muszta A, Kubista M. *Nucleic Acids Res.* 31(17) (2003) e105
- [9] Tichopad, A., Dilger, M., Schwarz, G., and Pfaffl, M.W., *Nucleic Acids Res.*, 31(20) (2003) e122
- [10] Vandecompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paep, A., and Speleman, F., *Genome Biology*, 3(7) (2002) 0034.1-0034.11
- [11] Andersen CL, Jensen JL, Orntoft TF. *Cancer Res* 64 (2004) 5245-5250

### Korrespondenzadresse

PD Dr. Michael W. Pfaffl  
 Lehrstuhl für Physiologie  
 Wissenschaftszentrum Weihenstephan  
 Technische Universität München  
 Weihenstephaner Berg 3  
 D-85354 Freising-Weihenstephan  
 Tel: +49-(0)8161-713511. Fax: -714204  
 eMail: michael.pfaffl@wzw.tum.de

Kennziffer 16 LW 02 · Informationen ordern? · [www.biocom.de](http://www.biocom.de)